Springer Series in Statistics

Springer Series in Statistics

Alho/Spencer: Statistical Demography and Forecasting
Andersen/Borgan/Gill/Keiding: Statistical Models Based on Counting Processes
Atkinson/Riani: Robust Diagnostic Regression Analysis
Atkinson/Riani/Ceriloi: Exploring Multivariate Data with the Forward Search
Berger: Statistical Decision Theory and Bayesian Analysis, 2nd edition
Borg/Groenen: Modern Multidimensional Scaling: Theory and Applications, 2nd edition
Brockwell/Davis: Time Series: Theory and Methods, 2nd edition
Bucklew: Introduction to Rare Event Simulation
Cappé/Moulines/Rydén: Inference in Hidden Markov Models
Chan/Tong: Chaos: A Statistical Perspective
Chen/Shao/Ibrahim: Monte Carlo Methods in Bayesian Computation
Coles: An Introduction to Statistical Modeling of Extreme Values
Devroye/Lugosi: Combinatorial Methods in Density Estimation
Diggle/Ribeiro: Model-based Geostatistics
Dudoit/Van der Laan: Multiple Testing Procedures with Applications to Genomics
Efromovich: Nonparametric Curve Estimation: Methods, Theory, and Applications
Eggermont/LaRiccia: Maximum Penalized Likelihood Estimation, Volume I: Density Estimation
Fahrmeir/Tutz: Multivariate Statistical Modeling Based on Generalized Linear Models, 2nd edition
Fan/Yao: Nonlinear Time Series: Nonparametric and Parametric Methods
Ferraty/Vieu: Nonparametric Functional Data Analysis: Theory and Practice
Ferreira/Lee: Multiscale Modeling: A Bayesian Perspective
Fienberg/Hoaglin: Selected Papers of Frederick Mosteller
Frühwirth-Schnatter: Finite Mixture and Markov Switching Models
Ghosh/Ramamoorthi: Bayesian Nonparametrics
Glaz/Naus/Wallenstein: Scan Statistics
Good: Permutation Tests: Parametric and Bootstrap Tests of Hypotheses, 3rd edition
Gouriéroux: ARCH Models and Financial Applications
Gu: Smoothing Spline ANOVA Models
Györfi/Kohler/Krzyźak/Walk: A Distribution-Free Theory of Nonparametric Regression
Haberman: Advanced Statistics, Volume I: Description of Populations
Hall: The Bootstrap and Edgeworth Expansion
Härdle: Smoothing Techniques: With Implementation in S
Harrell: Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis
Hart: Nonparametric Smoothing and Lack-of-Fit Tests
Hastie/Tibshirani/Friedman: The Elements of Statistical Learning: Data Mining, Inference, and Prediction
Hedayat/Sloane/Stufken: Orthogonal Arrays: Theory and Applications
Heyde: Quasi-Likelihood and Its Application: A General Approach to Optimal Parameter Estimation
Huet/Bouvier/Poursat/Jolivet: Statistical Tools for Nonlinear Regression: A Practical Guide with S-PLUS and R Examples, 2nd edition
Iacus: Simulation and Inference for Stochastic Differential Equations

(continued after index)

Stefano M. Iacus

Simulation and Inference for Stochastic Differential Equations

With R Examples

Springer

Stefano M. Iacus
Dept. Economics, Business and Statistics
University of Milan
Via Conservatorio, 7
20122 Milano
Italy
stefano.iacus@unimi.it

ISBN: 978-0-387-75838-1 e-ISBN: 978-0-387-75839-8
DOI: 10.1007/978-0-387-75839-8

Library of Congress Control Number:

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

To Ilia, Lucy, and Ludo

Preface

Stochastic differential equations model stochastic evolution as time evolves. These models have a variety of applications in many disciplines and emerge naturally in the study of many phenomena. Examples of these applications are physics (see, e.g., [176] for a review), astronomy [202], mechanics [147], economics [26], mathematical finance [115], geology [69], genetic analysis (see, e.g., [110], [132], and [155]), ecology [111], cognitive psychology (see, e.g., [102], and [221]), neurology [109], biology [194], biomedical sciences [20], epidemiology [17], political analysis and social processes [55], and many other fields of science and engineering. Although stochastic differential equations are quite popular models in the above-mentioned disciplines, there is a lot of mathematics behind them that is usually not trivial and for which details are not known to practitioners or experts of other fields. In order to make this book useful to a wider audience, we decided to keep the mathematical level of the book sufficiently low and often rely on heuristic arguments to stress the underlying ideas of the concepts introduced rather than insist on technical details. Mathematically oriented readers may find this approach inconvenient, but detailed references are always given in the text.

As the title of the book mentions, the aim of the book is twofold. The first is to recall the theory and implement methods for the simulation of paths of stochastic processes $\{X_t, t \geq 0\}$ solutions to stochastic differential equations (SDEs). In this respect, the title of the book is too ambitious in the sense that only SDEs with Gaussian noise are considered (i.e., processes for which the writing $\mathrm{d}X_t = S(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t$ has a meaning in the Itô sense). This part of the book contains a review of well-established results and their implementations in the R language, but also some fairly recent results on simulation.

The second part of the book is dedicated to the review of some methods of estimation for these classes of stochastic processes. While there is a well-established theory on estimation for continuous-time observations from these processes [149], the literature about discrete-time observations is dispersed (though vaste) in several journals. Of course, real data (e.g., from finance [47],

[88]) always lead to dealing with discrete-time observations $\{X_{t_i}, i = 1, \ldots, n\}$, and many of the results from the continuous-time case do not hold or cannot be applied (for example, the likelihood of the observations is almost always unavailable in explicit form). It should be noted that only the observations are discrete whilst the underlying model is continuous; hence most of the standard theory on discrete-time Markov processes does not hold as well.

Different schemes of observations can be considered depending on the nature of the data, and the estimation part of the problem is not necessarily the same for the different schemes. One case, which is considered "natural," is the fixed-Δ scheme, in which the time step between two subsequent observations X_{t_i} and $X_{t_i+\Delta_n}$ is fixed; i.e., $\Delta_n = \Delta$ (or is bounded away from zero) and independent from n. In this case, the process is observed on the time interval $[0, T = n\Delta]$ and the asymptotics considered as $n \to \infty$ (*large-sample asymptotics*). The underlying model might be ergodic or stationary and possibly homogeneous. For such a scheme, the time step Δ might have some influence on estimators because, for example, the transition density of the process is usually not known in explicit form and has to be approximated via simulations. This is the most difficult case to handle.

Another scheme is the "high frequency" scheme, in which the observational step size Δ_n decreases with n and two cases are possible: the time interval is fixed, say $[0, T = n\Delta_n]$, or $n\Delta_n$ increases as well. In the first case, neither homogeneity nor erogidicy are needed, but consistent estimators are not always available. On the contrary, in the "*rapidly increasing experimental design*," when $\Delta_n \to 0$ and $n\Delta_n \to \infty$ but $n\Delta_n^2 \to 0$, consistent estimators can be obtained along with some distributional results.

