Lecture Note for MAT7093: Stochastic Analysis

LI Liying

February 19, 2024

1 Introduction

In this section we will give some motivations to study Brownian motions and stochastic integrals.

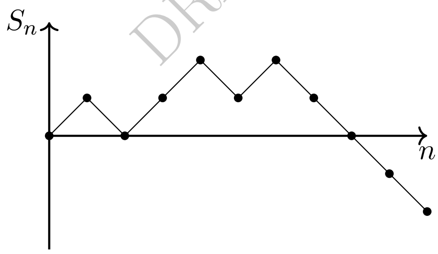
1.1 Stochastic processes

The well-known Central Limit Theorem (CLT) gives the universal behavior of the sum of many small independent variables: for i.i.d. r.v.'s X_i with $\mathsf{E}X_i = 0$, $\mathsf{E}X_i^2 = 1$, one has

$$\frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}} \Rightarrow_d \mathcal{N}(0,1).$$

Example 1.1 We can take X_i as the coin-flip results, $\mathsf{P}(X_i = \pm 1) = 1/2$.

Write the partial sum as $S_n = X_1 + X_2 + \cdots + X_n$. We can plot the $n \mapsto S_n$ as below:



The linear interpolation (n, S_n), $n \in \mathbb{N}$ can be formally written as

$$\tilde{S}_t = \begin{cases} S_n, & t = n \in \mathbb{N}, \\ (n+1-t)S_n + (t-n)S_{n+1}, & t \in (n, n+1). \end{cases}$$

Question *What is the limit of $t \mapsto \tilde{S}_t$ as (continuous) trajectories?*

The *Donsker's invariance principle*, a.k.a. the *Functional CLT*, states that in an appropriate sense, the limit is given by the *Brownian motion*, which is a "continuous stochastic process".

Theorem 1.1 (Functional CLT)

$$\Big(\frac{\tilde{S}_{[nt]}}{\sqrt{n}},\ t \geq 0\Big) \Rightarrow_d \Big(B_t,\ t \geq 0\Big),$$

where $(B_t)_{t\geq 0}$ is the Brownian motion (BM).

Remark 1.2 We will define rigorously what is a "continuous stochastic process" below.

Remark 1.3 The convergence "$\Rightarrow_d$" means convergence in distribution/law. Since our objects are random functions rather than random variables, this requires us to work on probability measures on infinite-dimensional spaces. We will do this in Section 1.2.

Using the CLT, we can obtain the finite-dimensional distribution (f.d.d.) for the BM. For fixed $t \geq 0$,

$$\mathcal{L}(B_t) = \lim_{n\to\infty} \mathcal{L}\Big(\frac{\tilde{S}_{[nt]}}{\sqrt{n}}\Big) = \lim_{n\to\infty} \mathcal{L}\Big(\frac{\tilde{S}_{[nt]}}{\sqrt{[nt]}} \cdot \sqrt{t}\Big) = \mathcal{N}(0, \sqrt{t}).$$

In general, for $0 < t_1 < t_2 < \cdots < t_m$, it is believable that

$$B_{t_1},\ B_{t_2 - t_1}, \cdots, B_{t_m} - B_{t_{m-1}}$$

should have the same distribution as independent $\mathcal{N}(0, \sqrt{t_1})$, $\mathcal{N}(0, \sqrt{t_2 - t_1}), \cdots, \mathcal{N}(0, \sqrt{t_m - t_{m-1}})$ r.v.'s.

Homework

1. Show that $\mathcal{L}(B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_m} - B_{t_{m-1}}) = \mathcal{N}(0, \mathrm{diag}\{\sqrt{t_{i+1} - t_i}\}_{0\leq i\leq m-1})$ is equivalent to the following two conditions:
   - $(B_{t_1}, B_{t_2}, \ldots, B_{t_m})$ is a centered Gaussian vector.
   - $\mathsf{E}B_{t_i}B_{t_j} = t_i \wedge t_j$. (Note: $x \wedge y = \min(x, y)$.)
2. Let $(B_t)_{t\geq 0}$ be the standard Brownian motion. Verified that the following processes all have the same f.d.d. as the standard Brownian motion:
   a) $(-B_t)_{t\geq 0}$.
   b) $(B_t^{\lambda})_{t\geq 0} := (\frac{1}{\lambda} B^{\lambda}_{\lambda^2 t})_{t\geq 0}$. (Fix $\lambda > 0$.)
   c) $(B_t^{(s)})_{t\geq 0} := (B_{t+s} - B_s)_{t\geq 0}$. (Fix $s > 0$.)
   d) $(tB_{1/t})_{t\geq 0}$ (with the convention $0 \cdot B_{1/0} = 0$).

Definition 1.1 *A stochastic process $(X_t)_{t\in T}$ ($T = \mathbb{Z}, \mathbb{R}$, etc) on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ is such that for every fixed $t \in T$,*

$$\omega \in \Omega \mapsto X_t(\omega)$$

is a measurable map from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Definition 1.2 *For a stochastic process $(X_t)_{t\in T}$, Its finite-dimensional distribution (f.d.d.) is the collection of all the laws*

$$\mathcal{L}(X_{t_1}, X_{t_2}, \ldots, X_{t_m}), \quad t_1, t_2, \ldots, t_m \in T.$$

It is well-defined by the measurability of all the sets of the form

$$\{(X_{t_1}, X_{t_2}, \ldots, X_{t_m}) \in A\}, \quad A \in \mathcal{B}(\mathbb{R}^m).$$

It is believable that a stochastic process is more or less completely determined by all its f.d.d.'s (which is in fact Komolgorov's Extension Theorem, see also XXX). With the definition of stochastic processes at hand, the next question is what makes a "continuous" stochastic process. Let T be an interval of $\mathbb{R}$ ($T = [a, b]$, $[0, \infty)$, etc). Then, a "continuous" process requires additionally the requirement that the map

$$t \mapsto X_t(\omega)$$

is *continuous* for P-a.e. ω.

Remark 1.4 For a generic stochastic process $(X_t)_{t\in\mathbb{R}}$, the sets

$$\mathcal{C} = \{\omega : t \mapsto X_t(\omega) \text{ is continuous.}\}$$

or for $t_0 \in T$,

$$\mathcal{C}_{t_0} = \{\omega : t \mapsto X_t(\omega) \text{ is continuous at } t = t_0.\}$$

are NOT measurable.

