Lecture notes for Numerik IVc - Numerics for Stochastic Processes, Wintersemester 2012/2013. Instructor: Prof. Carsten Hartmann

Scribe: H. Lie

Course outline

1. **Probability theory**
   (a) Some basics: stochastic processes, conditional probabilities and expectations, Markov chains
   References: [MS05, Kle06]

2. **Stochastic differential equations**
   (a) Brownian motion: properties of the paths, Strong Markov Property
   References: [MS05, Øks03, Arn73]
   (b) Stochastic integrals: Itô integrals, Itô calculus, Itô isometry
   References: [MS05, Øks03, Arn73]
   (c) SDEs: existence and uniqueness of solutions, numerical discretization, applications from physics, biology and finance
   References: [Øks03, Arn73, KP92]
   (d) Misc: Kolmogorov forward and backward PDEs, infinitesimal generators, semigroup theory, stopping times, invariant distributions, Markov Chain Monte Carlo methods for PDEs and SDEs
   References: [Øks03, Arn73, KP92]

3. **Filtering theory (if time permits)**
   (a) Linear filtering: conditional expectation, best approximation, Kalman-Bucy filter for SDEs

   Reference: [Jaz07, Øks03]

4. **Approximation of stochastic processes**
   (a) Spectral theory of Markov chains: infinitesimal generator, metastability, aggregation of Markov chains
   References: [HM05, Sar11]
   (b) Markov jump processes: applications from biology, physiocs and finance, Markov decision processes, control theory
   References: [GHL09]

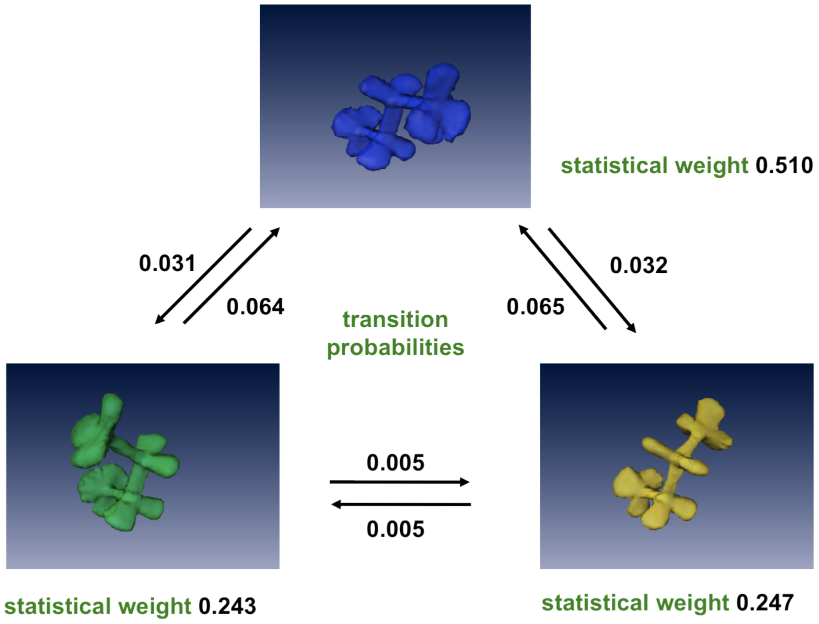

Figure 1: Simulation of a butane molecule and its approximation by a 3-state Markov chain (states in blue, green and yellow; solvent molecules not shown).

1 Day 1, 16.10.2012

1.1 Different levels of modelling

1.1.1 Time-discrete Markov chains

Time index set I is discrete, e.g. $I \subseteq \mathbb{N}$ and state space S is countable or finite, e.g. $S = \{s_1, s_2, s_3\}$ (see Figure 1). Key objects are transition probabilities. For a state space $S = \{1, \dots, n\}$, the transition probabilities p_{ij} satisfy

$$p_{ij} = \mathbb{P}\left(X_{t+1} = j \mid X_t = i\right)$$

and yield a row-stochastic matrix $P = (p_{ij})_{i,j \in S}$.

1.1.2 Markov jump processes

These are time-continuous, discrete state-space Markov chains. Time index set $I \subseteq \mathbb{R}_+$, S discrete. For a fixed time step $h > 0$, the transition probabilities are given by (see Figure 2)

$$\mathbb{P}\left(X_{t+h} = s_j \,|X_t = s_i\right) = h\ell_{ij} + o(h)$$

where $L = (\ell_{ij})_{i,j \in S}$ and P_h are matrices satisfying $P_h = \exp\left(hL\right)$.

Note: the matrix L is row sum zero, i.e. $\sum_j \ell_{ij} = 0$. The waiting times for the Markov chain in any state s_i are exponentially distributed in the sense that

$$\mathbb{P}\left(X_{t+s} = s_i,\ s \in [0, \tau) \,|X_t = s_i\right) = \exp\left(\ell_{ii}\tau\right)$$

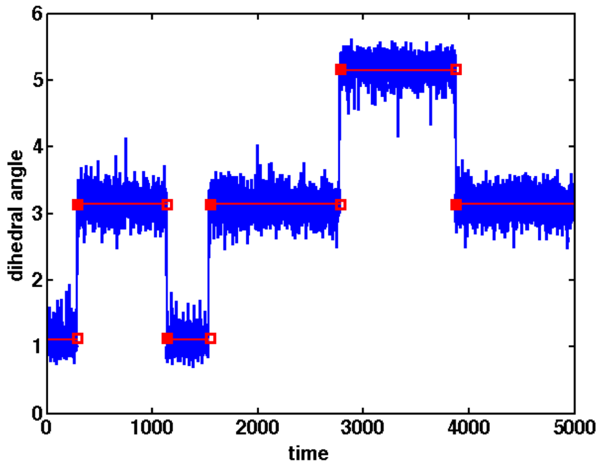

Figure 2: Simulation of butane: typical time series of the central dihedral angle (blue: metastable diffusion process, red: Markov jump process)

and the 'average waiting time' is $-\ell_{ii}$ (by definition of the exponential distribution).

Note: the spectrum of the matrix P_h is contained within the unit disk, i.e. for every eigenvalue λ of P_h, $|\lambda| \leq 1$. This property is a consequence of P_h being row-stochastic, i.e. that $\sum_j P_{h,ij} = 1$. Since $P_h = \exp(hL)$ it follows that

$$\sigma(P_h) \subset D := \left\{x \in \mathbb{R}^2 \mid |x| \leq 1\right\} \Leftrightarrow \sigma(L) \subset \mathbb{C}^- = \{y \in \mathbb{C} \,|\mathrm{Re}(y) \leq 0\}$$

Example 1.1. *Suppose one has a reversible reaction in which one has a large collection of N molecules of the same substance. The molecules can be either in state A or state B and the molecules can change between the two states. Let k^+ denote the rate of the reaction in which molecules change from state A to B and let k^- denote the rate at which molecules change from state B to A.*

For $t > 0$, consider the quantity

$$\mu_i^A(t) := \mathbb{P}\left(\textit{number of molecules in state } A \textit{ at time } t \textit{ is } i\right)$$

where $i = \{0, \ldots, N\}$. One can define quantities $\mu_i^B(t)$ in a similar way, and one can construct balance laws for these quantities, e.g.

$$\frac{d\mu_i^A(t)}{dt} = k^+\mu_{i+1}^A(t) + k^-\mu_{i-1}^A(t) - (k^+ + k^-)\mu_i^A(t).$$

The above balance law can be written in vector notation using a tridiagonal matrix L. By adding an initial condition one can obtain an initial value problem

$$\frac{d\mu^A(t)}{dt} = L^\top\mu^A(t), \quad \mu^A(0) = \mu_0.$$

The solution of the initial value problem above is

$$\mu^A(t) = \mu_0 \exp\left(tL^\top\right).$$

1.1.3 Stochastic differential equations (SDEs)

These are time-continuous, continuous state space Markov chains. SDEs may be considered to be ordinary differential equations (ODEs) with an additional noise term (cf. Figure 2). Let $b : \mathbb{R}^n \to \mathbb{R}^n$ be a smooth vector field and let $x(t)$ be a deterministic dynamical system governed by the vector field $b(\cdot)$. Then $x(t)$ evolves according to

$$\frac{dx}{dt} = b(x), \quad x(0) = x_0. \tag{1}$$

Now let $(B_t)_{t>0}$ be Brownian motion in $\mathbb{R}^d$, and let $(X_t)_{t>0}$ be a dynamical system in $\mathbb{R}^d$ which evolves according to the equation

$$\frac{dX_t}{dt} = b(X_t) + \frac{dB_t}{dt}. \tag{2}$$

The additional term $\frac{dB_t}{dt}$ represents 'noise', or random perturbations from the environment, but is not well-defined because the paths of Brownian motion are nowhere differentiable. Therefore, one sometimes writes

$$dX_t = b(X_t)dt + dB_t,$$

which is shorthand for

$$X_t = X_0 + \int_0^t b(X_t)dt + \int_0^t dB_t.$$

The most common numerical integration method for SDEs is the forward Euler method. If x is a C^1 function of time t, then

$$\left.\frac{dx}{dt}\right|_{t=s} = \lim_{h\to 0} \frac{x(s+h) - x(s)}{h}.$$

The forward Euler method for ODEs of the form (1) is given by

$$X_{t+h} = X_t + hb(X_t)$$

and for SDEs of the form (2) it is given by

$$X_{t+h} = X_t + hb(X_t) + \xi_h$$

where $0 < h \ll 1$ is the integration time step and the noise term ξ in the Euler method for SDEs is modeled by a mean-zero Gaussian random variable.

For stochastic dynamical systems which evolve according to SDEs as in (2), one can consider the probability that a system at some point $x \in \mathbb{R}^d$ will be in a set $A \subset \mathbb{R}^d$ after a short time $h > 0$:

$$\mathbb{P}\left(X_{t+h} \in A \,|X_t = x\right).$$

The associated transition probability density functions of these stochastic dynamical systems are Gaussian because the noise term in (2) is Gaussian.

What has been the generator matrix L in case of a Markov jump process is an infinite-dimensional operator acting on a suitable Banach space. Specifically,

$$Lf(x_0) = \lim_{t\to 0} \frac{\mathbb{E}_{x_0}[f(X_t)] - f(x_0)}{t},$$

provided that the limit exists. Here $f\colon \mathbb{R}^n \to \mathbb{R}$ is any measurable function and $\mathbb{E}_{x_0}[\cdot]$ denotes the expectation over all random paths of X_t satisfying $X_0 = x_0$. L is a second-order differential operator if f is twice differentiable.

2 Day 2, 23.10.2012

Preliminaries from probability theory

Let $(\Omega, \mathcal{E}, \mathbb{P})$ be a probability space, where Ω is a set and $\mathcal{E} \subseteq 2^{\Omega}$ is a σ-field or σ-algebra on Ω, and $\mathbb{P}$ is a probability measure (i.e., $\mathbb{P}$ is a nonnegative, countably additive measure on $(\Omega, \mathcal{E})$ with the property $\mathbb{P}(\Omega) = 1$).

2.1 Conditioning

Let $A \in \mathcal{E}$ be a set of nonzero measure, i.e. $\mathbb{P}(A) > 0$ and define $\mathcal{E}_A$ to be the set of all subsets of A which are elements of $\mathcal{E}$, i.e.

$$\mathcal{E}_A := \{E \subset A \mid E \in \mathcal{E}\}.$$

Definition 2.1 (Conditional probability, part I)**.** *For an event A and an event $E \in \mathcal{E}_A$, the* conditional probability *of E given A is*

$$\mathbb{P}(E|A) := \frac{\mathbb{P}(E \cap A)}{\mathbb{P}(A)}.$$

Remark 2.2. *Think of $\mathbb{P}_A := \mathbb{P}(\cdot\,|A)$ as a probability measure on the measurable space $(A, \mathcal{E}_A)$.*

Given a set $B \in \mathcal{E}$, the *characteristic* or *indicator* function $\chi_B : \Omega \to \{0, 1\}$ satisfies

$$\chi_B(x) = \begin{cases} 1 & x \in B \\ 0 & x \notin B. \end{cases}$$

Definition 2.3 (Conditional expectation, part I)**.** *Let $X : \Omega \to \mathbb{R}$ be a random variable with finite expectation with respect to $\mathbb{P}$. The* conditional expectation *of X given an event A is*

$$\mathbb{E}(X|A) = \frac{\mathbb{E}[X\chi_A]}{\mathbb{P}(A)}.$$

Remark 2.4. *We have*

$$\mathbb{E}(X|A) = \frac{1}{\mathbb{P}(A)} \int_A X d\mathbb{P} = \int X d\mathbb{P}_A.$$

Remark 2.5. *Observe that $\mathbb{P}(E|A) = \mathbb{E}[\chi_E|A]$.*

Up to this point we have only considered the case where A satisfies $\mathbb{P}(A) > 0$. We now consider the general case.

Definition 2.6 (Conditional expectation, part II)**.** *Let $X : \Omega \to \mathbb{R}$ be an integrable random variable with respect to $\mathbb{P}$ and let $\mathcal{F} \subset \mathcal{E}$ be any sub-sigma algebra of $\mathcal{E}$. The* conditional expectation of X given $\mathcal{F}$ *is a random variable $Y := \mathbb{E}[X|\mathcal{F}]$ with the following properties:*

- *Y is measurable with respect to $\mathcal{F}$: $\forall B \in \mathcal{B}(\mathbb{R})$, $Y^{-1}(B) \in \mathcal{F}$.*
- *We have*
$$\int_F X d\mathbb{P} = \int_F Y d\mathbb{P} \quad \forall F \in \mathcal{F}.$$

Remark 2.7. *The second condition in the last definition amounts to the projection property as can be seen by noting that*

$$\mathbb{E}\left[X\chi_F\right] = \int_F X d\mathbb{P} = \int_F Y d\mathbb{P} = \mathbb{E}\left[Y\chi_F\right] = \mathbb{E}\left[\mathbb{E}\left[X|\mathcal{F}\right]\chi_F\right].$$

By the Radon-Nikodym theorem [MS05], the conditional expectation exists and is unique up to $\mathbb{P}$-null sets.