Other interesting schemes of partially observed processes, missing at random [75], thresholded processes (see, e.g., [116], [118]), observations with errors (quantized or interval data, see, e.g., [66], [67], [97]), or large sample and "small diffusion" asymptotics have also recently appeared in the literature (see, e.g., [222], [217]). This book covers essentially the parametric estimation under the large-sample asymptotics scheme ($n\Delta_n \to \infty$) with either fixed $\Delta_n = \Delta$ or $\Delta_n \to 0$ with $n\Delta_n^k \to 0$ for some $k \geq 2$. The final chapter contains a miscellaneous selection of results, including nonparametric estimation, model selection, and change-point problems.

This book is intended for practitioners and is not a theoretical book, so this second part just recalls briefly the main results and the ideas behind the methods and implements several of them in the R language. A selection of the results has necessarily been made. This part of the book also shows the difference between the theory of estimation for discrete-time observations and the actual performance of such estimators once implemented. Further, the effect of approximation schemes on estimators is investigated throughout the text. Theoretical results are recalled as "Facts" and regularity conditions as "Assumptions" and numbered by chapter in the text.

So what is this book about?

This book is about ready to be used, R-efficient code for simulation schemes of stochastic differential equations and some related estimation methods based on discrete sampled observations from such models. We hope that the code presented here and the updated survey on the subject might be of help for practitioners, postgraduate and PhD students, and researchers in the field who might want to implement new methods and ideas using R as a statistical environment.

What this book is not about

This book is not intended to be a theoretical book or an exhaustive collection of all the statistical methods available for discretely observed diffusion processes. This book might be thought of as a companion book to some advanced theoretical publication (already available or forthcoming) on the subject. Although this book is not even a textbook, some previous drafts of it have been used with success in mathematical finance classes for the numerical simulation and empirical analysis of financial time series.

What comes with the book

All the algorithms presented in the book are written in pure R code but, because of the speed needed in real-world applications, we have rewritten some of the R code in the `C` language and assembled everything in a package called `sde` freely available on CRAN, the Comprehensive R Archive Network. R and `C` functions have the same end-user interface; hence all the code of the examples in the book will run smoothly regardless of the underlying coding language. A minimal knowledge of the R environment at the introductory level is assumed, although brief recalls to the main R concepts, limited to what is relevant to this text, are given at the end of the book. Some crucial aspects of implementation are discussed in the main body of the book to make them more effective.

What is missing?

This book essentially covers one-dimensional diffusion processes driven by the Wiener process. Today's literature is vast and wider than this choice. In particular, it focuses also on multidimensional diffusion processes and stochastic differential equations driven by Lévy processes. To keep the book self-contained and at an introductory level and to preserve some homogeneity within the text, we decided to restrict the field. This also allows simple and easy-to-understand R code to be written for each of the techniques presented.

Acknowledgments

I am grateful to Alessandro De Gregorio and Ilia Negri for careful reading of the manuscript, as well as to the anonymous referees for their constructive comments and remarks on several drafts of the book. All the remaining errors contained are of course my responsibility. I also wish to thank John Kimmel for his great support and patience with me during the last three years.

Milan, November 2007 *Stefano M. Iacus*

Contents

Preface VII

Notation XVII

1 Stochastic Processes and Stochastic Differential Equations 1
1.1 Elements of probability and random variables 1
1.1.1 Mean, variance, and moments 2
1.1.2 Change of measure and Radon-Nikodým derivative 4
1.2 Random number generation 5
1.3 The Monte Carlo method 5
1.4 Variance reduction techniques 8
1.4.1 Preferential sampling 9
1.4.2 Control variables 12
1.4.3 Antithetic sampling 13
1.5 Generalities of stochastic processes 14
1.5.1 Filtrations 14
1.5.2 Simple and quadratic variation of a process 15
1.5.3 Moments, covariance, and increments of stochastic processes 16
1.5.4 Conditional expectation 16
1.5.5 Martingales 18
1.6 Brownian motion 18
1.6.1 Brownian motion as the limit of a random walk 20
1.6.2 Brownian motion as $L^2[0,T]$ expansion 22
1.6.3 Brownian motion paths are nowhere differentiable 24
1.7 Geometric Brownian motion 24
1.8 Brownian bridge 27
1.9 Stochastic integrals and stochastic differential equations 29
1.9.1 Properties of the stochastic integral and Itô processes 32

1.10 Diffusion processes 33
1.10.1 Ergodicity 35
1.10.2 Markovianity 36
1.10.3 Quadratic variation 37
1.10.4 Infinitesimal generator of a diffusion process 37
1.10.5 How to obtain a martingale from a diffusion process . 37
1.11 Itô formula 38
1.11.1 Orders of differentials in the Itô formula 38
1.11.2 Linear stochastic differential equations 39
1.11.3 Derivation of the SDE for the geometric Brownian motion 39
1.11.4 The Lamperti transform 40
1.12 Girsanov's theorem and likelihood ratio for diffusion processes 41
1.13 Some parametric families of stochastic processes 43
1.13.1 Ornstein-Uhlenbeck or Vasicek process 43
1.13.2 The Black-Scholes-Merton or geometric Brownian motion model 46
1.13.3 The Cox-Ingersoll-Ross model 47
1.13.4 The CKLS family of models 49
1.13.5 The modified CIR and hyperbolic processes 49
1.13.6 The hyperbolic processes 50
1.13.7 The nonlinear mean reversion Aït-Sahalia model 50
1.13.8 Double-well potential 51
1.13.9 The Jacobi diffusion process 51
1.13.10 Ahn and Gao model or inverse of Feller's square root model 52
1.13.11 Radial Ornstein-Uhlenbeck process 52
1.13.12 Pearson diffusions 52
1.13.13 Another classification of linear stochastic systems . . . 54
1.13.14 One epidemic model 56
1.13.15 The stochastic cusp catastrophe model 57
1.13.16 Exponential families of diffusions 58
1.13.17 Generalized inverse gaussian diffusions 59

2 Numerical Methods for SDE 61
2.1 Euler approximation 62
2.1.1 A note on code vectorization 63
2.2 Milstein scheme 65
2.3 Relationship between Milstein and Euler schemes 66
2.3.1 Transform of the geometric Brownian motion 68
2.3.2 Transform of the Cox-Ingersoll-Ross process 68
2.4 Implementation of Euler and Milstein schemes: the `sde.sim` function 69
2.4.1 Example of use 70