To see this, recall that we can characterize the continuity of a function by sequential convergence, namely,

$$\lim_{t\to t_0} f(t) = f(t_0) \quad \Leftrightarrow \quad \forall t_n \to t_0, \ \lim_{n\to\infty} f(t_n) = f(t_0).$$

Although for any fixed sequence (t_n), the set

$$\{\omega : \lim_{n\to\infty} X_{t_n} = X_{t_0})\} = \bigcap_{m=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{\omega : |X_{t_n} - X_{t_0}| < \frac{1}{m}\}$$

is in $\mathcal{F}$ (hence measurable), there are uncountably many such sequence (t_n) such that $t_n \to t_0$.

Homework Let $(X_n)_{n\geq 1}$, X_∞ be r.v.'s defined on $(\Omega, \mathcal{F}, \mathsf{P})$. Show that

$$\{\omega : \lim_{n\to\infty} X_n = X_\infty\} = \bigcap_{m=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{\omega : |X_n - X_\infty| < \frac{1}{m}\}$$

Conclude that the left hand side belongs to $\mathcal{F}$.

Due to the potential measurability issue, the continuity of a stochastic process is somehow an "independent" property to consider, so additional efforts are always needed for the justification. There are generally two approaches: one is to use Komolgorov's Continuity Test (see XX), the other one is to directly consider building up probability measures on the space on continuous functions (Section 1.2)

But assuming that this can be done, we can now rigorously defined what a Brownian motion is.

Definition 1.3 *Two stochastic processes $X = (X_t)_{t\in T}$, $Y = (Y_t)_{t\in T}$, defined on $(\Omega, \mathcal{F}, \mathsf{P})$, are called modifications of each other if*

$$\mathsf{P}(X_t = Y_t) = 1, \quad \forall t \in T.$$

That is, X and Y have the same f.d.d.'s.

Definition 1.4 *Y is called a version of X, or indistinguishable from X, if for a.e. ω,*

$$X_t = Y_t, \quad \forall t \in T.$$

Clearly, when T is uncountable, the above two definitions are not equivalent.

Remark 1.5 It is tempting to write $\mathsf{P}(X_t = Y_t, \ \forall t \in T) = 1$. However, without additional assumptions on the processes X and Y, it is not clear whether the set $\{X_t = Y_t, \ \forall t \in T\}$ is measurable. If some statement holds for "a.e. ω", what is means is that it is true on an event $\tilde{\Omega}$ with $\mathsf{P}(\tilde{\Omega}) = 1$. It may still be true or not true for some ω in $\tilde{\Omega}^c$, but the point is that at least such exceptional points are contained in a set of zero probability. The issue could be resolved if additionally the probability space $(\Omega, \mathcal{F}, \mathsf{P})$ is assumed to be *complete*, in which case all subsets of zero-probability sets are all measurable.

Homework Let $(X_t)_{t\geq 0}$ be a stochastic process on $(\Omega, \mathcal{F}, \mathsf{P})$ such that $t \mapsto X_t(\omega)$ is continuous for almost every $\omega \in \Omega$. Construct a stochastic process $(Y_t)_{t\geq 0}$ which is a modification of $(X_t)_{t\geq 0}$, such that $t \mapsto Y_t(\omega)$ is NOT continuous for almost every $\omega \in \Omega$.

Definition 1.5 *The (1d, standard) Brownian motion $(B_t)_{t\geq 0}$ is a continuous stochastic process with f.d.d. given by*

$$\mathcal{L}(B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_m} - B_{t_{m-1}}) = \mathcal{N}(0, \mathrm{diag}\{\sqrt{t_{i+1} - t_i}\}_{0\leq i\leq m-1}), \quad 0 = t_0 < t_1 < \cdots < t_m. \tag{1}$$

In particular, $\mathsf{P}(B_0 = 0) = 1$.

The information of f.d.d. of BM indeed sheds some light on the continuity property. In fact, the continuity condition can be dropped in the above definition, if we allow ourselves to consider stochastic processes up to modifications. The next result is a consequence of the Kolmogorov's Continuity Test

Theorem 1.2 *If $(X_t)_{t\geq 0}$ has the f.d.d. given in (1), then $(X_t)_{t\geq 0}$ has a continuous modification.*

Idea of the proof. We can use the f.d.d. on $\mathbb{Q}_+$ to show that for a.e. ω, $t \mapsto B_t(\omega)$ is uniformly continuous on $\mathbb{Q}_+$, that is, $\forall \varepsilon > 0$, $\exists \delta = \delta(\varepsilon, \omega)$ such that

$$|X_{t_1}(\omega) - X_{t_2}(\omega)| < \delta, \quad \forall |t_1 - t_2| < \varepsilon, \ t_1, t_2 \in \mathbb{Q}_+.$$

Then we can extend the function $t \mapsto X_t(\omega)$ on $\mathbb{Q}_+$ to a continuous function on $\mathbb{R}+$. $\square$

The existence of a stochastic process with any given "reasonable" f.d.d.'s is guaranteed by Kolmogorov's Extension Theorem. Then, using the above theorem we manage to construct the BM as a continuous stochastic process. We will fill in the gaps later in this note.

1.2 Probability measures on metric spaces

Recall that X is a r.v. on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ if $X : \Omega \to \mathbb{R}$ is $\mathcal{B}(\mathbb{R})/\mathcal{F}$-measurable. The distribution of X is a measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, given by

$$\mathcal{L}(X)(A) = \mathsf{P} \circ X^{-1}(A) = \mathsf{P}(X \in A), \quad A \in \mathcal{B}(\mathbb{R}).$$

The measure $\mathcal{L}(X)$ is determined by $\mathsf{P}(X \leq a)$, $a \in \mathbb{R}$, since $\mathcal{B}(\mathbb{R}) = \sigma\big((-\infty, a], \ a \in \mathbb{R}\big)$.

We want to replace $\mathbb{R}$ by a general metric space (M, d), where M can be as large as the space of all continuous functions. Any stochastic process from a probability measure on the space of continuous functions will automatically be continuous. We start by some basic notions on probability measures on metric spaces.

