MFE 230D

Derivatives: Quantitative Methods

Lecture notes

Alberto Gennaro

https://github.com/GennaroAlberto/mfe230d-derivatives

Built October 31, 2026

Contents

About these notes **7**

I Analytical pricing **8**

1 The standard model, revisited **9**

1.1 Brownian motion and Itô's formula, in brief 9

1.2 The standard model 10

1.3 The Feynman–Kac formula 13

1.4 Dirac deltas and Green's functions 17

Further reading 20

2 The Black–Scholes formula three ways **21**

2.1 The formula by direct integration 22

2.2 The pricing equation and the heat equation 23

2.3 Solving the equation by Fourier transform 24

2.4 Solving the equation by Laplace transform 26

2.5 The Greeks and the relations between them 28

2.6 Implied volatility 29

Further reading 32

3 Arbitrage in the strike and the smile **33**

3.1 Call prices across strikes 33

3.2 The Breeden–Litzenberger formula 36

3.3 The smile 40

Further reading 44

4 Static replication and the Kolmogorov equations **45**
4.1 Static replication . . . 46
4.2 Power payoffs and the pre-wash . . . 49
4.3 Homogeneity and the Greeks' conservation laws . . . 51
4.4 The Kolmogorov equations . . . 53
Further reading . . . 56

II Monte Carlo methods **57**

5 Random numbers **58**
5.1 Pseudo-random numbers . . . 58
5.1.1 Linear congruential generators . . . 58
5.1.2 The lattice structure . . . 59
5.1.3 Generators in use . . . 61
5.2 Testing a generator . . . 62
5.3 Inversion . . . 63
5.3.1 Inversion by Newton's method: the semicircle law . . . 64
5.3.2 Inverting the normal distribution function . . . 65
5.4 Acceptance–rejection . . . 66
5.5 Normal variates . . . 67
5.6 Correlated Gaussian vectors . . . 69
Further reading . . . 71

6 Monte Carlo pricing and its error **72**
6.1 The Monte Carlo estimator . . . 72
6.2 The central limit theorem and confidence intervals . . . 74
6.2.1 Small samples and Student's t . . . 76
6.2.2 How accurate is the normal approximation? . . . 76
6.2.3 Heavy tails and infinite variance . . . 77
6.3 Batch means . . . 78
6.4 Cost, efficiency and bias . . . 79
6.5 Functions of expectations . . . 81
6.6 Monte Carlo against quadrature . . . 83

Further reading . . . 85

7 Simulating paths **86**

7.1 Exact simulation on a time grid . . . 86

7.1.1 Geometric Brownian motion with term structures . . . 86

7.1.2 Gaussian processes . . . 89

7.2 Three ways to build a Brownian path . . . 89

7.3 Discretising a stochastic differential equation . . . 91

7.4 Richardson extrapolation . . . 95

7.5 Time steps or paths? . . . 96

Further reading . . . 97

8 Antithetic variates, moment matching and control variates **99**

8.1 Setting and running examples . . . 99

8.2 Antithetic variates . . . 100

8.3 Moment matching . . . 103

8.4 Common random numbers . . . 105

8.5 Control variates . . . 106

8.5.1 Estimating the coefficient . . . 107

8.5.2 Several controls . . . 109

8.5.3 Nonlinear controls and hedges . . . 110

8.6 Combining the techniques: a laboratory . . . 111

Further reading . . . 112

9 Importance sampling **113**

9.1 The likelihood ratio . . . 113

9.2 Shifting and scaling the Gaussian . . . 115

9.3 Exponential tilting and large deviations . . . 117

9.4 Path space: Girsanov's theorem and the optimal drift . . . 119

9.5 Rare events and relative error . . . 121

9.6 Adaptive importance sampling: cross-entropy . . . 123

9.7 Importance sampling with control variates . . . 125

Further reading . . . 126

10 Stratified sampling and conditional Monte Carlo **127**
10.1 Stratified sampling . . . 127
10.1.1 Optimal allocation . . . 129
10.1.2 Error estimates . . . 131
10.2 Post-stratification . . . 132
10.3 Stratifying paths . . . 133
10.3.1 Latin hypercube sampling . . . 135
10.4 Conditional Monte Carlo . . . 136
10.4.1 Conditioning on the volatility path . . . 137
10.5 Laboratory . . . 139
Further reading . . . 140

11 Quasi-Monte Carlo **142**
11.1 Discrepancy and the Koksma–Hlawka inequality . . . 142
11.2 Low-discrepancy sequences . . . 145
11.2.1 Van der Corput and Halton . . . 145
11.2.2 Nets and the Sobol' sequence . . . 146
11.3 Randomised quasi-Monte Carlo . . . 149
11.4 Effective dimension and path constructions . . . 151
Further reading . . . 154

12 Multilevel Monte Carlo **155**
12.1 The multilevel identity . . . 156
12.2 Coupling fine and coarse paths . . . 157
12.3 Giles' complexity theorem . . . 161
12.4 The adaptive algorithm . . . 163
12.5 Unbiased randomised estimators . . . 165
Further reading . . . 167

13 Monte Carlo Greeks **169**
13.1 Finite differences . . . 169
13.2 Pathwise derivatives . . . 173
13.3 The likelihood ratio method . . . 176
13.4 The gamma . . . 179

13.5 Many Greeks at once: adjoints . . . 180
Further reading . . . 182

14 American options by simulation **183**
14.1 Bermudan options and optimal stopping . . . 183
14.2 Regression: the Longstaff–Schwartz algorithm . . . 186
14.3 Duality: upper bounds . . . 191
Further reading . . . 196

III Path-dependent and multi-asset payoffs **197**

15 Asian options **198**
15.1 Average-price options . . . 198
15.2 The geometric average . . . 199
15.3 Monte Carlo with the geometric control . . . 200
15.4 Moment matching . . . 203
15.5 Conditioning on the geometric average . . . 204
15.6 Accuracy and a reference price . . . 206
Further reading . . . 208

16 Barriers, lookbacks and the Brownian bridge **209**
16.1 The contracts . . . 209
16.2 The reflection principle with drift . . . 210
16.3 Closed forms by images . . . 211
16.4 Double barriers . . . 213
16.5 Lookback options . . . 214
16.6 Discrete monitoring . . . 215
16.6.1 Exact prices by recursion over the monitoring dates . . . 215
16.6.2 The continuity correction . . . 217
16.7 The Brownian bridge . . . 218
Further reading . . . 220

17 FX and multi-asset options **221**
17.1 Currency options . . . 221

17.2 Changing the numéraire . . . 222
17.3 Exchange options . . . 225
17.4 Correlated simulation and quantos . . . 226
17.5 Basket options . . . 228
Further reading . . . 232

IV Beyond Black–Scholes 233

18 Hedging in practice 234
18.1 Discrete delta hedging . . . 235
18.2 Hedging at the wrong volatility . . . 238
18.2.1 Implied or realised volatility? . . . 241
18.3 Transaction costs and Leland's volatility . . . 242
Further reading . . . 244

19 Stochastic volatility 246
19.1 The Heston model . . . 246
19.2 Affine structure and the characteristic function . . . 248
19.3 Pricing by Fourier inversion . . . 251
19.4 Simulation . . . 254
19.5 Smiles . . . 256
Further reading . . . 257

20 Trees and finite differences 258
20.1 Binomial trees . . . 258
20.2 Trinomial trees and the explicit scheme . . . 261
20.3 Finite differences: stability and order . . . 262
20.4 Boundaries and the choice of grid . . . 265
20.5 American options on a grid . . . 266
20.6 Path dependence and several factors . . . 268
Further reading . . . 269

About these notes

These are lecture notes for MFE 230D, *Derivatives: Quantitative Methods*, in the Master of Financial Engineering program at UC Berkeley Haas: analytical pricing beyond the Black–Scholes formula, Monte Carlo methods and their Greeks, path-dependent and multi-asset payoffs, and the numerical methods behind them.

They grew out of the discussion sessions I taught for the course, which is taught by Dmitry Livdan and builds on material developed by Eric Reiner. The text, the proofs, the demos and the wording of the exercises were written for this repository; any mistakes are mine.

How to use them. Each chapter has three parts. The notes, in this document, develop the theory with complete proofs wherever the argument teaches something, and with precise references where it does not. The exercises are collected in a separate document; solutions are not published, so that the course is not spoiled for the next class. The interactive demos, marked ▶, are Python notebooks that open in Google Colab in one click.

Prerequisites. Stochastic calculus for finance at the level of a first course: Brownian motion, Itô's formula, risk-neutral pricing and the Black–Scholes model. The companion notes for IEOR 222 cover all of it. Python with NumPy and SciPy.

Corrections. Please report errors, typos and unclear passages as issues on the GitHub repository.

Notation

$(\Omega, \mathcal{F}, \mathbb{P})$	probability space; $\mathbb{P}$ is the real-world (historical) measure
$\mathbb{Q}$, $\mathbb{E}^{\mathbb{Q}}$	a risk-neutral measure and its expectation
$(\mathcal{F}_t)_{t \geq 0}$	filtration: the information available at time t
W	a Brownian motion
$\mathcal{N}(m, v)$, $N(\cdot)$	normal law with mean m and variance v; standard normal CDF
x^+, $x \wedge y$	$\max(x, 0)$; $\min(x, y)$
$\mathbf{1}_A$	indicator of the event A
$[X, X]_t$	quadratic variation of X on $[0, t]$
r, σ, K, T	interest rate, volatility, strike, maturity

Part I

Analytical pricing

Chapter 1

The standard model, revisited

Every method in these notes starts from two facts about the Black–Scholes–Merton model with a dividend yield. A price is a discounted expectation under a risk-neutral measure, and the same price solves a parabolic partial differential equation. Monte Carlo methods compute the expectation (Part II); trees and finite-difference grids solve the equation (Chapter 20). The Feynman–Kac formula says that the two coincide, and the Green's function of the equation, the price of a claim that pays at a single terminal stock price, connects both to the static replication of Chapters 3 and 4.

The IEOR 222 notes build Brownian motion, Itô calculus and risk-neutral pricing from the ground up. Here we recall what the course uses, add the dividend yield, and prove what later chapters lean on: the pricing formula with a yield (Section 1.2), the Feynman–Kac formula for a diffusion with a state-dependent discount rate (Section 1.3), and the Green's function of the pricing equation, with the precise sense in which it solves the equation with a Dirac delta as terminal value (Section 1.4). Along the way we see that a pricing equation can have solutions that are not prices.

Throughout, $T > 0$ is a fixed horizon, N and N' are the standard normal distribution function and density, and $\mathcal{N}(m, v)$ is the normal law with mean m and variance v.

1.1 Brownian motion and Itô's formula, in brief

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with a filtration $(\mathcal{F}_t)_{0 \le t \le T}$ whose $\mathcal{F}_0$ contains the events of probability zero. A *Brownian motion* for $(\mathcal{F}_t)$ is a continuous adapted process W with $W_0 = 0$ whose increments $W_t - W_s$, $s < t$, are $\mathcal{N}(0, t-s)$ and independent of $\mathcal{F}_s$. Then $\mathrm{Cov}(W_s, W_t) = \min(s, t)$, and W_t has the density

$$k(t, x) = \frac{1}{\sqrt{2\pi t}}\, e^{-x^2/(2t)}, \qquad t > 0,\ x \in \mathbb{R}, \tag{1.1}$$

the *heat kernel*. Differentiating gives $\partial_x k = -\frac{x}{t} k$ and $\partial_t k = (\frac{x^2}{2t^2} - \frac{1}{2t})k = \frac{1}{2}\partial_{xx} k$, and as $t \downarrow 0$ the kernel concentrates at 0. Section 1.4 is built on these two facts.

Cumulative sums of independent $\mathcal{N}(0, h)$ increments have exactly the law of W on a grid of step h, which is how Monte Carlo builds paths (Chapter 7); steps of $\pm\sqrt{h}$ with probability $\frac{1}{2}$ give a walk

whose value at each fixed time converges in law to that of W (central limit theorem), which is the binomial model (Chapter 20; a trinomial variant is Exercise 1.1). The paths of W are nowhere differentiable, and along partitions of $[0,t]$ with mesh going to zero $\sum_j (W_{t_{j+1}} - W_{t_j})^2 \to t$ in L^2 (see the IEOR 222 notes): the *quadratic variation* $[W,W]_t = t$, source of the rules $\mathrm{d}W\,\mathrm{d}W = \mathrm{d}t$ and $\mathrm{d}W\,\mathrm{d}t = \mathrm{d}t\,\mathrm{d}t = 0$.

Itô integrals. Let H be progressively measurable with $\int_0^T H_s^2\,\mathrm{d}s < \infty$ almost surely. The Itô integral $M_t = \int_0^t H_s\,\mathrm{d}W_s$ is then a continuous *local* martingale. If $\mathbb{E}\int_0^T H_s^2\,\mathrm{d}s < \infty$, it is a square-integrable martingale and $\mathbb{E}[M_t^2] = \mathbb{E}\int_0^t H_s^2\,\mathrm{d}s$ (Itô's isometry). Without this condition M need not be a martingale; the hypotheses of the Feynman–Kac formula exclude that case.

Itô's formula. Let $I \subseteq \mathbb{R}$ be an open interval and X a continuous process with values in I and $X_t = X_0 + \int_0^t b_s\,\mathrm{d}s + \int_0^t \sigma_s\,\mathrm{d}W_s$, for adapted processes b, σ with $\int_0^T (|b_s| + \sigma_s^2)\,\mathrm{d}s < \infty$. A function f is in $C^{1,2}([0,T)\times I)$ if $\partial_t f$, $\partial_x f$, $\partial_{xx} f$ exist and are continuous on $[0,T)\times I$. For such f and $t < T$,

$$\mathrm{d}f(t,X_t) = \Big(\partial_t f + b_t\,\partial_x f + \frac{\sigma_t^2}{2}\,\partial_{xx} f\Big)(t,X_t)\,\mathrm{d}t + \sigma_t\,\partial_x f(t,X_t)\,\mathrm{d}W_t. \tag{1.2}$$

If A is a continuous adapted process of finite variation (no $\mathrm{d}W$ part), the product rule reads $\mathrm{d}(AY) = A\,\mathrm{d}Y + Y\,\mathrm{d}A$ for every Itô process Y. We use (1.2) and the product rule without proof; see Shreve (2004), Sections 4.3–4.4, or Karatzas and Shreve (1991), Section 3.3. Those versions are for f on $[0,\infty)\times\mathbb{R}$; for an open $I = (\alpha,\beta)$ and $t \le T' < T$, take $\alpha < \alpha_n < \beta_n < \beta$ with $\alpha_n \downarrow \alpha$, $\beta_n \uparrow \beta$, apply them to $f\chi_n$, with χ_n smooth, equal to 1 on $[\alpha_n,\beta_n]$ and vanishing near α and β, up to the exit time of X from (α_n,β_n), and note that this time exceeds T' for large n, since the compact set $X([0,T'])$ lies in I.

1.2 The standard model

Definition 1.1 (The standard model)**.** Let $r, y, \mu \in \mathbb{R}$, $\sigma > 0$ and $S_0 > 0$ be constants, let W be a Brownian motion under $\mathbb{P}$, and let $(\mathcal{F}_t)_{0\le t\le T}$ be the filtration generated by W and augmented by the $\mathbb{P}$-null events. The *standard model* is the market with

(i) a money-market account worth $B_t = e^{rt}$ at time t;

(ii) a stock that pays dividends continuously, at the rate $yS_t\,\mathrm{d}t$ per share, and whose price solves

$$\mathrm{d}S_t = (\mu - y)\,S_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}W_t, \qquad \text{that is,} \quad S_t = S_0 \exp\Big(\Big(\mu - y - \frac{\sigma^2}{2}\Big)t + \sigma W_t\Big); \tag{1.3}$$

(iii) the *risk-neutral measure* $\mathbb{Q}$ on $\mathcal{F}_T$, defined by $\mathbb{Q}(A) = \mathbb{E}[Z_T \mathbf{1}_A]$ with

$$Z_t = \exp\Big(-\theta W_t - \frac{\theta^2}{2}t\Big), \qquad \theta = \frac{\mu - r}{\sigma} \quad \text{(the market price of risk)},$$

and the process $W_t^{\mathbb{Q}} = W_t + \theta t$.

We write $\nu = r - y - \frac{\sigma^2}{2}$ for the drift of $\log S$ under $\mathbb{Q}$.

The constant μ is the expected rate of *total* return under $\mathbb{P}$: price appreciation, $\mu - y$, plus the dividend yield y. Paying dividends out lowers the drift of the price, not the return of the shareholder. Continuous dividends are an idealisation that fits an index well. For a currency, y is the foreign interest rate (Chapter 17); for a commodity it is a convenience yield net of storage costs. Time-dependent r, y and σ are treated in Chapter 7.

Lemma 1.2 (The two measures)**.** *In the standard model:*

(i) Z *is a* $\mathbb{P}$*-martingale with* $Z_T > 0$ *and* $\mathbb{E}[Z_T] = 1$*, so* $\mathbb{Q}$ *is a probability measure on* $\mathcal{F}_T$ *with the same null events as* $\mathbb{P}$*;*

(ii) $W^{\mathbb{Q}}$ *is a Brownian motion under* $\mathbb{Q}$ *for the filtration* $(\mathcal{F}_t)$*, and*

$$\mathrm{d}S_t = (r - y)\, S_t\, \mathrm{d}t + \sigma S_t\, \mathrm{d}W_t^{\mathbb{Q}}, \qquad S_t = S_0 \exp\left(\nu t + \sigma W_t^{\mathbb{Q}}\right); \tag{1.4}$$

(iii) for $0 \le s \le t \le T$ *and every real* q*,*

$$\mathbb{E}^{\mathbb{Q}}\left[S_t^q \,\middle|\, \mathcal{F}_s\right] = S_s^q \exp\left(q(r-y)(t-s) + \frac{q(q-1)}{2}\sigma^2(t-s)\right), \tag{1.5}$$

and the same holds under $\mathbb{P}$ *with* μ *in place of* r*. In particular* $e^{-(r-y)t}S_t$*, and the discounted value* $e^{-rt}\,e^{yt}S_t$ *of one share bought at time* 0 *with its dividends reinvested in shares, are* $\mathbb{Q}$*-martingales;*

(iv) $\mathbb{E}^{\mathbb{Q}}\left[\sup_{t\le T} S_t^q\right] \le 2S_0^q \exp\left((|q\nu + \tfrac{1}{2}q^2\sigma^2| + \tfrac{1}{2}q^2\sigma^2)T\right) < \infty$ *for every real* q*, and similarly under* $\mathbb{P}$*.*

Proof. (i) For $s < t$, $Z_t = Z_s \exp(-\theta(W_t - W_s) - \frac{\theta^2}{2}(t-s))$, and the increment is independent of $\mathcal{F}_s$ with $\mathbb{E}[e^{-\theta(W_t - W_s)}] = e^{\theta^2(t-s)/2}$ (the normal moment generating function). So $\mathbb{E}[Z_t \mid \mathcal{F}_s] = Z_s$, and $\mathbb{E}[Z_T] = Z_0 = 1$. Since $Z_T > 0$, $\mathbb{Q}(A) = 0$ if and only if $\mathbb{P}(A) = 0$.

(ii) By Girsanov's theorem (Karatzas and Shreve (1991), Theorem 3.5.1), applied to the martingale Z of (i), $W_t + \theta t$ is a $\mathbb{Q}$-Brownian motion for $(\mathcal{F}_t)$ on $[0,T]$; Novikov's condition $\mathbb{E}[\exp(\frac{1}{2}\theta^2 T)] < \infty$, which also implies (i), holds trivially because θ is constant. Substituting $\mathrm{d}W_t = \mathrm{d}W_t^{\mathbb{Q}} - \theta\,\mathrm{d}t$ in (1.3) turns the drift into $\mu - y - \sigma\theta = r - y$, and the exponent of (1.3) equals $\nu t + \sigma W_t^{\mathbb{Q}}$.

(iii) By (1.4), $S_t^q = S_s^q \exp\left(q\nu(t-s) + q\sigma(W_t^{\mathbb{Q}} - W_s^{\mathbb{Q}})\right)$, and the increment is $\mathcal{N}(0, t-s)$ under $\mathbb{Q}$ and independent of $\mathcal{F}_s$. The normal moment generating function gives the factor $\exp(q\nu(t-s) + \frac{1}{2}q^2\sigma^2(t-s))$, which is (1.5). Under $\mathbb{P}$ use (1.3) instead. Taking $q = 1$ gives the two martingales.

(iv) Write $S_t^q = S_0^q e^{at} M_t$ with $a = q\nu + \frac{1}{2}q^2\sigma^2$ and $M_t = \exp(q\sigma W_t^{\mathbb{Q}} - \frac{1}{2}q^2\sigma^2 t)$, a continuous $\mathbb{Q}$-martingale with $\mathbb{E}^{\mathbb{Q}}[M_T^2] = e^{q^2\sigma^2 T}$. Doob's maximal inequality (Karatzas and Shreve (1991), Theorem 1.3.8(iv)) gives $\mathbb{E}^{\mathbb{Q}}[\sup_{t\le T} M_t^2] \le 4\mathbb{E}^{\mathbb{Q}}[M_T^2]$, and then $\mathbb{E}^{\mathbb{Q}}[\sup_t S_t^q] \le S_0^q e^{|a|T}\left(\mathbb{E}^{\mathbb{Q}}[\sup_t M_t^2]\right)^{1/2}$. □

Trading with dividends. A self-financing strategy starts with wealth X_0, holds Δ_t shares at time t, and keeps the rest of its wealth, $X_t - \Delta_t S_t$, in the money-market account; dividends are paid into the account. Over $[t, t + \mathrm{d}t]$ the wealth gains the price change, the dividends and the

interest:

$$\mathrm{d}X_t = \Delta_t\,\mathrm{d}S_t + y\,\Delta_t S_t\,\mathrm{d}t + r\,(X_t - \Delta_t S_t)\,\mathrm{d}t. \tag{1.6}$$

We call Δ *admissible* if it is progressively measurable and $\mathbb{E}^{\mathbb{Q}}\int_0^T \Delta_t^2 S_t^2\,\mathrm{d}t < \infty$. By the product rule and (1.4),

$$\mathrm{d}(e^{-rt}X_t) = e^{-rt}\Delta_t(\mathrm{d}S_t - (r-y)S_t\,\mathrm{d}t) = e^{-rt}\sigma\Delta_t S_t\,\mathrm{d}W_t^{\mathbb{Q}}, \tag{1.7}$$

so $X_t = e^{rt}(X_0 + \int_0^t e^{-ru}\sigma\Delta_u S_u\,\mathrm{d}W_u^{\mathbb{Q}})$ is the unique solution of (1.6). The wealth does not depend on whether the integral against $\mathrm{d}S$ is built under $\mathbb{P}$ or under $\mathbb{Q}$: for simple integrands it is a finite sum, the same under both; a general Δ is approximated by simple ones whose integrals converge in probability (Karatzas and Shreve (1991), Section 3.2), and convergence in probability, with its limit, is preserved by equivalent measures. The real-world drift μ has disappeared from (1.7).

Proposition 1.3 (Risk-neutral pricing with a dividend yield)**.** *In the standard model, let $f : (0,\infty) \to \mathbb{R}$ be Borel with $\mathbb{E}^{\mathbb{Q}}|f(S_T)| < \infty$. For $0 \le t < T$ and $x > 0$ let*

$$v(t,x) = e^{-r(T-t)}\int_{\mathbb{R}} f\big(x\,e^{\nu(T-t)+\sigma\sqrt{T-t}\,z}\big)\,N'(z)\,\mathrm{d}z \tag{1.8}$$

wherever the integral converges absolutely, and let $v(T,x) = f(x)$.

(i) Price formula. *For every $t \in [0,T)$ the integral converges absolutely at $x = S_t$, $\mathbb{Q}$-almost surely, and for every $t \in [0,T]$*

$$V_t := e^{-r(T-t)}\,\mathbb{E}^{\mathbb{Q}}[f(S_T)\,|\,\mathcal{F}_t] = v(t,S_t) \qquad \text{almost surely.} \tag{1.9}$$

In particular $V_0 = e^{-rT}\mathbb{E}^{\mathbb{Q}}[f(S_T)] = v(0,S_0)$, and $e^{-rt}V_t$ is a $\mathbb{Q}$-martingale.

(ii) Replication and no arbitrage. *If Δ is admissible, the discounted wealth $e^{-rt}X_t$ is a $\mathbb{Q}$-martingale. Hence, if $X_T = f(S_T)$ almost surely, then $X_t = V_t$ for every $t \in [0,T]$ almost surely; and if $X_0 = 0$ and $X_T \ge 0$ almost surely, then $X_T = 0$ almost surely.*

(iii) Existence of the hedge. *If $\mathbb{E}^{\mathbb{Q}}[f(S_T)^2] < \infty$, there is an admissible Δ whose wealth started from $X_0 = V_0$ ends at $X_T = f(S_T)$.*

(iv) Path-dependent claims. *(ii) and (iii) hold for every $\mathcal{F}_T$-measurable claim ξ in place of $f(S_T)$ (integrable for (ii), with $\mathbb{E}^{\mathbb{Q}}[\xi^2] < \infty$ for (iii)), with $V_t = e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}[\xi \mid \mathcal{F}_t]$.*

So a square-integrable claim, path-dependent or not, can be replicated, every admissible replicating strategy is worth V_t at time t, and trading the claim at any other price against the replicating strategy is an arbitrage. The price depends on r, y and σ, not on μ.

Proof. (i) Write $S_T = S_t R$ with $R = \exp\big(\nu(T-t)+\sigma(W_T^{\mathbb{Q}} - W_t^{\mathbb{Q}})\big)$. Under $\mathbb{Q}$, S_t is $\mathcal{F}_t$-measurable and R is independent of $\mathcal{F}_t$, with the law of $e^{\nu(T-t)+\sigma\sqrt{T-t}\,Z}$, $Z \sim \mathcal{N}(0,1)$ (Theorem 1.2(ii)). By the independence lemma (Shreve (2004), Section 2.3), applied to the nonnegative functions f^+ and f^-, $\mathbb{E}^{\mathbb{Q}}[f^{\pm}(S_T) \mid \mathcal{F}_t] = e^{r(T-t)}v^{\pm}(t,S_t)$, where $v^{\pm} \in [0,\infty]$ is (1.8) with $f^{\pm}$ in place of f. Both are finite almost surely because $\mathbb{E}^{\mathbb{Q}}[f^{\pm}(S_T)] < \infty$; subtracting gives (1.9). The martingale property is the tower property: $e^{-rt}V_t = \mathbb{E}^{\mathbb{Q}}[e^{-rT}f(S_T) \mid \mathcal{F}_t]$.

(ii) By (1.7), $e^{-rt}X_t - X_0$ is the Itô integral of $e^{-ru}\sigma\Delta_u S_u$ against the $\mathbb{Q}$-Brownian motion $W^{\mathbb{Q}}$, and $\mathbb{E}^{\mathbb{Q}}\int_0^T e^{-2ru}\sigma^2\Delta_u^2 S_u^2\,\mathrm{d}u \le e^{2|r|T}\sigma^2\mathbb{E}^{\mathbb{Q}}\int_0^T \Delta_u^2 S_u^2\,\mathrm{d}u < \infty$, so it is a $\mathbb{Q}$-martingale. If

$X_T = f(S_T)$, then $e^{-rt}X_t = \mathbb{E}^{\mathbb{Q}}[e^{-rT}f(S_T) \mid \mathcal{F}_t] = e^{-rt}V_t$. If $X_0 = 0$ and $X_T \geq 0$, then $\mathbb{E}^{\mathbb{Q}}[e^{-rT}X_T] = X_0 = 0$, so $X_T = 0$ $\mathbb{Q}$-almost surely, hence $\mathbb{P}$-almost surely by Theorem 1.2(i).

(iii) The $\mathbb{Q}$-martingale $M_t = e^{-rt}V_t$ is square integrable. Since θ is constant, W and $W^{\mathbb{Q}}$ generate the same filtration, and the null events of $\mathbb{P}$ and $\mathbb{Q}$ coincide, so $(\mathcal{F}_t)$ is the augmented filtration of the $\mathbb{Q}$-Brownian motion $W^{\mathbb{Q}}$. By the martingale representation theorem (Karatzas and Shreve (1991), Section 3.4.D; Øksendal (2003), Theorem 4.3.4), $M_t = M_0 + \int_0^t \varphi_u \, \mathrm{d}W_u^{\mathbb{Q}}$ for a progressively measurable φ with $\mathbb{E}^{\mathbb{Q}} \int_0^T \varphi_u^2 \, \mathrm{d}u < \infty$. Then $\Delta_t = e^{rt}\varphi_t/(\sigma S_t)$ is admissible, and by (1.7) the wealth started from V_0 satisfies $e^{-rt}X_t = M_t$, so $X_T = V_T = f(S_T)$.

(iv) The proofs of (ii) and (iii) use only that $V_T = f(S_T)$ is $\mathcal{F}_T$-measurable, integrable or square integrable, and that $e^{-rt}V_t = \mathbb{E}^{\mathbb{Q}}[e^{-rT}f(S_T) \mid \mathcal{F}_t]$; they apply verbatim to ξ. □

Remark 1.4 (Forwards and the test case). For $f(x) = x$, (1.5) gives $V_t = e^{-r(T-t)}S_t e^{(r-y)(T-t)} = S_t e^{-y(T-t)}$: a claim to the share at T, without the dividends paid before T, is worth less than the share. The *forward price*, the delivery price F that makes a forward contract worth nothing, solves $e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}[S_T - F \mid \mathcal{F}_t] = 0$, so $F_{t,T} = S_t e^{(r-y)(T-t)} = \mathbb{E}^{\mathbb{Q}}[S_T \mid \mathcal{F}_t]$. For the course's test case $S_0 = K = 100$, $T = 1$, $r = 1\%$, $y = 2\%$, $\sigma = 20\%$, the forward is $100\, e^{-0.01} = 99.005$, below the spot because the yield exceeds the rate, and the call is worth 7.36429 (Chapter 2). A call with strike $K = 0$ pays S_T and is worth $S_0 e^{-yT}$.

Remark 1.5 (Edge cases). At $t = T$, (1.9) reads $V_T = f(S_T)$; with $y = 0$, $e^{-rt}S_t$ itself is a $\mathbb{Q}$-martingale. If $\sigma = 0$ (excluded by the definition), the stock is riskless with total return μ, absence of arbitrage forces $\mu = r$, and the price of $f(S_T)$ is $e^{-rT}f(S_0 e^{(r-y)T})$: for continuous f with $|f(x)| \leq a(x^q + x^{-q})$, the limit of (1.8) as $\sigma \downarrow 0$, by dominated convergence. The hypothesis $\mathbb{E}^{\mathbb{Q}}|f(S_T)| < \infty$ is not automatic: $\mathbb{E}^{\mathbb{Q}}[e^{S_T}] = \infty$, because a lognormal variable has no exponential moments, so the claim paying e^{S_T} has no finite price.

▶ Interactive demo: simulate the stock under $\mathbb{P}$ and under $\mathbb{Q}$ from the same normals, and price a call from real-world samples by reweighting them with Z_T.

1.3 The Feynman–Kac formula

Why should a price solve an equation? By Theorem 1.3, $e^{-rt}v(t, S_t)$ is a $\mathbb{Q}$-martingale. If v is smooth, Itô's formula (1.2) writes its differential as $e^{-rt}(\partial_t v + (r-y)x\,\partial_x v + \frac{1}{2}\sigma^2 x^2 \partial_{xx} v - rv)(t, S_t)\,\mathrm{d}t$ plus a $\mathrm{d}W^{\mathbb{Q}}$ term. A martingale has no drift, so the bracket vanishes along the path, and hence wherever the path can go. This heuristic assumes that v is smooth, the hard part. The Feynman–Kac formula goes the other way, as pricing by equations needs: a smooth solution is the expectation. The argument works for any diffusion.

Setting. Let $I = (\alpha, \beta)$ with $-\infty \leq \alpha < \beta \leq \infty$, and let $b, \sigma, r, c : [0, T] \times I \to \mathbb{R}$ be continuous: a drift, a volatility, a discount rate and a running payoff. Let $\mathbb{Q}$ be a probability measure (in applications, a risk-neutral measure), W a $\mathbb{Q}$-Brownian motion for $(\mathcal{F}_t)$, and X a continuous adapted process with values in I and

$$X_t = x_0 + \int_0^t b(s, X_s)\,\mathrm{d}s + \int_0^t \sigma(s, X_s)\,\mathrm{d}W_s, \qquad 0 \leq t \leq T,$$

for some $x_0 \in I$. Write $\mathcal{L}_t f = \frac{1}{2}\sigma(t,\cdot)^2 f'' + b(t,\cdot) f'$ and $D_t = \exp\big(-\int_0^t r(s,X_s)\,\mathrm{d}s\big)$.

Theorem 1.6 (Feynman–Kac). *Let $g : I \to \mathbb{R}$ be Borel, $E \subseteq I$ a Borel set with $\mathbb{Q}(X_T \in E) = 0$, and $u \in C^{1,2}([0,T) \times I)$ a solution of*

$$\partial_t u + \mathcal{L}_t u - r\,u + c = 0 \qquad \text{on } [0,T) \times I, \tag{1.10}$$

$$\lim_{(s,z)\to(T,x),\ s<T} u(s,z) = g(x) \qquad \text{for every } x \in I \setminus E. \tag{1.11}$$

Let $H_s = D_s\,\sigma(s,X_s)\,\partial_x u(s,X_s)$ for $s < T$, and assume (a) or (b):

(a) $\int_0^T H_s^2\,\mathrm{d}s < \infty$ almost surely, the Itô integral $\int_0^t H_s\,\mathrm{d}W_s$, $t \in [0,T]$, is a $\mathbb{Q}$-martingale (for instance, $\mathbb{E}^{\mathbb{Q}}\int_0^T H_s^2\,\mathrm{d}s < \infty$), and $\mathbb{E}^{\mathbb{Q}}\int_0^T D_s|c(s,X_s)|\,\mathrm{d}s < \infty$;

(b) $r \geq -r_0$ for a constant $r_0 \geq 0$, and there are a constant C and a continuous $\psi : I \to [1,\infty)$ with $|u(s,x)| + |c(s,x)| \leq C\psi(x)$ on $[0,T) \times I$ and $\mathbb{E}^{\mathbb{Q}}\big[\sup_{s\leq T}\psi(X_s)\big] < \infty$.

Then $D_T g(X_T)$ and $\int_0^T D_s c(s,X_s)\,\mathrm{d}s$ are integrable, and the process

$$Y_t = D_t\,u(t,X_t) + \int_0^t D_s\,c(s,X_s)\,\mathrm{d}s \quad (t<T), \qquad Y_T = D_T\,g(X_T) + \int_0^T D_s\,c(s,X_s)\,\mathrm{d}s,$$

is a $\mathbb{Q}$-martingale on $[0,T]$. In particular

$$u(0,x_0) = \mathbb{E}^{\mathbb{Q}}\Big[\int_0^T D_s\,c(s,X_s)\,\mathrm{d}s + D_T\,g(X_T)\Big], \tag{1.12}$$

and, in case (b), and in case (a) if r is bounded above, for every $t \in [0,T)$, almost surely,

$$u(t,X_t) = \mathbb{E}^{\mathbb{Q}}\Big[\int_t^T e^{-\int_t^s r(w,X_w)\,\mathrm{d}w} c(s,X_s)\,\mathrm{d}s + e^{-\int_t^T r(w,X_w)\,\mathrm{d}w} g(X_T)\,\Big|\,\mathcal{F}_t\Big]. \tag{1.13}$$

The theorem applies to a solution $X^{t,x}$ started at x at a time t (when there is one): with $\widetilde{W}_s = W_{t+s} - W_t$ and $\widetilde{\mathcal{F}}_s = \mathcal{F}_{t+s}$ it solves a shifted equation on $[0,T-t]$, and (1.12) gives $u(t,x)$ as the unconditional expectation of the variable in (1.13), with $X^{t,x}$ in place of X. The set E is there for discontinuous payoffs: no solution is continuous at (T,K) for the digital payoff $\mathbf{1}_{(K,\infty)}$, and $E = \{K\}$ is harmless when X_T has a density.

Proof. No drift. Fix $T' < T$. Itô's formula (1.2) on $[0,T']$ (where u is $C^{1,2}$ and X stays in I), the product rule with the finite-variation factor D, $\mathrm{d}D_t = -r(t,X_t)D_t\,\mathrm{d}t$, and (1.10) give

$$\mathrm{d}Y_t = D_t(\partial_t u + \mathcal{L}_t u - ru + c)(t,X_t)\,\mathrm{d}t + H_t\,\mathrm{d}W_t = H_t\,\mathrm{d}W_t, \qquad t \leq T'.$$

As $T' < T$ is arbitrary, almost surely $Y_t = Y_0 + \int_0^t H_s\,\mathrm{d}W_s$ for all $t < T$. On the event $\{X_T \notin E\}$, which has probability one, $Y_t \to Y_T$ as $t \uparrow T$: D, X and the running integral are continuous, and (1.11) applies along $(t,X_t) \to (T,X_T)$.

Case (a). The integral $M_t = \int_0^t H_s\,\mathrm{d}W_s$ is continuous on $[0,T]$, so letting $t \uparrow T$ gives $Y_T = Y_0 + M_T$ almost surely. As M is a martingale, Y_T is integrable and $\mathbb{E}^{\mathbb{Q}}[Y_T \mid \mathcal{F}_t] = Y_0 + M_t = Y_t$. Since $\int_0^T D_s|c(s,X_s)|\,\mathrm{d}s$ is integrable by assumption, so is $D_T g(X_T) = Y_T - \int_0^T D_s c(s,X_s)\,\mathrm{d}s$.

Case (b): localization. Now $D_t \leq e^{r_0 T}$, and $G = e^{r_0 T}(1+T)\,C\sup_{s\leq T}\psi(X_s)$ is integrable. Every Y_t, $t < T$, satisfies $|Y_t| \leq G$, and so does Y_T on $\{X_T \notin E\}$, because $|g(X_T)| = \lim_t |u(t,X_t)| \leq$

$C\psi(X_T)$ there. With α_n, β_n as in Section 1.1 and $x_0 \in (\alpha_1, \beta_1)$, let $T_n = T - T/(n+1)$, and let ϑ_n be the first time $t \le T_n$ at which $X_t \notin (\alpha_n, \beta_n)$, or T_n if there is none. It is a stopping time, as the hitting time of a closed set by a continuous adapted process (Karatzas and Shreve (1991), Section 1.2). For $s \le \vartheta_n$ the point (s, X_s) lies in the compact set $[0, T_n] \times [\alpha_n, \beta_n]$, on which σ, $\partial_x u$ and D are bounded, so $H\mathbf{1}_{[0,\vartheta_n]}$ is bounded, its integral is a martingale, and that integral is $Y_{t\wedge\vartheta_n} - Y_0$ (Karatzas and Shreve (1991), Section 3.2). Hence, for $t < T$,

$$Y_{t\wedge\vartheta_n} = \mathbb{E}^{\mathbb{Q}}[Y_{\vartheta_n} \,|\, \mathcal{F}_t].$$

For each outcome, $X([0,T])$ is a compact subset of I, contained in (α_n, β_n) for all large n; then $\vartheta_n = T_n \uparrow T$, so $Y_{t\wedge\vartheta_n} \to Y_t$ and, on $\{X_T \notin E\}$, $Y_{\vartheta_n} \to Y_T$. By dominated convergence (bound G), $Y_{\vartheta_n} \to Y_T$ in L^1, and since conditional expectation is a contraction in L^1, $\mathbb{E}^{\mathbb{Q}}[Y_{\vartheta_n} \mid \mathcal{F}_t] \to \mathbb{E}^{\mathbb{Q}}[Y_T \mid \mathcal{F}_t]$ in L^1. The left side converges almost surely to Y_t, so $Y_t = \mathbb{E}^{\mathbb{Q}}[Y_T \mid \mathcal{F}_t]$. Integrability of the two terms of Y_T follows from the bound G.

Conclusion. The martingale property at $t = 0$ is (1.12), since $Y_0 = u(0, x_0)$. For (1.13), subtract the $\mathcal{F}_t$-measurable $\int_0^t D_s c\,\mathrm{d}s$ from both sides of $Y_t = \mathbb{E}^{\mathbb{Q}}[Y_T \mid \mathcal{F}_t]$ and multiply by the $\mathcal{F}_t$-measurable D_t^{-1}; this is allowed when the product of D_t^{-1} with the integrable variable inside the expectation is integrable. In case (b) that product is bounded by G; in case (a) with $r \le \bar{r}$, $D_t^{-1} \le e^{|\bar{r}|T}$. □

Remark 1.7 (On the hypotheses).

(a) *Growth and moments.* Hypothesis (b) holds with $\psi \equiv 1$ when u and c are bounded and r is bounded below. For the stock of the standard model, $\psi(x) = x^p + x^{-p}$ works for every $p \ge 0$, by Theorem 1.2(iv). For $I = \mathbb{R}$ and coefficients that are Lipschitz in x uniformly in t, $\psi(x) = 1 + x^2$ works, because then $\mathbb{E}^{\mathbb{Q}}[\sup_{t\le T} X_t^2] < \infty$ (proved in the IEOR 222 notes).

(b) *Uniqueness.* If the hypotheses hold for the solutions started from every (t, x), two solutions of (1.10)–(1.11) in the class allowed by (b) are both equal to the same expectation, hence to each other: the theorem is also a uniqueness theorem. Without a restriction of this kind uniqueness fails. Tychonoff constructed a solution of the heat equation that vanishes at the initial time and is not identically zero; it grows faster than e^{ax^2} for every a, and uniqueness does hold among solutions bounded by Ce^{ax^2} (Karatzas and Shreve (1991), Section 4.3). In the variables of Section 1.4 this gives nonzero solutions of the pricing equation with zero terminal value. A solution can also fail by blowing up near the terminal time at a single point (Exercise 1.7).

(c) *Discounting.* The lower bound on r in (b) keeps D bounded. Case (a) has no such bound, and the next example needs this.

Remark 1.8 (The converse). That the expectation is a $C^{1,2}$ solution is the harder direction, which we do not prove in general. Karatzas and Shreve (1991), Theorem 5.7.6, proves (1.13) for solutions of polynomial growth, Lipschitz coefficients and a nonnegative discount rate; their Remark 5.7.8 recalls when a solution exists (σ^2 bounded below by a positive constant, bounded Hölder coefficients, a continuous payoff of polynomial growth), by the parametrix construction of Friedman (1964), Chapter 1. Øksendal (2003), Theorem 8.2.1, proves both directions for time-independent coefficients, a discount rate bounded below and a C^2 payoff with compact support. For the standard model, Theorem 1.13 proves the converse for every payoff with polynomial growth in ξ and $1/\xi$.

Example 1.9 (A bond with a random short rate)**.** Let the short rate be an Ornstein–Uhlenbeck process under $\mathbb{Q}$, $\mathrm{d}X_t = \kappa(\bar{x} - X_t)\,\mathrm{d}t + \gamma\,\mathrm{d}W_t$ with $\kappa, \gamma > 0$, on $I = \mathbb{R}$, and price the zero-coupon bond paying 1 at T by discounting at $r(t,x) = x$ (so $c = 0$, $g = 1$, $E = \emptyset$). Try $u(t,x) = \exp\big(A(T-t) - B(T-t)\,x\big)$. Substituting in (1.10) and collecting the terms in x and the constant terms gives $B' = 1 - \kappa B$ and $A' = -\kappa\bar{x}B + \frac{1}{2}\gamma^2 B^2$, with $A(0) = B(0) = 0$, which are solved by

$$B(\tau) = \frac{1 - e^{-\kappa\tau}}{\kappa}, \qquad A(\tau) = \Big(\bar{x} - \frac{\gamma^2}{2\kappa^2}\Big)\big(B(\tau) - \tau\big) - \frac{\gamma^2 B(\tau)^2}{4\kappa}. \tag{1.14}$$

Hypothesis (b) fails: on $I = \mathbb{R}$ the rate $r(t,x) = x$ is unbounded below, so D is unbounded. Hypothesis (a) holds. Here $H_s = -\gamma B(T-s)\exp\big(A(T-s) - B(T-s)X_s - \int_0^s X_w\,\mathrm{d}w\big)$, and since X is a Gaussian process, the exponent is a Gaussian variable whose mean and variance depend continuously on s; so $\mathbb{E}^{\mathbb{Q}}[H_s^2]$ is bounded on $[0,T]$ and $\mathbb{E}^{\mathbb{Q}}\int_0^T H_s^2\,\mathrm{d}s < \infty$. By (1.12), the bond is worth $u(0,x_0) = \mathbb{E}^{\mathbb{Q}}\big[\exp(-\int_0^T X_s\,\mathrm{d}s)\big]$. With $\kappa = 0.5$, $\bar{x} = 4\%$, $\gamma = 2\%$, $T = 5$ and $x_0 = 4\%$, $u(0,x_0) = 0.820253$. Our implementation (numpy `default_rng`, seed 2029), with 10^5 exact paths of X on 50 steps and the trapezoid rule for the integral, gives 0.820203 with standard error 1.6×10^{-4}. Figure 1.1 repeats the comparison for seven initial rates.

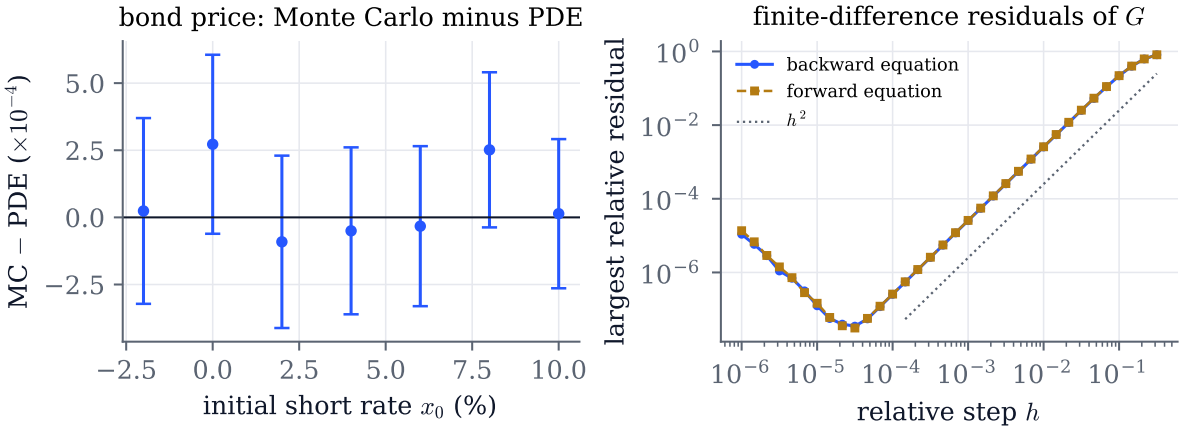

Figure 1.1: Two numerical checks. Left: the bond of Theorem 1.9 ($\kappa = 0.5$, $\bar{x} = 4\%$, $\gamma = 2\%$, $T = 5$), Monte Carlo estimate minus the solution (1.14) of the pricing equation, with 95% confidence intervals, for seven initial rates x_0. Each estimate uses 10^5 exact paths of X with 50 steps and its own seed (2026 to 2032); every interval contains zero. Right: the largest relative residual of the backward and forward equations (1.18)–(1.19) for the Green's function of the test case ($T - t = 0.25$, 61 points), computed by central differences with relative step h. It decreases like h^2 until rounding errors, which grow as h shrinks, take over.

For the standard model, Theorem 1.6 becomes the following.

> **Corollary 1.10** (The pricing equation of the standard model)**.** *Let $f : (0,\infty) \to \mathbb{R}$ be Borel, and let E be a Lebesgue-null Borel set. Let $v \in C^{1,2}([0,T) \times (0,\infty))$ solve*
>
> $$\partial_t v + (r - y)\,x\,\partial_x v + \frac{\sigma^2 x^2}{2}\,\partial_{xx} v - r\,v = 0 \qquad \text{on } [0,T) \times (0,\infty), \tag{1.15}$$
>
> *with $|v(t,x)| \le C(x^p + x^{-p})$ for constants $C, p \ge 0$ and $v(t,x) \to f(x_0)$ as $(t,x) \to (T,x_0)$, for*

every $x_0 \notin E$. Then $\mathbb{E}^{\mathbb{Q}}|f(S_T)| < \infty$, v is the function (1.8), and $v(t, S_t) = V_t$ is the price of Theorem 1.3 for every $t \in [0, T)$. In particular (1.15) has at most one solution with these properties.

Proof. Apply Theorem 1.6(b) under $\mathbb{Q}$ to the stock started at x at time t, which is $S_s^{t,x} = x \exp\left(\nu(s-t) + \sigma(W_s^{\mathbb{Q}} - W_t^{\mathbb{Q}})\right)$: here $I = (0, \infty)$, $b(t, x) = (r - y)x$, the volatility is σx, the discount rate is the constant r, $c = 0$ and $g = f$, and $\psi(x) = x^p + x^{-p} \geq 1$ satisfies the moment condition by Theorem 1.2(iv). The law of $S_T^{t,x}$ has a density, so $\mathbb{Q}(S_T^{t,x} \in E) = 0$. Equation (1.12) with the time shift gives $v(t, x) = e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}[f(S_T^{t,x})]$, which is (1.8), and the theorem started at time 0 gives the integrability of $f(S_T)$ and, by (1.13), $v(t, S_t) = V_t$. □

Remark 1.11 (The equation as a hedging argument). Rearranged, (1.15) reads $\partial_t v = r(v - x\,\partial_x v) + y\,x\,\partial_x v - \frac{1}{2}\sigma^2 x^2\,\partial_{xx} v$: for a position replicated by $\Delta = \partial_x v$ shares and $v - x\Delta$ in cash, the time decay equals the interest on the cash, plus the dividends on the shares, minus the gamma term, which pays for rebalancing in a moving market. This is the hedging derivation of Black, Scholes and Merton. Indeed, if v solves (1.15), Itô's formula and (1.7) give $e^{-rt}X_t$ and $e^{-rt}v(t, S_t)$ the same differential $e^{-rt}\sigma S_t\,\partial_x v\,\mathrm{d}W_t^{\mathbb{Q}}$ when $X_0 = v(0, S_0)$ and $\Delta_t = \partial_x v(t, S_t)$; so $X_t = v(t, S_t)$ for $t \leq T' < T$, whatever μ is.

▶ Interactive demo: price the bond of Theorem 1.9 by Monte Carlo and compare with the solution of the pricing equation, for the rate parameters, the starting rate and the number of paths and steps you choose.

1.4 Dirac deltas and Green's functions

A claim that pays 1 if S_T falls in $[\xi, \xi + h]$ costs about h times a number $G(t, S_t; T, \xi)$. Dividing by h and letting $h \downarrow 0$ produces a claim that pays "$\delta(S_T - \xi)$": infinitely much if $S_T = \xi$, nothing otherwise, with total weight one. This *Arrow–Debreu security* is not a function of S_T, but its price is a perfectly good function, and every other European price is a superposition of these. Two ideas make this precise.

Dirac deltas. For $a \in \mathbb{R}$, the *Dirac delta* δ_a maps a continuous function φ to $\varphi(a)$. It is the derivative of the step $H_a = \mathbf{1}_{(a,\infty)}$ and the second derivative of the ramp $R_a(x) = (x - a)^+$, in the following sense: for every $\varphi \in C^2(\mathbb{R})$ that vanishes outside a bounded interval,

$$-\int_{\mathbb{R}} H_a(x)\,\varphi'(x)\,\mathrm{d}x = \varphi(a), \qquad \int_{\mathbb{R}} R_a(x)\,\varphi''(x)\,\mathrm{d}x = \varphi(a), \tag{1.16}$$

by integration by parts (the boundary terms vanish because φ and φ' vanish outside a bounded interval). In the theory of distributions, the derivative of a generalised function L is *defined* by $L'(\varphi) = -L(\varphi')$, and (1.16) says that $H_a' = \delta_a$ and $R_a'' = \delta_a$ (Strichartz (2003), Chapters 1–2). The second view is as a limit: functions ϕ_ε *converge to* δ_a if $\int \phi_\varepsilon h \to h(a)$ for every bounded continuous h. The heat kernel is the standard example: $\int k(\varepsilon, x - a)h(x)\,\mathrm{d}x = \mathbb{E}[h(a + \sqrt{\varepsilon}Z)] \to h(a)$ as $\varepsilon \downarrow 0$, by dominated convergence. In probabilistic terms, $\mathcal{N}(a, \varepsilon)$ converges in law to the point mass at a. Below we test against functions of polynomial growth as well, which is what prices require.

The Green's function. By (1.4), given $S_t = x$ the variable $\log S_T$ is $\mathcal{N}(\log x + \nu\tau, \sigma^2\tau)$ under $\mathbb{Q}$, with $\tau = T - t$. So S_T has the density

$$p(\tau; x, \xi) = \frac{1}{\xi\sigma\sqrt{\tau}}\, N'\Big(\frac{\log(\xi/x) - \nu\tau}{\sigma\sqrt{\tau}}\Big) = \frac{1}{\xi}\, k(\sigma^2\tau,\ \log\xi - \log x - \nu\tau), \qquad x, \xi > 0, \tag{1.17}$$

and we define the *Green's function* of the pricing equation as the discounted density

$$G(t, x; T, \xi) = e^{-r(T-t)}\, p(T - t; x, \xi), \qquad 0 \le t < T,\ x, \xi > 0.$$

At $\xi = K$, the argument of N' in (1.17) is $-d_-$, so $G(t, x; T, K) = e^{-r\tau} N'(d_-)/(K\sigma\sqrt{\tau})$; in the test case $G(0, 100; 1, 100) = 0.019528$.

Proposition 1.12 (The Green's function of the pricing equation)**.** *Let $\sigma > 0$ and $r, y \in \mathbb{R}$.*

(i) Arrow–Debreu prices. *For every Borel $A \subseteq (0, \infty)$ and $t < T$, $e^{-r(T-t)}\mathbb{Q}(S_T \in A \mid \mathcal{F}_t) = \int_A G(t, S_t; T, \xi)\, \mathrm{d}\xi$: by Theorem 1.3, $G(t, S_t; T, \xi)\, \mathrm{d}\xi$ is the time-t price of the claim paying $\mathbf{1}_{[\xi, \xi + \mathrm{d}\xi]}(S_T)$.*

(ii) Backward equation, terminal value δ_ξ. *For fixed T and $\xi > 0$, the function $(t, x) \mapsto G(t, x; T, \xi)$ is infinitely differentiable on $[0, T) \times (0, \infty)$ and solves the pricing equation (1.15):*

$$\partial_t G + (r - y)\, x\, \partial_x G + \frac{\sigma^2 x^2}{2}\, \partial_{xx} G - rG = 0. \tag{1.18}$$

As $t \uparrow T$, $G(t, \cdot\,; T, \xi) \to \delta_\xi$: $\int_0^\infty G(t, x; T, \xi)\varphi(x)\, \mathrm{d}x \to \varphi(\xi)$ for every Borel φ that is continuous at ξ and satisfies $|\varphi(x)| \le a(x^q + x^{-q})$ for constants $a, q \ge 0$.

(iii) Forward equation, initial value δ_x. *For fixed t and $x > 0$, the function $(T, \xi) \mapsto G(t, x; T, \xi)$ is infinitely differentiable on $(t, \infty) \times (0, \infty)$ and solves the adjoint equation*

$$\partial_T G = \frac{1}{2}\, \partial_{\xi\xi}(\sigma^2\xi^2 G) - \partial_\xi((r - y)\, \xi\, G) - rG. \tag{1.19}$$

As $T \downarrow t$, $G(t, x; T, \cdot) \to \delta_x$ in the same sense as in (ii).

In neither limit does G converge to a function: as $t \uparrow T$, $G(t, x; T, \xi) \to 0$ for $x \ne \xi$, while $G(t, \xi; T, \xi) \to \infty$.

So G solves the pricing equation with terminal value δ_ξ, and the forward equation with initial value δ_x, "value" meaning the weak limits of (ii) and (iii), which are in particular limits in the sense of distributions.

Proof. (i) is (1.9) with $f = \mathbf{1}_A$, written with the density (1.17).

(ii) Pass to $w = \log x$ and $\tau = T - t$. Since $x\partial_x = \partial_w$ and $x^2\partial_{xx} = \partial_{ww} - \partial_w$, the pricing operator becomes $\partial_t + \nu\, \partial_w + \frac{1}{2}\sigma^2\partial_{ww} - r$, with constant coefficients. Now $G = \xi^{-1}e^{-r\tau}k(\sigma^2\tau, z)$ with $z = \log\xi - w - \nu\tau$, a smooth function of (t, w) for $\tau > 0$. With k_s, k_z, k_{zz} the partial derivatives of $k(s, z)$ at $(\sigma^2\tau, z)$, the chain rule gives

$$\partial_t G = \frac{e^{-r\tau}}{\xi}(rk - \sigma^2 k_s + \nu k_z), \qquad \partial_w G = -\frac{e^{-r\tau}}{\xi} k_z, \qquad \partial_{ww} G = \frac{e^{-r\tau}}{\xi} k_{zz},$$

so the operator applied to G equals $\xi^{-1}e^{-r\tau}\sigma^2(\frac{1}{2}k_{zz} - k_s) = 0$, by the heat equation $k_s = \frac{1}{2}k_{zz}$ of Section 1.1. For the limit, substitute $x = e^w$: as a function of w, $k(\sigma^2\tau, \log\xi - w - \nu\tau)$ is the

density of $V = \log\xi - \nu\tau + \sigma\sqrt{\tau}Z$, so

$$\int_0^\infty G\varphi\,\mathrm{d}x = \frac{e^{-r\tau}}{\xi}\,\mathbb{E}\big[e^V\varphi(e^V)\big] = e^{-r\tau}\mathbb{E}\big[e^{-\nu\tau+\sigma\sqrt{\tau}Z}\varphi(\xi e^{-\nu\tau+\sigma\sqrt{\tau}Z})\big].$$

As $\tau\downarrow 0$ the variable inside converges to $\varphi(\xi)$ for every value of Z, by continuity of φ at ξ, and for $\tau\le T$ it is bounded by $a\,e^{(1+q)(|\nu|T+\sigma\sqrt{T}|Z|)}(\xi^q+\xi^{-q})$, which is integrable. Dominated convergence gives the limit $\varphi(\xi)$.

(iii) Let $F = \xi G = e^{-r\tau}k(\sigma^2\tau, z)$ with $z = \eta - \log x - \nu\tau$, $\eta = \log\xi$, and $\partial_\xi = \xi^{-1}\partial_\eta$. Then $\partial_\xi(\xi G) = \xi^{-1}\partial_\eta F$ and $\partial_{\xi\xi}(\xi^2 G) = \partial_\xi(F + \partial_\eta F) = \xi^{-1}(\partial_\eta F + \partial_{\eta\eta}F)$, so the right side of (1.19) equals $\xi^{-1}\big(\tfrac12\sigma^2\partial_{\eta\eta}F - \nu\,\partial_\eta F - rF\big) = \xi^{-1}e^{-r\tau}\big(\tfrac12\sigma^2 k_{zz} - \nu k_z - rk\big)$. The left side is $\xi^{-1}\partial_\tau F = \xi^{-1}e^{-r\tau}(\sigma^2 k_s - \nu k_z - rk)$, and the two agree because $k_s = \tfrac12 k_{zz}$. For the limit, $\int G(t,x;T,\xi)h(\xi)\,\mathrm{d}\xi = e^{-r\tau}\mathbb{E}\big[h(xe^{\nu\tau+\sigma\sqrt{\tau}Z})\big]$, and dominated convergence applies as in (ii).

For the last claim, $k(\sigma^2\tau, z)\to 0$ as $\tau\downarrow 0$ for $z\neq 0$, while at $x=\xi$ the argument is $z=-\nu\tau$ and $k(\sigma^2\tau,-\nu\tau) = e^{-\nu^2\tau/(2\sigma^2)}/(\sigma\sqrt{2\pi\tau})\to\infty$. □

Superposing Arrow–Debreu securities builds every payoff, and it gives the converse of the Feynman–Kac formula in the standard model.

Corollary 1.13 (Prices as superpositions)**.** *Let $f:(0,\infty)\to\mathbb{R}$ be Borel with $|f(\xi)|\le a(\xi^q+\xi^{-q})$ for constants $a, q\ge 0$, and*

$$v(t,x) = \int_0^\infty G(t,x;T,\xi)\,f(\xi)\,\mathrm{d}\xi, \qquad 0\le t<T,\ x>0. \tag{1.20}$$

Then v is the function (1.8), so $v(t,S_t)$ is the price of $f(S_T)$; v is infinitely differentiable on $[0,T)\times(0,\infty)$ and solves (1.15); $|v(t,x)|\le C(x^q+x^{-q})$ for a constant C; and $v(t,x)\to f(x_0)$ as $(t,x)\to(T,x_0)$ at every point x_0 where f is continuous. The points of discontinuity of f form an F_σ, hence Borel, set; if it is Lebesgue-null, v is the only solution of (1.15) with these properties.

Proof. The change of variables $\xi = xe^{\nu\tau+\sigma\sqrt{\tau}z}$ turns (1.20) into (1.8). For smoothness, write $\phi(\eta) = f(e^\eta)$, so that $|\phi(\eta)|\le 2ae^{q|\eta|}$, and $v(t,e^w) = e^{-r\tau}\int_{\mathbb{R}}\phi(\eta)\,k(\sigma^2\tau,\eta-w-\nu\tau)\,\mathrm{d}\eta$. Every partial derivative of $(\tau,w)\mapsto k(\sigma^2\tau,\eta-w-\nu\tau)$ is a polynomial in $\eta-w-\nu\tau$ and $\tau^{-1/2}$ times k itself, because $k_z = -\frac{z}{s}k$ and $k_s = \frac12 k_{zz}$. For τ in a compact subset $[\tau_1,\tau_2]$ of $(0,T]$ and $|w|\le R$, such a derivative is bounded by $c(1+|\eta|^m)e^{-\eta^2/(4\sigma^2\tau_2)}$ for constants c, m (use $(\eta-\beta)^2\ge\frac12\eta^2-\beta^2$ with $\beta = w+\nu\tau$), which is integrable against $e^{q|\eta|}$. By dominated convergence and the mean value theorem we may differentiate under the integral sign as often as we like. So v is infinitely differentiable, and applying the operator under the integral, Theorem 1.12(ii) gives (1.15). The growth bound follows from (1.5): $|v(t,x)|\le ae^{-r\tau}\big(x^q\mathbb{E}^{\mathbb{Q}}[R^q] + x^{-q}\mathbb{E}^{\mathbb{Q}}[R^{-q}]\big)$ with $R = e^{\nu\tau+\sigma\sqrt{\tau}Z}$, and $\mathbb{E}^{\mathbb{Q}}[R^{\pm q}] = e^{\pm q\nu\tau+q^2\sigma^2\tau/2}$ is bounded for $\tau\le T$. The terminal limit is the dominated-convergence argument of Theorem 1.12(iii), with x restricted to $[x_0/2, 2x_0]$ for the bound. Uniqueness is Theorem 1.10. □

Remark 1.14 (Calls, digitals and deltas). The payoffs $\mathbf{1}_{(K,\infty)}$ and $(\cdot-K)^+$ in (1.20) give the digital call $D(K) = \int_K^\infty G\,\mathrm{d}\xi$ and the call $C(K) = \int_K^\infty(\xi-K)\,G\,\mathrm{d}\xi$, at fixed (t,x) and T. Differentiating in the strike, $C'(K) = -D(K)$ and $D'(K) = -G(t,x;T,K)$: the Arrow–Debreu price density is

the second strike derivative of the call price, the Breeden–Litzenberger formula (Theorem 3.5). As $T - t \downarrow 0$, the three prices converge to the delta, the step and the ramp of (1.16), written in the strike variable (Figure 1.2). The digital payoff jumps at K, and its price, which is $e^{-r\tau}N(d_-)$ (Chapter 2), is smooth before T but cannot be continuous at (T, K): this is the set $E = \{K\}$ of Theorem 1.6. The Green's function itself is the extreme case. As a function of (t, x) it solves the pricing equation and tends to 0 as $(t, x) \to (T, x_0)$ for every $x_0 \neq \xi$, so it satisfies (1.11) with $g = 0$ and $E = \{\xi\}$; yet it is not zero. So neither (a) nor (b) of Theorem 1.6 holds for it; indeed $\sup_{t<T} G(t, x; T, \xi) \to \infty$ as $x \to \xi$.

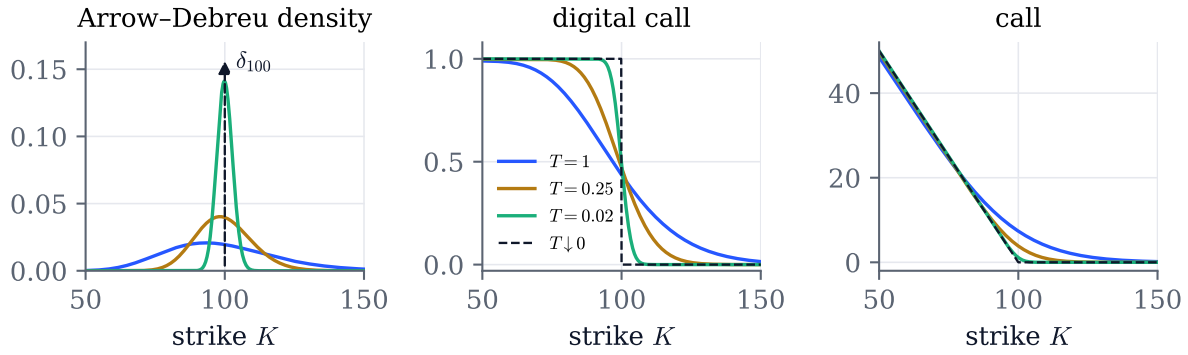

Figure 1.2: The Arrow–Debreu price density $G(0, 100; T, K)$, the digital call and the call, as functions of the strike K, in the standard model with $S_0 = 100$, $r = 1\%$, $y = 2\%$ and $\sigma = 20\%$, for maturities $T = 1$, 0.25 and 0.02. Each panel is minus the strike derivative of the next (Theorem 1.14). As $T \downarrow 0$ they approach the Dirac delta δ_{100} (arrow), the step $\mathbf{1}_{\{K<100\}}$ and the ramp $(100 - K)^+$ (dashed).

In code

`code/src/mfe230d/standard.py`: `lognormal_moment` is (1.5); `girsanov_price` reweights $\mathbb{P}$-samples by Z_T; `green` is G; `superposition_price` computes (1.20) by the trapezoid rule and reproduces the test-case call 7.36429; `backward_residual` and `forward_residual` evaluate (1.18)–(1.19) by central differences; `fk_vasicek_check` is the Monte Carlo side of Theorem 1.9, built on `code/src/mfe230d/paths.py` and `code/src/mfe230d/montecarlo.py`.

▶ Interactive demo: shrink the maturity and watch the three prices become a delta, a step and a ramp; check by finite differences that G solves both equations.

Further reading

Shreve (2004), Chapters 4–6, develops Itô calculus, risk-neutral pricing (with dividends in Section 5.5) and the Feynman–Kac formula at the level of these notes. Karatzas and Shreve (1991), Sections 3.5, 4.3 and 5.7, and Øksendal (2003), Chapter 8, give the measure-theoretic details, including the converse of the Feynman–Kac formula. The formula goes back to Feynman (1948) and Kac (1949); the pricing equation and its hedging derivation to Black and Scholes (1973) and Merton (1973), and the martingale view of pricing to Harrison and Pliska (1981).

Chapter 2

The Black–Scholes formula three ways

Every method in these notes is tested against the Black–Scholes formula. This chapter derives it three times, with the dividend yield of the standard model (Theorem 1.1), each time with a method that outlives the formula. The first integrates the payoff against the lognormal law of S_T, the computation that Monte Carlo replaces when there is no closed form (Part II). The second solves the pricing equation: a change of variables gives the heat equation, and a Fourier transform in the log-price an ordinary differential equation; with another characteristic function the same transform prices options in other models (Chapter 19). The third takes a Laplace transform in time and inverts it numerically. On the way we meet two hypotheses that are easy to forget: the call payoff has no Fourier transform without damping, and the pricing equation has many solutions unless we fix a growth class. The chapter ends with the Greeks and the scaling relations between them, and with the implied volatility: Newton's method on the price converges monotonically from the right starting point, and on the logarithm of the price, for out-of-the-money options and with a halving safeguard, from any starting point. The code is `code/src/mfe230d/transforms.py` and `code/src/mfe230d/impliedvol.py`; the closed forms are in `code/src/mfe230d/bs.py`.

Setting and notation. We work in the standard model of Theorem 1.1: under the risk-neutral measure $\mathbb{Q}$ the stock solves $\mathrm{d}S_t = (r-y)S_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}W_t^{\mathbb{Q}}$, with constant rate r, yield y and volatility $\sigma > 0$, and $\nu = r - y - \frac{\sigma^2}{2}$. For a valuation time $t < T$ we write $\tau = T - t$ for the time to maturity, and for a spot $x > 0$ and a strike $K > 0$

$$F = xe^{(r-y)\tau}, \qquad w = \sigma\sqrt{\tau}, \qquad m = \log\frac{F}{K}, \qquad d_\pm = \frac{\log(x/K) + (r - y \pm \frac{\sigma^2}{2})\tau}{\sigma\sqrt{\tau}} = \frac{m}{w} \pm \frac{w}{2}\,: \tag{2.1}$$

the forward price, the *total volatility*, the log-moneyness and the two arguments of the formula. N and N' are the standard normal distribution function and density. The course's test case is $S = K = 100$, $T = 1$, $r = 1\%$, $y = 2\%$, $\sigma = 20\%$, for which $F = 99.0050$, $m = -0.01$, $d_+ = 0.05$ and $d_- = -0.15$.

2.1 The formula by direct integration

By Theorem 1.3, a claim paying $f(S_T)$, with f Borel and $\mathbb{E}^{\mathbb{Q}}|f(S_T)| < \infty$, is worth $V_t = v(t, S_t)$ at time $t < T$, where

$$v(t,x) = e^{-r\tau} \int_{\mathbb{R}} f\big(xe^{\nu\tau + wz}\big)\, N'(z)\, \mathrm{d}z. \tag{2.2}$$

The whole computation rests on completing a square: for real a and z,

$$e^{az - a^2/2}\, N'(z) = N'(z-a), \qquad \text{hence} \qquad \int_b^{\infty} e^{az-a^2/2} N'(z)\, \mathrm{d}z = N(a-b) \quad (b \in [-\infty, \infty)). \tag{2.3}$$

Theorem 2.1 (The Black–Scholes formula)**.** *In the standard model, let $K > 0$, $0 \le t < T$ and $\tau = T - t$. The time-t prices of the European call and put with strike K and maturity T are $C_t = c(t, S_t)$ and $P_t = p(t, S_t)$, where, with $d_\pm = d_\pm(\tau, x)$ as in (2.1),*

$$c(t,x) = xe^{-y\tau} N(d_+) - Ke^{-r\tau} N(d_-), \qquad p(t,x) = Ke^{-r\tau} N(-d_-) - xe^{-y\tau} N(-d_+). \tag{2.4}$$

Both are $\phi\,[xe^{-y\tau} N(\phi d_+) - Ke^{-r\tau} N(\phi d_-)]$, with $\phi = 1$ for the call and $\phi = -1$ for the put. The cash-or-nothing *call and put, paying $\mathbf{1}_{\{S_T > K\}}$ and $\mathbf{1}_{\{S_T \le K\}}$, are worth $e^{-r\tau} N(d_-)$ and $e^{-r\tau} N(-d_-)$; the* asset-or-nothing *call and put, paying $S_T \mathbf{1}_{\{S_T > K\}}$ and $S_T \mathbf{1}_{\{S_T \le K\}}$, are worth $xe^{-y\tau} N(d_+)$ and $xe^{-y\tau} N(-d_+)$. In particular $c(t,x) - p(t,x) = xe^{-y\tau} - Ke^{-r\tau}$.*

Proof. All four binary payoffs are bounded by $1 + S_T$, which is $\mathbb{Q}$-integrable, so (2.2) applies to each, and the call and put are the differences $S_T \mathbf{1}_{\{S_T > K\}} - K \mathbf{1}_{\{S_T > K\}}$ and $K \mathbf{1}_{\{S_T \le K\}} - S_T \mathbf{1}_{\{S_T \le K\}}$. Fix $x > 0$. Since $w > 0$, $xe^{\nu\tau + wz} > K$ if and only if $z > \big(\log(K/x) - \nu\tau\big)/w = -d_-$. For the cash-or-nothing call, (2.2) gives $e^{-r\tau} \int_{-d_-}^{\infty} N'(z)\, \mathrm{d}z = e^{-r\tau} N(d_-)$. For the asset-or-nothing call, write $\nu\tau = (r - y)\tau - \frac{w^2}{2}$ and use (2.3) with $a = w$:

$$e^{-r\tau} \int_{-d_-}^{\infty} xe^{\nu\tau + wz} N'(z)\, \mathrm{d}z = xe^{-y\tau} \int_{-d_-}^{\infty} e^{wz - w^2/2} N'(z)\, \mathrm{d}z = xe^{-y\tau} N(d_- + w) = xe^{-y\tau} N(d_+).$$

The puts integrate over $z \le -d_-$ instead, which gives $e^{-r\tau} N(-d_-)$ and, since $\int_{-\infty}^{-d_-} N'(z - w)\, \mathrm{d}z = N(-d_- - w)$, the value $xe^{-y\tau} N(-d_+)$. Subtracting gives (2.4), and parity follows from $N(u) + N(-u) = 1$. □

In the test case the call is worth 7.36429 and the put 8.34941; the cash-or-nothing call is worth 0.436000 and the asset-or-nothing call 50.9643, and indeed $50.9643 - 100 \times 0.436000 = 7.3643$. Here $N(d_-) = \mathbb{Q}(S_T > K \mid S_t = x)$, and (2.3) is a change of measure to the share numeraire (Chapter 17), under which $N(d_+)$ is the probability of exercise. The prices depend on x, r and y only through F and $e^{-r\tau}$: $c = e^{-r\tau}[FN(d_+) - KN(d_-)]$ (Black's form); when $r = y$, $F = x$.

Remark 2.2 (Degenerate cases and limits). Fix $K > 0$; each limit follows from the behaviour of $d_\pm = m/w \pm w/2$. (a) $\sigma \downarrow 0$: $d_\pm \to +\infty$, $-\infty$ or 0 as $F > K$, $F < K$ or $F = K$, so $c \to e^{-r\tau}(F - K)^+ = (xe^{-y\tau} - Ke^{-r\tau})^+$, the discounted intrinsic value *at the forward* (the riskless price); at $F = K$, $0 \le c = e^{-r\tau} F(2N(w/2) - 1) \le e^{-r\tau} Fw/\sqrt{2\pi}$. (b) $\sigma \to \infty$: $c \to xe^{-y\tau}$ and $p \to Ke^{-r\tau}$, the upper no-arbitrage bounds (Theorem 3.1). (c) $c(t,x) \to (x_0 - K)^+$ as $(t,x) \to (T, x_0)$, for every $x_0 > 0$: for $x_0 \ne K$, $d_\pm \to \pm\infty$ with the sign of $\log(x_0/K)$; for $x_0 = K$, since $0 \le N(d_+) - N(d_-) \le w/\sqrt{2\pi}$, $|c| \le |xe^{-y\tau} - Ke^{-r\tau}| + Ke^{-r\tau}\sigma\sqrt{\tau}/\sqrt{2\pi} \to 0$. The same

holds for p. (d) $K \downarrow 0$: $c \to xe^{-y\tau}$, the price of the claim to S_T, which does not depend on σ. (e) $x \downarrow 0$: $c \to 0$ and $p \to Ke^{-r\tau}$; as $x \to \infty$, $c - (xe^{-y\tau} - Ke^{-r\tau}) = p \to 0$.

2.2 The pricing equation and the heat equation

By the Feynman–Kac formula (Theorem 1.6), a smooth solution of the pricing equation with polynomial growth is the risk-neutral price. Here we forget the probability and solve the equation. For a strike $K > 0$ the problem is to find $V \in C^{1,2}([0,T) \times (0,\infty))$ with

$$\partial_t V + (r-y)\, x\, \partial_x V + \frac{\sigma^2 x^2}{2}\, \partial_{xx} V - rV = 0 \qquad \text{on } [0,T) \times (0,\infty), \tag{2.5}$$

$$V(t,x) \to (x_0 - K)^+ \qquad \text{as } (t,x) \to (T,x_0),\ t < T, \text{ for every } x_0 > 0. \tag{2.6}$$

The coefficients grow with x; in the log-price they are constant, and in a frame that moves with the drift only the diffusion remains.

Lemma 2.3 (Reduction to the heat equation)**.** *For $V : [0,T) \times (0,\infty) \to \mathbb{R}$ define $u : (0, \sigma^2 T] \times \mathbb{R} \to \mathbb{R}$ by*

$$V(t,x) = Ke^{-r\tau}\, u(s,\xi), \qquad s = \sigma^2 \tau, \qquad \xi = \log\frac{x}{K} + \nu\tau, \qquad \tau = T - t. \tag{2.7}$$

The map $(t,x) \mapsto (s,\xi)$ is a bijection from $[0,T) \times (0,\infty)$ onto $(0,\sigma^2 T] \times \mathbb{R}$, infinitely differentiable in both directions, and it extends to a homeomorphism from $[0,T] \times (0,\infty)$ onto $[0,\sigma^2 T] \times \mathbb{R}$ that maps (T,x_0) to $(0, \log(x_0/K))$.

(i) $V \in C^{1,2}([0,T) \times (0,\infty))$ if and only if $u \in C^{1,2}((0,\sigma^2 T] \times \mathbb{R})$, and then

$$\partial_t V + (r-y)x\, \partial_x V + \frac{\sigma^2 x^2}{2}\partial_{xx} V - rV = \sigma^2 K e^{-r\tau}\Big(\frac{1}{2}\partial_{\xi\xi} u - \partial_s u\Big)(s,\xi).$$

So V solves (2.5) if and only if u solves the heat equation *$\partial_s u = \frac{1}{2}\partial_{\xi\xi} u$ on $(0,\sigma^2 T] \times \mathbb{R}$.*

(ii) V satisfies (2.6) if and only if u extends continuously to $[0,\sigma^2 T] \times \mathbb{R}$ with $u(0,\xi) = u_0(\xi) := (e^\xi - 1)^+$.

(iii) $|V(t,x)| \le A \exp\big(a(\log x)^2\big)$ for some constants A, a if and only if $|u(s,\xi)| \le A' \exp(a'\xi^2)$ for some constants A', a'.

Proof. The inverse map is $\tau = s/\sigma^2$, $x = K\exp(\xi - \nu s/\sigma^2)$. (i) By the chain rule, with $\partial_t \tau = -1$, $x\partial_x \xi = 1$ and $x^2\partial_{xx} = (x\partial_x)^2 - x\partial_x$,

$$\partial_t V = Ke^{-r\tau}(ru - \sigma^2 \partial_s u - \nu\, \partial_\xi u), \qquad x\, \partial_x V = Ke^{-r\tau}\partial_\xi u, \qquad x^2\partial_{xx} V = Ke^{-r\tau}(\partial_{\xi\xi} u - \partial_\xi u),$$

and $C^{1,2}$ regularity transfers in both directions. Substituting, the terms in u cancel and the coefficient of $\partial_\xi u$ is $-\nu + (r-y) - \frac{\sigma^2}{2} = 0$. (ii) Along the homeomorphism, $(t,x) \to (T,x_0)$ exactly when $(s,\xi) \to (0, \log(x_0/K))$, $e^{-r\tau} \to 1$, and $(x_0 - K)^+ = K(e^{\xi_0} - 1)^+$. (iii) $\log x = \xi - \nu\tau + \log K$ with $|\nu\tau - \log K| \le b := |\nu| T + |\log K|$, $(\xi \pm b)^2 \le 2\xi^2 + 2b^2$, and $e^{-|r|T} \le e^{-r\tau} \le e^{|r|T}$. □

Without a growth condition the heat equation does not determine its solutions: Tychonoff (1935) constructed a nonzero solution, continuous up to $s = 0$ and vanishing there, that grows faster

than $e^{a\xi^2}$ for every a (see also John (1982), Chapter 7); through (2.7) it gives nonzero solutions of (2.5) with zero terminal value. The class $e^{a\xi^2}$ excludes it.

Theorem 2.4 (Uniqueness for the heat equation). *Let $S > 0$ and let $u \in C^{1,2}((0,S]\times\mathbb{R}) \cap C([0,S]\times\mathbb{R})$ solve $\partial_s u = \frac{1}{2}\partial_{\xi\xi} u$ on $(0,S]\times\mathbb{R}$ with $u(0,\cdot) = 0$. If $|u(s,\xi)| \le Ae^{a\xi^2}$ on $[0,S]\times\mathbb{R}$ for constants $A, a \ge 0$, then $u = 0$.*

This is Evans (2010), Section 2.3.3, Theorem 7, for the equation $\partial_t \tilde{u} = \partial_{\xi\xi}\tilde{u}$, applied to $\tilde{u}(t,\xi) = u(2t,\xi)$ on $[0, S/2]$. In the price variable the class is $|V(t,x)| \le A\exp(a(\log x)^2)$ (Theorem 2.3(iii)); it contains every function bounded by $C(x^p + x^{-p})$, since $e^{p|\log x|} \le e^{p^2/(4a)} e^{a(\log x)^2}$, so it is larger than the polynomial class in which Theorem 1.6 gives uniqueness.

2.3 Solving the equation by Fourier transform

For an integrable $g : \mathbb{R} \to \mathbb{C}$ and real ω let $\hat{g}(\omega) = \int_{\mathbb{R}} e^{-i\omega\xi} g(\xi)\,\mathrm{d}\xi$. The transform turns ∂_ξ into multiplication by $i\omega$, so the heat equation becomes $\partial_s \hat{u} = -\frac{1}{2}\omega^2 \hat{u}$, solved by $\hat{u}(s,\omega) = \hat{u}_0(\omega) e^{-s\omega^2/2}$. The obstacle is the payoff. The function $u_0(\xi) = (e^\xi - 1)^+$ grows like e^ξ: it is not integrable, and it is not a tempered distribution either: since $u_0 \ge 0$, a tempered extension T would give, for $\varphi = e^{-\sqrt{1+\xi^2}/2}$ and a smooth cutoff $0 \le \chi \le 1$ equal to 1 near 0 and nonincreasing in $|\xi|$, $T(\varphi) = \lim_n T(\chi(\cdot/n)\varphi) = \lim_n \int u_0 \chi(\cdot/n)\varphi = \infty$ by monotone convergence. So it has no Fourier transform, even in the sense of distributions. The remedy is *damping*: for $a > 1$ the function $e^{-a\xi} u_0(\xi)$ is integrable, and its transform is the *generalised* transform of u_0 at complex frequencies,

$$\int_{\mathbb{R}} e^{-i\omega\xi} e^{-a\xi} u_0(\xi)\,\mathrm{d}\xi = \hat{u}_0(\omega - ia), \qquad \hat{u}_0(z) := \int_0^\infty e^{-iz\xi}(e^\xi - 1)\,\mathrm{d}\xi = \frac{1}{iz-1} - \frac{1}{iz} = -\frac{1}{z(z+i)}, \tag{2.8}$$

valid for $\operatorname{Im} z < -1$, where both integrals converge absolutely ($|e^{-iz\xi}| = e^{\xi \operatorname{Im} z}$). The heat kernel $k(s,\xi) = (2\pi s)^{-1/2} e^{-\xi^2/(2s)}$, the density of $\sqrt{s}Z$ with $Z \sim \mathcal{N}(0,1)$, has the transform

$$\int_{\mathbb{R}} e^{-iz\xi} k(s,\xi)\,\mathrm{d}\xi = \mathbb{E}\big[e^{-iz\sqrt{s}Z}\big] = e^{-sz^2/2} \qquad \text{for every } z \in \mathbb{C}, \tag{2.9}$$

because $\mathbb{E}[e^{\theta Z}] = e^{\theta^2/2}$ holds for real θ, and both sides are entire functions of θ (differentiate under the integral), so they agree on $\mathbb{C}$.

Proposition 2.5 (The formula by solving the pricing equation). *Let $K, \sigma > 0$ and $r, y \in \mathbb{R}$. There is exactly one $V \in C^{1,2}([0,T)\times(0,\infty))$ that solves (2.5)–(2.6) and satisfies $|V(t,x)| \le A\exp(a(\log x)^2)$ for some constants A, a: the call price c of Theorem 2.1. For every damping $a > 1$, every $t < T$ and every $x > 0$, with $s = \sigma^2\tau$ and $\xi = \log(x/K) + \nu\tau$,*

$$c(t,x) = \frac{Ke^{-r\tau}e^{a\xi}}{\pi} \int_0^\infty \operatorname{Re}\left[e^{i\omega\xi}\,\hat{u}_0(\omega - ia)\,e^{-s(\omega-ia)^2/2}\right]\mathrm{d}\omega. \tag{2.10}$$

Proof. By Theorem 2.3, with $S = \sigma^2 T$, we must show that there is exactly one $u \in C^{1,2}((0,S]\times\mathbb{R}) \cap C([0,S]\times\mathbb{R})$ with $\partial_s u = \frac{1}{2}\partial_{\xi\xi} u$, $u(0,\cdot) = u_0$ and $|u| \le A' e^{a'\xi^2}$, and identify it.

Uniqueness. The difference of two such solutions satisfies the hypotheses of Theorem 2.4 (the sum of two bounds $A_i e^{a_i\xi^2}$ is a bound of the same form), so it vanishes.

Existence, by Fourier transform. Fix $a > 1$ and, for $s > 0$, let

$$u(s,\xi) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{iz\xi}\,\hat{u}_0(z)\,e^{-sz^2/2}\,\mathrm{d}\omega, \qquad z = \omega - ia,$$

which is (2.10) written over $\mathbb{R}$ (the integrand at $-\omega$ is the conjugate of that at ω). Here $|e^{iz\xi}| = e^{a\xi}$, $|e^{-sz^2/2}| = e^{-s(\omega^2-a^2)/2}$ and $|\hat{u}_0(z)| \le 1/(a(a-1))$. A derivative of order j in ξ and l in s multiplies the integrand by $(iz)^j(-z^2/2)^l$, which for $s \ge s_0 > 0$ and $|\xi| \le R$ is bounded by the integrable $C(1+|\omega|)^{j+2l}e^{aR+Sa^2/2}e^{-s_0\omega^2/2}$. So we may differentiate under the integral: u is infinitely differentiable on $(0,S]\times\mathbb{R}$ and, since $-\frac{1}{2}z^2 = \frac{1}{2}(iz)^2$, solves the heat equation.

The solution is a convolution. The functions $f_1 = e^{-a\xi}k(s,\xi)$ and $f_2 = e^{-a\xi}u_0(\xi)$ are integrable, f_1 is bounded and continuous, so $f_1 * f_2$ is continuous, and $(f_1*f_2)(\xi) = e^{-a\xi}\int_{\mathbb{R}} k(s,\xi-\eta)u_0(\eta)\,\mathrm{d}\eta$. By (2.8), (2.9) and the convolution theorem (Rudin (1987), Theorem 9.2(c)), its transform is the integrable function $e^{-sz^2/2}\hat{u}_0(z)$; by the inversion theorem (Rudin (1987), Theorem 9.11) it is the inverse transform of that function at every point. Multiplying by $e^{a\xi}$,

$$u(s,\xi) = \int_{\mathbb{R}} k(s,\xi-\eta)\,u_0(\eta)\,\mathrm{d}\eta = \mathbb{E}\big[u_0(\xi+\sqrt{s}Z)\big]. \tag{2.11}$$

Initial value and growth. For $|\xi| \le R$ and $s \le S$, $0 \le u_0(\xi+\sqrt{s}Z) \le e^{R+\sqrt{S}|Z|}$, which is integrable; u_0 is continuous, so dominated convergence gives $u(s,\xi) \to u_0(\xi_0)$ as $(s,\xi) \to (0,\xi_0)$. Setting $u(0,\cdot) = u_0$ makes u continuous on $[0,S]\times\mathbb{R}$. Moreover $0 \le u(s,\xi) \le \mathbb{E}[e^{\xi+\sqrt{s}Z}] = e^{\xi+s/2} \le e^{S/2+1/4}e^{\xi^2}$.

Evaluation. The integrand $(e^{\xi+\sqrt{s}z}-1)^+$ is positive exactly when $z > -\xi/\sqrt{s}$, and (2.3) with $a = \sqrt{s}$ gives $u(s,\xi) = e^{\xi+s/2}N\big((\xi+s)/\sqrt{s}\big) - N\big(\xi/\sqrt{s}\big)$. With $s = w^2$ and $\xi = \log(x/K) + \nu\tau$ we have $\xi/\sqrt{s} = d_-$, $(\xi+s)/\sqrt{s} = d_+$ and $Ke^{-r\tau}e^{\xi+s/2} = xe^{-y\tau}$, so $Ke^{-r\tau}u = c$, the formula (2.4). □

The proof did not use the stock or the measure $\mathbb{Q}$: the call price is the only solution of the pricing equation in the class $e^{a(\log x)^2}$, and the Fourier transform finds it. The convolution (2.11) is Theorem 1.12 in other variables. In (2.10) the model enters only through $\phi_Y(-z)$, where $\phi_Y(u) = \mathbb{E}[e^{iuY}]$ is the characteristic function of the log-price increment $Y = \sqrt{s}Z$ (here $\phi_Y(-z) = e^{-sz^2/2}$; it is finite at $z = \omega - ia$ when $\mathbb{E}[S_T^a] < \infty$). Where ϕ_Y is known but the density is not, as in Lévy and affine stochastic-volatility models, the same integral of the transformed payoff prices the call (Lewis, 2001); Carr and Madan (1999) damp the price in the log-strike instead (Theorem 19.7).

Computing the integral. Figure 2.1(a) evaluates (2.10) by the trapezoidal rule with step h. As a function of ω the integrand has poles only at $\omega = ia$ and $\omega = i(a-1)$ (where $z = 0$ and $z = -i$), so it is analytic in the strip $|\operatorname{Im}\omega| < d$ for every $d < a-1$, and the error of the rule on the whole line is at most $2M_d/(e^{2\pi d/h}-1)$, where M_d bounds the integrals along horizontal lines in the strip (Trefethen and Weideman (2014), Theorem 5.1). To leading order, the error is the contribution of the nearest pole, $z = -i$, whose residue has modulus $e^{-(a-1)\xi+s/2}$: $Ke^{-r\tau}e^{\xi+s/2}e^{-2\pi(a-1)/h} = xe^{-y\tau}e^{-2\pi(a-1)/h}$. We do not prove that this term dominates, but the errors for four dampings fall on it. A larger damping allows a larger step but inflates the integrand by $e^{a\xi+sa^2/2}$, and rounding errors with it. Truncation at Ω costs about $e^{-s\Omega^2/2}$, so the number of nodes grows like $s^{-1/2}$ as $s \downarrow 0$.

2.4 Solving the equation by Laplace transform

The course takes a Laplace transform in time instead, for all maturities at once: $\bar{u}(\lambda,\xi) = \int_0^\infty e^{-\lambda s}u(s,\xi)\,\mathrm{d}s$. Formally, $\partial_s u$ becomes $\lambda\bar{u} - u_0$, and the heat equation the ordinary differential equation $\frac{1}{2}\partial_{\xi\xi}\bar{u} - \lambda\bar{u} = -u_0$, with homogeneous solutions $e^{\pm q\xi}$, $q = \sqrt{2\lambda}$; growth at $\pm\infty$ selects one on each side of the kink. For the Green's function, $u_0 = \delta_0$: a bounded solution of $\frac{1}{2}g'' = \lambda g$ off 0 whose slope jumps by -2 at 0. The next proposition confirms these formal computations.

Proposition 2.6 (The formula by Laplace transform)**.** *Let $q = \sqrt{2\lambda}$, with the principal square root.*

(i) For $\lambda > 0$ and $\xi \in \mathbb{R}$,

$$\int_0^\infty e^{-\lambda s}k(s,\xi)\,\mathrm{d}s = \frac{e^{-q|\xi|}}{q}. \tag{2.12}$$

As a function of ξ it is the only bounded continuous function that is C^2 off 0, solves $\frac{1}{2}g'' = \lambda g$ there, and has $g'(0+) - g'(0-) = -2$.

(ii) Let u be the heat solution (2.11) of the call, defined for all $s > 0$. For $\lambda > \frac{1}{2}$, $\bar{u}(\lambda,\xi) = \int_0^\infty e^{-\lambda s}u(s,\xi)\,\mathrm{d}s$ is finite and

$$\bar{u}(\lambda,\xi) = \begin{cases} \dfrac{e^{q\xi}}{q^2(q-1)}, & \xi \le 0, \\ \dfrac{e^{\xi}}{\lambda - \frac{1}{2}} - \dfrac{1}{\lambda} + \dfrac{e^{-q\xi}}{q^2(q+1)}, & \xi \ge 0. \end{cases} \tag{2.13}$$

It is the only C^1 function with $|\bar{u}| \le A(1+e^{\xi})$ that is C^2 off 0 and solves $\frac{1}{2}\bar{u}'' - \lambda\bar{u} = -u_0$ there.

(iii) The right side of (2.13) is analytic in λ on $\mathbb{C} \setminus ((-\infty,0] \cup \{\frac{1}{2}\})$ and equals the Laplace transform for $\operatorname{Re}\lambda > \frac{1}{2}$. For every $c > \frac{1}{2}$, $s > 0$ and ξ,

$$u(s,\xi) = \lim_{R\to\infty} \frac{1}{2\pi i}\int_{c-iR}^{c+iR} e^{\lambda s}\,\bar{u}(\lambda,\xi)\,\mathrm{d}\lambda. \tag{2.14}$$

Proof. (i) Write $b = |\xi|/\sqrt{2}$, so that $\lambda s + \xi^2/(2s) = (\sqrt{\lambda}s - b/\sqrt{s})^2 + 2b\sqrt{\lambda}$ and, with $s = v^2$,

$$\int_0^\infty e^{-\lambda s}k(s,\xi)\,\mathrm{d}s = \frac{e^{-2b\sqrt{\lambda}}}{\sqrt{2\pi}}\int_0^\infty 2e^{-(\sqrt{\lambda}v - b/v)^2}\,\mathrm{d}v.$$

Call the last integral $2J$. If $b = 0$, $J = \sqrt{\pi}/(2\sqrt{\lambda})$. If $b > 0$, the substitution $v = b/(\sqrt{\lambda}u)$ maps $\sqrt{\lambda}v - b/v$ to $-(\sqrt{\lambda}u - b/u)$, so $J = \int_0^\infty e^{-(\sqrt{\lambda}u - b/u)^2}\frac{b}{\sqrt{\lambda}u^2}\,\mathrm{d}u$; averaging the two expressions, $2J = \lambda^{-1/2}\int_0^\infty e^{-(\sqrt{\lambda}u - b/u)^2}\,\mathrm{d}(\sqrt{\lambda}u - b/u) = \sqrt{\pi/\lambda}$, since $\sqrt{\lambda}u - b/u$ increases from $-\infty$ to ∞. Either way the integral equals $e^{-2b\sqrt{\lambda}}(2\pi)^{-1/2}\sqrt{\pi/\lambda} = e^{-q|\xi|}/q$. This g has $\frac{1}{2}g'' = \frac{q^2}{2}g = \lambda g$ off 0 and $g'(0\pm) = \mp 1$. If g_1, g_2 both have the stated properties, $D = g_1 - g_2$ is C^1 (the jumps cancel), solves $\frac{1}{2}D'' = \lambda D$ on both sides, hence, being determined by $D(0)$ and $D'(0)$, is $\alpha e^{q\xi} + \beta e^{-q\xi}$ on $\mathbb{R}$; boundedness forces $\alpha = \beta = 0$.

(ii) By Tonelli's theorem (everything is nonnegative) and (i), $\bar{u}(\lambda,\xi) = \int_{\mathbb{R}}\frac{e^{-q|\xi-\eta|}}{q}(e^{\eta} - 1)^+\,\mathrm{d}\eta$, which is finite exactly when $q > 1$, that is $\lambda > \frac{1}{2}$. For $\xi \le 0$ only $\eta > 0 \ge \xi$ contributes:

$\frac{e^{q\xi}}{q}(\frac{1}{q-1}-\frac{1}{q})$. For $\xi \geq 0$, splitting at $\eta = \xi$ gives $e^{\xi}(\frac{1}{q(q-1)}+\frac{1}{q(q+1)})-\frac{2}{q^2}+e^{-q\xi}(\frac{1}{q^2}-\frac{1}{q(q+1)})$, which is (2.13) since $\frac{2}{q^2-1} = \frac{1}{\lambda-1/2}$. Both branches have value $\frac{1}{q^2(q-1)}$ and slope $\frac{1}{q(q-1)}$ at 0, and direct differentiation gives the equation. Uniqueness is proved as in (i): since $q > 1$, $\alpha e^{q\xi}+\beta e^{-q\xi} = O(1+e^{\xi})$ forces $\alpha = \beta = 0$.

(iii) With the principal root, q is analytic off $(-\infty, 0]$ with $\operatorname{Re} q > 0$; the right side of (2.13) has singularities only where $q \in \{0, 1\}$ or $\lambda \in \{0, \frac{1}{2}\}$. The Laplace transform is analytic on $\operatorname{Re}\lambda > \frac{1}{2}$ (it converges absolutely there, since $0 \leq u \leq e^{\xi+s/2}$) and agrees with the right side on $(\frac{1}{2}, \infty)$, hence on the whole half-plane. For (2.14), apply the Fourier inversion theorem for functions that are integrable and of bounded variation near the point (DLMF, Eq. (1.14.3)) to $f(\varsigma) = e^{-c\varsigma}u(\varsigma, \xi)\mathbf{1}_{\{\varsigma>0\}}$, which is integrable because $c > \frac{1}{2}$ and continuously differentiable near $\varsigma = s$. Its transform at ω is $\bar{u}(c+i\omega, \xi)$; inverting at $\varsigma = s$ and substituting $\lambda = c+i\omega$ gives (2.14). □

The course's variables. In $R = x/K$ and $\varsigma = \sigma^2\tau$ the course's transformed equation for the Green's function is of Euler type, with solutions $R^{\lambda_\pm}$, $\lambda_\pm = -\alpha \pm \sqrt{\alpha^2+2(\eta+\zeta)}$, $\alpha = \frac{r-y}{\sigma^2}-\frac{1}{2}$, $\eta = \frac{r}{\sigma^2}$. By Theorem 2.3 the Green's function is $e^{-\alpha z}e^{-(\eta+\alpha^2/2)\varsigma}k(\varsigma, z)$, $z = \log R$, whose transform $e^{-\alpha z-q|z|}/q$ at $\lambda = \zeta+\eta+\frac{\alpha^2}{2}$ is $R^{\lambda_\mp}/q$ for $R \gtrless 1$. Its Bromwich inversion, collapsing the line onto the cut, is Exercise 2.5.

Numerical inversion. Deforming the Bromwich line into a contour that wraps around the negative real axis, where $e^{\lambda s}$ is small, needs all singularities to its left; because of the pole at $\lambda = \frac{1}{2}$ we invert $\lambda \mapsto \bar{u}(\lambda+\frac{1}{2}, \xi)$, whose singularities lie on $(-\infty, 0]$, and multiply by $e^{s/2}$. The *fixed Talbot* rule of Abate and Valkó (2004) discretises Talbot's contour $\lambda(\theta) = \rho\theta(\cot\theta + i)$, $-\pi < \theta < \pi$ (Talbot, 1979), with M nodes and $\rho = 2M/(5s)$:

$$f(s) \approx \frac{\rho}{M}\Big[\frac{1}{2}\bar{f}(\rho)e^{\rho s}+\sum_{k=1}^{M-1}\operatorname{Re}\Big(e^{s\lambda_k}\bar{f}(\lambda_k)(1+i\beta_k)\Big)\Big], \qquad \lambda_k = \lambda(\theta_k),\ \theta_k = \frac{k\pi}{M}, \tag{2.15}$$

with $\beta_k = \theta_k + (\theta_k\cot\theta_k - 1)\cot\theta_k$. We do not prove the deformation or the error bound: Abate and Valkó (2004) report about $0.6M$ correct digits in exact arithmetic, and in double precision the factor $e^{\rho s} = e^{0.4M}$ amplifies rounding errors, so the error in Figure 2.1(b) falls until $M \approx 20$ and then grows. At the test case, $M = 20$ gives an error of 1.5×10^{-13}, and the Fourier rule with $a = 1.5$, $h = 0.05$ an error of 3×10^{-15}.

In code

`code/src/mfe230d/transforms.py`: `heat_variables` is (2.7); `fourier_call` evaluates (2.10) by the trapezoidal rule; `laplace_heat_call` is (2.13), `talbot` is (2.15), and `laplace_call` combines them.

▶ Interactive demo: Fourier and Laplace prices and their errors, for any inputs and node counts.

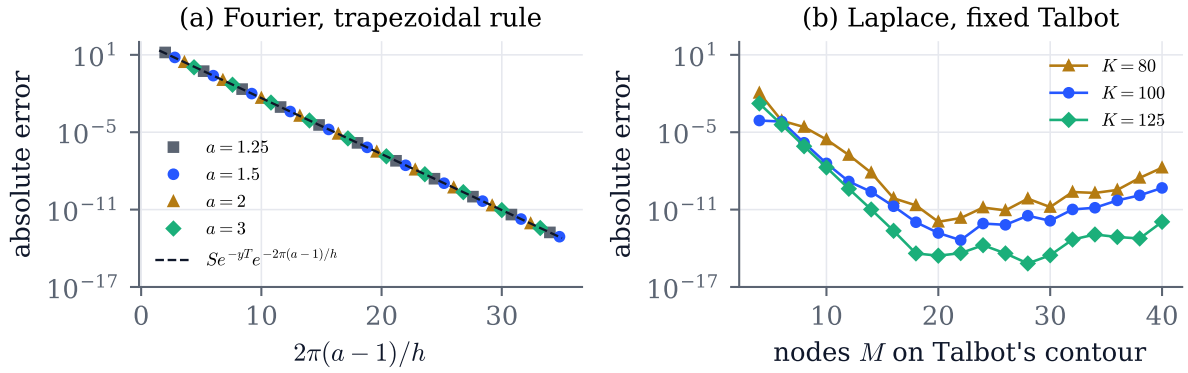

Figure 2.1: The test-case call by numerical inversion. (a) Fourier: absolute error of the trapezoidal rule for (2.10), with dampings $a = 1.25, 1.5, 2, 3$ and steps h chosen so that $2\pi(a-1)/h$ lies between 2 and 34; all four fall on the nearest-pole estimate $Se^{-yT}e^{-2\pi(a-1)/h}$ (dashed). (b) Laplace: absolute error of the fixed Talbot rule (2.15) applied to (2.13), against the number of nodes $M = 4, 6, \ldots, 40$, for $K = 80$, 100 and 125.

2.5 The Greeks and the relations between them

We use the conventions of `code/src/mfe230d/bs.py`. For a price $V(t,x)$ that also depends on K, r, y and σ: the delta $\Delta = \partial_x V$, the gamma $\Gamma = \partial_{xx} V$, the vega $\mathcal{V} = \partial_\sigma V$, the theta $\Theta = \partial_t V$ in calendar time (per year; at fixed t it equals $-\partial_T V$), and the rhos $\rho_r = \partial_r V$ and $\rho_y = \partial_y V$.

Proposition 2.7 (The Greeks of calls and puts)**.** *Let $K > 0$, $t < T$, $x > 0$ and $d_\pm = d_\pm(\tau, x)$.*

(i) $xe^{-y\tau}N'(d_+) = Ke^{-r\tau}N'(d_-)$.

(ii) The call c of (2.4) has

$$\Delta = e^{-y\tau}N(d_+), \qquad \Gamma = \frac{e^{-y\tau}N'(d_+)}{x\sigma\sqrt{\tau}}, \qquad \mathcal{V} = xe^{-y\tau}N'(d_+)\sqrt{\tau},$$
$$\Theta = -\frac{xe^{-y\tau}N'(d_+)\sigma}{2\sqrt{\tau}} + y\,xe^{-y\tau}N(d_+) - rKe^{-r\tau}N(d_-),$$
$$\rho_r = \tau Ke^{-r\tau}N(d_-), \qquad \rho_y = -\tau xe^{-y\tau}N(d_+), \qquad \partial_K c = -e^{-r\tau}N(d_-).$$

The put p has the same Γ and $\mathcal{V}$, and, by parity, $\Delta_p = \Delta - e^{-y\tau}$, $\partial_K p = \partial_K c + e^{-r\tau}$, $\Theta_p = \Theta - yxe^{-y\tau} + rKe^{-r\tau}$, $\rho_{r,p} = \rho_r - \tau Ke^{-r\tau}$ and $\rho_{y,p} = \rho_y + \tau xe^{-y\tau}$.

(iii) For $V = c$ and for $V = p$, with the Greeks of (ii):

(a) (homogeneity) $V = x\Delta + K\,\partial_K V$;

(b) (carry) $\rho_y = -\tau x\Delta$ and $\rho_r = -\tau K\,\partial_K V$, hence $\rho_r + \rho_y = -\tau V$;

(c) (vega and gamma) $\mathcal{V} = \sigma\tau x^2\Gamma$;

(d) (time) $\Theta = -(r\rho_r + y\rho_y + \frac{1}{2}\sigma\mathcal{V})/\tau$;

(e) (the pricing equation) $\Theta + (r-y)x\Delta + \frac{1}{2}\sigma^2x^2\Gamma = rV$.

Proof. (i) $d_+^2 - d_-^2 = (d_+ - d_-)(d_+ + d_-) = w \cdot \frac{2m}{w} = 2m$, so $N'(d_-)/N'(d_+) = e^{(d_+^2-d_-^2)/2} = F/K = xe^{(r-y)\tau}/K$.

(ii) Let θ be any of x, K, t, r, y, σ. By the chain rule and (i),

$$\partial_\theta c = \partial_\theta(xe^{-y\tau})N(d_+) - \partial_\theta(Ke^{-r\tau})N(d_-) + xe^{-y\tau}N'(d_+)\,\partial_\theta(d_+ - d_-),$$

and $d_+ - d_- = \sigma\sqrt{\tau}$ depends only on σ and t; this gives each first-order Greek, with $\partial_t\tau = -1$ for the theta. The gamma is $e^{-y\tau}N'(d_+)\partial_x d_+$, $\partial_x d_+ = 1/(x\sigma\sqrt{\tau})$. The put Greeks follow by differentiating the parity $p = c - xe^{-y\tau} + Ke^{-r\tau}$.

(iii) For the call, each identity is a substitution of (ii): $x\Delta + K\partial_K c = xe^{-y\tau}N(d_+) - Ke^{-r\tau}N(d_-) = c$; $-\tau x\Delta = \rho_y$ and $-\tau K\partial_K c = \rho_r$, whose sum is $-\tau(x\Delta + K\partial_K c) = -\tau c$ by (a); $\sigma\tau x^2\Gamma = xe^{-y\tau}N'(d_+)\sqrt{\tau} = \mathcal{V}$; and $-(r\rho_r + y\rho_y)/\tau = -rKe^{-r\tau}N(d_-) + yxe^{-y\tau}N(d_+)$, to which $-\sigma\mathcal{V}/(2\tau)$ adds the first term of Θ. The put differs from the call by $\varphi = xe^{-y\tau} - Ke^{-r\tau}$, which has $\Delta = e^{-y\tau}$, $\partial_K\varphi = -e^{-r\tau}$, $\Gamma = \mathcal{V} = 0$, $\rho_y = -\tau xe^{-y\tau}$, $\rho_r = \tau Ke^{-r\tau}$ and $\Theta = yxe^{-y\tau} - rKe^{-r\tau}$, and satisfies (a)–(d); by linearity so does p. Finally, (e) is a combination of (a)–(d): by (b) and (a), $-r\rho_r/\tau = r(V - x\Delta)$ and $-y\rho_y/\tau = yx\Delta$, and by (c) $-\sigma\mathcal{V}/(2\tau) = -\frac{1}{2}\sigma^2x^2\Gamma$; substituted in (d) they give (e). □

Each relation is a symmetry, and (e) is their sum. (a): the price is homogeneous of degree one in (x, K): quoting in cents multiplies spot, strike and price by 100 (Euler's theorem, Theorem 4.7). (b): r and y enter only through F and the discount factor. (d): τ enters only through $r\tau$, $y\tau$ and $\sigma^2\tau$. (c): raising the variance $\sigma^2\tau$ runs the heat equation of Theorem 2.3.

In the test case the call has $\Delta = e^{-yT}N(d_+) = 0.509643$, $\Gamma = 0.0195277$, $\mathcal{V} = 39.0554$, $\Theta = -3.32226$ per year, $\rho_r = 43.6000$ and $\rho_y = -50.9643$; the put has $\Delta = -0.470555$, $\Theta = -4.29260$, $\rho_r = -55.4049$ and $\rho_y = 47.0555$. For instance $\rho_r + \rho_y = -7.36429 = -TC$, and $-(0.01 \times 43.6000 - 0.02 \times 50.9643 + 0.1 \times 39.0554) = -3.32226 = \Theta$. The strike Greeks are the digital and the Arrow–Debreu prices: $-\partial_K c = e^{-rT}N(d_-) = 0.436000$ is the cash-or-nothing call of Theorem 2.1, and $\partial_{KK}c = e^{-rT}N'(d_-)/(K\sigma\sqrt{T})$ is the Green's function of Theorem 1.12 at K (Breeden–Litzenberger, Theorem 3.5). In the test case $\partial_{KK}c = 0.0195277 = \Gamma$: not a coincidence, since $S = K$ (Exercise 2.2).

Θ is per year and $\mathcal{V}$ per unit of σ. The call theta need not be negative: $\Theta - (yxe^{-y\tau} - rKe^{-r\tau}) \to 0$ as $x \to \infty$, so it is positive for large x when $y > 0$.

▶ Interactive demo: the Greeks against the spot, and the relations of Theorem 2.7.

2.6 Implied volatility

All inputs of the formula except σ are observable, and markets quote options by the volatility that reproduces their price. Write $D = e^{-rT}$ and $F = Se^{(r-y)T}$ for the time-0 discount factor and forward. With $m = \log(F/K)$ and the total volatility $w = \sigma\sqrt{T}$, Black's form gives $c(0, S) = DK\,c(m, w)$, where

$$c(m, w) = e^m N\Big(\frac{m}{w} + \frac{w}{2}\Big) - N\Big(\frac{m}{w} - \frac{w}{2}\Big), \qquad w > 0, \tag{2.16}$$

the call price in units of the discounted strike: one function of two variables instead of six.

Definition 2.8 (Implied volatility)**.** Let $S, K, T > 0$ and $r, y \in \mathbb{R}$. The *implied volatility* of a call price C is the unique $\sigma_{\text{imp}} > 0$, if there is one, whose Black–Scholes call price is C; for a put price, use the put formula. The *total implied volatility* $w_{\text{imp}} = \sigma_{\text{imp}}\sqrt{T}$ solves $c(m, w) = C/(DK)$.

Proposition 2.9 (Implied volatility: existence, uniqueness and Newton's method)**.** *Fix $m \in \mathbb{R}$ and write $c(w) = c(m, w)$, $d_\pm = m/w \pm w/2$.*

(i) c is infinitely differentiable and strictly increasing on $(0, \infty)$, with $c'(w) = N'(d_-) > 0$, $\lim_{w\downarrow 0} c(w) = (e^m - 1)^+$ and $\lim_{w\to\infty} c(w) = e^m$. Hence a call price C has an implied volatility if and only if

$$\left(Se^{-yT} - Ke^{-rT}\right)^+ < C < Se^{-yT}, \tag{2.17}$$

a put price P if and only if $(Ke^{-rT} - Se^{-yT})^+ < P < Ke^{-rT}$, and prices that satisfy parity have the same implied volatility.

(ii) $c''(w) = d_+ d_- N'(d_-)/w$. If $m \neq 0$, c is strictly convex on $(0, w^*]$ and strictly concave on $[w^*, \infty)$, where

$$w^* = \sqrt{2|m|}, \qquad \text{that is} \qquad \sigma^{*2} = \frac{2|\log(F/K)|}{T};$$

if $m = 0$, c is strictly concave on $(0, \infty)$. So w^* is where the vega c' is largest.

(iii) (Newton on the price; Manaster and Koehler (1982).) Let $\hat{c}$ be a value in $((e^m - 1)^+, e^m)$ and w_{imp} the solution of $c(w) = \hat{c}$. If $m \neq 0$, the iteration

$$w_{n+1} = w_n - \frac{c(w_n) - \hat{c}}{N'(d_-(w_n))}, \qquad w_0 = w^*,$$

stays in $(0, \infty)$ and converges monotonically to w_{imp}: increasing if $w_{\text{imp}} > w^*$, decreasing if $w_{\text{imp}} < w^*$. If $m = 0$, the same holds from $w_0 = \sqrt{2\pi}\,\hat{c}$, which lies below w_{imp}. In both cases the convergence is eventually quadratic.

(iv) (Newton on the log-price.) Let $m \leq 0$. Then $\log c$ is strictly concave in w on $(0, \infty)$. For every $w_0 > 0$, the iteration

$$w_{n+1} = w_n - \left(\log c(w_n) - \log \hat{c}\right)\frac{c(w_n)}{N'(d_-(w_n))},$$

replaced by $w_{n+1} = w_n/2$ when this is not positive, converges to w_{imp}. Its iterates fall to or below w_{imp} after finitely many steps, then increase monotonically, and the convergence is eventually quadratic.

Proof. (i) Differentiating (2.16), with $e^m N'(d_+) = N'(d_-)$ (the normalised form of Theorem 2.7(i)) and $\partial_w(d_+ - d_-) = 1$, gives $c' = N'(d_-)$: the vega in units of $DK\sqrt{T}$. As $w \downarrow 0$, $d_\pm$ tend to $+\infty$, $-\infty$ or 0 according as $m > 0$, $m < 0$ or $m = 0$, so $c \to e^m - 1$, 0 or $\frac{1}{2} - \frac{1}{2} = 0$; as $w \to \infty$, $d_\pm \to \pm\infty$ and $c \to e^m$. By continuity and strict monotonicity, $c(w) = \hat{c}$ has a solution if and only if $(e^m - 1)^+ < \hat{c} < e^m$, and then exactly one. With $\hat{c} = C/(DK)$, $DK(e^m - 1)^+ = (Se^{-yT} - Ke^{-rT})^+$ and $DKe^m = Se^{-yT}$, which is (2.17). A put price P has implied volatility σ if and only if $P + Se^{-yT} - Ke^{-rT}$ is the call price at σ (parity, Theorem 2.1),

which gives the put bounds and the last claim.

(ii) $\partial_w N'(d_-) = -d_- N'(d_-)\,\partial_w d_-$ and $\partial_w d_- = -m/w^2 - \frac{1}{2} = -d_+/w$. The product $d_+ d_- = m^2/w^2 - w^2/4$ is positive for $w < \sqrt{2|m|}$ and negative for $w > \sqrt{2|m|}$; if $m = 0$ it is $-w^2/4 < 0$.

(iii) We use a *tangent argument*. Let f be increasing and differentiable with $f' > 0$, $f(w_{\text{imp}}) = 0$, and $w_{n+1} = w_n - f(w_n)/f'(w_n)$, the zero of the tangent ℓ_n at w_n. If f is concave on an interval containing w_n and w_{imp}, with $w_n \le w_{\text{imp}}$, then $\ell_n \ge f$ there, so $\ell_n(w_n) \le 0 \le \ell_n(w_{\text{imp}})$ and $w_n \le w_{n+1} \le w_{\text{imp}}$. If f is convex on such an interval and $w_{\text{imp}} \le w_n$, then $\ell_n \le f$ there and $w_{\text{imp}} \le w_{n+1} \le w_n$. Apply this to $f = c - \hat{c}$. If $m \neq 0$ and $w_{\text{imp}} \ge w^*$, c is concave on $[w^*, \infty)$ and the iterates increase from w^*; if $w_{\text{imp}} < w^*$, c is convex on $(0, w^*]$ and they decrease, staying above $w_{\text{imp}} > 0$. If $m = 0$, $c(w) = \int_{-w/2}^{w/2} N'(u)\,\mathrm{d}u \le w/\sqrt{2\pi}$ since $N' \le 1/\sqrt{2\pi}$; at w_{imp} this gives $0 < w_0 = \sqrt{2\pi}\hat{c} \le w_{\text{imp}}$, and c is concave on $(0, \infty)$. In each case the iterates are monotone and bounded, so they converge to some $L > 0$, and letting $n \to \infty$ in the iteration gives $c(L) = \hat{c}$, so $L = w_{\text{imp}}$. For the rate, Taylor's formula gives $w_{n+1} - w_{\text{imp}} = c''(\eta_n)(w_n - w_{\text{imp}})^2/(2c'(w_n))$ with η_n between w_n and w_{imp}, and on the compact interval that contains the iterates $|c''|$ is bounded and c' is bounded away from 0.

(iv) For $m \le 0$, $c(0+) = 0$, so $c(w) = \int_0^w g(v)\,\mathrm{d}v$ with $g(v) = N'(m/v - v/2) = c'(v)$. Since $\log g(v) = \frac{m}{2} - \frac{m^2}{2v^2} - \frac{v^2}{8} - \log\sqrt{2\pi}$ is strictly concave, g'/g is strictly decreasing, and for $w > 0$

$$c(w)\,\frac{g'(w)}{g(w)} = \int_0^w g(v)\,\frac{g'(w)}{g(w)}\,\mathrm{d}v < \int_0^w g'(v)\,\mathrm{d}v = g(w) - g(0+) \le g(w),$$

that is, $cc'' < c'^2$: $G = \log c - \log\hat{c}$ is strictly concave, with $G' = c'/c > 0$ (the integral of a log-concave function is log-concave; Bagnoli and Bergstrom (2005), Theorem 1). Now apply the tangent argument to G, which is concave everywhere. If $w_n > w_{\text{imp}}$, the tangent $\ell_n \ge G$ has $\ell_n(w_n) > 0$ and $\ell_n(w_{\text{imp}}) \ge 0$, so the Newton point is at most w_{imp}; if it is not positive, $w_{n+1} = w_n/2$. Halving can happen only finitely many times in a row before an iterate drops to or below w_{imp}, so some $w_{n_0} \le w_{\text{imp}}$. If $w_n \le w_{\text{imp}}$, then $w_n \le w_{n+1} \le w_{\text{imp}}$, with no halving. From n_0 on, the iterates increase to a limit $L \in (0, w_{\text{imp}}]$ with $G(L) = 0$, so $L = w_{\text{imp}}$; the rate follows as in (iii), with G in place of c. □

Part (iv) is due to Choi et al. (2024): their Lemma 2 (the ratio c/c' of price to vega increases in w) is the concavity of $\log c$, and their Proposition 6 runs Newton on $\log c$ from a lower bound. The halving safeguard, hence convergence from every start, is added here; Jäckel (2015)'s objective $1/\log b$ for small prices is a close relative. The concavity holds in $w = \sigma\sqrt{T}$, hence in σ for fixed T, and in the total variance w^2 by composition with the concave increasing square root. It fails for $m > 0$, where $c(0+) = e^m - 1 > 0$: at $m = 0.5$, $w = 0.3$, $(\log c)'' = 1.73$.

Remark 2.10 (Newton on the price is not globally convergent). Part (iii) needs its starting point. At the money ($m = 0$, c concave), with $w_{\text{imp}} = 0.2$ ($\hat{c} = 2N(0.1) - 1 = 0.0797$) and $w_0 = 2$, the first step gives $w_1 = -0.492$; for $m = -0.5$, $w_{\text{imp}} = 0.4$ ($\hat{c} = 0.0155$) and $w_0 = 3$, it gives $w_1 = -1.904$ (Figure 2.2(a)). From both starts the iteration of (iv) converges.

Remark 2.11 (The algorithm). `code/src/mfe230d/impliedvol.py` proceeds as follows. (1) *Degenerate inputs.* If $T \le 0$, or $K = 0$ (price Se^{-yT} for every σ), there is no implied volatility; a price outside the bounds of (i) is an arbitrage (Theorem 3.1); a price at the lower bound is the limit $\sigma \downarrow 0$ (Theorem 2.2), reported as 0, and one at the upper bound is reported as $+\infty$. (2) *Out of the money.* If $m > 0$ we pass to the put by parity, $\hat{p} = \hat{c} - (e^m - 1)$, and use *put–call*

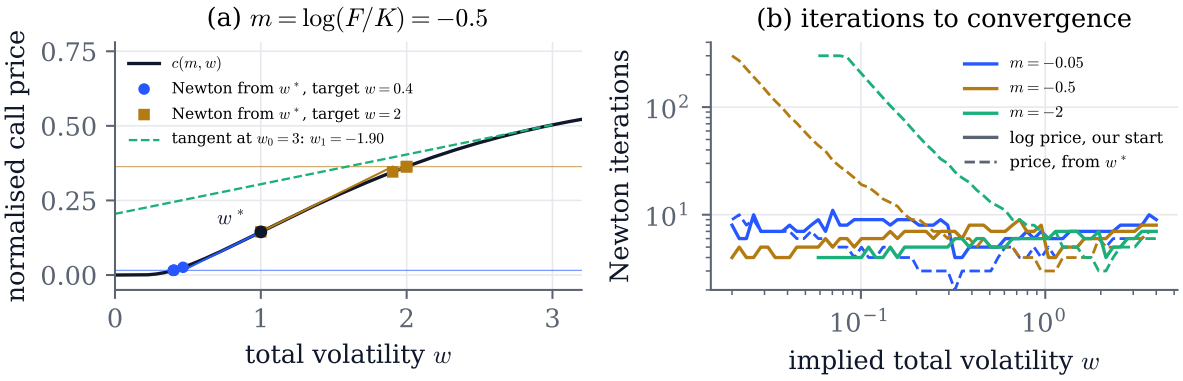

Figure 2.2: Newton's method for the implied volatility. (a) The normalised price $c(m, w)$ for $m = -0.5$, convex below $w^* = 1$ and concave above. Newton on the price from w^* decreases to the target $w = 0.4$ (blue) and increases to the target $w = 2$ (gold); horizontal lines mark the target prices. The tangent at $w_0 = 3$ (dashed) meets the level of the first target at $w_1 = -1.90$. (b) Iterations to convergence against the implied total volatility w (from 0.02 to 4), for $m = -0.05$, -0.5 and -2: Newton on the log-price with the starting point of Theorem 2.11 (solid) and Newton on the price from w^* (dashed, stopped at 300 iterations). Points with $c(m, w) < 10^{-300}$ are omitted.

symmetry, $\hat{p}(m, w) = N(-d_-) - e^m N(-d_+) = e^m\, c(-m, w)$ (as $-d_\pm(m) = d_\mp(-m)$), so we always solve $c(m, w) = \beta$ with $m \le 0$ and $0 < \beta < e^m$, where (iv) applies. The parity step loses relative accuracy when the time value is small: the price then carries little information about the volatility. (3) *Start.* At w^*, $d_+ = 0$ and $d_- = -w^*$, so $c(w^*) = \frac{1}{2}e^m - N(-w^*)$. If $\beta \ge c(w^*)$ we start at w^*, below the root; otherwise at $\min(w^*, w_g)$, where w_g solves $d_-(w) = -\sqrt{-2\log\beta}$ (the leading term of $\log c$ for small prices), on either side of the root, which (iv) allows. At $m = 0$ we start at $\sqrt{2\pi}\beta$ (Brenner and Subrahmanyam, 1988). (4) *Iteration.* Below w^* we evaluate $\log c = \log N'(d_-) + \log\left(R(-d_+) - R(-d_-)\right)$, where $R(u) = N(-u)/N'(u)$ is Mills' ratio, computed from the scaled complementary error function, to avoid underflow. We stop when an update is below $10^{-15}w$ or when an iterate below the root fails to increase (rounding).

Our implementation (numpy `default_rng`, seed 2026), on 12,454 random cases (m uniform on $[-2, 0]$, w log-uniform on $[0.02, 2]$, $c(m, w) > 10^{-14}e^m$), needs 6.3 iterations on average and 17 at most, with relative errors below 10^{-13}; Newton on the price from w^* needs 9.3 and 34, and many more for smaller prices (Figure 2.2(b)), though it is faster near the money. The test-case call ($m = -0.01$, $\beta = 0.0744 > c(w^*) = 0.0513$) takes 5 iterations.

▶ Interactive demo: Newton on the price and on the log-price, from a start you choose.

Further reading

The formula is due to Black and Scholes (1973) and Merton (1973). On implied volatility see Manaster and Koehler (1982), Higham (2004), Chapter 14, and Jäckel (2015), who reaches machine precision with two third-order (Householder) steps.

Chapter 3

Arbitrage in the strike and the smile

Fix a maturity and look at the price of a European call as a function of its strike. Before any model is chosen, absence of arbitrage already shapes this function: it is convex and nonincreasing, its slopes lie between $-e^{-rT}$ and 0, and it starts at the price of the stock without its dividends. Minus its first derivative is the price of a digital option, and its second derivative is the discounted risk-neutral density of S_T: the *Breeden–Litzenberger formula*, which turns a strip of option prices into a probability law. Chapter 1 met the same object as the Green's function of the pricing equation (Theorem 1.12).

Markets quote options through their Black–Scholes implied volatilities, and for equity indices these are far from flat: they fall as the strike rises. Since the strike enters a price through the implied volatility as well, every strike derivative picks up terms in the slope and the curvature of the smile. For an at-the-money digital the first of them, the *skew term*, is about a fifth of the price (Theorem 3.9).

The IEOR 222 notes prove the strike relations in a one-period market without dividends. Here we add the yield, prove the Breeden–Litzenberger formula in the generality that market data need (atoms in the law of S_T, call curves with kinks), characterize exactly which functions are call prices, and turn to methods: extracting a density from quotes and choosing the strike spacing (Theorem 3.6), pricing and replicating digitals under a smile (Theorems 3.8 and 3.10), and testing a parametric smile for arbitrage. Chapter 4 builds general payoffs out of the same strip of options.

Throughout, $T > 0$ is a fixed maturity, $r, y \in \mathbb{R}$, $S_0 > 0$, $F = S_0 e^{(r-y)T}$ is the forward price, and $\Delta_h^2 f(K) = \big(f(K-h) - 2f(K) + f(K+h)\big)/h^2$ is the second difference of a function f with step $h > 0$. One-sided derivatives are written $\partial^{\pm} f(K)$.

3.1 Call prices across strikes

The market. At time 0 the following claims on the date T are traded, in a frictionless market (any real quantity, no bid–ask spread):

- the zero-coupon bond, which pays 1 and costs e^{-rT};

- the *prepaid forward*, which pays S_T and costs S_0e^{-yT}; in the standard model of Theorem 1.1, e^{-yT} shares bought at time 0, with the dividends reinvested in shares, grow to one share at T;
- European calls and puts with strikes $K \geq 0$, which pay $(S_T - K)^+$ and $(K - S_T)^+$ and cost $C(K)$ and $P(K)$.

The terminal price S_T is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ with $S_T \geq 0$ (limited liability). A static portfolio holds fixed quantities of finitely many of these claims; its price and its payoff are the corresponding linear combinations. We use one of two hypotheses.

(A) *No static arbitrage:* no static portfolio with a payoff $V \geq 0$ almost surely has a negative price, or a price of zero together with $\mathbb{P}(V > 0) > 0$.

(B) *Pricing measure:* there is a probability measure $\mathbb{Q}$ equivalent to $\mathbb{P}$ with $\mathbb{E}^{\mathbb{Q}}[S_T] < \infty$ under which every traded payoff X costs $e^{-rT}\mathbb{E}^{\mathbb{Q}}[X]$. In particular $\mathbb{E}^{\mathbb{Q}}[S_T] = e^{rT}S_0e^{-yT} = F$, and $C(K) = e^{-rT}\mathbb{E}^{\mathbb{Q}}[(S_T - K)^+]$.

Both imply *positivity*:

(P) a static portfolio whose payoff is nonnegative almost surely has a nonnegative price.

Under (A) this is part of the definition; under (B) it holds because the expectation of a nonnegative variable is nonnegative, and $\mathbb{P}$ and $\mathbb{Q}$ have the same null sets. Applied to a payoff V and to $-V$, (P) gives the *law of one price*: if two portfolios have the same payoff almost surely, they have the same price. Hypothesis (B) implies (A) (the argument is the one for a one-period market in the IEOR 222 notes); the converse, the fundamental theorem of asset pricing, is not needed here. The proposition below uses (P) only, so it is *model-free*: it holds in every market free of static arbitrage, whatever the law of S_T.

Proposition 3.1 (Strike relations)**.** *Assume (P). Let $K \geq 0$, $0 \leq K_1 < K_2$, $\lambda \in [0,1]$ and $K_\lambda = \lambda K_1 + (1-\lambda)K_2$.*

*(i) (*Parity and bounds*)* $C(K) - P(K) = S_0e^{-yT} - Ke^{-rT}$, $C(0) = S_0e^{-yT}$, $P(0) = 0$, *and*

$$(S_0e^{-yT} - Ke^{-rT})^+ \leq C(K) \leq S_0e^{-yT}, \qquad (Ke^{-rT} - S_0e^{-yT})^+ \leq P(K) \leq Ke^{-rT}.$$

*(ii) (*Monotonicity and slopes*)* $-e^{-rT}(K_2 - K_1) \leq C(K_2) - C(K_1) \leq 0$ *and* $0 \leq P(K_2) - P(K_1) \leq e^{-rT}(K_2 - K_1)$.

*(iii) (*Convexity*)* $C(K_\lambda) \leq \lambda C(K_1) + (1-\lambda)C(K_2)$, *and the same for P.*

Consequently C is convex, nonincreasing and Lipschitz with constant e^{-rT} on $[0,\infty)$. It has one-sided derivatives at every $K > 0$, with $-e^{-rT} \leq \partial^- C(K) \leq \partial^+ C(K) \leq 0$, and a right derivative $\partial^+ C(0) \in [-e^{-rT}, 0]$. The put is convex and nondecreasing, with slopes in $[0, e^{-rT}]$, and $P'' = C''$ wherever either exists.

Proof. For $x \geq 0$ and K, K_1, K_2 as in the statement, the following hold:

$$(x-K)^+ - (K-x)^+ = x - K, \qquad 0 \leq (x-K)^+, \qquad x - K \leq (x-K)^+ \leq x, \tag{3.1}$$

$$(x-K_1)^+ - (x-K_2)^+ = \min\left((x-K_1)^+, K_2 - K_1\right) \in [0, K_2 - K_1], \tag{3.2}$$

$$\lambda(x-K_1)^+ + (1-\lambda)(x-K_2)^+ - (x-K_\lambda)^+ \geq 0. \tag{3.3}$$

The upper bound in (3.1) uses $x \geq 0$ and $K \geq 0$; (3.2) is checked on $x \leq K_1$, $K_1 < x < K_2$ and $x \geq K_2$; and (3.3) holds because $K \mapsto (x-K)^+$ is the maximum of the affine functions 0 and $x - K$, hence convex. At $x = S_T$ each line compares the payoffs of two static portfolios of calls, puts, bonds and the prepaid forward, and (P) with the law of one price turns it into the claimed relation between prices. For example, a call minus a put pays $S_T - K$, which is also the payoff of the prepaid forward minus K bonds; this is parity. A call with strike 0 pays S_T, so $C(0) = S_0 e^{-yT}$, and a put with strike 0 pays nothing. The *call spread* (3.2) pays between 0 and $K_2 - K_1$, the payoff of $K_2 - K_1$ bonds; this is (ii) for calls. The portfolio in (3.3) is a *butterfly*; this is (iii). By parity P differs from C by the affine function $K \mapsto Ke^{-rT} - S_0 e^{-yT}$, which turns the relations for calls into those for puts and leaves second derivatives unchanged.

For the derivatives, (iii) says that the slope $(C(b) - C(a))/(b-a)$ of a chord is nondecreasing in each endpoint, for $0 \leq a < b$. So as $h \downarrow 0$ the quotients $(C(K+h) - C(K))/h$ decrease and $(C(K) - C(K-h))/h$ increase, the first stays above the second, and by (ii) both lie in $[-e^{-rT}, 0]$. Their limits are $\partial^+ C(K) \geq \partial^- C(K)$ (only the first one at $K = 0$). □

Read backwards, each line of the proof is a trade. A call spread quoted at a negative price, or above $e^{-rT}(K_2 - K_1)$, is bought, respectively sold against $K_2 - K_1$ bonds; a butterfly quoted at a negative price is bought. The IEOR 222 notes work these trades out in detail.

Remark 3.2 (Edges and atoms). At $K = 0$ the call is worth $S_0 e^{-yT}$, not S_0: it is a claim to the share at T without the dividends paid before T. Under (B), the right slope at the origin is $\partial^+ C(0) = -e^{-rT}\mathbb{Q}(S_T > 0)$ (Theorem 3.5(i)): it equals the extreme value $-e^{-rT}$ exactly when $\mathbb{Q}(S_T = 0) = 0$, and a stock that can default with positive risk-neutral probability has a flatter call curve at the origin (Exercise 3.3). As $K \to \infty$, Theorem 3.1 only gives that C decreases to some limit ≥ 0; under (B) the limit is 0, by dominated convergence with $(S_T - K)^+ \leq S_T$, and the slopes $\partial^+ C(K) = -e^{-rT}\mathbb{Q}(S_T > K)$ tend to 0 as well. A kink of C at some $K > 0$, that is $\partial^- C(K) < \partial^+ C(K)$, is the signature of an atom of the law of S_T at K (again Theorem 3.5(i)), such as a cash takeover offer at the price K.

Remark 3.3 (Finitely many strikes, several maturities). Markets quote calls at finitely many strikes $0 < K_1 < \cdots < K_n$. With $K_0 = 0$ and $C(K_0) = S_0 e^{-yT}$, let $s_i = (C(K_i) - C(K_{i-1}))/(K_i - K_{i-1})$. Theorem 3.1 requires

$$-e^{-rT} \leq s_1 \leq s_2 \leq \cdots \leq s_n \leq 0 \qquad \text{and} \qquad C(K_n) \geq 0, \tag{3.4}$$

and conversely (3.4) implies every relation of the proposition between quoted strikes, because the piecewise-linear interpolation of the quotes is then convex, nonincreasing and has slopes in $[-e^{-rT}, 0]$. With equal spacing h, $s_{i+1} - s_i = h\,\Delta_h^2 C(K_i)$ is the price of a butterfly divided by h. These conditions are *necessary*. Whether quotes satisfying them can be produced by some law of S_T with mean F (a "model" in the sense of Davis and Hobson (2007), which chooses the null sets too) is a separate theorem: yes, except in the boundary case $s_n = 0 < C(K_n)$, a "weak arbitrage" (their Theorem 3.1 for one maturity, Theorem 4.2 for several; the IEOR 222 notes construct a discrete law). Under (B) with a given $\mathbb{P}$, the $\mathbb{Q}$-law of S_T must be equivalent to its $\mathbb{P}$-law, which can rule out more. If the $\mathbb{P}$-law has full support, so has the $\mathbb{Q}$-law, and then all inequalities in (3.4) are strict: $s_1 > -e^{-rT}$ because $\mathbb{Q}(S_T < K_1) > 0$, $s_{i+1} > s_i$ because the butterfly on K_{i-1}, K_i, K_{i+1} pays off on (K_{i-1}, K_{i+1}), and $s_n < 0 < C(K_n)$ because $\mathbb{Q}(S_T > K_n) > 0$. Conversely, if the $\mathbb{P}$-law has a positive density on $(0, \infty)$ (lognormal, say), strict inequalities suffice. Pick $\sigma_0 = -e^{-rT}$, $\sigma_i \in (s_i, s_{i+1})$ for $i < n$ and $\sigma_n \in (s_n, 0)$; the convex function through

the quotes with slopes $\sigma_0 < \sigma_1 < \cdots < \sigma_n < 0$, then 0, has one kink in each (K_{i-1}, K_i) and one in (K_n, ∞), and by Theorem 3.5(iv) it prices calls for a law with an atom at each kink. Replace each atom by a positive density on its cell with the same mass and mean (a small weight on a fixed positive density on the cell, the rest uniform near the atom, will do): the prices at the K_i do not change, since every $(x - K_i)^+$ is affine on each cell. With q the resulting density and p the $\mathbb{P}$-density of S_T, $\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P} = q(S_T)/p(S_T)$ gives (B). For a rectangular grid (the same strikes at every maturity), with zero rates and dividends, Carr and Madan (2005) show that nonnegative prices of call spreads, butterflies and calendar spreads suffice to exclude static arbitrage.

Remark 3.4 (Calendar spreads). Let calls with maturities $T_1 < T_2$ be priced by one pricing measure under which $X_t = e^{-(r-y)t} S_t$ is a martingale on $[0, T_2]$, as in the standard model. Then $C(K,T) = e^{-yT}\mathbb{E}^{\mathbb{Q}}[(X_T - \kappa)^+]$ with $\kappa = Ke^{-(r-y)T}$, and conditional Jensen, $\mathbb{E}^{\mathbb{Q}}[(X_{T_2} - \kappa)^+ \mid \mathcal{F}_{T_1}] \geq (X_{T_1} - \kappa)^+$, gives

$$e^{yT_2} C\big(\kappa e^{(r-y)T_2}, T_2\big) \;\geq\; e^{yT_1} C\big(\kappa e^{(r-y)T_1}, T_1\big), \qquad \kappa \geq 0 : \tag{3.5}$$

calls are nondecreasing in the maturity when the strike moves with the forward and prices are counted in prepaid forwards. In terms of implied volatilities (Section 3.3), the total implied variance $\sigma^2 T$ at a fixed ratio K/F is nondecreasing in T. At a fixed strike the call can lose value with maturity when $y > 0$ (Exercise 3.4).

3.2 The Breeden–Litzenberger formula

In this section (B) holds, and μ is the law of S_T under $\mathbb{Q}$, a probability measure on $[0, \infty)$ with mean F. It is convenient to define $C(K) = e^{-rT}\mathbb{E}^{\mathbb{Q}}[(S_T - K)^+]$ for every real K: for $K < 0$ this is $S_0 e^{-yT} - Ke^{-rT}$, the price of the prepaid forward and $-K$ bonds. The butterfly with strikes $K - h$, K, $K + h$ pays, for every x,

$$(x - K + h)^+ - 2(x - K)^+ + (x - K - h)^+ = h^2 \Lambda_h(x - K), \qquad \Lambda_h(u) = \frac{(h - |u|)^+}{h^2}, \tag{3.6}$$

a tent of height $1/h$ and area 1 centred at K (check the four intervals cut by $K - h$, K, $K + h$). Its price is $h^2 \Delta_h^2 C(K)$.

Theorem 3.5 (Breeden–Litzenberger). *Assume (B), and let μ be the law of S_T under $\mathbb{Q}$.*

*(i) (*Digitals*) For every $K > 0$,*

$$\partial^+ C(K) = -e^{-rT}\mathbb{Q}(S_T > K), \qquad \partial^- C(K) = -e^{-rT}\mathbb{Q}(S_T \geq K),$$

and $\partial^+ C(0) = -e^{-rT}\mathbb{Q}(S_T > 0)$. So $\partial^+ C(K) - \partial^- C(K) = e^{-rT}\mathbb{Q}(S_T = K)$, and C is differentiable at $K > 0$ if and only if $\mathbb{Q}(S_T = K) = 0$.

*(ii) (*Second differences and the law*) For $h > 0$ and $K \in \mathbb{R}$, $\Delta_h^2 C(K) = e^{-rT}\mathbb{E}^{\mathbb{Q}}[\Lambda_h(S_T - K)] \geq 0$. As $h \downarrow 0$:*

(a) $h\,\Delta_h^2 C(K) \to e^{-rT}\mathbb{Q}(S_T = K)$ for every K;

(b) $\int_{\mathbb{R}} \phi(K)\,\Delta_h^2 C(K)\,\mathrm{d}K \to e^{-rT}\mathbb{E}^{\mathbb{Q}}[\phi(S_T)]$ for every bounded continuous $\phi : \mathbb{R} \to \mathbb{R}$: the measures $\Delta_h^2 C(K)\,\mathrm{d}K$ converge weakly to $e^{-rT}\mu$.

Moreover $C'' = e^{-rT}\mu$ in the sense of distributions: $\int_{\mathbb{R}} C(K)\,\phi''(K)\,\mathrm{d}K = e^{-rT}\mathbb{E}^{\mathbb{Q}}[\phi(S_T)]$ for every $\phi \in C^2(\mathbb{R})$ that vanishes outside a bounded interval.

(iii) (Densities) *Let $I \subseteq (0,\infty)$ be an open interval. If μ has a density q on I, that is $\mathbb{Q}(S_T \in A) = \int_A q(x)\,\mathrm{d}x$ for Borel sets $A \subseteq I$, then C is differentiable on I, $C'(b) - C'(a) = e^{-rT}\int_a^b q(x)\,\mathrm{d}x$ for $a < b$ in I, and at every $K \in I$ where q is continuous*

$$C''(K) = \lim_{h\downarrow 0} \Delta_h^2 C(K) = e^{-rT} q(K).$$

Conversely, if C is differentiable on I and C' is absolutely continuous on I (for instance $C \in C^2(I)$), then μ has the density $e^{rT}C''$ on I.

(iv) (Which functions are call prices) *A function $c : [0,\infty) \to \mathbb{R}$ is of the form $c(K) = e^{-rT}\int (x-K)^+\,\nu(\mathrm{d}x)$ for a probability measure ν on $[0,\infty)$ with mean F if and only if c is convex, $c(0) = S_0 e^{-yT}$, $\partial^+ c(0) \geq -e^{-rT}$, and $c(K) \to 0$ as $K \to \infty$. Then ν is unique: $\nu((K,\infty)) = -e^{rT}\partial^+ c(K)$ for $K \geq 0$.*

Proof. (i) Let $K > 0$ and $0 < h \leq K$. By (3.2),

$$\frac{C(K+h) - C(K)}{h} = -e^{-rT}\mathbb{E}^{\mathbb{Q}}\Big[\frac{\min\left((S_T - K)^+, h\right)}{h}\Big],$$
$$\frac{C(K) - C(K-h)}{h} = -e^{-rT}\mathbb{E}^{\mathbb{Q}}\Big[\frac{\min\left((S_T - K + h)^+, h\right)}{h}\Big].$$

Both integrands take values in $[0,1]$; as $h \downarrow 0$ the first converges to $\mathbf{1}_{\{S_T > K\}}$ and the second to $\mathbf{1}_{\{S_T \geq K\}}$, for every outcome. Dominated convergence gives the two derivatives, and the first computation applies at $K = 0$ as well.

(ii) The identity is (3.6) at $x = S_T$, divided by h^2, and priced. (a) The variable $h\Lambda_h(S_T - K) = (1 - |S_T - K|/h)^+$ lies in $[0,1]$ and converges to $\mathbf{1}_{\{S_T = K\}}$; use dominated convergence. (b) The integrand $|\phi(K)|\Lambda_h(S_T - K)$ is jointly measurable and its double integral is at most $\sup|\phi|$, so by Fubini's theorem

$$\int_{\mathbb{R}} \phi(K)\,\Delta_h^2 C(K)\,\mathrm{d}K = e^{-rT}\mathbb{E}^{\mathbb{Q}}\Big[\int_{\mathbb{R}} \phi(S_T - u)\,\Lambda_h(u)\,\mathrm{d}u\Big].$$

For each outcome the inner integral is an average of ϕ over $[S_T - h, S_T + h]$, because Λ_h is a probability density on $[-h,h]$; it tends to $\phi(S_T)$ by continuity, and it is bounded by $\sup|\phi|$. Dominated convergence concludes. For the distributional identity, let ϕ vanish outside $[-M,M]$. Integrating by parts twice, with no boundary terms at $-M$, $\int_{\mathbb{R}}(x-K)^+\phi''(K)\,\mathrm{d}K = \int_{-M}^{x}(x-K)\phi''(K)\,\mathrm{d}K = \phi(x)$ for every x. Since $\int_{\mathbb{R}}|(x-K)^+\phi''(K)|\,\mathrm{d}K \leq 2M(|x| + M)\sup|\phi''|$ and $\mathbb{E}^{\mathbb{Q}}|S_T| < \infty$, Fubini's theorem lets us take $\mathbb{E}^{\mathbb{Q}}$ inside, which gives the identity.

(iii) μ has no atoms in I, so by (i) C is differentiable on I with $C'(K) = -e^{-rT}\mathbb{Q}(S_T > K)$, and $C'(b) - C'(a) = e^{-rT}\mathbb{Q}(a < S_T \leq b) = e^{-rT}\int_a^b q$. At a point K where q is continuous (in particular finite near K), the fundamental theorem of calculus gives $C''(K) = e^{-rT}q(K)$; and for h so small that $[K-h, K+h] \subset I$, $\Delta_h^2 C(K) = e^{-rT}\int q(x)\Lambda_h(x-K)\,\mathrm{d}x$, an average of q over $[K-h, K+h]$, which tends to $e^{-rT}q(K)$. Conversely, if C is differentiable on I, then μ has no atoms in I by (i), so $C'(K) = -e^{-rT}\mathbb{Q}(S_T > K)$ on I, and for $a < b$ in I, by absolute continuity, $\mathbb{Q}(a < S_T \leq b) = e^{rT}(C'(b) - C'(a)) = \int_a^b e^{rT}C''(x)\,\mathrm{d}x$; here $C'' \geq 0$ almost everywhere because C' is nondecreasing. The finite measures μ and $e^{rT}C''(x)\,\mathrm{d}x$ on $I = (\alpha,\beta)$ agree on the intervals $(a,b] \subset I$, hence, by continuity from below, on $(\alpha,b]$ for $b \in I$ and on I itself. Extended by zero outside I, they are finite measures on $\mathbb{R}$ with the same distribution function, so they coincide (Durrett (2019), Theorem 1.1.4: a measure on $\mathbb{R}$ is determined by its Stieltjes measure function).

(iv) *Necessity.* c is convex because each $K \mapsto (x-K)^+$ is, and integrating against ν preserves convexity; $c(0) = e^{-rT}\int x\,\nu(\mathrm{d}x) = e^{-rT}F$; $\partial^+c(0) = -e^{-rT}\nu((0,\infty)) \geq -e^{-rT}$ by (i), whose proof only used the law; and $c(K) \to 0$ by dominated convergence.

Sufficiency. First, three facts about a convex c on $[0,\infty)$ with $\partial^+c(0)$ finite. It is continuous: in the interior because chord slopes are monotone, hence locally bounded, and at 0 because the right difference quotient has the finite limit $\partial^+c(0)$. Its right derivative $g = \partial^+c$ exists everywhere, is nondecreasing, and is right-continuous: if $u_n \downarrow u$ and $v > u_1$, then $g(u) \leq g(u_n) \leq (c(v) - c(u_n))/(v-u_n)$, and letting $n \to \infty$ and then $v \downarrow u$ gives $\lim_n g(u_n) = g(u)$. And $c(b) - c(a) = \int_a^b g(u)\,\mathrm{d}u$: for $a = u_0 < \cdots < u_m = b$, convexity gives $g(u_j) \leq (c(u_{j+1}) - c(u_j))/(u_{j+1}-u_j) \leq g(u_{j+1})$, so $c(b)-c(a)$ lies between the lower and the upper Riemann sums of the monotone, hence Riemann-integrable, function g.

Now $g \leq 0$, since $g(u_0) > 0$ would force $c(u) \geq c(u_0) + g(u_0)(u-u_0) \to \infty$; and $g(u) \to 0$ as $u \to \infty$, since g increases to some $L \leq 0$, and $L < 0$ would force $c(u) \leq c(u_0) + L(u-u_0) \to -\infty$. So $G(u) = 1 + e^{rT}g(u)$ is nondecreasing, right-continuous, with $G(0) \geq 0$ by hypothesis and $G(u) \uparrow 1$; set $G = 0$ on $(-\infty, 0)$. It is then the distribution function of a probability ν on $[0,\infty)$ (Durrett (2019), Theorem 1.1.4), with $\nu((u,\infty)) = -e^{rT}g(u)$ for $u \geq 0$. Writing $(x-K)^+ = \int_K^\infty \mathbf{1}_{\{x>u\}}\,\mathrm{d}u$ and using Tonelli's theorem,

$$\int (x-K)^+\,\nu(\mathrm{d}x) = \int_K^\infty \nu((u,\infty))\,\mathrm{d}u = -e^{rT}\int_K^\infty g(u)\,\mathrm{d}u = e^{rT}\Big(c(K) - \lim_{v\to\infty} c(v)\Big) = e^{rT}c(K).$$

At $K = 0$ this gives the mean $e^{rT}c(0) = F$. Uniqueness: by (i), every ν that produces c has $\nu((K,\infty)) = -e^{rT}\partial^+c(K)$ for all $K \geq 0$, so its distribution function is G, and a measure is determined by its distribution function (Durrett (2019), Theorem 1.1.4). □

So the second strike derivative of the call is the discounted risk-neutral density when the density exists and is continuous, and in general it is the discounted law itself, as a measure. Prices at one maturity therefore identify the law of S_T under $\mathbb{Q}$ (not the law under $\mathbb{P}$, and not the law of the path), and part (iv) shows that Theorem 3.1 together with $C(K) \to 0$ is all that a single maturity can tell: every function with those properties is the call price curve of exactly one law of S_T with mean F. Whether that law is admissible for a given $\mathbb{P}$ is a further condition: under (B) it must be equivalent to the $\mathbb{P}$-law of S_T (Theorem 3.3). In the standard model, (iii) gives $C''(K) = G(0, S_0; T, K)$, the Green's function of Theorem 1.12.

