TOMSK POLYTECHNIC UNIVERSITY

STATISTICS FOR STOCHASTIC PROCESSES

Recommended for publishing as a study aid
by the Editorial Board of Tomsk Polytechnic University

Draftsmen **M.E. Semenov, G.V. Fedorov**

Tomsk Polytechnic University Publishing House
2023

UDC 519.246(075.8)=111
BBC 22.171я73=432.1
S81

Statistics for stochastic processes : study aid / draftsmen M.E. Semenov, G.V. Fedorov ; Tomsk Polytechnic University. – Tomsk : TPU Publishing House, 2023. – 80 p.

S81

ISBN 978-5-4387-1151-3

The study aid discusses the basic concepts of stochastic processes. Numerical schemes for solving stochastic differential equations are presented, parametric and non-parametric approaches for identifying model parameters are briefly considered, as well as techniques for reducing variation in Monte Carlo simulations and criteria for choosing the best model. All sections of the study aid include practical problems for solving in the R/Python language.

The study aid is based on the course "Statistics for stochastic processes" and can be useful for students of direction of training 01.03.02, 01.04.02 "Applied Mathematics and Computer Science".

UDC 519.246(075.8)=111
BBC 22.171я73=432.1

Reviewers

Candidate of Physical and Mathematical Sciences,
Associate Professor of the Tomsk State University
Department of Mathematical Analysis
E.A. Pchelintsev

Doctor of Physical and Mathematical Sciences,
Professor of the Sirius University of Science and Technology
Ya.I. Belopolskaya

ISBN 978-5-4387-1151-3

Contents

Course Introduction **5**

1. Preliminaries **8**
1.1. Generalities of stochastic processes 8
1.1.1. Filtrations 8
1.1.2. Simple and quadratic variation of a process 9
1.1.3. Moments, covariance, and increments of stochastic processes 10
1.1.4. Conditional expectation 11
1.1.5. Martingales 13
1.2. Brownian motion 14
1.2.1. Simulation of the trajectory of the Brownian motion 15
1.2.2. Brownian motion as the limit of a random walk . . 15
1.2.3. Brownian motion as Karhunen-Loeve expansion . . 16
1.2.4. Geometric Brownian motion 16
1.2.5. Brownian bridge 18
1.3. Stochastic integrals and SDEs 18
1.4. Properties of the stochastic integral and Ito processes . . . 21
1.5. Diffusion processes 22
1.5.1. Ergodicity 23
1.5.2. Markovianity 25
1.6. Ito formula 26
1.7. Practical problems 26

2. Some parametric families of stochastic processes **30**
2.1. Ornstein-Uhlenbeck process 30
2.2. The Black-Scholes-Merton model 32
2.3. The Cox-Ingersoll-Ross (CIR) model 34
2.4. The modified CIR model 35
2.5. The Chan-Karolyi-Longstaff-Sanders family of models . . 36
2.6. The nonlinear mean reversion Aït-Sahalia model 36
2.7. Double-well potential 37
2.8. The Jacobi diffusion process 37
2.9. Ahn and Gao (inverse of Feller's square root) model . . . 37
2.10. Radial Ornstein-Uhlenbeck process 38
2.11. Pearson diffusions 38
2.12. Practical Problems 39

3. Numerical methods for SDEs **41**
3.1. Approximation methods 41
3.1.1. Euler-Maruyama approximation 41
3.1.2. Milstein scheme 42
3.1.3. Predictor-corrector method 43
3.1.4. Kloden-Platen-Schurz-Sorensen method 43
3.1.5. Second Milstein scheme 44
3.2. Drawing from the transition density 44
3.3. Practical Problems 45

4. Parametric Estimation **47**
4.1. Exact likelihood inference 51
4.2. Pseudo-likelihood methods 53
4.2.1. Euler method 53
4.2.2. Local linearization methods 54
4.2.3. Approximated likelihood method 56
4.3. Practical Problems 56

5. Non-parametric estimation **58**
5.1. Stationary density estimation 59
5.2. Local-time and stationary density estimators 60
5.3. Estimation of diffusion and drift coefficients 61
5.4. Practical Problems 64

6. Variance reduction techniques **65**
6.1. Preferential sampling 65
6.2. Control variables 67
6.3. Antithetic sampling 67
6.4. Importance sampling 67
6.5. Practical Problems 68

7. Model identification **70**
7.1. AIC type of information criteria for diffusion processes . . 71
7.2. Practical Problems 73

8. Compensation Problems **75**

References **77**

Course Introduction

The course "Statistics for Stochastic Processes" presents a set of practical skills for a future quant researcher. After completing the course the student will have an opportunity to specify and implement appropriate mathematical model in Python language.

The Course Structure includes:

1) short introduction and preliminaries;
2) parametric families of stochastic processes;
3) numerical methods for stochastic differential equations (SDEs);
4) parametric and non-parametric estimations;
5) model identification via Akaike's information criterion.

We assume that a reader is familiar with elements of theory of probability, mathematical statistics, stochastic differential equations, and also has a basic knowledge of the Black-Scholes model and the Monte Carlo approach.

In the first stage, a student perform simple statistical simulation and visualise the paths of stochastic processes. Secondly, a student build a mathematical model on synthetic data with known parameters. Finally, a student create mathematical models on real data and choice the best one. For example, for 2-3 diffusion processes apply the Akaike's information criterion and choice the best one.

The guide consists of 8 different sections. Each section contains basic definitions, equation, short examples and practical problems for solution.

Students work individually or cooperate in small groups and after each lecture solves a different practical problems in a class and (or) home. Each day requires the participants to learn the theory, design an appropriate algorithm for simulation, analyze obtained results to understand the used mathematical and programming principles better.

Discipline Implementation Technology

The "Statistics for Stochastic Processes" course is based on a Problem-Based Learning (PBL) course and consists of four main parts:

1) to get receive its own practical problem set and a guide to its calculation; before each lesson, student study the obtained theoretical material and input dataset or model parameters for synthetic data simulation;

2) at the beginning of the class, students first defence the solutions before starting to new theoretical material and practical problems;

3) directly solving and presenting the steps of solution the practical problems;

4) at the end of each lesson, students receive a new practical problem set, which they prepare for the next lesson, etc.

Examination

The credit consists of the work of students during module (intensive case – two weeks, ordinary case – 8 weeks). The total amount of points for this course is 100. A student reaching at least 55 point passes the course.

To get a positive grade for the course you need:

1) attendance at tutorial classes;

2) work in class is evaluated according to the following criteria:

- theoretical knowledge of the topic, usage of additional literature;
- argumentation and logical explanation of the material;
- use of definitions and equations;
- attendance of the tutorials;

3) for each missed class, the student must solve compensation problems (Section 8);

4) performing practical problems, shared at github.com;

5) pass final exam with a grade.

Literature and Sources Recommendations

In order to obtain the basic notions of the stochastic calculus such that conditional expectations, predictable processes, martingales, stochastic integrals and Ito's formula one can read the Lecture notes in "Stochastic Modelling for the Financial Markets" [13] and book "Stochastic differential equations: An Introduction with Applications" [12]. The differences between behaviors of deterministic and stochastic models and classification tree one can see in the paper [10].

The book "Simulation and Inference for stochastic differential equations with R Examples" [7] by Iacus presents the main theoretical material and practical examples in R language, one can use the R examples as pseudo-code for your Python Language implementations (Sections 1-5). For further reading on non-parametric estimations (Section 5) we recommend the papers [1, 2, 15].

The book "Monte Carlo Methods in Financial Engineering" [5] by Glasserman provides a gentle introduction to Monte Carlo Methods through algorithms, real tasks and examples. One can use this book for solving practical problems from Section 6.

For the practical problems (Section 7) one can download economical data at Federal Reserve Bank of St. Louis and use the paper [9] to check the results. Some details on model identification via Akaike's information criterion (Section 7) one can find in the paper [16].

For future research and development, one can use an open source Python library *pymle* which includes the generators and estimation procedures, for a wide variety of models. The library is freely available at: https://github.com/jkirkby3/pymle. Also one can use the *yuima* [4] and *Sim.DiffProc* [6] projects in R language.

We would also like to mention books on the topic *stochastic differential equations* published in Russian [17–27].

We are grateful to Mr. Dmitry Redko (National Research Tomsk Polytechnic University, Russia) and Mr. Alexey Novohatsky (Sirius University of Science and Technology, Russia) for valuable comments and suggestions, solving practical problems, that improved the study aid and resulted in a better presentation of the material.

Supported by the Ministry of Science and Higher Education of the Russian Federation, (Agreement 075-10-2021-093, Project FMF-RND-2122).

1. Preliminaries

1.1. 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

$$X : \Gamma \times \Omega \to \mathbb{R}.$$

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

$$\{X(\bar{t}, \omega), \omega \in \Omega\} \quad (\text{respectively} \{X(\bar{n}, \omega), \omega \in \Omega\})$$

represents the set of possible *states* of the process at time $\bar{t}$ (respectively $\bar{n}$).

1.1.1. Filtrations

Consider the probability space $(\Omega, \mathcal{A}, P)$.

A *filtration* $\{F_t, t \geq 0\}$ 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 $F_s \subset F_t$ with $F_0 = \{\Omega, \emptyset\}$.

To each process $\{X(t), t \geq 0\}$ we can associate a σ-algebra denoted by

$$F_t = \sigma(X(s); 0 \leq s \leq t),$$

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 for every $0 \leq s \leq t$. This σ-algebra is the smallest set of subsets of Ω that makes it possible to assign probabilities to all the events related to the process X up to time t.

Given a stochastic process $\{X_t, t \geq 0\}$ and a filtration $\{F_t, t \geq 0\}$ (not necessarily the one generated by X), the process X is said to be *adapted* to $\{F_t, t \geq 0\}$ if for every $t \geq 0$, $X(t)$ is $\mathcal{F}_t$-*measurable*.

A stochastic process X defined on a filtered probability space

$$(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$$

is *progressively measurable* with respect to $(\mathcal{F}_t)_{t \geq 0}$, if the function

$$X(s, \omega) : [0, t] \times \Omega \to \mathbb{R}$$

is $\mathcal{B}([0, t]) \times \mathcal{F}_t$ measurable for every $t \geq 0$.

1.1.2. Simple and quadratic variation of a process

The notion of *total variation* or *first order variation* of a process $V(X)$ is linked to the differentiability of its paths.

Let $\Pi_n = \Pi_n([0, t]) = \{0 = t_0 < t_1 < \ldots < t_i < \ldots < t_n = t\}$ be any partition of the interval $[0, t]$ into n intervals and denote by

$$||\Pi_n|| = \max_{j=0,\ldots,n-1} (t_{j+1} - t_j)$$

the maximum step size of the partition Π_n, i. e. the mesh of the partition.

The *variation* of X is defined as

$$V_t(X) = p - \lim_{||\Pi_n|| \to 0} \sum_{k=0}^{n-1} |X(t_{k+1}) - X(t_k)|,$$

i. e. we have a convergence in probability. If X is differentiable, then

$$V_t(X) = \int_0^t |X_0'(u)| du.$$

If $V_t(X) < \infty$, then X is said to be of *bounded variation* on $[0, t]$. If this is true for all $t \geq 0$, then X is said to have bounded variation.

The *quadratic variation* $[X, X]_t$ at time t of a process X is defined as

$$[X, X]_t = p - \lim_{||\Pi_n|| \to 0} \sum_{k=0}^{n-1} |X(t_{k+1}) - X(t_k)|^2.$$

The limit exists for stochastic processes with continuous paths. In this case, the notation $\langle X, X \rangle_t$ is usually adopted.

Note that $V_t(X)$ and $[X, X]_t$ are stochastic processes as well.

1.1.3. Moments, covariance, and increments of stochastic processes

Remind, the base definitions from mathematical statistic. The mean (expected value) of a continuous random variable X with distribution function F is defined as

$$\mathbb{E}X = \int_\Omega X(\omega) dP(\omega) = \int_{\mathbb{R}} x dF(x)$$

provided that the integral is finite. If X has a density, then

$$\mathbb{E}X = \int_{\mathbb{R}} x f(x) dx$$

and the integral is the standard Riemann integral; otherwise integrals in dP or dF 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

$$\mathbb{E}X = \sum_{x \in I} x P(X = x),$$

where I is the set of possible values of X.

The variance is defined as

$$VarX = \mathbb{E}(X - \mathbb{E}X)^2 = \int_\omega (X(\omega) - \mathbb{E}X)^2 dP(\omega),$$

and the k-th moment is defined as

$$\mathbb{E}X^k = \int_\Omega X^k(\omega) dP(\omega).$$

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

$$\mathbb{E}g(X) = \int_{\omega} g(X(\omega))dP(\omega),$$

provided that the integral is finite.

The expected value and variance of a stochastic process are defined as

$$\mathbb{E}X_t = \int_{\Omega} X(t,\omega)dP(\omega), \quad t \in [0,T],$$

and

$$VarX_t = \mathbb{E}(X_t - \mathbb{E}X_t)^2, \quad t \in [0,T].$$

The k-th moment of X_t, $k \geq 1$, is defined, for all $t \in [0,T]$, as $\mathbb{E}X_t^k$. These quantities are well-defined when the corresponding integrals are finite.

The covariance function of the process for two time values s and t is defined as

$$Cov(X_s, X_t) = \mathbb{E}\{(X_s - \mathbb{E}X_s)(X_t - \mathbb{E}X_t)\}.$$

The quantity $X_t - X_s$ is called the *increment* of the process from s to t, $s < t$.

These quantities are useful in the description of stochastic processes that are usually introduced to model evolution subject to some stochastic shocks. There are different ways to introduce processes based on the characteristics one wants to model. A couple of the most commonly used approaches are the modeling of increments and (or) the choice of the covariance function.

1.1.4. Conditional expectation

The *conditional probability* of A given B is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) > 0.$$

In the same way, it is possible to introduce the *conditional distribution* of a random variable X with respect to the event B as

$$F_X(x|B) = \frac{P(\{X \leq x\} \cap B)}{P(B)}, \quad x \in \mathbb{R},$$

and the expectation with respect to this conditional distribution is naturally introduced as

$$\mathbb{E}[X|B] = \frac{\mathbb{E}(X\mathbb{1}_B)}{P(B)},$$

where $\mathbb{1}_B$ is the indicator function of the set B, which means

$$X\mathbb{1}_B(\omega) = 1,$$

if $\omega \in B$ and 0 otherwise. For discrete random variables, the conditional expectation takes the form

$$\mathbb{E}[X|B] = \sum_i x_i P(\{\omega : X(\omega) = x_i\} \cap B)P(B) = \sum_i x_i P(X = x_i|B).$$

For continuous random variables with density f_X, we have

$$\mathbb{E}[X|B] = \frac{1}{P(B)} \int_{\mathbb{R}} x\mathbb{1}_B(x) f_X(x)dx = \frac{1}{P(B)} \int_B x f_X(x)dx.$$

Consider now a discrete random variable Y that takes distinct values in the sets A_i, i. e., $A_i = A_i(\omega) = \{\omega : Y(\omega) = y_i\}$, $i = 1, 2, \ldots$, and assume that all $P(A_i)$ are positive. Let $\mathbb{E}|X| < \infty$. Then a new random variable Z can be defined as follows:

$$Z(\omega) = \mathbb{E}[X|Y](\omega) = \mathbb{E}[X|A_i(\omega)] = \mathbb{E}[X|Y(\omega) = y_i], \quad \omega \in A_i.$$

For each fixed $\omega \in A_i$ the conditional expectation $\mathbb{E}[X|Y]$ coincides with $\mathbb{E}[X|A_i]$, but, as a whole, it is a random variable itself because it depends on the events generated by Y.

If instead of a single element A_i we consider a complete σ-algebra of events (for example, the one generated by the generic random variable Y), we arrive at the general definition of conditional expectation: let X be a random variable such that $\mathbb{E}|X| < \infty$.

A random variable Z is called the *conditional expectation* of X with respect to the σ-algebra $\mathcal{F}$ if:

- Z is F-measurable;
- Z is such that $\mathbb{E}(Z\mathbb{1}_A) = \mathbb{E}(X\mathbb{1}_A)$ for every $A \in \mathcal{F}$.

The conditional expectation is unique and will be denoted as

$$Z = \mathbb{E}[X|\mathcal{F}].$$

With this notation, the equivalence above can be written as

$$\mathbb{E}(\mathbb{E}[X|\mathcal{F}]\mathbb{1}_A) = \mathbb{E}(X\mathbb{1}_A), \quad \text{for every} \quad A \in \mathcal{F}.$$

As we noted, the conditional expectation is a random variable, and the equality is only true up to null-measure sets. Among the properties of the conditional expectation, we note only the following.

Let X, X_1, X_2 be random variables and a, b two constants. Then

$$\mathbb{E}[a \cdot X_1 + b \cdot X_2|\mathcal{F}] = a \cdot \mathbb{E}[X_1|\mathcal{F}] + b \cdot \mathbb{E}[X_2|\mathcal{F}],$$

$$\mathbb{E}[X|\mathcal{F}_0] = \mathbb{E}X,$$

if $\mathcal{F}_0 = \{\Omega, \emptyset\}$. Moreover, if Y is $\mathcal{F}$-measurable, then

$$\mathbb{E}[Y \cdot X|\mathcal{F}] = Y \cdot \mathbb{E}[X|\mathcal{F}],$$

and choosing $X = 1$, it follows that $E[Y|\mathcal{F}] = Y$. Finally, choosing $A = \Omega$, it follows that

$$\mathbb{E}\{\mathbb{E}[X|\mathcal{F}]\} = \mathbb{E}X.$$

If X is independent of $\mathcal{F}$, it follows that $\mathbb{E}[X|\mathcal{F}] = \mathbb{E}X$ and, in particular, if X and Y are independent, we have

$$\mathbb{E}[X|Y] = \mathbb{E}[X|\sigma(Y)] = \mathbb{E}X,$$

where $\sigma(Y)$ is the σ-algebra generated by the random variable Y.

1.1.5. Martingales

Given a probability space $(\Omega, \mathcal{F}, P)$ and a filtration $\{\mathcal{F}_t, t \geq 0\}$ on $\mathcal{F}$, a *martingale* is a stochastic process $\{X_t, t \geq 0\}$ such that $\mathbb{E}|X_t| < \infty$ for all $t \geq 0$, it is *adapted* to a filtration $\{\mathcal{F}_t, t \geq 0\}$, and for each $0 \leq s \leq t < \infty$, it holds true that

$$\mathbb{E}[X_t|\mathcal{F}_s] = X_s,$$

i. e., X_s is the best predictor of X_t given $\mathcal{F}_s$. If in the definition above the equality "$=$" is replaced by "$\geq$", the process is called *submartingale*, and if it is replaced by "$\leq$", it is called *supermartingale*.

From the properties of the expected value operator it follows that if X is a martingale, then

$$\mathbb{E}X_s = \mathbb{E}\{\mathbb{E}[X_t|\mathcal{F}_s]\} = \mathbb{E}X_t,$$

which means that martingales have a constant mean for all $t \geq 0$.

In continuous time finance certain stochastic processes become martingales when properly stopped but are not true martingales.

A stochastic process $X = \{X_t, t \geq 0\}$ is a *local martingale* if there exists an increasing sequence $(\tau_n)_{n\in\mathbb{N}}$ of stopping times, that may depend on X, such that

$$\lim_{n\to\infty} \tau_n \overset{a.s.}{=} \infty$$

and each stopped process

$$X_{\tau_n} = \{X_t^{\tau_n} = X_{\min(t,\tau_n)}, t \geq 0\}$$

is an martingale.

If X is a local martingale, then the value X_s does, in general, not equal the conditional expectation $\mathbb{E}(X_t|\mathcal{F}_s)$ for $0 \leq s \leq t \leq \infty$. A local martingale that is not a martingale is called a *strict local martingale*.

1.2. Brownian motion

The very basic ingredient of a model describing stochastic evolution is the so-called *Brownian motion* (BM) or *Wiener process*. There are several alternative ways to characterize and define the Wiener process

$$W = \{W(t), t \geq 0\},$$

and one is the following: it is a Gaussian process with continuous paths and with independent increments such that $W(0) = 0$ with probability 1, $\mathbb{E}W(t) = 0$, and $Var(W(t) - W(s)) = t - s$ for all $0 \leq s \leq t$.