A metric space (M, d) is a set M equipped with a metric $d : M \times M \to \mathbb{R}_+$ which satisfies

- (symmetry) $d(x, y) = d(y, x)$;
- (positivity) $d(x, y) \geq 0$, and the equality holds only when $x = y$.
- (triangle inequality) $d(x, y) + d(y, z) \geq d(x, z)$.

Example 1.6 1. $M = \mathbb{Z}$, $d(x, y) = |x - y|$.

2. $M = \mathbb{R}^m$, with ℓ_p-distance

$$d_p(x, y) = \begin{cases} \Big[\displaystyle\sum_{i=1}^{m} (x_i - y_i)^p\Big]^{1/p}, & 1 < p < \infty, \\ \displaystyle\max_{1\leq i\leq m} |x_i - y_i|, & p = \infty. \end{cases}$$

3. $M = \mathcal{C}[0, 1]$, $d(x, y) = \displaystyle\sup_{t\in[0,1]} |x(t) - y(t)|$.

For a metric space, its Borel σ-algebra $\mathcal{B}(M)$ is the σ-algebra generated by all the open sets in M, or
the smallest σ-algebra containing all the open balls

$$B_r(x_0) = \{x : d(x, x_0) < r\}, \quad x_0 \in M,\ r > 0.$$

Definition 1.6 *Let (M, d) be a metric space. An M-value random element on $(\Omega, \mathcal{F}, \mathsf{P})$ is a measurable map from $(\Omega, \mathcal{F})$ to $(M, \mathcal{B}(M))$. The distribution of X is a probability measure on $(M, \mathcal{B}(M))$, given by*

$$(\mathsf{P} \circ X^{-1})(A) = \mathsf{P}(X \in A), \quad A \in \mathcal{B}(M). \tag{2}$$

The measure in (2) is determined its value on all open balls $B_r(x_0)$.

Example 1.7 Let X be a $\mathcal{C}[0,1]$-valued random element. Then $(X_t)_{t\in[0,1]}$ is a stochastic process.
In fact, for $t \in [0,1]$, we have the composition

$$\omega \mapsto X(\omega) \mapsto X_t(\omega),$$

where the first map is $\mathcal{B}(M)/\mathcal{F}$-measurable by the definition of random elements, and the second map is continuous since it is the evaluation map at given t of continuous functions and hence $\mathcal{B}(\mathbb{R})/\mathcal{B}(M)$-measurable. Therefore, the map $\omega \mapsto X_t(\omega)$ is $\mathcal{B}(\mathbb{R})/\mathcal{F}$-measurable.

Example 1.8 (Coordinate process) Let μ be a measure on $\big(\mathcal{C}(\mathbb{R}_+), \mathcal{B}(\mathcal{C}(\mathbb{R}_+))\big)$.

$$(\Omega, \mathcal{F}, \mathsf{P}) = \big(\mathcal{C}(\mathbb{R}_+), \mathcal{B}(\mathcal{C}(\mathbb{R}_+)), \mu\big), \quad X_t(\omega) = \omega_t,\ t \geq 0.$$

Then $(X_t)_{t\geq 0}$ is a continuous stochastic process.

A function $F : M \to \mathbb{R}$ is continuous if $d(x, x_0) \to 0$ implies $|F(x) - F(x_0)| \to 0$.

Definition 1.7 *Let $X^{(n)}$ and X be $\mathcal{C}[0,1]$-valued random elements defined on $(\Omega^{(n)}, \mathcal{F}^{(n)}, \mathsf{P}^{(n)})$ and $(\Omega, \mathcal{F}, \mathsf{P})$. We say that $X^{(n)}$ converge weakly (in distribution/law) to X, denoted by $X^{(n)} \Rightarrow_d X$, if for all bounded and continuous $F : \mathcal{C}[0,1] \to \mathbb{R}$,*

$$\lim_{n\to\infty} \mathsf{E}^{(n)} F(X^{(n)}) = \mathsf{E} F(X).$$

Remark 1.9 It is annoying to work with different probability spaces, but the good news is that the underlying probability spaces are not relevant for the notion of weak convergence. Let $\mu_n = \mathsf{P}^{(n)} \circ [X^{(n)}]^{-1}$ and $\mu = \mathsf{P} \circ X^{-1}$. Then μ_n, μ are all (probability) measures on $(\mathcal{C}[0,1], \mathcal{B}(\mathcal{C}[0,1]))$. By standard functional analysis terminologies, the above definition says that $\mu_n \to \mu$ in the weak-* topology (since measures on metric spaces form the dual space of bounded continuous functions). In probability it is just called weak convergence.

The Brownian motion will correspond to a measure on $\mathcal{C}[0,1]$, called the "Wiener measure". It is a probability measure on $\mathcal{C}[0,1]$ whose coordinate process has specific f.d.d.'s. To construct the Wiener measure directly:

- Functional CLT: need to understand (pre-)compact sets in $\mathcal{C}[0,1]$, and use the f.d.d.-information to verify tightness. A good read is [Bil99]).
- Gaussian measures on Banach spaces: more general, but still using the Gaussian information in an essential way. Such construction is needed for the study of stochastic PDEs, where the state space of the Gaussian processes is infinite-dimensional. This is a little beyond the scope of this course, and we will not go into more details here. Interesting readers can take a look at [PZ14, Chap. 2] or [Hai, Chap. 2-3].

With the Wiener measure at hand, we can now think of BM as random continuous functions. We conclude by mentioning the Hölder-continuity property of BM.