2.5 The constant elasticity of variance process and strange paths . . . 72
2.6 Predictor-corrector method . . . 72
2.7 Strong convergence for Euler and Milstein schemes . . . 74
2.8 KPS method of strong order $\gamma = 1.5$ . . . 77
2.9 Second Milstein scheme . . . 81
2.10 Drawing from the transition density . . . 82
2.10.1 The Ornstein-Uhlenbeck or Vasicek process . . . 83
2.10.2 The Black and Scholes process . . . 83
2.10.3 The CIR process . . . 83
2.10.4 Drawing from one model of the previous classes . . . 84
2.11 Local linearization method . . . 85
2.11.1 The Ozaki method . . . 85
2.11.2 The Shoji-Ozaki method . . . 87
2.12 Exact sampling . . . 91
2.13 Simulation of diffusion bridges . . . 98
2.13.1 The algorithm . . . 99
2.14 Numerical considerations about the Euler scheme . . . 101
2.15 Variance reduction techniques . . . 102
2.15.1 Control variables . . . 103
2.16 Summary of the function sde.sim . . . 105
2.17 Tips and tricks on simulation . . . 106

3 Parametric Estimation . . . 109
3.1 Exact likelihood inference . . . 112
3.1.1 The Ornstein-Uhlenbeck or Vasicek model . . . 113
3.1.2 The Black and Scholes or geometric Brownian motion model . . . 117
3.1.3 The Cox-Ingersoll-Ross model . . . 119
3.2 Pseudo-likelihood methods . . . 122
3.2.1 Euler method . . . 122
3.2.2 Elerian method . . . 125
3.2.3 Local linearization methods . . . 127
3.2.4 Comparison of pseudo-likelihoods . . . 128
3.3 Approximated likelihood methods . . . 131
3.3.1 Kessler method . . . 131
3.3.2 Simulated likelihood method . . . 134
3.3.3 Hermite polynomials expansion of the likelihood . . . 138
3.4 Bayesian estimation . . . 155
3.5 Estimating functions . . . 157
3.5.1 Simple estimating functions . . . 157
3.5.2 Algorithm 1 for simple estimating functions . . . 164
3.5.3 Algorithm 2 for simple estimating functions . . . 167
3.5.4 Martingale estimating functions . . . 172
3.5.5 Polynomial martingale estimating functions . . . 173

3.5.6 Estimating functions based on eigenfunctions 178
3.5.7 Estimating functions based on transform functions . . 179
3.6 Discretization of continuous-time estimators 179
3.7 Generalized method of moments 182
3.7.1 The GMM algorithm 184
3.7.2 GMM, stochastic differential equations, and Euler method 185
3.8 What about multidimensional diffusion processes? 190

4 Miscellaneous Topics 191
4.1 Model identification via Akaike's information criterion 191
4.2 Nonparametric estimation 197
4.2.1 Stationary density estimation 198
4.2.2 Local-time and stationary density estimators 201
4.2.3 Estimation of diffusion and drift coefficients 202
4.3 Change-point estimation 208
4.3.1 Estimation of the change point with unknown drift . . 212
4.3.2 A famous example 215

Appendix A: A brief excursus into R 217
A.1 Typing into the R console 217
A.2 Assignments 218
A.3 R vectors and linear algebra 220
A.4 Subsetting 221
A.5 Different types of objects 222
A.6 Expressions and functions 225
A.7 Loops and vectorization 227
A.8 Environments 228
A.9 Time series objects 229
A.10 R Scripts 231
A.11 Miscellanea 232

Appendix B: The sde Package 233
BM 234
cpoint 235
DBridge 236
dcElerian 237
dcEuler 238
dcKessler 238
dcOzaki 239
dcShoji 240
dcSim 241
DWJ 243
EULERloglik 243
gmm 245

HPloglik 247
ksmooth 248
linear.mart.ef 250
rcBS 251
rcCIR 252
rcOU 253
rsCIR 254
rsOU 255
sde.sim 256
sdeAIC 259
SIMloglik 261
simple.ef 262
simple.ef2 264

References 267

Index 279

Notation

Var : the variance operator

$\mathbb{E}$: the expected value operator

$N(\mu, \sigma^2)$: the Gaussian law with mean μ and variance σ^2

χ^2_d : chi-squared distribution with d degrees of freedom

t_d : Student t distribution with d degrees of freedom

$I_\nu(x)$: modified Bessel function of the first kind of order ν

$\mathbb{R}$: the real line

$\mathbb{N}$: the set of natural numbers $1, 2, \ldots$

$\xrightarrow{d}$: convergence in distribution

$\xrightarrow{p}$: convergence in probability

$\xrightarrow{a.s.}$: almost sure convergence

$\mathcal{B}(\mathbb{R})$: Borel σ-algebra on $\mathbb{R}$

$\chi_A, \mathbf{1}_A$: indicator function of the set A

$x \wedge y$: $\min(x, y)$

$x \vee y$: $\max(x, y)$

$(f(x))_+$: $\max(f(x), 0)$

$\Phi(z)$: cumulative distribution function of standard Gaussian law

$[x]$: integer part of x

$<X, X>_t, [X, X]_t$: quadratic variation process associated to X_t

$\mathrm{V}_t(X)$: simple variation of process X

$\propto$: proportional to

$f_{v_i}(v_1, v_2, \ldots, v_n) : \frac{\partial}{\partial v_i} f(v_1, v_2, \ldots, v_n)$

$f_{v_i,v_j}(v_1, v_2, \ldots, v_n) : \frac{\partial^2}{\partial v_i v_j} f(v_1, v_2, \ldots, v_n)$, etc.

$\partial_\theta f(v_1, v_2, \ldots, v_n; \theta) : \frac{\partial}{\partial \theta} f(v_1, v_2, \ldots, v_n; \theta)$

$\partial_\theta^k f(v_1, v_2, \ldots, v_n; \theta) : \frac{\partial^k}{\partial \theta^k} f(v_1, v_2, \ldots, v_n; \theta)$

$\Pi_n(A)$: partition of the interval $A = [a, b]$ in n subintervals of $[a = x_0, x_1)$, $[x_1, x_2)$, …, $[x_{n-1}, x_n = b]$

$||\Pi_n|| : \max_j |x_{j+1} - x_j|$

$C_0^2(\mathbb{R})$: space of functions with compact support and continuous derivatives up to order 2

$L^2([0, T])$: space of functions from $[0, T] \to \mathbb{R}$ endowed by the L^2 norm

$||f||_2$: the L^2 norm of f

W_t : Brownian motion or Wiener process

i.i.d. : independent and identically distributed

AIC : Akaike information criterion

CIR : Cox-Ingersoll-Ross

CRAN : the Comprehensive R Archive Network

CKLS : Chan-Karolyi-Longstaff-Sanders

EA : exact algorithm

GMM : generalized method of moments

MCMC : Markov chain Monte Carlo

MISE : mean integrated square error

1

Stochastic Processes and Stochastic Differential Equations

This chapter reviews basic material on stochastic processes and statistics as well as stochastic calculus, mostly borrowed from [170], [130], and [193]. It also covers basic notions on simulation of commonly used stochastic processes such as random walks and Brownian motion and also recalls some Monte Carlo concepts. Even if the reader is assumed to be familiar with these basic notions, we will present them here in order to introduce the notation we will use throughout the text. We will limit our attention mainly to one-dimensional, real random variables and stochastic processes. We also restrict our attention to parametric models with multidimensional parameters.