In practice, what is relevant for our purposes is that

$$W(t) - W(s) \sim \mathcal{N}(0, t - s), \quad \text{for } 0 \leq s \leq t$$

and that on any two disjoint intervals, say

$$(t_1, t_2), (t_3, t_4)$$

with

$$t_1 \leq t_2 \leq t_3 \leq t_4,$$

the increments $W(t_2) - W(t_1)$ and $W(t_4) - W(t_3)$ are independent.

1.2.1. Simulation of the trajectory of the Brownian motion

Given a fixed time increment $\Delta t > 0$, one can easily simulate a trajectory of the Wiener process in the time interval $[0, T]$. Indeed, for $W_{\Delta t}$ it holds true that

$$W(\Delta t) = W(\Delta t) - W(0) \sim \mathcal{N}(0, \Delta t) \sim \sqrt{\Delta t} \cdot \mathcal{N}(0, 1),$$

and the same is also true for any other increment $W(t + \Delta t) - W(t)$:

$$W(t + \Delta t) - W(t) \sim \mathcal{N}(0, \Delta t) \sim \sqrt{\Delta t} \cdot \mathcal{N}(0, 1).$$

We can simulate one such path as follows. Divide the interval $[0, T]$ into a grid such as $0 = t_1 < t_2 < \ldots < t_{N-1} < t_N = T$ with $t_{i+1} - t_i = \Delta t$. Set $i = 1$ and $W(0) = W(t_1) = 0$ and iterate the following algorithm:

1) generate a (new) random number z from the standard Gaussian distribution;
2) $i = i + 1$;
3) set $W(t_i) = W(t_{i-1}) + z \cdot \sqrt{\Delta t}$;
4) if $i \leq N$, iterate from step 1.

This method of simulation is valid only on the points of the grid, but in between any two points t_i and t_{i+1} the trajectory is usually approximated by linear interpolation.

1.2.2. Brownian motion as the limit of a random walk

One characterization of the Brownian motion says that it can be seen as the limit of a random walk in the following sense. Given a sequence of independent and identically distributed (i.i.d.) random variables

$$X_1, X_2, \ldots, X_n,$$

taking only two values $+1$ and -1 with equal probability and considering the partial sum,

$$S_n = X_1 + X_2 + \ldots + X_n.$$

Then, as $n \to \infty$,

$$P\left(\frac{S_{[nt]}}{\sqrt{n}} < x\right) \to P(W(t) < x),$$

where $[x]$ is the integer part of the real number x. Note that this result is a refinement of the central limit theorem that, in our case, asserts that

$$S_n/\sqrt{n} \to \mathcal{N}(0, 1).$$

1.2.3. Brownian motion as Karhunen-Loeve expansion

Another characterization of the Wiener process quite useful in statistics is the Karhunen-Loeve expansion of W.

The Karhunen-Loeve expansion is a powerful tool that is nothing but an $L^2([0,T])$ expansion of random processes in terms of a sequence of independent random variables. This is particularly useful for continuous time processes that are a collection of uncountably many random variables (such as the Wiener process which is indeed a collection of uncountably many Gaussian variables).

The Karhunen-Loeve expansion is in fact a series of only countably many terms and is useful for representing a process on some fixed interval $[0,T]$. We recall that $L^2([0,T])$, or simply L^2, is the space of functions from $[0,T]$ to $\mathbb{R}$ defined as

$$L^2 = \{f : [0,T] \to \mathbb{R} : ||f||^2 < \infty\},$$

where

$$||f||^2 = \left(\int_0^T |f(t)|^2 dt\right)^{\frac{1}{2}}.$$

Let us denote by $X(t)$ the trajectory of random process $X(t,\omega)$ for a given ω. The Wiener process $W(t)$ has trajectories belonging to $L^2([0,T])$ for almost all ω's, and the Karhunen-Loeve expansion for it takes the form

$$W(t) = W(t,\omega) = \sum_{i=0}^{\infty} Z_i(\omega)\phi_i(t), \quad 0 \le t \le T,$$

with

$$\phi_i(t) = \frac{2\sqrt{2T}}{(2i+1)\pi} \sin\left(\frac{(2i+1)\pi t}{2T}\right).$$

The functions $\phi_i(t)$ form a basis of orthogonal functions in $L^2([0,T])$ and Z_i a sequence of i.i.d. Gaussian random variables.

1.2.4. Geometric Brownian motion

A process used quite often in finance to model the dynamics of some asset is the so-called *geometric Brownian motion*. This process has the property of having independent multiplicative increments and is defined

as a function of the standard Brownian motion

$$S(t) = x \exp\left(\left(r - \frac{\sigma^2}{2}\right) t - \sigma W(t)\right), \quad t > 0,$$

with $S(0) = x$, $x \in \mathbb{R}$ is the initial value $\sigma > 0$ and r, where σ interpreted as the volatility, and r – the interest rate (two constants).

An equivalent way of simulating a trajectory of the geometric Brownian motion is by simulating the increments of S. Indeed,

$$S(t+\Delta t) = S(t) \exp\left(\left(r - \frac{\sigma^2}{2}\right)(t+\Delta t - t) + \right. \tag{1}$$
$$\left. + \sigma(W(t+\Delta t) - W(t))\right),$$

which simplifies to

$$S(t+\Delta t) = S(t) \exp\left(\left(r - \frac{\sigma^2}{2}\right)\Delta t + \sigma\sqrt{\Delta t} Z\right), \quad Z \sim \mathcal{N}(0,1).$$

Formula (1), which we will derive formally later, is a particular case of the generalized geometric Brownian motion, which is a process starting from x at time s whose dynamic is

$$Z_{s,x}(t) = x \exp\left(\left(r - \frac{\sigma^2}{2}\right)(t-s) + \sigma(W(t) - W(s)\right), \quad t \geq s.$$

Of course, $Z_{0,S(0)}(t) = S(t)$. In the same manner, we can consider the translated Brownian motion. Given a Brownian motion $W(t)$, we define a new process

$$W_{0,x}(t) = x + W(t)$$

with a constant x. Then $W_{0,x}(t)$ is a Brownian motion starting from x instead of 0. If we further want this to happen at some fixed time t_0 instead of at time 0, we need to translate the process further by $W(t_0)$. Thus,

$$W_{t_0,x}(t) = x + W(t) - W(t_0), \quad t \geq t_0,$$

is a Brownian motion starting at x at time t_0.

More precisely, this is the process

$$W_{t_0,x} = \{W(t), t_0 \leq t \leq T | W(t_0) = x\}$$

and, of course, $W_{0,W(0)}(t) = W(t)$. By the properties of the Brownian motion, $W_{t_0,x}(t)$ is equal in distribution to $x + W(t - t_0)$, and one way to simulate it is to simulate a standard Brownian motion, add the constant x, and then translate the time.

1.2.5. Brownian bridge

A *Brownian bridge* is a Brownian motion starting at x at time t_0 and passing through some point y at time T, $T > t_0$. It is defined as

$$W_{t_0,x}^{T,y}(t) = x + W(t - t_0) - \frac{t - t_0}{T - t_0}(W(T - t_0) - y + x).$$

More precisely, this is the process

$$\{W(t), t_0 \leq t \leq T | W(t_0) = x, W(T) = y\}.$$

1.3. Stochastic integrals and SDEs

Stochastic integrals and in particular Ito integrals are naturally introduced to correctly define a SDE.

Let us suppose we have the quantity $S(t)$, $t \geq 0$, which represents the value of an asset at time t. Consider now the increment

$$\Delta S = S(t + \Delta t) - S(t)$$

of S in a small time interval $[t, t + \Delta t)$.

The returns of the asset for which S is the dynamics are defined as the ratio $\Delta S/S$.

We can model the returns as

$$\frac{\Delta S}{S} = \text{deterministic contribution} + \text{stochastic contribution}.$$

The deterministic contribution might be assumed to be linked to the interest rate of non-risky activities and thus proportional to time with some constant rate μ, thus

$$\text{deterministic contribution} = \mu \Delta t.$$

Note, that μ can be made a function of either t or $S(t)$.

The stochastic contribution is assumed to be related to the variation of some source of noise and to the natural variability of the market (the volatility).

We denote by

$$\Delta X = X(t + \Delta t) - X(t)$$

the increment of the noisy process (i. e., the shocks) and make it proportional to the market volatility σ:

$$\text{stochastic contribution} = \sigma \Delta X.$$

Note, σ can also be made a function of t and/or S.

The natural hypothesis is to assume Gaussian behavior of the noise (i. e., $\Delta X \sim N(0,1)$) which implies the assumption of X being the Wiener process if the shocks are, in addition, supposed to be independent. Finally, we have

$$\frac{\Delta S}{S} = \mu \Delta t + \sigma \Delta W.$$

Now, the evil temptation is to consider the difference equation above for infinitesimal time intervals (i. e., for $\Delta t \to 0$) in order to obtain a (stochastic) differential equation of the form

$$\frac{S'(t)}{S(t)} = \mu + \sigma W'(t), \quad \text{namely} \quad S'(t) = \mu S(t) + \sigma S(t) W'(t). \tag{2}$$

The equation (2), unfortunately, has no mathematical meaning because the Wiener process $W(t)$ has nowhere differentiable paths.

We can rewrite equation (2) in differential form as

$$dS(t) = \mu S(t) dt + \sigma S(t) dW(t). \tag{3}$$

To make sense of Equation (3), we switch to its integral form

$$S(t) = S(0) + \mu \int_0^t S(u) du + \sigma \int_0^t S(u) dW(u). \tag{4}$$

Equation (4) introduces the *stochastic integral*

$$I(X) = \int_0^T X(u) dW(u)$$

with respect to the Brownian motion.

The definition of $I(X)$ is quite easy for *simple* (i. e., piece-wise constant) processes X, but it requires more attention for generic processes. Even if we can not go into the details of the construction of the stochastic integral, we will outline the basic steps in order to understand what $I(X)$ really means and find a way to simulate it.

The quantity being integrated, also called the *integrand*. For example, in $\int f(x)dx$, $f(x)$ is the *integrand*.

Given a generic integrand $f : [0,T] \times \Omega \to \mathbb{R}$, $I(f)$ is defined as the limit of the sequence of the integrals $I(f^{(n)})$, where $f^{(n)}$, called *simple processes*, are defined as

$$f^{(n)}(t,\omega) = f(t_j,\omega), \quad t_j \leq t \leq t_{j+1},$$

with $t_j \in \Pi_n([0,1])$ and such that $\Pi_n \to 0$ as $n \to \infty$. It is easy to show that $f^{(n)}$ converges to f in quadratic mean. Then $I(f^{(n)})$ is defined as

$$\begin{aligned} I(f^{(n)}) &= \sum_{j=0}^{n-1} f^{(n)}(t_j)\{W(t_{j+1}) - W(t_j)\} = \\ &= \sum_{j=0}^{n-1} f(t_j)\{W(t_{j+1}) - W(t_j)\}. \end{aligned} \tag{5}$$

Equation (5) does not converge in the usual sense, as W does not have finite variation. On the contrary, if we consider the mean square convergence, the limit exists. Indeed, for every n, we have that

$$\mathbb{E}\{I(f^{(n)})\}^2 = \sum_{j=0}^{n} \mathbb{E}(f(t_j))^2(t_{j+1} - t_j),$$

from which it follows that I in square mean sense, the limit being unique.

From the crude construction depicted above, few important things emerge as essential in the definition of $I(f^{(n)}) \to I(f)$.

First of all, it is required that f be a process adapted to the natural filtration of the Wiener process; i. e., f is $\mathcal{F}_t$-measurable for every t. This is required in Equation (5) in order to have a well-defined process and is the reason why, in the Ito sums of Equation (5), the function is calculated at the beginning (Ito type) of the interval $[t_j, t_{j+1})$ instead of in the middle (Stratonovich type).

Moreover, the behavior of the integrand process needs to compensate for the weirdness of the path of the Brownian motion. This second fact implies the technical condition $\mathbb{E}\int_0^t X^2(u)du < \infty$.

1.4. Properties of the stochastic integral and Ito processes

Let $\{X(t), 0 \leq t \leq T\}$ be a stochastic process adapted to the filtration $\mathcal{F}_t$ generated by the Brownian motion and such that

$$\int_0^T \mathbb{E}(X(s)^2)ds < +\infty.$$

The stochastic integral of X is defined as

$$I_t(X) = \int_0^t X_s dW_s = \lim_{||\Pi_n|| \to 0} \sum_{i=0}^{n-1} X(t_i)(W(t_{i+1}) - W(t_i)), \qquad (6)$$

where the convergence is in the square mean sense and $t_i \in \Pi_n$.

Properties of the Ito integral.

1. Ito isometry. If X is Ito integrable, then

$$\mathbb{E}\left(\int_0^T X(s)dW(s)\right) = 0$$

and

$$Var\left(\int_0^T X(s)dW(s)\right) = \int_0^T \mathbb{E}X^2(t)dW(t)).$$

2. Linearity. If X and Y are two Ito integrable processes and a and b two constants, then

$$\int_0^T (aX(t) + bY(t))dW(t) = a\int_0^T X(t)dW(t) + b\int_0^T Y(t)dW(t).$$

3. $\int_0^t X(s)dW(s) = \int_0^T \mathbb{1}(s \leq t)X(s)dW(s)$.

4. The process $M(t) = M(s) + \int_s^t X(\tau)dW(\tau)$ is a martingale.

An Ito process $\{X_t, 0 \leq t \leq T\}$ is a stochastic process that can be written in the form

$$X_t = X_0 + \int_0^t g(s)ds + \int_0^t h(s)dW(s),$$

where $g(t, \omega)$ and $h(t, \omega)$ are two *adapted* and *progressively measurable* random functions such that

$$P\left(\int_0^T |g(t, \omega)|dt < \infty\right) = 1 \quad \text{and} \quad P\left(\int_0^T h(t, \omega)^2 dt < \infty\right) = 1.$$

1.5. Diffusion processes

The class of processes that is considered in this study aid is that of diffusion process solutions to SDEs of the form

$$dX(t) = b(t, X(t))dt + \sigma(t, X(t))dW(t) \tag{7}$$

with some initial condition $X(0)$. As usual, Equation (7) is interpreted in the Ito sense; i. e.,

$$X(t) = X(0) + \int_0^T b(u, X(u))du + \int_0^T \sigma(u, X(u))dW(u). \tag{8}$$

The initial condition $X(0)$ can be random or not.

Random initial condition $X(0)$

If random, say $X(0) = Z$, it should be independent of the σ-algebra generated by W and satisfy the condition $\mathbb{E}|Z|^2 < \infty$.

The two deterministic functions $b(\cdot, \cdot)$ and $\sigma^2(\cdot, \cdot)$ are called respectively the *drift* and the *diffusion* coefficients of the SDE (7). Later, even when not mentioned, they are supposed to be measurable and such that

$$P\left(\int_0^T \sup_{|x| \leq \mathbb{R}} (|b(t, x)| + \sigma^2(t, x))dt < \infty\right) = 1$$

for all $T, \mathbb{R} \in [0, \infty)$ because equation (8) is an Ito process.

Assumption 1.1 (Global Lipschitz). For all $x, y \in \mathbb{R}$ and $t \in [0, T]$, there exists a constant $K < +\infty$ such that

$$|b(t, x) - b(t, y)| + |\sigma(t, x) - \sigma(t, y)| < K|x - y|.$$

Assumption 1.2 (Linear growth). For all x, $y \in \mathbb{R}$ and $t \in [0, T]$, there exists a constant $C < +\infty$ such that

$$|b(t, x)| + |\sigma(t, x)| < C(1 + |x|).$$

The linear growth condition controls the behaviour of the solution so that X_t does not explode in a finite time.

Fact 1.2 (Existence and uniqueness). Under Assumptions 1.1 and 1.2, the SDE (7) has a unique, continuous, and adapted strong solution such that

$$\mathbb{E}\left(\int_0^T |X_t|^2 dt\right) < \infty.$$

We call such a process X a *diffusion* process. The result above states that the solution X is of *strong* type. This essentially implies the pathwise uniqueness of the result. It is also possible to obtain weak solutions under different assumptions.

In many cases in statistics, conditions for weak solutions are enough because they imply that any two *weak* solutions $X(1)$ and $X(2)$ are not necessarily pathwise identical, while their distributions are, and this is enough for likelihood inference. Of course, strong solutions, are also weak solutions but the contrary is not necessarily true.

Non-random initial condition $X(s)$

The major part of this study aid will focus on the homogeneous version (time) of the SDE with nonrandom initial condition, say

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \qquad X_s = x. \tag{9}$$

We say X_t is the solution (9), if X_t is $\mathcal{F}_t$-measurable and with probability 1

$$X_t = x + \int_s^t b(X_u)du + \int_s^t \sigma(X_u)dW_u$$

holds.

1.5.1. Ergodicity

Diffusion processes possess the Markov property and may or may not be ergodic. The *ergodic* property implies that, for any measurable function $h(\cdot)$, the following result holds with probability 1:

$$\frac{1}{T}\int_0^T h(X_t)dt \to \int_{-\infty}^{+\infty} h(x)\pi(x)dx = \mathbb{E}h(\xi), \qquad T \to \infty,$$

where $\pi(\cdot)$ is called the *invariant* or *stationary* density of the diffusion process and ξ is some random variable with $\pi(\cdot)$ as density.

Diffusion processes have the nice property that the stationary distribution, when it exists, can be expressed in terms of the *scale measure* and the *speed measure*, defined respectively as

$$s(x) = \exp\left(-2\int_{x_0}^x \frac{b(y)}{\sigma^2(y)}dy\right) \quad \text{and} \quad m(x) = \frac{1}{\sigma^2(x)s(x)}.$$

In particular, the density of the invariant distribution $\pi(\cdot)$ is proportional, up to a normalizing constant, to the speed measure $m(\cdot)$; i. e.,

$$\pi(x) = \frac{m(x)}{M},$$

where $M = \int m(x)dx$.

Assumption 1.3. Let (l, r), with $-\infty < l < r < +\infty$, be the state space of the diffusion process X solution to Equation (9), and assume that

$$\int_l^r m(x)dx < \infty.$$

Let x^* be an arbitrary point in the state space of X such that

$$\int_{x^*}^r s(x)dx = \int_{x^*}^l s(x)dx = \infty.$$

If one or both of the integrals above are finite, the corresponding boundary is assumed to be instantaneously reflecting.

Under Assumption 1.3, the process X is *ergodic* and has an invariant distribution function.

Stationarity refers to the *distributions* of the random variables.

Specifically, in a stationary process, all the random variables have the same distribution function, and more generally, for every positive integer n and n time instants $t_1, t_2, \ldots, t_n$, the joint distribution of the n random variables

$$X(t_1), X(t_2), \ldots, X(t_n)$$

is the same as the joint distribution of

$$X(t_1 + \tau), X(t_2 + \tau), \ldots, X(t_n + \tau).$$

That is, if we shift all time instants by τ, the statistical description of the process does not change at all: the process is stationary.

Ergodicity, on the other hand, does not look at statistical properties of the random variables but at the *sample paths*, i. e. what you observe physically.

Referring back to the random variables, recall that random variables are mappings from a sample space to the real numbers; each outcome is mapped onto a real number, and different random variables will typically map any given outcome to different numbers.

1.5.2. Markovianity

A random process X_t is said to be *Markovian* if when the present state of the process X_t is known, then any additional information on its past history is totally irrelevant for a prediction of its future evolution.

From the Markovian property of the diffusion, it is also possible to define the *transition density* from value x at time s to value y at time t by

$$p(t, y|s, x)$$

or, when convenient, as $p(t - s, y|x)$.

For parametric models, we will later use the notation

$$p(t, y|s, x; \theta) \quad \text{or} \quad p_\theta(t, y|s, x)$$

and

$$p(t - s, y|x; \theta) \quad \text{or} \quad p_\theta(t - s, y|x),$$

respectively.