Definition 1.8 *Let $\alpha \in (0,1]$. A continuous function f is called (locally) α-Hölder if every x,*

$$\sup_{y \neq x} \frac{|f(x)-f(y)|}{|x-y|^\alpha} < \infty.$$

The α-Hölder continuous functions on $[0,T]$ form a complete metric space $\mathcal{C}^\alpha[0,1] \subset \mathcal{C}[0,1]$ under the norm:

$$|f|_{\mathcal{C}^\alpha} = \sup_x |f(x)| + \sup_{x \neq y} \frac{|f(x)-f(y)|}{|x-y|^\alpha}.$$

Theorem 1.3 *For $\alpha \in (0,1/2)$, the Wiener measure P^W is supported on α-Hölder continuous functions, that is,*

$$\forall \alpha \in (0,1/2), \quad \mathsf{P}^W(\omega \in \mathcal{C}^\alpha[0,1]) = 1.$$

Homework Let $\alpha \in (0,1]$. Show that

$$\Big\{\omega \in \mathcal{C}[0,1] : \sup\big\{\frac{|\omega(x)-\omega(y)|}{|x-y|^\alpha},\ x,y \in [0,1],\ x \neq y\big\} < \infty\Big\} \in \mathcal{B}\big(\mathcal{C}[0,1]\big).$$

1.3 Stochastic integrals and SDEs

Denote by $x(t)$ the position of a particle at time t. The *Langevin dynamics* of the particle is described by the equation

$$m\ddot{x}(t) = -\big(\nabla U\big)\big(x(t)\big) - \gamma \dot{x}(t) + c\eta(t).$$

The equation arises from Newton's second law:

- $m\ddot{x}(t)$ is the mass multiplied by the acceleration. It should be equal to the force, which is the right hand side of the equation.
- U is the potential, and $-\big(\nabla U\big)\big(x(t)\big)$ gives the potential force.
- $-\gamma\dot{x}(t)$ represents the friction which is usually proportional to the velocity $\dot{x}(t)$.
- $c\eta(t)$ is the random forcing, with c controlling its magnitude.

In an ideal physical model, $\eta(t)$ is the so-called *white noise*. As a "stochastic process", it should have at least the following two properties.

- **independence** $\eta(t)$ should be independent over disjoint intervals, namely, if I_1 and I_2 are two disjoint intervals of $\mathbb{R}$, then the two σ-fields

  $$\sigma\big(\eta(t),\ t \in I_1\big), \quad \sigma\big(\eta(t),\ t \in I_2\big)$$

  are independent.
- **stationarity** the one-dimensional distribution of $\eta(t)$ does not change:

  $$\mathcal{L}\big(\eta(t_1)\big) = \mathcal{L}\big(\eta(t_2)\big), \quad \forall t_1 \neq t_2.$$

Brownian motion in fact got its name from the botanist Robert Brown who observed the motion of pollen of plants through a microscope. For things like the pollen, the term $m\ddot{x}(t)$ is negligible compared to other terms since m is so small, the above equation can be approximated by the *overdamped Langevin dynamics*:

$$\dot{x}(t) = -(\nabla u)\big(x(t)\big) + \eta(t) \tag{3}$$

For simplicity, we also set all constants (c, γ, etc) to 1.

Free motion case. Let us set $U \equiv 0$ in (3). This means that no external potential (such the gravity) is taking effect. We can simply integrate (3) to obtain

$$x(t) = \int_0^t \eta(s)\, ds.$$

The function $t \mapsto x(t)$ is just the trajectory of a randomly moving light-weighted particle. Based on our assumption on the white noise $\eta(t)$, its antiderivative $x(t)$ will satisfy

- $t \mapsto x(t)$ is continuous; this is really a physical constraint.
- $x(t)$ has independent increments: for all $0 = t_0 < t_1 < \cdots < t_m$, $\{x(t_{i+1}) - x(t_i)\}_{1\le i\le m}$ are independent.
- The increments are centered Gaussian: $x(t) - x(s) \sim \mathcal{N}(0, \sigma^2_{t-s})$. This is because any increment can be written as i.i.d. sums of small r.v.'s:

  $$x(t) - x(s) = \sum_{i=1}^{N} x\Big(s + \frac{i(t-s)}{N}\Big) - x\Big(s + \frac{(i-1)(t-s)}{N}\Big).$$

  Moreover, due to stationarity, it only makes sense to have σ^2_{t-s} to be linear: $\sigma^2_{t-s} = K(t-s)$.

Up to a constant, the only process that satisfies all these conditions is the Brownian motion. This means the write noise $\eta(t)$ should be interpreted as the "derivative" of the Brownian motion. However, there is one fundamental issue of such interpretation:

Question *The Brownian motion is only α-Hölder continuous for $\alpha < 1/2$. In fact it is nowhere monotone and nowhere differentiable (we will see proofs of these statements later on). Then how should we define $\eta(t) = \dfrac{dB_t}{dt}$?*

The $U \not\equiv 0$ case. Let us consider a more general form

$$\dot{x}(t) = b\big(x(t)\big) + \eta(t), \tag{4}$$

where $b : \mathbb{R} \to \mathbb{R}$ is a sufficiently nice function. We are now entering the realm of the *stochastic differential equation* (SDE). It has a lot of applications in other fields, for example stable diffusion in text-to-image AI models. As we mentioned above, $\eta(t)$ is not a function. At best it could be defined as a generalized function (viewed as a linear functional acting on $\mathcal{C}_0^\infty(\mathbb{R})$). Due to the special structure of (4), this issue could be circumvented by considering the equivalent integral equation

$$x(t) = x(0) + \int_0^t b\big(x(s)\big)\, ds + B(t). \tag{5}$$

Now the noise enters the equation as a Brownian motion $B(t)$, which is a random continuous function. All terms in (5) make sense if $x(t)$ is a continuous function. Then standard fixed-point or Picard-iteration techniques can be applied here to construct a unique solution $x(t)$.

First variation of (4): the magnitude of the noise is time-dependent.
Let us consider

$$\ddot{x}(t) = b\big(x(t)\big) + f(t)\eta(t),$$

where $f(t)$ is a nice (say bounded and smooth) function. Inspired from the integral equation, it suffices to define the so-called *stochastic integral*

$$\int_0^t f(s)\eta(s)\, ds := \int_0^t f(s)\, dB(s) \tag{6}$$

The notation on the right hand side is to mimic that of the Riemann–Stieltjes integral. We recall its definition below.