1.1 Elements of probability and random variables

A probability space is a triple $(\Omega, \mathcal{A}, P)$ where Ω is the *sample space* of possible outcomes of a random experiment; $\mathcal{A}$ is a σ-*algebra*: i.e., $\mathcal{A}$ is a collection of sets such that *i)* the empty set $\emptyset$ is in $\mathcal{A}$; *ii)* if $A \in \mathcal{A}$, then the complementary set $\bar{A} \in \mathcal{A}$; *iii)* if $A_1, A_2, \ldots \in \mathcal{A}$, then

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}\,.$$

P is a probability measure on $(\Omega, \mathcal{A})$. In practice, $\mathcal{A}$ forms the collection of events for which a probability can be assigned. Given a probability space $(\Omega, \mathcal{A}, P)$, a *random variable* X is defined as a *measurable* function from Ω to $\mathbb{R}$,

$$X : \Omega \mapsto \mathbb{R}\,.$$

In the above, the term measurable intuitively means that it is always possible to calculate probabilities related to the random variable X. More precisely, denote by $\mathcal{B}(\mathbb{R})$ the Borel σ-algebra on $\mathbb{R}$ (i.e., the σ-algebra generated by the open sets of $\mathbb{R}$) and let X^{-1} be the inverse function of X. Then, X is *measurable* if

S.M. Iacus, *Simulation and Inference for Stochastic Differential Equations*,
doi: 10.1007/978-0-387-75839-8_1,

$$\forall A \in \mathcal{B}(\mathbb{R}), \quad \exists B \in \mathcal{A} : X^{-1}(A) = B;$$

i.e., such that it is always possible to measure the set of values assumed by X using the probability measure P on the original space Ω,

$$P(X \in A) = P(\{\omega \in \Omega : X(\omega) \in A\}) = P(\{\omega \in \Omega : \omega \in X^{-1}(A)\}) = P(B),$$

for $A \in \mathcal{B}(\mathbb{R})$ and $B \in \mathcal{A}$.

Distribution and density function

The function $F(x) = P(X \leq x) = P(X(\omega) \in (-\infty, x])$ is called the cumulative *distribution function*: it is a nondecreasing function such that $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to+\infty} F(x) = 1$, and F is right continuous. If F is absolutely continuous, its derivative $f(x)$ is called a *density function*, which is a Lebesgue integrable nonnegative function whose integral over the real line is equal to one. Loosely speaking, if $F(x)$ is the probability that the random variable X takes values less than or equal to x, the quantity $f(x)\mathrm{d}x$ can be thought of as the probability that the random variable takes values in the infinitesimal interval $[x, x + \mathrm{d}x)$. If the random variable takes only a countable set of values, then it is said to be discrete and its density at point x is defined as $P(X = x)$. In the continuous case, $P(X = x) = 0$ always.

Independence

Two random variables X and Y are *independent* if

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

for any two sets A and B in $\mathbb{R}$.

1.1.1 Mean, variance, and moments

The mean (or expected value) of a continuous random variable X with distribution function F is defined as

$$\mathbb{E}X = \int_\Omega X(\omega)\mathrm{d}P(\omega) = \int_{\mathbb{R}} x\mathrm{d}F(x)$$

provided that the integral is finite. If X has a density, then $\mathbb{E}X = \int_{\mathbb{R}} xf(x)\mathrm{d}x$ and the integral is the standard Riemann integral; otherwise integrals in $\mathrm{d}P$ or $\mathrm{d}F$ should be thought of as integrals in the abstract sense. If Ω is countable, the expected value is defined as

$$\mathbb{E}X = \sum_{\omega\in\Omega} X(\omega)P(\omega)$$

or, equivalently, when X is a discrete random variable, the expected value reduces to $\mathbb{E}X = \sum_{x\in I} xP(X = x)$, where I is the set of possible values of X. The variance is defined as

$$\operatorname{Var} X = \mathbb{E}\,(X - \mathbb{E}X)^2 = \int_\Omega (X(\omega) - \mathbb{E}X)^2 \mathrm{d}P(\omega)\,,$$

and the kth moment is defined as

$$\mathbb{E}X^k = \int_\Omega X^k(\omega)\mathrm{d}P(\omega)\,.$$

In general, for any measurable function $g(\cdot)$, $\mathbb{E}g(X)$ is defined as

$$\mathbb{E}g(X) = \int_\Omega g(X(\omega))\mathrm{d}P(\omega)\,,$$

provided that the integral is finite.

Types of convergence

Let $\{F_n\}_{n\in\mathbb{N}}$ be a sequence of distribution functions for the sequence of random variables $\{X_n\}_{n\in\mathbb{N}}$. Assume that

$$\lim_{n\to\infty} F_n(x) = F(x)$$

for all $x \in \mathbb{R}$ such that $F(\cdot)$ is continuous in x, where F is the distribution function of some random variable X. Then, the sequence X_n is said to *converge in distribution* to the random variable X, and this is denoted by $X_n \xrightarrow{d} X$. This only means that the distributions F_n of the random variables converge to another distribution F, but nothing is said about the random variables itself. So this convergence is only about the probabilistic behavior of the random variables on some intervals $(-\infty, x]$, $x \in \mathbb{R}$.

A sequence of random variables X_n is said to *converge in probability* to a random variable X if, for any $\epsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0.$$

This is denoted by $X_n \xrightarrow{p} X$ and it is a pointwise convergence of the probabilities. This convergence implies the convergence in distribution. Sometimes we use the notation

$$p - \lim_{n\to\infty} |X_n - X| \; = 0$$

for the convergence in probability. A stronger type of convergence is defined as the probability of the limit in the sense $P(\lim_{n\to\infty} X_n = X) = 1$ or, more precisely,

$$P(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1.$$

When this happens, X_n is said to converge to X *almost surely* and is denoted by $X_n \xrightarrow{a.s.} X$. Almost sure convergence implies convergence in probability.

A sequence of random variables X_n is said to *converge in the rth mean* to a random variable X if

$$\lim_{n\to\infty} \mathbb{E}|X_n - X|^r = 0, r \geq 1.$$

The convergence in the r-th mean implies the convergence in probability thanks to Chebyshev's inequality, and if X_n converges to X in the rth mean, then it also converges in the sth mean for all $r > s \geq 1$. *Mean square convergence* is a particular case of interest and corresponds to the case $r = 2$. This type of convergence will be used in Section 1.9 to define the Itô integral.