Definition 2.8 (Conditional probability, part II)**.** *Define the conditional probability of an event $E \in \mathcal{E}$ given A by $\mathbb{P}(E|A) := \mathbb{E}\left[\chi_E|A\right]$*

Exercise 2.9. *Let X, $Y : \Omega \to \mathbb{R}$ and scalars $a, b \in \mathbb{R}$. Prove the following properties of the conditional expectation:*

- *(Linearity):*
$$\mathbb{E}\left[aX + bY|A\right] = a\mathbb{E}\left[X|A\right] + b\mathbb{E}\left[Y|A\right].$$
- *(Law of total expectation):*
$$\mathbb{E}\left[X\right] = \mathbb{E}\left[X|A\right] + \mathbb{P}(A) + \mathbb{E}\left[X|A^c\right]\mathbb{P}(A^c)$$
- *(Law of total probability):*
$$\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c).$$

Example 2.10. *The following is a collection of standard examples.*

- *Gaussian random variables: Let X_1, X_2 be jointly Gaussian with distribution $N(\mu, \Sigma)$, where*
$$\mu = \begin{pmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_2] \end{pmatrix}, \quad \Sigma = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$
such that Σ is positive definite. The density of the distribution is
$$\rho(x) = \frac{1}{\sqrt{det(2\pi\Sigma)}} \exp\left[-\frac{1}{2}(x-\mu)^\top \Sigma (x-\mu)\right]$$
(Ex.: Compute the distribution of X_1 given that $X_2 = a$ for some $a \in \mathbb{R}$.)
- *(Conditioning as coarse-graining): Let $Z = \{Z_i\}_{i=1}^M$ be a partition of Ω, i.e. $\Omega = \cup_{i=1}^M Z_i$ with $Z_i \cap Z_j = \emptyset$ and define*
$$Y(\omega) = \sum_{i=1}^{M} \mathbb{E}\left[X\ |Z_i\right]\chi_{Z_i}(\omega).$$
Then $Y = \mathbb{E}\left[X|Z\right]$ is a conditional expectation (cf. Figure 3)
- *(Exponential waiting times): exponential waiting times are random variables $T : \Omega \to [0, \infty)$ with the* memoryless property*:*
$$\mathbb{P}\left(T > s + t\ |T > s\right) = \mathbb{P}\left(T > t\right).$$
This property is equivalent to the statement that T has an exponential distribution, i.e. that $\mathbb{P}(T > t) = \exp(-\lambda t)$ for a parameter value $\lambda > 0$.

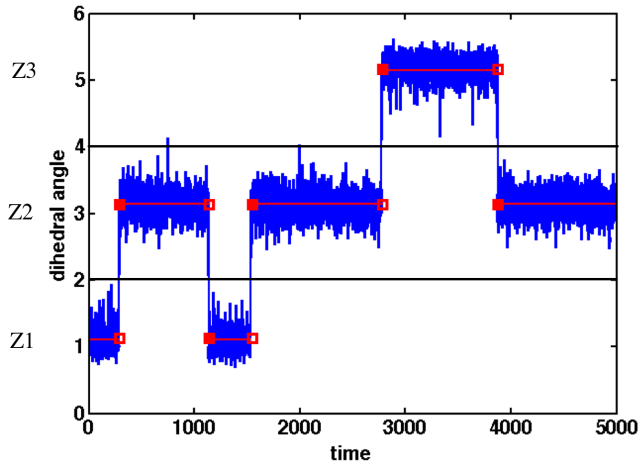

Figure 3: Simulation of butane, coarse-grained into three states Z_1, Z_2, Z_3.

2.2 Stochastic processes

Definition 2.11 (Stochastic process)**.** *A stochastic process* $X = \{X_t\}_{t\in I}$ *is a collection of random variables on a probability space* $(\Omega, \mathcal{E}, \mathbb{P})$ *indexed by a parameter* $t \in I \subseteq [0, \infty)$. *We call* X

- discrete *in time if* $I \subseteq \mathbb{N}_0$
- continuous *in time if* $I = [0, T]$ *for any* $T < \infty$.

How does one define probabilities for X? We provide a basic argument to illustrate the possible difficulties in defining the probability of a stochastic process in an unambiguous way. By definition of a stochastic process, $X_t = X_t(\omega)$ is measurable for every fixed $t \in I$, but if one has an event of the form

$$E = \{\omega \in \Omega \mid X_t(\omega) \in [a, b] \ \forall t \in I\}$$

how does one define the probability of this event? If t is discrete, the σ-additivity of $\mathbb{P}$ saves us, together with the measurability of X_t for every t. If, however, the process is time-continuous, X_t is defined only almost surely (a.s.) and we are free to change X_t on a set A_t with $\mathbb{P}(A_t) = 0$. By this method we can change X_t on $A = \cup_{t\in I} A_t$, The problem now is that $\mathbb{P}(A)$ need not be equal to zero even though $\mathbb{P}(A_t) = 0 \ \forall t \in I$. Furthermore, $\mathbb{P}(E)$ may not be uniquely defined. So what can we do? The solution to the question of how to define probabilities for stochastic processes is to use finite-dimensional distributions or marginals.

Definition 2.12. *(Finite dimensional distributions): Fix* $d \in \mathbb{N}$, $t_1, \ldots, t_d \in I$. *The* finite-dimensional distributions *of the stochastic process* X *for* $(t_1, \ldots, t_d)$ *are defined as*

$$\mu_{t_1,\ldots,t_d}(B) := \mathbb{P}_{(X_{t_k})_{k=1,\ldots,d}}(B) = \mathbb{P}\left(\{\omega \in \Omega \mid (X_{t_1}(\omega), \ldots, X_{t_d}(\omega)) \in B\}\right)$$

for $B \in \mathcal{B}(\mathbb{R}^d)$.

Here and in the following we use the shorthand notation $\mathbb{P}_Y := \mathbb{P} \circ Y^{-1}$ to denote the *push forward* of $\mathbb{P}$ by the random variable Y.

Theorem 2.13. *(Kolmogorov Extension Theorem): Fix $d \in \mathbb{N}$, $t_1, \ldots, t_d \in I$, and let $\mu_{t_1,\ldots,t_d}$ be a* consistent *family of finite-dimensional distributions, i.e.*

- *for any permutation π of $(1, \ldots, d)$,*

$$\mu_{t_1,\ldots,t_d}(B_1 \times \ldots B_d) = \mu_{(t_{\pi(1)},\ldots,t_{\pi(d)}}(B_{\pi(1)} \times \ldots \times B_{\pi(d)})$$

- *For $t_1, \ldots, t_{d+1} \in I$, we have that*

$$\mu_{t_1,\ldots,t_{d+1}}(B_1 \times \ldots B_d \times \mathbb{R}) = \mu_{t_1,\ldots,t_d}(B_1 \times \ldots \times B_d).$$

Then there exists a stochastic process $X = (X_t)_{t \in I}$ with $\mu_{t_1,\ldots t_d}$ as its finite-dimensional distribution.

Remark 2.14. *The Kolmogorov Extension Theorem does not guarantee uniqueness, not even $\mathbb{P}$-a.s. uniqueness, and, as we will see later on, such a kind of uniqueness would not be a desirable property of a stochastic process.*

Definition 2.15. *(Filtration generated by a stochastic process X): Let $\mathcal{F} = \{\mathcal{F}_t\}_{t \in I}$ with $\mathcal{F}_s \subset \mathcal{F}_t$ for $s < t$ be a filtration generated by $\mathcal{F}_t = \sigma(\{X_s \mid s \leq t\})$ is called the filtration generated by X.*

2.3 Markov processes

Definition 2.16. *A stochastic process X is a* Markov process *if*

$$\mathbb{P}(X_{t+s} \in A \mid \mathcal{F}_s) = \mathbb{P}(X_{t+s} \in A \mid X_s) \tag{3}$$

where

$$\mathbb{P}(\cdot | X_s) := \mathbb{P}(\cdot | \sigma(X_s)),$$
$$\mathbb{P}(E | \sigma(X_s)) := \mathbb{E}[\chi_E \mid \sigma(X_s)]$$

for some event E.

Remark 2.17. *If I is discrete, then X is a Markov process if*

$$\mathbb{P}(X_{n+1} \in A \mid X_0 = x_0, \ldots, X_n = x_n) = \mathbb{P}(X_{n+1} \in A \mid X_n = x_n)$$

Example 2.18. *Consider a Markov Chain $(X_t)_{t \in \mathbb{N}_0}$ on a continuous state space $S \subset \mathbb{R}$ and let $\mathcal{S}$ be a σ-algebra on S. Let the evolution of $(X_t)_{t \in \mathbb{N}_0}$ be described by the* transition kernel *$p(\cdot, \cdot) : S \times \mathcal{S} \to [0, 1]$ which gives the single-step transition probabilities:*

$$\begin{aligned} p(x, A) &:= \mathbb{P}(X_{t+1} \in A \mid X_t = x) \\ &= \int_A q(x, y) dy. \end{aligned}$$

In the above, $A \in \mathcal{B}(S)$ and $q = \frac{d\mathbb{P}}{d\lambda}$ is the density of the transition kernel with respect to Lebesgue measure. The transition kernel has the property that

$\forall x \in S$, $p(x,\cdot)$ is a probability measure on $\mathcal{S}$, while for every $A \in \mathcal{S}$, $p(\cdot, A)$ is a measurable function on S.

For a concrete example, consider the Euler-Maruyama discretization of an SDE for a fixed time step Δt,

$$X_{n+1} = X_n + \sqrt{\Delta t}\xi_{n+1}, \quad X_0 = 0,$$

where $(\xi_i)_{i\in\mathbb{N}}$ are independent, identically distributed (i.i.d) Gaussian $\mathcal{N}(0,1)$ random variables. The process $(X_i)_{i\in\mathbb{N}}$ is a Markov Chain on $\mathbb{R}$. The transition kernel $p(x,A)$ has the Gaussian transition density

$$q(x,y) = \frac{1}{\sqrt{2\pi\Delta t}} \exp\left[-\frac{1}{2}\frac{|y-x|^2}{\Delta t}\right].$$

Thus, if $X_n = x$, then the probability that $X_{n+1} \in A \subset \mathbb{R}$ is given by

$$\mathbb{P}\left(X_{n+1} \in A | X_n = x\right) = \int_A q(x,y)dy.$$

3 Day 3, 30.10.2012

Recapitulation:

- A stochastic process $X = (X_t)_{t\in I}$ is a collection of random variables $X_t : \Omega \to \mathbb{R}$ indexed by $t \in I$ (e.g. $I = [0,\infty)$) on some probability space $(\Omega, \mathcal{E}, \mathbb{P})$.
- A filtration $\mathcal{F} := (\mathcal{F}_t)_{t\in I}$ is a collection of increasing sigma-algebras satisfying $\mathcal{F}_t \subset \mathcal{F}_s$ for $t < s$. A stochastic process X is said to be *adapted to* $\mathcal{F}$ if $(X_s)_{s\le t}$ is $\mathcal{F}_t$-measurable. For example, if we define $\mathcal{F}_t := \sigma(X_s \ : \ s \le t)$, then X is adapted to $\mathcal{F}$.
- The probability distribution of a random variable X is given in terms of its finite dimensional distributions.

Example 3.1 (Continued from last week)**.** *Let $I = \mathbb{N}_0$ and consider a sequence $(X_n)_{n\in\mathbb{N}_0}$ of random variables $X_n = X_n^{\Delta t}$ governed by the relation*

$$X_{n+1}^{\Delta t} = X_n^{\Delta t} + \sqrt{\Delta t}\xi_{n+1}, \quad X_0^{\Delta t} = 0 a.s. \tag{4}$$

where $\Delta t > 0$, and $(\xi_k)_{k\in\mathbb{N}_0}$ are i.i.d. random variables with $\mathbb{E}[\xi_k] = 0$ and $\mathbb{E}\left[\xi_k^2\right] = 1$ (not necessarily Gaussian). To obtain a continuous-time stochastic process, the values of the stochastic process on non-integer time values may be obtained by linear interpolation (cf. Figure 4 below). We want to consider the limiting behaviour of the stochastic process in the limit as Δt goes to zero. Set $\Delta t = t/N$ for a fixed terminal time $t < \infty$ and let $N \to \infty$ ($\Delta t \to 0$). Then, by the central limit theorem,

$$X_N^{\Delta t} = \sqrt{\frac{t}{N}} \sum_{k=1}^{N} \xi_k \rightharpoonup \sqrt{t}Z \tag{5}$$

where $Z \sim \mathcal{N}(0,1)$, and "$\rightharpoonup$" means "convergence in distribution", i.e., weak convergence of the induced probability measure; equivalently, the limiting random variable is distributed according to $\mathcal{N}(0,t)$. In other words the limiting

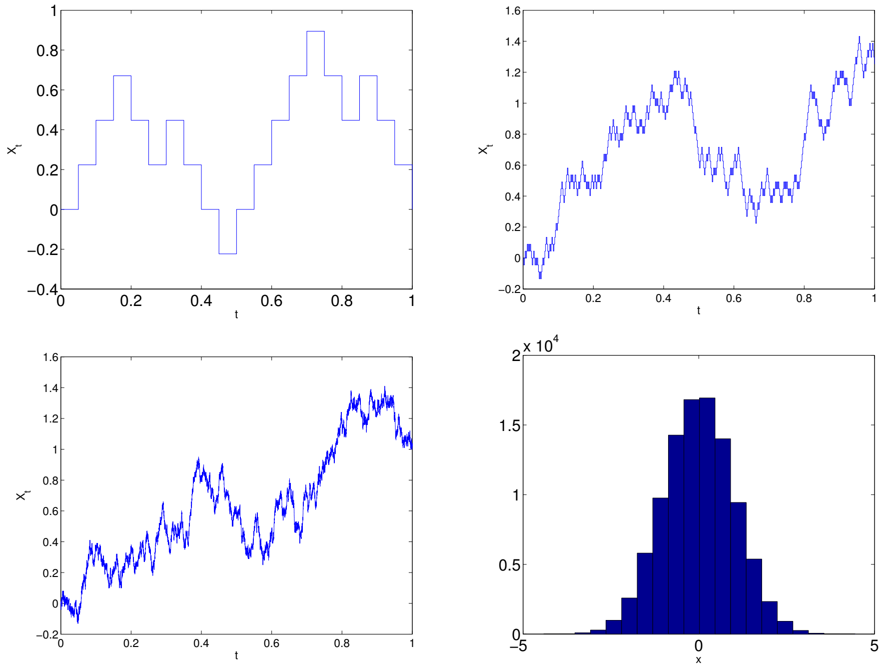

Figure 4: Sample paths of $(X_n^{\Delta t})_n$ for $\Delta t = 0.05,\ 0.002,\ 0.001$ over the unit time interval $[0, 1]$, with piecewise constant interpolation. The lower right plot shows the histogram (i.e., the unnormalized empirical distribution) of $(X_{1000}^{\Delta t})$ at time $t = 1$, averaged over 10 000 independent realizations.

distribution of the random variable $X_N^{\Delta t}$ for fixed $t = N\Delta t$ is the same as the distribution of a centered Gaussian random variable with variance t. As this is true for any $t > 0$, we can think of the limiting process as a continuous-time Markov process $B = (B_t)_{t>0}$ with Gaussian transition probabilities,

$$\begin{aligned}\mathbb{P}\left(B_{t+s} \in A \,|B_s = x\right) &= \int_A q_{s,t}(x,y)dy \\ &= \frac{1}{\sqrt{2\pi|t-s|}}\int_A \exp\left(-\frac{|y-x|^2}{2|t-s|}\right)dy.\end{aligned}$$

The stochastic process B is homogeneous *or* time-homogeneous *because the transition probability density $q_{s,t}(\cdot,\cdot)$ does not depend on the actual values of t and s, but only on their difference, i.e.,*

$$q_{s,t}(\cdot,\cdot) = \tilde{q}_{|s-t|}(\cdot,\cdot) \tag{6}$$

Remark 3.2. *The choice of exponent $1/2$ in $\sqrt{\Delta t} = (\Delta t)^{1/2}$ in (5) is unique. For $(\Delta t)^\alpha$ with $\alpha \in (0, \frac{1}{2})$, the limit of $X_n^{\Delta t}$ "explodes" in the sense that the variance of the process blows up, i.e., $\mathbb{E}[(X_N^{\Delta t})^2] \to \infty$ as $N \to \infty$. On the other hand, for $(\Delta t)^\alpha$ with $\alpha > 1/2$, $X_N^{\Delta t} \to 0$ in probability as $N \to \infty$.*

3.1 Brownian motion

Brownian motion is named after the British botanist, Robert Brown (1773-1858), who first observed the random motion of pollen particles suspended in water. Einstein called the Brownian process "Zitterbewegung" in his 1905 paper, *Über die von der molekularkinetischen Theorie der Wärme gefordete Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen.* The Brownian motion is a continuous-time stochastic process which is nowhere differentiable. It is also a *martingale* in the sense that on average, the particle stays in the same location at which it was first observed. In other words, the best estimate of where the particle will be after a time $t > 0$ is its initial location.