Definition 1.9 *Let g be a function of finite variation (i.e., $g = g^+ - g^-$, where both g^+ and g^- are increasing) and f be a continuous function. Then the* <u>*Riemann–Stieltjes integral*</u> $\int f dg$ *is defined as*

$$\int_a^b f(s)\, dg(s) := \lim_{|\Delta|\to 0} \sum_{i=1}^N f(\xi_i)\big(g(t_{i+1}) - g(t_i)\big), \tag{7}$$

where $\Delta : a = t_0 < t_1 < \cdots < t_N = b$ is a partition, $\xi_i \in (t_i, t_{i+1})$ is arbitrary, and $|\Delta| = \max|t_{i+1} - t_i|$. The limit does not depend on the sequence of partitions chosen.

Example 1.10 When $g(t) = t$, the Riemann–Stieltjes integral is just Riemann integral.

A nice thing about the Riemann–Stieltjes integral is that integration by parts holds.

Proposition 1.4 *Let f, g be functions of bounded variation. Then*

$$\int_a^b f(t)\, dg(t) = f(b)g(b) - f(a)g(a) - \int_a^b g(t)\, df(t).$$

Homework

1. Prove the Abel transformation (summation by parts): for two sequences (a_k) and (b_k),

$$\sum_{k=1}^n a_k(b_{k+1} - b_k) = a_{k+1}b_{n+1} - a_1 b_1 - \sum_{k=1}^n b_{k+1}(a_{k+1} - a_k).$$

2. Show the integration by parts holds for Riemann–Stieltjes integrals when both f and g have bounded variation.

Of course, the Brownian motion does not have bounded variation; such property is almost requiring differentiability. However, we can still use the idea of integration by parts to define simple stochastic integrals in the form of (6) by

$$\int_0^t f(s)\, dB_s := f(t)B_t - \int_0^t B_s\, df(s).$$

It requires only that f has bounded variation.

In fact, the integration-by-part formula suggests a trade-off between the regularities of f and g. A further generalization of Riemann–Stieltjes integral is the *Young's integral*, which says that (7) makes sense for $f \in \mathcal{C}^\alpha$, $g \in \mathcal{C}^\beta$ with $\alpha + \beta > 1$. Intuitively, the Riemann–Stieltjes integral corresponds roughly to the case $\alpha = 0$ and $\beta = 1$.

Second variation of (4): the magnitude of the noise is both time- and space-dependent. We are now consider the SDE

$$\ddot{x}(t) = b\big(x(t)\big) + \sigma\big(t, x(t)\big)\eta(t), \tag{8}$$

where both b, σ are smooth. Again, with the integral form of the SDE, it all boils down to defining the stochastic integral

$$\int_0^t \sigma\big(s, x(s)\big)\, dB_s. \tag{9}$$

We already know that $t \mapsto B_t$ is $\mathcal{C}^\alpha$ with $\alpha < 1/2$. We also note that $x(t)$ cannot be more regular than $B(t)$, and hence no matter how smooth the function σ is, the map $t \mapsto \sigma\big(t, x(t)\big)$ is at most $\mathcal{C}^\beta$ with $\beta < 1/2$. One such simple example is $\displaystyle\int_0^t B_s\, dB_s$. Therefore, it is hopeless to define (9) even as a Young's integral, since $\alpha + \beta < 1$. This is as far as classical analysis can take us to. It tells us that the stochastic integral (9) cannot be defined for a fixed realization of (B_t). In fact, it could only be defined (or constructed) as a new stochastic process with the help of and some new probabilistic tools.

To summarize, two central goals of this course are

1. Define the stochastic integral
$$\int_0^t Y_s\, dB_s$$
for very *irregular* stochastic processes $Y = (Y_t)_{t\geq 0}$.

   Note that if $Y \in \mathcal{C}^\beta$, $\beta > 1/2$, then the stochastic integral can be defined for every fixed realization of the Brownian motion, but it does not cover even the simplest case where $Y_t = B_t$ itself.

2. Develop a good solution theory for the SDE (8).

2 Construction and properties of Brownian motion

2.1 Gaussian r.v.'s and vectors

We begin with the definition of a (generalized) Gaussian r.v.

Definition 2.1 *A Gaussian r.v. X with $\mathcal{N}(\mu, \sigma^2)$-distribution ($\mu \in \mathbb{R}$, $\sigma \geq 0$) is characterized by one of the following:*

1) The characteristic function (ch.f.) if $\varphi_X(\xi) = \mathbb{E}e^{i\xi X} = e^{i\mu\xi - \frac{\sigma^2}{2}\xi^2}$.

2) $\mathcal{L}(X) = \mathcal{L}(\sigma \cdot Y) + \mu$, where $Y \sim \mathcal{N}(0,1)$ is the standard normal, a continuous r.v. with density $\dfrac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$.

3) If $\sigma \neq 0$ (non-degenerate case), then X is a continuous r.v. with density $\dfrac{1}{\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$; if $\sigma = 0$, then $\mathsf{P}(X = 0) = 1$.

Proposition 2.1 *1. If X is a Gaussian r.v. on $(\Omega, \mathcal{F}, \mathsf{P})$, then $X \in L^p(\Omega, \mathcal{F}, \mathsf{P})$, $\forall p \in (0, \infty)$. In particular, for $X \sim \mathcal{N}(\mu, \sigma^2)$, $\mathsf{E}X = 0$ and $\mathrm{Var}(X) = \sigma^2$.*

2. If $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $i = 1, 2, \ldots, n$ and X_i are independent, then $X_1 + X_2 + \cdots + X_n \sim \mathcal{N}(\mu_1 + \cdots + \mu_n, \sigma_1^1 + \cdots + \sigma_n^2)$.

3. *(Closedness in L^2) If $X_m \sim \mathcal{N}(\mu_m, \sigma_m^2)$ and $X_m \to X$ in $L^2(\Omega, \mathcal{F}, \mathsf{P})$, i.e., $\lim_{m\to\infty} \mathsf{E}(X_m - X)^2 = 0$, then $X \sim \mathcal{N}(\mu, \sigma^2)$ with $\mu = \lim_{m\to\infty} \mu_m$ and $\sigma = \lim_{m\to\infty} \sigma_m$.*

Proof. 1. Direct computation using the Gaussian density.