Atoms show up in data through part (ii)(a): the butterfly $h^2\Delta_h^2 C(K)$ centred at an atom is asymptotic to $e^{-rT}\mathbb{Q}(S_T = K)\,h$, not $O(h^2)$, so $\Delta_h^2 C$ has a spike of height close to $e^{-rT}\mathbb{Q}(S_T = K)/h$ (Figure 3.1). The extreme case is a riskless stock, $\mu = \delta_F$: then $C(K) = e^{-rT}(F-K)^+$, a ramp whose second derivative in the sense of (ii) is $e^{-rT}\delta_F$. In the standard model the law of S_T tends to δ_{S_0} as $T \downarrow 0$ (Theorem 1.12), and the call curve to the ramp $(S_0 - K)^+$.

▶ Interactive demo: move an atom of the law of S_T and the strike step h, and watch the kink in the call curve, the jump of the digital and the spike of the second difference.

From quotes to densities. Quotes carry errors: rounding to the tick, the bid–ask spread, stale prices. Second differences divide these errors by h^2, so the strike spacing has to balance them against the bias of a coarse grid.

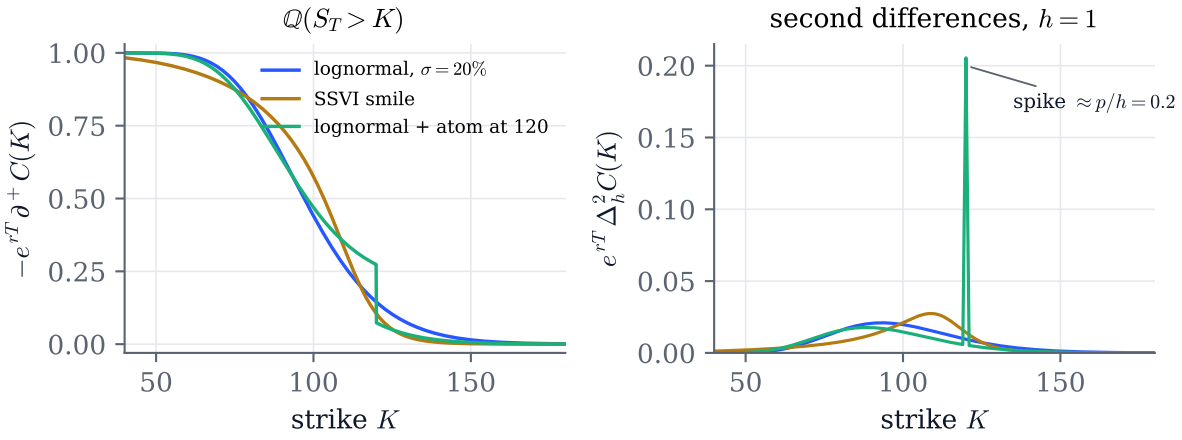

Figure 3.1: Three laws of S_T with the same mean $F = 99.005$ (test case, $T = 1$): lognormal with $\sigma = 20\%$; the SSVI smile of Section 3.3; and a takeover model in which $S_T = 120$ with probability $p = 0.2$ and is otherwise lognormal ($\sigma = 20\%$, mean 93.756). Left: $\mathbb{Q}(S_T > K)$ computed as $-e^{rT}$ times the right slope of the call price, which jumps by p at the atom. Right: $e^{rT}\Delta_h^2 C(K)$ with $h = 1$ on the strikes $40, 41, \dots, 180$, a density estimate for the first two laws and a spike of height p/h plus the continuous part at $K = 120$ for the third.

Proposition 3.6 (Densities from noisy quotes)**.** *Let $h > 0$, let C be four times continuously differentiable on $[K - h, K + h]$ with $|C''''| \le M_4$ there, and let $\tilde{C}$ be quotes with $|\tilde{C} - C| \le \varepsilon$ at $K - h$, K and $K + h$. Then*

$$|\Delta_h^2 \tilde{C}(K) - C''(K)| \le \frac{M_4 h^2}{12} + \frac{4\varepsilon}{h^2}. \tag{3.7}$$

For $M_4, \varepsilon > 0$ the right side is smallest at $h_* = (48\varepsilon/M_4)^{1/4}$, where it equals $2\sqrt{M_4\varepsilon/3}$.

Proof. Taylor's formula with integral remainder at $K \pm h$, added, gives $\Delta_h^2 C(K) - C''(K) = h^{-2}\int_0^h \frac{(h-u)^3}{6}(C''''(K+u) + C''''(K-u))\,du$, whose absolute value is at most $h^{-2} \cdot 2M_4 h^4/24 = M_4 h^2/12$. The errors contribute $|e(K-h) - 2e(K) + e(K+h)|/h^2 \le 4\varepsilon/h^2$, with $e = \tilde{C} - C$. The function $ah^2 + b/h^2$ with $a, b > 0$ is smallest at $h^4 = b/a$, where it equals $2\sqrt{ab}$. □

The bias is of order h^2 and the noise of order ε/h^2, so the best error is of order $\sqrt{\varepsilon}$: halving the price errors improves the density by a factor $\sqrt{2}$ only. As an example, take the SSVI prices of Figure 3.3 at $T = 1$, rounded to the cent ($\varepsilon = 0.005$). On $[50, 150]$, $\max|C''''| = 2.763 \times 10^{-4}$, so $h_* = 5.43$ and the bound is 1.36×10^{-3}, against a peak value 0.0272 of C''. On the strikes $60, 61, \dots, 140$ the largest error of $\Delta_1^2 \tilde{C}$ is 0.0169, more than half the peak; on $60, 65, \dots, 140$ that of $\Delta_5^2 \tilde{C}$ is 0.0010 (Figure 3.3, right). In practice one first fits an arbitrage-free smile to the quotes and differentiates the fit (Section 3.3); higher-order stencils change the exponents (Exercise 3.6).

3.3 The smile

Let $C^{\mathrm{BS}}(K,\sigma) = S_0 e^{-yT} N(d_+) - K e^{-rT} N(d_-)$ be the Black–Scholes call (Theorem 2.1), with $d_\pm = \big(\log(S_0/K) + (r - y \pm \sigma^2/2)T\big)/(\sigma\sqrt{T})$, and let $\mathcal{V} = S_0 e^{-yT} N'(d_+)\sqrt{T} = K e^{-rT} N'(d_-)\sqrt{T}$ be its vega (Theorem 2.7). Under (B), if $\mathbb{Q}(S_T < K) > 0$ and $\mathbb{Q}(S_T > K) > 0$, the bounds of Theorem 3.1(i) are strict (if $F \geq K$, the gap below is $P(K) > 0$; if $F < K$, it is $C(K) > 0$; the gap above is $e^{-rT}\mathbb{E}^{\mathbb{Q}}[\min(S_T, K)] > 0$), so by Theorem 2.9 there is exactly one $\sigma(K) > 0$ with $C(K) = C^{\mathrm{BS}}(K, \sigma(K))$: the implied volatility of Theorem 2.8. By parity the put has the same one. The map $K \mapsto \sigma(K)$ is the *smile* at maturity T. It is flat in the Black–Scholes model, and it is not flat in the market.

Remark 3.7 (The shape of equity-index smiles). Since the crash of October 1987, implied volatilities of S&P 500 index options decrease with the strike over most of the quoted range: out-of-the-money puts are expensive relative to the lognormal model (Rubinstein, 1994). The smile is therefore usually called a *skew*; at short maturities it often turns up again for far out-of-the-money calls. The skew is steepest at short maturities and flattens as the maturity grows: Gatheral (2006), Chapter 3, shows the SPX surface of 15 September 2005 (Figure 3.2) and its at-the-money skew $\partial\sigma/\partial \log K$ against the time to expiry (Figure 3.4). A traders' rule of thumb, used in the course, scales the at-the-money skew like $T^{-1/2}$, so that smiles of different maturities nearly collapse when plotted against $\log(K/S_0)/(\sigma_1\sqrt{T})$, with σ_1 the one-year at-the-money volatility. Over time the level of the surface moves with market stress more than its shape does. The risk-neutral density that Theorem 3.5 extracts from a skew has a fatter left tail and a thinner right tail than the lognormal one (Figures 3.2 and 3.3), consistent with the heavy tails, the gain/loss asymmetry and the negative correlation between returns and volatility (the leverage effect) documented for index returns by Cont (2001), Section 2 (see also the SPX returns in Gatheral (2006), Figures 1.1–1.3), and with the premium that investors pay for crash protection. Since quotes are proprietary, this chapter uses synthetic smiles of the same shape.

Every strike derivative of $C(K) = C^{\mathrm{BS}}(K, \sigma(K))$ has a direct term, at fixed volatility, and terms in the derivatives of the smile. For digitals this changes prices substantially.

Proposition 3.8 (Digitals under a smile). *Let $I \subseteq (0,\infty)$ be an open interval, $\sigma : I \to (0,\infty)$, and $C(K) = C^{\mathrm{BS}}(K, \sigma(K))$ for $K \in I$; below, $d_\pm$, $\mathcal{V}$ and σ are evaluated at $\sigma = \sigma(K)$.*

(i) If σ is differentiable at $K \in I$, then so is C, and

$$-C'(K) = e^{-rT} N(d_-) - \mathcal{V}\,\sigma'(K). \tag{3.8}$$

If moreover the prices come from a pricing measure (B), then $\mathbb{Q}(S_T = K) = 0$, and the cash-or-nothing call paying $\mathbf{1}_{\{S_T > K\}}$ and the asset-or-nothing call paying $S_T \mathbf{1}_{\{S_T > K\}}$ cost

$$D(K) = e^{-rT}\mathbb{Q}(S_T > K) = -C'(K),$$
$$A(K) = e^{-rT}\mathbb{E}^{\mathbb{Q}}\big[S_T \mathbf{1}_{\{S_T > K\}}\big] = C(K) + K\,D(K) = S_0 e^{-yT} N(d_+) - K\mathcal{V}\sigma'(K).$$

(ii) If σ is differentiable on I and twice differentiable at K, then $C''(K)$ exists and

$$C''(K) = \frac{e^{-rT} N'(d_-)}{K\sigma\sqrt{T}} \Big[1 + 2K\sqrt{T}\, d_+ \sigma' + K^2 T\, d_+ d_-\, \sigma'^2 + K^2 T \sigma\, \sigma''\Big]. \tag{3.9}$$

Under (B), if σ is twice continuously differentiable on I, the law of S_T has the continuous density $e^{rT} C''$ on I.

Proof. (i) C^{BS} is continuously differentiable in (K,σ), with $\partial_K C^{\mathrm{BS}} = -e^{-rT}N(d_-)$ (the terms in $\partial_K d_\pm$ cancel because $S_0 e^{-yT}N'(d_+) = Ke^{-rT}N'(d_-)$) and $\partial_\sigma C^{\mathrm{BS}} = \mathcal{V}$; the chain rule gives (3.8). Under (B), Theorem 3.5(i) gives $\mathbb{Q}(S_T = K) = 0$ and $D(K) = -\partial^+ C(K) = -C'(K)$. The identity $(S_T - K)^+ = S_T \mathbf{1}_{\{S_T>K\}} - K\mathbf{1}_{\{S_T>K\}}$ and the law of one price give $A = C + KD$, and $C^{\mathrm{BS}} + Ke^{-rT}N(d_-) = S_0 e^{-yT}N(d_+)$.

(ii) Differentiate $C' = \partial_K C^{\mathrm{BS}} + \partial_\sigma C^{\mathrm{BS}}\sigma'$ once more: $C'' = \partial_{KK}C^{\mathrm{BS}} + 2\,\partial_{K\sigma}C^{\mathrm{BS}}\sigma' + \partial_{\sigma\sigma}C^{\mathrm{BS}}\sigma'^2 + \mathcal{V}\sigma''$. From $\partial_K d_- = -1/(K\sigma\sqrt{T})$ and $\partial_\sigma d_- = -d_+/\sigma$,

$$\partial_{KK}C^{\mathrm{BS}} = \frac{e^{-rT}N'(d_-)}{K\sigma\sqrt{T}}, \qquad \partial_{K\sigma}C^{\mathrm{BS}} = \frac{e^{-rT}N'(d_-)\,d_+}{\sigma}, \qquad \partial_{\sigma\sigma}C^{\mathrm{BS}} = \frac{\mathcal{V}\,d_+d_-}{\sigma},$$

the last because $\mathcal{V} = Ke^{-rT}N'(d_-)\sqrt{T}$ and $\partial_\sigma N'(d_-) = N'(d_-)\,d_-d_+/\sigma$. Factoring out $e^{-rT}N'(d_-)/(K\sigma\sqrt{T})$ gives (3.9). Under (B) with $\sigma \in C^2(I)$, also $C \in C^2(I)$, and Theorem 3.5(iii) applies. □

The *skew term* $-\mathcal{V}\sigma'(K)$ in (3.8) is the error of pricing a digital with the Black–Scholes formula at the implied volatility of its own strike. For an equity skew, $\sigma' < 0$, so digital calls cost more and digital puts, $e^{-rT} - D(K)$, cost less than that formula says. Second-order terms enter the density: (3.9) contains the strike gamma, the cross derivative $\partial_{K\sigma}C^{\mathrm{BS}}$ (counted twice) and the volga. With $\sigma' = \sigma'' = 0$ the bracket equals 1, and (3.9) is the discounted lognormal density, the Green's function of Theorem 1.12.

Example 3.9 (The course's smile)**.** The course models the one-year index smile near the money by $\sigma(K) = \sigma_{\mathrm{atm}} + \xi\log(S_0/K)$, with $S_0 = 3900$, $T = 1$, $r = 3\%$, $y = 1.3\%$, $\sigma_{\mathrm{atm}} = 25\%$ and $\xi = 28\%$. At $K = S_0$, $\sigma'(K) = -\xi/K$, so the skew term is $\xi\, e^{-yT}N'(d_+)\sqrt{T}$ and

$$D(S_0) = e^{-rT}N(d_-) + \xi e^{-yT}N'(d_+)\sqrt{T} = 0.463167 + 0.108227 = 0.571394.$$

The skew term is 19% of the price; leaving out the factor e^{-yT}, a common slip, gives 0.572810. The asset-or-nothing call costs $A(S_0) = 413.04 + 3900 \times 0.571394 = 2641.47$, against 2219.39 with a flat 25% volatility. Figure 3.2 shows the digitals and the densities across strikes: the smile moves risk-neutral mass from the right tail to the left tail and shifts the mode to the right of the forward.

Testing a smile for arbitrage. A smile $\sigma : (0,\infty) \to (0,\infty)$ is arbitrage-free at one maturity when $C(K) = C^{\mathrm{BS}}(K, \sigma(K))$, with $C(0) = S_0 e^{-yT}$, satisfies the conditions of Theorem 3.5(iv). The conditions at $K = 0$ hold automatically: Black–Scholes prices satisfy $S_0e^{-yT} - Ke^{-rT} < C^{\mathrm{BS}} < S_0e^{-yT}$, so $C(K) \to C(0)$ and, once C is convex, $\partial^+C(0) \geq -e^{-rT}$. For a twice differentiable smile, what remains is that the bracket in (3.9) be nonnegative and that $C(K) \to 0$ as $K \to \infty$. Written in the log-moneyness $k = \log(K/F)$ and the total implied variance $w(k) = \sigma(K)^2T$, the bracket is the function

$$g(k) = \Big(1 - \frac{k\,w'(k)}{2w(k)}\Big)^2 - \frac{w'(k)^2}{4}\Big(\frac{1}{w(k)} + \frac{1}{4}\Big) + \frac{w''(k)}{2}$$

(Exercise 3.7), and a slice is free of butterfly arbitrage if and only if $g \geq 0$ and $d_+ \to -\infty$ as $k \to \infty$ (Gatheral and Jacquier (2014), eq. (2.1) and Lemma 2.2). Roper (2010), Theorems 2.9

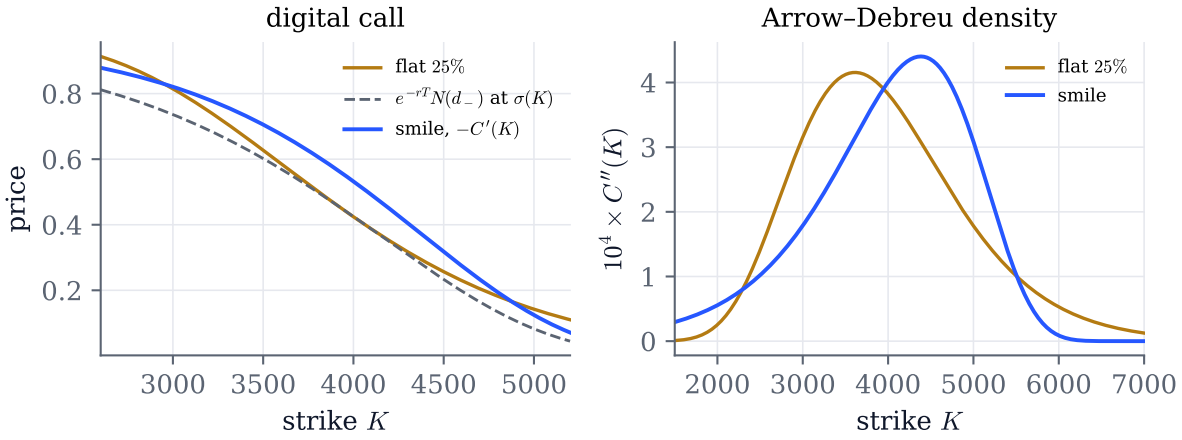

Figure 3.2: The course's smile $\sigma(K) = 25\% + 28\% \log(S_0/K)$ ($S_0 = 3900$, $T = 1$, $r = 3\%$, $y = 1.3\%$) against a flat 25% volatility. Left: digital call prices; the smile price $-C'(K)$ of (3.8), the Black–Scholes digital with a flat 25%, and the Black–Scholes digital $e^{-rT}N(d_-)$ at $\sigma(K)$, which omits the skew term. Right: the discounted density $C''(K)$ from (3.9) and the discounted lognormal density with $\sigma = 25\%$, both multiplied by 10^4.

and 2.15, shows for zero rates and dividends and for twice-differentiable positive smiles that these conditions on every slice, together with total variance nondecreasing in T and vanishing at $T = 0$, are sufficient and necessary for a surface free of static arbitrage; for constant r and y apply them to the calls on S_T/F_T, $F_T = S_0 e^{(r-y)T}$, whose prices are $C(K,T)/(S_0 e^{-yT})$ at the strikes K/F_T. Arbitrage also bounds the slope of a smile (Exercise 3.8).

The course's smile fails the test far from the money. Its bracket is positive for $84.33 < K < 9524.18$ and negative for $K < 84.33$, that is below 2.2% of the spot, where $\sigma = 132\%$: butterflies there have negative prices. Above $9524.18 = S_0 e^{\sigma_{\text{atm}}/\xi}$ the formula gives $\sigma \le 0$. The total variance $(\sigma_{\text{atm}} - \xi k')^2 T$, $k' = \log(K/S_0)$, grows like k'^2, while under any pricing measure $\limsup w(k)/|k| \le 2$ in each wing: Lee (2004)'s moment formula (his Theorems 3.2 and 3.4) identifies the limsup as an explicit function, with values in $[0, 2]$, of the highest finite moment of S_T (right wing) or of $1/S_T$ (left wing); see also Gatheral (2006), Chapter 7, "Extreme strikes". A near-the-money fit is not a smile on all of $(0, \infty)$, and digitals, densities and Chapter 4's replication integrals reach far into the wings.

A synthetic surface. The SSVI surface of Gatheral and Jacquier (2014) has the right shape and provably no arbitrage. With $\theta_T = \sigma_{\text{atm}}^2 T$ the at-the-money-forward total variance, $\rho \in (-1, 1)$ and a function $\phi > 0$, it sets

$$w(k,T) = \frac{\theta_T}{2}\Big(1 + \rho\,\phi(\theta_T)k + \sqrt{\big(\phi(\theta_T)k + \rho\big)^2 + 1 - \rho^2}\Big), \qquad \phi(\theta) = \frac{\eta}{\theta^\gamma (1+\theta)^{1-\gamma}} \tag{3.10}$$

(their Definition 4.1 and eq. (4.5)). We take the test case with $\sigma_{\text{atm}} = 20\%$, $\rho = -0.7$, $\eta = 1$, $\gamma = \frac{1}{2}$. By their Remark 4.4, $\eta(1 + |\rho|) \le 2$ (here 1.7) suffices for (4.5) to be free of static arbitrage; directly, their Theorem 4.2 excludes butterfly arbitrage if $\theta\phi(\theta)(1 + |\rho|) < 4$ and $\theta\phi(\theta)^2(1 + |\rho|) \le 4$ for all $\theta > 0$; here $\theta\phi(\theta) = \eta\sqrt{\theta/(1+\theta)} < 1$ and $\theta\phi(\theta)^2 = \eta^2/(1+\theta) < 1$, so

both left sides are below 1.7. Their Theorem 4.1 excludes calendar arbitrage if θ_T is nondecreasing and $0 \le \partial_\theta(\theta\phi) \le \phi\,(1+\sqrt{1-\rho^2})/\rho^2$; here $\partial_\theta(\theta\phi)/\phi = \frac{1}{2}/(1+\theta) \le \frac{1}{2}$, and the right bound is 3.50. The at-the-money skew is $\partial_k\sigma = \rho\,\phi(\theta_T)\sqrt{\theta_T}/(2\sqrt{T}) \approx \rho\eta/(2\sqrt{T})$ for small θ_T: it scales like $T^{-1/2}$, as in Theorem 3.7 (-0.697, -0.343 and -0.238 at $T = 0.25$, 1 and 2). At $T = 1$ and $K = 100$, $\sigma(K) = 19.657\%$, the call costs 7.23033 (against 7.36429 at a flat 20%), and the digital costs $0.436329 + 0.133830 = 0.570159$, against 0.436000 at a flat 20%; the density of S_T at 100 is 0.021125, against 0.019724 for the lognormal law.

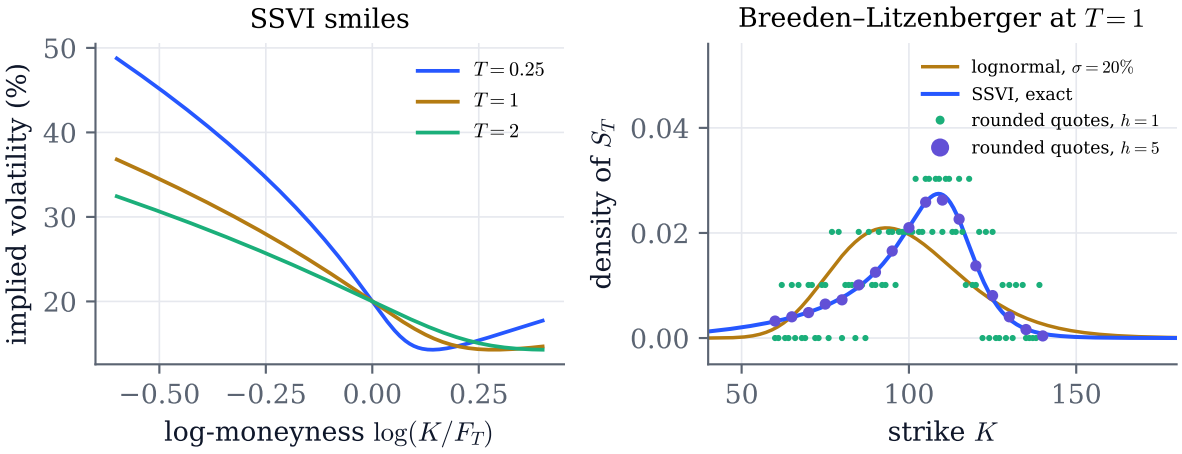

Figure 3.3: Left: implied volatilities of the SSVI surface (3.10) ($\sigma_{\text{atm}} = 20\%$, $\rho = -0.7$, $\eta = 1$, $\gamma = \frac{1}{2}$; test case) against $\log(K/F_T)$ for $T = 0.25$, 1, 2. Right: at $T = 1$, the density $e^{rT}C''(K)$ of S_T from (3.9), the lognormal density with $\sigma = 20\%$, and the estimates $e^{rT}\Delta_h^2\tilde{C}(K)$ from SSVI prices rounded to the cent, with $h = 1$ (strikes $60, 61, \dots, 140$) and $h = 5$ (strikes $60, 65, \dots, 140$).

▶ Interactive demo: change the SSVI parameters and the maturity, and compare the smile, the digital with and without its skew term, the density and Durrleman's function g; then find where the course's smile stops being arbitrage-free.

Digitals from call spreads. A digital is rarely listed, but call spreads are, and they bound it. Let $h > 0$ and let the strike grid be $h\mathbb{N} = \{0, h, 2h, \dots\}$; a *grid portfolio* holds bonds, the prepaid forward and calls with strikes in $h\mathbb{N}$.

Proposition 3.10 (Super- and sub-replication of a digital)**.** *Let $K \in h\mathbb{N}$ with $K \ge h$, and $\delta_K(x) = \mathbf{1}_{\{x>K\}}$.*

(i) Among the payoffs $f(x)$ of grid portfolios with $f \ge \delta_K$ on $[0,\infty)$, the smallest one pointwise is $f^+(x) = ((x-K+h)^+ - (x-K)^+)/h$; among those with $f \le \delta_K$ on $[0,\infty)$, the largest is $f^-(x) = ((x-K)^+ - (x-K-h)^+)/h$.

(ii) Under (P), the cheapest super-replicating grid portfolio costs $\bar{D}_h(K) = (C(K-h) - C(K))/h$ and the dearest sub-replicating one costs $\underline{D}_h(K) = (C(K) - C(K+h))/h$. If the digital is traded at the price $D(K)$ and (P) holds in the market that includes it, then $\underline{D}_h(K) \le D(K) \le \bar{D}_h(K)$.

(iii) Under (B), $\bar{D}_h(K) - \underline{D}_h(K) = h\,\Delta_h^2 C(K)$, which tends to $e^{-rT}\mathbb{Q}(S_T = K)$ as $h \downarrow 0$; if

the law of S_T has a density bounded by $\bar{q}$ on $[K-h, K+h]$, the gap is at most $e^{-rT}\bar{q}h$.

Proof. (i) The payoff f of a grid portfolio is continuous and affine on each interval $[jh, (j+1)h]$. If $f \geq \delta_K$, then $f \geq 1$ on (K, ∞), hence $f(K) \geq 1$ by continuity, and $f(K-h) \geq 0$; as f is affine on $[K-h, K]$, $f \geq f^+$ there, and $f \geq 0 = f^+$ on $[0, K-h]$ and $f \geq 1 = f^+$ on $[K, \infty)$. If $f \leq \delta_K$, then $f(K) \leq 0$ and $f(K+h) \leq 1$, so $f \leq f^-$ on $[K, K+h]$, and $f \leq 0 = f^-$ on $[0, K]$ and $f \leq 1 = f^-$ on $[K+h, \infty)$. Both $f^{\pm}$ are payoffs of call spreads on the grid. (ii) By (P), prices are monotone in payoffs, so the extremes of (i) are the cheapest and the dearest; they also bound the digital. (iii) The difference of the two spreads is h times the butterfly (3.6) divided by h^2; apply Theorem 3.5(ii)(a). With a density $q \leq \bar{q}$ on $[K-h, K+h]$, $h\,\Delta_h^2 C(K) = e^{-rT}\int q(x)\, h\Lambda_h(x-K)\,\mathrm{d}x \leq e^{-rT}\bar{q}\,h$, because Λ_h has integral 1. □

The gap is first order in h, and it does not close at an atom. For the SSVI smile at $K = 100$, the bounds are [0.5138, 0.6184] with $h = 5$ and [0.5595, 0.5804] with $h = 1$, around $D(100) = 0.5702$; Exercise 3.5 does the computation for the course's smile on its listed strike grids. A dealer who sells the digital at $\bar{D}_h(K)$ and buys the spread carries no risk at all, whatever the law of S_T; the price of this safety is the gap.

In code

The module is `code/src/mfe230d/smile.py`. There, `strike_checks` tests quotes against (3.4); `second_difference`, `fd_density`, `fd_density_error_bound` and `fd_optimal_step` implement Theorem 3.6; `smile_digital`, `smile_asset_digital` and `smile_density` implement Theorem 3.8; `course_smile`, `ssvi_smile`, `ssvi_conditions` and `durrleman_g` are the two synthetic smiles and their tests; `call_spread_bounds` is Theorem 3.10; `atom_call` prices calls on a law with an atom. Black–Scholes prices and implied volatilities come from `code/src/mfe230d/bs.py`, and `code/tests/test_smile.py` checks every number quoted in this chapter.

Further reading

Breeden and Litzenberger (1978) is the original source of the density formula, and Merton (1973) of the strike relations. Gatheral (2006) is the practitioner's account of the volatility surface; its Chapter 8 discusses digital options under a skew. Rubinstein (1994) reads risk-neutral laws from index options and builds trees that fit them. Davis and Hobson (2007), Carr and Madan (2005) and Roper (2010) characterize prices and surfaces free of static arbitrage, Lee (2004) bounds the wings of a smile, and Gatheral (2004) and Gatheral and Jacquier (2014) give the SVI and SSVI parameterizations.

Chapter 4

Static replication and the Kolmogorov equations

Chapter 3 read the risk-neutral law of S_T off the prices of calls across strikes. This chapter runs the argument the other way. A European payoff $f(S_T)$ is a portfolio of bonds, forwards, puts, calls and digitals, bought today and held to maturity: each kink of f is an option at one strike, each jump a digital, and the curvature a continuum of options with weights $f''(K)\,\mathrm{d}K$. This is the Carr–Madan formula. Its price uses no model, only option prices, provided the strip of options has a finite price (Section 4.1).

The rest of the chapter exploits structure. In the standard model a power S^p of the stock is again a geometric Brownian motion, so payoffs written in terms of S^p, or S^p times a payoff of S, are priced by the Black–Scholes formula after a change of inputs, which the course calls the *pre-wash* (Section 4.2). When a price does not depend on the unit of account, Euler's theorem on homogeneous functions turns this invariance into identities between Greeks: the hedge splits into its parts, the gammas sum to zero, and the spot gamma is the risk-neutral density at the strike (Section 4.3). Finally, the Arrow–Debreu price density of a diffusion solves two equations: the pricing equation in its initial point, and the adjoint equation in its final point (Section 4.4). The first prices one payoff for every spot; the second prices every payoff, all strikes at once, for one spot.

Setting. As in Chapter 3, $T > 0$ is a fixed maturity, $r, y \in \mathbb{R}$, $S_T \geq 0$, and $\mathbb{Q}$ is a pricing measure: every traded payoff X costs $e^{-rT}\mathbb{E}^{\mathbb{Q}}[X]$ at time 0. The bond paying 1 costs e^{-rT}, the prepaid forward paying S_T costs $S_0 e^{-yT}$, and $F = S_0 e^{(r-y)T} = \mathbb{E}^{\mathbb{Q}}[S_T] < \infty$. Calls, puts and digital calls with strike K cost

$$C(K) = e^{-rT}\mathbb{E}^{\mathbb{Q}}[(S_T - K)^+], \qquad P(K) = e^{-rT}\mathbb{E}^{\mathbb{Q}}[(K - S_T)^+], \qquad D(K) = e^{-rT}\mathbb{Q}(S_T > K),$$

and the digital paying $\mathbf{1}_{\{S_T \geq K\}}$ costs $\bar{D}(K) = e^{-rT}\mathbb{Q}(S_T \geq K)$. By Theorem 3.5(i), $D(K) = -\partial^+ C(K)$ and $\bar{D}(K) = -\partial^- C(K)$: the two differ only if S_T has an atom at K. A payoff built from infinitely many options is priced, like any other payoff, by $e^{-rT}\mathbb{E}^{\mathbb{Q}}$; it equals the superposition of the option prices when Fubini's theorem allows the exchange of $\mathbb{E}^{\mathbb{Q}}$ and the integral over strikes. Sections 4.2 and 4.3 work in the standard model of Theorem 1.1.

4.1 Static replication

A payoff that is piecewise linear in S_T (a call spread, a collar, a range forward, most capital-protected notes sold to retail investors) is a finite portfolio of bonds, shares and calls, with one call at each kink weighted by the change of slope there; the IEOR 222 notes prove this. A general payoff is a limit of such payoffs, in which the changes of slope become $f''(K)\,\mathrm{d}K$. The natural class is that of differences of convex functions, which contains the twice differentiable payoffs, the payoffs with kinks, and their sums.

The second derivative as a measure. Let f be convex on $(0,\infty)$. At every point f has finite one-sided derivatives $f'_- \le f'_+$, both nondecreasing; f'_+ is right-continuous; and f is absolutely continuous on compact intervals, with $f(b)-f(a)=\int_a^b f'_+(u)\,\mathrm{d}u$ (Revuz and Yor (1999), Appendix, Section 3). A nondecreasing right-continuous function is the distribution function of a unique measure (Durrett (2019), Section 1.1, applied on $[\varepsilon,\infty)$ for each $\varepsilon>0$), so there is a unique measure μ_f on $(0,\infty)$, finite on compact subintervals, with $\mu_f((a,b])=f'_+(b)-f'_+(a)$ for $0<a<b$; in particular $\mu_f(\{k\})=f'_+(k)-f'_-(k)$. If $f=f_1-f_2$ with f_1,f_2 convex, we set $\mu_f=\mu_{f_1}-\mu_{f_2}$, a signed measure that is finite on compact subintervals and depends on f only through f'_+. We write $f''(\mathrm{d}K)$ for $\mu_f(\mathrm{d}K)$ and $|f''|(\mathrm{d}K)$ for its total variation. If $f\in C^1$ and f' is absolutely continuous on compact intervals (for instance $f\in C^2$), then f is a difference of convex functions, since $f'=f'(1)+\int_1^{\cdot}(f'')^+-\int_1^{\cdot}(f'')^-$, and $f''(\mathrm{d}K)=f''(K)\,\mathrm{d}K$. A kink is an atom: $(x-k)^+$ and $(k-x)^+$ both have second derivative δ_k, the unit mass at k. So a payoff that is C^2 on each of the closed intervals $(0,k_1],[k_1,k_2],\dots,[k_n,\infty)$ (one-sided derivatives at the ends), with kinks at $k_1<\cdots<k_n$, has $f''(\mathrm{d}K)=f''(K)\,\mathrm{d}K+\sum_i\big(f'(k_i+)-f'(k_i-)\big)\delta_{k_i}(\mathrm{d}K)$.

Theorem 4.1 (Carr–Madan)**.** *Let $g:(0,\infty)\to\mathbb{R}$ be a difference of convex functions and $\kappa>0$.*

(i) (Spanning) *For every $x>0$,*

$$g(x)=g(\kappa)+g'_+(\kappa)(x-\kappa)+\int_{(0,\kappa]}(K-x)^+\,g''(\mathrm{d}K)+\int_{(\kappa,\infty)}(x-K)^+\,g''(\mathrm{d}K). \tag{4.1}$$

For each x at most one of the integrals is nonzero, and it extends over the bounded interval between x and κ.

(ii) (Jumps) *Let $f=g+\sum_{j=1}^m J_jH_j$ with $J_j\in\mathbb{R}$ and $H_j=\mathbf{1}_{(a_j,\infty)}$ or $H_j=\mathbf{1}_{[a_j,\infty)}$, $a_j>0$. Then $f(x)$ is the right side of (4.1) plus $\sum_j J_jH_j(x)$, for every $x>0$.*

(iii) (Edges) *(a) If $\int_{(0,\kappa]}K\,|g''|(\mathrm{d}K)<\infty$, then $g(0+)=\lim_{x\downarrow0}g(x)$ exists, and (4.1) holds at $x=0$ with $g(0)=g(0+)$. (b) If $|g''|((0,\kappa])<\infty$, then $g(0+)$ and $g'_+(0+)$ exist and, for every $x\ge0$, $g(x)=g(0+)+g'_+(0+)\,x+\int_{(0,\infty)}(x-K)^+\,g''(\mathrm{d}K)$: bonds, shares and calls. (c) If $\int_{[\kappa,\infty)}K\,|g''|(\mathrm{d}K)<\infty$, then $B=\lim_{k\to\infty}g'_+(k)$ and $A=\lim_{k\to\infty}\big(g(k)-k\,g'_+(k)\big)$ exist and, for every $x>0$, $g(x)=A+Bx+\int_{(0,\infty)}(K-x)^+\,g''(\mathrm{d}K)$: bonds, shares and puts.*

(iv) (Prices) *Let f be as in (ii), and assume*

$$\int_{(0,\kappa]}P(K)\,|g''|(\mathrm{d}K)+\int_{(\kappa,\infty)}C(K)\,|g''|(\mathrm{d}K)<\infty, \tag{4.2}$$

and $f(0) = g(0+)$ if $\mathbb{Q}(S_T = 0) > 0$. Then $f(S_T)$ is $\mathbb{Q}$-integrable and its price is

$$e^{-rT} g(\kappa) + g'_+(\kappa)(S_0 e^{-yT} - \kappa e^{-rT}) + \int_{(0,\kappa]} P(K)\, g''(\mathrm{d}K) + \int_{(\kappa,\infty)} C(K)\, g''(\mathrm{d}K) + \sum_{j=1}^{m} J_j D_j, \tag{4.3}$$

with $D_j = D(a_j)$ if $H_j = \mathbf{1}_{(a_j,\infty)}$ and $D_j = \bar{D}(a_j)$ otherwise. If g is convex or concave and $\mathbb{Q}(S_T = 0) = 0$, (4.2) is also necessary for $f(S_T)$ to be integrable.

Proof. (i) This is Taylor's formula with integral remainder, with the remainder written in options. Let $x > \kappa$. For $K \in (\kappa, x]$, $x - K = \int_\kappa^x \mathbf{1}_{\{u>K\}}\, \mathrm{d}u$, and Fubini's theorem (the integrand is bounded and $|g''|((\kappa, x]) < \infty$) gives

$$\int_{(\kappa,\infty)} (x-K)^+\, g''(\mathrm{d}K) = \int_{(\kappa,x]} (x-K)\, g''(\mathrm{d}K) = \int_\kappa^x g''((\kappa,u))\, \mathrm{d}u = \int_\kappa^x \left(g'_+(u) - g'_+(\kappa)\right) \mathrm{d}u.$$

In the last step $g''((\kappa,u)) = g''((\kappa,u])$ except at the countably many atoms of g'', a Lebesgue null set. The right side is $g(x) - g(\kappa) - g'_+(\kappa)(x-\kappa)$, and the put integral vanishes because $(K-x)^+ = 0$ for $K \le \kappa < x$. For $x < \kappa$, in the same way, $\int_{(0,\kappa]} (K-x)^+ g''(\mathrm{d}K) = \int_{(x,\kappa]} \int_x^\kappa \mathbf{1}_{\{u<K\}}\, \mathrm{d}u\, g''(\mathrm{d}K) = \int_x^\kappa \left(g'_+(\kappa) - g'_+(u)\right) \mathrm{d}u = g'_+(\kappa)(\kappa - x) - g(\kappa) + g(x)$, and the call integral vanishes. For $x = \kappa$ both integrals vanish.

(ii) Apply (i) to g and add the steps.

(iii) (a) For $0 < x < \kappa$, (4.1) reads $g(x) = g(\kappa) + g'_+(\kappa)(x-\kappa) + \int_{(0,\kappa]} (K-x)^+ g''(\mathrm{d}K)$. As $x \downarrow 0$ the integrand converges to K and is bounded by K, which is $|g''|$-integrable on $(0,\kappa]$; dominated convergence gives the limit, which is the right side of (4.1) at $x = 0$. (b) g'_+ has bounded variation on $(0,\kappa]$, so $g'_+(0+)$ exists, g'_+ is bounded near 0, and $g(0+)$ exists. Fix $x > 0$ and let $\kappa' \downarrow 0$ in (4.1) written with κ' in place of κ: $g(\kappa') \to g(0+)$, $g'_+(\kappa') \to g'_+(0+)$, the put integral is at most $\kappa' |g''|((0,\kappa']) \to 0$, and the call integral converges by dominated convergence, since $(x-K)^+ \le x \mathbf{1}_{\{K<x\}}$ and $|g''|((0,x)) < \infty$. At $x = 0$ the identity reads $g(0+) = g(0+)$. (c) The hypothesis gives $|g''|([\kappa,\infty)) < \infty$, so B exists. For $k > \kappa$, (i) at $x = k$ and $g'_+(k) = g'_+(\kappa) + g''((\kappa,k])$ give $g(k) - k g'_+(k) = g(\kappa) - \kappa g'_+(\kappa) - \int_{(\kappa,k]} K\, g''(\mathrm{d}K)$, which converges. Fix $x > 0$ and let $k \to \infty$ in (4.1) written with k in place of κ: the call integral vanishes as soon as $k > x$; $g(k) + g'_+(k)(x-k) \to A + Bx$; and the put integral converges by dominated convergence, with the bound $(K-x)^+ \le K$ on $[\kappa,\infty)$.

(iv) Write $X = S_T$. By Tonelli's theorem,

$$\mathbb{E}^{\mathbb{Q}}\Big[\int_{(0,\kappa]} (K-X)^+\, |g''|(\mathrm{d}K)\Big] = e^{rT} \int_{(0,\kappa]} P(K)\, |g''|(\mathrm{d}K) < \infty,$$

and likewise for the call integral. So both integrals in (4.1), evaluated at $x = X$, are integrable, and by Fubini's theorem their expectations are e^{rT} times the strike integrals in (4.3). The terms $g(\kappa) + g'_+(\kappa)(X-\kappa)$ are integrable because $\mathbb{E}^{\mathbb{Q}}[X] = F < \infty$, and the steps are bounded. On $\{X > 0\}$, (ii) expresses $f(X)$ as the sum of these terms. If $\mathbb{Q}(X = 0) > 0$, then $P(K) \ge e^{-rT} K\, \mathbb{Q}(X = 0)$, so (4.2) gives the condition of (iii)(a), and on $\{X = 0\}$ we use (iii)(a). Taking expectations and multiplying by e^{-rT} gives (4.3), with $e^{-rT}\mathbb{E}^{\mathbb{Q}}[g'_+(\kappa)(X-\kappa)] = g'_+(\kappa)(S_0 e^{-yT} - \kappa e^{-rT})$. If g is convex, both integrals in (4.1) are nonnegative, and almost surely their sum is $f(X)$ minus an integrable variable; if $f(X)$ is integrable, so is the sum, and by Tonelli's theorem its expectation is e^{rT} times the left side of (4.2). For concave g, apply this to $-f$. □

Remark 4.2 (Reading the formula).

(a) *The portfolio.* The payoff is replicated by $g(\kappa)$ bonds, $g'_+(\kappa)$ forwards struck at κ, puts with strikes $K \le \kappa$ and calls with strikes $K > \kappa$ in the quantities $g''(\mathrm{d}K)$, and J_j digitals. The options are all out of the money with respect to κ. With $\kappa = F$ the forwards cost nothing. An atom of g'' at κ is held as a put; using $g'_-(\kappa)$ instead of $g'_+(\kappa)$ moves it to the calls, and put–call parity shows that the two portfolios agree.

(b) *Payoffs that vanish on one side.* If $g = 0$ on $(0, \kappa]$, then $g'_+(\kappa) = g''(\{\kappa\})$, and the forward together with the put at κ is a call:

$$g(x) = g'_+(\kappa)\,(x-\kappa)^+ + \int_{(\kappa,\infty)} (x-K)^+\, g''(\mathrm{d}K), \qquad x > 0. \tag{4.4}$$

Symmetrically, if $g = 0$ on $[\kappa, \infty)$, then $g(x) = -g'_-(\kappa)(\kappa - x)^+ + \int_{(0,\kappa)}(K-x)^+ g''(\mathrm{d}K)$.

(c) *Integrability.* (4.2) holds in every model if $|g''|$ has finite mass near 0 and ∞ ($P(K) \le e^{-rT}K$, C is nonincreasing); otherwise it depends on the tails of S_T (Exercise 4.2).

Example 4.3 (The square-root call)**.** The course replicates $f(x) = (\sqrt{x} - \sqrt{K})^+$. It vanishes on $(0, K]$, has the right derivative $1/(2\sqrt{K})$ at the kink, and is concave above it, with $f''(\mathrm{d}k) = \frac{1}{2\sqrt{K}}\delta_K(\mathrm{d}k) - \frac{1}{4}k^{-3/2}\mathbf{1}_{\{k>K\}}\mathrm{d}k$. By (4.4),

$$(\sqrt{x} - \sqrt{K})^+ = \frac{(x-K)^+}{2\sqrt{K}} - \frac{1}{4}\int_K^\infty k^{-3/2}\,(x-k)^+\,\mathrm{d}k\,:$$

the tangent call at the kink, minus a strip of calls with higher strikes. Since $C(k) \le C(K)$ for $k \ge K$, the strip costs at most $C(K)/(2\sqrt{K})$, so (4.2) holds in every model, and

$$\text{price} = \frac{C(K)}{2\sqrt{K}} - \frac{1}{4}\int_K^\infty k^{-3/2}\,C(k)\,\mathrm{d}k.$$

With the course's inputs $S_0 = K = 10^4$, $T = 1$, $r = 1\%$, $y = -1\%$, $\sigma = 40\%$ and Black–Scholes call prices, the tangent call costs 8.43615 and the strip 1.07186 (Gauss–Legendre quadrature in $\log k$), so the square-root call costs 7.36429. That this is the test-case price is no accident (Theorem 4.5).

Finitely many strikes. Markets list finitely many strikes $k_0 < k_1 < \cdots < k_m$. The continuous piecewise-linear interpolant f_h of f at these strikes, extended linearly beyond them, is a finite portfolio: bonds, shares, and calls at $k_1, \dots, k_{m-1}$, weighted by the changes of slope. For spacing h the weight at k_i is $(f(k_{i+1}) - 2f(k_i) + f(k_{i-1}))/h \approx f''(k_i)\,h$, a Riemann sum for the strip. Where f is C^2, $|f - f_h| \le \frac{h^2}{8}\max|f''|$ between neighbouring strikes (the IEOR 222 notes prove this), so the price error is of order h^2, plus whatever the linear extrapolation outside $[k_0, k_m]$ costs. A jump behaves differently, because no continuous interpolant reproduces it. If f jumps by $J > 0$ at a listed strike a, the interpolant replaces $J\mathbf{1}_{[a,\infty)}$ by a call spread rising from 0 at $a - h$ to J at a, and the price error is

$$J\,e^{-rT}\mathbb{E}^{\mathbb{Q}}\Big[\Big(1 - \frac{a - S_T}{h}\Big)\mathbf{1}_{\{a-h<S_T<a\}}\Big] = Jh\int_0^1 (1-u)\,G(a-hu)\,\mathrm{d}u = \tfrac{1}{2}JhG(a) + o(h), \tag{4.5}$$

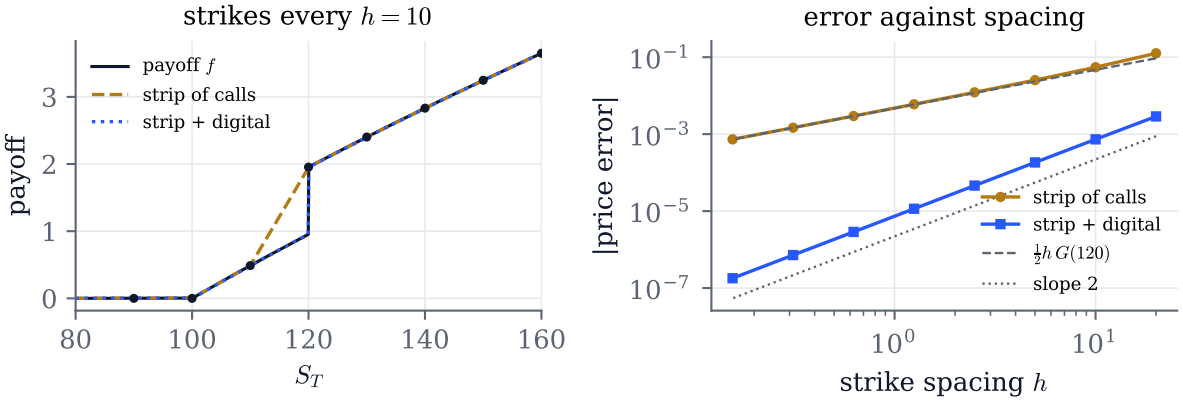

Figure 4.1: Static replication of a square-root call with a digital bonus, $f(x) = (\sqrt{x} - 10)^+ + \mathbf{1}_{\{x \geq 120\}}$, in the test case ($S_0 = 100$, $T = 1$, $r = 1\%$, $y = 2\%$, $\sigma = 20\%$), where it costs 0.488321. Left: f and two portfolios on the strikes $90, 100, \dots, 400$: the interpolant of f, in which the jump becomes a call spread on $[110, 120]$, and the interpolant of the square-root call plus one digital. Right: absolute price error of the two portfolios against the spacing h of the strikes $100 - h, 100, \dots, 400$, with the first-order prediction $\frac{1}{2}hG(120)$ of (4.5) (dashed) and a line of slope 2 (dotted).

if S_T has a density q near a; here $G = e^{-rT}q$ is the Arrow–Debreu price density (Theorem 1.12), and the last step assumes that G is continuous at a. The error is of first order in h. Holding the digital itself, as in Theorem 4.1(ii), restores the second order (Figure 4.1).

▶ Interactive demo: replicate a payoff with kinks and jumps on a strike grid of your choice, and watch the price error with and without the digitals.

4.2 Power payoffs and the pre-wash

In this section and the next the stock follows the standard model of Theorem 1.1: under $\mathbb{Q}$, $S_t = S_0 \exp(\nu t + \sigma W_t)$ with $\nu = r - y - \frac{\sigma^2}{2}$ and W a $\mathbb{Q}$-Brownian motion. For $t < T$ we write $\tau = T - t$, and $\mathrm{BS}(x, K, \tau; r, y, \sigma)$ and $\mathrm{BSP}(x, K, \tau; r, y, \sigma)$ for the call and put prices of Theorem 2.1 at spot x. For a Borel g, $v_g(t, x; r, y, \sigma)$ is the price function (1.8) of the payoff $g(S_T)$ in the standard model with inputs r, y, σ. For $p \in \mathbb{R}$ let

$$y_p = r - p\,(r - y) - \tfrac{1}{2}p(p-1)\sigma^2, \qquad \sigma_p = |p|\,\sigma. \tag{4.6}$$

Proposition 4.4 (Power payoffs and the pre-wash)**.** *In the standard model, let $p \in \mathbb{R}$.*

(i) The process S^p satisfies $S^p_t = S^p_0 \exp\big((r - y_p - \frac{1}{2}\sigma_p^2)t + \sigma_p W^{(p)}_t\big)$, where $W^{(p)} = W$ if $p \geq 0$ and $W^{(p)} = -W$ if $p < 0$ is a $\mathbb{Q}$-Brownian motion. So S^p moves like the stock of a standard model with rate r, yield y_p and volatility σ_p, and $\mathbb{E}^{\mathbb{Q}}[S^p_T \mid \mathcal{F}_t] = S^p_t e^{(r - y_p)\tau}$.

*(ii) (*Pre-wash*) Let $p \neq 0$ and let g be Borel with $\mathbb{E}^{\mathbb{Q}}|g(S^p_T)| < \infty$. The price at time t of $g(S^p_T)$ is $v_g(t, S^p_t; r, y_p, \sigma_p)$. In particular $(S^p_T - K)^+$ and $(K - S^p_T)^+$ cost*

$$\mathrm{BS}(S^p_t, K, \tau; r, y_p, \sigma_p) \qquad \text{and} \qquad \mathrm{BSP}(S^p_t, K, \tau; r, y_p, \sigma_p).$$

*(iii) (*Leveraged payoffs*) Let g be Borel with $\mathbb{E}^{\mathbb{Q}}|S_T^p g(S_T)| < \infty$, and let*

$$r'_p = y_p, \qquad y'_p = y - p(r - y + \tfrac{1}{2}(p+1)\sigma^2). \tag{4.7}$$

The price at time t of $S_T^p g(S_T)$ is $S_t^p\, v_g(t, S_t; r'_p, y'_p, \sigma)$. In particular $S_T^p (S_T - K)^+$ costs $S_t^p\, \mathrm{BS}(S_t, K, \tau; r'_p, y'_p, \sigma)$.

At the edges: $p = 1$ gives $y_1 = y$ and $\sigma_1 = \sigma$ in (ii), and $p = 0$ gives $r'_0 = r$ and $y'_0 = y$ in (iii); for $p = 0$ in (ii) the payoff is the constant $g(1)$, worth $e^{-r\tau} g(1)$, which is what the Black–Scholes formula gives as $\sigma_p \downarrow 0$ with forward 1. For $K = 0$, (ii) prices S_T^p itself at $S_t^p e^{-y_p \tau}$. For $p < 0$ the call $(S_T^p - K)^+$ pays when S_T is small: it is a put-like payoff on S.

Proof. (i) $S_t^p = S_0^p \exp(p\nu t + p\sigma W_t)$, with $p\sigma W_t = \sigma_p W_t^{(p)}$ and $p\nu = p(r-y) - \frac{1}{2}p\sigma^2 = r - y_p - \frac{1}{2}\sigma_p^2$ by (4.6). $-W$ is a Brownian motion, and (1.5) with $q = p$ gives the conditional mean.

(ii) Write $S_T^p = S_t^p R$ with $R = \exp\left(p\nu\tau + \sigma_p(W_T^{(p)} - W_t^{(p)})\right)$, independent of $\mathcal{F}_t$ and distributed as $\exp\left((r - y_p - \frac{1}{2}\sigma_p^2)\tau + \sigma_p\sqrt{\tau}Z\right)$, $Z \sim \mathcal{N}(0,1)$. This is the law of the gross return of the stock with inputs (r, y_p, σ_p), and the proof of Theorem 1.3(i), with S^p in place of S, gives $e^{-r\tau}\mathbb{E}^{\mathbb{Q}}[g(S_T^p) \mid \mathcal{F}_t] = v_g(t, S_t^p; r, y_p, \sigma_p)$.

(iii) Let $M_t = \exp(p\sigma W_t - \frac{1}{2}p^2\sigma^2 t)$, a positive $\mathbb{Q}$-martingale with $M_0 = 1$, and let $\mathrm{d}\mathbb{Q}^p = M_T\,\mathrm{d}\mathbb{Q}$ on $\mathcal{F}_T$. Since $p\nu + \frac{1}{2}p^2\sigma^2 = r - y_p$,

$$S_T^p = S_t^p\, e^{(r-y_p)\tau}\, \frac{M_T}{M_t}.$$

By Bayes' rule for conditional expectations (Shreve (2004), Lemma 5.2.2), $\mathbb{E}^{\mathbb{Q}}[M_T X \mid \mathcal{F}_t] = M_t\, \mathbb{E}^{\mathbb{Q}^p}[X \mid \mathcal{F}_t]$ for every $X \geq 0$, or X with $\mathbb{E}^{\mathbb{Q}}|M_T X| < \infty$. Hence $e^{-r\tau}\mathbb{E}^{\mathbb{Q}}[S_T^p g(S_T) \mid \mathcal{F}_t] = S_t^p\, e^{-y_p\tau}\, \mathbb{E}^{\mathbb{Q}^p}[g(S_T) \mid \mathcal{F}_t]$. By Girsanov's theorem (Karatzas and Shreve (1991), Theorem 3.5.1), $W_t^p = W_t - p\sigma t$ is a $\mathbb{Q}^p$-Brownian motion, and $S_T = S_t \exp\left((\nu + p\sigma^2)\tau + \sigma(W_T^p - W_t^p)\right)$. Under $\mathbb{Q}^p$ the stock therefore moves like that of a standard model with volatility σ and any rate and yield whose difference is $r - y + p\sigma^2$, in particular $r'_p = y_p$ and $y'_p = y_p - (r - y + p\sigma^2)$, which is (4.7). With that rate, $e^{-y_p\tau}\mathbb{E}^{\mathbb{Q}^p}[g(S_T) \mid \mathcal{F}_t] = e^{-r'_p\tau}\mathbb{E}^{\mathbb{Q}^p}[g(S_T) \mid \mathcal{F}_t] = v_g(t, S_t; r'_p, y'_p, \sigma)$, by the argument of (ii). □

The point is practical. A Black–Scholes routine discounts at its rate and grows the forward at rate minus yield; for a power payoff the yield y_p is chosen so that the routine finds the forward $S_t^p e^{(r-y_p)\tau}$ while still discounting at r, and a leveraged payoff also changes the discount rate. Feeding these inputs to the routine is the pre-wash; the same device prices geometric averages (Theorem 15.2) and exchange options (Theorem 17.6).

Example 4.5 (The square-root call, again)**.** For $p = \frac{1}{2}$, $\sigma_{1/2} = \sigma/2$ and $y_{1/2} = \frac{1}{2}(r + y + \frac{1}{4}\sigma^2)$, so by Theorem 4.4(ii) the call $(\sqrt{S_T} - \sqrt{K})^+$ costs $\mathrm{BS}(\sqrt{S_0}, \sqrt{K}, T; r, \frac{1}{2}(r + y + \frac{1}{4}\sigma^2), \frac{\sigma}{2})$. With the inputs of Theorem 4.3, $\sqrt{S_0} = \sqrt{K} = 100$, $\sigma/2 = 20\%$ and $y_{1/2} = \frac{1}{2}(0 + 0.04) = 2\%$: the call on $\sqrt{S}$ is exactly the test case, and costs 7.36429, as the strip of Theorem 4.3 found. Chapter 8 uses this payoff as a test bed for control variates (Theorem 8.13).

Post-wash. The Greeks returned by the routine are derivatives with respect to the washed inputs. The chain rule converts them back: with primes for the Greeks at $(S^p, K, \tau; r, y_p, \sigma_p)$,

$$\Delta = pS^{p-1}\Delta', \qquad \Gamma = p(p-1)S^{p-2}\Delta' + p^2 S^{2p-2}\Gamma',$$
$$\mathcal{V} = |p|\mathcal{V}' - p(p-1)\sigma\rho'_y, \qquad \rho_r = \rho'_r + (1-p)\rho'_y, \qquad \rho_y = p\rho'_y,$$

and $\Theta = \Theta'$, since $\partial y_p/\partial\sigma = -p(p-1)\sigma$, $\partial y_p/\partial r = 1-p$ and $\partial y_p/\partial y = p$.

The share measure $\mathbb{Q}^1$ of the proof, under which the stock is the numeraire, also gives a symmetry between calls and puts.

> **Corollary 4.6** (Put–call symmetry)**.** *In the standard model, for $t < T$ and $K > 0$, with $F_t = S_t e^{(r-y)\tau}$,*
>
> $$\mathrm{BS}(S_t, K, \tau; r, y, \sigma) = \frac{K}{F_t}\,\mathrm{BSP}\Big(S_t, \frac{F_t^2}{K}, \tau; r, y, \sigma\Big).$$
>
> *A call struck at K is worth K/F_t puts struck at the reflected strike F_t^2/K, the reflection of K in F_t on a logarithmic scale.*

Proof. Let $X = S_T/F_t$. Given $\mathcal{F}_t$, $\log X \sim \mathcal{N}(-\frac{1}{2}\sigma^2\tau, \sigma^2\tau)$ under $\mathbb{Q}$ and, by the proof of Theorem 4.4(iii) with $p = 1$, $\log X \sim \mathcal{N}(+\frac{1}{2}\sigma^2\tau, \sigma^2\tau)$ under $\mathbb{Q}^1$, where $\mathbb{E}^{\mathbb{Q}}[M_T\varphi \mid \mathcal{F}_t] = M_t\mathbb{E}^{\mathbb{Q}^1}[\varphi \mid \mathcal{F}_t]$ and $M_T/M_t = X$. So $1/X$ has under $\mathbb{Q}^1$ the law that X has under $\mathbb{Q}$, and

$$\mathbb{E}^{\mathbb{Q}}[(S_T-K)^+ \mid \mathcal{F}_t] = F_t\,\mathbb{E}^{\mathbb{Q}^1}\Big[\Big(1 - \frac{K}{F_t}\frac{1}{X}\Big)^+\Big|\mathcal{F}_t\Big] = F_t\,\mathbb{E}^{\mathbb{Q}}\Big[\Big(1 - \frac{K}{F_t}X\Big)^+\Big|\mathcal{F}_t\Big]$$
$$= \frac{K}{F_t}\,\mathbb{E}^{\mathbb{Q}}\Big[\Big(\frac{F_t^2}{K} - S_T\Big)^+\Big|\mathcal{F}_t\Big].$$

Multiply by $e^{-r\tau}$. □

Put–call symmetry is the basis of the static hedges of barrier options of Bowie and Carr (1994) and Carr and Chou (1997). If $r = y$, the forward equals the spot, and at the moment the stock touches a barrier H a call struck at K is worth K/H puts struck at H^2/K. For a down-and-in call with $0 < H \le K$ and $S_0 > H$, these puts are struck at $H^2/K \le H$, so they expire worthless if the barrier is never touched, and holding them and exchanging them for the call, at no cost, when it is touched replicates the option (Exercise 4.7); for $K < H$ the hedge needs further pieces. With $r \neq y$, Carr and Chou (1997) apply the symmetry to a power of S with zero drift (Carr and Lee (2009), Example 6.11), Carr et al. (1998) bound the price by hedges with flat barriers on the forward, and Carr and Lee (2009) extend put–call symmetry beyond the standard model. Chapter 17 develops changes of numeraire in general.

▶ Interactive demo: price $(S_T^p - K)^+$ for any power p by the pre-wash, and compare with a Monte Carlo estimate from the same model.

4.3 Homogeneity and the Greeks' conservation laws

Changing the unit of account, from dollars to cents say, multiplies every price and every input measured in money by the same factor $\lambda = 100$; rates, yields, volatilities and times are unchanged.

If a price is written as a function $V(z)$ of its money-valued inputs $z = (z_1, \dots, z_m)$ (spot, strike, barrier, cash dividend, ...), this says $V(\lambda z) = \lambda V(z)$. In the standard model the only money-valued inputs of a call are S and K, so $V(\lambda S, \lambda K) = \lambda V(S, K)$. A set $U \subseteq \mathbb{R}^m$ is a *cone* if $\lambda z \in U$ for all $z \in U$ and $\lambda > 0$, and $V : U \to \mathbb{R}$ is *positively homogeneous of degree* $k \in \mathbb{R}$ if $V(\lambda z) = \lambda^k V(z)$ for all $z \in U$ and $\lambda > 0$.

Proposition 4.7 (Homogeneity and conservation laws)**.** *Let $U \subseteq \mathbb{R}^m$ be an open cone and $V : U \to \mathbb{R}$.*

*(i) (*Euler's theorem*) If V is differentiable and positively homogeneous of degree k, then*

$$\sum_{i=1}^{m} z_i \, \partial_i V(z) = k \, V(z) \qquad \text{for all } z \in U. \tag{4.8}$$

Conversely, if V is differentiable on U and satisfies (4.8), then V is positively homogeneous of degree k.

*(ii) (*Gamma sum rule*) Let V be differentiable and homogeneous of degree 1. Each $\partial_j V$ is homogeneous of degree 0, and if $\partial_j V$ is differentiable, $\sum_i z_i \, \partial_i \partial_j V(z) = 0$. If $V \in C^2(U)$, the Hessian of V at z is symmetric and maps z to 0; in particular it is singular.*

*(iii) (*Conservation laws*) Let $V(z; \eta)$ depend on a real parameter η (a time, a volatility, a rate) and be differentiable and homogeneous of degree 1 in z for each η. If each $\partial_i V(z; \cdot)$ is differentiable at η, then $V(z; \cdot)$ is too, and*

$$\partial_\eta V(z; \eta) = \sum_{i=1}^{m} z_i \, \partial_\eta \partial_i V(z; \eta). \tag{4.9}$$

*(iv) (*Spot and strike*) If $V(S, K)$ is homogeneous of degree 1 and C^2 on $(0, \infty)^2$, then*

$$S \, \partial_S V + K \, \partial_K V = V, \qquad S \, \partial_{SS} V + K \, \partial_{SK} V = 0, \qquad S \, \partial_{SK} V + K \, \partial_{KK} V = 0,$$

and hence $S^2 \, \partial_{SS} V = -SK \, \partial_{SK} V = K^2 \, \partial_{KK} V$.

Proof. (i) Fix z. The function $\lambda \mapsto V(\lambda z)$ is differentiable by the chain rule, with derivative $\sum_i z_i \partial_i V(\lambda z)$; differentiating $V(\lambda z) = \lambda^k V(z)$ at $\lambda = 1$ gives (4.8). Conversely, $\phi(\lambda) = \lambda^{-k} V(\lambda z)$ has $\phi'(\lambda) = \lambda^{-k-1}\big(\sum_i \lambda z_i \, \partial_i V(\lambda z) - k V(\lambda z)\big) = 0$ by (4.8) at $\lambda z \in U$, so $\phi(\lambda) = \phi(1)$ for all $\lambda > 0$. (ii) Differentiating $V(\lambda z) = \lambda V(z)$ in z_j gives $\lambda \, \partial_j V(\lambda z) = \lambda \, \partial_j V(z)$. Then (i) with $k = 0$, applied to $\partial_j V$, gives the sum rule; for $V \in C^2$ the Hessian is symmetric (Schwarz), and the sum rule says that it maps z to 0. (iii) By (i), $V(z; \eta) = \sum_i z_i \, \partial_i V(z; \eta)$ for every η; differentiate in η. (iv) is (i) and (ii) with $z = (S, K)$; multiplying the second identity by S and the third by K gives the last. □

Remark 4.8 (When is a price homogeneous in spot and strike?). Suppose the time-t price of $f(S_T, K)$ is $V(S, K) = e^{-r\tau} \mathbb{E}^{\mathbb{Q}}[f(SR, K)]$, where the gross return $R = S_T / S_t \geq 0$ has a law that depends neither on S nor on K, and f is positively homogeneous of degree 1, like $(s - K)^+$, $(K - s)^+$ or $|s - K|$. Then $V(\lambda S, \lambda K) = e^{-r\tau} \mathbb{E}^{\mathbb{Q}}[\lambda f(SR, K)] = \lambda V(S, K)$. This covers the standard model, stochastic volatility models whose variance does not depend on the level of the stock (Chapter 19), and, with R the vector of returns, path-dependent payoffs such as Asian options. It fails when the law of returns depends on the level: in local-volatility models

$\mathrm{d}S_t = (r-y)S_t\,\mathrm{d}t + \sigma(t,S_t)S_t\,\mathrm{d}W_t$ with σ not constant in S, in the constant-elasticity model $\sigma(S) = \sigma_0 S^{\beta-1}$, $\beta \neq 1$, and in the Bachelier model. The price is still homogeneous in *all* its money-valued inputs (changing currency is always a symmetry), but these now include the local volatility (scaling S and K by λ replaces $\sigma(t,s)$ by $\sigma(t,s/\lambda)$) or a normal volatility, and (4.8) in (S,K) alone acquires an extra term (Exercise 4.4). For the local volatility of Figure 4.2 and $S = K = 100$, $T = 1$, the call costs 9.297 but $S\,\partial_S C + K\,\partial_K C = 1.815$: Euler's identity in (S,K) misses by 80% of the price.

Three consequences for hedging hold in every model whose call price $C(S,K)$ is homogeneous of degree one and C^2; in the standard model the Greeks are those of Theorem 2.7.

The call in its parts. $C = S\,\partial_S C + K\,\partial_K C$, and $-\partial_K C$ is the digital call (Theorem 3.5). Since $(S_T - K)^+ = S_T \mathbf{1}_{\{S_T>K\}} - K\mathbf{1}_{\{S_T>K\}}$, the asset-or-nothing call costs $C - K\partial_K C = S\,\partial_S C$: the value of the delta position. In the standard model the replicating portfolio holds $\partial_S C = e^{-y\tau}N(d_+)$ shares and $K\partial_K C = -Ke^{-r\tau}N(d_-)$ in cash, and Euler's identity says that it is worth the call. In the test case, $50.9643 - 43.6000 = 7.3643$.

Gamma is the density. By Theorem 4.7(iv) and the Breeden–Litzenberger formula (Theorem 3.5(iii)),

$$e^{-r\tau}q(K) = \partial_{KK}C = (S/K)^2\,\partial_{SS}C\,:$$

the spot gamma of the call struck at K, scaled by $(S/K)^2$, is the risk-neutral density at K. A trader's gamma across strikes is the market's density. In the test case $S = K$, and both equal 0.0195277 (Theorem 1.12). The gamma sum rule $S\,\partial_S(\partial_S C) + K\,\partial_S(\partial_K C) = 0$ says that, to first order in the move of the spot at a fixed time, rebalancing pays for itself: the shares bought, $S\,\mathrm{d}(\partial_S C)$, are paid for by the cash, $K\,\mathrm{d}(\partial_K C)$; the second-order remainder is the gamma profit $\frac{1}{2}\partial_{SS}C\,(\mathrm{d}S)^2$. In the standard model it reads $Se^{-y\tau}\partial_S N(d_+) = Ke^{-r\tau}\partial_S N(d_-)$, which is Theorem 2.7(i) because $\partial_S d_+ = \partial_S d_-$.

Conservation laws. By (4.9) every sensitivity to a parameter splits into the sensitivities of the two hedge ratios. With $\eta = \sigma$, the vega is S times the vanna $\partial_\sigma\partial_S C$ plus K times $\partial_\sigma\partial_K C$; in the test case the two terms are 29.2916 and 9.7639, and they add up to the vega 39.0554 up to rounding. With $\eta = t$, the theta is S times the charm $\partial_t\partial_S C$ plus $K\,\partial_t\partial_K C$: $-3.32226 = 0.04290 - 3.36516$.

▶ Interactive demo: check Euler's identity and "gamma is the density" across strikes in the standard model, and watch both fail in the local-volatility model of Figure 4.2.

4.4 The Kolmogorov equations

Theorem 1.12 showed that the Arrow–Debreu price density of the standard model solves the pricing equation in its initial point and an adjoint equation in its final point. Both facts hold for diffusions in general; this section states them, and proves the second, which explains where the adjoint comes from. Let $\bar{T} > 0$ and let $b, a, r : [0,\bar{T}] \times \mathbb{R} \to \mathbb{R}$ be a drift, a volatility and a discount rate. For $0 \le t \le \bar{T}$ and $x \in \mathbb{R}$, let $X^{t,x}$ solve $X_s = x + \int_t^s b(u,X_u)\,\mathrm{d}u + \int_t^s a(u,X_u)\,\mathrm{d}W_u$ on $[t,\bar{T}]$, with W

a $\mathbb{Q}$-Brownian motion, and let $D_s^{t,x} = \exp\big(-\int_t^s r(u, X_u^{t,x})\,\mathrm{d}u\big)$. Write $\mathcal{L}_t\varphi = b(t,\cdot)\varphi' + \frac{1}{2}a(t,\cdot)^2\varphi''$. For prices, $X = \log S$; in a local-volatility model $b = r - y - \frac{1}{2}\sigma^2$ and $a = \sigma$, evaluated at (t, e^x). Hypothesis (H) is: *b, a and r are bounded with bounded continuous partial derivatives of all orders on $[0,\bar T]\times\mathbb{R}$, and $a^2 \geq \lambda$ for a constant $\lambda > 0$ (uniform ellipticity).* Then b and a are Lipschitz in x, and $X^{t,x}$ exists and is unique (Karatzas and Shreve (1991), Section 5.2).

Theorem 4.9 (Kolmogorov's equations)**.** *Assume (H). There is a continuous function $G(t,x;T,\xi) \geq 0$, for $0 \leq t < T \leq \bar T$ and $x, \xi \in \mathbb{R}$, with the following properties.*

*(i) (*Discounted transition density*) For every bounded Borel f, $\mathbb{E}^{\mathbb{Q}}[D_T^{t,x} f(X_T^{t,x})] = \int_{\mathbb{R}} G(t,x;T,\xi) f(\xi)\,\mathrm{d}\xi$, and for $t < s < T$ the Chapman–Kolmogorov identity holds:*

$$\int_{\mathbb{R}} G(t,x;T,\xi) f(\xi)\,\mathrm{d}\xi = \int_{\mathbb{R}} G(t,x;s,\eta)\Big(\int_{\mathbb{R}} G(s,\eta;T,\xi) f(\xi)\,\mathrm{d}\xi\Big)\mathrm{d}\eta. \tag{4.10}$$

For $r = 0$, G is the transition density of X.

*(ii) (*Backward equation*) For fixed (T,ξ), G is $C^{1,2}$ in (t,x) on $[0,T)\times\mathbb{R}$ and*

$$\partial_t G + b(t,x)\,\partial_x G + \tfrac{1}{2}a(t,x)^2\,\partial_{xx} G - r(t,x)\,G = 0. \tag{4.11}$$

For every bounded continuous f, $u(t,x) = \int G(t,x;T,\xi) f(\xi)\,\mathrm{d}\xi$ tends to $f(x_0)$ as $(t,x) \to (T,x_0)$: G is the solution with terminal value δ_ξ.

*(iii) (*Forward, or Fokker–Planck, equation*) For fixed (t,x), G is $C^{1,2}$ in (T,ξ) on $(t,\bar T]\times\mathbb{R}$ and*

$$\partial_T G = \tfrac{1}{2}\,\partial_{\xi\xi}\big(a(T,\xi)^2\,G\big) - \partial_\xi\big(b(T,\xi)\,G\big) - r(T,\xi)\,G. \tag{4.12}$$

For every bounded continuous h, $\int G(t,x;T,\xi)h(\xi)\,\mathrm{d}\xi \to h(x)$ as $T \downarrow t$: G is the solution with initial value δ_x.

Proof. Existence. Under (H), Friedman (1964), Chapter 1, constructs by the parametrix method a fundamental solution Γ of the backward equation: a continuous function, $C^{1,2}$ in (t,x) for $t < T$, that solves (4.11), such that for every bounded continuous f the function $u(t,x) = \int \Gamma(t,x;T,\xi) f(\xi)\,\mathrm{d}\xi$ is bounded, solves (4.11) and tends to $f(x_0)$ as $(t,x) \to (T,x_0)$. The same chapter constructs the fundamental solution of the adjoint equation and shows that it is Γ with its two pairs of variables exchanged; in particular Γ is $C^{1,2}$ in (T,ξ) for $T > t$. (Friedman writes the equations forward in time; reverse time to match.)

Identification, (i) and (ii). Let f be bounded and continuous, and u as above. The Feynman–Kac formula (Theorem 1.6, case (b) with $\psi \equiv 1$, applied to $X^{t,x}$) gives $u(t,x) = \mathbb{E}^{\mathbb{Q}}[D_T^{t,x} f(X_T^{t,x})]$. The finite measures $A \mapsto \mathbb{E}^{\mathbb{Q}}[D_T^{t,x}\mathbf{1}_A(X_T^{t,x})]$ and $\Gamma(t,x;T,\xi)\,\mathrm{d}\xi$ give the same integral to every bounded continuous function, so they are equal; hence $G = \Gamma$ is the discounted transition density, $G \geq 0$, and (ii) holds. For (4.10), the conditional form (1.13) of the Feynman–Kac formula gives $\mathbb{E}^{\mathbb{Q}}[D_T^{t,x} f(X_T) \mid \mathcal{F}_s] = D_s^{t,x}\,u(s, X_s^{t,x})$; take expectations, and use (i) at time s for the bounded continuous function $u(s,\cdot)$. Both sides of (4.10) define finite measures in f, so the identity extends from bounded continuous to bounded Borel f.

The forward equation, (iii). Given the regularity of G in (T,ξ), the equation follows from Itô's formula, and the argument shows where the adjoint comes from. Fix (t,x), write $X = X^{t,x}$, $D = D^{t,x}$, and let $\varphi \in C^2(\mathbb{R})$ vanish outside a bounded interval. By Itô's formula and the

product rule, for $s \geq t$,

$$\mathrm{d}\big(D_s\varphi(X_s)\big) = D_s\big(\mathcal{L}_s\varphi - r(s,\cdot)\varphi\big)(X_s)\,\mathrm{d}s + D_s\,a(s,X_s)\,\varphi'(X_s)\,\mathrm{d}W_s.$$

The integrand of the $\mathrm{d}W$ term is bounded, because $D_s \leq e^{\|r\|_\infty \bar{T}}$ and a, φ' are bounded, so it is a martingale, and with $\psi_s = \mathcal{L}_s\varphi - r(s,\cdot)\varphi$,

$$m(T) := \mathbb{E}^{\mathbb{Q}}[D_T\varphi(X_T)] = \varphi(x) + \int_t^T \mathbb{E}^{\mathbb{Q}}\big[D_s\psi_s(X_s)\big]\,\mathrm{d}s.$$

The integrand is continuous in s, by dominated convergence along the continuous paths, so m is differentiable with $m'(T) = \mathbb{E}^{\mathbb{Q}}[D_T\psi_T(X_T)] = \int G(t,x;T,\xi)\psi_T(\xi)\,\mathrm{d}\xi$, by (i). On the other hand $m(T) = \int G(t,x;T,\xi)\varphi(\xi)\,\mathrm{d}\xi$, and since $\partial_T G$ is continuous and φ has compact support, $m'(T) = \int \partial_T G\,\varphi\,\mathrm{d}\xi$. Integrating by parts, twice in the second-order term and once in the first-order term, with no boundary terms because φ vanishes outside a bounded interval,

$$\int \partial_T G\,\varphi\,\mathrm{d}\xi = \int G\Big(\tfrac{1}{2}a^2\varphi'' + b\varphi' - r\varphi\Big)\mathrm{d}\xi = \int \Big(\tfrac{1}{2}\partial_{\xi\xi}(a^2 G) - \partial_\xi(bG) - rG\Big)\varphi\,\mathrm{d}\xi.$$

This holds for every such φ, and both brackets are continuous in ξ, so they are equal: this is (4.12). Finally, $\int G(t,x;T,\xi)h(\xi)\,\mathrm{d}\xi = \mathbb{E}^{\mathbb{Q}}[D_T h(X_T)] \to h(x)$ as $T \downarrow t$, by dominated convergence and the continuity of the paths. □

Remark 4.10 (The course's derivation). The course differentiates (4.10) in the intermediate time s, replaces $\partial_s G(s,\eta;T,\xi)$ by the backward equation, integrates by parts twice in η, and lets $T \downarrow s$ so that $G(s,\eta;T,\xi)$ becomes $\delta_\xi(\eta)$. As written this is heuristic: the boundary terms at $\eta \to \pm\infty$ must vanish and the limit must commute with the integral, which hold under (H) by Friedman's Gaussian bounds. The proof above uses a test function with compact support instead, which has no boundary terms.

Remark 4.11 (The standard model and beyond (H)). For the standard model in $x = \log S$, $b = \nu$, $a = \sigma$ and r are constants, (H) holds, and $G(t,x;T,\eta) = e^{-r\tau}k(\sigma^2\tau, \eta - x - \nu\tau)$ with the heat kernel k of (1.1). In the price variable the density acquires the Jacobian $1/\xi$, and the two equations become (1.18) and (1.19), verified directly in Theorem 1.12. Hypothesis (H) is sufficient, not necessary, but cannot simply be dropped: in the constant-elasticity model $\mathrm{d}S = \sigma_0 S^\beta\,\mathrm{d}W$, $0 < \beta < 1$, absorbed at zero, the volatility of $\log S$ is unbounded and $\mathbb{Q}(S_T = 0) > 0$, so the density on $(0,\infty)$ integrates to less than one.

Backward and forward, numerically. On a uniform grid in x, replace $\mathcal{L}_t - r$ by central differences, a tridiagonal matrix L_h, and step in time with Crank–Nicolson, with two implicit steps at each end to damp the irregular data (Chapter 20). One backward solve from a payoff gives its price at every spot; one forward solve with $L_h^\top$ from δ_{x_0}/h gives the density, hence the price of every payoff, for one spot. The two discretizations are exactly adjoint, so they agree up to rounding. In the test case, a grid of 801 points spanning eight standard deviations each way and 400 time steps price the call at 7.36386 from either side, an error of 4.3×10^{-4} that falls by a factor of four when the grid and the time steps are doubled. Figure 4.2 uses one forward solve to price all strikes in a local-volatility model, the starting point of Dupire's construction of local volatilities from option prices (Exercise 4.6).

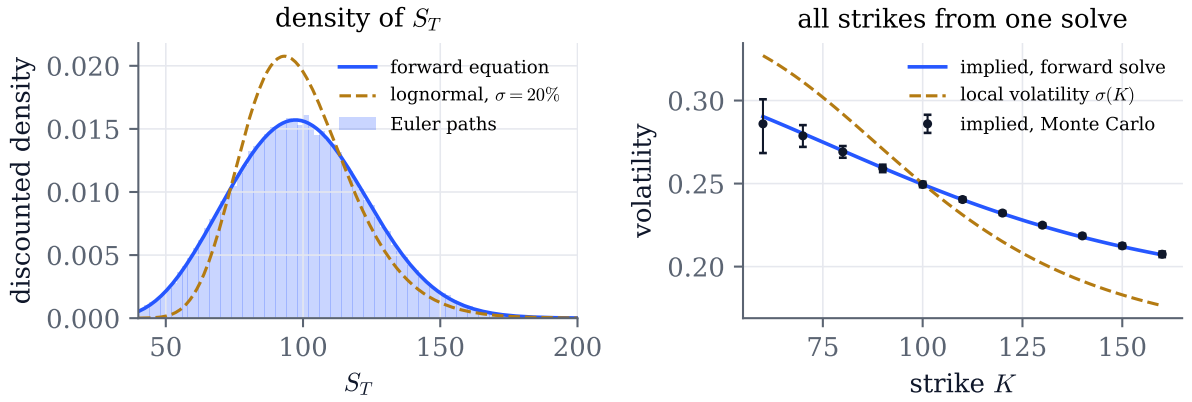

Figure 4.2: One forward solve in the local-volatility model $\sigma(S) = 0.25 - 0.10\tanh\left(2\log(S/100)\right)$, with $S_0 = 100$, $T = 1$, $r = 1\%$, $y = 2\%$ (grid of 801 points in $\log S$, 400 time steps). Left: the discounted density of S_T from the forward equation, against a histogram of 2×10^5 Euler paths with 250 steps (our implementation, numpy `default_rng`, seed 4), and the lognormal density of the test case. Right: Black–Scholes implied volatilities of the calls priced from that one density, for all strikes, the implied volatilities of the Monte Carlo prices with 95% confidence bars, and the local volatility at the strike. Near the money the implied skew is about half as steep as the local one.

In code

`code/src/mfe230d/replication.py`: `carr_madan_price` evaluates (4.3) by Gauss–Legendre quadrature in log-strike; `grid_replication` and `grid_price` build and price the finite strip; `power_call`, `power_greeks` and `leveraged_call` implement Theorem 4.4 and the post-wash on top of `code/src/mfe230d/bs.py`; `put_call_symmetry` is Theorem 4.6; `bs_second_order`, `bachelier_call` and `call_euler_terms` test Theorem 4.7; `backward_solve` and `forward_solve` are the two discretizations of Section 4.4. `code/tests/test_replication.py` checks every number quoted in this chapter.

▶ Interactive demo: change the local volatility and watch the density and the implied volatilities from one forward solve; check that a backward solve gives the same prices.

Further reading

Carr and Madan (1998) state the spanning formula and use it to trade volatility; Breeden and Litzenberger (1978) is the source of the density formula it inverts. Bowie and Carr (1994), Carr and Chou (1997), Carr et al. (1998) and Carr and Lee (2009) develop static hedging with put–call symmetry. Merton (1973) first proved that option prices are homogeneous in the stock price and the strike when returns do not depend on the price level. Kolmogoroff (1931) introduced the two equations; Friedman (1964) constructs their fundamental solutions, and Karatzas and Shreve (1991), Section 5.7, connects them with diffusions. Dupire (1994) turns the forward equation into a formula for the local volatility, and Gatheral (2006), Chapter 1, derives it.

Part II

Monte Carlo methods

Chapter 5

Random numbers

Every Monte Carlo estimator in this part of the notes averages functions of independent random variables. A computer produces none of these: it produces a deterministic sequence in $(0,1)$ designed to pass for an independent uniform sample, and every other law is obtained by transforming it. Sections 5.1 and 5.2 show how uniform pseudo-random numbers are made and tested, and why their defects are geometric. Sections 5.3 to 5.5 turn uniforms into other laws by inversion, by acceptance–rejection and, for the normal law, by the Box–Muller transform. Section 5.6 builds Gaussian vectors with a given covariance. Chapter 6 turns these draws into estimators. Throughout, $\mathcal{U}(0,1)$ is the uniform law on $(0,1)$, $U, U_1, U_2, \dots$ are independent with this law, and N, N' are the standard normal distribution function and density.

5.1 Pseudo-random numbers

A *pseudo-random number generator* consists of a finite set $\mathcal{S}$ of states, a transition map $f : \mathcal{S} \to \mathcal{S}$, an output map $g : \mathcal{S} \to [0,1)$ and a seed $s_0 \in \mathcal{S}$; it produces $s_{i+1} = f(s_i)$ and outputs $u_i = g(s_i)$. Determinism is a feature: the same seed reproduces a simulation exactly, which makes debugging possible and is what common random numbers exploit in Chapter 8. Finiteness has a price: as soon as $s_i = s_j$ with $i < j$, $s_{i+k} = s_{j+k}$ for all $k \geq 0$, so the sequence is eventually periodic, with period at most $|\mathcal{S}|$.

5.1.1 Linear congruential generators

Definition 5.1 (Linear congruential generator)**.** Fix integers $m \geq 2$ (the *modulus*), $a \in \{1, \dots, m-1\}$ (the *multiplier*), $c \in \{0, \dots, m-1\}$ (the *increment*) and a seed $x_0 \in \{0, \dots, m-1\}$. The *linear congruential generator* (LCG) is

$$x_{i+1} = (a\, x_i + c) \bmod m, \qquad u_i = x_i/m, \qquad i \geq 0.$$

It is *multiplicative* if $c = 0$. It has *full period* if $x_0, \dots, x_{m-1}$ are the m residues $0, \dots, m-1$ in some order.

The outputs lie in the grid $\{0, \frac{1}{m}, \dots, \frac{m-1}{m}\}$, the period is at most m, and a full-period generator

visits every residue once per period, whatever the seed. The value $u_i = 0$ can occur and must be removed or shifted before computing $\log u_i$ or $N^{-1}(u_i)$.

Example 5.2 (The course's generator)**.** Take $m = 16$, $a = 9$, $c = 5$ and $x_0 = 0$. Then $x_1 = 5$, $x_2 = 50 \bmod 16 = 2$, and the sequence is

$$0,\ 5,\ 2,\ 7,\ 4,\ 9,\ 6,\ 11,\ 8,\ 13,\ 10,\ 15,\ 12,\ 1,\ 14,\ 3,\ 0,\ \ldots$$

Each residue appears once before the sequence returns to 0: the period is full. Yet the sequence is visibly not random. Since $a^2 = 81 \equiv 1$, we have $x_{i+2} \equiv a^2 x_i + c(a+1) \equiv x_i + 2 \pmod{16}$: the terms of even index climb $0, 2, 4, \ldots$ and those of odd index $5, 7, 9, \ldots$. Small moduli exaggerate, but every defect in this section has this form: a linear relation among a few consecutive terms.

When does an LCG have full period? For $c \neq 0$ the answer is arithmetic.

> **Theorem 5.3** (Hull and Dobell)**.** *An LCG with modulus m, multiplier a and increment c has full period if and only if*
>
> *(i) c and m are coprime;*
>
> *(ii) every prime that divides m divides $a - 1$;*
>
> *(iii) if 4 divides m, then 4 divides $a - 1$.*

This is Theorem 1 of Hull and Dobell (1962); a proof is in Knuth (1997), Section 3.2.1.2, Theorem A. The necessity of the three conditions is elementary (Exercise 5.2); the sufficiency takes more number theory. In Theorem 5.2, $c = 5$ is odd, 2 is the only prime factor of 16, and $a - 1 = 8$ is divisible by 4. The theorem also shows how weak a guarantee a full period is: $a = 1$, $c = 1$ satisfies all three conditions and produces $x_i = x_0 + i \bmod m$.

Remark 5.4 (Multiplicative generators). If $c = 0$, the residue 0 is fixed and the period is at most $m - 1$. If a and the seed are coprime to m, then $x_k = a^k x_0 \equiv x_0$ if and only if $a^k \equiv 1$, so the period is the multiplicative order of a. For a prime m it is $m - 1$ exactly when a is a primitive root: $a^{(m-1)/q} \not\equiv 1 \pmod{m}$ for every prime $q \mid m - 1$ (Knuth (1997), Section 3.2.1.2, Theorems B and C). The "minimal standard" MINSTD of Park and Miller (1988) has $m = 2^{31} - 1$ and the primitive root $a = 7^5 = 16807$, hence period $2^{31} - 2$. For $m = 2^e$, $e \geq 4$, the period is at most 2^{e-2} (same reference); IBM's RANDU, $m = 2^{31}$ and $a = 2^{16} + 3 = 65539$, attains 2^{29} from odd seeds.

Remark 5.5 (Period and resolution). The period (draws before the sequence repeats) and the resolution (gap between neighbouring possible outputs) are different measures of coarseness. For an LCG with $u_i = x_i/m$ they are at most m and $1/m$; NumPy's default generator has period 2^{128} but returns multiples of 2^{-53} in $[0, 1)$. A long period does not make the outputs finer: 32-bit uniforms repeat values long before the period ends (Exercise 5.8), and the resolution limits how far into the tails a transformation can reach (Theorem 5.13).

5.1.2 The lattice structure

A Monte Carlo simulation of a d-dimensional quantity uses the outputs d at a time. For an LCG, the d-tuples of consecutive outputs are far from uniformly scattered in the cube.

Proposition 5.6 (Marsaglia's lattice)**.** *Let (x_i) be an LCG with modulus m, multiplier a and increment c, let $u_i = x_i/m$, fix $d \geq 2$, and set $P_i = (u_i, u_{i+1}, \ldots, u_{i+d-1}) \in [0,1)^d$ and $v = (1, a, a^2, \ldots, a^{d-1})$.*

(i) There is $b \in \mathbb{R}^d$, depending only on m, a, c and d, such that every P_i lies in $b + \Lambda$, where $\Lambda = \{(kv + mz)/m : \ k \in \mathbb{Z},\ z \in \mathbb{Z}^d\}$ is a lattice.

(ii) Let $h = (h_0, \ldots, h_{d-1}) \in \mathbb{Z}^d \setminus \{0\}$ satisfy $h \cdot v \equiv 0 \pmod m$. Then all the points P_i lie on at most $\|h\|_1 = \sum_j |h_j|$ of the parallel hyperplanes $\{x : h \cdot x = \kappa\}$, and consecutive hyperplanes of this family are $1/\|h\|_2$ apart.

(iii) There is such an h with $\max_j |h_j| \leq m^{1/d}$. Hence the d-tuples of every LCG with modulus m lie on at most $d\, m^{1/d}$ parallel hyperplanes.

Proof. (i) By induction on j, $x_{i+j} \equiv a^j x_i + c\, s_j \pmod m$ with $s_j = 1 + a + \cdots + a^{j-1}$ ($s_0 = 0$): the step from j to $j+1$ is $a(a^j x_i + c s_j) + c = a^{j+1} x_i + c s_{j+1}$. With $s = (s_0, \ldots, s_{d-1})$, this says $m P_i = x_i v + cs + m z_i$ for some $z_i \in \mathbb{Z}^d$, so $P_i \in cs/m + \Lambda$. The set Λ is a subgroup of $\frac{1}{m}\mathbb{Z}^d$ that contains $\mathbb{Z}^d$, hence a lattice.

(ii) For $y = (kv + mz)/m \in \Lambda$, $h \cdot y = k\,(h \cdot v)/m + h \cdot z$ is an integer. Hence $h \cdot P_i \in h \cdot b + \mathbb{Z}$ for every i: the points lie on the hyperplanes $h \cdot x = h \cdot b + n$, $n \in \mathbb{Z}$, and two consecutive ones are at distance $1/\|h\|_2$. For $x \in [0,1)^d$, each $h_j x_j$ ranges over $[0, h_j)$ or $(h_j, 0]$, so $h \cdot x$ ranges over an interval of length $\|h\|_1$ that contains at most one of its endpoints. Such an interval contains at most $\|h\|_1$ points of the set $h \cdot b + \mathbb{Z}$.

(iii) Let $K = \lfloor m^{1/d} \rfloor$, so that $(K+1)^d > m$. The $(K+1)^d$ vectors $g \in \{0, \ldots, K\}^d$ give at most m distinct residues $g \cdot v \bmod m$, so by the pigeonhole principle two different ones, g and g', give the same residue. Then $h = g - g' \neq 0$ satisfies $h \cdot v \equiv 0$ and $|h_j| \leq K$, so $\|h\|_1 \leq dK \leq d m^{1/d}$. □

Minkowski's convex body theorem, applied to the region $\{\|h\|_1 \leq R\}$ instead of a cube, sharpens (iii) to $\|h\|_1 \leq (d!\, m)^{1/d}$; this is the theorem of Marsaglia (1968). For $m = 2^{32}$ and $d = 10$ it gives about 41.6: the 10-tuples of any LCG modulo 2^{32} lie on at most 41 parallel hyperplanes. The hyperplanes of (ii) are furthest apart for the shortest admissible h. The *spectral test* (Knuth (1997), Section 3.3.4) computes

$$\nu_d = \min \{\|h\|_2 : \ h \in \mathbb{Z}^d \setminus \{0\},\ h \cdot v \equiv 0 \pmod m\}$$

and reports the spacing $1/\nu_d$: the open slab between two adjacent hyperplanes of that family contains no point. Good multipliers make $1/\nu_d$ small in several dimensions at once (Glasserman (2003), Section 2.1.4). A generator that does not visit all residues, such as a multiplicative one modulo 2^e, fills only part of the lattice, and the test is then run with a reduced modulus (Knuth (1997), Section 3.3.4).

Example 5.7 (RANDU)**.** For RANDU, $a = 2^{16} + 3$ and $m = 2^{31}$, so $a^2 = 2^{32} + 6 \cdot 2^{16} + 9 \equiv 6a - 9 \pmod{2^{31}}$ and $x_{i+2} \equiv 6x_{i+1} - 9x_i$. Thus $h = (9, -6, 1)$ is admissible in Theorem 5.6(ii): the number $9u_i - 6u_{i+1} + u_{i+2}$ is an integer in the open interval $(-6, 10)$, so it takes at most 15 values, and all triples of consecutive outputs lie on 15 parallel planes, $1/\sqrt{118} \approx 0.092$ apart (Figure 5.1(d)). An exhaustive search in the code confirms $\nu_3 = \sqrt{118}$, also with the reduced modulus 2^{30} that accounts for the odd states. In two dimensions RANDU is unremarkable: with the reduced modulus, its pairs lie on lines 4.3×10^{-5} apart. A defect can hide in a dimension nobody checked. In miniature, $m = 2^{12}$ and $a = 2^6 + 1$ give $(a-1)^2 \equiv 0$, so $h = (1, -2, 1)$ is

admissible: the pairs fill the square evenly ($\nu_2 = 63.0$; the best full-period multiplier for this modulus reaches 67.5), but the triples lie on at most four planes (Figure 5.1(b,c)).

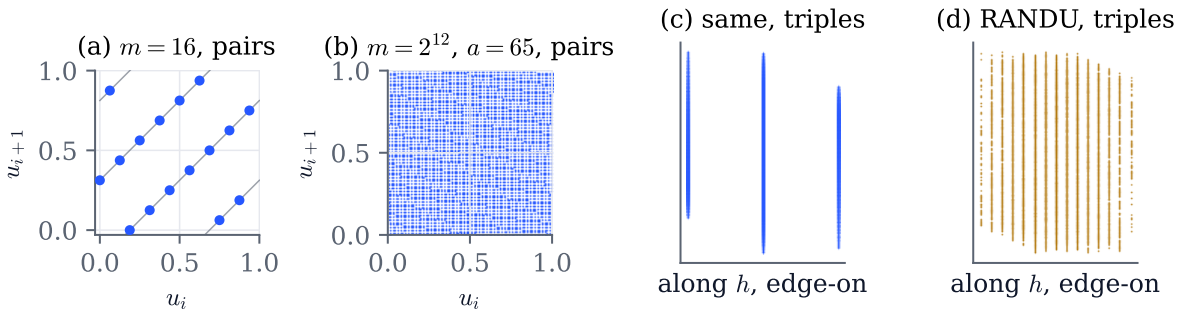

Figure 5.1: (a) The 16 pairs (u_i, u_{i+1}) of Theorem 5.2 lie on 4 lines, as Theorem 5.6(ii) predicts for $h = (2, -2)$. (b) The pairs of the generator $m = 2^{12}$, $a = 65$, $c = 1$ fill the square evenly, but (c) its triples, seen edge-on along $h = (1, -2, 1)$, lie on three planes. (d) 10^4 consecutive (overlapping) triples of RANDU seen along $h = (9, -6, 1)$: 15 planes.

▶ Interactive demo: choose m, a and c, check the Hull–Dobell conditions and the period, and look at the pairs and at the triples seen edge-on.

5.1.3 Generators in use

Generators in current use have far larger states than these examples, and better-mixed outputs. The *Mersenne Twister* MT19937 (Matsumoto and Nishimura, 1998), the default of Python's `random` module, is a linear recursion over the two-element field on 19937 bits, with period $2^{19937} - 1$ and equidistribution in up to 623 dimensions at 32-bit resolution; being linear, it fails the linear-complexity tests of the TestU01 batteries (L'Ecuyer and Simard, 2007). *MRG32k3a* (L'Ecuyer, 1999) combines two recursions of order three modulo primes close to 2^{32}, has period about 2^{191}, and is discussed in detail by Glasserman (2003), Section 2.1.5. *PCG* (O'Neill, 2014) is an LCG modulo 2^{64} or 2^{128} whose output is a permuted, truncated function of the state (32 bits out of 64 in PCG32: a shift-and-xor and a rotation driven by the leading bits) that hides the weak low-order bits (Exercise 5.1) and the lattice; NumPy's default generator is PCG64. The *xoshiro* family (Blackman and Vigna, 2021) consists of scrambled linear generators that are among the fastest available. Parallel simulations need non-overlapping streams. For an LCG, k steps compose to the affine map $x_{i+k} \equiv A_k x_i + C_k$, $A_k = a^k$, $C_k = c(1 + a + \cdots + a^{k-1})$ modulo m, which repeated squaring computes in $O(\log k)$ operations: jumping ahead 2^{64} draws is cheap.

In code

`code/src/mfe230d/rng.py` implements LCGs (vectorised by jump-ahead), the spectral test and PCG32; NumPy's PCG64 is the reference generator.

```
from mfe230d.rng import lcg_sequence, spectral_test, PCG32, uniforms
lcg_sequence(17, m=16, a=9, c=5, x0=0) # 0 5 2 7 4 9 6 11 8 13 10 15 12 1 14 3 0
spectral_test(a=65539, m=2**31, d=3) # (array([ 9, -6, 1]), 10.86...) = sqrt(118)
```

```
PCG32(initstate=42, initseq=54).random_raw(2) # 0xa15c02b7, 0x7b47f409 (
    reference)
u = uniforms(10**6, "pcg64", seed=1) # NumPy's PCG64, mapped into (0, 1)
```

5.2 Testing a generator

A generator cannot be proved random; it can only fail to look random. A statistical test fixes a statistic T of $u_1, \dots, u_n$ whose law under the null hypothesis

$$H_0: \ u_1, \dots, u_n \text{ are independent with law } \mathcal{U}(0,1)$$

is known, and reports the p-value $p = \mathbb{P}_{H_0}(T \geq t)$ at the observed value t. If the law of T under H_0 is continuous, p is uniform on $(0,1)$ under H_0 (Theorem 5.8(iii) below), so a perfect generator gives $p < 0.01$ once in a hundred tests. A defect shows as a p-value too extreme to be credible, or as p-values that are not uniform over repetitions (Knuth (1997), Section 3.3.1).

Moments. Under H_0, u_i has mean $\frac{1}{2}$ and variance $\frac{1}{12}$, so by the central limit theorem $Z = \sqrt{12n}\,(\bar{u}_n - \frac{1}{2})$ is approximately $\mathcal{N}(0,1)$ (Exercise 5.7 treats the higher moments).

Chi-square tests. Split $[0,1)$ into k cells of width $1/k$ and let O_j be the number of u_i in cell j. Pearson's statistic

$$X^2 = \sum_{j=1}^{k} \frac{(O_j - n/k)^2}{n/k}$$

converges in law under H_0 to the χ^2 law with $k-1$ degrees of freedom (Knuth (1997), Section 3.3.1.A); the approximation is adequate when $n/k \geq 5$ or so. The *serial test* applies it to the non-overlapping d-tuples $(u_{d(i-1)+1}, \dots, u_{di})$ in the k^d cells of $[0,1)^d$, with $k^d - 1$ degrees of freedom (overlapping tuples are dependent and need another limit; Knuth (1997), Section 3.3.2.B). It is the test that sees a lattice.

Kolmogorov–Smirnov. With F_n the empirical distribution function and $u_{(1)} \leq \cdots \leq u_{(n)}$ the order statistics,

$$D_n = \sup_{x \in [0,1]} |F_n(x) - x| = \max_{1 \leq i \leq n} \max\Big(\frac{i}{n} - u_{(i)},\ u_{(i)} - \frac{i-1}{n}\Big).$$

Under H_0, $\sqrt{n}\,D_n$ converges in law to a variable K with $\mathbb{P}(K > t) = 2\sum_{j \geq 1} (-1)^{j-1} e^{-2j^2 t^2}$ (Kolmogorov; see Knuth (1997), Section 3.3.1.B). By Theorem 5.8(iii) it applies to any continuous F after the map $x \mapsto F(x)$. It sees one-dimensional marginals only.

Serial correlation. The sample autocorrelation at lag k is $r_k = \sum_{i=1}^{n-k} (u_i - \bar{u}_n)(u_{i+k} - \bar{u}_n) / \sum_{i=1}^{n} (u_i - \bar{u}_n)^2$. For an i.i.d. sequence with finite variance, $\sqrt{n}\,(r_1, \dots, r_K)$ converges in law to $\mathcal{N}(0, I_K)$ (Bartlett's formula; Brockwell and Davis (1991), Theorem 7.2.2). About one lag in twenty falls outside the band $\pm 1.96/\sqrt{n}$ by chance.

Histograms with error bars. If $z_1, \dots, z_n$ are independent with a continuous law F, the count O_j in a bin $[e_{j-1}, e_j)$ is binomial with n trials and success probability $p_j = F(e_j) - F(e_{j-1})$: mean np_j, standard deviation $\sigma_j = \sqrt{np_j(1-p_j)}$. About 4.6% of the standardised counts $(O_j - np_j)/\sigma_j$ should exceed 2 in absolute value (Figure 5.5).

generator	period	$\sqrt{12n}(\bar{u} - \frac{1}{2})$	KS	χ^2, $k = 10^3$	$\sqrt{n}\, r_1$	serial $d = 2$	serial $d = 3$
RANDU	2^{29}	-0.71	0.54	0.46	0.91	0.32	$< 10^{-300}$
MINSTD	$2^{31} - 2$	-1.37	0.28	0.08	-0.64	0.28	0.09
LCG 2^{32}	2^{32}	0.97	0.41	0.49	0.90	0.74	0.89
PCG32	2^{64}	-0.17	0.86	0.86	-0.67	0.52	0.32
PCG64	2^{128}	-0.08	0.77	0.73	-0.88	0.35	0.12

Table 5.1: Tests on $n = 10^6$ uniforms from five generators: two z-scores, and p-values for Kolmogorov–Smirnov and the chi-square tests (100^2 cells in $d = 2$, 20^3 cells in $d = 3$). "LCG 2^{32}" is the full-period generator with $a = 1664525$, $c = 1013904223$. RANDU fails only the three-dimensional test, with $X^2 = 247{,}846$ on 7,999 degrees of freedom.

In Table 5.1 every generator passes the one-dimensional tests and the test on pairs, and RANDU fails the test on triples beyond doubt, as Theorem 5.7 predicts: its planes are 0.092 apart and the cells 0.05 wide, so 2000 of the 8000 cells are empty while others hold up to 129 triples, against 41.7 expected. The spectral test sees the defect without sampling: $1/\nu_3$ is 0.092 for RANDU, 1.6×10^{-3} for MINSTD and 6.6×10^{-4} for the LCG modulo 2^{32}, too fine for these cells. Batteries such as TestU01 (L'Ecuyer and Simard, 2007) run many such tests on far longer sequences. Passing them is necessary, not sufficient: what matters is whether a generator's defects interact with the problem at hand (Glasserman (2003), Section 2.1.1).

5.3 Inversion

From now on we assume a source of independent $\mathcal{U}(0,1)$ variables. The most direct way to obtain any other law on $\mathbb{R}$ is to invert its distribution function.

Proposition 5.8 (Inversion)**.** *Let F be a distribution function on $\mathbb{R}$ and define its generalized inverse $F^{-1}(u) = \inf\{x \in \mathbb{R} : \ F(x) \geq u\}$ for $u \in (0,1)$.*

(i) $F^{-1}(u)$ is a real number, F^{-1} is nondecreasing, and for all $u \in (0,1)$ and $x \in \mathbb{R}$,

$$F^{-1}(u) \leq x \iff u \leq F(x). \tag{5.1}$$

(ii) If $U \sim \mathcal{U}(0,1)$, then $X = F^{-1}(U)$ has distribution function F.

(iii) If F is continuous and X has distribution function F, then $F(X) \sim \mathcal{U}(0,1)$.

Proof. (i) Let $u \in (0,1)$ and $A_u = \{x : F(x) \geq u\}$. Since $F(x) \to 1$ as $x \to \infty$ and $u < 1$, A_u is not empty; since $F(x) \to 0$ as $x \to -\infty$ and $u > 0$, it is bounded below. So $F^{-1}(u) = \inf A_u$ is real. Because F is nondecreasing, A_u is a half-line that contains every point to the right of any of its points, and because F is right-continuous, it contains its infimum: $A_u = [F^{-1}(u), \infty)$. This is (5.1). If $u \leq u'$ then $A_{u'} \subseteq A_u$, so $F^{-1}(u) \leq F^{-1}(u')$.

(ii) By (5.1), $\mathbb{P}(F^{-1}(U) \le x) = \mathbb{P}(U \le F(x)) = F(x)$ for every $x \in \mathbb{R}$.

(iii) Fix $u \in (0,1)$ and let $x_u = F^{-1}(u)$. By (5.1), $\{F(X) \ge u\} = \{X \ge x_u\}$, so $\mathbb{P}(F(X) \ge u) = 1 - F(x_u-)$, where $F(x_u-)$ is the left limit. Now $F(x_u) \ge u$ by (i), and $F(x) < u$ for $x < x_u$, so $F(x_u-) \le u$; by continuity $F(x_u-) = F(x_u)$, hence $F(x_u) = u$ and $\mathbb{P}(F(X) \ge u) = 1 - u$ for all $u \in (0,1)$. This is the uniform law. □

Continuity in (iii) cannot be dropped (if X is Bernoulli, $F(X)$ takes two values), but (ii) needs no assumption: atoms of F become flat stretches of F^{-1}. Inversion uses one uniform per variate and is monotone in it, so $F^{-1}(U)$ and $G^{-1}(U)$ are comonotone and $F^{-1}(U)$ and $F^{-1}(1-U)$ are antithetic. Antithetic variates, common random numbers, stratification and quasi-Monte Carlo all rely on this (Theorems 8.2 and 8.11, Chapters 10 and 11). For the exponential law, $F^{-1}(u) = -\log(1-u)/\lambda$, and since $1-U$ is uniform, $-\log(U)/\lambda$ is exponential with rate λ. For a discrete law on $y_1 < y_2 < \cdots$ with $F_k = \mathbb{P}(X \le y_k)$, $F^{-1}(u) = y_k$ for $F_{k-1} < u \le F_k$: inversion is a table search (Exercise 5.3).

5.3.1 Inversion by Newton's method: the semicircle law

When F can be evaluated but not inverted, $F(x) = u$ is solved numerically, and the starting point decides whether the iteration is safe. The course's example is the semicircle law,

$$f(x) = \frac{2}{\pi}\sqrt{1-x^2}, \qquad F(x) = \frac{1}{2} + \frac{1}{\pi}\Big(x\sqrt{1-x^2} + \arcsin x\Big),$$

with the guess $x_0 = 2u - 1$ that inverts $F(x) \approx \frac{1+x}{2}$. The density vanishes at ± 1, where Newton's method could fail; from this start it does not.

Proposition 5.9 (Newton's method for the semicircle law)**.** *Let $u \in (\frac{1}{2}, 1)$, $x^* = F^{-1}(u) \in (0,1)$, and $x_{k+1} = x_k - (F(x_k) - u)/f(x_k)$ with $x_0 = 2u - 1$. Then $x_0 \ge x^*$,*

$$0 \le x_1 \le x_2 \le \cdots \le x^*, \qquad x_k \to x^*, \qquad |x_{k+1} - x^*| \le \frac{x^*}{2(1-x^{*2})}\,|x_k - x^*|^2 \quad (k \ge 1).$$

By the symmetry $F(-x) = 1 - F(x)$, the case $u \in (0, \frac{1}{2})$ is the mirror image, and $u = \frac{1}{2}$ gives $x_0 = x^* = 0$.

Proof. On $[0,1]$, f is decreasing, so F is concave: it lies above its chord, $F(x) \ge \frac{1+x}{2}$, and below its tangents, $F(y) \le F(x) + f(x)(y-x)$ for $x \in [0,1)$, $y \in [0,1]$.

Since $u = \frac{1+x_0}{2}$, the chord bound gives $F(x_0) \ge u$, so $x_0 \ge x^*$. Next, $x_1 = h(x_0)/f(x_0)$ with $h(x) = xf(x) - F(x) + \frac{1+x}{2}$. Now $h(0) = h(1) = 0$ and $h'(x) = \frac{1}{2} + xf'(x) = \frac{1}{2} - \frac{2}{\pi}\frac{x^2}{\sqrt{1-x^2}}$ is decreasing, so h increases and then decreases on $[0,1]$, and $h \ge 0$ there; thus $x_1 \ge 0$.

Now argue by induction. Let $x_k \in [0,1)$, which holds for $k = 0$. The tangent $T_k(y) = F(x_k) + f(x_k)(y - x_k)$ is increasing and $T_k(x_{k+1}) = u < 1 = F(1) \le T_k(1)$, so $x_{k+1} < 1$. Also $x_{k+1} \ge 0$: for $k = 0$ by the bound on h; for $k \ge 1$, if $x_k \le x^*$, then $F(x_k) \le u$ and $x_{k+1} = x_k + (u - F(x_k))/f(x_k) \ge x_k \ge 0$. The tangent bound then gives $F(x_{k+1}) \le T_k(x_{k+1}) = u$, that is, $x_{k+1} \le x^*$. Hence $0 \le x_1 \le x_2 \le \cdots \le x^*$. The limit $\ell \le x^* < 1$ satisfies $\ell = \ell - (F(\ell) - u)/f(\ell)$ by continuity, so $F(\ell) = u$ and $\ell = x^*$.

For the rate, Taylor's formula gives $F(x^*) = F(x_k) + f(x_k)(x^* - x_k) + \frac{1}{2}f'(\xi)(x^* - x_k)^2$ for some $\xi \in [x_k, x^*]$, so $x_{k+1} - x^* = \frac{f'(\xi)}{2f(x_k)}(x_k - x^*)^2$. On $[0, 1)$, f is decreasing and $|f'(x)| = \frac{2}{\pi}\frac{x}{\sqrt{1-x^2}}$ is increasing, so $|f'(\xi)| \le |f'(x^*)|$ and $f(x_k) \ge f(x^*)$, and $|f'(x^*)|/(2f(x^*)) = x^*/(2(1 - x^{*2}))$. □

The constant blows up as $x^* \to 1$: near the edges, where f is small, the quadratic regime starts late (Figure 5.2(a)). On a grid of $u \in [10^{-4}, 1 - 10^{-4}]$, the residual $|F(x_k) - u|$ is at most 2.6×10^{-4} after two steps and 3.4×10^{-6} after three, with medians 1.1×10^{-6} and 5×10^{-13}. Halley's method, which also uses f', iterates $x_{k+1} = x_k - 2gf/(2f^2 - gf')$ with $g = F(x_k) - u$ and everything evaluated at x_k; two Halley steps reach a median residual of 6×10^{-12}.

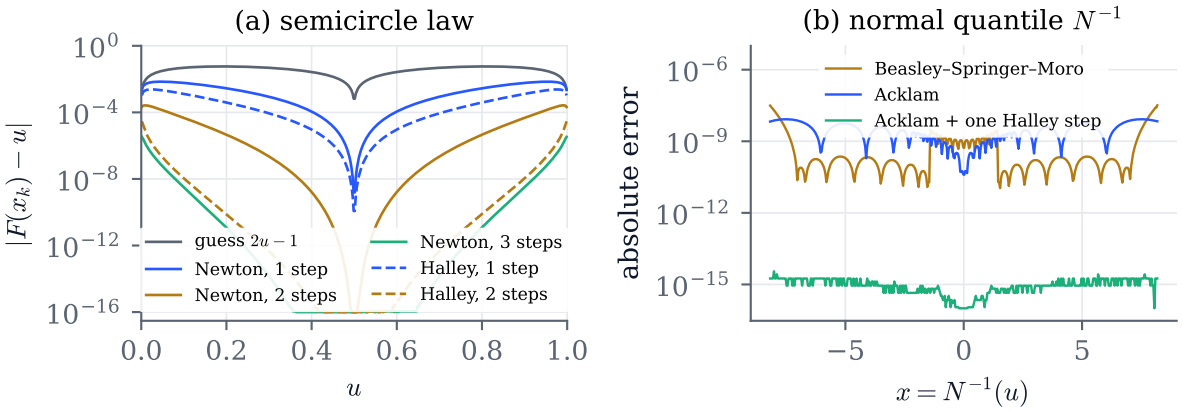

Figure 5.2: (a) Residual $|F(x_k) - u|$ of Newton's and Halley's iterations for the semicircle law from $x_0 = 2u - 1$, against u; convergence is slowest near $u = 0$ and $u = 1$, as Theorem 5.9 predicts. (b) Absolute errors of two approximations of N^{-1} and of Acklam's after one Halley step, against $x = N^{-1}(u)$. Both panels plot block maxima over blocks of 15 grid points.

5.3.2 Inverting the normal distribution function

N^{-1} has no closed form, and it is the inversion that finance uses most. Practical algorithms use a rational function of $u - \frac{1}{2}$ in the centre and of a variable that straightens the tail, such as $\sqrt{-2\log u}$ or $\log(-\log u)$, for small u; the upper tail follows from $N^{-1}(1 - u) = -N^{-1}(u)$. The *Beasley–Springer–Moro* algorithm combines the central approximation of Beasley and Springer (1977) ($|u - \frac{1}{2}| < 0.42$) with the tail polynomial in $\log(-\log u)$ of Moro (1995); Glasserman (2003), Section 2.3.2, prints the coefficients and reports a maximum absolute error of 3×10^{-9} for $|x| \le 7$ (we measure 3.0×10^{-9}, near $|x| = 1.4$ where the two pieces meet). *Acklam's algorithm* (Acklam, 2000) uses $qP_5(q^2)/Q_5(q^2)$ with $q = u - \frac{1}{2}$ for $0.02425 \le u \le 0.97575$ and $P_5(s)/Q_4(s)$ with $s = \sqrt{-2\log u}$ in the lower tail (P_k, Q_k polynomials of degree k), with relative error below 1.15×10^{-9} (we measure 1.13×10^{-9}). AS 241 (Wichura, 1988) reaches about 10^{-16}, and Giles (2011) approximates the related erfinv (with $N^{-1}(u) = \sqrt{2}\,\mathrm{erfinv}(2u - 1)$) in single precision for GPUs, with a branch that threads rarely take differently. A cheap approximation can be polished. Since $N''(x) = -xN'(x)$, one step of Halley's method on $N(x) = u$ reads

$$x \leftarrow x - \frac{t}{1 + xt/2}, \qquad t = \frac{N(x) - u}{N'(x)}. \tag{5.2}$$

It needs an accurate N, evaluated in the lower half ($u \le \frac{1}{2}$) where small probabilities keep their relative precision. One step from Acklam's value brings the error below 4×10^{-15} for $|x| \le 8.2$ (Figure 5.2(b)). An error of 10^{-9} is invisible next to a statistical error of 10^{-3}, but not to a finite-difference Greek computed with common random numbers.

5.4 Acceptance–rejection

When F cannot be inverted cheaply, a density can still be sampled by proposing from an easier density and keeping each proposal with a suitable probability (von Neumann, 1951).

Proposition 5.10 (Acceptance–rejection)**.** *Let f and g be probability densities on $\mathbb{R}^d$ with $f \le c\,g$ for a constant c. Let $(Y_k, U_k)_{k\ge1}$ be independent pairs, Y_k with density g and $U_k \sim \mathcal{U}(0,1)$ independent of Y_k, and call trial k* accepted *if $c\,U_k g(Y_k) \le f(Y_k)$. Let τ be the first accepted trial.*

(i) Each trial is accepted with probability $1/c$; in particular $c \ge 1$.

(ii) τ is geometric, $\mathbb{P}(\tau = k) = (1-\frac{1}{c})^{k-1}\frac{1}{c}$, with $\mathbb{E}[\tau] = c$ and $\mathrm{Var}(\tau) = c(c-1)$.

(iii) Y_τ has density f and is independent of τ. More generally, the values at the successive accepted trials are independent with density f.

Proof. Since $f \le cg$, f vanishes where g does, and Y_k falls in $\{g = 0\}$ with probability zero; there we set $f/(cg) = 0$. For a Borel set B, conditioning on Y_k (Fubini),

$$\mathbb{P}(Y_k \in B,\ \text{trial } k \text{ accepted}) = \int_B g(y)\,\mathbb{P}\Big(U \le \frac{f(y)}{c\,g(y)}\Big)\mathrm{d}y = \int_B g(y)\frac{f(y)}{c\,g(y)}\mathrm{d}y = \frac{1}{c}\int_B f(y)\,\mathrm{d}y,$$

using $f/(cg) \le 1$. With $B = \mathbb{R}^d$ this is (i), and $1/c \le 1$ because a probability is at most one. The trials are independent, so

$$\mathbb{P}(\tau = k,\ Y_\tau \in B) = \Big(1-\frac{1}{c}\Big)^{k-1}\frac{1}{c}\int_B f(y)\,\mathrm{d}y.$$

The right side factorises: summing over k shows that Y_τ has density f, $B = \mathbb{R}^d$ gives the geometric law of τ (mean c, variance $(1-\frac{1}{c})c^2$), and the product form is the independence. The successive accepted trials $\tau_1 < \tau_2 < \cdots$ cut the trials into blocks of independent pairs, and the same computation, block by block, gives a product of such factors for $(\tau_1, Y_{\tau_1}, \ldots, \tau_j - \tau_{j-1}, Y_{\tau_j})$. □

Geometrically, $(Y_k, c\,U_k g(Y_k))$ is uniform under the graph of cg and is kept if it falls under the graph of f; the acceptance probability is the ratio of the areas. The test only involves $f/(cg)$, so f is needed only up to a constant: if $\tilde f = Z f \le \chi$ with $A = \int \chi < \infty$, sampling from χ/A and accepting with probability $\tilde f/\chi$ has acceptance probability Z/A. The best constant is $c = \sup f/g$; it is finite only if the proposal has tails at least as heavy as the target (a normal proposal cannot serve a Cauchy target). Each accepted draw costs c proposals on average.

Example 5.11 (Envelopes for the semicircle law)**.** The flat envelope $\chi = 2/\pi$ on $[-1,1]$ has area $4/\pi$, so the acceptance probability is $\pi/4 \approx 0.785$ (proposal $X = 2U_1 - 1$, accepted if $U_2 \le \sqrt{1-X^2}$). Since f is concave, its tangents lie above it. The tangent at $\pm\sin\vartheta$, for $\vartheta \in (0, \frac{\pi}{2})$,

is $\frac{2}{\pi}(1 \mp x\sin\vartheta)/\cos\vartheta$, and it crosses the level $2/\pi$ at $\pm t$ with $t = (1-\cos\vartheta)/\sin\vartheta = \tan\frac{\vartheta}{2}$. The envelope $\chi_\vartheta(x) = \frac{2}{\pi}\min\left(1, (1-|x|\sin\vartheta)/\cos\vartheta\right)$ is piecewise linear (sampled by inverting a piecewise quadratic), with area

$$\int_{-1}^{1}\chi_\vartheta = \frac{4}{\pi}\Big(t + \frac{(1-t) - \frac{1}{2}(1-t^2)\sin\vartheta}{\cos\vartheta}\Big) = \frac{4}{\pi}\,\frac{1+t^2}{1+t},$$

using $\sin\vartheta = 2t/(1+t^2)$ and $\cos\vartheta = (1-t^2)/(1+t^2)$. The acceptance probability is therefore

$$p(\vartheta) = \frac{\pi}{4}\,\frac{1+t}{1+t^2}, \qquad t = \tan\frac{\vartheta}{2}. \tag{5.3}$$

Tangents at $\pm\frac{1}{2}$ ($\vartheta = \frac{\pi}{6}$, $t = 2-\sqrt{3}$) give $p = (3+\sqrt{3})\pi/16 \approx 0.929$. The derivative of $(1+t)/(1+t^2)$ vanishes where $t^2 + 2t - 1 = 0$, so the best tangent point is $t = \sqrt{2}-1$, that is $\vartheta = \frac{\pi}{4}$ and tangents at $\pm\frac{\sqrt{2}}{2}$, with $p = (1+\sqrt{2})\pi/8 \approx 0.948$. These are the course's two envelopes (Figure 5.3).

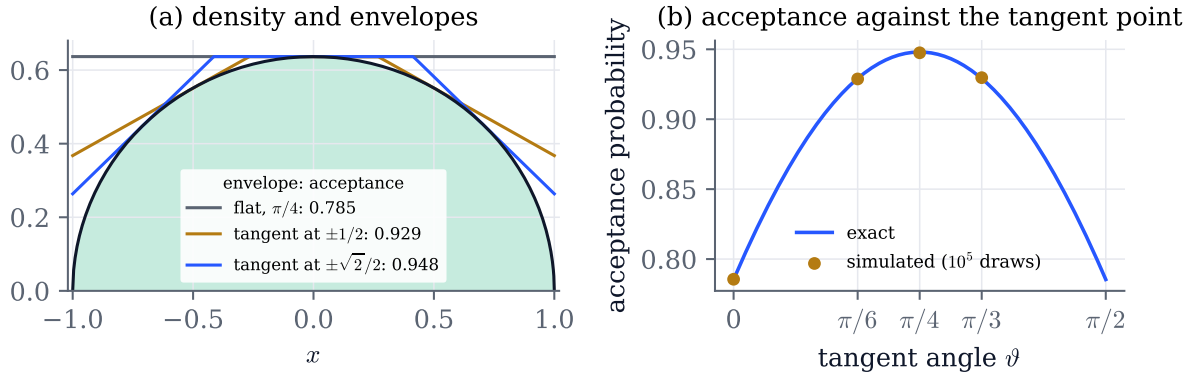

Figure 5.3: (a) The semicircle density with three envelopes and their acceptance probabilities. (b) The acceptance probability (5.3) against the tangent angle, with rates estimated from 10^5 accepted draws each.

▶ Interactive demo: move the tangent point of the envelope and compare the acceptance probability (5.3) with simulation.

5.5 Normal variates

For the normal law there are faster routes than a numerical N^{-1}. Both rest on the polar coordinates that compute $\int e^{-x^2/2}\mathrm{d}x = \sqrt{2\pi}$.

Proposition 5.12 (Box–Muller and the polar method). *(i) Let U_1, U_2 be independent $\mathcal{U}(0,1)$, $R = \sqrt{-2\log U_1}$ and $\Theta = 2\pi U_2$. Then $Z_1 = R\cos\Theta$ and $Z_2 = R\sin\Theta$ are independent $\mathcal{N}(0,1)$ (Box and Muller, 1958).*

(ii) Let (V_1, V_2) be uniform on the punctured disc $\{0 < v_1^2 + v_2^2 < 1\}$ and $S = V_1^2 + V_2^2$. Then $Z_i = V_i\sqrt{-2\log(S)/S}$, $i = 1, 2$, are independent $\mathcal{N}(0,1)$ (Marsaglia and Bray, 1964).

> *Such a pair is obtained by rejection: take $V_i = 2U_i - 1$ and accept if $0 < S < 1$, which happens with probability $\pi/4$.*

Proof. (i) $R^2 = -2\log U_1$ has $\mathbb{P}(R^2 > s) = \mathbb{P}(U_1 < e^{-s/2}) = e^{-s/2}$ for $s \geq 0$, so R has density $re^{-r^2/2}$ on $(0,\infty)$, and Θ is uniform on $(0,2\pi)$ and independent of R. The map $\phi(r,\theta) = (r\cos\theta, r\sin\theta)$ is a continuously differentiable bijection from $(0,\infty)\times(0,2\pi)$ onto the plane minus the half-line $\{(x,0) : x \geq 0\}$, a null set, with Jacobian determinant r. By the change-of-variables formula, $(Z_1, Z_2) = \phi(R,\Theta)$ has density, at $z = \phi(r,\theta)$,

$$\frac{1}{r}\cdot\frac{1}{2\pi}\,re^{-r^2/2} = \frac{1}{2\pi}e^{-(z_1^2+z_2^2)/2} = N'(z_1)\,N'(z_2),$$

the product of two standard normal densities.

(ii) Write $V = (\sqrt{S}\cos\Theta, \sqrt{S}\sin\Theta)$ with $\Theta \in (0,2\pi)$ (unique outside a null set). The disc has density $1/\pi$ and, in polar coordinates with $r = \sqrt{s}$, area element $r\,\mathrm{d}r\,\mathrm{d}\theta = \frac{1}{2}\mathrm{d}s\,\mathrm{d}\theta$, so (S,Θ) has density $\frac{1}{2\pi}$ on $(0,1)\times(0,2\pi)$: S and $\Theta/(2\pi)$ are independent $\mathcal{U}(0,1)$. Since $V_i/\sqrt{S}$ are $\cos\Theta$ and $\sin\Theta$, (Z_1, Z_2) is the Box–Muller pair built from $U_1 = S$, $U_2 = \Theta/(2\pi)$. Finally, $(2U_1 - 1, 2U_2 - 1)$ is uniform on $(-1,1)^2$, and given that it falls in the punctured disc, an event of probability $\pi/4$, it is uniform there (Theorem 5.10 with f uniform on the disc and g on the square). □

The polar method avoids the sine and cosine at the price of rejecting 21% of the pairs: it consumes on average $4/\pi \approx 1.27$ uniforms per normal, against one for Box–Muller and for inversion. It consumes a random number of uniforms, which breaks the structure of the low-discrepancy points of Chapter 11. Both methods are exact; with pseudo-random inputs their quality is that of the pairs (u_{2i}, u_{2i+1}). Two more exact samplers of the normal law, by rejection, are in Exercise 5.4 and Exercise 5.5.

Remark 5.13 (Tails and lattices). The resolution of the uniforms caps the normals: if $u_1 \geq \delta$, then $|z_i| \leq R \leq \sqrt{-2\log\delta}$. For 32-bit uniforms, $\delta \approx 2^{-32}$ and $|z| \leq 6.66$, a level that a true normal exceeds with probability 2.7×10^{-11}; for 53-bit uniforms the cap is 8.57. More seriously, Box–Muller transforms the lattice of pairs of an LCG into spirals: a line $u_2 = \alpha u_1 + \beta$ in the square becomes the curve $r = \sqrt{-2\log u_1}$, $\theta = 2\pi(\alpha u_1 + \beta)$. Neave (1973) showed that with multiplicative generators this distorts the tails: if $u_2 = au_1 \bmod 1$, then $u_2 = au_1$ whenever $u_1 < 1/a$, so the largest radii, which come from the smallest u_1, come with angles $2\pi a u_1$ close to 0. Figure 5.4(b) shows the effect for a full-period LCG with $m = 2^{12}$: all the normals lie on a few spiral arms inside the circle of radius $\sqrt{2\log(2m)}$.

Checking normal variates. Figure 5.5 runs the diagnostics of the course's homework on Box–Muller normals. With $n = 10^5$, one bin in 80 lies outside ±2 binomial standard deviations (about 3.6 expected), one lag in 20 falls outside the band for each of z_i and z_i^2, and 3 and 4 values fall below -4 and above 4 (about 3.2 expected on each side). The squares matter because variates can be uncorrelated but dependent, for instance with clustered magnitudes; the lag-one pairs mapped back by N should fill the square uniformly (Theorem 5.8(iii)).

▶ Interactive demo: feed Box–Muller with an LCG of your choice and watch the spirals; run the histogram and autocorrelation diagnostics for several generators.

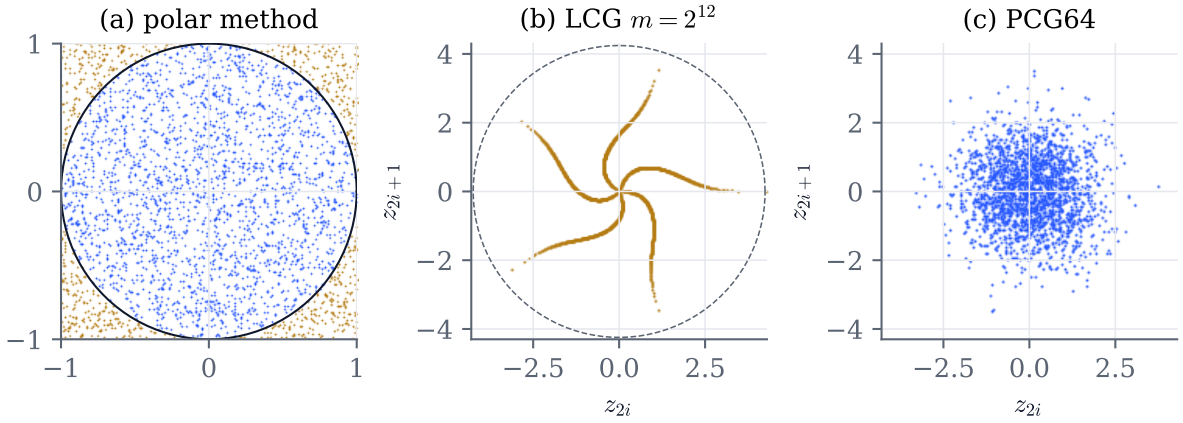

Figure 5.4: (a) The polar method: 3000 uniform pairs on $(-1,1)^2$; those inside the disc (cobalt) are kept. (b) Box–Muller applied to the 2048 consecutive pairs of the full-period LCG $m = 2^{12}$, $a = 1229$, $c = 1$ (uniforms $(x_i + \frac{1}{2})/m$); the dashed circle is the largest attainable radius. (c) Box–Muller applied to 2048 pairs from PCG64.

5.6 Correlated Gaussian vectors

A multi-asset model or a discretised path needs a Gaussian vector $X \sim \mathcal{N}(\mu, \Sigma)$ with a given covariance matrix Σ, symmetric and positive semidefinite. Independent standard normals and a square root of Σ suffice.

Proposition 5.14 (Cholesky and principal components)**.** *Let Σ be a symmetric positive semidefinite $d \times d$ matrix and $\mu \in \mathbb{R}^d$.*

(i) If A is a $d \times k$ matrix with $AA^\top = \Sigma$ and $Z \sim \mathcal{N}(0, I_k)$, then $\mu + AZ \sim \mathcal{N}(\mu, \Sigma)$.

(ii) (Cholesky.) *If Σ is positive definite, there is exactly one lower triangular L with positive diagonal and $LL^\top = \Sigma$. Column by column, for $j = 1, \dots, d$ and $i > j$,*

$$L_{jj} = \Big(\Sigma_{jj} - \sum_{k<j} L_{jk}^2\Big)^{1/2}, \qquad L_{ij} = \frac{1}{L_{jj}}\Big(\Sigma_{ij} - \sum_{k<j} L_{ik}L_{jk}\Big), \tag{5.4}$$

and every quantity under the square root is positive.

(iii) (Principal components.) *Let $\Sigma = V\Lambda V^\top$ with V orthogonal, columns $v_1, \dots, v_d$, and $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_d)$, $\lambda_1 \geq \cdots \geq \lambda_d \geq 0$. Then $A = V\Lambda^{1/2}$ satisfies $AA^\top = \Sigma$, so $X = \mu + \sum_{j=1}^{d} \sqrt{\lambda_j}\, Z_j v_j$ is $\mathcal{N}(\mu, \Sigma)$. Its truncation $X_r = \mu + \sum_{j \leq r} \sqrt{\lambda_j} Z_j v_j$ has $\mathbb{E}|X - X_r|^2 = \sum_{j>r} \lambda_j$, and no random vector Y with values in an affine subspace of dimension r does better: $\mathbb{E}|X - Y|^2 \geq \sum_{j>r} \lambda_j$.*

Proof. (i) $\mu + AZ$ is an affine image of a Gaussian vector, hence Gaussian, with mean μ and covariance $A\, I_k\, A^\top = \Sigma$; a Gaussian law is determined by its mean and covariance (its characteristic function is $e^{i\xi\cdot\mu - \frac{1}{2}\xi^\top \Sigma \xi}$).

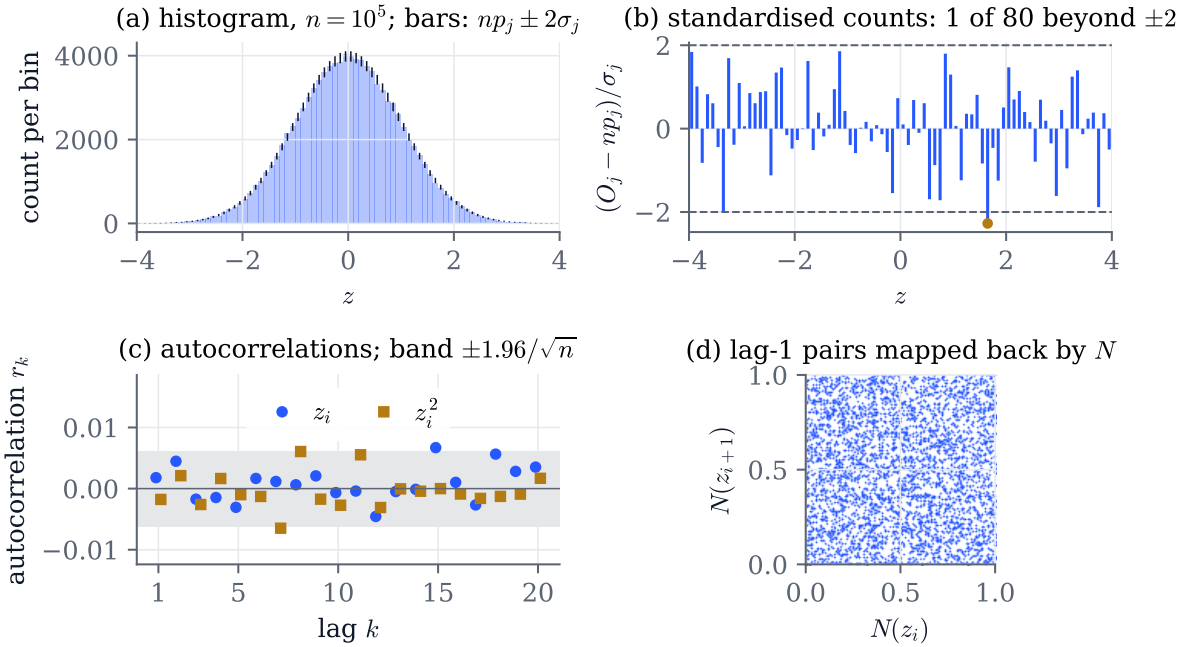

Figure 5.5: Diagnostics of $n = 10^5$ Box–Muller normals built from PCG32 uniforms: (a) histogram with bins of width 0.1 and the binomial intervals $np_j \pm 2\sigma_j$; (b) standardised counts $(O_j - np_j)/\sigma_j$; (c) autocorrelations of z_i and z_i^2 at lags 1 to 20; (d) the first 4000 lag-one pairs mapped back by N.

(ii) Induction on d; the case $d = 1$ is $L_{11} = \sqrt{\Sigma_{11}}$. Write $\Sigma = \begin{pmatrix} \Sigma_{11} & \sigma \\ \sigma^\top & s \end{pmatrix}$ with Σ_{11} of size $d-1$, positive definite, and look for $L = \begin{pmatrix} L_{11} & 0 \\ \ell^\top & \lambda \end{pmatrix}$. Then $LL^\top = \Sigma$ if and only if $L_{11}L_{11}^\top = \Sigma_{11}$, $L_{11}\ell = \sigma$ and $\ell^\top\ell + \lambda^2 = s$. By induction the first equation has exactly one solution with positive diagonal; L_{11} is then invertible, so $\ell = L_{11}^{-1}\sigma$ is unique, and $\lambda^2 = s - \sigma^\top\Sigma_{11}^{-1}\sigma$. This is positive: it equals $w^\top\Sigma w$ for $w = (-\Sigma_{11}^{-1}\sigma, 1) \neq 0$. So $\lambda > 0$ exists and is unique. Solving $L_{11}\ell = \sigma$ by forward substitution and writing out λ^2 gives (5.4) for the last row; the other rows come from the induction.

(iii) $AA^\top = V\Lambda V^\top = \Sigma$, and (i) applies. Next, $X - X_r = \sum_{j>r}\sqrt{\lambda_j}Z_j v_j$ with orthonormal v_j, so $\mathbb{E}|X - X_r|^2 = \sum_{j>r}\lambda_j\mathbb{E}Z_j^2 = \sum_{j>r}\lambda_j$. Let Y take values in $a + \mathcal{V}$, $\mathcal{V}$ a subspace of dimension r, and let P be the orthogonal projection onto $\mathcal{V}$. For $y \in a + \mathcal{V}$, Pythagoras gives $|x - y|^2 \geq |(I - P)(x - a)|^2$, so

$$\mathbb{E}|X - Y|^2 \geq \mathbb{E}|(I-P)(X-a)|^2 \geq \mathbb{E}|(I-P)(X-\mu)|^2 = \operatorname{tr}\left((I-P)\Sigma\right) = \sum_j \lambda_j - \operatorname{tr}(P\Sigma),$$

where the middle step drops $|(I-P)(\mu-a)|^2$ (the cross term has mean zero). Take an orthonormal basis $q_1, \ldots, q_r$ of $\mathcal{V}$; then $\operatorname{tr}(P\Sigma) = \sum_k q_k^\top\Sigma q_k = \sum_j \lambda_j w_j$ with $w_j = \sum_k (v_j \cdot q_k)^2$. Each $w_j \in [0, 1]$ by Bessel's inequality, and $\sum_j w_j = \sum_k |q_k|^2 = r$. Since $\lambda_j \geq \lambda_r$ for $j \leq r$ and $\lambda_j \leq \lambda_r$ for $j > r$,

$$\sum_j \lambda_j w_j - \sum_{j\leq r}\lambda_j = \sum_{j\leq r}\lambda_j(w_j - 1) + \sum_{j>r}\lambda_j w_j \leq \lambda_r\Big(\sum_j w_j - r\Big) = 0.$$

Hence $\operatorname{tr}(P\Sigma) \le \sum_{j\le r}\lambda_j$ and $\mathbb{E}|X-Y|^2 \ge \sum_{j>r}\lambda_j$. □

The Cholesky factor costs about $d^3/3$ operations once and $d^2/2$ per vector, and its j-th component uses only $Z_1,\dots,Z_j$; for a Brownian path on a grid it is the sum of independent increments (Exercise 5.6). The principal-components factor puts as much variance as possible into the first Z_j, which matters for quasi-Monte Carlo and stratification (Theorem 11.13, Chapter 7). If Σ is not positive definite, the recursion (5.4) meets a nonpositive pivot (otherwise it would produce a factor). A zero pivot means that a leading principal submatrix is singular: Σ may be singular and positive semidefinite (two perfectly correlated assets; the principal-components factor exists) or not positive semidefinite at all, which the eigenvalues decide. A negative pivot makes a leading principal minor negative, so it certifies a "correlation matrix" that no random vector can have.

Example 5.15 (Three assets)**.** For the correlation matrix with $\rho_{12}=0.5$, $\rho_{13}=0.3$, $\rho_{23}=0.6$, (5.4) gives $L_{21}=0.5$, $L_{22}=\sqrt{0.75}\approx 0.866$, $L_{31}=0.3$, $L_{32}=(0.6-0.15)/\sqrt{0.75}\approx 0.520$ and $L_{33}=\sqrt{1-0.09-0.27}=0.8$. Thus $X_3=0.3Z_1+0.520Z_2+0.8Z_3$. The eigenvalues are 1.944, 0.707 and 0.349: the first principal component carries 65% of the total variance 3, and the first two carry 88%. In the code, `cholesky` runs (5.4) and raises an error on a nonpositive pivot, and `pca_factor` returns $V\Lambda^{1/2}$.

▶ Interactive demo: choose three correlations, compare the Cholesky and principal-components factors, and see what happens when the matrix is not positive semidefinite.

Further reading

Glasserman (2003), Chapter 2, covers the same ground, with more on implementing LCGs and combined generators (Sections 2.1.3–2.1.5) and on normal variates (Section 2.3). Knuth (1997), Chapter 3, is the classic account of uniform generators and their tests. Devroye (1986) is the encyclopaedic reference for non-uniform variates (inversion and rejection: Chapter II, Sections 2 and 3). L'Ecuyer and Simard (2007) describe modern testing, and O'Neill (2014) explains what makes a generator fast and sound.

Chapter 6

Monte Carlo pricing and its error

Every price in these notes is a discounted expectation under a risk-neutral measure (Theorem 1.3). For a European claim with payoff $f(S_T)$ it is $e^{-rT}\mathbb{E}^{\mathbb{Q}}[f(S_T)]$; for a path-dependent or multi-asset claim the payoff depends on more than one random number, but the price is still an expectation. When that expectation has no closed form we estimate it by simulation: generate many independent scenarios under $\mathbb{Q}$, compute the discounted payoff in each, and average. This is the Monte Carlo method, brought to option pricing by Boyle (1977).

This chapter is about what the average is worth. A simulation reports three numbers: an estimate, its standard error and the time it took. The central limit theorem connects the first two: the error shrinks like $n^{-1/2}$ in the number n of scenarios, whatever the dimension of the problem and however irregular the payoff. That is the strength of the method (Section 6.6) and its weakness: one more correct digit costs a hundred times more work. The computing time enters through the efficiency of an estimator (Section 6.4), which is what the variance reduction methods of Chapters 8 to 12 improve. Much of the chapter is about when the reported standard error should not be believed: small samples, skewed payoffs, heavy tails, dependent samples and biased estimators.

Samples are drawn under some probability measure; in pricing it is the risk-neutral measure $\mathbb{Q}$, but to lighten notation we write $\mathbb{P}$ and $\mathbb{E}$ for whichever measure the samples come from. X is a real random variable, typically a discounted payoff, and $\mu = \mathbb{E}[X]$ is the number we want. $\mathcal{N}(m, v)$ is the normal law, N the standard normal distribution function, and t_k Student's t law with k degrees of freedom.

6.1 The Monte Carlo estimator

Our running example is the Black–Scholes model with a dividend yield. Under $\mathbb{Q}$,

$$S_T = S_0 \exp\left(\nu T + \sigma\sqrt{T}\, Z\right), \qquad Z \sim \mathcal{N}(0,1), \qquad \nu = r - y - \tfrac{\sigma^2}{2}, \tag{6.1}$$

and the call price is $C = e^{-rT}\mathbb{E}^{\mathbb{Q}}\left[(S_T - K)^+\right]$. From independent standard normals $Z_1, Z_2, \ldots$ (Chapter 5) we obtain independent copies $X_i = e^{-rT}\left(S_0 e^{\nu T + \sigma\sqrt{T}Z_i} - K\right)^+$ of the discounted payoff $X = e^{-rT}(S_T - K)^+$, whose mean is C.

Definition 6.1 (Monte Carlo estimator)**.** Let X be an integrable random variable with mean $\mu = \mathbb{E}[X]$, and let $X_1, X_2, \dots$ be independent random variables with the law of X. The *Monte Carlo estimator* of μ from n draws is

$$\hat{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

For $n \geq 2$, the *sample variance*, the *sample standard deviation* and the *standard error* are

$$\hat{s}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \hat{X}_n\right)^2, \qquad \hat{s}_n = \sqrt{\hat{s}_n^2}, \qquad \frac{\hat{s}_n}{\sqrt{n}}.$$

For $\delta \in (0,1)$ let $z_\delta = N^{-1}(1-\delta/2)$, so that $z_{0.05} = 1.96$. The interval

$$I_n = \left[\hat{X}_n - z_\delta \frac{\hat{s}_n}{\sqrt{n}},\ \hat{X}_n + z_\delta \frac{\hat{s}_n}{\sqrt{n}}\right] \tag{6.2}$$

is the *nominal* $(1-\delta)$ *confidence interval* for μ.

The word *nominal* is a warning: whether $\mathbb{P}(\mu \in I_n)$ is close to $1-\delta$ is a theorem with hypotheses, the subject of Section 6.2.

An expectation that we can simulate is an integral over a cube. If one draw of X consumes a fixed number d of independent uniform random numbers, then $X = g(U_1, \dots, U_d)$ for some function g, and

$$\mu = \int_{[0,1]^d} g(u)\,\mathrm{d}u. \tag{6.3}$$

For the call, $d = 1$ with $Z = N^{-1}(U)$; for a payoff observed at m dates on k assets, $d = mk$. The dimension d decides the fate of deterministic quadrature, but it enters the error of Monte Carlo only through the variance of g (Section 6.6).

Proposition 6.2 (Unbiasedness and consistency)**.** *Let X be integrable with mean μ.*

(i) $\mathbb{E}[\hat{X}_n] = \mu$ *for every* $n \geq 1$.

(ii) $\hat{X}_n \to \mu$ *almost surely as* $n \to \infty$.

If moreover $\mathbb{E}[X^2] < \infty$ *and* $\sigma^2 = \mathrm{Var}(X)$*, then:*

(iii) $\mathbb{E}[(\hat{X}_n - \mu)^2] = \mathrm{Var}(\hat{X}_n) = \sigma^2/n$*, and* $\mathbb{E}[\hat{s}_n^2] = \sigma^2$ *for* $n \geq 2$*;*

(iv) $\hat{s}_n \to \sigma$ *almost surely.*

Proof. (i) is linearity of expectation. (ii) is the strong law of large numbers, whose only hypothesis for independent identically distributed variables is $\mathbb{E}|X| < \infty$ (Durrett, 2019, Section 2.4).

(iii) Variances of independent variables add, so $\mathrm{Var}(\hat{X}_n) = n\sigma^2/n^2$, and $\mathbb{E}[(\hat{X}_n - \mu)^2] = \mathrm{Var}(\hat{X}_n)$ by (i). Expanding the squares around $\hat{X}_n$,

$$\sum_{i=1}^{n}(X_i - \hat{X}_n)^2 = \sum_{i=1}^{n}(X_i - \mu)^2 - n(\hat{X}_n - \mu)^2,$$

whose expectation is $n\sigma^2 - \sigma^2$. Dividing by $n-1$ gives $\mathbb{E}[\hat{s}_n^2] = \sigma^2$.

(iv) Apply the strong law to X and to X^2, which is integrable by assumption: $\frac{1}{n}\sum_i X_i^2 \to \mathbb{E}[X^2]$ and $\hat{X}_n \to \mu$ almost surely. Hence $\hat{s}_n^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_i X_i^2 - \hat{X}_n^2\right) \to \mathbb{E}[X^2] - \mu^2 = \sigma^2$, and $\hat{s}_n \to \sigma$ by continuity of the square root. □

Two things (iii) does not say. First, $\hat{s}_n$ is biased low: $\mathbb{E}[\hat{s}_n] \le \sqrt{\mathbb{E}[\hat{s}_n^2]} = \sigma$ by Jensen's inequality, although the bias vanishes as $n \to \infty$. Second, the estimate and its standard error come from the same sample, and in general they are not independent (for a normal sample they are). For a call, a sample that happens to contain few large payoffs has both a low mean and a small $\hat{s}_n$: the interval is short exactly when it is in the wrong place. Section 6.2 shows the consequence.

Example 6.3 (Bare-bones Monte Carlo for the test case)**.** Take the course's test case $S_0 = K = 100$, $T = 1$, $r = 1\%$, $y = 2\%$, $\sigma = 20\%$, for which $\mathbb{E}^{\mathbb{Q}}[S_T] = S_0 e^{(r-y)T} = 99.0050$ and $C = 7.36429$. The course's *bare-bones* Monte Carlo draws n standard normals, builds S_T from (6.1), and reports the estimate and the standard error for $\mathbb{E}^{\mathbb{Q}}[S_T]$, for $\mathbb{E}^{\mathbb{Q}}[(S_T - K)^+]$ and for C. With $n = 2^{20}$, our implementation (NumPy's `default_rng`, seed 2024) returns $\hat{C} = 7.3786$ with standard error 0.0123, an error of 1.17 standard errors, and 99.0263 with standard error 0.0196 for $\mathbb{E}^{\mathbb{Q}}[S_T]$. The exact standard deviation of the discounted call payoff is $\sigma = 12.570$ (Exercise 6.1), so the standard error at $n = 2^{20}$ is $12.570/1024 = 0.0123$, as reported. To be within one cent of C with probability 95% we need $(1.96 \times 12.570/0.01)^2 \approx 6.1 \times 10^6$ paths.