Definition 3.3. *(Brownian motion) The stochastic process* $B = (B_t)_{t>0}$ *with* $B_t \in \mathbb{R}$ *is called the 1-dimensional* Brownian motion *or the 1-dimensional* Wiener process *if it has the following properties:*

(i) $B_0 = 0$ $\mathbb{P}$-*a.s.*

(ii) B *has* independent increments, *i.e., for all* $s < t$, $(B_t - B_s)$ *is a random variable which is independent of* B_r *for* $0 \le r \le s$.

(iii) B *has* stationary, Gaussian increments, *i.e., for* $t > s$ *we have*[1]

$$B_t - B_s \overset{D}{=} B_{t-s} \tag{7a}$$

$$\overset{D}{=} \mathcal{N}(0, t-s). \tag{7b}$$

(iv) Trajectories of Brownian motion are continuous functions of time.

We now make precise some important notions:

Definition 3.4. *(Filtered probability space) A filtered probability space is a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *with a filtration* $(\mathcal{F}_t)_{t\ge 0}$ *such that* $\forall t \ge 0$,

$$\mathcal{F}_t \subset \mathcal{F}.$$

Remark 3.5. *One may write* $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ *to refer to a filtered probability space. However, if one is working with a particular stochastic process* X, *one may consider the sigma-algebra* $\mathcal{F}$ *on* Ω *to simply be the smallest sigma-algebra which contains the union of the* $\mathcal{F}_t^X$, *where* $\mathcal{F}_t^X := \sigma(X_s \ : \ s \le t)$. *In symbols, we define the sigma-algebra in the probability space to be*

$$\mathcal{F} := \vee_{t\ge 0}\mathcal{F}_t := \sigma\left(\cup_{t\ge 0}\mathcal{F}_t\right).$$

Definition 3.6 (Martingale)**.** *A stochastic process* $X = (X_t)_{t>0}$ *is a* martingale *with respect to a filtered probability space* $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ *if* X *satisfies the following properties:*

(i) X *is adapted to* $\mathcal{F}$, *i.e.* X_t *is measurable with respect to* $\mathcal{F}_t$ *for every* $t \ge 0$

(ii) X *is integrable:* $X \in L^1(\Omega, \mathbb{P})$, *i.e.*

$$\mathbb{E}\left[|X|\right] = \int_\Omega |X(\omega)| d\mathbb{P}(\omega) < \infty$$

[1] The notation "$X \overset{D}{=} Y$" means "X has the same distribution as Y".

(iii) X has the martingale property*:* $\forall t > s \geq 0$

$$\mathbb{E}\left[X_t | \mathcal{F}_s\right] = X_s.$$

Definition 3.7. *(Gaussian process) A 1-dimensional process $G = (G_t)_{t>0}$ is called a* Gaussian process *if for any collection $(t_1, \ldots, t_m) \subset I$ for arbitrary $m \in \mathbb{N}_0$, the random variable $(G_{t_1}, \ldots, G_{t_m})$ has a Gaussian distribution, i.e. it has a density*

$$f(g) = \frac{1}{\sqrt{det(2\pi\Sigma)}} \exp\left[-\frac{1}{2}(g-\mu)^\top \Sigma^{-1}(g-\mu)\right] \tag{8}$$

where $g = (g_1, \ldots, g_m)$, $\mu \in \mathbb{R}^m$ is a constant vector of means and $\Sigma = \Sigma^\top \in \mathbb{R}^{m\times m}$ is a symmetric positive semi-definite matrix.

Remark 3.8. *The Brownian motion process is a Gaussian process with the vector of means $\mu = 0$ and covariance matrix*

$$\Sigma = \begin{pmatrix} t_1 & 0 & \dots & 0 \\ 0 & t_2 - t_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & t_m - t_{m-1} \end{pmatrix} \tag{9}$$

The covariance matrix is diagonal due to the independence of the increments of Brownian motion.

Remark 3.9. *Some further remarks are in order.*

(a) Conditions (i)-(iii) define a consistent family of finite-dimensional distributions. Hence, the existence of the process B is guaranteed by the Kolmogorov Extension Theorem.

(b) Conditions (i)-(iii) imply that $\mathbb{E}[B_t] = 0$ and $\mathbb{E}[B_t B_s] = \min(t, s)$ $\forall s, t \in \mathbb{R}$. The proof is left as an exercise.

(c) The discrete process $(X_n^{\Delta t})_{n\in\mathbb{N}_0}$ converges in distribution to a Brownian motion $(B_t)_{t\geq 0}$ if the time discrete is linearly interpolated between two successive points. In other words, if we consider the continuous-time stochastic processes $(X_t^{\Delta t})_{t>0}$ (which is obtained by linear interpolation between the $X_N^{\Delta t}$) and B as random variables on the space of continuous trajectories ($C(\mathbb{R}_+)$ and $\mathcal{B}(C(\mathbb{R}_+))$), then the process $(X_t^{\Delta t})_{t>0}$ converges in distribution to B.

(d) We have that

$$\begin{aligned} \mathbb{E}\left[(B_t - B_s)^2\right] &= \mathbb{E}\left[(B_{t-s})^2\right] \text{ by (7a) in Definition 3.3} \\ &= |t - s| \text{ by (7b) in Definition 3.3.} \end{aligned}$$

(e) Brownian motion enjoys the following scaling invariance, also known as self-similarity of Brownian motion*: for every $t > 0$ and $\alpha > 0$,*

$$B_t \stackrel{D}{=} \alpha^{-1/2} B_{\alpha t}.$$

An alternative construction of Brownian motion

Observe that we have constructed Brownian motion by starting with the scaled random walk process and using the Kolmogorov Extension Theorem. Now we present an alternative method for constructing Brownian motion that is useful for numerics, called the Karhunen-Loève expansion of Brownian motion. We will consider this expansion for Brownian motion on the unit time interval $[0,1]$.

Let $\{\eta_k\}_{k\in\mathbb{N}}$ be a collection of independent, identically distributed (i.i.d) Gaussian random variables distributed according to $\mathcal{N}(0,1)$, and let $\{\phi_k(t)\}_{k\in\mathbb{N}}$ be an orthonormal basis of

$$L^2([0,1]) = \left\{ u : [0,1] \to \mathbb{R} \, : \, \int_0^1 |u(t)|^2 dt < \infty \right\}. \tag{10}$$

By construction, the basis functions satisfy

$$\langle \phi_i, \phi_j \rangle = \int_0^1 \phi_i(t)\phi_j(t) dt = \delta_{ij} \, ,$$

and we can represent any function $\forall f \in L^2([0,1])$ by

$$f(t) = \sum_{k\in\mathbb{N}} \alpha_k \phi_k(t)$$

for $\alpha_k = \langle f, \phi_k \rangle$. We have the following result.

Theorem 3.10. *(Karhunen-Loève): The process $(W_t)_{0\le t\le 1}$ defined by*

$$W_t = \sum_{k\in\mathbb{N}} \eta_k \int_0^t \phi_k(s) ds \tag{11}$$

is a Brownian motion.

Proof. We give only a sketch of the proof. For details, see the Appendix in [MS05], or [KS91]). The key components of the proof are to show the following:

(i) The infinite sum which defines the Karhunen-Loève expansion is absolutely convergent, uniformly on $[0,1]$.

(ii) It holds that $\mathbb{E}[W_t] = 0$ and $\mathbb{E}[W_t W_s] = \min(s,t)$.

□

4 Day 4, 06.11.2012

4.1 Brownian motion

From last week, we saw that the Brownian motion $(B_t)_{t\ge 0}$ is a continuous-time stochastic process on $\mathbb{R}$ with

- stationary, independent, Gaussian increments
- a.s. continuous paths. That is, for fixed ω, each $(B_t)_{t\ge0}(\omega)$ is a continuous trajectory in $\mathbb{R}$.

Moreover the scaled random walk defined by

$$X_{n+1}^{\Delta t} = X_n^{\Delta t} + \sqrt{\Delta t}\xi_{n+1}$$

with linear interpolation converges weakly (i.e. converges in distribution) to the Brownian motion process. Above, the $(\xi_n)_{n\in\mathbb{N}}$ are independent, identically distributed (i.i.d) normalized Gaussian random variables (i.e. ξ_n is Gaussian with mean zero and variance 1).

Remark 4.1. *Two remarks are in order.*

- *Continuity can be understood using the Lévy construction of Brownian motion on the set of dyadic rationals,*

$$D := \bigcup_{n\in\mathbb{N}} D_n, \quad D_n := \left\{ \frac{k}{2^n} \ : \ k = 0.\dots, 2^n \right\}.$$

The construction of Brownian motion on the unit time interval is as follows. Let $\{Z_t\}_{t\in D}$ be a collection of independent, normalized random variables defined on a probability space. Define the collection of functions $(F_n)_{n\in\mathbb{N}}$, where $F_n : [0,1] \to \mathbb{R}$ are given by

$$F_n(t) := \begin{cases} 0 & t \in D_{n-1} \\ 2^{-(j+1)/2} Z_t & t \in D_j \setminus D_{j-1} \\ \textit{lin. interp.} & \textit{in between.} \end{cases}$$

Then the process

$$B(t) = \sum_{n=1}^{\infty} F_n(t).$$

is indeed a Brownian motion on $[0,1]$. The Gaussianity of the $\{Z_t\}_{t\in D}$ leads to the stationary, independent Gaussian increments of the process $(B_t)_{t\in[0,1]}$. The continuity of the process follows from an application of the Borel-Cantelli Lemma, which states that there exists a random and almost surely finite number $N \in \mathbb{N}$ such that for all $n \geq N$ and $d \in D_n$, $|Z_d| < c\sqrt{n}$ holds. This boundedness condition implies that $\forall n \geq N$ we have a decay condition *for the F_n:*

$$\|F_n\|_\infty < c\sqrt{n}2^{-n/2}.$$

Therefore the sum $\sum_j F_j(\cdot)$ converges uniformly on $[0,1]$. As each F_j is continuous and the uniform limit of continuous functions is continuous, the process $(B_t)_{t\in[0,1]}$ is continuous. For more details, see [MP10].

- *The Hausdorff dimension* $\dim_{\mathcal{H}}$ *of Brownian motion paths depends on the dimension of the space $\mathbb{R}^d$ in which the Brownian motion paths live.*[2] *Let $B_{[0,1]} = \{B_t \in \mathbb{R}^d \colon t \in [0,1]\}$ be the graph of B_t over $I = [0,1]$. Then*

$$\dim_{\mathcal{H}} B_{[0,1]} = \begin{cases} 3/2 & d = 1 \\ 2 & d \geq 2\,. \end{cases}$$

[2] If you do not know what this is, just think of the box counting dimension that is an upper limit of the Hausdorff dimension.

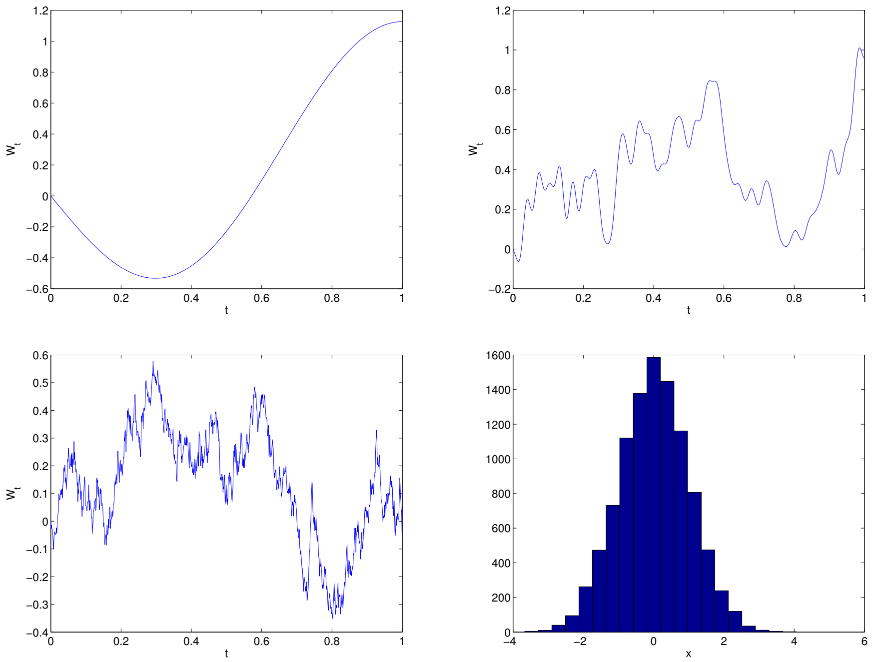

Figure 5: Sample paths of the Karhunen-Loève expansion of $(W_t$ for $M = 2, 64, 2048$ basis functions (you can guess which one is which). The lower right plot shows the unnormalized histogram of W_t at time $t = 1$, using $M = 64$ basis functions and averaged over 10 000 independent realizations.

> *The significance of this is as follows: if you consider Brownian motion paths confined to a smooth and compact two-dimensional domain and impose reflecting boundary conditions, then the Brownian motion paths will fill the domain in the limit as $t \to \infty$.*

4.2 Brownian bridge (Karhunen-Loève expansion of Brownian motion)

Theorem 4.2. *Let $\{\eta_k\}_{k\in\mathbb{N}}$ be i.i.d. normalized random variables and $\{\phi_k\}_{k\in\mathbb{N}}$ form a real orthonormal basis of $L^2([0,1])$. Then*

$$W_t = \sum_{k\in\mathbb{N}} \eta_k \int_0^t \phi_k(s)ds$$

is a Brownian motion on the interval $I = [0,1]$.

Exercise 4.3. *Show that, for the definition of $(W_t)_{t\in[0,1]}$ above, it holds that $\mathbb{E}[W_t W_s] = \min(s,t)$.*

Remark 4.4. *Unlike the scaled random walk construction of Brownian motion, no forward iterations are required here. This helps for the consideration of round-off errors in the construction of $(W_t)_{t\in[0,1]}$. Furthermore:*

- *Standard choices for the orthonormal basis $\{\phi_k\}_{k\in\mathbb{N}}$ are Haar wavelets or trigonometric functions. Hence the numerical error can be controlled by truncating the series and by the choice of the basis.*
- *To obtain a Brownian motion on any general time interval $[0,T]$, it suffices to use the scaling property, e.g.*

$$\begin{aligned} W_{[0,T]} &\overset{D}{=} \sqrt{T} W_{[0,1]/T} \\ &= \sqrt{T}\sum_{k\in\mathbb{N}} \eta_k \int_0^{t/T} \phi_k(s)ds. \end{aligned}$$

4.3 Application: filtering of Brownian motion

Suppose we know that $W_0 = 0$ and W_1 is equal to some constant ω. Without loss of generality, let $\omega = 0$. Suppose we wanted to generate a Brownian motion path which interpolated between the values $W_0 = 0$ and $W_1 = 0$.