2. Use the ch.f. of Gaussian r.v.'s.

3. $X_m \to X$ in L^2 implies the existence of both limits

$$\mu = \lim_{m\to\infty} \mu_m, \quad \sigma = \lim_{m\to\infty} \sigma_m.$$

Then $\varphi_{X_m}(\xi) \to \exp(i\mu\xi - \frac{\sigma^2\xi^2}{2})$, which is the ch.f. of $\mathcal{N}(\mu, \sigma^2)$-Gaussian. On the other hand, the L^2-convergence of X_m to X also implies that $X_m \to X$ in probability, and thus in distribution. Therefore, $\varphi_{X_m}(\xi) \to \varphi_X(\xi)$ and hence $\varphi_X(\xi) = \exp(i\mu\xi - \frac{\sigma^2\xi^2}{2})$ as desired.

For any $q > 0$, it is easy to get a uniform upper bound by direct computation:

$$\sup_m \mathsf{E}|X_m - X|^q \le C = C(\sup_m \mu_m, \sup_m \sigma_m).$$

By choosing $q > p$, we see that $|X_m - X|^p$ is uniformly integrable. Since $|X_m - X| \to 0$ in probability, this and uniform integrability imply (see [Dur07, Chap. 4.5]) that $\mathsf{E}|X_m - X|^p \to 0$. $\square$

Definition 2.2 *A random vector $X \in \mathbb{R}^d$ is Gaussian if for all $v \in \mathbb{R}^d$, $\langle v, X\rangle$ is a Gaussian r.v.*

Example 2.1 1. $X = (X_1, \ldots, X_d)$ where all X_i's are independent Gaussian random variables.

2. Let $X \in \mathbb{R}^d$ be Gaussian and Q be a $d \times d$ matrix. Then $Y = QX$ is Gaussian, since $\langle v, QX\rangle = \langle Q^T v, X\rangle$ for any vector v.

3. Let $(B_t)_{t\ge 0}$ be the Brownian motion. For any $0 \le t_1 < t_2 < \cdots < t_m$, both random vectors

$$(B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_m} - B_{t_{m-1}}), \quad (B_{t_1}, B_{t_2}, \ldots, B_{t_m})$$

are Gaussian.

Definition 2.3 *A stochastic process $(X_t)_{t\in T}$ is a <u>Gaussian process</u> if for any $t_1, t_2, \ldots, t_m \in T$, $(X_{t_1}, \ldots, X_{t_m})$ is a Gaussian vector.*

Example 2.2 The Brownian motion is a (centered) Gaussian process.

Theorem 2.2 *Each of the following is an equivalent definition for a random vector $X \in \mathbb{R}^d$ being Gaussian.*

1. *There exists $\mu_X \in \mathbb{R}^d$ and a non-negative quadratic form $Q : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ such that the ch.f. of X is*

$$\varphi_X(\xi) = \mathsf{E}e^{i\langle \xi, X\rangle} = e^{i\langle \mu_X, X\rangle - \frac{1}{2}Q(\xi,\xi)}.$$

2. *There exists $\mu_X \in \mathbb{R}^d$, an orthonormal basis (ONB) $\{b_1, \ldots, b_d\}$, and $\varepsilon_1 \ge \varepsilon_2 \ge \cdots \ge \varepsilon_r > 0 = \varepsilon_{r+1} = \cdots = \varepsilon_d$ such that*

$$X \overset{d}{=} Y = \mu_X + \sum_{i=1}^{r} \varepsilon\eta_i \cdot b_i, \quad \eta_i \overset{i.i.d.}{\sim} \mathcal{N}(0,1). \tag{10}$$

Proof. **From Definition to 1.**
Since $\langle \xi, X\rangle$ is Gaussian for every $\xi \in \mathbb{R}^d$, we have

$$\varphi_X(\xi) = \mathsf{E}e^{i\langle \xi, X\rangle} = e^{i\mathsf{E}\langle \xi, X\rangle - \frac{1}{2}\operatorname{Var}(\langle \xi, X\rangle)}.$$

We can take $\mu_X = \mathsf{E}X$ (coordinate-wise) so that $\mathsf{E}\langle \xi, X\rangle = \langle \xi, \mu_X\rangle$, and take

$$Q(\xi, \zeta) = \operatorname{Cov}(\langle \xi, X\rangle, \langle \zeta, X\rangle).$$

It is easy to check that $Q(\cdot, \cdot)$ is bilinear, symmetric, and defines a non-negative quadratic form on $\mathbb{R}^d$.

From 1. to 2.
Since Q is a non-negative quadratic form, it can be diagonalized in an ONB $\{b_1, b_2, \dots, b_d\}$ with eigenvalues $\varepsilon_i^2 \geq 0$:

$$Q(\xi, \zeta) = \sum_{i=1}^{d} (\varepsilon_i)^2 \langle \xi, b_i\rangle \langle \zeta, b_i\rangle.$$

(In matrix form, this is just $Q = B^T \Sigma B$ where $B = \{b_1, \dots, b_d\}$ and $\Sigma = \operatorname{diag}\{\varepsilon_1^2, \dots, \varepsilon_d^2\}$.) Without loss of generality we can take $\varepsilon_i \geq 0$ and order them from the largest to the smallest.

Suppose on some probability space we have i.i.d. $\mathcal{N}(0,1)$-Gaussian r.v.'s η_i and let Y be defined by (10). For all $v \in \mathbb{R}^d$,

$$\langle v, Y\rangle = \sum_{i=1}^{r} \varepsilon_i \langle v, b_i\rangle \eta_i$$

is a sum of independent Gaussian r.v.'s, and hence is Gaussian. This verifies that Y is a Gaussian vector. Also, we have

$$\mathsf{E}\langle v, Y\rangle = \langle v, \mu_X\rangle, \quad \operatorname{Var}(\langle v, Y\rangle) = \sum_{i=1}^{r} \varepsilon_i^2 \langle v, b_i\rangle^2 = Q(v, v).$$

So X and Y have the same ch.f., and hence $\mathcal{L}(X) = \mathcal{L}(Y)$ as desired.

From 2. to the definition of Gaussian vectors.
It is already done above. □

A Gaussian vector is non-degenerate if the quadratic form Q is non-degenerate, i.e., all eigenvalues are strictly positive.