1.1.2 Change of measure and Radon-Nikodým derivative

In some situations, for example in mathematical finance, it is necessary to reassign the probabilities to the events in Ω, switching from a measure P to another one $\tilde{P}$. This is done with the help of a random variable, say Z, which reweights the elements in Ω. This change of measure should be done set-by-set instead of ω-by-ω (see, e.g., [209]) as

$$\tilde{P}(A) = \int_A Z(\omega) \mathrm{d}P(\omega)\,, \tag{1.1}$$

where Z is assumed to be almost surely nonnegative and such that $\mathbb{E}Z = 1$. The new $\tilde{P}$ is then a true probability measure and, for any nonnegative random variable X, the equality

$$\tilde{\mathbb{E}}X = \mathbb{E}(XZ)$$

holds, where $\tilde{\mathbb{E}}X = \int_\Omega X(\omega)\mathrm{d}\tilde{P}(\omega)$. Two measures P and $\tilde{P}$ are said to be *equivalent* if they assign probability 0 to the same sets. The previous change of measure from P to $\tilde{P}$ trivially guarantee that the two measures are equivalent when Z is strictly positive. Another way to read the change of measure in (1.1) is to say that Z is the *Radon-Nikodým derivative* of $\tilde{P}$ with respect to P. Indeed, a formal differentiation of (1.1) allows us to write

$$Z = \frac{\mathrm{d}\tilde{P}}{\mathrm{d}P}\,. \tag{1.2}$$

Fact 1.1 (Theorem 1.6.7 [209]) *Let P and $\tilde{P}$ be two equivalent measures on $(\Omega, \mathcal{A})$. Then, there exists a random variable Z, almost surely positive, such that $\mathbb{E}Z = 1$ and*

$$\tilde{P}(A) = \int_A Z(\omega)\mathrm{d}P(\omega)$$

for every $A \in \mathcal{A}$.

The Radon-Nikodým derivative is an essential requirement in statistics because Z plays the role of the likelihood ratio in the inference for diffusion processes.

1.2 Random number generation

Every book on simulation points the attention of the reader to the quality of random number generators. This is of course one central point in simulation studies. R developers and R users are in fact quite careful in the implementations and use of random number generators. We will not go into details, but we just warn the reader about the possibilities available in R and what is used in the examples in this book.

The random number generator can be specified in R via the `RNGkind` function. The default generator of uniform pseudo random numbers is the *Mersenne-Twister* and is the one used throughout the book. Other methods available as of this writing are *Wichmann-Hill*, *Marsaglia-Multicarry*, *Super-Duper*, and two versions of *Knuth-TAOCP* random number generators. The user can implement and provide his own method as well. Specifically, for the normal random number generators, available methods are *Kinderman-Ramage*, *Buggy Kinderman-Ramage*, *Ahrens-Dieter*, *Box-Muller*, and the default *Inversion* method, as explained in [229]. For this case as well, the user can provide her own algorithm. For other than normal variates, R implements quite advanced pseudo random number generators. For each of these, the reader has to look at the manual page of the corresponding `r*` functions (e.g., `rgamma`, `rt`, `rbeta`, etc.).

For reproducibility of all the numerical results in the book we chose to use a fixed initialization seed before any listing of R code. We use everywhere the function `set.seed(123)`, and the reader should do the same if she wants to obtain the same results.

1.3 The Monte Carlo method

Suppose we are given a random variable X and are interested in the evaluation of $\mathbb{E}g(X)$ where $g(\cdot)$ is some known function. If we are able to draw n pseudo random numbers $x_1, \ldots, x_n$ from the distribution of X, then we can think about approximating $\mathbb{E}g(X)$ with the sample mean of the $g(x_i)$,

$$\mathbb{E}g(X) \simeq \frac{1}{n}\sum_{i=1}^{n} g(x_i) = \bar{g}_n \,. \tag{1.3}$$

The expression (1.3) is not just symbolic but holds true in the sense of the law of large numbers whenever $\mathbb{E}|g(X)| < \infty$. Moreover, the central limit theorem guarantees that

$$\bar{g}_n \xrightarrow{d} N\left(\mathbb{E}g(X), \frac{1}{n}\mathrm{Var}(g(X))\right),$$

where $N(m, s^2)$ denotes the distribution of the Gaussian random variable with expected value m and variance s^2. In the end, the number we estimate with

simulations will have a deviation from the true expected value $\mathbb{E}g(X)$ of order $1/\sqrt{n}$. Given that $P(|Z| < 1.96) \simeq 0.95$, $Z \sim N(0,1)$, one can construct an interval for the estimate $\bar{g}_n$ of the form

$$\left(\mathbb{E}g(X) - 1.96\frac{\sigma}{\sqrt{n}}, \mathbb{E}g(X) + 1.96\frac{\sigma}{\sqrt{n}}\right),$$

with $\sigma = \sqrt{\mathrm{Var}g(X)}$, which is interpreted that the Monte Carlo estimate of $\mathbb{E}g(X)$ above is included in the interval above 95% of the time. The confidence interval depends on $\mathrm{Var}g(X)$, and usually this quantity has to be estimated through the sample as well. Indeed, one can estimate it as the sample variance of Monte Carlo replications as

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(g(x_i) - \bar{g}_n\right)^2$$

and use the following 95% level Monte Carlo confidence interval[1] for $\mathbb{E}g(X)$:

$$\left(\bar{g}_n - 1.96\frac{\hat{\sigma}}{\sqrt{n}}, \bar{g}_n + 1.96\frac{\hat{\sigma}}{\sqrt{n}}\right).$$

The quantity $\hat{\sigma}/\sqrt{n}$ is called the *standard error*. The standard error is itself a random quantity and thus subject to variability; hence one should interpret this value as a "qualitative" measure of accuracy.

One more remark is that the rate of convergence $\sqrt{n}$ is not particularly fast but at least is independent of the smoothness of $g(\cdot)$. Moreover, if we need to increase the quality of our approximation, we just need to draw additional samples[2] instead of rerunning the whole simulation.

About Monte Carlo intervals length

In some cases, Monte Carlo intervals are not very informative if the variance of $Y = g(X)$ is too large. The next example, taken from [156], is one such case. Let $Y = g(X) = e^{\beta X}$ with $X \sim N(0,1)$, and assume we are interested in $\mathbb{E}g(X)$ with $\beta = 5$. The analytical value can be calculated as $e^{\beta^2/2} = 268337.3$, and the true standard deviation $\sigma = \sqrt{e^{2\beta^2} - e^{\beta^2}} = 72004899337$, quite a big number with respect to the mean of Y. Suppose we want to estimate $\mathbb{E}Y$ via the Monte Carlo method using 100000 replications and construct 95% confidence intervals using the true standard deviation σ and the estimated standard error. The following R code does the job.

[1] Again, this means that the interval covers the true value 95% of the time.