The transition density satisfies the *Kolmogorov forward equation*

$$\frac{\partial}{\partial t}p(t, y|s, x) = -\frac{\partial}{\partial y}b(y)p(t, y|s, x)) + \frac{1}{2}\frac{\partial^2}{\partial y^2}(\sigma^2(y)p(t, y|s, x)) \qquad (10)$$

and *Kolmogorov backward equation*

$$-\frac{\partial}{\partial s}p(t, y|s, x) = b(x)\frac{\partial}{\partial x}p(t, y|s, x)) + \sigma^2(x)\frac{1}{2}\frac{\partial^2}{\partial y^2}p(t, y|s, x). \qquad (11)$$

Letting $t \to \infty$ in the Kolmogorov forward equation (10), it is possible to obtain

$$\frac{d^2}{dx^2}(\sigma^2(x)\pi(x)) = 2\frac{d}{dx}(b(x)\pi(x)), \qquad (12)$$

where $\pi(x)$ is the stationary density. Equation (12) establishes a relationship between the drift $b(\cdot)$, the diffusion coefficient $\sigma(\cdot)$, and the invariant density $\pi(\cdot)$. Hence, in principle, given either of the two, one can obtain the third. For example, by integrating (12), we obtain

$$b(x) = \frac{1}{2\pi(x)}\frac{d}{dx}(\sigma^2(x)\pi(x)),$$

and integrating again, one gets

$$\sigma^2(x) = \frac{2}{\pi(x)}\int_0^x b(u)\pi(u)du.$$

1.6. Ito formula

The Ito formula can be seen as the stochastic version of a Taylor expansion of $f(X)$ stopped at the second order, where X is a diffusion process.

Ito's lemma says if $f(t,x)$ is a twice differentiable function on both t and x, then

$$f(t,X_t) = f(0,X_0) + \\ + \int_0^t f_t(u,X_u)d_u + \int_0^t f_x(u,X_u)dX_u + \frac{1}{2}\int_0^t f_{xx}(u,X_u)(dX_u)^2,$$

where

$$\frac{\partial f(t,x)}{\partial t} = f_t(t,x), \quad \frac{\partial f(t,x)}{\partial x} = f_x(t,x), \quad \frac{\partial^2 f(t,x)}{\partial x^2} = f_{xx}(t,x),$$

or, in differential form,

$$df(t,X_t) = f_t(t,X_t)dt + f_x(t,X_t)dX_t + \frac{1}{2}f_{xx}(t,X_t)(dX_t)^2.$$

If X_t is the Brownian motion, this simplifies to the following

$$f(t,W_t) = f(0,0) + \int_0^t f_t(u,W_u) + \frac{1}{2}f_{xx}(u,W_u)du + \int_0^t f_x(u,W_u)dW_u$$

or, in differential form,

$$df(t,W_t) = f_t(t,W_t) + \frac{1}{2}f_{xx}(t,W_t)dt + f_x(t,W_t)dW_t.$$

All these facts and relationships will be useful for both the simulation algorithms and in the topic on parametric and non-parametric inference.

1.7. Practical problems

1. Using the algorithm (Section 1.2.1) simulate a path of the Wiener process. Number of end-points of the grid including T is $N = 100$, length of the interval $[0,T]$ in time units is 1, time increment is $\Delta = T/N$. Plot the path. The expected result is in Figure 1.

2. Using the random walk algorithm (Section 1.2.2) simulate three paths of the Wiener process as the limit of a random walk for different $n = 10, 100, 1000$. Plot the paths in one figure, add legends. The expected result is in Figure 2.

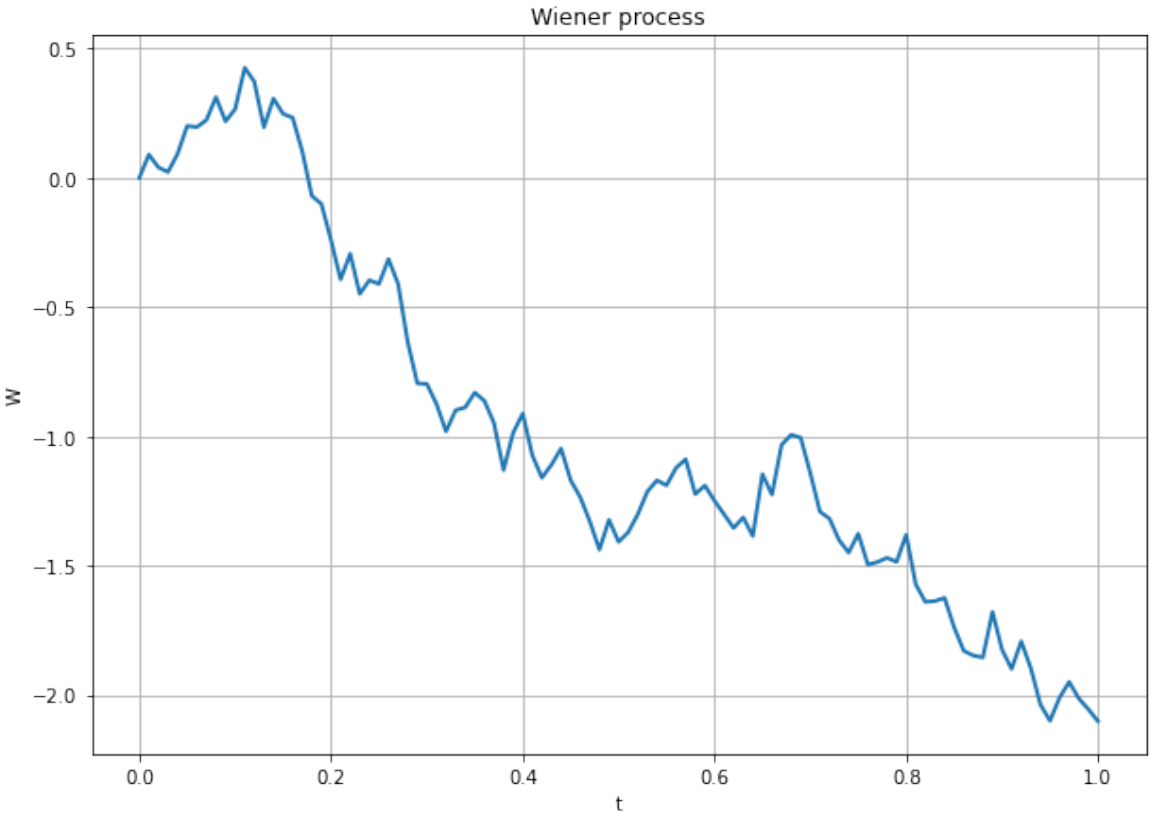

Figure 1. A simulated path of the Wiener process

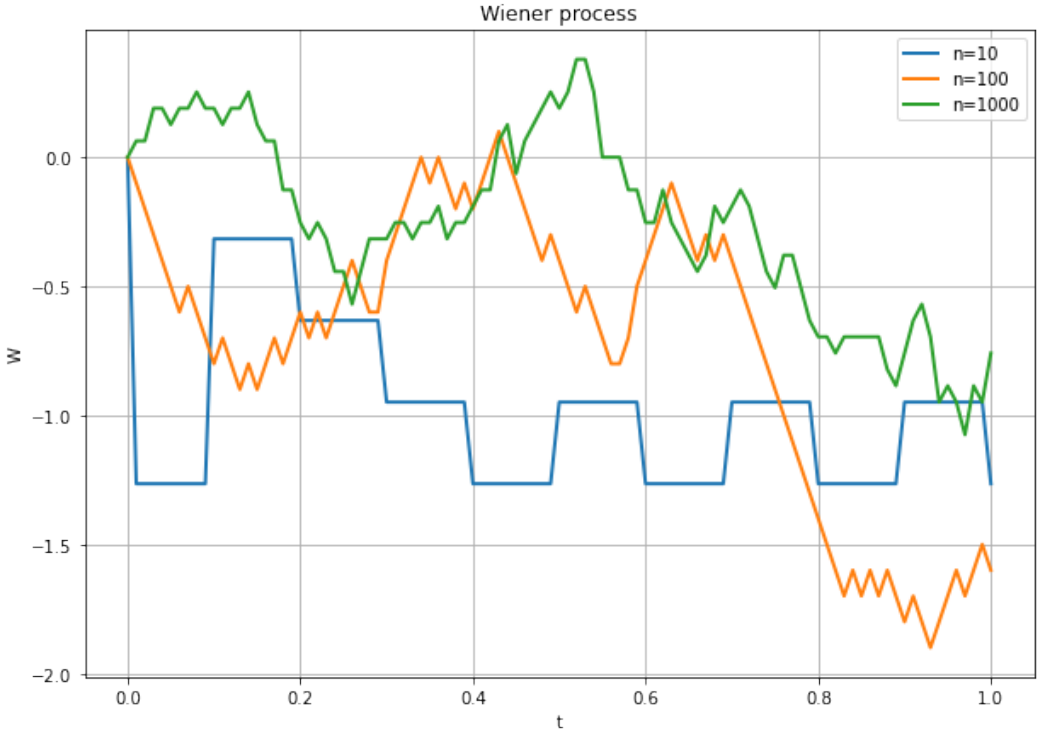

Figure 2. Path of the Wiener process as the limit of a random walk

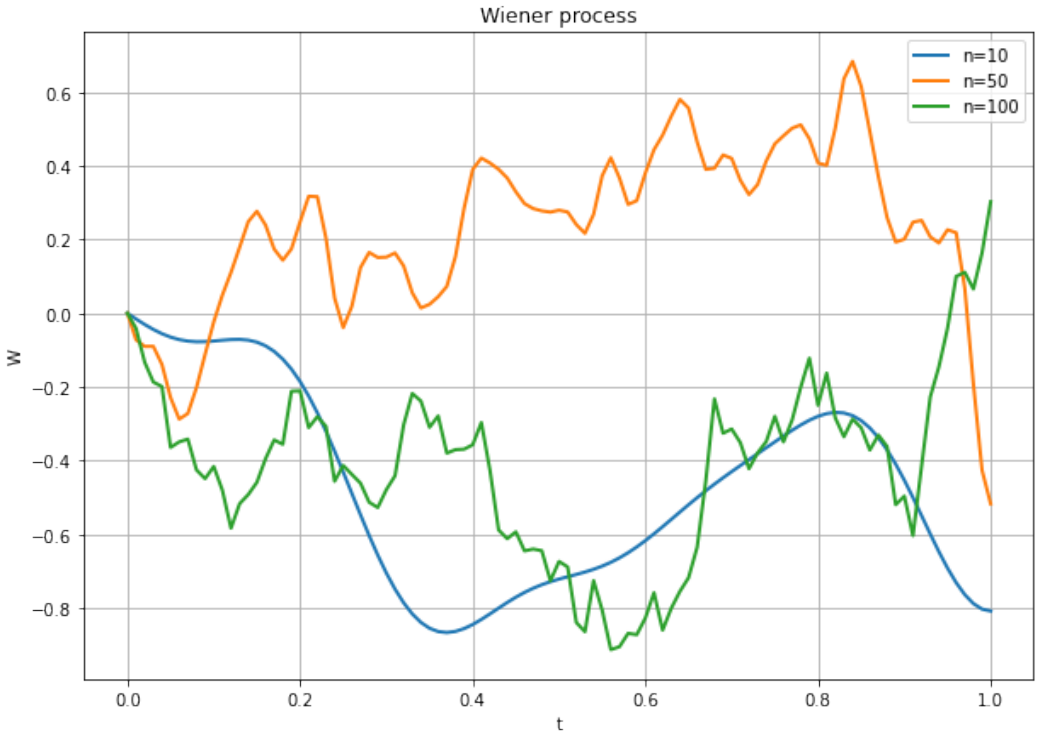

Figure 3. Karhunen-Loeve approximation of the path of the Wiener process with different terms in the expansion

3. Using the the Karhunen-Loeve expansion (Section 1.2.3) simulate three paths of the Wiener process with $n = 10, 50, 100$ terms. Plot the paths in one figure, add legends. The expected result is in Figure 3.

4. Plot a trajectory of the geometric Brownian motion (Section 1.2.4) obtained from the simulation of the path of the Wiener process, $r = 1$ (interest rate) and $\sigma = 0.5$ (volatility). The expected result is in Figure 4.

5. Plot a trajectory of the the Brownian bridge (Section 1.2.5) starting at x at time 0 and terminating its run at $y = -1$ at time T obtained from the simulation of the path of the Wiener process. The expected result is in Figure 5.

6. Let us consider a random walk $X = (X_n)_{n\geq 0}$:

$$X_n = \sum_{k=1}^{n} \xi_k, \quad \xi_k = \begin{cases} 1, & \text{with probability } \frac{1}{2}; \\ -1, & \text{with probability } \frac{1}{2}. \end{cases}$$

$$\mathbb{E}\xi_k = 0, \quad \operatorname{Var} \xi_k = 1.$$

Plot $(X_{t_i})_{t_i=\frac{i}{n}}$ so that t_i takes values in $[0, 1]$.

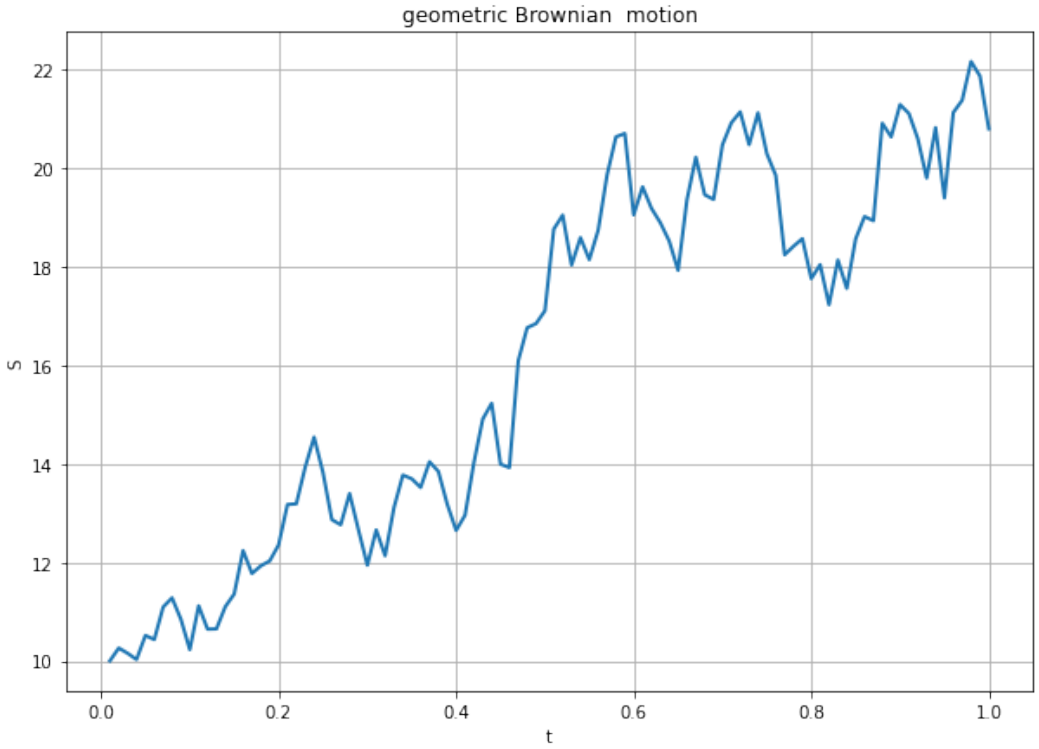

Figure 4. A trajectory of the geometric Brownian motion obtained from the simulation of the path of the Wiener process

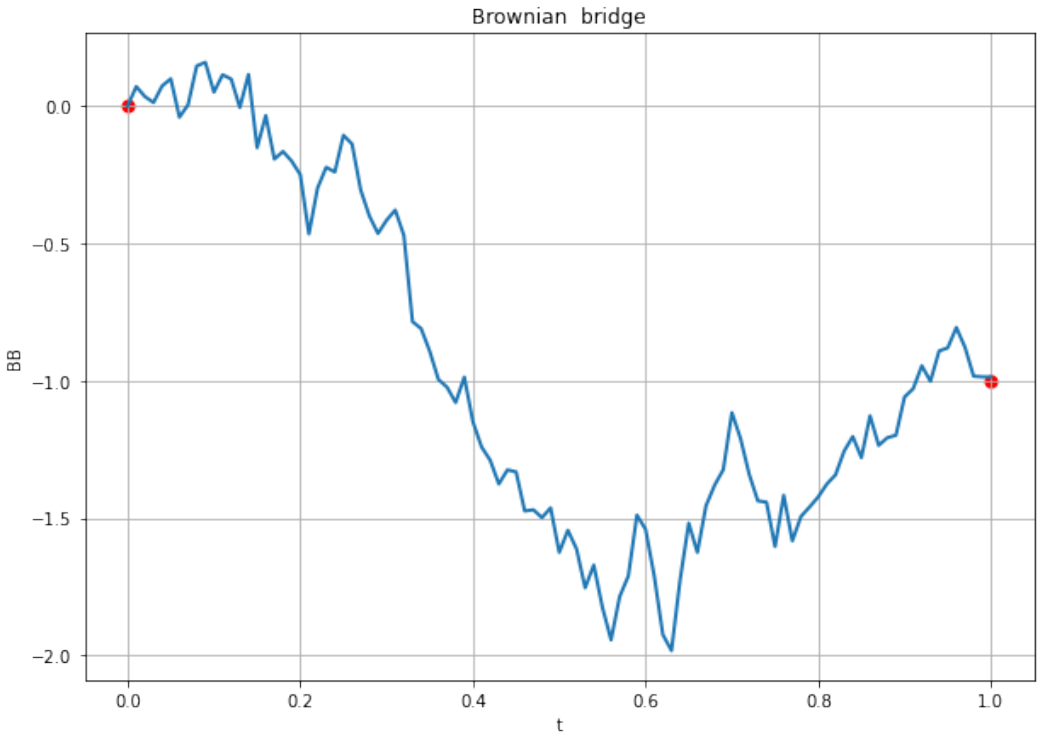

Figure 5. A simulated trajectory of the Brownian bridge starting at x at time $t = 0$ and terminating its run at $y = -1$ at time $T = 1$

2. Some parametric families of stochastic processes

We present some of the well-known and widely used stochastic process solutions to the general SDE

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t$$

with a quick review of their main properties.

When possible, we will describe each process in terms of its *first moments* and *covariance* function, in terms of the *stationary density* $\pi(x)$, and in terms of its *transition probability densities*

$$p(t-s, y|x; \theta) \quad \text{or} \quad p_\theta(t-s, y|x)$$

of X_t given some previous state of the process $X_s = x$.

In some cases, it will be simpler to express the *stationary density* $\pi(x)$ of a diffusion in terms of the *scale measure* $s(\cdot)$ and the *speed measure* $m(\cdot)$ of the diffusion.

2.1. Ornstein-Uhlenbeck process

The Ornstein-Uhlenbeck (OU) or Vasicek process is the unique solution to the following SDE

$$dX_t = (\theta_1 - \theta_2 X_t)dt + \theta_3 dW_t, \quad X_0 = x_0, \tag{13}$$

with $\theta_3 \in \mathbb{R}_+$ and $\theta_1, \theta_2 \in \mathbb{R}$.

The model $\theta_1 = 0$ was originally proposed by Ornstein and Uhlenbeck in the context of physics and then generalized by Vasicek to model interest rates.

For $\theta_2 > 0$, this is a *mean reverting* process, which means that the process tends to oscillate around some equilibrium state.

Another interesting property of the Ornstein-Uhlenbeck process is that, contrary to the Brownian motion, it is a process with finite variance for all $t \geq 0$.

More common in finance modeling parametrization of the Ornstein-Uhlenbeck process is

$$dX_t = \kappa(\alpha - X_t)dt + \sigma dW_t, \quad X_0 = x_0, \tag{14}$$

where σ is interpreted as the volatility, α is the long-run equilibrium value of the process, and κ is the speed of reversion.

As an application of the Ito lemma, we can show the explicit solution of Equation (13) by choosing $f(t,x) = xe^{\theta_2 t}$.