In code

```
z = np.random.default_rng(seed).standard_normal(n)
ST = S * np.exp((r - y - 0.5 * sigma**2) * T + sigma * np.sqrt(T) * z)
X = np.exp(-r * T) * np.maximum(ST - K, 0.0) # discounted payoffs
C_hat, se = X.mean(), X.std(ddof=1) / np.sqrt(n) # estimate, standard error
```

`bs_call_bbmc` in `code/src/mfe230d/montecarlo.py` does this for the three targets of Theorem 6.3; `mc_estimate` and `batch_estimate` return estimates, standard errors and confidence intervals for any array of samples.

6.2 The central limit theorem and confidence intervals

Theorem 6.4 (Central limit theorem for Monte Carlo)**.** *Let $\mathbb{E}[X^2] < \infty$ and $\sigma^2 = \mathrm{Var}(X) > 0$, and let $T_n = \sqrt{n}\,(\hat{X}_n - \mu)/\hat{s}_n$ on the event $\{\hat{s}_n > 0\}$ and $T_n = 0$ otherwise.*

(i) $\sqrt{n}\,(\hat{X}_n - \mu)/\sigma$ converges in distribution to $\mathcal{N}(0,1)$.

(ii) T_n converges in distribution to $\mathcal{N}(0,1)$, and the interval (6.2) has asymptotically its nominal coverage: $\mathbb{P}(\mu \in I_n) \to 1 - \delta$.

(iii) If X is normal, then T_n has the law t_{n-1} for every $n \ge 2$, and the interval (6.2) with z_δ replaced by the $(1-\delta/2)$-quantile $t_{n-1,\delta}$ of t_{n-1} has coverage exactly $1 - \delta$.

If instead $\mathbb{E}|X| < \infty$ but $\mathbb{E}[X^2] = \infty$, then $\hat{X}_n \to \mu$ almost surely as before, but $\hat{s}_n \to \infty$ almost surely.

Proof. (i) is the central limit theorem for independent identically distributed variables with finite variance (Durrett, 2019, Section 3.4.1).

(ii) By Theorem 6.2(iv), $\sigma/\hat{s}_n \to 1$ almost surely; in particular $\mathbb{P}(\hat{s}_n = 0) \to 0$. Off that event, $T_n = (\sigma/\hat{s}_n) \cdot \sqrt{n}(\hat{X}_n - \mu)/\sigma$ is the product of a sequence that converges in distribution to $\mathcal{N}(0,1)$ and one that converges in probability to 1, so $T_n \Rightarrow \mathcal{N}(0,1)$ by Slutsky's lemma (van der Vaart, 1998, Section 2.1, Lemma 2.8). On $\{\hat{s}_n > 0\}$ the events $\{\mu \in I_n\}$ and $\{|T_n| \le z_\delta\}$ coincide, and the normal distribution function is continuous, so $\mathbb{P}(\mu \in I_n) \to \mathbb{P}(|Z| \le z_\delta) = 1 - \delta$ for $Z \sim \mathcal{N}(0,1)$.

(iii) For a normal sample, $\hat{X}_n$ and $\hat{s}_n^2$ are independent, $\sqrt{n}(\hat{X}_n - \mu)/\sigma \sim \mathcal{N}(0,1)$ and $(n-1)\hat{s}_n^2/\sigma^2$ has the χ^2 law with $n-1$ degrees of freedom (Casella and Berger, 2002, Theorem 5.3.1). So T_n is a standard normal divided by the square root of an independent $\chi^2_{n-1}/(n-1)$ variable, which is the definition of t_{n-1}.

Finally, suppose that $\mathbb{E}|X| < \infty = \mathbb{E}[X^2]$. The strong law still gives $\hat{X}_n \to \mu$. For $M > 0$, the variables $\min(X_i^2, M)$ are bounded, so by the strong law $\liminf_n \frac{1}{n}\sum_i X_i^2 \ge \lim_n \frac{1}{n}\sum_i \min(X_i^2, M) = \mathbb{E}[\min(X^2, M)]$ almost surely. As $M \uparrow \infty$ through the integers, $\mathbb{E}[\min(X^2, M)] \uparrow \mathbb{E}[X^2] = \infty$ by monotone convergence, and a countable intersection of almost sure events is almost sure; hence $\frac{1}{n}\sum_i X_i^2 \to \infty$ almost surely. Since $\hat{X}_n^2 \to \mu^2 < \infty$, $\hat{s}_n^2 = \frac{n}{n-1}(\frac{1}{n}\sum_i X_i^2 - \hat{X}_n^2) \to \infty$. □

Part (i) says that $\hat{X}_n \approx \mu + \sigma Z/\sqrt{n}$ with Z standard normal. Three consequences shape the rest of Part II. The error is of order $\sigma/\sqrt{n}$, so each additional decimal digit costs a factor of 100 in the number of paths. The rate does not depend on the dimension d in (6.3) or on the smoothness of g: only $\sigma^2 < \infty$ is needed. And the constant σ is what can be improved: halving σ is worth four times as many paths, which is what variance reduction does (Chapter 8).

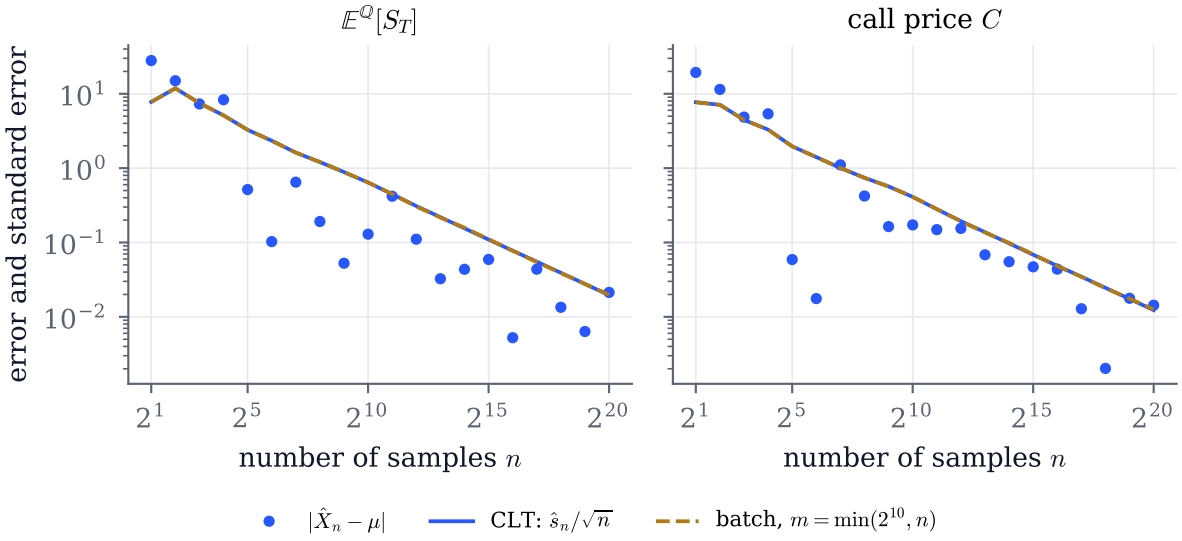

Figure 6.1: Bare-bones Monte Carlo on the test case (Theorem 6.3): one stream of 2^{20} normal draws (seed 2024), read at $n = 2, 4, \ldots, 2^{20}$. Dots: the actual errors for $\mathbb{E}^{\mathbb{Q}}[S_T]$ (left) and for the call price (right). Solid: the standard error $\hat{s}_n/\sqrt{n}$. Dashed: the batch standard error of Theorem 6.8 with $\min(2^{10}, n)$ batches, which coincides with $\hat{s}_n/\sqrt{n}$ for $n \le 2^{10}$. Both standard errors fall on a line of slope $-\frac{1}{2}$ and describe the size of the errors from a few dozen samples on.

▶ Interactive demo: rerun the bare-bones Monte Carlo of Figure 6.1 with other seeds, strikes

and sample sizes, and watch the errors scatter around the standard error.

6.2.1 Small samples and Student's t

Remark 6.5 (When to use t quantiles). For the sample sizes of pricing, the quantiles of t_{n-1} and of $\mathcal{N}(0,1)$ are indistinguishable: $t_{999,0.05} = 1.9623$ against $z_{0.05} = 1.9600$. Student's t matters where the number of independent observations is small: a handful of batches (Theorem 6.8), a few independent runs of an expensive model, or a comparison across a few seeds. With 10 observations, $t_{9,0.05} = 2.262$, and an interval built with 1.96 instead covers a normal mean with probability 91.8%, not 95%. Part (iii) of Theorem 6.4 is exact only for normal X; for a skewed payoff the t correction accounts for the randomness of $\hat{s}_n$ but not for the skewness.

6.2.2 How accurate is the normal approximation?

The central limit theorem is a statement about $n \to \infty$. How large n must be depends on the shape of the law of X, and the Berry–Esseen theorem quantifies it.

Remark 6.6 (Berry–Esseen). If $\mathbb{E}|X|^3 < \infty$, let $\rho = \mathbb{E}|X-\mu|^3$. There is an absolute constant C_{BE} such that for all $n \geq 1$ and $x \in \mathbb{R}$,

$$\left|\mathbb{P}\Big(\frac{\sqrt{n}(\hat{X}_n - \mu)}{\sigma} \leq x\Big) - N(x)\right| \leq \frac{C_{\mathrm{BE}}\,\rho}{\sigma^3\sqrt{n}}. \tag{6.4}$$

Durrett (2019), Section 3.4.4, proves it with $C_{\mathrm{BE}} = 3$; the best proved bound is $C_{\mathrm{BE}} < 0.4748$ (Shevtsova, 2011), and Esseen's lower bound 0.4097 shows that the constant cannot be much smaller (see the same paper). A bound of the same form, with another constant, holds for the studentised statistic T_n (Bentkus and Götze, 1996). The ratio ρ/σ^3 measures the asymmetry of the payoff. For the test case it is 1.70 for S_T and 2.48 for the at-the-money call. For an indicator of an event of probability p, such as a digital option,

$$\frac{\rho}{\sigma^3} = \frac{p(1-p)^3 + (1-p)p^3}{\big(p(1-p)\big)^{3/2}} = \frac{(1-p)^2 + p^2}{\sqrt{p(1-p)}} \approx \frac{1}{\sqrt{p}} \quad (p \to 0),$$

which is 9.85 for $p = 1\%$. The bound (6.4) is a worst case and is pessimistic, but its scaling is right: for a rare event the bound is about $C_{\mathrm{BE}}/\sqrt{np}$, so what must be large is the expected number of hits np, not n. In the extreme case where no scenario hits, $\hat{X}_n = 0$, $\hat{s}_n = 0$ and $I_n = \{0\}$: the interval claims certainty (Exercise 6.2).

Figure 6.2 (left) measures the coverage directly. For the at-the-money call the nominal 95% is reached, within the $\pm 1\%$ accuracy of the measurement, from about $n = 2^7$ on; with 8 samples it is about 82%. For the digital paying in 1% of the scenarios the coverage is below 30% for $n \leq 2^5$, where most samples contain no hit at all, and approaches 95% once np is in the tens. Coverage below the nominal level is the typical failure for right-skewed payoffs, and the mechanism is the one noted after Theorem 6.2: samples that miss the large payoffs underestimate the mean and the standard error at the same time.

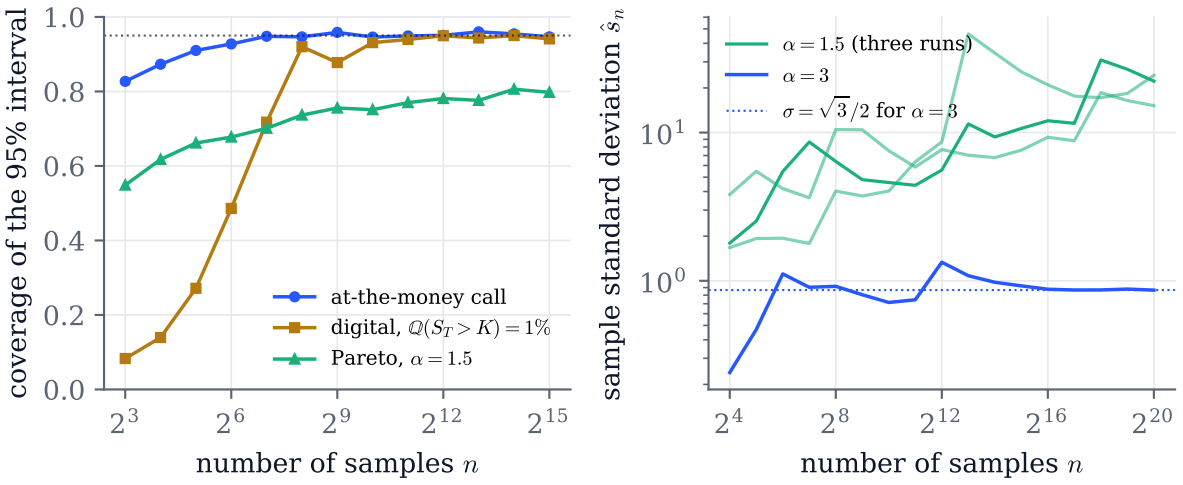

Figure 6.2: Left: coverage of the nominal 95% interval (6.2), estimated from 2000 independent samples of each size n, for the discounted at-the-money call of the test case, a digital paying 1 when $S_T > 154.54$ (probability 1% under $\mathbb{Q}$), and a Pareto variable with $\alpha = 1.5$ (infinite variance). Right: the sample standard deviation $\hat{s}_n$ along single runs, for Pareto variables with $\alpha = 1.5$ (three runs), for which $\hat{s}_n \to \infty$ (Theorem 6.4), and $\alpha = 3$, for which $\hat{s}_n \to \sigma$.

6.2.3 Heavy tails and infinite variance

The last statement of Theorem 6.4 describes what survives when $\mathbb{E}[X^2] = \infty$: unbiasedness and the strong law, but not the rate $n^{-1/2}$ or the standard error. The standard example is the Pareto law, $\mathbb{P}(X > x) = x^{-\alpha}$ for $x \geq 1$, which can be sampled by inversion as $X = U^{-1/\alpha}$ (Chapter 5). Its mean $\alpha/(\alpha - 1)$ is finite for $\alpha > 1$, and $\mathbb{E}[X^2] = 1 + \int_1^\infty 2x \cdot x^{-\alpha}\,\mathrm{d}x$ is infinite for $\alpha \leq 2$. In Figure 6.2 (right), $\hat{s}_n$ creeps upwards with n for $\alpha = 1.5$, jumping whenever a new extreme sample arrives; the reported standard error does not estimate the error.

Remark 6.7 (What breaks with infinite variance). Let X be Pareto with $1 < \alpha < 2$.

(a) *The rate.* $n^{1-1/\alpha}(\hat{X}_n - \mu)$ converges in distribution to a non-Gaussian stable law (Durrett, 2019, Section 3.8), so the error decays like $n^{-(1-1/\alpha)}$: like $n^{-1/3}$ for $\alpha = 1.5$, against $n^{-1/2}$ with finite variance. A heuristic shows why. The largest of n samples is of order $n^{1/\alpha}$, since $\mathbb{P}(\max_{i \leq n} X_i \leq t n^{1/\alpha}) = (1 - t^{-\alpha}/n)^n \to e^{-t^{-\alpha}}$, and a single sample of that size moves the average by $n^{1/\alpha - 1}$, which is much larger than $n^{-1/2}$ when $\alpha < 2$. The fluctuations of the average are dominated by its few largest terms: most runs miss the rare huge values that carry part of the mean, so $\hat{X}_n$ is usually below μ and occasionally far above it.

(b) *The interval.* T_n converges in distribution to $\mathcal{N}(0,1)$ if and only if X is in the domain of attraction of the normal law (Giné et al., 1997). For $\alpha < 2$ it is not; in fact T_n then converges in distribution to a non-normal law (Logan et al., 1973), so the coverage of I_n tends to a limit other than 95%. In Figure 6.2 it creeps up from 55% to about 80% for $\alpha = 1.5$, and it is still about 80% at $n = 2^{18}$. Infinite variance alone does not decide this. For $\alpha = 2$ the variance is infinite, yet $(\hat{X}_n - \mu)\sqrt{n/\log n}$ is asymptotically standard

normal,[1] so X is in the domain of attraction of the normal law, $T_n \Rightarrow \mathcal{N}(0,1)$, and the coverage tends to 95%, slowly.

(c) *Batching does not help*: every batch mean has infinite variance too.

In pricing, infinite variance arises from importance sampling with too narrow a sampling law (Theorem 9.3), and from payoffs that grow like a power of the underlying in models with moment explosions: in the Heston model, depending on the parameters, a moment $\mathbb{E}[S_T^p]$ of order $p > 1$ is finite only for maturities T below a critical time $T^*(p)$ (Andersen and Piterbarg, 2007), so a claim with a finite price can have a payoff with infinite variance (Chapter 19). The symptoms are those of Figure 6.2: a standard error that grows when it should shrink, and estimates that jump when one sample dominates the sum.

▶ Interactive demo: the coverage of the nominal interval for calls and digitals of any strike, the histogram of T_n against the normal density, and running means and standard deviations of Pareto samples for any α.

6.3 Batch means

The standard error $\hat{s}_n/\sqrt{n}$ is justified by independence of the samples and by the estimator being their plain average. Several methods of Part II violate one or the other. Moment matching rescales all the samples with their own sample moments, which couples them (Theorem 8.7); randomised quasi-Monte Carlo points are deliberately dependent (Theorem 11.11); and samples produced by a Markov chain are correlated. The remedy used throughout the course is to run the whole procedure m times independently and to measure the error by the spread of the m results.

Proposition 6.8 (Batch means)**.** *Let $Y_1, \dots, Y_m$, $m \geq 2$, be independent and identically distributed outputs of m independent runs (*batches*) of an estimator, each run using k samples, with $\mathbb{E}[Y_1^2] < \infty$. Let*

$$\bar{Y} = \frac{1}{m}\sum_{j=1}^{m} Y_j, \qquad \hat{V}_m = \frac{1}{m(m-1)}\sum_{j=1}^{m} (Y_j - \bar{Y})^2.$$

(i) $\mathbb{E}[\hat{V}_m] = \mathrm{Var}(\bar{Y})$*, whatever the dependence between the samples inside a batch. The* batch standard error *is* $\sqrt{\hat{V}_m}$.

(ii) For plain Monte Carlo, with $n = mk$ samples X_i as in Theorem 6.1 split into consecutive batches $Y_j = \frac{1}{k}\sum_{i=(j-1)k+1}^{jk} X_i$, one has $\bar{Y} = \hat{X}_n$ and $\mathbb{E}[\hat{V}_m] = \sigma^2/n = \mathbb{E}[\hat{s}_n^2/n]$.

(iii) Suppose that, as $k \to \infty$ with m fixed, $\sqrt{k}\,(Y_1 - \mu)$ converges in distribution to $\mathcal{N}(0, v)$ for some $v > 0$ (a bias of order $o(k^{-1/2})$ inside each batch is allowed). Then

$$\frac{\bar{Y} - \mu}{\sqrt{\hat{V}_m}} \;\Rightarrow\; t_{m-1},$$

[1] Truncate at $b_n = \sqrt{n \log n}$. With probability at least $1 - n\,b_n^{-2} = 1 - 1/\log n$ no sample exceeds b_n. The truncated variables $X\mathbf{1}_{\{X \leq b_n\}}$ have variance $\log n + O(\log\log n)$, satisfy Lindeberg's condition (Durrett, 2019, Section 3.4.2), and their mean differs from μ by $\mathbb{E}[X\mathbf{1}_{\{X > b_n\}}] = 2/b_n$, which is $o(b_n/n)$.

so the interval $\bar{Y} \pm t_{m-1,\delta}\sqrt{\hat{V}_m}$ *has asymptotic coverage* $1-\delta$*. In the setting of (ii) the hypothesis holds with* $v = \sigma^2$ *whenever* $0 < \sigma^2 < \infty$*.*

Proof. (i) $\hat{V}_m$ is $1/m$ times the sample variance of the independent identically distributed Y_j, which is unbiased for $\mathrm{Var}(Y_1)$ by Theorem 6.2(iii); and $\mathrm{Var}(\bar{Y}) = \mathrm{Var}(Y_1)/m$.

(ii) The batches use disjoint sets of independent samples, so the Y_j are independent, identically distributed, with variance σ^2/k; averaging them averages all n samples. By (i), $\mathbb{E}[\hat{V}_m] = \sigma^2/(km)$, and $\mathbb{E}[\hat{s}_n^2] = \sigma^2$.

(iii) Let $W_j = \sqrt{k}\,(Y_j - \mu)/\sqrt{v}$. The ratio in question is $g(W_1, \dots, W_m)$, where

$$g(w) = \frac{\sqrt{m}\,\bar{w}}{\left(\frac{1}{m-1}\sum_j (w_j - \bar{w})^2\right)^{1/2}}, \qquad \bar{w} = \frac{1}{m}\sum_j w_j,$$

because the factors $\sqrt{k}$ and $\sqrt{v}$ cancel. The W_j are independent, so the characteristic function of $(W_1, \dots, W_m)$ is the product of theirs, and each factor converges to $e^{-u_j^2/2}$; by Lévy's continuity theorem in $\mathbb{R}^m$ (Durrett, 2019, Section 3.10), $(W_1, \dots, W_m)$ converges in distribution to a vector $(Z_1, \dots, Z_m)$ of independent standard normals. The map g (set to 0 on the diagonal $\{w_1 = \cdots = w_m\}$, as T_n in Theorem 6.4) is continuous except on that diagonal, which has probability zero under the limit, so by the mapping theorem (Billingsley, 1999, Theorem 2.7) $g(W)$ converges in distribution to $g(Z)$, whose law is t_{m-1} by Theorem 6.4(iii). The coverage follows as in Theorem 6.4(ii), the t_{m-1} distribution function being continuous. For plain Monte Carlo, $\sqrt{k}(Y_1 - \mu) \Rightarrow \mathcal{N}(0, \sigma^2)$ is Theorem 6.4(i) applied to one batch. □

The course uses $m = 2^{10}$ batches. With independent samples, batching changes nothing on average, by (ii), but the batch standard error is noisier: it rests on $m-1$ degrees of freedom instead of $n-1$ (Exercise 6.3). With $m = 2^{10}$ its relative noise is about 2% and $t_{1023,0.05} = 1.9623$ is indistinguishable from 1.96; with $m = 10$, the quantile 2.262 makes the interval noticeably wider. Both m and k must be large: m for an accurate standard error, k for nearly normal batch means. And each batch must be an independent run of the *whole* procedure: moment matching, for instance, must be applied inside each batch, or the batches are coupled again. For plain Monte Carlo the two standard errors agree (Figure 6.1); with dependent samples the plain one can be wrong by a large factor (Exercise 6.4).

6.4 Cost, efficiency and bias

Variance alone does not rank estimators: a method that halves the variance but triples the cost per sample is worse than plain Monte Carlo run three times longer.

Definition 6.9 (Efficiency)**.** Consider an unbiased estimator obtained by averaging independent replications, each with variance σ^2 and expected computing cost $\tau > 0$. Its *efficiency* is $1/(\sigma^2\tau)$. The *relative efficiency* of method 2 with respect to method 1 is

$$\mathcal{E}_{2,1} = \frac{\sigma_1^2\,\tau_1}{\sigma_2^2\,\tau_2}.$$

A computing budget c buys c/τ_j replications of method j, hence an estimator with variance $\sigma_j^2\tau_j/c$. So $\mathcal{E}_{2,1}$ is the ratio of the variances at equal budget, and equally the ratio of the budgets needed for equal variance; method 2 is better if $\mathcal{E}_{2,1} > 1$. The work-normalised variance $\sigma^2\tau$ is the quantity to make small (Glasserman, 2003, Section 1.1.3). If method 2 cuts the standard error at equal n by a factor q and costs λ times more per replication, then $\mathcal{E}_{2,1} = 1/(\lambda q^2)$: cutting the error by a factor $q = \frac{1}{4}$ at double cost gives $\mathcal{E}_{2,1} = 1/(2 \times 0.25^2) = 8$, while $q = 0.8$ at double cost gives 0.78, a loss.

Remark 6.10 (Random costs and overheads). When the cost of a replication is random (paths stopped at a barrier, rejection sampling), a budget c completes a random number $N(c)$ of replications. If the pairs (X_i, τ_i) of output and cost are independent and identically distributed, $\tau_i > 0$ has a finite mean and $\sigma^2 < \infty$, then $\sqrt{c}\,(\hat{X}_{N(c)} - \mu) \Rightarrow \mathcal{N}(0, \sigma^2\mathbb{E}[\tau])$, which makes $\sigma^2\mathbb{E}[\tau]$ the right figure of merit; Glynn and Whitt (1992) prove this and justify the criterion from a decision-theoretic point of view (see also Asmussen and Glynn (2007, Section III.10)). Measured times also contain overheads: if a run of n samples costs $c_0 + c_1 n$, the cost per sample $c_1 + c_0/n$ makes short runs look inefficient. The course measured the efficiency of runs of $n = 2, \dots, 2^{20}$ samples relative to $n = 2^{20}$ and found it rising from about 10^{-3} to 1 around $n = 2^{14}$, which is this effect: timings compare methods only once $n \gg c_0/c_1$ (Exercise 6.8).

▶ Interactive demo: time the bare-bones Monte Carlo on your own machine and plot its efficiency against n.

An estimator can be biased: the payoff may be computed on a time grid (Chapter 7), a coefficient may be estimated from the same sample (Theorem 8.14), or the quantity may be a nonlinear function of means (Section 6.5). The right yardstick is then the mean squared error, of which the variance is only one part.

Proposition 6.11 (Mean squared error and the cost–error trade-off)**.** *Let $\hat{\theta}_n$ be the average of n independent copies of a random variable Y with $\mathbb{E}[Y^2] < \infty$, $\mathbb{E}[Y] = \theta + b$ and $\mathrm{Var}(Y) = v > 0$.*

(i) $\mathrm{MSE}(\hat{\theta}_n) := \mathbb{E}[(\hat{\theta}_n - \theta)^2] = b^2 + v/n$.

(ii) If $b \neq 0$, the probability that the nominal interval $\hat{\theta}_n \pm z_\delta \hat{s}_n/\sqrt{n}$, with $\hat{s}_n$ the sample standard deviation of the copies, contains θ tends to 0 as $n \to \infty$.

(iii) Suppose that a parameter $h > 0$ (a time step, say) controls bias and cost: the bias is $b(h) = c_b h^\beta$, the variance is v for every h, and one copy costs $c_\tau h^{-\gamma}$, with constants $c_b, c_\tau, \beta, \gamma > 0$. With a budget c and $n = ch^\gamma/c_\tau$ copies (ignoring rounding), the mean squared error $c_b^2 h^{2\beta} + v c_\tau h^{-\gamma}/c$ is minimised at

$$h^* = \Big(\frac{\gamma\, v\, c_\tau}{2\beta\, c_b^2\, c}\Big)^{1/(2\beta+\gamma)},$$

where the squared bias is the fraction $\gamma/(2\beta+\gamma)$ of the mean squared error, and the minimum is proportional to $c^{-2\beta/(2\beta+\gamma)}$. Equivalently, a root mean squared error ε costs a budget proportional to $\varepsilon^{-2-\gamma/\beta}$, against ε^{-2} for an unbiased estimator with fixed cost per sample.

Proof. (i) $\mathbb{E}[(\hat{\theta}_n - \theta)^2] = \mathrm{Var}(\hat{\theta}_n) + (\mathbb{E}[\hat{\theta}_n] - \theta)^2 = v/n + b^2$ by Theorem 6.2(iii).

(ii) By Theorem 6.2, $\hat{\theta}_n \to \theta + b$ and $\hat{s}_n \to \sqrt{v}$ almost surely. So $|\hat{\theta}_n - \theta| \to |b| > 0$ while

$z_\delta \hat{s}_n/\sqrt{n} \to 0$, and almost surely the interval misses θ for all large n. The coverage tends to 0 by bounded convergence.

(iii) Write the mean squared error as $\phi(h) = ah^{2\beta} + Bh^{-\gamma}$ with $a = c_b^2$ and $B = vc_\tau/c$. Then $\phi'(h) = h^{-\gamma-1}(2\beta a h^{2\beta+\gamma} - \gamma B)$ is negative for $h < h^*$ and positive for $h > h^*$, where $(h^*)^{2\beta+\gamma} = \gamma B/(2\beta a)$, which is the stated formula; so h^* is the global minimiser. At h^*, $ah^{2\beta} = \frac{\gamma}{2\beta} Bh^{-\gamma}$, so the squared bias is $\frac{\gamma/(2\beta)}{1+\gamma/(2\beta)} = \frac{\gamma}{2\beta+\gamma}$ of the total, and $\phi(h^*) = \frac{2\beta+\gamma}{\gamma}\, a\,(h^*)^{2\beta}$ is proportional to $c^{-2\beta/(2\beta+\gamma)}$. Setting $\phi(h^*) = \varepsilon^2$ and solving for c gives $c \propto \varepsilon^{-(2\beta+\gamma)/\beta} = \varepsilon^{-2-\gamma/\beta}$. □

Two lessons. First, with a fixed bias more paths do not help: the error levels off at $|b|$, and the confidence interval, which knows nothing about b, becomes confidently wrong. Bias must be controlled and reported separately. Second, bias costs rate: with the Euler scheme ($\beta = \gamma = 1$) the cost of accuracy ε grows like ε^{-3} instead of ε^{-2}, the allocation of Theorem 7.14; multilevel Monte Carlo recovers ε^{-2} up to logarithms (Theorem 12.6). In practice v depends on h as well; as long as it stays between two positive constants, the exponents are unchanged.

6.5 Functions of expectations

Some quantities are not expectations but smooth functions of expectations: the implied volatility of a simulated price, a standard deviation, or a ratio $\mathbb{E}[Y]/\mathbb{E}[X]$. Ratios arise as conditional expectations $\mathbb{E}[Y \mid A] = \mathbb{E}[Y\mathbf{1}_A]/\mathbb{P}(A)$, as prices measured in units of a simulated numeraire, and in self-normalised importance sampling (Chapter 9). We plug sample means into the function. The result is consistent by the strong law and continuity, but biased, and its error obeys a central limit theorem whose variance comes from the gradient. For a vector x, $|x|$ is the Euclidean norm, and for a matrix the operator norm.

Proposition 6.12 (Delta method and ratio estimator)**.** *Let X be a random vector in $\mathbb{R}^k$ with $\mathbb{E}|X|^2 < \infty$, mean μ and covariance matrix Σ, and let $\hat{X}_n$ be the average of n independent copies.*

(i) If $g : \mathbb{R}^k \to \mathbb{R}$ is differentiable at μ, then $\sqrt{n}(g(\hat{X}_n) - g(\mu)) \Rightarrow \mathcal{N}(0, \nabla g(\mu)^\top \Sigma\, \nabla g(\mu))$.

(ii) If moreover $\mathbb{E}|X|^4 < \infty$, and g is twice continuously differentiable with a Hessian H that is Lipschitz, $|H(x) - H(x')| \le L|x - x'|$ for all $x, x' \in \mathbb{R}^k$, then

$$\mathbb{E}[g(\hat{X}_n)] = g(\mu) + \frac{\operatorname{tr}\left(H(\mu)\Sigma\right)}{2n} + R_n, \qquad |R_n| \le \frac{L}{6}\,\mathbb{E}|\hat{X}_n - \mu|^3 = O(n^{-3/2}).$$

(iii) (Ratio estimator.) Let (X, Y) be a pair of real random variables with $X \ge a$ for a constant $a > 0$, $\mathbb{E}[X^4] < \infty$ and $\mathbb{E}[Y^2] < \infty$; let $\theta = \mathbb{E}[Y]/\mathbb{E}[X]$ and $\hat{R}_n = \bar{Y}_n/\bar{X}_n$, the ratio of the sample means. Then $\sqrt{n}(\hat{R}_n - \theta) \Rightarrow \mathcal{N}(0, \operatorname{Var}(Y - \theta X)/\mathbb{E}[X]^2)$ and

$$\mathbb{E}[\hat{R}_n] = \theta + \frac{\theta \operatorname{Var}(X) - \operatorname{Cov}(X, Y)}{n\,\mathbb{E}[X]^2} + O(n^{-3/2}).$$

Proof. (i) Differentiability at μ means that $g(x) = g(\mu) + \nabla g(\mu)^\top (x - \mu) + |x - \mu|\,\epsilon(x)$ with

$\epsilon(x) \to 0 = \epsilon(\mu)$ as $x \to \mu$. Hence

$$\sqrt{n}(g(\hat{X}_n) - g(\mu)) = \nabla g(\mu)^\top \sqrt{n}(\hat{X}_n - \mu) + |\sqrt{n}(\hat{X}_n - \mu)|\, \epsilon(\hat{X}_n).$$

By the multivariate central limit theorem (Durrett, 2019, Section 3.10), $\sqrt{n}(\hat{X}_n - \mu) \Rightarrow \mathcal{N}(0, \Sigma)$, so the first term converges in distribution to $\mathcal{N}(0, \nabla g(\mu)^\top \Sigma \nabla g(\mu))$, and $|\sqrt{n}(\hat{X}_n - \mu)|$ is bounded in probability. By the strong law $\epsilon(\hat{X}_n) \to 0$ almost surely, so the second term tends to 0 in probability, and Slutsky's lemma concludes.

(ii) Taylor's formula with integral remainder along the segment from μ to x gives $g(x) = g(\mu) + \nabla g(\mu)^\top (x - \mu) + \frac{1}{2}(x - \mu)^\top H(\mu)(x - \mu) + R(x)$ with

$$R(x) = \int_0^1 (1 - s)\, (x - \mu)^\top (H(\mu + s(x - \mu)) - H(\mu))(x - \mu)\, \mathrm{d}s,$$
$$|R(x)| \le \int_0^1 (1 - s) L s\, \mathrm{d}s\, |x - \mu|^3 = \frac{L}{6} |x - \mu|^3.$$

Every term is integrable at $x = \hat{X}_n$ because $\mathbb{E}|\hat{X}_n|^4 < \infty$. Take expectations: the linear term vanishes, and $\mathbb{E}[(\hat{X}_n - \mu)^\top H(\mu)(\hat{X}_n - \mu)] = \operatorname{tr}(H(\mu)\Sigma)/n$ since $\hat{X}_n$ has covariance Σ/n. For the order of the remainder, write $\xi_i = X_i - \mu$. Expanding the fourth power of a coordinate of $\sum_i \xi_i$, terms in which some index appears exactly once have mean zero, so $\mathbb{E}[(\sum_i \xi_{i,l})^4] = n\mathbb{E}[\xi_{1,l}^4] + 3n(n-1)\mathbb{E}[\xi_{1,l}^2]^2 \le 3n^2 \mathbb{E}[\xi_{1,l}^4]$ for each coordinate l. With $|x|^4 \le k \sum_l x_l^4$ this gives $\mathbb{E}|\hat{X}_n - \mu|^4 = O(n^{-2})$, and by Jensen's inequality $\mathbb{E}|\hat{X}_n - \mu|^3 \le (\mathbb{E}|\hat{X}_n - \mu|^4)^{3/4} = O(n^{-3/2})$.

(iii) Let $m = \mathbb{E}[X]$ and $D = Y - \theta X$, so that $\mathbb{E}[D] = 0$ and $\hat{R}_n - \theta = \bar{D}_n / \bar{X}_n$. The central limit theorem gives $\sqrt{n}\, \bar{D}_n \Rightarrow \mathcal{N}(0, \operatorname{Var} D)$, and $\bar{X}_n \to m$ almost surely, so the first claim follows from Slutsky's lemma. For the bias, the identity $\frac{1}{\bar{X}_n} = \frac{1}{m} - \frac{\bar{X}_n - m}{m^2} + \frac{(\bar{X}_n - m)^2}{m^2 \bar{X}_n}$ gives

$$\hat{R}_n - \theta = \frac{\bar{D}_n}{m} - \frac{\bar{D}_n (\bar{X}_n - m)}{m^2} + \frac{\bar{D}_n (\bar{X}_n - m)^2}{m^2 \bar{X}_n}.$$

The first term has mean 0. The second has mean $-\operatorname{Cov}(D, X)/(nm^2)$, because the cross terms between different samples are uncorrelated, and $\operatorname{Cov}(D, X) = \operatorname{Cov}(X, Y) - \theta \operatorname{Var}(X)$. In the third, $\bar{X}_n \ge a$, so by the Cauchy–Schwarz inequality its absolute mean is at most $(\mathbb{E}[\bar{D}_n^2]\, \mathbb{E}[(\bar{X}_n - m)^4])^{1/2} / (m^2 a)$, which is $O((n^{-1} n^{-2})^{1/2}) = O(n^{-3/2})$ by the fourth-moment bound in (ii). □

The assumption $X \ge a > 0$ in (iii) keeps $1/\bar{X}_n$ bounded. The same proof works when $X > 0$ with $\mathbb{E}[X^8] < \infty$ and $\mathbb{E}[X^{-4}] < \infty$, which covers a lognormal X such as S_T: bound the third term with Hölder's inequality (exponents $2, 4, 4$) instead of Cauchy–Schwarz, use $\bar{X}_n^{-4} \le \frac{1}{n} \sum_i X_i^{-4}$ (convexity of $x \mapsto x^{-4}$), and $\mathbb{E}[(\bar{X}_n - m)^8] = O(n^{-4})$. The last bound follows from the counting argument of (ii): in the expansion of $\mathbb{E}[(\sum_i \xi_i)^8]$ only index tuples in which every index appears at least twice survive, they involve at most four distinct indices, there are $O(n^4)$ of them, and each term is at most $\mathbb{E}[\xi_1^8]$ by Hölder's inequality. If instead $\mathbb{P}(X = 0) > 0$, as for $X = \mathbf{1}_A$, then $\bar{X}_n = 0$ with positive probability and $\hat{R}_n$ is not even defined for every sample; for more general conditions see Asmussen and Glynn (2007, Section III.3). The standard error of $\hat{R}_n$ is estimated from the residuals $Y_i - \hat{R}_n X_i$ (`ratio_estimate` in the code). The bias is of order $1/n$ and the error of order $n^{-1/2}$, so the bias is negligible for one large sample. It is not negligible if many small-sample ratios are averaged: the average of m batch ratios, each on k samples, has a bias of order $1/k$ and a standard error of order $1/\sqrt{mk}$, so form the ratio of the overall means instead.

Example 6.13 (How accurate is the standard error?)**.** The sample standard deviation is a smooth function of two sample means: $\sqrt{(n-1)/n}\,\hat{s}_n = g(\frac{1}{n}\sum_i X_i, \frac{1}{n}\sum_i X_i^2)$ with $g(m_1, m_2) = \sqrt{m_2 - m_1^2}$. Assume $\mathbb{E}[X^4] < \infty$, so that the vector (X, X^2) has a finite covariance matrix, and write $\mu_4 = \mathbb{E}[(X-\mu)^4]$. Since $\hat{s}_n$ does not change when a constant is added to X, take $\mu = 0$; then $\nabla g(0, \sigma^2) = (0, 1/(2\sigma))$ and part (i) gives

$$\sqrt{n}\,(\hat{s}_n - \sigma) \Rightarrow \mathcal{N}\Big(0, \frac{\mu_4 - \sigma^4}{4\sigma^2}\Big), \qquad \text{relative error of } \hat{s}_n \approx \sqrt{\frac{\kappa - 1}{4n}}, \quad \kappa = \frac{\mu_4}{\sigma^4}.$$

For normal X, $\kappa = 3$ and the relative error is $1/\sqrt{2n}$. For the at-the-money call of the test case, $\kappa = 8.95$ and the relative error is $1.41/\sqrt{n}$: with 100 paths the standard error is itself uncertain by about 14%, with 10^4 paths by 1.4%. Heavy tails hurt here first: if $\mathbb{E}[X^4] = \infty$ but $\mathbb{E}[X^2] < \infty$, the central limit theorem for $\hat{X}_n$ still holds, but this approximation for $\hat{s}_n$ does not, and the standard error itself is then typically much less accurate than $n^{-1/2}$ suggests.

6.6 Monte Carlo against quadrature

Why simulate at all? The course compares the ways of computing an expectation by their cost τ to reach an error ε. A deterministic quadrature rule with error of order $n^{-\alpha}$ for n points per dimension needs n^d points in dimension d, so $\tau \sim \varepsilon^{-d/\alpha}$. Lattice and finite-difference methods add a time dimension, $\tau \sim \varepsilon^{-(d+1)/\alpha}$ (Chapter 20). Monte Carlo needs ε^{-2} samples, each costing about d operations, so $\tau \sim d\,\varepsilon^{-2}$. The next proposition makes the comparison precise for functions with two bounded derivatives, and shows that the exponent $d/2$ is not a defect of one particular rule. Let

$$\mathcal{F}_d = \{f \in C^2([0,1]^d) : \ |\partial^a f(x)| \le 1 \text{ for all } x \text{ and all multi-indices } a \text{ with } |a| \le 2\},$$

and $I(f) = \int_{[0,1]^d} f(x)\,\mathrm{d}x$.

Proposition 6.14 (Monte Carlo against quadrature)**.** *Let $d \ge 1$.*

(i) (Product midpoint rule.) For $k \ge 1$, let $h = 1/k$ and let $Q_k f$ be the average of f over the $N = k^d$ centres of the cubes of side h that tile $[0,1]^d$. For every $f \in C^2([0,1]^d)$,

$$|I(f) - Q_k f| \le \frac{h^2}{24} \sum_{j=1}^{d} \sup_{[0,1]^d} |\partial_{jj} f|, \qquad \text{so} \quad |I(f) - Q_k f| \le \frac{d}{24} N^{-2/d} \quad \text{on } \mathcal{F}_d.$$

(ii) (No deterministic rule does better.) There is $c_d > 0$ such that for every $N \ge 1$, all nodes $x_1, \dots, x_N \in [0,1]^d$ and all weights $w_1, \dots, w_N \in \mathbb{R}$, some $f \in \mathcal{F}_d$ satisfies $|I(f) - \sum_{i=1}^N w_i f(x_i)| \ge c_d N^{-2/d}$.

(iii) (Monte Carlo.) For every f with $I(f^2) < \infty$, the average of f at N independent uniform points of $[0,1]^d$ is unbiased with root mean squared error $\sigma_f/\sqrt{N}$, where $\sigma_f^2 = I(f^2) - I(f)^2$. On $\mathcal{F}_d$, $\sigma_f \le 1$.

Consequently, guaranteeing an error ε uniformly on $\mathcal{F}_d$ takes at least $(c_d/\varepsilon)^{d/2}$ function values for any rule of the form in (ii), and at least $(c_d/2\varepsilon)^{d/2}$ for any deterministic rule (see the remark after the proof), while Monte Carlo reaches a root mean squared error ε with ε^{-2} of them, in every dimension.

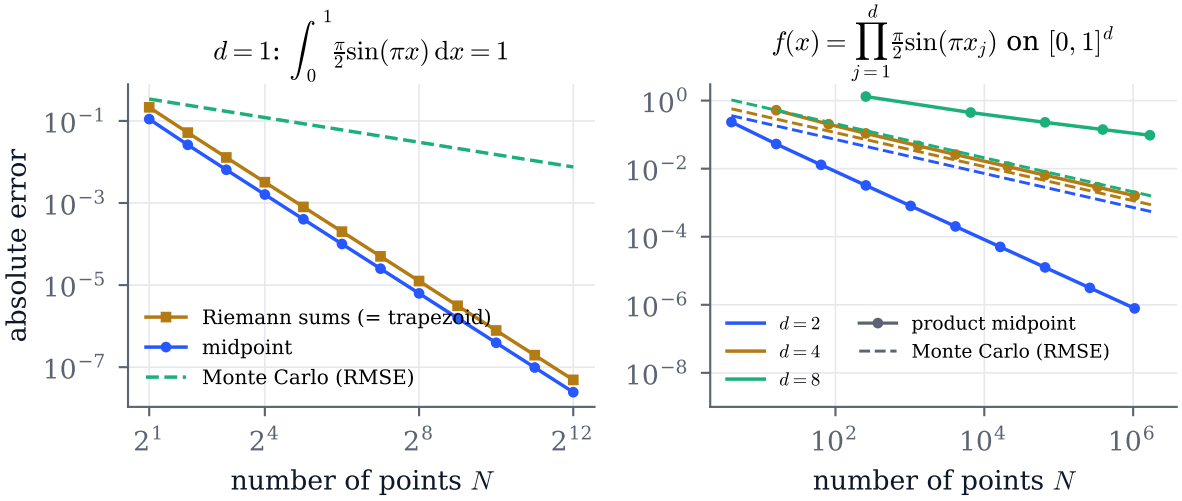

Figure 6.3: Left: the course's one-dimensional example $\int_0^1 \frac{\pi}{2}\sin(\pi x)\,dx = 1$. Because the integrand vanishes at both ends, the left and right Riemann sums ("forward" and "backward Euler") coincide with the trapezoid rule, and all rules are of order N^{-2} (Exercise 6.9); Monte Carlo, of order $N^{-1/2}$, is no match. Right: the product of d copies of the same integrand on $[0,1]^d$, whose integral is 1. Solid: the product midpoint rule with $N = k^d$ points, with slope $-2/d$. Dashed: the root mean squared error $\sigma_f/\sqrt{N}$ of Monte Carlo, with $\sigma_f^2 = (\pi^2/8)^d - 1$. At $d = 4$ the slopes agree and Monte Carlo has the smaller constant; at $d = 8$, Monte Carlo with 10^4 points is more accurate than the product rule with 1.7×10^6.

Proof. (i) One dimension first. If $u \in C^2[c - \frac{h}{2}, c + \frac{h}{2}]$, Taylor's formula at the centre gives $u(x) = u(c) + u'(c)(x-c) + \frac{1}{2}u''(\eta_x)(x-c)^2$; the linear term integrates to zero over the cell, so

$$\Big| \int_{c-h/2}^{c+h/2} u(x)\,dx - h\,u(c) \Big| \le \frac{1}{2} \sup|u''| \int_{-h/2}^{h/2} s^2\,ds = \frac{h^3}{24} \sup|u''|,$$

and summing over the k cells of $[0,1]$ bounds the error of the composite midpoint rule M by $\frac{h^2}{24}\sup|u''|$. In d dimensions, let I_j and M_j be integration and the midpoint rule in the variable x_j alone. Then $I = I_1 \cdots I_d$ and $Q_k = M_1 \cdots M_d$, and telescoping,

$$I - Q_k = \sum_{j=1}^{d} M_1 \cdots M_{j-1}\,(I_j - M_j)\,I_{j+1} \cdots I_d.$$

In the j-th term, $I_{j+1} \cdots I_d f$ is a C^2 function of $(x_1, \ldots, x_j)$ whose second derivative in x_j is bounded by $\sup|\partial_{jj} f|$ (differentiate under the integral sign); $I_j - M_j$ maps it to a function bounded by $\frac{h^2}{24}\sup|\partial_{jj} f|$; and $M_1, \ldots, M_{j-1}$ are averages, which do not increase the supremum. Sum over j. On $\mathcal{F}_d$ each supremum is at most 1, and $h^2 = N^{-2/d}$.

(ii) Let $\varphi(t) = e^{-1/(t(1-t))}$ for $t \in (0,1)$ and $\varphi = 0$ elsewhere: a C^∞ function on $\mathbb{R}$ that is positive exactly on $(0,1)$. Let $\psi(u) = \lambda \prod_{j=1}^{d} \varphi(u_j)$ with $\lambda > 0$ so small that $|\partial^a \psi| \le 1$ for $|a| \le 2$. Set $k = \lceil (2N)^{1/d} \rceil$ and $h = 1/k$, and tile $[0,1]^d$ by the $k^d \ge 2N$ open cubes of side h. They are disjoint, so each node lies in at most one of them, and at least $k^d - N \ge k^d/2$ cubes contain no

node. For each such cube, with lower corner z, put $\psi_z(x) = h^2\psi((x-z)/h)$, and let f be the sum of these ψ_z. Since $\partial^a\psi_z(x) = h^{2-|a|}(\partial^a\psi)((x-z)/h)$ and $h \le 1$, every ψ_z has derivatives of order at most 2 bounded by 1; they vanish, with all their derivatives, outside their own cube, so $f \in \mathcal{F}_d$. Now $f(x_i) = 0$ at every node, so the rule returns 0, while

$$I(f) \;\ge\; \frac{k^d}{2}\,h^{d+2}\,I(\psi) = \frac{h^2}{2}\,I(\psi) \;\ge\; \frac{I(\psi)}{8\,(2N)^{2/d}},$$

because $k \le (2N)^{1/d} + 1 \le 2(2N)^{1/d}$. This is the claim with $c_d = 2^{-3-2/d}I(\psi)$.

(iii) Unbiasedness and the variance are Theorem 6.2(i),(iii) for $X = f(U)$ with U uniform on $[0,1]^d$. On $\mathcal{F}_d$, $|f| \le 1$, so $\sigma_f^2 \le I(f^2) \le 1$. □

The proof of (ii) uses only the values of f at the nodes, so it extends, with the constant $c_d/2$, to any rule that looks at N values of f, linear or not: such a rule returns the same number A for f and for 0, so one of the errors $|I(f) - A|$ and $|A|$ is at least $I(f)/2$. If the rule chooses its nodes adaptively, run it on $f \equiv 0$ to see where it looks and build f there. The comparison of exponents is now clear: for functions with two bounded derivatives, deterministic rules converge at best like $N^{-2/d}$ and Monte Carlo like $N^{-1/2}$, so Monte Carlo wins asymptotically when $d > 4$, and the two rates are equal at $d = 4$ (Figure 6.3). For functions with r bounded derivatives the same two proofs, with rules of order r and bumps scaled by h^r, give the rate $N^{-r/d}$, and the threshold moves to $d = 2r$. But option payoffs have kinks (calls) and jumps (digitals), so r is small, while Monte Carlo needs only a finite variance.

Three caveats keep the comparison honest. The guarantees are of different kinds: the quadrature bound is deterministic, the Monte Carlo one holds in mean square. The Monte Carlo constant σ_f can grow with the dimension. And between the two lie methods that exploit more structure: quasi-Monte Carlo, whose error bound is of order $(\log N)^d/N$ for integrands of bounded variation in the sense of Hardy and Krause (Theorem 11.3), and sparse grids (Bungartz and Griebel, 2004) for functions with bounded mixed derivatives; how well they do in high dimension depends on the effective dimension of the integrand (Theorem 11.13).

▶ Interactive demo: product rules against Monte Carlo for a smooth integrand in the dimension of your choice.

Further reading

Glasserman (2003), Section 1.1 and Appendix A, covers the principles of this chapter in the setting of financial engineering. Asmussen and Glynn (2007), Chapter III, is a thorough treatment of output analysis: confidence intervals, smooth functions of expectations, sectioning (batching) of independent runs, bias–variance trade-offs and the simulation budget; their Chapter IV treats batch means for dependent output. Glynn and Whitt (1992) is the reference on efficiency. Durrett's book (Durrett, 2019) has the limit theorems we used, including stable laws and the Berry–Esseen theorem.

Chapter 7

Simulating paths

Chapter 6 priced claims that depend on the stock only at maturity: one normal draw per sample gives S_T exactly. An Asian option needs the average of the stock over a set of fixing dates, a barrier option needs to know whether the path crossed a level, and a model without a closed-form solution, such as a local or stochastic volatility model, cannot be sampled at T in one step. In all three cases we must simulate a *path*: the vector $(S_{t_1}, \dots, S_{t_m})$ on a grid of dates.

Two situations arise. When the joint law of the vector is known, as for geometric Brownian motion with time-dependent parameters or for any Gaussian process, it can be sampled exactly, with no error other than the statistical one, on any grid. Even then there is a choice: the same vector can be built from the same normals in different orders, and the order decides which normals carry most of the variance. This is irrelevant for plain Monte Carlo and decisive for stratification and quasi-Monte Carlo. When the law is not known, the stochastic differential equation must be discretised, and the scheme introduces a bias. We measure the bias (strong and weak orders), remove its leading term (Richardson extrapolation), and decide how to split a computing budget between finer time steps and more paths (Duffie and Glynn).

Throughout, $0 = t_0 < t_1 < \cdots < t_m = T$ is a time grid with steps $\Delta_j = t_j - t_{j-1}$, W is a Brownian motion under the risk-neutral measure $\mathbb{Q}$, and $Z_1, Z_2, \dots$ are independent $\mathcal{N}(0,1)$ variables. Path arrays in the code have one row per path.

7.1 Exact simulation on a time grid

7.1.1 Geometric Brownian motion with term structures

In the standard model of Theorem 1.1 the parameters are constants. Market data are not: the rate and the yield are read off a curve, and implied volatilities depend on the maturity. The simplest model consistent with such data lets r, y and σ depend on time. Let $r, y, \sigma : [0,T] \to \mathbb{R}$ be bounded Borel functions, $\sigma \geq 0$, and

$$\frac{\mathrm{d}S_t}{S_t} = \big(r(t) - y(t)\big)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}W_t, \qquad S_0 > 0. \tag{7.1}$$

For $0 \le s < t$ write

$$\bar r(s,t) = \frac{1}{t-s}\int_s^t r(u)\,\mathrm{d}u, \qquad \bar\sigma^2(s,t) = \frac{1}{t-s}\int_s^t \sigma(u)^2\,\mathrm{d}u,$$

and $\bar y(s,t)$ likewise. The *term rate* and the *term volatility* are $r_{0,t} = \bar r(0,t)$ and $\sigma_{0,t} = \bar\sigma(0,t)$; the *interval forward* quantities on the grid are $\bar r_j = \bar r(t_{j-1},t_j)$, $\bar y_j$ and $\bar\sigma_j^2 = \bar\sigma^2(t_{j-1},t_j)$, and we set $\bar\nu_j = \bar r_j - \bar y_j - \frac12\bar\sigma_j^2$. Forward variances are recovered from term volatilities by

$$\bar\sigma_j^2\,\Delta_j = t_j\,\sigma_{0,t_j}^2 - t_{j-1}\,\sigma_{0,t_{j-1}}^2, \tag{7.2}$$

so a term structure of volatility is consistent with (7.1) only if the total variance $t\,\sigma_{0,t}^2$ is nondecreasing in t (Exercise 7.1).

The integrals of a deterministic integrand against W are Gaussian. We record the fact with a proof, since it is the whole content of exact simulation.

Lemma 7.1 (Wiener integrals)**.** *Let $g:[0,T]\to\mathbb{R}$ be a deterministic function with $\int_0^T g^2 < \infty$, and $I_j = \int_{t_{j-1}}^{t_j} g(u)\,\mathrm{d}W_u$ for $j = 1,\dots,m$. Then $I_1,\dots,I_m$ are independent and $I_j \sim \mathcal{N}(0, \int_{t_{j-1}}^{t_j} g^2)$.*

Proof. If g is a step function, constant on the intervals of a partition that refines the grid, each I_j is a finite linear combination of increments of W over disjoint subintervals of $(t_{j-1},t_j]$. So $(I_1,\dots,I_m)$ is a linear image of independent Gaussian increments; distinct I_j use disjoint sets of increments, hence are independent, and $\mathrm{Var}(I_j)$ is the sum of g^2 times the lengths of the subintervals, that is $\int_{t_{j-1}}^{t_j} g^2$.

For general g, pick step functions $g_n \to g$ in $L^2[0,T]$ (they are dense) and let $I_j^{(n)}$ be the corresponding integrals. By Itô's isometry, $\mathbb{E}[(I_j^{(n)} - I_j)^2] = \int_{t_{j-1}}^{t_j}(g_n - g)^2 \to 0$. For $u \in \mathbb{R}^m$, using $|e^{ia} - e^{ib}| \le |a-b|$ and the Cauchy–Schwarz inequality,

$$\Big|\mathbb{E}\big[e^{i\sum_j u_j I_j^{(n)}}\big] - \mathbb{E}\big[e^{i\sum_j u_j I_j}\big]\Big| \le \sum_j |u_j|\,\mathbb{E}\big[(I_j^{(n)} - I_j)^2\big]^{1/2} \longrightarrow 0.$$

The left characteristic function equals $\prod_j \exp\big(-\tfrac12 u_j^2 \int_{t_{j-1}}^{t_j} g_n^2\big)$ by the step-function case, and converges to $\prod_j \exp\big(-\tfrac12 u_j^2 \int_{t_{j-1}}^{t_j} g^2\big)$. A characteristic function determines the law, also in $\mathbb{R}^m$ (Durrett, 2019, Section 3.10), so the components are independent Gaussians. □

Proposition 7.2 (Exact simulation of GBM)**.** *(i) The unique solution of (7.1) is $S_t = S_0 \exp\big(\int_0^t \nu(u)\,\mathrm{d}u + \int_0^t \sigma(u)\,\mathrm{d}W_u\big)$, with $\nu = r - y - \frac12\sigma^2$.*

(ii) For every grid, the recursion

$$\hat S_0 = S_0, \qquad \hat S_j = \hat S_{j-1}\exp\big(\bar\nu_j\Delta_j + \bar\sigma_j\sqrt{\Delta_j}\,Z_j\big), \quad j = 1,\dots,m, \tag{7.3}$$

produces a vector $(\hat S_0,\dots,\hat S_m)$ with the same joint law as $(S_{t_0},\dots,S_{t_m})$, whatever the step sizes.

(iii) In particular $\mathbb{E}^{\mathbb{Q}}[S_t] = S_0 e^{(r_{0,t}-y_{0,t})t}$, and the European call with maturity t has the Black–Scholes price with the term parameters $r_{0,t}$, $y_{0,t}$, $\sigma_{0,t}$ (if $\sigma_{0,t} = 0$, the formula is

read as its limit $\sigma \downarrow 0$, the discounted intrinsic value of the forward).

Proof. (i) Let $X_t = \int_0^t \nu + \int_0^t \sigma \, \mathrm{d}W$, an Itô process with $\mathrm{d}X = \nu \, \mathrm{d}t + \sigma \, \mathrm{d}W$ and $\mathrm{d}[X,X] = \sigma^2 \mathrm{d}t$. Itô's formula applied to $S_0 e^{X_t}$ gives $\mathrm{d}S = S\,(\nu + \frac{1}{2}\sigma^2)\,\mathrm{d}t + S\sigma \, \mathrm{d}W$, which is (7.1). Its coefficients are Lipschitz in S uniformly in t, because r, y, σ are bounded, and for such coefficients strong (pathwise) uniqueness holds (Karatzas and Shreve, 1991, Theorem 5.2.5).

(ii) By (i), $\log S_{t_j} - \log S_{t_{j-1}} = \bar{\nu}_j \Delta_j + I_j$ with $I_j = \int_{t_{j-1}}^{t_j} \sigma \, \mathrm{d}W$. By Theorem 7.1 the I_j are independent with $I_j \sim \mathcal{N}(0, \bar{\sigma}_j^2 \Delta_j)$, so the vector of log-increments has the law of $\left(\bar{\nu}_j \Delta_j + \bar{\sigma}_j \sqrt{\Delta_j} Z_j\right)_{j \le m}$. Both $(S_{t_j})_j$ and $(\hat{S}_j)_j$ are the same function (cumulative sums, then the exponential) of these vectors.

(iii) $\log S_t$ is normal with mean $\log S_0 + \int_0^t \nu$ and variance $\int_0^t \sigma^2 = \sigma_{0,t}^2 t$, so $\mathbb{E}^{\mathbb{Q}}[S_t] = S_0 \exp\left(\int_0^t (r - y)\right)$. A lognormal S_t with this mean and this log-variance is exactly the terminal law in the Black–Scholes model with constant parameters $r_{0,t}$, $y_{0,t}$, $\sigma_{0,t}$, and the discount factor $e^{-\int_0^t r} = e^{-r_{0,t} t}$ is the same. □

Three remarks. The recursion has no discretisation error: a grid is chosen only because the payoff looks at those dates. The parameters enter only through their averages over the steps; this is the precise meaning of the "effective" parameters of a time-stepping scheme. And nothing is said about the path between grid dates, which Section 7.2 fills in.

Example 7.3 (A term structure from the course)**.** Take $y = 0$, an instantaneous forward rate $r(t) = 0.02\,t$, a volatility $\sigma(t) = 20\%$ on $[0, \frac{1}{2})$ and 10% afterwards, and monthly dates $t_j = j/12$. The interval forward variances are $\bar{\sigma}_j^2 = 0.04$ for $j \le 6$ and 0.01 for $j > 6$; the interval forward rates are $\bar{r}_j = 0.01\,(t_{j-1} + t_j) = 0.01\,(2j-1)/12$. At one year the term rate is $r_{0,1} = 1\%$ and the term volatility $\sigma_{0,1} = \sqrt{0.025} = 15.81\%$, so the one-year at-the-money call is worth 6.7799 with $S_0 = K = 100$ (Theorem 7.2(iii)), and (7.3) reproduces this within its standard error (`code/tests/test_paths.py`).

Remark 7.4 (Frozen parameters). A tempting shortcut is to step with the parameters at the left end of each step, $r(t_{j-1})$ and $\sigma(t_{j-1})$: this is the Euler scheme of Section 7.3 applied to $\log S$. In Theorem 7.3 it replaces $\int_0^1 r = 0.01$ by the Riemann sum 0.00917, so the simulated forward is 100.92 instead of 101.01. Worse, if the convention $\sigma(\frac{1}{2}) = 20\%$ is used at the jump, the variance $\int_0^1 \sigma^2$ becomes 0.0275 instead of 0.025, and the call, discounted with the true factor $e^{-0.01}$, comes out at 7.0384 instead of 6.7799, a 4% error that no number of paths will remove (with the convention $\sigma(\frac{1}{2}) = 10\%$ the drift error alone gives 6.7337; discounting with the Riemann sum as well gives 7.0443 and 6.7393). With averages over the steps, the convention at the jump does not matter.

Remark 7.5 (Time change). Let $\Theta(t) = \int_0^t \sigma^2$. The process $M_t = \int_0^t \sigma \, \mathrm{d}W$ is centered Gaussian with $\mathrm{Cov}(M_s, M_t) = \Theta(s \wedge t)$ (Theorem 7.1), which is the covariance of $B_{\Theta(t)}$ for a Brownian motion B. So on any grid, M has the law of a standard Brownian motion run on the *variance clock* Θ; the code builds GBM paths this way. If moreover the drift is proportional to the variance rate, $r(t) - y(t) = \lambda \sigma(t)^2$, then $\log(S_t/S_0)$ has the law of $(\lambda - \frac{1}{2})\Theta(t) + B_{\Theta(t)}$, a Brownian motion with constant drift in the new clock, and constant-coefficient formulas, such as those for barriers in Chapter 16, apply in that clock. In Theorem 7.3, $\Theta(1) = 0.025$ and $\int_0^1 r = 0.01$ give $\lambda = 0.4$: a process with the same total drift and variance but drift proportional to variance, of the kind used

as a control variate for discretely monitored barriers, has monthly forward rates $0.4 \times 0.04 = 1.6\%$ for the first six months and $0.4 \times 0.01 = 0.4\%$ afterwards.

7.1.2 Gaussian processes

Theorem 7.2 is a case of a general recipe. If X is a Gaussian process, the vector $(X_{t_1}, \dots, X_{t_m})$ is $\mathcal{N}(\mu, C)$ with $\mu_i = \mathbb{E}X_{t_i}$ and $C_{ij} = \operatorname{Cov}(X_{t_i}, X_{t_j})$, and $\mu + AZ$ has this law for any matrix A with $AA^\top = C$ (Theorem 5.14). A dense factorisation costs $O(m^3)$ once and $O(m^2)$ per path. When X is also Markov the cost drops to $O(m)$ per path, because each X_{t_j} given the past is Gaussian with a mean that depends only on $X_{t_{j-1}}$.

Example 7.6 (Ornstein–Uhlenbeck)**.** Let $\mathrm{d}X_t = \kappa(\theta - X_t)\,\mathrm{d}t + \xi\,\mathrm{d}W_t$ with $\kappa > 0$, the Vasicek short-rate model. Applying Itô's formula to $e^{\kappa t}X_t$ gives $\mathrm{d}(e^{\kappa t}X_t) = e^{\kappa t}(\kappa\theta\,\mathrm{d}t + \xi\,\mathrm{d}W_t)$, hence

$$X_{t+h} = \theta + (X_t - \theta)e^{-\kappa h} + \xi \int_t^{t+h} e^{-\kappa(t+h-u)}\,\mathrm{d}W_u.$$

By Theorem 7.1 (and since the integral is independent of $\mathcal{F}_t$), the exact step is $X_{t+h} = \theta + (X_t - \theta)e^{-\kappa h} + \xi\sqrt{(1 - e^{-2\kappa h})/(2\kappa)}\,Z$, on any grid. The Euler scheme for the same equation is biased, and unstable for $\kappa h > 2$ (Exercise 7.3).

▶ Interactive demo: move the volatility jump and the number of steps, and compare exact stepping with frozen parameters against the closed-form price.

In code

code/src/mfe230d/paths.py: `gbm_paths(S0, t, r, y, sigma, n_paths, seed)` implements (7.3); `r`, `y`, `sigma` may be numbers, `PiecewiseConstant` curves or functions of time, and `effective_parameters` returns the averages $\bar r_j$, $\bar y_j$, $\bar\sigma_j^2$ and the clock $\Theta(t_j)$. `ou_paths` is the exact Ornstein–Uhlenbeck step. Every sampler accepts the normals `Z` instead of a seed, so that antithetic, stratified or quasi-random normals can be passed in (Chapters 8, 10 and 11).

7.2 Three ways to build a Brownian path

By Theorem 7.5, the Gaussian part of a GBM path is a Brownian motion sampled on a grid (of the variance clock). Its covariance is $C_{ij} = \min(t_i, t_j)$, and each square root A of C gives a construction $W = AZ$. Three are in use.

Increments. $W_{t_j} = W_{t_{j-1}} + \sqrt{\Delta_j}\,Z_j$. The matrix A is lower triangular with $A_{ij} = \sqrt{\Delta_j}$ for $j \le i$: the Cholesky factor of C. The normal Z_j moves the path from t_j on.

Brownian bridge. Sample the endpoint first, $W_T = \sqrt{T}\,Z_1$, then the midpoint given the two ends, then the quarter points given their neighbours, and so on. This rests on the law of a Brownian path between two known values.

Proposition 7.7 (Brownian bridge)**.** *Let $0 \le s < u < t$ and let $\mathcal{G}_{s,t} = \sigma(W_r : r \in [0,s] \cup [t,\infty))$.*

(i) The variable

$$B = W_u - \frac{(t-u)\,W_s + (u-s)\,W_t}{t-s}$$

is independent of $\mathcal{G}_{s,t}$ *and* $B \sim \mathcal{N}\big(0, \frac{(u-s)(t-u)}{t-s}\big)$.

(ii) Consequently, given $\mathcal{G}_{s,t}$, W_u *is normal with mean* $\frac{(t-u)W_s+(u-s)W_t}{t-s}$ *(linear interpolation) and variance* $\frac{(u-s)(t-u)}{t-s}$*: it depends on the rest of the path only through the two neighbouring values.*

(iii) Let $0 < t_1 < \cdots < t_m$ *and fill the points in any order. When* t_i *is filled, let* t_l *be the largest already-filled time below* t_i *(or* 0*, with* $W_0 = 0$*) and* t_r *the smallest already-filled time above it. Set* $W_{t_i} = W_{t_l} + \sqrt{t_i - t_l}\,Z$ *if there is no* t_r*, and otherwise*

$$W_{t_i} = \frac{(t_r - t_i)\,W_{t_l} + (t_i - t_l)\,W_{t_r}}{t_r - t_l} + \sqrt{\frac{(t_i - t_l)(t_r - t_i)}{t_r - t_l}}\,Z, \tag{7.4}$$

with a fresh normal Z *at each step. The resulting vector has the law of* $(W_{t_1}, \dots, W_{t_m})$.

Proof. (i) Write $a = u - s$, $b = t - u$, $D_1 = W_u - W_s$ and $D_2 = W_t - W_u$. Then $B = W_s + D_1 - W_s - \frac{a}{a+b}(D_1 + D_2) = \frac{bD_1 - aD_2}{a+b}$. The increments D_1, D_2 are independent with variances a, b, so B is centered normal with variance $\frac{b^2 a + a^2 b}{(a+b)^2} = \frac{ab}{a+b}$, and $\operatorname{Cov}(B, D_1 + D_2) = \frac{ba - ab}{a+b} = 0$. As $(B, D_1 + D_2)$ is Gaussian, B is independent of $D_1 + D_2 = W_t - W_s$.

Let $\mathcal{H}$ be generated by $(W_r)_{r \le s}$ and $(W_r - W_t)_{r \ge t}$. The pair (D_1, D_2) is independent of $\mathcal{H}$ (independent increments), and $\mathcal{G}_{s,t} = \sigma(\mathcal{H}, W_t - W_s)$ because $W_r = W_s + (W_t - W_s) + (W_r - W_t)$ for $r \ge t$. For bounded measurable f, g, k,

$$\mathbb{E}\big[f(B)\,g(H)\,k(W_t - W_s)\big] = \mathbb{E}[g(H)]\,\mathbb{E}[f(B)]\,\mathbb{E}[k(W_t - W_s)] = \mathbb{E}[f(B)]\,\mathbb{E}\big[g(H)\,k(W_t - W_s)\big],$$

where H is any finite vector of $\mathcal{H}$-measurable variables: the first equality uses that $(B, W_t - W_s)$ is a function of (D_1, D_2), independent of H, and then the independence of B and $W_t - W_s$; the second uses again that $W_t - W_s$ is independent of H. Such products determine independence from $\mathcal{G}_{s,t}$, by Dynkin's π–λ theorem (Durrett, 2019, Section 2.1).

(ii) W_u is the sum of the $\mathcal{G}_{s,t}$-measurable interpolation and the independent variable B; condition with the independence lemma (Shreve, 2004, Lemma 2.3.4).

(iii) By induction on the number k of filled points: the joint law of the first k values is that of the corresponding W's. The k-th point t_i is generated from its filled neighbours by (7.4). All previously filled times lie outside (t_l, t_r), so the σ-algebra $\mathcal{F}'$ they generate is contained in $\mathcal{G}_{t_l,t_r}$ and contains W_{t_l}, W_{t_r}. By (ii) and the tower property, the conditional law of W_{t_i} given $\mathcal{F}'$ is the normal law in (7.4); without a right neighbour it is $\mathcal{N}(W_{t_l}, t_i - t_l)$ by independent increments. A joint law is determined by the law of the first $k-1$ coordinates and the conditional law of the k-th given them. □

The conditional variance $\frac{(u-s)(t-u)}{t-s}$ vanishes at both ends and is largest, $\frac{t-s}{4}$, at the midpoint. The *pinned* bridge $W_u - \frac{u}{T}W_T$, $u \in [0, T]$, is independent of W_T with covariance $\min(u, v) - uv/T$. On a grid of $m = 2^k$ steps the dyadic order (`bridge_order` in the code) fills T, $T/2$, $T/4$ and $3T/4$, and so on; after 2^ℓ normals the path is known exactly at the dyadic points of level ℓ

(Figure 7.1, left). The construction costs $O(m)$ per path, like the increments, and works for any grid, uniform or not.

Principal components. Diagonalise $C = V\Lambda V^\top$ with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_m > 0$ and orthonormal eigenvectors v_i, and take $A = V\Lambda^{1/2}$: $W = \sum_i \sqrt{\lambda_i}\, v_i Z_i$. This costs $O(m^2)$ per path.

All three constructions produce exactly the same law, so for plain Monte Carlo they give estimators with the same variance; only their cost differs. They differ in how the variance is spread over the normals. Call

$$\pi_k(A) = \frac{\sum_{j\leq k} \|a_j\|^2}{\operatorname{tr} C}, \qquad a_j \text{ the } j\text{-th column of } A,$$

the share of the total variance $\sum_i \operatorname{Var}(W_{t_i}) = \operatorname{tr} C$ carried by the first k normals.

Proposition 7.8 (Principal components carry the most variance)**.** *For every $m \times m$ matrix A with $AA^\top = C$ and every $k \leq m$, $\pi_k(A) \leq \sum_{i\leq k} \lambda_i / \operatorname{tr} C$, with equality for $A = V\Lambda^{1/2}$.*

Proof. C is invertible (the increments over the grid are independent and nondegenerate), and $U = C^{-1/2}A$ satisfies $UU^\top = C^{-1/2}CC^{-1/2} = I$, so its columns u_j are orthonormal and $a_j = C^{1/2}u_j$. Then $\sum_{j\leq k} \|a_j\|^2 = \sum_{j\leq k} u_j^\top C u_j = \sum_i \lambda_i w_i$ with $w_i = \sum_{j\leq k} (v_i^\top u_j)^2$. Each w_i is the squared length of the projection of v_i on the span of $u_1, \ldots, u_k$, so $0 \leq w_i \leq 1$, and $\sum_i w_i = k$. A weighted sum $\sum_i \lambda_i w_i$ with these constraints is largest when the weight 1 sits on the k largest eigenvalues, which gives the bound. For $A = V\Lambda^{1/2}$, $\|a_j\|^2 = \lambda_j$. □

For a uniform grid with $m = 64$ steps, the shares are:

first k normals	$k = 1$	$k = 2$	$k = 4$	$k = 16$
increments	0.031	0.061	0.120	0.435
Brownian bridge	0.672	0.836	0.918	0.981
principal components	0.811	0.901	0.950	0.988

The first increment carries $2/(m+1)$ of the variance; the first bridge normal, which sets W_T, carries $\sum_i t_i^2/(T\sum_i t_i) = (2m+1)/(3m) \to \frac{2}{3}$; the first principal component tends to $8/\pi^2 \approx 0.81$ (Exercise 7.8). This ordering matters whenever the normals are not treated symmetrically: stratifying Z_1 stratifies W_T under the bridge construction (Theorem 10.7), and quasi-Monte Carlo points are most uniform in their first coordinates, which is why the bridge and PCA lower the effective dimension of path-dependent integrands (Theorem 11.13). The bridge law (ii) also answers questions about the path between two grid dates; the probability that it crosses a barrier is Theorem 16.9.

▶ Interactive demo: watch the bridge construction refine a path level by level, and compare the variance carried by the first normals under the three constructions.

7.3 Discretising a stochastic differential equation

Now let X solve

$$\mathrm{d}X_t = a(t, X_t)\,\mathrm{d}t + b(t, X_t)\,\mathrm{d}W_t, \qquad X_0 = x_0 \in \mathbb{R}, \tag{7.5}$$

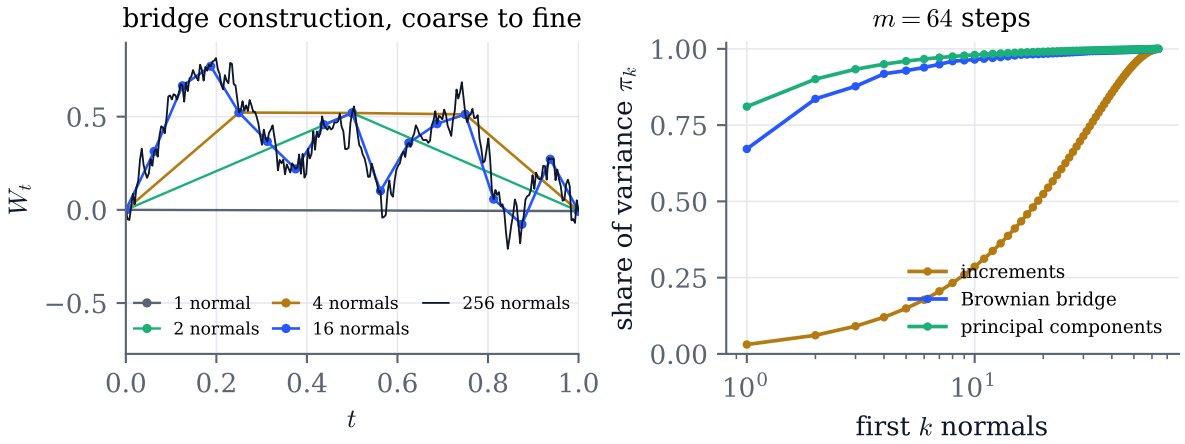

Figure 7.1: Left: one Brownian path built by the bridge construction on 2^8 steps, shown after 1, 2, 4, 16 and 256 normals (linear interpolation between the points already built). Right: share π_k of the variance carried by the first k normals, $m = 64$ (Theorem 7.8).

for which no exact sampler is available. Take the uniform grid $t_k = kh$, $h = T/n$, and write $\Delta W_k = W_{t_{k+1}} - W_{t_k}$. Over one step,

$$X_{t_{k+1}} = X_{t_k} + \int_{t_k}^{t_{k+1}} a(s, X_s)\,\mathrm{d}s + \int_{t_k}^{t_{k+1}} b(s, X_s)\,\mathrm{d}W_s.$$

Freezing both integrands at the left end gives the *Euler* (Euler–Maruyama) scheme

$$\bar{X}_0 = x_0, \qquad \bar{X}_{k+1} = \bar{X}_k + a(t_k, \bar{X}_k)\,h + b(t_k, \bar{X}_k)\,\Delta W_k. \tag{7.6}$$

The largest term it neglects comes from the variation of $b(s, X_s)$ inside the step. By Itô's formula, $b(s, X_s) = b(t_k, X_{t_k}) + \int_{t_k}^{s} (b\,\partial_x b)(r, X_r)\,\mathrm{d}W_r +$ a $\mathrm{d}r$-integral, and freezing $b\,\partial_x b$ at t_k leaves $\int_{t_k}^{t_{k+1}} (W_s - W_{t_k})\,\mathrm{d}W_s = \tfrac{1}{2}\big((\Delta W_k)^2 - h\big)$ (Itô's formula for $(W_s - W_{t_k})^2$). This term has standard deviation $h/\sqrt{2}$, while the other neglected terms are of order $h^{3/2}$. Keeping it gives the *Milstein* scheme

$$\tilde{X}_{k+1} = \tilde{X}_k + a(t_k, \tilde{X}_k)\,h + b(t_k, \tilde{X}_k)\,\Delta W_k + \tfrac{1}{2}\,(b\,\partial_x b)(t_k, \tilde{X}_k)\big((\Delta W_k)^2 - h\big). \tag{7.7}$$

When b does not depend on x (additive noise) the two schemes coincide.

Definition 7.9 (Strong and weak order)**.** A scheme Y^h with step h has *strong order* $\gamma > 0$ if $\mathbb{E}|X_T - Y_T^h| \le C\,h^\gamma$ for all $h \le h_0$, and *weak order* $\beta > 0$ for a class of functions f if $|\mathbb{E}f(Y_T^h) - \mathbb{E}f(X_T)| \le C_f\,h^\beta$ for all $h \le h_0$.

The strong error compares the two paths driven by the same Brownian motion; it is what matters when the path itself is used, as in hedging simulations or the couplings of multilevel Monte Carlo. The weak error compares laws, and is the bias of a Monte Carlo price. Since $|\mathbb{E}f(Y_T^h) - \mathbb{E}f(X_T)| \le L\,\mathbb{E}|Y_T^h - X_T|$ for an L-Lipschitz f, strong order γ implies weak order γ for such f; the weak order is often higher.

Theorem 7.10 (Euler–Maruyama). *Suppose there is K such that, for all $s,t \in [0,T]$ and $x,z \in \mathbb{R}$,*

(A1) $|a(t,x)-a(t,z)| + |b(t,x)-b(t,z)| \le K|x-z|$;

(A2) $|a(t,x)| + |b(t,x)| \le K(1+|x|)$;

(A3) $|a(s,x)-a(t,x)| + |b(s,x)-b(t,x)| \le K(1+|x|)\,|s-t|^{1/2}$.

(i) The Euler scheme has strong order $\frac{1}{2}$: $\mathbb{E}|X_T - \bar{X}^h_T| \le C\,h^{1/2}$ (Kloeden and Platen, 1992, Theorem 10.2.2).

(ii) Suppose moreover that a and b are infinitely differentiable on $[0,T]\times\mathbb{R}$ with bounded partial derivatives of every order ≥ 1, and that f is infinitely differentiable with f and all its derivatives of polynomial growth. Then

$$\mathbb{E}f(\bar{X}^h_T) - \mathbb{E}f(X_T) = C_f\,h + O(h^2) \tag{7.8}$$

for a constant C_f (Talay and Tubaro, 1990); in particular the weak order is 1. Weak order 1 holds already when a, b, f have four derivatives of polynomial growth (Kloeden and Platen, 1992, Theorem 14.5.1 with $\beta = 1$).

Talay and Tubaro (1990) state (ii) for time-independent coefficients in several dimensions; the pair (t, X_t) solves such an equation, and its Euler scheme is (7.6). The proof of (i) compares the scheme with the solution in mean square and closes with Gronwall's lemma. The proof of (ii) teaches more: it explains why the weak order of Euler is twice its strong order.

Proof of (ii), sketch. Let $u(t,x) = \mathbb{E}[f(X_T) \mid X_t = x]$. Under the hypotheses, u is smooth with derivatives of polynomial growth (Talay and Tubaro, 1990), and by the Markov property $\mathbb{E}u(t+h, X^{t,x}_{t+h}) = u(t,x)$ for the solution $X^{t,x}$ started at x at time t. Telescoping,

$$\mathbb{E}f(\bar{X}_n) - \mathbb{E}f(X_T) = \mathbb{E}u(t_n,\bar{X}_n) - u(0,x_0) = \sum_{k=0}^{n-1} \mathbb{E}\,\varepsilon(t_k,\bar{X}_k),$$
$$\varepsilon(t,x) = \mathbb{E}\big[u(t+h, x+\hat{D}) - u(t+h, x+D)\big],$$

where $\hat{D} = a(t,x)h + b(t,x)\Delta W$ is one Euler step and $D = X^{t,x}_{t+h} - x$ one exact step. Expand $u(t+h,\cdot)$ around x to third order: ε is a combination of the differences $\mathbb{E}\hat{D}^j - \mathbb{E}D^j$, $j=1,2,3$, plus a remainder of order $(1+|x|^q)h^2$, because $\partial_x^4 u$ grows polynomially and both steps have fourth moments of order h^2. For the Euler step, $\mathbb{E}\hat{D} = ah$, $\mathbb{E}\hat{D}^2 = b^2h + a^2h^2$, $\mathbb{E}\hat{D}^3 = 3ab^2h^2 + a^3h^3$, $\mathbb{E}\hat{D}^4 = O(h^2)$. For the exact step, Itô's formula applied to $(X_s - x)^j$ gives $\mathbb{E}D = ah + O(h^2)$, $\mathbb{E}D^2 = b^2h + O(h^2)$, $\mathbb{E}D^3 = O(h^2)$, $\mathbb{E}D^4 = O(h^2)$. So $|\varepsilon(t,x)| \le C(1+|x|^q)h^2$. The scheme has moments of all orders bounded uniformly in h and k (a discrete Gronwall argument using (A2)), and summing $n = T/h$ local errors of order h^2 gives an error of order h. □

The local errors of Euler have size h in mean square but h^2 in mean. Over $n = T/h$ steps, errors with zero mean add up in quadrature, to $\sqrt{n}\,h = O(h^{1/2})$: the strong order. Their means add up linearly, to $n\,h^2 = O(h)$: the weak order. The proof also shows that only the first three moments of the increments matter for the weak error, so ΔW_k can be replaced by $\pm\sqrt{h}$ coin flips without changing the weak order (Kloeden and Platen, 1992, Section 14.1).

Theorem 7.11 (Milstein). *Suppose (A1)–(A3) hold, that a and b are regular enough for the functions $b\,\partial_x b$, $L^0 a$, $L^1 a$, $L^0 b$, $L^0(b\,\partial_x b)$ and $L^1(b\,\partial_x b)$ to exist and be continuous, where $L^0 = \partial_t + a\,\partial_x + \frac{1}{2}b^2\partial_{xx}$ and $L^1 = b\,\partial_x$, that $b\,\partial_x b$ is Lipschitz in x uniformly in t, and that all these functions grow at most linearly in x, uniformly in t.*

(i) The Milstein scheme has strong order 1: $\mathbb{E}|X_T - \tilde{X}_T^h| \le C\,h$ (Kloeden and Platen, 1992, Theorem 10.3.5, a case of Theorem 10.6.3).

(ii) Under the hypotheses of Theorem 7.10(ii), its weak order is 1, and no better in general.

The hypotheses hold, for example, when a and b are affine in x with continuously differentiable coefficients in t (Brownian motion with drift, GBM, Ornstein–Uhlenbeck).

For (ii), the proof of Theorem 7.10(ii) applies verbatim: the extra term $\frac{1}{2}b\,\partial_x b\,((\Delta W)^2 - h)$ has mean zero and changes the second and third moments of the step by $O(h^2)$, so the local weak error is still of order h^2, and it is not smaller in general. Milstein improves the paths, not the order of the bias of prices.

For GBM, $\mathrm{d}S = \mu S\,\mathrm{d}t + \sigma S\,\mathrm{d}W$, both schemes multiply by a factor at each step: $1 + \mu h + \sigma\Delta W_k$ for Euler, plus $\frac{1}{2}\sigma^2((\Delta W_k)^2 - h)$ for Milstein. The exact factor is $e^{(\mu-\sigma^2/2)h+\sigma\Delta W_k} = 1 + \mu h + \sigma\Delta W_k + \frac{1}{2}\sigma^2((\Delta W_k)^2 - h) + O(h^{3/2})$, so the Milstein term is precisely the leading local error of Euler. Since GBM has an exact solution on the same Brownian path, $S_0 e^{(\mu-\sigma^2/2)T+\sigma W_T}$, it is the natural laboratory for both orders.

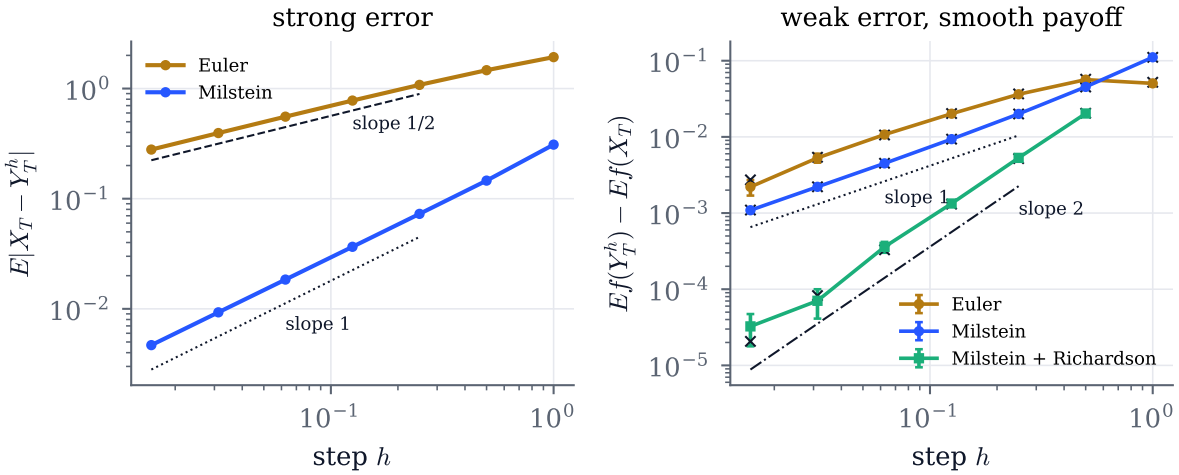

Figure 7.2: Euler and Milstein for GBM with the test-case parameters ($\mu = r - y = -1\%$, $\sigma = 20\%$, $T = 1$), 10^6 paths, seed 2026. Left: strong error $\mathbb{E}|X_T - Y_T^h|$ against the exact solution on the same path. Right: weak error $\mathbb{E}f(Y_T^h) - \mathbb{E}f(X_T)$ for $f(x) = 5\log(1 + e^{(x-100)/5})$, estimated from the differences on the same paths (± 2 standard errors), and the Richardson-extrapolated Milstein error (absolute value) of Theorem 7.13; crosses: exact values.

Figure 7.2 shows the measurement with $m = 1, 2, 4, \ldots, 64$ steps, all schemes driven by the same 10^6 Brownian paths (the coarse increments are sums of fine ones). At $h = 1/64$ the strong errors are 0.279 for Euler and 0.0047 for Milstein; the fitted slopes over $m \ge 4$ are 0.49 and 0.99. For the weak error we use the smooth call-like payoff $f(x) = 5\log(1 + e^{(x-K)/5})$, $K = 100$, and estimate $\mathbb{E}f(Y_T^h) - \mathbb{E}f(X_T)$ by the average of $f(Y_T^h) - f(X_T)$ over the paths, which is unbiased

for the difference and has a small variance because the two terms are close. For GBM the weak errors can also be computed without simulation, since $\log Y_T^h$ is a sum of m i.i.d. terms with an explicit law (`gbm_scheme_weak_error` convolves it numerically). The exact values are 0.01059, 0.00540, 0.00273 for Euler at $m = 16$, 32, 64, and 0.00451, 0.00221, 0.00110 for Milstein: both of order h, as the theorems predict, and the Monte Carlo estimates agree with them within 2.1 standard errors (Euler's standard errors are 0.0005, 0.0004, 0.0003; Milstein's are ten times smaller, because its strong error is smaller). The bias is small next to the price $\mathbb{E}f(X_T) = 8.206$: at $\sigma = 20\%$ a few steps per year are enough for this payoff.

Remark 7.12 (When the theorems do not apply). The hypotheses are not decoration.

- *Non-Lipschitz coefficients.* The square-root diffusion $\mathrm{d}v = \kappa(\theta - v)\,\mathrm{d}t + \xi\sqrt{v}\,\mathrm{d}W$ of the Heston model violates (A1) at $v = 0$, and the Euler step can make v negative, after which $\sqrt{v}$ is undefined. Fixes such as full truncation and Andersen's scheme are in Chapter 19.
- *Superlinear growth.* For a large class of equations with superlinearly growing coefficients and noise acting from the start, for example $\mathrm{d}X = -X^3\mathrm{d}t + \mathrm{d}W$, the moments of the Euler approximation at time T diverge as $h \to 0$, although those of X_T are finite (Hutzenthaler et al., 2011). (Without noise, Euler for $x' = -x^3$ started at x_0 stays bounded as soon as $hx_0^2 \le 2$; the noise is what occasionally throws the scheme far enough to explode.)
- *Positivity.* The Euler factor $1 + \mu h + \sigma\Delta W_k$ for GBM is negative with probability $N(-(1+\mu h)/(\sigma\sqrt{h}))$: negligible for moderate h, fatal for a payoff such as $\log S_T$ (Exercise 7.5).
- *Nonsmooth functionals.* A digital payoff is not even continuous. The expansion (7.8), hence weak order 1, still holds for bounded measurable f when the coefficients are smooth and the diffusion satisfies a nondegeneracy condition of Hörmander type (Bally and Talay, 1996). For a barrier or a maximum the error is of order $h^{1/2}$, because the path is only inspected on the grid; see Theorem 16.7 (for GBM), Theorem 16.8 (Euler for general diffusions) and Exercise 7.6.

▶ Interactive demo: measure the strong and weak errors of Euler and Milstein for GBM, for any volatility, drift and number of paths.

7.4 Richardson extrapolation

An error expansion such as (7.8) can be used, not just bounded.

Proposition 7.13 (Richardson extrapolation). *Let Y^h be a scheme and f a payoff such that, for some $c \in \mathbb{R}$ and $0 < p < q$,*

$$\mathbb{E}f(Y_T^h) = \mathbb{E}f(X_T) + c\,h^p + O(h^q) \qquad (h \to 0), \tag{7.9}$$

and set $R^h = (2^p f(Y_T^h) - f(Y_T^{2h}))/(2^p - 1)$.

(i) $\mathbb{E}R^h = \mathbb{E}f(X_T) + O(h^q)$.

(ii) If Y_T^h and Y_T^{2h} are computed on the same Brownian path, f is Lipschitz and $\mathbb{E}|Y_T^h - X_T|^2 \to 0$, then $\operatorname{Var} R^h \to \operatorname{Var} f(X_T)$. If they are computed on independent paths, $\operatorname{Var} R^h \to \frac{4^p+1}{(2^p-1)^2}\operatorname{Var} f(X_T)$, which is $5\operatorname{Var} f(X_T)$ for $p = 1$.

Proof. (i) $\mathbb{E}R^h = \frac{2^p(\theta+ch^p)-(\theta+c2^ph^p)}{2^p-1} + O(h^q) = \theta + O(h^q)$, with $\theta = \mathbb{E}f(X_T)$; the constant c cancels without being known.

(ii) With a Lipschitz constant L for f,

$$R^h - f(X_T) = \frac{2^p\big(f(Y_T^h) - f(X_T)\big) - \big(f(Y_T^{2h}) - f(X_T)\big)}{2^p - 1},$$

whose L^2 norm is at most $\frac{L}{2^p-1}\big(2^p\|Y_T^h - X_T\|_2 + \|Y_T^{2h} - X_T\|_2\big) \to 0$. Convergence in L^2 implies convergence of variances. On independent paths, $\operatorname{Var} R^h = (4^p \operatorname{Var} f(Y_T^h) + \operatorname{Var} f(Y_T^{2h}))/(2^p-1)^2$, and each variance tends to $\operatorname{Var} f(X_T)$ by the same L^2 convergence. □

The assumption is the expansion (7.9), not merely an error bound: a bound $|\text{error}| \le Ch^p$ allows an oscillating error that extrapolation can make worse. For Euler, (7.9) with $p = 1$, $q = 2$ is (7.8); the mean-square convergence in (ii) holds under the hypotheses of Theorems 7.10 and 7.11, whose proofs bound $\mathbb{E}|X_T - Y_T^h|^2$. The coupling in (ii) is what makes extrapolation cheap: the pair costs $m + m/2$ steps and has the variance of one path, while independent pairs pay a factor 5 in variance. The same coupling of a fine and a coarse path is the building block of multilevel Monte Carlo (Theorem 12.4).

In the experiment of Figure 7.2, Richardson applied to Milstein has exact errors -0.000329, -0.000083, -0.000021 at $m = 16$, 32, 64, against 0.00451, 0.00221, 0.00110 before extrapolation: they fall by a factor 4 per halving of h, so $q = 2$, and the Monte Carlo estimates of the figure follow them. For Euler the extrapolated errors are 0.0162, 0.0038, 0.00091, 0.00022, 0.000055 at $m = 4, \dots, 64$, also of order h^2; for $m \ge 16$ they are below the noise of 10^6 paths. If the error expansion is in powers of $h^{1/2}$, as for discretely monitored extremes, the exponent $p = 1$ is wrong and the extrapolation must use $p = \frac{1}{2}$ (Exercise 7.6).

7.5 Time steps or paths?

A discretised Monte Carlo estimator has two errors: the bias of the scheme, which falls with the step, and the statistical error, which falls with the number of paths. With a fixed computing budget, more steps per path means fewer paths. The mean-square error trades the two (Theorem 6.11).

Let $\hat\theta = \frac{1}{n}\sum_{i=1}^n f(Y_{T,i}^h)$ be computed from n independent paths with $m = T/h$ steps, each step costing κ, so that a budget C buys $n = C/(\kappa m)$ paths (we ignore integer rounding). With bias $b(h) = \mathbb{E}f(Y_T^h) - \mathbb{E}f(X_T)$ and variance $v(h) = \operatorname{Var} f(Y_T^h)$,

$$\operatorname{MSE}(h) = b(h)^2 + \frac{v(h)}{n} = b(h)^2 + \frac{v(h)\,\kappa T}{C\,h}. \tag{7.10}$$

Proposition 7.14 (Allocation of effort)**.** *Suppose that $b(h) = c\,h^\beta(1 + o(1))$ with $c \ne 0$, $\beta > 0$, that $v(h) \to v > 0$ as $h \to 0$, and that $|b|$ is bounded away from 0 on $[\varepsilon, h_0]$ for every $\varepsilon > 0$.*

(i) For the model $c^2h^{2\beta} + v\kappa T/(Ch)$ of (7.10), the optimal step is

$$h^* = \Big(\frac{v\,\kappa\,T}{2\beta\,c^2\,C}\Big)^{1/(2\beta+1)},$$

so the number of steps grows like $C^{1/(2\beta+1)}$ and the number of paths like $C^{2\beta/(2\beta+1)}$. At the optimum the variance equals 2β times the squared bias, and $\mathrm{MSE}^* = (1+2\beta)\,c^2(h^*)^{2\beta} = M^*C^{-2\beta/(2\beta+1)}$ with $M^* = (1+2\beta)\,c^{2/(2\beta+1)}\big(v\kappa T/(2\beta)\big)^{2\beta/(2\beta+1)}$.

(ii) For every choice of steps $h_C \in (0, h_0]$, $\liminf_{C\to\infty} C^{2\beta/(2\beta+1)}\,\mathrm{MSE}(h_C) \geq M^*$, and the choice $h_C = h^*$ attains this limit. The root-mean-square error therefore cannot fall faster than $C^{-\beta/(2\beta+1)}$.

Proof. (i) Let $g(h) = c^2h^{2\beta} + A/h$ with $A = v\kappa T/C$. Its derivative $g'(h) = h^{-2}\big(2\beta c^2h^{2\beta+1} - A\big)$ is negative for $h < h^*$ and positive for $h > h^*$, so h^* is the global minimiser. At h^*, $A/h = 2\beta c^2h^{2\beta}$; this equation says that the variance term is 2β times the squared bias, so $g(h^*) = (1+2\beta)c^2(h^*)^{2\beta}$, and substituting h^* gives M^*. The paths number $n = Ch^*/(\kappa T) \propto C^{1-1/(2\beta+1)}$.

(ii) Fix $\eta \in (0,1)$ and pass to a sequence $C_k \to \infty$ along which $C^{2\beta/(2\beta+1)}\mathrm{MSE}(h_C)$ converges to its liminf. Either $\liminf_k h_{C_k} > 0$: then $h_{C_k} \geq \varepsilon$ for some $\varepsilon > 0$ and all large k, the mean-square error is at least $\inf_{[\varepsilon,h_0]} b^2 > 0$, and the liminf is $+\infty$. Or a further subsequence has $h_{C_k} \to 0$: along it, for k large, $b(h_C)^2 \geq (1-\eta)c^2h_C^{2\beta}$ and $v(h_C) \geq (1-\eta)v$, so $\mathrm{MSE}(h_C) \geq (1-\eta)\,g(h_C) \geq (1-\eta)\,g(h^*) = (1-\eta)M^*C^{-2\beta/(2\beta+1)}$. As η is arbitrary, the liminf is at least M^*. With $h_C = h^* \to 0$, the same expansions with $1+\eta$ give a matching upper bound. □

Duffie and Glynn (1995) derived this allocation rule; the mean-square version above follows Glasserman (2003), Section 1.1.3, which also allows a cost per path growing like $h^{-\eta}$. The consequences, with the computing budget C:

method	bias order β	root-mean-square error
exact simulation (Theorem 7.2)	none	$C^{-1/2}$
Euler or Milstein, optimal allocation	1	$C^{-1/3}$
coupled Richardson extrapolation	2	$C^{-2/5}$
multilevel Monte Carlo (Theorem 12.6)	—	$C^{-1/2}$, up to a logarithm for Euler

For a first-order scheme, the rule of thumb is $n \propto m^2$: doubling the number of steps calls for four times as many paths. Spending the whole budget on paths with a fixed step leaves the bias in place, and the error stalls at $|b(h)|$.

For the payoff of Figure 7.2 the bias constant of Euler is $c \approx 0.17$ (m times the exact weak error is 0.169 and 0.173 at $m = 16$ and 32; the limit $h \to 0$ is 0.176), and $v = \mathrm{Var}\,f(X_T) = 154$. With a budget of $C = 10^6$ steps the optimal allocation is $m^* \approx 7$ steps and $n^* \approx 1.4 \times 10^5$ paths, for a root-mean-square error of 0.041; exact simulation with 10^6 paths at the same cost reaches 0.012. At $C = 10^8$, $m^* \approx 33$, $n^* \approx 3 \times 10^6$, and the errors are 0.0088 against 0.0012: the gap widens like $C^{1/6}$ (Figure 7.3). When an exact sampler exists, it should be used.

▶ Interactive demo: choose a budget and a bias constant and see the mean-square error split between bias and variance as the number of steps changes.

Further reading

Glasserman (2003) is the reference for this chapter: Section 3.1 for the three constructions of

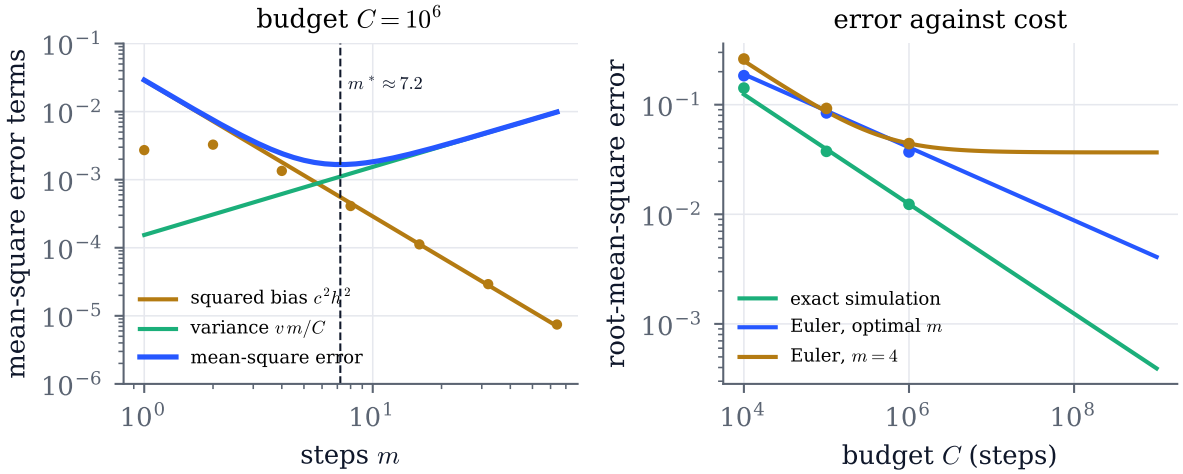

Figure 7.3: Allocation of effort for the payoff of Figure 7.2 with Euler, using the constants $c = 0.17$ and $v = 154$; cost in units of one step of one path. Left: squared bias, variance and mean-square error against the number of steps for a budget of 10^6 (dots: exact squared biases). Right: root-mean-square error against the budget for exact simulation, Euler with the optimal allocation, and Euler with 4 fixed steps (lines: model; dots: 100 independent replications each).

Brownian paths, Section 3.2 for GBM, Sections 6.1–6.2 for the Euler and Milstein schemes, convergence orders and extrapolation (Section 6.2.4), and Section 1.1.3 for the allocation of effort. Kloeden and Platen (1992) prove the convergence theorems in full, with the stochastic Taylor expansions behind every scheme (Chapter 5) and higher-order strong and weak schemes (Chapters 10–15). Talay and Tubaro (1990) and Bally and Talay (1996) are the sources for the weak error expansion, and Mil'shtein (1975) introduced the scheme that bears his name.

Chapter 8

Antithetic variates, moment matching and control variates

The plain Monte Carlo estimator of Chapter 6 has standard error $\sigma/\sqrt{n}$: one more correct digit costs a hundred times more samples. Independent sampling cannot beat the rate $n^{-1/2}$, but it can beat the constant. This chapter is about making σ^2, or rather the product of σ^2 and the cost per sample, smaller by using what we know about the problem.

Four ideas, each simple to state. *Antithetic variates* use the symmetry of the normal law: pair every draw Z with $-Z$. *Moment matching* forces the sample to have the moments of the model. *Common random numbers* correlate the estimators of two related prices, so that their difference is accurate even when each price is not. *Control variates* subtract a correlated quantity whose mean is known exactly, and include, as limiting cases, regression on several such quantities and hedging. Each comes with an exact variance formula, a condition under which it helps, and examples in which it does not. They combine, but not multiplicatively; Section 8.6 measures all of them on the course's test case.

8.1 Setting and running examples

We estimate $\mu = \mathbb{E}[f(Z)]$, where Z is a standard normal vector in $\mathbb{R}^d$ (one coordinate per time step or per asset) and $f(Z) \in L^2$ with $\sigma^2 = \operatorname{Var} f(Z) > 0$. The plain estimator $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^n f(Z_i)$ of Theorem 6.1 has variance σ^2/n. Every method below produces another unbiased (or asymptotically unbiased) estimator. We report its *standard-error ratio* to $\hat{\mu}_n$ at equal n, $\varrho = \sigma_2/\sigma$, and, if one of its samples costs c times a plain one, its efficiency relative to plain Monte Carlo (Theorem 6.9)

$$\mathcal{E}_{2,1} = \frac{\sigma^2\tau_1}{\sigma_2^2\tau_2} = \frac{1}{c\,\varrho^2}. \tag{8.1}$$

Following the course, we take the conservative value $c = 2$ for every method when a cost must be assumed; the notebook measures it.

Throughout, the test case is $S = K = 100$, $T = 1$, $r = 1\%$, $y = 2\%$, $\sigma = 20\%$, and

$$S_T = S\exp\left(\nu T + sZ\right), \qquad \nu = r - y - \tfrac{\sigma^2}{2} = -0.03, \qquad s = \sigma\sqrt{T} = 0.2,$$

with $Z \sim \mathcal{N}(0,1)$ (Theorem 7.2). The running examples are

(a) the *stock*, $f(Z) = e^{-rT} S_T$, with $\mu = Se^{-yT} = 98.0199$: trivial to price, and therefore a good test;

(b) the *call*, $f(Z) = e^{-rT}(S_T - K)^+$, with $\mu = 7.36429$ and $\sigma^2 = 158.006$ (for the undiscounted payoff, 161.198), so $\sigma = 12.570$;

(c) the *square-root call* of Theorem 8.13.

8.2 Antithetic variates

If $Z \sim \mathcal{N}(0, I_d)$ then $-Z$ has the same law. Draw $Z_1, \dots, Z_n$ independently and use each draw twice:

$$\hat{\mu}_n^{\mathrm{a}} = \frac{1}{n} \sum_{i=1}^{n} \frac{f(Z_i) + f(-Z_i)}{2}. \tag{8.2}$$

For a path driven by the increments $Z = (Z_1, \dots, Z_m)$, the antithetic path is driven by $-Z$: it is the original path reflected in its drift. With uniforms, U is paired with $1 - U$, which under inversion (Theorem 5.8) is the same as Z with $-Z$. The estimator averages n i.i.d. *pair averages*, so its standard error comes from the n pair averages, never from the $2n$ individual payoffs, which are dependent.

When does pairing help? The answer rests on an inequality that goes back to Chebyshev in one dimension, and in several dimensions is the simplest case of *association* (Esary et al., 1967).

> **Lemma 8.1** (Monotone functions are positively correlated)**.** *Let $Z = (Z_1, \dots, Z_d)$ have independent components, and let $g, h : \mathbb{R}^d \to \mathbb{R}$ be Borel functions, nondecreasing in each coordinate, with $g(Z), h(Z) \in L^2$. Then* $\mathrm{Cov}(g(Z), h(Z)) \geq 0$.

Proof. Assume first that g and h are bounded, and argue by induction on d. For $d = 1$, let Z' be an independent copy of Z. Since g and h are nondecreasing, $g(Z) - g(Z')$ and $h(Z) - h(Z')$ have the same sign, so $(g(Z) - g(Z'))(h(Z) - h(Z')) \geq 0$; taking expectations, $2\mathbb{E}[g(Z)h(Z)] - 2\mathbb{E}[g(Z)]\mathbb{E}[h(Z)] \geq 0$.

For $d \geq 2$, write $Z = (W, Z_d)$ with $W = (Z_1, \dots, Z_{d-1})$ and, for $w \in \mathbb{R}^{d-1}$,

$$G(w) = \mathbb{E}[g(w, Z_d)], \qquad H(w) = \mathbb{E}[h(w, Z_d)], \qquad Q(w) = \mathbb{E}[g(w, Z_d)h(w, Z_d)].$$

These are bounded and measurable (Fubini), and G, H are nondecreasing in each coordinate of w (integrate the pointwise inequalities). For fixed w, the functions $t \mapsto g(w,t)$ and $t \mapsto h(w,t)$ are nondecreasing, so the case $d = 1$ gives $Q(w) \geq G(w)H(w)$. Since W and Z_d are independent, Fubini's theorem gives $\mathbb{E}[g(Z)h(Z)] = \mathbb{E}[Q(W)]$, $\mathbb{E}[g(Z)] = \mathbb{E}[G(W)]$ and $\mathbb{E}[h(Z)] = \mathbb{E}[H(W)]$. Hence, by the induction hypothesis applied to G and H,

$$\mathbb{E}[g(Z)h(Z)] = \mathbb{E}[Q(W)] \geq \mathbb{E}[G(W)H(W)] \geq \mathbb{E}[G(W)]\,\mathbb{E}[H(W)] = \mathbb{E}[g(Z)]\,\mathbb{E}[h(Z)].$$

For general g, h, the truncations $g_k = \max(\min(g, k), -k)$ and h_k are bounded and still nondecreasing, so $\mathrm{Cov}(g_k(Z), h_k(Z)) \geq 0$; as $k \to \infty$, dominated convergence ($|g_k h_k| \leq |gh|$, $|g_k| \leq |g|$) gives the claim. □

Proposition 8.2 (Antithetic variates)**.** *Let $Z \overset{d}{=} -Z$, $f(Z) \in L^2$ with $\sigma^2 = \operatorname{Var} f(Z) > 0$, and $\rho = \operatorname{Corr}(f(Z), f(-Z))$. Write $f = f_{\mathrm{e}} + f_{\mathrm{o}}$ with the even and odd parts $f_{\mathrm{e}}(z) = \frac{1}{2}(f(z)+f(-z))$ and $f_{\mathrm{o}}(z) = \frac{1}{2}(f(z) - f(-z))$.*

(i) $f_{\mathrm{e}}(Z)$ and $f_{\mathrm{o}}(Z)$ are uncorrelated, so $\sigma^2 = \operatorname{Var} f_{\mathrm{e}}(Z) + \operatorname{Var} f_{\mathrm{o}}(Z)$, and

$$\rho = \frac{\operatorname{Var} f_{\mathrm{e}}(Z) - \operatorname{Var} f_{\mathrm{o}}(Z)}{\operatorname{Var} f_{\mathrm{e}}(Z) + \operatorname{Var} f_{\mathrm{o}}(Z)}.$$

(ii) $\hat{\mu}_n^{\mathrm{a}}$ is unbiased and

$$\operatorname{Var}(\hat{\mu}_n^{\mathrm{a}}) = \frac{\operatorname{Var} f_{\mathrm{e}}(Z)}{n} = \frac{\sigma^2}{n}\,\frac{1+\rho}{2}. \tag{8.3}$$

(iii) If the components of Z are independent and f is monotone in each coordinate (nondecreasing in some, nonincreasing in the others), then $\rho \le 0$: n antithetic pairs have at most the variance of $2n$ independent draws.

Proof. (i) Since $f(-Z) \overset{d}{=} f(Z)$, $\mathbb{E} f_{\mathrm{o}}(Z) = 0$ and $\mathbb{E}[f_{\mathrm{e}}(Z) f_{\mathrm{o}}(Z)] = \frac{1}{4}\mathbb{E}[f(Z)^2 - f(-Z)^2] = 0$. So $\sigma^2 = \operatorname{Var}(f_{\mathrm{e}} + f_{\mathrm{o}})(Z)$ splits into the two variances, and $\operatorname{Cov}(f(Z), f(-Z)) = \operatorname{Cov}((f_{\mathrm{e}} + f_{\mathrm{o}})(Z), (f_{\mathrm{e}} - f_{\mathrm{o}})(Z)) = \operatorname{Var} f_{\mathrm{e}}(Z) - \operatorname{Var} f_{\mathrm{o}}(Z)$. Divide by $\sigma^2 = \operatorname{Var} f(-Z)$.

(ii) The pair averages $f_{\mathrm{e}}(Z_i)$ are i.i.d. with mean μ. By (i), $\operatorname{Var} f_{\mathrm{e}} + \operatorname{Var} f_{\mathrm{o}} = \sigma^2$ and $\operatorname{Var} f_{\mathrm{e}} - \operatorname{Var} f_{\mathrm{o}} = \rho\sigma^2$, so $\operatorname{Var} f_{\mathrm{e}}(Z) = \sigma^2(1+\rho)/2$.

(iii) Choose signs $\varepsilon_j \in \{\pm 1\}$ such that $\tilde{f}(v) = f(\varepsilon_1 v_1, \dots, \varepsilon_d v_d)$ is nondecreasing in every coordinate, and let $V = (\varepsilon_j Z_j)_j$, which again has independent components. Then $f(Z) = \tilde{f}(V)$ and $f(-Z) = \tilde{f}(-V)$. Both $\tilde{f}$ and $v \mapsto -\tilde{f}(-v)$ are nondecreasing in every coordinate, so Theorem 8.1 gives $\operatorname{Cov}(\tilde{f}(V), -\tilde{f}(-V)) \ge 0$, that is $\rho \le 0$. □

The even–odd split says exactly what antithetic sampling does: it integrates the odd part of f with no error at all, and samples the even part as before. It helps when f is nearly odd about the centre of the distribution (nearly linear payoffs are the main example), and it is useless when f is even.

Cost and efficiency. A pair costs c times a single draw with $1 \le c \le 2$: $c \approx 2$ when evaluating the payoff dominates, and less when generating normals or paths dominates, since the antithetic path reuses them. By (8.1) and (8.3), the efficiency of n pairs against n plain draws is $2/(c(1+\rho))$; at $c = 2$ it is $1/(1+\rho)$, which is at least 1 under the hypotheses of (iii).

Example 8.3 (The stock)**.** For $f(z) = e^{-rT} S e^{\nu T} e^{sz}$ the even and odd parts are multiples of $\cosh(sz)$ and $\sinh(sz)$. With $\mathbb{E}\cosh(sZ) = e^{s^2/2}$, $\mathbb{E}\cosh^2(sZ) = \frac{1}{2}(e^{2s^2} + 1)$ and $\mathbb{E}\sinh^2(sZ) = \frac{1}{2}(e^{2s^2} - 1)$,

$$\operatorname{Var}\cosh(sZ) = \tfrac{1}{2}(e^{s^2} - 1)^2, \qquad \operatorname{Var}\sinh(sZ) = \tfrac{1}{2}(e^{2s^2} - 1), \qquad \text{so} \qquad \rho = -e^{-s^2}.$$

For $s = 0.2$: $\rho = -0.9608$, the standard-error ratio is $\sqrt{(1+\rho)/2} = \sqrt{(1 - e^{-s^2})/2} = 0.1400 \approx s/\sqrt{2}$, and the efficiency at $c = 2$ is $1/(1 - e^{-s^2}) = 25.5$. These are the course's "about 14%" and "about 25": over the range of sZ the exponential is nearly linear, and linear functions are integrated exactly.

For payoffs that vanish on part of the line, the antithetic correlation has a closed form. This is the course's exam formula.

Proposition 8.4 (Disjoint payoffs)**.** *In the setting of Theorem 8.2, suppose that $f(Z)f(-Z) = 0$ almost surely, and let $\mu = \mathbb{E}f(Z)$. Then $\mu^2 \le \sigma^2$ and*

$$\rho = -\frac{\mu^2}{\sigma^2}, \qquad \operatorname{Var}\left(\hat{\mu}_n^{\mathrm{a}}\right) = \frac{\sigma^2 - \mu^2}{2n}, \qquad \frac{\sqrt{\operatorname{Var}\,\hat{\mu}_n^{\mathrm{a}}}}{\sqrt{\operatorname{Var}\,\hat{\mu}_n}} = \frac{1}{\sqrt{2}}\sqrt{1 - \frac{\mu^2}{\sigma^2}},$$

and the efficiency at $c = 2$ is $\sigma^2/(\sigma^2 - \mu^2) \ge 1$ (infinite if $\mu^2 = \sigma^2$). In the Black–Scholes model the call $(S_T - K)^+$ and the digital call $\mathbf{1}_{\{S_T > K\}}$ satisfy the hypothesis if and only if $K \ge Se^{\nu T}$, the median of S_T; the put and the digital put if and only if $K \le Se^{\nu T}$.

Proof. $\operatorname{Cov}(f(Z), f(-Z)) = \mathbb{E}[f(Z)f(-Z)] - \mu^2 = -\mu^2$, so $\rho = -\mu^2/\sigma^2$; $\rho \ge -1$ gives $\mu^2 \le \sigma^2$, and the rest is (8.3) and $1/(1+\rho)$. For the call, $S_T(z)\,S_T(-z) = S^2 e^{2\nu T}$. If $K \ge Se^{\nu T}$ and $S_T(z) > K$, then $S_T(-z) < S^2 e^{2\nu T}/K \le K$, so the two payoffs are never positive together. If $K < Se^{\nu T}$, both are positive on $\{K < S_T < S^2 e^{2\nu T}/K\}$, a nonempty interval of values of S_T and therefore an event of positive probability. The digital has the same supports, and for puts the inequalities reverse. □

In the test case the median is $Se^{\nu T} = 97.04 < K$, so the at-the-money call is disjoint: with $\mu = 7.36429$ and $\sigma^2 = 158.006$, $\mu^2/\sigma^2 = 0.3432$, so $\rho = -0.3432$, the standard-error ratio is 0.5730 and the efficiency at $c = 2$ is 1.523, the course's "about 57%" and "about 1.5". For strikes below the median the supports overlap, $\mathbb{E}[f(Z)f(-Z)] > 0$, and ρ is larger than $-\mu^2/\sigma^2$ (Figure 8.1).

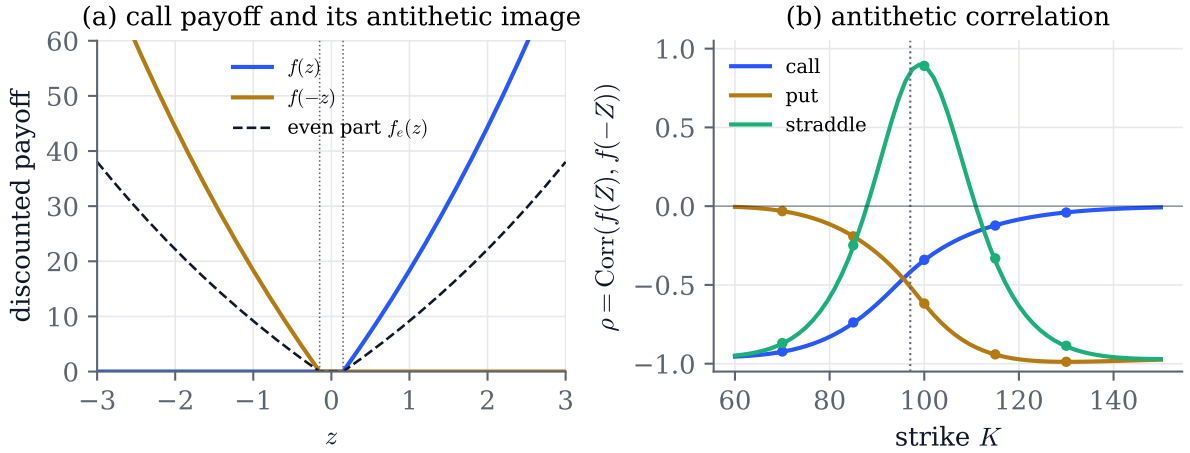

Figure 8.1: (a) The discounted at-the-money call payoff $f(z)$, its antithetic image $f(-z)$ and the even part f_{e}, the only part the antithetic estimator has to sample; the supports are disjoint (Theorem 8.4). (b) The antithetic correlation ρ against the strike, exact (lines, by quadrature) and from 20,000 pairs (dots); the dotted line is the median of S_T. Near the median the straddle $|S_T - K|$ is almost even in z: $\rho = 0.85$ at the median and 0.90 at its maximum, near $K = 99$.

Remark 8.5 (When pairing fails). Without monotonicity, Theorem 8.2(iii) gives no guarantee. If f is even, then $f_{\mathrm{o}} = 0$, $\rho = 1$, and a pair is worth one draw at up to twice the cost. Payoffs that

are nearly even in z (straddles, butterflies, digital ranges centred near the median of S_T) have ρ close to 1 (Exercise 8.2). At the other extreme, a deep out-of-the-money call is disjoint with $\mu^2 \ll \sigma^2$, so $\rho \approx 0$ and the efficiency tends to 1 at $c = 2$: rare events need importance sampling (Chapter 9). For paths, an arithmetic Asian call is nondecreasing in every increment, so (iii) applies; a knock-out barrier is not monotone, and the correlation must be measured.

▶ Interactive demo: the even and odd parts of any payoff, and the antithetic correlation against the strike for calls, puts, digitals and straddles.

8.3 Moment matching

Antithetic pairs make the sample symmetric. Moment matching goes further and transforms the whole sample so that its first moments are exactly those of the model.

Definition 8.6 (Moment matching)**.** Let $Z_1, \dots, Z_n$ be i.i.d. $\mathcal{N}(0,1)$, $\bar{Z}_n = \frac{1}{n}\sum_i Z_i$ and $\tilde{s}_n^2 = \frac{1}{n}\sum_i (Z_i - \bar{Z}_n)^2$ (divisor n).

(a) *Log matching*: $Z'_i = (Z_i - \bar{Z}_n)/\tilde{s}_n$. Then $\frac{1}{n}\sum_i Z'_i = 0$ and $\frac{1}{n}\sum_i Z'^2_i = 1$, so the simulated log-prices $\ln S + \nu T + sZ'_i$ have exactly the mean and the variance of $\ln S_T$.

(b) *Exponential-mean matching*: $Z'_i = \alpha_n + Z_i/\tilde{s}_n$ with $\alpha_n = \frac{s}{2} - \frac{1}{s}\ln\left(\frac{1}{n}\sum_i e^{sZ_i/\tilde{s}_n}\right)$. The sample variance of the Z'_i is still 1, and $\frac{1}{n}\sum_i e^{sZ'_i} = e^{s^2/2} = \mathbb{E}e^{sZ}$: the sample mean of $S'_{T,i}$ equals the forward $Se^{(r-y)T}$.

The moment-matched estimator is $\hat{\mu}_n^{\mathrm{mm}} = \frac{1}{n}\sum_i f(Z'_i)$.

In (b) the sample reprices the forward exactly, so the matched estimates of a call and a put with the same strike satisfy put–call parity exactly. The course also matches the "exponential variance", which requires solving a nonlinear equation for the scale by Newton's method. For paths, match each time step's normals across the sample.

Proposition 8.7 (Moment matching: structure and bias)**.** *In the setting of Theorem 8.6(a), with $n \geq 3$:*

(i) $\hat{\mu}_n^{\mathrm{mm}} = \mathbb{E}f(Z)$ *exactly for every quadratic* $f(z) = a + bz + cz^2$.

(ii) The vector $Z' = (Z'_1, \dots, Z'_n)$ is independent of $(\bar{Z}_n, \tilde{s}_n)$. Its components are identically distributed but not independent ($\sum_i Z'_i = 0$).

(iii) If $|f(z)| \leq Ce^{c|z|}$ for all z, then, with $\mathrm{He}_4(z) = z^4 - 6z^2 + 3$,

$$\mathbb{E}\hat{\mu}_n^{\mathrm{mm}} = \mathbb{E}f(Z) - \frac{\mathbb{E}[f(Z)\mathrm{He}_4(Z)]}{4n} + O(n^{-2}).$$

(iv) The sample variance of $f(Z'_1), \dots, f(Z'_n)$ does not estimate $n \operatorname{Var}\hat{\mu}_n^{\mathrm{mm}}$ in general. If the n draws are split into m batches of size $k = n/m$, each matched separately, the m batch estimates are i.i.d. with mean $\mathbb{E}\hat{\mu}_k^{\mathrm{mm}} = \mu + O(1/k)$. For fixed k and $m \to \infty$, the central limit theorem for these i.i.d. estimates gives an asymptotically valid confidence interval for $\mathbb{E}\hat{\mu}_k^{\mathrm{mm}}$; for fixed m and $k \to \infty$, Theorem 6.8(iii) gives one for μ provided $\sqrt{k}(\hat{\mu}_k^{\mathrm{mm}} - \mu)$ is asymptotically normal, which Theorem 8.8 indicates but we do not prove.

Proof. (i) $\frac{1}{n}\sum_i (a + bZ_i' + cZ_i'^2) = a + c = \mathbb{E}[a + bZ + cZ^2]$.

(ii) Complete $u_0 = n^{-1/2}(1,\dots,1)$ to an orthonormal basis $u_0,\dots,u_{n-1}$ of $\mathbb{R}^n$. The coordinates $G_j = \langle Z, u_j\rangle$ are i.i.d. $\mathcal{N}(0,1)$, being an orthogonal transformation of a standard normal vector. Now $\bar{Z}_n = G_0/\sqrt{n}$, $Z - \bar{Z}_n(1,\dots,1) = \sum_{j\ge 1} G_j u_j$, so with $G = (G_1,\dots,G_{n-1})$ we get $\tilde{s}_n = \|G\|/\sqrt{n}$ and $Z' = \sqrt{n}\sum_{j\ge1}(G_j/\|G\|)u_j$. In polar coordinates the density of G is a function of $\|G\|$ times the uniform measure on the unit sphere, so the direction $G/\|G\|$ is independent of the radius $\|G\|$; both are independent of G_0. Hence Z', a function of the direction, is independent of $(\bar{Z}_n, \tilde{s}_n)$. Permuting the Z_i permutes the Z_i', so they are exchangeable.

(iii) By exchangeability $\mathbb{E}\hat{\mu}_n^{\mathrm{mm}} = \mathbb{E}f(Z_1')$. By (ii), $Z_1' = \sqrt{n}\langle G/\|G\|, w\rangle$ with $w = (\langle e_1, u_j\rangle)_{j\ge1}$ and $\|w\|^2 = 1 - \frac{1}{n}$; by rotation invariance of G this has the law of $\sqrt{n-1}\,G_1/\|G\|$. As $G_1^2 \sim \chi_1^2$ and $\|G\|^2 - G_1^2 \sim \chi^2_{n-2}$ are independent, $G_1^2/\|G\|^2$ has the $\mathrm{Beta}(\frac12, \frac{n-2}{2})$ law, and Z_1' has the density $q_n(x) = \kappa_n\big(1 - \frac{x^2}{n-1}\big)^{(n-4)/2}$ on $|x| < \sqrt{n-1}$, and 0 outside. Write $q_n = \kappa_n\sqrt{2\pi}\,\varphi\, g_n$, with φ the standard normal density and $g_n(x) = (1-u)^{(n-4)/2}e^{x^2/2}$, $u = x^2/(n-1)$. From $\ln(1-u) = -u - \frac{u^2}{2} - \theta(u)$ with $0 \le \theta(u) \le u^3/(1-u)$,

$$\ln g_n(x) = \frac{3x^2}{2(n-1)} - \frac{(n-4)x^4}{4(n-1)^2} - \frac{n-4}{2}\theta(u).$$

For $|x| \le n^{1/4}$ and $n \ge 7$ we have $u \le \frac12$, so $\ln g_n(x) = h(x)/n + O((1+x^6)/n^2)$ with $h(x) = \frac32 x^2 - \frac14 x^4$, and $|h|/n \le 1$; hence $g_n(x) = 1 + h(x)/n + O((1+x^{12})/n^2)$ uniformly there. For $|x| > n^{1/4}$, both $\varphi g_n \le e^{-x^2(n-4)/(2(n-1))} \le e^{-x^2/4}$ and $\varphi(1+|h|/n)$ are dominated by a multiple of $e^{-x^2/4}$, and against $|f| \le Ce^{c|x|}$ these regions contribute $O(e^{-\sqrt{n}/8})$ once $n^{1/4} \ge 8c$. Therefore $J_n(f) := \int f\varphi g_n\,\mathrm{d}x = \mathbb{E}f(Z) + \mathbb{E}[f(Z)h(Z)]/n + O(n^{-2})$. Since $\mathbb{E}f(Z_1') = J_n(f)/J_n(1)$ and $J_n(1) = 1 + \mathbb{E}h(Z)/n + O(n^{-2}) = 1 + \frac{3}{4n} + O(n^{-2})$,

$$\mathbb{E}f(Z_1') = \mathbb{E}f(Z) + \frac{\mathbb{E}[f(Z)(h(Z) - \frac34)]}{n} + O(n^{-2}), \qquad h - \tfrac34 = -\tfrac14\mathrm{He}_4.$$

(iv) For $f(z) = z$ the estimator is exactly $0 = \mathbb{E}Z$, while the sample variance of the Z_i' is 1. Batches drawn from disjoint sets of i.i.d. normals are independent and identically distributed, with the bias $O(1/k)$ of (iii) and finite variance (the Z_i' of a batch are bounded by $\sqrt{k}$); the rest is the central limit theorem and Theorem 6.8(iii). □

Remark 8.8 (Moment matching is a control variate in disguise). Expanding $f(Z_i') \approx f(Z_i) + f'(Z_i)(Z_i' - Z_i)$ with $Z_i' - Z_i \approx -\bar{Z}_n - Z_i(\tilde{s}_n - 1)$ and $\tilde{s}_n - 1 \approx \frac12(\frac1n\sum_j Z_j^2 - 1)$, and using Stein's identities $\mathbb{E}f'(Z) = \mathbb{E}[Zf(Z)] =: a_1$ and $\mathbb{E}[Zf'(Z)] = \mathbb{E}[(Z^2-1)f(Z)] =: a_2$, one finds

$$\hat{\mu}_n^{\mathrm{mm}} \approx \frac1n\sum_{i=1}^n \Big(f(Z_i) - a_1 Z_i - \frac{a_2}{2}(Z_i^2 - 1)\Big),$$

so that, heuristically, $\sqrt{n}(\hat{\mu}_n^{\mathrm{mm}} - \mu)$ is asymptotically normal with variance $\mathrm{Var}\big(f(Z) - a_1 Z - \frac{a_2}{2}(Z^2-1)\big)$. Since Z and $Z^2 - 1$ are uncorrelated with variances 1 and 2, the coefficients a_1 and $a_2/2$ are exactly the optimal ones of Theorem 8.15: *asymptotically, log matching is the optimal control-variate estimator with the controls Z and Z^2*, obtained without estimating any coefficient (Exercise 8.3 proves this for $f(z) = z^3$). The bias, by contrast, is exact to first order by Theorem 8.7(iii): its constant $-\frac14\mathbb{E}[f(Z)\mathrm{He}_4(Z)]$ is -6 for z^4 and 1.75 for the call (Figure 8.3).

By this linearisation the test-case standard-error ratios are 0.0162 for the stock and 0.188 for the call, the course's "about 1.6%" and "about 20%" (and its "11% and 33% of the antithetic errors" are 0.0162/0.140 and 0.188/0.573). Exponential-mean matching reprices the stock exactly and gives 0.182 for the call.

Remark 8.9 (Batches: how many?). By Theorem 8.7(iv), with m batches the bias is about $c_f m/n$ while the standard error is of order $n^{-1/2}$, so their ratio grows like $m/\sqrt{n}$; with too few batches the error estimate itself is noisy ($m-1$ degrees of freedom). Keep m well below $\sqrt{n}$: with $n = 2^{20}$ and $m = 256$ the bias of the call estimate is about $1.75 \cdot 256/2^{20} \approx 4 \times 10^{-4}$, a fifth of its standard error 2.2×10^{-3}. How coverage collapses when m is too large is the subject of Exercise 8.4.

Remark 8.10 (Antithetic variates against moment matching). Antithetic sampling acts *locally*: each draw is paired with its mirror image, independently of the others, so the pair averages are i.i.d. and the usual standard error applies. The resulting sample is symmetric, so its mean and all its odd moments are exactly 0, but its variance is not matched. Moment matching acts *globally*: one affine map, computed from the whole sample, fixes the first two moments and introduces dependence and an $O(1/n)$ bias. The two can be combined, and the order matters: rescaling a symmetric sample keeps it symmetric, so antithetic sampling followed by scale matching fixes the mean, the variance and every odd moment; but exponential-mean matching shifts the sample and destroys its symmetry. With (randomised) quasi-Monte Carlo point sets (Chapter 11) the low moments are already integrated very accurately, and matching adds bias for little gain.

8.4 Common random numbers

Often the quantity of interest is a difference of two prices: the same option under two models or two volatilities, two strikes, or a bumped Greek (Chapter 13). Each price may be poorly estimated while their difference is estimated well, provided both use the same random numbers.

Proposition 8.11 (Common random numbers)**.** *Let U be a random vector with independent components, $f_1(U), f_2(U) \in L^2$ with variances σ_1^2, σ_2^2, and $\delta = \mathbb{E}f_1(U) - \mathbb{E}f_2(U)$.*

(i) With independent draws for the two terms, the estimator of δ from n draws of each has variance $(\sigma_1^2+\sigma_2^2)/n$; with common random numbers, $\frac{1}{n}\sum_i (f_1(U_i)-f_2(U_i))$ has variance $(\sigma_1^2 + \sigma_2^2 - 2\operatorname{Cov}(f_1(U), f_2(U)))/n$.

(ii) If, for each j, f_1 and f_2 are monotone in u_j in the same direction, then $\operatorname{Cov}(f_1(U), f_2(U)) \geq 0$ and common random numbers never increase the variance. If they are monotone in opposite directions in every coordinate, the covariance is ≤ 0 and common random numbers never decrease it; if moreover $U \overset{d}{=} -U$, pairing $f_1(U_i)$ with $f_2(-U_i)$ restores a nonnegative covariance without changing the mean.

(iii) Let $f(\theta, u)$ depend on a parameter, with $|f(\theta+h, u) - f(\theta, u)| \leq L(u)|h|$ and $\mathbb{E}L(U)^2 < \infty$. Then the common-random-number difference quotient $(f(\theta+h, U) - f(\theta, U))/h$ has variance at most $\mathbb{E}L(U)^2$ for every h, while with independent draws its variance $(\sigma^2(\theta+h) + \sigma^2(\theta))/h^2$ tends to infinity as $h \to 0$ whenever $\sigma^2(\theta) = \operatorname{Var} f(\theta, U) > 0$.

Proof. (i) is the variance of a difference. (ii) As in the proof of Theorem 8.2(iii), let $V_j = \varepsilon_j U_j$ with signs chosen so that $g_k(v) = f_k(\varepsilon_1 v_1, \ldots, \varepsilon_d v_d)$ is nondecreasing in every coordinate for $k = 1$; V has independent components and $f_k(U) = g_k(V)$. In the first case g_2 is nondecreasing

as well, in the second $-g_2$ is, and Theorem 8.1 gives $\operatorname{Cov}(g_1(V), g_2(V)) \geq 0$, respectively ≤ 0. For the pairing, $u \mapsto f_2(-u)$ moves in the same direction as f_1 in every coordinate, so the first case applies to f_1 and $f_2(-\cdot)$, and $f_2(-U) \stackrel{d}{=} f_2(U)$. (iii) The variance is at most the second moment, which is at most $\mathbb{E}L(U)^2$; the independent case is (i) divided by h^2. □

For the test case, the prices at $\sigma = 20\%$ and 21% differ by 0.39048. Both discounted payoffs are nondecreasing in z, and their correlation is 0.99999: the variance per draw of the difference is 0.535 with common normals and 334.85 with independent ones, a standard-error ratio of 0.040. Comparing a call with a put is the opposite case (Exercise 8.5). Finite-difference Greeks are the main application; Theorem 13.1 works out the optimal bump with and without common random numbers. Glasserman and Yao (1992) give general conditions under which common random numbers are guaranteed to help.

8.5 Control variates

Suppose that alongside the payoff $Y = f(Z)$ we can compute, at little extra cost, a *control* $X = g(Z)$ whose mean μ_X is known exactly: the stock, a vanilla option, a geometric average, a hedge. With i.i.d. copies (X_i, Y_i) and a coefficient β,

$$\hat{\mu}_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} \big(Y_i - \beta(X_i - \mu_X)\big). \tag{8.4}$$

We estimate the mean of Y by estimating the mean of the error $Y - \beta X$ and adding back the known part. The course uses $\beta = 1$.

Theorem 8.12 (The optimal control coefficient)**.** *Let $X, Y \in L^2$ with $\sigma_X^2 = \operatorname{Var} X > 0$, $\sigma_Y^2 = \operatorname{Var} Y$, $\sigma_{XY} = \operatorname{Cov}(X, Y)$ and $\rho = \operatorname{Corr}(X, Y)$.*

(i) For every β, $\hat{\mu}_n(\beta)$ is unbiased with $n \operatorname{Var} \hat{\mu}_n(\beta) = \sigma_Y^2 - 2\beta\sigma_{XY} + \beta^2\sigma_X^2$.

(ii) This is minimised uniquely at $\beta^* = \sigma_{XY}/\sigma_X^2$, where

$$n \operatorname{Var} \hat{\mu}_n(\beta^*) = \sigma_Y^2\,(1 - \rho^2).$$

(iii) If a sample with its control costs c times a plain sample, the efficiency of $\hat{\mu}_n(\beta^*)$ relative to plain Monte Carlo is $1/(c(1-\rho^2))$.

Proof. (i) $\mathbb{E}[Y_i - \beta(X_i - \mu_X)] = \mathbb{E}Y$, and the variance of $Y - \beta X$ expands as stated. (ii) Completing the square, $\sigma_Y^2 - 2\beta\sigma_{XY} + \beta^2\sigma_X^2 = \sigma_X^2(\beta - \beta^*)^2 + \sigma_Y^2 - \sigma_{XY}^2/\sigma_X^2$, and $\sigma_{XY}^2/\sigma_X^2 = \rho^2\sigma_Y^2$. (iii) is (8.1). □

Geometrically, $\beta^*(X - \mu_X)$ is the orthogonal projection of $Y - \mathbb{E}Y$ onto $X - \mu_X$ in L^2, and the control removes all of Y's variance except the part orthogonal to X. The factor $1 - \rho^2$ is unforgiving: $\rho = 0.9$ gives a variance factor 0.19, $\rho = 0.99$ gives 0.02. With $\beta = 1$ and $\sigma_X = \sigma_Y$ the variance factor is $2(1 - \rho)$, the course's formula; Exercise 8.6 asks when a fixed coefficient helps at all.

Example 8.13 (The square-root call)**.** Take $Y = e^{-rT}(\sqrt{S_T} - \sqrt{K})^+$. Since $\sqrt{S_T}$ is lognormal with volatility $\sigma/2$, the "pre-wash" map of Theorem 4.4 prices it as a Black–Scholes call on $\sqrt{S}$ with strike $\sqrt{K}$, volatility $\sigma/2$ and yield $\frac{1}{2}(r + y + \sigma^2/4)$. The course chooses devious inputs, $S = K = 10^4$, $T = 1$, $r = 1\%$, $y = -1\%$, $\sigma = 40\%$, which map exactly to the test case: the price is 7.36429. The control is the tangent of the payoff at the kink, $X = e^{-rT}(S_T - K)^+/(2\sqrt{K})$, that is $\frac{1}{200}$ of a vanilla call, with known mean $\mu_X = 8.43615$, about 15% above the price (Figure 8.2a). By quadrature, $\rho = 0.9960$, $\sigma_X^2 = 235.75$, $\sigma_Y^2 = 158.01$ and $\beta^* = 0.8154$. With the course's $\beta = 1$ the standard-error ratio is 0.2424, the course's "errors reduced by about 75%"; with β^* it is $\sqrt{1-\rho^2} = 0.0889$, a further factor 7.4 in efficiency. The course inferred $\rho \approx 0.97$ by reading the $\beta = 1$ result through $2(1-\rho) = 0.25^2$, which assumes $\sigma_X = \sigma_Y$; here $\sigma_X/\sigma_Y = 1.22$, and the true correlation is much higher. (The course writes the resulting efficiency as $2/0.25^2 = 8$; the formula is $1/(2 \cdot 0.25^2) = 8$.)

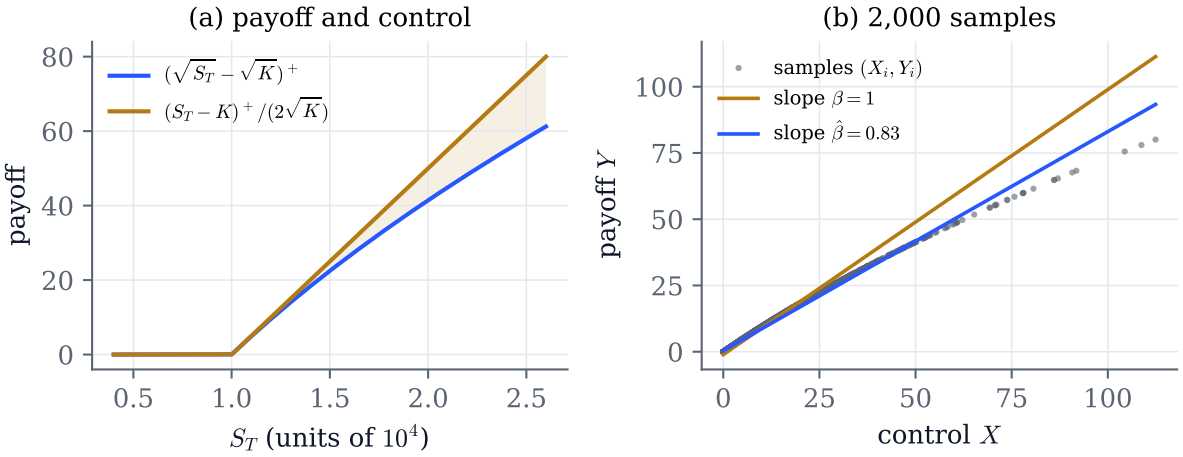

Figure 8.2: The square-root call with the course's devious inputs. (a) The payoff $(\sqrt{S_T} - \sqrt{K})^+$ and its control, the tangent call $(S_T - K)^+/(2\sqrt{K})$. (b) Two thousand discounted samples (X_i, Y_i) and the lines through the exact means (μ_X, μ_Y) with slope $\beta = 1$ and with the least-squares slope $\hat{\beta}$. Here Y is a function of X, so the points lie on a curve; what the control leaves is the deviation of that curve from a straight line.

8.5.1 Estimating the coefficient

β^* involves $\mathrm{Cov}(X, Y)$, which is unknown (if it were known, we would hardly need simulation). The natural estimator is the least-squares slope from the same sample,

$$\hat{\beta}_n = \frac{\sum_i (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)}{\sum_i (X_i - \bar{X}_n)^2}, \qquad \hat{\mu}_n^{\mathrm{cv}} = \bar{Y}_n - \hat{\beta}_n(\bar{X}_n - \mu_X),$$

when the denominator is positive; when it vanishes (all X_i equal) we set $\hat{\beta}_n = 0$. This is not an idle case: the tangent call of Theorem 8.13 vanishes with probability $N(0.15) = 0.56$, so all n controls are 0 with probability 0.56^n.

Proposition 8.14 (Estimated coefficient: bias and splitting)**.** *Let (X_i, Y_i) be i.i.d. copies of (X, Y) as in Theorem 8.12.*

(i) $\hat\beta_n \to \beta^*$ *almost surely, and* $\sqrt{n}\,(\hat\mu_n^{\mathrm{cv}} - \mathbb{E}Y) \Rightarrow \mathcal{N}(0, \sigma_Y^2(1-\rho^2))$*: estimating the coefficient costs nothing asymptotically.*

(ii) If $\mathbb{E}\hat\beta_n^2 < \infty$*, the bias is* $\mathbb{E}\hat\mu_n^{\mathrm{cv}} - \mathbb{E}Y = -\operatorname{Cov}(\hat\beta_n, \bar X_n)$*, and* $|\mathbb{E}\hat\mu_n^{\mathrm{cv}} - \mathbb{E}Y| \le \sigma_X \operatorname{sd}(\hat\beta_n)/\sqrt{n}$*; if moreover* $\operatorname{sd}(\hat\beta_n) = O(n^{-1/2})$*, the bias is* $O(1/n)$*.*

(iii) If X has a continuous law and $\mathbb{E}[Y \mid X] = a + bX$ is affine, then $\mathbb{E}[\hat\mu_n^{\mathrm{cv}} \mid X_1, \dots, X_n] = \mathbb{E}Y$; so $\hat\mu_n^{\mathrm{cv}}$ is unbiased whenever it is integrable.

(iv) (Splitting.) *Let $\hat\beta^{(1)}$ and $\hat\beta^{(2)}$ be the slopes computed on the two halves of the sample, and assume they are integrable. The estimator that applies each half's slope to the other half,*

$$\hat\mu_n^{\mathrm{split}} = \tfrac{1}{2}\Big[\bar Y^{(1)} - \hat\beta^{(2)}(\bar X^{(1)} - \mu_X)\Big] + \tfrac{1}{2}\Big[\bar Y^{(2)} - \hat\beta^{(1)}(\bar X^{(2)} - \mu_X)\Big],$$

is unbiased and has the same limit law as in (i).

Proof. (i) By the strong law, $\frac{1}{n}\sum(X_i - \bar X_n)(Y_i - \bar Y_n) \to \sigma_{XY}$ and $\frac{1}{n}\sum(X_i - \bar X_n)^2 \to \sigma_X^2$ almost surely, so $\hat\beta_n \to \beta^*$. Write

$$\sqrt{n}(\hat\mu_n^{\mathrm{cv}} - \mathbb{E}Y) = \sqrt{n}(\hat\mu_n(\beta^*) - \mathbb{E}Y) - (\hat\beta_n - \beta^*)\sqrt{n}(\bar X_n - \mu_X).$$

The first term converges in law to $\mathcal{N}(0, \sigma_Y^2(1-\rho^2))$ by the central limit theorem and Theorem 8.12; in the second, $\sqrt{n}(\bar X_n - \mu_X)$ converges in law and $\hat\beta_n - \beta^* \to 0$, so the product tends to 0 in probability and Slutsky's lemma (van der Vaart, 1998, Lemma 2.8) concludes.

(ii) $\mathbb{E}\hat\mu_n^{\mathrm{cv}} = \mathbb{E}Y - \mathbb{E}[\hat\beta_n(\bar X_n - \mu_X)]$ and $\mathbb{E}[\hat\beta_n(\bar X_n - \mu_X)] = \operatorname{Cov}(\hat\beta_n, \bar X_n)$ because $\mathbb{E}\bar X_n = \mu_X$. Cauchy–Schwarz bounds the covariance by $\operatorname{sd}(\hat\beta_n)\,\sigma_X/\sqrt{n}$. Both hypotheses can fail for small n: in the bivariate normal case $\operatorname{Var}\hat\beta_n = \sigma_Y^2(1-\rho^2)/(\sigma_X^2(n-3))$ for $n \ge 4$, but $\mathbb{E}\hat\beta_3^2 = \infty$ when $|\rho| < 1$ (Exercise 8.7).

(iii) Write $Y_i = a + bX_i + \varepsilon_i$ with $\mathbb{E}[\varepsilon_i \mid X_i] = 0$; by independence across i, $\mathbb{E}[\varepsilon_i \mid X_1, \dots, X_n] = 0$. Almost surely the X_i are distinct, and $\hat\beta_n = b + \sum_i (X_i - \bar X_n)\varepsilon_i / \sum_i (X_i - \bar X_n)^2$ and

$$\hat\mu_n^{\mathrm{cv}} = a + b\mu_X + \bar\varepsilon_n - (\bar X_n - \mu_X)\frac{\sum_i (X_i - \bar X_n)\varepsilon_i}{\sum_i (X_i - \bar X_n)^2},$$

which is $\mathbb{E}Y = a + b\mu_X$ plus a linear combination of the ε_i with coefficients that are functions of $X_1, \dots, X_n$; its conditional mean is $\mathbb{E}Y$.

(iv) Given the second half, $\hat\beta^{(2)}$ is a constant and the first half is still an i.i.d. sample independent of it, so the first bracket has conditional mean $\mathbb{E}Y$ (for any fixed β, $\hat\mu(\beta)$ is unbiased); it is integrable since $\mathbb{E}|\hat\beta^{(2)}(\bar X^{(1)} - \mu_X)| = \mathbb{E}|\hat\beta^{(2)}|\,\mathbb{E}|\bar X^{(1)} - \mu_X|$. The same holds for the second bracket. The limit law follows from (i) applied to each half, since each half's slope converges to β^*. □

A first-order expansion of the ratio $\hat\beta_n$ suggests the size of the bias in (ii): heuristically, $n(\mathbb{E}\hat\mu_n^{\mathrm{cv}} - \mathbb{E}Y) \to -\mathbb{E}[(X - \mu_X)^2\varepsilon]/\sigma_X^2$ with $\varepsilon = Y - \mathbb{E}Y - \beta^*(X - \mu_X)$, which vanishes in case (iii); Nelson (1990) discusses this bias and its remedies. For the call with the stock as control the constant is -8.18 (Figure 8.3). The bias is a fraction $O(n^{-1/2})$ of the standard error and is negligible for large samples, but it matters for small batches (for instance in the regressions inside American

option pricing, Chapter 14). The standard error of $\hat{\mu}_n^{\mathrm{cv}}$ is estimated by the residual standard deviation with $n-2$ degrees of freedom. A pilot sample, used only to estimate β and then discarded, removes the bias like splitting does. The exact small-sample variance in the normal case is computed in Exercise 8.7.