Definition 4.5. *A continuous, mean-zero Gaussian process $(BB_t)_{t\geq 0}$ is called a* Brownian bridge *to ω if it has the same distribution as $(W_t)_{t\in[0,1]}$ conditional on the terminal value $W_1 = \omega$. Equivalently, $(BB_t)_{t\geq 0}$ is a Brownian bridge if*

$$\mathrm{Cov}\left[BB_t BB_s\right] = \min(s,t) - st.$$

Lemma 4.6. *If $(W_t)_{t\in[0,1]}$ is a Brownian motion, then $BB_t = W_t - tW_1$ is a Brownian bridge.*

Proof. Observe that

$$\mathbb{E}\left[BB_t\right] = \mathbb{E}\left[W_t - tW_1\right] = 0 - t\cdot 0 = 0,$$

so that $(BB_t)_{t\in[0,1]}$ is indeed mean-zero. The process $(BB_t)_{t\in[0,1]}$ inherits continuity from the process $(W_t)_{t\in[0,1]}$. The covariance process is given by

$$\begin{aligned} \mathrm{Cov}(BB_t BB_s) &= \mathbb{E}\left[BB_t BB_s\right] = \mathbb{E}\left[(W_t - tW_1)(W_s - sW_1)\right] \\ &= \mathbb{E}\left[W_t W_s\right] - t\underbrace{\mathbb{E}\left[W_1 W_s\right]}_{=\min(s,1)} - s\underbrace{\mathbb{E}\left[W_1 W_t\right]}_{=\min(t,1)} + ts\mathbb{E}\left[W_1 W_1\right] \\ &= \min(t,s) - ts - st + ts\,. \end{aligned}$$

□

4.3.1 How does one simulate a Brownian bridge?

First approach: forward iteration, using Euler's method. The time interval is $[0,1]$ and we have a time step of $\Delta t := 1/N$, so we have $(N+1)$ discretized time nodes $(t_n = n\Delta t)_{n=0,\ldots,N}$ and $(N+1)$ values $(Y_n^{\Delta t})_{n=0,\ldots,N}$. Let $\{\xi_n\}_{n=0,\ldots,N-1}$ be a collection of i.i.d. normalized random variables. Forward iteration gives

$$Y_{n+1}^{\Delta t} = Y_n^{\Delta t}\left(1 - \frac{\Delta t}{1-t_n}\right) + \sqrt{\Delta t}\xi_{n+1}.$$

It holds that $1 - t_{N-1} = \Delta t$ by definition of $\Delta t = 1/N$. Therefore from the formula above we have

$$Y_N^{\Delta t} = \sqrt{\Delta t}\xi_{N+1}.$$

Therefore $Y_N^{\Delta t}$ is a mean zero Gaussian random variable with variance Δt. While this implies that $Y_N^{\Delta t}$ should converge in probability to the value 0 as the step size $\Delta t \to 0$, the forward iteration approach is not optimal because the random variable ξ_{N+1} is continuous, so

$$\mathbb{P}\left(Y_N^{\Delta t} = 0\right) = 0.$$

Therefore this construction of the Brownian bridge to the value $\omega = 0$ will in general not yield processes which are at 0 at time $t = 1$. As a matter of fact, $Y_N^{\Delta t}$ is unbounded and can be arbitrarily far away from zero.

Second approach: Recall the Karhunen-Loéve construction of Brownian motion and choose trigonometric functions as an orthonormal basis. Then the process $(W_t)_{t\in[0,1]}$ given by

$$W_t(\omega) = \sqrt{2}\sum_{k=1}^{M} \eta_k(\omega)\frac{\sin((k-\frac{1}{2})\pi t}{(k-\frac{1}{2})\pi}$$

is a Brownian motion and we can define the Brownian bridge to ω at $t = 1$ by

$$BB_t = W_t - t(W_1 - \omega).$$

Remark 4.7. *It holds that*

$$\begin{aligned} BB_t &= \sqrt{2}\sum_{k\in\mathbb{N}} \eta_k \frac{\sin(k\pi t)}{k\pi} \\ &= \sum_{k\in\mathbb{N}} \eta_k \sqrt{\lambda_k}\psi_k(t), \end{aligned}$$

where $\{\lambda_k, \psi_k\}_{k\in\mathbb{N}} = \left\{\sqrt{2}/k\pi, \sin(k\pi t)\right\}_{k\in\mathbb{N}}$ *is the eigensystem of the* covariance operator $T : L^1([0,1]) \to L^1([0,1])$ *of the process* $(BB_t)_{t\in[0,1]}$*, defined by*

$$(Tu)(t) = \int_0^1 \underbrace{\mathrm{Cov}(BB_t BB_s)}_{=\min(t,s)-st} u(s)ds\,,$$

i.e.,

$$T\psi_k(\cdot) = \lambda_k\psi_k(\cdot).$$

The second approach works for any *stochastic process which has finite variance over a finite time interval. For details, see [Xiu10].*

5 Day 5, 13.11.2012 (Lecturer: Stefanie W.)

5.1 Stochastic Integration (Itô integral)

Recall that Brownian motion $(B_t)_{t>0}$ is a stochastic process with the following properties:

- $B_0 = 0$ $\mathbb{P}$-a.s.
- $\forall 0 \leq t_0 < t_1 < t_2 < \ldots < t_n$, the increments $B_{t_i} - B_{t_{i-1}}$ are independent for $i = 1, \ldots, n$ and Gaussian with mean 0 and variance $t_i - t_{i-1}$.

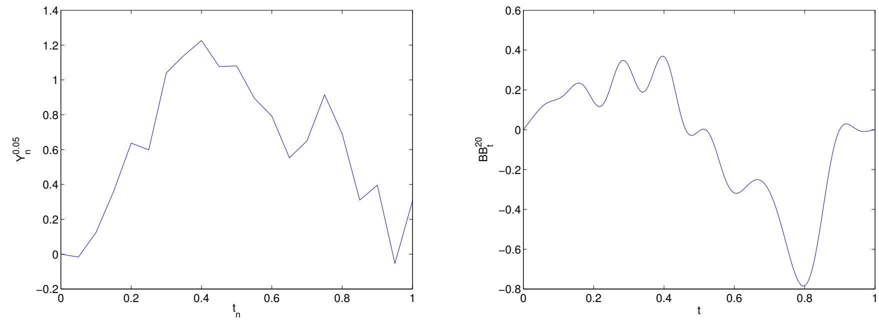

Figure 6: Sample paths of the Brownian bridge approximation, using the Euler scheme with $\Delta t = 0.05$ (left panel) and Karhunen-Loève expansion with $M = 20$ basis functions (right panel).

- $t \mapsto B_t(\omega)$ is continuous $\mathbb{P}$-a.s. but is $\mathbb{P}$-a.s. nowhere differentiable.

One of the motivations for the development of the stochastic integral lies in financial mathematics, where one wishes to determine the price of an asset that evolves randomly. The French mathematician Louis Bachelier is generally considered one of the first people to model random asset prices. In his PhD thesis, Bachelier considered the following problem. Let the value S_t of an asset at time $t > 0$ be modeled by

$$S_t = \sigma B_t$$

where $\sigma > 0$ is a scalar that describes the volatility of the stock price. Let $f(t)$ be the amount of money an individual invests in the asset in some infinitesimal time interval $[t, t + dt]$. Then the wealth of the individual at the end of a time interval $[0, T]$ is given by

$$\int_0^T f(t) dS_t = \sigma \int_0^T f(t) dB_t.$$

However, it is not clear what the expression 'dB_t' means. In this section, we will consider what an integral with respect to dB_t means, and we will also consider the case when the function f depends not only on time but on the random element ω.

The first idea is to rewrite

$$\int f(t) dB_t = \int f(t) \frac{dB_t}{dt} dt$$

but as Brownian motion is almost surely nowhere differentiable, we cannot write $\frac{dB_t}{dt}$.

The second idea is to proceed as in the definition of the Lebesgue integral: start with simple step functions and later extend the definition to more general functions by the Itô Isometry.

Step 1: Consider simple functions

$$f(t) = \sum_{i=1}^{n} a_i \chi_{(t_i, t_{i+1}]}(t)$$

where χ_A is the indicator function of a set A satisfying

$$\chi_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A. \end{cases}$$

Observe that f takes a finite number n of values. By the theory of Lebesgue integration, we know that the set of these simple functions is dense in $L^2([0,\infty))$. We also know that the usual Riemann integral of such a function f corresponds to the area under the graph of f, with

$$\int_0^\infty f(t)dt = \sum_i a_i(t_{i+1} - t_i)$$

Step 2: We now extend the method above to stochastic integral with respect to Brownian motion:

$$\int f(t)dB_t = \sum a_i(B_{t_{i+1}} - B_{t_i}).$$

Remark 5.1. *By the equation above, it follows that the integral $\int f(t)dB_t$ is a random variable, since the B_{t_i} are random variables. Since increments of Brownian motion are independent and Gaussian, the integral $\int f(t)dB_t$ is normally distributed with zero mean. What about its variance?*

Lemma 5.2. (Itô Isometry for simple functions) *For a simple function $f(t) = \sum_i a_i \chi_{(t_i,t_{i+1}]}(t)$, it holds that*

$$\mathbb{E}\left[\left(\int_0^\infty f(t)dB_t\right)^2\right] = \int_0^\infty (f(t))^2\, dt.$$

Proof.

$$\begin{aligned} \operatorname{var}\left(\int_0^\infty f(t)dB_t\right) &= \operatorname{var}\left(\sum_i a_i\left(B_{t_{i+1}} - B_{t_i}\right)\right) \\ &= \sum_{i=1}^n a_i^2 \operatorname{var}\left(B_{t_{i+1}} - B_{t_i}\right) \\ &= \sum_{i=1}^n a_i^2\left(t_{i+1} - t_i\right) \\ &= \sum_{i=1}^n a_i^2 \int_0^\infty \chi_{(t_i,t_{i+1}]}dt \\ &= \int \sum_{i=1}^n a_i^2 \chi_{(t_i,t_{i+1}]}dt \\ &= \int (f(t))^2\, dt. \end{aligned}$$

Therefore

$$\operatorname{var}\left(\int f(t)dB_t\right) = \mathbb{E}\left[\left(\int f(t)dB_t\right)^2\right] - \left(\underbrace{\mathbb{E}\left[\int f(t)dB_t\right]}_{=0}\right)^2$$
$$= \mathbb{E}\left[\left(\int f(t)dB_t\right)^2\right].$$

□

Step 3: Now we extend the definition of the integral to $L^2([0,\infty))$. The main result is the following

Theorem 5.3. (Itô integral for $L^2([0,\infty))$ functions) *The definition of the Itô integral can be extended to elements $f \in L^2([0,\infty))$ by setting*

$$\int_0^\infty f(t)dB_t := \lim_{n\to\infty} \int_0^\infty f_n(t)dB_t$$

where the sequence $(f_n)_{n\in\mathbb{N}}$ is a sequence of a simple functions satisfying $f_n \to f$ in $L^2([0,\infty))$, i.e.

$$\|f_n - f\|_{L^2([0,\infty))} = \left(\int_0^\infty (f_n - f)^2(t)dt\right)^{1/2} \underset{n\to\infty}{\longrightarrow} 0.$$

By the Itô isometry, we can show that $(\int f_n(t)dB_t)_{n\in\mathbb{N}}$ is a Cauchy sequence in the weighted L^2 space

$$L^2(\Omega,\mathbb{P}) := \left\{F\colon \Omega \to \mathbb{R} :\ \|F\|_{L^2(\Omega,\mathbb{P})} < \infty\right\}$$

of measurable functions F where

$$\|F\|^2_{L^2(\Omega,\mathbb{P})} = \int |F|^2(\omega)\, d\mathbb{P}(\omega).$$

To show that the sequence $(\int f_n(t)dB_t)_{n\in\mathbb{N}}$ is a Cauchy sequence, let $(f_l)_{l\in\mathbb{N}}$ be a sequence of functions converging to f in $L^2([0,\infty))$ and consider for $m,n \in \mathbb{N}$

$$\begin{aligned}
&\left\|\int f_n(t)dB_t - \int f_m(t)dB_t\right\|_{L^2(\Omega,\mathbb{P})} \\
&= \left(\mathbb{E}\left[\left(\int f_n(t)dB_t - \int f_m(t)dB_t\right)^2\right]\right)^{1/2} \\
&= \left(\mathbb{E}\left[\left(\int f_n(t) - f_m(t)dB_t\right)^2\right]\right)^{1/2} \\
&= \left(\int (f_n(t) - f_m(t))^2\, dt\right)^{1/2} \quad \textit{(Itô isometry)} \\
&= \|f_n - f_m\|_{L^2([0,\infty))} \\
&\le \|f_n - f\|_{L^2([0,\infty))} + \|f_m - f\|_{L^2([0,\infty))}
\end{aligned}$$

and using that $\|f_n - f\|_{L^2([0,\infty))}$ *and* $\|f_m - f\|_{L^2([0,\infty))} \to 0$ *as* $m, n \to \infty$, *the result follows.*

Since $L^2(\Omega, \mathbb{P})$ is complete, the limit exists and is in the same space. Moreover, by the Itô isometry, the limit is independent of the sequence $(f_n)_{n\in\mathbb{N}}$ used to approximate f in $L^2([0,\infty))$ (see [Dur96] for an example). □

Example 5.4. *Consider the random variable* $\int_0^\infty \exp(-t)dB_t$. *How is it distributed? Using the Itô Isometry, the random variable is Gaussian with mean zero and variance* $\frac{1}{2} = \int_0^\infty \exp(-2t)dt$.

Corollary 5.5. *The Itô Isometry holds as well for* $f \in L^2([0,\infty))$, *not just simple functions.*

Step 4: Now we consider functions f which depend both on the random element ω as well as time t. That is, we consider stochastic integrals of stochastic processes $f : [0,\infty) \times \Omega \to \mathbb{R}$ with the following properties:

(i) f is $\mathcal{B} \times \mathcal{F}$-measurable, where $\mathcal{B}$ is the Borel sigma-algebra on $[0,\infty)$ and $\mathcal{F}$ is a given sigma-algebra on Ω.

(ii) $f(t,\omega)$ is adapted with respect to $\mathcal{F}_t$, where $\mathcal{F}_t := \sigma\left(B_s \ : \ s \le t\right)$

(iii) $\mathbb{E}\left[\int |f(t,\omega)|^2 dt\right] < \infty$.