[2] A warning note: Of course one should take care of the seed of the random number generator to avoid duplicated samples. If we have already run n replications and we want to add n' new samples, we cannot simply rerun the algorithm for a length of n' with the same original seed because in this case we are just replicating the first n' samples among the n original ones, hence inducing bias without increasing accuracy.

```
> # ex1.01.R
> set.seed(123)
> n <- 1000000
> beta <-5
> x <- rnorm(n)
> y <- exp(beta*x)
>
> # true value of E(Y)
> exp(beta^2/2)
[1] 268337.3
> # MC estimation of E(Y)
> mc.mean <- mean(y)
> mc.mean
[1] 199659.2
> mc.sd <- sd(y)
> true.sd <- sqrt(exp(2*beta^2) - exp(beta^2))
>
> # MC conf. interval based on true sigma
> mc.mean - true.sd*1.96/sqrt(n)
[1] -140929943
> mc.mean + true.sd*1.96/sqrt(n)
[1] 141329262
>
> # MC conf. interval based on estimated sigma
> mc.mean - mc.sd*1.96/sqrt(n)
[1] 94515.51
> mc.mean + mc.sd*1.96/sqrt(n)
[1] 304802.9
>
> plot(1:n,cumsum(y)/(1:n),type="l",axes=F,xlab="n",
+   ylab=expression(hat(g)[n]),ylim=c(0,350000))
> axis(1,seq(0,n,length=5))
> axis(2,seq(0,350000,length=6))
> abline(h=268337.3)  # true value
> abline(h=mc.mean-mc.sd*1.96/sqrt(n),lty=3) # MC conf interval
> abline(h=mc.mean+mc.sd*1.96/sqrt(n),lty=3)
> abline(h=mc.mean,lty=2) # MC estimate
> box()
```

Running this code in R, we obtain the two intervals

$$(-140929943; 141329262) \quad \text{using } \sigma$$

and

$$(94515.51; 304802.9) \quad \text{using } \hat{\sigma}$$

with an estimated value of $\mathbb{E}g(X)$, $\hat{g}_n = 199659.2$. As one can see, the confidence interval based on σ contains the true value of $\mathbb{E}g(X)$ but is too large and hence meaningless. The confidence interval based on $\hat{\sigma}$ is smaller but still large. The first effect is due to the big variance of $g(X)$, while the second is due to the fact that the sample variance underestimates the true one ($\hat{\sigma} = 53644741$). The reason is that, in this particular case, the state of asymptotic normality after $n = 1000000$ replications is not yet reached (the reader is invited to look at this with a `plot(density(y))`) and thus the estimator $\hat{\sigma}$ is not necessarily an unbiased estimator of the true σ. Looking at Figure 1.1 one can expect that the Monte Carlo confidence interval for smaller values of n (the reader can try with $n = 100000$) does not even contain the true value.

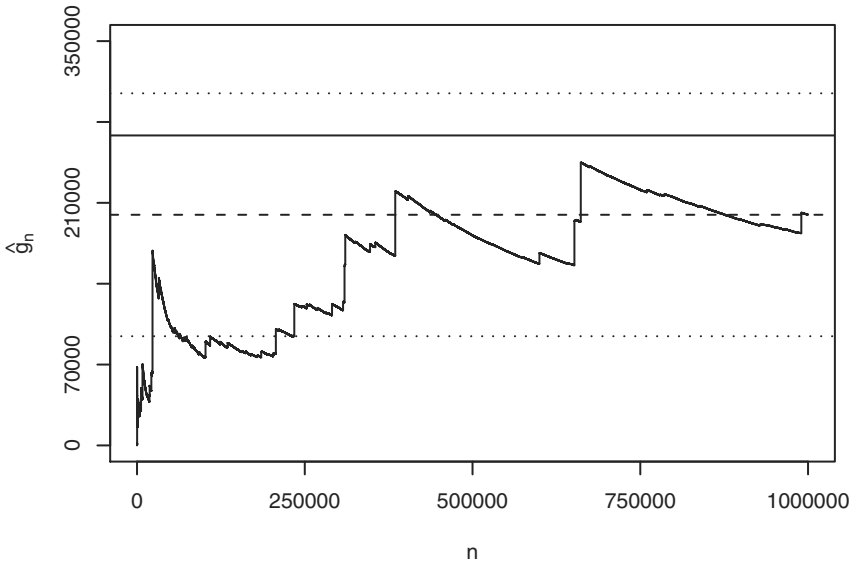

Fig. 1.1. The very slow convergence of the Monte Carlo estimate associated with a target true value with too high variability (see Section 1.3). The solid line is the target value, dotted lines are upper and lower limits of the Monte Carlo 95% confidence interval, and the dashed line is the estimated value $\hat{g}_n$.

1.4 Variance reduction techniques

The example in the last section gives evidence that in order to have less variability in Monte Carlo methods and hence use a smaller number of replications in simulations, one needs to try to reduce variability with some workaround. There are several methods of variance reduction for Monte Carlo estimators. We review here just the ones that can be applied in our context, but interesting reviews on methods for other classes of problems and processes can be found, for example, in [156] and [125]. Here we just show the basic ideas, while applications to stochastic differential equations are postponed to Section 2.15. We do not include the treatment of sequences with *low discrepancy*[3] because this is beyond the scope of this book.

[3] Discrepancy is a measure of goodness of fit for uniform random variates in high dimensions. Low-discrepancy sequences are such that numerical integration on this grid of points allows for a direct variance reduction. The reader can refer to the review paper [153].

1.4.1 Preferential sampling

The idea of this method is to express $\mathbb{E}g(X)$ in a different form in order to reduce its variance. Let $f(\cdot)$ be the density of X; thus

$$\mathbb{E}g(X) = \int_{\mathbb{R}} g(x)f(x)\mathrm{d}x.$$

Introduce now another strictly positive density $h(\cdot)$. Then,

$$\mathbb{E}g(X) = \int_{\mathbb{R}} \frac{g(x)f(x)}{h(x)}h(x)\mathrm{d}x$$

and

$$\mathbb{E}g(X) = \mathbb{E}\left(\frac{g(Y)f(Y)}{h(Y)}\right) = \mathbb{E}\tilde{g}(Y)\,,$$

with Y a random variable with density $h(\cdot)$, and denote $\tilde{g}(\cdot) = g(\cdot)f(\cdot)/h(\cdot)$. If we are able to determine an $h(\cdot)$ such that $\mathrm{Var}\tilde{g}(Y) < \mathrm{Var}g(X)$, then we have reached our goal. But let us calculate $\mathrm{Var}\tilde{g}(Y)$,

$$\mathrm{Var}\tilde{g}(Y) = \mathbb{E}\tilde{g}(Y)^2 - (\mathbb{E}\tilde{g}(Y))^2 = \int_{\mathbb{R}} \frac{g^2(x)f^2(x)}{h(x)}\mathrm{d}x - (\mathbb{E}g(X))^2\,.$$