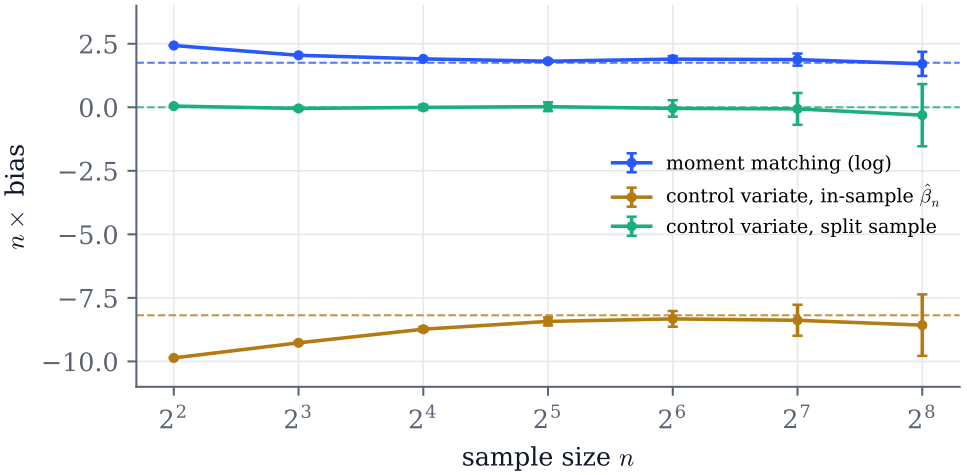

Figure 8.3: n times the bias of three estimators of the test-case call, from 6.4 million normal draws at each sample size n (bars: two standard errors). Moment matching (Theorem 8.7) and the control variate with in-sample coefficient (Theorem 8.14) have bias of order $1/n$, with the predicted constants 1.75 and -8.18 (dashed); the split-sample estimator is unbiased.

8.5.2 Several controls

With several controls $X = (X^1, \dots, X^q)$ of known mean $\mu_X \in \mathbb{R}^q$, use $\hat{\mu}_n(\beta) = \frac{1}{n}\sum_i (Y_i - \beta^\top (X_i - \mu_X))$ with $\beta \in \mathbb{R}^q$.

Theorem 8.15 (Multiple controls as regression)**.** *Let $Y \in L^2$ with $\sigma_Y^2 = \operatorname{Var} Y > 0$, $X \in L^2(\mathbb{R}^q)$, with $\Sigma_{XX} = \operatorname{Cov}(X)$ positive definite and $\Sigma_{XY} = \operatorname{Cov}(X, Y) \in \mathbb{R}^q$. Set $\beta^* = \Sigma_{XX}^{-1}\Sigma_{XY}$ and $R^2 = \Sigma_{XY}^\top \Sigma_{XX}^{-1}\Sigma_{XY}/\sigma_Y^2$.*

(i) For every $\beta \in \mathbb{R}^q$,

$$n \operatorname{Var} \hat{\mu}_n(\beta) = \sigma_Y^2(1 - R^2) + (\beta - \beta^*)^\top \Sigma_{XX}(\beta - \beta^*),$$

so β^* is the unique optimum, with variance factor $1 - R^2$, and $0 \le R^2 \le 1$.

(ii) $R^2 = \max_{a \in \mathbb{R}^q \setminus \{0\}} \operatorname{Corr}(Y, a^\top X)^2$. If $\Sigma_{XY} \neq 0$ the maximum is attained exactly at the nonzero multiples of β^*; if $\Sigma_{XY} = 0$, every a gives $0 = R^2$. In particular $R^2 \ge \operatorname{Corr}(Y, X^k)^2$ for each k, and adding a control never increases the optimal variance.

(iii) If $S_{XX} = \sum_i (X_i - \bar{X}_n)(X_i - \bar{X}_n)^\top$ is invertible and $\hat{\beta}_n = S_{XX}^{-1} \sum_i (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)$, then $\bar{Y}_n - \hat{\beta}_n^\top (\bar{X}_n - \mu_X)$ is the intercept $\hat{a}$ of the least-squares fit of Y_i on $(1, X_i - \mu_X)$, that is, the fitted regression evaluated at $X = \mu_X$. As $n \to \infty$, $\sqrt{n}(\hat{a} - \mathbb{E}Y) \Rightarrow \mathcal{N}(0, \sigma_Y^2(1 - R^2))$.

Proof. (i) $\operatorname{Var}(Y-\beta^\top X)=\sigma_Y^2-2\beta^\top\Sigma_{XY}+\beta^\top\Sigma_{XX}\beta$, and completing the square with $\Sigma_{XY}=\Sigma_{XX}\beta^*$ gives the formula, since $\beta^{*\top}\Sigma_{XX}\beta^*=\sigma_Y^2R^2$. The quadratic form is positive definite, so the optimum is unique; at β^* the variance $\sigma_Y^2(1-R^2)$ is nonnegative, and $R^2\ge 0$ since Σ_{XX}^{-1} is positive definite.

(ii) In the inner product $\langle u,v\rangle_\Sigma=u^\top\Sigma_{XX}v$, $a^\top\Sigma_{XY}=\langle a,\beta^*\rangle_\Sigma$, so by Cauchy–Schwarz

$$\operatorname{Corr}(Y,a^\top X)^2=\frac{(a^\top\Sigma_{XY})^2}{\sigma_Y^2\,a^\top\Sigma_{XX}a}\le\frac{\langle\beta^*,\beta^*\rangle_\Sigma}{\sigma_Y^2}=R^2,$$

with equality if and only if a and β^* are linearly dependent, that is, when $\beta^*\neq 0$ (equivalently $\Sigma_{XY}\neq 0$), if and only if a is a multiple of β^*; if $\Sigma_{XY}=0$ both sides vanish for every a. A larger set of controls maximises over a larger set of combinations.

(iii) For fixed b, $\sum_i(Y_i-a-b^\top(X_i-\mu_X))^2$ is minimised by $a=\bar Y_n-b^\top(\bar X_n-\mu_X)$; substituting leaves $\sum_i\left((Y_i-\bar Y_n)-b^\top(X_i-\bar X_n)\right)^2$, minimised by $b=\hat\beta_n$ (normal equations). The limit law follows as in Theorem 8.14(i), with the strong law for S_{XX}/n and the multivariate central limit theorem. □

In terms of correlations, $1-R^2=1-\rho_{YX}^\top\rho_{XX}^{-1}\rho_{XY}$, the form used in the course. Two cautions. First, each estimated coefficient costs a little: when (X,Y) is jointly normal, the variance of $\hat a$ is $\sigma_Y^2(1-R^2)/n$ times the loss factor $(n-2)/(n-q-2)$ (Lavenberg et al., 1982), so q must stay small compared with n. Second, nearly collinear controls make S_{XX} ill-conditioned and $\hat\beta_n$ erratic, even though the fitted intercept may remain accurate; drop redundant controls. For the square-root call, adding the stock to the tangent call raises R^2 from 0.9921 to 0.9934 and lowers the standard-error ratio from 0.0889 to 0.0815. The arithmetic Asian option with geometric, vanilla and strip controls (Theorem 15.4) is the course's showcase of multiple controls.

8.5.3 Nonlinear controls and hedges

Two further ways to use known means lead back to linear controls.

Proposition 8.16 (Nonlinear controls)**.** *(i) (*Delta method.*) Let h be defined and continuously differentiable on a neighbourhood U of $(\mu_X,\mathbb{E}Y)$, with $h(\mu_X,y)=y$ whenever $(\mu_X,y)\in U$, and set $h(\bar X_n,\bar Y_n)=\bar Y_n$ when $(\bar X_n,\bar Y_n)\notin U$. Then $h(\bar X_n,\bar Y_n)\to\mathbb{E}Y$ almost surely and*

$$\sqrt{n}\big(h(\bar X_n,\bar Y_n)-\mathbb{E}Y\big)\Rightarrow\mathcal{N}\Big(0,\operatorname{Var}\left(Y-\beta_h(X-\mu_X)\right)\Big),\qquad \beta_h=-\partial_x h(\mu_X,\mathbb{E}Y)\,:$$

a nonlinear control is asymptotically the linear control with coefficient β_h, and no better than β^*.

*(ii) (*Hedges as controls.*) Let $V(t,s)$ be the Black–Scholes price of a European claim with a Lipschitz payoff $F(S_T)$ (calls and puts qualify). For a grid $0=t_0<\cdots<t_m=T$ let $\tilde S_t=e^{-(r-y)t}S_t$ and*

$$M_m=\sum_{k=0}^{m-1}e^{-yt_k}\,\partial_s V(t_k,S_{t_k})\,(\tilde S_{t_{k+1}}-\tilde S_{t_k}).$$

Then $\mathbb{E}M_m=0$, and with $Y=e^{-rT}F(S_T)$, $\mathbb{E}[(Y-M_m-V(0,S))^2]\to 0$ as $\max_k(t_{k+1}-t_k)\to 0$.

Proof. (i) By the strong law, $(\bar{X}_n, \bar{Y}_n) \in U$ for all large n, almost surely, so the convention outside U does not affect the limits; continuity of h gives the almost sure limit. By the multivariate central limit theorem and the delta method (van der Vaart, 1998, Theorem 3.1), $\sqrt{n}(h(\bar{X}_n, \bar{Y}_n) - h(\mu_X, \mathbb{E}Y))$ converges in law to $\nabla h^\top N$ with $N \sim \mathcal{N}(0, \mathrm{Cov}(X, Y))$, that is, to the law of $\partial_x h\, X + \partial_y h\, Y$ centred. Differentiating $h(\mu_X, y) = y$ gives $\partial_y h = 1$, and $\partial_x h = -\beta_h$. The last claim is Theorem 8.12(ii).

(ii) Under $\mathbb{Q}$, $\tilde{S}_t = S\exp(\sigma W_t - \sigma^2 t/2)$ is a square-integrable martingale and the weights $e^{-yt_k}\partial_s V(t_k, S_{t_k})$ are bounded and $\mathcal{F}_{t_k}$-measurable, so every term of M_m has mean zero. Writing $V(t,s) = e^{-r(T-t)}\mathbb{E}F\big(se^{\nu(T-t)+\sigma\sqrt{T-t}Z}\big)$ and differentiating under the Gaussian integral, V is $C^{1,2}$ on $[0,T)\times(0,\infty)$ and solves the Black–Scholes equation; if L is the Lipschitz constant of F, $|\partial_s V| \le Le^{-y(T-t)}$, the delta is continuous, and $V(t,s) \to F(s)$ as $t \uparrow T$ by dominated convergence. Itô's formula gives, as in the delta-hedging argument of the IEOR 222 notes, $\mathrm{d}(e^{-rt}V(t,S_t)) = e^{-rt}\partial_s V\,\sigma S_t\,\mathrm{d}W_t = h_t\,\mathrm{d}\tilde{S}_t$ with $h_t = e^{-yt}\partial_s V(t, S_t)$. Letting $t \uparrow T$ (the integrand is bounded, and $V(t,S_t) \to F(S_T)$ since V is jointly continuous up to T by the same argument), $Y = V(0,S) + \int_0^T h_t\,\mathrm{d}\tilde{S}_t$. Hence $Y - V(0,S) - M_m = \int_0^T (h_t - h_{\eta(t)})\,\mathrm{d}\tilde{S}_t$ with $\eta(t) = t_k$ on $[t_k, t_{k+1})$, and by the Itô isometry

$$\mathbb{E}\big[(Y - M_m - V(0,S))^2\big] = \mathbb{E}\int_0^T (h_t - h_{\eta(t)})^2\sigma^2\tilde{S}_t^2\,\mathrm{d}t \longrightarrow 0$$

by dominated convergence: the integrand is bounded by $4\sup|h|^2\sigma^2\tilde{S}_t^2$, which is integrable, and tends to 0 for every $t < T$ along every path, by continuity of h. □

Part (i) explains the classical ratio estimator $\bar{Y}_n\mu_X/\bar{X}_n$ ($\mu_X \neq 0$, with U a neighbourhood not meeting $\{x = 0\}$), for which $\beta_h = \mathbb{E}Y/\mu_X$: for the call with the stock as control this is 0.075, far from $\beta^* = 0.560$, so the ratio estimator barely helps. Part (ii) says that the ideal control is the claim's own hedge: a control variate is a hedge whose cost is known. With the Black–Scholes delta of the call itself as weights, the standard-error ratio for the test-case call is about 0.48, 0.15 and 0.075 for $m = 1$, 12 and 52 rebalancing dates (20,000 paths), a variance falling asymptotically like $1/m$ as in Theorem 18.3. In practice the delta of a nearby tractable claim is used: the static tangent call of Theorem 8.13 is the one-date version, and the asymptotic square-root call of Exercise 8.8 uses the vanilla call with the coefficient $F'(K+) = 1$ read off the replication.

8.6 Combining the techniques: a laboratory

Table 8.1 runs every method of this chapter on the same $n = 2^{20}$ normal draws (seed 1) and sets the measured standard-error ratios beside the exact asymptotic ones, computed by quadrature from the formulas above. Moment matching uses 256 separately matched batches, and its measured ratios carry the noise of a batch standard error, about 4%.

Remark 8.17 (Why combinations are not multiplicative). Antithetic sampling removes the odd part f_o; a control applied afterwards regresses the even part of Y on the even part of X, and only the correlation of the even parts matters. For the call with the stock as control, the even parts have correlation 0.9645, against 0.882 for the raw payoffs, and the combination reaches 0.151, better than the product $0.573 \times 0.472 = 0.270$ of the separate ratios. The opposite also happens: the control Z itself, which helps alone, is useless after pairing, because its even part

target	method	exact	laboratory	efficiency
stock	antithetic	0.140	0.140	25.5
	moment matching (log)	0.016	0.017	1.9×10^3
	exponential matching; control S_T	0	0	exact
call	antithetic	0.573	0.574	1.5
	moment matching (log)	0.188	0.180	14
	moment matching (exponential)	0.182	0.174	15
	control S_T, $\hat{\beta}_n$	0.472	0.472	2.2
	antithetic and control S_T	0.151	0.152	22
square-root call	control, $\beta = 1$	0.242	0.242	8.5
	control, $\hat{\beta}_n$	0.089	0.089	63
	tangent call and stock	0.081	0.082	75

Table 8.1: Standard-error ratios to plain Monte Carlo at equal n, exact (asymptotic, by quadrature) and measured in the laboratory ($n = 2^{20}$, seed 1), for the test case and the square-root call, and the efficiency $1/(2\varrho^2)$ of (8.1) at the course's conservative cost $c = 2$. Antithetic ratios compare n pairs with n draws. The exact entries for moment matching use the linearisation of Theorem 8.8.

is zero; the variance it would remove is odd, and pairing has already removed it. Likewise, log matching already contains the optimal controls Z and Z^2 (Theorem 8.8), so little is left for a further control: exponential-mean matching, which also fixes the mean of the stock, improves the ratio only from 0.188 to 0.182. The lesson of the table is the course's: almost any combination helps, but the gains must be measured, not multiplied.

Course examples reproduced. The antithetic ratios 14% (stock) and 57% (call) with efficiencies 25 and 1.5; the moment-matching ratios 1.6% and 20% (efficiencies 1.9×10^3 and 14 here, about 2000 and 13 in the course); the square-root call 7.36429, its control mean 8.43615 and the 75% error reduction at $\beta = 1$; and the call-payoff variance 161.2.

▶ Interactive demo: moment matching with batches, the bias of order $1/n$, control variates with a slider for the coefficient, common random numbers for a volatility bump, and the laboratory of Table 8.1 with measured execution times.

Further reading

Glasserman (2003) is the standard reference: Sections 4.1 (control variates, with multiple, small-sample and nonlinear controls in 4.1.2–4.1.4), 4.2 (antithetic variates), 4.5.1 (moment matching) and 7.1 (common random numbers for Greeks). Antithetic variates go back to Hammersley and Morton (1956); the association inequality behind Theorem 8.1 is developed in Esary et al. (1967). Lavenberg and Welch (1981) survey control variates with estimated coefficients; the loss factor is due to Lavenberg et al. (1982), and Nelson (1990) treats the bias and its remedies. Glasserman and Yao (1992) study common random numbers. Boyle et al. (1997) review variance reduction for security pricing.

Chapter 9

Importance sampling

A Monte Carlo estimator spends its samples where the pricing measure puts its mass. For an at-the-money call that is roughly the right place. For a call struck three standard deviations out of the money, more than 99.8% of the paths pay nothing, and the price is decided by the few that do. *Importance sampling* draws the samples from another law, chosen to put them where the payoff lives, and removes the bias by weighting each sample with the likelihood ratio of the two laws. The same ratio drives the likelihood-ratio Greeks of Chapter 13; here it is used the other way round.

The method can cut the variance by orders of magnitude, and it can also make it infinite while the estimator still looks well behaved; every gain below therefore comes with its failure mode. The code is `code/src/mfe230d/importance.py`.

Throughout, the test case is $S = K = 100$, $T = 1$, $r = 1\%$, $y = 2\%$, $\sigma = 20\%$, with $\log(S_T/S) \sim \mathcal{N}(\nu T, \sigma^2 T)$ under $\mathbb{Q}$, $\nu = r - y - \sigma^2/2 = -0.03$, and call price 7.36429. In standard units $S_T = S\exp(\nu T + sZ)$ with $s = \sigma\sqrt{T}$ and $Z \sim \mathcal{N}(0,1)$, and the call pays exactly when $Z > d$, where

$$d = \frac{\log(K/S) - \nu T}{\sigma\sqrt{T}} = -d_-$$

is the distance from the centre of the law of Z to the strike, in standard deviations.

9.1 The likelihood ratio

Let X be a random vector in $\mathbb{R}^k$ with density q under $\mathbb{Q}$, and let $f : \mathbb{R}^k \to \mathbb{R}$ be Borel measurable with $\mathbb{E}^{\mathbb{Q}}|f(X)| < \infty$; we want $\mu = \mathbb{E}^{\mathbb{Q}}[f(X)]$. In pricing, X is the terminal log-price or the vector of normals that drives a simulated path, and f includes the discount factor. Let $\tilde{q}$ be another probability density on $\mathbb{R}^k$, the *sampling density*, and write $\widetilde{\mathbb{Q}}$ and $\widetilde{\mathbb{E}}$ for probabilities and expectations when X has density $\tilde{q}$. The *likelihood ratio*, or *weight*, is the Borel function

$$L(x) = \frac{q(x)}{\tilde{q}(x)} \ \text{ if } \tilde{q}(x) > 0, \qquad L(x) = 0 \ \text{ otherwise.}$$

Everything below holds verbatim for discrete laws, with densities with respect to counting measure (probability mass functions).

Proposition 9.1 (The importance-sampling estimator)**.** *Assume the* support condition

$$\tilde{q}(x) > 0 \quad \textit{for every } x \textit{ with } f(x)q(x) \neq 0. \tag{9.1}$$

(i) $f(X)L(X)$ *is* $\widetilde{\mathbb{Q}}$*-integrable and* $\widetilde{\mathbb{E}}[f(X)L(X)] = \mathbb{E}^{\mathbb{Q}}[f(X)] = \mu$.

(ii) If $X_1, X_2, \ldots$ *are independent with density* $\tilde{q}$*, the* importance-sampling estimator $\hat{\mu}_n^{\text{is}} = \frac{1}{n}\sum_{i=1}^n f(X_i)L(X_i)$ *is unbiased and converges to* μ *almost surely.*

(iii) Its variance is σ_{is}^2/n*, where*

$$\sigma_{\text{is}}^2 + \mu^2 = \widetilde{\mathbb{E}}[f(X)^2L(X)^2] = \mathbb{E}^{\mathbb{Q}}[f(X)^2L(X)] = \int_{\mathbb{R}^k} \frac{f(x)^2q(x)^2}{\tilde{q}(x)}\,\mathrm{d}x \;\in [0,\infty], \tag{9.2}$$

with the integral taken over $\{\tilde{q} > 0\}$.

(iv) Without (9.1)*,* $\widetilde{\mathbb{E}}[f(X)L(X)] = \mathbb{E}^{\mathbb{Q}}[f(X)\mathbf{1}_{\{\tilde{q}(X)>0\}}]$*: the estimator converges to the wrong number, and nothing in the samples reveals it.*

Proof. For any Borel $g \geq 0$, or g with $\mathbb{E}^{\mathbb{Q}}|g(X)| < \infty$,

$$\widetilde{\mathbb{E}}[g(X)L(X)] = \int_{\{\tilde{q}>0\}} g(x)\frac{q(x)}{\tilde{q}(x)}\,\tilde{q}(x)\,\mathrm{d}x = \int_{\{\tilde{q}>0\}} g(x)q(x)\,\mathrm{d}x = \mathbb{E}^{\mathbb{Q}}[g(X)\mathbf{1}_{\{\tilde{q}(X)>0\}}].$$

With $g = |f|$ this shows integrability, and with $g = f$ it gives (iv). Under (9.1), $f(x)q(x) = 0$ on $\{\tilde{q} = 0\}$, so the indicator can be dropped: this is (i). (ii) follows from (i) and the strong law of large numbers for the i.i.d. integrable variables $f(X_i)L(X_i)$. For (iii), apply the display to $g = f^2L \geq 0$ (Tonelli): $\widetilde{\mathbb{E}}[f^2L^2] = \mathbb{E}^{\mathbb{Q}}[f^2L\,\mathbf{1}_{\{\tilde{q}>0\}}] = \mathbb{E}^{\mathbb{Q}}[f^2L]$, since $L = 0$ off $\{\tilde{q} > 0\}$; both sides may be infinite. Subtract μ^2. □

If $0 < \sigma_{\text{is}}^2 < \infty$, Theorem 6.4 gives confidence intervals with the sample standard deviation $\hat{s}_n$ of the weighted samples; if $\sigma_{\text{is}}^2 = 0$, every sample equals μ; if $\sigma_{\text{is}}^2 = \infty$, the estimator still converges, but $\hat{s}_n \to \infty$ almost surely and the intervals mean nothing (Theorem 6.7). The weight is always integrable, $\widetilde{\mathbb{E}}[L] = \mathbb{Q}(\tilde{q}(X) > 0) \leq 1$ ($g = 1$ in the proof), but its variance can be infinite.

The middle expression in (9.2) is the key to everything that follows: the second moment of the payoff under the target law, weighted by L. Where the sampling density is much smaller than the target density, L is large and the payoff is punished; where it is larger, $L < 1$ and the payoff is discounted. A good sampling density is large where $|f|q$ is large. The extra cost of a sample is one evaluation of L, usually one exponential, so the efficiency gain over plain Monte Carlo (Theorem 6.9) is close to the variance ratio $\sigma^2/\sigma_{\text{is}}^2$. Under $\widetilde{\mathbb{Q}}$, discounted prices are not martingales: the sampling law is a computational device, not a pricing measure.

Proposition 9.2 (The zero-variance law)**.** *Assume* $0 < \mathbb{E}^{\mathbb{Q}}|f(X)| < \infty$*. For every sampling density satisfying* (9.1)*,*

$$\sigma_{\text{is}}^2 \;\geq\; \left(\mathbb{E}^{\mathbb{Q}}|f(X)|\right)^2 - \mu^2,$$

with equality if and only if $\tilde{q} = |f|q/\mathbb{E}^{\mathbb{Q}}|f(X)|$ *almost everywhere. In particular, if* $f \geq 0$*, the density* $\tilde{q}^* = fq/\mu$ *gives* $f(X)L(X) = \mu$ *for* $\widetilde{\mathbb{Q}}$*-almost every sample, and it is the only sampling density with zero variance.*

Proof. By Cauchy–Schwarz and Theorem 9.1(i) applied to $|f|$, $\widetilde{\mathbb{E}}[(fL)^2] \geq \big(\widetilde{\mathbb{E}}|fL|\big)^2 = \big(\mathbb{E}^{\mathbb{Q}}|f|\big)^2$, which is the inequality. Equality holds if and only if $|f(X)|L(X)$ is $\widetilde{\mathbb{Q}}$-almost surely equal to a constant c, and then $c = \widetilde{\mathbb{E}}|fL| = \mathbb{E}^{\mathbb{Q}}|f| > 0$. On $\{\tilde{q} > 0,\ fq = 0\}$ we would have $|f|L = 0 \neq c$, so this set has $\widetilde{\mathbb{Q}}$-probability zero; on $\{fq \neq 0\}$, which lies in $\{\tilde{q} > 0\}$ by (9.1), $|f|q/\tilde{q} = c$, that is $\tilde{q} = |f|q/c$. Hence $\tilde{q} = |f|q/c$ almost everywhere; conversely this density gives $|f|L = c$. For $f \geq 0$, $\mathbb{E}^{\mathbb{Q}}|f| = \mu$ and the bound is 0. □

The zero-variance density needs μ, the number we are computing, so it is never used as such. What it gives is the shape to aim for: *sample in proportion to payoff times density.* For a call it is the risk-neutral density of S_T on $\{S_T > K\}$, multiplied by $S_T - K$ (Figure 9.1). A payoff that changes sign cannot be estimated with zero variance by one sampling law; writing $f = f^+ - f^-$ and estimating the two parts with two laws removes the obstruction.

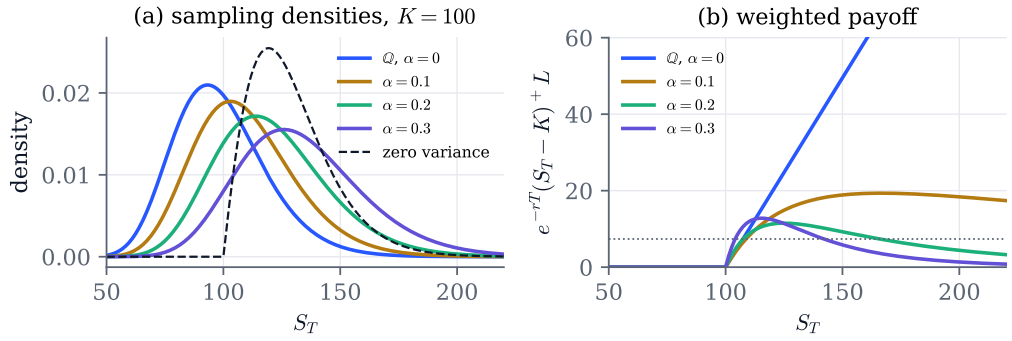

Figure 9.1: The course's experiment at $K = 100$. (a) The risk-neutral density of S_T and the sampling densities with drift shifts $\alpha = 0.1, 0.2, 0.3$ ($\beta = 1$); dashed, the zero-variance density of Theorem 9.2. (b) The weighted payoff $e^{-rT}(S_T - K)^+ L$, whose $\widetilde{\mathbb{Q}}$-mean is the price 7.36 (dotted) in every case: the shift turns the unbounded payoff into a bounded hump, much flatter under the sampling law.

9.2 Shifting and scaling the Gaussian

In lecture the target is the law of the terminal log-return $x_T = \log(S_T/S)$, which is $\mathcal{N}(\nu T, \sigma^2 T)$ under $\mathbb{Q}$, and the sampling law is $\mathcal{N}((\nu + \alpha)T, \beta^2\sigma^2 T)$: α shifts the drift of the log-price and β rescales its volatility. For $\beta = 1$ the weight is

$$L(x_T) = \exp\Big(\frac{\alpha\,[(\alpha + 2\nu)T - 2x_T]}{2\sigma^2}\Big), \tag{9.3}$$

as one checks by expanding the two Gaussian exponents. In standard units the target is $\mathcal{N}(m, s^2)$ with $m = \nu T$, $s = \sigma\sqrt{T}$, and the sampling law is $\mathcal{N}(m + a, \beta^2 s^2)$ with $a = \alpha T$, a shift of $c = a/s$ standard deviations.

Proposition 9.3 (Mean shift and scale)**.** *Let the target law be $\mathcal{N}(m, s^2)$ and the sampling law $\mathcal{N}(m + a, \beta^2 s^2)$, with $s > 0$, $\beta > 0$, $a \in \mathbb{R}$, and $c = a/s$.*

(i) The weight has second moment

$$\widetilde{\mathbb{E}}[L^2] = \mathbb{E}^{\mathbb{Q}}[L] = \frac{\beta^2}{\sqrt{2\beta^2-1}} \exp\Big(\frac{c^2}{2\beta^2-1}\Big) \quad \text{if } \beta > \tfrac{1}{\sqrt{2}}, \qquad \widetilde{\mathbb{E}}[L^2] = \infty \quad \text{if } \beta \le \tfrac{1}{\sqrt{2}}.$$

For $\beta = 1$ this is e^{c^2}.

(ii) (Mirror formula.) If $\beta > 1/\sqrt{2}$, then for every Borel f,

$$\widetilde{\mathbb{E}}[f(X)^2 L(X)^2] = \widetilde{\mathbb{E}}[L^2]\, \mathbb{E}[f(\tilde{X})^2], \qquad \tilde{X} \sim \mathcal{N}\Big(m - \frac{a}{2\beta^2-1},\ \frac{\beta^2 s^2}{2\beta^2-1}\Big). \tag{9.4}$$

(iii) The importance-sampling estimator of $\mathbb{E}^{\mathbb{Q}}[f(X)]$ has finite variance for every f of at most exponential growth ($|f(x)| \le Ce^{\kappa|x|}$ for some constants C, κ) if and only if $\beta > 1/\sqrt{2}$.

Proof. Write $u = x - m$ and $\gamma = 2\beta^2 - 1$. The target density is $(2\pi s^2)^{-1/2} e^{-u^2/(2s^2)}$ and the sampling density $(2\pi\beta^2 s^2)^{-1/2} e^{-(u-a)^2/(2\beta^2 s^2)}$, so

$$\frac{q(x)^2}{\tilde{q}(x)} = \frac{\beta}{s\sqrt{2\pi}} \exp\Big(-\frac{u^2}{s^2} + \frac{(u-a)^2}{2\beta^2 s^2}\Big) = \frac{\beta}{s\sqrt{2\pi}} \exp\Big(-\frac{\gamma u^2 + 2au - a^2}{2\beta^2 s^2}\Big).$$

If $\gamma > 0$, complete the square, $\gamma u^2 + 2au = \gamma(u + a/\gamma)^2 - a^2/\gamma$, and use $1 + 1/\gamma = 2\beta^2/\gamma$:

$$\frac{q(x)^2}{\tilde{q}(x)} = \frac{\beta}{s\sqrt{2\pi}}\, e^{c^2/\gamma} \exp\Big(-\frac{(x - m + a/\gamma)^2}{2\beta^2 s^2/\gamma}\Big) = \frac{\beta^2}{\sqrt{\gamma}}\, e^{c^2/\gamma}\, \tilde{p}(x),$$

where $\tilde{p}$ is the density of $\tilde{X}$. By (9.2), $\widetilde{\mathbb{E}}[f^2 L^2] = \int f^2 q^2/\tilde{q}\,\mathrm{d}x$; taking $f \equiv 1$ gives (i) and a general f gives (ii). If $\gamma < 0$ the exponent tends to $+\infty$ quadratically as $|u| \to \infty$, and if $\gamma = 0$ it is affine in u, $(a^2 - 2au)/(2\beta^2 s^2)$, whose exponential is not integrable over $\mathbb{R}$; in both cases $\widetilde{\mathbb{E}}[L^2] = \infty$. (iii) If $\beta > 1/\sqrt{2}$, then $\mathbb{E}[f(\tilde{X})^2] \le C^2 \mathbb{E}[e^{2\kappa|\tilde{X}|}] < \infty$ for a Gaussian $\tilde{X}$, so (ii) gives a finite variance. If $\beta \le 1/\sqrt{2}$, the payoff $f \equiv 1$, of exponential growth with $\kappa = 0$, has infinite variance by (i). □

The mirror formula has a simple reading for $\beta = 1$: $\widetilde{\mathbb{E}}[(fL)^2] = e^{c^2}\, \mathbb{E}^{\mathbb{Q}}[f(X-a)^2]$. Shifting the sampling law by a towards the payoff costs the factor e^{c^2}, and in exchange the second moment is that of the payoff evaluated at the target shifted by a *the other way*, where a call or a digital is much smaller. The best shift balances the two. For the call, $\mathbb{E}[((Se^Y - K)^+)^2]$ with $Y \sim \mathcal{N}(\mu_Y, v)$ equals

$$S^2 e^{2\mu_Y + 2v} N(\delta_2) - 2KSe^{\mu_Y + v/2} N(\delta_1) + K^2 N(\delta_0), \qquad \delta_j = \frac{\mu_Y + jv - \log(K/S)}{\sqrt{v}},$$

so the variance of the importance-sampling estimator is an explicit function of (α, β), although its minimiser is not. Figure 9.2 reproduces the variance-ratio contours of the course. With $\beta = 1$, the optimal shift and the variance ratio are $\alpha^* = 0.206$ and 9.03 at $K = 90$, $\alpha^* = 0.258$ and 10.58 at $K = 100$, and $\alpha^* = 0.314$ and 14.29 at $K = 110$: an order of magnitude even at the money, and more the further out of the money the strike. With β free at $K = 100$, the ratio keeps increasing as β decreases, to 30.9 at $\beta = 0.75$ and 34.6 as $\beta \downarrow 1/\sqrt{2}$ (with $\alpha^* \approx 0.270$): the best scale sits on the edge of the region where the weight has finite variance.

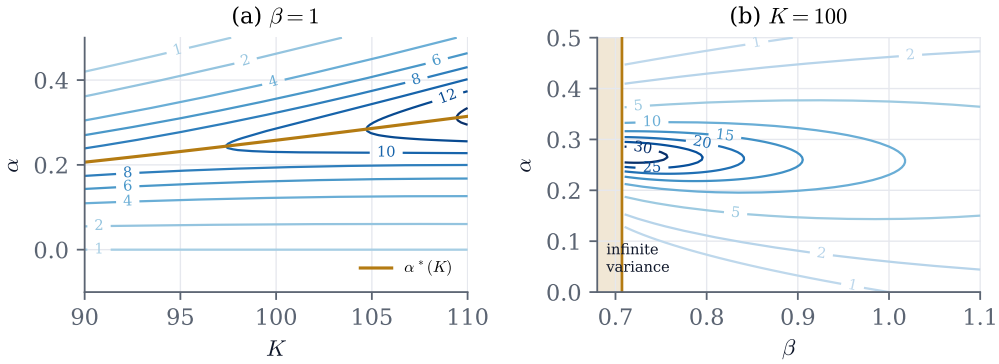

Figure 9.2: Variance of plain Monte Carlo over variance of importance sampling for the test-case call, computed exactly from Theorem 9.3. (a) Against the strike and the drift shift, $\beta = 1$, with the optimal shift $\alpha^*(K)$. (b) Against the scale and the shift at $K = 100$; to the left of $\beta = 1/\sqrt{2}$ the variance is infinite.

Remark 9.4 (The share measure and the limits of α^*). With $\alpha = \sigma^2$ and $\beta = 1$, (9.3) gives $S_T L = S \exp\left(x_T + (\sigma^2/2 + \nu)T - x_T\right) = Se^{(r-y)T}$ for every sample: the forward is estimated with zero variance. The sampling law, under which $\log S_T$ has drift $r - y + \sigma^2/2$, is the share measure of Theorem 17.3, and its density with respect to $\mathbb{Q}$ is $S_T/\mathbb{E}^{\mathbb{Q}}[S_T]$, the zero-variance density of Theorem 9.2 for $f(x) = Se^x$. Numerically $\alpha^* \to \sigma^2$ as $K \to 0$ ($\alpha^* = 0.0501$ at $K = 20$, 0.0404 at $K = 1$), and as $K \to \infty$, heuristically, $c^* = d + \frac{3}{2d} + o(1/d)$ (checked numerically in our code), the leading terms of the lecture's expansion.

Remark 9.5 (Thin tails are dangerous, fat tails are safe). For every $\beta < 1$ the weight $L(x)$ grows like $\exp(\kappa x^2)$, $\kappa > 0$, in both tails, while $L\tilde{q} = q$ always decays. What fails for $\beta < 1/\sqrt{2}$ is the integrand of the second moment, $L^2\tilde{q} = q^2/\tilde{q}$, which then grows like $\exp(\kappa' x^2)$. At $\beta = 1/\sqrt{2}$ exactly, the weight has infinite variance, but the call does not: $q^2/\tilde{q}$ is then the exponential of an affine function of x, and $\widetilde{\mathbb{E}}[(fL)^2] < \infty$ if and only if $\alpha > \sigma^2$ (compare the growth e^{2x} of f^2 with the decay of $q^2/\tilde{q}$). For $\beta > 1$, on the other hand, the exponent $-u^2/(2s^2) + (u-a)^2/(2\beta^2 s^2)$ of L (with $u = x - m$) is a concave quadratic in u, so the weight is *bounded*, and $\sigma_{\text{is}}^2 + \mu^2 = \mathbb{E}^{\mathbb{Q}}[f^2 L] \le \sup L\, \mathbb{E}^{\mathbb{Q}}[f^2]$: whatever is finite under $\mathbb{Q}$ stays finite. (For $\beta = 1$ the exponent is affine and L is bounded on a half-line only.) The same computation with $q^4/\tilde{q}^3$ shows that the fourth moment of the call estimator is infinite whenever $\beta^2 < 3/4$ (and, for the n-th moment of the weight, $\beta^2 > 1 - 1/n$; Exercise 9.1). For $1/\sqrt{2} < \beta < \sqrt{3}/2$, then, $\hat{s}_n$ converges, but its own fluctuations have infinite variance, and the error bars are least trustworthy exactly where the course's optimum lies. Figure 9.5(b) shows the running $\hat{s}_n$ for $\beta = 1$, 0.75 and 0.55: the first two settle at their exact values 3.88 and 2.27 (against 12.57 for plain Monte Carlo), the third never does.

▶ Interactive demo: move α, β and K; watch an infinite variance.

9.3 Exponential tilting and large deviations

The mean shift is a special case of a general recipe. Let X be a real random variable whose cumulant generating function $\psi(\theta) = \log \mathbb{E}^{\mathbb{Q}}[e^{\theta X}]$ is finite on an open interval $\Theta \ni 0$.

Definition 9.6 (Exponential tilting)**.** For $\theta \in \Theta$, the *exponentially tilted* (or *Esscher*) law of X has density $q_\theta(x) = e^{\theta x - \psi(\theta)} q(x)$; its weight is $L = e^{-\theta X + \psi(\theta)}$. For a sum $S_n = X_1 + \cdots + X_n$ of independent copies, each tilted, the weight is $e^{-\theta S_n + n\psi(\theta)}$.

On Θ, ψ is infinitely differentiable, with $\psi'(\theta) = \mathbb{E}_\theta[X]$ and $\psi''(\theta) = \mathrm{Var}_\theta(X)$, the mean and variance under the tilted law: for θ in Θ and $\varepsilon > 0$ with $\theta \pm \varepsilon \in \Theta$, $|x|^j e^{\theta x} \le C_j(e^{(\theta+\varepsilon)x} + e^{(\theta-\varepsilon)x})$, so dominated convergence allows differentiation under $\mathbb{E}^{\mathbb{Q}}$. If X is not almost surely constant, $\psi'' > 0$ and ψ' is strictly increasing: tilting with $\theta > 0$ raises the mean. Examples: for $\mathcal{N}(m, s^2)$, $\psi(\theta) = m\theta + s^2\theta^2/2$ and the tilted law is $\mathcal{N}(m + \theta s^2, s^2)$, the mean shift of Section 9.2 with $a = \theta s^2$; for a Poisson count with mean λ, $\psi(\theta) = \lambda(e^\theta - 1)$ and the tilted law is Poisson with mean λe^θ. Other families, and a law that cannot be tilted, are in Exercise 9.2.

Proposition 9.7 (The saddle-point tilt for sums)**.** *Let $X_1, X_2, \ldots$ be i.i.d. copies of a non-constant X with ψ finite on an open interval $\Theta \ni 0$. Let $a > \mathbb{E}^{\mathbb{Q}}[X]$, and suppose that $\psi'(\theta_a) = a$ for some $\theta_a \in \Theta$. Set $I(a) = a\theta_a - \psi(\theta_a)$ and $p_n = \mathbb{Q}(S_n \ge na)$. Then $\theta_a > 0$ and $I(a) > 0$, and:*

(i) (Chernoff bound.) $p_n \le e^{-nI(a)}$.

(ii) The estimator $Y = \mathbf{1}_{\{S_n \ge na\}} e^{-\theta_a S_n + n\psi(\theta_a)}$, with S_n sampled under the tilted law, satisfies $0 \le Y \le e^{-nI(a)}$ surely, and $\widetilde{\mathbb{E}}[Y^2] \le e^{-nI(a)} p_n \le e^{-2nI(a)}$.

(iii) $\frac{1}{n}\log p_n \to -I(a)$, and $\log \widetilde{\mathbb{E}}[Y^2] / \log(p_n^2) \to 1$, while plain Monte Carlo has relative variance $(1 - p_n)/p_n = e^{nI(a) + o(n)}$.

Proof. ψ' is strictly increasing and $\psi'(\theta_a) = a > \mathbb{E}^{\mathbb{Q}}[X] = \psi'(0)$, so $\theta_a > 0$. By strict convexity, $0 = \psi(0) > \psi(\theta_a) - \theta_a\psi'(\theta_a) = -I(a)$. (ii) On $\{S_n \ge na\}$, $\theta_a > 0$ gives $e^{-\theta_a S_n} \le e^{-n\theta_a a}$, so $Y \le e^{-n(a\theta_a - \psi(\theta_a))} = e^{-nI(a)}$; hence $Y^2 \le e^{-nI(a)} Y$ and $\widetilde{\mathbb{E}}[Y^2] \le e^{-nI(a)} \widetilde{\mathbb{E}}[Y] = e^{-nI(a)} p_n$, by Theorem 9.1(i). (i) $p_n = \widetilde{\mathbb{E}}[Y] \le e^{-nI(a)}$. (iii) The function $\theta \mapsto a\theta - \psi(\theta)$ is concave with a critical point at θ_a, so $I(a) = \sup_\theta(a\theta - \psi(\theta))$, the Legendre transform of ψ at a. By Cramér's theorem (Dembo and Zeitouni, 1998, Theorem 2.2.3), $\liminf_n \frac{1}{n}\log \mathbb{Q}(S_n/n > a) \ge -\inf_{x>a} I(x)$. The function I is convex, $I \ge 0$ (take $\theta = 0$) and $I(\mathbb{E}^{\mathbb{Q}}X) = 0$ (Jensen: $\psi(\theta) \ge \theta\,\mathbb{E}^{\mathbb{Q}}X$), so I is nondecreasing on $[\mathbb{E}^{\mathbb{Q}}X, \infty)$; it is finite near a (every $x = \psi'(\theta)$ with θ near θ_a is attained), hence continuous at a. So $\inf_{x>a} I(x) = I(a)$ and the right side is $-I(a)$; with (i), $\frac{1}{n}\log p_n \to -I(a)$. Jensen's inequality gives $\widetilde{\mathbb{E}}[Y^2] \ge p_n^2$, and (ii) gives $\widetilde{\mathbb{E}}[Y^2] \le e^{-nI(a)} p_n$; dividing the logarithms by $\log p_n^2 < 0$,

$$1 \;\ge\; \frac{\log \widetilde{\mathbb{E}}[Y^2]}{\log p_n^2} \;\ge\; \frac{-nI(a) + \log p_n}{2\log p_n} \;\longrightarrow\; 1. \qquad \square$$

The tilt θ_a is a *saddle point*: it makes the rare level the mean of the sampling law, $\mathbb{E}_{\theta_a}[S_n/n] = a$, and no more. Given the rare event, the summands behave approximately like tilted variables (Asmussen and Glynn, 2007, Sections VI.2 and VI.5). As a jump-count example, let S_n be the number of jumps of a Poisson process with intensity 2 per year over $n = 10$ years, and $p_n = \mathbb{Q}(S_n \ge 40) = 5.32 \times 10^{-5}$. Then $\theta_a = \log 2$, $I(a) = 4\log 2 - 2 = 0.773$, the bound is $e^{-nI(a)} = 4.41 \times 10^{-4}$, and plain Monte Carlo has relative error 137 per sample. Tilted counts (Poisson with mean 4 per year) give 2.17 (our implementation, `default_rng`, seed 7, 10^5 samples): $(137/2.17)^2 \approx 4000$ times fewer samples.

9.4 Path space: Girsanov's theorem and the optimal drift

A simulated path is a function of a vector $Z \in \mathbb{R}^m$ of independent normals, one per date (or one per date and factor). Shifting Z to $\mathcal{N}(\mu, I_m)$ multiplies m independent one-dimensional weights, so by Theorem 9.3 with $\beta = 1$

$$L(Z) = \exp\big(-\mu^\top Z + \tfrac{1}{2}|\mu|^2\big), \qquad \widetilde{\mathbb{E}}[L^2] = e^{|\mu|^2}. \tag{9.5}$$

The continuous-time meaning of this weight is Girsanov's theorem.

Theorem 9.8 (Path-space importance sampling)**.** *Let B be a standard Brownian motion on $[0,T]$ under $\widetilde{\mathbb{Q}}$, with a filtration for it, and let θ be an adapted process that is either bounded or deterministic with $\int_0^T \theta_t^2\,\mathrm{d}t < \infty$. Set*

$$\tilde{W}_t = B_t + \int_0^t \theta_u\,\mathrm{d}u, \qquad L_T = \exp\Big(-\int_0^T \theta_u\,\mathrm{d}B_u - \frac{1}{2}\int_0^T \theta_u^2\,\mathrm{d}u\Big).$$

(i) For every Borel functional G on $C[0,T]$ with $\mathbb{E}^{\mathbb{Q}}|G(W)| < \infty$ for a $\mathbb{Q}$-Brownian motion W: $\mathbb{E}^{\mathbb{Q}}[G(W)] = \widetilde{\mathbb{E}}\big[G(\tilde{W})\,L_T\big]$.

(ii) If θ is deterministic, $\widetilde{\mathbb{E}}[L_T^2] = \exp\big(\int_0^T \theta_u^2\,\mathrm{d}u\big)$.

(iii) If $\theta_u = \mu_j/\sqrt{\Delta t}$ on $(t_{j-1}, t_j]$, $t_j = j\Delta t$, $\Delta t = T/m$, then $L_T = \exp(-\mu^\top Z + \tfrac{1}{2}|\mu|^2)$ with $Z_j = (\tilde{W}_{t_j} - \tilde{W}_{t_{j-1}})/\sqrt{\Delta t}$, and the Z_j are independent $\mathcal{N}(\mu_j, 1)$ under $\widetilde{\mathbb{Q}}$: the discrete mean shift (9.5) is this change of measure exactly, with no discretisation error in the weight.

(iv) Under $\widetilde{\mathbb{Q}}$, $S_t = S\exp(\nu t + \sigma\tilde{W}_t)$ solves $\mathrm{d}S_t = (r - y + \sigma\theta_t)S_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}B_t$.

A feedback drift $\theta_t = h(t, \tilde{W}_t)$ makes $\tilde{W}$ the solution of $\mathrm{d}\tilde{W} = h(t,\tilde{W})\,\mathrm{d}t + \mathrm{d}B$, and θ is adapted only if this equation has a strong solution, for instance when h is Lipschitz in x; if h is also bounded, the theorem applies.

Proof. (i) We use Girsanov's theorem in the form of Shreve (2004), Theorem 5.2.3, with $\Theta = \theta$ and $W = B$: if $\widetilde{\mathbb{E}}\int_0^T \theta_u^2 L_u^2\,\mathrm{d}u < \infty$, where L_u is the same exponential on $[0,u]$, then $\widetilde{\mathbb{E}}[L_T] = 1$ and $\tilde{W}$ is a Brownian motion under the probability $\hat{\mathbb{Q}}$ with $\mathrm{d}\hat{\mathbb{Q}} = L_T\,\mathrm{d}\widetilde{\mathbb{Q}}$. The hypothesis holds: if $|\theta| \le C$, then $L_u^2 = M_u\exp(\int_0^u \theta^2) \le M_u e^{C^2 u}$ with $M_u = \exp(-2\int_0^u \theta\,\mathrm{d}B - 2\int_0^u \theta^2)$ a positive local martingale, hence a supermartingale with $\widetilde{\mathbb{E}}M_u \le 1$; if θ is deterministic, (ii) below on $[0,u]$ gives $\widetilde{\mathbb{E}}L_u^2 = e^{\int_0^u \theta^2}$. The law of a Brownian motion on $C[0,T]$ is determined by its finite-dimensional distributions, so $G(\tilde{W})$ has the same law under $\hat{\mathbb{Q}}$ as $G(W)$ under $\mathbb{Q}$, and $\mathbb{E}^{\mathbb{Q}}[G(W)] = \mathbb{E}_{\hat{\mathbb{Q}}}[G(\tilde{W})] = \widetilde{\mathbb{E}}[L_T G(\tilde{W})]$. (ii) For deterministic θ, $\int_0^T \theta\,\mathrm{d}B \sim \mathcal{N}(0, v)$ with $v = \int_0^T \theta^2$ (Shreve, 2004, Theorem 4.4.9), so $\widetilde{\mathbb{E}}[L_T^2] = \widetilde{\mathbb{E}}[e^{-2\int\theta\,\mathrm{d}B}]e^{-v} = e^{2v}e^{-v}$. (iii) $\int_0^T \theta\,\mathrm{d}B = \sum_j \mu_j(B_{t_j} - B_{t_{j-1}})/\sqrt{\Delta t} = \sum_j \mu_j(Z_j - \mu_j)$, because $\tilde{W}_{t_j} - \tilde{W}_{t_{j-1}} = B_{t_j} - B_{t_{j-1}} + \mu_j\sqrt{\Delta t}$, and $\frac{1}{2}\int\theta^2 = \frac{1}{2}|\mu|^2$; so $L_T = \exp(-\mu^\top Z + |\mu|^2 - \frac{1}{2}|\mu|^2)$. The Brownian increments are independent $\mathcal{N}(0,\Delta t)$ under $\widetilde{\mathbb{Q}}$. (iv) Itô's formula: $\mathrm{d}S = S\big((\nu + \tfrac{\sigma^2}{2})\,\mathrm{d}t + \sigma\,\mathrm{d}\tilde{W}\big)$ and $\mathrm{d}\tilde{W} = \mathrm{d}B + \theta\,\mathrm{d}t$. □

So a drift added to the Brownian motion is a drift $\sigma\theta$ added to the asset: to price an out-of-the-money call, simulate a market that rallies, and weight. A volatility cannot be changed this

way. If each of m increments is rescaled by $\beta \neq 1$, the weight has $\widetilde{\mathbb{E}}[L^2] = \big(\beta^2/\sqrt{2\beta^2-1}\big)^m$ by Theorem 9.3, and $\beta^4 > 2\beta^2 - 1$ makes this explode as the grid is refined; in the limit the two path laws are mutually singular, because the realised quadratic variation of $\log S$ is $\sigma^2 T$ almost surely under one and $\beta^2\sigma^2 T$ under the other. The course's β is a device for one coordinate, such as the terminal value.

Remark 9.9 (The zero-variance law in path space is a feedback drift). Let $\mathrm{d}X_t = b(t,X_t)\,\mathrm{d}t + \sigma(t,X_t)\,\mathrm{d}W_t$ under $\mathbb{Q}$, with b and σ continuous, let $g > 0$, and suppose that $u \in C^{1,2}([0,T)\times\mathbb{R})$ is positive, continuous on $[0,T]\times\mathbb{R}$, solves $u_t + bu_x + \frac{1}{2}\sigma^2 u_{xx} = 0$ with $u(T,\cdot) = g$, and that $\int \sigma u_x(t,X_t)\,\mathrm{d}W_t$ is a martingale. Itô's formula and the equation give $\mathrm{d}u(t,X_t) = \sigma u_x\,\mathrm{d}W_t$, so $u(t,X_t) = \mathbb{E}^{\mathbb{Q}}[g(X_T)\mid \mathcal{F}_t]$ is the price process (the Feynman–Kac representation of Theorem 1.6, with zero rate), and Itô's formula for $\log u$ gives

$$g(X_T) = u(0,X_0)\exp\Big(\int_0^T \theta_t^*\,\mathrm{d}W_t - \frac{1}{2}\int_0^T (\theta_t^*)^2\,\mathrm{d}t\Big), \qquad \theta_t^* = \sigma(t,X_t)\,\partial_x \log u(t,X_t).$$

The exponential is the martingale $u(t,X_t)/u(0,X_0)$, so Girsanov's theorem for a martingale density (Karatzas and Shreve, 1991, Theorem 3.5.1) applies: the zero-variance law of Theorem 9.2, with density $g(X_T)/u(0,X_0)$ with respect to $\mathbb{Q}$, is the change of measure under which W acquires the drift θ^*. It needs u, the price function itself; the drift of the next proposition freezes it along a single path.

Proposition 9.10 (The Glasserman–Heidelberger–Shahabuddin drift)**.** *Let $G:\mathbb{R}^m \to [0,\infty)$ be Borel with $p = \mathbb{E}[G(Z)] \in (0,\infty)$ for $Z \sim \mathcal{N}(0,I_m)$, let $F = \log G$ (with $F = -\infty$ where $G = 0$), and $H(z) = F(z) - \frac{1}{2}|z|^2$. For $\mu \in \mathbb{R}^m$ let $Y_\mu = G(Z)\exp(-\mu^\top Z + \frac{1}{2}|\mu|^2)$ with $Z \sim \mathcal{N}(\mu, I_m)$ under $\widetilde{\mathbb{Q}}$.*

(i) $\widetilde{\mathbb{E}}[Y_\mu] = p$ and $\widetilde{\mathbb{E}}[Y_\mu^2] = \mathbb{E}\big[G(Z)^2 e^{-\mu^\top Z + |\mu|^2/2}\big]$ for $Z \sim \mathcal{N}(0,I_m)$.

(ii) If $G(z) = e^{\kappa + v^\top z}$, then v is the unique maximiser of H and $Y_v = p$ surely: zero variance.

(iii) If F is concave and μ maximises H, then $F(z) \le F(\mu) + \mu^\top(z-\mu)$ for all z. Consequently $0 \le Y_\mu \le e^{H(\mu)}$ surely, $p \le e^{H(\mu)}$, and $\widetilde{\mathbb{E}}[Y_\mu^2]/p^2 \le e^{H(\mu)}/p$.

At a maximiser where G is positive and differentiable, $\nabla F(\mu) = \mu$, that is $\nabla G(\mu) = G(\mu)\,\mu$.

Proof. (i) is Theorem 9.1 with (9.5). (ii) $H(z) = \kappa + \frac{1}{2}|v|^2 - \frac{1}{2}|z-v|^2$ is uniquely maximised at v, and $Y_v = e^{\kappa + v^\top Z - v^\top Z + |v|^2/2} = e^{\kappa + |v|^2/2} = \mathbb{E}[e^{\kappa + v^\top Z}] = p$. (iii) $H(\mu)$ is finite since $p > 0$ forces $G > 0$ somewhere. Fix z with $F(z) > -\infty$ (otherwise there is nothing to prove) and $t \in (0,1]$. Concavity gives $F(\mu + t(z-\mu)) \ge (1-t)F(\mu) + tF(z)$, and maximality gives $H(\mu + t(z-\mu)) \le H(\mu)$. Together,

$$t\big(F(z) - F(\mu)\big) \le \tfrac{1}{2}|\mu + t(z-\mu)|^2 - \tfrac{1}{2}|\mu|^2 = t\,\mu^\top(z-\mu) + \tfrac{t^2}{2}|z-\mu|^2.$$

Divide by t and let $t \downarrow 0$. Then $Y_\mu = \exp\big(F(Z) - \mu^\top Z + \frac{1}{2}|\mu|^2\big) \le \exp\big(F(\mu) - \frac{1}{2}|\mu|^2\big) = e^{H(\mu)}$, so $p = \widetilde{\mathbb{E}}[Y_\mu] \le e^{H(\mu)}$ and $\widetilde{\mathbb{E}}[Y_\mu^2] \le e^{H(\mu)}\widetilde{\mathbb{E}}[Y_\mu] = e^{H(\mu)}p$. The first-order condition is $\nabla H(\mu) = 0$. □

The drift μ is the *most important path*: the point that maximises payoff times density, $G(z)e^{-|z|^2/2}$. Writing $Y_\mu = e^{H(\mu)}\exp\big(F(Z) - F(\mu) - \nabla F(\mu)^\top(Z-\mu)\big)$ shows what the shift achieves: the linear part of $\log G$ around μ is removed exactly, and only its curvature contributes variance. The drift is asymptotically optimal in a small-noise limit (Theorem 9.11); Glasserman et al. (1999)

combine it with stratification along the directions where the curvature is largest (Chapter 10). A call on a lognormal asset has a concave $\log G$, since $u \mapsto \log(e^u - K)$ is concave on $\{e^u > K\}$; so do a digital on a convex set and the geometric Asian call (Exercise 9.7). For the arithmetic Asian, $\log G$ need not be concave, and the drift is found numerically.

Remark 9.11 (Small-noise optimality). For $\varepsilon > 0$ let $p_\varepsilon = \mathbb{E}\big[\exp\big(F(\sqrt{\varepsilon}Z)/\varepsilon\big)\big]$, $Z \sim \mathcal{N}(0, I_m)$: noise of size $\sqrt{\varepsilon}$ and a log-payoff of order $1/\varepsilon$, the large-deviations regime. Sample with the drift $\mu/\sqrt{\varepsilon}$, where μ maximises H, and call the estimator Y_ε. *Assume* that F is concave and finite and continuous at μ. Then $\varepsilon \log p_\varepsilon \to H(\mu)$ and $\varepsilon \log \widetilde{\mathbb{E}}[Y_\varepsilon^2] \to 2H(\mu)$: the estimator is logarithmically efficient in ε. Indeed $F_\varepsilon(z) = F(\sqrt{\varepsilon}z)/\varepsilon$ is concave with $F_\varepsilon(z) - \frac{1}{2}|z|^2 = H(\sqrt{\varepsilon}z)/\varepsilon$, maximised at $\mu/\sqrt{\varepsilon}$, so (iii) gives $p_\varepsilon \le e^{H(\mu)/\varepsilon}$ and $\widetilde{\mathbb{E}}[Y_\varepsilon^2] \le e^{H(\mu)/\varepsilon} p_\varepsilon$. Conversely, given $\delta > 0$ take $r \in (0, \delta]$ with $F \ge F(\mu) - \delta$ on the ball $B(\mu, r)$; shifting Z by $\mu/\sqrt{\varepsilon}$ as in (9.5), $\mathbb{Q}(\sqrt{\varepsilon}Z \in B(\mu, r)) \ge e^{-(|\mu|^2/2 + |\mu| r)/\varepsilon}\, \mathbb{Q}(|Z| < r/\sqrt{\varepsilon})$, so $\liminf \varepsilon \log p_\varepsilon \ge H(\mu) - \delta(1 + |\mu|)$. Jensen's inequality $\widetilde{\mathbb{E}}[Y_\varepsilon^2] \ge p_\varepsilon^2$ closes the argument. Glasserman et al. (1999) establish this kind of optimality for smooth payoffs that need not be log-concave; see Glasserman (2003), Section 4.6.2, for the statement and its hypotheses.

Example 9.12 (Arithmetic Asian call)**.** Take the test case with $m = 16$ equally spaced fixings $t_j = j/16$ and payoff $e^{-rT}\big(\frac{1}{16}\sum_j S_{t_j} - K\big)^+$. Maximising H by a quasi-Newton method from a point where the payoff is positive gives the drifts of Figure 9.3(a): they decrease along the path, because an early normal moves every later fixing of the average, and their size $|\mu|$ is 1.10, 2.11, 3.17 and 4.13 for $K = 100$, 120, 140, 160. The variance falls by factors 9.2, 42, 618 and 1.9×10^4 (Table 9.2).

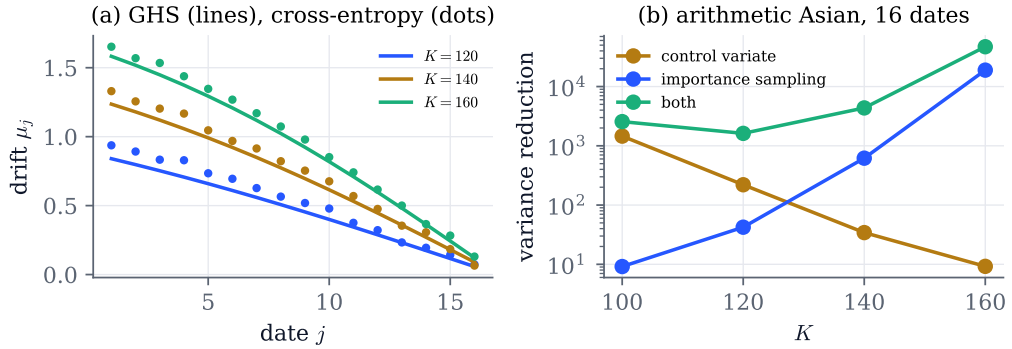

Figure 9.3: Arithmetic Asian call, test case, 16 fixings. (a) The GHS drift (lines) and the cross-entropy drift after 8 iterations of 10^4 paths (dots; Section 9.6). (b) Variance reduction factors of the control variate, of importance sampling and of both together (Table 9.2).

9.5 Rare events and relative error

When the price p is small, the absolute error says little; what matters is the *relative error per sample* $\sqrt{\mathrm{Var}(Y)}/p = \sqrt{\widetilde{\mathbb{E}}[Y^2]/p^2 - 1}$. A relative precision ε at 95% confidence needs about $(1.96/\varepsilon)^2\, \widetilde{\mathbb{E}}[Y^2]/p^2$ samples. For plain Monte Carlo of a probability, $\widetilde{\mathbb{E}}[Y^2]/p^2 = 1/p$.

Definition 9.13 (Efficiency for rare events)**.** Let Y_d be unbiased estimators of p_d, with $p_d \to 0$ as $d \to \infty$. The family has *bounded relative error* if $\limsup_{d\to\infty} \widetilde{\mathbb{E}}[Y_d^2]/p_d^2 < \infty$, and is *logarithmically efficient* if $\log \widetilde{\mathbb{E}}[Y_d^2]/\log(p_d^2) \to 1$ (Asmussen and Glynn, 2007, Section VI.1).

Since $\widetilde{\mathbb{E}}[Y_d^2] \geq p_d^2$, the ratio of logarithms never exceeds 1 once $p_d < 1$: logarithmic efficiency says that the second moment decays at the best possible exponential rate, while the relative error may still grow, but more slowly than any power of $1/p_d$. Theorem 9.7 is an example.

Lemma 9.14 (Mills ratio)**.** *For $x > 0$,* $\dfrac{x}{1+x^2} N'(x) \leq N(-x) \leq \dfrac{N'(x)}{x}$.

Proof. $N(-x) = \int_x^\infty N'(t)\,\mathrm{d}t \leq \int_x^\infty \frac{t}{x} N'(t)\,\mathrm{d}t = \frac{N'(x)}{x}$, since $(N')'(t) = -tN'(t)$. And $\frac{\mathrm{d}}{\mathrm{d}t}(-N'(t)/t) = (1+t^{-2})N'(t)$, so $N'(x)/x = \int_x^\infty (1+t^{-2})N'(t)\,\mathrm{d}t \leq (1+x^{-2})N(-x)$. □

Proposition 9.15 (Deep out-of-the-money digitals)**.** *Let $d > 0$ and $p_d = \mathbb{Q}(Z > d) = N(-d)$: the undiscounted value of a digital call with $d = -d_-$.*

(i) Plain Monte Carlo has $\widetilde{\mathbb{E}}[Y^2]/p_d^2 = 1/p_d$, between $\sqrt{2\pi}\, d e^{d^2/2}$ and $\sqrt{2\pi}\,\frac{1+d^2}{d} e^{d^2/2}$: exponential growth in d^2.

(ii) The mean shift to the strike, $Z \sim \mathcal{N}(d, 1)$ under $\widetilde{\mathbb{Q}}$, has

$$\frac{2\sqrt{2\pi}\, d^3}{1+4d^2} \leq \frac{\widetilde{\mathbb{E}}[Y^2]}{p_d^2} = \frac{e^{d^2} N(-2d)}{N(-d)^2} \leq \frac{\sqrt{2\pi}\,(1+d^2)^2}{2d^3},$$

both bounds being $\sqrt{\pi/2}\, d\,(1+o(1))$: logarithmically efficient, but the relative error grows like $(\pi/2)^{1/4}\sqrt{d}$.

(iii) Sampling $Z = d + E/d$ with E standard exponential gives $Y \leq N'(d)/d$ surely and $\widetilde{\mathbb{E}}[Y^2]/p_d^2 \leq (1+d^2)/d^2$: the relative error per sample is at most $1/d$, so it vanishes as $d \to \infty$.

Proof. (i) is Theorem 9.14 applied to $1/p_d$. (ii) By the mirror formula with $\beta = 1$ and $a = c = d$, $\widetilde{\mathbb{E}}[Y^2] = e^{d^2}\mathbb{Q}(Z - d > d) = e^{d^2} N(-2d)$. Theorem 9.14 at $2d$ and at d, with $e^{d^2} N'(2d) = e^{-d^2}/\sqrt{2\pi}$ and $N'(d)^2 = e^{-d^2}/(2\pi)$, gives

$$\frac{e^{d^2} N(-2d)}{N(-d)^2} \leq \frac{e^{-d^2}/(2d\sqrt{2\pi})}{d^2 e^{-d^2}/(2\pi(1+d^2)^2)}, \qquad \frac{e^{d^2} N(-2d)}{N(-d)^2} \geq \frac{2d\, e^{-d^2}/((1+4d^2)\sqrt{2\pi})}{e^{-d^2}/(2\pi d^2)},$$

which simplify to the stated bounds. Logarithmic efficiency follows because $\log p_d^2 = -d^2 + O(\log d)$ and $\log \widetilde{\mathbb{E}}[Y^2] = -d^2 + O(\log d)$. (iii) The sampling density is $\tilde{q}(z) = d e^{-d(z-d)}$ on (d, ∞), which contains the support of $f = \mathbf{1}_{(d,\infty)}$. Since $N'(z) = N'(d) e^{-d(z-d) - (z-d)^2/2}$, the weight is $L(z) = N'(d) e^{-(z-d)^2/2}/d \leq N'(d)/d$, and $Y = L(Z)$. Then $\widetilde{\mathbb{E}}[Y^2] \leq \frac{N'(d)}{d}\widetilde{\mathbb{E}}[Y] = \frac{N'(d)}{d} p_d$, and Theorem 9.14 gives $N'(d)/(dp_d) \leq (1+d^2)/d^2$. □

The exponential law in (iii) is, to first order, the conditional law of $Z - d$ given $Z > d$: it approximates the zero-variance law of Theorem 9.2 in shape, not just in location, which a Gaussian cannot do. For the call, with the same d, the relative second moments grow like $2\sqrt{2\pi}\, d e^{d^2/2}$ for plain Monte Carlo and like $\sqrt{\pi/8}\, d$ for the shift to the strike (Exercise 9.5). For the digital, written as $G = \mathbf{1}_{[d,\infty)}$ (same price), $\log G$ is concave, the GHS drift is the boundary

point d, and the bound of Theorem 9.10(iii) reads $\widetilde{\mathbb{E}}[Y^2]/p_d^2 \le e^{-d^2/2}/N(-d) \le \sqrt{2\pi}(1+d^2)/d$, about twice the exact value.

Table 9.1 compares, at two strikes, the shifts chosen by three criteria: the GHS drift, the variance-optimal shift, and the cross-entropy shift of Section 9.6, which for a digital is $\mathbb{E}[Z \mid Z > d] = N'(d)/N(-d)$. They differ by amounts that shrink as d grows, and their relative errors are within 3% of each other; the payoff shape matters more than the criterion. Figure 9.4 shows the whole range of strikes.

		GHS		optimal		cross-entropy		plain
		shift	rel. err.	shift	rel. err.	shift	rel. err.	rel. err.
$K = 150$, $d = 2.177$	digital	2.177	1.587	2.379	1.563	2.531	1.577	8.18
	call	2.581	1.010	2.731	0.998	2.844	1.005	11.43
$K = 200$, $d = 3.616$	digital	3.616	2.018	3.747	2.007	3.861	2.015	81.71
	call	3.880	1.312	3.995	1.305	4.092	1.310	115.26

Table 9.1: Mean shifts (in standard deviations of Z) and exact relative errors per sample for the test case with strikes 150 and 200 (digital prices 0.0146 and 1.48×10^{-4}, call prices 0.165 and 1.52×10^{-3}). The exponential sampler of Theorem 9.15(iii) has relative errors 0.234 and 0.119 for the digitals.

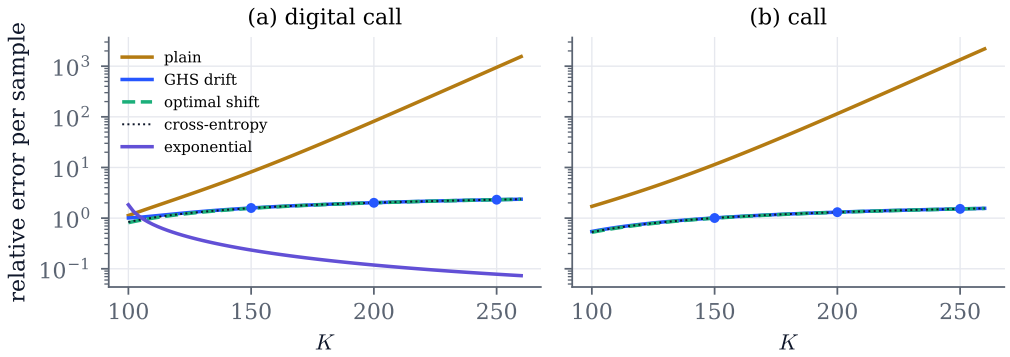

Figure 9.4: Exact relative error per sample against the strike for the test case: (a) digital call, (b) call. The GHS, variance-optimal and cross-entropy shifts are nearly indistinguishable; dots are Monte Carlo estimates for the GHS shift (10^5 samples).

9.6 Adaptive importance sampling: cross-entropy

The variance-optimal parameter minimises $\widetilde{\mathbb{E}}[(fL)^2]$, itself an expectation to be estimated with weights that may be wild. The *cross-entropy method* (Rubinstein, 1997; de Boer et al., 2005) replaces the variance by a distance to the zero-variance law $g^* = fq/\mu$ that, for exponential families, is minimised in closed form.

Proposition 9.16 (Cross-entropy for Gaussian mean shifts)**.** *Let $f \ge 0$ be Borel, $Z \sim \mathcal{N}(0, I_m)$ under $\mathbb{Q}$, $\mu = \mathbb{E}^{\mathbb{Q}}[f(Z)] \in (0,\infty)$, $\mathbb{E}^{\mathbb{Q}}[f(Z)|Z|^2] < \infty$, and $g^* = fq/\mu$. For $v \in \mathbb{R}^m$ let $\tilde{q}_v$ be the density of $\mathcal{N}(v, I_m)$ and $L_v = q/\tilde{q}_v$.*

(i) The cross-entropy $-\int g^* \log \tilde{q}_v$ *equals* $\frac{m}{2}\log(2\pi) + \frac{1}{2}\mathbb{E}_{g^*}|Z - v^*|^2 + \frac{1}{2}|v - v^*|^2$, *with* $v^* = \mathbb{E}_{g^*}[Z] = \mathbb{E}^{\mathbb{Q}}[f(Z)Z]/\mu$. *It is uniquely minimised at* $v = v^*$; *when the Kullback–Leibler divergence* $\int g^* \log(g^*/\tilde{q}_v)$ *is finite, it differs from the cross-entropy by a constant, and* v^* *minimises it too.*

(ii) For every $u \in \mathbb{R}^m$, $v^* = \widetilde{\mathbb{E}}_u[fL_uZ]/\widetilde{\mathbb{E}}_u[fL_u]$, *and if* $Z_1, Z_2, \dots$ *are independent* $\mathcal{N}(u, I_m)$, *the ratio* $\hat{v}_n = \sum_i f(Z_i)L_u(Z_i)Z_i / \sum_i f(Z_i)L_u(Z_i)$ *converges to* v^* *almost surely.*

Proof. (i) $-\log \tilde{q}_v(z) = \frac{m}{2}\log(2\pi) + \frac{1}{2}|z - v|^2$, and $\mathbb{E}_{g^*}|Z - v|^2 = \mathbb{E}_{g^*}|Z - v^*|^2 + |v - v^*|^2$ because $\mathbb{E}_{g^*}[Z - v^*] = 0$; all terms are finite by the moment assumption. (ii) By Theorem 9.1(i), applied coordinatewise to fZ_j and to f, $\mathbb{E}^{\mathbb{Q}}[fZ] = \widetilde{\mathbb{E}}_u[fL_uZ]$ and $\mu = \widetilde{\mathbb{E}}_u[fL_u]$. The numerator and denominator of $\hat{v}_n$, divided by n, converge almost surely by the strong law, and the limit of the denominator is $\mu > 0$. □

When f is rare, most $f(Z_i)$ vanish at $u = 0$ and $\hat{v}_n$ is undefined or noisy. The multilevel algorithm (de Boer et al., 2005, Section 2.1 and Algorithm 2.1) climbs to the target through intermediate levels: sample from $\mathcal{N}(v_t, I)$, set the level γ_t to the smaller of the target and the $(1-\rho)$-quantile of the simulated scores (for example the average A of an Asian), update $v_{t+1} = \hat{v}_n$ with the payoff at level γ_t, and stop once γ_t has reached the target and v_t has settled. For the Asian example, with $\rho = 0.1$ and 10^4 paths per iteration, the target level is reached at the second or third iteration, and after 8 iterations the ratio of the norms of the cross-entropy and GHS drifts is 1.23, 1.13, 1.07, 1.05 for $K = 100, 120, 140, 160$ (Figure 9.3(a); our implementation, seed K): close in the tail, larger near the money. For the digital, $v^* = \mathbb{E}[Z \mid Z > d] = N'(d)/N(-d)$ overshoots the GHS drift d by at most $1/d$, by Theorem 9.14.

Adaptive schemes have risks that a fixed drift does not.

(a) *Noisy updates.* Each update is a weighted average. If the current law is far from g^*, a few samples carry all the weight and successive iterates wander; the effective sample size below, computed with the update weights $f(Z_i)L(Z_i)$, is the alarm.

(b) *Reuse bias.* Iteration t samples from $\mathcal{N}(v_t, I)$ with v_t fixed by earlier iterations, so its weighted mean is conditionally unbiased, and an equal-weight pool of the iterations' estimates is unbiased (a sum of martingale differences). Bias appears when a parameter chosen from a sample is applied to that same sample, for instance $f(Z_i + \hat{v})e^{-\hat{v}^\top Z_i - |\hat{v}|^2/2}$ with $\hat{v}$ computed from the Z_i, or when the best of several candidates is reported with its in-sample error: the analogue of Theorem 8.14.

(c) *Scale.* The full Gaussian family also updates the covariance, to $\mathrm{Cov}_{g^*}(Z)$. A smaller covariance means thinner tails, and Theorem 9.3 shows how that can make the variance infinite without any sign in the sample; in path space, Section 9.4 shows that rescaling many coordinates degenerates. Estimating $m(m+1)/2$ parameters from weighted samples is itself noisy.

Proposition 9.17 (Effective sample size)**.** *Let* $L_1, L_2, \dots \geq 0$ *be i.i.d. weights with* $\widetilde{\mathbb{E}}[L] = 1$, *and* $\mathrm{ESS}_n = \left(\sum_{i \leq n} L_i\right)^2 / \sum_{i \leq n} L_i^2$ *(on the event that some* $L_i > 0$*).*

(i) $1 \leq \mathrm{ESS}_n \leq n$, *with* $\mathrm{ESS}_n = n$ *if and only if all the* L_i *are equal.*

(ii) $\mathrm{ESS}_n/n \to 1/\widetilde{\mathbb{E}}[L^2]$ *almost surely, with* $1/\infty = 0$.

(iii) For the mean shift $\mathcal{N}(\mu, I_m)$, $1/\widetilde{\mathbb{E}}[L^2] = e^{-|\mu|^2}$; *for a rescaling* $\mathcal{N}(0, \beta^2 I_m)$ *with* $\beta > 1/\sqrt{2}$, *it is* $\left(\sqrt{2\beta^2 - 1}/\beta^2\right)^m$.

Proof. (i) Cauchy–Schwarz gives $(\sum L_i)^2 \le n \sum L_i^2$, with equality iff the L_i are equal; and $(\sum L_i)^2 \ge \sum L_i^2$ for nonnegative terms. (ii) By the strong law, $\frac{1}{n}\sum L_i \to 1$; if $\widetilde{\mathbb{E}}[L^2] < \infty$, also $\frac{1}{n}\sum L_i^2 \to \widetilde{\mathbb{E}}[L^2]$. If $\widetilde{\mathbb{E}}[L^2] = \infty$, then for every M, $\frac{1}{n}\sum L_i^2 \ge \frac{1}{n}\sum \min(L_i^2, M) \to \widetilde{\mathbb{E}}[\min(L^2, M)]$, which tends to ∞ as $M \to \infty$ by monotone convergence. (iii) The weights are products of m independent one-dimensional weights; use Theorem 9.3(i). □

The effective sample size of a mean shift depends on $|\mu|^2$ only. A drift of fixed size spread over many dates, like a Girsanov drift with fixed $\int \theta^2$ on a finer and finer grid, does not degenerate as the dimension grows; a fixed shift c per coordinate gives ne^{-mc^2}, and a rescaling decays geometrically in m (Figure 9.5(a)). The effective sample size describes the weights, not the estimator of a given payoff. For the Asian at $K = 160$, $e^{-|\mu|^2} \approx 4 \times 10^{-8}$, and yet the estimator is excellent: the weights are deliberately small on the paths that pay and large on paths that do not. For one payoff the diagnostic is the relative standard error of fL; the effective sample size matters when one weighted sample prices many payoffs, and inside the adaptive updates.

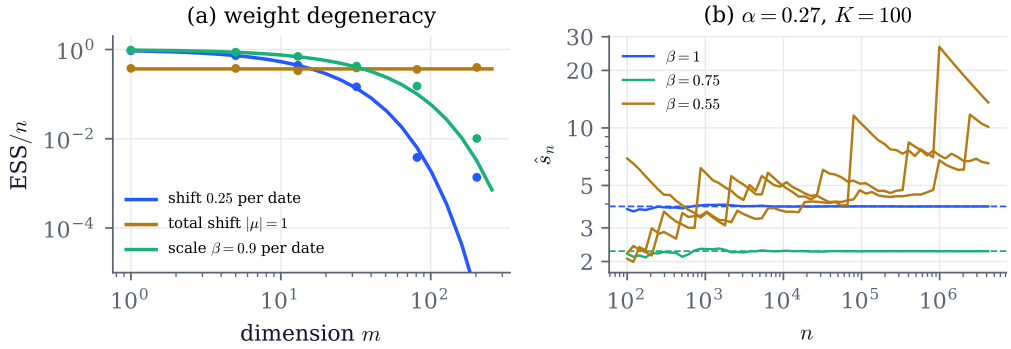

Figure 9.5: (a) ESS/n against the dimension (Theorem 9.17): lines exact, dots from 10^4 draws, which overestimate it when the weights are heavy-tailed. (b) Running sample standard deviation of the call estimator with $\alpha = 0.27$: for $\beta = 1$ and 0.75 it settles at the exact values (dashed); for $\beta = 0.55 < 1/\sqrt{2}$ the variance is infinite and three seeds never settle (our implementation, `default_rng`, seeds 1–3).

9.7 Importance sampling with control variates

Importance sampling decides where to sample; a control variate removes what is predictable in the samples. A control keeps its known mean after weighting.

Proposition 9.18 (Controls under the sampling law)**.** *Let* g *be Borel with known* $\gamma = \mathbb{E}^{\mathbb{Q}}[g(X)]$, *and let* $\tilde{q}$ *satisfy* (9.1) *for both* f *and* g, *with* $f(X)L(X)$ *and* $g(X)L(X)$ *in* $L^2(\mathbb{Q})$.

(i) For every b, $Y_b = f(X)L(X) - b(g(X)L(X) - \gamma)$ *has* $\widetilde{\mathbb{E}}[Y_b] = \mu$. *Its variance is minimised at* $b^* = \widetilde{\mathrm{Cov}}(fL, gL)/\widetilde{\mathrm{Var}}(gL)$, *where it equals* $(1 - \tilde{\rho}^2)\widetilde{\mathrm{Var}}(fL)$, $\tilde{\rho}$ *being the correlation of* fL *and* gL *under* $\widetilde{\mathbb{Q}}$ *(provided* $\widetilde{\mathrm{Var}}(gL) > 0$*).*

> *(ii) With $g \equiv 1$ the control is the weight itself, with mean 1 when $\{q > 0\} \subseteq \{\tilde{q} > 0\}$. The* self-normalised *estimator $\hat{\mu}_n^{\mathrm{sn}} = \sum_i f(X_i)L(X_i) / \sum_i L(X_i)$ equals the control-variate estimator with control L and coefficient $b = \hat{\mu}_n^{\mathrm{sn}}$.*

Proof. (i) $\widetilde{\mathbb{E}}[gL] = \gamma$ by Theorem 9.1(i), and the rest is Theorem 8.12 under $\widetilde{\mathbb{Q}}$. (ii) With $\bar{Y} = \frac{1}{n}\sum f(X_i)L(X_i)$ and $\bar{L} = \frac{1}{n}\sum L(X_i)$, $\bar{Y} - \hat{\mu}_n^{\mathrm{sn}}(\bar{L} - 1) = \hat{\mu}_n^{\mathrm{sn}}$ because $\hat{\mu}_n^{\mathrm{sn}}\bar{L} = \bar{Y}$. □

The correlation that matters is the one between the *weighted* quantities under $\widetilde{\mathbb{Q}}$. The linear control S_T is weak for deep out-of-the-money calls even under $\mathbb{Q}$ (exact $\rho^2 = 0.10$ at $K = 150$ and 0.003 at $K = 200$, against 0.78 at $K = 100$), and after the GHS shift it is weaker still ($\tilde{\rho}^2 = 0.003$ and 8×10^{-7}), because the drift removes the linear part of $\log G$ (Theorem 9.10). A nonlinear control that tracks the payoff on the important paths survives, such as the geometric Asian call for the Asian (Theorem 15.4). In Table 9.2 the control dominates near the money, importance sampling in the tail, and together they win at every strike. The self-normalised estimator suits weights known up to a constant (Exercise 9.4).

		standard error, $n = 2^{17}$				variance reduction		
K	price	plain	control	IS	both	control	IS	both
100	4.4963	2.0×10^{-2}	5.3×10^{-4}	6.7×10^{-3}	4.0×10^{-4}	1458	9.2	2560
120	0.36113	5.7×10^{-3}	3.9×10^{-4}	8.8×10^{-4}	1.4×10^{-4}	220	42	1621
140	0.012755	1.0×10^{-3}	1.7×10^{-4}	4.1×10^{-5}	1.5×10^{-5}	34	618	4359
160	2.7333×10^{-4}	1.4×10^{-4}	4.7×10^{-5}	1.0×10^{-6}	6.6×10^{-7}	9.2	1.9×10^{4}	4.7×10^{4}

Table 9.2: Arithmetic Asian call, test case, 16 fixings: plain Monte Carlo, the geometric control, importance sampling with the GHS drift, and both (control fitted in sample with `variance.control_variate`). Our implementation (`default_rng`, `SeedSequence(2026)`, `importance.asian_laboratory`); price: combined estimate. Only 9 of 2^{17} plain paths pay at $K = 160$, so the variances of plain Monte Carlo and of the control alone are estimated from the importance sample ($\mathbb{E}^{\mathbb{Q}}[h] = \widetilde{\mathbb{E}}[hL]$, optimal coefficient).

▶ Interactive demo: Asian drifts, effective sample sizes, this laboratory.

Further reading

Figures 9.1 and 9.2 and Theorem 9.4 reproduce the course's slides. Glasserman (2003) covers the likelihood ratio and tilting in Section 4.6.1 and the GHS drift in Section 4.6.2; Asmussen and Glynn (2007) treat importance sampling in Section V.1 and rare events in Chapter VI (efficiency VI.1, tilting VI.2, cross-entropy VI.8). Cross-entropy is due to Rubinstein (1997) (tutorial: de Boer et al. (2005)); Dembo and Zeitouni (1998), Chapter 2, proves Cramér's theorem.

Chapter 10

Stratified sampling and conditional Monte Carlo

The methods of Chapter 8 use a symmetry of the input law or a quantity with a known mean. This chapter uses the structure of the variance itself. For any random variable ξ computed along with the payoff Y, the law of total variance splits

$$\operatorname{Var} Y = \underbrace{\mathbb{E}\big[\operatorname{Var}(Y \mid \xi)\big]}_{\text{within}} + \underbrace{\operatorname{Var}\big(\mathbb{E}[Y \mid \xi]\big)}_{\text{between}}. \tag{10.1}$$

Stratified sampling cuts the range of ξ into pieces and fixes in advance how many draws land in each piece; it removes the variance between the pieces, and in the limit of many pieces the whole between part. *Conditional Monte Carlo* computes $\mathbb{E}[Y \mid \xi]$ in closed form and averages that instead of Y; it removes the within part. The two are complementary, and the art is in the choice of ξ: the terminal value of a Brownian path, a direction in the space of normals, every coordinate at once (Latin hypercube sampling), or the whole volatility path of a stochastic volatility model, given which the Black–Scholes formula applies.

Throughout, Y is a real random variable with $\mathbb{E}[Y^2] < \infty$, mean μ and variance σ^2, and $\hat{\mu}_n$ is the plain estimator from n independent copies, with variance σ^2/n (Theorem 6.1). In pricing, Y is a discounted payoff and the samples are drawn under $\mathbb{Q}$; as in Chapter 6 we write $\mathbb{P}$ and $\mathbb{E}$ for the sampling measure. The test case is $S = K = 100$, $T = 1$, $r = 1\%$, $y = 2\%$, $\sigma = 20\%$, with $S_T = S\exp(\nu T + \sigma\sqrt{T}Z)$, $\nu = r - y - \frac{\sigma^2}{2}$, and the discounted call has mean 7.36429 and variance 158.006.

10.1 Stratified sampling

Let ξ be a random element of a measurable space E (the *stratification variable*) and let $A_1, \dots, A_H$ be a measurable partition of E into *strata*, with $p_h = \mathbb{P}(\xi \in A_h)$. A stratum with $p_h = 0$ is almost surely never visited and contributes nothing to μ; we discard such strata and assume $p_h > 0$ for every h. Let

$$\mu_h = \mathbb{E}[Y \mid \xi \in A_h] = \frac{\mathbb{E}[Y\mathbf{1}_{A_h}(\xi)]}{p_h}, \qquad \sigma_h^2 = \mathbb{E}\big[(Y-\mu_h)^2 \mid \xi \in A_h\big]$$

be the mean and the variance of Y in stratum h. Given integers $n_h \geq 1$ with $\sum_h n_h = n$, draw $Y_{h,1}, \dots, Y_{h,n_h}$ independently from the conditional law of Y given $\xi \in A_h$, independently across strata, and set

$$\hat{\mu}_n^{\mathrm{st}} = \sum_{h=1}^{H} p_h \bar{Y}_h, \qquad \bar{Y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} Y_{h,i}. \tag{10.2}$$

Two things are needed: the probabilities p_h exactly, and a way to sample in a stratum. Both are easy when ξ is a uniform. Suppose $Y = g(U, U')$, where U is uniform on $(0,1)$ and independent of the rest U' of the randomness, and take strata $I_h = (a_{h-1}, a_h]$ with $0 = a_0 < \cdots < a_H = 1$, so $p_h = a_h - a_{h-1}$. If V is uniform and independent of U', then $g(a_{h-1} + p_h V, U')$ has the conditional law of Y given $U \in I_h$: for bounded measurable F, Fubini's theorem gives $\mathbb{E}[F(Y)\mathbf{1}_{I_h}(U)] = \int_{I_h} \mathbb{E}[F(g(u,U'))]\,\mathrm{d}u = p_h \mathbb{E}[F(g(a_{h-1} + p_h V, U'))]$. With $Z = N^{-1}(U)$ (Theorem 5.8) this stratifies a normal; equiprobable strata $a_h = h/H$ are the usual choice. Defining the strata in u-space makes every p_h exact even when the variable built from U has atoms or gaps.

Theorem 10.1 (Stratified sampling)**.** *In the setting above:*

(i) $\mu = \sum_h p_h \mu_h$ *and* $\sigma^2 = \sum_h p_h \sigma_h^2 + \sum_h p_h (\mu_h - \mu)^2$ *(within plus between).*

(ii) If $n_h \geq 1$ *for every* h, $\hat{\mu}_n^{\mathrm{st}}$ *is unbiased and* $\operatorname{Var} \hat{\mu}_n^{\mathrm{st}} = \sum_h p_h^2 \sigma_h^2 / n_h$. *If* $n_h = 0$ *for the strata in a set* E_0 *and their terms are dropped, the bias is* $-\sum_{h \in E_0} p_h \mu_h$.

(iii) (Proportional allocation.) *If* $n_h = n p_h$ *for every* h *(integers), then*

$$\operatorname{Var} \hat{\mu}_n^{\mathrm{st}} = \frac{1}{n} \sum_h p_h \sigma_h^2 = \frac{\sigma^2}{n} - \frac{1}{n} \sum_h p_h (\mu_h - \mu)^2 \leq \frac{\sigma^2}{n},$$

with equality if and only if $\mu_h = \mu$ *for every* h.

(iv) Splitting strata further never increases $\sum_h p_h \sigma_h^2$.

(v) (Many strata.) *Let* $\xi = U$ *be uniform and* $I_h = ((h-1)/H, h/H]$. *As* $H \to \infty$, $\sum_h p_h \sigma_h^2 \to \mathbb{E}[\operatorname{Var}(Y \mid U)] = \sigma^2 - \operatorname{Var}(\mathbb{E}[Y \mid U])$. *In particular, if* Y *is a function of* U *alone and* $H = H_n \to \infty$ *with proportional allocation, then* $n \operatorname{Var} \hat{\mu}_n^{\mathrm{st}} \to 0$.

Proof. (i) $\mathbb{E}Y = \sum_h \mathbb{E}[Y \mathbf{1}_{A_h}(\xi)] = \sum_h p_h \mu_h$. Expanding $Y - \mu = (Y - \mu_h) + (\mu_h - \mu)$ on $\{\xi \in A_h\}$, the cross term has expectation $2(\mu_h - \mu)\mathbb{E}[(Y - \mu_h)\mathbf{1}_{A_h}(\xi)] = 0$, so $\mathbb{E}[(Y-\mu)^2 \mathbf{1}_{A_h}(\xi)] = p_h \sigma_h^2 + p_h(\mu_h - \mu)^2$; sum over h.

(ii) The $\bar{Y}_h$ are independent, with means μ_h and variances σ_h^2/n_h; use (i). Without draws in the strata of E_0 the estimator misses their terms $p_h \mu_h$ in (i).

(iii) Substitute $n_h = n p_h$ in (ii) and use (i). The difference is $\frac{1}{n} \sum_h p_h (\mu_h - \mu)^2$, which vanishes exactly when every μ_h equals μ, since every $p_h > 0$.

(iv) If A_h is split into pieces $A_{h,k}$, apply (i) to the conditional law given $\xi \in A_h$: σ_h^2 is at least the within-piece part $\sum_k (p_{h,k}/p_h) \sigma_{h,k}^2$. Multiply by p_h and sum over h.

(v) Let $g(U) = \mathbb{E}[Y \mid U]$ and let $\mathcal{G}_H$ be generated by the events $\{U \in I_h\}$. On $\{U \in I_h\}$, $\mathbb{E}[Y \mid \mathcal{G}_H] = \mu_h$, so $\sum_h p_h \sigma_h^2 = \mathbb{E}[(Y - \mathbb{E}[Y \mid \mathcal{G}_H])^2]$. Since $\mathcal{G}_H \subset \sigma(U)$, $\mathbb{E}[Y \mid \mathcal{G}_H] = \mathbb{E}[g(U) \mid \mathcal{G}_H] =: (P_H g)(U)$, where $P_H g$ is the average of g over each cell I_h. The variable $Y - g(U)$ is

orthogonal to every square-integrable function of U, so

$$\sum_h p_h \sigma_h^2 = \mathbb{E}[(Y - g(U))^2] + \|g - P_H g\|^2_{L^2(0,1)}.$$

The first term is $\mathbb{E}[\mathrm{Var}(Y \mid U)]$. For the second, P_H is an orthogonal projection, so $\|P_H k\| \le \|k\|$; for a continuous c on $[0,1]$, $|c - P_H c| \le \omega_c(1/H)$, the modulus of continuity of c at $1/H$. Hence $\|g - P_H g\| \le \|g - c\| + \|c - P_H c\| + \|P_H(c - g)\| \le 2\|g - c\| + \omega_c(1/H)$. Continuous functions are dense in $L^2(0,1)$, so the second term tends to 0. □

Under plain sampling the number of draws that land in A_h is random, binomial with mean np_h; proportional allocation fixes it at its mean, and the between-strata variance, which came from that randomness, disappears. The word "never" in (iii) belongs to proportional allocation. With other allocations stratification can lose: take two equiprobable strata with $\mu_1 = \mu_2$, $\sigma_1 = 0$ and $\sigma_2 > 0$, so that $\sigma^2 = \sigma_2^2/2$; the allocation $n_1 = n - 1$, $n_2 = 1$ gives the variance $\sigma_2^2/4 > \sigma^2/n$ for every $n \ge 3$. Part (v) says that with many strata on a uniform, stratification removes all the variance that the uniform explains. By Theorem 5.8 the same holds for equiprobable strata of any scalar $\xi = F^{-1}(U)$ with a continuous distribution function F, since then $U = F(\xi)$ and $\sigma(\xi) = \sigma(U)$. No rate is claimed: how fast $\sum_h p_h \sigma_h^2$ approaches its limit depends on the smoothness of $u \mapsto \mathbb{E}[Y \mid U = u]$ (Exercise 10.3).

Example 10.2 (The test-case call)**.** Stratify $Z = N^{-1}(U)$ into H equiprobable strata. The call pays off when $Z > z^* = 0.15$, so every stratum below z^* has $\mu_h = \sigma_h = 0$: with $H = 2$ the lower half is worthless, and the variance factor $\sum_h p_h \sigma_h^2/\sigma^2$ is 0.657. With $H = 10$, 100 and 1000 it is 0.1007, 0.0081 and 0.00073, standard-error ratios 0.317, 0.090 and 0.027 (Figure 10.1). The stratum moments are computed by Gauss–Legendre quadrature on each stratum (`stratum_moments`), which avoids the cancellation of the closed forms in narrow strata. A stratified draw costs one inversion. On our machine `rng.norm_inv` costs about seven times NumPy's normal generator, and the whole stratified estimator about 4.5 times the plain one with NumPy normals (machine-dependent; `laboratory_costs` and the notebook time it), so the efficiency (Theorem 6.9) is about $1/(4.5 \times 0.1007) \approx 2.2$ with 10 strata, not 10, and about 27 with 100.

The course's experiment. The course's notebook prices a deep out-of-the-money call, $S = 100$, $K = 150$, $T = 1$, $r = 5\%$, $y = 0$, $\sigma = 20\%$ (price 0.35963), with $n = 10^5$ draws in 10, 50 or 100 equiprobable strata, and compares the variances of 100 repeated estimates. The exact variance factors are 0.856, 0.318 and 0.158. Only 3.0% of the mass lies above the strike ($z^* = 1.877$), so with 10 strata the whole payoff sits in the top stratum, and stratification removes only the randomness of the number of draws that reach it; Neyman allocation repairs this (Exercise 10.6). Repeating the course's protocol in our implementation (`course_otm_experiment`, numpy `default_rng`, seed 5), the variances of the 100 estimates are 9.61, 8.05, 2.64 and 1.46 (in units of 10^{-5}) for plain sampling and 10, 50, 100 strata, against the exact 8.08, 6.92, 2.57 and 1.28; a variance estimated from 100 estimates has a relative standard error of about 14%.

10.1.1 Optimal allocation

Proportional allocation ignores the σ_h. Strata where the payoff varies more deserve more draws.

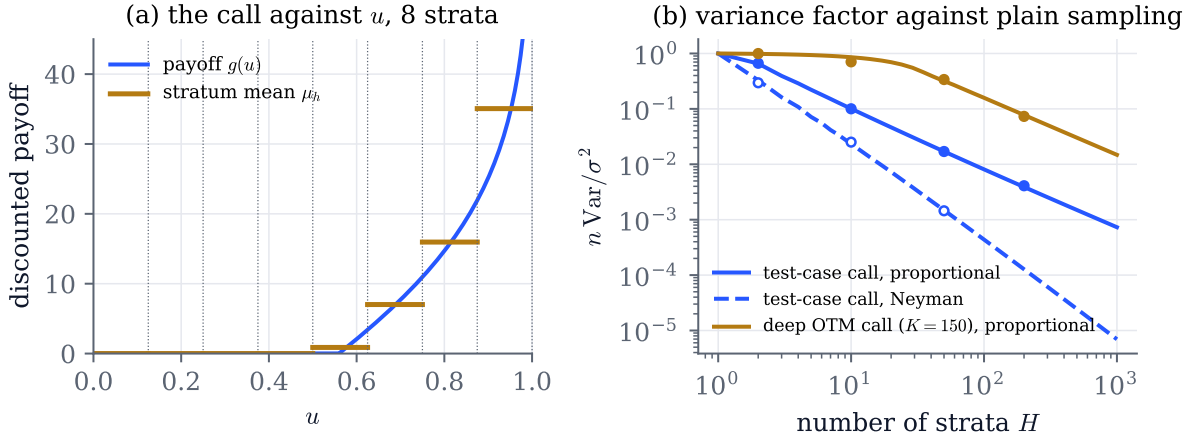

Figure 10.1: (a) The discounted test-case call as a function of the uniform u, $g(u) = e^{-rT}(S_T(N^{-1}(u)) - K)^+$, and its means on 8 equiprobable strata; the curve leaves the frame near $u = 1$. Proportional stratification samples only the deviations from the step function. (b) Variance factors $n\,\mathrm{Var}\,\hat{\mu}_n^{\mathrm{st}}/\sigma^2$ against the number H of equiprobable strata: lines are exact (Theorems 10.1 and 10.3), dots are measured over 400 independent replications of $n = 2000$ draws, seed 1010 (filled: proportional; open: Neyman, with at least one draw per stratum). The deep out-of-the-money call is the course's example.

Proposition 10.3 (Neyman allocation)**.** *Let* $\bar{\sigma} = \sum_h p_h\sigma_h > 0$.

(i) For all real $n_h > 0$ *with* $\sum_h n_h = n$,

$$\sum_h \frac{p_h^2\sigma_h^2}{n_h} \geq \frac{\bar{\sigma}^2}{n},$$

with equality if and only if $n_h = np_h\sigma_h/\bar{\sigma}$ *for every* h. *If every* $\sigma_h > 0$, *this* Neyman allocation *is the unique minimiser. If* $\sigma_h = 0$ *for* $k \geq 1$ *strata, the bound is not attained by an allocation with all* $n_h > 0$*; one draw in each such stratum and the Neyman allocation of the other* $n - k$ *draws give* $\bar{\sigma}^2/(n-k)$.

(ii) $\dfrac{\sigma^2}{n} - \dfrac{\bar{\sigma}^2}{n} = \dfrac{1}{n}\sum_h p_h(\mu_h - \mu)^2 + \dfrac{1}{n}\sum_h p_h(\sigma_h - \bar{\sigma})^2$*; the second term is the gain over proportional allocation, zero if and only if all* σ_h *are equal.*

(iii) If the $n_h \geq 1$ *are computed from a pilot run independent of the draws used in* (10.2)*, the estimator is unbiased, and conditionally on the pilot the allocation, though random, is fixed, so that Theorem 10.4(i)–(ii) below apply.*

Proof. (i) By the Cauchy–Schwarz inequality,

$$\bar{\sigma}^2 = \Big(\sum_h \frac{p_h\sigma_h}{\sqrt{n_h}}\sqrt{n_h}\Big)^2 \leq \Big(\sum_h \frac{p_h^2\sigma_h^2}{n_h}\Big)\Big(\sum_h n_h\Big),$$

with equality if and only if the vectors $(p_h\sigma_h/\sqrt{n_h})_h$ and $(\sqrt{n_h})_h$ are proportional, that is $n_h = cp_h\sigma_h$ for a constant c, which $\sum_h n_h = n$ fixes as $n/\bar{\sigma}$. If some $\sigma_h = 0$ this would require

$n_h = 0$. Strata with $\sigma_h = 0$ contribute nothing to the variance whatever $n_h \geq 1$, and the bound applied to the others with $n - k$ draws gives $\bar{\sigma}^2/(n-k)$.

(ii) By Theorem 10.1(i), $\sigma^2 - \bar{\sigma}^2 = \sum_h p_h(\mu_h - \mu)^2 + \sum_h p_h\sigma_h^2 - \bar{\sigma}^2$, and $\sum_h p_h\sigma_h^2 - \bar{\sigma}^2 = \sum_h p_h(\sigma_h - \bar{\sigma})^2$ is the variance of σ_h under the weights p_h.

(iii) Conditionally on the pilot run the n_h are constants and the main draws keep their law, so Theorem 10.1(ii) and Theorem 10.4 hold conditionally; for the mean, take expectations. □

The σ_h are estimated from a pilot run. If the pilot draws are reused in the final estimate, the allocation depends on them, and both the estimate and its variance estimate are in general biased (typically at order $1/n$, which we do not prove). Rounding to integers and keeping $n_h \geq 2$ costs little when $np_h\sigma_h/\bar{\sigma} \gg 1$. For the test-case call with 10 strata the Neyman factor is 0.0234 against 0.1007 for proportional allocation (standard-error ratios 0.153 and 0.317), and with 100 strata 0.00044 against 0.0081. Half of the 10 strata lie below z^*; Neyman gives them nothing in the limit, and they could even be skipped since $\mu_h = 0$ is known. Neyman allocation is a form of importance sampling (Chapter 9) restricted to a partition: it oversamples the strata where the payoff varies, and the weights $p_h n/n_h$ implicit in (10.2) are its likelihood ratios, constant on each stratum. With unequal costs per stratum the rule changes (Exercise 10.2).

10.1.2 Error estimates

Proposition 10.4 (Standard errors for stratified estimators)**.** *(i) If $n_h \geq 2$ for every h and $\hat{s}_h^2$ is the sample variance in stratum h, then $\hat{v} = \sum_h p_h^2 \hat{s}_h^2/n_h$ is an unbiased estimator of $\operatorname{Var}\hat{\mu}_n^{\mathrm{st}}$.*

(ii) Let H be fixed and $n_h/n \to q_h > 0$ as $n \to \infty$, and let $v = \sum_h p_h^2\sigma_h^2/q_h > 0$. Then $\sqrt{n}(\hat{\mu}_n^{\mathrm{st}} - \mu) \Rightarrow \mathcal{N}(0, v)$ and $(\hat{\mu}_n^{\mathrm{st}} - \mu)/\sqrt{\hat{v}} \Rightarrow \mathcal{N}(0,1)$.

(iii) R independent replications of any stratified estimator (for instance with one draw per stratum) are i.i.d. with mean μ, and Theorem 6.4 applies to their average as $R \to \infty$.

Proof. (i) $\mathbb{E}\hat{s}_h^2 = \sigma_h^2$ for $n_h \geq 2$. (ii) By the central limit theorem in each stratum, independently across h, $\sqrt{n_h}(\bar{Y}_h - \mu_h) \Rightarrow \mathcal{N}(0, \sigma_h^2)$, a point mass if $\sigma_h = 0$. As $\sqrt{n}(\hat{\mu}_n^{\mathrm{st}} - \mu) = \sum_h p_h\sqrt{n/n_h}\,\sqrt{n_h}(\bar{Y}_h - \mu_h)$, the characteristic functions multiply and the limit is $\mathcal{N}(0, v)$. By the strong law $n\hat{v} \to v$ almost surely; Slutsky's lemma (van der Vaart, 1998, Lemma 2.8) concludes. (iii) is immediate. □

With one draw per stratum no $\hat{s}_h$ exists; (iii) is then the way to a confidence interval, at the price of fewer strata per replication. Merging pairs of adjacent strata gives a conservative alternative (Exercise 10.4).

▶ Interactive demo: stratify a call or a put at any strike with any number of strata, compare proportional and Neyman allocation, and see where the draws go.

10.2 Post-stratification

Sometimes the stratum of a draw can be observed but not prescribed: the paths are simulated forward, and only afterwards do we see in which range their terminal value fell. If the p_h are known, the draws can still be reweighted. Let $X_1, \dots, X_n$ be i.i.d. copies of (ξ, Y), N_h the number of draws with $\xi \in A_h$, $\bar{Y}_h$ the mean of their Y, and $\mathcal{E} = \{N_h \geq 1 \text{ for every } h\}$. The *post-stratified estimator* is

$$\hat{\mu}_n^{\text{post}} = \sum_h p_h \bar{Y}_h \ \text{ on } \mathcal{E}, \qquad \hat{\mu}_n^{\text{post}} = \hat{\mu}_n \ \text{ (the plain mean) on } \mathcal{E}^c.$$

Lemma 10.5. *If N is binomial with parameters n and $p \in (0,1]$, then*

$$\mathbb{E}[N^{-1}\mathbf{1}_{\{N\geq 1\}}] = \frac{1}{np} + \frac{1-p}{n^2p^2} + O(n^{-3}).$$

Proof. Let $q = 1-p$ and $j \geq 1$. Since $\binom{n}{k}\frac{k!}{(k+j)!} = \frac{n!}{(n+j)!}\binom{n+j}{k+j}$, the binomial theorem gives

$$\mathbb{E}\Big[\frac{1}{(N+1)\cdots(N+j)}\Big] = \frac{1-R_j}{(n+1)\cdots(n+j)\,p^j},$$

where R_j is the probability that a binomial$(n+j,p)$ variable is below j, so $0 \leq R_j \leq (n+j)^j q^n$. For $N \geq 1$,

$$\frac{1}{N} = \frac{1}{N+1} + \frac{1}{(N+1)(N+2)} + \frac{2}{N(N+1)(N+2)},$$

and $0 \leq \frac{2}{N(N+1)(N+2)} \leq \frac{8}{(N+1)(N+2)(N+3)}$, whose expectation is $O(n^{-3})$. Restricting the first two expectations to $\{N \geq 1\}$ changes them by at most q^n. Hence $\mathbb{E}[N^{-1}\mathbf{1}_{\{N\geq 1\}}] = \frac{1}{(n+1)p} + \frac{1}{(n+1)(n+2)p^2} + O(n^{-3})$; finally $\frac{1}{(n+1)p} = \frac{1}{np} - \frac{1}{n^2p} + O(n^{-3})$ and $\frac{1}{(n+1)(n+2)p^2} = \frac{1}{n^2p^2} + O(n^{-3})$. □

Proposition 10.6 (Post-stratification)**.** *In this setting, with every $p_h > 0$:*

(i) Given the stratum labels of the n draws, on $\mathcal{E}$ the estimator has conditional mean μ and conditional variance $\sum_h p_h^2\sigma_h^2/N_h$.

(ii) $|\mathbb{E}\hat{\mu}_n^{\text{post}} - \mu| \leq \frac{\sigma}{\sqrt{n}}\big(\sum_h (1-p_h)^n\big)^{1/2}$: the bias is exponentially small.

(iii) $\mathbb{E}[(\hat{\mu}_n^{\text{post}} - \mu)^2] = \frac{1}{n}\sum_h p_h\sigma_h^2 + \frac{1}{n^2}\sum_h (1-p_h)\sigma_h^2 + O(n^{-3})$.

(iv) $\sqrt{n}(\hat{\mu}_n^{\text{post}} - \mu) \Rightarrow \mathcal{N}(0, \sum_h p_h\sigma_h^2)$.

Proof. (i) The pairs (label, draw) are i.i.d., and given its label h a draw has the conditional law of (ξ, Y) given $\xi \in A_h$; so, given all labels, the draws are independent with these laws. On $\mathcal{E}$, $\bar{Y}_h$ is then a mean of N_h independent draws from stratum h; apply Theorem 10.1(ii) with $n_h = N_h$.

(ii) $\mathcal{E}$ is determined by the labels, so by (i) $\mathbb{E}[(\hat{\mu}_n^{\text{post}} - \mu)\mathbf{1}_{\mathcal{E}}] = 0$, and $|\mathbb{E}\hat{\mu}_n^{\text{post}} - \mu| = |\mathbb{E}[(\hat{\mu}_n - \mu)\mathbf{1}_{\mathcal{E}^c}]| \leq \frac{\sigma}{\sqrt{n}}\mathbb{P}(\mathcal{E}^c)^{1/2}$ by Cauchy–Schwarz, with $\mathbb{P}(\mathcal{E}^c) \leq \sum_h (1-p_h)^n$.

(iii) By (i), $\mathbb{E}[(\hat{\mu}_n^{\text{post}} - \mu)^2] = \sum_h p_h^2\sigma_h^2\mathbb{E}[N_h^{-1}\mathbf{1}_{\mathcal{E}}] + \mathbb{E}[(\hat{\mu}_n - \mu)^2\mathbf{1}_{\mathcal{E}^c}]$. Given the labels, the plain mean has conditional mean $\sum_h \frac{N_h}{n}\mu_h$ and conditional variance $\sum_h N_h\sigma_h^2/n^2$, so $\mathbb{E}[(\hat{\mu}_n - \mu)^2 \mid \text{labels}] \leq$

$\max_h \sigma_h^2 + \max_h(\mu_h - \mu)^2$, and the last term is $O(\mathbb{P}(\mathcal{E}^c))$. Also $0 \le \mathbb{E}[N_h^{-1}\mathbf{1}_{\{N_h \ge 1\}}] - \mathbb{E}[N_h^{-1}\mathbf{1}_{\mathcal{E}}] \le \mathbb{P}(\mathcal{E}^c)$. As $\mathbb{P}(\mathcal{E}^c) \le H \max_h(1-p_h)^n = O(n^{-3})$, Theorem 10.5 with N_h binomial(n, p_h) gives $\sum_h p_h^2 \sigma_h^2 (\frac{1}{np_h} + \frac{1-p_h}{n^2 p_h^2}) + O(n^{-3})$, which is the claim.

(iv) On $\mathcal{E}$, whose probability tends to 1, $\hat{\mu}_n^{\text{post}} = \varphi(\bar{a}, \bar{b})$ with $\varphi(a,b) = \sum_h p_h a_h / b_h$, $\bar{a}_h = \frac{1}{n}\sum_i Y_i \mathbf{1}_{A_h}(\xi_i)$ and $\bar{b}_h = N_h/n$. The vectors $(Y\mathbf{1}_{A_h}(\xi), \mathbf{1}_{A_h}(\xi))_h$ are i.i.d. and square integrable, so by the multivariate central limit theorem and the delta method (van der Vaart, 1998, Theorem 3.1), $\sqrt{n}(\varphi(\bar{a}, \bar{b}) - \mu)$ converges to a centred normal law with the variance of $\sum_h (\partial_{a_h}\varphi\, Y\mathbf{1}_{A_h}(\xi) + \partial_{b_h}\varphi\, \mathbf{1}_{A_h}(\xi))$ at $(a_h, b_h) = (p_h\mu_h, p_h)$, where $\partial_{a_h}\varphi = 1$ and $\partial_{b_h}\varphi = -\mu_h$. That variable is $\sum_h \mathbf{1}_{A_h}(\xi)(Y - \mu_h)$, whose variance is $\sum_h p_h \sigma_h^2$. □

Post-stratification is therefore asymptotically as good as proportional stratification; it loses the term $\frac{1}{n^2}\sum_h (1-p_h)\sigma_h^2$, the price of the random counts, which is negligible when every np_h is large. It breaks down when strata are so many that some are often empty. It needs no conditional sampling, so it keeps whatever fast generator the plain estimator uses. It is also a control variate in disguise, with the stratum indicators as controls (Exercise 10.5). For the test-case call with 10 strata and $n = 2 \times 10^5$, (iii) gives the standard-error ratio 0.3174, larger than for proportional allocation only in the sixth digit; the measured 0.311 (Table 10.1) is below both by sampling noise.

10.3 Stratifying paths

A path-dependent payoff depends on many normals, $Y = f(Z)$ with $Z \sim \mathcal{N}(0, I_d)$. Stratifying all of them on a product grid needs H^d cells, which is hopeless beyond a few dimensions. Two ways around this: stratify one well-chosen direction, or stratify every coordinate separately (Section 10.3.1).

Proposition 10.7 (Stratifying a direction; terminal stratification). *(i) Let $Z \sim \mathcal{N}(0, I_d)$, $v \in \mathbb{R}^d$ a unit vector and $\xi = v^\top Z$. Then $\xi \sim \mathcal{N}(0,1)$, and $Z - v\xi$ is independent of ξ with law $\mathcal{N}(0, I_d - vv^\top)$. Consequently, if ξ' has the conditional law of ξ given $\xi \in A$ and $Z' \sim \mathcal{N}(0, I_d)$ is independent of ξ', then $v\xi' + (I_d - vv^\top)Z'$ has the conditional law of Z given $v^\top Z \in A$.*

(ii) Let $0 < t_1 < \cdots < t_m = T$. In the Brownian-bridge construction of Theorem 7.7(iii), fill $t_m = T$ first, so that the first normal is $Z_1 = W_T/\sqrt{T}$. If Z_1 is drawn from its conditional law given $Z_1 \in A$ and $Z_2, \dots, Z_m$ are independent standard normals, the construction produces a vector with the conditional law of $(W_{t_1}, \dots, W_{t_m})$ given $W_T/\sqrt{T} \in A$. With the increments construction, $W_T/\sqrt{T} = v^\top Z$ with $v_j = \sqrt{(t_j - t_{j-1})/T}$, and (i) gives the same law.

(iii) Hence, for $f(W_{t_1}, \dots, W_{t_m}) \in L^2$ and H equiprobable strata of W_T with proportional allocation, the stratified estimator is unbiased with variance $\frac{1}{n}\sum_h p_h \sigma_h^2$, and $n \operatorname{Var} \hat{\mu}_n^{\text{st}} \to \mathbb{E}[\operatorname{Var}(f \mid W_T)]$ as $H \to \infty$.

Proof. (i) $(\xi, Z - v\xi)$ is a linear image of Z, hence Gaussian, and $\operatorname{Cov}(Z - v\xi, \xi) = v - v\,v^\top v = 0$, so the two are independent; $\operatorname{Var}\xi = v^\top v = 1$ and $\operatorname{Cov}(Z - v\xi) = I_d - 2vv^\top + vv^\top = I_d - vv^\top$. As $I_d - vv^\top$ is a symmetric projection, $(I_d - vv^\top)Z' \sim \mathcal{N}(0, I_d - vv^\top)$. For bounded measurable

F, independence gives $\mathbb{E}[F(Z)\mathbf{1}_A(\xi)] = \mathbb{E}[F(v\xi + R)\mathbf{1}_A(\xi)]$ with $R = Z - v\xi$ independent of ξ; dividing by $\mathbb{P}(\xi \in A)$ this is $\mathbb{E}[F(v\xi' + R')]$ with ξ' and $R' \stackrel{d}{=} R$ independent.

(ii) The construction is a fixed linear map, $W = AZ$, and $Z_1 = W_T/\sqrt{T}$ because the first step sets $W_T = \sqrt{T}Z_1$. The vector $(Z_1, \dots, Z_m)$ has independent components, so conditioning on $Z_1 \in A$ changes the law of Z_1 only; apply the map A. For the increments construction $W_T = \sum_j \sqrt{t_j - t_{j-1}}Z_j$.

(iii) follows from (ii), Theorem 10.1(iii) and (v), with $\xi = N(W_T/\sqrt{T})$. □

For geometric Brownian motion with deterministic parameters, the paths of Theorem 7.2 built by the bridge on the variance clock have $\ln S_T$ set by the first normal, so terminal stratification stratifies S_T; for a European payoff the limit in (iii) is 0. In the code, the function `terminal_stratified_normals` draws Z_1 in its strata and the rest freely, and `paths.gbm_paths` with `method="bridge"` builds the paths (Figure 10.2a).

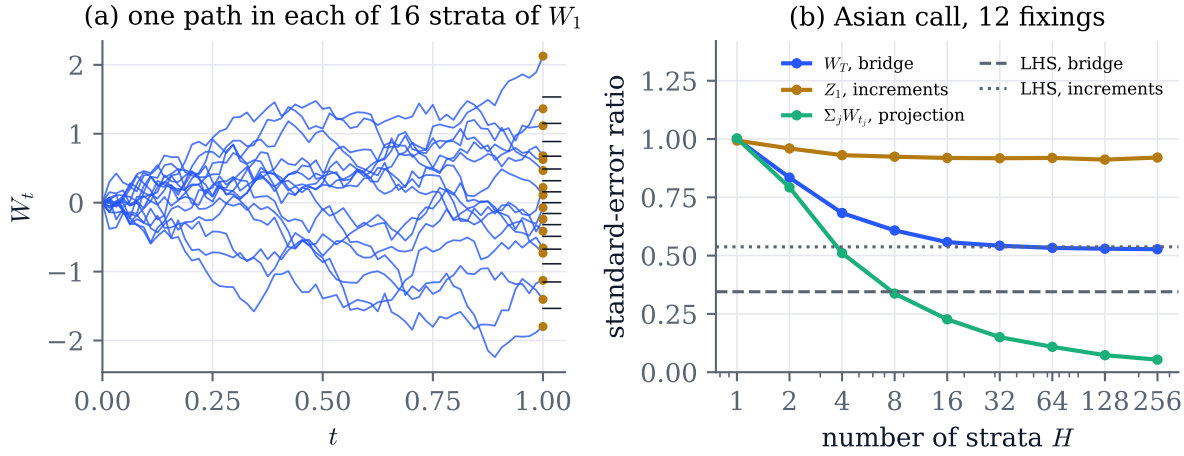

Figure 10.2: (a) Sixteen Brownian paths on 64 steps, one in each of 16 equiprobable strata of W_1 (ticks: stratum boundaries), completed by the Brownian bridge. (b) The arithmetic Asian call on 12 monthly fixings in the test case: standard-error ratio to plain sampling at equal n against the number H of equiprobable strata, with proportional allocation and $n = 2^{17}$ paths per point (seed 1021). Stratified variables: W_T through the bridge (blue), the first increment Z_1 of the increments construction (gold), and the direction of $\sum_j W_{t_j}$ (Theorem 10.7(i); green). Horizontal lines: Latin hypercube sampling with the two constructions, from 2000 independent designs of 2^{10} points.

Example 10.8 (An Asian call)**.** The course's Asian example averages the stock at $t_j = j/12$, $j = 1, \dots, 12$, with strike 100 in the test case. The course reports 4.56058, a Monte Carlo estimate from 10^8 paths (standard error about 0.0008); our geometric-control-variate estimate (`asian_reference`, two runs of 2^{25} paths, seeds 2026 and 7) is 4.56122 and 4.56127, each with standard error 0.00003, against the exact value 4.5612893 computed in Theorem 15.11 (the first run is 2.0 standard errors low); the course's estimate is within its standard error. The plain standard deviation per path is 7.49. With 64 strata of W_T the standard-error ratio is 0.530 (Table 10.1): terminal stratification removes about 72% of the variance of the payoff,

and more strata add little (Figure 10.2b), so this is nearly all of the part that W_T explains. Stratifying the first normal of the increments construction instead fixes only $W_{1/12}$, and the ratio is 0.92. The average of the path matters more than its end: stratifying the direction $v \propto L^\top \mathbf{1}$ of $\sum_j W_{t_j}$, where L is the increments matrix, gives 0.108 with 64 strata and about 0.055 with 256 (Figure 10.2b). That direction is, to within a cosine of 0.9999, the optimal importance-sampling drift of the payoff (Theorem 9.10), the direction along which Glasserman et al. (1999) stratify.

10.3.1 Latin hypercube sampling

Instead of one direction finely, Latin hypercube sampling stratifies every coordinate, each into n strata, with n points in total.

Definition 10.9 (Latin hypercube sample)**.** Let $\pi_1, \dots, \pi_d$ be independent uniformly random permutations of $\{1, \dots, n\}$ and U_{ij}, $i \le n$, $j \le d$, i.i.d. uniforms on $(0,1)$, independent of the permutations. The points $V_i = (V_{i1}, \dots, V_{id})$ with $V_{ij} = (\pi_j(i) - 1 + U_{ij})/n$ form a *Latin hypercube sample*, and $\hat\mu_n^{\mathrm{LHS}} = \frac{1}{n}\sum_i f(V_i)$ estimates $\mu = \int_{[0,1]^d} f$. Normals are obtained as $N^{-1}(V_{ij})$.

Every column has exactly one point in each interval $[(k-1)/n, k/n)$, and the permutations pair the columns at random (Figure 10.3a). For $f \in L^2([0,1]^d)$ write $f_j(u_j) = \mathbb{E}[f(U) \mid U_j = u_j] - \mu$ for U uniform on the cube, the *additive part* $f_{\mathrm{add}} = \mu + \sum_j f_j$ and the residual $e = f - f_{\mathrm{add}}$. The $f_j(U_j)$ are independent and centred, and $e(U)$ is uncorrelated with each of them, so $\sigma^2 = \sum_j \operatorname{Var} f_j(U_j) + \int e^2$.

Proposition 10.10 (Latin hypercube sampling)**.** *Let $f \in L^2([0,1]^d)$ with variance σ^2 under the uniform law.*

(i) Each V_i is uniform on $[0,1]^d$, so $\hat\mu_n^{\mathrm{LHS}}$ is unbiased.

(ii) If f is additive, $f = \mu + \sum_j f_j$, then $\operatorname{Var}\hat\mu_n^{\mathrm{LHS}} = \sum_{j=1}^d \frac{1}{n^2}\sum_{k=1}^n \sigma_{j,k}^2$, with $\sigma_{j,k}^2$ the variance of $f_j(U)$ given $U \in [(k-1)/n, k/n)$: each additive term is integrated as by one-dimensional stratification with one point per stratum, and $n \operatorname{Var}\hat\mu_n^{\mathrm{LHS}} \to 0$.

(iii) (Stein, 1987) For every $f \in L^2$, $\operatorname{Var}\hat\mu_n^{\mathrm{LHS}} = \frac{1}{n}\int e^2 + o(1/n)$: asymptotically, Latin hypercube sampling removes the variance of the additive part.

(iv) (McKay et al., 1979) If f is monotone in each coordinate, $\operatorname{Var}\hat\mu_n^{\mathrm{LHS}} \le \sigma^2/n$ for every n. (Owen, 1997b) For every $f \in L^2$ and $n \ge 2$, $\operatorname{Var}\hat\mu_n^{\mathrm{LHS}} \le \sigma^2/(n-1)$.

Proof. (i) For fixed i, $\pi_j(i)$ is uniform on $\{1, \dots, n\}$ and independent of U_{ij}, so $\mathbb{P}(V_{ij} \le x) = \frac{1}{n}\sum_k \mathbb{P}(k - 1 + U \le nx) = x$; the pairs $(\pi_j(i), U_{ij})$ are independent across j.

(ii) Fix j and list the points of column j by stratum: the point in $[(k-1)/n, k/n)$ is $(k - 1 + U_{\pi_j^{-1}(k)j})/n$. Given π_j, the $U_{\pi_j^{-1}(k)j}$, $k = 1, \dots, n$, are i.i.d. uniforms, and their conditional law does not depend on π_j. So $\frac{1}{n}\sum_i f_j(V_{ij})$ is the stratified estimator of $\int f_j$ with n equiprobable strata and one point each, whose variance is $\frac{1}{n^2}\sum_k \sigma_{j,k}^2$ by Theorem 10.1(ii). Different columns are built from independent ingredients, so the variances add. Finally $n \operatorname{Var}\hat\mu_n^{\mathrm{LHS}} = \sum_j \sum_k \frac{1}{n}\sigma_{j,k}^2 \to 0$ by Theorem 10.1(v) applied to $Y = f_j(U)$, for which $\operatorname{Var}(Y \mid U) = 0$.

(iii) is Stein (1987); see also Owen (2013, Proposition 10.1), where it is stated for $\int f^2 < \infty$.

(iv) The first statement is the theorem stated in Section 2.1 of McKay et al. (1979) and proved in their appendix (Section 8.3). The second is due to Owen (1997b), see Owen (2013, Proposition 10.4), stated there for $n > 2$; for $n = 2$ it is immediate, since $\operatorname{Var}\frac{f(V_1)+f(V_2)}{2} \le \frac{1}{2}(\operatorname{Var} f(V_1) + \operatorname{Var} f(V_2)) = \sigma^2$ by Cauchy–Schwarz. □

The n points of a design are dependent, so the variance is estimated from independent replications of the whole design, as in Theorem 10.4(iii). The bound $\sigma^2/(n-1)$ cannot be improved to σ^2/n: for a function with no additive part Latin hypercube sampling can be slightly worse than plain sampling (Exercise 10.7). For the Asian call of Theorem 10.8, Latin hypercube sampling of the 12 normals gives standard-error ratios of 0.54 ± 0.01 with the increments construction and 0.354 ± 0.006 with the bridge (2000 designs of 2^{10} points), which puts the most important variables first and makes the payoff more nearly additive in them (Table 10.1). Latin hypercube sampling balances only the one-dimensional projections of the design; the randomised quasi-Monte Carlo point sets of Chapter 11 also balance the higher-dimensional ones.

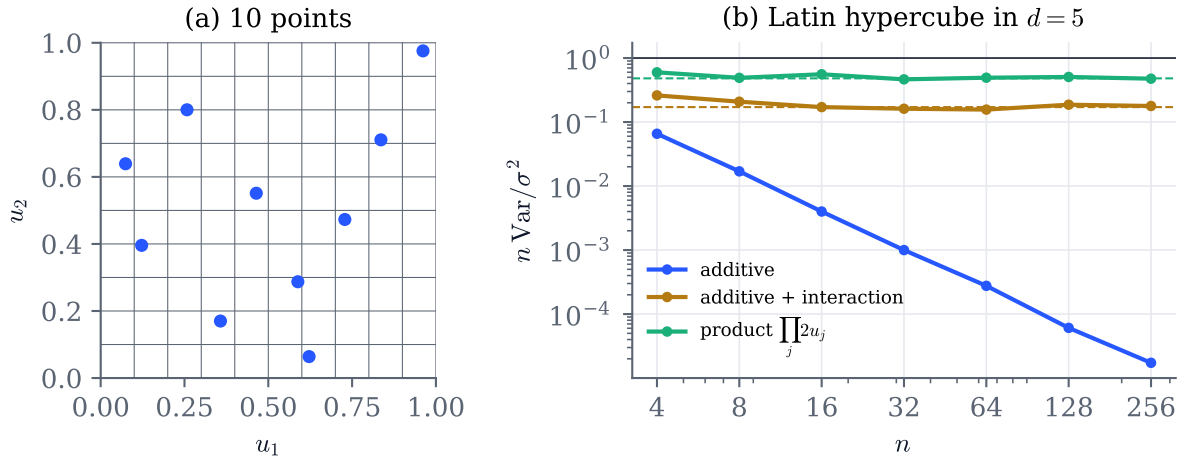

Figure 10.3: (a) A Latin hypercube sample of 10 points in $[0,1]^2$: one point in each row and each column of the grid. (b) $n \operatorname{Var} \hat{\mu}_n^{\mathrm{LHS}}/\sigma^2$ against n in dimension 5, from 400 independent designs for each n (seed 1040), for $\sum_j e^{u_j}$ (additive), the same plus $6(u_1 - \frac{1}{2})(u_2 - \frac{1}{2})$, and $\prod_j 2u_j$. Dashed: the limits $\int e^2/\sigma^2$ of Theorem 10.10(iii), 0.17 and 0.48; the solid line is plain sampling.

▶ Interactive demo: stratify Brownian paths on their endpoint or on any linear combination of the path, and draw Latin hypercube designs.

10.4 Conditional Monte Carlo

If part of the randomness can be integrated out in closed form, it should be.

Theorem 10.11 (Conditional Monte Carlo)**.** *Let $Y \in L^2$, let $\mathcal{G}$ be a sub-σ-algebra, and $Y^{\mathrm{c}} = \mathbb{E}[Y \mid \mathcal{G}]$.*

(i) $\mathbb{E}Y^{\mathrm{c}} = \mathbb{E}Y$ and $\operatorname{Var} Y = \operatorname{Var} Y^{\mathrm{c}} + \mathbb{E}[\operatorname{Var}(Y \mid \mathcal{G})]$, where $\operatorname{Var}(Y \mid \mathcal{G}) = \mathbb{E}[(Y - Y^{\mathrm{c}})^2 \mid \mathcal{G}]$. Hence $\operatorname{Var} Y^{\mathrm{c}} \le \operatorname{Var} Y$, with equality if and only if $Y = Y^{\mathrm{c}}$ almost surely, that is, if and

only if Y is almost surely equal to a $\mathcal{G}$-measurable variable.

(ii) *If $\mathcal{G}_1 \subset \mathcal{G}_2$, then* $\operatorname{Var}\mathbb{E}[Y \mid \mathcal{G}_1] \le \operatorname{Var}\mathbb{E}[Y \mid \mathcal{G}_2]$.

(iii) *The average $\hat{\mu}^{\mathrm{c}}_n$ of n i.i.d. copies of Y^{c} is unbiased with variance* $\operatorname{Var} Y^{\mathrm{c}}/n$*; if a copy of Y^{c} costs λ times a copy of Y, its efficiency relative to plain sampling is* $\operatorname{Var} Y/(\lambda \operatorname{Var} Y^{\mathrm{c}})$*, which may be below* 1.

Proof. (i) $\mathbb{E}Y^{\mathrm{c}} = \mathbb{E}Y$ by the tower property. $Y - Y^{\mathrm{c}}$ is orthogonal in L^2 to every square-integrable $\mathcal{G}$-measurable variable, in particular to $Y^{\mathrm{c}} - \mathbb{E}Y$, so $\mathbb{E}(Y - \mathbb{E}Y)^2 = \mathbb{E}(Y^{\mathrm{c}} - \mathbb{E}Y)^2 + \mathbb{E}(Y - Y^{\mathrm{c}})^2$, and $\mathbb{E}(Y - Y^{\mathrm{c}})^2 = \mathbb{E}[\operatorname{Var}(Y \mid \mathcal{G})]$. The last term vanishes if and only if $Y = Y^{\mathrm{c}}$ almost surely. (ii) Apply (i) to $\mathbb{E}[Y \mid \mathcal{G}_2]$ and $\mathcal{G}_1$, using $\mathbb{E}[\mathbb{E}[Y \mid \mathcal{G}_2] \mid \mathcal{G}_1] = \mathbb{E}[Y \mid \mathcal{G}_1]$. (iii) is Theorem 6.9. □

In statistics this is the Rao–Blackwell theorem (Casella and Berger, 2002, Theorem 7.3.17), where $\mathcal{G}$ is generated by a sufficient statistic so that Y^{c} does not depend on the unknown parameter. Here $\mathcal{G}$ is ours to choose, and the only requirement is that Y^{c} be computable. Part (ii) is the trade-off: conditioning on less information removes more variance but leaves a harder conditional expectation. Equation (10.1) shows how conditioning and stratification divide the work. For a scalar ξ with a continuous law and many equiprobable strata, stratifying on ξ gives $n \operatorname{Var} \to \mathbb{E}[\operatorname{Var}(Y \mid \xi)]$ (Theorem 10.1(v)), and conditioning on ξ gives $n \operatorname{Var} = \operatorname{Var}(\mathbb{E}[Y \mid \xi])$; the two add up to σ^2. Stratify on what cannot be integrated, integrate out what can, and, when $\mathbb{E}[Y \mid \xi] = g(\xi)$ is available, do both: stratifying ξ in $g(\xi)$ has variance $o(1/n)$. Further examples in these notes are the Brownian-bridge correction for barriers, which integrates out the path between monitoring dates (Theorem 16.9), conditioning a two-asset payoff on one asset (Exercise 10.8) and Curran's conditioning on the geometric average (Theorem 15.10). Conditional values are smooth in the parameters even when the payoff is not, which is what pathwise Greeks need (Theorem 13.4).

10.4.1 Conditioning on the volatility path

In a stochastic volatility model, given the whole path of the variance the log-price is Gaussian, and the Black–Scholes formula gives the conditional price (Hull and White, 1987; Romano and Touzi, 1997; Willard, 1997).

Proposition 10.12 (Mixing: conditioning on the volatility path)**.** *Let W and $W^{\perp}$ be independent Brownian motions under $\mathbb{Q}$, $(\mathcal{F}^W_t)$ the filtration generated by W, and $V \ge 0$ an $(\mathcal{F}^W_t)$-progressively measurable process with $Q_T = \int_0^T V_t \,\mathrm{d}t < \infty$ almost surely. Let $\rho \in [-1, 1]$, $\bar{\rho} = \sqrt{1 - \rho^2}$, $I_T = \int_0^T \sqrt{V_t}\,\mathrm{d}W_t$, and*

$$S_T = S \exp\Big((r - y)T - \tfrac{1}{2} Q_T + \rho I_T + \bar{\rho} \int_0^T \sqrt{V_t}\,\mathrm{d}W^{\perp}_t \Big),$$

the solution of $\mathrm{d}S_t = S_t\big((r - y)\,\mathrm{d}t + \sqrt{V_t}(\rho\,\mathrm{d}W_t + \bar{\rho}\,\mathrm{d}W^{\perp}_t)\big)$. Set

$$\tilde{S} = S \exp\left(\rho I_T - \tfrac{1}{2}\rho^2 Q_T\right), \qquad \tilde{\sigma} = \bar{\rho}\sqrt{Q_T/T}.$$

(i) There is a standard normal N, independent of $\mathcal{F}^W_T$, with $\ln S_T = \ln \tilde{S} + (r - y - \tfrac{1}{2}\tilde{\sigma}^2)T + \tilde{\sigma}\sqrt{T}N$.

(ii) For every Borel f with $\mathbb{E}|f(S_T)| < \infty$, $\mathbb{E}[f(S_T) \mid \mathcal{F}_T^W] = \int_{\mathbb{R}} f(\tilde{S}e^{(r-y-\tilde{\sigma}^2/2)T+\tilde{\sigma}\sqrt{T}z})N'(z)\,\mathrm{d}z$. In particular the conditional price of a call is the Black–Scholes price at spot $\tilde{S}$ and volatility $\tilde{\sigma}$, $e^{-rT}\mathbb{E}[(S_T-K)^+ \mid \mathcal{F}_T^W] = C_{\mathrm{BS}}(\tilde{S},K,T,r,y,\tilde{\sigma})$, and likewise for a put; when $\tilde{\sigma}=0$ (for $\rho=\pm1$ or $Q_T=0$) it is the discounted payoff at the forward $\tilde{S}e^{(r-y)T}$.

*(iii) (*Hull and White.*) If $\rho=0$, the call price is $\mathbb{E}[C_{\mathrm{BS}}(S,K,T,r,y,\sqrt{Q_T/T})]$: a mixture of Black–Scholes prices over the law of the realised volatility (Exercise 10.9 expands it for small volatility of variance).*

*(iv) (*On a grid.*) If $\ln S$ is advanced by the log-Euler step $\Delta\ln S = (r-y-\frac{1}{2}V_{t_k})\Delta_k + \sqrt{V_{t_k}}(\rho\Delta W_k + \bar{\rho}\Delta W_k^{\perp})$ with V_{t_k} functions of W only, (i) and (ii) hold exactly with Q_T and I_T replaced by $\sum_k V_{t_k}\Delta_k$ and $\sum_k \sqrt{V_{t_k}}\Delta W_k$.*

Proof. (i) Let $\mathcal{G}_t = \mathcal{F}_T^W \vee \sigma(W_s^{\perp} : s \le t)$, completed. Since $W^{\perp}$ is independent of W and has independent increments, $W_t^{\perp} - W_s^{\perp}$ is independent of $\mathcal{G}_s$, so $W^{\perp}$ is a $(\mathcal{G}_t)$-Brownian motion. Let $h_t = \sqrt{V_t/Q_T}$ on $\{Q_T>0\}$ and $h_t = 1/\sqrt{T}$ on $\{Q_T=0\}$. It is measurable in (t,ω) with respect to $\mathcal{B}([0,T])\otimes\mathcal{F}_T^W$, hence $(\mathcal{G}_t)$-progressive, and $\int_0^T h_t^2\,\mathrm{d}t = 1$. Let $N = \int_0^T h\,\mathrm{d}W^{\perp}$ in the filtration $(\mathcal{G}_t)$. For $u\in\mathbb{R}$, Itô's formula shows that $M_t = \exp\left(iu\int_0^t h\,\mathrm{d}W^{\perp} + \frac{u^2}{2}\int_0^t h^2\,\mathrm{d}s\right)$ is a local martingale; it is bounded by $e^{u^2/2}$, hence a martingale, and $\mathbb{E}[M_T \mid \mathcal{G}_0] = 1$, that is, $\mathbb{E}[e^{iuN} \mid \mathcal{G}_0] = e^{-u^2/2}$. A deterministic conditional characteristic function means that N is standard normal and independent of $\mathcal{G}_0 \supset \mathcal{F}_T^W$. Finally, $\int_0^T \sqrt{V}\,\mathrm{d}W^{\perp} = \sqrt{Q_T}\,N$ almost surely: $\sqrt{V} = \sqrt{Q_T}\,h$, the $\mathcal{G}_0$-measurable factor $\sqrt{Q_T}$ can be taken out of the integral (check it on simple integrands), and the integral of the $(\mathcal{F}_t^W \vee \mathcal{F}_t^{W^{\perp}})$-progressive process $\sqrt{V}$ is the same in either filtration, since $W^{\perp}$ is a Brownian motion for both and the integral is the limit in probability of the integrals of the same simple processes. Substitute in $\ln S_T$: $\bar{\rho}\sqrt{Q_T} = \tilde{\sigma}\sqrt{T}$ and $-\frac{1}{2}Q_T + \rho I_T = \ln(\tilde{S}/S) - \frac{1}{2}\tilde{\sigma}^2 T$.

(ii) $\tilde{S}$ and $\tilde{\sigma}$ are $\mathcal{F}_T^W$-measurable and N is independent of $\mathcal{F}_T^W$; apply the independence lemma (Shreve, 2004, Lemma 2.3.4). For the call, the integral is the Black–Scholes expectation with spot $\tilde{S}$ and volatility $\tilde{\sigma}$. (iii) is (ii) with $\tilde{S}=S$ and $\tilde{\sigma}=\sqrt{Q_T/T}$, after taking expectations. (iv) Given $\mathcal{F}_T^W$ the sum $\sum_k \sqrt{V_{t_k}}\Delta W_k^{\perp}$ is a combination, with $\mathcal{F}_T^W$-measurable coefficients, of the independent $\Delta W_k^{\perp} \sim \mathcal{N}(0,\Delta_k)$, hence conditionally $\mathcal{N}(0,\sum_k V_{t_k}\Delta_k)$; the rest is as before. □

The conditional estimator averages $C_{\mathrm{BS}}(\tilde{S},K,T,r,y,\tilde{\sigma})$ over simulated variance paths and needs no normal for $W^{\perp}$. By Theorem 10.11 it removes the part of the variance due to $W^{\perp}$, which is large when $|\rho|$ is small. At $\rho=\pm1$ there is nothing left to integrate. In the Heston model (Theorem 19.1), $\mathrm{d}V_t = \kappa(\theta - V_t)\,\mathrm{d}t + \xi\sqrt{V_t}\,\mathrm{d}W_t$; this equation has a pathwise unique strong solution, adapted to the filtration of W (Karatzas and Shreve, 1991, Sections 5.2 and 5.3), and integrating it gives

$$I_T = \frac{1}{\xi}(V_T - V_0 - \kappa\theta T + \kappa Q_T),$$

so the conditional estimator needs only V_T and Q_T (Willard, 1997; Broadie and Kaya, 2006).

A laboratory. Take the test case with $V_0=\theta=0.04$, $\kappa=2$ and $\xi=0.3$, and the at-the-money put, whose payoff is bounded, so that every variance is finite whatever the moments of S_T. We sample V exactly at 64 equal steps, by noncentral chi-square transitions (Glasserman, 2003,

Section 3.4.1), take Q_T by the trapezoidal rule and I_T from the identity above. The plain estimator draws one more normal per path and builds S_T from (i); the conditional estimator uses (ii) on the same variance paths. Both estimate the same discretised price, and the conditional one is the Rao–Blackwellisation of the other. The Fourier prices of Theorem 19.7 are 8.1534 for $\rho = 0$ and 7.9480 for $\rho = -0.7$. Over 40 independent runs of 5000 paths (our implementation, numpy `default_rng`, seeds 777 to 816), the errors of the conditional estimate divided by its standard error have mean 0.16 and standard deviation 1.04 for $\rho = 0$, and 0.15 and 1.01 for $\rho = -0.7$: no discretisation bias is visible, and the standard errors are honest. The standard-error ratio of the conditional to the plain estimator is 0.163 for $\rho = 0$ and 0.728 for $\rho = -0.7$ (4×10^5 paths, Table 10.1); the cost per path is dominated by the 64 variance steps and is the same for both, so the efficiencies are about 37 and 1.9. Figure 10.4 shows why: for $\rho = 0$ the conditional value is a smooth function of the realised variance, while the plain payoff scatters around it.

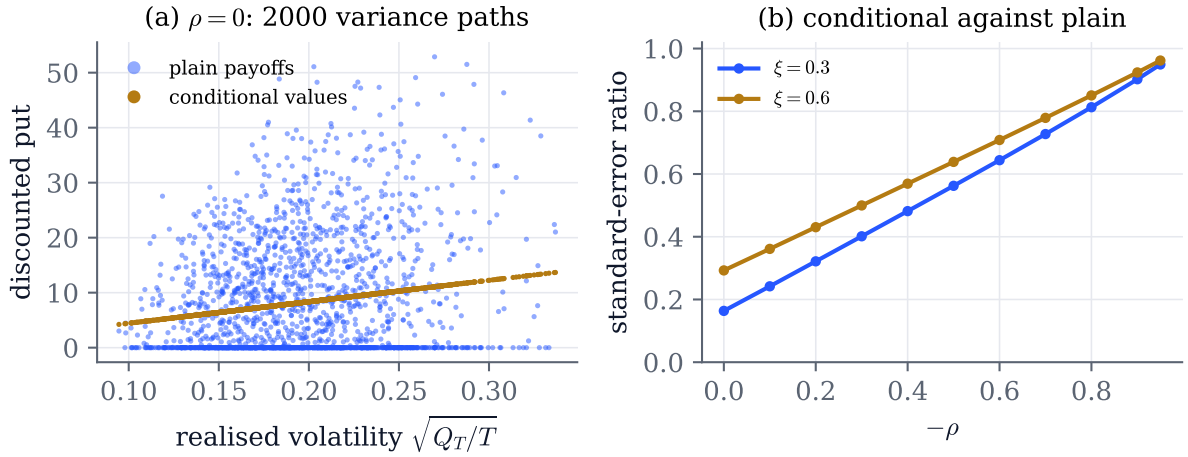

Figure 10.4: Conditional Monte Carlo for the at-the-money put in the Heston model ($V_0 = \theta = 0.04$, $\kappa = 2$, 64 exact variance steps). (a) $\rho = 0$, $\xi = 0.3$: plain payoffs and conditional values $P_{\mathrm{BS}}(S, K, T, r, y, \sqrt{Q_T/T})$ for 2000 variance paths, against the realised volatility. (b) Standard-error ratio of the conditional to the plain estimator against $-\rho$, from 5×10^4 variance paths (the same for every ρ; seed 1031), for $\xi = 0.3$ and $\xi = 0.6$.

▶ Interactive demo: conditional Monte Carlo in the Heston model: move ρ and ξ and watch the plain payoffs collapse onto the conditional values.

10.5 Laboratory

Table 10.1 collects the methods of the chapter. One-dimensional rows use $n = 2 \times 10^5$ draws (the test-case call) and $n = 10^5$ (the course's deep out-of-the-money call); the exact ratios come from Theorems 10.1, 10.3 and 10.6 with stratum moments by quadrature. The ratios are produced by `stratified.laboratory()` (numpy `default_rng`, seed 10). The costs are medians of wall-clock times of the whole estimator against its plain competitor, measured by `laboratory_costs()` on our machine: they vary by tens of percent from run to run and more across machines, and the notebook measures them on yours.

target	method	exact	measured	cost	efficiency
call	10 strata, proportional	0.317	0.318	4.5	2.2
	10 strata, Neyman	0.153	0.152	5	8.6
	10 strata, post-stratified	0.317	0.311	3	3.3
	100 strata, proportional	0.090	0.085	4.5	27
deep OTM call	10 strata, proportional	0.925	0.939	4.5	0.26
	100 strata, proportional	0.398	0.386	4.5	1.4
Asian call	64 strata of W_T (bridge)		0.530	1.6	2.2
	64 strata of the first increment		0.917	1.2	1.0
	64 post-strata of W_T		0.530	1.5	2.4
	64 strata of $\sum_j W_{t_j}$		0.108	1.5	57
	Latin hypercube, increments		0.54 ± 0.01	4.2	0.8
	Latin hypercube, bridge		0.354 ± 0.006	4.8	1.7
Heston put	conditional, $\rho = 0$		0.163	1.0	37
	conditional, $\rho = -0.7$		0.728	1.0	1.9

Table 10.1: Standard-error ratios ϱ to plain Monte Carlo at equal n, exact and measured; cost per draw relative to the plain estimator (NumPy normals; increments construction for the Asian call), on our machine; and the efficiency $1/(\varrho^2 \times \text{cost})$, from the exact ratio where there is one. Measured ratios divide the method's standard error by $\sigma/\sqrt{n}$, with σ exact for the calls, from 2^{20} plain paths for the Asian call, and from the plain estimator on the same paths for the Heston put. The Asian rows use $n = 2^{16}$ paths, Latin hypercube 2000 designs of 2^{10} points; the Heston rows use 4×10^5 variance paths, whose simulation dominates the cost of both estimators.

Costs change the picture. Stratification by inversion pays in one dimension only with enough strata, and loses for the deep out-of-the-money call with 10 strata, whose equiprobable strata are wasted where the payoff vanishes; post-stratification keeps the fast generator. Latin hypercube sampling needs 12 inversions per path and barely covers that cost here, while stratifying the right direction costs little and gains a factor 57. Conditioning costs one Black–Scholes evaluation per path and helps as much as the integrated-out noise contributes.

Course examples reproduced. The stratified-sampling experiment of the course's notebook, with its exact variances; the call price 7.36429; and the course's estimate 4.56058 of its Asian call, confirmed within its standard error.

Further reading

Glasserman (2003) is the reference: Section 4.3.1 for stratified sampling and optimal allocation, Section 4.3.2 for stratified paths, Section 4.3.3 for post-stratification and Section 4.4 for Latin hypercube sampling; Asmussen and Glynn (2007), Sections V.4 and V.7, treat conditional Monte Carlo and stratification. Neyman (1934) introduced the optimal allocation; Cochran (1977), Chapter 5, is the classical account of stratified sampling. Latin hypercube sampling is due to McKay et al. (1979); its asymptotics are in Stein (1987) and Owen (1992), and Owen (2013), Chapter 10, gives a modern account. The mixing formula goes back to Hull and White (1987) for

independent volatility; Romano and Touzi (1997) treat correlated volatility, and Willard (1997) turned it into a conditional Monte Carlo method.

Chapter 11

Quasi-Monte Carlo

Monte Carlo integrates with random points, and its error $\sigma/\sqrt{n}$ (Theorem 6.4) is the price of that randomness: independent points leave gaps and clusters by chance. *Quasi-Monte Carlo* (QMC) replaces them by deterministic points built to fill the unit cube as evenly as possible. For integrands that are smooth enough, the error then decreases almost like $1/n$ instead of $1/\sqrt{n}$, in any fixed dimension. The catch lies in "smooth enough" and "fixed dimension": option payoffs have kinks and jumps, the map from uniforms to normals is unbounded, and a path has as many dimensions as it has dates. This chapter separates what is proved from what is observed, and measures the rest.

Section 11.1 measures evenness by the discrepancy and bounds the error by the Koksma–Hlawka inequality. Section 11.2 builds the classical low-discrepancy sequences of van der Corput, Halton and Sobol'. Section 11.3 randomises them, which restores unbiasedness and the error estimates of Chapter 6. Section 11.4 explains, through the effective dimension, why the Brownian-bridge and principal-component constructions of Chapter 7 make QMC so effective for Asian options.

Throughout, λ is the volume (Lebesgue measure) on $[0,1)^d$, $I(F) = \int_{[0,1)^d} F(u)\,\mathrm{d}u$, and for points $P = (x_1, \dots, x_n)$ of $[0,1)^d$ (repetitions allowed), $Q_P(F) = \frac{1}{n}\sum_{i=1}^{n} F(x_i)$ is the equal-weight rule. We write $D = \{1, \dots, d\}$; for $u \subseteq D$, $x_u = (x_j)_{j\in u}$, and $(x_u{:}1)$ is the point whose j-th coordinate is x_j for $j \in u$ and 1 for $j \notin u$.

11.1 Discrepancy and the Koksma–Hlawka inequality

Let $Z \sim \mathcal{N}(0, I_d)$ and let $f(Z)$ be a discounted payoff, with one coordinate per date or per asset. If U is uniform on $(0,1)^d$, the vector $N^{-1}(U) = (N^{-1}(U_1), \dots, N^{-1}(U_d))$ has independent standard normal coordinates (Theorem 5.8), so

$$\mathbb{E}[f(Z)] = \int_{(0,1)^d} F(u)\,\mathrm{d}u, \qquad F = f \circ N^{-1}. \tag{11.1}$$

Monte Carlo evaluates $Q_P(F)$ at independent uniform points; QMC chooses the points. The dimension d is the number of uniforms that one sample consumes, so it must be fixed: samplers that consume a random number of uniforms, such as rejection, are excluded. The map from

uniforms to normals should also be monotone in each coordinate, as inversion is, so that evenly spread uniforms give evenly spread normals.