Proposition 2.3 *A non-degenerate Gaussian vector $X \in \mathbb{R}^d$ has density*

$$p(x) = \frac{1}{(2\pi)^{d/2}} \frac{1}{\sqrt{\det(Q)}} e^{-\frac{1}{2}(x-\mu_X)^T Q^{-1}(x-\mu_X)},$$

where $Q = (Q_{ij}) = \big(\operatorname{Cov}(X_i, X_j)\big)$ is the covariance matrix.

The proposition can be proven by direct computation. It can also be used as a definition of non-degenerate Gaussian random vectors.

2.2 Construction of Brownian motion

2.2.1 *Gaussian measures on Banach spaces

A Gaussian vector is a map from $X : \{1, 2, \ldots, d\} \to \mathbb{R}$, i.e., an element in $\mathbb{R}^d = \mathbb{R}^{\{1,\ldots,d\}}$. To construct a Gaussian process on $[0, 1]$, a straightforward generalization is to replace the finite index set $\{1, \ldots, d\}$ by the interval $[0, 1]$. In infinite dimension there are many possible ways to choose the topology. So instead of working on the space of all functions $\mathbb{R}^{[0,1]}$ which has no special structure and is too large to study, we can work on more structured spaces like $\mathcal{C}[0, 1]$, $L^p[0, 1]$, etc. We will sketch how to define BM in this way. This idea will be useful for some heuristic computations later on.

Definition 2.4 (Gaussian measure on Banach spaces) *Let E be a separable Banach space. We say that an E-valued random element X has Gaussian distribution, if for any linear functional $\ell \in E^*$, $\ell(X)$ is a Gaussian r.v.*

Example 2.3 For the Brownian motion $X = (B_t)_{t\in[0,1]}$, $E = \mathcal{C}[0, 1]$, E^* is the space of all finite signed measures on $[0, 1]$. Then for $\mu \in E^*$, $\mathsf{E}\langle \mu, X\rangle = 0$ and

$$\mathrm{Var}(\langle \mu, X\rangle) = \int_0^1 \int_0^1 (s \wedge t)\, \mu(ds)\mu(dt).$$

Homework Let $f(t) = \mu((t, 1])$ Show that

$$\int_0^1 \int_0^1 (s \wedge t)\, \mu(ds)\mu(dt) = \int_0^1 |f(t)|^2\, dt.$$

Note: this is not surprising since integration by parts gives us

$$\int_0^1 B_t\, \mu(dt) = \int_0^1 B_t d(-f(t)) = \int_0^1 f(t)\, dB_t.$$

Hint: you can perform the computation by first assuming μ has continuous density, i.e., $f(t) = \int_t^1 g(s)\, ds$ for $g \in \mathcal{C}[0, 1]$. Then explain why all the steps are still valid for the general case using integration by parts for Riemann–Stieltjes integrals.

2.2.2 Gaussian white noise

Recall that our goal is to construct a centered Gaussian process $(B_t)_{t\in[0,1]}$ with covariance $\mathsf{E}B_tB_s = t \wedge s$. Surprisingly, it is convenient to first define the simplest stochastic integral $G(f) = \int_0^1 f(t)dB_t$, then take

$$B_t = \int_0^1 \mathbb{1}_{[0,t]}(s)\, ds$$

as the definition of Brownian motion.

The following discussion shows that the natural class of functions to define $G(f)$ is $L^2[0, 1]$, and for such f, $G(f)$ is in fact a Gaussian r.v.

First: f piecewise constant

Suppose that $[0, 1]$ is partitioned into $0 = t_0 < t_1 < \cdots < t_m = 1$ and $f(s) = \sum_{i=0}^{m-1} f_i \mathbb{1}_{[t_i,t_{i+1})}(s)$. Then in light of the Riemann–Stieltjes integral, it only makes sense to define $G(f)$ as

$$G(f) := \sum_{i=0}^{m-1} f_i \cdot (B_{t_{i+1}} - B_{t_i}). \tag{11}$$

We did not specify $f(1)$, but it does not enter the definition of (11) anyway, so it is safe to ignore it. The r.v. in (11) is a sum of i.i.d. Gaussian r.v.'s, so it is also Gaussian. It has zero mean, and a variance

$$\mathrm{Var}\big(G(f)\big) = \sum_{i=0}^{m-1} f_i^2(t_{i+1}-t_i) = \int_0^1 |f(t)|^2 dt$$

Second: difference of $G(f_1)$ and $G(f_2)$ for piecewise constant f_i.
Without loss of generality we can assume that f_1 and f_2 has the same partition of $[0,1]$, since otherwise we can enlarge their partitions to a common partition by including all the endpoints. Then, a similar computation yields that $G(f_1) - G(f_2)$ is also a centered Gaussian, with variance

$$\mathsf{E}|G(f_1) - G(f_2)|^2 = |f_1 - f_2|^2_{L^2[0,1]}.$$

Last: general $f \in L^2[0,1]$
Every function $f \in L^2[0,1]$ can be approximated by piecewise functions f_n in $L^2[0,1]$. One way to see is to first approximate any $L^2[0,1]$ function by continuous functions, then to approximate continuous functions by piecewise constant functions. Suppose that $f_n \to f$ in $L^2[0,1]$ and f_n are all piecewise constant. Note that

$$|G(f_n) - G(f_m)|_{L^2(\Omega,\mathcal{F},\mathsf{P})} = \mathsf{E}|G(f_n) - G(f_m)|^2 = |f_n - f_m|_{L^2[0,1]}$$

Since $f_n \to f$, (f_n) is a Cauchy sequence in $L^2[0,1]$, and hence $\big(G(f_n)\big)$ is a Cauchy sequence in $L^2(\Omega,\mathcal{F},\mathsf{P})$. But $L^2(\Omega,\mathcal{F},\mathsf{P})$ is a complete metric space, so every Cauchy sequence has a limit; let us call the limit $G(f)$. Note that all $G(f_n)$ are Gaussian, so by closedness of Gaussian r.v.'s, the limit $G(f)$ is also Gaussian.