If $g()$ is strictly positive, by choosing $h(x) = g(x)f(x)/\mathbb{E}g(X)$, we obtain $\mathrm{Var}\tilde{g}(Y) = 0$, which is nice only in theory because, of course, we don't know $\mathbb{E}g(X)$. But the expression of $h(x)$ suggests a way to obtain a useful approximation: just take $\tilde{h}(x) = |g(x)f(x)|$ (or something close to it), then normalize it by the value of its integral, and use

$$h(x) = \frac{\tilde{h}(x)}{\int_{\mathbb{R}} \tilde{h}(x)\mathrm{d}x}\,.$$

Of course this is simple to say and hard to solve in specific problems, as integration should be done analytically and not using the Monte Carlo technique again. Moreover, the choice of $h(\cdot)$ changes from case to case. We show an example, again taken from [156], which is quite interesting and is a standard application of the method in finance. Suppose we want to calculate $\mathbb{E}g(X)$ with $g(x) = \max(0, K - e^{\beta x}) = (K - e^{\beta x})_+$, K and β constants, and $X \sim N(0,1)$. This is the price of a *put option* in the Black and Scholes framework [36, 162], and the explicit solution, which is known, reads as

$$\mathbb{E}\left(K - e^{\beta X}\right)_+ = K\Phi\left(\frac{\log(K)}{\beta}\right) - e^{\frac{1}{2}\beta^2}\Phi\left(\frac{\log(K)}{\beta} - \beta\right),$$

where Φ is the cumulative distribution function of the standard Gaussian law; i.e., $\Phi(x) = P(Z < z)$ with $Z \sim N(0,1)$. The true value, in the case $K = \beta = 1$, is $\mathbb{E}g(X) = 0.2384217$. Let's see what happens in Monte Carlo simulations.

```
> # ex1.02.R
> set.seed(123)
> n <- 10000
> beta <-1
> K <- 1
> x <- rnorm(n)
> y <- sapply(x, function(x) max(0,K-exp(beta*x)))
>
> # the true value
> K*pnorm(log(K)/beta)-exp(beta^2/2)*pnorm(log(K)/beta-beta)
[1] 0.2384217
>
> t.test(y[1:100]) # first 100 simulations

    One Sample t-test

data:  y[1:100]
t = 7.701, df = 99, p-value = 1.043e-11
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.1526982 0.2586975
sample estimates:
mean of x
0.2056978

> t.test(y[1:1000]) # first 1000 simulations

    One Sample t-test

data:  y[1:1000]
t = 24.8772, df = 999, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.2131347 0.2496388
sample estimates:
mean of x
0.2313868

> t.test(y) # all simulation results

    One Sample t-test

data:  y
t = 80.3557, df = 9999, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.2326121 0.2442446
sample estimates:
mean of x
0.2384284
```

Results of simulations are reported in Table 1.1 (a). Note that on the algorithm we have used the function `sapply` instead of the (apparently) more natural (but wrong) line of code

```
y <- max(0, K-exp(beta*x))
```

as this will only return one value, actually the maximum value among 0 and all the $y_i = K - e^{\beta * x_i}$. This is one place where, vector-wise functions need to be used in the correct order. Note also the use of `t.test`, which actually performs both estimation and construction of the confidence intervals for $Y = \mathbb{E}g(X)$.

We now try to rewrite $\mathbb{E}g(X)$ as $\mathbb{E}g'(Y)$ (where g' is a function different from g) in order to reduce its variance. Indeed, $\mathbb{E}g(X)$ can be rewritten as

Table 1.1. Evaluation of the price of a put option with the Monte Carlo method. The true value is 0.2384217, with (b) and without (a) applying the variance reduction technique of Section 1.4.1.

n	$\hat{g}_n$	95% conf. interval
100	0.206	(0.153 ; 0.259)
1000	0.231	(0.213 ; 0.250)
10000	0.238	(0.232 ; 0.244)

(a)

n	$\hat{g}_n$	95% conf. interval
100	0.234	(0.222 ; 0.245)
1000	0.236	(0.233 ; 0.240)
10000	0.238	(0.237 ; 0.239)

(b)

$$\int_{\mathbb{R}} \frac{(1-e^{\beta x})_+}{\beta|x|}\beta|x|\frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}}\mathrm{d}x,$$

setting $K=1$ and noticing that $e^x - 1 \simeq x$ for x close to 0. By the change of variable $x=\sqrt{y}$ for $x>0$ and $x=-\sqrt{y}$ for $x<0$, the integral above can be rewritten as

$$\int_0^\infty \frac{\left(1-e^{\beta\sqrt{y}}\right)_+ + \left(1-e^{-\beta\sqrt{y}}\right)_+}{\sqrt{2\pi}\sqrt{y}}\frac{e^{-\frac{1}{2}y}}{2}\mathrm{d}y,$$

from which we remark that $f(y)=\lambda e^{-\lambda y}$, with $\lambda=\frac{1}{2}$, is the density of the exponential distribution. Therefore,

$$\mathbb{E}g(X) = \mathbb{E}\left(\frac{\left(1-e^{\beta\sqrt{Y}}\right)_+ + \left(1-e^{-\beta\sqrt{Y}}\right)_+}{\sqrt{2\pi}\sqrt{Y}}\right)$$

can be evaluated as the expected value of a function of the exponential random variable Y. The following algorithm executes the calculation, and results are reported in Table 1.1 (b), from which the reduction in variance is quite evident.

```
> # ex1.03.R
> set.seed(123)
> n <- 10000
> beta <-1
> K <- 1
>
> x <- rexp(n,rate=0.5)
> h <- function(x) (max(0,1-exp(beta*sqrt(x))) +
+   max(0,1-exp(-beta*sqrt(x))))/sqrt(2*pi*x)
> y <- sapply(x, h)
>
> # the true value
> K*pnorm(log(K)/beta)-exp(beta^2/2)*pnorm(log(K)/beta-beta)
[1] 0.2384217
>
> t.test(y[1:100]) # first 100 simulations
> t.test(y[1:1000]) # first 1000 simulations
> t.test(y) # all simulation results
```

1.4.2 Control variables

The very simple case of variance reduction via *control variables* is as follows. Suppose that we want to calculate $\mathbb{E}g(X)$. If we can rewrite it in the form

$$\mathbb{E}g(X) = \mathbb{E}(g(X) - h(X)) + \mathbb{E}h(X)\,,$$

where $\mathbb{E}h(X)$ can be calculated explicitly and $g(X) - h(X)$ has variance less than $g(X)$, then by estimating $\mathbb{E}(g(X) - h(X))$ via the Monte Carlo method, we obtain a reduction in variance.