Indeed,

$$f_t(t,x) = \theta_2 f(t,x), \quad f_x(t,x) = e^{\theta_2 t}, \quad f_{xx}(t,x) = 0.$$

Therefore,

$$\begin{aligned} X_t e^{\theta_2 t} = f(t, X_t) = \\ = f(0, X_0) + \int_0^t \theta_2 X_u e^{\theta_2 u} du + \int_0^t e^{\theta_2 u} dX_u = \\ = x_0 + \int_0^t \theta_2 X_u e^{\theta_2 u} du + \int_0^t e^{\theta_2 u} \{(\theta_1 - \theta_2 X_u) du + \theta_3 dW_u\} = \\ = x_0 + \frac{\theta_1}{\theta_2}(e^{\theta_2 t} - 1) + \theta_3 \int_0^t e^{\theta_2 u} dW_u, \end{aligned}$$

from which we obtain

$$X_t = \frac{\theta_1}{\theta_2} + \left(x_0 - \frac{\theta_1}{\theta_2} \right) e^{-\theta_2 t} + \theta_3 \int_0^t e^{-\theta_2 (t-u)} dW_u \tag{15}$$

or, in the second parametrization,

$$X_t = \mu + (x_0 - \mu) e^{-\theta t} + \sigma \int_0^t e^{-\theta (t-u)} dW_u.$$

The invariant law

For $\theta_2 > 0$, the process X_t is ergodic, and its invariant law is Gaussian

$$X_t \sim \mathcal{N}\left(\frac{\theta_1}{\theta_2}, \frac{\theta_3^2}{2\theta_2} \right). \tag{16}$$

The covariance function is

$$Cov(X_t, X_s) = \frac{\theta_3^2}{2\theta_2} e^{-\theta_2 |t-s|}.$$

Sometimes it is convenient to rewrite the process as the scaled time-transformed Wiener process

$$X_t = \frac{\theta_1}{\theta_2} + \frac{\theta_3 e^{-2\theta_2 t}}{2\sqrt{\theta_2}} W(e^{2\theta_2 t}).$$

The conditional law

For any $t \geq 0$, the Ornstein-Uhlenbeck process has a Gaussian transition (or conditional) density

$$p_\theta(t, X_t|X_0 = x_0),$$

with mean and variance respectively

$$m(t,x) = \mathbb{E}_\theta(X_t|X_0 = x_0) = \frac{\theta_1}{\theta_2} + \left(x_0 - \frac{\theta_1}{\theta_2}\right) e^{-\theta_2 t}$$

and

$$v(t,x) = Var_\theta(X_t|X_0 = x_0) = \frac{\theta_3^2}{2\theta_2}(1 - e^{-2\theta_2 t}).$$

The conditional covariance function is

$$Cov(X_s, X_t|X_0 = x_0) = \frac{\theta_3^2}{2\theta_2} e^{-2\theta_2(s+t)} \left(e^{2\theta_2 \min(s,t)} - 1\right)$$

and its scaled time-transformed Wiener representation is

$$X_t = \frac{\theta_1}{\theta_2} + \left(x_0 - \frac{\theta_1}{\theta_2}\right) e^{-\theta_2 t} + \frac{\theta_3}{\sqrt{2\theta_2}} W\left(e^{2\theta_2 t} - 1\right) e^{-2\theta_2 t}.$$

Example of simulation of the stochastic integral. It can be seen that for $\theta_1 = 0$ the trajectory of X_t is essentially a negative exponential perturbed by the stochastic integral. One way of simulating trajectories of the Ornstein-Uhlenbeck process is indeed via the simulation of the stochastic integral.

The result for the Ornstein-Uhlenbeck process (14):

$$dX_t = -\theta X_t + \sigma dW_t$$

with $X(0) = 10$, $\theta = 5$, and $\sigma = 3.5$ is shown in Figure 6.

2.2. The Black-Scholes-Merton model

The process is the solution to the SDE

$$dX_t = \theta_1 X_t dt + \theta_2 X_t dW_t, \quad X_0 = x_0,$$

with $\theta_2 > 0$. The parameter θ_1 is interpreted as the constant interest rate and θ_2 as the volatility of risky activities. The explicit solution is

$$X_t = x_0 e^{(\theta_1 - \frac{\theta_2^2}{2})t + \theta_2 W_t}. \tag{17}$$

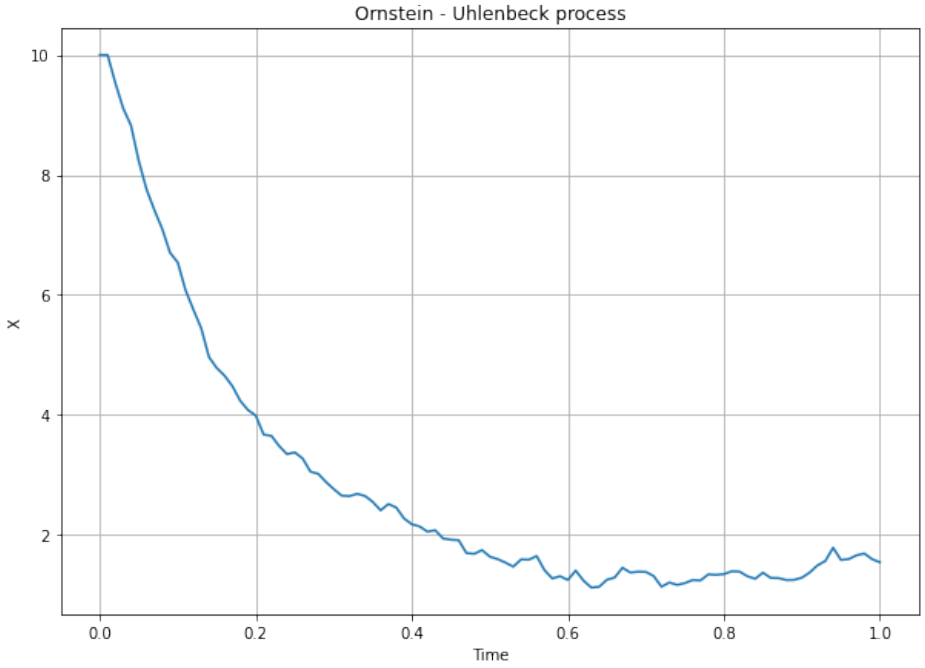

Figure 6. Simulated path of the Ornstein-Uhlenbeck process $dX_t = -\theta X_t + \sigma dW_t$ with $X(0) = 10$, $\theta = 5$, and $\sigma = 3.5$

The conditional law

The conditional density function is log-normal with the mean and variance of its logarithm transform (i. e., the log-mean and log-variance) given by

$$\mu = \log(x_0) + \left(\theta_1 - \frac{\theta_2^2}{2}\right) t, \qquad \sigma^2 = \theta_2^2 t,$$

with mean and variance

$$m(t, x_0) = e^{\mu + \frac{1}{2}\sigma^2} = x_0 e^{\theta_1 t},$$
$$v(t, x_0) = e^{2\mu + \sigma^2}\left(e^{\sigma^2} - 1\right) = x_0^2 e^{2\theta_1 t}\left(e^{\theta_2^2 t - 1}\right).$$

Hence

$$p_\theta(t, y|x_0) = \frac{1}{y\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log y - \mu)^2}{2\sigma^2}\right) =$$
$$= \frac{1}{y\theta_2\sqrt{2\pi t}} \exp\left(-\frac{(\log y - \log(x_0) + \left(\theta_1 - \frac{\theta_2^2}{2}\right) t)^2}{2\theta_2^2 t}\right).$$

No stationary distribution

$$\log X_t \sim \mathcal{N}((\mu - \sigma^2/2)t, \sigma^2 t).$$

2.3. The Cox-Ingersoll-Ross (CIR) model

The CIR (square-root) process is the solution to the SDE

$$dX_t = (\theta_1 - \theta_2 X_t)dt + \theta_3\sqrt{X_t}dW_t, \quad X_0 = x_0 > 0,$$

sometimes parametrized as

$$dX_t = \theta(\beta - X_t)dt + \sigma\sqrt{X_t}dW_t, \quad X_0 = x_0 > 0,$$

where $\theta_1,\ \theta_2,\ \theta_3 \in \mathbb{R}_+$.

If $2\theta_1 > \theta_3^2$, the process is strictly positive otherwise it is nonnegative, which means that it can reach the state 0. Of course, the last case is not admitted in finance when the CIR process is used to model interest rates. The SDE for CIR process has the explicit solution

$$X_t = \left(X_0 - \frac{\theta_1}{\theta_2}\right)e^{-\theta_2 t} + \theta_3 e^{-\theta_2 t}\int_0^t e^{\theta_2 u}\sqrt{X_u}dW_u. \tag{18}$$

The conditional distribution

For the Cox-Ingersoll-Ross process, the mean of the conditional density is that of the Ornstein-Uhlenbeck process

$$m(t, x_0) = \frac{\theta_1}{\theta_2} + \left(x_0 - \frac{\theta_1}{\theta_2}\right)e^{-\theta_2 t},$$

and the variance is

$$v(t, x_0) = x_0\frac{\theta_3^2(e^{-\theta_2 t} - e^{-2\theta_2 t})}{\theta_2} + \frac{\theta_1\theta_3^2}{2\theta_2^2}(1 - e^{-2\theta_2 t}).$$

The covariance function is given by

$$Cov(X_s, X_t) = x_0\frac{\theta_3^2}{\theta_2}\left(e^{-\theta_2 t} - e^{-\theta_2(s+t)}\right) + \frac{\theta_1\theta_3^2}{2\theta_2^2}\left(e^{-\theta_2(t-s)} - e^{-\theta_2(t+s)}\right).$$

The stationary law

The stationary distribution of the CIR process is a Gamma law with shape parameter $\alpha = 2\frac{\theta_1}{\theta_3^2}$ and scale parameter $\beta = \frac{\theta_3^2}{2\theta_2}$:

$$X \sim Gamma\left(2\frac{\theta_1}{\theta_3^2}, \frac{\theta_3^2}{2\theta_2}\right). \tag{19}$$

Hence the stationary law has mean equal to $m = \frac{\theta_1}{\theta_2}$ and variance $v = \frac{\theta_1\theta_3^2}{2\theta_2^2}$.

The covariance function in the stationary case is given by

$$Cov(X_s, X_t) = \frac{\theta_1 \theta_3^2}{2\theta_2^2} e^{-\theta_2(t-s)}.$$

A confidence interval of the CIR

Let $v(s,t) = p(t, y|s, x) = x_s$. We known [11] that

$$p(t, y|s, x) \sim \zeta \chi_k^2(\lambda),$$

with degrees of freedom $k = \frac{4}{\sigma^2}\theta\beta$, $\zeta = \frac{\sigma^2}{4}\frac{1-e^{-\theta(t-s)}}{\theta}$ and non-centrality parameter $\lambda = \frac{4\theta e^{-\theta(t-s)}}{\sigma^2(1-e^{-\theta(t-s)})}x_s$. Therefore, the random variable

$$\frac{v(s,t)}{\zeta} \sim \chi_k^2(\lambda).$$

We make the approximation of the chi-square by the standard normal distribution. So, the random variable z is given by

$$z = \frac{\frac{v(s,t)}{\zeta} - (k+\lambda)}{\sqrt{2(k+2\lambda)}} \sim \mathcal{N}(0,1), \quad \text{when } k \to \infty \text{ or } \lambda \to \infty.$$

A $100(1-\alpha)\%$ conditional confidence interval for z is given by

$$P(-\xi \leq z \leq \xi) = 1 - \alpha.$$

From this, we can obtain a confidence interval of $v(s,t)$ with following form

$$(v_{lower}(s,t), v_{upper}(s,t)) = \zeta(k + \lambda \pm \xi\sqrt{2(k+2\lambda)}) \tag{20}$$

with $\xi = \Phi^{-1}_{\mathcal{N}(0,1)}(1-\alpha/2)$.

2.4. The modified CIR model

The modified Cox-Ingersoll-Ross process is solution to

$$dX_t = -\theta_1 X_t d_t + \theta_2 \sqrt{1 + X_t^2} dW_t$$

with $\theta_1 + \theta_2^2/2 > 0$. This is needed to make the process positive recurrent.

This process has a stationary distribution whose density $\pi(\cdot)$ is proportional to

$$\pi(x) \propto \frac{1}{(1+x^2)^{1+\theta_1/\theta_2^2}}.$$

Setting $\nu = 1 + 2\theta_1/\theta_2^2$, then $X_t \sim t(\nu)/\sqrt{\nu}$, where $t(\cdot)$ is the Student distribution with ν degrees of freedom. Applying the Lamperti transform

$$F(x) = \int_0^x \frac{1}{\theta_2\sqrt{1+u^2}}du = \frac{\operatorname{arcsinh}(x)}{\theta_2} = \frac{1}{\theta_2}\log(x + \sqrt{1+x^2})$$

to the SDE, we have that $Y_t = F(X_t)$ satisfies the SDE

$$dF(X_t) = -(\theta_1/\theta_2 + \theta_2/2)\frac{X_t}{1+X_t^2}d_t + dW_t,$$

which we can rewrite in terms of the Y_t process as

$$dY_t = -(\theta_1/\theta_2 + \theta_2/2)\tanh(\theta_2 Y_t) + dW_t$$

with $Y_0 = F(X_0)$.

2.5. The Chan-Karolyi-Longstaff-Sanders family of models

The Chan-Karolyi-Longstaff-Sanders (CKLS) process solves the SDE

$$dX_t = (\theta_1 + \theta_2 X_t)dt + \theta_3 X_t^{\theta_4} dW_t.$$

This CKLS model is a further extension of the Cox-Ingersoll-Ross model and hence embeds all previous models. For details, see Table 1.4 in the book [7].

The CKLS model does not admit an explicit transition density unless $\theta_1 = 0$ or $\theta_4 = \frac{1}{2}$.

It takes values in $(0, +\infty)$, if $\theta_1, \theta_2 > 0$, and $\theta_4 > \frac{1}{2}$. In all cases, θ_3 is assumed to be positive.

2.6. The nonlinear mean reversion Aït-Sahalia model

This model satisfies the nonlinear SDE

$$dX_t = (\alpha_{-1}X_t^{-1} + \alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2)dt + \beta_1 X_t^{\rho} dW_t.$$

In general, there are no analytical solution of this SDE, but approximate transition densities can be obtained via Hermite polynomial expansion [7]. This model was proposed by Aït-Sahalia to model interest rates and later a further generalization was proposed that includes more structure in the diffusion coefficient. The second model is of the form

$$dX_t = (\alpha_{-1}X^{-1}t + \alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2)dt + \sqrt{\beta_0 + \beta_1 X_t + \beta_2 X_t^{\beta_3}}dW_t.$$

Some natural constraints on the parameters are needed in order to have a meaningful specification of the model. Moreover, Markovianity is granted only under additional constraints. The book [7] summarizes relations between coefficients.

It is clear that the CKLS model is just a particular case of the Aït-Sahalia model.

2.7. Double-well potential

This model is interesting because of the fact that its density has a bimodal shape. The process satisfies the SDE

$$dX_t = (X_t - X_t^3)dt + dW_t.$$

This model is challenging in the sense that the standard Euler approximation could not be expected to work due to the high non-linearity of the SDE and high non-Gaussianity of its finite-dimensional distributions.

2.8. The Jacobi diffusion process

The Jacobi diffusion process is the solution to the SDE

$$dX_t = -\theta\left(X_t - \frac{1}{2}\right)dt + \sqrt{\theta X_t(1-X_t)}dW_t$$

for $\theta > 0$. It has an invariant distribution that is uniform on $(0,1)$. The peculiar thing is that, given any twice differentiable distribution function F, the transformed diffusion $Y_t = F^{-1}(X_t)$ has an invariant density $\pi(\cdot)$ that is the density of F: $\pi = F'$.

2.9. Ahn and Gao (inverse of Feller's square root) model

Let the process X_t is the solution to the stochastic differential equation

$$dX_t = X_t(\theta_1 - (\theta_3^3 - \theta_1\theta_2)X_t)dt + \theta_3 X_t^{\frac{3}{2}}dW_t.$$

The conditional distribution of this process is related to that of the Cox-Ingersoll-Ross model as

$$p_\theta(t, y|x_0) = \frac{1}{y^2}p_\theta^{CIR}\left(t, \frac{1}{y}\middle|\frac{1}{x_0}\right),$$

where the conditional density

$$p_{\theta}^{CIR}(t, y|x_0) = ce^{-(u+v)} \left(\frac{u}{v}\right)^{q/2} I_q(2\sqrt{uv}),$$

with $u = cx_0e^{-\theta_2 t}$, $v = cy$, $q = 2\theta_1/\theta_3^2 - 1$, where $I_q(\cdot)$ is the modified Bessel function of the first kind of order q:

$$I_q(x) = \sum_{k=0}^{\infty} (x/2)^{2k+q} \frac{1}{k!\Gamma(k+q+1)}, \quad x \in \mathbb{R},$$

and $\Gamma(z) = \int_0^{\infty} x^{z-1}e^{-x}dx$ is the gamma function, $z \in \mathbb{R}_+$.

2.10. Radial Ornstein-Uhlenbeck process

The radial Ornstein-Uhlenbeck process is the solution to the SDE

$$dX_t = (\theta X^{-1}t - X_t)dt + dW_t$$

for $\theta > 0$. This model is a special case of the Aït-Sahalia model but still interesting because some distributional results are known.

In particular, the conditional distribution has the explicit form

$$p_{\theta}(t, y|x_0) = \frac{(\frac{y}{\theta})^{\theta}\sqrt{xy}e^{-y^2+(\theta+1/2)t}}{\sinh t} \cdot \exp\left(-\frac{x^2+y^2}{e^{2t}-1}\right) I_{\theta-1/2}\left(\frac{xy}{\sinh t}\right),$$

where I_{ν} is the modified Bessel function of order ν.

2.11. Pearson diffusions

A class that further generalizes the Cox-Ingersoll-Ross and Ornstein-Uhlenbeck processes is the class of Pearson diffusion. Its name is due to the fact that, when a stationary solution exists for this model, its invariant density belongs to the Pearson system. The Pearson system allows for a big variety of distributions which can take positive and/or negative values, and can be bounded, symmetric or skewed, and heavy or light tailed.

Pearson diffusions solve the SDE

$$dX_t = -\theta(X_t - \mu)dt + \sqrt{2\theta(aX_t^2 + bX_t + c)}dW_t$$

with $\theta > 0$ and a, b, and c such that the diffusion coefficient is well-defined, i. e., the square root can be extracted, for all the values of the state space of X_t.

Pearson diffusions are characterized as having a mean reverting linear drift and a squared diffusion coefficient that is a second-order polynomial of the state of the process.

A further nice property of these models is that they are closed under translation and scale transformations: if X_t is an ergodic Pearson diffusion then also $\tilde{X}_t = \gamma X_t + \delta$ is a Pearson diffusion satisfying the SDE with parameters

$$\tilde{a} = a, \quad \tilde{b} = b\gamma - 2a\delta, \quad \tilde{c} = c\gamma^2 - b\gamma\delta + a\delta^2, \quad \tilde{\theta} = \theta \text{ and } \tilde{\mu} = \gamma\mu + \delta.$$

The parameter γ may also be negative, and in that case the state space of $\tilde{X}_t$ will change its sign.

The scale and the speed measures of these processes have the forms

$$s(x) = \exp\left(\int_{x_0}^{x} \frac{u-\mu}{au^2+bu+c} du\right)$$

and

$$m(x) = \frac{1}{2\theta s(x)(ax^2+bx+c)},$$

where x_0 is some value such that $ax_0^2 + bx_0 + c > 0$.

Additional details of Pearson diffusions one can find in [7, 14].

2.12. Practical Problems

1. Use the stochastic integral (6), $N = 100$ to simulate and plot a path with $X_0 = 10$ of

- the Ornstein-Uhlenbeck process, $\theta = (-5.0, 3.5)$;
- the Cox-Ingersoll-Ross process, $\theta = (2.00, 0.20, 0.15)$;
- the Black-Scholes-Merton process, $\theta = (1.0, 0.2)$.

2. Compute a confidence interval (20) of the CIR process from the previous item. Plot the path and a confidence interval. The expected result is in Figure 7.

3. For the Ornstein-Uhlenbeck process

$$dX_t = \theta(\mu - X_t)dt + \sigma dW_t, \quad \theta = 1, \quad \mu = 1.2, \quad \sigma = 0.3$$

generate three paths: $X_0 = 0.0, 2.0, \mathcal{N}\left(\mu, \frac{\sigma^2}{2\theta}\right)$ and plot in graph. Define a path with the stationary distribution from the graph. The expected result is in Figure 8.

4. Design a process that stays in the interval $[0, 1]$ and mean-reverts around $1/2$, generate a path and plot it in graph. Prove your solution.