Definition 11.1 (Discrepancy)**.** Let $P = (x_1, \dots, x_n)$ be points of $[0,1)^d$. For $a \in [0,1]^d$, let $A_P(a) = \#\{i : x_i \in [0,a)\}$ count the points in the anchored box $[0,a) = \prod_{j=1}^d [0,a_j)$, and let

$$\delta_P(a) = \frac{A_P(a)}{n} - \lambda([0,a)) = \frac{A_P(a)}{n} - \prod_{j=1}^d a_j$$

be the *local discrepancy*. The *star discrepancy* of P is $D_n^*(P) = \sup_{a\in[0,1]^d} |\delta_P(a)|$. For an infinite sequence, D_n^* is the star discrepancy of its first n points. The sequence is *uniformly distributed* if $D_n^* \to 0$, and it is a *low-discrepancy sequence* if $D_n^* = O((\log n)^d/n)$.

The local discrepancy is the integration error of Q_P on the indicator of a box: $\delta_P(a) = Q_P(\mathbf{1}_{[0,a)}) - I(\mathbf{1}_{[0,a)})$. In one dimension, $D_n^*(P) = \sup_x |\hat{F}_n(x) - x|$ is the Kolmogorov–Smirnov distance between the empirical distribution function $\hat{F}_n$ of the points and the uniform law, the statistic of the test used in Chapter 5; for independent uniform points $\sqrt{n}\, D_n^*$ converges in law to Kolmogorov's distribution (Knuth, 1997, Section 3.3.1.B), so random points have discrepancy of order $n^{-1/2}$. No point set does better than order $1/n$: in one dimension $D_n^* \geq \frac{1}{2n}$ (Exercise 11.1), and in dimension $d \geq 2$ every set of n points has $D_n^* \geq c_d (\log n)^{(d-1)/2}/n$ for a constant $c_d > 0$ (Roth, 1954). For infinite sequences in $[0,1)$, $nD_n^* \geq c \log n$ for infinitely many n (Schmidt, 1972).

Variation and the error bound

The error bound needs a notion of variation. Call $f : [0,1]^d \to \mathbb{R}$ *smooth* if all its mixed partial derivatives $\partial^u f = \partial^{|u|} f / \prod_{j\in u} \partial x_j$, $u \subseteq D$, exist and are continuous on the closed cube. For smooth f, the *variation in the sense of Hardy and Krause* is

$$V_{HK}(f) = \sum_{\emptyset \neq u \subseteq D} \int_{[0,1]^{|u|}} \left|\partial^u f(a_u{:}1)\right| \mathrm{d}a_u, \tag{11.2}$$

one term per face of the cube that contains the corner $(1, \dots, 1)$. In one dimension it is the total variation $\int_0^1 |f'|$. The general definition, through sums of alternating differences over grids, agrees with (11.2) for smooth f (Owen, 2005).

Lemma 11.2 (Expansion from the corner)**.** *If f is smooth, then for every $x \in [0,1]^d$*

$$f(x) = \sum_{u \subseteq D} (-1)^{|u|} \int_{[0,1]^{|u|}} \mathbf{1}\{x_u < a_u\}\, \partial^u f(a_u{:}1)\, \mathrm{d}a_u, \tag{11.3}$$

where $\mathbf{1}\{x_u < a_u\} = \prod_{j\in u} \mathbf{1}\{x_j < a_j\}$ and the term $u = \emptyset$ is $f(1, \dots, 1)$. In particular $|f(x) - f(1, \dots, 1)| \leq V_{HK}(f)$.

Proof. For a function g of one variable with a continuous derivative, $g(x) = g(1) - \int_x^1 g'(a)\, \mathrm{d}a = g(1) - \int_0^1 \mathbf{1}\{x < a\}\, g'(a)\, \mathrm{d}a$. Apply this in x_1 to f, with the other coordinates fixed. Each of the two resulting terms is a function of $x_2, \dots, x_d$ (integrated against a_1 in the second term) whose partial derivative in x_2 is continuous; apply the identity in x_2 to it, under the integral sign, and so on. After the d-th step, each coordinate j has either been set to 1 or replaced by

$-\int_0^1 \mathbf{1}\{x_j < a_j\}(\cdot)\,\mathrm{d}a_j$ with one more derivative ∂_j, which is (11.3). All integrands are continuous on compact sets, so the iterated integrals are integrals over $[0,1]^{|u|}$ (Fubini). The bound follows by taking absolute values in the terms $u \neq \emptyset$. □

Theorem 11.3 (Koksma–Hlawka). *Let $f : [0,1]^d \to \mathbb{R}$ have bounded variation $V_{HK}(f)$ in the sense of Hardy and Krause. For all points $P = (x_1, \dots, x_n)$ of $[0,1)^d$,*

$$|Q_P(f) - I(f)| \le V_{HK}(f)\, D_n^*(P). \tag{11.4}$$

If f is smooth, the error has the exact representation

$$Q_P(f) - I(f) = \sum_{\emptyset \neq u \subseteq D} (-1)^{|u|} \int_{[0,1]^{|u|}} \delta_P(a_u{:}1)\, \partial^u f(a_u{:}1)\, \mathrm{d}a_u. \tag{11.5}$$

The inequality is due to Koksma for $d = 1$ and to Hlawka in general; see Niederreiter (1992), Chapter 2. We prove it for smooth f, which shows where the variation comes from.

Proof for smooth f. Average (11.3) over the points, and integrate it over x. For $j \notin u$ the points satisfy $x_{ij} < 1$, so $\frac{1}{n}\sum_i \mathbf{1}\{x_{i,u} < a_u\} = A_P(a_u{:}1)/n$, while $\int \mathbf{1}\{x_u < a_u\}\,\mathrm{d}x = \prod_{j\in u} a_j = \lambda([0,(a_u{:}1)))$. The terms $u = \emptyset$ cancel, and the difference of the other terms is (11.5) (the sums are finite and the integrands bounded, so averaging commutes with the integrals). Each $[0,(a_u{:}1))$ is an anchored box, so $|\delta_P(a_u{:}1)| \le D_n^*(P)$, and (11.4) follows with $V_{HK}(f)$ given by (11.2). □

The bound separates the integrand from the points: the variation measures the integrand, the discrepancy the points. In one dimension it cannot be improved: the indicator $\mathbf{1}_{[0,a)}$ has variation 1 and error $\delta_P(a)$, whose supremum over a is $D_n^*(P)$.

Remark 11.4 (What the inequality says about option payoffs). Very little, for three reasons.

(a) A function of bounded Hardy–Krause variation is bounded: Theorem 11.2 gives $|f(x) - f(1,\dots,1)| \le V_{HK}(f)$ for smooth f, and the same inclusion–exclusion argument, with differences in place of integrals, gives it in general. The integrand $F = f \circ N^{-1}$ of (11.1) is unbounded as soon as the payoff is unbounded in Z (the stock, a call, an Asian call), so $V_{HK}(F) = \infty$ and (11.4) is empty.

(b) Bounded payoffs are not safe either. For a hyperplane $a \cdot x = c$ that cuts the cube, a jump $\mathbf{1}\{a \cdot x > c\}$ has finite variation if only one a_j is nonzero (a digital in one coordinate has variation 1) and infinite variation as soon as two are; a kink $(a \cdot x - c)^+$ has finite variation if at most two a_j are nonzero (its mixed differences in those two coordinates have the constant sign of their product and telescope) and infinite variation as soon as three are (Owen, 2005); the cases $\mathbf{1}\{x_1 + x_2 \le 1\}$, $(x_1 + x_2 - 1)^+$ and $(x_1 + x_2 + x_3 - 1)^+$ are Exercise 11.3. A digital on an average of two or more dates, or a put on an average of three or more, jumps or kinks across a curved surface of this kind after the map N^{-1}. A put on the average of two dates has infinite variation too, but because of N^{-1} itself: with $u_j = N(z_j)$, the term $u = \{1,2\}$ of (11.2) becomes $\int |\partial_{z_1}\partial_{z_2} f|\,\mathrm{d}z_1\mathrm{d}z_2$ with Lebesgue measure on $\mathbb{R}^2$, which diverges as $z_1 \to -\infty$.

(c) Even for smooth integrands the bound is hard to use: V_{HK} involves $2^d - 1$ integrals of mixed derivatives, computing D_n^* exactly is infeasible beyond small n and d, and $(\log n)^d/n$ falls below $n^{-1/2}$ only when $n^{1/2} > (\log n)^d$, which for $d = 10$ and $n \ge 3$ means $n > 10^{39}$.

What does follow without any variation is convergence.

Proposition 11.5 (Convergence without a rate)**.** *Let $x_1, x_2, \dots$ be uniformly distributed in $[0,1)^d$, and let $F : [0,1)^d \to \mathbb{R}$ be Riemann integrable: for every $\varepsilon > 0$ there are step functions $g \le F \le h$ (finite linear combinations of indicators of boxes $[a,b) \subseteq [0,1)^d$) with $I(h-g) \le \varepsilon$. Then $Q_n(F) \to I(F)$, where Q_n uses the first n points.*

Proof. A box $B = [a,b) \subseteq [0,1)^d$ is a signed combination of 2^d anchored boxes, by inclusion–exclusion: $\mathbf{1}_B = \sum_{v \subseteq D} (-1)^{|v|} \mathbf{1}_{[0,c^v)}$ with $c^v_j = a_j$ for $j \in v$ and $c^v_j = b_j$ otherwise. Hence $|Q_n(\mathbf{1}_B) - \lambda(B)| \le 2^d D^*_n \to 0$, and $Q_n(g) \to I(g)$ for every step function g. With $g \le F \le h$ as in the statement, $\limsup_n Q_n(F) \le \lim_n Q_n(h) = I(h) \le I(F) + \varepsilon$ and likewise $\liminf_n Q_n(F) \ge I(F) - \varepsilon$. Let $\varepsilon \downarrow 0$. □

A bounded payoff whose discontinuities lie on a smooth surface, such as a digital on an average after the map N^{-1}, is Riemann integrable: on a fine grid, the cells that meet the closure of the surface have small total volume, and on the others the oscillation of F is small. So QMC converges for it, at a rate that no theorem gives. For an unbounded $F \ge 0$, uniform distribution alone does not even give convergence: replacing the points of indices $n = 2^k$ by points so close to a point where F is unbounded that $F(x_n) \ge n^2$ changes D^*_n by at most $(\log_2 n + 1)/n$, but makes $Q_n(F) \ge n$ along $n = 2^k$. In one dimension a truncation argument recovers a rate for monotone integrands (Exercise 11.4), and Owen (2006) gives conditions on the growth of the mixed derivatives near the boundary of the cube under which QMC keeps a rate close to $1/n$; they require mixed derivatives, which kinked payoffs do not have. Randomised QMC (Section 11.3) needs only $F \in L^2$.

11.2 Low-discrepancy sequences

11.2.1 Van der Corput and Halton

Let $b \ge 2$ be an integer. Every integer $i \ge 0$ has a unique expansion $i = \sum_{k \ge 0} a_k(i)\, b^k$ with digits $a_k(i) \in \{0, \dots, b-1\}$, finitely many of them nonzero. The *radical inverse* mirrors the digits at the radix point,

$$\phi_b(i) = \sum_{k \ge 0} a_k(i)\, b^{-k-1} \in [0,1),$$

and the *van der Corput sequence* in base b is $\phi_b(0), \phi_b(1), \phi_b(2), \dots$; in base 2 it starts $0, \frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{5}{8}, \frac{3}{8}, \frac{7}{8}$. The *Halton sequence* (Halton, 1960) uses one base per coordinate, $x_i = (\phi_{b_1}(i), \dots, \phi_{b_d}(i))$, with pairwise coprime bases, usually the first d primes.

Proposition 11.6 (Van der Corput and Halton)**.** *Let $b \ge 2$.*

(i) For every $k \ge 0$, any b^k consecutive points $\phi_b(i)$, $j \le i < j + b^k$, lie one in each interval $[l\, b^{-k}, (l+1)\, b^{-k})$, $0 \le l < b^k$. The first b^k points are exactly the multiples of b^{-k}.

(ii) For every $n \ge 1$, the first n points of the van der Corput sequence satisfy

$$D^*_n \le \frac{(b-1)\big(\lfloor \log_b n \rfloor + 1\big)}{n},$$

and $D^*_n = 1/n$ when n is a power of b.

(iii) Let $b_1, \dots, b_d$ be pairwise coprime and $k_1, \dots, k_d \geq 0$. Any $B = \prod_j b_j^{k_j}$ consecutive points of the Halton sequence lie one in each box $\prod_j [l_j b_j^{-k_j}, (l_j+1) b_j^{-k_j})$.

Proof. (i) The tail $\sum_{r \geq k} a_r(i) b^{-r-1}$ is smaller than $\sum_{r \geq k} (b-1) b^{-r-1} = b^{-k}$, because the digits of an integer are eventually 0. Hence the interval that contains $\phi_b(i)$ has index $\lfloor b^k \phi_b(i) \rfloor = \sum_{r<k} a_r(i)\, b^{k-1-r}$: the first k digits of i in reverse order. It depends on i only through i mod $b^k = \sum_{r<k} a_r(i) b^r$, and digit reversal is a bijection of $\{0, \dots, b^k - 1\}$. Consecutive integers $j, \dots, j + b^k - 1$ take every residue modulo b^k once. For $i < b^k$ the tail vanishes, so $\phi_b(i) = l\, b^{-k}$.

(ii) Let $K = \lfloor \log_b n \rfloor$ and $n = \sum_{k=0}^{K} c_k b^k$ in base b. Split the indices $0, \dots, n-1$ into consecutive blocks: c_K blocks of length b^K, then c_{K-1} of length b^{K-1}, and so on down to c_0 blocks of length 1; there are $\sum_k c_k \leq (b-1)(K+1)$ blocks. Fix $a \in [0,1]$ and a block of length b^k. By (i) it has one point in each interval of length b^{-k}; $[0, a)$ contains the $\lfloor a b^k \rfloor$ intervals to the left of a and meets at most one more, so the block puts $\lfloor a b^k \rfloor$ or $\lfloor a b^k \rfloor + 1$ points in $[0, a)$, which differs from $a b^k$ by less than 1. Summing over blocks, $|A_P(a) - na| < (b-1)(K+1)$; divide by n. When $n = b^k$ the points are l/n, $A_P(a) = \lceil an \rceil$, and $\delta_P(a) = (\lceil an \rceil - an)/n$ takes values in $[0, 1/n)$ with supremum $1/n$.

(iii) By (i), the box that contains x_i is determined by the residues i mod $b_j^{k_j}$, $j = 1, \dots, d$, through digit reversals, which are bijections. The moduli $b_j^{k_j}$ are pairwise coprime, so by the Chinese remainder theorem i mod $B \mapsto (i \bmod b_j^{k_j})_{j \leq d}$ is a bijection. B consecutive integers take every residue modulo B once. □

Part (ii) is of the best possible order for a sequence in one dimension, by Schmidt's theorem; sharper constants are in Niederreiter (1992), Chapter 3. In base 2, the largest value of $n D_n^*$ for $n \leq 4096$ is 4.78, at $n = 3413$, where (ii) allows 12. Part (iii), with the decomposition into blocks of (ii) in each coordinate, gives the bound

$$D_n^* < \frac{d}{n} + \frac{1}{n} \prod_{j=1}^{d} \Big(\frac{b_j - 1}{2 \log b_j} \log n + \frac{b_j + 1}{2} \Big)$$

for the Halton sequence (Niederreiter, 1992, Chapter 3): a low-discrepancy sequence, but with a constant $\prod_j (b_j - 1)/(2 \log b_j)$ that grows superexponentially in d when the b_j are the first d primes. The reason is visible in the coordinates with large bases. For $i < b$, $\phi_b(i) = i/b$: the coordinate creeps up in b equal steps. Two coordinates with nearby large bases creep up together, and the first points of their projection lie on a few lines (Figure 11.1, right; Exercise 11.5). Randomly permuting the digits in each base, as the scramble of Section 11.3 does in base 2, breaks the lines; shifting the points does not.

11.2.2 Nets and the Sobol' sequence

Uniformity in all coordinates at once is easiest to organise in a single base. The following notions are due to Sobol' and, in this generality, to Niederreiter (Niederreiter, 1992, Chapter 4).

Definition 11.7 (Nets and (t, d)-sequences)**.** Let $b \geq 2$ and $0 \leq t \leq m$ be integers. An *elementary interval* in base b is a box $E = \prod_{j=1}^{d} [l_j b^{-k_j}, (l_j + 1) b^{-k_j})$ with integers $k_j \geq 0$

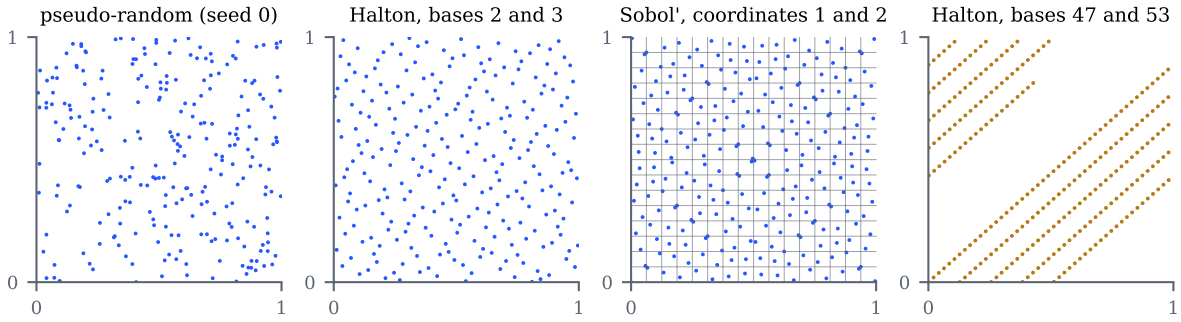

Figure 11.1: The first 256 points of four sequences, projected on two coordinates. Pseudo-random points (numpy `default_rng`, seed 0) leave gaps and clusters. The Halton points in bases 2 and 3 and the Sobol' points in coordinates 1 and 2 fill the square evenly; the grey grid shows that each of the 16×16 squares of side $\frac{1}{16}$ holds exactly one Sobol' point. The Halton coordinates 15 and 16, with bases 47 and 53, lie on a few parallel lines.

and $0 \le l_j < b^{k_j}$; its volume is $b^{-\sum_j k_j}$. A (t, m, d)-*net* in base b is a set of $n = b^m$ points of $[0,1)^d$ such that every elementary interval of volume b^{t-m} contains exactly b^t points. A sequence $x_0, x_1, \dots$ is a (t, d)-*sequence* in base b if, for every $m \ge t$ and $q \ge 0$, the points x_i with $qb^m \le i < (q+1)b^m$ form a (t, m, d)-net.

A net with $t = 0$ is balanced in every elementary interval of volume $1/n$: every square of side $\frac{1}{16}$ in Figure 11.1, but also every rectangle of $\frac{1}{2} \times \frac{1}{128}$ or $\frac{1}{256} \times 1$. By Theorem 11.6(i), the van der Corput sequence in base b is a $(0, 1)$-sequence in base b.

Definition 11.8 (Sobol' sequence)**.** For each coordinate $j \ge 2$, take a primitive polynomial over the field $\mathbb{F}_2$ of degree $s = s_j$, $p_j(z) = z^s + c_1 z^{s-1} + \cdots + c_{s-1} z + 1$, a different one for each coordinate, and odd integers $m_1, \dots, m_s$ with $m_k < 2^k$. Extend them by

$$m_k = 2c_1 m_{k-1} \oplus 2^2 c_2 m_{k-2} \oplus \cdots \oplus 2^{s-1} c_{s-1} m_{k-s+1} \oplus 2^s m_{k-s} \oplus m_{k-s}, \qquad k > s,$$

where $\oplus$ is the bitwise exclusive or of binary expansions, and define the *direction numbers* $v_{k,j} = m_k / 2^k$. For coordinate 1, $s_1 = 1$ and $m_k = 1$ for all k. The i-th point of the *Sobol' sequence*, for $i = \sum_{k \ge 1} i_k 2^{k-1}$ with bits $i_k \in \{0, 1\}$, has coordinates

$$x_{i,j} = i_1 v_{1,j} \oplus i_2 v_{2,j} \oplus i_3 v_{3,j} \oplus \cdots, \qquad j = 1, \dots, d.$$

The recursion keeps every m_k odd and smaller than 2^k: the term m_{k-s} is odd, all the others are even, and each is smaller than 2^k. Coordinate 1 is the van der Corput sequence in base 2, since $x_{i,1} = \sum_k i_k 2^{-k} = \phi_2(i)$. The construction is due to Sobol' (1967); its quality depends on the choice of the initial numbers m_k. We use those of Joe and Kuo (2008), chosen to improve the two-dimensional projections, for the first 128 coordinates (see also Glasserman (2003), Section 5.2.3).

Proposition 11.9 (Sobol' points form digital nets)**.** *Fix $m \ge 1$, and for each coordinate j let C_j be the $m \times m$ matrix over $\mathbb{F}_2$ whose k-th column holds the first m binary digits of $v_{k,j}$.*

(i) For $0 \le i < 2^m$, $x_{i,j}$ has at most m binary digits, and they are $C_j(i_1, \dots, i_m)^\top$ (arithmetic

modulo 2). Each C_j is upper triangular with ones on the diagonal. Consequently, in every block $q2^m \le i < (q+1)2^m$, each coordinate has one point in each interval $[l2^{-m}, (l+1)2^{-m})$.

(ii) Every block $q2^m \le i < (q+1)2^m$ is a (t, m, d)-net in base 2 if and only if, for every $(k_1, \dots, k_d)$ with $k_j \ge 0$ and $\sum_j k_j = m - t$, the first k_1 rows of C_1, the first k_2 rows of C_2, ..., and the first k_d rows of C_d are linearly independent over $\mathbb{F}_2$.

(iii) If the polynomials p_j are distinct, the Sobol' sequence is a (t, d)-sequence in base 2 with $t \le \sum_{j=1}^{d} (s_j - 1)$.

Proof. (i) The bitwise exclusive or adds binary digits modulo 2 without carries, so digit r of $x_{i,j}$ is $\sum_k i_k$ (digit r of $v_{k,j}$) modulo 2. Since m_k is odd and $m_k < 2^k$, the digits of $v_{k,j} = m_k / 2^k$ beyond the k-th are 0 and the k-th is 1: column k of C_j vanishes below row k and has a 1 in row k. For $i < 2^m$ only $i_1, \dots, i_m$ can be nonzero, which gives the first claim. In the block of index q, write $i = i' + q2^m$ with $i' < 2^m$; the bits of $q2^m$ contribute the same vector e_j to the first m digits of every point, so those digits are $C_j (i'_1, \dots, i'_m)^\top + e_j$. As i' runs over $\{0, \dots, 2^m - 1\}$, this runs over all 2^m digit strings, because C_j is invertible.

(ii) A point lies in the elementary interval with parameters (k_j, l_j) if and only if, for each j, the first k_j digits of its j-th coordinate are the k_j digits of l_j. By (i) this is a system of $\sum_j k_j = m - t$ linear equations over $\mathbb{F}_2$ in the m unknown bits of i', whose matrix M stacks the first k_j rows of each C_j, and whose right-hand side is the vector of digits of the l_j minus the shift of (i); as the l_j vary, it runs over all of $\mathbb{F}_2^{m-t}$. If the rows of M are independent, M is onto and every system has exactly $2^{m-(m-t)} = 2^t$ solutions. If they are dependent, M is not onto, so some right-hand side has no solution: some elementary interval of volume 2^{t-m} is empty, and the block is not a (t, m, d)-net.

(iii) is due to Sobol' (1967); see Dick and Pillichshammer (2010), Chapter 8. □

By (ii), the exact t-value of the first 2^m points is the smallest t that passes the rank test, which the code computes by Gaussian elimination over $\mathbb{F}_2$ and checks by counting points in every elementary interval. For the first 2^{10} points, the first 2, 3, 5 and 8 coordinates have $t = 0$, 1, 3 and 5, against the bounds 0, 1, 5 and 15 of (iii). Nets have low discrepancy: for a (t, m, d)-net in base b, $D_n^* \le C_{d,b}\, b^t (\log n)^{d-1}/n + O(b^t (\log n)^{d-2}/n)$, and for the first n points of a (t, d)-sequence, for every $n \ge 2$, $D_n^* \le C'_{d,b}\, b^t (\log n)^d/n + O(b^t (\log n)^{d-1}/n)$ (Niederreiter, 1992, Chapter 4; Theorem 4.17 for sequences). In two dimensions the star discrepancy can be computed exactly, and at $n = 2048$ it is 0.020 for pseudo-random points, 0.0037 for Halton and 0.0024 for Sobol' (Figure 11.2, left).

Remark 11.10 (The origin, and n not a power of two). Both sequences start at $x_0 = 0$, which N^{-1} sends to $-\infty$. A deterministic run with the normal map must drop the first point, which spoils the net property of the first 2^m points, or move all the points: by (i) the first 2^m Sobol' points have coordinates that are multiples of 2^{-m}, and adding 2^{-m-1} to every coordinate keeps each point in all the elementary intervals of side at least 2^{-m} that contained it, so the result is still a (t, m, d)-net, with no coordinate at 0. Randomisation settles the issue. Second, the balance of Theorem 11.7 concerns blocks of 2^m points with $m \ge t$. For other n, the first n points split into aligned blocks of sizes 2^k (as in the proof of Theorem 11.6(ii)), which are (t, k, d)-nets only for $k \ge t$; the blocks with $k < t$ carry no such guarantee. The discrepancy bound for sequences still holds, but the faster rates of Section 11.3 are lost (Figure 11.2, right).

In code

code/src/mfe230d/qmc.py: `van_der_corput(n, b)`, `halton(n, d)` and `sobol(n, d)` (natural order, direction numbers of Joe and Kuo for up to 128 coordinates; `scipy.stats.qmc.Sobol` returns the same points in Gray-code order, which the tests check); `star_discrepancy` (exact, practical for $d \le 2$); `t_value(generator_matrices(d, m))` and `is_net`.

11.3 Randomised quasi-Monte Carlo

A deterministic rule has no standard error, and its error bound is rarely usable (Theorem 11.4). Randomisation makes every point uniform while keeping the structure of the set. Three randomisations of points $x_1, \dots, x_n$ are in use; each draws its randomness once and applies it to all the points.

(a) *Random shift* (Cranley and Patterson, 1976): $y_i = x_i + U \bmod 1$ (coordinatewise fractional part), with U uniform on $[0,1)^d$.

(b) *Random digital shift* in base 2: $y_i = x_i \oplus U$, digit by digit.

(c) *Nested uniform scramble* (Owen, 1995), in base 2: in each coordinate, independently, digit k of y is digit k of x flipped by a random bit $e(x^{(1)}, \dots, x^{(k-1)})$ attached to the first $k-1$ digits of x; all the bits e are independent fair coin tosses. Geometrically: swap the two halves of $[0,1)$ with probability $\frac{1}{2}$, then independently swap the two halves of each half, and so on.

Proposition 11.11 (Randomised QMC)**.** *Let $y_1, \dots, y_n$ be obtained from $x_1, \dots, x_n \in [0,1)^d$ by (a), (b) or (c), let $F \in L^2([0,1)^d)$ with $\sigma^2 = \operatorname{Var} F(U)$ for U uniform, and let $\hat{\mu} = \frac{1}{n}\sum_i F(y_i)$.*

(i) Each y_i is uniformly distributed on $[0,1)^d$. Hence $\mathbb{E}\hat{\mu} = I(F)$, and $\operatorname{Var}\hat{\mu} \le \sigma^2$.

(ii) The bound $\operatorname{Var}\hat{\mu} \le \sigma^2$, n times the Monte Carlo variance, is attained for some point sets and integrands under (a) and (b), and under (c) when all the points coincide.

(iii) Under (b) and (c), if $x_1, \dots, x_n$ is a (t, m, d)-net in base 2, then so is $y_1, \dots, y_n$, with probability one.

(iv) Let $\hat{\mu}^{(1)}, \dots, \hat{\mu}^{(R)}$, $R \ge 2$, be computed from independent randomisations, with mean $\bar{\mu}_R$ and sample variance s_R^2. Then $\bar{\mu}_R$ is unbiased, $\operatorname{Var}\bar{\mu}_R = \operatorname{Var}\hat{\mu}/R$, and $\mathbb{E}[s_R^2/R] = \operatorname{Var}\bar{\mu}_R$. If the replicates are normally distributed, $\bar{\mu}_R \pm t_{R-1,1-\alpha/2}\, s_R/\sqrt{R}$ covers $I(F)$ with probability exactly $1-\alpha$; without normality, its coverage tends to $1-\alpha$ as $R \to \infty$.

(v) Let $x_1, \dots, x_n$ be a (t, m, d)-net in base 2, $n = 2^m$, randomised by (c). Then $\operatorname{Var}\hat{\mu} \le 2^t 3^d \sigma^2 / n$. For the first 2^m points of a (t, d)-sequence, $\operatorname{Var}\hat{\mu} = o(1/n)$ as $m \to \infty$, and if moreover F is smooth, $\operatorname{Var}\hat{\mu} = O\big(n^{-3}(\log n)^{d-1}\big)$.

Proof. (i) Under (a), $u \mapsto x + u \bmod 1$ is a translation of the torus and preserves volume. Under (b) and (c), recall that U is uniform on $[0,1)$ if and only if its binary digits are independent fair bits (outside the null set of dyadic rationals). Under (b), each digit of y is a fixed digit of x flipped by an independent fair bit; under (c), digit k of y is digit k of x flipped by the bit attached to one node at depth $k-1$, and different depths use different, independent bits. In both cases the

digits of each coordinate of y are independent fair bits, and different coordinates are independent. So $\mathbb{E}F(y_i) = I(F)$, and by the Cauchy–Schwarz inequality $\operatorname{Var}\hat{\mu} = n^{-2}\sum_{i,k}\operatorname{Cov}(F(y_i), F(y_k)) \le n^{-2}\sum_{i,k}\sigma^2 = \sigma^2$.

(ii) For (a), take $d = 1$, $x_i = (i-1)/n$ and $F(y) = \cos(2\pi n y)$: every $F(y_i) = \cos(2\pi n U)$, so $\hat{\mu} = \cos(2\pi n U)$ has variance $\frac{1}{2} = \sigma^2$. For (b), take $n = 2^m$ and $F(y) = (-1)^{y^{(m+1)}}$, the sign given by digit $m+1$: that digit is 0 for every $x_i = (i-1)/n$, so it equals $U^{(m+1)}$ for every y_i, and $\hat{\mu} = \pm 1$ with variance $1 = \sigma^2$. If all points coincide, so do the y_i, and $\hat{\mu} = F(y_1)$.

(iii) Under (b) and (c), the first k digits of each coordinate of y are a bijective function of the first k digits of the same coordinate of x (for (c), the composition of the flips along the path from the root). So each elementary interval is mapped onto an elementary interval of the same shape, and the counts of Theorem 11.7 are preserved. The digit expansions involved are proper with probability one.

(iv) The replicates are independent copies of $\hat{\mu}$, which is in L^2 by (i). Unbiasedness of $\bar{\mu}_R$ and of the sample variance for independent identically distributed variables give the first three claims; the normal case is the classical Student interval, and the limit as $R \to \infty$ is Theorem 6.8 with the replicates as batches.

(v) The rate $o(1/n)$ for every $F \in L^2$ is due to Owen (1997b), as is a bound by a small multiple of σ^2/n for $t = 0$; the bound $2^t 3^d \sigma^2/n$ is Owen (1998), Lemma 4. The rate for smooth F is due to Owen (1997a); Owen (2008), Section 5.1 and Theorem 3, repairs an error in its proof and weakens its hypothesis to the smoothness assumed here. □

Three comments. First, $\hat{\mu}$ is a sum of dependent terms, and the replicates of (iv), not the spread of the $F(y_i)$, measure its error: this is the batch-means principle of Theorem 6.8. For a fixed number of replicates, the Student interval rests on approximate normality of each replicate, which Loh (2003) proved for scrambled $(0, m, d)$-nets and smooth integrands. Second, the budget: with $N = nR$ evaluations and $\operatorname{Var}\hat{\mu} \approx cn^{-\alpha}$, $\operatorname{Var}\bar{\mu}_R \approx cn^{1-\alpha}/N$, so for $\alpha > 1$ one should take n large and R as small as a stable error estimate allows, say 10 to 30. Third, in one dimension the nested scramble of the first 2^m van der Corput points is a stratified sample with 2^m strata (Chapter 10): one point in each interval $[l2^{-m}, (l+1)2^{-m})$, uniform there and independent of the others (Exercise 11.6). In d dimensions, by Theorem 11.9(i), each one-dimensional projection of the scrambled first 2^m Sobol' points is stratified in this way, as in Latin hypercube sampling (Theorem 10.10), and the net also balances the elementary intervals in several coordinates at once.

Figure 11.2 shows the three regimes. For the smooth integrand in $d = 5$, at $n = 2^{16}$, Monte Carlo has root mean squared error 5.3×10^{-3}, the scrambled Sobol' points 8.8×10^{-5}, and the deterministic Halton and Sobol' rules have errors 2.8×10^{-5} and 1.9×10^{-4} (our implementation: numpy `default_rng`, seeds as in the caption). The deterministic errors jump around with n; the randomised one decreases steadily, with a fitted slope of -0.99 in log–log scale between $n = 2^8$ and 2^{16} (from -1.03 to -0.95 over the four groups of 32 scrambles; the replicate errors are heavy-tailed, so fewer scrambles give noticeably different values): the rate $n^{-3/2}$ of (v) is asymptotic, and its logarithmic factor $(\log n)^{(d-1)/2}$ still grows fast at these n when $d = 5$. In $d = 2$ the asymptotic regime is visible: $n \times \text{RMSE}$ falls from 0.36 at $n = 2^8$ to 0.10 at $n = 2^{12}$ along powers of two, while between them it is of order 1 (median 1.2 for $2^{11} < n < 2^{12}$), the order $1/n$ that the smallest blocks of Theorem 11.10 impose.

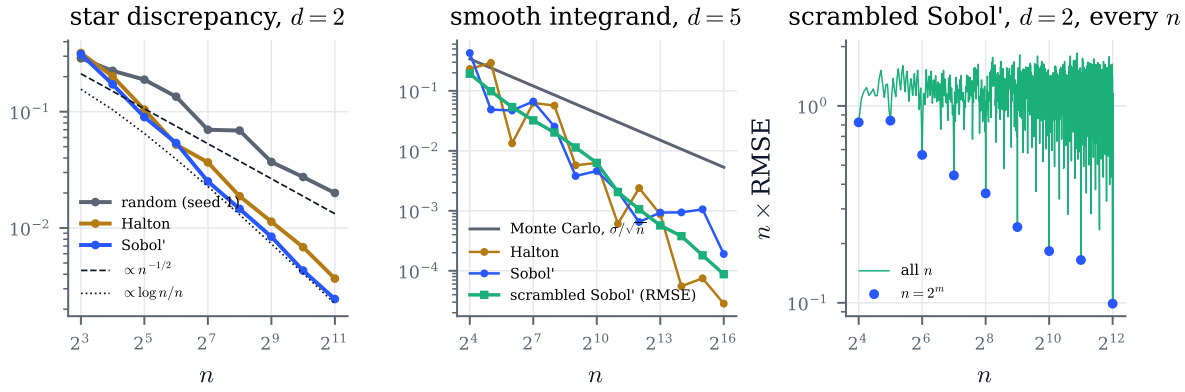

Figure 11.2: Left: exact star discrepancy of the first n points in two dimensions, for pseudo-random points (seed 1), Halton (bases 2, 3) and Sobol' (coordinates 1, 2), with reference slopes $n^{-1/2}$ and $\log n/n$. Centre: the product of sines $\prod_{j=1}^{5} \frac{\pi}{2} \sin(\pi x_j)$ of Chapter 6, whose integral is 1: root mean squared error $\sigma/\sqrt{n}$ of Monte Carlo, absolute errors of the Halton and Sobol' rules, and the root mean squared error of nested-scrambled Sobol' points over 128 independent scrambles (seed 7). Right: n times the root mean squared error of scrambled Sobol' points for the same integrand in $d = 2$, over 64 scrambles (seed 8), for every n from 16 to 4096; the dots mark $n = 2^m$.

▶ Interactive demo: scramble the Sobol' points and watch the net survive; then compare Monte Carlo, QMC and scrambled QMC on smooth, kinked and discontinuous integrands, in the dimension of your choice, with replicate-based error bars.

11.4 Effective dimension and path constructions

A monthly Asian option is an integral in 12 dimensions, a weekly one in 52. The bounds of Section 11.2 are useless there, and yet QMC prices such options remarkably well. The explanation of Caflisch et al. (1997) is that the integrands are *effectively* low-dimensional, provided the normals are mapped to the path in the right order. It rests on the analysis-of-variance (ANOVA) decomposition. Every $F \in L^2([0,1)^d)$ can be written uniquely as $F = \sum_{u \subseteq D} F_u$, where F_u depends only on x_u, $F_\emptyset = I(F)$, and $\int_0^1 F_u \, \mathrm{d}x_j = 0$ for $j \in u$; the terms are orthogonal, so $\sigma^2 = \sum_{u \neq \emptyset} \sigma_u^2$ with $\sigma_u^2 = I(F_u^2)$, and $\mathbb{E}[F(U) \mid U_v] = \sum_{u \subseteq v} F_u(U_u)$ for every $v \subseteq D$ (Sobol', 2001). In particular

$$S_v := \frac{\operatorname{Var} \mathbb{E}[F(U) \mid U_v]}{\sigma^2} = \frac{1}{\sigma^2} \sum_{\emptyset \neq u \subseteq v} \sigma_u^2 \tag{11.6}$$

is the share of the variance carried by the coordinates in v and their interactions.

> **Lemma 11.12** (Pick and freeze)**.** *Let U, U' be independent and uniform on $[0,1)^d$, $v \subseteq D$, and let $U^{(v)} = (U_v, U'_{D \setminus v})$ keep the coordinates of U in v and take the others from U'. For $F \in L^2$, $\operatorname{Var} \mathbb{E}[F(U) \mid U_v] = \operatorname{Cov}\left(F(U), F(U^{(v)})\right)$.*

Proof. Given U_v, the variables $F(U)$ and $F(U^{(v)})$ are independent with the same conditional

law, because they are the same function of U_v and of the independent vectors $U_{D\setminus v}$ and $U'_{D\setminus v}$. With $h(U_v) = \mathbb{E}[F(U) \mid U_v]$, the tower property gives $\mathbb{E}[F(U)F(U^{(v)})] = \mathbb{E}[h(U_v)^2]$; subtract $I(F)^2 = (\mathbb{E}h(U_v))^2$. □

The lemma turns each share into a covariance that plain Monte Carlo estimates (Sobol', 2001). In normal coordinates it reads the same with Z, Z' in place of U, U'.

Remark 11.13 (Effective dimension). Following Caflisch et al. (1997), with a proportion p such as 0.99, F has *truncation dimension* d_T, the smallest k with $S_{\{1,\dots,k\}} \geq p$, and *superposition dimension* d_S, the smallest s with $\sum_{0<|u|\leq s} \sigma_u^2 \geq p\sigma^2$. The first says how many leading coordinates matter, the second how many coordinates interact; $\sum_j S_{\{j\}}$, the share of the additive part, is the case $s = 1$. Both depend on how the uniforms are turned into the path, although the price does not: the increments, Brownian-bridge and principal-component constructions of Chapter 7 give three integrands F on $[0,1)^d$ with the same integral. The bridge and principal components put most of the variance of the path into the first normals (Theorems 7.7 and 7.8), and QMC points are best in their first coordinates: the leading Sobol' coordinates use the polynomials of lowest degree, hence the smallest bounds on t (Theorem 11.9(iii)), and the leading Halton coordinates the smallest bases. For scrambled nets, part of this argument can be made exact. The variance of (c) splits as $\operatorname{Var}\hat{\mu} = n^{-1}\sum_{u\neq\emptyset}\sum_r \Gamma_{u,r}\sigma_{u,r}^2$, where the $\sigma_{u,r}^2 \geq 0$ split each σ_u^2 further by resolution r and the gains $\Gamma_{u,r}$ depend only on the points (Owen, 1997b); for a (t,m,d)-net in base 2, $0 \leq \Gamma_{u,r} \leq 2^t 3^d$ (Owen, 1998). The terms with $u \subseteq \{1,\dots,k\}$ make up the scrambled-net variance for the k-dimensional integrand $\mathbb{E}[F \mid U_1,\dots,U_k]$, which is $o(1/n)$ by Theorem 11.11(v); the others add at most $2^t 3^d(1 - S_{\{1,\dots,k\}})\sigma^2/n$. This much is a theorem. That a small $1 - S_{\{1,\dots,k\}}$ makes QMC accurate at a given n is a heuristic: nothing says how large n must be before the k-dimensional part is integrated accurately, and the constant $2^t 3^d$ is astronomically large for $d = 64$. It is the measurements below that support it.

For the *geometric* average the shares are exact. On the dates $t_1 < \cdots < t_m$, with $W = AZ$ for the construction matrix A,

$$\log G = \log S + \nu\bar{t} + \frac{\sigma}{m}\sum_{i=1}^{m} W_{t_i} = \log S + \nu\bar{t} + \frac{\sigma}{m}\,|A^\top \mathbf{1}|\,(w\cdot Z), \qquad w = \frac{A^\top \mathbf{1}}{|A^\top \mathbf{1}|},$$

with $\bar{t}$ the mean date. So the geometric Asian call is $g(w \cdot Z)$ for a fixed function g and a unit vector w that depends on the construction. For such a *ridge* payoff, $(w\cdot Z, w\cdot Z^{(v)})$ is a standard bivariate normal pair with correlation $\rho_v^2 = \sum_{j\in v} w_j^2$. Write $\kappa_g(c) = \operatorname{Corr}\left(g(X), g(X')\right)$ for standard normal X, X' with $\operatorname{Corr}(X, X') = c$; then Theorem 11.12 gives

$$S_v = \kappa_g(\rho_v^2), \qquad \rho_v^2 = \sum_{j\in v} w_j^2. \tag{11.7}$$

For a linear payoff $\kappa_g(c) = c$; in general $\kappa_g(c) \leq c$ (Exercise 11.7). For the call, $\mathbb{E}[g(X') \mid X]$ is a Black–Scholes-type formula and κ_g is a one-dimensional integral. Table 11.1 gives the result for $m = 64$ equally spaced dates over one year in the test case.

The first normal of the bridge sets W_T, which carries about $\frac{3}{4}$ of the variance of the average; the first principal component carries 98.6% of it. The call loses a little of each share to nonlinear interactions, much more under the increments, where the average mixes all 64 normals with comparable weights and only 68% of the variance is additive. The arithmetic average is not a

construction	w_1^2	$S_{\{1\}}$	$S_{\{1,2\}}$	$S_{\{1,\dots,4\}}$	$S_{\{1,\dots,8\}}$	d_T	additive share
increments	0.046	0.032	0.063	0.127	0.253	53	0.684
Brownian bridge	0.756	0.685	0.914	0.978	0.994	7	0.860
principal components	0.986	0.979	0.997	1.000	1.000	2	0.989

Table 11.1: Variance shares of the geometric Asian call with $m = 64$ dates (test case), by (11.7) and quadrature: the linear share w_1^2 of the first normal, the shares S_v of the first k normals, the truncation dimension d_T for $p = 0.99$, and the share $\sum_j S_{\{j\}}$ of the additive part.

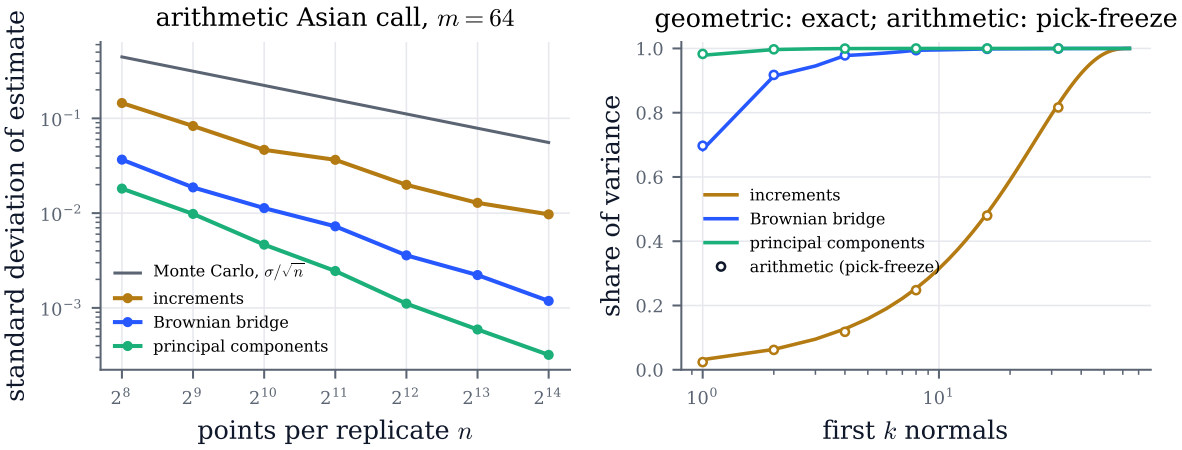

Figure 11.3: The arithmetic Asian call with $m = 64$ dates, test case. Left: standard deviation of the estimate with n nested-scrambled Sobol' points, over 64 independent scrambles (seed 11), under the three path constructions, against the Monte Carlo standard error $\sigma/\sqrt{n}$ ($\sigma = 7.12$, from the second moment of the payoff over all 2^{20} scrambled points). Right: share of the variance carried by the first k normals: exact for the geometric Asian call (lines, (11.7)), and estimated by pick-freeze for the arithmetic one (circles, 2^{15} draws, seed 5).

ridge payoff, but its shares, estimated by Theorem 11.12, are close to those of the geometric call (Figure 11.3, right).

Example 11.14 (Asian options by randomised QMC)**.** With $m = 64$ dates, $n = 2^{14}$ scrambled Sobol' points and 64 scrambles (our implementation, seed 11), the variance of the estimate is smaller than the Monte Carlo variance σ^2/n with the same n by a factor of 33 with increments, 2.2×10^3 with the bridge and 3.0×10^4 with principal components (Figure 11.3, left). The replicate errors are heavy-tailed, and the factors are noisy: over the four groups of 16 scrambles they range from 24 to 50, from 1.4×10^3 to 3.8×10^3, and from 1.5×10^4 to 6.2×10^4. From $n = 2^8$ to 2^{14} the factor of the bridge grows from 1.5×10^2 and that of principal components from 6.0×10^2, while that of the increments levels off around 30. The ranking is that of Table 11.1. Principal components cost $O(m^2)$ operations per path against $O(m)$ for the other two, which does not change it here. For the course's monthly example ($m = 12$), $n = 2^{14}$ points and 16 scrambles (seed 2026) give 4.56133 ± 0.00011 for the arithmetic Asian call with principal components and 4.56111 ± 0.00038 with the bridge, and 4.40291 ± 0.00011 for the geometric one, whose exact value is 4.40286. The course's estimate 4.56058 came from 10^8 plain paths, with a standard error of 7.5×10^{-4}, seven times that of principal components with 2.6×10^5 paths; the two estimates

differ by one of its standard errors.

The constructions are not uniformly ranked. A payoff that depends only on the first date, W_{t_1}, is one-dimensional under the increments and spread over several normals by the bridge (Exercise 11.8); Papageorgiou (2002) constructs integrands for which the bridge does worse than the increments. The effective dimension belongs to the pair (payoff, construction), and it can be measured before pricing with Theorem 11.12. The use of the bridge for QMC is due to Moskowitz and Caflisch (1996), and that of principal components to Acworth et al. (1998).

In code

`code/src/mfe230d/qmc.py`: `random_shift`, `digital_shift` and `owen_scramble` randomise points; `rqmc_estimate(f, n, d, R)` returns the replicate mean, its standard error and a Student interval; `to_normals` applies N^{-1} (the accurate quantile of Chapter 5); `asian_call_payoffs` builds the paths with `paths.gbm_paths`; `lognormal_call_share`, `ridge_shares` and `pick_freeze_shares` compute the variance shares.

▶ Interactive demo: choose the number of dates and the construction, and compare the variance shares of the first normals with the accuracy of randomised QMC for the Asian call.

Course examples reproduced. The course announced low-discrepancy sequences without covering them, so it has no worked examples on this topic. We reproduce the numbers of its Asian example that the chapter touches: the exact geometric Asian call 4.40286, and the bare-bones estimate 4.56058 of the arithmetic one, which agrees with our randomised QMC value within its standard error.

Further reading

Glasserman (2003), Chapter 5, covers QMC for finance: discrepancy and the Koksma–Hlawka bound in Section 5.1, the Halton, Faure and Sobol' constructions in Section 5.2, lattice rules, which we have skipped, in Section 5.3, randomised QMC in Section 5.4, and numerical examples with the bridge and principal components in Section 5.5. Niederreiter (1992), Chapters 2–4, is the classical treatment of discrepancy, error bounds and (t, m, s)-nets, and Dick and Pillichshammer (2010) the modern one of digital nets, including their randomisation in Chapter 13. Owen (2005) explains the Hardy–Krause variation and why kinks and jumps defeat it.

Chapter 12

Multilevel Monte Carlo

Chapter 7 ended with an uncomfortable trade-off. When the solution of a stochastic differential equation cannot be sampled exactly, a discretised Monte Carlo estimator pays twice: the bias of the scheme calls for small time steps, the statistical error for many paths, and with a first-order scheme the cost of a root-mean-square error ε grows like ε^{-3} instead of the ε^{-2} of exact sampling (Theorems 6.11 and 7.14). Multilevel Monte Carlo closes most of this gap, and in favourable cases all of it.

The idea fits in one line. Write the price computed with the finest step as the price computed with a very coarse step plus a sum of corrections, each the difference between the payoffs computed with two consecutive steps *on the same Brownian path*. The coarse price is cheap, so it can be estimated from many paths. A correction is expensive per path, but the two paths in it are close, so its variance is small and few paths suffice. Most of the work is done where it is cheap.

Giles (2008b) introduced the method for path simulation, after the multilevel methods of Heinrich (2001) for parametric integration and the two-level "statistical Romberg" estimator of Kebaier (2005). We prove the identity it rests on and the best allocation of samples across levels (Section 12.1), measure how fast the corrections shrink under the Euler and Milstein couplings, also for discontinuous payoffs (Section 12.2), prove Giles' complexity theorem with its three regimes and compare it with the allocation of Theorem 7.14 (Section 12.3), describe the adaptive algorithm used in practice (Section 12.4), and end with the randomised estimators of Rhee and Glynn (2015), which remove the bias altogether when the corrections shrink fast enough (Section 12.5).

Throughout, X solves (7.5) on $[0,T]$ under $\mathbb{Q}$ with the Brownian motion W; P is a square-integrable discounted payoff, a function of X_T or of the path of X; $\theta = \mathbb{E}[P]$ is its price; and P_ℓ is the payoff computed from a scheme with step $h_\ell = T2^{-\ell}$, $\ell \in \mathbb{N}$. The running example is the test case: geometric Brownian motion $\mathrm{d}S = \mu S\,\mathrm{d}t + \sigma S\,\mathrm{d}W$ with $\mu = r - y = -1\%$, $\sigma = 20\%$, $T = 1$, and $P = e^{-rT}(S_T - K)^+$ with $S_0 = K = 100$, so $\theta = 7.36429$. Costs are counted in time steps: one step of one path costs 1. A warning on notation: Theorem 7.9 called the strong and the weak orders γ and β. From now on, as in the literature on multilevel methods, α, β and γ are the rates at which the bias, the variance of the corrections and the cost change from one level to the next, and the strong order of a scheme is written λ.

12.1 The multilevel identity

For a finest level L, linearity of expectation gives

$$\mathbb{E}[P_L] = \mathbb{E}[P_0] + \sum_{\ell=1}^{L} \mathbb{E}\big[P_\ell - P_{\ell-1}\big]. \tag{12.1}$$

This is a triviality. It becomes a method when every term on the right is estimated from its own independent samples, and every difference is computed from a fine and a coarse approximation that are strongly dependent.

Definition 12.1 (Coupled corrections)**.** Let $P_0, P_1, \dots \in L^2$. A *coupled correction* at level $\ell \geq 1$ is a random variable $Y_\ell = P_\ell^{\mathrm{f}} - P_{\ell-1}^{\mathrm{c}}$, where the *fine* member P_ℓ^{f} and the *coarse* member $P_{\ell-1}^{\mathrm{c}}$ are square-integrable random variables on one probability space with

$$\mathbb{E}\big[P_\ell^{\mathrm{f}}\big] = \mathbb{E}[P_\ell], \qquad \mathbb{E}\big[P_{\ell-1}^{\mathrm{c}}\big] = \mathbb{E}[P_{\ell-1}];$$

usually P_ℓ^{f} has the law of P_ℓ and $P_{\ell-1}^{\mathrm{c}}$ the law of $P_{\ell-1}$. At level 0, $Y_0 = P_0$. We write $V_\ell = \mathrm{Var}(Y_\ell)$ and $\tau_\ell > 0$ for the expected cost of one sample of Y_ℓ.

Only the means of the two members are constrained; their joint law is ours to choose, and the whole art is to choose it so that V_ℓ is small.

Proposition 12.2 (The multilevel estimator)**.** *Let $Y_0, Y_1, \dots$ be coupled corrections, $L \in \mathbb{N}$ and $n_0, \dots, n_L \geq 1$ integers. Let $Y_\ell^{(i)}$, $1 \leq i \leq n_\ell$, $0 \leq \ell \leq L$, be independent, with $Y_\ell^{(i)}$ distributed as Y_ℓ, and*

$$\hat{\theta}_L = \sum_{\ell=0}^{L} \hat{Y}_\ell, \qquad \hat{Y}_\ell = \frac{1}{n_\ell} \sum_{i=1}^{n_\ell} Y_\ell^{(i)}. \tag{12.2}$$

(i) $\mathbb{E}[\hat{\theta}_L] = \mathbb{E}[P_L]$, *whatever the joint laws of the fine and coarse members.*

(ii) $\mathrm{Var}(\hat{\theta}_L) = \sum_{\ell=0}^{L} V_\ell/n_\ell$, *and for $\theta = \mathbb{E}[P]$ with $P \in L^1$, $\mathbb{E}\big[(\hat{\theta}_L - \theta)^2\big] = \big(\mathbb{E}[P_L] - \theta\big)^2 + \sum_{\ell=0}^{L} V_\ell/n_\ell$.*

(iii) The expected cost is $\mathcal{C} = \sum_{\ell=0}^{L} n_\ell \tau_\ell$. For all real $n_\ell > 0$,

$$\Big(\sum_{\ell=0}^{L} \frac{V_\ell}{n_\ell}\Big)\Big(\sum_{\ell=0}^{L} n_\ell \tau_\ell\Big) \;\geq\; S_L^2, \qquad S_L = \sum_{\ell=0}^{L} \sqrt{V_\ell \tau_\ell},$$

and if every $V_\ell > 0$, equality holds if and only if n_ℓ is proportional to $\sqrt{V_\ell/\tau_\ell}$.

(iv) Let $v > 0$. Every choice of the n_ℓ with $\mathrm{Var}(\hat{\theta}_L) \leq v$ costs at least S_L^2/v, and the integers

$$n_\ell = \max\Big(1, \Big\lceil \frac{S_L}{v} \sqrt{\frac{V_\ell}{\tau_\ell}} \Big\rceil\Big) \tag{12.3}$$

give $\mathrm{Var}(\hat{\theta}_L) \leq v$ at a cost $\mathcal{C} \leq S_L^2/v + \sum_{\ell=0}^{L} \tau_\ell$.

Proof. (i) By Theorem 12.1, $\mathbb{E}[Y_\ell] = \mathbb{E}[P_\ell] - \mathbb{E}[P_{\ell-1}]$ for $\ell \geq 1$; only the two means enter. Each $\hat{Y}_\ell$ is unbiased for $\mathbb{E}[Y_\ell]$, and the sum telescopes as in (12.1).

(ii) The $\hat{Y}_\ell$ are built from disjoint sets of independent samples, so they are independent and their variances add; $\mathrm{Var}(\hat{Y}_\ell) = V_\ell/n_\ell$ by Theorem 6.2(iii). The mean squared error is the squared bias plus the variance, as in Theorem 6.11(i).

(iii) Apply the Cauchy–Schwarz inequality to $a_\ell = \sqrt{V_\ell/n_\ell}$ and $b_\ell = \sqrt{n_\ell \tau_\ell}$: $(\sum a_\ell^2)(\sum b_\ell^2) \geq (\sum a_\ell b_\ell)^2 = S_L^2$, with equality if and only if $a = tb$ for some $t \geq 0$ (as $b \neq 0$). If every $V_\ell > 0$, then $t > 0$, and $\sqrt{V_\ell/n_\ell} = t\sqrt{n_\ell \tau_\ell}$ means $n_\ell = t^{-1}\sqrt{V_\ell/\tau_\ell}$.

(iv) The lower bound is (iii) with $\mathrm{Var}(\hat{\theta}_L) \leq v$. For (12.3), $n_\ell \geq v^{-1} S_L \sqrt{V_\ell/\tau_\ell}$ gives $V_\ell/n_\ell \leq v\sqrt{V_\ell \tau_\ell}/S_L$ (trivially if $V_\ell = 0$), and summing, $\mathrm{Var}(\hat{\theta}_L) \leq v$. And $n_\ell \leq v^{-1} S_L \sqrt{V_\ell/\tau_\ell} + 1$, so $\sum n_\ell \tau_\ell \leq v^{-1} S_L \sum \sqrt{V_\ell \tau_\ell} + \sum \tau_\ell$. □

Plain Monte Carlo on the finest level reaches the variance v at the cost $\mathrm{Var}(P_L)\, 2^L/v$, one fine path of 2^L steps per sample. The multilevel estimator replaces the work-normalised variance $\mathrm{Var}(P_L)\, 2^L$ (Theorem 6.9) by S_L^2, which is much smaller when $V_\ell \ll \mathrm{Var}(P_\ell)$ at the expensive levels. The optimal n_ℓ spends the amount $n_\ell \tau_\ell \propto \sqrt{V_\ell \tau_\ell}$ on level ℓ; where that amount concentrates decides the complexity (Section 12.3). Two degenerate cases are worth checking. If $V_\ell = 0$, then Y_ℓ equals its mean almost surely: one sample is exact, the continuous optimum $n_\ell = 0$ is not admissible because $\mathbb{E}[Y_\ell]$ must still be added, and (12.3) takes $n_\ell = 1$, which is the reason for the term $\sum \tau_\ell$ in (iv); this happens for an equation with constant coefficients, Brownian motion with drift, for which the Euler scheme is exact on every grid, so that the fine and coarse values of X_T coincide (with additive noise but a state-dependent drift, as for Ornstein–Uhlenbeck, it is not, and $V_\ell > 0$). If $L = 0$, the estimator is plain Monte Carlo with one coarse step.

Remark 12.3 (A chain of control variates). With $L = 1$, $\hat{\theta}_1 = \frac{1}{n_1}\sum_i \left(P_1^{\mathrm{f}} - P_0^{\mathrm{c}}\right)^{(i)} + \frac{1}{n_0}\sum_j P_0^{(j)}$ is the control-variate estimator (8.4) for P_1, with the control P_0^{c}, the coefficient 1, and the unknown control mean $\mathbb{E}[P_0]$ replaced by an independent estimate from n_0 cheap samples. Any coefficient b in place of 1, used in both sums, keeps the estimator unbiased; its variance $\big(\mathrm{Var}(P_1) - 2b\,\mathrm{Cov}(P_1^{\mathrm{f}}, P_0^{\mathrm{c}}) + b^2\,\mathrm{Var}(P_0)\big)/n_1 + b^2\,\mathrm{Var}(P_0)/n_0$ is smallest for $b^* = \mathrm{Cov}(P_1^{\mathrm{f}}, P_0^{\mathrm{c}})/\big(\mathrm{Var}(P_0)(1 + n_1/n_0)\big)$, the coefficient of Theorem 8.12 shrunk to account for the noise in the estimate of the control mean. When the levels are close, the correlation of the two members is near 1, their variances are near $\mathrm{Var}(P)$, and $n_1 \ll n_0$, so $b^* \approx 1$: the simple choice $b = 1$ loses little. A multilevel estimator is a chain of such controls: each level controls the one above it, and its own mean is estimated with the help of the level below. Computing both members from one Brownian path is common random numbers (Theorem 8.11); with independent paths, $V_\ell = \mathrm{Var}(P_\ell) + \mathrm{Var}(P_{\ell-1})$ does not shrink and the method loses (Exercise 12.1).

12.2 Coupling fine and coarse paths

Let Y^h denote the Euler scheme (7.6) or the Milstein scheme (7.7) with step h, driven by the increments of W on its grid. One sample of Y_ℓ, $\ell \geq 1$, uses one draw of the 2^ℓ fine increments $\Delta W_1, \ldots, \Delta W_{2^\ell}$ of W over steps of length h_ℓ: the fine member is the payoff of Y^{h_ℓ} driven by them, the coarse member the payoff of $Y^{h_{\ell-1}}$ driven by the sums $\Delta W_{2k-1} + \Delta W_{2k}$, $k \leq 2^{\ell-1}$, which are the increments of the same W over the coarse grid (`paths.coarsen`). Both members approximate the payoff of the same X. The sums are independent $\mathcal{N}(0, h_{\ell-1})$ variables, so the coarse member is computed by the level-$(\ell-1)$ scheme from inputs with the right law: it has the

law of $P_{\ell-1}$, and Y_ℓ is a coupled correction. (Drawing the coarse increments first and splitting each one with the Brownian bridge of Theorem 7.7 gives the same joint law.)

Proposition 12.4 (Variance of the corrections: Euler against Milstein)**.** *Let Y_ℓ be built as above from the Euler or the Milstein scheme, and suppose that for some $p \geq 1$ there are constants $A_p, \lambda > 0$ with*

$$\big(\mathbb{E}|X_T - Y_T^h|^p\big)^{1/p} \leq A_p\, h^\lambda \qquad \text{for all } h = T2^{-\ell},\ \ell \in \mathbb{N}. \tag{12.4}$$

(i) (Lipschitz payoffs.) Let $P = f(X_T)$ and $P_\ell = f(Y_T^{h_\ell})$ with $|f(x) - f(z)| \leq \Lambda|x - z|$. If (12.4) holds with $p = 2$, then

$$V_\ell \leq \mathbb{E}[Y_\ell^2] \leq \Lambda^2 A_2^2\,(1 + 2^\lambda)^2\, h_\ell^{2\lambda}.$$

(ii) (Digital payoffs.) Let $P = \mathbf{1}_{\{X_T > K\}}$ and $P_\ell = \mathbf{1}_{\{Y_T^{h_\ell} > K\}}$. If X_T has a density bounded by ρ on $[K - \delta_0, K + \delta_0]$ for some $\delta_0 > 0$, then for every ℓ with $h_\ell^{\lambda p/(p+1)} \leq \delta_0$,

$$V_\ell \leq \mathbb{E}[Y_\ell^2] = \mathbb{Q}\big(\mathbf{1}_{\{Y_T^{h_\ell} > K\}} \neq \mathbf{1}_{\{Y_T^{h_{\ell-1}} > K\}}\big) \leq \big(2\rho + (1 + 2^{\lambda p})A_p^p\big)\, h_\ell^{\lambda p/(p+1)}.$$

(iii) For the scalar equation (7.5), the bound (12.4) holds with $p = 2$, $\lambda = \frac{1}{2}$ for the Euler scheme under the hypotheses (A1)–(A3) of Theorem 7.10, and with $p = 2$, $\lambda = 1$ for the Milstein scheme under the hypotheses of Theorem 7.11. Consequently, for Lipschitz payoffs $V_\ell = O(h_\ell)$ with Euler and $V_\ell = O(h_\ell^2)$ with Milstein. (For a system driven by several Brownian motions with non-commuting diffusion vector fields, the Milstein scheme without the Lévy areas has strong order only $\frac{1}{2}$, and then $\beta = 1$ (Giles and Szpruch, 2014).)

Proof. (i) $|Y_\ell| \leq \Lambda|Y_T^{h_\ell} - Y_T^{h_{\ell-1}}| \leq \Lambda(|Y_T^{h_\ell} - X_T| + |X_T - Y_T^{h_{\ell-1}}|)$. Both schemes are driven by the same W as X, so (12.4) applies to each term, and Minkowski's inequality gives $\|Y_\ell\|_2 \leq \Lambda A_2(h_\ell^\lambda + h_{\ell-1}^\lambda) = \Lambda A_2(1 + 2^\lambda)h_\ell^\lambda$. Finally $V_\ell \leq \mathbb{E}[Y_\ell^2]$.

(ii) Y_ℓ takes values in $\{-1, 0, 1\}$, so $\mathbb{E}[Y_\ell^2]$ is the probability that the two indicators differ. Let $0 < \delta \leq \delta_0$. If $|X_T - K| > \delta$ and both schemes are within δ of X_T, then both lie strictly on the side of K where X_T lies, and the indicators agree. Hence, by Markov's inequality and (12.4),

$$\mathbb{E}[Y_\ell^2] \leq \mathbb{Q}(|X_T - K| \leq \delta) + \frac{\mathbb{E}|Y_T^{h_\ell} - X_T|^p + \mathbb{E}|Y_T^{h_{\ell-1}} - X_T|^p}{\delta^p} \leq 2\rho\delta + (1 + 2^{\lambda p})A_p^p\, \frac{h_\ell^{\lambda p}}{\delta^p}.$$

The choice $\delta = h_\ell^{\lambda p/(p+1)}$ balances the two terms and gives the bound.

(iii) Kloeden and Platen (1992) prove Theorem 7.10(i) and Theorem 7.11(i) by bounding $\mathbb{E}|X_T - Y_T^h|^2$ (Theorems 10.2.2 and 10.3.5 there; see also the comment after the proof of Theorem 7.13), which is (12.4) with $p = 2$; then apply (i). □

So the variance of the corrections decays at the rate $\beta = 2\lambda$, twice the strong order: $\beta = 1$ for Euler, $\beta = 2$ for Milstein. This is where the strong order, irrelevant for the bias of prices (Section 7.3), pays off. The coupling is essential to the proof: the two schemes are compared with the same X_T.

For digital payoffs the variance decays more slowly, because a small difference between the fine and coarse values changes the payoff by a full unit whenever X_T is near the strike. With $p = 2$, (ii) gives $\beta \geq \frac{1}{3}$ for Euler and $\beta \geq \frac{2}{3}$ for Milstein. For an autonomous equation with Lipschitz coefficients, such as GBM, the Euler bound (12.4) holds for every $p \geq 2$ (Giles et al., 2009, equation (1.8), referring to Kloeden and Platen, 1992), and letting $p \to \infty$ in (ii) gives $V_\ell = O(h_\ell^{1/2-\delta})$ for every $\delta > 0$; this is the argument of Giles et al. (2009), Section 3, and Avikainen (2009) proves the sharp form of the inequality behind it (Theorem 3.4 of the preprint arXiv:0712.3635). Giles (2015), Table 5.2, lists $O(h^{1/2} \log h)$ for the Euler digital. One cannot hope for much more with Euler: for GBM, the supremum over strikes $K \geq K_0$ of the strong error of the binary payoff is of exact order $h^{1/2}$ (Theorem 7.2 of the preprint), whereas corrections with $\mathbb{E}[Y_\ell^2] \leq C h_\ell^\beta$ uniformly in $K \geq K_0$ for some $\beta > \frac{1}{2}$ would give $\mathbb{E}|P - P_L| = \|P - P_L\|_2^2 \leq \big(\sum_{\ell > L} \|Y_\ell\|_2\big)^2 = O(h_L^\beta)$. For Milstein, Avikainen (2009) proves that the strong error of the binary payoff is $O(h^{1-\delta})$ for every $\delta > 0$ (scalar equation with Lipschitz coefficients and Lipschitz first derivatives in x, and X_T with a bounded density; Theorem 3.7 of the preprint). Since $\mathbb{E}[Y_\ell^2] \leq \mathbb{E}|P_\ell - P| + \mathbb{E}|P_{\ell-1} - P|$ for indicators, $V_\ell = O(h_\ell^{1-\delta})$; Giles et al. (2019) analyse the Milstein corrections in detail.

Example 12.5 (Exact level variances for geometric Brownian motion)**.** Take Euler for GBM and the payoff $P = DS_T$, $D = e^{-rT}$, whose level variances can be computed exactly. Group the fine steps in pairs. Over one coarse step the fine path is multiplied by $f_1 f_2$ with $f_i = a + \sigma \Delta W_i$, $a = 1 + \mu h$, $h = h_\ell$, and the coarse path by $F = A + \sigma(\Delta W_1 + \Delta W_2)$, $A = 1 + 2\mu h$. The $n = 2^{\ell-1}$ coarse steps contribute independent factors, so with $u = \mathbb{E}[(f_1 f_2)^2] = (a^2 + \sigma^2 h)^2$, $w = \mathbb{E}[f_1 f_2 F] = a^2 A + 2a\sigma^2 h$ and $v = \mathbb{E}[F^2] = A^2 + 2\sigma^2 h$,

$$V_\ell = D^2 S_0^2 \Big[u^n - 2w^n + v^n - \big(a^{2n} - A^n\big)^2 \Big].$$

Here $u - 2w + v = \mathbb{E}[(f_1 f_2 - F)^2] = \sigma^4 h^2 + 2\mu^2 \sigma^2 h^3 + \mu^4 h^4$ and $w - v = (A\mu^2 + 2\mu\sigma^2) h^2$. Writing $u^n - 2w^n + v^n = v^n \big[(1 + x)^n - 2(1 + y)^n + 1\big]$ with $x = (u - v)/v$, $y = (w - v)/v$, both $O(h^2)$, and expanding the powers with $n = T/(2h)$ gives $n(x - 2y) + O(n^2 h^4) = \frac{T}{2}\sigma^4 h + O(h^2)$, while $(a^{2n} - A^n)^2 = O(h^2)$ and $v^n \to e^{(2\mu + \sigma^2)T}$. Hence

$$V_\ell = \tfrac{1}{2}\, \sigma^4 T\, \mathbb{E}[P^2]\, h_\ell + O(h_\ell^2), \qquad \mathbb{E}[P^2] = D^2 S_0^2 e^{(2\mu + \sigma^2)T}.$$

For the test case $\mathbb{E}[P^2] = 10^4$ exactly, so $V_\ell = 8 h_\ell + O(h_\ell^2)$: $V_1 = 3.9306$, $V_4 = 0.49890$, $V_7 = 0.062483$. The rate $\beta = 1$ of Theorem 12.4 is attained, not just a bound. For Milstein the same computation gives V_ℓ of exact order h_ℓ^2, with a constant that receives a second contribution (Exercise 12.5).

Figure 12.1 shows these rates on the test case. For the call, V_ℓ halves at each level with Euler (fitted slope 0.99 for $\ell \geq 4$; $V_{10} = 4.6 \times 10^{-3}$) and is divided by four with Milstein (slope 1.99; $V_{10} = 2.0 \times 10^{-7}$), while $\mathrm{Var}(P_\ell)$ stays near 158: at level 10 a correction has a variance 3×10^4 (Euler) and almost 10^9 (Milstein) times smaller than the payoff. The means fall like $2^{-\ell}$ for both schemes (fitted slope 1.06 for Milstein, $\ell \geq 2$; the Euler means at $\ell \geq 8$ are within two standard errors of 0), the weak order $\alpha = 1$ of Theorems 7.10 and 7.11. For the digital, Euler's V_ℓ falls only by $\sqrt{2}$ per level (fitted slope 0.50 for $\ell \geq 4$), and Milstein's by factors between 1.5 and 2.3 for $\ell \geq 4$ (fitted slope 0.85, resting on about 36 nonzero corrections out of 5×10^5 at $\ell = 7$). A digital correction is almost always 0, and its kurtosis $\mathbb{E}[(Y_\ell - \mathbb{E}Y_\ell)^4]/V_\ell^2$ is large: 260 for Euler and 1.4×10^4 for Milstein at $\ell = 7$, against 7.6 for the Euler call. This matters for estimating V_ℓ (Theorem 12.9).

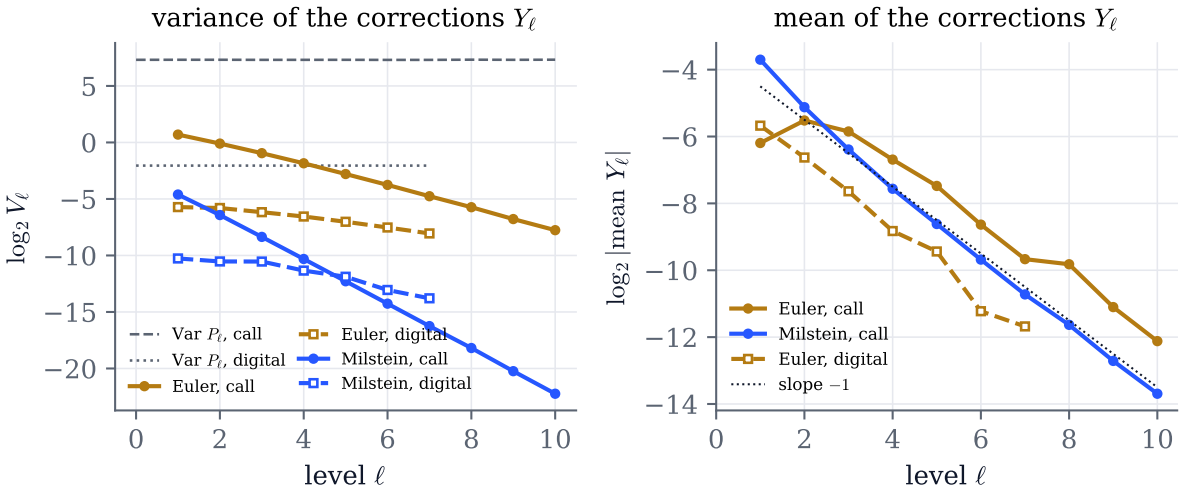

Figure 12.1: Level diagnostics for the test case: Euler and Milstein for GBM, the discounted at-the-money call and digital, levels $\ell = 0, \dots, 10$ (call) and $0, \dots, 7$ (digital); 5×10^5 coupled paths per level for $\ell \le 7$, on which both payoffs are evaluated, and 4×10^4 for $\ell = 8, 9, 10$. Our implementation (NumPy `default_rng`, seed $[2027, s, \ell]$ for scheme s and level ℓ). Left: $\log_2 V_\ell$ for $\ell \ge 1$, and $\log_2 \mathrm{Var}(P_\ell)$ for Milstein (grey). Right: $\log_2$ of the absolute sample mean of Y_ℓ, $\ell \ge 1$; the Milstein digital is omitted, its means being below their standard errors from $\ell = 2$ on.

▶ Interactive demo: level variances and means for the call, the digital and an Asian option, with Euler or Milstein, for any volatility.

Path-dependent payoffs

For a payoff that depends on the path, each level needs a rule that turns a discrete path into a payoff, and the telescoping requires only that the coarse member at level ℓ have the mean of the fine member at level $\ell - 1$. The simplest choice applies the same rule on every grid.

Asian options. For $A = \frac{1}{T}\int_0^T S_t \, \mathrm{d}t$, use the trapezoidal rule on the grid of each path. The payoff is Lipschitz in the path values, but the coarse and fine averages also differ by the difference of the two quadrature errors, whose variance is of order h_ℓ^2. For the discounted test-case Asian call $(A - K)^+$ (our implementation, seeds $[4, \ell]$, $\max(4 \times 10^5 \cdot 2^{-\ell}, 2 \times 10^4)$ paths), the ratio $V_\ell / V_{\ell+1}$ stays between 3.8 and 4.2 for $1 \le \ell \le 7$ with Milstein. With Euler the quadrature term dominates at coarse levels (ratios 3.2 to 3.4 for $\ell \le 3$), and the asymptotic ratio 2 of the strong order $\frac{1}{2}$ emerges only slowly (2.3 and 2.2 for $\ell = 6, 7$). These are the rates $\beta = 1$ and 2 listed by Giles (2015), Section 5.

Extremes and barriers. A maximum, a lookback or a barrier inspected only on the grid has a bias of order $h^{1/2}$ (Theorem 16.7 for GBM, Theorem 16.8 for Euler on general diffusions), so $\alpha = \frac{1}{2}$, and a knock-out indicator is discontinuous, so that with Euler $V_\ell = O(h_\ell^{1/2-\delta})$ for barriers and $O(h_\ell^{1-\delta})$ for lookbacks (Giles et al., 2009). The cure is to look between grid points with the Brownian bridge: use the probability that the bridge over each step crosses the barrier (Theorem 16.9), or sample the extreme of the bridge. This restores $\alpha = 1$, and with Milstein it

gives $\beta \approx \frac{3}{2}$ for barriers and $\beta \approx 2$ for lookbacks (Giles, 2008a). There, the coarse member uses the value of the fine path at the midpoint of each coarse step, which changes its law but not its mean, so the identity (12.1) is preserved; this freedom is why Theorem 12.1 constrains only means.

12.3 Giles' complexity theorem

Theorem 12.6 (Giles). *Let $P \in L^2$ and let Y_ℓ be coupled corrections with variances V_ℓ and expected costs $\tau_\ell > 0$. Suppose that there are constants $\alpha, \beta, \gamma, c_1, c_2, c_3 > 0$ with*

$$\alpha \geq \tfrac{1}{2}\min(\beta, \gamma)$$

such that for every $\ell \in \mathbb{N}$

$$(B)\ \ \big|\mathbb{E}[P_\ell] - \mathbb{E}[P]\big| \leq c_1 2^{-\alpha \ell}, \qquad (V)\ \ V_\ell \leq c_2 2^{-\beta \ell}, \qquad (C)\ \ \tau_\ell \leq c_3 2^{\gamma \ell}.$$

(i) There is a constant c_4, depending only on $\alpha, \beta, \gamma, c_1, c_2, c_3$, such that for every $\varepsilon \in (0, e^{-1})$ the finest level $L = \max\big(0, \lceil \alpha^{-1} \log_2(\sqrt{2}\, c_1/\varepsilon) \rceil\big)$ and the sample sizes (12.3) with $v = \varepsilon^2/2$ give $\mathbb{E}[(\hat{\theta}_L - \theta)^2] \leq \varepsilon^2$ at an expected cost

$$\mathcal{C} \leq \begin{cases} c_4\, \varepsilon^{-2}, & \beta > \gamma, \\ c_4\, \varepsilon^{-2} (\log \varepsilon)^2, & \beta = \gamma, \\ c_4\, \varepsilon^{-2 - (\gamma - \beta)/\alpha}, & \beta < \gamma. \end{cases}$$

(ii) Conversely, suppose that $|\mathbb{E}[P_\ell] - \mathbb{E}[P]| \geq c_1' 2^{-\alpha \ell}$, $V_\ell \geq c_2' 2^{-\beta \ell}$ and $\tau_\ell \geq c_3' 2^{\gamma \ell}$ for every ℓ, with constants $c_i' > 0$. Then there are $c_4' > 0$ and $\varepsilon_0 > 0$ such that, for $\varepsilon < \varepsilon_0$, every estimator (12.2) (any L, any $n_\ell \geq 1$) with mean squared error at most ε^2 costs at least $c_4' \varepsilon^{-2}$, $c_4' \varepsilon^{-2} (\log \varepsilon)^2$ or $c_4' \varepsilon^{-2 - (\gamma - \beta)/\alpha}$ in the three cases.

Proof. (i) *Bias.* If $L = 0$ then $\sqrt{2} c_1 \leq \varepsilon$; otherwise $L \geq \alpha^{-1} \log_2(\sqrt{2} c_1/\varepsilon)$. In both cases (B) gives $|\mathbb{E}[P_L] - \theta| \leq c_1 2^{-\alpha L} \leq \varepsilon/\sqrt{2}$.

Variance and cost. By Theorem 12.2(iv) with $v = \varepsilon^2/2$, the variance is at most $\varepsilon^2/2$, so the mean squared error is at most ε^2 by Theorem 12.2(ii), and

$$\mathcal{C} \leq 2\varepsilon^{-2} S_L^2 + \sum_{\ell=0}^{L} \tau_\ell, \qquad S_L \leq \sqrt{c_2 c_3} \sum_{\ell=0}^{L} 2^{(\gamma - \beta)\ell/2} \tag{12.5}$$

by (V) and (C). Two consequences of the choice of L and of $\varepsilon < e^{-1} < 1$ will be used. First, $L \leq 1 + \alpha^{-1} \log_2^+(\sqrt{2} c_1/\varepsilon)$, so with $c_5 = \max(1, \sqrt{2} c_1)^{1/\alpha}$,

$$2^L \leq 2 c_5\, \varepsilon^{-1/\alpha}.$$

Second, $\log(1/\varepsilon) > 1$, so $L + 1 \leq 2 + \alpha^{-1} \log_2 c_5^\alpha + \alpha^{-1} \log_2(1/\varepsilon) \leq c_6 \log(1/\varepsilon)$ for a constant c_6. The last sum in (12.5), which comes from the integer ceilings, is at most $c_3 2^{\gamma L}/(1 - 2^{-\gamma}) \leq c_7 \varepsilon^{-\gamma/\alpha}$.

Case $\beta > \gamma$. The geometric series gives $S_L \leq \sqrt{c_2 c_3}/(1 - 2^{-(\beta - \gamma)/2})$, so $2\varepsilon^{-2} S_L^2 \leq c\, \varepsilon^{-2}$. Here $\min(\beta, \gamma) = \gamma$, so $\gamma/\alpha \leq 2$ and $\varepsilon^{-\gamma/\alpha} \leq \varepsilon^{-2}$.

Case $\beta = \gamma$. Every term of S_L equals $\sqrt{c_2 c_3}$, so $S_L \le \sqrt{c_2 c_3}\,(L+1) \le \sqrt{c_2 c_3}\,c_6 \log(1/\varepsilon)$ and $2\varepsilon^{-2} S_L^2 \le c\,\varepsilon^{-2}(\log\varepsilon)^2$. Again $\gamma/\alpha \le 2$, and $\varepsilon^{-2} \le \varepsilon^{-2}(\log\varepsilon)^2$.

Case $\beta < \gamma$. Now the series is dominated by its last term: $S_L \le \sqrt{c_2 c_3}\,2^{(\gamma-\beta)L/2}/(1-2^{-(\gamma-\beta)/2})$, and with $2^L \le 2c_5\varepsilon^{-1/\alpha}$, $2\varepsilon^{-2}S_L^2 \le c\,\varepsilon^{-2-(\gamma-\beta)/\alpha}$. Here $\min(\beta,\gamma) = \beta \le 2\alpha$, which is the same as $\gamma/\alpha \le 2 + (\gamma-\beta)/\alpha$, so the ceiling term is of no larger order.

(ii) The mean squared error is at least the squared bias $(\mathbb{E}[P_L]-\theta)^2 \ge c_1'^2 4^{-\alpha L}$, so $2^{-\alpha L} \le \varepsilon/c_1'$, that is $L \ge \alpha^{-1}\log_2(c_1'/\varepsilon)$. It is also at least the variance, so $\mathrm{Var}(\hat\theta_L) \le \varepsilon^2$, and by Theorem 12.2(iv) $\mathcal{C} \ge \varepsilon^{-2} S_L^2$ with $S_L \ge \sqrt{c_2' c_3'}\,\sum_{\ell\le L} 2^{(\gamma-\beta)\ell/2}$. If $\beta > \gamma$, keep the term $\ell = 0$: $\mathcal{C} \ge c_2' c_3' \varepsilon^{-2}$. If $\beta = \gamma$, $S_L \ge \sqrt{c_2' c_3'}(L+1) \ge \sqrt{c_2' c_3'}\,\alpha^{-1}\log_2(c_1'/\varepsilon)$, which is at least a constant times $\log(1/\varepsilon)$ for small ε. If $\beta < \gamma$, keep the term $\ell = L$: $S_L^2 \ge c_2' c_3' 2^{(\gamma-\beta)L} \ge c_2' c_3' (c_1'/\varepsilon)^{(\gamma-\beta)/\alpha}$. □

This is Theorem 3.1 of Giles (2008b), stated there with $\gamma = 1$, $\alpha \ge \frac12$ and a general refinement factor M ($h_\ell = TM^{-\ell}$, which only changes the constants); the present form is Theorem 2.1 of Cliffe et al. (2011), proved in their Appendix A, and Giles (2015), Section 2.1 (Theorem 1 in the author's version). Part (ii) shows that the three rates are not artefacts of the proof: for a hierarchy whose bias, variances and costs behave exactly like powers of 2, no choice of L and n_ℓ does better. Where the work goes explains the three regimes. With $n_\ell\tau_\ell \propto \sqrt{V_\ell\tau_\ell} \propto 2^{(\gamma-\beta)\ell/2}$, the work concentrates on the coarsest levels if $\beta > \gamma$, so the cost is that of an unbiased estimator with bounded cost per sample; it is spread evenly over the $L+1 \approx \alpha^{-1}\log_2(1/\varepsilon)$ levels if $\beta = \gamma$, hence the squared logarithm; and it concentrates on the finest levels if $\beta < \gamma$.

Remark 12.7 (The condition $\alpha \ge \frac12\min(\beta,\gamma)$). It is used only to bound the term $\sum_\ell \tau_\ell$ in (12.5): every level up to L must be sampled at least once, and the bias fixes how many levels there are. It cannot be dropped. Under the lower bounds of (ii), any estimator with mean squared error ε^2 costs at least $\sum_{\ell\le L}\tau_\ell \ge c_3' 2^{\gamma L} \ge c_3'(c_1'/\varepsilon)^{\gamma/\alpha}$, and if $\alpha < \frac12\min(\beta,\gamma)$ this exceeds the bound in (i) by a power of $1/\varepsilon$. For the Euler and Milstein couplings with smooth or Lipschitz payoffs, $\alpha = 1$ and $\gamma = 1$, and the condition holds. For a lookback monitored on the grid with Euler, $\alpha = \frac12$ and $\beta = 1-\delta$ (Section 12.2), and it holds only just.

For plain Monte Carlo with a bias $c\,h^\alpha$ and a cost $h^{-\gamma}$ per path, Theorem 6.11(iii) gives a cost of order $\varepsilon^{-2-\gamma/\alpha}$. The table uses the rates β of Section 12.2, $\gamma = 1$, and for α the weak orders of Theorems 7.10 and 7.11 (for the digital, those of the expansion of Bally and Talay, 1996 recalled in Theorem 7.12, also observed in Figure 12.1); the rate for the Milstein digital is Avikainen's, recalled after Theorem 12.4 (a sample of Y_ℓ costs $2^\ell + 2^{\ell-1} = 3\cdot 2^{\ell-1}$ steps):

payoff and scheme	α	β	multilevel	plain Monte Carlo
Lipschitz, Euler	1	1	$\varepsilon^{-2}(\log\varepsilon)^2$	ε^{-3}
Lipschitz, Milstein	1	2	ε^{-2}	ε^{-3}
digital, Euler	1	$\frac12$	$\varepsilon^{-5/2}$	ε^{-3}
digital, Milstein	1	$1-\delta$	$\varepsilon^{-2-\delta}$	ε^{-3}
barrier on the grid, Euler	$\frac12$	$\frac12-\delta$	$\varepsilon^{-3-2\delta}$	ε^{-4}

For the digital the slower decay of V_ℓ moves Euler from the borderline regime $\beta = \gamma$ into the regime $\beta < \gamma$: the cost of the finest levels dominates, and halving ε multiplies the cost by $2^{5/2} \approx 5.7$ instead of about 4. Milstein restores $\beta \approx \gamma$, and smoothing the payoff by a conditional expectation over the last step pushes β above γ (Exercise 12.7; Giles, 2015, Section 5).

Against the optimal allocation of Chapter 7

Theorem 7.14 chose the best single step for a given budget. Theorem 12.2 chooses the best spread of the same budget over a hierarchy of steps. On the laboratory of Figure 7.3 (GBM with $\mu = -1\%$, $\sigma = 20\%$, the smooth payoff $f(x) = 5\log(1 + e^{(x-100)/5})$, cost in steps), the multilevel root-mean-square error for a budget $\mathcal{C}$ is $\min_L \left(b_L^2 + S_L^2/\mathcal{C}\right)^{1/2}$ (continuous sample sizes, as in Theorem 12.2(iii)), with the exact biases b_L (`paths.gbm_scheme_weak_error`) and the level variances estimated with $\max(4 \times 10^5 \cdot 2^{-\ell}, 2 \times 10^4)$ paths (our implementation, seeds $[3, \ell]$ for Euler and $[4, \ell]$ for Milstein; finest levels $L = 4, 8$ for Euler and $6, 10$ for Milstein). For a fair comparison, the single-level columns use the same exact biases: the error is $\min_m \left(b_m^2 + v\, m/\mathcal{C}\right)^{1/2}$ over integer numbers of steps m, with $v = 154$ (best $m = 7, 34$ for Euler and $5, 19$ for Milstein; the model of Theorem 7.14 with $c = 0.17$ gives 0.041 and 0.0088 for Euler):

budget (steps)	exact	Euler	Milstein	MLMC, Euler	MLMC, Milstein
10^6	0.0124	0.0400	0.0318	0.0227	0.0136
10^8	0.00124	0.00885	0.00659	0.00307	0.00137

Forty independent replications of each multilevel estimator at 10^6 steps have root-mean-square errors 0.026 (Euler) and 0.012 (Milstein) (Figure 12.2, right). At equal budget, the multilevel estimators cut the error of the best single-level estimator with the same scheme by factors of 1.8 and 2.9 (Euler) and 2.3 and 4.8 (Milstein) at 10^6 and 10^8 steps, and the error of multilevel Milstein is only about 10% above that of exact simulation: in terms of cost, it pays a factor of about 1.2 for not knowing the law of S_T.

12.4 The adaptive algorithm

In practice V_ℓ and the bias are unknown, and the number of levels is decided on the fly. The bias test rests on the following observation.

Lemma 12.8 (Estimating the remaining bias)**.** *If* $\mathbb{E}[P_\ell] - \mathbb{E}[P] = c\, 2^{-\alpha\ell}(1 + \eta_\ell)$ *with* $c \neq 0$, $\alpha > 0$ *and* $\eta_\ell \to 0$, *then*

$$\mathbb{E}[P_L] - \mathbb{E}[P] = -\frac{\mathbb{E}[Y_L]}{2^\alpha - 1}\,(1 + o(1)) \qquad (L \to \infty).$$

Proof. $\mathbb{E}[Y_L] = c2^{-\alpha L}(1 + \eta_L - 2^\alpha(1 + \eta_{L-1})) = -c2^{-\alpha L}(2^\alpha - 1)(1 + o(1))$, and $\mathbb{E}[P_L] - \mathbb{E}[P] = c2^{-\alpha L}(1 + o(1))$; divide. □

Algorithm (Giles, 2008b, Section 5; Giles, 2015, Section 3.1). Given ε, rates α and β (known from the theory, or fitted to the level means and variances), and an initial sample size n_{init}:

1. Set $L = 2$ and draw n_{init} samples on levels $0, 1, 2$.
2. Estimate V_ℓ by the sample variance of the samples drawn so far on each level, and set n_ℓ by (12.3) with $v = \varepsilon^2/2$; draw the missing samples. Repeat until no level needs more than 1% additional samples.

3. Estimate the remaining bias by $\max_{j=0,1,2} 2^{-j\alpha}|\hat{Y}_{L-j}|/(2^\alpha - 1)$ (levels $L - j \geq 1$). If it is at most $\varepsilon/\sqrt{2}$, stop and return $\hat{\theta}_L$ with the standard error $(\sum_\ell \hat{V}_\ell/n_\ell)^{1/2}$. Otherwise set $L \leftarrow L + 1$, initialise $\hat{V}_L = \hat{V}_{L-1}2^{-\beta}$, and go to step 2.

Step 3 applies Theorem 12.8 to the sample means, extrapolating from the two levels below L as well, so that a correction whose mean happens to be close to zero does not stop the algorithm too early. The algorithm is a heuristic, and Giles (2015) says so: the lemma is asymptotic, the regime it assumes may not have been reached, and the sample means are noisy, so nothing guarantees a mean squared error below ε^2. Its standard error ignores the bias, which the stopping rule only estimates. Our implementation (`multilevel.mlmc`) also prevents $\hat{V}_\ell$ from falling below half of $\hat{V}_{\ell-1}2^{-\beta}$ for $\ell \geq 2$.

Remark 12.9 (Kurtosis and the variance estimates). The sample variance of n i.i.d. copies of Y has a relative standard deviation of about $\sqrt{(\kappa - 1)/n}$, where κ is the kurtosis of Y: twice the relative error of the sample standard deviation computed in Theorem 6.13. For a digital correction, Y_ℓ is nonzero with a small probability q_ℓ and $\kappa \approx 1/q_\ell$, so the variance estimate is reliable only when $n_\ell q_\ell$, the number of nonzero corrections seen, is large. With 10^3 initial samples and $\kappa = 1.4 \times 10^4$ (Figure 12.1), the estimate of V_ℓ is often exactly 0: the floor in the implementation, or an initial sample size of order κ, is needed. A smoothed payoff has a much smaller kurtosis.

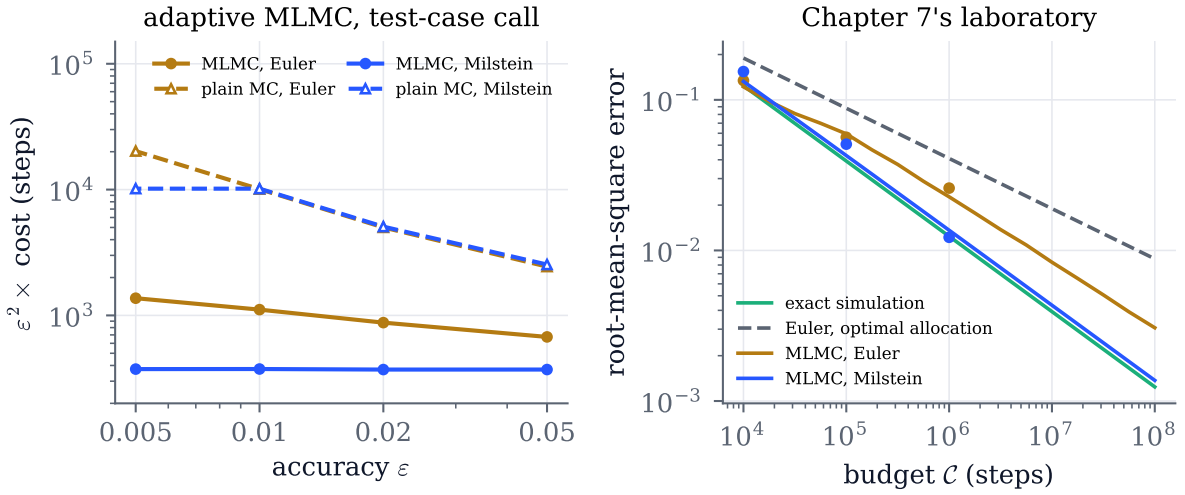

Figure 12.2: Left: the adaptive algorithm on the test-case call ($\alpha = 1$, $\beta = 1$ for Euler and 2 for Milstein, $n_{\text{init}} = 1000$, seed 11; one run per accuracy): ε^2 times the cost in steps, for $\varepsilon = 0.05$, 0.02, 0.01, 0.005. Triangles: plain Monte Carlo on the finest level L chosen by the same run, with $\lceil 2\,\mathrm{Var}(P_L)/\varepsilon^2 \rceil$ paths of 2^L steps ($\mathrm{Var}(P_L)$ from Figure 12.1). Right: Chapter 7's laboratory (Figure 7.3): root-mean-square error against the budget for exact simulation, Euler with the optimal allocation of Theorem 7.14 ($c = 0.17$, $v = 154$), and the best multilevel estimators (lines: model of Section 12.3; dots: 40 replications at 10^4, 10^5 and 10^6 steps).

On the test-case call (Figure 12.2, left; our implementation, seed 11), the algorithm stops at $L = 3, 4, 5, 6$ with Euler and $L = 3, 4, 5, 5$ with Milstein for $\varepsilon = 0.05, 0.02, 0.01, 0.005$. With Milstein, ε^2 times the cost stays at 371–375 steps: the regime $\beta > \gamma$, where the

cost is that of an unbiased estimator. With Euler it grows from 675 to 1370, the slow logarithmic growth of the regime $\beta = \gamma$. At $\varepsilon = 0.005$ the Euler run uses $(n_0, \dots, n_6) = (23497780, 1547621, 827227, 434917, 226194, 114576, 57780)$ samples, returns 7.3703 with standard error 0.0035, and costs 5.5×10^7 steps; plain Monte Carlo on level 6 would cost 15 times more, and the gap grows like $1/(\varepsilon(\log \varepsilon)^2)$. For Milstein the gap is 27. For the digital with Euler ($\beta = \frac{1}{2}$, $n_{\text{init}} = 2000$), averaged over eight runs (seeds 11 to 18), $\varepsilon^2 \times$cost is 10.8, 17.4 and 32.5 at $\varepsilon = 0.004, 0.002, 0.001$, with root-mean-square errors 0.0033, 0.0022, 0.0010. Unlike for the call it grows, as in the regime $\beta < \gamma$ of Theorem 12.6, where it grows like $\varepsilon^{-1/2}$; at these accuracies the growth is irregular (single runs vary by a factor of 2), because the finest level chosen by the algorithm jumps from one run to another.

▶ Interactive demo: run the adaptive algorithm on the call or an Asian option for any accuracy and scheme, see how it spreads the samples over the levels, and compare its cost with plain Monte Carlo.

In code

code/src/mfe230d/multilevel.py: `GBMLevels(payoff, scheme)` returns, for a level ℓ, coupled fine and coarse payoffs computed on shared Brownian increments; `SDELevels` does the same for any scalar equation through `paths.euler` and `paths.milstein`. `level_statistics` produces the diagnostics of Figure 12.1, `mlmc` is the adaptive algorithm, `optimal_samples` implements (12.3), `gbm_level_exact` the exact moments of Theorem 12.5, and `single_term` and `coupled_sum` the estimators of Section 12.5.

12.5 Unbiased randomised estimators

When $\beta > \gamma$, the multilevel cost ε^{-2} is that of an unbiased estimator with bounded cost per sample. McLeish (2011) and Rhee and Glynn (2015) showed that such an estimator exists when, in addition, the bias vanishes fast enough ($2\alpha > \gamma$): draw the level at random, so that every level, however fine, has a chance to be sampled, and reweight. There is then no finest level and no bias.

Proposition 12.10 (Single-term randomised estimator)**.** *Let $Y_0, Y_1, \dots$ be coupled corrections with $\mathbb{E}[P_\ell] \to \mathbb{E}[P]$, let $m_j = \mathbb{E}[Y_j^2]$, and let J be an $\mathbb{N}$-valued random variable, independent of the corrections, with $p_j = \mathbb{Q}(J = j) > 0$ for every j. Draw J, then one sample of Y_J, and set $Z = Y_J/p_J$.*

(i) $\mathbb{E}[Z^2] = \sum_{j \ge 0} m_j/p_j$ (finite or not), and the expected cost of one sample is $\mathbb{E}[\tau_J] = \sum_j p_j \tau_j$.

(ii) If $\sum_j m_j/p_j < \infty$, then Z is integrable, $\mathbb{E}[Z] = \mathbb{E}[P]$, and $\operatorname{Var}(Z) = \sum_j m_j/p_j - \mathbb{E}[P]^2$.

(iii) (Geometric randomisation.) Let $p_j = (1 - 2^{-r})2^{-rj}$ with $r > 0$, and suppose that $m_j \le c\, 2^{-\beta' j}$ and $\tau_j \le c\, 2^{\gamma j}$. Then $\operatorname{Var}(Z) < \infty$ if $r < \beta'$, and $\mathbb{E}[\tau_J] < \infty$ if $r > \gamma$. If instead $m_j \ge c' 2^{-\beta' j}$ and $\tau_j \ge c' 2^{\gamma j}$ with $c' > 0$, then $\operatorname{Var}(Z) = \infty$ for $r \ge \beta'$ and $\mathbb{E}[\tau_J] = \infty$ for $r \le \gamma$. Under (B) and (V) of Theorem 12.6, one can take $\beta' = \min(\beta, 2\alpha)$.

(iv) (No free lunch.) For every distribution (p_j) with all $p_j > 0$,

$$\mathbb{E}[Z^2]\,\mathbb{E}[\tau_J] \;\ge\; \Big(\sum_{j \ge 0} \sqrt{m_j \tau_j}\Big)^2 .$$

> *In particular, if $m_j \geq c' 2^{-\beta' j}$ and $\tau_j \geq c' 2^{\gamma j}$ with $\beta' \leq \gamma$, then no randomisation gives both a finite variance and a finite expected cost.*

Proof. (i) Conditioning on J and using its independence of the corrections, by Tonelli's theorem $\mathbb{E}[Z^2] = \sum_j p_j \mathbb{E}[Y_j^2]/p_j^2 = \sum_j m_j/p_j$; the cost is τ_j on $\{J = j\}$.

(ii) By the Cauchy–Schwarz inequality,

$$\mathbb{E}|Z| = \sum_j \mathbb{E}|Y_j| \leq \sum_j \sqrt{m_j/p_j}\,\sqrt{p_j} \leq \Big(\sum_j \frac{m_j}{p_j}\Big)^{1/2} \Big(\sum_j p_j\Big)^{1/2} < \infty.$$

So Z is integrable and, by Fubini's theorem and Theorem 12.2(i), $\mathbb{E}[Z] = \sum_j \mathbb{E}[Y_j] = \lim_L \sum_{j \leq L} \mathbb{E}[Y_j] = \lim_L \mathbb{E}[P_L] = \mathbb{E}[P]$. The variance follows from (i).

(iii) $m_j/p_j \leq c\, 2^{(r-\beta')j}/(1-2^{-r})$ and $p_j \tau_j \leq c\, 2^{(\gamma - r)j}$ give convergent geometric series when $r < \beta'$, respectively $r > \gamma$. With the lower bounds, the terms are bounded below by $c' 2^{(r-\beta')j}/(1-2^{-r})$ and $c'(1-2^{-r})2^{(\gamma-r)j}$, which do not tend to 0 when $r \geq \beta'$, respectively $r \leq \gamma$; then $\mathbb{E}[Z^2] = \infty$, and $\mathrm{Var}(Z) = \infty$ because $\mathbb{E}[P]$ is finite. Finally, $m_j = V_j + \mathbb{E}[Y_j]^2$ and $|\mathbb{E}[Y_j]| \leq |\mathbb{E}[P_j] - \mathbb{E}[P]| + |\mathbb{E}[P_{j-1}] - \mathbb{E}[P]| \leq c_1(1 + 2^{\alpha}) 2^{-\alpha j}$, so $m_j \leq c_2 2^{-\beta j} + c_1^2 (1 + 2^{\alpha})^2 4^{-\alpha j} \leq c\, 2^{-\min(\beta, 2\alpha) j}$.

(iv) The Cauchy–Schwarz inequality with $a_j = \sqrt{m_j/p_j}$ and $b_j = \sqrt{p_j \tau_j}$, as in Theorem 12.2(iii). If $\beta' \leq \gamma$, then $\sqrt{m_j \tau_j} \geq c' 2^{(\gamma - \beta')j/2} \geq c'$, the series diverges, and one of the two factors on the left is infinite. □

So a finite variance and a finite expected cost are compatible exactly when the second moments $m_j = V_j + \mathbb{E}[Y_j]^2$ shrink faster than the costs grow: if $m_j \asymp 2^{-\beta' j}$ and $\tau_j \asymp 2^{\gamma j}$, by (iii) and (iv) some randomisation works if and only if $\beta' > \gamma$, and then every geometric one with $r \in (\gamma, \beta')$ does. Under (B) and (V), $\beta' = \min(\beta, 2\alpha)$, and the condition $\min(\beta, 2\alpha) > \gamma$ is stronger than the $\beta > \gamma$ of Theorem 12.6: with a bias $c\, 2^{-\gamma \ell/2}$, variances $2^{-\beta \ell}$, $\beta > \gamma$, and costs $2^{\gamma \ell}$, multilevel Monte Carlo costs ε^{-2}, but $m_j \geq \mathbb{E}[Y_j]^2 \asymp 2^{-\gamma j}$ and (iv) rules out every randomisation. When the condition holds, the estimator averages i.i.d. samples with random costs, and by the central limit theorem with random costs (Theorem 6.10) its error for a budget $\mathcal{C}$ is of order $(\mathrm{Var}(Z)\mathbb{E}[\tau_J]/\mathcal{C})^{1/2}$: the square-root rate of exact simulation, with no bias to estimate and a valid confidence interval (Theorem 6.4). Milstein with a Lipschitz payoff qualifies ($\beta = 2$, $\alpha = \gamma = 1$); Euler, with $\beta = \gamma = 1$, does not, and Rhee and Glynn (2015), Section 4, analyse what can still be done then. Equality in (iv) holds for $p_j \propto \sqrt{m_j/\tau_j}$ (Exercise 12.6 treats geometric laws); Rhee and Glynn (2015), Theorems 3 and 4, find the randomisations that minimise the work–variance product $\mathrm{Var}(Z)\mathbb{E}[\tau_J]$.

Remark 12.11 (The coupled-sum estimator). Rhee and Glynn's main estimator uses one path for all the levels up to the random one: $\bar{Z} = \sum_{j=0}^{J} Y_j/\mathbb{Q}(J \geq j)$, where $Y_j = P_j - P_{j-1}$ are all computed from one Brownian path with 2^J steps, coarsened level by level, and J is independent of the path. If $P_j \to P$ in L^2 and $\sum_{j \geq 1} \mathbb{E}[(P_{j-1} - P)^2]/\mathbb{Q}(J \geq j) < \infty$, then $\bar{Z}$ is square integrable and unbiased for $\mathbb{E}[P]$, and its second moment is given explicitly (Rhee and Glynn, 2015, Theorem 1). Level 0 is always included, since $\mathbb{Q}(J \geq 0) = 1$, so a cheap and noisy level 0 is not inflated by a factor $1/p_0$ as in the single-term estimator. The expected cost is $\sum_j 2^j \mathbb{Q}(J \geq j)$ steps.

Figure 12.3 shows the window on the test case. At $r = 1.5$ (our implementation, seed $[5, 150]$, 4×10^5 samples), the single-term estimator returns 7.4075 with standard error 0.0267, at an

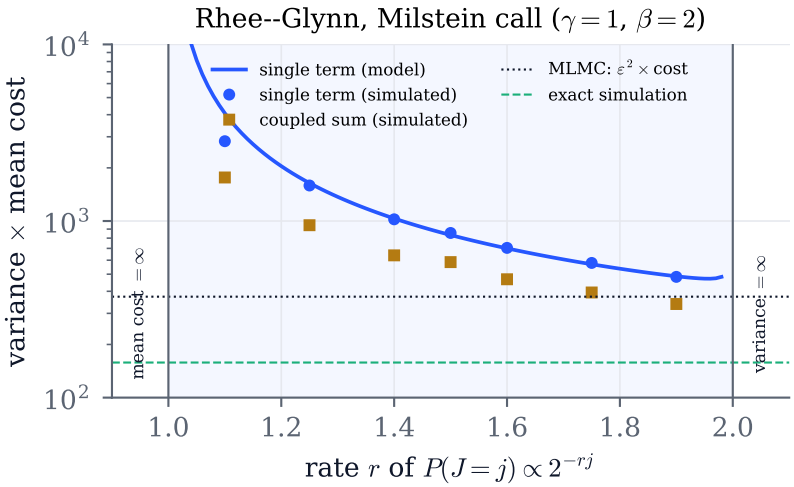

Figure 12.3: Randomised estimators for the test-case call with Milstein, $\mathbb{Q}(J = j) \propto 2^{-rj}$: variance times mean cost per sample, in steps. Line: single-term model of Theorem 12.10(i), with m_j estimated from Figure 12.1 for $j \le 10$ and continued geometrically with $\beta = 2$, $\gamma = 1$. Dots: simulations with 4×10^5 samples per value of r (seeds $[5, 100r]$ and $[6, 100r]$). Shaded: the window $\gamma < r < \beta$ of Theorem 12.10(iii). Horizontal lines: ε^2 times the cost of the adaptive multilevel runs of Figure 12.2 (mean over ε), and the variance 158.0 of exact simulation at one step per sample.

average cost of 3.01 steps per sample (the expected cost is 2.987); the coupled-sum estimator (seed $[6, 150]$) returns 7.3238 with standard error 0.0199 at an average cost of 3.71 steps (expected cost $1/(1 - 2^{-1/2}) = 3.414$; see the first caveat below). The fourteen estimates in the figure lie between -2.4 and $+1.7$ standard errors of 7.36429. As r decreases to $\gamma = 1$, the cost of the rare fine levels dominates and the product blows up; at the other end the variance diverges only at $r = \beta = 2$, and so slowly that the curve barely turns up, because m_j falls by four per level from $m_1 \approx 0.047$ while m_0/p_0, about 300, dominates. Near $r = 1.75$–1.9 the coupled sum costs 340–390 steps per unit of variance, against 373 for ε^2 times the cost of adaptive multilevel Monte Carlo, which spends half of ε^2 on the bias. Two caveats: for $r \le 2$ the cost per sample has infinite variance ($\mathbb{E}[\tau_J^2] = \sum_j p_j \tau_j^2$ diverges), so the realised cost of a run fluctuates a lot, although the central limit theorem with random costs needs only $\mathbb{E}[\tau_J] < \infty$; and the sample variance of Z converges slowly when $\mathbb{E}[Z^4] = \infty$, which for the single-term estimator happens when $r \ge \frac{4}{3}$, since $\mathbb{E}[Z^4] = \sum_j \mathbb{E}[Y_j^4]/p_j^3$ and $\mathbb{E}[Y_j^4]$ is here of the order of $m_j^2 \approx 4^{-2j}$.

▶ Interactive demo: randomised estimators for the call with Euler or Milstein: running means, confidence intervals and running costs for any rate r.

Further reading

Giles (2015) is the standard survey: the complexity theorem and its variants (Sections 2.1–2.4, including randomised and multi-index methods), the algorithm (Section 3.1), and the analysis of path-dependent payoffs with Euler and Milstein (Section 5), with references. Giles (2008b) is

the original paper, Giles (2008a) the Milstein version with the Brownian-bridge treatment of lookbacks, barriers and digitals, and Giles et al. (2009) and Avikainen (2009) the analysis of nonsmooth payoffs with Euler. Kebaier (2005) studied the two-level estimator and its central limit theorem (Exercise 12.4), and Ben Alaya and Kebaier (2015) proved a central limit theorem for the multilevel Euler estimator. In several dimensions, where Milstein needs Lévy areas, the antithetic coupling of Giles and Szpruch (2014) recovers $\beta = 2$ without them for smooth payoffs (about $\frac{3}{2}$ for a call). Rhee and Glynn (2015) and McLeish (2011) are the sources for Section 12.5. Glasserman (2003), which predates multilevel methods, covers the ingredients: discretisation schemes and their orders (Sections 6.1–6.2) and the allocation of effort (Section 1.1.3).

Chapter 13

Monte Carlo Greeks

A price is one number; a hedge needs its derivatives. The delta and the gamma decide the hedge and how often to rebalance it (Chapter 18), the vega and the rhos measure the exposure to volatility and rates, and a risk system recomputes all of them every day for every position. When the price comes from a simulation, so must the Greeks. The course compares three ways to get them: *bumping* an input and pricing again, differentiating the simulated payoff path by path (the course's *analytical differencing*, called the *pathwise* method in the literature), and differentiating the probability density of the simulated state (the *likelihood-ratio* method). The gamma needs mixtures of the last two, and books with many inputs need the *adjoint* method, which computes every first-order Greek of a path at a small multiple of the cost of its price.

The three approaches trade bias, variance and regularity in different ways. Bumping is biased, and done naively its variance explodes as the bump shrinks. The pathwise method is unbiased with a small variance, but it needs a payoff that is Lipschitz in the parameter. The likelihood-ratio method needs only a smooth density, but its variance is usually larger, and it grows with the number of time steps and as $\sigma\sqrt{T} \to 0$. The code is `code/src/mfe230d/greeks.py`.

Throughout, a Greek is the derivative $\alpha'(\theta)$ of $\alpha(\theta) = \mathbb{E}[Y(\theta)]$, where $Y(\theta)$ is a discounted payoff and θ one input. The running example is the test case $S = K = 100$, $T = 1$, $r = 1\%$, $y = 2\%$, $\sigma = 20\%$, with

$$S_T = S\,G, \qquad G = \exp\left(\nu T + \sigma\sqrt{T}\,Z\right), \qquad \nu = r - y - \tfrac{\sigma^2}{2}, \qquad Z \sim \mathcal{N}(0,1)$$

under $\mathbb{Q}$, call price 7.36429, and Black–Scholes Greeks (`bs.greeks`, Theorem 2.7) $\Delta = e^{-yT}N(d_+) = 0.509643$, $\Gamma = 0.0195277$, vega $\mathcal{V} = 39.0554$, theta $\partial_t V = -3.32226$ per year, $\rho_r = 43.6000$ and $\rho_y = -50.9643$. The cash-or-nothing digital call pays $\mathbf{1}_{\{S_T > K\}}$; its price is $e^{-rT}N(d_-) = 0.436000$ and its delta $e^{-rT}N'(d_-)/(S\sigma\sqrt{T}) = 0.0195277$, equal to the gamma of the call because $S\,\Gamma_{\text{call}} = K\,\Delta_{\text{dig}}$ (from $Se^{-yT}N'(d_+) = Ke^{-rT}N'(d_-)$).

13.1 Finite differences

Bumping is the natural first approach: it treats the pricer as a black box and works for any model. The course bumps the spot by a relative amount, $S_\pm = S(1 \pm \varepsilon)$, and the other inputs by absolute amounts (one basis point for r and y, one volatility point for σ). With C_0 and $C_\pm$

the Monte Carlo prices at S and $S_\pm$, the forward, backward and central deltas and the central gamma are

$$\Delta_+ = \frac{C_+ - C_0}{\varepsilon S}, \qquad \Delta_- = \frac{C_0 - C_-}{\varepsilon S}, \qquad \Delta_c = \frac{C_+ - C_-}{2\varepsilon S}, \qquad \Gamma_c = \frac{C_+ + C_- - 2C_0}{(\varepsilon S)^2}. \tag{13.1}$$

Two choices matter: the size $h = \varepsilon S$ of the bump, and whether the three prices use the same random numbers (*common random numbers*, CRN; Theorem 8.11) or independent ones. The next proposition, for a generic input θ, quantifies both.

Proposition 13.1 (Finite differences: bias, variance and the optimal bump)**.** *Let I be an open interval and, for $\theta \in I$, let $Y(\theta) = g(\theta, U)$ with U a random vector and $\mathbb{E}[Y(\theta)^2] < \infty$; write $\alpha(\theta) = \mathbb{E}[Y(\theta)]$ and $\sigma^2(\theta) = \mathrm{Var}(Y(\theta))$. Fix θ and $h > 0$ with $[\theta - h, \theta + h] \subset I$, and let $\hat\Delta_n^F$, $\hat\Delta_n^C$ and $\hat\Gamma_n$ be the averages over n independent replications of the forward difference $(Y(\theta+h) - Y(\theta))/h$, the central difference $(Y(\theta+h) - Y(\theta-h))/(2h)$ and the second difference $(Y(\theta+h) - 2Y(\theta) + Y(\theta-h))/h^2$.*

*(i) (*Bias*, whatever the sampling.) If α is C^2 near θ, then $\mathbb{E}\hat\Delta_n^F - \alpha'(\theta) = \frac{1}{2}\alpha''(\theta)h + o(h)$. If α is C^3, then $\mathbb{E}\hat\Delta_n^C - \alpha'(\theta) = \frac{1}{6}\alpha'''(\theta)h^2 + o(h^2)$. If α is C^4, then $\mathbb{E}\hat\Gamma_n - \alpha''(\theta) = \frac{1}{12}\alpha^{(4)}(\theta)h^2 + o(h^2)$.*

*(ii) (*Independent sampling*: each evaluation within a replication uses its own independent copy of U.) Then*

$$\mathrm{Var}\,\hat\Delta_n^F = \frac{\sigma^2(\theta+h) + \sigma^2(\theta)}{nh^2}, \qquad \mathrm{Var}\,\hat\Delta_n^C = \frac{\sigma^2(\theta+h) + \sigma^2(\theta-h)}{4nh^2},$$
$$\mathrm{Var}\,\hat\Gamma_n = \frac{\sigma^2(\theta+h) + 4\sigma^2(\theta) + \sigma^2(\theta-h)}{nh^4}.$$

If σ^2 is continuous and positive at θ, these are of exact order $1/(nh^2)$, $1/(nh^2)$ and $1/(nh^4)$.

*(iii) (*Common random numbers*: one U per replication.) Then $\mathrm{Var}\,\hat\Delta_n^C = \mathrm{Var}\,(Y(\theta+h) - Y(\theta-h))/(4nh^2)$, and similarly for the others.*

(a) If there are $\delta > 0$ and a random variable κ with $\mathbb{E}\kappa^2 < \infty$ and $|Y(\theta_1) - Y(\theta_2)| \le \kappa|\theta_1 - \theta_2|$ almost surely for all $\theta_1, \theta_2 \in (\theta - \delta, \theta + \delta)$ (a Lipschitz *payoff), then $\mathrm{Var}\,\hat\Delta_n^F$ and $\mathrm{Var}\,\hat\Delta_n^C$ are at most $\mathbb{E}\kappa^2/n$ for every $h < \delta$.*

(b) If $\mathbb{E}[(Y(\theta_1) - Y(\theta_2))^2] \le c|\theta_1 - \theta_2|$ for θ_1, θ_2 near θ (a payoff with a jump, Theorem 13.2), then $\mathrm{Var}\,\hat\Delta_n^F \le c/(nh)$ and $\mathrm{Var}\,\hat\Delta_n^C \le c/(2nh)$.

(c) If $|Y(\theta+h) - 2Y(\theta) + Y(\theta-h)| \le h\kappa \mathbf{1}_{A_h}$ for a random $\kappa \ge 0$ and events A_h with $\mathbb{E}[\kappa^2 \mathbf{1}_{A_h}] \le ch$ (a payoff with a kink, Theorem 13.3), then $\mathrm{Var}\,\hat\Gamma_n \le c/(nh)$.

*(iv) (*Optimal bump.*) Suppose that, as $h \downarrow 0$, the bias is $bh^\beta(1 + o(1))$ and n times the variance is $vh^{-\gamma}(1 + o(1))$, with constants $b \neq 0$, $v > 0$, $\beta > 0$, $\gamma > 0$ that do not depend on n. Let $c^* = \big(\gamma v/(2\beta b^2)\big)^{1/(2\beta+\gamma)}$. For every sequence $h_n \to 0$,*

$$\liminf_{n\to\infty} n^{2\beta/(2\beta+\gamma)}\, \mathbb{E}[(\hat\Delta_n - \alpha'(\theta))^2] \;\ge\; \Big(1 + \frac{2\beta}{\gamma}\Big) b^2 (c^*)^{2\beta},$$

with equality (and a limit) for $h_n = c^* n^{-1/(2\beta+\gamma)}$; here $\hat\Delta_n$ is the estimator built with the bump h_n, and $\alpha'(\theta)$ is replaced by $\alpha''(\theta)$ for $\hat\Gamma_n$. In case (iii)(a), the mean squared error is at most $b^2h^{2\beta}(1 + o(1)) + \mathbb{E}\kappa^2/n$, which is $O(1/n)$ for every $h_n = O(n^{-1/(2\beta)})$.

Proof. (i) By Taylor's theorem with the Lagrange remainder, $\alpha(\theta+h)=\alpha(\theta)+\alpha'(\theta)h+\frac{1}{2}\alpha''(\xi)h^2$ with ξ between θ and $\theta+h$, and $\alpha''(\xi)\to\alpha''(\theta)$ by continuity. For the central difference, expand $\alpha(\theta\pm h)$ to third order: the even terms cancel and $\alpha(\theta+h)-\alpha(\theta-h)=2\alpha'(\theta)h+\frac{1}{6}(\alpha'''(\xi_+)+\alpha'''(\xi_-))h^3$. For the second difference, expand to fourth order: the odd terms cancel and $\alpha(\theta+h)+\alpha(\theta-h)-2\alpha(\theta)=\alpha''(\theta)h^2+\frac{1}{24}(\alpha^{(4)}(\xi_+)+\alpha^{(4)}(\xi_-))h^4$. In every case the expectation of the estimator is the same difference of α, by linearity, whatever the dependence between the evaluations.

(ii) The replications are independent and, within one, the evaluations are independent, so the variances add: $\operatorname{Var}(aY_1+bY_2+cY_3)=a^2\sigma_1^2+b^2\sigma_2^2+c^2\sigma_3^2$, and averaging over n replications divides by n.

(iii) The first claim is the variance of an average of i.i.d. terms. (a) With $|Y(\theta+h)-Y(\theta-h)|\le 2h\kappa$, the variance of one central quotient is at most its second moment, $\le\mathbb{E}\kappa^2$; the forward quotient is the same. (b) Likewise, the variance of one quotient is at most ch/h^2 (forward) or $2ch/(4h^2)$ (central). (c) The second moment of one second difference is at most $h^2\mathbb{E}[\kappa^2\mathbf{1}_{A_h}]/h^4\le c/h$.

(iv) The mean squared error is the squared bias plus the variance (Theorem 6.11(i)). Write $h_n=c_n n^{-1/(2\beta+\gamma)}$. Then $n^{2\beta/(2\beta+\gamma)}h_n^{2\beta}=c_n^{2\beta}$ and $n^{2\beta/(2\beta+\gamma)}h_n^{-\gamma}/n=c_n^{-\gamma}$, so

$$n^{2\beta/(2\beta+\gamma)}\,\mathrm{MSE}=b^2c_n^{2\beta}(1+o(1))+vc_n^{-\gamma}(1+o(1))\;\ge\;(1-\eta)\,\phi(c_n)\;\ge\;(1-\eta)\,\phi(c^*)$$

for every $\eta\in(0,1)$ and n large, where $\phi(c)=b^2c^{2\beta}+vc^{-\gamma}$. As in the proof of Theorem 6.11(iii), ϕ decreases on $(0,c^*]$ and increases on $[c^*,\infty)$, so c^* is its global minimiser, and $vc^{*-\gamma}=\frac{2\beta}{\gamma}b^2c^{*2\beta}$ gives $\phi(c^*)=(1+2\beta/\gamma)b^2c^{*2\beta}$. With $c_n=c^*$ the two $o(1)$ terms vanish in the limit. The last claim combines the bias with the bound of (iii)(a). □

The proposition gives the familiar table of rates (Glasserman, 2003, Sections 7.1.1–7.1.2), due to Glynn (1989) and L'Ecuyer and Perron (1994):

	β	γ	best bump	best MSE	RMSE
forward, independent	1	2	$n^{-1/4}$	$n^{-1/2}$	$n^{-1/4}$
central, independent	2	2	$n^{-1/6}$	$n^{-2/3}$	$n^{-1/3}$
forward, CRN, payoff with a jump	1	1	$n^{-1/3}$	$n^{-2/3}$	$n^{-1/3}$
central, CRN, payoff with a jump	2	1	$n^{-1/5}$	$n^{-4/5}$	$n^{-2/5}$
central or forward, CRN, Lipschitz payoff	1 or 2	—	small enough	n^{-1}	$n^{-1/2}$
second difference, independent	2	4	$n^{-1/8}$	$n^{-1/2}$	$n^{-1/4}$
second difference, CRN, payoff with a kink	2	1	$n^{-1/5}$	$n^{-4/5}$	$n^{-2/5}$

In the table, only common random numbers with a Lipschitz payoff reach the Monte Carlo rate $n^{-1/2}$; higher-order differences come closer without reaching it (Exercise 13.2). L'Ecuyer and Perron (1994) prove the rate $n^{-1/2}$ under the conditions that also make the pathwise method of Section 13.2 valid. The last line is Theorem 13.3. Everything assumes $b\neq 0$. If $b=0$ the bias is of higher order and the rates improve; if b is merely small, the asymptotic rates set in only for very large n (Figure 13.1(c)).

Example 13.2 (The delta of a digital)**.** Let $Y(S)=e^{-rT}\mathbf{1}_{\{SG>K\}}$ with common draws. Then $Y(S+h)-Y(S-h)=e^{-rT}\mathbf{1}_{\{K/(S+h)<G\le K/(S-h)\}}$, so

$$\mathbb{E}\big[(Y(S+h)-Y(S-h))^2\big]=e^{-2rT}\big(N(z_{-h})-N(z_{+h})\big),\qquad z_{\pm h}=\frac{\log(K/(S\pm h))-\nu T}{\sigma\sqrt{T}}.$$

By the mean value theorem,

$$N(z_{-h}) - N(z_{+h}) = N'(\xi_h)(z_{-h} - z_{+h}), \qquad z_{-h} - z_{+h} = \frac{1}{\sigma\sqrt{T}} \log \frac{S+h}{S-h} = \frac{2h}{S\sigma\sqrt{T}}(1 + o(1)),$$

with $\xi_h \to -d_-$. So the second moment is $2he^{-rT}\Delta_{\text{dig}}(1 + o(1))$: case (iii)(b) with $\gamma = 1$ and $v = e^{-rT}\Delta_{\text{dig}}/2$. The bias of the central difference is $\frac{1}{6}\alpha'''(S)h^2$, with $\alpha'''(S) = -4.82 \times 10^{-5}$ for the test case, and (iv) gives $h^* = 2.06$ at $n = 10^6$, a 2% bump; minimising the exact mean squared error (bias from the closed form, variance by quadrature, `greeks.optimal_bump`) gives $h^* = 2.07$ and a root mean squared error of 7.39×10^{-5}, 0.38% of the delta. Without common random numbers the best is 1.35×10^{-4}, and the likelihood-ratio estimator of Section 13.3 has standard error $0.0290/\sqrt{n} = 2.90 \times 10^{-5}$ with no bias at all: at equal accuracy it needs 6.5 times fewer draws than the best bump.

Remark 13.3 (The bumped gamma is a kernel estimator). For the call with common draws, $(x+h)^+ - 2x^+ + (x-h)^+ = (h - |x|)^+$ gives, sample by sample,

$$\frac{Y(S+h) - 2Y(S) + Y(S-h)}{h^2} = e^{-rT} G\, k_h(S - K/G), \qquad k_h(u) = \frac{(h - |u|)^+}{h^2},$$

and Δ_c is $e^{-rT}G$ times the ramp $\int_{-\infty}^{S-K/G} \mathbf{1}_{\{|u|<h\}}\, \mathrm{d}u/(2h)$. Each draw has a *critical spot* K/G, the initial price at which it finishes at the money. The bumped gamma $\hat{\Gamma}_n = \frac{1}{n}\sum_i e^{-rT} G_i k_h(S - K/G_i)$ is a kernel estimator, with the triangular kernel k_h of width $2h$ and integral 1, of the density of the critical spots weighted by $e^{-rT}G$, which is the gamma. Since the tent is at most h, each second difference is at most $he^{-rT}G\mathbf{1}_{A_h}$ with $A_h = \{|S - K/G| < h\}$, and $\mathbb{E}[e^{-2rT}G^2\mathbf{1}_{A_h}] = O(h)$ because K/G has a bounded density near S and $G \le 2K/S$ on A_h for $h < S/2$: Theorem 13.1(iii)(c) bounds the variance by a multiple of $1/(nh)$, and the bias is $O(h^2)$, the last line of the table. The course's region-by-region analysis is this decomposition: regions II and III, where $K/(1+\varepsilon) < S_T < K/(1-\varepsilon)$, are the two sides of the tent, and outside them a draw contributes nothing to Γ_c. The number of draws in the tent is binomial, with mean $n\,\mathbb{Q}(|S - K/G| < h)$: 4.0 for the course's $n = 1024$ and $\varepsilon = 0.001$, 40 for $\varepsilon = 0.01$ and 388 for $\varepsilon = 0.1$. With four draws the estimate is a handful of tents and zero in between, the "butterflies" of Figure 13.2(b). The deltas are sums of ramps, which become steps as $h \to 0$: the "$1/n$ quantization" of the course.

Figure 13.1(a) shows the two regimes for the call delta. With independent draws the error grows like $1/\varepsilon$ for small bumps, and the best forward difference ($\varepsilon = 4.4\%$) has error 0.061. With common random numbers the central difference has error 0.0058 for every bump up to 3%, ten times smaller, which is worth a hundred times more paths; the forward difference stays there up to 0.1%, after which its bias of order h takes over. A bump of 10^{-4} is as good as a bump of 10^{-2}: with a Lipschitz payoff and common draws, small bumps cost nothing. Rounding does not interfere at these sizes, since in double precision a difference quotient loses about $10^{-16}|Y|/h$, far below the statistical error. Figure 13.2 reproduces the course's experiment: $n = 1024$ draws, reused for every S_0. The forward and backward deltas bracket the central one, $\Delta_+ - \Delta_- = \varepsilon S\,\Gamma_c \ge 0$ sample by sample because the payoff is convex, and the gammas are biased for $\varepsilon = 0.1$ and useless for $\varepsilon = 0.001$.

▶ Interactive demo: move the bump, the number of draws and the payoff, and watch the two regimes of the error.

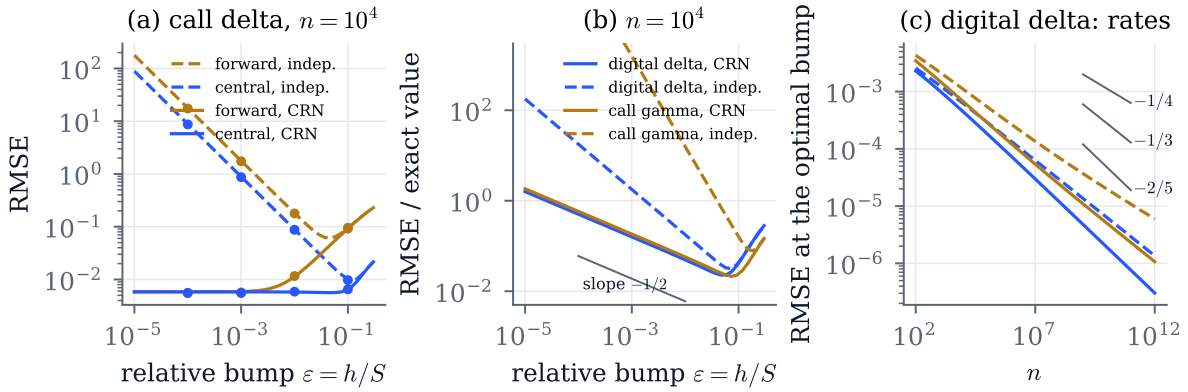

Figure 13.1: Finite differences for the test case, exact curves (bias from the closed forms, variance by quadrature). (a) Root mean squared error of the call delta against the relative bump, $n = 10^4$; dots: 200 independent runs each (our implementation, `default_rng`, seeds 100–103). With CRN the error is flat at the pathwise standard error 0.0058. (b) Error relative to the exact value for the central digital delta and the second-difference call gamma, $n = 10^4$; with CRN the two curves almost coincide (their exact values are equal and both variances are of order $1/(nh)$). (c) Error at the optimal bump against n for the digital delta; reference slopes $-1/4$, $-1/3$, $-2/5$. The forward curves approach their slopes slowly because the digital's $\alpha'' = -4.9 \times 10^{-5}$ is small at the test case.

13.2 Pathwise derivatives

Let the bump go to zero with common random numbers. If $\theta \mapsto Y(\theta)$ is differentiable along each path, the difference quotients converge to the derivative $Y'(\theta)$ of the payoff along the path, and the question is whether $\mathbb{E}[Y'(\theta)] = \alpha'(\theta)$: whether the derivative and the expectation can be interchanged. Almost sure differentiability is not enough (Theorem 13.5); what is needed is a dominating Lipschitz bound.

Theorem 13.4 (Pathwise derivatives). *Let I be an open interval, $\theta_0 \in I$, and $(Y(\theta))_{\theta \in I}$ random variables with $\mathbb{E}|Y(\theta_0)| < \infty$. Assume*

(PW1) for almost every ω, the map $\theta \mapsto Y(\theta, \omega)$ is differentiable at θ_0, with derivative $Y'(\theta_0)$;

(PW2) there are $\delta > 0$ and a random variable κ with $\mathbb{E}\kappa < \infty$ such that, almost surely, $|Y(\theta_1) - Y(\theta_2)| \le \kappa|\theta_1 - \theta_2|$ for all $\theta_1, \theta_2 \in (\theta_0 - \delta, \theta_0 + \delta)$.

Then $Y(\theta)$ is integrable for $|\theta - \theta_0| < \delta$, α is differentiable at θ_0, and $\alpha'(\theta_0) = \mathbb{E}[Y'(\theta_0)]$. If moreover $\mathbb{E}\kappa^2 < \infty$, then $\operatorname{Var} Y'(\theta_0) \le \mathbb{E}\kappa^2$ and the common-random-number difference quotients converge to $Y'(\theta_0)$ in L^2 as $h \to 0$.

Sufficient conditions *(Glasserman, 2003, Section 7.2.2, conditions (A1)–(A4)). Let $Y(\theta) = f(X(\theta))$ with $X(\theta)$ a random vector in $\mathbb{R}^m$ and $f : \mathbb{R}^m \to \mathbb{R}$ Borel. Then (PW1) and (PW2) hold if: (a) $X(\theta)$ is almost surely differentiable at θ_0; (b) $\mathbb{P}(X(\theta_0) \in D_f) = 1$, where D_f is the set of points at which f is differentiable; (c) f is Lipschitz, $|f(x) - f(x')| \le L_f \sum_i |x_i - x_i'|$; (d) there are random variables κ_i with $\mathbb{E}\kappa_i < \infty$ and $|X_i(\theta_1) - X_i(\theta_2)| \le \kappa_i|\theta_1 - \theta_2|$ on a neighbourhood of θ_0. Then $\kappa = L_f \sum_i \kappa_i$ and $Y'(\theta_0) = \nabla f(X(\theta_0)) \cdot X'(\theta_0)$. In particular (PW2) holds, by the mean value theorem, whenever $\theta \mapsto Y(\theta)$ is almost surely continuously*

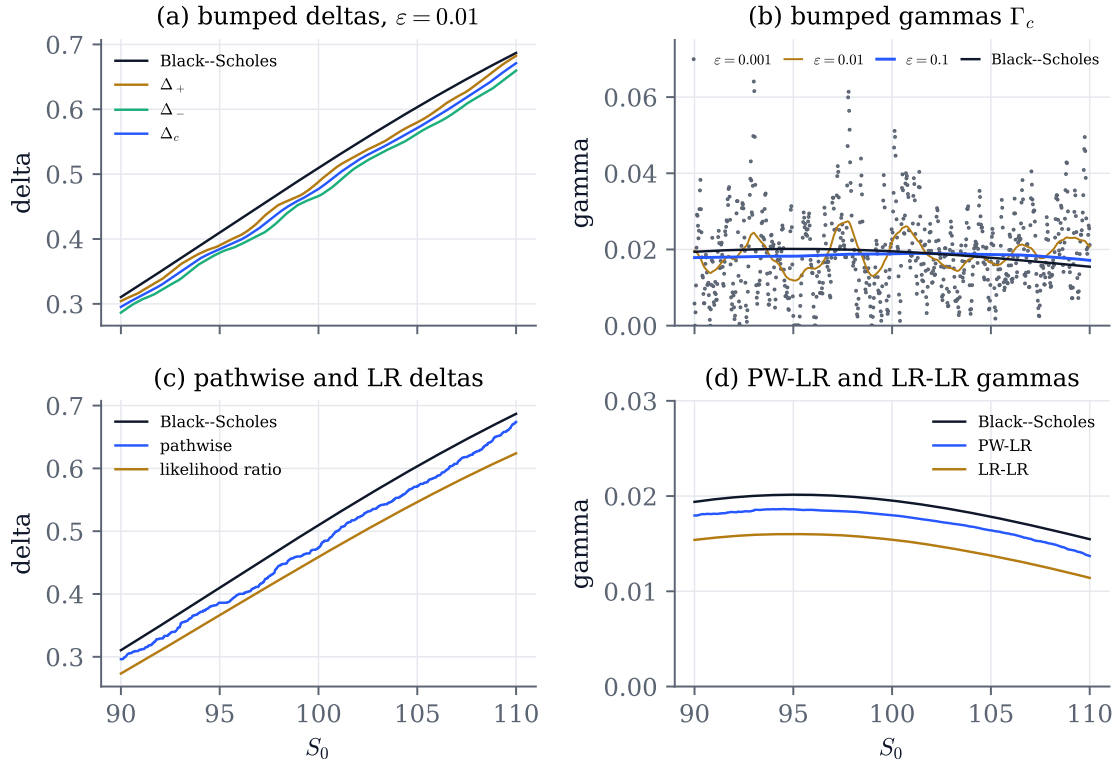

Figure 13.2: The course's shoot-out on the test-case call, $n = 1024$ draws reused for every S_0 (our implementation, `default_rng`, seed 1024). (a) Bumped deltas (13.1), $\varepsilon = 0.01$. (b) Bumped gammas Γ_c for $\varepsilon = 0.1$, 0.01 and 0.001 (dots; values above 0.075 are cut). (c) Pathwise and likelihood-ratio deltas, (13.2) and (13.5). (d) The PW-LR and LR-LR gammas of Theorem 13.8. Black: Black–Scholes.

differentiable on the neighbourhood with $\mathbb{E}\sup_{|\theta-\theta_0|<\delta}|Y'(\theta)| < \infty$.

Proof. For $|\theta - \theta_0| < \delta$, $|Y(\theta)| \le |Y(\theta_0)| + \kappa\delta$ is integrable. Let $h_k \to 0$ with $0 < |h_k| < \delta$. By (PW1) the quotients $Q_k = (Y(\theta_0+h_k)-Y(\theta_0))/h_k$ converge to $Y'(\theta_0)$ almost surely, and by (PW2) $|Q_k| \le \kappa$, integrable. By dominated convergence, $(\alpha(\theta_0 + h_k) - \alpha(\theta_0))/h_k = \mathbb{E}Q_k \to \mathbb{E}Y'(\theta_0)$; the sequence being arbitrary, $\alpha'(\theta_0) = \mathbb{E}Y'(\theta_0)$. Also $|Y'(\theta_0)| = \lim|Q_k| \le \kappa$ almost surely, so $\operatorname{Var} Y'(\theta_0) \le \mathbb{E}\kappa^2$; and $(Q_k - Y'(\theta_0))^2 \le 4\kappa^2$, so $\mathbb{E}(Q_k - Y'(\theta_0))^2 \to 0$ by dominated convergence again; the central quotients are the average of a forward and a backward one.

For the sufficient conditions: on the almost sure event where (a) and (b) hold, f is differentiable at $X(\theta_0)$ and X at θ_0, and the chain rule gives (PW1) with the stated derivative. On the event where (d) holds, $|f(X(\theta_1))-f(X(\theta_2))| \le L_f\sum_i|X_i(\theta_1)-X_i(\theta_2)| \le L_f\sum_i\kappa_i|\theta_1-\theta_2|$. Finally, if $\theta \mapsto Y(\theta)$ is C^1 on the interval, the mean value theorem gives $|Y(\theta_1) - Y(\theta_2)| \le \sup|Y'|\,|\theta_1 - \theta_2|$. □

Common random numbers and the pathwise method are two views of one thing: under (PW2) with $\mathbb{E}\kappa^2 < \infty$, the bumped quotients with common draws converge in L^2 to the pathwise estimator as $h \to 0$, and differ from it in mean by the bias of order h or h^2. This is why the error in Figure 13.1(a) is flat.

Black–Scholes. Differentiating $S_T = S\exp(\nu T + \sigma\sqrt{T}Z)$ gives $\partial_S S_T = G$, $\partial_\sigma S_T = S_T(\sqrt{T}Z - \sigma T)$, $\partial_r S_T = TS_T$, $\partial_y S_T = -TS_T$ and $\partial_T S_T = S_T(\nu + \sigma Z/(2\sqrt{T}))$. For the call, $f'(s) = \mathbf{1}_{\{s>K\}}$ off the strike, and the pathwise estimators (the course's analytical differencing) are

$$\begin{aligned} \Delta^{\mathrm{pw}} &= e^{-rT}\mathbf{1}_{\{S_T>K\}}G, & \mathcal{V}^{\mathrm{pw}} &= e^{-rT}\mathbf{1}_{\{S_T>K\}}S_T(\sqrt{T}Z - \sigma T), \\ \rho_r^{\mathrm{pw}} &= TKe^{-rT}\mathbf{1}_{\{S_T>K\}}, & \rho_y^{\mathrm{pw}} &= -Te^{-rT}\mathbf{1}_{\{S_T>K\}}S_T = -TS\,\Delta^{\mathrm{pw}}, \\ \partial_K^{\mathrm{pw}} &= -e^{-rT}\mathbf{1}_{\{S_T>K\}}, & \Theta^{\mathrm{pw}} &= re^{-rT}(S_T - K)^+ - e^{-rT}\mathbf{1}_{\{S_T>K\}}S_T\Big(\nu + \frac{\sigma Z}{2\sqrt{T}}\Big), \end{aligned} \tag{13.2}$$

where ρ_r^{pw} combines the discount factor, $-Te^{-rT}(S_T - K)^+$, with the drift, $Te^{-rT}\mathbf{1}_{\{S_T>K\}}S_T$, and Θ^{pw} estimates $\partial_t V = -\partial_T V$. The sufficient conditions hold for every input when $S, \sigma, T > 0$: f is 1-Lipschitz, $\mathbb{Q}(S_T = K) = 0$ because S_T has a density, and for the input in a small neighbourhood (with T bounded away from 0) each derivative of S_T is bounded by an integrable variable of the form $c_1 e^{c_2|Z|}(1 + |Z|)$. So all six are unbiased, as Table 13.1 confirms; the put is the same with $f'(s) = -\mathbf{1}_{\{s<K\}}$. Two of them are the course's "pathwise proportionality": $\rho_y^{\mathrm{pw}} = -TS\Delta^{\mathrm{pw}}$ and $\rho_r^{\mathrm{pw}} = -TK\partial_K^{\mathrm{pw}}$ hold draw by draw, because the discounted payoff $(Se^{-yT}e^{-\sigma^2 T/2 + \sigma\sqrt{T}Z} - Ke^{-rT})^+$ depends on S and y only through Se^{-yT}, and on K and r only through Ke^{-rT} (Exercise 13.1). The bumped ρ_y and Δ of the course move together for the same reason.

The edges deserve a look. At $S = 0$ write $\Delta^{\mathrm{pw}} = e^{-rT}\mathbf{1}_{\{SG>K\}}G$, which is 0 for $K > 0$, the limit of $e^{-yT}N(d_+)$. For $K = 0$ the payoff is the share, $\Delta^{\mathrm{pw}} = e^{-rT}G$ has mean e^{-yT} and variance $e^{-2yT}(e^{\sigma^2 T} - 1)$, small when $\sigma^2 T$ is. As $\sigma\sqrt{T} \to 0$ the hypothesis $\mathbb{Q}(S_T = K) = 0$ survives for every $\sigma > 0$, but at $\sigma = 0$ with K equal to the forward it fails, and indeed the price $(Se^{-yT} - Ke^{-rT})^+$ has a kink there. As $T \downarrow 0$ the theta estimator contains $\sigma Z/(2\sqrt{T})$. At and in the money its variance grows like $1/T$ ($T\operatorname{Var}\Theta^{\mathrm{pw}}$ tends to $\frac{1}{4}S^2\sigma^2\operatorname{Var}(Z^+) = 34.1$ at $S = 100$ and to $\frac{1}{4}S^2\sigma^2 = 121$ at $S = 110$), while out of the money it tends to 0 (37.6, 0.019 at $T = 1$, 0.01 for $S = 90$); the theta itself grows like $1/\sqrt{T}$ at the money and stays bounded away from it.

Remark 13.5 (Why digitals fail, and three remedies). For the digital, $Y(S) = e^{-rT}\mathbf{1}_{\{SG>K\}}$ is differentiable at S with derivative 0 for almost every draw, so the pathwise estimator is identically zero, while the delta is 0.0195. Condition (PW2) is what fails. For a draw with critical spot K/G at distance $u = |S - K/G| < \delta$ from S, the quotient $(Y(S') - Y(S))/(S' - S)$ equals $e^{-rT}/|S' - S|$ for S' just across K/G, so any admissible κ must exceed e^{-rT}/u; and u has a density bounded below near 0, so $\mathbb{E}[1/u; u < \delta] = \infty$. Indeed no integrable κ can exist, since otherwise the theorem would give the delta 0. The same happens for barrier options, whose payoff jumps when the path touches the barrier, and for the pathwise derivative of the call's pathwise delta, which jumps at the strike (Section 13.4). Three remedies:

(a) *Smoothing*: replace the payoff by a Lipschitz approximation, such as the call spread $((S_T - K + \eta)^+ - (S_T - K)^+)/\eta$, and accept a bias of order η; the variance of its pathwise delta grows like $1/\eta$, the trade-off of Theorem 13.1(iv) again.

(b) *Likelihood ratios*, which differentiate the density instead of the payoff (Section 13.3).

(c) *Conditional expectation*: replace Y by $\mathbb{E}[Y \mid \mathcal{G}]$ for a σ-algebra $\mathcal{G}$ such that the conditional payoff is a smooth function of the input; this keeps the mean, lowers the variance (Theorem 10.11) and makes the pathwise method valid. For a payoff on a simulated path, conditioning on everything but the last step replaces the indicator of the last step by a Black–Scholes digital price (Exercise 13.3).

Along a path. For an Asian call on the arithmetic average $A = \frac{1}{m}\sum_j S_{t_j}$ of m equally spaced fixings, $t_j = jT/m$, simulated exactly with `paths.gbm_paths` (Theorem 7.2) from $S_{t_j} = S\exp(\nu t_j + \sigma W_{t_j})$,

$$\Delta^{\mathrm{pw}} = e^{-rT}\mathbf{1}_{\{A>K\}}\frac{A}{S}, \qquad \mathcal{V}^{\mathrm{pw}} = e^{-rT}\mathbf{1}_{\{A>K\}}\frac{1}{m}\sum_{j=1}^{m} S_{t_j}(W_{t_j} - \sigma t_j). \tag{13.3}$$

The sufficient conditions hold as for the call (A has a density). For the geometric average A^{g} the same formulas, with $\partial_\sigma A^{\mathrm{g}} = A^{\mathrm{g}} \cdot \frac{1}{m}\sum_j (W_{t_j} - \sigma t_j)$, can be checked against the closed form (Theorem 15.2): with $m = 12$ the exact delta and vega are 0.486882 and 22.4010, and 4×10^5 paths give 0.4866(9) and 22.42(6) (our implementation, seed 3; standard errors in units of the last digit). The arithmetic Asian gives 0.4951(9) and 24.07(7) from the same paths.

When the path is produced by a discretisation scheme, the pathwise derivative is propagated along with it. For the Euler scheme $X_{k+1} = X_k + a(X_k)h + b(X_k)\Delta W_k$ of Theorem 7.10, the derivative $D_k = \partial X_k/\partial x_0$ obeys

$$D_{k+1} = D_k(1 + a'(X_k)h + b'(X_k)\Delta W_k), \qquad D_0 = 1, \tag{13.4}$$

and $\partial_{x_0} f(X_N) = f'(X_N)D_N$. If $|a'|, |b'| \le L$ and f is Lipschitz, each step is Lipschitz in its input with constant $1 + Lh + L|\Delta W_k|$, so (PW2) holds with $\kappa = L_f \prod_k (1 + Lh + L|\Delta W_k|)$, whose mean is the product of the means of independent factors. If b does not vanish, X_N has a density (given X_{N-1} it is Gaussian), so (PW1) holds for every f that is differentiable except at finitely many points, such as calls, puts and their combinations. For geometric Brownian motion, $D_N = X_N/x_0$ exactly. This is *forward mode*: one derivative per input, computed alongside the path. Section 13.5 runs it backwards.

13.3 The likelihood ratio method

The pathwise method moves the payoff and keeps the law. The likelihood-ratio method keeps the payoff and moves the law. Let X have density g_θ on $\mathbb{R}^k$, let the payoff f not depend on θ, and write $\alpha(\theta) = \mathbb{E}_\theta[f(X)] = \int f\, g_\theta \,\mathrm{d}x$.

Theorem 13.6 (Likelihood ratio method)**.** *Let $\theta_0 \in I$, an open interval, $f : \mathbb{R}^k \to \mathbb{R}$ Borel, and assume:*

(LR1) there is $\delta > 0$ such that, for every x, $\theta \mapsto g_\theta(x)$ is differentiable on $U = (\theta_0 - \delta, \theta_0 + \delta)$, with derivative $\dot g_\theta(x)$;

(LR2) there is a Borel $H \ge 0$ with $|\dot g_\theta(x)| \le H(x)$ for all $\theta \in U$ and x, and $\int_{\mathbb{R}^k} |f(x)|H(x)\,\mathrm{d}x < \infty$.