Consider simple stochastic processes of the form

$$f(t,\omega) = \sum_{i=1}^{n} a_i(\omega)\chi_{(t_i,t_{i+1}]}(t).$$

Then

$$\int f(t,\omega)dB_t = \sum_{i=1}^{n} a_i(\omega)\left(B_{t_{i+1}} - B_{t_i}\right).$$

Example 5.6. *Fix* $n \in \mathbb{N}$, *fix a time step* $\Delta t := 2^{-n}$ *and define the time nodes* $t_i := i\Delta t$ *for* $i = 0, 1, 2, \ldots$. *Let* $(B_t)_{t>0}$ *be the standard Brownian motion. Define the following processes on* $[0,\infty)$:

$$\begin{aligned} f_1(t,\omega) &= \sum_{i\in\mathbb{N}} B_{t_i}(\omega)\chi_{[t_i,t_{i+1})}(t) \\ f_2(t,\omega) &= \sum_{i\in\mathbb{N}} B_{t_{i+1}}(\omega)\chi_{[t_i,t_{i+1})}(t). \end{aligned}$$

Now fix $T > 0$ *and* N *such that* $T = t_N = N\Delta t = N2^{-n}$ *and compute the expected values of the integrals of* f_1 *and* f_2 *over* $[0,T]$. *By the independent increments property of Brownian motion (or the martingale property of Brownian motion), we have*

$$\mathbb{E}\left[\int_0^T f_1(t,\omega)dB_t\right] = \sum_{i=0}^{N-1} \mathbb{E}\left[B_{t_i}\left(B_{t_{i+1}} - B_{t_i}\right)\right] = 0.$$

Using the fact above with linearity of expectation, we also have

$$
\begin{aligned}
\mathbb{E}\left[\int_0^T f_2(t,\omega)dB_t\right] &= \sum_{i=0}^{N-1} \mathbb{E}\left[B_{t_{i+1}}\left(B_{t_{i+1}} - B_{t_i}\right)\right] - 0 \\
&= \sum_{i=0}^{N-1} \left(\mathbb{E}\left[B_{t_{i+1}}\left(B_{t_{i+1}} - B_{t_i}\right)\right] - \mathbb{E}\left[B_{t_i}\left(B_{t_{i+1}} - B_{t_i}\right)\right]\right) \\
&= \sum_{i=0}^{N-1} \mathbb{E}\left[\left(B_{t_{i+1}} - B_{t_i}\right)^2\right] \\
&= \sum_{i=0}^{N-1} t_{i+1} - t_i = T.
\end{aligned}
$$

In the case of Riemann integration of deterministic integrals, letting $n \to \infty$ would lead to the result that both integrals above are equal. We see that for stochastic integration, this is not the case; even if we let $n \to \infty$, the expectations of the Itô integrals would not be equal. This is because the choice of endpoint of the interval matters in stochastic integration. Choosing the left endpoint (i.e. choosing B_{t_i}) for f_1 and the right endpoint (i.e. $B_{t_{i+1}}$for f_2 leads to different expectations. Note also that taking the right endpoint in f_2 leads to f_2 not being adapted, since $B_{t_{i+1}}$ is not measurable with respect to $\mathcal{F}_t$ for $t < t_{i+1}$. Therefore, by property (ii) above, we may not integrate f_2 with respect to dB_t in the way we have just described.

6 Day 6, 20.11.2012

6.1 The Itô Integral, continued

We extend the Itô integral to the case

$$
I[f](\omega) = \int_0^t f(s,\omega)dB_s(\omega)\,,
$$

where B_t is one-dimensional Brownian motion. One aim is to understand

$$
dX_t = b(t,X_t)dt + \sigma(t,X_t)dB_t\,,\ X_0 = x
$$

that is SDE shorthand for

$$
X_t = x + \int_0^t b(s,X_s)\,ds + \int_0^t \sigma(s,X_s)\,dB_s\,.
$$

A second objective later on will be to analyze discretizations of SDEs, such as

$$
X_{n+1} - X_n = b(t_n,X_n)\Delta t + \sigma(t_n,X_n)\Delta B_n.
$$

We begin with a couple of definitions.

Definition 6.1. *We call $\|\cdot\|_{\mathcal{V}}$ the norm defined by*

$$
\|f\|_{\mathcal{V}}^2 = \mathbb{E}\left[\int_s^t |f(u,\cdot)|^2 du\right] = \int_\Omega \int_s^t |f(u,\omega)|^2 du\ d\mathbb{P}(\omega)\,.
$$

Definition 6.2 (Cf. the considerations at the bottom of p. 21)**.** *Let $\mathcal{V} = \mathcal{V}(s,t)$ be the class of functions $f : [0,\infty) \times \Omega \to \mathbb{R}$ with*

(i) $(t,\omega) \mapsto f(t,\omega)$ *is* $\mathcal{B} \times \mathcal{F}$*-measurable*[3]

(ii) $f(t,\cdot)$ *is* $\mathcal{F}_t$*-adapted*

(iii) $\|f\|_{\mathcal{V}} < \infty$.

Definition 6.3. *A* simple function $\varphi : [0,\infty) \times \Omega \to \mathbb{R}$ *is a function of the form*

$$\varphi(t,\omega) = \sum_j e_j(\omega)\chi_{[t_j,t_{j+1})}(t)$$

where each e_j is $\mathcal{F}_{t_j}$-measurable and $\{\mathcal{F}_t\}_{t\geq 0}$ with $\mathcal{F}_t = \sigma(B_s \colon s \leq t)$ is the filtration generated by Brownian motion.

Definition 6.4. *The Itô integral for a simple function φ is defined by*

$$I[\varphi](\omega) = \int_0^t \varphi(s,\omega)dB_s(\omega) = \sum_j e_j(\omega)(B_{t_{j+1}} - B_{t_j}).$$

Lemma 6.5. Itô Isometry*: If $\varphi(t,\omega)$ is a bounded, simple function, then*

$$\mathbb{E}\left[\left(\int_s^t \varphi(u,\omega)dB_u(\omega)\right)^2\right] = \mathbb{E}\left[\int_s^t |\varphi(u,\omega)|^2 du\right].$$

Proof. Define $\Delta B_j := B_{t_{j+1}} - B_{t_j}$. Then $I[\varphi] = \sum_j e_j \Delta B_j$. By independence of $e_i e_j \Delta B_i$ from ΔB_j when $i \neq j$, we have that

$$\mathbb{E}\left[e_i e_j \Delta B_i \Delta B_j\right] = \begin{cases} 0 & i \neq j \\ \mathbb{E}\left[e_j^2\right](t_{j+1} - t_j) & i = j. \end{cases}$$

Therefore,

$$\begin{aligned} \mathbb{E}\left[\left(\int_s^t \varphi(u,\cdot)dB_u\right)^2\right] &= \sum_{i,j} \mathbb{E}\left[e_i e_j \Delta B_i \Delta B_j\right] \\ &= \sum_j \mathbb{E}\left[e_j^2\right](t_{j+1} - t_j) = \mathbb{E}\left[\int_s^t \varphi(u,\cdot)du\right]. \end{aligned}$$

□

Now we will extend the Itô integral to $\mathcal{V} = \mathcal{V}(s,t)$, by extending the Itô integral to progressively larger classes of functions.

Step 1: Let $g \in \mathcal{V}$ be a uniformly bounded function which is continuous for each ω. Then there exists a sequence of simple functions $(\varphi_n)_{n\in\mathbb{N}}$ such that

$$\|\varphi_n - g\|_{\mathcal{V}} \to 0$$

as $n \to \infty$.

[3] Here again: $\mathcal{B} = \mathcal{B}([0,\infty))$ is the σ-algebra od Borel sets over $[0,\infty)$.

Proof. Choose $\varphi_n(t,\omega) = \sum_j g(t_j,\omega)\chi_{[t_j,t_{j+1})}(t)$. Then $\varphi_n \to g$ in $L^2([s,t])$ for each $\omega \in \Omega$, and hence $\|\varphi_n - g\|_{\mathcal{V}}^2 \to 0$, i.e.,

$$\mathbb{E}\left[\int_s^t (\varphi_n - g)^2 \, du\right] = \int_\Omega \left(\int_s^t (\varphi_n - g)^2 \, du\right) d\mathbb{P} \to 0$$

as $n \to \infty$. □

Step 2: Let $h \in \mathcal{V}$ be bounded. Then there exists a bounded sequence of functions $(g_n)_{n\in\mathbb{N}} \subset \mathcal{V}$ such that each g_n is continuous in t for each ω and for each $n \in \mathbb{N}$, such that $\|g_n - h\|_{\mathcal{V}} \to 0$.

Proof. Suppose that $|h(t,\omega)| \le M < \infty$. For each n, let ψ_n be defined by

(i) $\psi_n(x) = 0$ for $x \in (-\infty, -\frac{1}{n}] \cup [0,\infty)$

(ii) $\int_{\mathbb{R}} \psi_n(x)dx = 1$.

Now define the functions $g_n : [0,\infty) \times \Omega \to \mathbb{R}$ by

$$g_n(t,\omega) = \int_0^t \psi_n(s-t)h(s,\omega)ds.$$

Then it holds that $g_n \to h$ in $L^2([s,t])$, i.e., $\|g_n - h\|_{L^2([s,t])} \to 0$ as $n \to \infty$. As h is bounded, we can apply the bounded convergence theorem to obtain

$$\mathbb{E}\left[\int_s^t (g_n - h)^2 \, du\right] \to 0$$

as $n \to \infty$. □

Remark 6.6. *In the limit as $n \to \infty$, the $\psi_n(x)$ become more sharply peaked at $x = 0$—in other words, they approach a Dirac delta distribution:*

$$h(t,\omega) = \int_0^\infty \delta(s-t)h(s,\omega)ds\,.$$

Step 3: Let $f \in \mathcal{V}$. Then there exists a sequence of functions $(h_n)_{n\in\mathbb{N}} \subset \mathcal{V}$ such that h_n is bounded for each $n \in \mathbb{N}$ and $\|h_n - f\|_{\mathcal{V}} \to 0$ as $n \to \infty$.

Proof. Define

$$h_n(t,\omega) = \begin{cases} -n & f(t,\omega) < -n \\ f(t,\omega) & -n \le f(t,\omega) \le n \\ n & f(t,\omega) > n. \end{cases}$$

Then the assertion follows by the dominated convergence theorem. □

By Steps 1-3, $f \in \mathcal{V}$ can be approximated by sequences of simply functions φ_n in the sense that

$$\|f - \varphi_n\|_{\mathcal{V}} \to 0\,.$$

Therefore we can define

$$I[f](\omega) = \int_s^t f(u,\omega) dB_u(\omega) = \lim_{n\to\infty} \int_s^t \varphi_n(u,\omega) dB_u(\omega)\,,$$

where by the Itô isometry, the limit exists in $L^2(\Omega, \mathbb{P})$ because

$$\left(\int \varphi_n(u,\omega) dB_u(\omega) \right)_{n\in\mathbb{N}}$$

is a Cauchy sequence in $L^2(\Omega, \mathbb{P})$; see p. 21.

Definition 6.7. *Let $f \in \mathcal{V} = \mathcal{V}(s,t)$. Then the* Itô integral of f *is defined by*

$$\lim_{n\to\infty} \int_s^t \varphi_n(u,\omega) dB_u(\omega)$$

where $(\varphi_n)_{n\in\mathbb{N}}$ is a sequence of simple functions with $\varphi_n \to f$ in $\mathcal{V}$, i.e.,

$$\mathbb{E}\left[\int_s^t \left(f(u,\omega) - \varphi(u,\omega) \right)^2 du \right] \to 0 \ \text{as}\ n \to \infty.$$

Corollary 6.8. (Itô isometry) *For all $f \in \mathcal{V} = \mathcal{V}(s,t)$, we have*

$$\mathbb{E}\left[\left(\int_s^t f(u,\omega) dB_u(\omega) \right)^2 \right] = \mathbb{E}\left[\int_s^t |f(u,\omega)|^2 \, du \right].$$

Theorem 6.9. *Let $f, g \in \mathcal{V}(0,t)$ and $0 \le s \le u \le t$. Then (a.s.):*

(i)

$$\int_s^t f(\tau,\omega) dB_\tau = \int_s^u f(\tau,\omega) dB_\tau + \int_u^t f(\tau,\omega) dB_\tau\,.$$

(ii)

$$\int_s^t (\alpha f + \beta g dB_u = \alpha \int_s^t f dB_u + \beta \int_s^t g dB_u \quad \forall \alpha, \beta \in \mathbb{R}$$

(iii)

$$\mathbb{E}\left[\int_0^t f(s,\omega) dB_s \right] = 0$$

(iv)

$$\int_s^t f(u,\omega) dB_u$$

is $\mathcal{F}_t$-measurable.

Proof. Exercise. □

Example 6.10. *Consider the linear SDE for constants $A, B \in \mathbb{R}$*

$$dX_t = AX_t dt + BdW_t, \quad X_0 = x,$$

which means

$$X_t = x + A\int_0^t X_s ds + B\int_0^t dW_s.$$

One can show that the solution to the linear SDE can be expressed using the variation-of-constants-formula

$$X_t = e^{At}x + \int_0^t e^{A(t-s)} B dW_s.$$

The solution $(X_t)_{t>0}$ *is a Gaussian process, so it is completely specified by its mean and variance*

$$\begin{aligned}
\mathbb{E}\left[X_t\right] &= e^{At}x \text{ by property (iv) above} \\
\mathbb{E}\left[\left(X_t - \mathbb{E}\left[X_t\right]\right)^2\right] &= \mathbb{E}\left[\left(\int_0^t e^{A(t-s)} B dW s\right)^2\right] \\
&= \mathbb{E}\left[\int_0^t \left(e^{A(t-s)} B\right)^2 ds\right] \text{ by Itô isometry} \\
&= \int_0^t e^{2A(t-s)} B^2\, ds \\
&= \frac{B^2}{2A}\left(e^{2At} - 1\right)
\end{aligned}$$

Remark 6.11. *The main things to remember are that the approximation procedure for defining the Itô integral reduces to the Itô isometry for elementary functions* $\varphi_n \to f$ *(convergence in* $\mathcal{V}$*) and that the limiting integral* $I[f]$ *is in* $L^2(\Omega, \mathbb{P})$*. Specifically, we have proved that* $I[\varphi_n] \to I[f]$ *in* $L^2(\Omega, \mathbb{P})$*, i.e.,*

$$\mathbb{E}\left[(I[\varphi_n] - I[f])^2\right] = \int_\Omega \left(I[\varphi_n](\omega) - I[f](\omega)\right)^2 d\mathbb{P}(\omega) \to 0$$

as $n \to \infty$*.*

7 Day 7, 27.11.2012

7.1 The Itô Integral—recapitulation

The Itô integral for functions $f \in \mathcal{V} \cong L^2(\Omega \times [0,T], \mathbb{P} \otimes \lambda)$ is defined by

$$I[f](\omega) = \int_0^T f(\omega, s) dB_s(\omega) = \lim_{n\to\infty} \sum_j e_j^n \left(B_{t_{j+1}^n} - B_{t_j^n}\right),$$

with convergence in $L^2(\Omega, \mathbb{P})$. Here the $(e_j^n)_{n,j\in\mathbb{N}}$ is a sequence of random variables that are measurable with respect to $\sigma(B_s\colon s \le t_j^n)$, and

$$(\varphi_n)_{n\in\mathbb{N}}, \quad \varphi_n(\omega, t) = \sum_j e_j^n(\omega)\chi_{[t_j^n, t_{j+1}^n)}(t)$$

is a sequence of simple functions such that $\|\varphi_n - f\|_{\mathcal{V}} \to 0$.