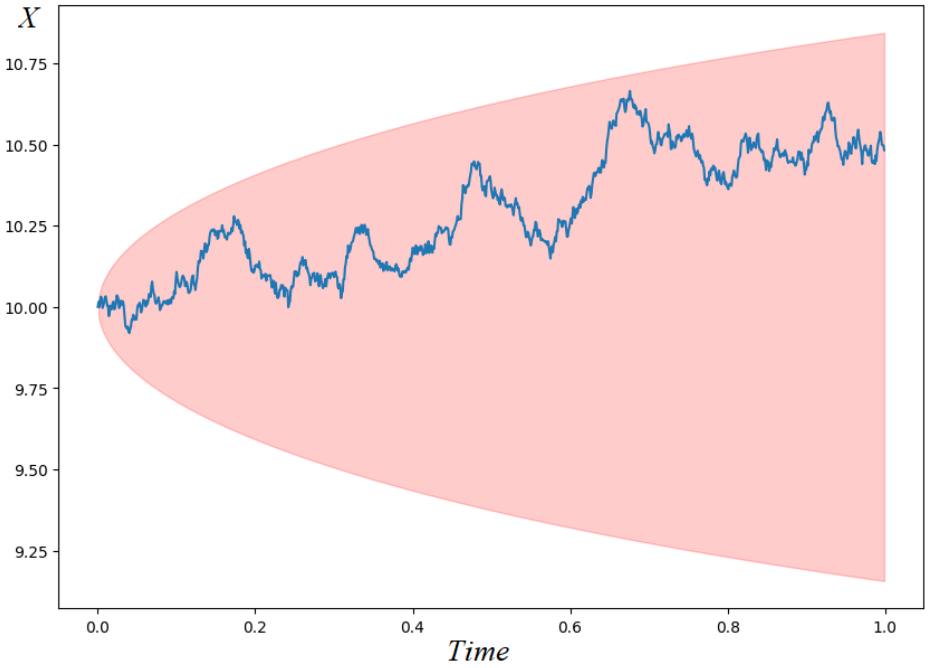

Figure 7. A confidence interval of the CIR with $X_0 = 10$, parameters $\theta = (2., 0.2, 0.15)$, and $\alpha = 0.05$ %

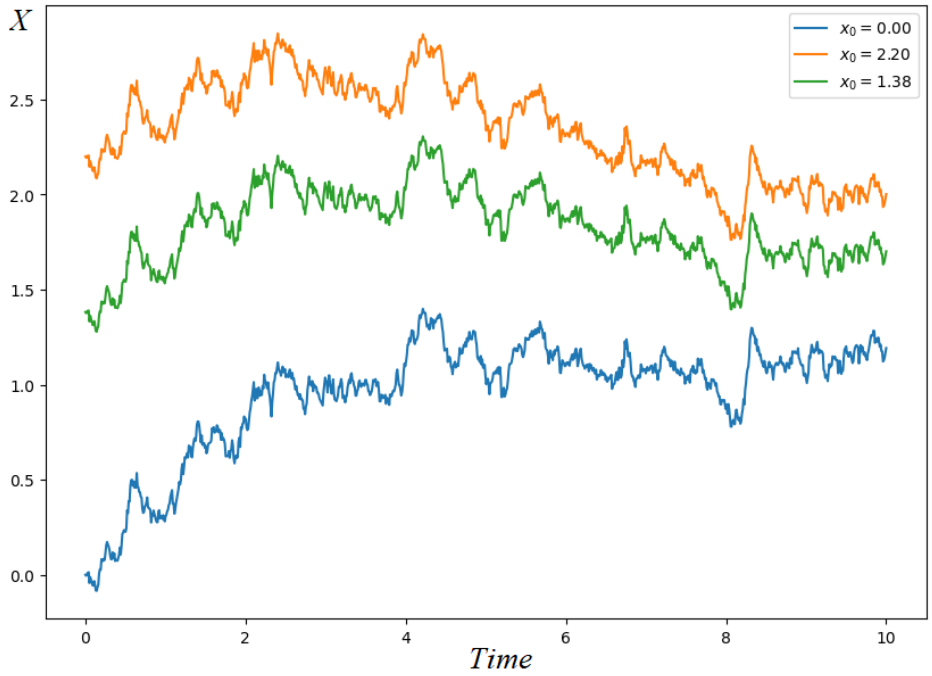

Figure 8. Three paths of the Ornstein-Uhlenbeck process with $X_0 = 0.0, 2.0,$ and $\mathcal{N}\left(\mu, \frac{\sigma^2}{2\theta}\right) = 1.38$

3. Numerical methods for SDEs

There are two main objectives in the simulation of the trajectory of a process solution of a SDE: either interest is in the whole trajectory or in the expected value of some functional of the process (moments, distributions, etc) which usually are not available in explicit analytical form.

Simulation methods are usually based on discrete approximations of the continuous solution to a SDE. The methods of approximation are classified according to their different properties. Mainly two criteria of optimality are used in the literature: the strong and the weak (orders of) convergence.

Order of strong convergence

A time-discretized approximation Y_δ of a continuous-time process Y, with δ the maximum time increment of the discretization, is said to have *order* γ *of strong convergence* to Y if for any fixed time horizon T it holds true that

$$\mathbb{E}|Y_\delta(T) - Y(T)| \leq C\delta^\gamma, \quad \forall \delta < \delta_0,$$

with $\delta_0 > 0$ and C a constant not depending on δ. This kind of criterion is similar to the one used in the approximation of the trajectories of non-stochastic dynamical systems.

Order of weak convergence

Along with the strong convergence, the weak convergence can be defined. Y_δ is said to converge weakly of order β to Y if for any fixed horizon T and any $2(\beta+1)$ continuous differentiable function g of polynomial growth, it holds true that

$$|\mathbb{E}g(Y(T)) - \mathbb{E}g(Y_\delta(T))| \leq C\delta^\beta, \quad \forall \delta < \delta_0,$$

with $\delta_0 > 0$ and C a constant not depending on δ.

3.1. Approximation methods

As a first step we shall consider the discretization of time and present the Euler and Milstein scheme.

3.1.1. Euler-Maruyama approximation

One of the most used schemes of approximation is the Euler method, originally used to approximate solutions to deterministic differential equations.

The idea is the following: given an Ito process $\{X_t, 0 \leq t \leq T\}$ solution of the SDE

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \quad X_s = x$$

and the discretization

$$\Pi_N = \Pi_N([0, T])$$

of the interval $[0, T]$, $0 \leq s = t_0 < t_1 < \ldots < t_N = T$.

The Euler approximation of X is a continuous stochastic process Y satisfying the iterative scheme

$$Y_{i+1} = Y_i + b(t_i, Y_i)(t_{i+1} - t_i) + \sigma(t_i, Y_i)(W_{i+1} - W_i), \tag{21}$$

for $i = 0, 1, \ldots, N-1$, with $Y_0 = X_0$. We have simplified the notation setting $Y(t_i) = Y_i$ and $W(t_i) = W_i$.

Usually the time increment $\Delta t = t_{i+1} - t_i$ is taken to be constant, i. e., $\Delta t = 1/N$. In between any two time points t_i and t_{i+1}, the process can be defined differently. One natural approach is to consider linear interpolation so that $Y(t)$ is defined as

$$Y(t) = Y_i + \frac{t - t_i}{t_{i+1} - t_i}(Y_{i+1} - Y_i), \quad t \in [t_i, t_{i+1}).$$

From Equation (21), one can see that to simulate the process Y one only needs to simulate the increment of the Wiener process. The Euler scheme has order of strong convergence $\gamma = \frac{1}{2}$.

3.1.2. Milstein scheme

The Milstein scheme makes use of Ito's lemma to increase the accuracy of the approximation by adding the second-order term.

Denoting by σ_x the partial derivative of $\sigma(t, x)$ with respect to x, the Milstein approximation looks like

$$\begin{aligned} Y_{i+1} = Y_i + b(t_i, Y_i)(t_{i+1} - t_i) + \sigma(t_i, Y_i)(W_{i+1} - W_i) + \\ + \frac{1}{2}\sigma(t_i, Y_i)\sigma_x(t_i, Y_i)\{(W_{i+1} - W_i)^2 - (t_{i+1} - t_i)\}, \end{aligned} \tag{22}$$

or, in more symbolic form,

$$Y_{i+1} = Y_i + b\Delta t + \sigma\Delta W_t + \frac{1}{2}\sigma\sigma_x\{(\Delta W_t)^2 - \Delta t\}.$$

This scheme has orders of strong and weak convergence equal to 1.

3.1.3. Predictor-corrector method

Both schemes of discretization (21),(22) consider the coefficients b and σ as not varying during the time interval Δt, which is of course untrue for a generic SDE, as b and σ can depend on both the time t and the state of the process X_t.

One way to recover the varying nature of these coefficients is to average their values in some way. Since the coefficients depend on X_t and we are simulating X_t, the method we present here just tries to approximate the states of the process first. This method is of weak convergence order 1.

The predictor-corrector algorithm

First consider the simple approximation (the predictor)

$$\tilde{Y}_{i+1} = Y_i + b(t_i, Y_i)\Delta t + \sigma(t_i, Y_i)\sqrt{\Delta t}Z.$$

Then choose two weighting coefficients α and η in $[0, 1]$, and calculate the corrector as

$$\begin{aligned} Y_{i+1} = Y_i + (\alpha\tilde{b}(t_{i+1}, \tilde{Y}_{i+1}) + (1-\alpha)\tilde{b}(t_i, Y_i))\Delta t + \\ + (\eta\sigma(t_{i+1}, \tilde{Y}_{i+1}) + (1-\eta)\sigma(t_i, Y_i))\sqrt{\Delta t}Z \end{aligned}$$

with

$$\tilde{b}(t_i, Y_i) = b(t_i, Y_i) - \eta\sigma(t_i, Y_i)\sigma_x(t, Y_i).$$

Note that the predictor-corrector method falls back to the standard Euler method for $\alpha = \eta = 0$.

3.1.4. Kloden-Platen-Schurz-Sorensen method

By adding more terms to the Ito-Taylor expansion, one can achieve a order of strong convergence higher than 1.

Denoting by b_x, b_{xx}, σ_x, σ_{xx} the partial derivative of the drift and diffusion coefficient $b(t, x)$, $\sigma(t, x)$ with respect to x. In particular, the following Kloden-Platen-Schurz-Sorensen (KPS) scheme has order of strong convergence $\gamma = 1.5$:

$$\begin{aligned} Y_{i+1} = Y_i + b\Delta t + \sigma\Delta W_t + \frac{1}{2}\sigma\sigma_x\{(\Delta W_t)^2 - \Delta t\} + \\ + \sigma b_x \Delta U_t + \frac{1}{2}\{bb_x + \frac{1}{2}\sigma^2 b_{xx}\}\Delta t^2 + \\ + \{b\sigma_x + \frac{1}{2}\sigma^2\sigma_{xx}\}\{\Delta W_t\Delta t - \Delta U_t\} + \frac{1}{2}\sigma(\sigma\sigma_x)_x\{\frac{1}{3}(\Delta W_t)^2 - \Delta t\}\Delta W_t, \end{aligned}$$

where

$$\Delta U_t = \int_{t_0}^{t_{i+1}} \int_{t_i}^{s} dW_u ds$$

is a Gaussian random variable with zero mean and variance $\frac{1}{3}\Delta t^3$ and such that $\mathbb{E}(\Delta U_t \Delta W_t) = \frac{1}{2}\Delta t^2$. All the pairs $(\Delta W_t, \Delta U_t)$ are mutually independent for all t_i's.

3.1.5. Second Milstein scheme

The second Milstein scheme has weak second-order convergence in contrast to the weak first order convergence of the Euler scheme. This scheme requires partial (first and second) derivatives of both drift and diffusion coefficients

$$Y_{i+1} = Y_i + \left(b - \frac{1}{2}\sigma\sigma_x\right)\Delta t + \sigma Z\sqrt{\Delta t} + \frac{1}{2}\sigma\sigma_x \Delta t Z^2 +$$
$$+ \Delta t^{3/2}\left(\frac{1}{2}b\sigma_x + \frac{1}{2}b_x\sigma + \frac{1}{4}\sigma^2\sigma_{xx}\right)Z + \Delta t^2\left(\frac{1}{2}bb_x + \frac{1}{4}b_{xx}\sigma^2\right).$$

3.2. Drawing from the transition density

All the methods presented so far are based on the discretized version of the SDE. In the case where a transition density of X_t given some previous value X_s, $s < t$, is known in explicit form, direct simulation from this can be done.

Unfortunately, the transition density is known for very few processes, and these cases are the ones for which exact likelihood ratio can be obtained.

In these fortunate cases, the algorithm for simulating processes is very easy to implement.

We suppose that a random number generator is available for the transition density of the process X_t:

$$p_\theta(\Delta, y|x) = Pr(X_{t+\Delta} \in dy | X_t = x).$$

If this generator is not available one can always use one of the standard methods to draw from known densities, such as the rejection method.

Rejection Method

1. Simulate the value of Y, having probability mass function q_j.

2. Generate a random number U.

3. If $U < p_Y/cq_y$, set $X = Y$ and stop. Otherwise, return to Step 1.

Example. Suppose we wanted to simulate the value of a random variable $1, 2, \ldots, 10$ that takes one of the values with respective probabilities

$$0.11, 0.12, 0.09, 0.08, 0.12, 0.10, 0.09, 0.09, 0.10, 0.10.$$

We will use the rejection method with q being the discrete uniform density on $1, 2, \ldots, 10$. That is, $q_j = 1/10$, $j = 1, 2, \ldots, 10$. For this choice of $\{q_j\}$ we can choose by $c = \max p_j/q_j = 1.2$ and so the algorithm would be as follows:

1. Generate a random number U_1 and set $Y = int(10U_1) + 1$.

2. Generate a second random number U_2.

3. If $U_2 < p_Y/0.12$, set $X = Y$ and stop. Otherwise return to Step 1.

The constant 0.12 in Step 3 arises since $cq_Y = 1.2/10 = -0.112$. On average, this algorithm requires only 1.2 iterations to obtain the generated value of X.

3.3. Practical Problems

1. Using different approximations

- the Euler approximation algorithm;
- the 1st, 2nd Milstein schemes;
- the predictor-corrector method (set default $\alpha = \eta = 1/2$);
- KPS method,

simulate and plot the trajectory with different levels of discretization,

$$N = \{2, 4, 8, 128, 256, 512, 8192, 16384, 32768\}$$

being the number of subintervals of $[0, 1]$ of the Brownian motion

$$dX_t = \theta_1 X_t dt + \theta_2 X_t dW_t, \quad X_0 = 1, \text{ with } (\theta_1, \theta_2) = (2, 0.5).$$

Expected result for speed of increased levels of discretization, N in Figure 9.

2. Repeat the problem 1 for the Cox-Ingersoll-Ross process

$$dX_t = (\theta_1 - \theta_2 X_t)dt + \theta_3\sqrt{X_t}dW_t, X_0 = 10, (\theta_1, \theta_2, \theta_3) = (6, 3, 2).$$

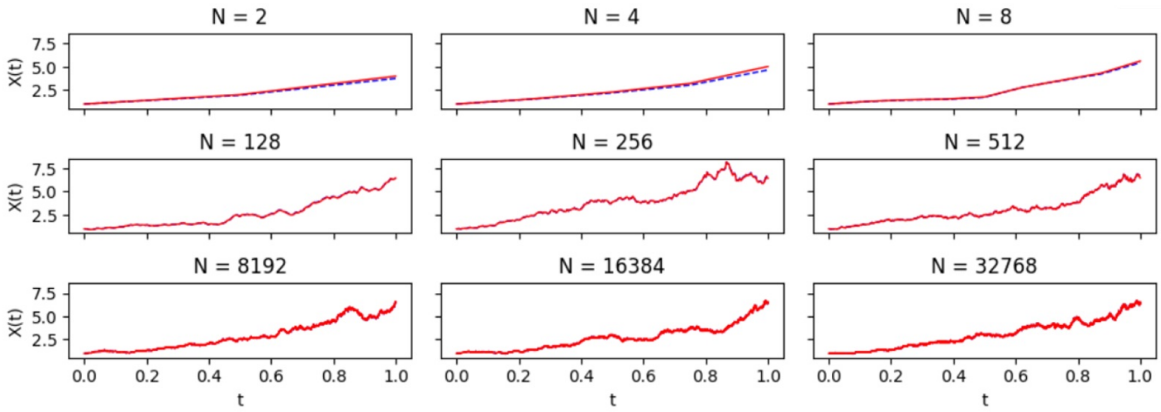

Figure 9. A simulated path of the Brownian motion for increased levels of discretization, N being the number of subintervals of $[0, 1]$: Euler (red solid line) and 1st Milstein (blue dotted line) schemes

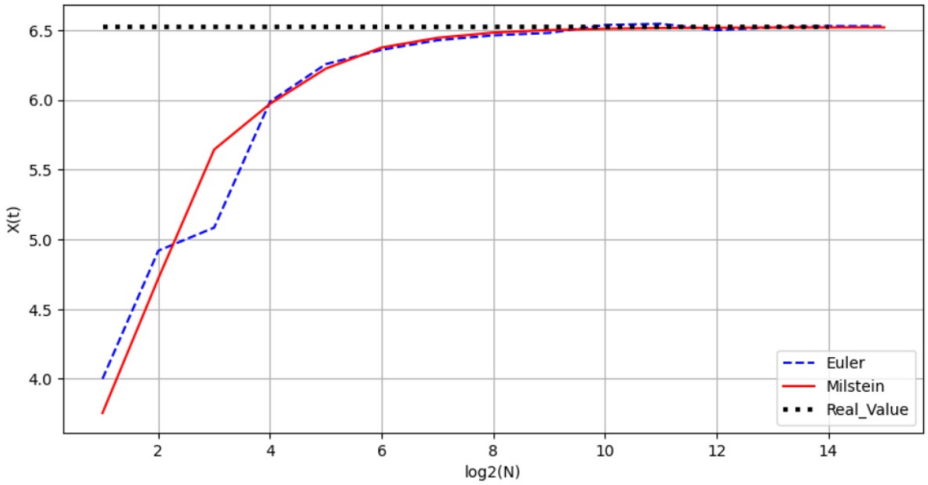

Figure 10. Speed of convergence of Euler and 1st Milstein schemes to the true value (dotted line) as a function of the discretization step $\Delta t = 1/N$

3. Compare the speed of convergence of Euler and 1st Milstein schemes (Brownian motion). Expected result in Figure 10.

4. Write a random number generator function for the processes:

- Ornstein-Uhlenbeck process (15),
- Geometric Brownian motion process (17),
- Cox-Ingersoll-Ross process (18),

then set parameters, generate and save time series into *.csv file.

5. For the processes from previous item generate trajectories and plot

- cumulative distribution (cdf),
- density function (pdf),
- calculate the 95%-quantile.

4. Parametric Estimation

In this section we consider parametric estimation problems for diffusion processes sampled at discrete times. We can imagine different schemes of observation.

1. *Large sample scheme*: In this scheme, the time Δ between two consecutive observations is fixed and the number of observations n increases. In this case, the window of observation $[0, n\Delta = T]$ also increases with n. In this framework, which is considered the most natural, additional assumptions on the underlying continuous model are required such as stationarity and/or ergodicity.

2. *High-frequency scheme*: In this case, $\Delta = \Delta_n$ goes to zero as n increases, and the window of observation $[0, n\Delta_n = T]$ is fixed. Neither stationarity nor ergodicity is needed.

3. *Rapidly increasing design*: Δ_n shrinks to zero as n grows, but the window of observation $[0, n\Delta_n]$ increases with n; i. e.,

$$n\Delta_n \to \infty.$$

In this case, stationarity or ergodicity is needed. Further, the mesh Δ_n should go to zero at a prescribed rate $n\Delta_n^k \to 0$, $k \geq 2$. For high values of k, this is not a severe constraint because this means that Δ_n goes to zero but slowly.

A random process is a collection of random variables, one for each time instant under consideration. Typically this may be

1) continuous time: $-\infty < t < \infty$;
2) discrete time: all integers n, or all time instants nT where T is the sample interval.

Stationarity refers to the *distributions* of the random variables.

Specifically, in a stationary process, all the random variables have the same distribution function, and more generally, for every positive integer n and n time instants $t_1, t_2, \ldots, t_n$, the joint distribution of the n random variables

$$X(t_1), X(t_2), \ldots, X(t_n)$$

is the same as the joint distribution of

$$X(t_1 + \tau), X(t_2 + \tau), \ldots, X(t_n + \tau).$$

That is, if we shift all time instants by τ, the statistical description of the process does not change at all: the process is stationary.

Ergodicity, on the other hand, is not defined by statistical properties of the random processes but by their *sample paths*, i. e. what you observe physically.

Referring back to the random variables, recall that random variables are mappings from a sample space to the real numbers; each outcome is mapped onto a real number, and different random variables will typically map any given outcome to different numbers.