If $\mathbb{E}_{\theta_0}|f(X)| < \infty$, then α is differentiable at θ_0 and

$$\alpha'(\theta_0) = \mathbb{E}_{\theta_0}[f(X)\,\ell_{\theta_0}(X)], \qquad \ell_\theta(x) = \frac{\dot g_\theta(x)}{g_\theta(x)} \ \text{ on } \{g_\theta > 0\}, \quad \ell_\theta = 0 \text{ elsewhere.}$$

The score *ℓ_{θ_0} does not involve f, so one simulation under g_{θ_0} gives every Greek of every payoff; the estimator $f(X)\ell_{\theta_0}(X)$ has finite variance if and only if $\mathbb{E}_{\theta_0}[f(X)^2\ell_{\theta_0}(X)^2] < \infty$.*

Proof. For $0 < |h| < \delta$ and every x, the mean value theorem gives $|g_{\theta_0+h}(x) - g_{\theta_0}(x)|/|h| \le H(x)$, so the integrands $f(x)(g_{\theta_0+h}(x) - g_{\theta_0}(x))/h$ are dominated by the integrable $|f|H$ and converge to $f(x)\dot{g}_{\theta_0}(x)$. By dominated convergence, $\alpha'(\theta_0) = \int f\dot{g}_{\theta_0}\,\mathrm{d}x$. Where $g_{\theta_0}(x) = 0$, the nonnegative function $\theta \mapsto g_\theta(x)$ has a minimum at θ_0, so $\dot{g}_{\theta_0}(x) = 0$; elsewhere $\dot{g}_{\theta_0} = \ell_{\theta_0} g_{\theta_0}$. Hence $\int f\dot{g}_{\theta_0}\,\mathrm{d}x = \int f\ell_{\theta_0} g_{\theta_0}\,\mathrm{d}x$. □

Condition (LR2) is a domination condition. It makes the difference quotients

$$f(X)\Big(\frac{g_{\theta_0+h}(X)}{g_{\theta_0}(X)} - 1\Big)\Big/h, \qquad 0 < |h| < \delta,$$

uniformly integrable under g_{θ_0}. When the support $\{g_\theta > 0\}$ does not depend on θ, these quotients have mean $(\alpha(\theta_0+h) - \alpha(\theta_0))/h$, and uniform integrability together with almost sure convergence is all that the interchange needs (Durrett, 2019, Section 4.6). A moving support is not in itself an obstacle: a location family $\psi(x-\theta)$ with ψ a smooth, compactly supported density satisfies (LR1)–(LR2). The uniform law on $[0,\theta]$, on the other hand, violates (LR1): its density is not differentiable in θ at $x = \theta$.

The likelihood ratio method is importance sampling read the other way round (Chapter 9). If $g_{\theta_0} > 0$ wherever $g_\theta > 0$, then $\alpha(\theta) = \mathbb{E}_{\theta_0}[f(X)g_\theta(X)/g_{\theta_0}(X)]$ by Theorem 9.1, with sampling law g_{θ_0} and target g_θ; the score is the derivative of that weight at $\theta = \theta_0$.

Black–Scholes. The log-price $\log S_T$ has the density of $\mathcal{N}(\log S + \nu T, \sigma^2 T)$. Differentiating its logarithm in each input and writing the result in terms of $Z = (\log S_T - \log S - \nu T)/(\sigma\sqrt{T})$, the estimators are, with $Y = e^{-rT} f(S_T)$ for any payoff f,

$$\begin{aligned}
\Delta^{\mathrm{lr}} &= Y\frac{Z}{S\sigma\sqrt{T}}, & \Gamma^{\mathrm{lr}} &= Y\frac{Z^2 - 1 - \sigma\sqrt{T}Z}{S^2\sigma^2 T}, & \mathcal{V}^{\mathrm{lr}} &= Y\Big(\frac{Z^2-1}{\sigma} - \sqrt{T}Z\Big),\\
\rho_y^{\mathrm{lr}} &= -Y\frac{\sqrt{T}Z}{\sigma} = -TS\Delta^{\mathrm{lr}}, & \rho_r^{\mathrm{lr}} &= -TY + Y\frac{\sqrt{T}Z}{\sigma}, & \Theta^{\mathrm{lr}} &= rY - Y\Big(\frac{\nu Z}{\sigma\sqrt{T}} + \frac{Z^2-1}{2T}\Big).
\end{aligned} \tag{13.5}$$

The gamma uses $\partial_S^2 g/g$, computed in Section 13.4; ρ_r^{lr} adds the discount factor's $-TY$ to the score of the drift. For the digital, $\Delta^{\mathrm{lr}} = e^{-rT}\mathbf{1}_{\{S_T>K\}}Z/(S\sigma\sqrt{T})$ is unbiased with per-sample standard deviation 0.0290 (Theorem 13.2). The domination (LR2) holds for every payoff of polynomial growth in s and $1/s$: for the mean, for instance, $\sup_{|m'-m|<\delta}|\partial_m q(x-m')| \le C(1+|x-m|)e^{-((|x-m|-\delta)^+)^2/(2\sigma^2T)}$ for the Gaussian density q.

Table 13.1 compares the three methods on the test case. Every estimate is within two standard errors of the closed form. The bumps with common random numbers reproduce the pathwise estimates and their errors, as Theorem 13.4 predicts. The likelihood-ratio estimators have standard deviations 2.2 to 3.5 times larger, that is 5 to 12 times more paths for the same accuracy; the course's observation that the likelihood-ratio ρ_y has errors "more than twice" the pathwise ones is the ratio $128.71/58.22 = 2.21$; the exact ratio grows with the spot, from 1.86 at $S_0 = 90$ to 2.81 at $S_0 = 110$, and exceeds 2 for $S_0 \gtrsim 95$. In Figure 13.2(c) the pathwise delta has the $1/n$ steps of Theorem 13.3, while the likelihood-ratio delta is smooth in S_0, since each draw contributes the smooth $e^{-rT}(SG-K)^+Z/(S\sigma\sqrt{T})$, but further from the truth.

Remark 13.7 (The weight degenerates as $\sigma\sqrt{T} \to 0$). The delta weight $Z/(S\sigma\sqrt{T})$ has standard deviation $1/(S\sigma\sqrt{T})$. Let f be continuous at the forward $F = Se^{(r-y)T}$, with $|f(s)| \le C(1+s^p)$.

	Black–Scholes	pathwise	likelihood ratio	bump, CRN	sd per sample	
					pathwise	LR
Δ	0.509643	0.5108(18)	0.5093(41)	0.5108(18)	0.582	1.287
Γ	0.0195277	0.01954(11)*	0.01941(39)	0.01977(36)	0.0363*	0.1234
$\mathcal{V}$	39.0554	39.08(23)	38.82(79)	39.08(23)	72.67	246.70
ρ_r	43.6000	43.71(16)	43.55(37)	43.71(16)	49.15	117.08
ρ_y	−50.9643	−51.08(18)	−50.93(41)	−51.08(18)	58.22	128.71
$\partial_t V$	−3.32226	−3.324(21)	−3.299(75)	−3.324(21)	6.73	23.45

Table 13.1: The six Greeks of the test-case call from the same $n = 10^5$ draws (our implementation, `default_rng`, seed 13, `greeks.greeks_table`); standard errors in units of the last digit. Bumps with common random numbers: 1% of S (second difference for Γ), of σ and of T, and one basis point for r and y. The per-sample standard deviations are exact (quadrature). *For the gamma, the PW-LR estimator of Theorem 13.8; its LR-PW estimator gives 0.019557(92), with standard deviation 0.0290.

Then, as $\sigma \downarrow 0$ with S, T, r, y fixed,

$$\sigma^2 T \operatorname{Var}(\Delta^{\mathrm{lr}}) \longrightarrow e^{-2rT} f(F)^2 / S^2.$$

Indeed $\sigma^2 T \operatorname{Var}(\Delta^{\mathrm{lr}}) = e^{-2rT}\mathbb{E}[f(S_T)^2 Z^2]/S^2 - \left(\sigma\sqrt{T}\,\Delta\right)^2$; as $\sigma \downarrow 0$, $S_T \to F$ for every Z, and for $\sigma \le 1$ the integrand is dominated by $C'(1 + e^{2p\sqrt{T}|Z|})Z^2$, with C' depending on S, T, r, y, p, so the first term converges to $e^{-2rT} f(F)^2/S^2$ by dominated convergence; the same argument gives $\sigma\sqrt{T}\Delta = e^{-rT}\mathbb{E}[f(S_T)Z]/S \to e^{-rT} f(F)\mathbb{E}[Z]/S = 0$. So whenever $f(F) \neq 0$ the standard deviation grows like $1/(\sigma\sqrt{T})$, while the pathwise delta of a payoff with Lipschitz constant L_f has variance at most $e^{-2rT} L_f^2 \mathbb{E}[G^2]$, bounded. For the share ($K = 0$) the two exact variances are $e^{-2yT}(e^{\sigma^2 T}(1 + 4\sigma^2 T)/(\sigma^2 T) - 1) = 28.04$ and $e^{-2yT}(e^{\sigma^2 T} - 1) = 0.0392$ at the test case, a ratio of 715. The same happens as $T \downarrow 0$, and the gamma weight is worse, of order $1/(\sigma^2 T)$. At the forward, $f(F) = 0$ for the call and the limit is finite (Exercise 13.5).

Along a path. For a payoff of the whole path, the law of the path is a product of Gaussian steps: with $\Delta t = T/m$ and $\log(S_{t_j}/S_{t_{j-1}}) = \nu\Delta t + \sigma\sqrt{\Delta t} Z_j$, the input S enters only the first step and σ enters every step, so the scores are

$$\Delta^{\mathrm{lr}} = Y \frac{Z_1}{S\sigma\sqrt{t_1}}, \qquad \mathcal{V}^{\mathrm{lr}} = Y \sum_{j=1}^{m} \Big(\frac{Z_j^2 - 1}{\sigma} - \sqrt{\Delta t}\, Z_j \Big), \tag{13.6}$$

valid for any payoff of the fixings, digital or barrier included. Both degrade as the grid is refined (Figure 13.3(a)). The vega weight is a sum of m independent terms of variance $2/\sigma^2 + \Delta t$, so its variance grows linearly in m; in continuous time the laws of the paths for two volatilities are mutually singular, because the quadratic variation of $\log S$ reveals σ, and there is no likelihood-ratio vega at all. The delta weight has standard deviation $1/(S\sigma\sqrt{t_1})$, which grows as the first fixing approaches 0. This weight is Girsanov's (Theorem 9.8): moving $\log S$ by ϵ moves every fixing, and the cheapest drift that achieves this by time t_1 is the constant $\epsilon/(\sigma t_1)$ on $[0, t_1]$ (by Cauchy–Schwarz it has the smallest $\int \theta^2$ among drifts with $\int_0^{t_1} \theta = \epsilon/\sigma$, hence the smallest second moment of the weight), and its weight differentiates to $W_{t_1}/(S\sigma t_1) = Z_1/(S\sigma\sqrt{t_1})$. For $m = 1$ the Asian is the call, and the likelihood-ratio standard deviations are 2.2 (delta) and 3.4 (vega)

times the pathwise ones (Table 13.1); for the arithmetic Asian with $m = 128$ they are 8.9 and 17 times (10^5 paths, seed 2026).

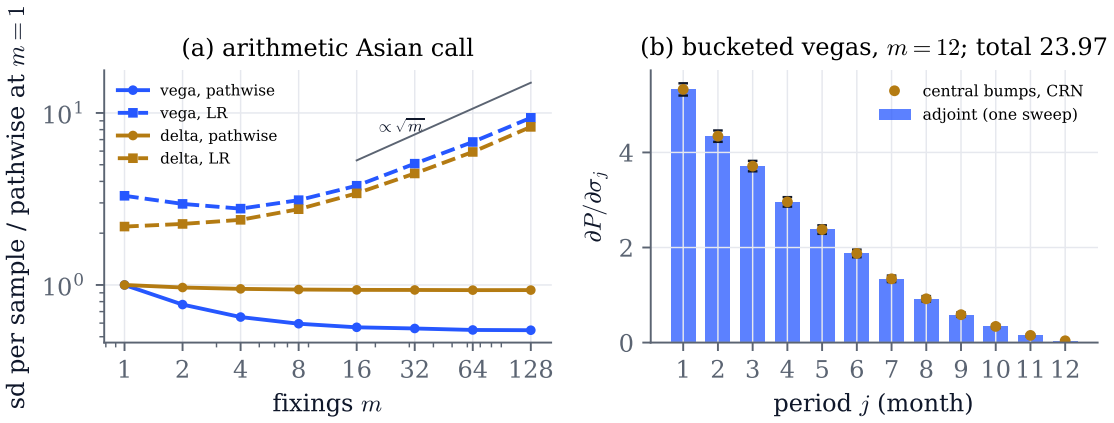

Figure 13.3: Asian call on the test case. (a) Per-sample standard deviations of the pathwise (13.3) and likelihood-ratio (13.6) deltas and vegas against the number of fixings, divided by the pathwise value at $m = 1$ (10^5 paths per m, seed 2026). (b) Bucketed vegas $\partial P/\partial\sigma_j$ for $m = 12$ from one adjoint sweep (bars, with 95% intervals) and from 24 central bumps of 10^{-4} with common random numbers (dots), same 10^5 paths (seed 2027); the total is the parallel vega.

13.4 The gamma

A second derivative is harder. For the call, the pathwise delta $e^{-rT}\mathbf{1}_{\{SG>K\}}G$ jumps when the critical spot K/G crosses S, so its own pathwise derivative is 0 almost everywhere: the "pathwise gamma" is identically zero (Theorem 13.5), and the second difference with common draws is a kernel estimator with variance of order $1/(nh)$ (Theorem 13.3). The likelihood-ratio gamma of (13.5) is unbiased but noisy. The course's way out is to differentiate twice by different methods: once pathwise, where the payoff is smooth enough, and once by likelihood ratio, where it is not.

Proposition 13.8 (Mixed estimators of the gamma)**.** *Let $S, \sigma, T > 0$, $S_T = SG$ as above, and let $f : (0,\infty) \to \mathbb{R}$ be continuous, and continuously differentiable except at finitely many points, with $|f(s)| + |sf'(s)| \le C(s^p + s^{-p})$ wherever $f'(s)$ exists, for some constants C, p. Then $V(S) = e^{-rT}\mathbb{E}f(S_T)$ is twice differentiable, and its gamma $\Gamma = V''(S)$ is the expectation of each of*

$$\text{(LR-LR)} \quad e^{-rT} f(S_T)\,\frac{Z^2 - 1 - \sigma\sqrt{T}Z}{S^2\sigma^2 T},$$

$$\text{(PW-LR)} \quad e^{-rT} f'(S_T)S_T\,\frac{Z - \sigma\sqrt{T}}{S^2\sigma\sqrt{T}},$$

$$\text{(LR-PW)} \quad e^{-rT}\big(f'(S_T)S_T - f(S_T)\big)\frac{Z}{S^2\sigma\sqrt{T}},$$

all with finite variance. PW-LR applies the likelihood ratio to the pathwise delta, LR-PW differentiates the likelihood-ratio delta pathwise. The LR-LR estimator needs only f Borel with $|f(s)| \le C(s^p + s^{-p})$, digitals included; the other two are biased for the digital.

Proof. Let $u = \log S$, $X = \nu T + \sigma\sqrt{T}Z$, so that $S_T = e^{u+X}$, and q the density of X. For a Borel k with $|k(s)| \le C(s^p + s^{-p})$ let $\phi_k(u) = \mathbb{E}[k(e^{u+X})] = \int k(e^x)q(x-u)\,\mathrm{d}x$. The family $u \mapsto q(x-u)$ satisfies (LR1) and (LR2), and so does $u \mapsto \partial_u q(x-u)$ in place of $q(x-u)$, since $\partial_u^j q(x-u)$ is a polynomial in $x-u$ times $q(x-u)$ and $k(e^x)$ grows at most like $e^{p|x|}$. By Theorem 13.6, and the same dominated convergence argument for the second derivative,

$$\phi_k'(u) = \frac{\mathbb{E}[k(S_T)Z]}{\sigma\sqrt{T}}, \qquad \phi_k''(u) = \frac{\mathbb{E}[k(S_T)(Z^2-1)]}{\sigma^2 T},$$

because $\partial_u \log q(x-u) = (x-u-\nu T)/(\sigma^2 T) = Z/(\sigma\sqrt{T})$ and $\partial_u^2 q/q = (Z^2-1)/(\sigma^2 T)$. Next, f is continuous and piecewise C^1, so $f(e^{u'+X}) - f(e^{u+X}) = \int_u^{u'} f'(e^{v+X})e^{v+X}\,\mathrm{d}v$, and $|f(e^{u'+X}) - f(e^{u+X})| \le |u'-u|\,\kappa$ for $|u'-u| < 1$, with $\kappa = \sup_{|v-u|<1} C(e^{p(v+X)} + e^{-p(v+X)})$ integrable, also after multiplication by $|Z|$; and f is differentiable at S_T almost surely, since S_T has a density and f fails to be differentiable at finitely many points only. By Theorem 13.4, $\phi_f'(u) = \mathbb{E}[f'(S_T)S_T]$ and $\frac{\mathrm{d}}{\mathrm{d}u}\mathbb{E}[f(S_T)Z] = \mathbb{E}[f'(S_T)S_T Z]$.

Since $\partial_S = S^{-1}\partial_u$, $\Gamma = e^{-rT}S^{-2}(\phi_f'' - \phi_f')(u)$. Writing $\phi_f'' - \phi_f'$ with both derivatives by likelihood ratio gives LR-LR. Writing $\phi_f' = \phi_k$ with $k(s) = sf'(s)$ (pathwise) and then $\phi_f'' = \phi_k' = \mathbb{E}[k(S_T)Z]/(\sigma\sqrt{T})$ (likelihood ratio, which needs only the growth of k) gives PW-LR. Writing $\phi_f' = \mathbb{E}[f(S_T)Z]/(\sigma\sqrt{T})$ (likelihood ratio) and differentiating it pathwise gives $\phi_f'' = \mathbb{E}[f'(S_T)S_T Z]/(\sigma\sqrt{T})$, hence LR-PW. Each estimator is a polynomial in Z times a function of S_T of polynomial growth in S_T and $1/S_T$, so it has finite variance. For the digital, f is not continuous and $f' = 0$ off K: PW-LR is then identically 0 and LR-PW has mean $-\Delta/S$. □

For the test-case call the per-sample standard deviations are 0.1234 (LR-LR), 0.0363 (PW-LR) and 0.0290 (LR-PW): the mixed estimators need 12 and 18 times fewer paths than the pure likelihood ratio. The second difference with common draws and a 1% bump has standard deviation 0.112 and a bias of 0.02%: comparable to LR-LR, at three evaluations per draw. For the call, $f'(s)s - f(s) = K\mathbf{1}_{\{s>K\}}$, and LR-PW is K/S times the likelihood-ratio delta of the digital: the gamma of a call is the delta of a digital in disguise. Figure 13.2(d) is the course's picture: with 1024 draws the mixed gamma is already smooth in S_0, and LR-LR converges as well, more slowly. Both mixed estimators inherit the likelihood-ratio weight, so they degenerate like $1/(\sigma\sqrt{T})$, not like the $1/(\sigma^2 T)$ of LR-LR.

▶ Interactive demo: pathwise, likelihood-ratio and mixed Greeks against the spot, for any number of draws.

13.5 Many Greeks at once: adjoints

A book may depend on hundreds of inputs: a volatility per maturity, a forward rate per period. Bumping costs one or two repricings per input, and so does the pathwise method in forward mode (13.4), one derivative at a time. The *adjoint*, or reverse, mode computes the derivative of one output with respect to all the inputs in one backward sweep.

Remark 13.9 (Adjoint algorithmic differentiation). Write the computation of a path's discounted payoff P from the inputs $x_1, \dots, x_p$ as a sequence of elementary operations $v_k = \varphi_k(v_j : j \prec k)$, each using earlier values. Forward mode propagates tangents $\dot v_k = \sum_{j\prec k} \partial_j \varphi_k\, \dot v_j$ from a seed

$\dot{x}$, one input direction per sweep. Reverse mode propagates adjoints $\bar{v}_j = \partial P/\partial v_j$ backwards, $\bar{v}_j = \sum_{k \succ j} \bar{v}_k \, \partial_j \varphi_k$, from $\bar{P} = 1$; at the end, $\bar{x}_i = \partial P/\partial x_i$ for every i. Each elementary operation is visited once in each sweep, and its local partial derivatives cost a bounded multiple of the operation itself. In the operation-count model of Griewank and Walther (2008, Section 4.6), the price and its whole gradient together cost at most four times the price alone, whatever p. Giles and Glasserman (2006) quote this bound and use the adjoint to compute the deltas and vegas of the LIBOR market model, with an advantage over forward mode that grows in proportion to the number of forward rates; Capriotti (2011) develops the method for pricing code in general. Three caveats. (a) The backward sweep needs the intermediate values of the forward sweep, stored (the "tape") or recomputed; in Monte Carlo this is one path at a time, so the memory is proportional to the number of time steps. (b) Applied path by path, the adjoint computes *exactly* the pathwise estimator, to rounding: it inherits the hypotheses of Theorem 13.4, so digitals and barriers must be smoothed first, and a gamma needs a second-order method such as those of Theorem 13.8. (c) The bound is on operations; actual running times also depend on memory traffic and on the implementation.

For the Asian call with a volatility σ_j on each period $(t_{j-1}, t_j]$, the forward sweep is $X_0 = \log S$, $X_j = X_{j-1} + (r - y - \sigma_j^2/2)\Delta t + \sigma_j \sqrt{\Delta t} Z_j$, $S_{t_j} = e^{X_j}$, $A = \frac{1}{m}\sum_j S_{t_j}$, $P = e^{-rT}(A - K)^+$. Differentiating each line backwards gives $\bar{A} = e^{-rT} \mathbf{1}_{\{A > K\}}$ and, for $j = m, \dots, 1$,

$$\bar{X}_j = \bar{A} \, \frac{S_{t_j}}{m} + \bar{X}_{j+1} \quad (\bar{X}_{m+1} = 0), \qquad \frac{\partial P}{\partial \sigma_j} = \bar{X}_j (\sqrt{\Delta t} Z_j - \sigma_j \Delta t),$$

then $\Delta = \bar{X}_1 / S$, $\rho_y = -\Delta t \sum_j \bar{X}_j$ and $\rho_r = -TP + \Delta t \sum_j \bar{X}_j$: all $m + 3$ sensitivities for one forward and one backward pass, against $2(m + 3)$ repricings for central bumps. With a flat volatility the bucketed vegas add up to the pathwise vega (13.3). In our implementation the adjoint agrees with the $m + 3$ forward-mode sweeps to 10^{-12} on every path and with central bumps of 10^{-4} to 10^{-3}, and Figure 13.3(b) shows the twelve bucketed vegas: the early periods matter most, because a volatility early in the year moves every later fixing.

In code

```
x = np.full(n, np.log(S)) # forward sweep, storing the tape X
for j in range(m):
    x = x + (r - y - sig[j]**2 / 2) * dt + sig[j] * sq * Z[:, j]
    X[:, j] = x
Sj = np.exp(X); A = Sj.mean(axis=1); P = disc * np.maximum(A - K, 0.0)
Abar, acc = disc * (A > K), 0.0 # reverse sweep
for j in reversed(range(m)):
    acc = acc + Abar * Sj[:, j] / m
    Xbar[:, j] = acc
vega = Xbar * (sq * Z - sig * dt); delta = Xbar[:, 0] / S
```

`asian_adjoint` in code/src/mfe230d/greeks.py; `asian_tangent` is the forward mode, one input direction per sweep. No automatic-differentiation library is used: the backward sweep is the chain rule written by hand.

▶ Interactive demo: Asian Greeks against the number of fixings, and bucketed vegas by adjoint and by bumping, with timings.

Further reading

Figure 13.2 and Table 13.1 reproduce the course's bumping study and its analytical-differencing, likelihood-ratio and mixed estimators. Glasserman (2003), Chapter 7, is the reference for everything here: finite differences in Section 7.1, the pathwise method in Section 7.2, likelihood ratios and the gamma in Section 7.3. Broadie and Glasserman (1996) introduced the pathwise and likelihood-ratio estimators to option pricing, and Asmussen and Glynn (2007, Chapter VII) treat derivative estimation in simulation in general. The Malliavin-calculus weights of Fournié et al. (1999) generalise the likelihood ratio to models without explicit densities. Adjoints are due to Giles and Glasserman (2006) in this context; Capriotti (2011) and Griewank and Walther (2008) cover algorithmic differentiation in depth.

Chapter 14

American options by simulation

An American option can be exercised at any time up to its maturity; a Bermudan option, on a finite set of dates. Pricing either means solving an optimal stopping problem under the risk-neutral measure: at each date the holder compares the payoff of exercising now with the value of keeping the option alive. In the binomial model of the IEOR 222 notes this is a backward induction over the nodes of a tree, and the trees and finite differences of Chapter 20 do the same on a grid. Grids are efficient in one or two state variables and hopeless in ten (Theorem 6.14). Monte Carlo handles many state variables, but it runs forward in time, while the exercise decision at a date requires the conditional expectation of the future, which one simulated path does not reveal.

Three ideas make simulation work, and each is a section of this chapter. Section 14.1 reduces the problem on a finite set of dates to a dynamic programming recursion whose only unknowns are the continuation values, conditional expectations of a function of the next state. Section 14.2 estimates them by regression across simulated paths, as Longstaff and Schwartz (2001) proposed: this produces an exercise policy, and any policy gives a *lower* bound for the price. Section 14.3 turns the problem around: every martingale gives an *upper* bound, the optimal one gives the price exactly, and a good policy gives a good martingale (Rogers, 2002; Haugh and Kogan, 2004; Andersen and Broadie, 2004). The two bounds bracket the price and measure how good the policy is. The words "low" and "high" need care, and we will be precise about which estimator is biased in which direction.

Throughout, S is the Black–Scholes stock with yield y under the risk-neutral measure $\mathbb{Q}$ (Theorem 1.1), with a filtration $(\mathcal{F}_t)$ for its Brownian motion; all expectations are under $\mathbb{Q}$, and we write $\mathbb{E}$ for $\mathbb{E}^{\mathbb{Q}}$. The running examples are the put of Longstaff and Schwartz (2001), with $K = 40$, $r = 6\%$, $y = 0$, $\sigma \in \{20\%, 40\%\}$, $T \in \{1, 2\}$, $S_0 \in \{36, 38, \dots, 44\}$ and 50 exercise dates per year, and the course's test case $S_0 = K = 100$, $T = 1$, $r = 1\%$, $y = 2\%$, $\sigma = 20\%$.

14.1 Bermudan options and optimal stopping

Fix exercise dates $0 = t_0 < t_1 < \cdots < t_m = T$ and write $\mathcal{F}_k := \mathcal{F}_{t_k}$. A Bermudan claim with payoff function g pays $g(S_{t_k})$ if it is exercised at t_k; let

$$Z_k = e^{-rt_k} g(S_{t_k}), \qquad k = 0, \dots, m,$$

be the payoff discounted to time 0. More generally, Z may be any adapted sequence. A *stopping time* is a random variable τ with values in $\{0,\dots,m\}$ such that $\{\tau = j\} \in \mathcal{F}_j$ for every j: the decision to stop at t_j uses only the information available at t_j. Write $\mathcal{T}_k$ for the set of stopping times with $k \le \tau \le m$. The price of the Bermudan claim is

$$U_0 = \sup_{\tau \in \mathcal{T}_0} \mathbb{E}[Z_\tau], \tag{14.1}$$

the value of the holder's optimal stopping problem. That this is the arbitrage-free price, the smallest initial capital from which a self-financing portfolio dominates the payoff at every date, is shown for finite probability spaces in the IEOR 222 notes and in general discrete time by Föllmer and Schied (2016), Chapter 6; Shreve (2004), Section 8.4, uses the corresponding definition for the American put. Here we take (14.1) as the definition.

On a finite set of dates, integrability of $\max_k |Z_k|$ is the same as integrability of each Z_k, because $\max_k |Z_k| \le \sum_k |Z_k|$; for calls and puts it holds since $\mathbb{E}[S_t] < \infty$. The next lemma is optional stopping on a finite horizon, conditionally on $\mathcal{F}_k$.

Lemma 14.1 (Optional stopping)**.** *Let $(Y_j)_{k \le j \le m}$ be a supermartingale with respect to $(\mathcal{F}_j)$ and $\tau \in \mathcal{T}_k$. Then Y_τ is integrable and $\mathbb{E}[Y_\tau \mid \mathcal{F}_k] \le Y_k$, with equality if Y is a martingale.*

Proof. $|Y_\tau| \le \sum_j |Y_j|$ is integrable, and $Y_\tau - Y_k = \sum_{j=k+1}^{m} \mathbf{1}_{\{\tau \ge j\}}(Y_j - Y_{j-1})$. The event $\{\tau \ge j\} = \Omega \setminus \{\tau \le j-1\}$ is in $\mathcal{F}_{j-1}$, so by the tower property and taking out what is known, $\mathbb{E}[\mathbf{1}_{\{\tau \ge j\}}(Y_j - Y_{j-1}) \mid \mathcal{F}_k] = \mathbb{E}\big[\mathbf{1}_{\{\tau \ge j\}} \mathbb{E}[Y_j - Y_{j-1} \mid \mathcal{F}_{j-1}] \,\big|\, \mathcal{F}_k\big] \le 0$, with equality for a martingale. □

Theorem 14.2 (Bermudan dynamic programming)**.** *Let Z be adapted with $\mathbb{E}[\max_k |Z_k|] < \infty$. Define $U_m = Z_m$ and, for $k = m-1, \dots, 0$,*

$$C_k = \mathbb{E}[U_{k+1} \mid \mathcal{F}_k], \qquad U_k = \max(Z_k, C_k). \tag{14.2}$$

*(i) (*Snell envelope.*) Each U_k is integrable, $U \ge Z$, and U is a supermartingale. Every supermartingale V with $V \ge Z$ satisfies $V \ge U$.*

*(ii) (*Optimal stopping.*) For every k and every $\tau \in \mathcal{T}_k$, $\mathbb{E}[Z_\tau \mid \mathcal{F}_k] \le U_k$, with equality for $\tau_k^* = \min\{j \ge k : \ j = m \text{ or } Z_j \ge C_j\}$. In particular $U_0 = \max_{\tau \in \mathcal{T}_0} \mathbb{E}[Z_\tau] = \mathbb{E}[Z_{\tau_0^*}]$ is the price (14.1).*

*(iii) (*A sufficient condition.*) If $\tau \in \mathcal{T}_k$, $Z_\tau = U_\tau$, and the stopped process $(U_{j \wedge \tau})_{k \le j \le m}$ is a martingale, then $\mathbb{E}[Z_\tau \mid \mathcal{F}_k] = U_k$: τ is optimal from t_k on.*

*(iv) (*Stopping-time form.*) $\tau_m^* = m$, $\tau_k^* = k$ on $\{Z_k \ge C_k\}$ and $\tau_k^* = \tau_{k+1}^*$ elsewhere; and $C_k = \mathbb{E}[Z_{\tau_{k+1}^*} \mid \mathcal{F}_k]$.*

*(v) (*Doob decomposition.*) Let $M_0^* = A_0^* = 0$ and*

$$M_{k+1}^* - M_k^* = U_{k+1} - C_k, \qquad A_{k+1}^* - A_k^* = U_k - C_k. \tag{14.3}$$

Then M^* is a martingale, A^* is nondecreasing and predictable (A_{k+1}^* is $\mathcal{F}_k$-measurable), and $U_k = U_0 + M_k^* - A_k^*$.

*(vi) (*Markov form.*) If $Z_k = e^{-rt_k} g(S_{t_k})$ with g measurable and $|g(s)| \le a(1+s)$, then $U_k = e^{-rt_k} V_k(S_{t_k})$ and $C_k = e^{-rt_k} c_k(S_{t_k})$, where $V_m = g$ and, with $\Delta_k = t_{k+1} - t_k$ and $\xi \sim \mathcal{N}(0,1)$,*

$$c_k(s) = e^{-r\Delta_k}\, \mathbb{E}\big[V_{k+1}\big(s e^{\nu\Delta_k + \sigma\sqrt{\Delta_k}\,\xi}\big)\big], \qquad V_k = \max(g, c_k). \tag{14.4}$$

Proof. (i) Let $Z^\star = \max_j |Z_j|$. By backward induction $|U_k| \le \mathbb{E}[Z^\star \mid \mathcal{F}_k]$: this holds for $k = m$, and if it holds for $k+1$ then $|C_k| \le \mathbb{E}[|U_{k+1}| \mid \mathcal{F}_k] \le \mathbb{E}[Z^\star \mid \mathcal{F}_k]$ and $|Z_k| \le \mathbb{E}[Z^\star \mid \mathcal{F}_k]$ because Z_k is $\mathcal{F}_k$-measurable. So U is integrable and adapted, $U \ge Z$, and $U_k \ge C_k = \mathbb{E}[U_{k+1} \mid \mathcal{F}_k]$ is the supermartingale property. If $V \ge Z$ is a supermartingale, then $V_m \ge Z_m = U_m$, and $V_{k+1} \ge U_{k+1}$ implies $V_k \ge \mathbb{E}[V_{k+1} \mid \mathcal{F}_k] \ge C_k$ and $V_k \ge Z_k$, so $V_k \ge U_k$.

(iii) The stopped process is a martingale, so $U_k = \mathbb{E}[U_{m\wedge\tau} \mid \mathcal{F}_k] = \mathbb{E}[U_\tau \mid \mathcal{F}_k] = \mathbb{E}[Z_\tau \mid \mathcal{F}_k]$.

(ii) By Theorem 14.1 and (i), $U_k \ge \mathbb{E}[U_\tau \mid \mathcal{F}_k] \ge \mathbb{E}[Z_\tau \mid \mathcal{F}_k]$. For τ_k^*: if $\tau_k^* = j < m$ then $Z_j \ge C_j$, so $U_j = Z_j$; and $U_m = Z_m$. The increments of the stopped process are $U_{(j+1)\wedge\tau_k^*} - U_{j\wedge\tau_k^*} = \mathbf{1}_{\{\tau_k^* > j\}}(U_{j+1} - U_j)$, and on $\{\tau_k^* > j\} \in \mathcal{F}_j$ we have $Z_j < C_j$, hence $U_j = C_j$ and $\mathbb{E}[\mathbf{1}_{\{\tau_k^*>j\}}(U_{j+1} - U_j) \mid \mathcal{F}_j] = \mathbf{1}_{\{\tau_k^*>j\}}(C_j - U_j) = 0$. So (iii) applies. As $\mathcal{F}_0$ is trivial, the maximum over $\mathcal{T}_0$ is attained at τ_0^*.

(iv) The recursion is the definition of τ_k^* read from date k. By (ii) at $k+1$ and the tower property, $C_k = \mathbb{E}\big[\mathbb{E}[Z_{\tau_{k+1}^*} \mid \mathcal{F}_{k+1}] \mid \mathcal{F}_k\big] = \mathbb{E}[Z_{\tau_{k+1}^*} \mid \mathcal{F}_k]$.

(v) $\mathbb{E}[U_{k+1} - C_k \mid \mathcal{F}_k] = 0$, $U_k - C_k \ge 0$ is $\mathcal{F}_k$-measurable, and the sum of the two increments in (14.3) is $U_{k+1} - U_k$.

(vi) By backward induction. Suppose $U_{k+1} = e^{-rt_{k+1}} V_{k+1}(S_{t_{k+1}})$ with $|V_{k+1}(s)| \le a'(1+s)$. Write $S_{t_{k+1}} = S_{t_k} e^{\nu\Delta_k + \sigma(W_{t_{k+1}} - W_{t_k})}$; the increment of W is independent of $\mathcal{F}_k$ and S_{t_k} is $\mathcal{F}_k$-measurable, so the independence lemma (Shreve, 2004, Lemma 2.3.4) gives $C_k = e^{-rt_k} c_k(S_{t_k})$ with c_k as in (14.4), which is finite with $|c_k(s)| \le a'(1 + se^{(r-y)\Delta_k})e^{-r\Delta_k}$. Then $U_k = e^{-rt_k} \max(g, c_k)(S_{t_k})$, and V_k again grows at most linearly. □

The process U is the *Snell envelope* of Z; in the Markov form, V_k is the value of the option at t_k and c_k its *continuation value*, the value of holding it for one more period. Part (iv) is the form of the recursion that simulation will use: it moves stopping times, not values, backwards, and the continuation value is the conditional mean of the discounted payoff received by following the optimal rule from the next date on. On continuous time, with its additional measurability issues, see Peskir and Shiryaev (2006), Chapter I, and for the financial formulation Glasserman (2003), Section 8.1.

Proposition 14.3 (From Bermudan to American)**.** *Let g be continuous with $|g(s)| \le a(1+s)$, $Z_t = e^{-rt} g(S_t)$ for $t \in [0,T]$, and let the American price be $A = \sup_\tau \mathbb{E}[Z_\tau]$ over stopping times τ of $(\mathcal{F}_t)$ with values in $[0,T]$. For a set of dates Π as above, let $B(\Pi)$ be the Bermudan price (14.1). Then $B(\Pi) \le B(\Pi') \le A$ whenever $\Pi \subseteq \Pi'$, and $B(\Pi_n) \to A$ for every sequence (Π_n) whose mesh tends to 0.*

Proof. A stopping time with values in Π is one with values in Π', and one with values in $[0,T]$: the sets over which the suprema are taken increase. For the limit, $\mathbb{E}[\sup_{t\le T} |Z_t|] < \infty$: $|Z_t| \le a(1 + S_t)$ and $S_t = S_0 e^{(r-y)t}\mathcal{E}_t \le S_0 e^{|r-y|T}\mathcal{E}_t$ with the martingale $\mathcal{E}_t = e^{\sigma W_t - \sigma^2 t/2}$, whose supremum over $[0,T]$ is square-integrable by Doob's maximal inequality (Karatzas and Shreve, 1991, Theorem 1.3.8). Fix a stopping time τ and let $\tau_n = \min\{t \in \Pi_n : t \ge \tau\}$. Since $\{\tau_n \le t\} = \{\tau \le t\} \in \mathcal{F}_t$ for $t \in \Pi_n$, τ_n is a stopping time with values in Π_n, so $B(\Pi_n) \ge \mathbb{E}[Z_{\tau_n}]$. As the mesh tends to 0, $\tau_n \to \tau$, so $Z_{\tau_n} \to Z_\tau$ by continuity of the paths, and dominated convergence (with the bound $\sup_t |Z_t|$) gives $\mathbb{E}[Z_{\tau_n}] \to \mathbb{E}[Z_\tau]$. Hence $\liminf_n B(\Pi_n) \ge \mathbb{E}[Z_\tau]$ for every τ, so $\liminf_n B(\Pi_n) \ge A$. □

The proposition gives no rate. On the Longstaff–Schwartz put with $S_0 = 40$, $\sigma = 20\%$, $T = 1$, our grid (Theorem 14.4) gives 2.0664 (European), 2.1991, 2.2930, 2.3141, 2.3185 for $m = 1, 2, 10, 50, 250$ equally spaced dates, and the American price is 2.3196: the gap falls from 0.027 to 0.0055 to 0.0011 as m goes from 10 to 50 to 250, roughly like $1/m$ in this example.

Remark 14.4 (Reference values). Our reference prices use no simulation (`bermudan_reference` and `american_binomial` in the code of this chapter). Bermudan prices come from (14.4) on a grid in $\log s$ of mesh h: the trapezoid rule against the exact Gaussian transition density (an FFT convolution), extrapolated in h; the error is below 10^{-6} here, and a binomial tree that exercises only at the 50 dates agrees. American prices come from the binomial tree with the Black–Scholes value in its last step and Richardson extrapolation (Broadie and Detemple, 1996): with 2000 steps in Table 14.1, quoted to 10^{-3}, and 4000 steps elsewhere, with an error of about 10^{-5}. Grids are the subject of Chapter 20.

Remark 14.5 (Degenerate cases). (a) *Calls without dividends are never exercised early.* If $y = 0$, $r \geq 0$ and $K \geq 0$, Jensen's inequality and the martingale property of $e^{-rt}S_t$ give $\mathbb{E}[Z_{k+1} \mid \mathcal{F}_k] \geq \left(e^{-rt_k}S_{t_k} - e^{-rt_{k+1}}K\right)^+ \geq Z_k$ for $Z_k = e^{-rt_k}(S_{t_k} - K)^+$: Z is a submartingale. So $V_k = \mathbb{E}[Z_m \mid \mathcal{F}_k]$ is a martingale with $V \geq Z$, and Theorem 14.2(i) gives $U_0 \leq V_0 = \mathbb{E}[Z_m] \leq U_0$. The Bermudan call, and by Theorem 14.3 the American call, is worth the European call. With $y > 0$ this fails: on the test case, where $y > r$, the Bermudan call with 50 dates is worth 7.4567 and the American call 7.4597, against 7.36429 for the European call.

(b) *The test-case put.* It is worth 8.349406 as a European option, 8.349408 as a Bermudan option with 50 dates and as an American option: early exercise is worth 2×10^{-6}. The reason is where the exercise region lies. Since $c_k(s) \geq e^{-r\Delta}\mathbb{E}[K - S_{t_{k+1}} \mid S_{t_k} = s] = Ke^{-r\Delta} - se^{-y\Delta}$ with $\Delta = t_{k+1} - t_k$, exercise at t_k is suboptimal whenever $s > \bar{s} = K(1 - e^{-r\Delta})/(1 - e^{-y\Delta})$, which is close to $Kr/y = 50$ here. The boundary rises from 44.9 at t_1 to $\bar{s} = 50.005$ just before maturity, prices that S rarely reaches within a year. A put deep enough in the money is still exercised.

(c) *Exercise at t_0.* $U_0 = \max(Z_0, C_0)$, and immediate exercise is optimal when $Z_0 \geq C_0$, as for a put deep in the money. All simulated paths share S_{t_0}, so no regression is possible at t_0: C_0 is estimated by a sample mean (Exercise 14.2).

(d) *Out of the money.* If $g \geq 0$ then $C_k \geq 0$, so stopping where $Z_k = 0$ is never better than continuing, and a rule may as well continue there.

14.2 Regression: the Longstaff–Schwartz algorithm

The recursion of Theorem 14.2(iv) needs the continuation value C_k at the simulated state. Estimating it by a fresh simulation from every node of every path, as in the random tree of Broadie and Glasserman (1997), costs a number of paths exponential in m. The idea of Carrière (1996), Tsitsiklis and Van Roy (2001) and Longstaff and Schwartz (2001) is to estimate the function c_k in one go, by regressing across paths: since $C_k = \mathbb{E}[Z_{\tau^*_{k+1}} \mid S_{t_k}]$, the realised discounted cash flows of the paths are noisy observations of it, and least squares onto functions of S_{t_k} recovers their conditional mean. Let $\psi = (\psi_1, \ldots, \psi_J)$ be a vector of basis functions.

The algorithm. Simulate N training paths $S^{(i)}$ at the dates t_k exactly (Theorem 7.2), with discounted payoffs $Z_k^{(i)}$, and set $Y^{(i)} = Z_m^{(i)}$. For $k = m-1, \dots, 1$:

(1) let $I_k = \{i : Z_k^{(i)} > 0\}$ be the paths in the money at t_k;

(2) compute $\hat\beta_k \in \arg\min_\beta \sum_{i \in I_k} \left(Y^{(i)} - \beta^\top \psi(S_{t_k}^{(i)})\right)^2$ and set $\hat C_k(s) = \hat\beta_k^\top \psi(s)$;

(3) for $i \in I_k$ with $Z_k^{(i)} \geq \hat C_k(S_{t_k}^{(i)})$, replace $Y^{(i)}$ by $Z_k^{(i)}$.

Finally $\hat c_0 = \frac{1}{N}\sum_i Y^{(i)}$, and the *in-sample estimate* is $\hat V_N^{\text{in}} = \max(Z_0, \hat c_0)$. The coefficients define an exercise rule: for a path S,

$$\hat\tau = \begin{cases} 0 & \text{if } Z_0 > 0 \text{ and } Z_0 \geq \hat c_0, \\ \min\left\{k \geq 1 : \ k = m \text{ or } \left(Z_k > 0 \text{ and } Z_k \geq \hat C_k(S_{t_k})\right)\right\} & \text{otherwise.} \end{cases} \tag{14.5}$$

The *out-of-sample estimate* freezes $\hat\beta_1, \dots, \hat\beta_{m-1}$ and $\hat c_0$, simulates n new paths independent of the training ones, and averages $Z_{\hat\tau^{(i)}}^{(i)}$: $\hat V_n^{\text{out}} = \frac{1}{n}\sum_{i=1}^n Z_{\hat\tau^{(i)}}^{(i)}$.

Step (3) is Theorem 14.2(iv) with $\hat C_k$ in place of C_k: after it, $Y^{(i)}$ is the discounted cash flow of path i under the estimated rule from t_k on, and the regression at t_{k-1} uses these cash flows, not the fitted values. Step (2) is the least-squares regression of Theorem 8.15(iii), now used to estimate a conditional expectation rather than an unconditional one. Regressing on in-the-money paths only is the choice of Longstaff and Schwartz (2001): the decision matters only there (Theorem 14.5(d)), and the fit is not spent on the region $S \geq K$ where the put is out of the money.

In code

```
Y = D[m] * g(S[:, m]) # discounted cash flows, time-0 units
for k in range(m - 1, 0, -1):
    Zk = D[k] * g(S[:, k]); idx = np.flatnonzero(Zk > 0)
    X = basis(S[idx, k]) # scaled basis, e.g. x = S / K
    beta, *_ = np.linalg.lstsq(X, Y[idx], rcond=None)
    ex = Zk[idx] >= X @ beta; Y[idx[ex]] = Zk[idx[ex]]
```

`lsm_fit`, `policy_cashflows` (the frozen rule) and `lsm_price` (both estimates, and a control variate) in `code/src/mfe230d/american.py`.

Proposition 14.6 (Longstaff–Schwartz: which way the estimates are biased)**.** *Let $g \geq 0$, $\mathbb{E}[Z_k] < \infty$ for every k, and let $\mathcal{G}$ be the σ-algebra generated by the training paths.*

*(i) (*Out of sample: low.*) Let the evaluation paths each have the law of S, and let them, as a family, be independent of $\mathcal{G}$. For every value of the training data, the rule (14.5) is a stopping time in $\mathcal{T}_0$. Its value $\hat L = \mathbb{E}[Z_{\hat\tau} \mid \mathcal{G}]$ satisfies $\hat L \leq U_0$, the out-of-sample estimator is conditionally unbiased for it, $\mathbb{E}[\hat V_n^{\text{out}} \mid \mathcal{G}] = \hat L$, and hence*

$$\mathbb{E}[\hat V_n^{\text{out}}] \leq U_0.$$

If the evaluation paths are i.i.d. (or i.i.d. antithetic pairs, whose averages are then the samples), then $\hat V_n^{\text{out}} \to \hat L$ almost surely as $n \to \infty$, and when $\mathbb{E}[Z_k^2] < \infty$ the central limit theorem (Theorem 6.4) applies conditionally on $\mathcal{G}$. If $\hat C_k(S_{t_k}) = C_k$ almost surely on $\{Z_k > 0\}$ for $1 \leq k \leq m-1$ and $\{Z_0 > 0, Z_0 \geq \hat c_0\} = \{Z_0 > 0, Z_0 \geq C_0\}$, then $\hat L = U_0$.

*(ii) (*In sample: no guarantee.*) Suppose that at every date the matrix* $(\psi_j(S_{t_k}^{(i)}))_{i\in I_k,\, j\le J}$ *has rank* $|I_k|$*, so that least squares interpolates the cash flows. Then after the backward pass* $Y^{(i)} = \max_{1\le k\le m} Z_k^{(i)}$ *for every* i*, and*

$$\mathbb{E}[\hat{V}_N^{\text{in}}] \;\ge\; \max\left(Z_0, \mathbb{E}[\max_{1\le k\le m} Z_k]\right) \;\ge\; U_0.$$

The last inequality is an equality if $Z_0 \ge \mathbb{E}[\max_{k\ge 1} Z_k]$*; otherwise it is strict unless the optimal rule from* t_1 *on stops, almost surely, at a date where the path attains* $\max_{k\ge 1} Z_k$*.*

Proof. (i) Fix the training data, so that $\hat{\beta}_k$ and $\hat{c}_0$ are numbers. Then $\{\hat{\tau} = k\}$ is determined by $S_{t_0}, \dots, S_{t_k}$, so $\hat{\tau} \in \mathcal{T}_0$, and $\mathbb{E}[Z_{\hat{\tau}}] \le U_0$ by Theorem 14.2(ii). The evaluation paths are independent of $\mathcal{G}$, so by the independence lemma $\hat{L}$ is this expectation evaluated at the training data, hence $\hat{L} \le U_0$; the same argument for each evaluation path gives $\mathbb{E}[\hat{V}_n^{\text{out}} \mid \mathcal{G}] = \hat{L}$, and taking expectations gives the inequality. For i.i.d. paths (or pairs) the terms are conditionally i.i.d. given $\mathcal{G}$, so the strong law and the central limit theorem apply conditionally. For the equality case, use Theorem 14.2(iii): at $\hat{\tau}$ we have $Z = U$ (where the rule stops before m, $Z_k \ge C_k$); and on $\{\hat{\tau} > j\}$ either $Z_j > 0$ and $Z_j < C_j$, or $Z_j = 0 \le C_j$ (as $g \ge 0$), and in both cases $U_j = C_j$, so the stopped Snell envelope is a martingale as in the proof of Theorem 14.2(ii).

(ii) Rank $|I_k|$ means the fitted values can match any vector of responses, so the least-squares residuals vanish: $\hat{C}_k(S_{t_k}^{(i)})$ is the current $Y^{(i)}$ for $i \in I_k$. Step (3) then replaces $Y^{(i)}$ by $Z_k^{(i)}$ exactly when $Z_k^{(i)} \ge Y^{(i)}$, that is, $Y^{(i)} \leftarrow \max(Z_k^{(i)}, Y^{(i)})$; for $i \notin I_k$, $Z_k^{(i)} = 0 \le Y^{(i)}$ and the same formula holds. By induction, $Y^{(i)} = \max_{j\ge k} Z_j^{(i)}$ after date k. The map $x \mapsto \max(Z_0, x)$ is convex, so by Jensen's inequality $\mathbb{E}[\hat{V}_N^{\text{in}}] \ge \max(Z_0, \mathbb{E}[\max_{k\ge 1} Z_k])$. Finally $\mathbb{E}[\max_{k\ge 1} Z_k] \ge \mathbb{E}[Z_{\tau_1^*}] = C_0$ by Theorem 14.2(iv), and $\max(Z_0, C_0) = U_0$. If $Z_0 < \mathbb{E}[\max_{k\ge 1} Z_k]$, the two maxima differ unless $C_0 = \mathbb{E}[\max_{k\ge 1} Z_k]$, that is, unless $\max_{k\ge 1} Z_k = Z_{\tau_1^*}$ almost surely. □

Part (ii) is an extreme case (our code skips the regression at dates with at most J in-the-money paths, so it never interpolates), but it shows the mechanism. In-sample, each path's own future enters the fit that decides when that path stops: a path whose future cash flow is high pulls its fitted continuation value up and is kept alive, and one whose future is low is exercised. That is foresight, and it biases upwards. Out of sample, the fitted function is just some function, the rule just some stopping time, and its value can only be below the optimum. Figure 14.1 shows both effects on the Longstaff–Schwartz put with $S_0 = 40$, $\sigma = 20\%$, $T = 1$. With $N = 100$ training paths and the basis $1, x, x^2$ ($J = 3$, $x = S/K$), the in-sample estimate is on average 0.28 above the price 2.3141, while the value of the fitted policy is on average 0.157 below it; each replication gives a policy worth less than the price, as (i) requires. The in-sample bias shrinks as N grows; with $J = 3$ it is -0.011 at $N = 10^4$, with a standard error of 0.005, as the in-sample estimate approaches the value of the limiting policy, which is below the price. More basis functions make the in-sample estimate worse for small N, and the policy better only once N is large. Longstaff and Schwartz (2001) found in-sample and out-of-sample estimates "virtually identical" with 10^5 paths and four basis functions, which Table 14.1 confirms; with few paths or many basis functions only the out-of-sample estimate is a lower bound.

Remark 14.7 (Convergence). Clément et al. (2002) analyse a variant that regresses on all paths, without the restriction $Z_k > 0$, and exercises where Z_k is at least the fitted value; it is not the rule (14.5). Let $X_k = S_{t_k}$, $Z_k = f_k(X_k)$ with $\mathbb{E}[Z_k^2] < \infty$, and let $\psi_1, \psi_2, \dots$ be functions such

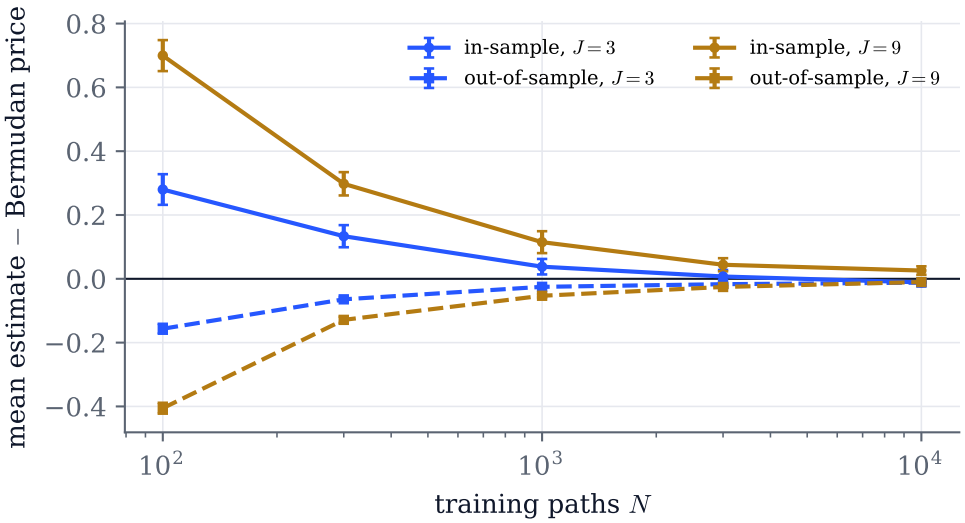

Figure 14.1: Longstaff–Schwartz put, $S_0 = K = 40$, $\sigma = 20\%$, $T = 1$, 50 dates (price 2.3141). For each number N of training paths, LSM is fitted on 150, 80, 40, 30 and 20 independent sets of N paths (for $N = 100, 300, 1000, 3000, 10^4$) with the monomials $1, x, \dots, x^{J-1}$ in $x = S/K$, $J = 3$ and $J = 9$. Solid: mean in-sample estimate minus the price. Dashed: mean value of the fitted policy minus the price, the value computed exactly on the grid, which is the conditional mean of the out-of-sample estimator given the training paths. Bars: two standard errors across the sets. Our implementation (NumPy `default_rng`, seeds $[8, J-1, N, j]$).

that, at every date $1 \le k \le m-1$, $\psi_1(X_k), \dots, \psi_J(X_k)$ are linearly independent in L^2 for every J. Let $\tau^{[J]}$ be their rule built from the exact L^2 projections onto the first J functions (the regression with infinitely many paths). Their Theorem 3.1: if moreover the whole sequence is total in $L^2(\sigma(X_k))$, then as $J \to \infty$, $\mathbb{E}[Z_{\tau_k^{[J]}} \mid \mathcal{F}_k] \to \mathbb{E}[Z_{\tau_k^*} \mid \mathcal{F}_k]$ in L^2 for every k, so the value of $\tau^{[J]}$ tends to U_0. Their Theorem 3.2, which needs linear independence but not totality: for fixed J, if the projected continuation value ties with the payoff with probability zero at every date, the in-sample estimate converges almost surely to the value of $\tau^{[J]}$ as $N \to \infty$, which is at most U_0 by Theorem 14.2(ii); they also prove a central limit theorem for this convergence. The two limits are taken one after the other. How fast N must grow with J is studied by Glasserman and Yu (2004): for polynomial bases, the number of paths needed for convergence in the worst case grows exponentially in the degree for Brownian motion, and faster for geometric Brownian motion. See also Glasserman (2003), Section 8.6.

Remark 14.8 (Bases and conditioning). The basis should be scaled. Longstaff and Schwartz (2001) use a constant and the weighted Laguerre functions $e^{-x/2}$, $e^{-x/2}(1-x)$ and $e^{-x/2}(1-2x+\frac{x^2}{2})$; we use them in the scaled variable $x = S/K$, because unscaled columns of very different sizes make the regression matrix badly conditioned. On the in-the-money prices at $t = 0.5$ of the put with $S_0 = 36$, $\sigma = 20\%$ (10^5 paths, seed 0), the condition number of the regression matrix is 2.6×10^7 for $1, S, S^2, S^3$ and 5.9×10^3 for $1, x, x^2, x^3$; 1.1×10^8 and 1.4×10^5 for the Laguerre basis in S and in x; and 2.7×10^9 for the monomials of degree 8 in x. Solving the normal equations squares these numbers; `numpy.linalg.lstsq` works with the matrix itself (by a singular value decomposition) and returns a least-squares solution even when the matrix is rank-deficient. Our

code skips the regression, and never exercises, at a date with at most J usable paths: any rule that depends only on the path up to t_k is a stopping time, so Theorem 14.6(i) is not affected.

Table 1 of Longstaff and Schwartz. Table 14.1 reproduces the table with which Longstaff and Schwartz (2001) introduced the method, with their setup: 10^5 paths (5×10^4 and their antithetic images), 50 dates per year, the four basis functions above. Our closed-form European prices agree with theirs to the three printed decimals. Two findings about their columns. First, their "finite-difference American" prices agree with our *Bermudan* prices to within 0.0005 in the 15 rows with $\sigma = 20\%$ or $T = 1$, and lie between our Bermudan and American prices, up to the last printed digit, in the five rows with $\sigma = 40\%$, $T = 2$; the American prices exceed the Bermudan ones by 0.003 to 0.009. They do not say whether their grid allowed exercise at every time step; the comparison suggests that their column approximates the Bermudan price, which is also what LSM estimates. Second, their standard errors are close to the naive formula $\hat{s}_n/\sqrt{n}$ over all 10^5 paths (we get 0.0092, 0.0086, 0.0206 for the first, ninth and last rows; they print 0.010, 0.009, 0.021). But antithetic pairs are negatively correlated: the correlation ρ of the two cash flows of a pair is between -0.68 and -0.56 in these rows, and the standard error from the 5×10^4 pair averages (Theorem 8.2), which we report, is $\sqrt{1+\rho}$ times the naive one, that is 0.57 to 0.66 times. All our in-sample estimates are within 1.3 of our standard errors of the Bermudan price, and the out-of-sample estimates within 2.4, with mean deviations of -0.2 and -0.3 standard errors: the bias is small compared with the noise at this sample size.

Figure 14.2 looks inside the algorithm for the first row. At $t = 0.5$, panel (a) shows the realised cash flows that step (2) regresses (minus the exercise value $K - s$, so that the decision is the sign), the fitted continuation value and the exact one from the grid. The fit is within a few cents of the truth over the range where paths are, and it crosses the exercise value close to where the truth does. Panel (b) compares the exercise boundaries $b(t_k)$, the largest price at which each rule exercises. The LSM boundaries are noisy and miss the steep rise near maturity, where few paths are close to the boundary. The value lost is small. Heuristically, payoff and continuation value are equal at the exact boundary, so exercising on the wrong side of it by a distance δ loses an amount of order δ^2 per date. The LSM policy is worth 4.4744 against the price 4.4778 (Table 14.2).

Example 14.9 (Control variates, and the test case)**.** The out-of-sample estimator averages independent copies of $Z_{\hat{\tau}}$, so every method of Chapter 8 applies to it. The discounted European payoff $X = e^{-rT} g(S_T)$ has a known mean, the Black–Scholes price, and is computed along the same paths; we estimate its coefficient on the training paths, a pilot sample independent of the evaluation paths, so that the controlled estimator stays unbiased for $\hat{L}$ and Theorem 14.6(i) is not disturbed (Theorem 8.14). On the first row of Table 14.1 with plain paths (seed 3) the correlation is 0.56 and the variance ratio 0.68: deep in the money, many paths are exercised early and their cash flow is weakly related to S_T. On the test-case put (seed 1, 10^5 paths, the Laguerre basis in $x = S/100$) the correlation is 0.93 and the variance ratio 0.13: the control estimate is 8.3024 with standard error 0.0110, against 8.2766 and 0.0308 without it. Both are well below the price 8.34941, and this is not noise: the fitted policy is worth 8.3079 (computed on the grid) and exercises 35% of the paths before maturity, where the optimal rule exercises essentially never (Theorem 14.5(b)). The early exercises happen in the last quarter of the year at prices around 80 (between 76 and 88 for 90% of them), where the time value of the put, measured by the European put, is between 2 and 19 cents: comparable with the error of the fit, which

			ours, no simulation		Longstaff–Schwartz		ours, LSM (s.e.)	
S_0	σ	T	Bermudan	American	"FD"	LSM (s.e.)	in-sample	out-of-sample
36	0.2	1	4.478	4.487	4.478	4.472 (.010)	4.482 (.006)	4.470 (.006)
36	0.2	2	4.840	4.848	4.840	4.821 (.012)	4.839 (.007)	4.841 (.007)
36	0.4	1	7.101	7.109	7.101	7.091 (.020)	7.104 (.008)	7.088 (.008)
36	0.4	2	8.507	8.514	8.508	8.488 (.024)	8.517 (.010)	8.507 (.010)
38	0.2	1	3.250	3.257	3.250	3.244 (.009)	3.247 (.005)	3.253 (.005)
38	0.2	2	3.745	3.751	3.745	3.735 (.011)	3.745 (.006)	3.751 (.006)
38	0.4	1	6.148	6.155	6.148	6.139 (.019)	6.138 (.008)	6.145 (.008)
38	0.4	2	7.668	7.675	7.670	7.669 (.022)	7.675 (.010)	7.674 (.010)
40	0.2	1	2.314	2.320	2.314	2.313 (.009)	2.309 (.005)	2.310 (.005)
40	0.2	2	2.885	2.890	2.885	2.879 (.010)	2.879 (.006)	2.877 (.006)
40	0.4	1	5.312	5.318	5.312	5.308 (.018)	5.309 (.009)	5.305 (.009)
40	0.4	2	6.917	6.923	6.920	6.921 (.022)	6.906 (.010)	6.919 (.010)
42	0.2	1	1.617	1.621	1.617	1.617 (.007)	1.624 (.006)	1.607 (.006)
42	0.2	2	2.212	2.217	2.212	2.206 (.010)	2.207 (.007)	2.219 (.007)
42	0.4	1	4.582	4.588	4.582	4.588 (.017)	4.578 (.010)	4.559 (.010)
42	0.4	2	6.244	6.250	6.248	6.243 (.021)	6.231 (.011)	6.254 (.011)
44	0.2	1	1.110	1.113	1.110	1.118 (.007)	1.111 (.006)	1.112 (.006)
44	0.2	2	1.690	1.693	1.690	1.675 (.009)	1.690 (.007)	1.690 (.007)
44	0.4	1	3.948	3.953	3.948	3.957 (.017)	3.950 (.011)	3.931 (.011)
44	0.4	2	5.641	5.647	5.647	5.622 (.021)	5.634 (.012)	5.644 (.012)

Table 14.1: American puts with $K = 40$, $r = 6\%$ and 50 exercise dates per year. Bermudan: grid of Theorem 14.4; American: smoothed binomial tree with Richardson extrapolation. Longstaff–Schwartz: their "finite difference American" and LSM columns, as printed. Ours: 10^5 antithetic paths and the four Laguerre basis functions for both training and evaluation, standard errors from pair averages; our implementation (NumPy `default_rng`, seeds $[2001, i, 0]$ for training and $[2001, i, 1]$ for evaluation in row $i = 0, \dots, 19$).

then exercises paths whose continuation value it underestimates. The policy loses 0.04, some 10^4 times the early-exercise premium, and the in-sample estimate, 8.2729, does not reveal it. For the test-case call the same setup gives 7.4949 in sample and 7.4237 (0.0330) out of sample, around the policy value 7.4462 and below the Bermudan price 7.4567.

▶ Interactive demo: the regression of step (2) at any date, the in-sample and out-of-sample estimates for any number of paths and basis, and the exercise boundaries.

14.3 Duality: upper bounds

A lower bound alone does not tell how far below the price it is. Upper bounds come from a dual problem. Its simplest case is perfect foresight: $\max_k Z_k \geq Z_\tau$ for every τ, so $\mathbb{E}[\max_k Z_k] \geq U_0$. This bound is poor because the holder who sees the future exercises each path at its best date. Subtracting a martingale with mean zero before taking the maximum removes part of that foresight without changing the inequality.

Theorem 14.10 (Duality). *Let Z be adapted with $\mathbb{E}[\max_k |Z_k|] < \infty$, U its Snell envelope, and M^* the martingale of (14.3).*

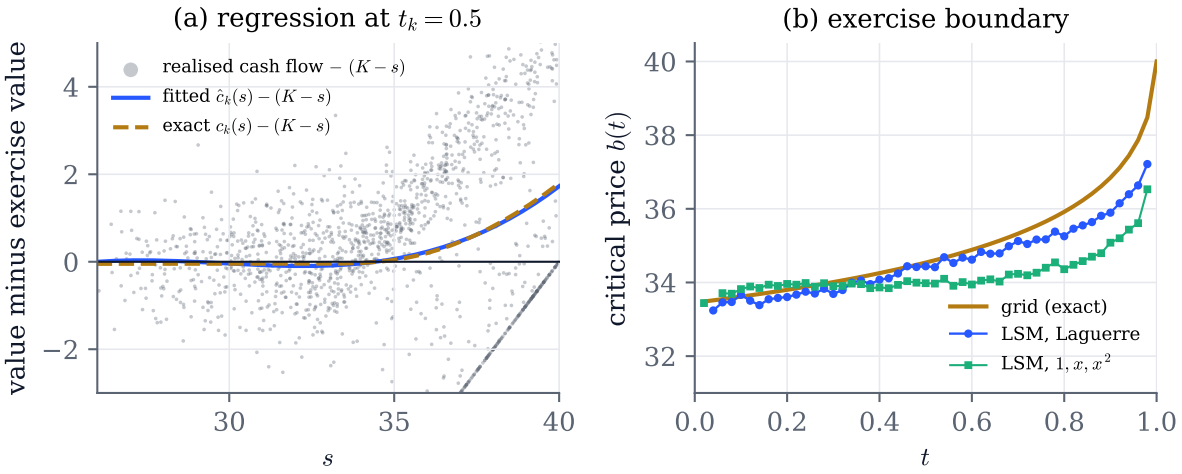

Figure 14.2: Longstaff–Schwartz put with $S_0 = 36$, $\sigma = 20\%$, $T = 1$, 50 dates; LSM on 10^5 antithetic training paths (seed $[14, 0]$). (a) At $t_k = 0.5$: 1500 randomly chosen in-the-money realised cash flows (in time-t_k units) minus $K - s$, the fitted continuation value with the four Laguerre functions and the exact one, both minus $K - s$; exercise where the curve is below zero. (b) Exercise boundary: the critical price from the grid, and the largest price at which the LSM rule exercises, with the Laguerre basis and with $1, x, x^2$, $x = s/K$.

(i) (Upper bounds.) *For every martingale M with $M_0 = 0$, and more generally every supermartingale with $M_0 = 0$,*

$$U_0 \le \mathbb{E}\Big[\max_{0\le k\le m}(Z_k - M_k)\Big].$$

(ii) (The bound is attained, without variance.) $\max_k(Z_k - M_k^*) = U_0$ *almost surely. Hence* $U_0 = \min_M \mathbb{E}[\max_k(Z_k - M_k)]$ *over martingales with* $M_0 = 0$, *attained at* M^*.

(iii) (The gap.) *If* $\mathbb{E}[Z_k^2] < \infty$ *for every* k *and* M *is a square-integrable martingale with* $M_0 = 0$, *then*

$$0 \le \mathbb{E}\Big[\max_k(Z_k - M_k)\Big] - U_0 \le \mathbb{E}\Big[\max_k |M_k - M_k^*|\Big] \le 2\,\|M_m - M_m^*\|_{L^2}.$$

Proof. (i) Every M_k is integrable, so $|\max_k(Z_k - M_k)| \le \sum_k(|Z_k| + |M_k|)$ is integrable. For $\tau \in \mathcal{T}_0$, Theorem 14.1 gives $\mathbb{E}[M_\tau] \le M_0 = 0$, so $\mathbb{E}[Z_\tau] \le \mathbb{E}[Z_\tau - M_\tau] \le \mathbb{E}[\max_k(Z_k - M_k)]$. Take the supremum over τ.

(ii) By Theorem 14.2(v), $Z_k - M_k^* \le U_k - M_k^* = U_0 - A_k^* \le U_0$ for every k, so $\max_k(Z_k - M_k^*) \le U_0$ surely; by (i) its mean is at least U_0. A random variable bounded above by U_0 with mean at least U_0 equals U_0 almost surely.

(iii) $Z_k - M_k = (Z_k - M_k^*) + (M_k^* - M_k) \le U_0 + \max_j |M_j^* - M_j|$ by (ii). Under the hypothesis, U and hence M^* are square-integrable (by the proof of Theorem 14.2(i), $|U_k| \le \mathbb{E}[\max_j |Z_j| \mid \mathcal{F}_k]$), so $N = M^* - M$ is a square-integrable martingale, and by the Cauchy–Schwarz inequality and Doob's L^2 maximal inequality (Durrett, 2019, Section 4.4), $\mathbb{E}[\max_k |N_k|] \le \mathbb{E}[\max_k N_k^2]^{1/2} \le$

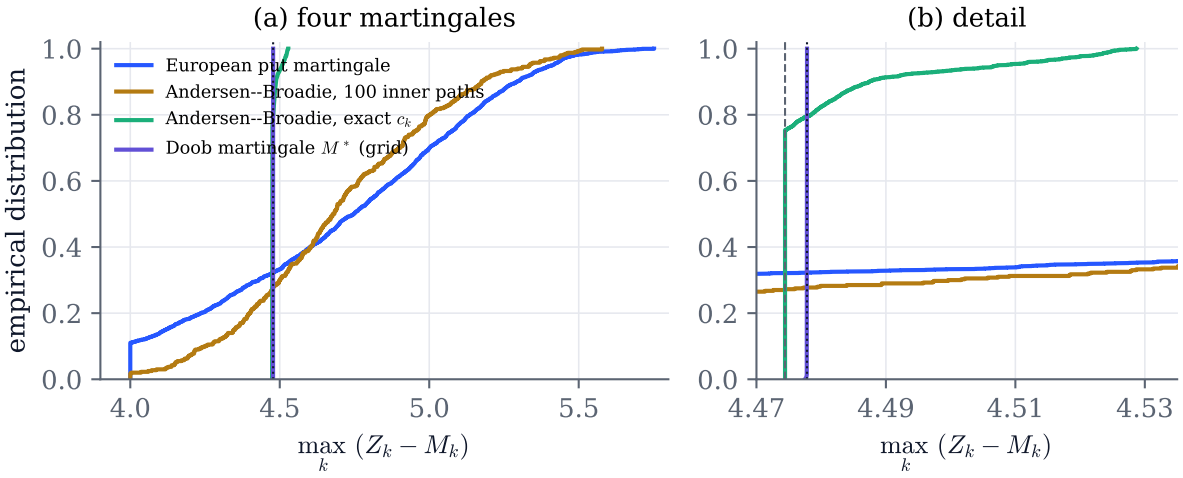

Figure 14.3: Empirical distribution functions of the dual samples $\max_k(Z_k - M_k)$ for the setting of Table 14.2: the European put martingale (2000 outer paths), the Andersen–Broadie martingale of the LSM policy estimated with 100 inner paths (400 outer paths), the same martingale with exact continuation values, and the Doob martingale of the grid value function (2000 outer paths). Dotted: the price U_0. Dashed in (b): the value $L_0 = 4.4744$ of the policy. The nested sizes are kept small to keep the figure cheap; Table 14.2 reports a run with 500 inner paths.

$2\,\mathbb{E}[N_m^2]^{1/2}$. □

The theorem is due to Rogers (2002), in continuous time, and Haugh and Kogan (2004), in discrete time. Rogers states it over martingales M with $M_0 = 0$ and $\mathbb{E}[\sup_t |M_t|] < \infty$, with $\sup_{t\le T}(Z_t - M_t)$ in place of the maximum, and Haugh and Kogan over supermartingales, with the bound written $\pi_0 + \mathbb{E}[\max_k(Z_k - \pi_k)]$. Part (ii) is the striking part: the optimal martingale M^* is the martingale part of the discounted option price, the gains of the seller's hedging portfolio, and with it every single path returns the price. This suggests that a martingale close to M^* gives an upper bound with small variance, which makes the dual estimator cheap where it works. Part (iii) says the gap is controlled by how well M approximates M^* in mean square, which is how martingales are built in practice: from an approximation of the value process (Haugh and Kogan, 2004), or from an exercise policy (Andersen and Broadie, 2004).

Martingales in closed form. $M = 0$ gives perfect foresight. Another martingale with $M_0 = 0$ that needs no simulation is the discounted European price minus its initial value, $M_k = e^{-rt_k}P(t_k, S_{t_k}) - P(0, S_0)$ with $P(t, s)$ the Black–Scholes price of the European claim with payoff g at T; for a put it removes the part of the foresight that a European put already prices. On the first row of Table 14.1, Table 14.2 shows that perfect foresight gives 7.78 for a price of 4.48, and the European martingale 4.732: better, but 5.7% high. Exercise 14.6 improves it.

Proposition 14.11 (Andersen–Broadie)**.** *Let Z be adapted with $\mathbb{E}[\max_k |Z_k|] < \infty$. Let a policy be given by events $A_k \in \mathcal{F}_k$ (exercise at t_k on A_k), $k < m$, with $A_m = \Omega$, and let $\tau_k = \min\{j \ge k : \omega \in A_j\}$. Let*

$$L_k = \mathbb{E}[Z_{\tau_k} \mid \mathcal{F}_k], \qquad \bar{L}_k = \mathbb{E}[Z_{\tau_{k+1}} \mid \mathcal{F}_k] \ (k < m), \qquad M_0 = 0, \quad M_{k+1} = M_k + L_{k+1} - \bar{L}_k.$$

(i) *$L_k = Z_k$ on A_k and $L_k = \bar{L}_k$ off A_k; M is a martingale; and with $H = \mathbb{E}[\max_k(Z_k - M_k)]$,*

$$L_0 \le U_0 \le H.$$

(ii) *For every k, $Z_k - M_k = L_0 + (Z_k - L_k) + \sum_{j<k} \mathbf{1}_{A_j}(\bar{L}_j - Z_j)$. Hence $\max_k(Z_k - M_k) \ge L_0$ surely, and*

$$0 \le H - L_0 \le \mathbb{E}\Big[\sum_{k<m} \Big(\mathbf{1}_{A_k^c}(Z_k - \bar{L}_k)^+ + \mathbf{1}_{A_k}(\bar{L}_k - Z_k)^+\Big)\Big].$$

In particular, if $A_k = \{Z_k \ge \bar{L}_k\}$ for every $k < m$, then $L_0 = U_0 = H$.

(iii) *(*Nested simulation is biased high.*) Let $\hat{L}_k$ $(k < m)$ be random variables with $\mathbb{E}[|\hat{L}_k|] < \infty$ and $\mathbb{E}[\hat{L}_k \mid \mathcal{F}_T] = \bar{L}_k$, and let $\hat{M}$ be defined like M with $\hat{L}_k$ in place of $\bar{L}_k$, both in the increments and in L_k off A_k. Then*

$$\mathbb{E}\Big[\max_k(Z_k - \hat{M}_k)\Big] \;\ge\; H, \qquad \mathbb{E}\Big[\max_k(Z_k - \hat{M}_k)\Big] - H \le \sum_{k<m} \mathbb{E}[|\hat{L}_k - \bar{L}_k|].$$

(iv) *(*A confidence interval.*) Let $\hat{V}^{\mathrm{low}}$ be the average of n_1 independent copies of Z_{τ_0} and $\hat{V}^{\mathrm{up}}$ the average of n_2 independent copies of $\max_k(Z_k - \hat{M}_k)$ (outer paths with their own inner simulations), with standard errors $\hat{\sigma}_1/\sqrt{n_1}$ and $\hat{\sigma}_2/\sqrt{n_2}$, and both variances finite and positive. Then, as $n_1, n_2 \to \infty$, the interval $[\hat{V}^{\mathrm{low}} - z_\delta\hat{\sigma}_1/\sqrt{n_1},\ \hat{V}^{\mathrm{up}} + z_\delta\hat{\sigma}_2/\sqrt{n_2}]$ contains U_0 with probability at least $1 - \delta + o(1)$.*

Proof. (i) On A_k, $\tau_k = k$; off A_k, $\tau_k = \tau_{k+1}$; both are $\mathcal{F}_k$-measurable events, which gives the formula for L_k. By the tower property $\mathbb{E}[L_{k+1} \mid \mathcal{F}_k] = \mathbb{E}[Z_{\tau_{k+1}} \mid \mathcal{F}_k] = \bar{L}_k$, so M has increments with conditional mean zero, and it is integrable because $|L_k|, |\bar{L}_k| \le \mathbb{E}[\max_j |Z_j| \mid \mathcal{F}_k]$. Then $L_0 = \mathbb{E}[Z_{\tau_0}] \le U_0 \le H$ by Theorem 14.2(ii) and Theorem 14.10(i).

(ii) By (i), $\bar{L}_j = L_j + \mathbf{1}_{A_j}(\bar{L}_j - Z_j)$, so $M_k = \sum_{j<k}(L_{j+1} - L_j) - \sum_{j<k}\mathbf{1}_{A_j}(\bar{L}_j - Z_j) = L_k - L_0 - \sum_{j<k}\mathbf{1}_{A_j}(\bar{L}_j - Z_j)$, which is the identity. At $k = \tau_0$ the last two terms vanish ($Z = L$ there, and no A_j occurs before), so the maximum is at least L_0. For the bound, $Z_k - L_k$ is 0 on A_k and $Z_k - \bar{L}_k$ off it, and for real numbers $\max_k (a_k + \sum_{j<k} b_j) \le \sum_k a_k^+ + \sum_j b_j^+$. If $A_k = \{Z_k \ge \bar{L}_k\}$ every term of the bound vanishes, so $H \le L_0$, and (i) gives equality throughout.

(iii) Off A_{j+1} the increment of $\hat{M}$ uses $\hat{L}_{j+1}$ for L_{j+1}. With $\delta_j = \hat{L}_j - \bar{L}_j$ for $j < m$ and $\delta_m = 0$, the difference $\varepsilon_k = \hat{M}_k - M_k$ is

$$\varepsilon_k = \sum_{j<k} (\mathbf{1}_{A_{j+1}^c}\delta_{j+1} - \delta_j) = \mathbf{1}_{A_k^c}\delta_k - \delta_0 - \sum_{0<j<k} \mathbf{1}_{A_j}\delta_j \qquad (k \ge 1),$$

and $\mathbb{E}[\varepsilon_k \mid \mathcal{F}_T] = 0$ because the events A_j are in $\mathcal{F}_T$. The function $x \mapsto \max_k(Z_k - M_k - x_k)$ is convex, so by the conditional Jensen inequality $\mathbb{E}[\max_k(Z_k - \hat{M}_k) \mid \mathcal{F}_T] \ge \max_k(Z_k - M_k)$; take expectations. For the second claim, $|\max_k(Z_k - \hat{M}_k) - \max_k(Z_k - M_k)| \le \max_k |\varepsilon_k| \le \sum_{j<m} |\delta_j|$, since each δ_j appears at most once in ε_k.

(iv) The two estimators are independent, $\mathbb{E}[\hat{V}^{\mathrm{low}}] = L_0 \le U_0$ and $\mathbb{E}[\hat{V}^{\mathrm{up}}] \ge H \ge U_0$ by (i) and (iii). If U_0 is below the interval, then $\hat{V}^{\mathrm{low}} - z_\delta\hat{\sigma}_1/\sqrt{n_1} > L_0$, an event whose probability tends to $\delta/2$ by Theorem 6.4(ii); likewise above. The union of the two events has probability at most $\delta + o(1)$. □

The policy of interest is the LSM rule, with A_k the exercise region of (14.5) and the training paths held fixed (conditionally on $\mathcal{G}$, as in Theorem 14.6). Then $\bar{L}_k$ is the continuation value of the LSM policy at the simulated state, an expectation that the regression did not compute. The algorithm of Andersen and Broadie (2004) estimates it, along each *outer* path and at each date, by the mean of n_{in} *inner* paths started from the current state and stopped by the policy; these estimates are conditionally unbiased, so (iii) applies. If $\hat{L}_k$ is the mean of n_{in} inner samples with conditional variance v_k, then $\mathbb{E}|\hat{L}_k - \bar{L}_k| \le (\mathbb{E}[v_k]/n_{\text{in}})^{1/2}$, and the bias is at most $\sum_k (\mathbb{E}[v_k])^{1/2}/\sqrt{n_{\text{in}}}$, of order $m/\sqrt{n_{\text{in}}}$. The cost is of order $n_{\text{out}}\, n_{\text{in}}\, m^2/2$ simulated steps. Part (ii) explains the quality of the bound: the gap is made of the policy's own mistakes, continuing where its continuation value is below the payoff or exercising where it is above, measured against its own continuation values. A self-consistent policy has no gap, and a policy that errs rarely and by little has a small one.

Table 14.2 and Figure 14.3 put the pieces together for the put with $S_0 = 36$. The LSM policy of Figure 14.2 is worth 4.4744, 0.0034 below the price. Its Andersen–Broadie martingale, with continuation values computed exactly on the grid, gives the upper bound 4.4786 with a standard error of 0.0002 from 2000 outer paths: the duality gap of the policy is 0.0042, and 75% of the outer paths return exactly L_0, which part (ii) predicts for every path on which the policy makes no mistake. Nested simulation is another matter: with 100 inner paths per date the estimator is 4.705, and with 500 it is 4.537. The inner noise enters through the maximum over 50 dates and dominates the gap; reducing it takes many inner paths or variance reduction inside the inner simulation, at a cost that grows like m^2. We keep the nested runs small, with 400 and 200 outer paths, so that tests and figures stay cheap; the standard errors are those of these runs. With the out-of-sample LSM estimate and the nested estimator with 500 inner paths, part (iv) gives the 95% interval $[4.461, 4.558]$, which contains the price; with exact continuation values it would be $[4.461, 4.479]$. The Doob martingale of the grid value function, finally, illustrates Theorem 14.10(ii): its 2000 paths return 4.47781 with a standard deviation of 1.6×10^{-5}, the error of the grid and of the linear interpolation.

		estimate	s.e.	paths
price	grid (Bermudan, 50 dates)	4.4778		
lower	LSM policy, exact value (grid)	4.4744		
	LSM policy, out of sample	4.4793	0.0094	10^5
upper	$M = 0$ (perfect foresight)	7.779	0.010	10^5
	European put martingale	4.7322	0.0014	10^5
	Andersen–Broadie, 100 inner paths	4.705	0.017	400
	Andersen–Broadie, 500 inner paths	4.537	0.011	200
	Andersen–Broadie, exact continuation values	4.4786	0.0002	2000
	Doob martingale of the grid value function	4.47781	$< 10^{-6}$	2000

Table 14.2: Lower and upper bounds for the Longstaff–Schwartz put with $S_0 = 36$, $\sigma = 20\%$, $T = 1$, 50 dates. The LSM policy is the one of Figure 14.2. The two closed-form bounds use 10^5 paths of their own; the other upper bounds share 2000 outer paths (the nested estimators use the first 400 or 200). Our implementation (NumPy `default_rng`, seeds $[14, j]$; see `duality_study`).

▶ Interactive demo: upper bounds from perfect foresight, the European martingale, the Andersen–Broadie martingale with nested simulation and with exact continuation values, and the Doob

martingale, for any strike and volatility.

Further reading

Glasserman (2003), Chapter 8, is the reference for this chapter: the problem in Section 8.1, regression in Section 8.6, duality in Section 8.7. Longstaff and Schwartz (2001) introduced the algorithm used here, after Carrière (1996) and alongside Tsitsiklis and Van Roy (2001), who regress values instead of cash flows. Clément et al. (2002) prove its convergence and Glasserman and Yu (2004) study the balance of paths and basis functions. Duality is due to Rogers (2002) and Haugh and Kogan (2004), and the primal–dual algorithm to Andersen and Broadie (2004). Peskir and Shiryaev (2006), Chapter I, and Föllmer and Schied (2016), Chapter 6, treat optimal stopping and American claims in general.

Part III

Path-dependent and multi-asset payoffs

Chapter 15

Asian options

An *Asian option* pays a function of the average of the underlying price over a schedule of *fixing dates*. The average-price call on oil or on a currency pays $(A-K)^+$, where A is the mean of daily or monthly prices. Averaging matches the exposure of a producer or an importer who trades at many dates, makes the payoff hard to move by trading at a single close, and usually makes the option cheaper than the European call with the same maturity: always when $r \geq y$ (Exercise 15.1), but not when the yield exceeds the rate and the strike is low (in the test case, 39.068 against 38.647 at $K = 60$).

The arithmetic average of lognormal prices is not lognormal, and its call has no closed form. The *geometric* average is lognormal, and this chapter uses it three ways: as an exact price (Section 15.2), as a control variate (Section 15.3), and as a variable to condition on (Section 15.5), which yields lower bounds and very accurate approximations. Moment matching (Section 15.4) is the other classical route. Section 15.6 computes the course's monthly example to nine digits and compares everything with it. Averages observed continuously, and the pricing equation they satisfy, are treated in the IEOR 222 notes; here the averages are discrete, as in traded contracts.

15.1 Average-price options

Under $\mathbb{Q}$ the stock follows the model (7.1) of Section 7.1, $\mathrm{d}S_t/S_t = (r(t)-y(t))\,\mathrm{d}t + \sigma(t)\,\mathrm{d}W_t$, with bounded Borel functions r, y and $\sigma \geq 0$. A contract fixes dates $0 \leq t_1 < \cdots < t_m \leq T$, a payment date T and weights $w_j > 0$ with $\sum_j w_j = 1$ (equal weights $1/m$ unless stated), and defines

$$A = \sum_{j=1}^{m} w_j S_{t_j}, \quad G = \prod_{j=1}^{m} S_{t_j}^{w_j}, \quad C = D\,\mathbb{E}^{\mathbb{Q}}[(A-K)^+], \quad C^G = D\,\mathbb{E}^{\mathbb{Q}}[(G-K)^+], \quad D = e^{-\int_0^T r}.$$

C is the arithmetic average-price call and C^G its geometric twin; P denotes the arithmetic put, paying $(K-A)^+$. A fixing at $t_1 = 0$ is allowed: it fixes the known price $S_0 = S$. Write $F_j = \mathbb{E}^{\mathbb{Q}}[S_{t_j}] = S\,e^{\int_0^{t_j}(r-y)}$ for the forwards and $\Theta(t) = \int_0^t \sigma^2$ for the variance clock. By Theorem 7.2(i), $\log S_t = \log S + \int_0^t (r - y - \frac{1}{2}\sigma^2) + M_t$ with $M_t = \int_0^t \sigma\,\mathrm{d}W$, a centred Gaussian process with $\mathrm{Cov}(M_s, M_t) = \Theta(s \wedge t)$ (Theorem 7.5). Hence the vector $(\log S_{t_1}, \ldots, \log S_{t_m})$ is

Gaussian, with

$$\mathbb{E}\log S_{t_j} = \log F_j - \tfrac{1}{2}\Theta(t_j), \qquad \operatorname{Cov}(\log S_{t_i}, \log S_{t_j}) = C_{ij} := \Theta(t_i \wedge t_j). \tag{15.1}$$

We use one Gaussian identity repeatedly: if (U, V) is a Gaussian vector, then

$$\mathbb{E}[e^U \mathbf{1}_{\{V>x\}}] = \mathbb{E}[e^U]\,\mathbb{P}(V + \operatorname{Cov}(U,V) > x), \qquad x \in \mathbb{R}. \tag{15.2}$$

Indeed, under the probability with density $e^U/\mathbb{E}e^U$, the Gaussian moment generating function at the complex argument $(1, i\lambda)$ gives the characteristic function of V,

$$\mathbb{E}[e^{U+i\lambda V}]/\mathbb{E}[e^U] = \exp\left(i\lambda(\mathbb{E}V + \operatorname{Cov}(U,V)) - \tfrac{1}{2}\lambda^2 \operatorname{Var} V\right),$$

so V is normal with its mean shifted by $\operatorname{Cov}(U,V)$ and the same variance.

Proposition 15.1 (Parity, bounds and degenerate cases). *(i)* $C - P = D\big(\sum_j w_j F_j - K\big)$.

(ii) $G \le A$, *and* $D\big(\sum_j w_j F_j - K\big)^+ \le C \le D\sum_j w_j \mathbb{E}^{\mathbb{Q}}[(S_{t_j} - K)^+]$ *and* $C^G \le C \le C^G + D\,\mathbb{E}^{\mathbb{Q}}[A - G]$.

(iii) If $K \le 0$, *then* $C = D(\sum_j w_j F_j - K)$. *If* $\sigma = 0$ *almost everywhere on* $[0, t_m]$, *then* $C = D(\sum_j w_j F_j - K)^+$. *If* $m = 1$, *then* $C = e^{-\int_{t_1}^T r}$ *times the Black–Scholes price of a call with maturity* t_1 *and the term parameters* $r_{0,t_1}, y_{0,t_1}, \sigma_{0,t_1}$ *(Theorem 7.2(iii)).*

Proof. (i) $x^+ - (-x)^+ = x$ with $x = A - K$. (ii) log is concave, so Jensen's inequality for the weights gives $\log A \ge \sum_j w_j \log S_{t_j} = \log G$. The map $x \mapsto (x-K)^+$ is convex, nondecreasing and 1-Lipschitz. Convexity gives $(\mathbb{E}A - K)^+ \le \mathbb{E}(A-K)^+$ and $(A-K)^+ \le \sum_j w_j (S_{t_j} - K)^+$; monotonicity and the Lipschitz property give $0 \le (A-K)^+ - (G-K)^+ \le A - G$. Take expectations. (iii) If $K \le 0$, $A - K > 0$. If $\sigma = 0$ on $[0, t_m]$, $\Theta(t_j) = 0$ and $A = \sum_j w_j F_j$ is deterministic. If $m = 1$, $A = S_{t_1}$. □

The upper bound in (ii) is the *strip*: w_j calls maturing at t_j, each with its payment deferred to T. *Seasoned options.* Once some fixings are known, say s_j for j in a set $\mathcal{P}$, write $A = a + WA'$ with $a = \sum_{\mathcal{P}} w_j s_j$, $W = \sum_{j \notin \mathcal{P}} w_j$ and A' the average of the remaining fixings with weights w_j/W. Then $(A-K)^+ = W(A' - K')^+$ with $K' = (K-a)/W$: the option is W fresh Asian options on the remaining dates with the shifted strike K', valued by restarting the model at the valuation date. If $K' \le 0$ the call will be exercised for sure and, by (iii), it is a forward on the average: linear in the spot, with no gamma. A fixing at $t_1 = 0$ is the case of a known first fixing.

15.2 The geometric average

Proposition 15.2 (The geometric average with term structures). *Let* $t_0 = 0$, $\Delta\Theta_k = \Theta(t_k) - \Theta(t_{k-1})$ *and* $W_k = \sum_{j \ge k} w_j$.

(i) $\log G \sim \mathcal{N}(\mu_G, v_G)$ *with*

$$\mu_G = \sum_{j=1}^m w_j\big(\log F_j - \tfrac{1}{2}\Theta(t_j)\big), \qquad v_G = \sum_{i,j=1}^m w_i w_j\,\Theta(t_i \wedge t_j) = \sum_{k=1}^m W_k^2\,\Delta\Theta_k.$$

(ii) Let $F_G = \mathbb{E}^{\mathbb{Q}}[G] = e^{\mu_G + v_G/2}$. *If* $v_G > 0$ *and* $K > 0$,

$$C^G = D(F_G\, N(d_+) - K\, N(d_-)), \quad d_\pm = \frac{\log(F_G/K) \pm \frac{1}{2} v_G}{\sqrt{v_G}}, \quad \frac{\partial C^G}{\partial S} = \frac{D\, F_G\, N(d_+)}{S}.$$

If $v_G = 0$, $C^G = D(F_G - K)^+$; *if* $K \le 0$, $C^G = D(F_G - K)$.

(iii) With equal weights, constant r, y, σ *and* $t_j = jT/m$: $\mu_G = \log S + \nu T \frac{m+1}{2m}$ *and* $v_G = \sigma^2 T \frac{(m+1)(2m+1)}{6m^2}$.

Proof. (i) $\log G = \sum_j w_j \log S_{t_j}$ is a linear function of the Gaussian vector (15.1). For the second form of v_G, write $\Theta(t_i \wedge t_j) = \sum_{k \le i \wedge j} \Delta\Theta_k$ and exchange the sums: $\sum_{i,j} w_i w_j \sum_k \Delta\Theta_k \mathbf{1}_{\{k \le i\}} \mathbf{1}_{\{k \le j\}}$ equals $\sum_k \Delta\Theta_k W_k^2$. (ii) If $v_G = 0$, $G = F_G$; if $K \le 0$, $(G-K)^+ = G - K$. Otherwise $\{G > K\} = \{\log G > \log K\}$ has probability $N(d_-)$, and (15.2) with $U = V = \log G$ gives $\mathbb{E}[G \mathbf{1}_{\{G>K\}}] = F_G\, \mathbb{P}(\log G + v_G > \log K) = F_G N(d_+)$. For the delta, F_G is proportional to S and D, v_G do not depend on S; the derivative of $F \mapsto FN(d_+) - KN(d_-)$ is $N(d_+) + (FN'(d_+) - KN'(d_-))\partial_F d_+ = N(d_+)$, since $FN'(d_+) = KN'(d_-)$. (iii) The mean of the dates is $T(m+1)/(2m)$, and $\sum_{i,j=1}^m \min(i,j) = \sum_{k=1}^m (m-k+1)^2 = m(m+1)(2m+1)/6$. □

So the geometric call is the Black–Scholes call with maturity T, rate $r_{0,T}$, volatility $\sigma_G = \sqrt{v_G/T}$ and the yield y_G defined by $Se^{(r_{0,T} - y_G)T} = F_G$; for $m = 1$ it is the Black–Scholes call itself. As $m \to \infty$ in (iii), $v_G \to \sigma^2 T/3$: the continuous geometric average of the IEOR 222 notes. The course's rule of thumb, that an arithmetic Asian option behaves like an option with volatility $\sigma/\sqrt{3}$, transfers this limit to the arithmetic average; it is a heuristic, with no error bound behind it.

Example 15.3 (The course's monthly Asian)**.** In the test case with $t_j = j/12$, (iii) gives $\sigma_G = 0.2\sqrt{13 \cdot 25/864} = 12.27\%$ and $C^G = 4.40286$, the course's value. The delta is 0.486882 and the vega, the derivative under a parallel shift $\sigma(\cdot) + \varepsilon$ of the volatility curve (so that $\partial_\varepsilon \Theta(t) = 2\int_0^t \sigma$), is 22.4010: the closed forms against which Chapter 13 checks its pathwise estimators. Term structures enter only through F_j and $\Theta(t_j)$. With the curves of Theorem 7.3 ($y = 0$, $r(t) = 0.02t$, $\sigma = 20\%$ on $[0, \frac{1}{2})$ and 10% after) and $S = K = 100$, $C^G = 4.67732$, and the arithmetic call is worth 4.78053 (Section 15.6). A flat volatility equal to the one-year term volatility, 15.81%, with the same rates, gives 4.02796, 16% too little: by (i) the variance of month k enters with the weight W_k^2, which is largest for the early months, where this curve is most volatile.

15.3 Monte Carlo with the geometric control

Prices at the fixing dates are sampled exactly by Theorem 7.2, with one normal per interval between dates (none for a fixing at 0), so the only error is statistical. Three payoffs with known means serve as controls: the geometric call $X^{\mathrm{g}} = D(G-K)^+$ (Theorem 15.2), the vanilla call $X^{\mathrm{v}} = D(S_{t_m} - K)^+$ and the strip $X^{\mathrm{s}} = D\sum_j w_j (S_{t_j} - K)^+$, whose means are Black–Scholes prices with term parameters. Kemna and Vorst (1990) introduced the geometric control.

Proposition 15.4 (The geometric control variate)**.** *Let* $K > 0$, $v_G > 0$, $Y = D(A-K)^+$, *and* $c_j = \sum_i w_i C_{ij} = \mathrm{Cov}(\log S_{t_j}, \log G)$.

(i) $0 \le Y - X^{\mathrm{g}} \le D(A-G)$ *on every path, and*

$$\operatorname{Var}(Y - X^{\mathrm{g}}) \le D^2 \mathbb{E}[(A-G)^2] = D^2 \Big(\sum_{i,j} w_i w_j F_i F_j e^{C_{ij}} - 2F_G \sum_j w_j F_j e^{c_j} + F_G^2 e^{v_G} \Big).$$

(ii) With $\rho = \operatorname{Corr}(Y, X^{\mathrm{g}})$ *and the optimal coefficient* $\beta^* = \operatorname{Cov}(Y, X^{\mathrm{g}})/\operatorname{Var} X^{\mathrm{g}}$*, the control-variate estimator has variance* $(1-\rho^2)\operatorname{Var} Y/n$ *(Theorem 8.12), and* $1 - \rho^2 \le D^2 \mathbb{E}[(A-G)^2]/\operatorname{Var} Y$*.*

(iii) If $(X^{\mathrm{g}}, X^{\mathrm{v}}, X^{\mathrm{s}})$ *has an invertible covariance matrix, the optimal combination of the three controls has variance factor* $1 - R^2 \le 1 - \rho^2$*, and the regression estimator, with estimated coefficients, attains it as* $n \to \infty$ *(Theorem 8.15).*

Proof. (i) The pathwise inequality is Theorem 15.1(ii), and $\operatorname{Var}(Y - X^{\mathrm{g}}) \le \mathbb{E}[(Y - X^{\mathrm{g}})^2]$. For the moments, $\mathbb{E}[S_{t_i} S_{t_j}] = F_i F_j e^{C_{ij}}$ and $\mathbb{E}[G^2] = F_G^2 e^{v_G}$ by (15.1); and $\log S_{t_j} + \log G$ is normal with variance $C_{jj} + 2c_j + v_G$, so $\mathbb{E}[S_{t_j} G] = F_j F_G e^{c_j}$. (ii) $\operatorname{Var} X^{\mathrm{g}} > 0$ because G is lognormal with $v_G > 0$, so it charges both sides of K. By Theorem 8.12(ii) the optimal variance is the minimum over β of $\operatorname{Var}(Y - \beta X^{\mathrm{g}})$, which is at most its value at $\beta = 1$; combine with (i). (iii) is Theorem 8.15: parts (i)–(ii) for the optimal combination and part (iii) for the regression estimator. □

The bound (i) does not depend on the strike: it measures how far apart the two averages are, which depends only on how dispersed the fixings are, and in the test case it keeps the residual variance below 0.197 at every strike. Relative to $\operatorname{Var} Y$, which vanishes out of the money, the gain degrades: the ratio $D^2\mathbb{E}[(A-G)^2]/\operatorname{Var} Y$ is 3.5×10^{-3}, 0.042, 1.26 and 59 at $K = 100$, 120, 140 and 160 (vacuous for the last two), while $1 - \rho^2$ is 6.9×10^{-4}, 4.3×10^{-3}, 0.027 and 0.097. It also degrades as volatility or maturity grows (Figure 15.1). The hypothesis in (iii) fails for $m = 1$, where $X^{\mathrm{g}} = X^{\mathrm{v}} = X^{\mathrm{s}} = Y$, and for $m = 2$ with $t_1 = 0$, where $X^{\mathrm{s}} = \frac{D}{2}(S - K)^+ + \frac{1}{2} X^{\mathrm{v}}$.

Example 15.5 (The course's control-variate experiment)**.** In the test case with $m = 12$ and $K = 100$, $\operatorname{Var} Y = 56.040$, $\rho = 0.999656$ and $\beta^* = 1.02940$, to the digits shown (Theorem 15.7); the course estimated 0.999655 and 1.02941 from 10^8 paths. Before any simulation, (i) gives $D^2\mathbb{E}[(A-G)^2] = 0.1971$, so $1 - \rho^2 \le 3.5 \times 10^{-3}$; the actual value is 6.88×10^{-4}, and $\operatorname{Var}(Y - X^{\mathrm{g}})$ is 0.0842 at $\beta = 1$ and 0.0386 at β^*. The standard error shrinks by $\sqrt{1-\rho^2} = 0.0262$. Our implementation, with 2^{20} paths (`default_rng`, seed 2026), gives 4.5567 ± 0.0073 plain, 4.56146 ± 0.00019 with the geometric control and an in-sample coefficient, and 4.56137 ± 0.00014 with the three controls, whose $R^2 = 0.99962$ and coefficients $(0.9286, 0.0107, 0.0856)$ agree with the course's 0.999616 and $(0.928489, 0.010793, 0.085712)$. The two extra controls remove only 44% of the variance left by the first.

The geometric average also tells us how to integrate part of the path out exactly: everything but its own direction.

Proposition 15.6 (Integrating out the geometric direction)**.** *Let* $v_G > 0$ *and* $\xi = (\log G - \mu_G)/\sqrt{v_G}$*. There are a standard normal vector* $Z' \in \mathbb{R}^{m-1}$ *independent of* ξ *and an* $m \times (m-1)$ *matrix* M *with* $\log S_{t_j} = \mathbb{E} \log S_{t_j} + b_j \xi + (MZ')_j$*, where* $b_j = c_j/\sqrt{v_G} \ge 0$*. Given* Z'*,*

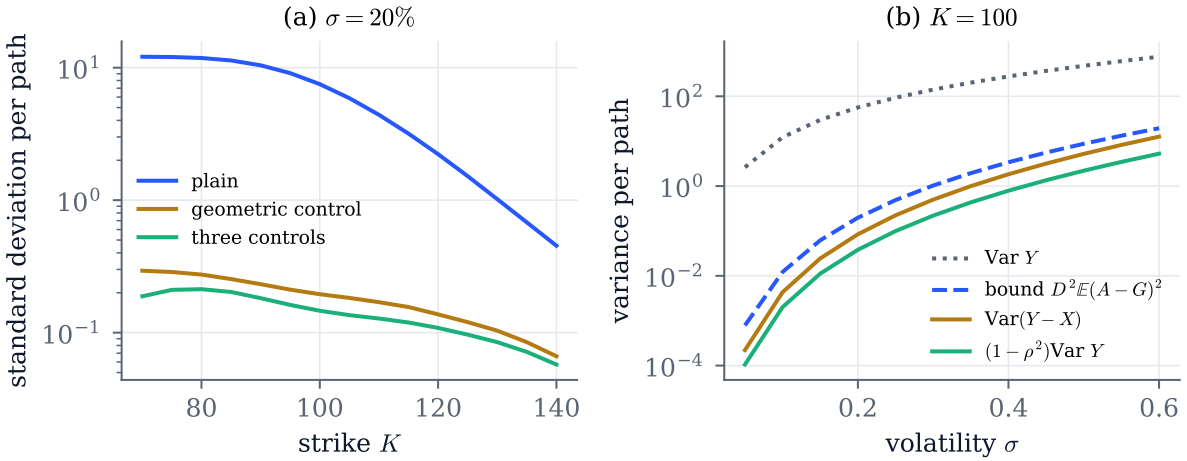

Figure 15.1: Control variates for the monthly arithmetic Asian call (test case, $m = 12$). (a) Standard deviation per path against the strike: plain, with the geometric control and with the three controls (in-sample coefficients). (b) At $K = 100$, against σ: the variance of the payoff Y, the bound $D^2\mathbb{E}(A-G)^2$ of Theorem 15.4(i), and the variances of $Y - X^{\mathrm{g}}$ and of the optimal residual. Simulated quantities use the same 2^{16} paths (seed 15) at every point.

$A = \sum_j a_j e^{b_j \xi}$ with $a_j = w_j \exp\left(\mathbb{E}\log S_{t_j} + (MZ')_j\right)$, and

$$\mathbb{E}\left[(A-K)^+ \mid Z'\right] = \sum_j a_j e^{b_j^2/2} N(b_j - \xi^*) - K\,N(-\xi^*),$$

where ξ^* solves $\sum_j a_j e^{b_j \xi^*} = K$ ($\xi^* = -\infty$ if $\sum_{\{j:b_j=0\}} a_j \geq K$). This conditional value is an unbiased estimator of C/D with no larger variance than $(A-K)^+$ (Theorem 10.11).

Proof. By (15.1) and Theorem 7.1, $\log S_{t_j} - \mathbb{E}\log S_{t_j} = (BZ)_j$ with $Z \sim \mathcal{N}(0, I_m)$ and $B_{jk} = \sqrt{\Delta\Theta_k}\mathbf{1}_{\{k\leq j\}}$, so that $BB^\top = (C_{ij})$. Then $\log G - \mu_G = u \cdot Z$ with $u = B^\top w$ and $|u|^2 = w^\top BB^\top w = v_G$. Let $e = u/|u|$ and let the columns of Q be an orthonormal basis of $e^\perp$. The orthogonal map $Z \mapsto (e \cdot Z, Q^\top Z)$ makes $\xi = e \cdot Z$ and $Z' = Q^\top Z$ independent standard normals, and $Z = e\xi + QZ'$; so $M = BQ$ and $b = Be = BB^\top w/\sqrt{v_G} = c/\sqrt{v_G}$, with $c_j \geq 0$ because $C_{ij} \geq 0$. Given Z', A is a nondecreasing continuous function of ξ, so $(A-K)^+ = (A-K)\mathbf{1}_{\{\xi>\xi^*\}}$, and $\mathbb{E}[e^{b\xi}\mathbf{1}_{\{\xi>x\}}] = e^{b^2/2}N(b-x)$ by (15.2). □

In the test case the conditional value has standard deviation 0.1082 per draw (from its second moment, by randomised QMC), against 0.196 for the optimal geometric control and 7.49 for the payoff. It is also a smooth function of Z' (by the implicit function theorem, ξ^* is smooth where it is finite), which is what randomised QMC needs (Theorem 11.11); Section 15.6 uses it. This is conditioning in the opposite direction to Curran's (Section 15.5): there we condition on G and must approximate; here we condition on everything but G and integrate exactly.

Remark 15.7 (Exact correlations). Given Z', $(A-K)^{+2} = (A-K)^2\mathbf{1}_{\{\xi>\xi^*\}}$ and, since $G \leq A$, $(A-K)^+(G-K)^+ = (A-K)(G-K)\mathbf{1}_{\{\xi>\xi_G\}}$ with $\xi_G = (\log K - \mu_G)/\sqrt{v_G} \geq \xi^*$. Both are sums of terms $e^{\beta\xi}$ restricted to a half-line, so their conditional expectations are closed forms like

the one in Theorem 15.6. Averaging them by randomised QMC gives $\operatorname{Var} Y$, $\operatorname{Cov}(Y, X^{\mathrm{g}})$, ρ and β^* to six or more digits, as quoted in Theorem 15.5.

15.4 Moment matching

The oldest approximations replace the law of A by a tractable law with the same first moments, which are known in closed form.

Proposition 15.8 (Moment matching)**.** *Let* $\operatorname{Var} A > 0$.

(i) (Moments.) $\mathbb{E}[A^p] = \sum_{j_1,\dots,j_p} \prod_a w_{j_a} F_{j_a} \exp\big(\sum_{a<b} C_{j_a j_b}\big)$. *Equivalently, let* $R_k = S_{t_k}/S_{t_{k-1}}$ *with* $S_{t_0} = S$, *and* $\Gamma_m = w_m$, $\Gamma_k = w_k + R_{k+1}\Gamma_{k+1}$. *Then* $A = SR_1\Gamma_1$, R_{k+1} *is independent of* Γ_{k+1}, *and*

$$\mathbb{E}[\Gamma_k^p] = \sum_{q=0}^{p} \binom{p}{q} w_k^{p-q}\, \mathbb{E}[R_{k+1}^q]\, \mathbb{E}[\Gamma_{k+1}^q], \qquad \mathbb{E}[R_k^q] = (F_k/F_{k-1})^q\, e^{q(q-1)\Delta\Theta_k/2},$$

with $F_0 = S$: *all moments up to order* p *cost* $O(mp^2)$ *operations.*

(ii) (Two moments.) *The lognormal law with the mean and variance of* A *has log-variance* $v_2 = \log\big(\mathbb{E}[A^2]/(\mathbb{E}A)^2\big)$. *Replacing* A *by it gives* $C_2 = D\big(\mathbb{E}A\, N(d_+) - K\, N(d_-)\big)$, $d_\pm$ *as in Theorem 15.2(ii) with* $(\mathbb{E}A, v_2)$ *in place of* (F_G, v_G).

(iii) (Three moments.) *The third central moment* μ_3 *of* A *is positive, and the equation* $\zeta^3 + 3\zeta = \gamma := \mu_3/(\operatorname{Var} A)^{3/2}$ *has exactly one real root,* $\zeta > 0$. *With* $b^2 = \log(1+\zeta^2)$, $\lambda = \sqrt{\operatorname{Var} A}/\zeta$, $\tau = \mathbb{E}A - \lambda$ *and* $J = \lambda e^{b\epsilon - b^2/2}$, $\epsilon \sim \mathcal{N}(0,1)$, *the shifted lognormal variable* $\tau + J$ *has the first three moments of* A. *Replacing* A *by it gives* $C_3 = D\big(\lambda N(d_+) - (K-\tau)N(d_-)\big)$, *with* $d_\pm$ *computed from* $(\lambda, K-\tau, b^2)$, *if* $K > \tau$, *and* $C_3 = D(\mathbb{E}A - K)$ *otherwise.*

Proof. (i) Expand the product. For fixed $j_1, \dots, j_p$, $\sum_a \log S_{t_{j_a}}$ is normal with variance $\sum_a C_{j_a j_a} + 2\sum_{a<b} C_{j_a j_b}$, so $\mathbb{E}\prod_a S_{t_{j_a}} = \prod_a F_{j_a} \exp(\sum_{a<b} C_{j_a j_b})$. For the recursion, $A = S(w_1 R_1 + w_2 R_1 R_2 + \dots) = SR_1\Gamma_1$ by induction. Γ_{k+1} is a function of $R_{k+2}, \dots, R_m$, which are independent of R_{k+1} because the $\log R_k$ are increments over disjoint intervals (Theorem 7.1); expand $(w_k + R_{k+1}\Gamma_{k+1})^p$ binomially. Finally $\log R_k \sim \mathcal{N}(\log(F_k/F_{k-1}) - \frac{1}{2}\Delta\Theta_k, \Delta\Theta_k)$. (ii) A lognormal variable with log-variance v has second moment e^v times its squared mean; the price is Theorem 15.2(ii) for that law. (iii) For lognormal S_i, S_j, S_k with $\alpha = e^{C_{ij}}$, $\beta = e^{C_{ik}}$, $\gamma' = e^{C_{jk}}$, all ≥ 1, the joint third cumulant is

$$\mathbb{E}[S_i S_j S_k] - \sum \mathbb{E}[S_i S_j]\mathbb{E}[S_k] + 2\mathbb{E}S_i \mathbb{E}S_j \mathbb{E}S_k = F_i F_j F_k(\alpha\beta\gamma' - \alpha - \beta - \gamma' + 2) \geq 0,$$

because $\alpha\beta\gamma' - \alpha - \beta - \gamma' + 2 = (\alpha-1)(\beta-1)(\gamma'-1) + (\alpha-1)(\beta-1) + (\alpha-1)(\gamma'-1) + (\beta-1)(\gamma'-1)$. Hence $\mu_3 = \sum_{i,j,k} w_i w_j w_k \kappa_3(S_i, S_j, S_k)$ is at least its diagonal part, which is positive when some $\Theta(t_i) > 0$, that is, when $\operatorname{Var} A > 0$. With $u = e^{b^2}$, $\operatorname{Var} J = \lambda^2(u-1)$ and $\mathbb{E}(J-\lambda)^3 = \lambda^3(u-1)^2(u+2)$, so the skewness of J is $\zeta(\zeta^2+3)$ with $\zeta = \sqrt{u-1}$. The map $\zeta \mapsto \zeta^3 + 3\zeta$ has a positive derivative and increases from 0 to ∞ on $[0,\infty)$, so $\zeta^3 + 3\zeta = \gamma$ has exactly one real root; with λ chosen to match the variance and τ the mean, the shift leaves central moments unchanged. The price is (ii) with the strike $K - \tau$; if $K \leq \tau$ the payoff is $\tau + J - K > 0$. □

For $m = 1$, A is lognormal, so $C_2 = C$, and $C_3 = C$ because γ is then the lognormal skewness and $\tau = 0$. Neither is a bound: in the test case $C_2 - C$ and $C_3 - C$ each take both signs as K varies (Table 15.1).

The two-moment approximation (ii) is Levy (1992)'s. Turnbull and Wakeman (1991) compute the moments by the backward recursion of (i), and use the lognormal of (ii) as the leading term of an Edgeworth series. The root of $\zeta^3 + 3\zeta = \gamma$ is Cardano's $\zeta = \sqrt[3]{\gamma/2 + \sqrt{\gamma^2/4 + 1}} + \sqrt[3]{\gamma/2 - \sqrt{\gamma^2/4 + 1}}$. The next terms of the series, in the third and fourth cumulants κ_3, κ_4 of A, take the form of Jarrow and Rudd (1982): two integrations by parts against the corrected density $\ell - \frac{\kappa_3 - \kappa_3^\ell}{3!}\ell''' + \frac{\kappa_4 - \kappa_4^\ell}{4!}\ell''''$ give

$$C_{\mathrm{TW}} = C_2 + D\Big(- \frac{\kappa_3 - \kappa_3^\ell}{3!}\,\ell'(K) + \frac{\kappa_4 - \kappa_4^\ell}{4!}\,\ell''(K)\Big), \tag{15.3}$$

where ℓ is the matched lognormal density and κ_n^ℓ its cumulants. The corrected function integrates to one and has the first four moments of A, but it need not be a density: in the test case it stays above $0.81\,\ell$, while at $\sigma = 40\%$ it is negative near $x = 52$ (down to -1.3×10^{-3}). The correction reduces the error of C_2 near the money, but it is worse than C_2 at low strikes ($K \le 72.5$ in Figure 15.2(a)), and at $\sigma = 40\%$ at most strikes (-0.238 against $+0.141$ at $K = 80$); in Table 15.1 it beats C_3 in only three of the six columns.

15.5 Conditioning on the geometric average

Curran (1994) split the call according to the value of G. Because $A \ge G$, the part where $G > K$ is a linear payoff, priced in closed form; only the part where $G \le K$ needs the conditional law of A given G, and that is where all approximations live.

Lemma 15.9 (The conditional law given G)**.** *Let $v_G > 0$ and $\Lambda = \log G$. Given $\Lambda = x$, the vector (15.1) is Gaussian with means $\mathbb{E}\log S_{t_j} + c_j(x - \mu_G)/v_G$ and covariances $C_{ij} - c_i c_j / v_G$. Hence*

$$h(x) := \mathbb{E}[A \mid \Lambda = x] = \sum_j h_j(x), \qquad h_j(x) = w_j F_j\, e^{c_j(x-\mu_G)/v_G - c_j^2/(2v_G)},$$
$$\varsigma^2(x) := \mathrm{Var}(A \mid \Lambda = x) = \sum_{i,j} h_i(x) h_j(x)\,(e^{C_{ij} - c_i c_j/v_G} - 1),$$

and h is continuous and strictly increasing, with $h(+\infty) = \infty$ and $h(-\infty) = \sum_{\{j : c_j = 0\}} w_j F_j$.

Proof. For a Gaussian vector (U, V) with $\mathrm{Var}\, V > 0$, the conditional law of U given $V = x$ is normal with mean $\mathbb{E}U + \mathrm{Cov}(U,V)(x - \mathbb{E}V)/\mathrm{Var}\, V$ and variance $\mathrm{Var}\, U - \mathrm{Cov}(U,V)^2/\mathrm{Var}\, V$ (regress U on V; the residual is uncorrelated with V, hence independent of it). Apply this to the linear combinations of the $\log S_{t_j}$, with $\mathrm{Cov}(\log S_{t_j}, \Lambda) = c_j$, and take lognormal means of S_{t_j} and of $S_{t_i} S_{t_j}$. Each h_j is nondecreasing since $c_j \ge 0$, and the last one is strictly increasing and unbounded because its coefficient is positive: c_j is nondecreasing in j and $\sum_j w_j c_j = v_G > 0$. □

Proposition 15.10 (Conditioning on the geometric average)**.** *Let $K > 0$, $v_G > 0$, $\Lambda = \log G$, and $L(x) = D\,\mathbb{E}\big[(A - K)\mathbf{1}_{\{\Lambda > x\}}\big]$ for $x \in [-\infty, \infty)$.*

(i) (Curran's decomposition.) $C = L(\log K) + D\,\mathbb{E}\big[(A-K)^+\mathbf{1}_{\{G\le K\}}\big]$, *and for finite* x

$$L(x) = D\Big(\sum_j w_j F_j\, N\Big(\frac{\mu_G + c_j - x}{\sqrt{v_G}}\Big) - K\, N\Big(\frac{\mu_G - x}{\sqrt{v_G}}\Big)\Big),$$

and $L(-\infty) = D\big(\sum_j w_j F_j - K\big)$.

(ii) (Lower bounds.) $L(x) \le C$ *for every* x. *Let* x^* *be the root of* $h(x) = K$ *if* $h(-\infty) < K$, *and* $x^* = -\infty$ *otherwise. Then, for every* x,

$$L(x) \le L(x^*) = D\,\mathbb{E}\big[(\mathbb{E}[A \mid G] - K)^+\big] \le C.$$

(iii) (Error bound.) $0 \le C - L(x^*) \le \frac{D}{2}\,\mathbb{E}\big[\varsigma(\Lambda)\mathbf{1}_{\{G<K\}}\big]$.

Proof. (i) On $\{G > K\}$, $A \ge G > K$, so $(A-K)^+ = A-K$ there. For the closed form, (15.2) with $U = \log S_{t_j}$ and $V = \Lambda$ gives $\mathbb{E}[S_{t_j}\mathbf{1}_{\{\Lambda>x\}}] = F_j\mathbb{P}(\Lambda + c_j > x)$; and $\mathbb{P}(\Lambda > x) = N((\mu_G - x)/\sqrt{v_G})$.

(ii) $(A-K)\mathbf{1}_{\{\Lambda>x\}} \le (A-K)^+$ gives $L(x) \le C$. By the tower property, $L(x) = D\,\mathbb{E}[(h(\Lambda) - K)\mathbf{1}_{\{\Lambda>x\}}]$. By Theorem 15.9, $h(\Lambda) - K < 0$ on $\{\Lambda < x^*\}$ and $h(\Lambda) - K \ge 0$ on $\{\Lambda \ge x^*\}$, so $(h(\Lambda)-K)\mathbf{1}_{\{\Lambda>x\}} \le (h(\Lambda)-K)^+$, with equality almost surely when $x = x^*$, because $\mathbb{P}(\Lambda = x^*) = 0$. Hence $L(x) \le D\,\mathbb{E}[(h(\Lambda) - K)^+] = L(x^*)$. Finally $(h(\Lambda) - K)^+ \le \mathbb{E}[(A-K)^+ \mid \Lambda]$ by the conditional Jensen inequality.

(iii) Let $H = \mathbb{E}[(A-K)^+ \mid \Lambda] - (h(\Lambda) - K)^+ \ge 0$, so that $C - L(x^*) = D\,\mathbb{E}H$. For any integrable U, $u^+ = \frac{1}{2}(u + |u|)$ gives $\mathbb{E}U^+ - (\mathbb{E}U)^+ = \frac{1}{2}(\mathbb{E}|U| - |\mathbb{E}U|) \le \frac{1}{2}\mathbb{E}|U - \mathbb{E}U| \le \frac{1}{2}\sqrt{\operatorname{Var} U}$. Applied to the conditional law of $U = A - K$ given $\Lambda = x$ (explicit by Theorem 15.9), this gives $H \le \frac{1}{2}\varsigma(\Lambda)$. On $\{G \ge K\}$, $A - K \ge 0$, so $H = 0$ there. □

The second term in (i) depends on the conditional law of A given G, which is neither lognormal nor known in closed form; a formula that replaces it is an approximation, unless it is proved to be a bound, as in (ii).

The first term $L(\log K)$ is the part of Curran's decomposition that is exact; his published formula uses a cheap threshold, examined in Exercise 15.4. $L(x^*)$, obtained by moving the positive part outside the conditional expectation, is what the course calls Curran's *naive* approximation; it is in fact a lower bound, the largest of the family $L(x)$. Rogers and Shi (1995) derived the analogous lower bound for continuous averaging, conditioning on the integral of the Brownian path, together with an error bound of the type (iii). The bound (iii) is a guarantee, not a measurement: in the test case it allows 0.060, while the actual gap is 4.5×10^{-4}.

The course's *refined* version keeps $L(\log K)$ exact and approximates the second term by assuming that, given $G = g < K$, $A - G$ is lognormal with its conditional mean $h(\log g) - g$ and variance $\varsigma^2(\log g)$ from Theorem 15.9 (the course freezes both at $g = K$). The conditional term is then a Black formula with strike $K - g$, integrated against the normal density of $\log G$ by quadrature. It is not a bound; it is remarkably accurate (Figure 15.2).

15.6 Accuracy and a reference price

To judge the approximations we need prices accurate to much better than 10^{-6}. The recursion of Theorem 15.8(i) provides them. Since R_{k+1} is independent of Γ_{k+1},

$$\mathbb{P}(\log \Gamma_k \le u) = \mathbb{E}\big[\mathbb{P}\big(\log \Gamma_{k+1} \le \log(e^u - w_k) - \log R_{k+1} \,\big|\, R_{k+1}\big)\big], \qquad u > \log w_k, \tag{15.4}$$

a one-dimensional Gaussian integral for each u, starting from $\Gamma_m = w_m$. At the end, $C = D\,\mathbb{E}[c(\Gamma_1)]$, where $c(z) = \mathbb{E}[(SR_1 z - K)^+]$ is a Black formula (or, if $t_1 = 0$, $c(z) = (Sz - K)^+$). We store each distribution function on a grid of 4001 points with cubic interpolation and use a 64-point Gauss–Hermite rule for (15.4); the recursion is exact, so the only errors are those of the grid and the quadrature, and a grid four times finer with a rule twice as long changes the test-case price by 2×10^{-11}. Recursions of this type, with fast Fourier transforms for the convolutions, go back to Carverhill and Clewlow (1990); our implementation reproduces the benchmark prices of Černý and Kyriakou (2011), Table 4 (thirteen fixings including $t = 0$), to all seven decimals.

Example 15.11 (A reference value for the monthly Asian)**.** In the test case with $m = 12$, the recursion gives $C = 4.5612893$. Independently, Theorem 15.6 with 2^{14} scrambled Sobol' points and 32 independent scrambles (`SeedSequence(99)`) gives 4.5612897 ± 0.0000026. The estimates in these notes and in the course agree with it within about two standard errors: the course's 10^8 plain paths, 4.56058 ± 0.00075, are 0.9 standard errors low; its control-variate estimates 4.56133 and 4.56132 ± 0.000015 are about 2 high; the two runs of Theorem 10.8, 4.56122 and 4.56127, each ± 0.00003, are 2.0 and 0.6 low; and the randomised QMC value 4.56133 ± 0.00011 of Theorem 11.14 is 0.4 high. Deviations of two standard errors are unremarkable among a dozen estimates; the reference settles the fifth decimal.

	$\sigma = 20\%$					$\sigma = 40\%$
K	80	90	100	110	120	100
reference C	19.414851	10.679692	4.561289	1.492200	0.383703	9.354588
two moments C_2	164.6	315.8	67.5	−216.1	−210.0	660.3
three moments C_3	−10.0	−1.5	16.3	8.4	−3.4	109.0
Edgeworth (15.3)	4.8	−98.4	1.9	66.1	17.3	86.3
first term $L(\log K)$	−16.7	−17.2	−11.8	−14.4	−16.1	−94.4
lower bound $L(x^*)$	−4.6	−6.5	−4.5	−5.3	−5.2	−36.0
conditional lognormal	−0.03	0.03	0.01	−0.02	−0.01	0.16

Table 15.1: The monthly Asian call ($m = 12$, $S = 100$, $T = 1$, $r = 1\%$, $y = 2\%$): reference prices from the recursion (15.4), and the errors of the approximations, in units of 10^{-4}.

The ranking in Table 15.1 and Figure 15.2 is the course's. Over the strikes of Figure 15.2, from 60 to 160, two moments miss by up to 0.032, with both signs, and three moments by up to 0.0017; the Edgeworth correction lies in between (up to 0.0098). The first term of Curran's decomposition alone is no better than three moments (0.0019; at $\sigma = 40\%$, 0.0147 against 0.0130); the bound $L(x^*)$, within 6.7×10^{-4}, is about 2.5 times better; and only the conditional lognormal, within 4×10^{-6}, is better by orders of magnitude. All errors grow with $\sigma\sqrt{T}$, as the last column shows: the dispersion of the fixings, which the bound of Theorem 15.4(i) measures, is what every method must account for.

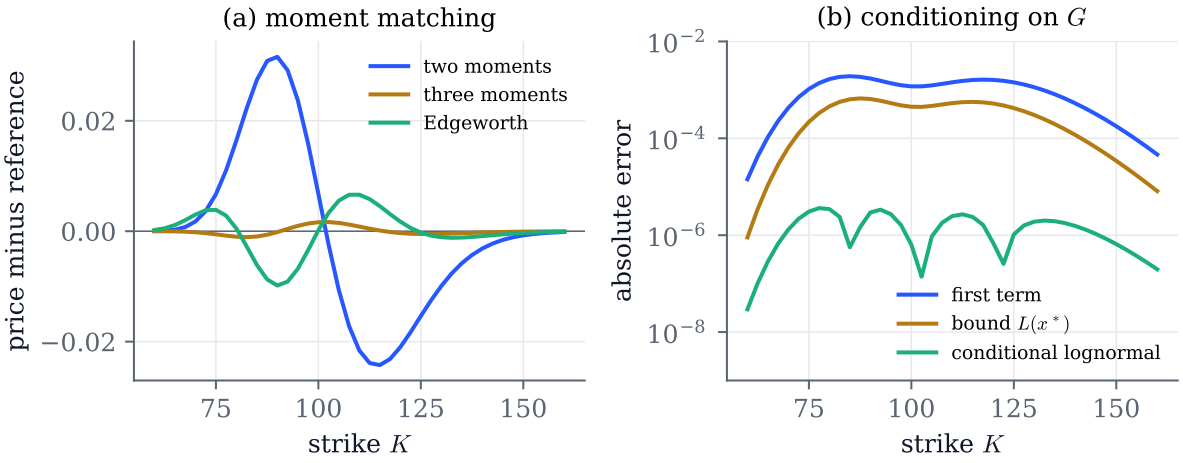

Figure 15.2: Errors of the approximations for the monthly Asian call against the strike (test case, $m = 12$), relative to the recursion (15.4). (a) Price minus reference for the two-moment, three-moment and Edgeworth approximations. (b) Absolute error of the first term $L(\log K)$, of the lower bound $L(x^*)$ and of the conditional lognormal approximation.

▶ Interactive demo: arithmetic, geometric and European calls for any number of fixings and volatility; the control variates of Figure 15.1; and the errors of every approximation against the reference, with sliders for σ, T and m.

In code

`code/src/mfe230d/asian.py`: `fixing_law` returns F_j and C_{ij} for any fixing dates and term structures (`paths.PiecewiseConstant`); `geometric_price` and `geometric_greeks` implement Theorem 15.2; `mc_price` and `cv_laboratory` the control variates, through `variance.control_variate` and `variance.regression_control_variates`; `conditional_samples` and `rqmc_reference` Theorem 15.6; `arithmetic_moments`, `lognormal_price`, `shifted_lognormal_price` and `edgeworth_price` Section 15.4; `threshold_bound`, `conditioning_lower_bound`, `conditional_upper_bound` and `curran_refined` Theorem 15.10; `recursion_price` (15.4); `seasoned_strike` the strike shift.

Course examples reproduced. The geometric Asian call 4.40286; the correlation 0.999655 and coefficient 1.02941 of the geometric control, now exact; the three-control regression ($R^2 = 0.999616$, coefficients 0.928, 0.011, 0.086, standard error 1.5×10^{-5} at 10^8 paths, which our 2^{20}-path run implies); the course's estimates of the arithmetic price, all within about two standard errors of the reference; and the ranking and size of the errors of the two-moment, three-moment and Curran approximations plotted in the course.

Further reading

Glasserman (2003) treats control variates in Section 4.1, and multiple controls, as in Theorem 15.4(iii), in Section 4.1.2. Kemna and Vorst (1990) price the geometric average and propose it as a control; Turnbull and Wakeman (1991) and Levy (1992) give the moment approximations, Curran (1994) the decomposition, and Rogers and Shi (1995) the conditioning bounds for continuous averaging, where they also reduce the price to a parabolic equation in two variables. Černý and Kyriakou (2011) improve the convolution method for discrete averages and give benchmark prices.

Chapter 16

Barriers, lookbacks and the Brownian bridge

A barrier option is a European option that is cancelled, or activated, when the underlying reaches a level; a lookback option pays a function of the highest or lowest price. Both depend on the path through its extremes, and for geometric Brownian motion the joint law of the endpoint and an extreme is known in closed form (the reflection principle). Reflecting a density in the barrier, the *method of images*, prices every continuously monitored barrier and lookback; an infinite sequence of reflections handles two barriers.

Real contracts are monitored at dates. Discrete monitoring changes prices at order $h^{1/2}$ in the spacing h of the dates: for the course's monthly down-and-out call it raises the price by 42%. A shift of the barrier by $e^{\pm 0.5826\,\sigma\sqrt{h}}$ captures the leading term (Broadie, Glasserman and Kou); exact discrete prices follow from a recursion over the dates. In Monte Carlo the same $h^{1/2}$ is the bias of checking a continuous barrier on the time grid, and the Brownian bridge removes it.

Throughout, W is a Brownian motion under the risk-neutral measure $\mathbb{Q}$ with a filtration $(\mathcal{F}_t)$, the parameters r, y and $\sigma > 0$ are constants, and

$$S_t = S_0 e^{X_t}, \qquad X_t = \nu t + \sigma W_t, \qquad \nu = r - y - \tfrac{1}{2}\sigma^2, \qquad \lambda = \frac{2\nu}{\sigma^2}. \tag{16.1}$$

We write $\overline{X}_T = \max_{t\le T} X_t$, $\underline{X}_T = \min_{t\le T} X_t$ (the maximum and minimum exist because the paths are continuous), $\overline{S}_T = S_0 e^{\overline{X}_T}$ and $\underline{S}_T = S_0 e^{\underline{X}_T}$, and $p_T(x) = (\sigma\sqrt{T})^{-1} N'\big((x-\nu T)/(\sigma\sqrt{T})\big)$ for the density of X_T. Numbers refer to the course's test case $S_0 = K = 100$, $T = 1$, $r = 1\%$, $y = 2\%$, $\sigma = 20\%$, for which $\nu = -0.03$ and $\lambda = -1.5$.

16.1 The contracts

Fix a payoff $g(S_T)$, here a call $(S_T - K)^+$ or a put $(K - S_T)^+$, a barrier $B > 0$ and a *monitoring set* $\mathcal{T} \subseteq [0,T]$: either $[0,T]$ (continuous monitoring) or dates $t_1 < \cdots < t_m$, by default $t_j = jh$ with $h = T/m$; as in the course, time 0 is not a monitoring date. A *down* barrier, normally below the spot, is *hit* if $S_t \le B$ for some $t \in \mathcal{T}$; an *up* barrier, normally above it, is hit if $S_t \ge B$ for some $t \in \mathcal{T}$. The *out* option pays $g(S_T)$ if the barrier is not hit, the *in* option pays $g(S_T)$ if it is.

Two directions, in or out, call or put: these are the eight standard barrier options of Rubinstein and Reiner (1991), who priced them all. A *rebate* R may be added: we pay it at T to the holder of an out option that has been knocked out, or of an in option that has never been knocked in (paying it at the hitting time is Exercise 16.4). Since $\mathbf{1}_{\{\text{hit}\}} + \mathbf{1}_{\{\text{not hit}\}} = 1$ path by path,

$$V^{\text{in}} + V^{\text{out}} = e^{-rT}\mathbb{E}^{\mathbb{Q}}[g(S_T)] + Re^{-rT} \tag{16.2}$$

in any model, provided the two options have the same barrier, the same monitoring set and the same rebate; with different monitoring sets the identity fails (Exercise 16.1).

Some cases are trivial. If the spot is on the far side of the barrier at inception, a continuously monitored out option is knocked out at once, and so it is when $S_0 = B$ (Theorem 16.1(iii)); with dates $t_1 > 0$ it is not: with $S_0 = B = 100$ and monthly dates the down-and-out call is worth 3.23135 (Section 16.6). If $K \geq B$, the up-and-out call pays nothing, and so does the down-and-out put if $K \leq B$.

Lookbacks use the extremes over $\mathcal{T} \cup \{0\}$: the floating-strike put pays $\overline{S}_T - S_T$ and the call $S_T - \underline{S}_T$; the fixed-strike call pays $(\overline{S}_T - K)^+$ and the put $(K - \underline{S}_T)^+$.

16.2 The reflection principle with drift

Theorem 16.1 (Reflection principle with drift)**.** *Let $T > 0$ and use the notation (16.1).*

(i) For every stopping time τ of the filtration generated by W, the process $W^\tau_t = 2W_{t\wedge\tau} - W_t$, equal to W up to τ and reflected about W_τ afterwards, is a Brownian motion.

(ii) For a level $b \neq 0$ and a Borel set A on the same side of b as 0,

$$\begin{aligned} \mathbb{Q}(\overline{X}_T \geq b,\ X_T \in A) &= e^{\lambda b}\int_A p_T(x-2b)\,\mathrm{d}x \quad \text{if } b>0,\ A \subseteq (-\infty, b), \\ \mathbb{Q}(\underline{X}_T \leq b,\ X_T \in A) &= e^{\lambda b}\int_A p_T(x-2b)\,\mathrm{d}x \quad \text{if } b<0,\ A \subseteq (b, \infty). \end{aligned} \tag{16.3}$$

(iii) With $s = \sigma\sqrt{T}$: for $b \geq 0$, $\mathbb{Q}(\overline{X}_T \geq b) = N(\frac{\nu T - b}{s}) + e^{\lambda b}N(\frac{-\nu T - b}{s})$, and for $b \leq 0$, $\mathbb{Q}(\underline{X}_T \leq b) = N(\frac{b - \nu T}{s}) + e^{\lambda b}N(\frac{b + \nu T}{s})$. These functions of b are continuous, so the extremes have no atoms on either side of 0, and $\mathbb{Q}(\overline{X}_T > 0) = \mathbb{Q}(\underline{X}_T < 0) = 1$.

(iv) For $b < 0$ the knock-out density $q_T(x) = \mathbb{Q}(\underline{X}_T > b,\ X_T \in \mathrm{d}x)/\mathrm{d}x$ *is $p_T(x) - e^{\lambda b}p_T(x-2b)$ for $x > b$ and 0 for $x \leq b$; for $b > 0$ the same with $\overline{X}_T < b$ and $x < b$. It is positive on the surviving side and vanishes at $x = b$.*

Proof. (i) On $[0, T']$, $W^\tau = W^{\tau\wedge T'}$, so we may assume τ bounded. The *strong Markov property* (Mörters and Peres, 2010, Theorem 2.16, for stopping times of the right-continuous filtration generated by W, which include ours; Karatzas and Shreve, 1991, Section 2.6 for general filtrations) says that $\beta_t = W_{\tau+t} - W_\tau$ is a Brownian motion independent of $\mathcal{F}_\tau$; it needs continuous paths and a stopping time, not merely a random time. The stopped path $W_{\cdot\wedge\tau}$ and τ are $\mathcal{F}_\tau$-measurable, and

$$W_t = W_{t\wedge\tau} + \beta_{(t-\tau)^+}, \qquad W^\tau_t = W_{t\wedge\tau} - \beta_{(t-\tau)^+}.$$

Since $-\beta$ is also a Brownian motion independent of $\mathcal{F}_\tau$, the triples $(W_{\cdot\wedge\tau}, \tau, \beta)$ and $(W_{\cdot\wedge\tau}, \tau, -\beta)$ have the same law, and so do W and W^τ, which are the same measurable function of them.

(ii) *No drift, $\sigma = 1$, $b > 0$.* Let $\tau_b = \inf\{t : W_t = b\}$, a stopping time because $\{\tau_b \le t\} = \{\max_{u\le t} W_u \ge b\}$ by continuity. The paths of W and W^{τ_b} agree up to τ_b, so each reaches b by time T exactly when $\tau_b \le T$, and on that event $W_T^{\tau_b} = 2b - W_T$. By (i), the pair (maximum on $[0,T]$, value at T) has the same law for both processes (it is a measurable function of the path), hence for $A \subseteq (-\infty, b)$

$$\mathbb{Q}(\overline{W}_T \ge b,\, W_T \in A) = \mathbb{Q}(\tau_b \le T,\, 2b - W_T \in A) = \mathbb{Q}(2b - W_T \in A),$$

because $2b - W_T \in A$ forces $W_T > b$, hence $\tau_b \le T$. The right side is $\int_A \varphi_T(x - 2b)\,\mathrm{d}x$ with φ_T the $\mathcal{N}(0,T)$ density.

Drift. Let $\mu = \nu/\sigma$, so $X/\sigma = W + \mu t$. For a bounded Borel functional F of the path on $[0,T]$, Theorem 9.8(i), applied to $G(w) = F(w)\, e^{\mu w_T - \mu^2 T/2}$, gives $\mathbb{E}^{\mathbb{Q}}[F(W + \mu\,\cdot)] = \mathbb{E}^{\mathbb{Q}}\big[F(W)\, e^{\mu W_T - \mu^2 T/2}\big]$. With $F = \mathbf{1}\{\max_{[0,T]} w \ge b/\sigma,\ w_T \in A/\sigma\}$ and the driftless case (extended from sets to nonnegative functions of W_T),

$$\mathbb{Q}(\overline{X}_T \ge b, X_T \in A) = \int_{A/\sigma} e^{\mu w - \mu^2 T/2}\, \varphi_T\Big(w - \frac{2b}{\sigma}\Big)\mathrm{d}w = \int_{A/\sigma} e^{2\mu b/\sigma}\, \varphi_T\Big(w - \frac{2b}{\sigma} - \mu T\Big)\mathrm{d}w,$$

by completing the square. Substituting $x = \sigma w$ and using $2\mu b/\sigma = \lambda b$ gives (16.3). For $b < 0$ apply this to $-X$, whose drift is $-\nu$ and whose λ is $-\lambda$, at the level $-b$.

(iii) $\mathbb{Q}(\overline{X}_T \ge b) = \mathbb{Q}(X_T \ge b) + \mathbb{Q}(\overline{X}_T \ge b, X_T < b)$, and by (ii) the second term is $e^{\lambda b}\mathbb{Q}(X_T + 2b < b) = e^{\lambda b}\mathbb{Q}(X_T < -b)$. The formula is continuous in $b > 0$, so $\mathbb{Q}(\overline{X}_T = b) = 0$, and letting $b \downarrow 0$ gives $\mathbb{Q}(\overline{X}_T > 0) = N(\nu T/s) + N(-\nu T/s) = 1$. Likewise for the minimum.

(iv) Subtract (16.3) from the law of X_T. At $x = b$ the exponents of $p_T(b)$ and $e^{\lambda b}p_T(-b)$ differ by $\lambda b - \big[(b + \nu T)^2 - (b - \nu T)^2\big]/(2\sigma^2 T) = \lambda b - 2b\nu/\sigma^2 = 0$. Positivity on the surviving side is Theorem 16.9(i): the ratio of the two terms is a probability $e^{-2b(b-x)/(\sigma^2 T)} < 1$. □

The image $e^{\lambda b}p_T(x - 2b)$ is the density of a mirror particle started at $2b$, weighted so that it cancels the source on the barrier (Figure 16.1, left); Girsanov's theorem supplies the weight. Equivalently, q_T solves the forward Kolmogorov equation $\partial_T q = \frac{1}{2}\sigma^2 q'' - \nu q'$ on the surviving side with $q_T(b) = 0$ (Exercise 16.2), the course's derivation; the probabilistic proof replaces the uniqueness theorem that route needs. By (iii), $\mathbb{Q}(\underline{S}_T \le 95) = 0.82717$ for the test case.

16.3 Closed forms by images

The *surviving side* of a barrier B is $\mathcal{A} = (B, \infty)$ for a down barrier and $\mathcal{A} = (0, B)$ for an up barrier. For a payoff $g \ge 0$ let

$$G(x) = e^{-rT}\mathbb{E}^{\mathbb{Q}}\big[g(xe^{X_T})\, \mathbf{1}_{\mathcal{A}}(xe^{X_T})\big], \tag{16.4}$$

the Black–Scholes price of the European claim $g(S_T)\mathbf{1}_{\{S_T \in \mathcal{A}\}}$ when the spot is x. For calls and puts, G is a combination of vanilla prices $C(x,k)$, $P(x,k)$ and of cash-or-nothing prices $D^c(x,k)$, $D^p(x,k)$ (paying 1 if $S_T > k$, respectively $S_T < k$), all with the parameters T, r, y, σ:

claim	$K \geq B$	$K < B$
call, down barrier	$C(x,K)$	$C(x,B) + (B-K)D^c(x,B)$
call, up barrier	0	$C(x,K) - C(x,B) - (B-K)D^c(x,B)$
put, down barrier	$P(x,K) - P(x,B) - (K-B)D^p(x,B)$	0
put, up barrier	$P(x,B) + (K-B)D^p(x,B)$	$P(x,K)$

(For instance $(s-K)^+\mathbf{1}_{\{s>B\}} = (s-B)^+ + (B-K)\mathbf{1}_{\{s>B\}}$ when $K < B$.)

Proposition 16.2 (Single barriers by images)**.** *Let $g \geq 0$ be Borel with $\mathbb{E}^{\mathbb{Q}}[g(S_T)] < \infty$, monitor continuously, and in (i) and (ii) let $S_0 \in \mathcal{A}$.*

(i) The out option paying $g(S_T)$ costs

$$V^{\text{out}} = G(S_0) - \Big(\frac{B}{S_0}\Big)^{\lambda} G\Big(\frac{B^2}{S_0}\Big), \qquad \lambda = \frac{2(r-y)}{\sigma^2} - 1. \tag{16.5}$$

(ii) The in option costs $e^{-rT}\mathbb{E}^{\mathbb{Q}}[g(S_T)] - V^{\text{out}}$. A rebate R at T adds $Re^{-rT}\pi$ to the out option and $Re^{-rT}(1-\pi)$ to the in option, where $\pi = \mathbb{Q}(\text{barrier hit})$ is given by Theorem 16.1(iii) at $b = \log(B/S_0)$.

(iii) With the table above, (i) and (ii) price the eight standard barrier options. If $S_0 \notin \mathcal{A}$ (including $S_0 = B$), the out option is worth Re^{-rT} and the in option is the vanilla.

Proof. (i) Take a down barrier, $b = \log(B/S_0) < 0$. On survival $X_T > b$, so the payoff is $g(S_T)\mathbf{1}_{\{X_T>b\}} - g(S_T)\mathbf{1}_{\{X_T>b\}}\mathbf{1}_{\{\underline{X}_T \leq b\}}$. The first term has price $G(S_0)$. For the second, Theorem 16.1(ii), extended from sets to nonnegative functions, gives

$$\mathbb{E}^{\mathbb{Q}}\big[g(S_T)\mathbf{1}_{\{X_T>b,\ \underline{X}_T\leq b\}}\big] = e^{\lambda b}\int_b^{\infty} g(S_0e^x)p_T(x-2b)\,\mathrm{d}x = e^{\lambda b}\,\mathbb{E}^{\mathbb{Q}}\big[g(S_0e^{2b}e^{X_T})\mathbf{1}_{\{X_T+2b>b\}}\big],$$

and $S_0e^{2b} = B^2/S_0$, $S_0e^{X_T+2b} > B$ if and only if $(B^2/S_0)e^{X_T} > B$, and $e^{\lambda b} = (B/S_0)^{\lambda}$. So the discounted second term is $(B/S_0)^{\lambda}G(B^2/S_0)$. Up barriers are identical with $X_T < b$. (ii) is (16.2), and the rebate is paid on an event of probability π or $1-\pi$. (iii) The table is checked case by case, as in the example below it; the degenerate case is Theorem 16.1(iii). □

The image term is the same European claim on a fictitious stock started at the mirror point B^2/S_0 (the reflection of S_0 in B on the log scale), weighted by $(B/S_0)^{\lambda}$. For the test case, the down-and-out call with $B = 95$ costs 3.75412 against 7.36429 for the call, and the down-and-in call 3.61017; the up-and-out call with $B = 115$ costs only 0.44998, because its payoff is largest just below the barrier, where survival is least likely. Figure 16.1 (right) compares the knock-out density with simulated survivors. With $K \geq B$ and $y = 0$, (16.5) for the down-and-out call reads $C(S_0,K) - (B/S_0)^{2r/\sigma^2-1}C(B^2/S_0,K)$, the formula of Merton (1973).

Remark 16.3 (What the images need). The proof used constant coefficients twice: the reflected path must be a Brownian motion with the same drift up to the weight $e^{\lambda b}$, and $p_T(x-2b)$ must be a translate of p_T. With time-dependent r, y or σ, a mirror source of constant weight no longer vanishes on the barrier at every time, and the method fails in general. One exception: if $r(t) - y(t) - \frac{1}{2}\sigma(t)^2$ is proportional to $\sigma(t)^2$, the log price is a Brownian motion with constant drift run on the variance clock $\int_0^t \sigma^2$ (Theorem 7.5), and the formulas hold in that clock. Exponential barriers reduce to constant ones (Exercise 16.3). Otherwise one simulates (Section 16.7) or uses grids (Chapter 20).

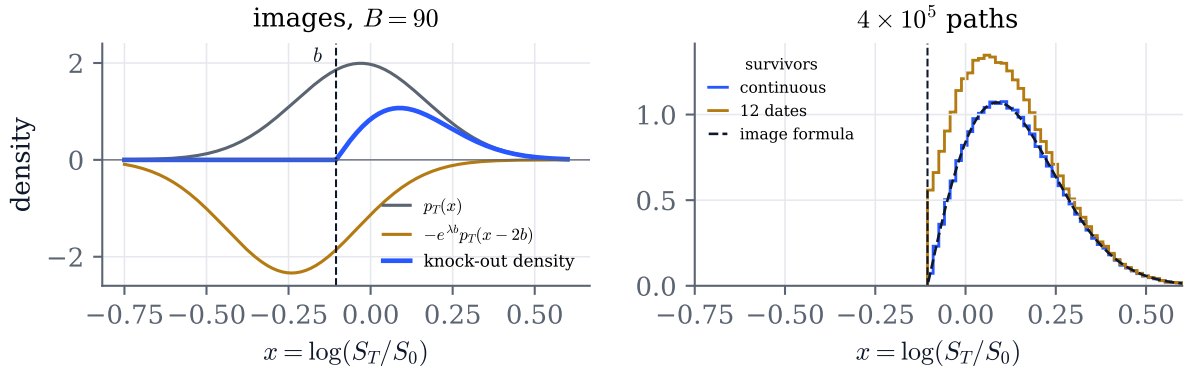

Figure 16.1: Test case, $B = 90$. Left: the density p_T of X_T, the negative image and their sum on $x > b$, the knock-out density (Theorem 16.1(iv)). Right: histograms of X_T over the surviving paths among 4×10^5 (our implementation, numpy `default_rng`, seed 16), with continuous monitoring (exact, by the minimum of the bridge from 0 to X_T, Theorem 16.9) and with 12 monthly dates; dashed, the image formula for the knock-out density.

16.4 Double barriers

A double knock-out option pays $g(S_T)$ if $L < S_t < U$ for all $t \le T$, where $L < S_0 < U$. Let $\ell = \log(L/S_0) < 0 < u = \log(U/S_0)$ and $w = u - \ell = \log(U/L)$. One image cancels the source on one barrier but not on the other; reflecting the image in the other barrier, and so on, produces an infinite series.

Proposition 16.4 (Double barriers). *(i) For $x \in (\ell, u)$,*

$$\frac{\mathbb{Q}(\ell < X_t < u \ \forall t \le T,\ X_T \in \mathrm{d}x)}{\mathrm{d}x} = q(x) = \sum_{k \in \mathbb{Z}} e^{\lambda k w} \Big[p_T(x - 2kw) - e^{\lambda u}\, p_T(x - 2u - 2kw) \Big]. \tag{16.6}$$

(ii) Let q_n be the sum over $|k| \le n$. For $x \in (\ell, u)$ and $n \ge 0$,

$$|q(x) - q_n(x)| \le \frac{4\, e^{|\nu| w/\sigma^2}}{\sigma\sqrt{2\pi T}} \sum_{j \ge n} e^{-2 j^2 w^2/(\sigma^2 T)},$$

so the series converges faster than geometrically, at the rate $e^{-2n^2 w^2/(\sigma^2 T)}$.

(iii) With G the price (16.4) of $g(S_T)\mathbf{1}_{\{L < S_T < U\}}$, the double knock-out costs

$$V = \sum_{k \in \mathbb{Z}} \Big(\frac{U}{L}\Big)^{\lambda k} \Big[G\Big(S_0 \Big(\frac{U}{L}\Big)^{2k}\Big) - \Big(\frac{U}{S_0}\Big)^{\lambda} G\Big(\frac{U^2}{S_0}\Big(\frac{U}{L}\Big)^{2k}\Big) \Big],$$

and truncating at $|k| \le n$ costs at most $e^{-rT} w \sup_{(L,U)} g$ times the bound in (ii).