The Itô integral provides the solution to the stochastic differential equation

$$dX_t(\omega) = b(X_t(\omega), t)dt + \sigma(X_t(\omega), t)dB_t(\omega), \quad X_0 = x. \tag{12}$$

Specifically, assuming that $(X_t)_{t\geq 0}$ is adapted to the filtration generated by B_t, i.e., X_t is measurable with respect to $\sigma(B_s\colon s\leq t)$, we have

$$X_t = x + \int_0^t b(X_s, s)\, ds + \int_0^t \sigma(X_s, s)\, dB_s\,.$$

7.2 Functions of bounded (quadratic) variation

What we now need is a theory of differentiation that is useful in solving equations such as (12) and which can explain properties of the Itô integral, such as

$$\int_0^t B_t dB_t = \frac{1}{2}B_t^2 - \frac{1}{2}t.$$

Exercise 7.1. *Prove the above equation.*

Definition 7.2. *Let $T > 0$. A sequence $(\Delta_n)_{n\in\mathbb{N}}$ of partitions of $[0,T]$, with*

$$\Delta_n = \{t_0^n, \dots, t_{k^n}^n\} \subset [0,T]\,, \quad 0 = t_0^n < t_1^n < \cdots < t_{k^n}^n = T$$

is called a refinement of partitions *of $[0,T]$ if the sequence satisfies*

$$\Delta_{n+1} \supset \Delta_n \quad \& \quad |\Delta_n| := \max_i |t_i^n - t_{i-1}^n| \to 0 \quad \text{as } n \to 0.$$

Example 7.3. *An example of a refinement of partitions is the sequence of dyadic partitions. Let $T = 1$, and define*

$$\Delta_n = \left\{ \frac{j}{2^n} : j = 0, 1, \dots, 2^n - 1, 2^n \right\}.$$

Definition 7.4. *A function $f : [0,T] \to \mathbb{R}$ is of* bounded variation *(BV) if*

$$\sup_{n\in\mathbb{N}} |f(t_i^n) - f(t_{i-1}^n)| < \infty$$

for all refinements of partitions $(\Delta_n)_{n\in\mathbb{N}}$.

Definition 7.5. *A function $f : [0,T] \to \mathbb{R}$ is of* quadratic variation *(QV) if its quadratic variation*

$$\langle f\rangle_t := \sup_n \sum_{t_i^n \leq t} |f(t_i^n) - f(t_{i-1}^n)|^2 < \infty.$$

is finite for every $t \in [0,T]$ and over all refinements of partitions.

Remark 7.6. *We make some remarks which we will not prove, with the exception of (iii).*

(i) Continuously differentiable functions are BV functions.

(ii) If one integrates against a BV function, the resulting Riemann-Stieltjes integral is independent of the refinement of partitions.

(iii) Continuous BV functions have zero QV:

$$\begin{aligned}
0 \le \langle f \rangle_t &= \sup_{n\in\mathbb{N}} \sum_{t_i^n \le t} |f(t_i^n) - f(t_{i-1}^n)|^2 \\
&= \lim_{n\to\infty} \sum_{t_i^n \le t} |f(t_i^n) - f(t_{i-1}^n)|^2 \\
&\le \max_i |f(t_i^n) - f(t_{i-1}^n)| \lim_{n\to\infty} \sum_{t_i^n \le t} |f(t_i^n) - f(t_{i-1}^n)| \\
&\le C \lim_{n\to\infty} |f(t_i^n) - f(t_{i-1}^n)| \text{ (since } f \text{ is a BV function)} \\
&= 0 \text{ (by continuity of } f) \; .
\end{aligned}$$

(iv) The quadratic variation of Brownian motion at time t is equal to t:

$$\langle B \rangle_t = t.$$

Brownian motion is not of bounded variation.

(v) Given an interval $[0, T]$ for $T > 0$ and a function $f : [0, T] \to \mathbb{R}$ of QV, the quadratic variation $\langle f \rangle_t$ is a BV function of time, which follows from the fact that $\langle \cdot \rangle$ is monotonic as a function of time.

Theorem 7.7 (Itô's formula I)**.** *Let $F \in C^{2,1}(\mathbb{R}, [0, T])$ and let $X = B \in C([0, T])$ be Brownian motion. Then*

$$F(X_t, t) = F(0, 0) + \int_0^t \frac{\partial F}{\partial x}(X_s, s)\, dX_s + \int_0^t \left(\frac{1}{2} \frac{\partial^2}{\partial x^2} + \frac{\partial}{\partial s} \right) F(X_s, s)\, ds \, .$$

Proof. For convenience, we will drop the time-dependence of F, so that $F(x, s) = F(x)$. By Taylor's theorem,

$$F(X_{t_i^n}) - F(X_{t_{i-1}^n}) = F'(X_{t_{i-1}^n})(X_{t_i^n} - X_{t_{i-1}^n}) + \frac{1}{2} F''(\xi_i^n)(X_{t_i^n} - X_{t_{i-1}^n})^2$$

for a number $\xi_i^n \in (X_{t_i^n}, X_{t_{i-1}^n})$. Then

$$F(X_t) - F(X_0) = \underbrace{\sum_{t_i^n \le t} F'(X_{t_{i-1}^n})(X_{t_i^n} - X_{t_{i-1}^n})}_{I_n} + \underbrace{\frac{1}{2} \sum_{t_i^n \le t} F''(\xi_i^n)(X_{t_i^n} - X_{t_{i-1}^n})^2}_{Q_n} \, .$$

We consider the two sums separately. As for the first sum, we observe that I_n is a discrete version of the Itô integral, and therefore

$$\left\| I_n - \int_0^t \frac{\partial F}{\partial x}(X_s, s)\, dX_s \right\|^2_{L^2(\Omega,\mathbb{P})} \to 0$$

as $n \to \infty$. Using (v) from the preceding remark, we know that the quadratic variation of $(X_t)_{t\in[0,T]}$ is itself a BV function of time. Therefore Q_n converges to the standard Riemann-Stieltjes integral,

$$\frac{1}{2} \int_0^t F''(X_s)\, d\langle X \rangle_s = \frac{1}{2} \int_0^t F''(X_s)\, ds \, ,$$

where, again, convergence is in $L^2(\Omega, \mathbb{P})$. ∎

Corollary 7.8 (Itô's Formula II). *Let $B = (B_t)_{t\geq 0}$ be d-dimensional Brownian motion and let $X = (X_t)_{t\geq 0}$ be the n-dimensional solution to the Itô stochastic differential equation*

$$dX_t = b(X_t, t)dt + \sigma(X_t, t)dB_t\,,$$

where $b\colon \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$, $\sigma\colon \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^{n\times d}$. Let $F \in C^{2,1}(\mathbb{R}^n, [0,\infty))$. Then $Y_t := F(X_t, t)$ solves the Itô equation

$$\begin{aligned} dY_t &= \underbrace{\nabla_x F(X_t,t)\cdot dX_t + \frac{\partial F}{\partial t}(X_t,t)dt}_{BV\ part\ (by\ chain\ rule)} + \underbrace{\frac{1}{2}dX_t\cdot\nabla_x^2 F(X_t,t)dX_t}_{QV\ part} \\ &= \left(\frac{\partial F}{\partial t} + \sum_{i=1}^n \frac{\partial F}{\partial x_i}b_i + \frac{1}{2}\left(\sigma\sigma^\top : \nabla_x^2 F\right)\right)(X_t,t)\,dt + \left(\sigma^\top\nabla_x F\right)(X_t,t)\cdot dB_t \end{aligned}$$

where $A\colon B = (A^T B)$ denotes the matrix inner product, and we have obtained the last equation by substituting $dX_t = b(X_t,t)dt + \sigma(X_t,t)dB_t$, using the rules

$$dtdt = dtdB_t^i = dB_t^j dt = 0 \;\text{ and }\; dB_t^i dB_t^j = \delta_{ij}dt \quad (i,j = 1,\ldots,d)\,,$$

where B_t^i denotes the i-th component of B_t.

Remark 7.9. *The matrix family $a(\cdot,\cdot) := \sigma\sigma^\top : \mathbb{R}^n\times\mathbb{R} \to \mathbb{R}^{n\times n}$ is sometimes called the* diffusion matrix.

Remark 7.10. *Note that, for functions that have a quadratic variation, Itô's formula is what is chain rule for functions of bounded variation. Dropping the dependence on (X_t,t) for the moment, one may rewrite the last equation as*

$$dY_t = \left(\frac{\partial F}{\partial t} + \sum_{i=1}^n \frac{\partial F}{\partial x_i}b_i + \frac{1}{2}\sum_{i,j=1}^n a_{ij}\frac{\partial^2 F}{\partial x_i\partial x_j}\right)dt + \sum_{i=1}^d\left(\sum_{j=1}^n \sigma_{ij}\frac{\partial F}{\partial x_j}\right)dB_t^i\,.$$

Historical remarks

Itô's original work was published in 1951.[4] However it was recently revealed that in 1940 Wolfgang Döblin, brother of novelist Alfred Döblin, French-German mathematician and student of Maurice Fréchet and Paul Lévy, sent a sealed letter to the Académie Française, while he was on the German front with the French army (as a telephone operator). Döblin committed suicide before he was captured by the German troops and burned all his mathematical notes. According to Döblin's last will, the letter to the Académie Française was opened in the year 2000 and found to contain a proof of Itô's lemma.[5]

7.3 Geometric Brownian motion

Consider the *Geometric Brownian motion* $S = (S_t)_{t\geq 0}$ that is the solution of the Itô stochastic differential equation (SDE)

$$dS_t = \mu S_t dt + \sigma S_t dB_t\,, \quad S_0 > 0.$$

[4] Kiyoshi Itô, 1915–2008, Japanese Mathematician; the famous lemma appeared in [K. Itô, On stochastic differential equations, *Memoirs AMS* **4**, 1–51, 1951].

[5] For a summary of Döblin's work, see [B. Bru and M. Yor, Comments on the life and mathematical legacy of Wolfgang Doeblin, Finance Stochast. **6**, 4–47, 2002]

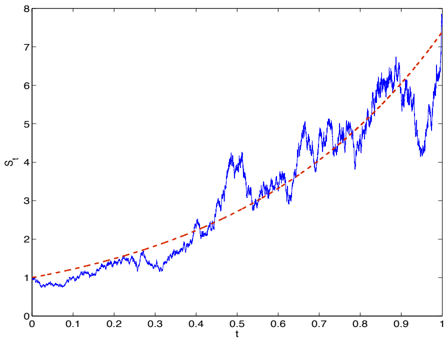

Figure 7: Typical realization of Geometric Brownian Motion $(S_t)_{t\in[0,1]}$ for $\mu = 2$ and $\sigma = 1$. The red dashed line shows the mean $\mathbb{E}[S_t]$.

We claim that the solution to the SDE is

$$S_t = S_0 \exp\left[\left(\mu - \frac{\sigma^2}{2}\right) t + \sigma B_t\right].$$

This can be seen as follows: using Itô's formula for $F(x) = \log x$, we find

$$\begin{aligned} Y_t &= \log S_t \\ \Rightarrow dY_t &= \frac{dS_t}{S_t} - \frac{\sigma^2}{2} \frac{S_t^2}{S_t^2} dt \\ &= \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma dB_t \end{aligned}$$

and therefore

$$\begin{aligned} Y_t &= Y_0 + \int_0^t \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma \int_0^t dB_t \\ &= Y_0 + \left(\mu - \frac{\sigma^2}{2}\right) t + \sigma B_t \end{aligned}$$

which proves that S_t follows the log-normal distribution with mean $\left(\mu - \frac{\sigma^2}{2}\right) t$ and variance σ^2. Moreover,

$$\mathbb{E}\left[S_t\right] = \exp\left(\mu t\right) .$$

Remark 7.11. *The geometric Brownian motion is sometimes used to model the growth of one's wealth subject to some positive interest rate $\mu > 0$ and random fluctuations due to market conditions, represented by the volatility-modified Brownian motion term σB_t. Good to know that $\mathbb{E}[S_t] = \exp(\mu t)$.*

It is also known, however, that the Brownian motion satisfies the Law of the

Iterated Logarithm (see, e.g., [Øks03, Thm. 5.1.2])

$$\limsup_{t\to\infty} \frac{B_t}{\sqrt{2t\log\log t}} = +1$$
$$\liminf_{t\to\infty} \frac{B_t}{\sqrt{2t\log\log t}} = -1\,,$$

which states that Brownian grows sublinearly. Since

$$S_t = S_0 \exp\left[\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B_t\right]$$

it follows that depending on different values of μ and σ, the wealth process $(S_t)_{t\geq 0}$ is dominated by the linear drift term. Indeed S, can exhibit rather different behaviours in the limit as $t \to \infty$, depending on the values of μ and σ:

(a) If $\mu < \frac{\sigma^2}{2}$, then $S_t \to 0$ as $t \to \infty$.

(b) If $\mu = \frac{\sigma^2}{2}$, then $\limsup_{t\to\infty} S_t = \infty$, $\liminf_{t\to\infty} S_t = 0$.

(c) If $\mu > \frac{\sigma^2}{2}$, then $S_t \nearrow \infty$ as $t \to \infty$.

(All the statements above hold $\mathbb{P}$-almost surely.) The mind-blowing aspect of geometric Brownian motion is the seemingly contradictory property that, even though the expected value grows exponentially with time for every volatility value σ, the process will hit zero with probability 1 whenever the volatility is sufficiently large, i.e. when $\sigma > \sqrt{2\mu}$. That is, even though the expected wealth grows exponentially (and thus never hits zero), for $\mathbb{P}$-almost all ω, all the wealth will vanish due to fluctuations in the long time limit, i.e., every single market player goes broke with probability one. Think about it!

8 Day 8, 04.12.2012

8.1 Short reminder: Itô's Formula

Given $F \in C^2(\mathbb{R}^n)$ and an Itô SDE

$$dX_t = b(X_t,t)dt + \sigma(X_t,t)dB_t\,,$$

the process $(Y_t)_{t\geq 0}$, $Y_t := F(X_t)$ solves the Itô SDE

$$\begin{aligned} dY_t &= \nabla F(X_t)\cdot dX_t + \frac{1}{2}dX_t \cdot \nabla^2 F(X_t)dX_t \\ &= \left(\nabla F \cdot b + \frac{1}{2}\sigma\sigma^\top : \nabla^2 F\right)dt + \left(\sigma^\top \nabla F\right)\cdot dB_t \end{aligned}$$

where $u\cdot v = u^\top v$ denotes the usual inner product between vectors and $A : B = \mathrm{tr}(A^\top B)$ is the inner product between matrices. Furthermore, in the step from the first to the second line, we have used the rule that

$$dtdt = dtdB_t^i = dB_t^i dt = 0 \quad dB_t^i dB_t^j = \delta_{ij}dt\,, \quad i,j = 1,\dots,n\,.$$

If we define the second-order differential operator (the infinitesimal generator of the stochastic "flow" X_t that will be introduced later on)

$$\mathcal{L}\varphi = \frac{1}{2}\sigma\sigma^\top : \nabla^2\varphi + b \cdot \nabla\varphi\,,$$

Itô's formula may be rewritten as

$$dY_t = (\mathcal{L}F)\,(X_t,t)dt + \left(\sigma^\top \nabla F\right)(X_t,t) \cdot dB_t.$$

Remark 8.1. *The fact that the usual chain does not apply for Itô processes has to do with the definition of the corresponding stochastic integral. We shall briefly comment on what makes Itô integral special.*

(i) The Martingale Representation Theorem *states that every $\mathcal{F}_t$-martingale $(X_t)_{t\geq 0}$ (i.e., X_t is adapted to the filtration generated by B_t and satisfies $X_s = \mathbb{E}\left[X_t|\mathcal{F}_s\right]$) can be written as the integral*

$$X_t = X_0 + \int_0^t \phi(\omega,s)dB_s$$

for a function $\phi \in \mathcal{V}(0,t)$, that is uniquely determined. Conversely, every Itô integral of the form

$$\int_0^t \phi_s dB_s$$

is a martingale with respect to $(\mathcal{F}_t)_{t\geq 0}$

(ii) The Stratonovich integral *is another stochastic integral, distinct from the Itô integral, that is based on the midpoint rule, i.e.,*

$$\int \psi(\omega,s) \circ dB_s = \lim_{n\to\infty} \sum_j \psi\left(\omega, t_{j+\frac{1}{2}}\right)\left(B_{t_{j+1}} - B_{t_j}\right)$$

where we emphasize the different notation using the "$\circ$" symbol and where

$$t_{j+\frac{1}{2}} := \frac{t_j + t_{j+1}}{2}.$$

The Stratonovich integral has the property that

$$\mathbb{E}\left[\int \psi_s \circ dB_s\right] \neq 0\,,$$

hence the Stratonovich integral is not a martingale, unlike the Itô integral. Furthermore, the thus defined integral does not satisfy the Itô Isometry. On the other hand, the usual chain rule applies.