The underlying continuous model

Consider the one-dimensional, time-homogeneous SDE

$$dX_t = b(X_t, \theta)dt + \sigma(X_t, \theta)dW_t, \tag{23}$$

where $\theta \in \Theta \subset \mathbb{R}^p$ is the multidimensional parameter and θ_0 is the true parameter to be estimated.

The functions

$$b : \mathbb{R} \times \Theta \to \mathbb{R} \quad \text{and} \quad \sigma : \mathbb{R} \times \Theta \to (0, \infty)$$

are known and such that the solution of Equation (23) exists.

The state space of the process is denoted by

$$I = (l, r), \quad \text{and} \quad -\infty \leq l < r \leq +\infty$$

is an open set and the same for all θ. Moreover, for any $\theta \in \Theta$ and any random variable ξ with support in I, equation (23) has a unique strong solution for $X_0 = \xi$.

When ergodicity is required, additional assumptions to guarantee the existence of the invariant distribution $\pi_\theta(\cdot)$ should be imposed. In this case, the solution of Equation (23) with $X_0 = \xi \sim \pi_\theta$ is strictly stationary and ergodic.

We have seen different sets of sufficient conditions, and when $\pi_\theta(\cdot)$ exists it has the form

$$\pi_\theta(x) = \frac{1}{M(\theta)\sigma^2(x, \theta)s(x, \theta)},$$

where

$$s(x, \theta) = \exp\left(-2\int_{x_0}^{x} b(y, \theta)\sigma^2(y, \theta)dy\right)$$

for some $x_0 \in I$ and $M(\theta)$ the normalizing constant.

Remain, the function s is called the *scale measure* and

$$m(x) = \pi_\theta(x) \cdot M(\theta)$$

is called the *speed measure*.

The distribution of X with $X_0 \sim \pi_\theta$ is denoted by P_θ, and under P_θ, $X_t \sim \pi_\theta$ for all t.

We further denote by $p_\theta(t, \cdot|x)$ the conditional density (or transition density) of X_t given $X_0 = x$.

As X is time-homogeneous, $p_\theta(t, \cdot|x)$ is just the density of X_{s+t} conditional on $X_s = x$ for all $t \geq 0$.

In some cases, we will use the notation $p(t, \cdot|x, \theta)$. As already mentioned, the transition probabilities in most of the cases are not known in explicit analytic form. On the contrary, the invariant density is easier to obtain (up to the normalizing constant).

Introduce the *infinitesimal generator* of the diffusion X

$$\mathcal{L}_\theta f(x, \theta) = b(x, \theta) f_x(x, \theta) + \frac{1}{2}\sigma^2(x, \theta) f_{xx}(x, \theta).$$

Here $f(\cdot)$ is a twice continuous differentiable function

$$f : \mathbb{R} \times \Theta \to \mathbb{R},$$

where $f_x(\cdot)$ and $f_{xx}(\cdot)$ are the first and second partial derivatives of $f(\cdot)$ with respect to argument x.

It is quite straightforward to estimate the parameters efficiently. In particular, elements of vector θ (at least the subset of parameters concerning the diffusion part of Equation (23)) can be calculated rather than estimated from the quadratic variation of the process since, for all $t \geq 0$,

$$\langle X, X \rangle_t = \lim_{n \to \infty} \sum_{k=1}^{2^n} (X_{t \wedge k/2^n} - X_{t \wedge (k-1)/2^n})^2 = \int_0^t \sigma^2(X_s, \theta) ds \qquad (24)$$

as $n \to \infty$ in probability under P_θ.

Once the diffusion coefficient is known: $\sigma(x, \theta) = \sigma(x)$ the likelihood function of X is given by

$$L_T(\theta) = \exp\left(\int_0^T \frac{b(X_s, \theta)}{\sigma(X_s)} dW_s - \frac{1}{2} \int_0^T \frac{b^2(X_s, \theta)}{\sigma^2(X_s)} ds \right).$$

Therefore, θ can be estimated by maximizing $L_T(\theta)$.

The discrete-time observations

We assume that the process is observed at discrete times $t_i = i\Delta_i$, $i = 0, 1, \ldots, n$, and $T = n\Delta_n$. In some cases, the sampling rate has to be constant $\Delta_i = \Delta$ or such that $\max_i \Delta_i < \Delta$ for some fixed Δ; in other cases, Δ_n varies and it is assumed that $n\Delta_n^k \to 0$ for some power $k \geq 2$. The asymptotics is considered as $n \to \infty$, which is equivalent to $T \to \infty$.

In the following, we will denote $X_{t_i} = X_{i\Delta_i}$ just by X_i to simplify the writing.

We further denote by $\mathcal{F}_n = \sigma\{X_{t_i}, i \geq n\}$ the σ-field generated by the first n observations with $\mathcal{F}_0$ the trivial σ-field.

The discrete counterpart of $L_T(\theta)$ we are interested in, conditional on X_0, is given by

$$L_n(\theta) = \prod_{i=1}^{n} p_\theta(\Delta, X_i|X_{i-1})p_\theta(X_0),$$

which can be derived using the Markov property of X.

We denote by $\ell_n(\theta) = \log L_n(\theta)$ the log-likelihood function

$$\ell_n(\theta) = \log L_n(\theta) = \sum_{i=1}^{n} \ell_i(\theta) + \log(p_\theta(X_0)) =$$

$$= \sum_{i=1}^{n} \log p_\theta(\Delta, X_i|X_{i-1}) + \log(p_\theta(X_0)).$$

Usually $p_\theta(X_0)$ is not known unless the process is assumed to be in a stationary regime but, even in this case, it is not always easy to determine $p_\theta(X_0)$. If the number of observations increases with time, one can assume that the relative weight of $p_\theta(X_0)$ in the whole likelihood $L_n(\theta)$ decreases, so we will assume that $p_\theta(X_0) = 1$ from now on without mentioning it any further.

In the following, we will use a dot "$\dot{}$" for single time times differentiation with respect to the vector θ and $\partial_{\theta_i} f(\cdot)$ for $\frac{\partial}{\partial_{\theta_i}} f(\cdot)$ to keep formulas compact but still understandable. When the transition density is differentiable, we can define the score (vector) function

$$\dot{\ell}_n(\theta) = \sum_{i=1}^{n} \dot{\ell}_i(\theta) = \sum_{n}^{i=1} (\partial_{\theta_1}\ell_i(\theta), \ldots, \partial_{\theta_p}\ell_i(\theta))^\top$$

and the Fisher information matrix for θ,

$$i_n(\theta) = \sum_{i=1}^{n} \mathbb{E}_\theta\{\dot{\ell}_i(\theta)\dot{\ell}_i(\theta)^\top\},$$

where $^\top$ denotes the transposition operator. Since $p_\theta(t, \cdot|x)$ is usually not known explicitly, so are $L_n(\theta)$ and all the derived quantities. There are different ways to deal with this problem and we will show some options in what follows. Still, there are quite important models for which the transition density is known in explicit form.

The maximum likelihood estimator (MLE) of θ is defined to be the maximizer of the following constrained optimization problem:

$$\hat{\theta}_n := \arg\max_{\theta \in \Theta} L_n(\theta),$$

and we will refer to this as the Exact MLE, as it utilizes the exact transition density. Hence we start with exact likelihood inference for some models and make the following assumptions.

4.1. Exact likelihood inference

Assumption 3.1 (Global Lipschitz assumption). There exists a constant K independent of θ such that

$$|b(x, \theta) - b(y, \theta)| + |\sigma(x, \theta) - \sigma(y, \theta)| < K|x - y|.$$

Assumption 3.2 (Linear growth assumption). For all x, there exists a constant K independent of θ such that

$$|b(x, \theta)| + |\sigma(x, \theta)| < K(1 + |x|).$$

Assumption 3.3 (Positiveness of diffusion coefficient).

$$\inf_x \sigma^2(x, \theta) > 0.$$

Assumption 3.4 (Bounded moments). For all $k > 0$, all moments of order k of the diffusion process exist and are such that

$$\sup_t \mathbb{E}|X_t|^k < \infty.$$

Assumption 3.5 (Smoothness of the coefficients). The two coefficients b and σ are smooth in x (eventually up to order 3) and have polynomial growth in x uniformly on θ.

Estimation for Ornstein-Uhlenbeck Process

Recall, from the Section 2.1: the Ornstein-Uhlenbeck (or Vasicek) process is the unique solution to the following SDE (14):

$$dX_t = \kappa(\alpha - X_t)dt + \sigma dW_t, \quad X_0 = x_0,$$

where σ is interpreted as the volatility, α is the long-run equilibrium value of the process, and κ is the speed of reversion.

The conditional distribution of X_t given X_{t-1} is

$$\mathcal{N}\left(X_{t-1}e^{-\kappa\Delta} + \alpha(1 - e^{-\kappa\Delta}), \frac{1}{2}\sigma^2\kappa^{-1}(1 - e^{-2\kappa\Delta})\right),$$

where Δ is the *sampling interval* and can be either fixed or very small corresponding to high-frequency data, and the stationary distribution is $\mathcal{N}(\alpha, \frac{1}{2}\sigma^2\kappa^{-1})$.

The conditional mean and variance of X_t given X_{t-1} are

$$\mathbb{E}(X_t|X_{t-1}) = X_{t-1}e^{-\kappa\Delta} + \alpha(1 - e^{-\kappa\Delta}) =: \mu(X_{t-1})$$

and

$$Var(X_t|X_{t-1}) = \frac{1}{2}\sigma^2\kappa^{-1}(1 - e^{-2\kappa\Delta}) := \nu(X_{t-1}).$$

Let $\Phi(x)$ be the density function of the standard normal distribution $\mathcal{N}(0, 1)$. Then, the likelihood function of $\theta = (\kappa, \alpha, \sigma^2)$ is

$$L(\theta) = \Phi\left(\frac{\sqrt{2\kappa}}{\sigma}(X_0 - \alpha)\right)\prod_{t=1}^{n}\Phi\left(\frac{\{X_t - \mu(X_{t-1})\}}{\sqrt{\nu(X_{t-1})}}\right).$$

The maximum likelihood estimators (MLE) are

$$\hat{\kappa} = -\frac{1}{\delta}\log(\hat{\beta}_1), \quad \hat{\alpha} = \hat{\beta}_2 \quad \text{and} \quad \hat{\sigma}^2 = 2\hat{\kappa}\hat{\beta}_3(1 - \hat{\beta}_1^2)^{-1}$$

where

$$\hat{\beta}_1 = \frac{n^{-1}\sum_{i=1}^{n}X_iX_{i-1} - n^{-2}\sum_{i=1}^{n}X_i\sum_{i=1}^{n}X_{i-1}}{n^{-1}\sum_{i=1}^{n}X_{i-1}^2 - n^{-2}(\sum_{i=1}^{n}X_{i-1})^2},$$

$$\hat{\beta}_2 = \frac{n^{-1}\sum_{i=1}^{n}(X_i - \hat{\beta}_1X_{i-1})}{1 - \hat{\beta}_1}$$

and

$$\hat{\beta}_3 = n^{-1} \sum_{i=1}^{n} \{X_i - \hat{\beta}_1 X_{i-1} - \hat{\beta}_2(1 - \hat{\beta}_1)\}^2.$$

In the paper [15] the parameter's estimation was carried out under two asymptotic regimes. In the first regime (large sample scheme) it was assumed that $n \to \infty$ while Δ is a fixed constant and in the second regime (rapidly increasing design) that $n \to \infty$, $\Delta \to 0$, $T = n\Delta \to \infty$ and $T\Delta^{1/k} \to \infty$ for some $k > 2$.

Example. Particular Ornstein-Uhlenbeck model specification. If $\theta_1 = 0$, then the SDE (13) becomes

$$dX_t = -\theta_2 X_t dt + \theta_3 dW_t$$

and the only parameters of interest are θ_2 and θ_3. In this case, if Δ is fixed, the maximum likelihood estimator of θ_2 and θ_3 are available in explicit form [7]:

$$\hat{\theta}_{2,n} = -\frac{1}{\Delta} \log \left(\frac{\sum_{i=1}^{n} X_{i-1} X_i}{\sum_{i=1}^{n} X_{i-1}^2} \right), \tag{25}$$

$$\hat{\theta}_{3,n} = \sqrt{\frac{2\hat{\theta}_{2,n}}{n(1 - e^{-2\Delta\hat{\theta}_{2,n}})} \sum_{i=1}^{n} (X_i - X_{i-1} e^{-\Delta\hat{\theta}_{2,n}})^2}. \tag{26}$$

4.2. Pseudo-likelihood methods

Another way of obtaining estimators is to use some approximation scheme. These approximation schemes do not approximate the transition density directly but the path of the process.

4.2.1. Euler method

Consider a process solution of the general SDE:

$$dX_t = b(X_t, \theta)dt + \sigma(X_t, \theta)dW_t, \quad t \geq 0, \quad X_0 = x_0, \tag{27}$$

The Euler scheme has the form:

$$X_{t+\Delta t} - X_t = b(X_t, \theta)\Delta t + \sigma(X_t, \theta)(W_{t+\Delta t} - W_t). \tag{28}$$

The increments $X_{t+\Delta t} - X_t$ are independent Gaussian random variables with mean: $b(X_t, \theta)\Delta t$, and variance $\sigma^2(X_t, \theta)\Delta t$.

The transition density of the process can be written as

$$p_\theta(t, y|x) = \frac{1}{\sqrt{2\pi \cdot t \cdot \sigma^2(x, \theta)}} \exp\left(-\frac{(y - x - b(x, \theta)t)^2}{2 \cdot t \cdot \sigma^2(x, \theta)}\right),$$

and the log-likelihood is:

$$\ell_n(\theta) = -\frac{1}{2}\left(\sum_{i=1}^{n} \frac{(X_i - X_{i-1} - b(X_{i-1}, \theta)\Delta)^2}{\sigma^2 \Delta t} + n \log(2\pi\sigma^2\Delta t)\right).$$

The $\ell_n(\theta)$ is also called the *locally Gaussian approximation.*

4.2.2. Local linearization methods

Another approach to approximate the solution of a SDE is to use a local linearization method.

Ozaki method. Consider the homogeneous SDE

$$dX_t = b(X_t)dt + \sigma dW_t, \qquad t \geq 0, \quad X_0 = x_0, \tag{29}$$

where σ is supposed to be constant.

The construction of the method starts from the corresponding deterministic dynamical system $\frac{dx_t}{dt} = b(x_t)$, where x_t has to be a smooth function of t in the sense that it is two times differentiable with respect to t. Then we have

$$\frac{d^2x_t}{dt^2} = b_x(x_t)\frac{dx_t}{dt}. \tag{30}$$

Suppose now that $b_x(x)$ is constant in the interval $[t, t+\Delta t)$, and hence by iterated integration of both sides of the equation (30), first from t to $u \in [t, t+\Delta t)$ and then from t to $t+\Delta t$, we obtain the difference equation

$$x_{t+\Delta t} = x_t + \frac{b(x_t)}{b_x(x_t)}\left(e^{b_x(x_t)\Delta t} - 1\right). \tag{31}$$

Now we return back to the original SDE (29). So, suppose $b(x)$ is approximated by the linear function $K_t \cdot x$, where K_t is constant in the interval $[t, t + \Delta t)$. The solution to the SDE is

$$X_{t+\Delta t} = X_t e^{K_t \Delta t} + \sigma \int_t^{t+\Delta t} e^{K_t(t+\Delta t-u)} dW_u.$$

From the assumption on the conditional expectation of $X_{t+\Delta t}$ given X_t

$$\mathbb{E}(X_{t+\Delta_t}|X_t) = X_t e^{K_t \Delta_t},$$

we ask for the following equality to hold:

$$X_t e^{K_t \Delta t} = X_t + \frac{b(X_t)}{b_x(X_t)} \left(e^{b_x(X_t)\Delta t} - 1 \right),$$

and finally:

$$K_t = \frac{1}{\Delta t} \log \left(1 + \frac{b(X_t)}{X_t b_x(X_t)} \left(e^{b_x(X_t)\Delta t} - 1 \right) \right).$$

Notice that K_t depends on t only through the state of the process $X_t = x$ and later we use K_x instead of K_t.

The transition density of $X_{t+\Delta t}$ for given X_t is Gaussian with mean and variance respectively:

$$x + \frac{b(x)}{b_x(x)} \left(e^{b_x(x)\Delta t} - 1 \right) \quad \text{and} \quad \sigma^2 \cdot \frac{e^{2K_x \Delta t} - 1}{2K_x}, \tag{32}$$

where

$$K_x = \frac{1}{\Delta t} \log \left(1 + \frac{b(x)}{x b_x(x)} \left(e^{b_x(x)\Delta t} - 1 \right) \right).$$

Shoji-Ozaki method. Consider the SDE

$$dX_t = b(t, X_t)dt + \sigma(X_t)dW_t, \quad t \geq 0, \quad X_0 = x_0, \tag{33}$$

where the drift is allowed to depend on the time variable t, and also the diffusion coefficient can be varied. It is possible to transform Equation (33) into one with a constant diffusion coefficient using the Lamperti transform. So one can start by considering the nonhomogeneous SDE

$$dX_t = b(t, X_t)dt + \sigma dW_t, \tag{34}$$

which is different from Equation (29) in that the drift function also depends on variable t. The main point is that the equation (34) is approximated locally on $[s, s + \Delta s)$. For the details, one can read the book [7].

The transition density of $X_{s+\Delta s}$ given X_s is Gaussian with mean and variance

$$\mathrm{A}(X_s) = 1 + \frac{b(s, X_s)}{L_s \mathrm{L}_s} \left(e^{\mathrm{X}_s \Delta s} - 1 \right) + \frac{\mathrm{M}_s}{X_s \mathrm{L}_s^2} \left(e^{\mathrm{L}_s \Delta s} - 1 - \mathrm{L}_s \Delta s \right),$$

$$\mathrm{B}(X_s) = \sigma\sqrt{\left(e^{2\mathrm{L}_s\Delta s} - 1\right)/(2\mathrm{L}_s)},$$

with

$$\mathrm{L}_s = b_x(s, X_s) \quad \text{and} \quad \mathrm{M}_s = \sigma^2/2 \cdot b_{xx}(s, X_s) + b_t(s, X_s).$$

4.2.3. Approximated likelihood method

In this section, we present method that differ from the previous in that they do not try to approximate the paths of a diffusion but instead provide direct approximation of the likelihood.

Kessler method. The main idea is to use a higher-order Ito-Taylor expansion to approximate the mean and variance.

Consider the SDE

$$dX_t = b(X_t, \theta)dt + \sigma(X_t, \theta)dW_t, \quad t \geq 0, \quad X_0 = x_0. \tag{35}$$

The transition density of $X_{t+\Delta t}$ given X_t by Kessler method is Gaussian with mean and variance:

$$\mathrm{m}_x = x + b(t,x)\Delta t + \left(b(t,x)b_x(t,x) + \frac{1}{2}\sigma^2(t,x)b_{xx}(t,x)\right)\frac{(\Delta t)^2}{2},$$

$$\mathrm{V}_x = x^2 + (2b(t,x)x + \sigma^2(t,x))\Delta t +$$

$$+\Bigg(2b(t,x)\left(b_x(t,x)x + b(t,x) + \sigma(t,x)\sigma_x(t,x)\right) + \sigma^2(t,x)\Bigg(b_{xx}(t,x)x +$$

$$+2b_x(t,x) + \sigma_x^2(t,x) + \sigma(t,x)\sigma_{xx}(t,x)\Bigg)\Bigg)\frac{(\Delta t)^2}{2} - \mathrm{m}_x^2.$$

4.3. Practical Problems

1. Evaluate the conditional density of the Ornstein-Uhlenbeck process

$$dX_t = (3 - X_t)dt + 2dW_t, \quad X_0 = 1, \quad N = 1000, \quad \Delta = 1$$

and calculate the maximum likelihood estimation.

2. Find the maximum likelihood estimators numerically for Ornstein-Uhlenbeck process

$$dX_t = -\theta_2 X_t dt + \theta_3 dW_t, \quad \theta = (0, 3, 2), \quad N = 1000, \quad \Delta = 1$$

and compare with explicit estimations (25)-(26).

3. Consider the Chan-Karolyi-Longstaff-Sanders (CKLS) model

$$dX_t = (\theta_1 + \theta_2 X_t)dt + \theta_3 X_t^{\theta_4} dW_t, \qquad X_0 = 2$$

with $\theta_1 = 1$, $\theta_2 = 2$, $\theta_3 = 0.5$, $\theta_4 = 0.3$.