Definition 2.5 (Gaussian white noise) *Let $(E,\mathcal{E})$ be a measurable space, μ be a σ-finite measure on $(E,\mathcal{E})$. Denote by $H = L^2(E,\mathcal{E},\mu)$. A Gaussian white noise (with intensity μ) is an isometry (i.e., preserving the inner product between two inner product spaces) from H to $L^2(\Omega,\mathcal{F},\mathsf{P})$ with values being (centered) Gaussian r.v.'s. The isometry is given by*

$$G : f \mapsto G(f) \sim \mathcal{N}(0, |f|^2_H).$$

Theorem 2.4 *If the Hilbert space $H = L^2(E,\mathcal{E},\mu)$ is separable there exists a probability space $(\Omega,\mathcal{F},\mathsf{P})$ such that the Gaussian white noise $G : H \to L^2(\Omega,\mathcal{F},\mathsf{P})$ exists.*

Remark 2.4 A Hilbert space is an inner product space which is also complete. One can think of a Hilbert space as an infinite-dimensional Euclidean space. All L^2-spaces are Hilbert space by definition. "Separable" means that there is a dense countable set, which is true when $H = L^2([0,1])$.

In proving the theorem, the only thing we will use about a separable Hilbert space is the existence of a ONB.

Proposition 2.5 *If H is a separable Hilbert space, then there exist $(f_n)_{n\geq 1} \subset H$, such that*

- $\langle f_n, f_m\rangle = \mathbb{1}_{n=m}$.
- *(basis) for every $f \in H$, it can be written as*

$$f = \sum_{n=1}^{\infty} \langle f_n, f\rangle f_n,$$

where the infinite sum is converging in H.

Such collection $(f_n)_{n\geq 1}$ *is called an orthonormal basis of* H*.*

Proof of Theorem 2.4. Pick an ONB $(f_n)_{n\geq 1}$ for $H = L^2(E, \mathcal{E}, \mu)$. Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space on which there are i.i.d. $\mathcal{N}(0,1)$-r.v.'s ξ_n, $n \geq 1$. Let us define

$$G_N(f) = \sum_{n=1}^{N} \xi_n \langle f_n, f\rangle.$$

Then $G_N(f)$, $N \geq 1$, each being a sum of independent Gaussians, are all Gaussian. Also, for $N < N'$,

$$\mathsf{E}|G_N(f) - G_{N'}(f)|^2 = \sum_{N\leq n<N'} |\langle f_n, f\rangle|^2.$$

Since $f \in H = L^2(E, \mathcal{E}, \mu)$ and $|f|_H^2 = \sum_{n=1}^{\infty} |\langle f_n, f\rangle|^2 < \infty$, $\{G_N(f)\}_{N\geq 1}$ is Cauchy in $L^2(\Omega, \mathcal{F}, \mathsf{P})$. So $G(f) := \lim_{N\to\infty} G_N(f)$ exists in $L^2(\Omega, \mathcal{F}, \mathsf{P})$. Since $G(f)$ is the L^2-limit of Gaussians, it is also Gaussian; moreover, it has distribution $\mathcal{N}(0, |f|_H^2)$. □

Example 2.5 $H = L^2(\mathbb{R}_{\geq 0}, \mathcal{B}(\mathbb{R}_{\geq 0}), dt)$. Then $B_t = G(\mathbb{1}_{[0,t]})$ is a centered Gaussian process, with covariance

$$\mathsf{E}B_tB_s = \int_0^\infty \mathbb{1}_{[0,t]}(r)\mathbb{1}_{[0,s]}(r)\, dr = s \wedge t.$$

That is, $(B_t)_{t\geq 0}$ has the same f.d.d. as the Brownian motion.

The definition of Gaussian white noise only shows B_t is Gaussian for a fixed t. To see that any f.d.d. is jointly Gaussian, we need to use the fact that all isometries between Hilbert spaces are linear, so for any $t_1 < \cdots < t_m$ and $v_1, \ldots, v_m$,

$$v_1B_{t_1} + \cdots + v_mB_{t_m} = G\Big(\sum_{i=1}^{m} v_i\mathbb{1}_{[0,t_i]}\Big)$$

is indeed Gaussian. The covariance computation is a consequence of applying the following polarization identity to the inner product spaces $L^2(\Omega, \mathcal{F}, \mathsf{P})$ and $L^2[0,1]$:

$$4\langle f, g\rangle = \langle f+g, f+g\rangle - \langle f-g, f-g\rangle.$$

3 Notations

3.1 Abbreviations

i.i.d.	independent, identically distributed
r.v.	random variable
f.d.d.	finite-dimensional distribution
ch.f.	characteristic function

3.2 Sets

$\mathbb{Z}$	set of integers
$\mathbb{N}$	set of natural numbers $\{0, 1, 2, \ldots\}$
$\mathbb{Q}$	set of rational numbers
$\mathbb{R}$	set of real numbers
$\mathbb{R}_+$ (resp. $\mathbb{R}_-$)	set of non-negative (resp. non-positive) real numbers

3.3 Relations

$\Rightarrow_d$ or $\Rightarrow$	convergence in distribution/law

3.4 Functional spaces

$\mathcal{C}[a,b]$	continuous function defined on the interval $[a,b]$
$\mathcal{C}^\alpha[a,b]$	α-Hölder continuous function defined on the interval $[a,b]$

3.5 Operations

$a \wedge b$	$\min(a,b)$
$a \vee b$	$\max(a,b)$
$\langle a,b \rangle$	inner product in a Euclidean space/Hilbert space (or) a linear functional a in the dual space $\mathcal{X}^*$ acting on an element b in a Banach space $\mathcal{X}$

3.6 Miscellaneous

$\mathcal{L}(X)$	distribution/law of a random variable/element X.

References

[Bil99] Patrick Billingsley. Convergence of Probability Measures. Wiley Series in Probability and Statistics. Probability and Statistics. Wiley, New York, 2nd ed edition, 1999.

[Dur07] Richard Durrett. Probability: Theory and Examples. Duxbury Advanced Series. Thomson Brooks/Cole, Belmont, Calif, 3. ed. edition, 2007.

[Hai] Martin Hairer. An Introduction to Stochastic PDEs. https://www.hairer.org/notes/SPDEs.pdf.

[PZ14] Giuseppe Da Prato and Jerzy Zabczyk. Stochastic Equations in Infinite Dimensions. Cambridge University Press, 2014.