Proof. (i) Take first $\nu = 0$, $\sigma = 1$, so $X = W$, $a = \ell < 0 < u$, $w = u - a$. Let σ_1 be the first hitting time of u, σ_2 the first hitting time of a after σ_1, σ_3 that of u after σ_2, and so on; define

$\rho_1, \rho_2, \dots$ in the same way starting with a. A continuous path alternates finitely often between a and u on $[0,T]$; if it reaches u first and then $k-1$ more alternating levels, exactly k of the σ_n and, as $\rho_j = \sigma_{j+1}$, $k-1$ of the ρ_n are at most T; so

$$\mathbf{1}\{\text{exit before } T\} = \sum_{n\ge 1} (-1)^{n+1} \big(\mathbf{1}_{\{\sigma_n \le T\}} + \mathbf{1}_{\{\rho_n \le T\}}\big),$$

with partial sums in $\{0,1,2\}$, which allows dominated convergence. Now fix $A \subseteq (a,u)$ and write $c_i = u$ for odd i and $c_i = a$ for even i. Reflecting the path after σ_n about c_n (Theorem 16.1(i); the reflected path has the same $\sigma_1, \dots, \sigma_n$) gives $\mathbb{Q}(\sigma_n \le T, W_T \in A) = \mathbb{Q}(\sigma_n \le T, W_T \in A_1)$ with $A_1 = 2c_n - A$, a set beyond c_n. A path that is at c_{n-1} at time σ_{n-1} and ends beyond c_n must reach c_n in between, so the event $\{\sigma_n \le T\}$ may be replaced by $\{\sigma_{n-1} \le T\}$. Reflecting after σ_{n-1} about c_{n-1} moves the set beyond c_{n-1}, and so on down to σ_1, after which the set lies above u and $\{\sigma_1 \le T\}$ can be dropped. Two consecutive reflections, about a and then u, translate by $2w$, so $\mathbb{Q}(\sigma_{2k} \le T, W_T \in A) = \mathbb{Q}(W_T \in A + 2kw)$ and $\mathbb{Q}(\sigma_{2k+1} \le T, W_T \in A) = \mathbb{Q}(W_T \in 2u + 2kw - A)$, and symmetrically for the ρ's with a in place of u and $-w$ in place of w. Collecting densities, with $2a = 2u - 2w$,

$$q(x) = \sum_{k \in \mathbb{Z}} \big[\varphi_T(x - 2kw) - \varphi_T(x - 2u - 2kw)\big],$$

where the series of Gaussians converges absolutely, which justifies the interchanges. The drift enters through the factor $e^{\mu x - \mu^2 T/2}$ of the proof of Theorem 16.1(ii), which again turns each translate $\varphi_T(x-c)$ into $e^{\mu c}\varphi_T(x - c - \mu T)$; rescaling by σ gives (16.6).

(ii) In the driftless unit-variance coordinates, $|x| < w$ on (a,u). For $|k| = j+1 \ge 1$ both $|x - 2kw|$ and $|x - 2u - 2kw|$ are at least $2jw$ (for instance, for $k = -j-1$, $x - 2u + 2(j+1)w > a - 2u + 2(j+1)w = 2jw + (w-u) + (a - u + w)$, and $w - u = -a > 0$, $a - u + w = 0$), so each of the four terms with $|k| = j+1$ is at most $\varphi_T(2jw)$. The Girsanov factor is at most $e^{|\mu| w}$ on (a,u). Rescaling gives (ii). (iii) Integrate $e^{-rT} g(S_0 e^x)$ against (16.6) term by term as in Theorem 16.2: the translate by $2kw$ is the spot $S_0 (U/L)^{2k}$, the translate by $2u + 2kw$ the spot $(U^2/S_0)(U/L)^{2k}$. □

Kunitomo and Ikeda (1992) derived the series for exponentially curved barriers. For the test case with $L = 80$, $U = 120$, the terms with $|k| \le 2$ give the density to machine precision ($q - q_1$ is 3.6×10^{-6} at most, the bound 2.9×10^{-3}), the probability of staying in the corridor for a year is 0.38057, and the double knock-out call costs 1.00574. The rate $e^{-2n^2 w^2/(\sigma^2 T)}$ deteriorates when $\sigma^2 T$ is large compared with w^2, and then an eigenfunction expansion of the same density is the better series (Exercise 16.5).

16.5 Lookback options

For a lookback the extreme itself is the payoff, and its moments follow from the law of Theorem 16.1(iii). Let $s = \sigma\sqrt{T}$ and, for real κ, c and d,

$$J(\kappa, c, d) = \int_c^\infty e^{\kappa z} N\Big(\frac{d - z}{s}\Big) \mathrm{d}z = \frac{e^{\kappa d + \kappa^2 s^2/2} N\big(\frac{d-c}{s} + \kappa s\big) - e^{\kappa c} N\big(\frac{d-c}{s}\big)}{\kappa}$$

for $\kappa \neq 0$ (integrate by parts), and $J(0, c, d) = s\, N'\big(\frac{d-c}{s}\big) + (d - c) N\big(\frac{d-c}{s}\big)$, the limit $\kappa \to 0$.

Proposition 16.5 (Lookback options). *(i) For $c \geq 0$ and real κ, $\mathbb{E}^{\mathbb{Q}}\big[e^{\kappa \max(c,\overline{X}_T)}\big] = e^{\kappa c} + \kappa\big[J(\kappa,c,\nu T) + J(\kappa+\lambda,c,-\nu T)\big]$, and $\mathbb{E}^{\mathbb{Q}}\big[(e^{\overline{X}_T} - e^c)^+\big] = J(1,c,\nu T) + J(1+\lambda,c,-\nu T)$.*

(ii) The floating-strike lookback put, at inception, costs

$$LP = S_0\Big[e^{-yT}N(d_+)\Big(1 + \frac{\sigma^2}{2(r-y)}\Big) - e^{-yT} + e^{-rT}N(-d_-)\Big(1 - \frac{\sigma^2}{2(r-y)}\Big)\Big],$$

with $d_\pm = (r - y \pm \tfrac{1}{2}\sigma^2)\sqrt{T}/\sigma$, for $r \neq y$, and the limit of this expression as $y \to r$ when $y = r$. When the maximum so far is $S^+ \geq S_0$, the price is $S_0 e^{-rT}\mathbb{E}^{\mathbb{Q}}[e^{\max(c,\overline{X}_T)}] - S_0 e^{-yT}$ with $c = \log(S^+/S_0)$.

(iii) The fixed-strike lookback call costs $S_0 e^{-rT}\mathbb{E}^{\mathbb{Q}}[(e^{\overline{X}_T} - e^c)^+]$ with $c = \log(K/S_0)$ if $K \geq S_0$, and $LP + S_0 e^{-yT} - Ke^{-rT}$ if $K < S_0$.

The lookback call and the fixed-strike put follow from (i) applied to $-X$ (drift $-\nu$), since $\underline{X}_T = -\max_{t\leq T}(-X_t)$.

Proof. (i) For f continuously differentiable and monotone,

$$f\big(\max(c,\overline{X}_T)\big) = f(c) + \int_c^\infty f'(z)\mathbf{1}_{\{\overline{X}_T > z\}}\,\mathrm{d}z.$$

Take expectations with Tonelli's theorem (f' has one sign) and insert $\mathbb{Q}(\overline{X}_T > z) = N(\frac{\nu T - z}{s}) + e^{\lambda z}N(\frac{-\nu T - z}{s})$ (Theorem 16.1(iii)). With $f(z) = e^{\kappa z}$ the two terms give $\kappa J(\kappa,c,\nu T)$ and $\kappa J(\kappa+\lambda,c,-\nu T)$; with $f(z) = (e^z - e^c)^+$ on $z \geq c$ the same computation without the constant. The formula for J is one integration by parts and the identity $\int_\alpha^\infty e^{\kappa s v}N'(v)\,\mathrm{d}v = e^{\kappa^2 s^2/2}N(\kappa s - \alpha)$. (ii) The payoff is $S_0 e^{\max(c,\overline{X}_T)} - S_T$, and $e^{-rT}\mathbb{E}^{\mathbb{Q}}[S_T] = S_0 e^{-yT}$. Take $c = 0$ and $\kappa = 1$. Since $\nu T + \frac{1}{2}s^2 = (r-y)T$, $\nu T/s = d_-$ and $1 + \lambda = 2(r-y)/\sigma^2$, we have $(1+\lambda)(-\nu T + \frac{1}{2}(1+\lambda)s^2) = (r-y)T$ and $-\nu T/s + (1+\lambda)s = d_+$, so

$$J(1,0,\nu T) = e^{(r-y)T}N(d_+) - N(d_-), \qquad J(1+\lambda,0,-\nu T) = \frac{\sigma^2}{2(r-y)}\big[e^{(r-y)T}N(d_+) - N(-d_-)\big],$$

and collecting terms gives the formula. The price is continuous in y by dominated convergence ($\overline{X}_T$ is continuous in ν and $e^{\overline{X}_T} \leq e^{|\nu|T + \sigma\max_t W_t}$, which is integrable by (i) with $\nu = 0$), and $J(\kappa,c,d)$ is continuous at $\kappa = 0$. (iii) If $K \geq S_0$, the payoff is $S_0(e^{\overline{X}_T} - e^c)^+$; if $K < S_0$, it is $\overline{S}_T - K = (\overline{S}_T - S_T) + S_T - K$. □

For the test case the floating-strike put costs 17.23062 and the call 14.27527; the fixed-strike call and put at $K = 100$ cost 16.24550 and 15.26039, about twice the vanilla prices. The floating-strike options were priced by Goldman et al. (1979), the fixed-strike ones by Conze and Viswanathan (1991).

16.6 Discrete monitoring

16.6.1 Exact prices by recursion over the monitoring dates

With dates $t_j = jh$, the knock-out indicator is a function of the vector $(X_{t_1}, \dots, X_{t_m})$, whose increments are independent $\mathcal{N}(\nu h, \sigma^2 h)$ (Theorem 7.2). The price is therefore an m-dimensional

Gaussian integral; the course expresses it through m-variate normal distribution functions, which are expensive beyond a few dates. A recursion integrates one date at a time.

Proposition 16.6 (Discrete barriers by backward recursion)**.** *Let $I \subseteq \mathbb{R}$ be an open interval (the surviving set of $\log(S_t/S_0)$, for instance (b,∞), $(-\infty,b)$ or (ℓ,u)), and $g \geq 0$ with $\mathbb{E}^{\mathbb{Q}}[g(S_T)] < \infty$. Define $v_m(x) = g(S_0e^x)\mathbf{1}_I(x)$ and, for $j = m-1,\dots,1$, $v_j(x) = \mathbf{1}_I(x)\int_I p_h(x'-x)\,v_{j+1}(x')\,\mathrm{d}x'$. The option paying $g(S_T)$ if $X_{t_j} \in I$ for $j = 1,\dots,m$ costs*

$$V_m = e^{-rT}\int_I p_h(x')\,v_1(x')\,\mathrm{d}x',$$

and, if $m \geq 2$, $v_{m-1}(x) = \mathbf{1}_I(x)\,e^{rh}G_h(S_0e^x)$, where G_h is the Black–Scholes price (16.4) with maturity h and surviving set S_0e^I.

Proof. By backward induction, $v_j(X_{t_j}) = \mathbb{E}^{\mathbb{Q}}\big[g(S_T)\prod_{i\geq j}\mathbf{1}_I(X_{t_i})\,\big|\,\mathcal{F}_{t_j}\big]$: for $j = m$ this is the definition, and given $\mathcal{F}_{t_j}$ the increment $X_{t_{j+1}} - X_{t_j}$ is independent with density p_h, so the independence lemma and the tower property give the step. The last line is $\mathbb{E}^{\mathbb{Q}}[v_1(X_{t_1})]$ discounted, and v_{m-1} is a one-step European price. □

The payoff kink and the barrier enter only v_{m-1}, in closed form, so every integrand is smooth on I, and composite Gauss–Legendre rules (panels of width $\sigma\sqrt{h}$, 8 nodes each, I truncated at 8 standard deviations $\sigma\sqrt{T}$) converge fast; one step is then a matrix–vector product, the same for every date. For the course's monthly down-and-out call ($B = 95$, $m = 12$) this gives 5.329789841605 in milliseconds with 232 nodes, and doubling the nodes leaves twelve digits unchanged. It reproduces all 112 entries of the course's table of discrete down-and-out calls ($B = 85,\dots,100$, $m = 1,2,3,4,6,12,\infty$) to the five decimals shown, and bivariate normals check $m = 2$ independently. Table 16.1 shows how slowly the discrete prices approach the continuous one.

Table 16.1: Discretely monitored down-and-out calls on the test case: exact prices (Theorem 16.6), continuity-corrected prices (Theorem 16.7) and the relative error of the latter; $m = \infty$ is Theorem 16.2. The last rows are the up-and-out call with $K = 100$, $B = 115$.

		$m = 4$	$m = 12$	$m = 52$	$m = 252$	$m = \infty$
$B = 95$	exact	6.22816	5.32979	4.57782	4.14873	3.75412
	BGK	6.06446	5.30339	4.57963	4.14875	
	error	−2.6%	−0.50%	+0.04%	+0.001%	
$B = 99$	exact	5.27574	3.67565	2.26089	1.48813	0.87818
	BGK	4.61914	3.30925	2.14258	1.47501	
	error	−12%	−10%	−5.2%	−0.9%	
UOC	exact	1.11608	0.83710	0.63269	0.53166	0.44998
	BGK	1.35690	0.90999	0.64809	0.53468	
	error	+22%	+8.7%	+2.4%	+0.57%	

16.6.2 The continuity correction

Proposition 16.7 (Continuity correction; Broadie, Glasserman and Kou)**.** *Let S be as in (16.1) with constant $r, y, \sigma > 0$, let $h = T/m$, $t_j = jh$, and $\beta = -\zeta(\frac{1}{2})/\sqrt{2\pi} = 0.58260\ldots$, where ζ is the Riemann zeta function.*

(i) Barriers. *Let $V(B)$ be the price of one of the eight standard barrier options with continuous monitoring and barrier $B \neq S_0$, and $V_m(B)$ the price of the same option monitored at $t_1, \ldots, t_m$. As $m \to \infty$,*

$$V_m(B) = V\left(Be^{\beta\sigma\sqrt{h}}\right) + o(\sqrt{h}) \qquad \text{for an up barrier } (B > S_0),$$
$$V_m(B) = V\left(Be^{-\beta\sigma\sqrt{h}}\right) + o(\sqrt{h}) \qquad \text{for a down barrier } (B < S_0):$$

the discrete contract is priced by the continuous formula with the barrier moved away from the spot *by the factor $e^{\beta\sigma\sqrt{h}}$.*

(ii) The size of the effect. *$V_m(B) - V(B) = \pm\beta\sigma\sqrt{h}\, BV'(B) + o(\sqrt{h})$, with $+$ for up and $-$ for down barriers. Whenever $V'(B) \neq 0$, the continuous price, and the mean of any simulation that checks a continuous barrier only on a grid of step h, are off by an amount of exact order $h^{1/2}$.*

(iii) Extremes and lookbacks. *With $\widetilde{M}_m = \max_{0\le j\le m} X_{t_j}$, $\mathbb{E}^{\mathbb{Q}}[\widetilde{M}_m] = \mathbb{E}^{\mathbb{Q}}[\overline{X}_T] - \beta\sigma\sqrt{h} + O(h)$; and for $K > S_0$ the fixed-strike lookback call on $S_0 e^{\widetilde{M}_m}$ costs $e^{-\beta\sigma\sqrt{h}}\, V^{\mathrm{lb}}(Ke^{\beta\sigma\sqrt{h}}) + o(\sqrt{h})$, where $V^{\mathrm{lb}}(k)$ is the continuous fixed-strike lookback call with strike k. So a maximum or a lookback inspected only on the grid also has a bias of exact order $h^{1/2}$.*

Part (i) is Theorem 1.1 of Broadie et al. (1997) for the down-and-in and down-and-out calls and the up-and-in and up-and-out puts, and Kou (2003) extends it to all eight options (with a simpler proof). Part (iii) is Lemma 3 and Theorem 4 of Broadie et al. (1999); their Theorem 4 also shifts the prior maximum, which is irrelevant here because $K > S_0$. Floating-strike lookbacks (their Theorem 3) are not part of the proposition. BGK 1999, footnote 3, note that a continuous yield is accommodated by $\nu = r - y$; the 1997 proof uses only that the drift is constant. Part (ii) follows from (i) by a Taylor expansion, as V is continuously differentiable in B (differentiate (16.5)). Exercise 7.6 derives the expansion of $\mathbb{E}^{\mathbb{Q}}[\widetilde{M}_m]$ without drift from Spitzer's identity.

Remark 16.8 (Beyond geometric Brownian motion). Theorem 16.7 assumes exact grid values of a geometric Brownian motion. For a general diffusion killed on leaving an open set and simulated by the Euler scheme with the killing checked only at the grid dates, the weak error is also of order $h^{1/2}$, and this order is exact; with the killing decided by the Brownian bridge between dates (the *continuous* Euler scheme) it is of order h, under conditions on the payoff (Gobet, 2000).

Why $\beta\sigma\sqrt{h}$ (heuristic, not a proof). At the first date after a crossing the grid value has overshot the barrier, by $\beta\sigma\sqrt{h}$ on average when the barrier is far away in units of $\sigma\sqrt{h}$ (BGK 1997, Lemma 4.2): the discrete path sees a barrier that much further away.

When it fails. The correction is asymptotic in $\log(S_0/B)/(\sigma\sqrt{h})$: with $B = 99$ the barrier is 0.17 monthly standard deviations from the spot, and the corrected price is 10% too low (Table 16.1; Broadie et al. (1997), Section 2, note the same near the spot). And the $o(\sqrt{h})$ term can be large: the up-and-out call, whose payoff jumps from $B - K = 15$ to 0 at the barrier, is overpriced by 8.7% with monthly dates, the error decaying like h (Figure 16.2). Unequal spacing, few dates

and term structures are outside the theorem; Theorem 16.6 handles all three.

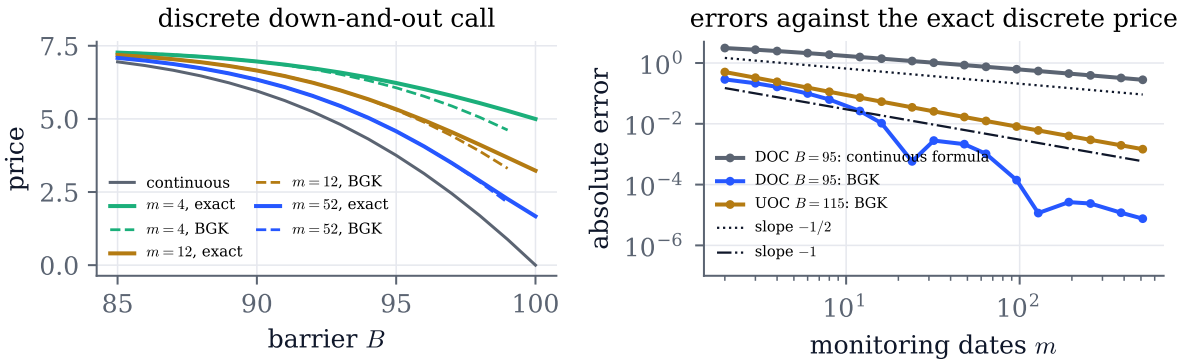

Figure 16.2: Test case. Left: down-and-out call against the barrier, exact for $m = 4, 12, 52$ dates (solid), BGK (dashed; not at $B = S_0$) and continuous (grey). Right: absolute errors against m of the continuous and BGK prices of the down-and-out call with $B = 95$ (the BGK error changes sign between $m = 16$ and 24) and of the BGK price of the up-and-out call with $B = 115$; reference slopes $-\frac{1}{2}$ and -1.

16.7 The Brownian bridge

Checking a continuous barrier only on the time grid of a simulation estimates the discrete price $V_m(B)$, with a bias of order $h^{1/2}$ (Theorem 16.7(ii)): +1.58 with monthly steps and still +0.39 with daily ones for the down-and-out call with $B = 95$. The Brownian bridge (Theorem 7.7) removes it: given the grid, the path between two dates is a bridge whose extremes have a known law.

Proposition 16.9 (Crossing probabilities of the bridge)**.** *Let X be as in (16.1), $0 = t_0 < t_1 < \cdots < t_m = T$ a grid (not necessarily uniform), $h_j = t_j - t_{j-1}$, and $\mathcal{G} = \sigma(X_{t_0}, \dots, X_{t_m})$.*

(i) For $s \geq 0$, $h > 0$ and a level b,

$$\mathbb{Q}\Big(\max_{s \leq t \leq s+h} X_t \geq b \;\Big|\; X_s = x,\ X_{s+h} = z \Big) = \exp\Big(-\frac{2(b-x)(b-z)}{\sigma^2 h} \Big) \quad \text{if } x, z < b,$$

and $\mathbb{Q}(\min_{[s,s+h]} X \leq b \mid X_s = x, X_{s+h} = z) = \exp\big(-2(x-b)(z-b)/(\sigma^2 h) \big)$ if $x, z > b$; the probability is 1 if x or z is beyond b. It does not depend on the drift ν. For the stock, the crossing probability of a barrier B over a step from S to S' is $\exp\big(-2\log(B/S)\log(B/S')/(\sigma^2 h) \big)$.

(ii) Given $\mathcal{G}$, the events $\{\max_{[t_{j-1},t_j]} X \geq b\}$, $j = 1, \dots, m$, are independent with the probabilities p_j of (i) with $h = h_j$; in particular $\mathbb{Q}(\overline{X}_T < b \mid \mathcal{G}) = \prod_{j=1}^m (1 - p_j)$, and likewise for minima.

(iii) Given $\mathcal{G}$, the step maxima $M_j = \max_{[t_{j-1},t_j]} X$ are independent, and if $U_1, \dots, U_m$ are independent uniform variables on $(0, 1]$, independent of $\mathcal{G}$, then

$$\widehat{M}_j = \tfrac{1}{2}\Big(X_{t_{j-1}} + X_{t_j} + \sqrt{(X_{t_j} - X_{t_{j-1}})^2 - 2\sigma^2 h_j \log U_j} \Big)$$

have the same joint law with $\mathcal{G}$ as the M_j (for minima, take the minus sign before the square root). So $(X_{t_0},\dots,X_{t_m},\max_j \widehat{M}_j)$ has the law of $(X_{t_0},\dots,X_{t_m},\overline{X}_T)$.

(iv) For a continuously monitored up-and-out claim with payoff $Y = e^{-rT}g(S_T)\mathbf{1}_{\{\overline{X}_T<b\}}$, $\mathbb{E}^{\mathbb{Q}}[g(S_T)^2]<\infty$, the estimator $\widehat{Y} = e^{-rT}g(S_T)\prod_j(1-p_j)$, computed from an exact sample of the grid values, is unbiased for $\mathbb{E}^{\mathbb{Q}}[Y]$ and $\operatorname{Var}\widehat{Y} \le \operatorname{Var} Y$; likewise for down barriers. Refining the grid (adding dates to it) can only increase $\operatorname{Var}\widehat{Y}$.

Proof. (i) Restarting at s (the process $X_{s+\cdot}-X_s$ is independent of $\mathcal{F}_s$ with the law of X) reduces to $s=0$, $x=0$ and the level $b'=b-x>0$. For $z<b'$, Theorem 16.1(ii) gives $\mathbb{Q}(\overline{X}_h\ge b', X_h\in \mathrm{d}z) = e^{\lambda b'}p_h(z-2b')\,\mathrm{d}z$, and dividing by $p_h(z)$,

$$\exp\Big(\lambda b' - \frac{(z-2b'-\nu h)^2-(z-\nu h)^2}{2\sigma^2 h}\Big) = \exp\Big(\frac{2\nu b'}{\sigma^2}+\frac{2b'(z-\nu h)-2b'^2}{\sigma^2 h}\Big) = \exp\Big(-\frac{2b'(b'-z)}{\sigma^2 h}\Big),$$

which is $(b-x)(b-(x+z))$ in the original variables. With $\pi(z)$ this expression for $z<b'$ and $\pi(z)=1$ for $z\ge b'$, $\mathbb{E}^{\mathbb{Q}}[\mathbf{1}\{\overline{X}_h\ge b'\}\varphi(X_h)] = \mathbb{E}^{\mathbb{Q}}[\pi(X_h)\varphi(X_h)]$ for bounded Borel φ, which is the statement. Minima follow by symmetry.

(ii)–(iii) For $t\in[t_{j-1},t_j]$ let $Y^j_t = X_t - \frac{(t_j-t)X_{t_{j-1}}+(t-t_{j-1})X_{t_j}}{h_j}$. The drift cancels, $Y^j_t = \sigma(W_t -$ its interpolation), and by Theorem 7.7(i) Y^j_t is independent of $\mathcal{G}_{t_{j-1},t_j} = \sigma(W_r : r\notin(t_{j-1},t_j))$, which contains $\mathcal{G}$ and every Y^k_u, $k\ne j$. All these variables are linear in W, hence jointly Gaussian, and they are uncorrelated across the families $\mathcal{G}$, $Y^1,\dots,Y^m$; so the processes $Y^1,\dots,Y^m$ and the grid vector are independent. Now $M_j = F(X_{t_{j-1}},X_{t_j},Y^j)$ for a measurable F (a supremum over rational times, by continuity), and the independence lemma gives the conditional independence, with conditional probabilities equal to versions of those in (i). In (iii), $\widehat{M}_j$ solves $\exp(-2(\widehat{M}_j-X_{t_{j-1}})(\widehat{M}_j-X_{t_j})/(\sigma^2 h_j)) = U_j$ with $\widehat{M}_j\ge\max(X_{t_{j-1}},X_{t_j})$, so by inversion (Theorem 5.8) its conditional law is that of M_j in (i).

(iv) By (ii), $\widehat{Y} = \mathbb{E}^{\mathbb{Q}}[Y\mid\mathcal{G}]$, and Theorem 10.11(i) applies. On a coarser grid contained in the finer one, the σ-algebra $\mathcal{G}$ is smaller, and Theorem 10.11(ii) gives the last claim. □

In code

`code/src/mfe230d/barriers.py`: `barrier_price`, `double_barrier_price`, `lookback_floating`, `lookback_fixed`, `discrete_barrier_price`, `bgk_price`, `crossing_probability`, `bridge_extreme`, and `barrier_mc_samples` with `monitor="grid"`, `"bridge"` or `"bridge-sample"`.

Figure 16.3 compares three estimators of the continuous down-and-out call ($B=95$, exact 3.75412). The bridge estimators are unbiased for every m, *including* $m=1$: one normal for S_T and the weight $1-\exp(-2\log(S_0/B)\log(S_T/B)/(\sigma^2T))$ on $\{S_T>B\}$ give an unbiased estimate of the continuously monitored price. The weighted estimator has the smallest variance with one step (standard deviation 8.06 per path, against 9.88 with 12 steps and 10.6 for the sampled minima; our implementation, numpy `default_rng`, seed 7, 2×10^5 paths), as (iv) predicts. Steps are needed when the payoff looks at intermediate dates or the parameters change in time: Theorem 16.9 holds step by step when r, y, σ are constant on each step, with σ_j in place of σ (same proof): the bridge survives the term structures that defeat the images.

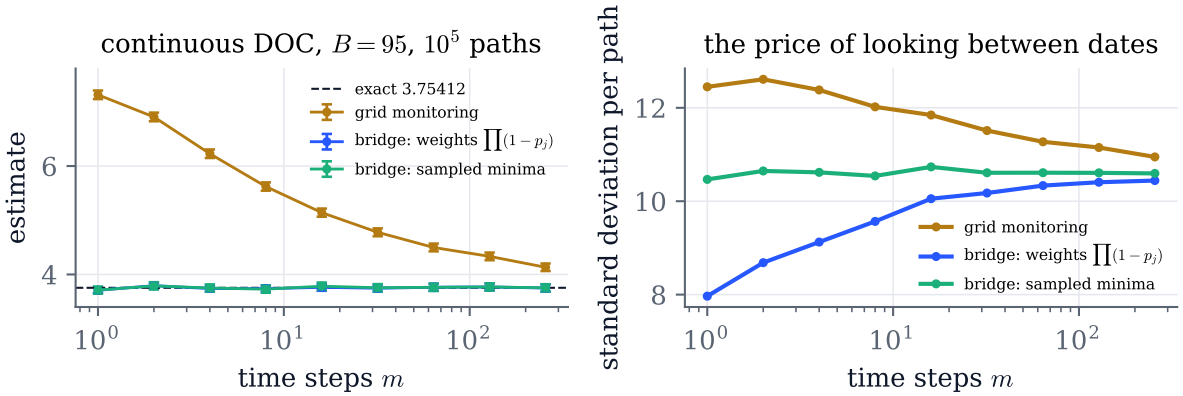

Figure 16.3: Continuous down-and-out call, test case, $B = 95$, from m exact steps (10^5 paths per point; our implementation, numpy `default_rng`, seeds $[16, i]$). Left: estimates ± 2 standard errors with grid monitoring and the two bridge estimators of Theorem 16.9; dashed, the closed form. Right: standard deviation per path.

The course's discrete down-and-out call by simulation. The monthly down-and-out call ($B = 95$, price 5.32979) needs no bridge, since it looks only at the dates; but the continuous price, known in closed form, is a natural control variate. Table 16.2 reproduces the course's study, with coefficients estimated on the same paths (Theorems 8.12 and 8.15). The continuous control is the sampled payoff (Theorem 16.9(iii)), as in the course, or its conditional expectation given the grid ((iv)), which is better correlated with the discrete payoff.

Table 16.2: Monthly down-and-out call, $B = 95$ (exact 5.32979), 10^6 paths or antithetic pairs (our implementation, numpy `default_rng`, seed 2026). VRF: variance reduction factor.

estimator	estimate	s.e.	VRF	course VRF
plain	5.33642	0.01191	1	1
antithetic pairs (Theorem 8.2; $\rho = -0.201$)	5.33197	0.00752	2.51	2.51
control: continuous DOC, sampled, $\beta^* = 0.947$	5.33882	0.00637	3.49	3.54
control: continuous DOC, bridge weights, $\beta^* = 1.101$	5.33621	0.00491	5.89	
control: vanilla call, $\beta^* = 0.825$	5.32051	0.00582	4.19	4.19
both controls (sampled DOC, vanilla)	5.32770	0.00464	6.59	6.64
both controls (bridge weights, vanilla)	5.32900	0.00399	8.91	

All estimates are within 1.6 standard errors of the exact price. Antithetic pairs cost two paths, so their efficiency gain (Theorem 6.9) is only about 1.25.

▶ Interactive demo: images against simulated survivors, exact discrete prices against the continuity correction, and grid against bridge Monte Carlo, for any parameters.

Further reading

Glasserman (2003), Section 6.4, treats barrier crossings by Brownian interpolation, and Shreve (2004), Chapter 7, barrier and lookback prices at the level of these notes.

Chapter 17

FX and multi-asset options

A currency option is the simplest option on two assets: the foreign money market and the domestic one. Its price depends on which currency we count in, and the answer to "what is it worth?" must not. This invariance is a change of *numéraire*, the unit in which prices are counted, and it is the one tool that runs through the chapter. It prices currency options from either side of the market (Sections 17.1 and 17.2), turns the option to exchange one asset for another into a Black–Scholes call (Section 17.3), and explains the drift adjustment of quanto contracts (Section 17.4). Payoffs that do not reduce to one dimension, such as baskets, are priced by simulating correlated assets, with the geometric basket as a control variate, as for the Asian options of Chapter 15 (Section 17.5).

We write calls on lognormal variables with *Black's function*: for $F > 0$, $v > 0$ and $K > 0$,

$$\operatorname{Black}(F, K, v) = F\,N(d_+) - K\,N(d_-), \qquad d_\pm = \frac{\log(F/K) \pm \frac{1}{2}v}{\sqrt{v}}, \tag{17.1}$$

and $\operatorname{Black}(F, K, 0) = (F - K)^+$, $\operatorname{Black}(F, K, v) = F - K$ for $K \le 0$. If L is lognormal with mean F and $\operatorname{Var} \log L = v$, then $\mathbb{E}(L - K)^+ = \operatorname{Black}(F, K, v)$: this is the computation behind Theorem 2.1, and the Black–Scholes call is $\operatorname{BS}(S, K, T; r, y, \sigma) = e^{-rT} \operatorname{Black}(Se^{(r-y)T}, K, \sigma^2 T)$. Black's function is positively homogeneous of degree one in (F, K).

17.1 Currency options

A currency market has a domestic and a foreign money market, with constant rates r_d and r_f, and an *exchange rate* X_t, the domestic price of one unit of foreign currency. With the dollar domestic and the pound foreign, X is the USD price of one GBP, quoted as GBP/USD ("cable"). A pound deposited at the rate r_f is worth $X_t e^{r_f t}$ dollars at time t, so for a dollar investor the pound is an asset that pays a continuous dividend at the rate r_f. Under a risk-neutral measure $\mathbb{Q}$ for that investor (Theorem 1.3), $e^{-r_d t} X_t e^{r_f t}$ is a martingale. The model of Garman and Kohlhagen (1983) is the standard model of Theorem 1.1 for this asset:

$$\frac{\mathrm{d}X_t}{X_t} = (r_d - r_f)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t \quad \text{under } \mathbb{Q}, \qquad \sigma > 0, \tag{17.2}$$

that is, $S = X$, $r = r_d$ and $y = r_f$: the foreign rate is the yield.

Proposition 17.1 (Garman–Kohlhagen)**.** *In the model (17.2), let $K > 0$, $T > 0$, $F = X_0 e^{(r_d - r_f)T}$ be the forward exchange rate and $d_\pm = \big(\log(F/K) \pm \frac{1}{2}\sigma^2 T\big)/(\sigma\sqrt{T})$.*

(i) A call on one unit of foreign currency with strike K, paying $(X_T - K)^+$ units of domestic currency, and the put, paying $(K - X_T)^+$, cost in domestic currency

$$C = X_0 e^{-r_f T} N(d_+) - K e^{-r_d T} N(d_-) = e^{-r_d T}\,\mathrm{Black}(F, K, \sigma^2 T),$$
$$P = K e^{-r_d T} N(-d_-) - X_0 e^{-r_f T} N(-d_+),$$

and C/X_0, P/X_0 in foreign currency. Parity reads $C - P = e^{-r_d T}(F - K)$; at the money forward, $K = F$, $C = P = e^{-r_d T} F\big(2N(\frac{1}{2}\sigma\sqrt{T}) - 1\big)$. If $r_d = r_f$, then $F = X_0$ and at the money forward is at the money spot.

(ii) The replicating portfolio of the call holds, at time 0, $\Delta_f = \partial C/\partial X_0 = e^{-r_f T} N(d_+)$ units of foreign currency and $\Delta_d = C - X_0\,\partial C/\partial X_0 = -K e^{-r_d T} N(d_-)$ units of domestic currency, both in the money markets; $C = X_0\Delta_f + \Delta_d$.

Proof. (i) is Theorem 2.1 with $S = X$, $r = r_d$, $y = r_f$; the ATMF value uses $d_\pm = \pm\frac{1}{2}\sigma\sqrt{T}$. (ii) The delta is that of Theorem 2.7; it is a number of units of the asset, here foreign currency, held with its "dividends" (the foreign interest) reinvested, as in the replication of Theorem 1.3(ii). The rest of the portfolio is cash, $C - X_0\Delta_f = -K e^{-r_d T} N(d_-)$. The price is homogeneous of degree one in (X_0, K), so $\Delta_d = K\,\partial C/\partial K$ by Euler's theorem (Theorem 4.7). □

Example 17.2 (The course's cable call)**.** With $X_0 = 1.2000$, $r_d = r_{\mathrm{USD}} = 3\%$, $r_f = r_{\mathrm{GBP}} = 2.5\%$, $\sigma = 11\%$ and $T = 1$, the forward is $F = 1.2\,e^{0.005} = 1.206015$, and the ATMF call costs $C = 0.0513343$ USD, that is 0.0427786 GBP, or 4.2565% of the strike. Its hedge holds $\Delta_f = 0.509044$ GBP and $\Delta_d = -0.559519$ USD, and indeed $1.2 \times 0.509044 - 0.559519 = 0.051334$. These are the course's 0.05133, 0.04278, 0.5090 and −0.5595.

Currency markets quote differently from equity markets. Options are struck relative to the forward (ATMF rather than at the spot), premiums are quoted in either currency, and implied volatilities are quoted against a delta rather than a strike. The course's GBP/USD and EUR/USD smiles (CME settlement prices, 22 July 2022), plotted against the normalised strike $\log(K/F)/(\sigma_{\mathrm{atm}}\sqrt{T})$, fall from low strikes to a minimum above the forward and then rise again: a skew, as for equity indices (Theorem 3.7), with a smile on the upside. Neither currency is special: by Theorem 17.3(v) below, the USD/GBP smile is the mirror image of the GBP/USD smile, $\sigma_{\mathrm{USD/GBP}}(m) = \sigma_{\mathrm{GBP/USD}}(-m)$ in the log-moneyness $m = \log(K/F)$.

17.2 Changing the numéraire

In this section W is a d-dimensional Brownian motion under $\mathbb{Q}$ on $[0, T]$, $(\mathcal{F}_t)$ is its augmented natural filtration, and $B_t = e^{rt}$ is the money market. A *price process* is an adapted process V such that V/B is a $\mathbb{Q}$-martingale on $[0, T]$: the value of a traded asset that pays no dividends, of a self-financing portfolio, or of a claim. A stock that pays the dividend yield y is not one, but $e^{yt}S_t$, one share at 0 with its dividends reinvested, is. Counting prices in units of a positive price process N instead of money is a *change of numéraire*; Theorem 3.4 already counted prices in prepaid forwards.

Proposition 17.3 (Change of numéraire)**.** *Let N be a price process with $N_t > 0$ for all $t \le T$, almost surely.*

(i) $Z_T = \dfrac{N_T B_0}{N_0 B_T}$ *is positive with* $\mathbb{E}^{\mathbb{Q}} Z_T = 1$*, so* $\mathrm{d}\mathbb{Q}^N = Z_T \, \mathrm{d}\mathbb{Q}$ *defines a probability measure on* $\mathcal{F}_T$ *equivalent to* $\mathbb{Q}$*, and* $\mathbb{E}^{\mathbb{Q}}[Z_T \mid \mathcal{F}_t] = N_t B_0/(N_0 B_t)$*.*

(ii) (Numéraire invariance.) *An adapted process V is a price process if and only if V/N is a $\mathbb{Q}^N$-martingale. For every $\mathcal{F}_T$-measurable H with $\mathbb{E}^{\mathbb{Q}}[|H|/B_T] < \infty$, equivalently $\mathbb{E}^{\mathbb{Q}^N}[|H|/N_T] < \infty$,*

$$B_t \, \mathbb{E}^{\mathbb{Q}}\Big[\frac{H}{B_T}\,\Big|\,\mathcal{F}_t\Big] = N_t \, \mathbb{E}^{\mathbb{Q}^N}\Big[\frac{H}{N_T}\,\Big|\,\mathcal{F}_t\Big], \qquad 0 \le t \le T.$$

(iii) (Drifts.) *Suppose $N_t/B_t = N_0 \exp(\theta \cdot W_t - \frac{1}{2}|\theta|^2 t)$ for a constant $\theta \in \mathbb{R}^d$. Then $W_t^N = W_t - \theta t$ is a $\mathbb{Q}^N$-Brownian motion. If $\mathrm{d}S_t/S_t = \mu_t \, \mathrm{d}t + \sigma_S \cdot \mathrm{d}W_t$ with $\sigma_S \in \mathbb{R}^d$, then $\mathrm{d}S_t/S_t = (\mu_t + \sigma_S \cdot \theta)\,\mathrm{d}t + \sigma_S \cdot \mathrm{d}W_t^N$: the drift rises by $\sigma_S \cdot \theta$, the instantaneous covariance of $\log S$ and $\log N$, and the volatility does not change.*

(iv) (The share measure.) *If $\mathrm{d}S_t/S_t = (r - y)\,\mathrm{d}t + \sigma_S \cdot \mathrm{d}W_t$ with constants y and σ_S, then $N_t = e^{yt} S_t$ satisfies (iii) with $\theta = \sigma_S$,*

$$\frac{\mathrm{d}\mathbb{Q}^S}{\mathrm{d}\mathbb{Q}} = \frac{S_T}{\mathbb{E}^{\mathbb{Q}}[S_T]},$$

and under $\mathbb{Q}^S := \mathbb{Q}^N$ the log-price has drift $r - y + \frac{1}{2}|\sigma_S|^2$ instead of $\nu = r - y - \frac{1}{2}|\sigma_S|^2$.

(v) (Currency put–call symmetry.) *In a currency market as in Section 17.1, with no assumption on the dynamics of X, let $N_t = X_t e^{r_f t}$ (the foreign money market in domestic currency) be a positive price process and $\mathbb{Q}_f = \mathbb{Q}^N$, with $r = r_d$. A payoff of $H \ge 0$ units of foreign currency at T is worth $e^{-r_f T}\mathbb{E}^{\mathbb{Q}_f}[H]$ in foreign currency at 0, and for $K > 0$*

$$e^{-r_d T}\mathbb{E}^{\mathbb{Q}}\big[(X_T - K)^+\big] = X_0 K \cdot e^{-r_f T}\mathbb{E}^{\mathbb{Q}_f}\Big[\Big(\frac{1}{K} - \frac{1}{X_T}\Big)^+\Big]:$$

a call on one unit of foreign currency struck at K is worth $X_0 K$ times the foreign price of a put on one unit of domestic currency struck at $1/K$. In the model (17.2), $1/X$ has drift $r_f - r_d$ and volatility σ under $\mathbb{Q}_f$, so the put is the Garman–Kohlhagen put with the roles of r_d and r_f exchanged.

Proof. (i) $Z_T > 0$, and $\mathbb{E}^{\mathbb{Q}} Z_T = 1$ because N/B is a martingale; equivalence follows from $Z_T > 0$. The martingale property gives $\mathbb{E}^{\mathbb{Q}}[Z_T \mid \mathcal{F}_t] = N_t B_0/(N_0 B_t) =: Z_t$.

(ii) By Bayes' rule for conditional expectations (Shreve (2004), Lemma 5.2.2), for $s \le T$ and $\mathcal{F}_s$-measurable Y with $\mathbb{E}^{\mathbb{Q}^N}|Y| < \infty$, $\mathbb{E}^{\mathbb{Q}^N}[Y \mid \mathcal{F}_t] = \mathbb{E}^{\mathbb{Q}}[Z_s Y \mid \mathcal{F}_t]/Z_t$ for $t \le s$. With $Y = V_s/N_s$, $Z_s Y = (B_0/N_0)V_s/B_s$, so

$$\mathbb{E}^{\mathbb{Q}^N}\Big[\frac{V_s}{N_s}\,\Big|\,\mathcal{F}_t\Big] = \frac{B_t}{N_t}\,\mathbb{E}^{\mathbb{Q}}\Big[\frac{V_s}{B_s}\,\Big|\,\mathcal{F}_t\Big], \qquad \mathbb{E}^{\mathbb{Q}^N}\Big|\frac{V_s}{N_s}\Big| = \frac{B_0}{N_0}\,\mathbb{E}^{\mathbb{Q}}\Big|\frac{V_s}{B_s}\Big|.$$

Hence V/N is $\mathbb{Q}^N$-integrable if and only if V/B is $\mathbb{Q}$-integrable, and $V_t/N_t = \mathbb{E}^{\mathbb{Q}^N}[V_s/N_s \mid \mathcal{F}_t]$ if

and only if $V_t/B_t = \mathbb{E}^{\mathbb{Q}}[V_s/B_s \mid \mathcal{F}_t]$. The pricing identity is the display with $s = T$ and $V_T = H$.

(iii) The density process $Z_t = \exp(\theta \cdot W_t - \frac{1}{2}|\theta|^2 t)$ is the one of Girsanov's theorem (Karatzas and Shreve (1991), Theorem 3.5.1), so W^N is a $\mathbb{Q}^N$-Brownian motion, and substituting $\mathrm{d}W_t = \mathrm{d}W_t^N + \theta\,\mathrm{d}t$ gives the dynamics. The covariation of $\log S$ and $\log N$ is $\sigma_S \cdot \theta\,\mathrm{d}t$, since $\mathrm{d}\log N_t = (r - \frac{1}{2}|\theta|^2)\,\mathrm{d}t + \theta \cdot \mathrm{d}W_t$.

(iv) $N_t/B_t = S_t e^{(y-r)t} = S_0 \exp(\sigma_S \cdot W_t - \frac{1}{2}|\sigma_S|^2 t)$, so $\theta = \sigma_S$ and $Z_T = S_T e^{-(r-y)T}/S_0 = S_T/\mathbb{E}^{\mathbb{Q}}[S_T]$. By (iii) the drift of $\mathrm{d}S/S$ becomes $r - y + |\sigma_S|^2$, and that of $\log S$ is $r - y + \frac{1}{2}|\sigma_S|^2$.

(v) A foreign payoff H is the domestic payoff $X_T H$; by (ii), which holds for nonnegative payoffs whether or not their prices are finite (Bayes' rule does), its domestic price is $N_0 \mathbb{E}^{\mathbb{Q}_f}[X_T H/N_T] = X_0 e^{-r_f T}\mathbb{E}^{\mathbb{Q}_f}[H]$, which is $e^{-r_f T}\mathbb{E}^{\mathbb{Q}_f}[H]$ units of foreign currency. The identity follows from $(X_T - K)^+ = X_T K(1/K - 1/X_T)^+$, a foreign payoff of $K(1/K - 1/X_T)^+$. In the model (17.2), $N_t/B_t = X_0 \exp(\sigma W_t - \frac{1}{2}\sigma^2 t)$, so by (iii) $\mathrm{d}X/X = (r_d - r_f + \sigma^2)\,\mathrm{d}t + \sigma\,\mathrm{d}W^N$ under $\mathbb{Q}_f$, and Itô's formula gives $\mathrm{d}(1/X)/(1/X) = (r_f - r_d)\,\mathrm{d}t + \sigma\,\mathrm{d}(-W^N)$, with $-W^N$ a $\mathbb{Q}_f$-Brownian motion. □

The hypothesis that N/B is a true martingale matters: a positive local martingale is a supermartingale, and if $\mathbb{E}^{\mathbb{Q}}[N_T/B_T] < N_0/B_0$ then $\mathbb{Q}^N$ has total mass below one. In (iii) and (iv) the exponential has constant coefficients, and it is a martingale. Geman et al. (1995) developed the method in this generality; Shreve (2004), Section 9.2 (Theorem 9.2.2) and Sections 9.3.3 and 9.3.7, gives it for the foreign measure and the put–call duality of (v). Under the share measure $N(d_+)$ is the probability that the call finishes in the money, as noted after Theorem 2.1, and the density in (iv) is the one used by Theorem 9.4.

Example 17.4 (The course's London trader)**.** A trader in London counts in pounds. For her the dollar is the foreign currency, worth $X_0' = 1/1.2 = 0.833333$ GBP, with rate $r_{\text{GBP}} = 2.5\%$ and yield $r_{\text{USD}} = 3\%$, and by (v) the cable call of Theorem 17.2 is $K = 1.206015$ dollar puts struck at $1/K$ pounds. The Garman–Kohlhagen put is worth 0.0354710 GBP, so the call is worth $K \times 0.0354710 = 0.0427786$ GBP, the course's value, which is 0.0513343 USD at the spot rate. Her hedge is the same portfolio: the put's delta, -0.463940 dollars per put, times K gives -0.559519 USD, and the rest is 0.509044 GBP in cash, as in Theorem 17.2. Both traders hold the same assets, pounds and dollars; only the unit of account differs.

The two forms of put–call symmetry are one identity. In the model (17.2), X_T/F has under $\mathbb{Q}_f$ the law that F/X_T has under $\mathbb{Q}$ (both log-normal with log-mean $\frac{1}{2}\sigma^2 T$ and log-variance $\sigma^2 T$), and substituting this into (v) turns the foreign put into K/F domestic puts struck at F^2/K: this is Theorem 4.6, whose applications to static hedging of barriers are discussed there.

Remark 17.5 (Monte Carlo in another numéraire). By (ii), $V_0 = N_0\,\mathbb{E}^{\mathbb{Q}^N}[H/N_T]$: one may simulate under $\mathbb{Q}^N$ and average $N_0 H/N_T$. This is importance sampling with the likelihood ratio $\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{Q}^N = N_0 B_T/(N_T B_0)$ (Theorem 9.1). Under the share measure the call becomes $S_0 e^{-yT}\,\mathbb{E}^{\mathbb{Q}^S}[(1 - K/S_T)^+]$, an expectation of a payoff between 0 and 1. In the test case, with 2^{20} draws (our implementation, `numpy default_rng`, seed 3), the standard deviation per draw is 9.95 under $\mathbb{Q}^S$ against 12.57 under $\mathbb{Q}$, a variance ratio of 1.6; Exercise 17.2 shows that the first stays bounded as σ grows while the second does not.

17.3 Exchange options

From here on, d assets follow, under $\mathbb{Q}$,

$$\frac{\mathrm{d}S_{i,t}}{S_{i,t}} = (r - y_i)\,\mathrm{d}t + \sigma_i\,\mathrm{d}W_{i,t}, \qquad \mathrm{d}\langle W_i, W_j\rangle_t = \rho_{ij}\,\mathrm{d}t, \qquad i,j = 1,\dots,d, \tag{17.3}$$

with $\sigma_i > 0$ and a correlation matrix $\rho = (\rho_{ij})$, symmetric positive semidefinite with unit diagonal. Concretely, $W = A\bar{W}$ for a standard k-dimensional Brownian motion $\bar{W}$ and a $d \times k$ matrix A with $AA^\top = \rho$ (Theorem 5.14), and $(\mathcal{F}_t)$ is generated by $\bar{W}$. The log-prices $\log S_{i,T} = \log S_{i,0} + \nu_i T + \sigma_i W_{i,T}$, $\nu_i = r - y_i - \frac{1}{2}\sigma_i^2$, form a Gaussian vector with $\mathbb{E}S_{i,T} = F_i := S_{i,0}e^{(r-y_i)T}$ and covariance $\Sigma_{ij} = \rho_{ij}\sigma_i\sigma_j T$.

Theorem 17.6 (Margrabe's formula)**.** *Let $d = 2$, $\rho = \rho_{12} \in [-1,1]$, $T > 0$, and*

$$\sigma_{\mathrm{ex}} = \sqrt{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2} \;\in\; [|\sigma_1 - \sigma_2|,\, \sigma_1 + \sigma_2].$$

(i) The option to exchange asset 2 for asset 1 at T, paying $(S_{1,T} - S_{2,T})^+$, is worth, at 0,

$$M = \mathrm{Black}\left(S_{1,0}e^{-y_1T},\, S_{2,0}e^{-y_2T},\, \sigma_{\mathrm{ex}}^2 T\right) = \mathrm{BS}(S_{1,0}, S_{2,0}, T;\, y_2,\, y_1,\, \sigma_{\mathrm{ex}}) :$$

the Black–Scholes call with spot $S_{1,0}$, strike $S_{2,0}$, rate y_2, yield y_1 and volatility σ_{ex}. The interest rate r does not enter.

(ii) $\sigma_{\mathrm{ex}} = 0$ if and only if $\rho = 1$ and $\sigma_1 = \sigma_2$; then $M = (S_{1,0}e^{-y_1T} - S_{2,0}e^{-y_2T})^+$. In all cases σ_{ex} decreases strictly in ρ, from $\sigma_1 + \sigma_2$ at $\rho = -1$ to $|\sigma_1 - \sigma_2|$ at $\rho = 1$.

(iii) If $\sigma_{\mathrm{ex}} > 0$, M is homogeneous of degree one in $(S_{1,0}, S_{2,0})$ with $\partial M/\partial S_{1,0} = e^{-y_1T}N(d_+)$ and $\partial M/\partial S_{2,0} = -e^{-y_2T}N(d_-)$, $d_\pm = \left(\log\frac{S_{1,0}e^{-y_1T}}{S_{2,0}e^{-y_2T}} \pm \frac{1}{2}\sigma_{\mathrm{ex}}^2 T\right)/(\sigma_{\mathrm{ex}}\sqrt{T})$, and $M = S_{1,0}\,\partial_1 M + S_{2,0}\,\partial_2 M$: the delta hedge holds the two assets and no cash.

Proof. (ii) $\sigma_{\mathrm{ex}}^2 = (\sigma_1 - \sigma_2)^2 + 2(1-\rho)\sigma_1\sigma_2$, which vanishes exactly when both terms do, and is affine in ρ with slope $-2\sigma_1\sigma_2 < 0$.

(i) Write $W_i = a_i \cdot \bar{W}$, with a_i the i-th row of A, so that $|a_i| = 1$ and $a_1 \cdot a_2 = \rho$; this holds for every k, including the rank-one factors at $\rho = \pm 1$. The numéraire $N_t = e^{y_2 t}S_{2,t}$ has $N_t/B_t = S_{2,0}\exp(\sigma_2 a_2 \cdot \bar{W}_t - \frac{1}{2}\sigma_2^2 t)$, so it satisfies Theorem 17.3(iii), with $\bar{W}$ in the role of W and $\theta = \sigma_2 a_2$. Under $\mathbb{Q}^N$, $\bar{W}^N = \bar{W} - \theta t$ is a standard Brownian motion; the drift of $\mathrm{d}S_1/S_1$ rises by $\sigma_1 a_1 \cdot \theta = \rho\sigma_1\sigma_2$ and that of $\mathrm{d}S_2/S_2$ by σ_2^2. Hence $R = S_1/S_2$ satisfies

$$\log R_T = \log R_0 + (y_2 - y_1 - \tfrac{1}{2}\sigma_{\mathrm{ex}}^2)T + u \cdot \bar{W}_T^N, \qquad u = \sigma_1 a_1 - \sigma_2 a_2,$$

because $\rho\sigma_1\sigma_2 - \frac{1}{2}\sigma_1^2 - \frac{1}{2}\sigma_2^2 = -\frac{1}{2}\sigma_{\mathrm{ex}}^2$; and $|u|^2 = \sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2 = \sigma_{\mathrm{ex}}^2$. So under $\mathbb{Q}^N$, R_T is lognormal with mean $R_0 e^{(y_2-y_1)T}$ and log-variance $\sigma_{\mathrm{ex}}^2 T$ (a constant if $\sigma_{\mathrm{ex}} = 0$). By Theorem 17.3(ii) with $H = (S_{1,T} - S_{2,T})^+$, for which $\mathbb{E}^{\mathbb{Q}}H \le \mathbb{E}^{\mathbb{Q}}S_{1,T} < \infty$,

$$\begin{aligned} M &= N_0\,\mathbb{E}^{\mathbb{Q}^N}\left[\frac{(S_{1,T} - S_{2,T})^+}{e^{y_2 T}S_{2,T}}\right] = S_{2,0}e^{-y_2T}\,\mathbb{E}^{\mathbb{Q}^N}\left[(R_T - 1)^+\right] \\ &= S_{2,0}e^{-y_2T}\,\mathrm{Black}\left(R_0 e^{(y_2-y_1)T}, 1, \sigma_{\mathrm{ex}}^2 T\right), \end{aligned}$$

and homogeneity of Black gives the first form; the second is the identity after (17.1) with $r = y_2$, $y = y_1$. For $\sigma_{\text{ex}} = 0$ this is the formula in (ii).

(iii) The derivatives are Δ and $\partial_K c$ of Theorem 2.7 for the inputs of (i), and the last identity is Euler's theorem (Theorem 4.7). □

With $y_1 = y_2 = 0$ this is the formula of Margrabe (1978). In the course's words, the numéraire change is a *dimensional reduction*: writing the price as $S_2\, D(S_1/S_2)$ turns the two-asset pricing equation into a one-asset equation for D, the price measured in units of asset 2. The same device is the pre-wash of Theorem 4.4. Options on the better or the worse of two assets reduce to exchange options (Exercise 17.3); the spread $(S_{1,T} - S_{2,T} - K)^+$ with $K \neq 0$ does not (Exercise 17.7).

Cross rates and the volatility triangle. For a dollar investor, the pound and the euro are two assets of the form (17.3): $S_1 = X_1$, the USD price of a pound, with yield r_{GBP}, and $S_2 = X_2$, the USD price of a euro, with yield r_{EUR}. The cross rate $X_{12} = X_1/X_2$ is the EUR price of a pound, and a GBP/EUR call struck at K pays $(X_{12,T} - K)^+$ euros, that is $(X_{1,T} - KX_{2,T})^+$ dollars: an exchange option. Its volatility is $\sigma_{12} = \sigma_{\text{ex}} = \left(\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2\right)^{1/2}$, with ρ the correlation of GBP/USD and EUR/USD returns, both quoted against the dollar. The three volatilities are the sides of a triangle with angle $\arccos\rho$ opposite σ_{12} (the law of cosines), so $|\sigma_1 - \sigma_2| \le \sigma_{12} \le \sigma_1 + \sigma_2$, and a quoted cross volatility implies the correlation $\rho = (\sigma_1^2 + \sigma_2^2 - \sigma_{12}^2)/(2\sigma_1\sigma_2)$. Dividing the dollar price by $X_{2,0}$ gives the Garman–Kohlhagen price of a euro-based investor with volatility σ_{12}: the law of one price at work.

Example 17.7 (The course's cross-currency inputs)**.** With $X_1 = 1.20$, $X_2 = 1.02$, $\sigma_1 = 11\%$, $\sigma_2 = 10\%$, $r_{\text{USD}} = 3\%$, $r_{\text{GBP}} = 2.5\%$ and $r_{\text{EUR}} = 1\%$, the GBP/EUR spot is 1.176471 and its one-year forward is $1.176471\, e^{0.01-0.025} = 1.158955$. Correlations in $[-1, 1]$ allow $\sigma_{12} \in [1\%, 21\%]$. The ATMF GBP/EUR call, paying $(1 \text{ GBP} - 1.158955 \text{ EUR})^+$, is worth between 0.004669 and 0.097871 USD; a quoted $\sigma_{12} = 8\%$ implies $\rho = 0.7136$ and the value 0.037343 USD (Figure 17.1). A quote outside $[1\%, 21\%]$ admits no correlation and is an arbitrage within the model; Exercise 17.6 asks what a static portfolio of four options (the cross call, the GBP/USD call, and the EUR/USD call and put) can do with it.

17.4 Correlated simulation and quantos

Simulating correlated assets. Under (17.3), a vector of terminal prices is sampled exactly in one step,

$$S_{i,T} = S_{i,0} \exp\left(\nu_i T + \sigma_i \sqrt{T}\,(AZ)_i\right), \qquad Z \sim \mathcal{N}(0, I_k), \quad AA^{\top} = \rho,$$

by Theorem 5.14(i), and a path is sampled by repeating this over each interval with independent Z's (Theorem 7.2). The Cholesky factor costs about $d^3/3$ operations once and $d^2/2$ per draw; a lower triangular factor with positive diagonal exists if and only if ρ is positive definite. At $\rho_{12} = \pm 1$ the second pivot $1 - \rho_{12}^2$ vanishes and the positive-pivot recursion stops, while the principal-components factor of Theorem 5.14(iii), which needs only semidefiniteness, still works; we use it in Figure 17.1. A matrix of pairwise correlations collected from different sources need

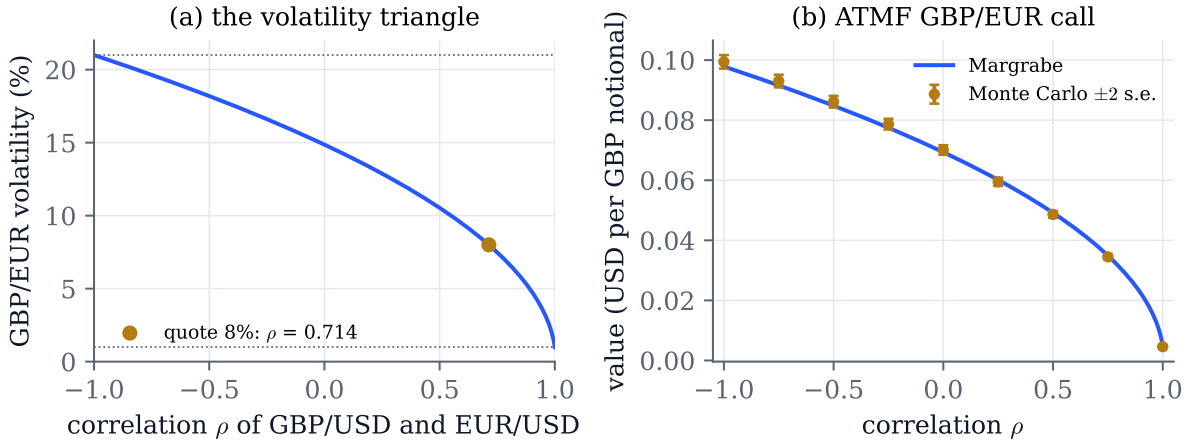

Figure 17.1: The cross rate GBP/EUR with $\sigma_1 = 11\%$ (GBP/USD), $\sigma_2 = 10\%$ (EUR/USD) and the rates of Theorem 17.7. (a) The cross volatility against the correlation ρ of GBP/USD and EUR/USD, between the dotted levels $|\sigma_1 - \sigma_2|$ and $\sigma_1 + \sigma_2$; the dot is the quote 8%, implying $\rho = 0.714$. (b) The USD value of the ATMF GBP/EUR call by Theorem 17.6, with Monte Carlo estimates and ± 2 standard errors, from the same 2^{14} normal pairs at every ρ (seed 17) and the principal-components factor, which also works at $\rho = \pm 1$.

not be semidefinite at all, and then no model has it. Antithetic draws negate the whole vector Z; moment matching must match the sample covariance as well as the means, which is harder (Theorem 8.7).

Quantos. A *quanto* pays a foreign quantity in domestic currency at a rate fixed in advance: an option on a foreign index, struck in foreign currency, whose payoff is converted at the rate $\bar{X}$, whatever the exchange rate at maturity. Let S be the foreign-currency price of a foreign stock with dividend yield y, X the exchange rate as in Section 17.1, and $\mathbb{Q}$ the domestic risk-neutral measure: the domestic money market, the foreign money market $X_t e^{r_f t}$ and the foreign stock with reinvested dividends, worth $X_t S_t e^{yt}$, all have $\mathbb{Q}$-martingale prices once discounted at r_d. Assume

$$\frac{\mathrm{d}X_t}{X_t} = \mu_X \,\mathrm{d}t + \sigma_X \,\mathrm{d}W_{X,t}, \qquad \frac{\mathrm{d}S_t}{S_t} = \mu_S \,\mathrm{d}t + \sigma_S \,\mathrm{d}W_{S,t}, \qquad \mathrm{d}\langle W_X, W_S\rangle_t = \rho \,\mathrm{d}t, \tag{17.4}$$

with constants μ_X, μ_S, $\sigma_X, \sigma_S > 0$ and $\rho \in [-1, 1]$, where $W_X = a_X \cdot \bar{W}$ and $W_S = a_S \cdot \bar{W}$ for a standard k-dimensional $\mathbb{Q}$-Brownian motion $\bar{W}$ generating the filtration and unit vectors with $a_X \cdot a_S = \rho$, as in (17.3). As in Theorem 1.1, $\mathbb{Q}$ is equivalent to the real-world measure $\mathbb{P}$ on $\mathcal{F}_T$.

Proposition 17.8 (Quantos and composites)**.** *In this market, with $T > 0$ and $K > 0$:*

(i) $\mu_X = r_d - r_f$ *and* $\mu_S = r_f - y - \rho\sigma_S\sigma_X$*. Under the foreign measure $\mathbb{Q}_f$ of Theorem 17.3(v), $\mathrm{d}S/S = (r_f - y)\,\mathrm{d}t + \sigma_S \,\mathrm{d}W_S^f$ with W_S^f a $\mathbb{Q}_f$-Brownian motion.*

(ii) (Quanto.) *The payoff $\bar{X}(S_T - K)^+$ in domestic currency costs*

$$\bar{X} e^{-r_d T}\,\mathrm{Black}\left(F^q, K, \sigma_S^2 T\right) = \bar{X}\,\mathrm{BS}(S_0, K, T;\, r_d,\, y_q,\, \sigma_S), \qquad F^q = S_0 e^{(r_f - y - \rho\sigma_S\sigma_X)T},$$

with $y_q = y + r_d - r_f + \rho\sigma_S\sigma_X$.

(iii) (Composite.) *The payoff* $(X_T S_T - K)^+$, *a foreign stock struck in domestic currency, costs* $\mathrm{BS}(X_0 S_0, K, T;\, r_d,\, y,\, \sigma_c)$ *with* $\sigma_c^2 = \sigma_S^2 + 2\rho\sigma_S\sigma_X + \sigma_X^2$.

(iv) (Foreign strike, spot conversion.) *The payoff* $X_T(S_T - K)^+$ *costs* $X_0\,\mathrm{BS}(S_0, K, T;\, r_f,\, y,\, \sigma_S)$.

Proof. (i) By Itô's formula, $X_t e^{(r_f - r_d)t}$ and $X_t S_t e^{(y - r_d)t}$ are exponentials of Brownian motions with constant coefficients, with drift rates $\mu_X + r_f - r_d$ and $\mu_X + \mu_S + \rho\sigma_S\sigma_X + y - r_d$ (the cross term is $\mathrm{d}\langle \log X, \log S\rangle = \rho\sigma_S\sigma_X\,\mathrm{d}t$). Such an exponential is a martingale if and only if its drift rate is zero (its mean is $e^{(\text{drift rate})\,t}$ times its initial value), which gives μ_X and then μ_S. Under $\mathbb{Q}_f$, Theorem 17.3(iii) with $N_t/B_t = X_0 \exp(\sigma_X W_{X,t} - \tfrac{1}{2}\sigma_X^2 t)$, that is $\theta = \sigma_X a_X$, raises the drift of $\mathrm{d}S/S$ by $\sigma_S a_S \cdot \theta = \rho\sigma_S\sigma_X$. (ii) Under $\mathbb{Q}$, S_T is lognormal with mean F^q and log-variance $\sigma_S^2 T$; and $\mathrm{BS}(S_0, K, T; r_d, y_q, \sigma_S) = e^{-r_d T}\,\mathrm{Black}(S_0 e^{(r_d - y_q)T}, K, \sigma_S^2 T)$ with $r_d - y_q = r_f - y - \rho\sigma_S\sigma_X$. (iii) By (i), $\log(X_T S_T)$ is normal with variance $\sigma_c^2 T$, and $\mathbb{E}^{\mathbb{Q}}[X_T S_T] = X_0 S_0 e^{(r_d - y)T}$ because $X_t S_t e^{(y - r_d)t}$ is a martingale. (iv) By Theorem 17.3(v) the foreign-currency price is $e^{-r_f T}\mathbb{E}^{\mathbb{Q}_f}[(S_T - K)^+] = \mathrm{BS}(S_0, K, T; r_f, y, \sigma_S)$ by (i); convert at X_0. □

The *quanto adjustment* $-\rho\sigma_S\sigma_X$ lowers the drift of a foreign asset that tends to rise with the foreign currency. The correlation is that of the returns of S, in foreign currency, and of X, quoted as domestic per foreign. Quadratic covariation is the limit in probability of sums of products of increments, and limits in probability are the same under equivalent measures; since $\mathbb{P}$, $\mathbb{Q}$ and $\mathbb{Q}_f$ are equivalent, ρ is the same under all three, and it can be estimated from historical data; quoting the rate the other way round changes its sign. The adjustment does not vanish when $r_d = r_f$, and it is the only trace of the exchange rate in the price: the quanto does not depend on X_0, and σ_X enters only through $\rho\sigma_X$. Exercise 17.4 prices the remaining contract of the course's list, a currency option whose notional is a foreign share.

Example 17.9 (A quanto call)**.** Take a foreign index with $S_0 = K = 100$ GBP, $y = 2\%$ and $\sigma_S = 20\%$, the cable parameters of Theorem 17.2, and $\rho = -0.3$. Then $y_q = 1.84\%$, the quanto forward is 101.1668 against the foreign forward 100.5013, and the quanto call with $\bar{X} = 1$ USD per GBP costs 8.35441 USD; with $\rho = 0$ and $\rho = +0.3$ it would cost 7.99516 and 7.64664. The ordinary call costs 8.03523 GBP, that is 9.64228 USD at the spot rate (iv), and the composite call struck at 120 USD costs 9.79110 USD, with $\sigma_c = 19.72\%$ (iii). Our implementation, simulating (S_T, X_T) under $\mathbb{Q}$ with 2^{20} draws (`numpy default_rng`, seed 2026), gives 8.3457, 9.6313 and 9.7784, with standard errors between 0.013 and 0.016, all within one standard error of the formulas.

17.5 Basket options

A *basket call* pays $(A - K)^+$, where $A = \sum_i w_i S_{i,T}$ is a weighted sum of the prices at maturity in the model (17.3): an option on an index of a few stocks, or on a currency basket. It is the cross-sectional analogue of the Asian option of Chapter 15: there the summands are one asset at several dates, with covariances $\sigma^2 \min(t_i, t_j) \geq 0$; here they are several assets at one date, with the covariances $\Sigma_{ij} = \rho_{ij}\sigma_i\sigma_j T$ of any sign. The sum of lognormals is not lognormal, and with $d \geq 2$ assets no closed form is known, except in degenerate cases. Write $D = e^{-rT}$, $Y = D(A - K)^+$, $C = \mathbb{E}^{\mathbb{Q}} Y$ and $\Phi = \sum_i w_i F_i = \mathbb{E}^{\mathbb{Q}} A$.

Proposition 17.10 (Basket options). *(i) (*The geometric basket.*) Let all $w_i > 0$, $a_i = w_i F_i/\Phi$, and $G = \Phi \prod_i (S_{i,T}/F_i)^{a_i}$. Then $\log G \sim \mathcal{N}(\mu_G, v_G)$ with $\mu_G = \log\Phi - \frac{1}{2}\sum_i a_i \Sigma_{ii}$ and $v_G = a^\top \Sigma a$; $G \le A$ on every path; and $X^{\mathrm{g}} = D(G-K)^+$ has mean $D\,\mathrm{Black}(\mathbb{E}^{\mathbb{Q}}G, K, v_G) \le C$, where $\mathbb{E}^{\mathbb{Q}}G = e^{\mu_G + v_G/2} \le \Phi$.*

*(ii) (*The geometric control.*) Under (i), $0 \le Y - X^{\mathrm{g}} \le D(A-G)$ and, with $c = \Sigma a$,*

$$\begin{aligned}\operatorname{Var}(Y - X^{\mathrm{g}}) &\le D^2 \mathbb{E}^{\mathbb{Q}}\big[(A-G)^2\big] \\ &= D^2\Big(\sum_{i,j} w_i w_j F_i F_j e^{\Sigma_{ij}} - 2\,\mathbb{E}^{\mathbb{Q}}G \sum_i w_i F_i e^{c_i} + (\mathbb{E}^{\mathbb{Q}}G)^2 e^{v_G}\Big).\end{aligned}$$

If $v_G > 0$ and $\operatorname{Var} Y > 0$, the control-variate estimator with the optimal coefficient has variance $(1 - \operatorname{Corr}(Y, X^{\mathrm{g}})^2) \operatorname{Var} Y/n$ (Theorem 8.12), and $1 - \operatorname{Corr}(Y, X^{\mathrm{g}})^2 \le D^2\mathbb{E}^{\mathbb{Q}}[(A-G)^2]/\operatorname{Var} Y$, whose numerator does not depend on K.

*(iii) (*Moment matching.*) $\mathbb{E}^{\mathbb{Q}}A = \Phi$ and $\mathbb{E}^{\mathbb{Q}}A^2 = \sum_{i,j} w_i w_j F_i F_j e^{\Sigma_{ij}}$. If $\Phi > 0$, the lognormal law with these two moments has log-variance $v_2 = \log(\mathbb{E}^{\mathbb{Q}}A^2/\Phi^2)$, and the* two-moment approximation *is $C_2 = D\,\mathrm{Black}(\Phi, K, v_2)$. It equals C if A is lognormal (for instance if $d = 1$, or if $\rho_{ij} = 1$ and all σ_i are equal), and if all $w_i \ge 0$ and $K \le 0$. It is not a bound: in the three-asset example below, $C_2 - C$ changes sign with K (Table 17.1).*

*(iv) (*Mixed signs.*) If some weights are positive and some negative and Σ is positive definite, then $\mathbb{Q}(A < 0) > 0$. The geometric basket of (i)–(ii) is then undefined, and no lognormal law has the law of A: when $\Phi > 0$, C_2 can still be computed, but it puts no mass where $A < 0$, and none of the exact cases of (iii) applies; if $\Phi \le 0$, no lognormal law even has the mean of A. Plain Monte Carlo is unaffected.*

Proof. (i) $\log G = \log\Phi + \sum_i a_i(\log S_{i,T} - \log F_i)$ is a linear function of a Gaussian vector, with $\mathbb{E}(\log S_{i,T} - \log F_i) = -\frac{1}{2}\Sigma_{ii}$ and variance $a^\top\Sigma a$. The weights a_i are positive with sum one, and log is concave, so the weighted arithmetic–geometric mean inequality gives $\prod_i x_i^{a_i} \le \sum_i a_i x_i$ for $x_i = S_{i,T}/F_i$; multiplied by Φ, this is $G \le \sum_i w_i S_{i,T} = A$. The mean of X^{g} is the identity after (17.1) (or $D(\mathbb{E}^{\mathbb{Q}}G - K)^+$ if $v_G = 0$); it is at most C because $(G-K)^+ \le (A-K)^+$, and $\mathbb{E}^{\mathbb{Q}}G \le \mathbb{E}^{\mathbb{Q}}A$. (ii) $x \mapsto (x-K)^+$ is nondecreasing and 1-Lipschitz, so $0 \le (A-K)^+ - (G-K)^+ \le A - G$, and $\operatorname{Var}(Y - X^{\mathrm{g}}) \le \mathbb{E}^{\mathbb{Q}}[(Y - X^{\mathrm{g}})^2]$. For the moments, $\mathbb{E}^{\mathbb{Q}}[S_{i,T}S_{j,T}] = F_i F_j e^{\Sigma_{ij}}$, $\mathbb{E}^{\mathbb{Q}}G^2 = (\mathbb{E}^{\mathbb{Q}}G)^2 e^{v_G}$, and $\log S_{i,T} + \log G$ is normal with variance $\Sigma_{ii} + 2c_i + v_G$, where $c_i = \operatorname{Cov}(\log S_{i,T}, \log G) = (\Sigma a)_i$, so $\mathbb{E}^{\mathbb{Q}}[S_{i,T}G] = F_i\,\mathbb{E}^{\mathbb{Q}}G\,e^{c_i}$. If $v_G > 0$ then X^{g} has positive variance, and Theorem 8.12(ii) applies; its optimal variance is at most the variance at coefficient one. (iii) The moments follow as in (ii). A lognormal L with mean Φ has $\mathbb{E}L^2 = \Phi^2 e^{\operatorname{Var}\log L}$, which gives v_2. If A is lognormal, $C_2 = C$ by (17.1); under the stated condition all $S_{i,T}/F_i$ are equal to one lognormal variable, so $A = \Phi\, S_{1,T}/F_1$. If all $w_i \ge 0$ and $K \le 0$, then $A - K \ge 0$ and both sides equal $D(\Phi - K)$. The sign changes are in Table 17.1. (iv) The vector $(\log S_{i,T})$ has a positive Gaussian density on $\mathbb{R}^d$, so every open subset of $(0,\infty)^d$ has positive probability. With $w_1 > 0 > w_2$, say, the open set where $w_1 S_{1,T} + \sum_{i\ge 3} |w_i| S_{i,T} < |w_2| S_{2,T}$ is nonempty and lies in $\{A < 0\}$. A lognormal variable is positive, and $G > 0$. □

The geometric basket with the weights a_i, proportional to the forward values $w_i F_i$ rather than to w_i, is the choice of Gentle (1993), as it is usually described: A/Φ and G/Φ are then the arithmetic and geometric means of the same ratios $S_{i,T}/F_i$, which agree to first order when the

ratios are close to one. Equal weights $w_i/\sum_j w_j$ would give a geometric mean of the prices themselves, which is far below A when the prices differ in scale. Gentle's *approximation* prices $(G-K')^+$ with the strike $K'=K-(\Phi-\mathbb{E}^{\mathbb{Q}}G)$ shifted by the difference of the means; like C_2, which is due to Levy (1992), it is a heuristic, with no error bound, and it errs in both directions: it underprices in Theorem 17.12, but with three equicorrelated assets ($\rho_{ij}=0.9$, volatilities 5%, 5%, 80%, $T=2$, $K=60$, other inputs as in Theorem 17.12) it gives 38.265 against $C=37.343$, and $C_2=40.992$. The bound in (ii) does not depend on the strike, as in Theorem 15.4, but it depends on how dispersed the ratios $S_{i,T}/F_i$ are, and dispersion grows with low or negative correlations (Exercise 17.8).

Remark 17.11 (Conditioning, and a reference price). Write the log-prices as $m+LZ$ with $LL^\top=\Sigma$ and $Z\sim\mathcal{N}(0,I_d)$, and split $Z=e\,\xi+QZ'$, where e is the unit vector along $L^\top a$ and the columns of Q are an orthonormal basis of $e^\perp$, so that $\log G-\mu_G=\sqrt{v_G}\,\xi$, with $\xi\sim\mathcal{N}(0,1)$ independent of $Z'\in\mathbb{R}^{d-1}$. Given Z', $A=\sum_i\alpha_i e^{b_i\xi}$ with $\alpha_i>0$ and $b=Le=c/\sqrt{v_G}$; it is a convex function of ξ, so it exceeds K outside an interval, whose endpoints are found by bisection, and $\mathbb{E}^{\mathbb{Q}}[(A-K)^+\mid Z']$ is a sum of terms $e^{b_i^2/2}N(\cdot)$, as for the Asian options of Chapter 15. This conditional value is an unbiased estimator with no larger variance (Theorem 10.11) and a smooth function of Z' when all $b_i>0$. In the example below ($d=3$), a product Gauss–Hermite rule in Z' with 8×8 nodes already gives ten digits, and 16×16 nodes agree with 64×64 to 10^{-14}. Conditioning on G instead gives the lower bound $D\,\mathbb{E}^{\mathbb{Q}}[(\mathbb{E}^{\mathbb{Q}}[A\mid G]-K)^+]\le C$ by the conditional Jensen inequality, whatever the covariances, with $\mathbb{E}^{\mathbb{Q}}[A\mid\log G=x]=\sum_i w_iF_i\exp\left(c_i(x-\mu_G)/v_G-c_i^2/(2v_G)\right)$. If every $c_i\ge0$, this function is increasing in x, it crosses K once, and the bound has the closed form of Theorem 15.10(ii), as Curran (1994) proposed for "portfolio options".

Example 17.12 (A three-asset basket)**.** Take three assets with the test-case inputs $S_{i,0}=100$, $r=1\%$, $y_i=2\%$, $T=1$, volatilities 20%, 25% and 30%, the correlations 0.5, 0.3, 0.6 of Theorem 5.15, equal weights $w_i=\frac{1}{3}$ and $K=100$. Then $\Phi=99.0050$, $\sqrt{v_G}=20.23\%$ and $\sqrt{v_2}=20.30\%$, against volatilities averaging 25%: diversification. The reference price of Theorem 17.11 is $C=7.4655807$. Our implementation, with 2^{20} draws (`numpy default_rng`, seed 2026), gives 7.46603 ± 0.01256 by plain Monte Carlo; the geometric control, with the in-sample coefficient 1.0498 and correlation 0.99679, gives 7.46541 ± 0.00101, 12 times more precise; and three controls (the geometric call, the forward DA and the strip $D\sum_i w_i(S_{i,T}-K)^+$) give 7.46585 ± 0.00071, with $R^2=0.99678$ (Theorem 8.15). Before any simulation, (ii) bounds $\mathrm{Var}(Y-X^{\mathrm{g}})$ by 2.888, about 1.7% of $\mathrm{Var}\,Y\approx165.5$. The two-moment approximation gives $C_2=7.48004$, Gentle's 7.36489, the conditioning bound of Theorem 17.11 7.45406, and the geometric call, a lower bound, 6.87976.

K	80	90	100	110	120
reference C	20.132481	12.804691	7.465581	4.023289	2.027888
two moments C_2	+319.2	+319.7	+144.6	−68.7	−199.0
Gentle	−325.6	−700.3	−1006.9	−1085.2	−945.9
conditioning bound	−76.8	−105.0	−115.2	−110.5	−95.5
geometric call	−9647.6	−7881.5	−5858.2	−3991.3	−2517.6

Table 17.1: The three-asset basket call of Theorem 17.12: reference prices from Theorem 17.11, and the errors of the approximations and bounds, in units of 10^{-4}.

Table 17.1 and Figure 17.2 show the pattern of Chapter 15 with larger numbers. The control removes between 98.8% and 99.6% of the variance for strikes from 80 to 120, less as the option

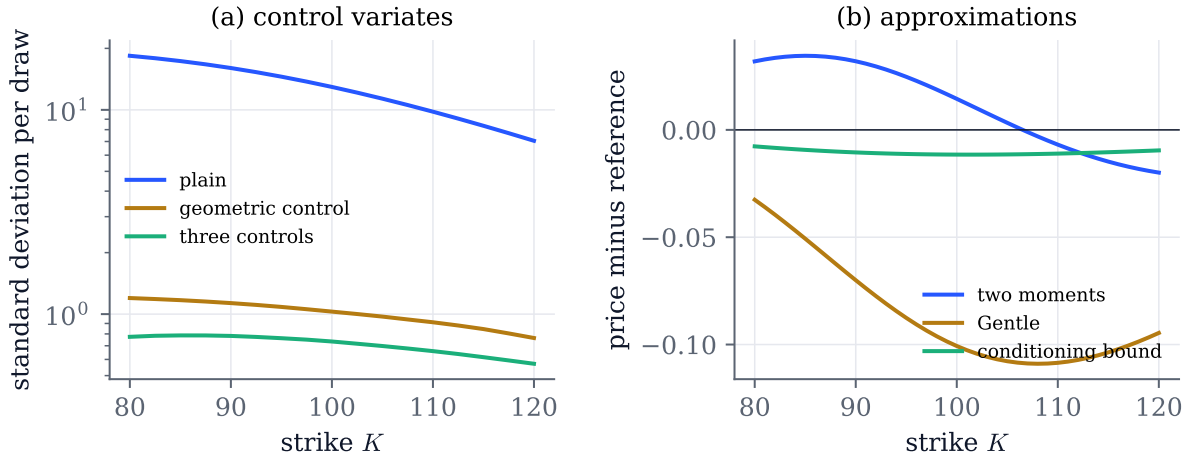

Figure 17.2: The three-asset basket call of Theorem 17.12 against the strike. (a) Standard deviation per draw: plain Monte Carlo, the geometric control and the three controls, all with in-sample coefficients, from the same 2^{16} draws (seed 17). (b) Errors relative to the reference price of the two-moment approximation, of Gentle's approximation and of the conditioning lower bound; the geometric call, 0.25 to 0.96 below the reference, is off the scale.

moves out of the money. Two-moment matching errs by up to 0.03 with both signs, overpricing in the money and underpricing out of it; Gentle's shift underprices by 0.03 to 0.11. For a basket the summands are less alike than the fixings of an average, so approximations that treat A as nearly lognormal, or nearly geometric, are less accurate, and the control is weaker: $1 - \rho^2$ is 6.4×10^{-3} here against 6.9×10^{-4} for the monthly Asian of Chapter 15. Beyond three or four assets, product quadrature in Z' becomes too expensive (Theorem 6.14), while Monte Carlo with the geometric control, or the conditional estimator of Theorem 17.11 with randomised quasi-Monte Carlo, keeps its cost per draw of order d^2.

▶ Interactive demo: the cable call in either currency, with simulation under the domestic and the foreign measures; the cross-rate triangle and exchange options against the correlation; quantos against the correlation; and the basket's control variates and approximations, with sliders for the volatilities, the correlations and the strike.

In code

code/src/mfe230d/multiasset.py: `black` is (17.1); `garman_kohlhagen`, `gk_hedge` and `currency_symmetry` implement Theorem 17.1 and Theorem 17.3(v), on top of code/src/mfe230d/bs.py; `numeraire_drift` and `call_samples_share_measure` Theorem 17.3(iii)–(iv) and Theorem 17.5; `terminal_prices` samples (17.3) with `rng.cholesky` or `rng.pca_factor`; `margrabe`, `exchange_vol`, `implied_correlation` and `cross_atmf_call` Theorem 17.6 and Theorem 17.7; `quanto_option`, `composite_option`, `foreign_option` and `fx_equity_terminal` Theorem 17.8; `geometric_basket_option`, `basket_cv_bound`, `basket_lognormal`, `gentle_call` and `basket_mc` (through `variance.control_variate` and `variance.regression_control_variates`) Theorem 17.10; `basket_reference` and `curran_lower_bound` Theorem 17.11.

Course examples reproduced. The ATMF cable call (0.05133 USD, 0.04278 GBP) and its two deltas (0.5090 GBP, −0.5595 USD); the London trader's thought experiment, with the same price and the same hedge in pounds, and put–call symmetry; the cross-rate volatility $\sigma_{\text{eff}}^2 = \sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2$, the Frankfurt trader's thought experiment and the triangle of admissible volatilities; the Cholesky recipe for correlated normals (via Theorem 5.15); and the course's "pre-washed" inputs for quantos, composites and foreign options.

Further reading

Shreve (2004), Chapter 9, treats numéraires, the domestic and foreign risk-neutral measures, Siegel's paradox (Section 9.3.4) and put–call duality, and Geman et al. (1995) is the original account of the change of numéraire. Garman and Kohlhagen (1983) and Margrabe (1978) are short and readable; Stulz (1982) prices options on the maximum or the minimum of two assets with the bivariate normal distribution. Glasserman (2003) samples multivariate normals in Section 2.3.3 and correlated geometric Brownian motions in Section 3.2.3, and treats control variates in Section 4.1. Levy (1992), Gentle (1993) and Curran (1994) give the basket approximations of Section 17.5.

Part IV

Beyond Black–Scholes

Chapter 18

Hedging in practice

The Black–Scholes price is the initial wealth of a strategy that replicates the claim exactly, under three idealisations: the hedger trades continuously, knows the volatility, and trades for free. A dealer who sells an option and hedges it gives up all three, and this chapter measures what each one costs. Rebalancing at m dates leaves an error whose standard deviation falls like $m^{-1/2}$, with a constant set by the gamma of the claim (Section 18.1). Hedging with the wrong volatility leaves an error that does not vanish however often one trades: the gamma-weighted gap between the hedging variance and the realised variance, whose sign is known for convex payoffs (Section 18.2). Proportional transaction costs make frequent trading expensive; Leland's enlarged volatility pays for them on average, and its limits are now precisely understood (Section 18.3). The laboratory is the course's hedging homework, posed there with $S = K = 3900$, $\sigma = 25\%$, $r = 3\%$, $y = 1.3\%$ and 100,000 paths, and rerun here on the test case with 50,000 paths.

Setting. The market is the standard model of Theorem 1.1: under $\mathbb{P}$, $\mathrm{d}S_t = (\mu - y)S_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}W_t$, where μ is the total expected return, and $\theta = (\mu - r)/\sigma$; the risk-neutral measure $\mathbb{Q}$ is the case $\mu = r$. A European claim pays $F(S_T)$ and is worth $V(t, S_t)$ at time t, with V given by Theorem 1.3; for $t < T$ we write $\Delta = \partial_s V$ and $\Gamma = \partial_{ss} V$ (for calls and puts, Theorem 2.7), and $\tau = T - t$. One share bought at time 0, with its dividends reinvested in the stock, is worth $e^{yt}S_t$ at time t; its discounted value

$$\tilde{S}_t = e^{-(r-y)t}S_t \quad \text{satisfies} \quad \mathrm{d}\tilde{S}_t = \tilde{S}_t\big((\mu - r)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t\big) = \sigma\tilde{S}_t\,\mathrm{d}W_t^{\mathbb{Q}}. \tag{18.1}$$

Unless stated otherwise, F is *Lipschitz*: $|F(a) - F(b)| \le L|a - b|$. Calls, puts and call spreads qualify; digital options do not. Then, as in the proof of Theorem 8.16(ii), $V(t,s) = e^{-r\tau}\int F(e^u)\,p_\tau(u - \log s)\,\mathrm{d}u$ with p_τ the $\mathcal{N}(\nu\tau, \sigma^2\tau)$ density, so V is infinitely differentiable on $[0,T) \times (0,\infty)$ (the derivatives fall on the Gaussian kernel) and solves the Black–Scholes equation there; $|\Delta| \le Le^{-y\tau}$; V is continuous on $[0,T] \times (0,\infty)$ with $V(T,\cdot) = F$; and, with $H_t = e^{-yt}\Delta(t, S_t)$,

$$Y := e^{-rT}F(S_T) = V(0,S_0) + \int_0^T H_t\,\mathrm{d}\tilde{S}_t \qquad \text{almost surely.} \tag{18.2}$$

Theorem 8.16 proves this under $\mathbb{Q}$; it holds $\mathbb{P}$-almost surely because the two measures have the same null sets, and the stochastic integral is the same under both (the remark after (1.7)).

18.1 Discrete delta hedging

Fix $m \geq 1$, let $h = T/m$, $t_j = jh$, and let $\eta(t) = t_j$ and $\eta^+(t) = t_{j+1}$ for $t \in [t_j, t_{j+1})$. The seller receives $V_0 = V(0, S_0)$, buys $\delta_j = \Delta(t_j, S_{t_j})$ shares at t_j, reinvests the dividends in the stock as they are paid, so that she holds $\delta_j e^{y(t-t_j)}$ shares at $t \in [t_j, t_{j+1}]$, and keeps the rest of her wealth in the money market; this is the convention of the course's homework and of Chapter 8. Her wealth at $t \in [t_j, t_{j+1}]$ is $\delta_j e^{y(t-t_j)} S_t + e^{r(t-t_j)}(X_{t_j} - \delta_j S_{t_j})$, so $e^{-rt}X_t - e^{-rt_j}X_{t_j} = e^{-yt_j}\delta_j(\tilde{S}_t - \tilde{S}_{t_j})$, and her *discounted hedging error* is

$$e_m := e^{-rT}\big(X_T - F(S_T)\big) = V_0 + M_m - Y, \qquad M_m = \sum_{j=0}^{m-1} e^{-yt_j}\delta_j(\tilde{S}_{t_{j+1}} - \tilde{S}_{t_j}). \tag{18.3}$$

M_m is the control variate of Theorem 8.16(ii). Since $M_m = \int_0^T H_{\eta(t)}\,\mathrm{d}\tilde{S}_t$, (18.2) gives

$$e_m = \int_0^T (H_{\eta(t)} - H_t)\,\mathrm{d}\tilde{S}_t. \tag{18.4}$$

The error is the stochastic integral of the drift of the hedge ratio within each period. Two lemmas describe that drift and average it.

Lemma 18.1 (The dynamics of the hedge)**.** *For $t < T$, $H_t = e^{-yt}\Delta(t, S_t)$ satisfies*

$$\mathrm{d}H_t = b_t(\mathrm{d}W_t^{\mathbb{Q}} - \sigma\,\mathrm{d}t) = b_t(\mathrm{d}W_t + (\theta - \sigma)\,\mathrm{d}t), \qquad b_t = \sigma e^{-yt} S_t\,\Gamma(t, S_t).$$

Proof. Differentiating $\partial_t V + \frac{1}{2}\sigma^2 s^2 \partial_{ss} V + (r-y)s\,\partial_s V - rV = 0$ in s gives $\partial_t\Delta + \frac{1}{2}\sigma^2 s^2 \partial_s\Gamma + (r - y + \sigma^2)s\Gamma - y\Delta = 0$. By Itô's formula under $\mathbb{Q}$, where $\mathrm{d}S = (r-y)S\,\mathrm{d}t + \sigma S\,\mathrm{d}W^{\mathbb{Q}}$, $\mathrm{d}H = e^{-yt}(\partial_t\Delta - y\Delta + (r-y)S\Gamma + \frac{1}{2}\sigma^2 S^2\partial_s\Gamma)\mathrm{d}t + e^{-yt}\sigma S\Gamma\,\mathrm{d}W^{\mathbb{Q}} = -\sigma b\,\mathrm{d}t + b\,\mathrm{d}W^{\mathbb{Q}}$. Finally $W_t^{\mathbb{Q}} = W_t + \theta t$. □

For the call, $H_t = e^{-yT}N(d_+)$; the lemma says that it drifts downwards under $\mathbb{Q}$, at a rate proportional to the gamma.

Lemma 18.2 (Averaging a sawtooth)**.** *If φ is integrable on $(0, T)$, then $\frac{1}{h}\int_0^T \varphi(u)(\eta^+(u) - u)\,\mathrm{d}u \to \frac{1}{2}\int_0^T \varphi(u)\,\mathrm{d}u$ as $m \to \infty$.*

Proof. For $\varphi = \mathbf{1}_{[a,b)}$, each period inside $[a, b)$ contributes $\int_{t_j}^{t_{j+1}}(t_{j+1} - u)\,\mathrm{d}u = h^2/2$, there are $(b-a)/h + O(1)$ of them, and the at most two periods that meet $[a, b)$ only partly contribute at most h^2 each; so the claim holds, with an error $O(h)$, and by linearity it holds for step functions. Continuous functions are dense in $L^1(0, T)$ (Rudin (1987), Theorem 3.14), and each is a uniform limit of step functions, so for $\varepsilon > 0$ there is a step function φ_ε with $\|\varphi - \varphi_\varepsilon\|_1 < \varepsilon$. Since $0 \leq \eta^+(u) - u \leq h$, replacing φ by φ_ε changes the left side by at most ε and the right side by at most $\varepsilon/2$. □

Proposition 18.3 (The error of discrete delta hedging)**.** *Let F be Lipschitz, and assume the integrability condition*

$$I_\mu := \int_0^T \varphi_\mu(t)\,\mathrm{d}t < \infty, \qquad \varphi_\mu(t) = \mathbb{E}\big[e^{-2rt}S_t^4\,\Gamma(t, S_t)^2\big], \tag{18.5}$$

the expectation being under $\mathbb{P}$, with drift μ. There is a constant C, independent of m (it depends on μ, through θ and I_μ), such that:

(i) $\mathbb{E}^{\mathbb{Q}}[e_m] = 0$ *for every* m, *and* $|\mathbb{E}[e_m]| \le C(\mu - r)^2 h$;

(ii) as $m \to \infty$,

$$\mathbb{E}[e_m^2] = \frac{\sigma^4 h}{2} I_\mu + o(h) = 2h \int_0^T \mathbb{E}\Big[e^{-2rt}\big(\tfrac{1}{2}\sigma^2 S_t^2 \Gamma(t, S_t)\big)^2\Big] \mathrm{d}t + o(h); \tag{18.6}$$

(iii) for the call and the put, $\varphi_\mu(t) \le C'(T-t)^{-1/2}$ *on* $[0, T)$ *for some constant* C', *so* (18.5) *holds for every* μ.

Proof. Let $G_t = H_{\eta(t)} - H_t$; $|G_t| \le 2Le^{-yT}$ because $|H_t| \le e^{-yt} L e^{-y(T-t)}$. By (18.4) and (18.1), $e_m = A_m + (\mu - r)B_m$ with $A_m = \int_0^T G_t \sigma \tilde{S}_t \, \mathrm{d}W_t$ and $B_m = \int_0^T G_t \tilde{S}_t \, \mathrm{d}t$. Note that $\tilde{S}_t^2 b_t^2 = \sigma^2 e^{-2rt} S_t^4 \Gamma(t, S_t)^2$, so $\mathbb{E}[\tilde{S}_t^2 b_t^2] = \sigma^2 \varphi_\mu(t)$.

The martingale part. By the Itô isometry, $\mathbb{E}[A_m^2] = \int_0^T \sigma^2 \mathbb{E}[G_t^2 \tilde{S}_t^2] \, \mathrm{d}t$. Let $\hat{\mathbb{P}}$ have density $\exp(2\sigma W_T - 2\sigma^2 T)$ with respect to $\mathbb{P}$ on $\mathcal{F}_T$. By Girsanov's theorem $\hat{W}_t = W_t - 2\sigma t$ is a $\hat{\mathbb{P}}$-Brownian motion, and since $\tilde{S}_t^2 = \tilde{S}_0^2 e^{ct} \exp(2\sigma W_t - 2\sigma^2 t)$ with $c = 2(\mu - r) + \sigma^2$, $\mathbb{E}[X \tilde{S}_t^2] = \tilde{S}_0^2 e^{ct} \hat{\mathbb{E}}[X]$ for every $\mathcal{F}_t$-measurable $X \ge 0$. By Theorem 18.1, for $t_j \le t < T$,

$$H_t - H_{t_j} = \underbrace{\int_{t_j}^t b_u \, \mathrm{d}\hat{W}_u}_{=: \mathcal{I}} + \underbrace{(\sigma + \theta) \int_{t_j}^t b_u \, \mathrm{d}u}_{=: \mathcal{J}},$$

where $\hat{\mathbb{E}} \int_0^T b_u^2 \, \mathrm{d}u = \tilde{S}_0^{-2} \int_0^T e^{-cu} \sigma^2 \varphi_\mu(u) \, \mathrm{d}u < \infty$ by (18.5). So $\hat{\mathbb{E}}[\mathcal{I}^2] = \int_{t_j}^t \hat{\mathbb{E}}[b_u^2] \, \mathrm{d}u$, and by the Cauchy–Schwarz inequality $\hat{\mathbb{E}}[\mathcal{J}^2] \le (\sigma + |\theta|)^2 h \, \hat{\mathbb{E}}[\mathcal{I}^2]$. Since $|\hat{\mathbb{E}}(\mathcal{I} + \mathcal{J})^2 - \hat{\mathbb{E}}\mathcal{I}^2| \le 2\|\mathcal{I}\| \|\mathcal{J}\| + \|\mathcal{J}\|^2$, $\hat{\mathbb{E}}[G_t^2]$ lies between $1 - \varepsilon_h$ and $1 + \varepsilon_h$ times $\hat{\mathbb{E}}[\mathcal{I}^2]$, with $\varepsilon_h = 2(\sigma + |\theta|)\sqrt{h} + (\sigma + |\theta|)^2 h$. Integrating, exchanging the order of integration (Fubini–Tonelli) and using $\sigma^2 \tilde{S}_0^2 e^{cu} \hat{\mathbb{E}}[b_u^2] = \sigma^2 \mathbb{E}[\tilde{S}_u^2 b_u^2] = \sigma^4 \varphi_\mu(u)$,

$$\mathbb{E}[A_m^2] = (1 + O(\varepsilon_h)) \int_0^T \sigma^4 \varphi_\mu(u) \, \kappa(u) \, \mathrm{d}u, \qquad \kappa(u) = \int_u^{\eta^+(u)} e^{c(t-u)} \, \mathrm{d}t.$$

As $e^{-|c|h}(\eta^+(u) - u) \le \kappa(u) \le e^{|c|h}(\eta^+(u) - u)$, Theorem 18.2 gives $\mathbb{E}[A_m^2] = \frac{\sigma^4 h}{2} I_\mu + o(h)$.

The drift part. Write $B_m = \sum_j D_j$ with $D_j = \int_{t_j}^{t_{j+1}} G_t \tilde{S}_t \, \mathrm{d}t$ and $\mathcal{F}_j = \mathcal{F}_{t_j}$. The terms $D_j - \mathbb{E}[D_j \mid \mathcal{F}_j]$ are orthogonal, and by the Cauchy–Schwarz inequality $\mathbb{E}[D_j^2] \le h \int_{t_j}^{t_{j+1}} \mathbb{E}[G_t^2 \tilde{S}_t^2] \, \mathrm{d}t$, so their sum has second moment at most $h \, \mathbb{E}[A_m^2]/\sigma^2 = O(h^2)$. For the conditional means, the product rule, Theorem 18.1 and (18.1) give, for $t_j \le u < T$,

$$\mathrm{d}\big((H_u - H_{t_j})\tilde{S}_u\big) = \theta(\tilde{S}_u b_u - \sigma G_u \tilde{S}_u)\mathrm{d}u + (\tilde{S}_u b_u - \sigma G_u \tilde{S}_u)\mathrm{d}W_u,$$

and the $\mathrm{d}W$-integral has conditional mean zero (its integrand is square integrable by (18.5) and G is bounded). Hence, with $Z_u = \tilde{S}_u |b_u| + \sigma |G_u| \tilde{S}_u$, $|\mathbb{E}[D_j \mid \mathcal{F}_j]| \le |\theta| h \, \mathbb{E}\big[\int_{t_j}^{t_{j+1}} Z_u \, \mathrm{d}u \,\big|\, \mathcal{F}_j\big]$, and by the triangle and Cauchy–Schwarz inequalities

$$\Big\| \sum_j \mathbb{E}[D_j \mid \mathcal{F}_j] \Big\|_2 \le |\theta| h \sum_j \Big(h \int_{t_j}^{t_{j+1}} \mathbb{E} Z_u^2 \, \mathrm{d}u \Big)^{1/2} \le |\theta| h \Big(T \int_0^T \mathbb{E} Z_u^2 \, \mathrm{d}u \Big)^{1/2} = O(h),$$

because $\int_0^T \mathbb{E}Z_u^2\,\mathrm{d}u \le 2\sigma^2 I_\mu + 2\mathbb{E}[A_m^2]$ is bounded. So $\|B_m\|_2 = O(h)$ and $|\mathbb{E}B_m| \le C_1|\theta|h$.

Conclusion. (i) $\mathbb{E}[A_m] = 0$, so $\mathbb{E}[e_m] = (\mu - r)\mathbb{E}[B_m]$, of size at most $C_1(\mu - r)^2 h/\sigma$, where C_1 depends on μ only through θ, I_μ and $\sup_m \mathbb{E}[A_m^2] \le 4L^2\sigma^2 \int_0^T \mathbb{E}\tilde{S}_t^2\,\mathrm{d}t$; for calls and puts the bound of (iii) shows that C_1 stays bounded for μ in bounded sets. Under $\mathbb{Q}$, $\mu = r$. (ii) $\|e_m - A_m\|_2 = |\mu - r|\|B_m\|_2 = O(h)$, so $|\mathbb{E}e_m^2 - \mathbb{E}A_m^2| \le \|e_m - A_m\|_2(\|e_m\|_2 + \|A_m\|_2) = O(h^{3/2})$; the second form of (18.6) is the identity $\sigma^4 S^4\Gamma^2 = 4(\frac{1}{2}\sigma^2 S^2\Gamma)^2$. (iii) By Theorem 2.7, $S^4\Gamma^2 = S^2 e^{-2y\tau}N'(d_+)^2/(\sigma^2\tau) \le e^{2|y|T}S^2/(2\pi\sigma^2\tau)$. On $[0, T/2]$, $\tau \ge T/2$ and φ_μ is bounded. On $[T/2, T)$, $\log S_t \sim \mathcal{N}(m_t, v_t)$ with $v_t = \sigma^2 t \ge \sigma^2 T/2$; the identity $e^{2x}\frac{e^{-(x-m_t)^2/2v_t}}{\sqrt{2\pi v_t}} = e^{2m_t+2v_t}\frac{e^{-(x-m_t-2v_t)^2/2v_t}}{\sqrt{2\pi v_t}}$ bounds the density of $\log S_t$, weighted by e^{2x}, by $e^{2m_t+2v_t}/\sqrt{2\pi v_t}$, and $N'(d_+)^2 = e^{-(x-c_t)^2/\sigma^2\tau}/(2\pi)$, with $x = \log S_t$ and c_t deterministic, integrates to $\sigma\sqrt{\pi\tau}/(2\pi)$. So $\mathbb{E}[S_t^2 N'(d_+)^2] \le C\sqrt{\tau}$ and $\varphi_\mu(t) \le C'/\sqrt{\tau}$. □

Reading the proposition. To halve the error, rebalance four times as often. The constant is the time integral of the squared "cash gamma" where the stock is likely to be: near expiry, $S^4\Gamma^2$ is of order $1/\tau$ within a distance of order $\sigma\sqrt{\tau}$ of the strike and negligible elsewhere, and the stock is there with probability of order $\sqrt{\tau}$, which turns the blow-up into the integrable $\tau^{-1/2}$ of (iii). Completing the square in the Gaussian integral of the proof gives φ_μ in closed form (`code/src/mfe230d/hedging.py`, `gamma_moment`); for the test case, $I_r = 60{,}391.4$, and the leading standard deviation of the error, in time-T money, is $e^{rT}\sigma^2\sqrt{TI_r/2}\,m^{-1/2} = 7.02\,m^{-1/2}$: 0.97 for weekly and 0.44 for daily rebalancing, against a premium of 7.36. Condition (18.5) is not a technicality. It fails for the digital option, whose gamma is of order $1/\tau$ at a distance of order $\sigma\sqrt{\tau}$ from the strike (and of order $\tau^{-1/2}$ at the strike itself; Exercise 18.2); there the mean squared error is only of order $h^{1/2}$ (Gobet and Temam, 2001, with the correction note of Gobet, 2014).

The expansion is classical. For calls and puts, Bertsimas et al. (2000) show that $\sqrt{m}\,e_m$ converges in law to a mixture of centred normal laws, described by the same integrand; Toft (1996) gives closed forms, for a finite number of dates, for the expected hedging error and the variance of the cash flows of a discretely rebalanced option hedge, with a drift $\mu \ne r$ and with a cost-adjusted hedging volatility; and Zhang (1999) proves that $m\,\mathbb{E}[e_m^2]$ converges to a nonzero limit for calls and puts. The IEOR 222 notes derive the same constant for smooth payoffs from a gamma–theta Taylor expansion; that route needs bounded higher derivatives, which the call does not have, and the change to $\hat{\mathbb{P}}$ above avoids it.

How accurate is the leading term? Under $\mathbb{Q}$ the variance can be computed exactly, by one-dimensional quadrature (Exercise 18.8). For the test-case call it gives standard deviations 6.040, 3.223, 1.925, 0.949 and 0.437 for $m = 1, 4, 12, 52, 252$, so that $\sqrt{m}$ times the standard deviation increases towards 7.02 from below (Figure 18.1(b)). Divided by the standard deviation 12.570 of the discounted payoff, the exact values for $m = 1$, 12 and 52 are 0.4757, 0.1516 and 0.0748: the standard-error ratios of the hedge used as a control variate in Chapter 8, simulated there as 0.48, 0.15 and 0.075. The leading term predicts 0.553, 0.160 and 0.0767: it overstates the error by 16% for a static hedge ($m = 1$) and by 2.6% weekly.

Table 18.1 is the course's experiment: n paths, no hedge, a static hedge, and quarterly, monthly, weekly and daily rebalancing, under $\mathbb{Q}$ and under three real-world drifts. Under $\mathbb{Q}$ the mean is

	$\mu = r$ (under $\mathbb{Q}$)					mean / sd under $\mathbb{P}$		
m	mean	sd	exact sd	leading	MAE	$\mu = 10\%$	$\mu = 20\%$	$\mu = -10\%$
none	0.010	12.651	12.696	–	9.342	−5.667 / 16.830	−14.453 / 21.352	4.286 / 8.013
1	−0.000	6.017	6.040	7.021	4.749	−0.830 / 7.238	−3.694 / 9.462	−1.086 / 6.041
4	0.010	3.207	3.223	3.510	2.459	−0.194 / 3.367	−0.855 / 3.490	−0.257 / 3.046
12	−0.001	1.942	1.925	2.027	1.466	−0.062 / 1.965	−0.266 / 1.911	−0.097 / 1.801
52	−0.001	0.956	0.949	0.974	0.715	−0.017 / 0.954	−0.064 / 0.903	−0.019 / 0.888
252	−0.004	0.437	0.437	0.442	0.326	−0.006 / 0.436	−0.014 / 0.410	−0.008 / 0.409

Table 18.1: The course's discrete-hedging experiment on the test case: the seller's error $X_T - (S_T - K)^+$ in time-T money, for no hedge and for m rebalancing dates. Our implementation (numpy `default_rng`, seed 2026, $n = 50{,}000$ paths, the same normals in every column). "exact sd" is computed by quadrature (Exercise 18.8), "leading" is $e^{rT}\sqrt{\sigma^4 h I_r/2}$. The standard errors of the means are $\mathrm{sd}/\sqrt{n}$: at most 0.096, and 0.002 for $m = 252$ under $\mathbb{Q}$; the means under $\mathbb{Q}$ are within 2.3 standard errors of 0.

exactly zero, by (i), and the simulated means are consistent with it: the largest, −0.0043 for $m = 252$, is 2.2 standard errors from zero, and among six rows one such deviation occurs with probability about 0.15. The standard deviations match the exact ones; the mean absolute error is about 0.75 of the standard deviation (it would be 0.80 for a normal law), a first sign of the heavy tails visible in Figure 18.1(a), which come from the paths that end near the strike. Without a hedge the seller carries the whole risk, and under $\mathbb{P}$ the mean is minus the risk premium of the call. The drift enters the hedged error in two ways, both small. Its mean is of order $(\mu - r)^2 h$ and, to leading order, negative (Exercise 18.5), and its variance changes through I_μ: the limits of $\sqrt{m}$ times the standard deviation are 7.02, 7.00, 6.56 and 6.55 for $\mu = r$, 10%, 20% and −10%, because a strong drift in either direction carries the stock away from the strike, where the gamma is. The hedger does not need to know μ.

Remark 18.4 (Dividends paid into the money market). If the dividends are paid into the money market and the seller holds exactly δ_j shares on $[t_j, t_{j+1})$, as in (1.6), her discounted gain over the period is $\delta_j \int_{t_j}^{t_{j+1}} e^{-yu}\,\mathrm{d}\tilde{S}_u$ instead of $e^{-yt_j}\delta_j(\tilde{S}_{t_{j+1}} - \tilde{S}_{t_j})$. The difference $\int_0^T \delta_{\eta(u)}(e^{-yu} - e^{-y\eta(u)})\,\mathrm{d}\tilde{S}_u$ has an integrand bounded by $L|y|e^{|y|T}h$, hence an L^2 norm $O(h)$, so Theorem 18.3(ii) is unchanged. The mean under $\mathbb{P}$ is no longer of order $(\mu - r)^2 h$: the extra term has mean $(\mu - r)\mathbb{E}\int_0^T \delta_{\eta(u)}(e^{-yu} - e^{-y\eta(u)})\tilde{S}_u\,\mathrm{d}u$, linear in $\mu - r$, and (i) holds with $(\mu - r)^2$ replaced by $|\mu - r|(|y| + |\mu - r|)$; the mean is still 0 under $\mathbb{Q}$.

▶ Interactive demo: rebalance the call hedge at m dates of your choice, under any drift, volatility and strike; watch the error distribution narrow like $m^{-1/2}$ and its constant change with the drift and the moneyness.

18.2 Hedging at the wrong volatility

The volatility of the hedge must be chosen before the volatility of the stock is known. In this section the stock follows

$$\mathrm{d}S_t = (\mu_t - y)S_t\,\mathrm{d}t + \sigma_t S_t\,\mathrm{d}W_t,$$

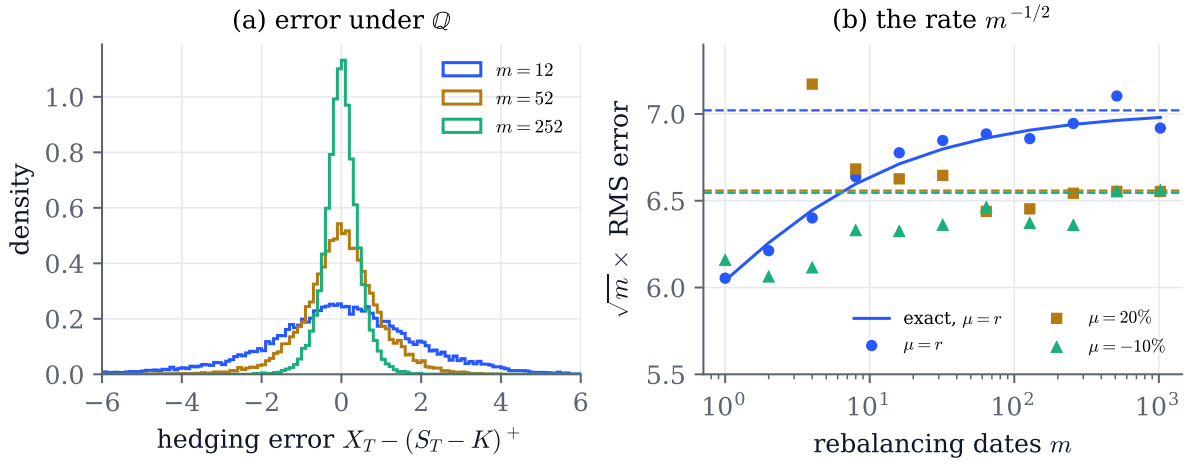

Figure 18.1: Discrete delta hedging of the test-case call. (a) Histograms of the seller's error $X_T - (S_T - K)^+$ under $\mathbb{Q}$ for $m = 12$, 52 and 252 (20,000 paths, seed 1812). (b) $\sqrt{m}$ times the root-mean-square error against m: exact values under $\mathbb{Q}$ (line, by quadrature), and simulations (10,000 paths for each m, seed $18 + m$) for $\mu = r$, 20% and -10%; the dashed lines are the limits $e^{rT}\sigma^2\sqrt{TI_\mu/2}$ of Theorem 18.3, 7.02 for $\mu = r$ and, almost on top of each other, 6.56 and 6.55 for $\mu = 20\%$ and -10%.

where μ_t and σ_t are progressively measurable, $|\mu_t| \le \bar{\mu}$ and $0 \le \sigma_t \le \bar{\sigma}$: the realised volatility may be random, depend on the path, and even vanish for a while. The dealer uses a constant *hedging volatility* $\sigma_h > 0$: V_h is the Black–Scholes price function at volatility σ_h (with the same r and y), and Δ_h, Γ_h its delta and gamma. She sells the claim for $V_h(0, S_0)$, holds $\Delta_h(t, S_t)$ shares at time t, and her wealth follows (1.6): $\mathrm{d}X_t = \Delta_h\,\mathrm{d}S_t + y\Delta_h S_t\,\mathrm{d}t + r(X_t - \Delta_h S_t)\,\mathrm{d}t$, $X_0 = V_h(0, S_0)$.

Proposition 18.5 (The P&L of hedging at the wrong volatility)**.** *Let F be Lipschitz. Then $s^2|\Gamma_h(t,s)| \le C_h s(1 + (T-t)^{-1/2})$ for a constant C_h, and almost surely*

$$e^{-rT}\big(X_T - F(S_T)\big) = \frac{1}{2}\int_0^T e^{-rt}\,\Gamma_h(t, S_t)\,S_t^2\,(\sigma_h^2 - \sigma_t^2)\,\mathrm{d}t, \tag{18.7}$$

the integral converging absolutely. A buyer who pays $V_h(0, S_0)$ and holds $-\Delta_h$ shares ends with the opposite, $\frac{1}{2}\int_0^T e^{-rt}\Gamma_h S_t^2(\sigma_t^2 - \sigma_h^2)\,\mathrm{d}t$.

Proof. The bound. Write $R = e^{\nu_h\tau + \sigma_h\sqrt{\tau}Z}$, $\nu_h = r - y - \frac{1}{2}\sigma_h^2$, $Z \sim \mathcal{N}(0,1)$. F is differentiable almost everywhere with $|F'| \le L$, and $\Delta_h(t,s) = e^{-r\tau}\mathbb{E}[F'(sR)R]$. Substituting $u = \log s + \nu_h\tau + \sigma_h\sqrt{\tau}z$ in this Gaussian integral and differentiating the kernel in s gives $s\,\Gamma_h(t,s) = e^{-r\tau}\mathbb{E}\big[F'(sR)R\,(Z/(\sigma_h\sqrt{\tau}) - 1)\big]$, so $s^2|\Gamma_h| \le sLe^{-r\tau}\mathbb{E}[R(1 + |Z|/(\sigma_h\sqrt{\tau}))]$, which has the stated form. (For the call, $s^2\Gamma_h = se^{-y\tau}N'(d_+)/(\sigma_h\sqrt{\tau}) \le s/(\sigma_h\sqrt{2\pi\tau})$.)

The identity. Fix $T' < T$. On $[0, T']$, V_h is $C^{1,2}$ and solves $\partial_t V_h - rV_h = -(r-y)s\Delta_h - \frac{1}{2}\sigma_h^2 s^2\Gamma_h$. By Itô's formula,

$$\mathrm{d}\big(e^{-rt}V_h(t,S_t)\big) = e^{-rt}\Big((\mu_t - r)S_t\Delta_h + \tfrac{1}{2}(\sigma_t^2 - \sigma_h^2)S_t^2\Gamma_h\Big)\mathrm{d}t + e^{-rt}\Delta_h\sigma_t S_t\,\mathrm{d}W_t,$$

while $\mathrm{d}(e^{-rt}X_t) = e^{-rt}\Delta_h(\mathrm{d}S_t - (r-y)S_t\,\mathrm{d}t) = e^{-rt}\Delta_h S_t((\mu_t - r)\,\mathrm{d}t + \sigma_t\,\mathrm{d}W_t)$. Subtracting, the $\mathrm{d}W$ terms and the drift cancel, and $e^{-rT'}(X_{T'} - V_h(T', S_{T'})) = \frac{1}{2}\int_0^{T'} e^{-rt}\Gamma_h S_t^2(\sigma_h^2 - \sigma_t^2)\,\mathrm{d}t$. Let $T' \uparrow T$: X is continuous, $V_h(T', S_{T'}) \to F(S_T)$ by the continuity of V_h on $[0,T]\times(0,\infty)$, and the integrand is bounded by $C_h(\sigma_h^2 + \bar\sigma^2)e^{|r|T}\sup_{u\le T} S_u\,(1 + (T-t)^{-1/2})$, which is integrable on $[0,T]$ along every path. □

No probability is involved: (18.7) holds path by path, whatever the drift. The hedger's P&L accrues at the rate $\frac{1}{2}\Gamma_h S^2$, the "cash gamma", times the spread between the variance she charged and the variance the stock realises. With $\sigma_t \equiv \sigma_h$ the hedge replicates, as it should.

Corollary 18.6 (Robustness)**.** *Let F be Lipschitz and $\sigma_h > 0$. The following are equivalent:*

(a) F is convex;

(b) for every model as above with $\sigma_t \le \sigma_h$ for all t, the dealer superhedges: $X_T \ge F(S_T)$ almost surely.

Under (a), $\sigma_t \ge \sigma_h$ for all t implies $X_T \le F(S_T)$ almost surely.

Proof. (a)⇒(b) and the last claim: for each z, $s \mapsto F(se^{\nu_h\tau + \sigma_h\sqrt{\tau}z})$ is convex, so $V_h(t,\cdot)$, an average of such functions, is convex and $\Gamma_h \ge 0$; apply (18.7). (b)⇒(a): if every $V_h(t,\cdot)$, $t<T$, were convex, so would be their pointwise limit F. So if F is not convex, $\Gamma_h(t_0,s_0) < 0$ for some $t_0 \in (0,T)$ and s_0 (a negative value at $t=0$ persists for small $t>0$, by continuity), and by continuity $\Gamma_h < 0$ on $[t_0, t_0+\varepsilon]\times[s_0e^{-a}, s_0e^{a}]$ for some $a>0$ and every small enough $\varepsilon > 0$; take ε so small that also $t_0 + \varepsilon < T$ and $|r-y|\varepsilon < a/2$. Consider the model with $\mu_t = r$, $\sigma_t = 0$ on $[t_0, t_0+\varepsilon]$ and $\sigma_t = \sigma_h$ elsewhere. On the event $\{S_{t_0} \in [s_0e^{-a/2}, s_0e^{a/2}]\}$, which has positive probability because $\sigma_t = \sigma_h > 0$ on $[0,t_0)$ gives S_{t_0} a positive density, the stock moves deterministically in $[s_0e^{-a}, s_0e^{a}]$ during $[t_0, t_0+\varepsilon]$, and the right side of (18.7) reduces to $\frac{1}{2}\sigma_h^2\int_{t_0}^{t_0+\varepsilon} e^{-rt}\Gamma_h S_t^2\,\mathrm{d}t < 0$. □

This is the robustness of the Black–Scholes formula of El Karoui et al. (1998): a dealer who overestimates the volatility of a convex claim is safe, whatever the realised volatility does, as long as the hedger's model keeps the price convex in the spot. Their Theorem 6.2 is the superhedging statement. In the Black–Scholes model with constant σ_h convexity is automatic; their Theorem 5.2 proves it for diffusion models whose volatility depends on time and the spot, and their Section 4 shows that it can fail when the model volatility depends on the past of the path. Hobson (1998) gives a proof by coupling. The identity (18.7) itself is the formula that traders know as the P&L of a delta-hedged option (Carr, 1997; Ellersgaard et al., 2017).

Degenerate cases. The bound of the proposition blows up like $1/\sigma_h$: at $\sigma_h = 0$, $V_h(t,s) = e^{-r\tau}F(se^{(r-y)\tau})$ is not twice differentiable for a call, the delta is a step function (hold $e^{-y\tau}$ shares when the forward is above the strike: the *stop-loss start-gain* strategy), and the dealer, who charged only the intrinsic value of the forward, loses a local time of the forward at the strike instead of $\frac{1}{2}\int \Gamma_h S^2\sigma_t^2\,\mathrm{d}t$ (Carr and Jarrow, 1990). The proposition needs $\sigma_h > 0$. At the other end, if the payoff has $s^2\Gamma_h$ independent of s (the log contract), the P&L depends on the path only through $\int_0^T \sigma_t^2\,\mathrm{d}t$: the dealer is short realised variance (Exercise 18.4).

18.2.1 Implied or realised volatility?

Suppose the stock has a constant volatility σ_r, the market prices the call at an implied volatility $\sigma_i < \sigma_r$, and a trader buys it for $V_i(0, S_0)$. She can hedge at either volatility, with different consequences (Ahmad and Wilmott, 2005).

Hedging at the realised volatility. By Theorem 18.5 with $\sigma_h = \sigma_r$, the hedge replicates the payoff, so her discounted P&L at T is the deterministic amount $V_r(0, S_0) - V_i(0, S_0)$, whatever the path and the drift. Marked to market at the implied volatility, her P&L at time t is $(V_r - V_i)(0, S_0) - e^{-rt}(V_r - V_i)(t, S_t)$: random, and possibly negative before T.

Hedging at the implied volatility. By Theorem 18.5 with $\sigma_h = \sigma_i$, her discounted P&L is $\frac{1}{2}(\sigma_r^2 - \sigma_i^2)\int_0^T e^{-rt}\Gamma_i(t, S_t)S_t^2\,\mathrm{d}t > 0$. It accrues without a $\mathrm{d}W$ term, smoothly, but how much accrues depends on the path: most when the stock stays near the strike close to expiry, little when it runs away early.

Proposition 18.7 (The expected P&L of the implied hedge; Ahmad and Wilmott, 2005)**.** *Let S be geometric Brownian motion with volatility σ_r and total return μ, and Γ_i the gamma of the call (or put) at volatility $\sigma_i > 0$. For $0 \le t \le T$, with $\log S_t \sim \mathcal{N}(m_t, v_t)$, $c_t = \log K - (r - y + \frac{1}{2}\sigma_i^2)\tau$ and $w_t = \sigma_i^2\tau$,*

$$\mathbb{E}[S_t^2\Gamma_i(t, S_t)] = e^{-y\tau}\,\frac{\exp\left(\bar{x}_t + \frac{1}{2}s_t^2 - \frac{(c_t - m_t)^2}{2(v_t + w_t)}\right)}{\sqrt{2\pi(v_t + w_t)}}, \qquad \bar{x}_t = \frac{c_t v_t + m_t w_t}{v_t + w_t}, \quad s_t^2 = \frac{v_t w_t}{v_t + w_t},$$

(at $t = T$, $S_T^2\Gamma_i$ is read as K^2 times the density of S_T at K, the limit as $t \uparrow T$), and the expected discounted P&L of the implied hedge is $\frac{1}{2}(\sigma_r^2 - \sigma_i^2)\int_0^T e^{-rt}\mathbb{E}[S_t^2\Gamma_i(t, S_t)]\,\mathrm{d}t$. If $\mu = r$, this equals $V_r(0, S_0) - V_i(0, S_0)$.

Proof. $S^2\Gamma_i = e^{-y\tau}e^x N'((x - c_t)/\sqrt{w_t})/\sqrt{w_t}$ with $x = \log S$, and $N'((x - c)/\sqrt{w})/\sqrt{w}$ is the $\mathcal{N}(c, w)$ density at x. The product of the $\mathcal{N}(c, w)$ and $\mathcal{N}(m, v)$ densities is the $\mathcal{N}(m, v+w)$ density at c times the $\mathcal{N}(\bar{x}, s^2)$ density, and $\int e^x$ against the latter is $e^{\bar{x}+s^2/2}$. The expectation of the P&L follows by Tonelli's theorem (the integrand is nonnegative). If $\mu = r$, take $\mathbb{Q}$-expectations in (18.7) for the seller at σ_i: her discounted wealth is a $\mathbb{Q}$-martingale, since its integrand $e^{-rt}\sigma_r\Delta_i(t, S_t)S_t$ is square integrable, so $\mathbb{E}^{\mathbb{Q}}[e^{-rT}X_T] = V_i(0, S_0)$, while $\mathbb{E}^{\mathbb{Q}}[e^{-rT}F(S_T)] = V_r(0, S_0)$. □

With $\sigma_i = 20\%$, $\sigma_r = 25\%$ and the rest of the test case, the realised hedge locks in $e^{rT}(V_r - V_i) = 1.970$ (time-T money; $V_r = 9.315$). The implied hedge earns on average 1.970, 1.987, 1.984, 1.899 and 1.839 for $\mu = r$, 5%, 10%, 20% and −10%: in the real world the expected profit depends on the drift, through the time the stock spends near the strike. For $\mu = 10\%$, its continuous-time standard deviation is 0.828 (20,000 paths, daily Riemann sums), and its smallest value on these paths is 0.216. Figure 18.2 shows both hedges rebalanced daily: the realised hedge, 1.972 on average, is no longer exactly locked in, but its standard deviation, 0.545, is the discrete-hedging error of Section 18.1 at volatility 25%; the implied hedge has mean 1.983 and standard deviation 0.998. The realised hedge has the smaller final risk and the noisier path, and it needs a volatility that nobody knows in advance; Ellersgaard et al. (2017) find, in simulations and data, that hedging at the implied volatility is often the better practical choice.

▶ Interactive demo: choose the implied and the realised volatility and the drift; compare the mark-to-market paths and the final P&L of the two hedges with the expected value of

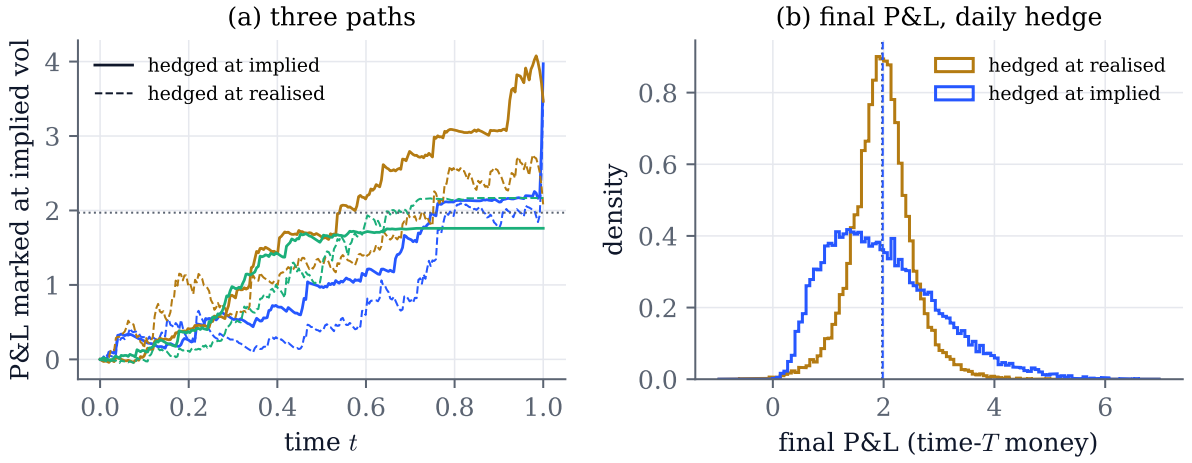

Figure 18.2: A call bought at the implied volatility 20% on a stock whose volatility is 25% ($\mu = 10\%$, rest of the test case), delta hedged daily at the implied or at the realised volatility. (a) The P&L marked to market at the implied volatility, in time-T money, along three paths (one colour per path; solid: hedged at implied, dashed: hedged at realised); the dotted line is the locked-in value 1.970. (b) Histograms of the final P&L over 20,000 paths (seed 18); dotted: 1.970; dashed: the expected P&L 1.984 of the implied hedge (Theorem 18.7).

Theorem 18.7.

18.3 Transaction costs and Leland's volatility

Let a trade of q shares at the price S cost $\frac{k}{2}|q|S$: $k \geq 0$ is the proportional *round-trip* cost of Leland (1985), and the course's one-way cost k' is $k/2$. The hedger of Section 18.1 now uses a hedging volatility $\hat{\sigma} > 0$: she sells the claim for $\hat{V}_0 = \hat{V}(0, S_0)$, the Black–Scholes price at $\hat{\sigma}$, holds $\hat{\delta}_j = \hat{\Delta}(t_j, S_{t_j})$ shares from t_j, and at each t_j, $1 \leq j \leq m-1$, pays $\frac{k}{2}|\hat{\delta}_j - e^{yh}\hat{\delta}_{j-1}|S_{t_j}$ (dividends are reinvested, so she arrives at t_j with $e^{yh}\hat{\delta}_{j-1}$ shares). As in Leland's paper and the course's homework, the trades that set up the hedge at t_0 and settle it at T are not charged; with constant k they would add a cost of order k that does not depend on m.

Each rebalance moves the delta by about $\Gamma\,\Delta S$, of order $\sqrt{h}$, so the cost per period is of order $k\sqrt{h}$ and the total cost of order $k/\sqrt{h}$: without an adjustment, frequent hedging ruins the hedger. The next lemma makes the cost per period precise.

> **Lemma 18.8** (The cost of one rebalance)**.** *Let $\hat{\Delta}$, $\hat{\Gamma}$ be the delta and gamma of the call (or the put) at a fixed volatility $\hat{\sigma} > 0$, let $t < T$ and $S_t = s$, and let the stock have volatility σ and any drift μ. As $h \downarrow 0$,*
>
> $$\mathbb{E}\Big[\tfrac{1}{2}S_{t+h}\,\big|\hat{\Delta}(t+h, S_{t+h}) - e^{yh}\hat{\Delta}(t,s)\big|\Big] = \tfrac{1}{2}s^2\hat{\Gamma}(t,s)\,\sigma\sqrt{2h/\pi} + O(h).$$

Proof. Let $t' = (t+T)/2$ and $h \leq t' - t$. On $[t, t'] \times (0, \infty)$, $\partial_t\hat{\Delta}$ and $x^2\partial_x\hat{\Gamma}$ are bounded, since they are products of $N'(d_\pm)$ with polynomials in $d_\pm$ and $\tau^{-1/2}$ (Theorem 2.7). With $R = S_{t+h}/s - 1$,

Taylor's formula in t and then in x gives $\hat\Delta(t+h,S_{t+h}) - e^{yh}\hat\Delta(t,s) = \hat\Gamma(t,s)sR + \rho$ with $|\rho| \le C(h + R^2\max(1,(1+R)^{-2}))$, the last factor bounding s^2/ξ^2 for ξ between s and S_{t+h}. Since $||a+\rho|-|a|| \le |\rho|$ and the moments of $1+R = e^{(\mu-y-\sigma^2/2)h+\sigma\sqrt{h}Z}$ and of $(1+R)^{-1}$ are bounded, the left side equals $\frac12 s^2\hat\Gamma(t,s)\mathbb{E}[(1+R)|R|]+O(h)$, and $\mathbb{E}[(1+R)|R|] = \sigma\sqrt{h}\,\mathbb{E}|Z|+O(h) = \sigma\sqrt{2h/\pi}+O(h)$ because $R = \sigma\sqrt{h}Z + O(h)$ in L^2. □

Leland's argument (heuristic). By the lemma, the expected cost of the rebalance at the end of a period is about $k\cdot\frac12 s^2\hat\Gamma\sigma\sqrt{2h/\pi}$. By Theorem 18.5, a hedger who charges the variance $\hat\sigma^2$ in a market of variance σ^2 earns, over the same period, $\frac12(\hat\sigma^2-\sigma^2)s^2\hat\Gamma h$. For a convex claim, $\hat\Gamma \ge 0$, and the two balance at every date and every spot exactly when

$$\hat\sigma^2 = \sigma^2\Big(1+\sqrt{\frac{2}{\pi}}\,\frac{k}{\sigma\sqrt{h}}\Big), \tag{18.8}$$

Leland's volatility; $A = \sqrt{2/\pi}\,k/(\sigma\sqrt{h})$ is the *Leland number*. The argument is a heuristic: the lemma is applied with $\hat\sigma$ depending on h, near expiry its $O(h)$ is not uniform, and the discrete-hedging error of Section 18.1 comes on top. What is true is the following.

Theorem 18.9 (Leland's volatility)**.** *In the standard model with $y=0$, let $F(s)=(s-K)^+$, let $k_m = k_0 m^{-\alpha}$ with $k_0>0$ and $0\le\alpha\le\frac12$ (equivalently $k_m = k_0(h/T)^\alpha$), and let $\hat\sigma_m$ be (18.8) with $k=k_m$ and $h=T/m$. Let $X_T^{(m)}$ be the terminal wealth of the hedger who sells the call for its price at $\hat\sigma_m$, holds $\hat\Delta_m(t_j,S_{t_j})$ shares and pays the costs k_m.*

(i) If $0<\alpha\le\frac12$, then $X_T^{(m)}\to(S_T-K)^+$ in probability as $m\to\infty$.

(ii) If $\alpha=0$ (a fixed cost k_0), then $X_T^{(m)}$ converges in probability to $(S_T-K)^+ + \ell(S_T)$ for an explicit function ℓ that is not identically zero: Leland's strategy does not replicate the call.

The theorem is not proved here. For $\alpha=\frac12$ the volatility $\hat\sigma = \sigma(1+\sqrt{2/\pi}\,k_0/(\sigma\sqrt{T}))^{1/2}$ does not depend on m; Leland (1985) conjectured the convergence, and Lott (1993) gave the first rigorous proof, in this case. Kabanov and Safarian (1997) extended (i) to $0<\alpha<\frac12$ and computed the limit of (ii) explicitly (their Theorem 2), which refuted Leland's claim that a constant cost is replicated in the limit. For $\alpha=\frac12$ and convex, piecewise smooth payoffs, Denis and Kabanov (2010) show that the mean squared error is $A/m+o(1/m)$ and identify A. Section 2 of Nguyen and Pergamenshchikov (2017) surveys these results (its Theorems 2.2, 2.3, 2.4 and 2.7 are, in this order, the results of Lott, of Kabanov and Safarian for $0<\alpha\le\frac12$, of Denis and Kabanov, and the limit for $\alpha=0$), including Pergamenshchikov's rate $m^{-1/4}$ for the fluctuations when $\alpha=0$ and the fact that the limit ℓ is nonpositive (the call is asymptotically underhedged); Kabanov and Safarian (2009), Chapter 1, is a book-length treatment. The cited results are stated without dividends, and typically with $r=0$, to which a constant rate reduces by discounting. Two edge cases: for $k=0$, $\hat\sigma=\sigma$ and the theorem is Theorem 18.3; for $\alpha>\frac12$, outside the theorem, $k_m/\sqrt{h}\to0$, so $\hat\sigma_m\to\sigma$, and heuristically the expected costs, of order $k_m/\sqrt{h}$, tend to zero. Leland's balance needs $\hat\Gamma\ge0$: for a book whose gamma changes sign it fails, and the correct adjustment leads to a nonlinear pricing equation (Exercise 18.6).

Table 18.2 runs the course's experiment. Hedging at σ, the seller loses the costs, 2.25 weekly and 4.99 daily on average, against 2.27 and 5.00 from summing the lemma, $e^{rT}k\mathcal{C}/\sqrt{h}$ with

			hedged at σ		Leland, $\mu = r$		Leland, mean / sd under $\mathbb{P}$		
m	$\hat{\sigma}$	$\hat{V}_0$	mean	sd	mean	sd	$\mu = 10\%$	$\mu = 20\%$	$\mu = -10\%$
1	0.2078	7.670	0.031	5.988	0.339	5.983	−0.447 / 7.167	−3.289 / 9.343	−0.805 / 6.040
4	0.2154	7.964	−0.497	3.336	0.128	3.336	−0.080 / 3.487	−0.784 / 3.610	−0.181 / 3.150
12	0.2260	8.377	−1.027	2.058	0.058	2.037	−0.012 / 2.070	−0.269 / 2.019	−0.062 / 1.915
52	0.2510	9.355	−2.246	1.337	0.013	1.063	−0.005 / 1.066	−0.076 / 1.013	−0.016 / 0.998
252	0.3011	11.303	−4.992	1.980	−0.007	0.555	−0.011 / 0.553	−0.028 / 0.524	−0.018 / 0.517

Table 18.2: The course's transaction-cost experiment: round-trip cost $k = 2\%$ (the homework's one-way $k' = 1\%$), test case, seller's error $X_T - (S_T - K)^+$ in time-T money. "Hedged at σ": premium and deltas at $\sigma = 20\%$; "Leland": premium $\hat{V}_0$ and deltas at $\hat{\sigma}$ of (18.8). Our implementation (numpy `default_rng`, seed 7, $n = 50{,}000$ paths); standard errors of the means at most 0.042.

$\mathcal{C} = \frac{1}{2}\sigma\sqrt{2/\pi}\int_0^T e^{-rt}\mathbb{E}[S_t^2\Gamma(t,S_t)]\,dt = 15.58$ (by Theorem 18.7 with $\sigma_i = \sigma_r = \sigma$). The loss grows like $\sqrt{m}$, and more frequent hedging even increases the standard deviation, through the randomness of the costs. With Leland's volatility the premium rises from 7.36 to 9.36 weekly and 11.30 daily, the mean error is within 0.06 of zero for $m \geq 12$, and the standard deviation falls, although more slowly than without costs (1.06 against 0.95 weekly, 0.555 against 0.437 daily). For $m = 1$ there is no rebalance, hence no cost, and the adjustment only overcharges. The strategy is as robust to the drift as the plain hedge: the means under $\mathbb{P}$ have the size of the drift effect of Table 18.1.

What of the failure in Theorem 18.9(ii)? Figure 18.3(b) refines the grid to $m = 10^5$. With $k = 2\%$ the mean error stays within a few hundredths of zero (-0.020 ± 0.004 at $m = 10^4$, where $\hat{\sigma} = 0.60$) while the standard deviation keeps falling (0.175). With $k = 10\%$ the mean is still drifting down (-0.23 at $m = 10^3$, -0.31 at 10^4, -0.36 at 10^5, each ± 0.016 at most), and the standard deviation, 0.34 at $m = 10^5$, falls only slowly (Pergamenshchikov's rate is $m^{-1/4}$): the limit of Kabanov and Safarian is nonzero and lies below -0.35. For the costs of a liquid market, it is a small effect.

▶ Interactive demo: set the cost k and the number of dates m; compare the error of the plain hedge with that of Leland's strategy, and push m up to see where the adjustment stops working.

In code

`code/src/mfe230d/hedging.py` implements the experiments of this chapter: `hedge_experiment` (and `hedge_stream`, memory-light for large m) simulates the seller's hedge with dividends reinvested, transaction costs and any hedging volatility; `gamma_integral` and `asymptotic_sd` compute the constant of Theorem 18.3; `gamma_s2_moment` and `expected_vol_pnl` the formulas of Theorem 18.7; `leland_vol` and `one_period_cost` (18.8) and Theorem 18.8. The tests in `code/tests/test_hedging.py` reproduce every number quoted here.

Further reading

Bertsimas et al. (2000) is a readable introduction to discrete hedging, and Toft (1996) adds

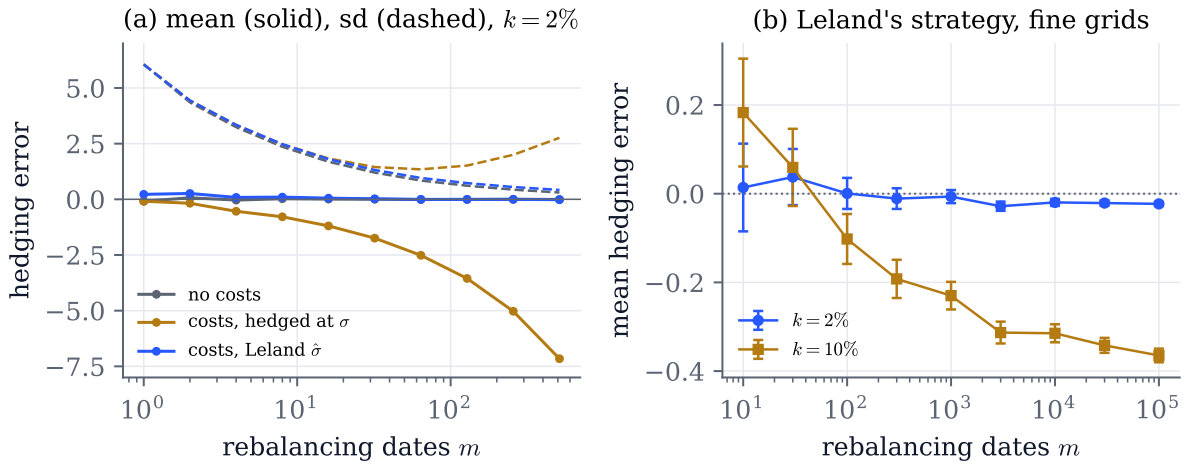

Figure 18.3: Transaction costs on the test-case call. (a) Mean (solid) and standard deviation (dashed) of the seller's error $X_T - (S_T - K)^+$ against m, with round-trip cost $k = 2\%$, hedging at σ or at Leland's $\hat{\sigma}$, and without costs (10,000 paths for each m, seed $180 + m$). (b) The mean error of Leland's strategy on finer grids, for $k = 2\%$ and $k = 10\%$, with ± 2 standard errors (2,000 paths for each m, seed $1985 + m$).

transaction costs and the trade-off between cost and risk (Exercise 18.7); Kamal and Derman (1999) give the traders' rule of thumb of Exercise 18.3. For irregular payoffs, Gobet and Temam (2001) compute the rates. El Karoui et al. (1998) is the reference on robustness, and Ahmad and Wilmott (2005) and Ellersgaard et al. (2017) discuss which volatility to hedge with. Kabanov and Safarian (2009), Chapter 1, covers Leland's approach with complete proofs.

Chapter 19

Stochastic volatility

In the Black–Scholes model every option on the same underlying has the same implied volatility. Market smiles say otherwise (Theorem 3.7): equity-index volatility is higher for low strikes, and it moves, mostly against the index. The *Heston model* makes the variance a mean-reverting diffusion correlated with the price. It keeps exactly the structure that made Black–Scholes tractable in Chapter 2: the characteristic function of the log-price is known in closed form, so European options are priced by one Fourier integral.

Sections 19.1 to 19.3 show that the model is well posed, derive the characteristic function, and settle two issues that trip up implementations: the branch of the complex logarithm and moments that explode at finite maturities. Section 19.4 simulates the model and measures the bias of two schemes against the Fourier price; Section 19.5 shows the smiles the model produces.

19.1 The Heston model

Definition 19.1 (Heston model)**.** Let W and $W^{\perp}$ be independent Brownian motions under $\mathbb{Q}$ with a filtration $(\mathcal{F}_t)$, let $\rho \in [-1,1]$, $\bar\rho = \sqrt{1-\rho^2}$ and $W^S = \rho W + \bar\rho W^{\perp}$. The *Heston model* with parameters $\kappa > 0$ (speed of mean reversion), $\theta > 0$ (long-run variance), $\xi > 0$ (volatility of variance), ρ (correlation) and $V_0 \geq 0$ is

$$\mathrm{d}V_t = \kappa(\theta - V_t)\,\mathrm{d}t + \xi\sqrt{V_t}\,\mathrm{d}W_t, \qquad \mathrm{d}S_t = S_t\big((r-y)\,\mathrm{d}t + \sqrt{V_t}\,\mathrm{d}W_t^S\big), \qquad S_0 = S > 0. \tag{19.1}$$

We write $X_t = \ln S_t$, $Q_t = \int_0^t V_s\,\mathrm{d}s$ and $F = Se^{(r-y)T}$ for the forward.

Given V, $S_t = S\exp\big((r-y)t - \tfrac12 Q_t + \int_0^t \sqrt{V}\,\mathrm{d}W^S\big)$. With $\rho < 0$ falls in the price come with rises in volatility, the leverage effect behind the equity skew. The variance is the square-root diffusion of Cox et al. (1985), sampled exactly in Chapter 10.

Proposition 19.2 (Well-posedness)**.** *(i) The variance equation has a pathwise unique strong solution, and $V_t \geq 0$ for all t, almost surely. Hence (19.1) has a unique strong solution.*

(ii) (Feller.) *If $V_0 > 0$ and $2\kappa\theta \geq \xi^2$, then $V_t > 0$ for all $t \geq 0$, almost surely. If $V_0 = 0$ and $2\kappa\theta \geq \xi^2$, then $V_t > 0$ for all $t > 0$, almost surely. If $2\kappa\theta < \xi^2$, V reaches 0 with positive probability.*

(iii) (Moments.) *For $0 \le s \le t$, with $e = e^{-\kappa(t-s)}$,*

$$\mathbb{E}[V_t \mid \mathcal{F}_s] = \theta + (V_s - \theta)e, \qquad \operatorname{Var}(V_t \mid \mathcal{F}_s) = V_s \frac{\xi^2 e(1-e)}{\kappa} + \theta \frac{\xi^2 (1-e)^2}{2\kappa}. \tag{19.2}$$

(iv) (Martingale.) *For every $\rho \in [-1,1]$, $(e^{-(r-y)t}S_t)_{t\ge 0}$ is a true martingale.*

Proof. (i) With $\sqrt{v^+}$ in place of $\sqrt{v}$, the coefficients are continuous with linear growth, which gives a weak solution (Karatzas and Shreve (1991), Theorem 5.4.22 and Remark 5.4.23). The drift is Lipschitz and $|\sqrt{v^+} - \sqrt{w^+}| \le \sqrt{|v-w|}$, the Yamada–Watanabe condition for pathwise uniqueness (Proposition 5.2.13); weak existence and pathwise uniqueness give a strong solution (Corollary 5.3.23). The process 0 solves the same equation with the smaller drift $-\kappa v$, so the comparison theorem (Proposition 5.2.18) gives $V \ge 0$, and $\sqrt{V^+} = \sqrt{V}$. The price is then the explicit exponential above.

(ii) For $v > 0$ let $s'(v) = v^{-q}e^{2\kappa v/\xi^2}$ with $q = 2\kappa\theta/\xi^2$, and $s(v) = \int_1^v s'$. Then $\kappa(\theta - v)s' + \frac{1}{2}\xi^2 v s'' = 0$, so for $0 < a < V_0 < b$ and $\tau = \inf\{t : V_t \notin (a,b)\}$, Itô's formula makes $s(V_{t\wedge\tau})$ a bounded local martingale, hence a martingale. The exit time τ is finite almost surely because the diffusion coefficient is bounded below on $[a,b]$ (Section 5.5), so optional stopping gives $\mathbb{Q}(V \text{ hits } a \text{ before } b) = (s(b) - s(V_0))/(s(b) - s(a))$. If $q \ge 1$, $s(a) \to -\infty$ as $a \downarrow 0$, so V reaches 0 before b with probability 0, for every b; its paths are continuous, so the hitting time of b tends to ∞ as $b \to \infty$, and V never reaches 0. If $V_0 = 0$, the law of V_s is a multiple of a chi-square law with $4\kappa\theta/\xi^2 > 0$ degrees of freedom (Glasserman, 2003, Section 3.4.1), so $V_s > 0$ almost surely; applying the first case after each time $s = 1/n$ (Markov property) and taking the union over n gives $V_t > 0$ for all $t > 0$. The converse for $q < 1$, where $s(0+) > -\infty$, is Feller's test (Feller, 1951); see Karatzas and Shreve (1991), Section 5.5.

(iii) Given V_s, $V_t = c\,\chi'^2_d(\lambda)$, a noncentral chi-square variable with $d = 4\kappa\theta/\xi^2$ degrees of freedom and noncentrality $\lambda = V_s e/c$, where $c = \xi^2(1-e)/(4\kappa)$ (Glasserman, 2003, Section 3.4.1). Its mean is $c(d+\lambda)$ and its variance $c^2(2d + 4\lambda)$; substitute.