Use the Euler method and

- generate the sample data X_{t_i} with time step $\Delta t = 10^{-4}$,
- estimate drift and diffusion coefficients,
- compute confidence intervals for all parameters in a fitted SDE.

4. Consider the Vasicek model

$$dX_t = \theta_1(\theta_2 - X_t)dt + \theta_3 dW_t, \qquad X_0 = 5$$

with $\theta_1 = 3$, $\theta_2 = 2$ and $\theta_3 = 0.5$.

Use the Ozaki method and

- generate the sample data X_{t_i}, time step $\Delta t = 10^{-2}$,
- estimate drift and diffusion coefficients,
- compute confidence intervals for all parameters in a fitted SDE.

5. Consider the model

$$dX_t = a(t)X_t dt + \theta_2 X_t dW_t, \qquad X_0 = 10$$

with $a(t) = \theta_1 t$, $\theta_1 = -2$, $\theta_2 = 0.2$.

Use the Shoji-Ozaki method and

- generate the sample data X_{t_i} time step $\Delta t = 10^{-3}$,
- estimate drift and diffusion coefficients,
- compute confidence intervals for all parameters in a fitted SDE.

6. Consider the Hull-White (extended Vasicek) model

$$dX_t = a(t)(b(t) - X_t)dt + \sigma(t)dW_t, \qquad X_0 = 2$$

with $a(t) = \theta_1 t$, $b(t) = \theta_2\sqrt{t}$, the volatility depends on time $\sigma(t) = \theta_3 t$. Generate sample data of X_t with time step $\Delta t = 10^{-3}$ and $\theta_1 = 3$, $\theta_2 = 1$ and $\theta_3 = 0.3$, then use the Kessler method and

- estimate drift and diffusion coefficients,
- compute confidence intervals for all parameters in a fitted SDE.

5. Non-parametric estimation

When there is no specific reason to specify a parametric form for either the diffusion or the drift coefficient or both, non-parametric methods help in the identification of the diffusion model.

In this section, we review, without going too much into details, some non-parametric techniques.

The main inference problems are related to invariant density function estimation and/or drift and diffusion coefficients.

Let us consider the ergodic diffusion process X solution to

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \tag{36}$$

where $b(\cdot)$ and $\sigma(\cdot)$ satisfy the usual assumptions of regularity and Assumption 1.3 holds true.

Remain, **Assumption 1.3**. Let (l, r), with $-\infty \leq l \leq r \leq +\infty$, be the state space of the diffusion process X solution to equation (36) and assume that

$$\int_l^r m(x)dx < \infty.$$

Let x^* be an arbitrary point in the state space of X such that

$$\int_{x^*}^r s(x)dx = \int_{x^*}^l s(x)dx = \infty.$$

If one or both of the integrals above are finite, the corresponding boundary is assumed to be *instantaneously reflecting*.

Under Assumption 1.3, the process X is ergodic and has an invariant distribution function. Our primary interest is now the invariant density $\pi(x)$. As in the i.i.d. case, a simple kernel type estimator can be used.

Let K be a non-negative function such that

$$\int K(u)du = 1$$

and K is bounded and twice continuously differentiable on $\mathbb{R}$. K and its derivatives are supposed to be in $L^2(\mathbb{R})$. Such a function K is called a *kernel* of order $r > 1$ if there exists an integer r such that

$$\int_{-\infty}^{+\infty} x^i K(x)dx = 0, \quad i = 1, 2, \ldots, r-1,$$

and

$$\int_{-\infty}^{+\infty} x^r K(x)dx \neq 0, \quad \int_{-\infty}^{+\infty} |x|^r |K(x)|dx < \infty.$$

We assume K to be of order 2 and we further define

$$K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right)$$

and notice that

$$\lim_{h\to\infty} K_h(u) = \delta(u),$$

where δ is the Dirac delta.

5.1. Stationary density estimation

The estimator

$$\hat{\pi}_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right) = \frac{1}{n} \sum_{i=1}^{n} K_{h_n}(x - X_i) \tag{37}$$

is the kernel estimator of $\pi(x)$. Usually the Gaussian kernel

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right) \tag{38}$$

is used, but any other reasonable kernel can be considered without loss in the optimality results.

The more critical choice is known to be the *bandwidth* h_n. The bandwidth h_n is a shrinking sequence with n; i. e., $h_n \to 0$ as $n \to \infty$.

For general m-dimensional densities, the bandwidth is usually chosen according to Scott's rule, which assumes h_n to be proportional to

$$h_n \propto d \cdot n^{-\frac{1}{m+4}},$$

where d is the standard deviation of the time series.

Under the assumption $\lim_{n\to\infty} nh^{4.5} = 0$ and mild regularity conditions on the time series dependence in data (Assumption A3), the kernel (Assumption A4) and bandwidth (Assumption A5) are given in paper [1], the stationary density estimator $\hat{\pi}_n$ behaves as in the i.i.d. setting. In particular, we have

$$\sqrt{nh_n}(\hat{\pi}_n(x) - \pi(x)) \xrightarrow{d} \mathcal{N}\left(0, \pi(x) \int_{-\infty}^{+\infty} K^2(u)du\right).$$

5.2. Local-time and stationary density estimators

A relationship between h_n and Δ_n was established in paper [2]. The result is given in terms of *local-time* estimation also for nonstationary processes.

In this case, the local-time estimator generalizes the stationary density estimator. We consider only diffusion processes without jumps, and hence the local time is defined as

$$L_X(T,x) = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \int_0^T \mathbb{1}_{[x,x+\epsilon)}(X_s) d\langle X, X\rangle_s, \tag{39}$$

where $\langle X, X\rangle_s$ is the quadratic variation process (24) and x is the state space of the process. Remain the Equation (24)

$$\langle X, X\rangle_t = \lim_{n\to\infty} \sum_{k=1}^{2^n} (X_{t\wedge k/2^n} - X_{t\wedge(k-1)/2^n})^2 = \int_0^t \sigma^2(X_s, \theta) ds,$$

where $t \wedge k/2^n = \min(t, k/2^n)$.

The local time is intuitively the amount of time a process sojourns in a neighborhood of x between time 0 and T. We know from Equation (24) that, for continuous diffusion processes

$$d\langle X, X\rangle_t = \sigma^2(X_t) dt.$$

From this it follows that the local time can be transformed into the so-called *chronological local* time, defined as

$$\bar{L}_X(T,x) = \frac{1}{\sigma^2(x)} L_X(T,x). \tag{40}$$

The difference between L_X and $\bar{L}_X$ is that the local time in Equation (39) is the amount of time expressed in time units of the quadratic variation process, while the chronological local time in Equation (40) is expressed in terms of real time units, which is the time we deal with in estimation.

The relationship in Equation (40) is interesting because it is related to the occupation measure of the process X:

$$\eta_A^T = \int_0^T \mathbb{1}_A(X_s) ds = \int_A \bar{L}_X(T,x) dx.$$

Fact 4.1 (Almost Sure Convergence to the Chronological Time [2]). If $h_n \to 0$ and $n \to \infty$ with fixed $T = \bar{T}$ in such a way that

$$\frac{1}{h_n}\sqrt{\Delta_n \log \frac{1}{\Delta_n}} = o(1),$$

then

$$\hat{\bar{L}}_X(T,x) = \frac{\Delta_n}{h_n}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right) \xrightarrow{a.s.} \bar{L}_X(T,x).$$

The result above shows the limiting quantity is a random object, which is not what happens in standard kernel density estimation, and this is not surprising.

The relation with kernel density estimation is given in the next result.

Fact 4.2 [2]. If $h_n \to 0$, $T = n\Delta_n \to \infty$ as $n \to \infty$ such that

$$\frac{T}{h_n}\sqrt{\Delta_n \log \frac{1}{\Delta_n}} = o(1),$$

then

$$\frac{\hat{\bar{L}}_X(T,x)}{T} = \frac{\hat{\bar{L}}_X(T,x)}{n\Delta_n} = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right) = \hat{\pi}_n(x) \xrightarrow{a.s.} \pi(x).$$

Notice that $\hat{\bar{L}}_X(T,x)/T$ is also an estimator of the expected local time.

5.3. Estimation of diffusion and drift coefficients

From the Section 1.5.2, the drift and diffusion coefficients are related to the stationary density $\pi(\cdot)$ via the forward and backward Kolmogorov equations.

Remain, the transition density satisfies the *Kolmogorov forward equation*

$$\frac{\partial}{\partial t}p(t,y|s,x) = -\frac{\partial}{\partial y}b(y)p(t,y|s,x)) + \frac{1}{2}\frac{\partial^2}{\partial y^2}(\sigma^2(y)p(t,y|s,x)) \qquad (41)$$

and *Kolmogorov backward equation*

$$-\frac{\partial}{\partial s}p(t,y|s,x) = b(x)\frac{\partial}{\partial x}p(t,y|s,x)) + \sigma^2(x)\frac{1}{2}\frac{\partial^2}{\partial y^2}p(t,y|s,x). \qquad (42)$$

Letting $t \to \infty$ in the Kolmogorov forward equation (10), it is possible to obtain

$$\frac{d^2}{dx^2}(\sigma^2(x)\pi(x)) = 2\frac{d}{dx}(b(x)\pi(x)), \tag{43}$$

where $\pi(x)$ is the stationary density. Equation (12) establishes a relationship between the drift $b(\cdot)$, the diffusion coefficient $\sigma(\cdot)$, and the invariant density $\pi(\cdot)$. By integrating (12), we obtain

$$b(x) = \frac{1}{2\pi(x)}\frac{d}{dx}(\sigma^2(x)\pi(x)),$$

and integrating again, one gets

$$\sigma^2(x) = \frac{2}{\pi(x)}\int_0^x b(u)\pi(u)du.$$

Given as the estimator the kernel estimator (37) of $\pi(\cdot)$, if $b(\cdot)$ is known or has a parametric form for which consistent estimators for the parameters exist, it is possible to use the estimator

$$\hat{\sigma}_n^2(x) = \frac{2}{\hat{\pi}_n(x)}\int_0^x b(u)\hat{\pi}_n(u)du,$$

where $b(x)$ can eventually be replaced by $b(x;\hat{\theta}_n)$ if it has the parametric form $b = b(x,\theta)$, where $\hat{\theta_n}$ is a $\sqrt{n}$-consistent estimator of θ.

It is possible to estimate the diffusion coefficient directly without information on the drift b or by nonparametric estimation of the drift. This approach requires a high-frequency asymptotic. We start with the Florens-Zmirou estimator

$$\hat{\sigma}_n^2(x) = \frac{\sum_{i=0}^{n-1} K\left(\frac{x-X_i}{h_n}\right)(X_{i+1}-X_i)^2}{\Delta_n \sum_{i=0}^{n-1} K\left(\frac{x-X_i}{h_n}\right)}, \tag{44}$$

The Florens-Zmirou estimator (44) require $\Delta_n \to 0$ at some proper rate in order to have consistency and asymptotic normality. In the same way, a non-parametric drift estimator can be obtained as follows:

$$\hat{b}_n(x) = \frac{\sum_{i=0}^{n-1} K\left(\frac{x-X_i}{h_n}\right)(X_{i+1}-X_i)}{\Delta_n \sum_{i=0}^{n-1} K\left(\frac{x-X_i}{h_n}\right)}. \tag{45}$$

Mixing sequence. There has been much research on stochastic models that have a well defined, specific structure – for example, Markov chains, Gaussian processes, or linear models, including ARMA (autoregressive – moving average) models. However, it became clear in the middle of the last century that there was a need for a theory of statistical inference (e.g. central limit theory) that could be used in the analysis of time series that did not seem to "fit" any such specific structure but which did seem to have some "asymptotic independence" properties. That motivated the development of a broad theory of "strong mixing conditions" to handle such situations. For details one can read the paper [3].

The core idea is that there are processes where as time elapses, future events can be regarded as "increasingly independent" from the past events. The different definitions of mixing refer to different formalisations of "increasingly independent".

Assumption 4.4.

$$\lim_{x\to 0} \sigma(x)\pi(x) = 0 \quad \text{or} \quad \lim_{x\to\infty} \sigma(x)\pi(x) = 0$$

and

$$\lim_{x\to 0} \left| \frac{\sigma(x)}{2b(x) - \sigma(x)\sigma_x(x)} \right| < \infty \quad \text{or} \quad \lim_{x\to\infty} \left| \frac{\sigma(x)}{2b(x) - \sigma(x)\sigma_x(x)} \right| < \infty.$$

The conditions in Assumption 4.4 imply geometric ergodicity, which in turn implies the following mixing condition on the observed data [1].

Assumption 4.5. The observed data X_i, $i = 1, 2, \ldots, n$, is a strictly stationary β-mixing sequence satisfying $k^\delta \beta_k \to 0$ and $k \to \infty$ for some $\delta > 1$.

Assumption 4.6. As $n \to \infty$ and $h_n \to 0$, we have

$$\sqrt{nh_n^{2r+1}} \to 0 \quad \textit{and} \quad nh_n \to \infty \quad \textit{and} \quad nh_n^3 \to \infty,$$

where r is the order of the kernel $K(\cdot)$.

Fact 4.3 [1]. Suppose Assumptions 4.5 and 4.6 hold true and

$$\sigma^2(x) > 0.$$

Assume that the drift $b(x)$ is known (or $b(x, \theta)$ unknown up to a finite dimensional parameter θ) and $\sigma(x)$ is differentiable with continuous derivatives on $(0, \infty)$ of order greater than or equal to 2. Then

$$\sqrt{nh_n}(\hat{\sigma}_n^2(x) - \sigma^2(x)) \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma^4(x)}{\pi(x)} \int_{-\infty}^{+\infty} K^2(u)du\right).$$

The result above is interesting in real-life applications only when the drift $b(x)$ has at least a parametric form $b(x) = b(x, \theta)$, but it is hardly reasonable to assume that $\sigma(x)$ is unknown and $b(x)$ is completely known.

5.4. Practical Problems

1. Simulate a Cox-Ingersoll-Ross model

$$dX_t = (\theta_1 - \theta_2 X_t)dt + \theta_3\sqrt{X_t}dW_t$$

with $\theta = (6, 2, 1)$, and choose the bandwidth according to Scott's rule and implement the non-parametric kernel estimation for stationary density of the CIR model. Plot in a graph the true stationary density against the estimated one for

- $n = 1000$, $\Delta_n = 1$,
- $n = 15,000$, $\Delta_n = 0.01$.

2. Implement a non-parametric drift (45) and diffusion (44) estimators for the model from previous item and plot two graphs. Use the

- uniform kernel $K(u) = 1/2$,
- Gaussian kernel (38),
- Epanechnikov (parabolic) kernel $K(u) = \frac{3}{4}(1 - u^2)$.

6. Variance reduction techniques

The method for increasing the efficiency of Monte Carlo simulation draw on two broad strategies for reduction variance:

1) taking advantage of tractable features of a model to adjust or correct simulation output;

2) reduction the variability in simulations inputs.

Let us enumerate some of the well-known methods: preferential sampling, control variates, antithetic methods, Latin hypercube sampling, moment matching methods, and importance sampling [5]. The schematic comparison of variance reduction techniques one can see in Figure 11.

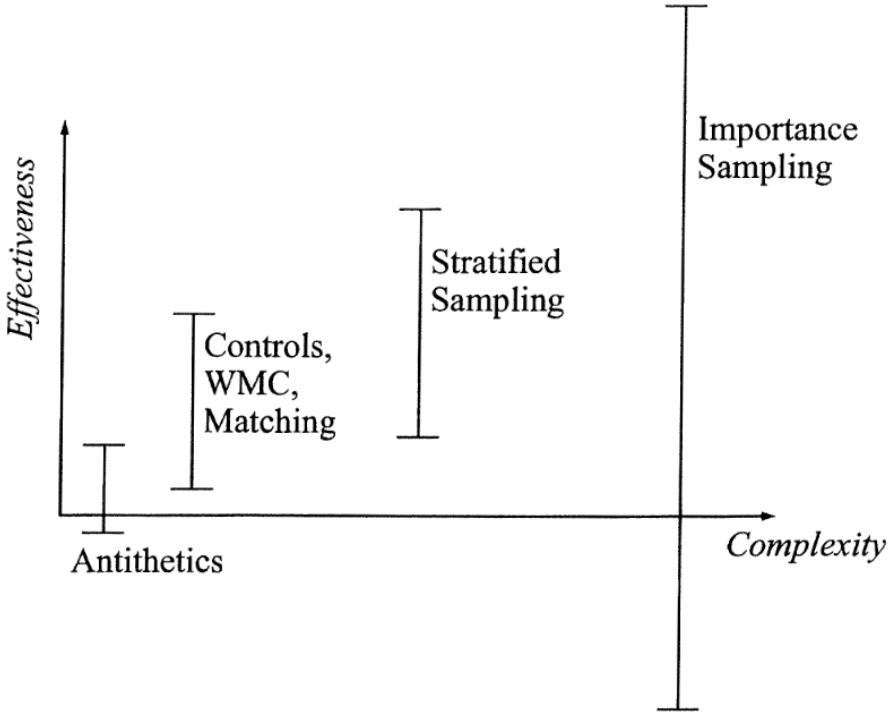

Figure 11. Schematic comparison of variance reduction techniques [5]

6.1. 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

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

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

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

and

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

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 $Var\tilde{g}(Y) < Varg(X)$, then we have reached our goal. But let us calculate $Var\tilde{g}(Y)$,

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

If $g(\cdot)$ is strictly positive, by choosing $h(x) = g(x)f(x)/\mathbb{E}g(X)$, we obtain $Var\tilde{g}(Y) = 0$, which is nice only in theory because, of course, we do not 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

$$h(x) = \frac{\tilde{h}(x)}{\int_{\mathbb{R}} \tilde{h}(x)dx}.$$

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.

Example. Calculate the price of a put option in the Black and Scholes framework

$$p = \mathbb{E}\{\max(0, K - e^{\beta x})\},$$

K and β constants, and $X \sim \mathcal{N}(0, 1)$.

The explicit solution, which is known, reads as

$$\mathbb{E}\{\max(0, K - e^{\beta x})\} = K\Phi\left(\frac{\log K}{\beta}\right) - e^{\frac{1}{2}\beta^2}\Phi\left(\frac{\log K}{\beta} - \beta\right),$$

$\Phi(\cdot)$ is the cumulative distribution function of the standard Gaussian law. The true value, in the case $K = \beta = 1$, is $p(x) = 0.2384217$.

6.2. 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.

6.3. 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

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

with $X \sim \mathcal{U}(0,1)$. The transformation $x \mapsto 1 - x$ leaves the measure unchanged, i. e. $1 - X \sim \mathcal{U}(0,1)$, and we can rewrite

$$I = \frac{1}{2}\int_0^1 (g(x) + g(1-x))dx =$$

$$= \frac{1}{2}\mathbb{E}(g(X) + g(1-X)) = \frac{1}{2}\mathbb{E}(g(X) + g(h(X))).$$

Therefore, we have a variance reduction if

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

which is equivalent to saying that $Cov(g(X), g(h(X))) < 0$. If $h(x)$ is a monotonic function of x (as in the example above), this is always the case.

6.4. Importance sampling

The idea of importance sampling is explained best in case of estimating the probability of an event A. The underlying sample space is $(\Omega, \mathcal{F})$ for

which $A \in \mathcal{F}$, and the probability measure P on this space is given by the specific simulation model. In a simulation experiment for estimating $P(A)$, the Monte Carlo sampling estimator would be

$$\hat{\ell}_N = \sum_{i=1}^{N} \mathbb{1}_A^{(i)},$$

where $\mathbb{1}_A^{(1)}, \ldots, \mathbb{1}_A^{(N)}$ are i.i.d. indicator functions of event A generated under P. On average in only one out of $1/P(A)$ generated samples the event A occurs, and thus for rare events (where $P(A)$ is extremely small) this procedure fails. Suppose that there is an alternative probability measure P^* on the same $(\Omega, \mathcal{F})$ such that A occurs much more often, and P is absolutely continuous with respect to P^*, meaning

$$\forall F \in \mathcal{F} : P(F) > 0 \implies P^*(F) > 0.$$

Then according to the Radon-Nikodym theorem, it holds that there is a measurable function L on Ω such that

$$\int_F dP = \int_F L dP^* \quad \forall F \in \mathcal{F}.$$

The function L is called *likelihood ratio* and usually written as

$$L = dP/dP^*$$

the alternative probability measure P^* is said to be the importance sampling probability measure, or the change of measure. Thus, by weighting the occurrence $\mathbb{1}_A$ of event A with the associated likelihood ratio, simulation under the change of measure yields an unbiased importance sampling estimator

$$\hat{\ell}_N^* = \sum_{i=1}^{N} L^{(i)} \mathbb{1}_A^{(i)}.$$

6.5. Practical Problems

In the Black-Scholes framework 1. Generate $X \sim \mathcal{N}(0, 1)$ and evaluate $\mathbb{E}g(Y)$, where

$$Y = g(X) = \exp\{\beta X\}$$

with the Monte Carlo method using $100,000$ replications and construct 95 % confidence intervals using the true standard deviation σ and the estimated standard error.