(iii) The Stratonovich integral is used for integrating Stratonovich SDEs

$$dX_t = b(X_t,t)dt + \sigma(X_t,t) \circ dB_t\,,$$

where, again, the "$\circ$" indicates that the SDE has to be interpreted in the Stratonovich sense (i.e., integrated using the Stratonovich integral). One can also convert Stratonovich SDEs to Itô SDEs using the conversion rule

$$dX_t = b(X_t,t)dt + \sigma(X_t,t) \circ dB_t = \left(b + \frac{1}{2}\sigma\nabla\sigma\right)(X_t,t)dt + \sigma(X_t,t)dB_t.$$

8.2 Stochastic Differential Equations (Itô SDEs)

We now want to find possible solutions $(X_t)_{t\geq 0} \subset \mathbb{R}^n$ for Itô SDEs of the form

$$dX_t = b(X_t, t)dt + \sigma(X_t, t)dB_t. \tag{13}$$

where $b : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ and $\sigma : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^{n\times d}$ are measurable functions.

Definition 8.2 (Strong solution)**.** *Let $T > 0$. A process $(X_t)_{t\in[0,T]}$ is called a* strong solution *of (13) if the map $t \mapsto X_t$ is almost surely continuous and adapted to the filtration generated by the Brownian motion $(B_t)_{t\in[0,T]}$ and if it holds for $\mathbb{P}$-almost all ω (i.e., if it holds $\mathbb{P}$-almost surely) that*

$$X_t(\omega) = X_0(\omega) + \int_0^t b(X_s(\omega), s)ds + \int_0^t \sigma(X_s, \omega)dB_s(\omega) \quad (\omega \textit{ fixed}).$$

Definition 8.3 (Uniqueness)**.** *The solution of (13) is called* unique *or* pathwise unique *if*

$$\mathbb{P}(X_0 = \tilde{X}_0) \quad \Rightarrow \quad \mathbb{P}(X_t = \tilde{X}_t) \ \forall t \in [0, T]$$

for any two solutions $(X_t)_{t\in[0,T]}$ and $(\tilde{X}_t)_{t\in[0,T]}$ of (13).

Theorem 8.4 (Existence and uniqueness)**.** *Let $T > 0$ and b, σ in (13) satisfy*

(i) (Global Lipschitz condition): $\forall x, y \in \mathbb{R}^n$, $\forall t \in [0, T]$,

$$|b(x, t) - b(y, t)| + |\sigma(x, t) - \sigma(y, t)| \leq L|x - y|$$

for some constant $0 < L < \infty$.[6]

(ii) (Sublinear growth condition): $\forall x, \in \mathbb{R}^n$ and $\forall t \in [0, T]$,

$$|b(x, t)| + |\sigma(x, t)| \leq G(1 + |x|)$$

for some $0 < G < \infty$.

Given that the above conditions hold, if we have $\mathbb{E}\left[X_0^2\right] < \infty$, then (13) has a pathwise unique, strong solution for any $T > 0$.

Proof. See [Øks03, Thm 5.2.1]. The main elements of the proof of the above theorem are the Itô Isometry and a Picard-Lindelöf-like fixed-point iteration, just as in case of ordinary differential equations. □

Example 8.5. *We now consider some examples which use Itô's formula.*

(i) (Geometric Brownian motion)*: the Itô SDE in this example is*

$$dS_t = \mu S_t dt + \sigma S_t dB_t, \quad S_0 > 0$$

and the solution to this SDE is

$$S_t = S_0 \exp\left[\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B_t\right]$$

[6] Here the norm on the matrix terms $\sigma(\cdot,\cdot)$ is arbitrary and may be taken to be, e.g., the Frobenius norm $|\sigma| = (\sum_{i,j} |\sigma_{ij}|^2)^{1/2}$.

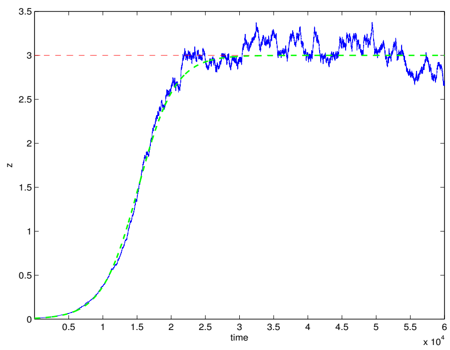

Figure 8: Typical realization of the logistic growth model with $r = 0.5$, $C = 3$ and $\sigma = 0.1$; for comparison, the dashed green curve shows the deterministic model with $\sigma = 0$; the red straight line shows the capacity bound C.

(ii) (Ornstein-Uhlenbeck process)*: the OU process is a Gaussian process whose evolution is given by*

$$dX_t = AX_t dt + BdW_t\,, \quad X_0 = x_0$$

for $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times d}$ *(i.e.,* W_t *is* d*-dimensional Brownian motion). The solution to the above SDE is*

$$X_t = e^{At}x_0 + \int_0^t e^{A(t-s)}B\,dW_s$$

(iii) (Brownian Bridge)*: the Brownian bridge* $(BB_t)_{t\in[0,1]}$ *is a Gaussian process with the property* $BB_0 = BB_1 = 0$*. The associated SDE is*

$$dBB_t = \frac{1-BB_t}{1-t}dt + dB_t\,, \quad BB_0 = 0.$$

The solution to the above SDE is

$$BB_t = (1-t)\int_0^t \frac{dB_s}{1-s}\,.$$

The proof is left as an exercise.

(iv) (Logistic growth model)*: Consider the ODE*

$$\frac{dz}{dt} = rz\,(C-z)\,, \quad z(0) > 0$$

that describes logistic growth of a population where $r > 0$ *is the (initial) growth rate and* $C > 0$ *is the capacity bound. If we add random perturbations we obtain the Itô SDE*

$$dZ_t = rZ_t(C-Z_t)dt + \sigma Z_t dB_t\,, \quad Z_0 > 0.$$

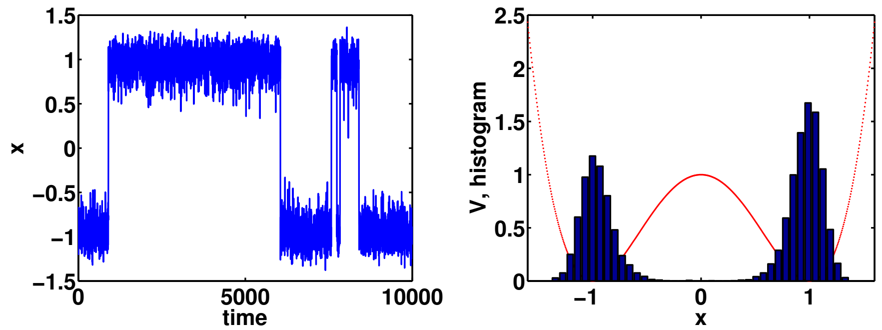

Figure 9: Typical realization of (14) with a bistable potential V; the solution has been computed using the Euler method (see the next section).

which describes logistic growth in a random environment (see Figure 8). This is an interesting example because the drift coefficient is not globally Lipschitz—even worse, the drift term can become unbounded. Despite this, one can obtain an analytic solution to the SDE above, given by

$$\hat{Z}_t = \frac{\exp\left[\left(rC - \frac{\sigma^2}{2}\right)t + \sigma B_t\right]}{\hat{Z}_0^{-1} + \int_0^t \exp\left[\left(rC - \frac{\sigma^2}{2}\right)s + \sigma B_s\right]ds}.$$

8.3 Central issues for numerical methods for SDEs

We now consider some key properties which we shall use in evaluating the quality of a numerical method for solving Itô SDEs of the form

$$dX_t = -\nabla V(X_t)dt + \sqrt{2\epsilon}dB_t \tag{14}$$

with a smooth function $V\colon \mathbb{R} \to \mathbb{R}$ (13) that will not meet the requirements of the existence and uniqueness theorem in general.

(i) There are various choices for stable numerical schemes for solving equations like (14). As we will see below, one such choice is

$$\tilde{X}_{n+1} - \tilde{X}_n = -\Delta t \nabla(\tilde{X}_n) + \sqrt{2\epsilon\Delta t}\,\xi_{n+1}\,,$$

where $\Delta t > 0$ and the ξ_n are suitable i.i.d. random variables, e.g., standard normal or uniform on the set $\{-1, 1\}$, such that $\tilde{X}_n \approx X(t_n)$ on a sufficiently fine grid $0 = t_0 < t_1 < t_3 < \ldots$ with $\Delta t = t_{n+1} - t_n$.

(ii) The numerical scheme under (i) can be shown to yield an approximation to the continuous SDE on any *finite* time interval, but diverges when $n \to \infty$. On the other hand, we may be interested in the limiting behaviour, specifically in the stationary distribution of the process (if it exists). In our case, (an under certain technical assumption on V), the process satisfies

$$(\mathbb{P} \circ X_t^{-1})(A) \to \int_A e^{-V/\epsilon}dx$$

as $t \to \infty$ and for all Borel sets $A \subset \mathbb{R}$ (assuming that the integral on the right hand side is properly normalized). For the continuous process convergence of the distribution can be shown to hold in L^1, but it may be very slow if ϵ in (14) is small. For the discrete approximation this question of convergence does not have an easy answer, for the standard numerical schemes are not asymptotically stable and the numerical discretization introduces a bias in the stationary distribution (see Figure 9).

(iii) Can we compute functionals of paths of X_t? For example, can we compute quantities, such as

$$\mathbb{E}\left[\phi(X_T)\right], \quad \mathbb{E}\left[\int_0^T \psi(X_t, t)dt\right]$$

for bounded continuous functions ϕ, ψ and $T > 0$, or can we compute

$$\mathbb{E}[\tau | X_0 = x]$$

with τ being some random stopping time (e.g., a first hitting time of a set $E \subset \mathbb{R}$). Questions dealing with such functionals, but also the long-term stability issue under (ii) will lead us to Markov Chain Monte-Carlo (MCMC) methods for PDEs and the celebrated Feynman-Kac formula.

9 Day 9, 11.12.2012

9.1 Stochastic Euler Method

We motivate the ideas in this method by considering the deterministic initial value problem

$$\frac{dx}{dt} = b(x, t), \quad x(0) = x_0,$$

for $t \in [0, T]$. The initial value problem has the solution

$$x(t) = x_0 + \int_0^t b(x(s), s)ds, \quad t \in [0, T]$$

and we can approximate the true solution, employing a suitable quadrature rule for the integral, e.g., the "rectangle rule":

$$\begin{aligned} x(t_{n+1}) &= x(t_n) + \int_{t_n}^{t_{n+1}} b(x(s), s)ds \\ &\approx x(t_n) + \int_{t_n}^{t_{n+1}} b(x(t_n), t_n)ds \\ &= x(t_n) + b(x(t_n), t_n)\,(t_{n+1} - t_n) . \end{aligned}$$

Given a sufficiently fine grid of time nodes $0 = t_0 < t_1 < \ldots < t_N = T$ with fixed time step $\Delta t = t_{n+1} - t_n$, we recognize the forward Euler method

$$x_{n+1} = x_n + \Delta t b(x_n, t_n) .$$

The discretization error induced by the Euler scheme can be shown to satisfy

$$\sup_{n=1,\ldots,N} |x(t_n) - x_n| \le C\Delta t.$$

for a $0 < C < \infty$ that is independent of Δt.

9.2 Simple quadrature rule for the Itô integral

Now let $(X_t)_{t\in[0,T]}$ solve the Itô SDE

$$dX_t = b(X_t,t)dt + \sigma(X_t,t)dB_t, \quad X_0 = x$$

with $b : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ and $\sigma : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^{n\times m}$ for $(B_t)_{t>0}$ a m-dimensional Brownian motion. We wish to approximate $(X_t)_{t>0}$ on the uniform grid

$$\{0 = t_0 < t_1 < \ldots < t_N = T\}\ , \quad \Delta t = t_{n+1} - t_n\ .$$

The "rectangle rule" for the solution between $[t_n, t_{n+1}] \subset [0,T]$ is

$$\begin{aligned} X_{t_{n+1}} &= X_{t_n} + \int_{t_n}^{t_{n+1}} b(X_s,s)ds + \int_{t_n}^{t_{n+1}} \sigma(X_s,s)dB_s \\ &\approx X_{t_n} + \int_{t_n}^{t_{n+1}} b(X_{t_n},t_n)ds + \int_{t_n}^{t_{n+1}} \sigma(X_{t_n},t_n)dB_s \\ &= X_{t_n} + \Delta t b(X_{t_n},t_n) + \underbrace{\left(B_{t_{n+1}} - B_{t_n}\right)}_{=:\Delta B_n \sim N(0,\Delta t)} \sigma(X_{t_n},t_n). \end{aligned}$$

Definition 9.1. (Euler-Maruyama scheme)*: For $n = 0,\ldots,N-1$, the* Euler-Maruyama scheme *or* Euler's method *gives the n-th iterate as*

$$\tilde{X}_{n+1} = \tilde{X}_n + \Delta t b(\tilde{X}_n,t_n) + \sigma(\tilde{X}_n,t_n)\Delta B_n, \quad \tilde{X}_0 = x$$

Remark 9.2. *A few remarks are in order.*

(i) Euler's method is consistent with the definition of the Itô integral, in that is evaluates the integrand at the left endpoint of the interval.

(ii) The Euler method gives the values of the numerical path at the time nodes. A numerical path is obtained by linear interpolation: for $t \in [t_n, t_{n+1}]$,

$$\begin{aligned} \tilde{X}_t(\omega) &= \tilde{X}_n(\omega) + \frac{(t-t_n)}{\Delta t}\left(\tilde{X}_{n+1}(\omega) - \tilde{X}_n(\omega)\right) \\ &= \tilde{X}_n + (t-t_n)\, b(\tilde{X}_n,t_n) + \frac{t-t_n}{\Delta t}\sigma(\tilde{X}_n,t_n)\left(B_{t_{n+1}} - B_{t_n}\right). \end{aligned}$$

Note that $\tilde{X}_t$, $t \le t_{n+1}$ depends on $B_{t_{n+1}}$, i.e., the interpolant $\tilde{X}_t$ is not non-anticipating.