Call-put parity example

Continuing with the example of the previous section, consider the price of a *call option*

$$c(X) = \mathbb{E}\left(e^{\beta X} - K\right)_+ .$$

It is easy to show that $c(X) - p(X) = e^{\frac{1}{2}\beta^2} - K$, where p is the price of the put option. Hence we can write $c(X) = p(X) + e^{\frac{1}{2}\beta^2} - K$. It is also known (see, e.g., [154]) that the variance of $p(X)$ is less than the variance of $c(X)$. Thus we obtained an estimator of $c(X)$ with reduced bias. The exact formula for $c(X)$ is also known and reads as

$$\mathbb{E}\left(e^{\beta X} - K\right)_+ = e^{\frac{1}{2}\beta^2}\Phi\left(\beta - \frac{\log(K)}{\beta}\right) - K\Phi\left(-\frac{\log(K)}{\beta}\right) .$$

The following R code shows this empirically, and the results are reported in Table 1.2.

```
> # ex1.04.R
> set.seed(123)
> n <- 10000
> beta <-1
> K <- 1
>
> x <- rnorm(n)
> y <- sapply(x, function(x) max(0,exp(beta*x)-K))
>
> # the true value
> exp(beta^2/2)*pnorm(beta-log(K)/beta)-K*pnorm(-log(K)/beta)
>
> t.test(y[1:100]) # first 100 simulations
> t.test(y[1:1000]) # first 1000 simulations
> t.test(y) # all simulation results
>
> set.seed(123)
> x <- rexp(n,rate=0.5)
> h <- function(x) (max(0,1-exp(beta*sqrt(x))) +
+   max(0,1-exp(-beta*sqrt(x))))/sqrt(2*pi*x)
> y <- sapply(x, h)
>
> # variance reduction
> # CALL = PUT + e^{0.5*beta^2} - K
> z <- y +exp(0.5*beta^2) - K
>
> t.test(z[1:100]) # first 100 simulations
> t.test(z[1:1000]) # first 1000 simulations
> t.test(z) # all simulation results
```

Table 1.2. Evaluation of the price of a call option with the Monte Carlo method. The true value is 0.887143 with (b) and without (a) applying the variance reduction technique of Section 1.4.2.

n	$\hat{g}_n$	95% conf. interval
100	0.858	(0.542 ; 1.174)
1000	0.903	(0.780 ; 1.026)
10000	0.885	(0.844 ; 0.925)

(a)

n	$\hat{g}_n$	95% conf. interval
100	0.882	(0.871 ; 0.894)
1000	0.885	(0.881 ; 0.889)
10000	0.887	(0.886 ; 0.888)

(b)

Table 1.3. Evaluation of the price of a put option with the Monte Carlo method. The true value is 0.2384217 with (b) and without (a) applying the variance reduction technique of Section 1.4.3.

n	$\hat{g}_n$	95% conf. interval
100	0.206	(0.153 ; 0.259)
1000	0.231	(0.213 ; 0.250)
10000	0.238	(0.232 ; 0.244)

(a)

n	$\hat{g}_n$	95% conf. interval
100	0.226	(0.202 ; 0.250)
1000	0.235	(0.226 ; 0.242)
10000	0.238	(0.235 ; 0.240)

(b)

1.4.3 Antithetic sampling

The idea of antithetic sampling can be applied when it is possible to find transformations of X that leave its measure unchanged (for example, if X is Gaussian, then $-X$ is Gaussian as well). Suppose that we want to calculate

$$I = \int_0^1 g(x)\mathrm{d}x = \mathbb{E}g(X)\,,$$

with $X \sim U(0,1)$. The transformation $x \mapsto 1-x$ leaves the measure unchanged (i.e., $1 - X \sim U(0,1)$), and I can be rewritten as

$$I = \frac{1}{2}\int_0^1 (g(x)+g(1-x))\mathrm{d}x = \frac{1}{2}\mathbb{E}(g(X)+g(1-X)) = \frac{1}{2}\mathbb{E}(g(X)+g(h(X)))\,.$$

Therefore, we have a variance reduction if

$$\mathrm{Var}\left(\frac{1}{2}\left(g(X)+g(h(X))\right)\right) < \mathrm{Var}\left(\frac{1}{2}g(X)\right),$$

which is equivalent to saying that $\mathrm{Cov}(g(X), g(h(X))) < 0$. If $h(x)$ is a monotonic function of x (as in the example above), this is always the case. This way of proceeding has the effect of reducing the variance but also increasing the accuracy of the calculation of the mean.[4] Going back to the example of

[4] It does not correct higher-order moment estimation, though.

the calculation of the price of a put option, one should calculate it using X and $-X$ and then averaging as follows:

```
> # ex1.05.R
> set.seed(123)
> n <- 10000
> beta <-1
> K <- 1
> x <- rnorm(n)
> y1 <- sapply(x, function(x) max(0,K-exp(beta*x)))
> y2 <- sapply(-x, function(x) max(0,K-exp(beta*x)))
>
> y <- (y1+y2)/2
> # the true value
> K*pnorm(log(K)/beta)-exp(beta^2/2)*pnorm(log(K)/beta-beta)
>
> t.test(y[1:100]) # first 100 simulations
> t.test(y[1:1000]) # first 1000 simulations
> t.test(y) # all simulation results
```

The results are reported in Table 1.3. Notice that we have applied this method to the naive Monte Carlo estimator and not the one built on the exponential distribution Y, as in that case $-Y$ is no longer an exponential distribution.

1.5 Generalities of stochastic processes

Let $(\omega, \mathcal{A}, P)$ a probability space. A real valued *stochastic process* is a family of random variables $\{X_\gamma, \gamma \in \Gamma\}$ defined on $\Omega \times \Gamma$ taking values in $\mathbb{R}$. Thus, the random variables of the family (measurable for every $\gamma \in \Gamma$) are functions of the form

$$X(\gamma, \omega) : \Gamma \times \Omega \mapsto \mathbb{R}\,.$$

For $\Gamma = \mathbb{N}$, we have a *discrete-time* process, and for $\Gamma \subset \mathbb{R}$ we have a *continuous-time* process. We are mainly interested in continuous-time processes with $\Gamma = [0, \infty)$, and we always think of $[0, \infty)$ as the time axis. We will denote a continuous-time stochastic process as $X = \{X_t, t \geq 0\}$. Sometimes, to avoid multiple subscripts, we will also adopt the usual notation $X(t)$ to denote X_t. For a fixed value of ω, say $\bar{\omega}$, $\{X(t, \bar{\omega}), t \geq 0\}$ (respectively $\{X(n, \bar{\omega}), n \in \mathbb{N}\}$ for the discrete case) is called the *path* or *trajectory* of the process and represents one possible evolution of the process. For a fixed t, say $\bar{t}$, the set of values $\{X(\bar{t}, \omega), \omega \in \Omega\}$ (respectively $\{X(\bar{n}, \omega), \omega \in \Omega\}$) represents the set of possible states of the process at time $\bar{t}$ (respectively n).

1.5.1 Filtrations

Consider the probability space $(\Omega, \mathcal{A}, P)$. A *filtration* $\{\mathcal{F}_t, t \geq 0\}$ is an increasing family of sub-σ-algebras of $\mathcal{A}$ indexed by $t \geq 0$; i.e., for each $s, t \geq 0$ such that $s < t$, we have $\mathcal{F}_s \subset \mathcal{F}_t$ with $\mathcal{F}_0 = \{\Omega, \emptyset\}$. To each process $\{X(t), t \geq 0\}$ and for each t, we can associate a σ-algebra denoted by $\mathcal{F}_t = \sigma(X(s); 0 \leq s \leq t)$, which is the σ-algebra generated by the process X up to time t; i.e., the smallest σ-algebra of $\mathcal{A}$ that makes $X(s, \omega)$ measurable