2. Plot a trajectory and add to a graph: (a) target value, (b) upper and lower limits of the Monte Carlo 95 % confidence interval.

3. Evaluate of the price of a *put* and *call* options (a) without and (b) with applying in Monte Carlo simulation

- the *preferential sampling* technique;
- the *control variables* technique;
- the *antithetic sampling* technique;
- the *importance sampling* technique.

4. Provide numerical (table) and graphical comparison for mean; and variance reduction for a *put* and *call* options (a) without and (b) with applying different variance reduction techniques. Calculate 95 % confidence interval for different $n = 100, 1000, 10000$.

7. Model identification

Consider a diffusion process solution to the SDE

$$dX_t = b(X_t, \alpha)dt + \sigma(X_t, \beta)dW_t \tag{46}$$

with some initial condition $X_0 = x_0$, where the parameter $\theta = (\alpha, \beta)$ is such that $\theta \in \Theta_\alpha \times \Theta_\beta = \Theta, \theta_\alpha \subset \mathbb{R}^p, \Theta_\beta \subset \mathbb{R}^q$, and Θ convex.

As usual, $b(\cdot, \cdot)$ and $\sigma(\cdot, \cdot)$ are two known (up to α and β) regular functions such that a solution of Equation (46) exists. The process X_t is also assumed to be ergodic for every θ with invariant distribution $\pi_\theta(\cdot)$.

Observations are assumed to be equally spaced and such that the discretization step Δ_n shrinks as the number of observations increases:

$$\Delta_n \to 0, \quad n\Delta_n = T \to \infty$$

under the rapidly increasing design; i. e.,

$$n\Delta_n^2 \to 0 \quad \text{as} \quad n \to \infty.$$

The aim is to try to identify the underlying continuous model on the basis of discrete observations using an information criterion that is a function of the dimension of the parameter space.

The Akaike information criterion (AIC) dates back to 1973 and is constructed in such a way that it searches the best model embedded in a wider class of models. It is a likelihood-based method that, roughly speaking, is defined as minus twice the log-likelihood plus twice the dimension of the parameter space. So it is based on the idea that an overspecified model (high dimension of parameter space = too many parameters in the SDE) is less valuable than a correctly specified one. Given a class of competing models, the best model is the one that minimizes the AIC criterion.

The main assumption is that the true model is currently included among the competing ones; otherwise there is a misspecification problem.

Let $\ell_n(\theta)$ be the log-likelihood of the process. Then the AIC statistic is defined as

$$AIC = -2\ell_n\left(\hat{\theta}_n^{(ML)}\right) + 2\dim(\Theta),$$

where $\hat{\theta}_n^{(ML)}$ is the true maximum likelihood estimator. Since, as we have seen, there are only a few models for which the explicit expression of

$l_n(\theta)$ is known, in most of the cases one of the approximated likelihood methods presented early is needed.

However, since the transition density $p(\cdot)$ of the diffusion process X does not generally have an explicit form, we can not directly obtain the log likelihood function ℓ_n and the maximum likelihood estimator $\hat{\theta}_n^{(ML)}$. That is why we need to obtain both:

1) an approximation of the log-likelihood function ℓ_n;

2) an asymptotically efficient estimator $\hat{\theta}_n$.

in order to construct AIC type of information criteria for diffusion processes.

7.1. AIC type of information criteria for diffusion processes

In order to obtain AIC type of information criteria for diffusion processes, we consider two kinds of functions [16]. One is an approximate log likelihood function unbased on a result of Dacunha-Castelle and Florens-Zmirou (1986). The other is a contrast (discrepancy) function g_n based on a locally Gaussian approximation. The approximate log-likelihood function u_n is used as an approximation of the log likelihood function and an asymptotically efficient estimator is derived from the contrast function g_n. The essential point is that in general we can not use the contrast function g_n as an approximation of the log-likelihood function.

The approximate log-likelihood function

$$u_n(\theta) = \sum_{k=1}^{n} u(\Delta_n, X_{i-1}, X_i, \theta), \tag{47}$$

where

$$u(t, x, y, \theta) = -\frac{1}{2}\log(2\pi t) - \log\sigma(y, \beta) - \\ -\frac{S^2(x, y, \beta)}{2t} + H(x, y, \theta) + t\tilde{g}(x, y, \theta),$$

with

$$S(x, y, \theta) = \int_x^y \frac{1}{\sigma(u, \beta)} du,$$

$$H(x, y, \theta) = \int_x^y \frac{B(u, \theta)}{\sigma(u, \beta)} du,$$

$$\tilde{g}(x, y, \theta) = -\frac{1}{2}\left(C(x, \theta) + C(y, \theta) + \frac{1}{3}B(x, \theta)B(y, \theta)\right),$$

$$C(x,\theta) = \frac{1}{2}B^2(x,\theta) + \frac{1}{2}B_x(x,\theta)\sigma(x,\beta),$$
$$B(x,\theta) = \frac{b(x,\alpha)}{\sigma(x,\beta)} - \frac{1}{2}\sigma_x(x,\beta).$$

Moreover, the following *contrast* function is defined in order to obtain an asymptotically efficient estimator to plug into the AIC statistic:

$$g_n(\theta) = \sum_{k=1}^{n} g(\Delta_n, X_{i-1}, X_i, \theta),$$

where

$$g(t,x,y,\theta) = -\frac{1}{2}\log(2\pi t) - \log\sigma(x,\beta) - \frac{(y-x-tb(x,\alpha))^2}{2t\sigma^2(x,\beta)}.$$

The maximum contrast estimator is then defined as

$$\theta_n^{(C)} = \arg\sup_{\theta} g_n(\theta).$$

Further, define the functions

$$s(x,\beta) = \int_0^x \frac{1}{\sigma(u,\beta)}du,$$
$$\tilde{B}(x,\theta) = B(s^{-1}(x,\beta),\theta),$$
$$\tilde{h}(x,\theta) = \tilde{B}^2(x,\theta) + \tilde{B}_x(x,\theta),$$

and denote by $\theta_0 = (\alpha_0, \beta_0)$ the true value of the parameter θ. We now introduce the set of assumptions that should be verified by the model in order to obtain the good properties for the AIC statistic.

Assumption 7.1. The coefficients are such that:

1) equation (46) has a unique strong solution on $[0,T]$;

2) $\inf_{x,\beta}\sigma^2(x,\beta) > 0$;

3) X is ergodic for every θ with invariant law μ_θ and all moments of μ_θ are finite;

4) for all $m \geq 0$ and for all θ, $\sup_t \mathbb{E}_\theta|X_t|^m < \infty$;

5) for every θ, $b(x,\alpha)$ and $\sigma(x,\beta)$ are twice continuously differentiable with respect to x and the derivatives are of polynomial growth in x uniformly in θ;

6) $b(x, \alpha)$ and $\sigma(x, \beta)$ and all their partial derivatives with respect to x up to order 2 are three times differentiable with respect to θ for all x and are of polynomial growth in x, uniformly in θ.

Assumption 7.2. The function $\tilde{h}(\cdot)$ is such that:

1) $\tilde{h}(x, \theta) = O(|x|^2)$ as $x \to \infty$;

2) $\sup_\theta \sup_x |\tilde{h}^3(x, \theta)| \leq M < \infty$;

3) there exists $\gamma > 0$ such that for every θ and $j = 1, 2$,

$$|\tilde{B}^j(x, \theta)| = O(|\tilde{B}(x, \theta)|^\gamma) \quad \text{as} \quad |x| \to \infty.$$

Assumption 7.3. Almost surely with respect to $\pi_\theta(\cdot)$ and for all x, $b(x, \alpha) = b(x, \alpha_0)$ implies $\alpha = \alpha_0$ and $\sigma(x, \beta) = \sigma(x, \beta_0)$ implies $\beta = \beta_0$.

These assumptions imply the existence of a good estimator and the validity of the approximation of the log-likelihood function, but they also imply that the estimator $\hat{\theta}_n^{(C)}$ is asymptotically efficient [16] and that the following version of the AIC, which will be used in practice, converges to the true AIC statistic (based on the true likelihood and calculated at the true maximum likelihood estimator):

$$AIC = -2u_n\left(\hat{\theta}_n^{(C)}\right) + 2\dim(\Theta).$$

The same result holds true if $\hat{\theta}_n^{(C)}$ is replaced by the approximated maximum likelihood estimator, say $\hat{\theta}_n^{(AML)}$, obtained by direct maximization of (47),

$$AIC = -2u_n\left(\hat{\theta}_n^{(AML)}\right) + 2\dim(\Theta).$$

In most of the cases, though, the estimator $\hat{\theta}_n^{(C)}$ is easier to obtain numerically than $\hat{\theta}_n^{(AML)}$ because g_n is simpler than u_n.

Conversely, it is not a good idea to use g_n instead of u_n to build the AIC statistic because the simple Gaussian contrast is in general too rough an approximation of the conditional density as we discussed early.

Numerical evidence about the discrepancy of g_n and u_n from the true likelihood was shown in paper [16].

7.2. Practical Problems

1. **Constant Maturity Interest Rates.** Fit 2-3 models to a sample of historical interest rates over the period Jan 1, 1962 to till now. Plot the historical daily time series and three time-discretization benchmarks:

- Kessler method,
- Shoji-Ozaki method,
- Euler method

using MLE approach. The parameter estimates display in Table for each method. Calculate the AIC and select the best model. Compare your results with paper [9].

Dataset: Market Yield on U.S. Treasury Securities at 10-Year Constant Maturity, Quoted on an Investment Basis (DGS10),
https://fred.stlouisfed.org/series/DGS10

2. **USD/Euro Exchange Rates.** Fit a time series of USD/EUR exchange rates over the period Jan 1, 1999 to till now (daily observations) by 2-3 models. Plot the historical daily time series and three time-discretization benchmarks:

- Kessler method,
- Shoji-Ozaki method,
- Euler method

using MLE approach. The parameter estimates display in Table for each method. Calculate the AIC and select the best model. Compare your results with paper [9].

Dataset: U.S./Euro Foreign Exchange Rate [DEXUSEU], retrieved from FRED, Federal Reserve Bank of St. Louis;
https://fred.stlouisfed.org/series/DEXUSEU

3. **U.S. Interest Rates** Fit a time series of IRates over the period from 06/1964 to 12/1989 (monthly data) by 2-3 models. Plot the historical daily time series and three time-discretization benchmarks:

- Kessler method,
- Shoji-Ozaki method,
- Euler method

using MLE approach. The parameter estimates display in Table for each method. Calculate the AIC and select the best model. Compare your results with paper [4].

Dataset: In order to obtain the data one can open browser, type the address https://rdrr.io/snippets/ and run the code in R language:

```
library(Ecdat)
data(Irates)
rates = Irates[, "r1"]; rates
plot(rates)
```

8. Compensation Problems

1. **Stochastic processes.** Let

$$X(t) = e^{-\kappa t}X_0 + \theta(1 - e^{-\kappa t}) + \sigma e^{-\kappa t}\int_0^t e^{\kappa s}\sqrt{X(t)}dW(s)$$

be extract solution for CIR model. Compute the moments of the extract solution $\mathbb{E}\{X^n(t)\}$ using the general formula [8]

$$\mathbb{E}\{X^n(t)\} = \sum_{j=0}^{[n/2]} A^{n-2j}(t)B^{2j}(t)\left(\frac{1}{2\kappa}(e^{2\kappa t} - 1)\right)^{2j}, \quad \forall n \in \mathbb{N},$$

where $A(t) = e^{-\kappa t X_0} + \theta(1 - e^{-\kappa t})$, $B(t) = \sigma e^{-\kappa t}$, $[n/2]$ denotes the greatest integer less than or equal to $\frac{n}{2}$ and $\kappa = 1$, $\theta = 0.45$, $\sigma = 1$.

2. **Numerical methods.** Define the default parameters by himself and approximate a solution by predictor-corrector method and KPS method for

- CKLS process,
- Feller Root process,
- BM process,
- Hyperbolic process,
- Jacobi process,
- Modified CIR process,
- Pearson process,
- Radial OU process.

3. **Numerical methods.** Compare a speed of convergence of Euler and Milstein schemes to the true value as a function of the discretization step $\Delta t = 1/N$. Plot in graph.

4. **Numerical methods.** Define the default parameters by himself and approximate conditional law of a diffusion process

- Ornstein-Uhlenbeck process,
- Cox-Ingersoll-Ross process,
- Black-Scholes-Merton (geometric Brownian motion model)

by Taylor, Heun, Improved 3-stage Runge-Kutta scheme,
https://hal.archives-ouvertes.fr/hal-00629841/document.

5. **Parametric Estimation**. Take one process and estimate the parameters by pseudo-likelihood method (Euler, Elerian, and Kessler method),

then (a) get the confidence intervals for parameters, (b) get the variance-covariance matrix.

6. **Non-parametric Estimation**. Implement a non-parametric diffusion (44) and drift (45) and estimators for the Ornstein-Uhlenbeck model, $X(0) = 10$, $\theta = (0, 3, 2)$ and plot two graphs. Use the

- uniform kernel $K(u) = 1/2$,
- Gaussian kernel (38),
- Epanechnikov (parabolic) kernel $K(u) = \frac{3}{4}(1 - u^2)$.

7. **Variance reduction techniques**. Evaluate of the price of a *put* and *call* options (a) without and (b) with applying in Monte Carlo simulation [5]: (a) Latin hypercube sampling; (b) moment matching method.

References

[1] Ait-Sahalia Y. Nonparametric Pricing of Interest Rate Derivative Securities / Y. Ait-Sahalia // Econometrica. – 1996. – Vol. 64, № 3. – P. 527-560.

[2] Bandi F.B. Fully Nonparametric Estimation of Scalar Diffusion Models / F.B. Bandi, P. Phillips // Econometrica. – 2003. – Vol. 71, № 3. – P. 241—283.

[3] Bradley R. C. Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions / R. C. Bradley // Probability Surveys. – 2005. – Vol. 2. – P. 107—144.

[4] Brouste A. The YUIMA Project: A Computational Framework for Simulation and Inference of Stochastic Differential Equations / A. Brouste // Journal of Statistical Software. – 2014. – Vol. 57, № 4. – P. 1–51.

[5] Glasserman P. Monte Carlo Methods in Financial Engineering / P. Glasserman. – New York : Springer, 2003. – 596 p.

[6] Guidoum A.C. Performing Parallel Monte Carlo and Moment Equations Methods for Ito and Stratonovich Stochastic Differential Systems: R Package Sim.DiffProc /A.C. Guidoum, K. Boukhetala // Journal of Statistical Software. – 2020. – Vol. 96, № 1. – P. 1–34.

[7] Iacus S.M. Simulation and Inference for Stochastic Differential Equations with R Examples / S.M. Iacus. – New York : Springer, 2008. – 284 p.

[8] Jafari M.A. The Moments for Solution of the Cox-Ingersoll-Ross Interest Rate Model / M.A. Jafari, S. Abbasian // Journal of Finance and Economics. – 2017. – Vol. 5, № 1. – P. 34—37.

[9] Kirkby J.L. Maximum likelihood estimation of diffusions by continuous time Markov chain / J.L. Kirkby // Computational Statistics & Data Analysis. – 2022. – Vol. 168. – P. 107408.

[10] Mubayi A. Studying Complexity and Risk Through Stochastic Population Dynamics: Persistence, Resonance, and Extinction in Ecosys-

tems. Integrated Population Biology and Modeling // Handbook of Statistics / A. Mubayi. – Amsterdam : Elsevier, 2019. – P. 157—193.

[11] Nafidi A. Inference in the stochastic Cox-Ingersoll-Ross diffusion process with continuous sampling: Computational aspects and simulation / A. Nafidi, E.A. Abdenbi. 2021. URL: https://arxiv.org/abs/2103.15678.

[12] Oksendal B. Ito Integrals. Stochastic Differential Equations: An Introduction with Applications / B. Oksendal. – Berlin, Heidelberg : Springer, 2003. – P. 21—42.

[13] Pergamenshchikov S.M. Stochastic Modelling for the Financial Markets. Lecture notes / S.M. Pergamenshchikov, E.A. Pchelintsev. – Tomsk : TSU Publishing House, 2007. – 40 p.

[14] Forman J.L. The Pearson Diffusions: A Class of Statistically Tractable Diffusion Processes / J.L. Forman, M. Sorensen // Social Science Research Network. – 2007.– Article ID 1150110. – P. 1–34.

[15] Tang C.Y. Parameter estimation and bias correction for diffusion processes, and a nonparametric approach to census population size estimation / C.Y. Tang, S.X. Chen // Journal of Econometrics. – 2009. – Vol. 149. – P. 65-81.

[16] Uchida M. AIC for ergodic diffusion processes from discrete observations / M. Uchida, N. Yoshida // Semantic Scholar. – 2005. – Corpus ID: 125037954.

[17] Гихман И.И. Введение в теорию случайных процессов / И.И. Гихман, А.В. Скороход. – М. : Наука, 1977. – 568 с.

[18] Гихман И.И. Теория случайных процессов. В 3-х томах / И.И. Гихман, А.В. Скороход. – М. : Наука, Т. 1, 1971. – 548 с.; Т. 2, 1973. – 639 с.; Т. 3, 1975. – 496 с.

[19] Гихман И.И. Стохастические дифференциальные уравнения / И.И. Гихман, А.В. Скороход. – Киев : Наукова думка, 1968. – 348 с.

[20] Икеда Н. Стохастические дифференциальные уравнения и диффузионные процессы / Н. Икеда, С. Ватанабе. – М. : Наука, 1986. – 445 с.

[21] Оксендаль Б. Стохастические дифференциальные уравнения. Введение в теорию и приложения / Б. Оксендаль. – М. : Мир, АСТ, 2003. – 408 с.

[22] Булинский А.В. Теория случайных процессов / А.В. Булинский, А.Н. Ширяев. – М. : Физматлит, 2003.

[23] Мильштейн Г.Н. Численное интегрирование стохастических дифференциальных уравнений / Г.Н. Мильштейн. – Свердловск : Изд-во Урал. ун-та, 1988. – 224 с.

[24] Белопольская Я.И. Стохастические дифференциальные уравнения. Приложения к задачам математической физики и финансовой математики / Я.И. Белопольская. – М. : Изд-во Лань, 2021. – 308 с.

[25] Лекции по случайным процессам / А.В. Гасников, Э.А. Горбунов, С.А. Гуз и др. – М. : URSS, 2022. – 304 с.

[26] Крицкий О.Л. Стохастические дифференциальные уравнения : монография / О.Л. Крицкий ; Томский политехнический университет. – Томск : Изд-во Томского политехнического университета, 2023. – 133 с.

[27] Медведев, Г. А. Временная структура доходности в диффузионных моделях процентных ставок / Г. А. Медведев ; – Минск : БГУ, 2018. – 203 с.

Educational Edition

Национальный исследовательский
Томский политехнический университет

СТАСТИТИКА СЛУЧАЙНЫХ ПРОЦЕССОВ

Учебное пособие
Издательство Томского политехнического университета, 2023
На английском языке

Составители
СЕМЕНОВ Михаил Евгеньевич
ФЕДОРОВ Глеб Владимирович

Published in author's version

Science Editor
Doctor of Physical and Mathematical Sciences, Professor
G.V. Kuznecov

Cover design *A.I. Sidorenko*

Printed in the TPU Publishing House in full accordance with the quality of the given make up page

Signed for the press 18.08.2023. Format 60x84/16. Paper "Snegurochka".
Print CANON. Arbitrary printer's sheet 4,65. Publisher's signature 4,20.
Order 213-23. Size of print run 100.

PUBLISHING HOUSE
TOMSK POLYTECHNIC UNIVERSITY