(iii) Sometimes one wishes to refine the partition for a specific realization, i.e. using the same path ω. For example, by halving the time step from from Δt to $\Delta t/2$, one obtains new grid points

$$t_{n+\frac{1}{2}} = t_n + \frac{\Delta t}{2}$$

The values of the refined Brownian motion can be computed by the rule

$$B_{t_{n+\frac{1}{2}}}(\omega) = \frac{1}{2}\left[B_{t_n}(\omega) + B_{t_{n+1}}(\omega)\right] + \frac{1}{2}\sqrt{\Delta t}\,\xi_n(\omega)$$

where the $\xi_n \sim N(0,1)$ are i.i.d. standard normal random variables.

9.3 Convergence of Euler's method

We can guess that $\tilde{X} \approx X(t_n)$, but in which sense?

Example 9.3. *Let $X_t = B_t$ and $Y_t = -B_t$; then $X_t \sim Y_t$ (that is, $(X_t)_{t>0}$ and $(Y_t)_{t>0}$ have the same distribution) but $|X_t - Y_t| = 2|B_t|$ is unbounded for all t. Hence pathwise comparisons may not be very informative.*

Definition 9.4. (Strong convergence)*: Let X_{t_n} denotes the value of the true solution $(X_t)_{t\in[0,T]}$ of our SDE at the time t_n. A numerical scheme $(\tilde{X}_n)_n = (\tilde{X}_n^{\Delta t})_{n=0,\dots,N-1}$ is called* strongly convergent *of order $\gamma > 0$ if*

$$\max_{n=0,\dots,N-1} \mathbb{E}[|\tilde{X}_n - X_{t_n}|] \leq C\Delta t^\gamma$$

where $0 < C < \infty$ is independent of Δt but can depend on the length $T = N\Delta t$ of the time interval.

Definition 9.5. (Weak convergence)*: A numerical scheme $(\tilde{X}_n)_n$ is called* weakly convergent *of order $\delta > 0$ if*

$$\max_{n=0,\dots,N-1} \left|\mathbb{E}[f(\tilde{X}_n)] - \mathbb{E}[f(X_{t_n})]\right| \leq D\Delta t^\delta$$

for all functions f in a suitably chosen class of functions (e.g. the space of continuous, bounded functions $C_b(\mathbb{R}^n$, or the space of polynomials of degree k). The constant D is independent of Δt, but may depend function being considered.

Remark 9.6. *A mnemonic for the difference between strong and weak convergence is that strong convergence is about the mean of the error, while weak convergence is about the error of the mean.*

Lemma 9.7. *Let f be globally Lipschitz. Then strong convergence implies weak convergence. The converse does not hold*

Proof. Since f is globally Lipschitz, $\exists\ 0 < L < \infty$ such that

$$|f(x) - f(y)| \leq L|x - y| \quad \forall x, y \in \mathbb{R}^n.$$

Then

$$\begin{aligned}\left|\mathbb{E}[f(\tilde{X}_n)] - \mathbb{E}[f(X_{t_n})]\right| &= \left|\mathbb{E}[f(\tilde{X}_n) - f(X_{t_n})]\right| \\ &\leq \mathbb{E}[|f(\tilde{X}_n) - f(X_{t_n})|] \\ &\leq L\mathbb{E}[|\tilde{X}_n - X_{t_n}|].\end{aligned}$$

Now consider the converse statement. Let $\tilde{X}_n = -X_{t_n}$, with $\mathbb{E}[\tilde{X}_n] = 0$ and $\mathbb{E}[|\tilde{X}_n|] \neq 0$. Then

$$|\mathbb{E}[\tilde{X}_n - \mathbb{E}[X_{t_n}]| = |0 - 0| = 0\,,$$

but

$$\mathbb{E}[|\tilde{X}_n - X_{t_n}|] = 2\mathbb{E}[|\tilde{X}_n|] \neq 0\,,$$

which concludes the proof. □

The next theorem states that Euler's method is strongly (and hence weakly) convergent.

Theorem 9.8. *Let $T > 0$ and b, σ in (13) satisfy*

(i) (Global Lipschitz condition): $\forall x, y \in \mathbb{R}^n$, $\forall t \in [0, T]$,

$$|b(x,t) - b(y,t)| + |\sigma(x,t) - \sigma(y,t)| \leq L|x - y|$$

for some constant $0 < L < \infty$.

(ii) (Growth condition): $\forall x, \in \mathbb{R}^n$ and $\forall t \in [0, T]$,

$$|b(x,t)|^2 + |\sigma(x,t)|^2 \leq G(1 + |x|^2)$$

for some $0 < G < \infty$.

Then Euler's method is strongly convergent of order $\gamma = 1/2$ and weakly convergent of order $\delta = 1$.

The proof is essentially based on the integral version of Gronwall's lemma (as is common when Lipschitz conditions are involved):

Lemma 9.9 (Gronwall Lemma)**.** *Let $y\colon [0, T] \to \mathbb{R}$ be non-negative and integrable such that*

$$y(t) \leq A + B \int_0^t y(s)\, ds\,, \quad 0 \leq t \leq T\,,$$

for some constants $A, B > 0$. Then

$$y(t) \leq Ae^{Bt}\,, \quad 0 \leq t \leq T\,.$$

Now we can prove that the Euler-Maruyama scheme convergences.

Proof of Theorem 9.8. We will show only strong convergence and leave the weak convergence part as an exercise. Without loss of generality, we assume that $b(x,t) = b(x)$ and $\sigma(x,t) = \sigma(x)$ are independent of t and that $x \in \mathbb{R}$ is scalar—this will greatly simplify the notation. Since $L^2(\Omega, P) \subset L^1(\Omega, P)$, i.e.,

$$\mathbb{E}[|\tilde{X}_n - X_{t_n}|] \leq \sqrt{\mathbb{E}[|\tilde{X}_n - X_{t_n}|^2]}\,,$$

it suffices to prove that

$$\mathbb{E}[|\tilde{X}_n - X_{t_n}|^2] \leq C^2 \Delta t$$

for sufficiently small Δt.

Now let $\tau \in [0, T)$ and define $n_\tau \in \mathbb{N}$ by $\tau \in [t_{n_\tau}, t_{n_\tau+1})$ with $t_k = k\Delta t$. Further let $\bar{X}_\tau = \tilde{X}_{n_\tau}$ be the piecewise constant interpolant of the time-discrete

Markov chain $\tilde{X}_0$, $\tilde{X}_1$, $\tilde{X}_2, \ldots$. Then

$$\begin{aligned}
\bar{X}_\tau - X_\tau &= \tilde{X}_{n_\tau} - \left(x + \int_0^\tau b(X_s)\,ds + \int_0^\tau \sigma(X_s)\,dB_s \right) \\
&= \sum_{i=0}^{n_\tau - 1} (\tilde{X}_{i+1} - \tilde{X}_i) - \int_0^\tau b(X_s)\,ds - \int_0^\tau \sigma(X_s)\,dB_s \\
&= \sum_{i=0}^{n_\tau - 1} (b(\tilde{X}_i)\Delta t + \sigma(\tilde{X}_i)\Delta B_{i+1}) - \int_0^\tau b(X_s)\,ds - \int_0^\tau \sigma(X_s)\,dB_s \\
&= \int_0^{t_{n_\tau}} b(\bar{X}_s)\,ds + \int_0^{t_{n_\tau}} \sigma(\bar{X}_s)\,dB_s - \int_0^\tau b(X_s)\,ds - \int_0^\tau \sigma(X_s)\,dB_s \\
&= \underbrace{\int_0^{t_{n_\tau}} (b(\bar{X}_s) - b(X_s))\,ds + \int_0^{t_{n_\tau}} (\sigma(\bar{X}_s) - \sigma(X_s))\,dB_s}_{\text{discretization error}} \\
&\quad \underbrace{- \left(\int_{t_{n_\tau}}^\tau b(X_s)\,ds + \int_{t_{n_\tau}}^\tau \sigma(X_s)\,dB_s \right)}_{\text{interpolation error}}
\end{aligned}$$

Squaring both sides of the equality and taking the expectation, it follows with the inequality $(a+b+c+d)^2 \le 4(a^2+b^2+c^2+d^2)$,

$$\begin{aligned}
\mathbb{E}[|\bar{X}_\tau - X_\tau|^2] \le 4\mathbb{E}&\left[\left(\int_0^{t_{n_\tau}} (b(\bar{X}_s) - b(X_s))\,ds \right)^2 \right] \\
&+ 4\mathbb{E}\left[\left(\int_0^{t_{n_\tau}} (\sigma(\bar{X}_s) - \sigma(X_s))\,dB_s \right)^2 \right] \\
&+ 4\mathbb{E}\left[\left(\int_{t_{n_\tau}}^\tau b(X_s)\,ds \right)^2 \right] + 4\mathbb{E}\left[\left(\int_{t_{n_\tau}}^\tau \sigma(X_s)\,dB_s \right)^2 \right]
\end{aligned}$$

We will now estimate the right hand side of the inequality term by term, using Lipschitz and growth conditions:

1. Noting the inner Riemann integral can be interpreted as a scalar product between the functions $g(s) = 1$ and $f(s) = b(\bar{X}_s) - b(X_s)$, we find

$$\begin{aligned}
\mathbb{E}\left[\left(\int_0^{t_{n_\tau}} (b(\bar{X}_s) - b(X_s))\,ds \right)^2 \right] &\underbrace{\le t_{n_\tau} \int_0^{t_{n_\tau}} \mathbb{E}[|b(\bar{X}_s) - b(X_s)|^2]\,ds}_{\text{Cauchy-Schwarz \& Fubini}} \\
&\underbrace{\le TL^2 \int_0^{t_{n_\tau}} \mathbb{E}[|\bar{X}_s - X_s|^2]\,ds}_{\text{Lipschitz bound}} \,.
\end{aligned}$$

2. For the stochastic integral, the Itô isometry implies

$$\mathbb{E}\left[\left(\int_0^{t_{n_\tau}} (\sigma(\bar{X}_s) - \sigma(X_s))\, dB_s\right)^2\right] \underbrace{= \int_0^{t_{n_\tau}} \mathbb{E}[|\sigma(\bar{X}_s) - \sigma(X_s)|^2]\, ds}_{\text{Itô isometry}}$$
$$\underbrace{\leq L^2 \int_0^{t_{n_\tau}} \mathbb{E}[|\bar{X}_s - X_s|^2]\, ds}_{\text{Lipschitz bound}} .$$

3. For the interpolation error coming from the drift, we have

$$\mathbb{E}\left[\left(\int_{t_{n_\tau}}^{\tau} b(X_s)\, ds\right)^2\right] \underbrace{\leq (\tau - t_{n_\tau}) \int_{t_{n_\tau}}^{\tau} \mathbb{E}[|b(X_s)|^2]\, ds}_{\text{Cauchy-Schwarz \& Fubini}}$$
$$\underbrace{\leq \Delta t\, G \int_{t_{n_\tau}}^{\tau} (1 + \mathbb{E}[|X_s|^2])\, ds}_{\text{sublinear growth}}$$
$$\leq M_1 (\Delta t)^2 ,$$

for a constant $0 < M_1 < \infty$. In the last step we have used that $\mathbb{E}[|X_s|^2]$ is finite by the assumptions on the coefficients b and σ.

4. Finally, using the Itô isometry again, we can bound the remaining stochastic integral by

$$\mathbb{E}\left[\left(\int_{t_{n_\tau}}^{\tau} \sigma(X_s)\, dB_s\right)^2\right] \underbrace{= \int_{t_{n_\tau}}^{\tau} \mathbb{E}[|\sigma(X_s)|^2]\, ds}_{\text{Itô isometry}}$$
$$\underbrace{\leq G \int_{t_{n_\tau}}^{\tau} (1 + \mathbb{E}[|X_s|^2])\, ds}_{\text{sublinear growth}}$$
$$\underbrace{\leq M_2 \Delta t}_{\mathbb{E}[|X_s|^2] < \infty} ,$$

for a constant $0 < M_2 < \infty$.

Setting $y(t) = \mathbb{E}[|\bar{X}_t - X_t|^2]$, the assertion follows from Gronwall's lemma with $A = M\Delta t$ for a $M > (M_1 \Delta t + M_2)$ and $B = L^2(1 + T)$. □

Remark 9.10. *Note that the (strong) error bound for Euler's method grows exponentially with T, hence becomes essentially of order one if $T = \mathcal{O}(-\log \Delta t)$.*

Remark 9.11. *We shall briefly comment on some implementation issues.*

(i) The standard implementation of Euler's method is

$$\tilde{X}_{n+1} = \tilde{X}_n + \Delta t b(\tilde{X}_n, t_n) + \sqrt{\Delta t} \sigma(\tilde{X}_n, t_n) \xi_{n+1}$$

where the ξ_n are standard normal, i.i.d. random variables.

(ii) *The simplified Euler method uses i.i.d. $\xi'_n \sim U(\{\pm 1\})$ random variables, i.e. $\mathbb{P}(\xi'_n = 1) = \mathbb{P}(\xi'_n = -1) = 1/2$. The advantage of the simplified Euler scheme is that it is much faster to generate $U(\{\pm 1\})$ random variables than $N(0,1)$ random variables. The disadvantage of the simplified Euler scheme is that it gives only weak convergence.*

Exercise 9.12. *Show that the simplified Euler method is weakly convergent.*

References

[Arn73] Ludwig Arnold. *Stochastische Differentialgleichungen: Theorie und Anwendung.* Wiley & Sons, 1973.

[Dur96] Richard Durrett. *Stochastic Calculus: A Practical Introduction.* CRC Press, 1996.

[GHL09] Xianping Guo and Onésimo Hernández-Lerma. *Continuous-Time Markov Decision Processes: Theory and Applications.* Springer, Heidelberg, 2009.

[HM05] Wilhelm Huisinga and Eike Meerbach. *Markov chains for everybody.* Lecture Notes, Fachbereich Mathematik und Informatik, Freie Universität Berlin, 2005.

[Jaz07] Andrew H. Jazwinski. *Stochastic Processes and Filtering Theory.* Dover, New York, 2007.

[Kle06] Achim Klenke. *Wahrscheinlichkeitstheorie.* Springer, Berlin, 2006.

[KP92] Peter E. Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations.* Springer, Berlin, 1992.

[KS91] Ioannis Karatzas and Steven Shreve. *Brownian Motion and Stochastic Calculus.* Springer, New York, 1991.

[MP10] Peter Mörters and Yuval Peres. *Brownian Motion (Cambridge Series in Statistical and Probabilistic Mechanics).* Cambridge University Press, 2010.

[MS05] David Meintrup and Stefan Schäffler. *Stochastik.* Springer, Berlin, 2005.

[Øks03] Bernt K. Øksendal. *Stochastic Differential Equations: An Introduction With Applications.* Springer, Berlin, 2003.

[Sar11] Marco Sarich. *Multiscale Stochastic Processes.* Lecture Notes, Fachbereich Mathematik und Informatik, Freie Universität Berlin, 2011.

[Xiu10] Dongbin Xiu. *Numerical Methods for Stochastic Computations: A Spectral Method Approach.* Princeton University Press, 2010.