(iv) $Z_t = e^{-(r-y)t}S_t/S = \exp\big(\int_0^t \sqrt{V}\,\mathrm{d}W^S - \frac{1}{2}Q_t\big)$ is a positive local martingale, hence a supermartingale, and it suffices to show $\mathbb{E}Z_T \ge 1$ for each T. Let $\tau_n = \inf\{t : V_t \ge n\}$. By Novikov's condition $Z_{t\wedge\tau_n}$ is a martingale, and under $\mathbb{Q}_n$, $\mathrm{d}\mathbb{Q}_n/\mathrm{d}\mathbb{Q} = Z_{T\wedge\tau_n}$, Girsanov's theorem makes $W^n_t = W_t - \rho\int_0^{t\wedge\tau_n}\sqrt{V_s}\,\mathrm{d}s$ a Brownian motion on $[0,T]$. Up to τ_n, V solves $\mathrm{d}V = (\kappa\theta - (\kappa - \rho\xi)V)\,\mathrm{d}t + \xi\sqrt{V}\,\mathrm{d}W^n$. Let $\tilde{V}$ be the solution of this equation on all of $[0,T]$, driven by W^n; it exists by the argument of (i), whatever the sign of $\kappa - \rho\xi$. The process equal to V up to τ_n and continued after τ_n by the strong solution started from V_{τ_n} and driven by W^n is another solution on $[0,T]$, so pathwise uniqueness gives $V = \tilde{V}$ up to τ_n. Hence

$$\mathbb{E}Z_T \ge \mathbb{E}\big[Z_{T\wedge\tau_n}\mathbf{1}_{\{\tau_n > T\}}\big] = \mathbb{Q}_n(\tau_n > T) = \mathbb{Q}_n\Big(\sup_{t\le T}\tilde{V}_t < n\Big) = p_n,$$

where p_n is the probability that the solution of this equation driven by any Brownian motion stays below n on $[0,T]$, which does not depend on the Brownian motion (pathwise uniqueness implies uniqueness in law). Its paths are continuous, so $p_n \to 1$, and $\mathbb{E}Z_T \ge 1$. A supermartingale with constant mean is a martingale. □

Property (iv) can fail in other stochastic volatility models (Sin, 1998); Heston is safe because

under the share measure the variance is again a square-root diffusion (see also Keller-Ressel, 2011, Corollary 2.7).

Remark 19.3 (The physical measure and the market price of volatility risk). Under $\mathbb{P}$ let $\mathrm{d}V = \kappa^{\mathbb{P}}(\theta^{\mathbb{P}} - V)\,\mathrm{d}t + \xi\sqrt{V}\,\mathrm{d}B$. The variance is not traded, so the market is incomplete: each market price of risk λ_V for B gives a risk-neutral measure, under which the drift of V is $\kappa^{\mathbb{P}}(\theta^{\mathbb{P}} - V) - \xi\sqrt{V}\lambda_V$. Heston (1993), Section 1, takes a premium proportional to V ($\lambda_V = \kappa_V\sqrt{V}$ in the course notes), which keeps the form (19.1) with $\kappa = \kappa^{\mathbb{P}} + \xi\kappa_V$ and $\kappa\theta = \kappa^{\mathbb{P}}\theta^{\mathbb{P}}$ (Exercise 19.1). Option prices identify only the $\mathbb{Q}$-parameters.

Proposition 19.4 (The pricing equation). *Let f be a payoff with $|f(s)| \le A(1+s)$, and*

$$\mathcal{L}u = (r-y)s\,\partial_s u + \kappa(\theta - v)\,\partial_v u + \tfrac{1}{2}vs^2\,\partial_{ss}u + \rho\xi vs\,\partial_{sv}u + \tfrac{1}{2}\xi^2 v\,\partial_{vv}u.$$

(i) (Verification.) *Let u be continuous on $[0,T]\times(0,\infty)\times[0,\infty)$ with $u(T,s,v) = f(s)$ and $|u(t,s,v)| \le A'(1+s+v)$, and $C^{1,2}$ on an open set containing $[0,T)\times(0,\infty)\times[0,\infty)$, where it solves*

$$\partial_t u + \mathcal{L}u - ru = 0. \tag{19.3}$$

Then $u(t,S_t,V_t) = e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}[f(S_T)\mid\mathcal{F}_t]$ for $t \le T$; in particular $u(0,S,V_0)$ is the price. At $v=0$ the equation reads $\partial_t u + (r-y)s\,\partial_s u + \kappa\theta\,\partial_v u - ru = 0$.

(ii) (Converse.) *Ekström and Tysk (2010) show, for a class of models that contains Heston's and under conditions on f stated there, that the price is the unique classical solution of (19.3) with terminal value f in a suitable growth class, the equation holding up to $v=0$ in the degenerate form above. When Feller's condition fails, 0 is attainable and this degenerate equation is the boundary condition at $v=0$, needed for uniqueness; when it holds, no condition at $v=0$ is needed.*

Proof. (i) By Itô's formula and (19.3), $M_t = e^{-rt}u(t,S_t,V_t)$ is a local martingale on $[0,T)$; let $\tau_n \uparrow T$ localise it. The families $\{S_\tau\}$ and $\{V_\tau\}$, over stopping times $\tau \le T$, are uniformly integrable: $e^{-(r-y)\tau}S_\tau = \mathbb{E}[e^{-(r-y)T}S_T \mid \mathcal{F}_\tau]$ by Theorem 19.2(iv) and optional sampling, and by (19.2) the process $e^{\kappa t}(V_t - \theta)$ is a martingale, so $|V_\tau| \le \theta + e^{\kappa T}\mathbb{E}[V_T + \theta \mid \mathcal{F}_\tau]$; conditional expectations of one integrable variable are uniformly integrable. Hence $\{M_\tau\}$ is uniformly integrable, and letting $n\to\infty$ in $M_{t\wedge\tau_n} = \mathbb{E}[M_{\tau_n}\mid\mathcal{F}_t]$ gives $M_t = \mathbb{E}[M_T\mid\mathcal{F}_t]$ with $M_T = e^{-rT}f(S_T)$. At $v=0$ the second-order coefficients vanish. □

The course derives (19.3) by hedging with the stock and other options; *one* option with nonzero sensitivity to V suffices (Exercise 19.1). It is Heston (1993)'s equation (6).

19.2 Affine structure and the characteristic function

The coefficients of (19.1) in (X,V) are *affine* in the state: the drift $(r-y-\tfrac{1}{2}V,\ \kappa\theta-\kappa V)$ and the instantaneous covariance $V\left(\begin{smallmatrix}1 & \rho\xi\\ \rho\xi & \xi^2\end{smallmatrix}\right)$ are affine functions of V and do not depend on X. For such processes the conditional characteristic function is the exponential of an affine function of the state (Duffie et al., 2000, 2003). We guess $\mathbb{E}[e^{iuX_T}\mid\mathcal{F}_t] = \exp\big(iuX_t + iu(r-y)\tau + C(u,\tau) + D(u,\tau)V_t\big)$, $\tau = T-t$, and substitute in (19.3) written in $x = \ln s$ without the term $-ru$ (a conditional

expectation is a martingale). The terms without v and those proportional to v must vanish separately, which gives the *Riccati equations*

$$\partial_\tau D = \tfrac{1}{2}\xi^2 D^2 - \beta(u)D - \tfrac{1}{2}(u^2+iu), \qquad \partial_\tau C = \kappa\theta D, \qquad C(u,0) = D(u,0) = 0, \tag{19.4}$$

with $\beta(u) = \kappa - i\rho\xi u$. These are the equations of the course's notes, which transform the call price in $\ln(S/K)$ and stop after solving for D. The quadratic $\frac{1}{2}\xi^2 D^2 - \beta D - \frac{1}{2}(u^2+iu)$ has roots $(\beta \pm d)/\xi^2$, where

$$d(u) = \sqrt{\beta(u)^2 + \xi^2(u^2+iu)}, \qquad g(u) = \frac{\beta - d}{\beta + d}, \tag{19.5}$$

and $\sqrt{\cdot}$ and Log denote principal branches ($\operatorname{Re}\sqrt{z} \geq 0$, $\operatorname{Arg} \in (-\pi, \pi]$). We write $\varphi(u) = \mathbb{E}[e^{iuX_T}]$, the convention $\varphi_Y(z) = \mathbb{E}[e^{izY}]$ of Chapter 2.

Theorem 19.5 (The characteristic function)**.** *Let $T \geq 0$ and, for $u \in \mathbb{C}$ and $t \geq 0$,*

$$D(u,t) = \frac{\beta - d}{\xi^2}\,\frac{1 - e^{-dt}}{1 - ge^{-dt}}, \qquad C(u,t) = \frac{\kappa\theta}{\xi^2}\Big((\beta - d)t - 2\operatorname{Log}\frac{1 - ge^{-dt}}{1 - g}\Big). \tag{19.6}$$

(i) (Real arguments.) *For $u \in \mathbb{R}$, (19.4) has a unique solution (C, D) on $[0,\infty)$ (by (ii) and (iii) it is (19.6) when $\rho\xi \leq 2\kappa$), $\operatorname{Re} D \leq 0$, and*

$$\varphi(u) = \exp\big(iu(\ln S + (r-y)T) + C(u,T) + D(u,T)V_0\big). \tag{19.7}$$

(ii) (Stable form.) *If $\operatorname{Re} d(u) > 0$ and $|g(u)| < 1$, then $\operatorname{Re}(1 - ge^{-dt}) > 0$ for all $t \geq 0$, and (19.6) solves (19.4) on $[0,\infty)$. The set of such u is open, and on it C and D are analytic in u.*

(iii) (When the stable form applies.) *For real u, $\operatorname{Re} d(u) > 0$, and $|g(u)| < 1$ if $\rho\xi \leq 2\kappa$. At $u - i$, u real, the same holds if $\rho\xi < \kappa$; there (19.7) gives $\varphi(u - i) = \mathbb{E}[S_T e^{iuX_T}]$. In particular, for $\rho \leq 0$ the stable form with principal branches is exact at every point used by Theorem 19.7(i) and (iii).*

(iv) (Moments.) *For real a let $T^*(a) = \sup\{T : \mathbb{E}^{\mathbb{Q}}[S_T^a] < \infty\}$. Then $T^*(a) = \infty$ for $a \in [0,1]$. Otherwise, with $\kappa_a = \kappa - \rho\xi a$ and $\Delta_a = \kappa_a^2 - \xi^2 a(a-1)$,*

$$T^*(a) = \begin{cases} \infty, & \Delta_a \geq 0,\ \kappa_a > 0, \\ \dfrac{2}{\gamma_a}\Big(\dfrac{\pi}{2} + \arctan\dfrac{\kappa_a}{\gamma_a}\Big), \quad \gamma_a = \sqrt{-\Delta_a}, & \Delta_a < 0, \\ \dfrac{1}{\sqrt{\Delta_a}}\ln\dfrac{\kappa_a - \sqrt{\Delta_a}}{\kappa_a + \sqrt{\Delta_a}}, & \Delta_a > 0,\ \kappa_a < 0, \end{cases} \tag{19.8}$$

($2/|\kappa_a|$ if $\Delta_a = 0 > \kappa_a$), and T^* is nonincreasing on $[1,\infty)$. Moments of order $a > 1$ can thus be infinite at a finite maturity.

Proof. (i) The right side of (19.4) is a polynomial in D, so there is a unique maximal solution on some $[0, t^*)$, and if $t^* < \infty$ then $|D| \to \infty$ at t^*. Write $D = a + ib$. The real part of the right side is

$$\tfrac{1}{2}\xi^2(a^2 - b^2) - \kappa a - \rho\xi u b - \tfrac{1}{2}u^2 = a(\tfrac{1}{2}\xi^2 a - \kappa) - \tfrac{1}{2}(\xi b + \rho u)^2 - \tfrac{1}{2}(1-\rho^2)u^2 \leq a(\tfrac{1}{2}\xi^2 a - \kappa).$$

Since $a(0) = 0$ and 0 solves $y' = y(\frac{1}{2}\xi^2 y - \kappa)$, the comparison theorem for scalar ordinary differential equations gives $a \le 0$ on $[0, t^*)$. If $t^* < \infty$, then $w = 1/D$, defined near t^*, satisfies $\partial_t w = -\frac{1}{2}\xi^2 + \beta w + \frac{1}{2}(u^2 + iu)w^2 \to -\frac{1}{2}\xi^2$ and $w \to 0$ as $t \uparrow t^*$, so $w(t) = \frac{1}{2}\xi^2(t^* - t)(1 + o(1))$ and $\operatorname{Re} D = \operatorname{Re} w/|w|^2 > 0$ near t^*, a contradiction. So $t^* = \infty$, and $\operatorname{Re} C = \kappa\theta \int \operatorname{Re} D \le 0$. Now let $M_t = \exp\left(iuX_t + iu(r - y)(T - t) + C(u, T - t) + D(u, T - t)V_t\right)$. Itô's formula and (19.4) show that the drift of M vanishes, and $|M_t| \le 1$; a bounded local martingale is a martingale, so $\varphi(u) = \mathbb{E}M_T = M_0$.

(ii) $|e^{-dt}| = e^{-t\operatorname{Re} d} \le 1$, so $|ge^{-dt}| < 1$ and $\operatorname{Re}(1 - ge^{-dt}) > 0$, as well as $\operatorname{Re}(1 - g) > 0$. The principal logarithms of both are continuous and their arguments lie in $(-\frac{\pi}{2}, \frac{\pi}{2})$, so the Log of the ratio is the difference of the Logs and is continuous in t. With $r_\pm = (\beta \pm d)/\xi^2$ we have $gr_+ = r_-$, and (19.6) gives $D - r_- = r_- e^{-dt}(g - 1)/(1 - ge^{-dt})$ and $D - r_+ = (r_- - r_+)/(1 - ge^{-dt})$, so $(D - r_-)/(D - r_+) = ge^{-dt}$. Differentiating, $\partial_t D\,(r_- - r_+)/(D - r_+)^2 = -dge^{-dt}$, that is $\partial_t D = \frac{1}{2}\xi^2(D - r_+)(D - r_-)$, which is (19.4); and $D(u, 0) = 0$. Differentiating C,

$$\partial_t C = \frac{\kappa\theta}{\xi^2}\,\frac{(\beta - d)(1 - ge^{-dt}) - 2gde^{-dt}}{1 - ge^{-dt}} = \kappa\theta D,$$

because $(\beta - d)(1 - g) = 2gd$. The conditions are open in u ($\operatorname{Re} d > 0$ means $\beta^2 + \xi^2(u^2 + iu) \notin (-\infty, 0]$), and compositions of analytic functions are analytic.

(iii) For real u, $\Delta = \beta^2 + \xi^2(u^2 + iu)$ has $\operatorname{Re}\Delta = \kappa^2 + \xi^2(1 - \rho^2)u^2 > 0$, so $\operatorname{Re} d > 0$. Next, $|g| < 1$ if and only if $\operatorname{Re}(\beta\bar{d}) > 0$. With $d = p + iq$, $2pq = \operatorname{Im}\Delta = \xi u(\xi - 2\kappa\rho)$ and $p^2 \ge \operatorname{Re}\Delta$, so

$$2p\operatorname{Re}(\beta\bar{d}) = 2p(\kappa p - \rho\xi uq) = 2\kappa p^2 - \rho\xi^2 u^2(\xi - 2\kappa\rho) \ge 2\kappa^3 + \xi^2 u^2(2\kappa - \rho\xi) > 0$$

when $\rho\xi \le 2\kappa$. At $u - i$, $\beta = \kappa' - i\rho\xi u$ with $\kappa' = \kappa - \rho\xi$, and $u^2 + iu$ becomes $u^2 - iu$; the same computation gives $2p\operatorname{Re}(\beta\bar{d}) \ge 2\kappa'^3 + \xi^2 u^2(2\kappa - \rho\xi) > 0$ when $\kappa' > 0$. For the identity, the share measure $\mathrm{d}\mathbb{Q}^S/\mathrm{d}\mathbb{Q} = S_T/F$ is a probability by Theorem 19.2(iv); under it (Girsanov) X has drift $r - y + \frac{1}{2}V$ and V has drift $\kappa\theta - \kappa' V$, and the argument of (i), which did not use the sign of κ, gives $\mathbb{E}^S[e^{iuX_T}]$ by the Riccati equations of $u - i$. Multiply by F.

(iv) For real a the Riccati equation of $u = -ia$ is real, $\partial_\tau D = \frac{1}{2}\xi^2 D^2 - \kappa_a D + \frac{1}{2}a(a - 1)$. For $a \in [0, 1]$ its roots have product $a(a - 1)/\xi^2 \le 0$, so D stays between 0 and the nonpositive root. Otherwise separating variables gives the blow-up time (19.8) (Exercise 19.4). For $T < T^*(a)$, M_t defined as in (i) with a in place of iu is a positive local martingale, hence a supermartingale, and $\mathbb{E}[S_T^a] = \mathbb{E}M_T \le M_0 < \infty$. That $\mathbb{E}[S_T^a] = \infty$ for $T > T^*(a)$, with the whole formula (19.8), is Andersen and Piterbarg (2007), Proposition 3.1; Keller-Ressel (2011), Theorem 4.1 and (6.2), proves it for affine models when $\kappa > \rho\xi$. Finally, by Lyapunov's inequality $\mathbb{E}[S_T^b] \le 1 + \mathbb{E}[S_T^a]$ for $1 \le b \le a$, so T^* is nonincreasing on $[1, \infty)$. □

Remark 19.6 (The little Heston trap). Heston (1993), equation (17), uses $g_H = 1/g$ and e^{+dt}: $C = \frac{\kappa\theta}{\xi^2}\left((\beta + d)t - 2\operatorname{Log}\frac{1 - g_H e^{dt}}{1 - g_H}\right)$. Since $\frac{1 - g_H e^{dt}}{1 - g_H} = e^{dt}\,\frac{1 - ge^{-dt}}{1 - g}$, this is algebraically (19.6); but $\operatorname{Log}(e^{dt}z) = dt + \operatorname{Log} z$ only while $t\operatorname{Im} d + \operatorname{Arg} z \in (-\pi, \pi]$. For real u, $\operatorname{Im} d$ is 0 at $u = 0$ and tends to $(\xi - 2\kappa\rho)/(2\bar{\rho})$ as $u \to \infty$ ($|\rho| < 1$), so when $\operatorname{Im} d \ne 0$ the argument wraps around at long maturities, C jumps by $\pm 4\pi i\kappa\theta/\xi^2$ and φ by the factor $e^{\pm 4\pi i\kappa\theta/\xi^2} \ne 1$ unless $2\kappa\theta/\xi^2$ is an integer (Albrecher et al. (2007), Proposition 1). Their Theorem 3 shows that the stable form does not cross the cut along the line of Theorem 19.7(ii); Kahl and Jäckel (2005) repair Heston's form by counting rotations; Lord and Kahl (2010) prove the principal branch correct in the whole

strip where φ is finite if $\rho\xi \le \kappa$, and in part of it if $\kappa < \rho\xi < 2\kappa$ (Theorem 3 of Lord and Kahl (2006)). Part (iii) is the case we need, with a short proof.

Figure 19.1(a) shows the trap for the 2005 Eurostoxx calibration of Fang and Oosterlee (2008) and Albrecher et al. (2007) ($V_0 = 0.0175$, $\kappa = 1.5768$, $\theta = 0.0398$, $\xi = 0.5751$, $\rho = -0.5711$; Feller's ratio 0.38) at $T = 10$. For the at-the-money call with $r = 2.5\%$, $y = 0$, the stable form gives 18.77, 24.50 and 43.17 at $T = 4$, 6 and 15, the values of Albrecher et al. (2007), Table 1; Heston's form gives 18.77, 23.92 and 31.99 in our Carr–Madan implementation ($\alpha = 0.75$, step 0.01; their wrong values, 23.90 and 32.57, differ because the error depends on where the nodes fall relative to the jumps).

Degenerate cases. The theorem holds as stated for $\rho = \pm 1$, without Feller's condition and for $V_0 = 0$. For $\xi = 0$ (Black–Scholes with a deterministic variance, Exercise 19.2) the closed form, which divides by ξ^2, is replaced in the code by its limit.

19.3 Pricing by Fourier inversion

Proposition 19.7 (Fourier pricing). *Let $k = \ln K$, $C(K) = e^{-rT}\mathbb{E}^{\mathbb{Q}}[(S_T - K)^+]$ and $P(K)$ the put; $C - P = Se^{-yT} - Ke^{-rT}$.*

(i) (Gil-Pelaez.) *For every K with $\mathbb{Q}(S_T = K) = 0$, which holds for every K when $|\rho| < 1$,*

$$C = Se^{-yT}P_1 - Ke^{-rT}P_2,$$
$$P_2 = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \operatorname{Re}\frac{e^{-iuk}\varphi(u)}{iu}\,\mathrm{d}u, \qquad P_1 = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \operatorname{Re}\frac{e^{-iuk}\varphi(u-i)}{iuF}\,\mathrm{d}u, \tag{19.9}$$

where $P_2 = \mathbb{Q}(S_T > K)$, $P_1 = \mathbb{Q}^S(S_T > K)$, the integrands are bounded near 0 and the integrals converge at ∞. Moreover $|\varphi(u)| \le L(\frac{1}{2}\bar\rho^2 u^2)$, where $L(\lambda) = \mathbb{E}^{\mathbb{Q}}[e^{-\lambda Q_T}]$ is explicit and $\ln L(\lambda) = -\sqrt{2\lambda}\,(V_0 + \kappa\theta T)/\xi + O(1)$ as $\lambda \to \infty$.

(ii) (Carr–Madan.) *Let $\alpha > 0$. The damped price $c(k) = e^{\alpha k}C(e^k)$ is integrable on $\mathbb{R}$ if and only if $\mathbb{E}^{\mathbb{Q}}[S_T^{1+\alpha}] < \infty$, that is, $T < T^*(1+\alpha)$ in (19.8). Then its transform is*

$$\psi(v) = \int_{\mathbb{R}} e^{ivk}c(k)\,\mathrm{d}k = \frac{e^{-rT}\varphi(v-(\alpha+1)i)}{\alpha^2 + \alpha - v^2 + i(2\alpha+1)v},$$
$$C(K) = \frac{e^{-\alpha k}}{\pi}\int_0^\infty \operatorname{Re}\left(e^{-ivk}\psi(v)\right)\mathrm{d}v. \tag{19.10}$$

(iii) (COS.) *Let $Y = \ln(S_T/K)$, $a < 0 < b$, $\omega_k = k\pi/(b-a)$, and let $V_k = \frac{2}{b-a}\int_a^0 K(1-e^y)\cos(\omega_k(y-a))\,\mathrm{d}y$ be the cosine coefficients of the put payoff. With $\varphi_Y(u) = \varphi(u)e^{-iuk}$ and $\sum'$ a sum whose first term is halved, let*

$$P_N = e^{-rT}\sum_{k=0}^{N-1}{}' \operatorname{Re}\left(\varphi_Y(\omega_k)e^{-i\omega_k a}\right)V_k.$$

If Y has a density f that is square integrable on $[a,b]$, and $\varepsilon = \mathbb{Q}(Y \notin [a,b])$, then

$$|P - P_N| \le e^{-rT}\Big(K\,\mathbb{Q}(Y<a) + \varepsilon\sum_{k<N}{}'|V_k| + \frac{b-a}{2}\Big|\sum_{k\ge N}A_kV_k\Big|\Big), \tag{19.11}$$

where $A_k = \frac{2}{b-a}\int_a^b f(y)\cos(\omega_k(y-a))\,\mathrm{d}y$*; the last term tends to* 0 *as* $N \to \infty$.

Proof. (i) *Gil-Pelaez's formula.* For any random variable Y and continuity point x of its distribution function, $\mathbb{Q}(Y > x) = \frac{1}{2} + \lim \frac{1}{\pi}\int_\epsilon^U \mathrm{Im}\,(e^{-iux}\mathbb{E}e^{iuY})\,\frac{\mathrm{d}u}{u}$ as $\epsilon \downarrow 0$, $U \to \infty$ (Gil-Pelaez, 1951). Indeed the integral equals $\mathbb{E}\big[\frac{1}{\pi}\int_\epsilon^U \sin(u(Y-x))\frac{\mathrm{d}u}{u}\big]$ by Fubini (the integrand is bounded on $[\epsilon, U]$). The inner integral is bounded uniformly in ϵ, U and Y, and tends to $\frac{1}{2}\operatorname{sgn}(Y-x)$, so by dominated convergence the limit is $\frac{1}{2}(\mathbb{Q}(Y > x) - \mathbb{Q}(Y < x)) = \mathbb{Q}(Y > x) - \frac{1}{2}$. Note that $\mathrm{Im}(e^{-iux}\varphi)/u = \mathrm{Re}(e^{-iux}\varphi/(iu))$. Apply this to $Y = X_T$ under $\mathbb{Q}$ and under $\mathbb{Q}^S$, whose characteristic function is $\varphi(u-i)/F$ by Theorem 19.5(iii); and $e^{-rT}\mathbb{E}[S_T\mathbf{1}_{\{S_T>K\}}] = e^{-rT}F\,\mathbb{Q}^S(S_T > K) = Se^{-yT}P_1$. Given $\mathcal{F}_T^W$, X_T is normal with variance $\bar\rho^2 Q_T$ (Theorem 10.12(i)), and $Q_T > 0$ almost surely: on $\{Q_T = 0\}$, $V = 0$ on $[0,T]$ and the variance equation would give $V_T = \kappa\theta T > 0$. So for $|\rho| < 1$, X_T has no atoms. Near $u = 0$ the integrands are bounded because X_T has finite first moments under both measures: by (19.8), $T^*(a) \to \infty$ as $a \to 0$ and as $a \to 1$, so $\mathbb{E}[S_T^a] < \infty$ for a in a neighbourhood of $[0,1]$. For the bound,

$$|\varphi(u)| = |\mathbb{E}\,\mathbb{E}[e^{iuX_T} \mid \mathcal{F}_T^W]| \le \mathbb{E}\,e^{-\frac{1}{2}u^2\bar\rho^2 Q_T} = L(\tfrac{1}{2}\bar\rho^2 u^2),$$

and $L(\lambda) = Ae^{-BV_0}$ is the bond price of Cox et al. (1985) with short rate λV: with $\gamma = \sqrt{\kappa^2 + 2\xi^2\lambda}$, $B = 2\lambda(e^{\gamma T}-1)/((\gamma+\kappa)(e^{\gamma T}-1)+2\gamma) = \sqrt{2\lambda}/\xi + O(1)$ and $\ln A = -\kappa\theta T\gamma/\xi^2 + O(1) = -\kappa\theta T\sqrt{2\lambda}/\xi + O(1)$.

(ii) For $s > 0$, $\int_{\mathbb{R}} e^{(\alpha+iv)k}(s-e^k)^+\,\mathrm{d}k = s^{\alpha+1+iv}/((\alpha+iv)(\alpha+1+iv))$, and for $v = 0$ the integrand is nonnegative. By Tonelli, $\int c = e^{-rT}\mathbb{E}[S_T^{1+\alpha}]/(\alpha(\alpha+1))$, which proves the equivalence; then Fubini, with $\mathbb{E}[S_T^{1+\alpha}e^{ivX_T}] = \varphi(v-(\alpha+1)i)$ and $(\alpha+iv)(\alpha+1+iv) = \alpha^2+\alpha-v^2+i(2\alpha+1)v$, gives ψ. Since $|\psi(v)| \le \psi(0)\alpha(\alpha+1)/|(\alpha+iv)(\alpha+1+iv)|$ is integrable and c is continuous, the inversion theorem (Rudin (1987), Theorem 9.11) gives $c(k) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-ivk}\psi(v)\,\mathrm{d}v$; c is real, so $\psi(-v) = \overline{\psi(v)}$ and the integral is twice the integral of the real part over $(0,\infty)$.

(iii) Let $g(y) = K(1-e^y)^+$, so $P = e^{-rT}\int gf$. The functions $\cos(\omega_k(y-a))$, $k \ge 0$, are an orthogonal basis of $L^2[a,b]$ with squared norms $b-a$ for $k = 0$ and $\frac{b-a}{2}$ otherwise, so Parseval's identity gives $\int_a^b gf = \frac{b-a}{2}{\sum_k}' A_kV_k$. Outside $[a,b]$, gf vanishes for $y > b > 0$ and $g \le K$. Next, $F_k = \frac{2}{b-a}\mathrm{Re}(\varphi_Y(\omega_k)e^{-i\omega_k a}) = \frac{2}{b-a}\int_{\mathbb{R}} f(y)\cos(\omega_k(y-a))\,\mathrm{d}y$, so $|A_k - F_k| \le \frac{2}{b-a}\varepsilon$, and $P_N = e^{-rT}\frac{b-a}{2}{\sum_{k<N}}' F_kV_k$. Combine. The series $\sum A_kV_k$ converges by Parseval, so its tail tends to 0. □

Gil-Pelaez's form (19.9) is the one Heston (1993), equation (18), gives. Carr and Madan's transform (19.10) is their equation (6), Carr and Madan (1999, Section 3.1); the damping plays the role of $a = 1+\alpha$ in the transform of the payoff of Section 2.3, and Theorem 19.5(iv) is what fixes the admissible range: $0 < \alpha < a_+(T) - 1$, where $a_+(T) = \sup\{a \ge 1 : \mathbb{E}[S_T^a] < \infty\}$. In the test case below $a_+(1) = 38.50$; for $\kappa = 1$, $\xi = 1$, $\rho = 0.8$, $T^*(2) = 1.77$, so at $T = 2$ no damping $\alpha \ge 1$ is admissible, and the call payoff, whose square grows like S_T^2, has infinite variance: plain Monte Carlo then has no central limit theorem (Theorem 6.4). Carr and Madan (1999) suggest a quarter of the upper bound for α.

The put coefficients are explicit (Fang and Oosterlee, 2008, (20)–(25)): $V_k = \frac{2K}{b-a}(\psi_k - \chi_k)$ with $\psi_0 = -a$, $\psi_k = \sin(-\omega_k a)/\omega_k$ $(k \ge 1)$ and $\chi_k = \big(\cos(\omega_k a) - e^a + \omega_k\sin(-\omega_k a)\big)/(1+\omega_k^2)$; calls follow by parity, as Fang and Oosterlee (2008) recommend. In (19.11) the first two terms are truncation errors, and $\sum|V_k| < \infty$ since $V_k = O(k^{-2})$. For the last term Fang and Oosterlee

(2008), Lemma 4.2 and (47), state exponential convergence for smooth densities; strictly, the coefficients of a C^2 density are only $O(k^{-2})$, with a constant proportional to $|f'(a)| + |f'(b)|$, which is of the order of the truncation error. So the interval matters more than N. Fang and Oosterlee (2008) take $[c_1 \pm 12\sqrt{c_2}]$ for Heston, with c_1, c_2 the mean and variance of Y. We use the moments of Theorem 19.5(iv) instead: by Chernoff's inequality $\mathbb{Q}(Y > b) \le e^{-pb}\mathbb{E}[e^{pY}]$ and $\mathbb{Q}(Y < a) \le e^{pa}\mathbb{E}[e^{-pY}]$ wherever the moments are finite, and we take the narrowest $[a, b]$ for which the best bounds are below 10^{-14}.

Quadrature. The integrands in (i) and (ii) have even real parts ($\varphi(-u) = \overline{\varphi(u)}$) and extend analytically to a strip, so the midpoint rule on $(0, \infty)$, the trapezoidal rule on $\mathbb{R}$, converges geometrically, as in Section 2.3. With step h the rule inverts the periodisation, with period $2\pi/h$, of the function of k being recovered: for (i) the distribution of X_T, for (ii) the damped price c, which decays only like $e^{\alpha k}$ as $k \to -\infty$ and $e^{-(a_+ - 1 - \alpha)k}$ as $k \to \infty$; a small damping needs a fine step. The truncation point is set by the decay of φ, about $e^{-\bar\rho(V_0 + \kappa\theta T)u/\xi}$ by (i): slow for large ξ, short maturities and small variance, and lost as $|\rho| \to 1$.

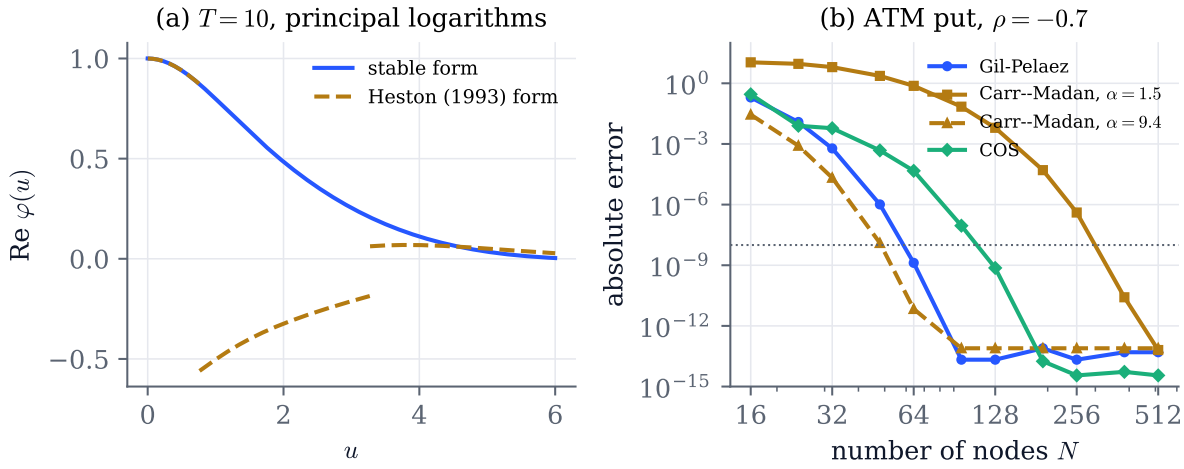

Figure 19.1: (a) $\mathrm{Re}\,\varphi(u)$ for $S = 1$, $r = y = 0$, $T = 10$ and the Eurostoxx parameters of Fang and Oosterlee (2008), from (19.6) and from Heston's form, both with principal logarithms. (b) Absolute error of the at-the-money put in the test case with $\rho = -0.7$ against the number of nodes N: Gil-Pelaez and Carr–Madan ($\alpha = 1.5$ and $\alpha = 9.4$) with the midpoint rule on $[0, 125]$, and COS on the Chernoff interval $[-4.53, 1.35]$. The dotted line is 10^{-8}.

The laboratory. Take the test case $S = K = 100$, $T = 1$, $r = 1\%$, $y = 2\%$ with $V_0 = \theta = 0.04$, $\kappa = 2$, $\xi = 0.3$ (Feller's ratio 1.78), the parameters of Chapter 10. The three methods, each run to convergence, give the at-the-money put

$$P = 8.153429952752 \quad (\rho = 0), \qquad P = 7.948021924422 \quad (\rho = -0.7),$$

and agree to 10^{-13}: these are the prices 8.1534 and 7.9480 of Chapter 10. In Figure 19.1(b), Gil-Pelaez reaches 10^{-8} with 64 nodes and COS with 128 terms; Carr–Madan needs 384 nodes with $\alpha = 1.5$ and 64 with $\alpha = 9.4$, a quarter of $a_+(1) - 1$. We also reproduce the Eurostoxx prices

of Fang and Oosterlee (2008), Section 5.3 ($r = y = 0$): 22.318945791 at $T = 10$, to all printed digits, and 5.785155434 at $T = 1$, 1.6×10^{-8} below their reference computed by Carr–Madan's method. On their interval of twelve exact standard deviations, our COS put, and the call by parity, stall at an error of 3.9×10^{-5} at $T = 1$, whatever N: the left tail of Y is heavier than that. The call priced directly on the same interval converges to 8×10^{-9}, and their table (Section 5.3) reports errors down to 3×10^{-7} with parity, a discrepancy we cannot resolve. The Chernoff interval $[-7.40, 2.56]$ removes the error for both.

▶ Interactive demo: the characteristic function in both forms, and the three inversions for any damping.

19.4 Simulation

Path-dependent payoffs (Chapters 15 and 16) and American options (Theorem 14.6) need paths of (S, V). The variance can be sampled exactly by its noncentral chi-square transitions (Theorem 19.2(iii); `stratified.cir_exact`), and given its path the log-price is normal (Theorem 10.12), with moments determined by Q_t and $\int_0^t \sqrt{V}\,\mathrm{d}W = (V_t - V_0 - \kappa\theta t + \kappa Q_t)/\xi$. Broadie and Kaya (2006) sample V_T, then Q_T given (V_0, V_T) by inverting its conditional characteristic function (built from modified Bessel functions, with branch issues of their own, Lord and Kahl, 2006, Section 5.3), then S_T: exact but slow. Chapter 10 instead averages the Black–Scholes value given (V_T, Q_T), which removes the noise of $W^\perp$ (Theorem 10.11), with Q_T by the trapezoidal rule on 64 exact variance steps.

Discretisation schemes are faster. Euler's scheme does not apply as it stands: its variance can become negative, and Theorem 7.10 does not cover $\sqrt{v}$, which is not Lipschitz at 0. Lord et al. (2010) write the fixes in one form, with $\tilde{V}_0 = V_0$, step h and $\Delta W_k \sim \mathcal{N}(0, h)$:

$$\tilde{V}_{k+1} = f_1(\tilde{V}_k) + \kappa(\theta - f_2(\tilde{V}_k))h + \xi\sqrt{f_3(\tilde{V}_k)}\,\Delta W_k, \qquad V_k = f_3(\tilde{V}_k).$$

Absorption takes $f_1 = f_2 = f_3 = (\cdot)^+$; reflection $f_1 = f_2 = f_3 = |\cdot|$; partial truncation $f_1 = f_2 = \mathrm{id}$, $f_3 = (\cdot)^+$; and their *full truncation* $f_1 = \mathrm{id}$, $f_2 = f_3 = (\cdot)^+$: the auxiliary $\tilde{V}$ may go negative, but the drift and the diffusion only ever see $\tilde{V}^+$.

Proposition 19.8 (Full truncation)**.** *Let $h = T/m$, $t_k = kh$, and let the full-truncation scheme for V be combined with the log-Euler step $\ln \bar{S}_{k+1} = \ln \bar{S}_k + (r - y - \frac{1}{2}V_k)h + \sqrt{V_k}\,(\rho\Delta W_k + \bar{\rho}\Delta W_k^\perp)$, $\bar{S}_0 = S$.*

(i) (Exact forward.) *For every h, $\mathbb{E}[\bar{S}_{k+1} \mid \mathcal{F}_{t_k}] = e^{(r-y)h}\bar{S}_k$, so $\mathbb{E}\bar{S}_m = F$.*

(ii) (Conditioning.) *Given W, $\ln \bar{S}_m$ is normal with mean $\ln S + (r - y)T - \frac{1}{2}\bar{Q} + \rho\bar{I}$ and variance $\bar{\rho}^2\bar{Q}$, where $\bar{Q} = \sum_k V_k h$ and $\bar{I} = \sum_k \sqrt{V_k}\Delta W_k$; the Black–Scholes value at spot $Se^{\rho\bar{I} - \rho^2\bar{Q}/2}$ and volatility $\bar{\rho}\sqrt{\bar{Q}/T}$ is an unbiased estimator of the scheme's price, with no larger variance.*

(iii) (Proved convergence.) *Lord et al. (2010), Section 3, prove that the full-truncation approximation of V converges strongly in L^1 as $h \to 0$, and that Monte Carlo prices of vanilla and barrier options converge for the Euler (not log-Euler) discretisation of S, stopped when S or V leaves a bounded set.*

(iv) (Not proved.) *No rate is known in general. Cozma and Reisinger (2020), Theorem 1.1,*

prove strong order $\frac{1}{2}$ in L^p for the variance when $2\kappa\theta/\xi^2 > 3$ and $2 \le p < 2\kappa\theta/\xi^2 - 1$, a regime that excludes every parameter set of this chapter. No weak order is proved; Lord et al. (2010) estimate it numerically, close to 1 in their Heston test.

Proof. (i) Given $\mathcal{F}_{t_k}$, V_k is known and $G_k = \sqrt{V_k}(\rho\Delta W_k + \bar\rho\Delta W_k^\perp) \sim \mathcal{N}(0, V_k h)$, so $\mathbb{E}[e^{-V_k h/2 + G_k} \mid \mathcal{F}_{t_k}] = 1$. Iterate with the tower property. (ii) V_k is a function of the increments of W only; this is Theorem 10.12(iv), and Theorem 10.11 for the variance. □

Full truncation preserves the forward exactly but not the law of V, and the error in that law near zero is large when Feller's condition fails badly (Figure 19.2(b)). Andersen (2008) replaces the Euler step by a draw from a law that matches the first two conditional moments (19.2) exactly, and the log-price step by one that uses the identity for $\int \sqrt{V}\,\mathrm{d}W$ above.

Proposition 19.9 (The quadratic-exponential scheme)**.** *Fix $h > 0$ and $v \ge 0$, let m and s^2 be the conditional mean and variance (19.2) of V_{t+h} given $V_t = v$, and $\psi = s^2/m^2$. Let $Z \sim \mathcal{N}(0,1)$ and U uniform on $(0,1)$.*

(i) (Quadratic branch.) *If $\psi \le 2$, let $b^2 = 2\psi^{-1} - 1 + \sqrt{2\psi^{-1}}\sqrt{2\psi^{-1} - 1}$ and $a = m/(1+b^2)$. Then $a(b+Z)^2$ has mean m and variance s^2.*

(ii) (Exponential branch.) *If $\psi \ge 1$, let $p = (\psi - 1)/(\psi + 1) \in [0,1)$ and $\beta = (1-p)/m$. Then the variable equal to 0 if $U \le p$ and to $\beta^{-1}\ln\frac{1-p}{1-U}$ otherwise has the law $p\,\delta_0 + (1-p)\,\mathrm{Exp}(\beta)$, mean m and variance s^2.*

(iii) (Switching.) *For any $\psi_c \in [1,2]$, the step that uses (i) when $\psi \le \psi_c$ and (ii) otherwise matches the conditional mean and variance of V_{t+h} exactly, from every state $v \ge 0$ and for every h.*

(iv) (Martingale correction.) *Let the log-price step be*

$$\ln \hat S_{t+h} = \ln \hat S_t + (r-y)h + K_0 + K_1 v + K_2 \hat V_{t+h} + \sqrt{K_3 v + K_4 \hat V_{t+h}}\, Z_S,$$

with $Z_S \sim \mathcal{N}(0,1)$ independent of the variance step, $\gamma_1 = \gamma_2 = \frac{1}{2}$, $K_1 = \gamma_1 h(\frac{\kappa\rho}{\xi} - \frac{1}{2}) - \frac{\rho}{\xi}$, $K_2 = \gamma_2 h(\frac{\kappa\rho}{\xi} - \frac{1}{2}) + \frac{\rho}{\xi}$, $K_3 = \gamma_1 h\bar\rho^2$, $K_4 = \gamma_2 h\bar\rho^2$, and $A = K_2 + \frac{1}{2}K_4$. Then $\mathbb{E}[\hat S_{t+h} \mid \hat S_t, v] = e^{(r-y)h}\hat S_t$ if and only if $K_0 = -\ln M - (K_1 + \frac{1}{2}K_3)v$, where $M = \mathbb{E}[e^{A\hat V_{t+h}}]$ is

$$M = \frac{\exp\left(Ab^2 a/(1-2Aa)\right)}{\sqrt{1-2Aa}} \text{ on (i), if } Aa < \tfrac{1}{2}; \qquad M = p + \frac{\beta(1-p)}{\beta - A} \text{ on (ii), if } A < \beta;$$

and $M = \infty$ otherwise. If $\rho \le 0$, then $A \le 0$ and both conditions hold.

Proof. (i) $\mathbb{E}(b+Z)^2 = 1 + b^2$ and $\mathrm{Var}(b+Z)^2 = 4b^2 + 2$, so the mean is $a(1+b^2) = m$ and the variance matches if and only if $x = b^2 \ge 0$ solves $(4x+2)/(1+x)^2 = \psi$, that is $\psi x^2 + (2\psi - 4)x + \psi - 2 = 0$, whose roots, real for $\psi \le 2$, are $2\psi^{-1} - 1 \pm \sqrt{2\psi^{-1}}\sqrt{2\psi^{-1}-1}$; the larger is nonnegative. (ii) $\mathbb{Q}(\hat V > x) = \mathbb{Q}(U > 1 - (1-p)e^{-\beta x}) = (1-p)e^{-\beta x}$ for $x \ge 0$. The mean is $(1-p)/\beta = m$; the second moment is $2(1-p)/\beta^2$, so the variance is $(1-p)(1+p)/\beta^2 = \psi m^2$. (iii) follows. (iv) Given v and $\hat V_{t+h}$, the log-price is normal, so $\mathbb{E}[\hat S_{t+h} \mid \hat S_t, v, \hat V_{t+h}] = \hat S_t e^{(r-y)h}\exp\left(K_0 + (K_1 + \frac{1}{2}K_3)v + A\hat V_{t+h}\right)$; take the expectation over $\hat V_{t+h}$. On (i), the Gaussian integral $\mathbb{E}e^{\lambda(b+Z)^2} = (1-2\lambda)^{-1/2}e^{\lambda b^2/(1-2\lambda)}$ for $\lambda < \frac{1}{2}$ (and ∞ otherwise),

with $\lambda = Aa$; on (ii), $\mathbb{E}e^{A\hat{V}} = p+(1-p)\beta/(\beta-A)$ for $A < \beta$. Finally $A = \frac{\rho}{\xi}(1+\kappa\gamma_2 h)-\frac{1}{2}\gamma_2 h\rho^2 \le 0$ for $\rho \le 0$, while $a > 0$ and $\beta > 0$. □

The K_j come from the exact identity $\ln S_{t+h} = \ln S_t + (r-y)h - \frac{1}{2}\int V + \frac{\rho}{\xi}(V_{t+h} - V_t - \kappa\theta h + \kappa\int V) + \bar{\rho}\int\sqrt{V}\,\mathrm{d}W^{\perp}$ with $\int_t^{t+h} V \approx h(\gamma_1 V_t + \gamma_2 V_{t+h})$; uncorrected, $K_0 = -\rho\kappa\theta h/\xi$ (Andersen (2008), Sections 3.2 and 4.3 of the preprint, with $\psi_c = 1.5$). Proved are moment matching and the exact forward; the third moment of V and the joint law of $(V_{t+h}, \int V)$ are not matched. Andersen proves weak consistency (preprint, Section 4.4, Proposition 10); beyond that no convergence theorem for the scheme is known to us. For $\rho > 0$ and long steps (iv) can fail: then $\mathbb{E}\hat{S}_{t+h} = \infty$ after one step. Given the variance path the QE log-price is normal, so the conditioning of Theorem 19.8(ii) applies too.

The bias in the laboratory. Figure 19.2 measures the bias of the at-the-money put against its Fourier price for $h = 1, \frac{1}{2}, \ldots, \frac{1}{32}$. Each run uses the conditional estimator, which has the scheme's mean, with the forward $\mathbb{E}[\bar{S}_m \mid W]e^{-(r-y)T}$, of mean S by Theorem 19.8(i) and Theorem 19.9(iv), as a control variate with an estimated coefficient (Theorem 8.12; its $O(1/n)$ bias, Theorem 8.14, is negligible); together they cut the standard error of the plain payoff by a factor of about five. Our implementation (numpy `default_rng`, seeds $[19, i, j]$ for step i and scheme j) uses 2^{19} paths in the test case and 2^{18} in the second case. In the test case ($\rho = -0.7$, reference 7.948022, standard errors about 0.003), full truncation is biased by 0.397, 0.305, 0.085, 0.025, 0.011 and 0.003, a decay of at least order one from $h = \frac{1}{4}$ on, while QE is off by 0.020 at $h = 1$ and by at most 0.011 at smaller steps. With the parameters of Andersen (2008)'s Case I ($V_0 = \theta = 0.04$, $\kappa = 0.5$, $\xi = 1$, $\rho = -0.9$, Feller's ratio 0.04; here with $T = 1$ and the test-case r, y; reference 4.576954), full truncation is still 0.54 too high at $h = \frac{1}{32}$, 12% of the price, while QE's bias is 0.094 at $h = 1$, -0.021 at $h = \frac{1}{4}$ and -0.001 at $h = \frac{1}{32}$: there QE at the coarsest step beats full truncation at the finest. Multilevel estimators also need coupled fine and coarse paths (Theorem 12.4), immediate for Euler but not for QE.

19.5 Smiles

Figure 19.3 inverts Heston prices into Black–Scholes implied volatilities (Theorem 2.9) at $T = 1$. The correlation sets the skew: $\rho < 0$ gives the downward-sloping equity smile, $\rho > 0$ its mirror image, and $\rho = 0$ a smile with its minimum near the money. The volatility of variance ξ sets the curvature and, with $\rho \neq 0$, amplifies the skew. Mean reversion κ damps both: over a maturity of a few $1/\kappa$ the variance forgets its starting point, its integrated value Q_T concentrates, and the smile flattens. For short maturities a small-time expansion gives the at-the-money skew $\partial\sigma_{\mathrm{imp}}/\partial \ln K \to \rho\xi/(4\sqrt{V_0})$ (Gatheral, 2006, Chapter 3); we use it only as a heuristic, and in the test case with $\rho = -0.7$ it predicts -0.2625, while finite differences of Fourier prices give -0.261 at $T = 0.01$ and -0.146 at $T = 1$.

Calibration reverses the map by least squares on implied volatilities. Identification is the hard part: at one maturity V_0, κ and θ enter mostly through the expected integrated variance, and κ and ξ trade off in the curvature (Exercise 19.7); fits often violate Feller's condition (Figure 19.2(b)).

▶ Interactive demo: smiles for any parameters, and both schemes for any step.

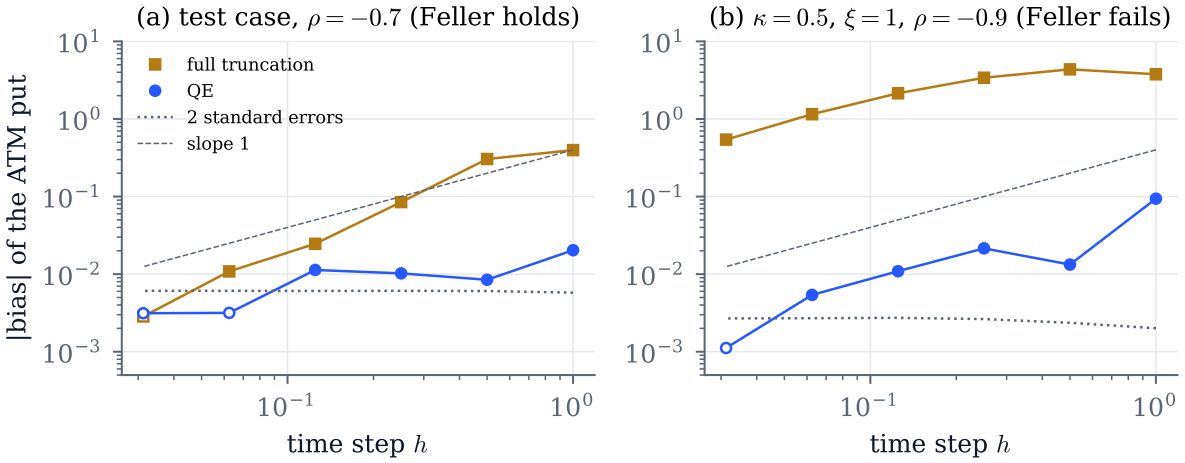

Figure 19.2: Absolute bias of the at-the-money put ($S = K = 100$, $T = 1$, $r = 1\%$, $y = 2\%$) against the time step h, for full truncation (log-Euler for S) and QE with the martingale correction, against the Fourier price. (a) Test case, $\rho = -0.7$, 2^{19} paths per point. (b) $V_0 = \theta = 0.04$, $\kappa = 0.5$, $\xi = 1$, $\rho = -0.9$, 2^{18} paths per point. Open markers: |bias| below twice its standard error. Dotted: twice the standard error of the QE estimates. Dashed: slope one.

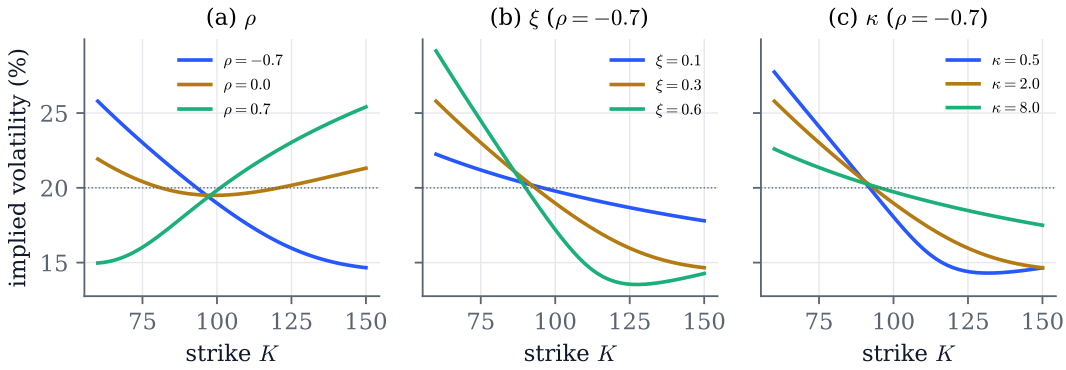

Figure 19.3: Implied volatility of Heston prices against the strike, $T = 1$, test-case S, r, y, $V_0 = \theta = 0.04$; out-of-the-money options. Base parameters $\kappa = 2$, $\xi = 0.3$, $\rho = -0.7$; each panel varies one of them. (a) $\rho \in \{-0.7, 0, 0.7\}$; (b) $\xi \in \{0.1, 0.3, 0.6\}$; (c) $\kappa \in \{0.5, 2, 8\}$. Dotted: $\sqrt{V_0} = 20\%$.

In code

code/src/mfe230d/heston.py: `charfn`, `moment_bound` (Theorem 19.5); `price` and the three inversions (Theorem 19.7); `qe_step`, `simulate`, `bias_study`.

Further reading

Gatheral (2006), Chapters 2 and 3, treats the model and its implied volatility surface.

Chapter 20

Trees and finite differences

Monte Carlo samples paths forward in time. Grids compute the same expectation backwards: they discretize the state, replace the conditional expectation over one time step by a weighted average of neighbouring values, and step from the payoff to the price today. A *tree* takes the weights from a discrete-time model that converges to the diffusion; a *finite-difference scheme* from difference quotients in the pricing equation. The two meet in the explicit scheme, a trinomial tree up to its discounting. Grids give the price for all spots at once, the Greeks almost for free, and early exercise for one maximum per node. This is why Chapter 14 took its American reference values from a smoothed binomial tree, studied in Section 20.1, and its Bermudan ones from a backward induction on a grid in log price with the exact Gaussian transition density, a quadrature cousin of the schemes of Section 20.3. Their cost grows exponentially with the number of state variables (Theorem 6.14), so they are the method of choice in one or two dimensions and of no use in ten.

Throughout, S is the stock with yield y of Theorem 1.1 under $\mathbb{Q}$, with log-drift $\nu = r - y - \sigma^2/2$, and $r \geq 0$. The price at time to maturity τ of the European payoff $g(S_T)$, as a function of the log price $x = \log S$, is $u(\tau, x) = e^{-r\tau}\mathbb{E}^{\mathbb{Q}}[g(e^{x+\nu\tau+\sigma\sqrt{\tau}Z})]$ with $Z \sim \mathcal{N}(0,1)$, and by Theorem 1.6 (whose code checks it with difference quotients) it solves

$$\partial_\tau u = \tfrac{1}{2}\sigma^2 \partial_{xx} u + \nu\, \partial_x u - ru, \qquad u(0,x) = g(e^x), \tag{20.1}$$

whenever g is regular enough for the equation to make sense. The running example is the course's test case $S_0 = K = 100$, $T = 1$, $r = 1\%$, $y = 2\%$, $\sigma = 20\%$, with call 7.36429 and put 8.34941.

20.1 Binomial trees

Definition 20.1 (Binomial tree)**.** Fix $n \geq 1$ steps of length $\Delta t = T/n$ and factors $0 < d < u$. The tree has the nodes $S_0 u^j d^{i-j}$, $0 \leq j \leq i \leq n$, at time $i\Delta t$, and the *risk-neutral probability* $q = (e^{(r-y)\Delta t} - d)/(u-d)$. The values $V_n(j) = g(S_0 u^j d^{n-j})$ are propagated backwards by

$$V_i(j) = e^{-r\Delta t}\big[q\, V_{i+1}(j+1) + (1-q)\, V_{i+1}(j)\big], \tag{20.2}$$

and the tree price is $V_0(0)$; for an American option each $V_i(j)$ is replaced by its maximum with the payoff $g(S_0 u^j d^{i-j})$. The *Cox–Ross–Rubinstein* (CRR) tree has $u = e^{\sigma\sqrt{\Delta t}}$ and $d = 1/u$ (Cox et al., 1979); the Jarrow–Rudd tree has $u, d = e^{\nu\Delta t \pm \sigma\sqrt{\Delta t}}$ (Jarrow and Rudd, 1983).

The tree is a model in its own right when $q \in (0,1)$, that is when $d < e^{(r-y)\Delta t} < u$: this is the no-arbitrage condition of the one-period model, and $V_i(j)$ is then the price in the binomial model whose stock moves by the factor u or d and pays dividends at the rate y (the IEOR 222 notes treat $y = 0$). For the European payoff, (20.2) unrolls into $V_0(0) = e^{-rT}\sum_{j=0}^{n} \binom{n}{j} q^j (1-q)^{n-j} g(S_0 u^j d^{n-j})$, which our code evaluates in $O(n)$ operations. For the CRR tree the terminal log prices are $\log S_0 + (2j-n)\delta$ with $\delta = \sigma\sqrt{T/n}$, a lattice of spacing 2δ, and we write

$$\theta_n = \operatorname{frac}\Big(\frac{\log(K/S_0) + n\delta}{2\delta}\Big) \in [0,1)$$

for the position of $\log K$ between two consecutive terminal nodes: $\theta_n = 0$ when K is a node. At $S_0 = K$, $\theta_n = 0$ for even n and $\theta_n = \tfrac{1}{2}$ for odd n.

Proposition 20.2 (Binomial trees)**.** *Let $\sigma > 0$ and let C_n be the CRR price of a European payoff g.*

(i) Validity. *$q \in (0,1)$ if and only if $n > (r-y)^2 T/\sigma^2$; if $r = y$, $q = 1/(1+u)$ for every n. As $n \to \infty$, $q = \tfrac{1}{2} + \tfrac{\nu}{2\sigma}\sqrt{\Delta t} + O(\Delta t)$.*

(ii) Convergence. *If g is continuous with $|g(s)| \leq a(1+s)$, then $C_n \to e^{-rT}\mathbb{E}^{\mathbb{Q}}[g(S_T)]$. The same holds for a bounded g whose discontinuities form a finite set (digitals).*

(iii) Oscillation. *For the call with strike $K > 0$, $C_n - C = O(1/n)$, where C is the Black–Scholes price, but $n(C_n - C)$ has no limit in general:*

$$n\,(C_n - C) = A + B\,\theta_n(1-\theta_n) + o(1), \qquad B = 2\sigma\sqrt{T}\,Ke^{-rT}N'(d_-), \tag{20.3}$$

with a constant A depending on $(S_0, K, T, r, y, \sigma)$ only.

(iv) Better trees. *(a) The BBS tree of Broadie and Detemple (1996) replaces the values one step before maturity by the Black–Scholes prices with maturity Δt; BBSR is the extrapolation $2C_n^{\mathrm{BBS}} - C_{n/2}^{\mathrm{BBS}}$. (b) The tree of Leisen and Reimer (1996) takes n odd, $q = h_n^{-1}(d_-)$, $q' = h_n^{-1}(d_+)$ with the Peizer–Pratt inversion $h_n^{-1}(z) = \tfrac{1}{2} + \tfrac{1}{2}\operatorname{sign}(z)\big[1 - \exp\big(-\big(\tfrac{z}{n+1/3+0.1/(n+1)}\big)^2(n+\tfrac{1}{6})\big)\big]^{1/2}$ (Peizer and Pratt, 1968), $u = e^{(r-y)\Delta t}q'/q$ and $d = (e^{(r-y)\Delta t} - qu)/(1-q)$. Its call prices converge at order $1/n^2$, without oscillation.*

Proof. (i) $q \in (0,1)$ if and only if $e^{-\sigma\sqrt{\Delta t}} < e^{(r-y)\Delta t} < e^{\sigma\sqrt{\Delta t}}$, that is $|r-y|\Delta t < \sigma\sqrt{\Delta t}$, that is $n > (r-y)^2 T/\sigma^2$. If $r = y$, $q = (1-u^{-1})/(u-u^{-1}) = 1/(1+u)$. With $\delta = \sigma\sqrt{\Delta t}$, the numerator of q is $\delta - \tfrac{1}{2}\delta^2 + (r-y)\Delta t + O(\Delta t^{3/2}) = \delta + \nu\Delta t + O(\Delta t^{3/2})$ and the denominator is $2\delta + O(\delta^3)$.

(ii) $\log(S_n/S_0) = X_n = \xi_1 + \cdots + \xi_n$ with independent $\xi_i = \pm\delta$, $\mathbb{Q}(\xi_i = \delta) = q$. By (i), $\mathbb{E}[\xi_i] = (2q-1)\delta = \nu\Delta t + O(\Delta t^{3/2})$ and $\mathbb{E}[\xi_i^2] = \sigma^2\Delta t$, so, as $|\xi_i| \leq \delta$, $\mathbb{E}[e^{is\xi_i}] = 1 + is\nu\Delta t - \tfrac{1}{2}s^2\sigma^2\Delta t + O(\Delta t^{3/2})$

for each real s, and $\mathbb{E}[e^{isX_n}] = \mathbb{E}[e^{is\xi_1}]^n \to e^{is\nu T - s^2\sigma^2 T/2}$. By Lévy's continuity theorem X_n converges in law to $\mathcal{N}(\nu T, \sigma^2 T)$, so S_n to S_T. Likewise $\mathbb{E}[S_n^2] = S_0^2(qu^2 + (1-q)u^{-2})^n = S_0^2(1 + (2\nu + 2\sigma^2)\Delta t + O(\Delta t^{3/2}))^n$ is bounded in n, so $(g(S_n))_n$ is bounded in L^2, hence uniformly integrable. By the mapping theorem, $g(S_n)$ converges in law to $g(S_T)$ when the discontinuities of g have probability zero under the law of S_T, which has a density (Billingsley, 1999, Theorem 2.7); uniform integrability then gives convergence of the means (Billingsley, 1999, Theorem 3.5).

(iii) and (iv) are cited. Diener and Diener (2004) compute the expansion for the CRR tree; Walsh (2003), Corollary 4.1, proves (20.3) with this B (his $2\sigma^2 T\tilde{K}\hat{p}$, with $\tilde{K} = Ke^{-rT}$) for n even in the CRR tree of the discounted price, where θ_n is the position of the discounted strike. For (iv)(b), Leisen and Reimer (1996) observed the order $1/n^2$ and conjectured it, and Joshi (2010) proved it; part (iv)(a) is a definition, discussed after Figure 20.1. □

Remark 20.3 (Where the oscillation comes from). By a local central limit theorem the tree puts mass close to $H\,f(x_j)$ on the terminal node x_j, where $H = 2\delta$ and f is a smooth density close to that of $\log S_T$. The call is then $e^{-rT}H\sum_j F(x_j)$ with $F(x) = f(x)(e^x - K)^+$, a function that vanishes at $\log K$ with slope $Kf(\log K)$ to the right. The Euler–Maclaurin formula for a lattice whose first point lies αH above c, $H\sum_{m\ge 0} F(c + (m+\alpha)H) = \int_c^\infty F - HB_1(\alpha)F(c) - \frac{H^2}{2}B_2(\alpha)F'(c) + O(H^3)$, with the Bernoulli polynomials $B_1(\alpha) = \alpha - \frac{1}{2}$ and $B_2(\alpha) = \alpha^2 - \alpha + \frac{1}{6}$, applied at $c = \log K$ with $\alpha = 1 - \theta_n$ and $B_2(1-\theta_n) = B_2(\theta_n)$, leaves $e^{-rT}\frac{H^2}{2}\theta_n(1-\theta_n)Kf(\log K)$ besides terms that do not depend on θ_n; with $H^2 = 4\sigma^2 T/n$ and $f(\log K) = N'(d_-)/(\sigma\sqrt{T})$ this is the B of (20.3). The argument does not control the error of the local expansion (Walsh's proof does, by Skorokhod embedding), but it shows where B comes from, and it applies to the CRR tree as defined here: on the test case $B = 15.6222$, and fitting $n(C_n - C)$ on 1 and $\theta_n(1-\theta_n)$ over $n = 1000, \dots, 2000$ gives $A = -1.964$, $B = 15.6222$; for $K = 110$, $B = 14.2815$ against fits of 14.2859 ($n = 1000, \dots, 2000$) and 14.2800 ($n = 8000, \dots, 8400$). For a digital the same computation leaves the first-order term $-HB_1(\alpha)F(c)$: its error oscillates at order $n^{-1/2}$ (Exercise 20.3).

Parity, smoothing, extrapolation. At $S_0 = K$ the two parities give two smooth sequences, $nC_n - nC \to A = -1.964$ for n even and $A + B/4 = 1.942$ for n odd: the CRR price is below the Black–Scholes price for even n and above it for odd n (Figure 20.1(a)), and Richardson extrapolation across consecutive n fails. Away from the money θ_n wanders through $[0, 1)$, and doubling n can increase the error. The remedies of (iv) attack the kink of the payoff. The BBS tree, whose last step is smooth on the scale of the lattice, does not remove the oscillation, but on the test case it shrinks the parity effect from $B/4 = 3.906$ to 0.023 in units of $1/n$ ($nC_n^{\mathrm{BBS}} - nC$ is 0.6498 at $n = 400$ and 0.6266 at $n = 401$). At the money BBSR is then of order $1/n^2$ for n divisible by 4 (so that n and $n/2$ have the same parity) and of order $1/n$ for $n \equiv 2 \pmod 4$ (n times its error is 0.061 at $n = 102$, 0.047 at $n = 1602$). For $K = 110$, n times the BBS error moves between 0.467 and 0.488, and BBSR is first order (n^2 times its error: 5.8, −1.6, −28.0 at $n = 400$, 800, 1600). The Leisen–Reimer tree places the strike between the two middle terminal nodes for every odd n (we checked this for strikes from 80 to 130) and chooses q and q' so that the binomial probabilities of finishing above the strike, under the two measures of the call formula, approximate $N(d_-)$ and $N(d_+)$: $n^2(C_n - C) = -0.3955$ at $n = 201$ and -0.3978 at $n = 801$ (Figure 20.1(b)). Neither survives early exercise unchanged: on the first Longstaff–Schwartz put of Table 20.1, the Leisen–Reimer error halves when n doubles (-4.8×10^{-3} at $n = 101$,

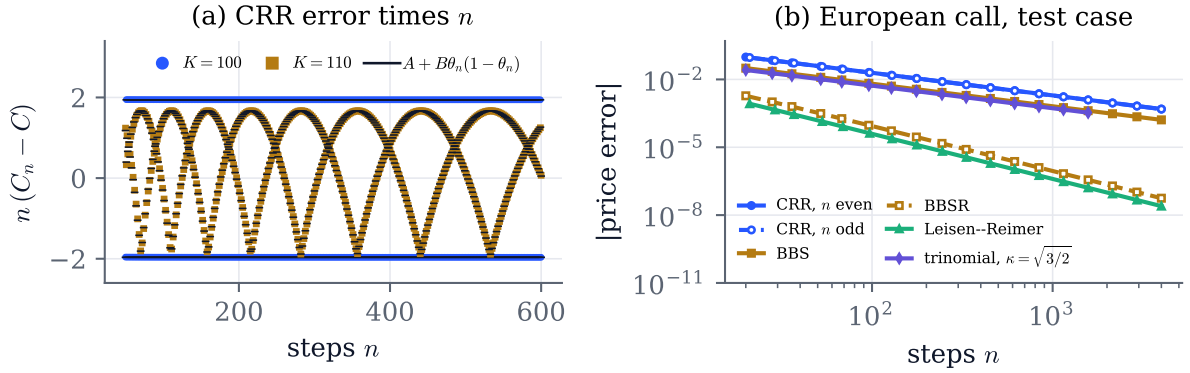

Figure 20.1: CRR trees on the test case. (a) $n(C_n - C)$ for the calls with $K = 100$ (two branches: n even and odd) and $K = 110$; the black marks are $A + B\theta_n(1-\theta_n)$ with B from (20.3) and A fitted on $n = 1000, \dots, 2000$. (b) Absolute errors of the European call: CRR for n even and odd, BBS for n even, BBSR and Leisen–Reimer, and the trinomial tree of Section 20.2 with $\kappa = \sqrt{3/2}$; n runs over multiples of 4 between 20 and 4000 ($n+1$ for the odd-n trees; up to 2000 for the trinomial tree).

-2.9×10^{-4} at $n = 1601$), because the American price is not smooth across the exercise boundary.

American options on trees. With the projection $V_i(j) \leftarrow \max(V_i(j), g)$ the CRR tree converges to the American price; for the put, Lamberton (2020), Theorem 1.1 (arXiv version), bounds the error by a constant times $(\log n)^{3/2}/n$ ($(\log n)/n$ if $y > r$). The American references of Chapter 14 are BBSR trees (its code and ours agree to rounding); Table 20.1 compares them with the grids of Section 20.5.

20.2 Trinomial trees and the explicit scheme

A trinomial tree moves the log price by $+h$, 0 or $-h$ in each step. Following Kamrad and Ritchken (1991) (and the three-jump process of Boyle (1986)), take $h = \kappa\sigma\sqrt{\Delta t}$ with a *stretch* $\kappa > 0$ and

$$p_\pm = \frac{1}{2\kappa^2} \pm \frac{\nu\sqrt{\Delta t}}{2\kappa\sigma}, \qquad p_0 = 1 - \frac{1}{\kappa^2}, \tag{20.4}$$

with backward induction $V \leftarrow e^{-r\Delta t}(p_+ V_+ + p_0 V_0 + p_- V_-)$ on the nodes $\log S_0 + mh$, $|m| \le i$, at step i. The tree needs no boundary condition: in n steps it reaches n nodes away.

Proposition 20.4 (Trinomial trees). *Let $\kappa > 0$, $\Delta t = T/n$ and $h = \kappa\sigma\sqrt{\Delta t}$.*

(i) Moments and positivity. *An increment ξ has $\mathbb{E}[\xi] = \nu\Delta t$ and $\mathbb{E}[\xi^2] = \sigma^2\Delta t$ exactly. The three weights are probabilities if and only if $\kappa \ge 1$ and $h|\nu| \le \sigma^2$. If $\nu = 0$, then $\mathbb{E}[\xi^4] = \kappa^2\sigma^4\Delta t^2$, which equals the Gaussian value $3\sigma^4\Delta t^2$ if and only if $\kappa = \sqrt{3}$, and then $(p_+, p_0, p_-) = (\frac{1}{6}, \frac{2}{3}, \frac{1}{6})$.*

(ii) The explicit scheme. *The explicit finite-difference scheme for (20.1) with time step $k = \Delta t$ and mesh h (Section 20.3) is $U_j^{n+1} = p_+ U_{j+1}^n + p_0 U_j^n + p_- U_{j-1}^n - rk\,U_j^n$ with*

the weights (20.4): the tree with the discount e^{-rk} replaced by a subtraction of rkU_j^n, which differs by $O(k^2)$ per step. Its parameter $\sigma^2 k/(2h^2)$ equals $1/(2\kappa^2)$, so $\kappa \geq 1$ is the stability condition $\sigma^2 k \leq h^2$ of Theorem 20.6(i).

(iii) Convergence. *If $\kappa \geq 1$ and $h|\nu| \leq \sigma^2$, the tree is a positive operator with $\|V\|_\infty$ nonincreasing from step to step, and if $\partial_\tau^i u$ ($i \leq 2$) and $\partial_x^j u$ ($j \leq 4$) are continuous and bounded on $[0,T]\times\mathbb{R}$, its error is at most $cT\Delta t$ uniformly in the nodes, with c depending on these bounds and on κ. If $\kappa < 1$, the backward induction multiplies the alternating sequence $(-1)^m$ by a factor tending to $2/\kappa^2 - 1 > 1$ in modulus as $\Delta t \to 0$.*

Proof. (i) $\mathbb{E}[\xi] = h(p_+ - p_-) = \kappa\sigma\sqrt{\Delta t}\,\nu\sqrt{\Delta t}/(\kappa\sigma)$ and $\mathbb{E}[\xi^2] = h^2(p_+ + p_-) = h^2/\kappa^2$. $p_0 \geq 0$ if and only if $\kappa \geq 1$, and $p_\pm \geq 0$ if and only if $\sigma \geq \kappa|\nu|\sqrt{\Delta t}$, that is $\sigma^2 \geq |\nu|h$. If $\nu = 0$, $\mathbb{E}[\xi^4] = h^4/\kappa^2 = \kappa^2\sigma^4\Delta t^2$. (ii) With $p_\pm^{\mathrm{fd}} = \sigma^2/(2h^2) \pm \nu/(2h)$ and $p_0^{\mathrm{fd}} = -\sigma^2/h^2 - r$ (Section 20.3), the explicit step is $U^{n+1} = U^n + k(p_-^{\mathrm{fd}}U_{j-1}^n + p_0^{\mathrm{fd}}U_j^n + p_+^{\mathrm{fd}}U_{j+1}^n)$, and $kp_\pm^{\mathrm{fd}} = \sigma^2\Delta t/(2\kappa^2\sigma^2\Delta t) \pm \nu\Delta t/(2\kappa\sigma\sqrt{\Delta t}) = p_\pm$, $1 + kp_0^{\mathrm{fd}} = p_0 - rk$. (iii) Under the conditions the weights are nonnegative with sum $e^{-r\Delta t} \leq 1$, which gives the first claims. Taylor's formula with $\mathbb{E}[\xi]$, $\mathbb{E}[\xi^2]$ from (i), $|\mathbb{E}[\xi^3]| = h^2|\mathbb{E}[\xi]|$ and $\mathbb{E}[\xi^4] = h^2\sigma^2\Delta t$ gives $e^{-r\Delta t}\mathbb{E}[u(\tau, x+\xi)] = u(\tau+\Delta t, x) + O(\Delta t^2)$, with a constant depending on the bounds; a positive step with weights summing to at most one adds these local errors without amplifying them, and $n\Delta t = T$. If $\kappa < 1$, one step maps $(-1)^m$ to $(e^{-r\Delta t}(p_0 - p_+ - p_-))(-1)^m$, and $p_0 - p_+ - p_- = 1 - 2/\kappa^2 < -1$. □

The identity (ii) is the synthesis of Brennan and Schwartz (1978): an explicit scheme is a jump process (our tests check it to 10^{-11}). For $\kappa = 0.9$ the backward induction returns -3×10^{64} for the call after 400 steps. On the test case $n(C_n - C)$ stays within 0.01 of -1.97, -0.53 and -2.06 for $\kappa = 1$, $\sqrt{3/2}$ and $\sqrt{3}$ and n from 400 to 1600 (the strike stays a node, so there is no oscillation): the choice $\kappa = \sqrt{3}$ of the course's trinomial walk, best for one step of a smooth payoff (see the exercises of Chapter 1), is not the best for a kink.

20.3 Finite differences: stability and order

A finite-difference grid separates what a tree ties together: the mesh h and the time step k. Take J intervals of a domain $[x_0, x_J]$ in log price, nodes $x_j = x_0 + jh$, times $\tau_n = nk$, $N = T/k$, and replace the derivatives in (20.1) by central differences:

$$(L_hV)_j = p_-V_{j-1} + p_0V_j + p_+V_{j+1}, \qquad p_\pm = \frac{\sigma^2}{2h^2} \pm \frac{\nu}{2h}, \qquad p_0 = -\frac{\sigma^2}{h^2} - r.$$

Definition 20.5 (The θ-scheme)**.** For $\theta \in [0,1]$, the θ-scheme for (20.1) is

$$\frac{U_j^{n+1} - U_j^n}{k} = \theta\,(L_hU^{n+1})_j + (1-\theta)\,(L_hU^n)_j, \qquad 0 < j < J, \tag{20.5}$$

with $U_j^0 = g(e^{x_j})$ and boundary values U_0^n, U_J^n (Section 20.4). It is *explicit* for $\theta = 0$, *implicit* for $\theta = 1$, and *Crank–Nicolson* for $\theta = \frac{1}{2}$ (Crank and Nicolson, 1947). We write $\lambda = \sigma^2 k/(2h^2)$. In matrix form, $BV^{n+1} = CV^n + kF^n$ with $B = I - \theta kA$, $C = I + (1-\theta)kA$, A tridiagonal with rows (p_-, p_0, p_+), and F^n carrying the boundary values; the *amplification operator* is $G = B^{-1}C$.

For $\theta > 0$ each step solves a tridiagonal system in $O(J)$ operations. To separate the discretization from the truncation of the domain, the theorem considers the scheme on the infinite grid $h\mathbb{Z}$, with the norms $\|V\|_\infty = \sup_j |V_j|$ and $\|V\|_{2,h} = (h\sum_j V_j^2)^{1/2}$. The scheme is *stable* in a norm, along a family of grids with $k \to 0$, if there is C_s with $\|G^m\| \le C_s$ and $\|G^m B^{-1}\| \le C_s$ for all $mk \le T$; the *truncation error* is $T_j^n = (u_j^{n+1} - u_j^n)/k - \theta(L_h u^{n+1})_j - (1-\theta)(L_h u^n)_j$ with $u_j^n = u(\tau_n, x_j)$.

Theorem 20.6 (Explicit, implicit and Crank–Nicolson schemes)**.** *Let $\sigma > 0$, $r \ge 0$ and $\theta \in [0,1]$, and consider the θ-scheme on $h\mathbb{Z}$.*

(i) Von Neumann stability. *The scheme maps $e^{\mathrm{i}j\xi h}$ to $g(\xi)e^{\mathrm{i}j\xi h}$, with*

$$g(\xi) = \frac{1+(1-\theta)z}{1-\theta z}, \qquad z = k\Big(-\frac{2\sigma^2}{h^2}\sin^2\frac{\xi h}{2} - r + \mathrm{i}\frac{\nu}{h}\sin\xi h\Big), \tag{20.6}$$

and it is stable in $\|\cdot\|_{2,h}$ if and only if $|g(\xi)| \le 1 + ck$ for all ξ and some c. If $\theta \ge \frac12$, $|g(\xi)| \le 1$ for all ξ, k and h: implicit and Crank–Nicolson are unconditionally *stable in $\|\cdot\|_{2,h}$. If $\theta < \frac12$, along grids with λ fixed the scheme is stable if and only if $(1-2\theta)\lambda \le \frac12$; the explicit scheme if and only if $\sigma^2 k \le h^2$.*

(ii) Consistency. *If $\partial_\tau^i u$ ($i \le 3$) and $\partial_x^j u$ ($j \le 4$) are continuous and bounded by M on $[0,T]\times\mathbb{R}$, then $|T_j^n| \le M[(\frac12+\theta)k + (\frac{\sigma^2}{24} + \frac{|\nu|}{6})h^2]$, and for $\theta = \frac12$, $|T_j^n| \le M[\frac16 k^2 + (\frac{\sigma^2}{24} + \frac{|\nu|}{6})h^2]$: orders $O(k+h^2)$ and $O(k^2+h^2)$.*

(iii) Convergence. *If the scheme is stable in a norm, then $\|u^n - U^n\| \le C_s T \max_{m<n}\|T^m\|$ for $nk \le T$. Conversely, a consistent scheme for a well-posed linear problem that is not stable does not converge for every initial value (Lax and Richtmyer, 1956).*

(iv) Positivity. *If $h|\nu| \le \sigma^2$ and $(1-\theta)k(\sigma^2/h^2 + r) \le 1$, then B^{-1} and G map nonnegative sequences to nonnegative sequences, $\|B^{-1}\|_\infty \le 1$ and $\|G\|_\infty \le 1$. Without the second condition positivity can fail: for Crank–Nicolson applied to $\partial_\tau u = \frac12\sigma^2\partial_{xx}u$, G is positive if and only if $\lambda \le \frac32$.*

Proof. (i) $(L_h e^{\mathrm{i}\cdot\xi h})_j = (\frac{\sigma^2}{2h^2}(e^{\mathrm{i}\xi h} - 2 + e^{-\mathrm{i}\xi h}) + \frac{\nu}{2h}(e^{\mathrm{i}\xi h} - e^{-\mathrm{i}\xi h}) - r)e^{\mathrm{i}j\xi h}$, which is $(z/k)e^{\mathrm{i}j\xi h}$; solve (20.5) for U^{n+1}. For $V \in \ell^2(\mathbb{Z})$, the Fourier series $\hat V(\xi) = h\sum_j V_j e^{-\mathrm{i}j\xi h}$ satisfies Parseval's identity $\|V\|_{2,h}^2 = \frac{1}{2\pi}\int_{-\pi/h}^{\pi/h}|\hat V|^2$, and the scheme multiplies $\hat V$ by g, so $\|G^m\|_{2,h} = \max_\xi |g(\xi)|^m$. If $|g| \le 1 + ck$ this is at most e^{cT} for $mk \le T$; conversely, $\max|g|^{\lfloor T/k\rfloor} \le C_s$ forces $\max|g| \le C_s^{1/\lfloor T/k\rfloor} = 1 + O(k)$. Also B^{-1} multiplies by $1/(1-\theta z)$, of modulus at most 1 because $\operatorname{Re} z \le 0$. Write $z = x + \mathrm{i}w$. Then

$$|1+(1-\theta)z|^2 - |1-\theta z|^2 = 2x + (1-2\theta)|z|^2.$$

If $\theta \ge \frac12$, this is at most $2x \le 0$, so $|g| \le 1$. Let $\theta < \frac12$ and $s = \sin^2(\xi h/2)$, so that $x = -4\lambda s - rk$ and $w^2 = (\nu k/h)^2\, 4s(1-s) \le 2\nu^2\lambda k/\sigma^2$. If $(1-2\theta)\lambda \le \frac12$, then $(1-2\theta)(4\lambda s + rk) \le 2s + rk$ and

$$2x + (1-2\theta)|z|^2 \le (4\lambda s + rk)(2s + rk - 2) + w^2 \le (4\lambda + rk)rk + 2\nu^2\lambda k/\sigma^2,$$

while $|1-\theta z| \ge 1$; so $|g|^2 \le 1 + c'k$. If $(1-2\theta)\lambda > \frac12$, at $\xi h = \pi$ we have $w = 0$ and $g = \frac{1-(1-\theta)(4\lambda+rk)}{1+\theta(4\lambda+rk)} \to \frac{1-4(1-\theta)\lambda}{1+4\theta\lambda} < -1$ as $k \to 0$, so $|g| \le 1 + ck$ fails.

(ii) By Taylor's formula, $\frac{1}{h^2}(f(x+h) - 2f(x) + f(x-h)) = f''(x) + \frac{h^2}{12}f''''(\eta)$ and $\frac{1}{2h}(f(x+h) - f(x-h)) = f'(x) + \frac{h^2}{6}f'''(\eta')$ for some η, η', so $L_h u$ differs from $\frac12\sigma^2\partial_{xx}u + \nu\partial_x u - ru = \partial_\tau u$ by

at most $Mh^2(\frac{\sigma^2}{24}+\frac{|\nu|}{6})$. In time, $\frac{1}{k}(u^{n+1}-u^n)=\partial_\tau u^n+\frac{k}{2}\partial_{\tau\tau}u(\zeta)$ and $\theta\partial_\tau u^{n+1}+(1-\theta)\partial_\tau u^n=\partial_\tau u^n+\theta k\partial_{\tau\tau}u(\zeta')$, which differ by at most $(\frac{1}{2}+\theta)kM$. For $\theta=\frac{1}{2}$, expanding about $\tau_n+\frac{k}{2}$ instead, the two quantities differ from $\partial_\tau u(\tau_n+\frac{k}{2})$ by $\frac{k^2}{24}\partial_\tau^3 u$ and $\frac{k^2}{8}\partial_\tau^3 u$ at intermediate points.

(iii) The exact values satisfy $Bu^{n+1}=Cu^n+kT^n$, the scheme $BU^{n+1}=CU^n$, so $e^n=u^n-U^n$ satisfies $e^{n+1}=Ge^n+kB^{-1}T^n$ and $e^0=0$. Hence $e^n=k\sum_{m<n}G^{n-1-m}B^{-1}T^m$ and $\|e^n\|\le nk\,C_s\max_m\|T^m\|$. For the converse see Strikwerda (2004), Theorem 1.5.1, proved in Chapter 10.

(iv) B has diagonal $b_0=1+\theta k(\sigma^2/h^2+r)$ and off-diagonal entries $-\theta kp_\pm\le 0$ (as $h|\nu|\le\sigma^2$ means $p_\pm\ge 0$). Write $B=b_0(I-N)$, where $N\ge 0$ entrywise and $\|N\|_\infty=\theta k\sigma^2h^{-2}/b_0<1$; the Neumann series $B^{-1}=b_0^{-1}\sum_m N^m$ is positive, with $\|B^{-1}\|_\infty\le 1/(b_0(1-\|N\|_\infty))=1/(1+\theta kr)\le 1$. C has entries $(1-\theta)kp_\pm\ge 0$ and $1-(1-\theta)k(\sigma^2/h^2+r)\ge 0$, with row sums $1-(1-\theta)kr\le 1$. So $G=B^{-1}C$ is positive with $\|G\|_\infty\le 1$. For Crank–Nicolson and $\nu=r=0$, $B+C=2I$, so $G=2B^{-1}-I$. The entries of B^{-1} off the diagonal are positive (Neumann series) and its diagonal entry is the mean of $1/(1-\frac{1}{2}z)=1/(1+2\lambda\sin^2\frac{\varphi}{2})$ over $\varphi=\xi h\in[-\pi,\pi]$:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\mathrm{d}\varphi}{1+\lambda-\lambda\cos\varphi}=\frac{1}{\sqrt{(1+\lambda)^2-\lambda^2}}=\frac{1}{\sqrt{1+2\lambda}},$$

by the standard integral $\int_{-\pi}^{\pi}(\alpha-\beta\cos\varphi)^{-1}\mathrm{d}\varphi=2\pi(\alpha^2-\beta^2)^{-1/2}$, $\alpha>|\beta|$. So the diagonal entry of G is $2/\sqrt{1+2\lambda}-1$, nonnegative if and only if $\lambda\le\frac{3}{2}$. □

With (ii), part (iii) gives the rates in the maximum norm whenever the scheme is positive: $|u_j^n-U_j^n|\le TM[(\frac{1}{2}+\theta)k+(\frac{\sigma^2}{24}+\frac{|\nu|}{6})h^2]$. In $\|\cdot\|_{2,h}$, $\|T^m\|_{2,h}$ is of the same order when the derivatives of (ii) are also square-integrable in x (write the Taylor remainders in integral form and use the Cauchy–Schwarz inequality). Von Neumann analysis is in Section 2.2 of Strikwerda (2004).

Remark 20.7 (Stable is not positive). The implicit scheme is positive for every k; Crank–Nicolson only for λ of order one, $k\lesssim h^2/\sigma^2$, the very restriction it was chosen to escape. No scheme does better: Bolley and Crouzeix (1978), Theorem 1, show that no linear one-step time discretization of order two or more preserves positivity for all step sizes (their sufficient condition for Crank–Nicolson is $\lambda\le 1$). On a finite grid with zero boundary values the row next to the boundary binds: there one needs $\lambda\le 4-2\sqrt{2}\approx 1.17$ (Exercise 20.5; for $J\ge 11$ our grids agree to ten digits). A negative weight is a negative state price. The discrete gamma of a call at maturity is a spike at the strike, and one Crank–Nicolson step with $\lambda>\frac{3}{2}$ makes its centre negative (Figure 20.2(b)). The other condition, $h|\nu|\le\sigma^2$, fails as $\sigma\to 0$, when the equation becomes a transport equation: the positive alternative to central differences, upwinding, is only first order, and by Godunov's theorem no linear positive scheme for transport is better (Godunov, 1959). CRR trees meet the same wall at $n\le(r-y)^2T/\sigma^2$ (Exercise 20.4).

Remark 20.8 (Kinks and Rannacher's start). The theorem assumes a smooth solution; a call is not smooth at $(0,\log K)$. Its kink feeds the highest frequencies, which the implicit scheme damps at each step, but Crank–Nicolson multiplies the mode $\xi h=\pi$ by $(1-2\lambda)/(1+2\lambda)\to -1$ (with $r=\nu=0$): with k proportional to h, the natural choice for a second-order scheme, λ grows like $1/h$ and the sawtooth survives. On the test case with $k=h/\sigma$, the Crank–Nicolson price error is of first order (3.8×10^{-3}, 2.0×10^{-3}, 1.0×10^{-3} for $J=800,1600,3200$ intervals over $\pm6\sigma\sqrt{T}$) and the error of its gamma at S_0 grows (-0.15, -0.31, -0.60, while the gamma is 0.0195). Rannacher (1984)'s remedy replaces the first Crank–Nicolson steps by implicit ones.

Giles and Carter (2006) prove that four implicit half-steps in place of the first two steps restore second order in the maximum norm for the value, delta and gamma of calls and puts; on the grids above the price errors become -4.0×10^{-4}, -1.0×10^{-4}, -2.5×10^{-5} and the gamma errors 1.5×10^{-6}, 3.9×10^{-7}, 9.6×10^{-8}. Short maturities are the worst case: for $T=0.05$ ($J=800$ over $\pm10\sigma\sqrt{T}$, $N=20$) the plain gamma at S_0 is 2.93, the Rannacher gamma 0.0892, the true one 0.0891. Placing the strike on a node or midway between two, and averaging discontinuous payoffs over a cell, complete the treatment (Pooley et al., 2003); the solvers of Chapter 4 use a variant (two implicit steps at each end), and the BBS tree is its analogue.

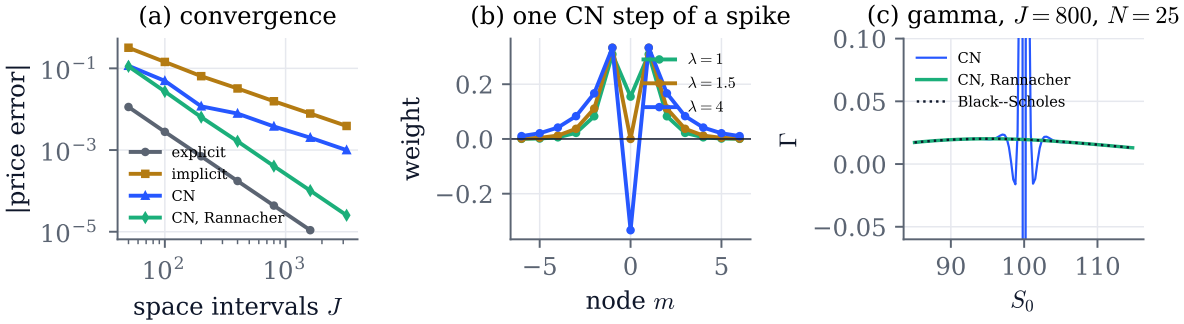

Figure 20.2: (a) Absolute error of the test-case call at S_0 on grids of J intervals over $\log S_0\pm6\sigma\sqrt{T}$: explicit with $\lambda=0.4$, implicit, Crank–Nicolson and Crank–Nicolson with Rannacher's start, the last three with $k\approx h/\sigma$. (b) One Crank–Nicolson step of $\partial_\tau u=\frac12\sigma^2\partial_{xx}u$ applied to a unit spike at node 0, for $\lambda=1$, $\frac32$, 4: the centre weight is $2/\sqrt{1+2\lambda}-1$. (c) The gamma at $t=0$ of the test-case call against the spot, on the grid $J=800$, $N=25$ ($\lambda=89$): plain Crank–Nicolson (between -1.49 and 0.68 near the strike, clipped), with Rannacher's start (on top of the Black–Scholes gamma, within 8×10^{-6}).

20.4 Boundaries and the choice of grid

A grid must stop somewhere. On $D=(x_\ell,x_u)$ the scheme needs values at the two ends, and the question is how much a wrong value there costs at the spot. Let $X_s=x+\nu s+\sigma W_s$, $\vartheta=\inf\{s\ge0:X_s\notin D\}$, and, for bounded measurable boundary data $\beta(\tau,x_\ell)$, $\beta(\tau,x_u)$,

$$u_D(\tau,x)=\mathbb{E}^{\mathbb{Q}}\Big[e^{-r\tau}g(e^{X_\tau})\mathbf{1}_{\{\vartheta>\tau\}}+e^{-r\vartheta}\beta(\tau-\vartheta,X_\vartheta)\mathbf{1}_{\{\vartheta\le\tau\}}\Big], \tag{20.7}$$

the price of the option that pays β when the log price leaves D first. When the Dirichlet problem for (20.1) on D with data g and β has a bounded classical solution, Itô's formula and optional stopping at $\vartheta\wedge\tau$, as in the proof of Theorem 1.6, show that it is u_D; this is the function that the grids of Section 20.3 approximate.

Proposition 20.9 (Far-field boundary conditions)**.** *Let g be measurable with $|g(s)|\le a(1+s)$, $r\ge0$, $x\in D$, and suppose $|\beta(t,x_b)-u(t,x_b)|\le\varepsilon$ for $t\in[0,\tau]$ and $x_b\in\{x_\ell,x_u\}$. Then*

$$|u_D(\tau,x)-u(\tau,x)|\le\varepsilon\,\mathbb{Q}(\vartheta\le\tau)\le\varepsilon\Big[e^{-(x_u-x-\nu^+\tau)^2/(2\sigma^2\tau)}+e^{-(x-x_\ell-\nu^-\tau)^2/(2\sigma^2\tau)}\Big],$$

where the bound holds when $x_u-x\ge\nu^+\tau$ and $x-x_\ell\ge\nu^-\tau$ (a term with $x_\ell=-\infty$ or $x_u=\infty$

is absent). For the call with the asymptotic values $\beta(t,x_\ell)=0$ and $\beta(t,x_u)=e^{x_u-yt}-Ke^{-rt}$, ε is the larger of the call price at e^{x_ℓ} and the put price at e^{x_u} over $t\le\tau$.

Proof. By the Markov property, $M_s=e^{-rs}u(\tau-s,X_s)=\mathbb{E}^{\mathbb{Q}}[e^{-r\tau}g(e^{X_\tau})\mid\mathcal{F}_s]$, $s\le\tau$, is a martingale, closed by an integrable variable (the growth bound and $\mathbb{E}^{\mathbb{Q}}[e^{X_\tau}]<\infty$); optional stopping at the bounded time $\vartheta\wedge\tau$ gives $u(\tau,x)=\mathbb{E}^{\mathbb{Q}}\big[e^{-r\tau}g(e^{X_\tau})\mathbf{1}_{\{\vartheta>\tau\}}+e^{-r\vartheta}u(\tau-\vartheta,X_\vartheta)\mathbf{1}_{\{\vartheta\le\tau\}}\big]$. Subtracting from (20.7), $u_D-u=\mathbb{E}^{\mathbb{Q}}\big[e^{-r\vartheta}(\beta-u)(\tau-\vartheta,X_\vartheta)\mathbf{1}_{\{\vartheta\le\tau\}}\big]$, which is at most $\varepsilon\,\mathbb{Q}(\vartheta\le\tau)$ in absolute value since $r\ge0$. The path leaves through x_u before τ only if $\max_{s\le\tau}\sigma W_s\ge x_u-x-\nu^+\tau=:m$, which by the reflection principle (Theorem 16.1) has probability $2N(-m/(\sigma\sqrt{\tau}))\le e^{-m^2/(2\sigma^2\tau)}$; the last inequality holds for $m\ge0$ because $e^{-z^2/2}-2N(-z)$ vanishes at 0 and at ∞ and increases and then decreases on $[0,\infty)$. The lower end is symmetric. For the call, $u-\beta$ at x_u is the put price, by put–call parity. □

The bound is Gaussian in the distance to the boundary in units of $\sigma\sqrt{\tau}$, and it multiplies the error of the boundary data: on $\pm m\sigma\sqrt{T}$ around the spot even the crude choice $\beta=g$ costs at most about $\max|g-u|\cdot 2e^{-m^2/2}$, $3\times10^{-8}\max|g-u|$ for $m=6$, the default of our code. Figure 20.3(b) measures the truncation error of the Crank–Nicolson grid. For positive schemes the same argument runs on the Markov chain of the scheme, whose steps are kernels with nonnegative weights summing to at most one; Kangro and Nicolaides (2000) prove the exponential decay for the Black–Scholes equation itself.

The course's example. The PINN of the course's discussion notebook prices the call with $K=40$, $r=7\%$, $\sigma=60\%$, $T=1$ on $S\in[0,120]$ with the value $S_{\max}-K=80$ at the top, whereas the asymptotic value is 82.70 and the price 83.08. On a Crank–Nicolson grid in S (1200 intervals, 600 steps) the value 80 makes the price too low by 0.036 at $S_0=40$ and by 0.75 at $S_0=80$; the asymptotic value reduces these to 0.0007 and 0.050. The proposition, with $\varepsilon=3.08$ and 0.378 and the exact exit probabilities 0.047 and 0.44, bounds them by 0.15, 1.35, 0.018 and 0.17. A grid in S needs no condition at $S=0$, where the equation reduces to $\partial_\tau V=-rV$, but its local λ grows like S^2, so its explicit version needs $k\lesssim h_S^2/(\sigma S_{\max})^2$ for a mesh h_S; in log price λ is the same everywhere.

20.5 American options on a grid

The American price is the value of an optimal stopping problem (Chapter 14), and it solves the linear complementarity problem $u\ge g$, $\partial_\tau u-\mathcal{L}u\ge0$, $(u-g)(\partial_\tau u-\mathcal{L}u)=0$, with $\mathcal{L}$ the right side of (20.1) (Jaillet et al., 1990). Each step of the θ-scheme becomes a finite-dimensional complementarity problem: given $q=CV^n+kF^n$ and the payoff vector g, find $z=V^{n+1}$ with

$$z\ge g,\qquad Bz-q\ge0,\qquad (z-g)\cdot(Bz-q)=0, \tag{20.8}$$

inequalities entry by entry. With $\Phi_j(z)=\max\big(g_j,(q_j-\sum_{l\ne j}B_{jl}z_l)/B_{jj}\big)$, *projected SOR* (PSOR, Cryer (1971)) sweeps through the nodes with a relaxation $\omega\in(0,2)$, $z_j\leftarrow\max\big(g_j,z_j+\omega(\Phi_j^0(z)-z_j)\big)$, where Φ_j^0 is Φ_j without the maximum and uses the entries already updated; $\omega=1$ is *projected Gauss–Seidel*, $z_j\leftarrow\Phi_j(z)$. The algorithm of Brennan and Schwartz (1977) eliminates the super-diagonal of B from the top, $d'_{J-1}=B_{J-1,J-1}$, $q'_{J-1}=q_{J-1}$ and, for $j=J-2,\dots,1$,

$$m_j=\frac{B_{j,j+1}}{d'_{j+1}},\quad d'_j=B_{jj}-m_jB_{j+1,j},\quad q'_j=q_j-m_jq'_{j+1}, \tag{20.9}$$

and then, upwards for $j = 1, \dots, J-1$ with $B_{1,0}z_0 := 0$, $z_j = \max\left(g_j, (q'_j - B_{j,j-1}z_{j-1})/d'_j\right)$: one elimination and one projected substitution, $O(J)$ operations.

Proposition 20.10 (American options on a grid)**.** *Let $\theta \in (0,1]$, $r \geq 0$, $h|\nu| \leq \sigma^2$, and let $B = I - \theta k A$ be the matrix of one step on the interior nodes $1, \dots, J-1$. Then $B_{jl} \leq 0$ for $l \neq j$ and every row sum of B is at least $1 + \theta k r$.*

(i) For all q, g, (20.8) has exactly one solution z^*. It is the smallest supersolution*: $z^* \leq s$ for every s with $s \geq g$ and $Bs \geq q$; and $z^* = \Phi(z^*)$.*

(ii) Projected Gauss–Seidel started at $z^{(0)} = g$ produces a nondecreasing sequence, bounded by every supersolution, that converges to z^*, for every order of the sweep. If $h|\nu| < \sigma^2$, PSOR converges to z^* for every $\omega \in (0,2)$ and every starting point.

(iii) If z^* has its exercise region at the bottom, that is, for some j^*, $z^*_j = g_j$ for $j \leq j^*$ and $(Bz^* - q)_j = 0$ for $j > j^*$, then (20.9) returns z^*. The mirror image, eliminating from the bottom, holds for an exercise region at the top.

(iv) For $\rho > 0$, the penalized equation $Bz = q + \rho(g - z)^+$ has a unique solution z_ρ, the policy iteration of Forsyth and Vetzal (2002) finds it in finitely many steps (their Theorem 6.1), and $z_\rho \leq z^* \leq z_\rho + \|(Bg - q)^+\|_\infty/\rho$.

Proof. The off-diagonal entries are $-\theta k p_\pm \leq 0$, and the row sums are $1 + \theta k r$ for interior rows and larger for the first and the last, which miss a neighbour.

(i) A solution z satisfies $z_j > g_j \Rightarrow (Bz)_j = q_j$ and $z_j = g_j \Rightarrow (Bz)_j \geq q_j$; dividing by $B_{jj} > 0$, this is $z_j = \Phi_j(z)$, and conversely. Let s be a supersolution and suppose $w = z - s$ has a positive maximum at j_0. Then $z_{j_0} > s_{j_0} \geq g_{j_0}$, so $(Bz)_{j_0} = q_{j_0} \leq (Bs)_{j_0}$ and $(Bw)_{j_0} \leq 0$; but $(Bw)_{j_0} \geq (\sum_l B_{j_0 l})w_{j_0} > 0$ because $B_{j_0 l} \leq 0$ and $w_l \leq w_{j_0}$. So $z \leq s$. A solution is a supersolution, so two solutions are each below the other: there is at most one, and it is the smallest supersolution. Existence follows from (ii).

(ii) Φ is continuous and nondecreasing in each entry, because $B_{jl} \leq 0$. A sweep replaces the entries one by one by Φ_j of the current vector, so the sweep map is nondecreasing too. Since $\Phi(g) \geq g$, the first sweep does not decrease any entry, and by induction the sequence is nondecreasing. If s is a supersolution, $\Phi(s) \leq s$ entrywise, so $z^{(i)} \leq s$ implies $z^{(i+1)} \leq s$; and $s = c\mathbf{1}$ is one as soon as $c \geq \max(\max_j g_j, \max_j q_j, 0)$, since $B\mathbf{1} \geq \mathbf{1}$. The bounded nondecreasing sequence converges to a fixed point of Φ, which solves (20.8). For PSOR: if $p_\pm > 0$, $D = \mathrm{diag}((p_-/p_+)^{j/2})$ makes $P = D^{-1}BD$ symmetric, with off-diagonal entries $-\theta k\sqrt{p_- p_+}$, and positive definite by Gershgorin, since $2\sqrt{p_- p_+} \leq p_- + p_+$. PSOR for B, q, g is PSOR for P, $D^{-1}q$, $D^{-1}g$, which converges for $\omega \in (0,2)$ by Cryer (1971).

(iii) By induction from the top, $d'_j \geq 1 + |B_{j,j-1}|$ and $m_j \leq 0$: $d'_{J-1} = B_{J-1,J-1}$, and if $d'_{j+1} \geq 1 + |B_{j+1,j}|$, then $0 \leq m_j B_{j+1,j} \leq |B_{j,j+1}|$, so $d'_j \geq B_{jj} - |B_{j,j+1}| \geq 1 + |B_{j,j-1}|$. Row j of the eliminated system, $B_{j,j-1}z_{j-1} + d'_j z_j = q'_j$, is row j of $Bz = q$ plus $|m_j|$ times row $j+1$ of the eliminated system. So the residuals $r'_j = B_{j,j-1}z^*_{j-1} + d'_j z^*_j - q'_j$ satisfy $r'_{J-1} = r_{J-1}$ and $r'_j = r_j + |m_j| r'_{j+1}$, where $r = Bz^* - q \geq 0$ vanishes above j^*. Hence $r'_j = 0$ for $j > j^*$ and $r'_j \geq 0$ for $j \leq j^*$. For $j > j^*$ this says $z^*_j = (q'_j - B_{j,j-1}z^*_{j-1})/d'_j$, which is at least g_j; for $j \leq j^*$, $z^*_j = g_j \geq (q'_j - B_{j,j-1}z^*_{j-1})/d'_j$. In both cases $z^*_j = \max\left(g_j, (q'_j - B_{j,j-1}z^*_{j-1})/d'_j\right)$, so the upward recursion, started at $j = 1$, reproduces z^*.

(iv) Existence, uniqueness and termination are Theorem 6.1 of Forsyth and Vetzal (2002); the bracket is Exercise 20.7. □

The hypothesis of (iii) is not automatic. For the put in the implicit scheme it is justified by Jaillet et al. (1990); our code checks the residual of (20.8): below 10^{-12} for the options of Table 20.1, but 5×10^{-3} and 7×10^{-3}, with the exercise region assumed at the top and at the bottom, for an American straddle with $r = y = 5\%$, which has an exercise region at each end (Exercise 20.8). Bermudan exercise replaces the complementarity problem by a maximum after the linear solve at the dates. Convergence of the discrete American prices to the American price holds for positive schemes by the theory of monotone schemes of Barles and Souganidis (1991), and for the implicit scheme by Jaillet et al. (1990); Crank–Nicolson on our grids, with λ from 3 to 23, is outside both, and its convergence is observed, not proved here.

	European	grid	corrected	BBSR	LR	trinomial	Ch. 14
test case, put	8.349406	8.349393	8.349408	8.349408	8.349408	8.349154	8.349408
test case, call	7.364290	7.459694	7.459709	7.459739	7.459766	7.459608	7.4597
LS 36, 0.2, 1	3.844308	4.486662	4.486666	4.486679	4.486560	4.486642	4.487
LS 40, 0.2, 1	2.066401	2.319556	2.319562	2.319579	2.319535	2.319522	2.320
LS 44, 0.4, 2	5.201995	5.646682	5.646696	5.646747	5.646698	5.646838	5.647

Table 20.1: American puts and calls. Grid: Crank–Nicolson with Rannacher's start, $J = 3200$ intervals over $\pm 6\sigma\sqrt{T}$, $N = 1600$ steps, Brennan–Schwartz (PSOR and the penalty method agree to 2×10^{-6}); corrected: plus the Black–Scholes price minus the European price on the same grid. BBSR: $n = 4000$ (the tree of Chapter 14, which used 2000 steps for Table 14.1); LR: Leisen–Reimer, $n = 4001$; trinomial: $\kappa = \sqrt{3/2}$, $n = 2000$. LS S_0, σ, T: the puts of Longstaff and Schwartz, $K = 40$, $r = 6\%$, $y = 0$. Chapter 14 quotes the test-case values and Table 14.1 to the digits shown.

The numbers. Table 20.1 reproduces the converged American values of Chapter 14: the test-case put 8.349408 (early-exercise premium 2.0×10^{-6} on the grid, as in Chapter 14), the test-case call 7.4597 ($y > r$, so early exercise of the call pays), and the Longstaff–Schwartz rows. With exercise at 50 dates per year the grid gives the Bermudan values of Chapter 14 within 10^{-4} (4.477807 against 4.477810 for the first LS row). Brennan–Schwartz costs one pass, the penalty method about 1.1 tridiagonal solves per step on average (at most 5), and PSOR with $\omega = 1.2$ up to 16, 36 and 70 sweeps per step (15, 34, 65 on average) for $J = 400$, 800, 1600 with $N = J/2$: an iterative solver slows down as λ grows. The American errors fall by about 3 when J and N double (order about 1.6), against 4 for the European ones: the price is not twice differentiable across the exercise boundary of Figure 20.3(a).

20.6 Path dependence and several factors

Barriers. A continuously monitored knock-out is a Dirichlet condition: the grid ends at the barrier, a node, with value zero. A discretely monitored one sets the values beyond the barrier to zero at the monitoring dates; we place the barrier midway between two nodes and repeat Rannacher's start after each date. On the course's monthly down-and-out call ($B = 95$)

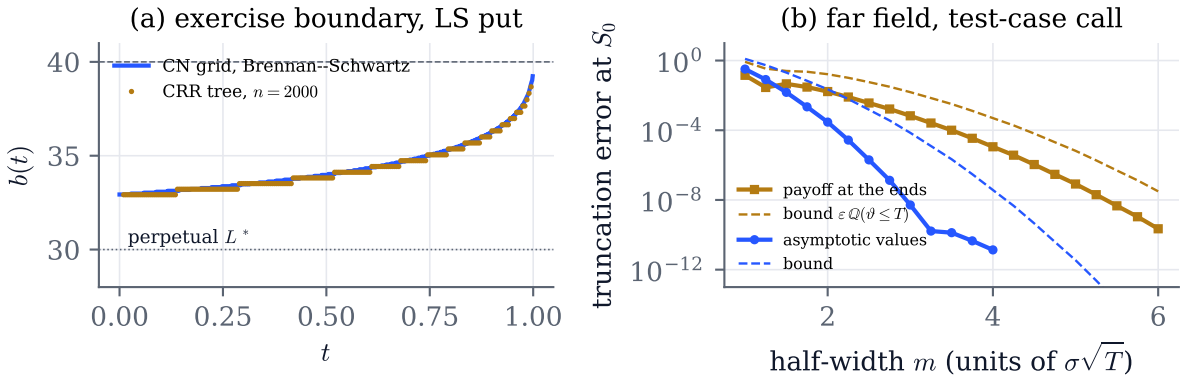

Figure 20.3: (a) Exercise boundary of the Longstaff–Schwartz put ($S_0 = 36$, $K = 40$, $r = 6\%$, $\sigma = 20\%$, $T = 1$): grid ($J = 1600$, $N = 800$, Brennan–Schwartz; largest exercised node) and CRR tree (2000 steps, every tenth shown); K dashed, $L^* = 2rK/(2r + \sigma^2) = 30$ dotted. (b) Test-case call: the Crank–Nicolson price ($k = 1/400$, $h = 0.003$) on $\log S_0 \pm m\sigma\sqrt{T}$ minus that on $\log S_0 \pm 10\sigma\sqrt{T}$, for asymptotic and payoff boundary values, and the bounds of Theorem 20.9 (dashed); differences below 10^{-11} (rounding) are not shown.

the Crank–Nicolson grid with $J = 1600$, $N = 960$ gives 5.32995 against the exact 5.32979 of Chapter 16, and 3.75411 against 3.75412 for continuous monitoring (compare Theorem 16.7). In a tree the error oscillates in n unless a layer of nodes lies on the barrier (Boyle and Lau, 1994).

Other path dependence and several factors. A running maximum or average becomes an extra state variable. In two dimensions (Heston, Theorem 19.4; two assets) the implicit step is a sparse system rather than a tridiagonal one; alternating-direction implicit methods split it into tridiagonal solves along each axis (Peaceman and Rachford, 1955), with the mixed derivative of correlated factors treated explicitly (Craig and Sneyd, 1988; in 't Hout and Foulon, 2010). For a target error ε and a smooth solution, the explicit scheme needs $h \approx \varepsilon^{1/2}$ and $k \approx h^2$, Crank–Nicolson with Rannacher's start $h \approx k \approx \varepsilon^{1/2}$; in d space dimensions they cost $\varepsilon^{-(d+2)/2}$ and $\varepsilon^{-(d+1)/2}$ operations, the course's rule $\varepsilon^{-(d+1)/\alpha}$ for lattices with $\alpha = 2$, against ε^{-2} for Monte Carlo (Theorem 6.14).

In code

`code/src/mfe230d/grids.py` implements the trees, `fd_solve` (the θ-scheme with barriers and American exercise by `psor`, `brennan_schwartz` or `penalty`) and the stability tools; `code/tests/test_grids.py` checks every number quoted here.

▶ Interactive demo: trees, stability, the far field and the American put.

Further reading

Finance: Wilmott et al. (1995); stability: Strikwerda (2004).

Bibliography

Joseph Abate and Peter P. Valkó. Multi-precision Laplace transform inversion. *International Journal for Numerical Methods in Engineering*, 60(5):979–993, 2004.

Peter J. Acklam. An algorithm for computing the inverse normal cumulative distribution function. Web note, archived at https://web.archive.org/web/20070505093933/http://home.online.no/~pjacklam/notes/invnorm/, 2000.

Peter A. Acworth, Mark Broadie, and Paul Glasserman. A comparison of some Monte Carlo and quasi Monte Carlo techniques for option pricing. In Harald Niederreiter, Peter Hellekalek, Gerhard Larcher, and Peter Zinterhof, editors, *Monte Carlo and Quasi-Monte Carlo Methods 1996*, volume 127 of *Lecture Notes in Statistics*, pages 1–18. Springer, New York, 1998. doi: 10.1007/978-1-4612-1690-2_1.

Riaz Ahmad and Paul Wilmott. Which free lunch would you like today, sir?: Delta hedging, volatility arbitrage and optimal portfolios. *Wilmott Magazine*, pages 64–79, November 2005.

Hansjörg Albrecher, Philipp Mayer, Wim Schoutens, and Jurgen Tistaert. The little Heston trap. *Wilmott Magazine*, (1):83–92, 2007.

Leif Andersen. Simple and efficient simulation of the Heston stochastic volatility model. *Journal of Computational Finance*, 11(3):1–42, 2008.

Leif Andersen and Mark Broadie. Primal-dual simulation algorithm for pricing multidimensional American options. *Management Science*, 50(9):1222–1234, 2004. doi: 10.1287/mnsc.1040.0258.

Leif B. G. Andersen and Vladimir V. Piterbarg. Moment explosions in stochastic volatility models. *Finance and Stochastics*, 11(1):29–50, 2007. doi: 10.1007/s00780-006-0011-7.

Søren Asmussen and Peter W. Glynn. *Stochastic Simulation: Algorithms and Analysis*, volume 57 of *Stochastic Modelling and Applied Probability*. Springer, New York, 2007.

Rainer Avikainen. On irregular functionals of SDEs and the Euler scheme. *Finance and Stochastics*, 13(3):381–401, 2009.

Mark Bagnoli and Ted Bergstrom. Log-concave probability and its applications. *Economic Theory*, 26(2):445–469, 2005.

Vlad Bally and Denis Talay. The law of the Euler scheme for stochastic differential equations I: Convergence rate of the distribution function. *Probability Theory and Related Fields*, 104(1): 43–60, 1996.

Guy Barles and Panagiotis E. Souganidis. Convergence of approximation schemes for fully nonlinear second order equations. *Asymptotic Analysis*, 4(3):271–283, 1991.

J. D. Beasley and S. G. Springer. Algorithm AS 111: The percentage points of the normal distribution. *Applied Statistics*, 26(1):118–121, 1977.

Mohamed Ben Alaya and Ahmed Kebaier. Central limit theorem for the multilevel Monte Carlo Euler method. *The Annals of Applied Probability*, 25(1):211–234, 2015.

Vidmantas Bentkus and Friedrich Götze. The Berry–Esseen bound for Student's statistic. *Annals of Probability*, 24(1):491–503, 1996. doi: 10.1214/aop/1042644728.

Dimitris Bertsimas, Leonid Kogan, and Andrew W. Lo. When is time continuous? *Journal of Financial Economics*, 55(2):173–204, 2000.

Patrick Billingsley. *Convergence of Probability Measures*. Wiley, New York, 2nd edition, 1999.

Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

David Blackman and Sebastiano Vigna. Scrambled linear pseudorandom number generators. *ACM Transactions on Mathematical Software*, 47(4):Article 36, 2021.

Catherine Bolley and Michel Crouzeix. Conservation de la positivité lors de la discrétisation des problèmes d'évolution paraboliques. *RAIRO Analyse numérique*, 12(3):237–245, 1978.

Jonathan Bowie and Peter Carr. Static simplicity. *Risk*, 7(8):45–49, 1994.

G. E. P. Box and Mervin E. Muller. A note on the generation of random normal deviates. *Annals of Mathematical Statistics*, 29(2):610–611, 1958.

Phelim Boyle, Mark Broadie, and Paul Glasserman. Monte Carlo methods for security pricing. *Journal of Economic Dynamics and Control*, 21(8–9):1267–1321, 1997.

Phelim P. Boyle. Options: A Monte Carlo approach. *Journal of Financial Economics*, 4(3): 323–338, 1977.

Phelim P. Boyle. Option valuation using a three-jump process. *International Options Journal*, 3: 7–12, 1986.

Phelim P. Boyle and Sok Hoon Lau. Bumping up against the barrier with the binomial method. *Journal of Derivatives*, 1(4):6–14, 1994.

Douglas T. Breeden and Robert H. Litzenberger. Prices of state-contingent claims implicit in option prices. *Journal of Business*, 51(4):621–651, 1978.

Michael J. Brennan and Eduardo S. Schwartz. The valuation of American put options. *Journal of Finance*, 32(2):449–462, 1977.

Michael J. Brennan and Eduardo S. Schwartz. Finite difference methods and jump processes arising in the pricing of contingent claims: a synthesis. *Journal of Financial and Quantitative Analysis*, 13(3):461–474, 1978.

Menachem Brenner and Marti G. Subrahmanyam. A simple formula to compute the implied standard deviation. *Financial Analysts Journal*, 44(5):80–83, 1988.

Mark Broadie and Jérôme Detemple. American option valuation: New bounds, approximations, and a comparison of existing methods. *The Review of Financial Studies*, 9(4):1211–1250, 1996.

Mark Broadie and Paul Glasserman. Estimating security price derivatives using simulation. *Management Science*, 42(2):269–285, 1996.

Mark Broadie and Paul Glasserman. Pricing American-style securities using simulation. *Journal of Economic Dynamics and Control*, 21(8–9):1323–1352, 1997. doi: 10.1016/S0165-1889(97)00029-8.

Mark Broadie and Özgür Kaya. Exact simulation of stochastic volatility and other affine jump diffusion processes. *Operations Research*, 54(2):217–231, 2006. doi: 10.1287/opre.1050.0247.

Mark Broadie, Paul Glasserman, and Steven Kou. A continuity correction for discrete barrier options. *Mathematical Finance*, 7(4):325–349, 1997.

Mark Broadie, Paul Glasserman, and Steven Kou. Connecting discrete and continuous path-dependent options. *Finance and Stochastics*, 3(1):55–82, 1999. doi: 10.1007/s007800050052.

Peter J. Brockwell and Richard A. Davis. *Time Series: Theory and Methods*. Springer, New York, 2nd edition, 1991.

Hans-Joachim Bungartz and Michael Griebel. Sparse grids. *Acta Numerica*, 13:147–269, 2004. doi: 10.1017/S0962492904000182.

Russel E. Caflisch, William Morokoff, and Art B. Owen. Valuation of mortgage-backed securities using Brownian bridges to reduce effective dimension. *Journal of Computational Finance*, 1 (1):27–46, 1997.

Luca Capriotti. Fast Greeks by algorithmic differentiation. *Journal of Computational Finance*, 14(3):3–35, 2011.

Peter Carr. FAQ's in option pricing theory. Working paper, 1997. Dated 27 March 1997.

Peter Carr and Andrew Chou. Breaking barriers. *Risk*, 10(9):139–145, 1997.

Peter Carr and Robert A. Jarrow. The stop-loss start-gain paradox and option valuation: A new decomposition into intrinsic and time value. *Review of Financial Studies*, 3(3):469–492, 1990.

Peter Carr and Roger Lee. Put–call symmetry: Extensions and applications. *Mathematical Finance*, 19(4):523–560, 2009.

Peter Carr and Dilip Madan. Towards a theory of volatility trading. In Robert Jarrow, editor, *Volatility: New Estimation Techniques for Pricing Derivatives*, pages 417–427. Risk Books, 1998.

Peter Carr and Dilip Madan. Option valuation using the fast Fourier transform. *Journal of Computational Finance*, 2(4):61–73, 1999.

Peter Carr and Dilip B. Madan. A note on sufficient conditions for no arbitrage. *Finance Research Letters*, 2(3):125–130, 2005.

Peter Carr, Katrina Ellis, and Vishal Gupta. Static hedging of exotic options. *Journal of Finance*, 53(3):1165–1190, 1998.

Jacques F. Carrière. Valuation of the early-exercise price for options using simulations and nonparametric regression. *Insurance: Mathematics and Economics*, 19(1):19–30, 1996.

Andrew Carverhill and Les Clewlow. Flexible convolution. *Risk*, 3(4):25–29, 1990.

George Casella and Roger L. Berger. *Statistical Inference*. Duxbury, Pacific Grove, CA, 2nd edition, 2002.

Aleš Černý and Ioannis Kyriakou. An improved convolution algorithm for discretely sampled Asian options. *Quantitative Finance*, 11(3):381–389, 2011. doi: 10.1080/14697680903397667.

Jaehyuk Choi, Jeonggyu Huh, and Nan Su. Tighter 'uniform bounds for Black–Scholes implied volatility' and the applications to root-finding. *Operations Research Letters*, 57:107189, 2024.

Emmanuelle Clément, Damien Lamberton, and Philip Protter. An analysis of a least squares regression method for American option pricing. *Finance and Stochastics*, 6(4):449–471, 2002. doi: 10.1007/s007800200071.

K. Andrew Cliffe, Michael B. Giles, Robert Scheichl, and Aretha L. Teckentrup. Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients. *Computing and Visualization in Science*, 14(1):3–15, 2011.

William G. Cochran. *Sampling Techniques*. Wiley, New York, 3rd edition, 1977.

Rama Cont. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001.

Antoine Conze and Viswanathan. Path dependent options: The case of lookback options. *Journal of Finance*, 46(5):1893–1907, 1991.

John C. Cox, Stephen A. Ross, and Mark Rubinstein. Option pricing: A simplified approach. *Journal of Financial Economics*, 7(3):229–263, 1979.

John C. Cox, Jonathan E. Ingersoll, and Stephen A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407, 1985.

Andrei Cozma and Christoph Reisinger. Strong order 1/2 convergence of full truncation Euler approximations to the Cox–Ingersoll–Ross process. *IMA Journal of Numerical Analysis*, 40(1): 358–376, 2020. doi: 10.1093/imanum/dry067.

Ian J. D. Craig and Alfred D. Sneyd. An alternating-direction implicit scheme for parabolic equations with mixed derivatives. *Computers & Mathematics with Applications*, 16(4):341–350, 1988.

John Crank and Phyllis Nicolson. A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type. *Proceedings of the Cambridge Philosophical Society*, 43(1):50–67, 1947.

R. Cranley and T. N. L. Patterson. Randomization of number theoretic methods for multiple integration. *SIAM Journal on Numerical Analysis*, 13(6):904–914, 1976. doi: 10.1137/0713071.

Colin W. Cryer. The solution of a quadratic programming problem using systematic overrelaxation. *SIAM Journal on Control*, 9(3):385–392, 1971.

Michael Curran. Valuing Asian and portfolio options by conditioning on the geometric mean price. *Management Science*, 40(12):1705–1711, 1994.

Mark H. A. Davis and David G. Hobson. The range of traded option prices. *Mathematical Finance*, 17(1):1–14, 2007.

Pieter-Tjerk de Boer, Dirk P. Kroese, Shie Mannor, and Reuven Y. Rubinstein. A tutorial on the cross-entropy method. *Annals of Operations Research*, 134(1):19–67, 2005. doi: 10.1007/s10479-005-5724-z.

Amir Dembo and Ofer Zeitouni. *Large Deviations Techniques and Applications*, volume 38 of *Applications of Mathematics*. Springer, New York, 2nd edition, 1998.

Emmanuel Denis and Yuri Kabanov. Mean square error for the Leland–Lott hedging strategy: Convex pay-offs. *Finance and Stochastics*, 14(4):625–667, 2010.

Luc Devroye. *Non-Uniform Random Variate Generation*. Springer, New York, 1986.

Josef Dick and Friedrich Pillichshammer. *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*. Cambridge University Press, Cambridge, 2010. doi: 10.1017/CBO9780511761188.

Francine Diener and Marc Diener. Asymptotics of the price oscillations of a European call option in a tree model. *Mathematical Finance*, 14(2):271–293, 2004.

DLMF. NIST digital library of mathematical functions. https://dlmf.nist.gov/, Release 1.2.8 of 2026-09-15, 2026. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds.

Darrell Duffie and Peter W. Glynn. Efficient Monte Carlo simulation of security prices. *The Annals of Applied Probability*, 5(4):897–905, 1995.

Darrell Duffie, Jun Pan, and Kenneth Singleton. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6):1343–1376, 2000.

Darrell Duffie, Damir Filipović, and Walter Schachermayer. Affine processes and applications in finance. *The Annals of Applied Probability*, 13(3):984–1053, 2003.

Bruno Dupire. Pricing with a smile. *Risk*, 7(1):18–20, 1994.

Rick Durrett. *Probability: Theory and Examples*. Cambridge University Press, Cambridge, 5th edition, 2019.

Erik Ekström and Johan Tysk. The Black–Scholes equation in stochastic volatility models. *Journal of Mathematical Analysis and Applications*, 368(2):498–507, 2010. doi: 10.1016/j.jmaa.2010.04.014.

Nicole El Karoui, Monique Jeanblanc-Picqué, and Steven E. Shreve. Robustness of the Black and Scholes formula. *Mathematical Finance*, 8(2):93–126, 1998.

Simon Ellersgaard, Martin Jönsson, and Rolf Poulsen. The fundamental theorem of derivative trading: Exposition, extensions and experiments. *Quantitative Finance*, 17(4):515–529, 2017.

J. D. Esary, F. Proschan, and D. W. Walkup. Association of random variables, with applications. *The Annals of Mathematical Statistics*, 38(5):1466–1474, 1967.

Lawrence C. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2nd edition, 2010.

Fang Fang and Cornelis W. Oosterlee. A novel pricing method for European options based on Fourier-cosine series expansions. *SIAM Journal on Scientific Computing*, 31(2):826–848, 2008.

William Feller. Two singular diffusion problems. *Annals of Mathematics*, 54(1):173–182, 1951.

Richard P. Feynman. Space-time approach to non-relativistic quantum mechanics. *Reviews of Modern Physics*, 20(2):367–387, 1948.

Hans Föllmer and Alexander Schied. *Stochastic Finance: An Introduction in Discrete Time*. De Gruyter, Berlin, 4th edition, 2016.

Peter A. Forsyth and Kenneth R. Vetzal. Quadratic convergence for valuing American options using a penalty method. *SIAM Journal on Scientific Computing*, 23(6):2095–2122, 2002.

Eric Fournié, Jean-Michel Lasry, Jérôme Lebuchoux, Pierre-Louis Lions, and Nizar Touzi. Applications of Malliavin calculus to Monte Carlo methods in finance. *Finance and Stochastics*, 3(4):391–412, 1999.

Avner Friedman. *Partial Differential Equations of Parabolic Type*. Prentice-Hall, Englewood Cliffs, NJ, 1964.

Mark B. Garman and Steven W. Kohlhagen. Foreign currency option values. *Journal of International Money and Finance*, 2(3):231–237, 1983.

Jim Gatheral. A parsimonious arbitrage-free implied volatility parameterization with application to the valuation of volatility derivatives. Presentation at Global Derivatives & Risk Management, Madrid, May 2004.

Jim Gatheral. *The Volatility Surface: A Practitioner's Guide*. Wiley, Hoboken, NJ, 2006.

Jim Gatheral and Antoine Jacquier. Arbitrage-free SVI volatility surfaces. *Quantitative Finance*, 14(1):59–71, 2014.

Hélyette Geman, Nicole El Karoui, and Jean-Charles Rochet. Changes of numéraire, changes of probability measure and option pricing. *Journal of Applied Probability*, 32(2):443–458, 1995.

David Gentle. Basket weaving. *Risk*, 6(6):51–52, 1993.

J. Gil-Pelaez. Note on the inversion theorem. *Biometrika*, 38(3–4):481–482, 1951.

Michael Giles and Paul Glasserman. Smoking adjoints: Fast Monte Carlo Greeks. *Risk*, 19(1): 88–92, 2006.

Michael B. Giles. Improved multilevel Monte Carlo convergence using the Milstein scheme. In Alexander Keller, Stefan Heinrich, and Harald Niederreiter, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2006*, pages 343–358. Springer, Berlin, 2008a.

Michael B. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008b.

Michael B. Giles. Multilevel Monte Carlo methods. *Acta Numerica*, 24:259–328, 2015.

Michael B. Giles and Rebecca Carter. Convergence analysis of Crank–Nicolson and Rannacher time-marching. *Journal of Computational Finance*, 9(4):89–112, 2006.

Michael B. Giles and Lukasz Szpruch. Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation. *The Annals of Applied Probability*, 24(4): 1585–1620, 2014.

Michael B. Giles, Desmond J. Higham, and Xuerong Mao. Analysing multi-level Monte Carlo for options with non-globally Lipschitz payoff. *Finance and Stochastics*, 13(3):403–413, 2009.

Michael B. Giles, Kristian Debrabant, and Andreas Rößler. Analysis of multilevel Monte Carlo path simulation using the Milstein discretisation. *Discrete and Continuous Dynamical Systems, Series B*, 24(8):3881–3903, 2019.

Mike Giles. Approximating the erfinv function. In Wen-mei W. Hwu, editor, *GPU Computing Gems, Jade Edition*, chapter 10. Morgan Kaufmann, Boston, 2011.

Evarist Giné, Friedrich Götze, and David M. Mason. When is the Student t-statistic asymptotically standard normal? *Annals of Probability*, 25(3):1514–1531, 1997. doi: 10.1214/aop/1024404523.

Paul Glasserman. *Monte Carlo Methods in Financial Engineering*, volume 53 of *Applications of Mathematics*. Springer, New York, 2003.

Paul Glasserman and David D. Yao. Some guidelines and guarantees for common random numbers. *Management Science*, 38(6):884–908, 1992.

Paul Glasserman and Bin Yu. Number of paths versus number of basis functions in American option pricing. *The Annals of Applied Probability*, 14(4):2090–2119, 2004. doi: 10.1214/105051604000000846.

Paul Glasserman, Philip Heidelberger, and Perwez Shahabuddin. Asymptotically optimal importance sampling and stratification for pricing path-dependent options. *Mathematical Finance*, 9(2):117–152, 1999. doi: 10.1111/1467-9965.00065.

Peter W. Glynn. Optimization of stochastic systems via simulation. In Edward A. MacNair, Kenneth J. Musselman, and Philip Heidelberger, editors, *Proceedings of the 1989 Winter Simulation Conference*, pages 90–105, New York, 1989. ACM.

Peter W. Glynn and Ward Whitt. The asymptotic efficiency of simulation estimators. *Operations Research*, 40(3):505–520, 1992. doi: 10.1287/opre.40.3.505.

Emmanuel Gobet. Weak approximation of killed diffusion using Euler schemes. *Stochastic Processes and their Applications*, 87(2):167–197, 2000.

Emmanuel Gobet. A correction note to "Discrete time hedging errors for options with irregular payoffs". *Finance and Stochastics*, 18(2):483–485, 2014.

Emmanuel Gobet and Emmanuel Temam. Discrete time hedging errors for options with irregular payoffs. *Finance and Stochastics*, 5(3):357–367, 2001.

Sergei K. Godunov. A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics. *Matematicheskii Sbornik*, 47(89)(3):271–306, 1959. In Russian.

M. Barry Goldman, Howard B. Sosin, and Mary Ann Gatto. Path dependent options: "buy at the low, sell at the high". *Journal of Finance*, 34(5):1111–1127, 1979.

Andreas Griewank and Andrea Walther. *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*. SIAM, Philadelphia, 2nd edition, 2008.

John H. Halton. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, 2:84–90, 1960. doi: 10.1007/BF01386213.

J. M. Hammersley and K. W. Morton. A new Monte Carlo technique: Antithetic variates. *Mathematical Proceedings of the Cambridge Philosophical Society*, 52(3):449–475, 1956.

J. Michael Harrison and Stanley R. Pliska. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications*, 11(3):215–260, 1981.

Martin B. Haugh and Leonid Kogan. Pricing American options: A duality approach. *Operations Research*, 52(2):258–270, 2004. doi: 10.1287/opre.1030.0070.

Stefan Heinrich. Multilevel Monte Carlo methods. In Svetozar Margenov, Jerzy Waśniewski, and Plamen Yalamov, editors, *Large-Scale Scientific Computing*, volume 2179 of *Lecture Notes in Computer Science*, pages 58–67. Springer, Berlin, 2001.

Steven L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343, 1993.

Desmond J. Higham. *An Introduction to Financial Option Valuation: Mathematics, Stochastics and Computation*. Cambridge University Press, Cambridge, 2004.

David G. Hobson. Volatility misspecification, option pricing and superreplication via coupling. *Annals of Applied Probability*, 8(1):193–205, 1998.

John Hull and Alan White. The pricing of options on assets with stochastic volatilities. *The Journal of Finance*, 42(2):281–300, 1987. doi: 10.1111/j.1540-6261.1987.tb02568.x.

T. E. Hull and A. R. Dobell. Random number generators. *SIAM Review*, 4(3):230–254, 1962.

Martin Hutzenthaler, Arnulf Jentzen, and Peter E. Kloeden. Strong and weak divergence in finite time of Euler's method for stochastic differential equations with non-globally Lipschitz continuous coefficients. *Proceedings of the Royal Society A*, 467(2130):1563–1576, 2011.

Karel J. in 't Hout and Sven Foulon. ADI finite difference schemes for option pricing in the Heston model with correlation. *International Journal of Numerical Analysis and Modeling*, 7 (2):303–320, 2010.

Peter Jäckel. Let's be rational. *Wilmott*, 2015(75):40–53, 2015.

Patrick Jaillet, Damien Lamberton, and Bernard Lapeyre. Variational inequalities and the pricing of American options. *Acta Applicandae Mathematicae*, 21(3):263–289, 1990.

Robert Jarrow and Andrew Rudd. Approximate option valuation for arbitrary stochastic processes. *Journal of Financial Economics*, 10(3):347–369, 1982.

Robert A. Jarrow and Andrew Rudd. *Option Pricing*. Irwin Series in Finance. Dow Jones-Irwin, Homewood, IL, 1983.

Stephen Joe and Frances Y. Kuo. Constructing Sobol' sequences with better two-dimensional projections. *SIAM Journal on Scientific Computing*, 30(5):2635–2654, 2008. doi: 10.1137/070709359.

Fritz John. *Partial Differential Equations*, volume 1 of *Applied Mathematical Sciences*. Springer, New York, 4th edition, 1982.

Mark S. Joshi. Achieving higher order convergence for the prices of European options in binomial trees. *Mathematical Finance*, 20(1):89–103, 2010.

Yuri Kabanov and Mher Safarian. *Markets with Transaction Costs: Mathematical Theory*. Springer Finance. Springer, Berlin, 2009.

Yuri M. Kabanov and Mher M. Safarian. On Leland's strategy of option pricing with transactions costs. *Finance and Stochastics*, 1(3):239–250, 1997.

Mark Kac. On distributions of certain Wiener functionals. *Transactions of the American Mathematical Society*, 65(1):1–13, 1949.

Christian Kahl and Peter Jäckel. Not-so-complex logarithms in the Heston model. *Wilmott Magazine*, pages 94–103, September 2005.

Michael Kamal and Emanuel Derman. Correcting Black–Scholes. *Risk*, pages 82–85, January 1999.

Bardia Kamrad and Peter Ritchken. Multinomial approximating models for options with k state variables. *Management Science*, 37(12):1640–1652, 1991.

Raul Kangro and Roy Nicolaides. Far field boundary conditions for Black–Scholes equations. *SIAM Journal on Numerical Analysis*, 38(4):1357–1368, 2000.

Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, New York, 2nd edition, 1991.

Ahmed Kebaier. Statistical Romberg extrapolation: A new variance reduction method and applications to option pricing. *The Annals of Applied Probability*, 15(4):2681–2705, 2005.

Martin Keller-Ressel. Moment explosions and long-term behavior of affine stochastic volatility models. *Mathematical Finance*, 21(1):73–98, 2011.

A. G. Z. Kemna and A. C. F. Vorst. A pricing method for options based on average asset values. *Journal of Banking & Finance*, 14(1):113–129, 1990.

Peter E. Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, Berlin, 1992.

Donald E. Knuth. *The Art of Computer Programming, Volume 2: Seminumerical Algorithms*. Addison-Wesley, Reading, MA, 3rd edition, 1997.

Andrei Kolmogoroff. Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Mathematische Annalen*, 104:415–458, 1931.

S. G. Kou. On pricing of discrete barrier options. *Statistica Sinica*, 13(4):955–964, 2003.

Naoto Kunitomo and Masayuki Ikeda. Pricing options with curved boundaries. *Mathematical Finance*, 2(4):275–298, 1992.

Damien Lamberton. On the binomial approximation of the American put. *Applied Mathematics and Optimization*, 82(2):687–720, 2020. doi: 10.1007/s00245-018-9545-2. Preprint arXiv:1802.05614.

Stephen S. Lavenberg and Peter D. Welch. A perspective on the use of control variables to increase the efficiency of Monte Carlo simulations. *Management Science*, 27(3):322–335, 1981.

Stephen S. Lavenberg, Thomas L. Moeller, and Peter D. Welch. Statistical results on control variables with application to queueing network simulation. *Operations Research*, 30(1):182–202, 1982.

Peter D. Lax and Robert D. Richtmyer. Survey of the stability of linear finite difference equations. *Communications on Pure and Applied Mathematics*, 9(2):267–293, 1956.

Pierre L'Ecuyer. Good parameters and implementations for combined multiple recursive random number generators. *Operations Research*, 47(1):159–164, 1999.

Pierre L'Ecuyer and Gaétan Perron. On the convergence rates of IPA and FDC derivative estimators. *Operations Research*, 42(4):643–656, 1994.

Pierre L'Ecuyer and Richard Simard. TestU01: A C library for empirical testing of random number generators. *ACM Transactions on Mathematical Software*, 33(4):Article 22, 2007.

Roger W. Lee. The moment formula for implied volatility at extreme strikes. *Mathematical Finance*, 14(3):469–480, 2004.

Dietmar P. J. Leisen and Matthias Reimer. Binomial models for option valuation – examining and improving convergence. *Applied Mathematical Finance*, 3(4):319–346, 1996.

Hayne E. Leland. Option pricing and replication with transactions costs. *Journal of Finance*, 40 (5):1283–1301, 1985.

Edmond Levy. Pricing European average rate currency options. *Journal of International Money and Finance*, 11(5):474–491, 1992.

Alan L. Lewis. A simple option formula for general jump-diffusion and other exponential Lévy processes. Working paper, OptionCity.net, 2001. SSRN 282110.

B. F. Logan, C. L. Mallows, S. O. Rice, and L. A. Shepp. Limit distributions of self-normalized sums. *Annals of Probability*, 1(5):788–809, 1973. doi: 10.1214/aop/1176996846.

Wei-Liem Loh. On the asymptotic distribution of scrambled net quadrature. *The Annals of Statistics*, 31(4):1282–1324, 2003. doi: 10.1214/aos/1059655914.

Francis A. Longstaff and Eduardo S. Schwartz. Valuing American options by simulation: A simple least-squares approach. *The Review of Financial Studies*, 14(1):113–147, 2001.

Roger Lord and Christian Kahl. Why the rotation count algorithm works. Discussion Paper TI 2006-065/2, Tinbergen Institute, 2006.

Roger Lord and Christian Kahl. Complex logarithms in Heston-like models. *Mathematical Finance*, 20(4):671–694, 2010.

Roger Lord, Remmert Koekkoek, and Dick van Dijk. A comparison of biased simulation schemes for stochastic volatility models. *Quantitative Finance*, 10(2):177–194, 2010. doi: 10.1080/14697680802392496.

Klaus Lott. *Ein Verfahren zur Replikation von Optionen unter Transaktionskosten in stetiger Zeit*. PhD thesis, Universität der Bundeswehr München, 1993.

Steven Manaster and Gary Koehler. The calculation of implied variances from the Black–Scholes model: A note. *Journal of Finance*, 37(1):227–230, 1982.

William Margrabe. The value of an option to exchange one asset for another. *Journal of Finance*, 33(1):177–186, 1978.

G. Marsaglia and T. A. Bray. A convenient method for generating normal variables. *SIAM Review*, 6(3):260–264, 1964.

George Marsaglia. Random numbers fall mainly in the planes. *Proceedings of the National Academy of Sciences of the USA*, 61(1):25–28, 1968.

Makoto Matsumoto and Takuji Nishimura. Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*, 8(1):3–30, 1998.

M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21 (2):239–245, 1979. doi: 10.1080/00401706.1979.10489755.

Don McLeish. A general method for debiasing a Monte Carlo estimator. *Monte Carlo Methods and Applications*, 17(4):301–315, 2011.

Robert C. Merton. Theory of rational option pricing. *Bell Journal of Economics and Management Science*, 4(1):141–183, 1973.

G. N. Mil'shtein. Approximate integration of stochastic differential equations. *Theory of Probability and Its Applications*, 19(3):557–562, 1975. Russian original: Teoriya Veroyatnostei i ee Primeneniya 19(3), 583–588, 1974.

Boris Moro. The full Monte. *Risk*, 8(2):57–58, 1995.

Peter Mörters and Yuval Peres. *Brownian Motion*. Cambridge University Press, Cambridge, 2010.

Bradley Moskowitz and Russel E. Caflisch. Smoothness and dimension reduction in quasi-Monte Carlo methods. *Mathematical and Computer Modelling*, 23(8–9):37–54, 1996. doi: 10.1016/0895-7177(96)00038-6.

Henry R. Neave. On using the Box–Muller transformation with multiplicative congruential pseudo-random number generators. *Applied Statistics*, 22(1):92–97, 1973.

Barry L. Nelson. Control variate remedies. *Operations Research*, 38(6):974–992, 1990.

Jerzy Neyman. On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97(4):558–625, 1934. doi: 10.2307/2342192.

Thai Huu Nguyen and Serguei Pergamenshchikov. Approximate hedging problem with transaction costs in stochastic volatility markets. *Mathematical Finance*, 27(3):832–865, 2017.

Harald Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*, volume 63 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, Philadelphia, 1992. doi: 10.1137/1.9781611970081.

Bernt Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Springer, Berlin, 6th edition, 2003.

Melissa E. O'Neill. PCG: A family of simple fast space-efficient statistically good algorithms for random number generation. Technical Report HMC-CS-2014-0905, Harvey Mudd College, Claremont, CA, 2014.

Art B. Owen. A central limit theorem for Latin hypercube sampling. *Journal of the Royal Statistical Society: Series B*, 54(2):541–551, 1992.

Art B. Owen. Randomly permuted (t, m, s)-nets and (t, s)-sequences. In Harald Niederreiter and Peter Jau-Shyong Shiue, editors, *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, volume 106 of *Lecture Notes in Statistics*, pages 299–317. Springer, New York, 1995. doi: 10.1007/978-1-4612-2552-2_19.

Art B. Owen. Scrambled net variance for integrals of smooth functions. *The Annals of Statistics*, 25(4):1541–1562, 1997a. doi: 10.1214/aos/1031594731.

Art B. Owen. Monte Carlo variance of scrambled net quadrature. *SIAM Journal on Numerical Analysis*, 34(5):1884–1910, 1997b. doi: 10.1137/S0036142994277468.

Art B. Owen. Scrambling Sobol' and Niederreiter–Xing points. *Journal of Complexity*, 14(4): 466–489, 1998. doi: 10.1006/jcom.1998.0487.

Art B. Owen. Multidimensional variation for quasi-Monte Carlo. In Jianqing Fan and Gang Li, editors, *Contemporary Multivariate Analysis and Design of Experiments*, pages 49–74. World Scientific, Singapore, 2005. doi: 10.1142/9789812567765_0004.

Art B. Owen. Halton sequences avoid the origin. *SIAM Review*, 48(3):487–503, 2006. doi: 10.1137/S0036144504441573.

Art B. Owen. Local antithetic sampling with scrambled nets. *The Annals of Statistics*, 36(5): 2319–2343, 2008. doi: 10.1214/07-AOS548.

Art B. Owen. Monte Carlo theory, methods and examples. https://artowen.su.domains/mc/, 2013.

Anargyros Papageorgiou. The Brownian bridge does not offer a consistent advantage in quasi-Monte Carlo integration. *Journal of Complexity*, 18(1):171–186, 2002. doi: 10.1006/jcom.2001.0631.

Stephen K. Park and Keith W. Miller. Random number generators: Good ones are hard to find. *Communications of the ACM*, 31(10):1192–1201, 1988.

Donald W. Peaceman and Henry H. Rachford. The numerical solution of parabolic and elliptic differential equations. *Journal of the Society for Industrial and Applied Mathematics*, 3(1): 28–41, 1955.

David B. Peizer and John W. Pratt. A normal approximation for binomial, F, beta, and other common, related tail probabilities, I. *Journal of the American Statistical Association*, 63(324): 1416–1456, 1968.

Goran Peskir and Albert Shiryaev. *Optimal Stopping and Free-Boundary Problems*. Lectures in Mathematics ETH Zürich. Birkhäuser, Basel, 2006.

David M. Pooley, Kenneth R. Vetzal, and Peter A. Forsyth. Convergence remedies for non-smooth payoffs in option pricing. *Journal of Computational Finance*, 6(4):25–40, 2003.

Rolf Rannacher. Finite element solution of diffusion problems with irregular data. *Numerische Mathematik*, 43(2):309–327, 1984.

Daniel Revuz and Marc Yor. *Continuous Martingales and Brownian Motion*, volume 293 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, 3rd edition, 1999.

Chang-han Rhee and Peter W. Glynn. Unbiased estimation with square root convergence for SDE models. *Operations Research*, 63(5):1026–1043, 2015.

L. C. G. Rogers. Monte Carlo valuation of American options. *Mathematical Finance*, 12(3): 271–286, 2002. doi: 10.1111/1467-9965.02010.

L. C. G. Rogers and Z. Shi. The value of an Asian option. *Journal of Applied Probability*, 32(4): 1077–1088, 1995.

Marc Romano and Nizar Touzi. Contingent claims and market completeness in a stochastic volatility model. *Mathematical Finance*, 7(4):399–412, 1997. doi: 10.1111/1467-9965.00038.

Michael Roper. Arbitrage free implied volatility surfaces. Preprint, School of Mathematics and Statistics, The University of Sydney, 2010.

Klaus F. Roth. On irregularities of distribution. *Mathematika*, 1(2):73–79, 1954. doi: 10.1112/S0025579300000541.

Mark Rubinstein. Implied binomial trees. *Journal of Finance*, 49(3):771–818, 1994.

Mark Rubinstein and Eric Reiner. Breaking down the barriers. *Risk*, 4(8):28–35, 1991.

Reuven Y. Rubinstein. Optimization of computer simulation models with rare events. *European Journal of Operational Research*, 99(1):89–112, 1997. doi: 10.1016/S0377-2217(96)00385-2.

Walter Rudin. *Real and Complex Analysis*. McGraw-Hill, New York, 3rd edition, 1987.

Wolfgang M. Schmidt. Irregularities of distribution. VII. *Acta Arithmetica*, 21:45–50, 1972. doi: 10.4064/aa-21-1-45-50.

Irina Shevtsova. On the absolute constants in the Berry–Esseen type inequalities for identically distributed summands. arXiv preprint arXiv:1111.6554, 2011.

Steven E. Shreve. *Stochastic Calculus for Finance II: Continuous-Time Models*. Springer, New York, 2004.

Carlos A. Sin. Complications with stochastic volatility models. *Advances in Applied Probability*, 30(1):256–268, 1998.

Ilya M. Sobol'. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, 7(4):86–112, 1967. doi: 10.1016/0041-5553(67)90144-9.

Ilya M. Sobol'. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1–3):271–280, 2001. doi: 10.1016/S0378-4754(00)00270-6.

Michael Stein. Large sample properties of simulations using Latin hypercube sampling. *Technometrics*, 29(2):143–151, 1987. doi: 10.1080/00401706.1987.10488205.

Robert S. Strichartz. *A Guide to Distribution Theory and Fourier Transforms*. World Scientific, River Edge, NJ, 2003.

John C. Strikwerda. *Finite Difference Schemes and Partial Differential Equations*. SIAM, Philadelphia, 2nd edition, 2004.

René M. Stulz. Options on the minimum or the maximum of two risky assets: Analysis and applications. *Journal of Financial Economics*, 10(2):161–185, 1982.

Denis Talay and Luciano Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Analysis and Applications*, 8(4):483–509, 1990.

A. Talbot. The accurate numerical inversion of Laplace transforms. *IMA Journal of Applied Mathematics*, 23(1):97–120, 1979.

Klaus Bjerre Toft. On the mean-variance tradeoff in option replication with transactions costs. *Journal of Financial and Quantitative Analysis*, 31(2):233–263, 1996.

Lloyd N. Trefethen and J. A. C. Weideman. The exponentially convergent trapezoidal rule. *SIAM Review*, 56(3):385–458, 2014.

John N. Tsitsiklis and Benjamin Van Roy. Regression methods for pricing complex American-style options. *IEEE Transactions on Neural Networks*, 12(4):694–703, 2001.

S. M. Turnbull and L. M. Wakeman. A quick algorithm for pricing European average options. *Journal of Financial and Quantitative Analysis*, 26(3):377–389, 1991.

A. Tychonoff. Théorèmes d'unicité pour l'équation de la chaleur. *Matematicheskii Sbornik*, 42 (2):199–216, 1935.

A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, Cambridge, 1998.

John von Neumann. Various techniques used in connection with random digits. In *Monte Carlo Method*, volume 12 of *National Bureau of Standards Applied Mathematics Series*, pages 36–38. U.S. Government Printing Office, Washington, DC, 1951.

John B. Walsh. The rate of convergence of the binomial tree scheme. *Finance and Stochastics*, 7 (3):337–361, 2003.

Michael J. Wichura. Algorithm AS 241: The percentage points of the normal distribution. *Applied Statistics*, 37(3):477–484, 1988.

Gregory A. Willard. Calculating prices and sensitivities for path-independent derivative securities in multifactor models. *The Journal of Derivatives*, 5(1):45–61, 1997.

Paul Wilmott, Sam Howison, and Jeff Dewynne. *The Mathematics of Financial Derivatives: A Student Introduction*. Cambridge University Press, Cambridge, 1995.

R. Zhang. *Couverture approchée des options européennes*. PhD thesis, École Nationale des Ponts et Chaussées, 1999.