

Lecture Note for MAT7093: Stochastic Analysis

LI Liying

February 27, 2024

1 Introduction

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

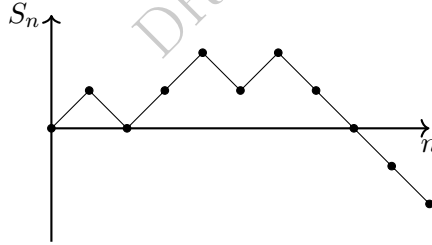
1.1 Stochastic processes

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 $\mathbb{E}X_i = 0$, $\mathbb{E}X_i^2 = 1$, 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 results of independent coin flips, so $\mathbb{P}(X_i = \pm 1) = 1/2$.

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



The plotted trajectory, which linearly interpolates between (n, S_n) , $n \in \mathbb{N}$, can be 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)

$$\left(\frac{\tilde{S}_{nt}}{\sqrt{n}}, t \geq 0 \right) \Rightarrow_d \left(B_t, t \geq 0 \right),$$

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. As we are studying random functions rather than random variables, we need to work on probability measures on functional spaces, which are infinite-dimensional and quite different from finite-dimensional spaces like \mathbb{R}^d . We will return to this in [Section 1.2](#).

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

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

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

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

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

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

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

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

Remark 1.4 As a notation, we may simply write “ X_t is $\mathcal{B}(\mathbb{R})/\mathcal{F}$ -measurable”.

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

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

It follows from [Definition 1.1](#) that all the sets

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

are measurable, and hence f.d.d. of a stochastic process is well-defined.

Homework (Transformation of BM)

1. Prove the equivalency of the following two conditions: for $0 = t_0 \leq t_1 < \dots < t_m$,

$$\begin{aligned} & \mathcal{L}(B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_m} - B_{t_{m-1}}) = \mathcal{N}(0, \text{diag}\{t_{i+1} - t_i\}_{0 \leq i \leq m-1}) \\ \Leftrightarrow & (B_{t_1}, B_{t_2}, \dots, B_{t_m}) \text{ is a centered Gaussian vector with covariance } \mathbb{E}B_{t_i}B_{t_j} = t_i \wedge t_j. \end{aligned} \quad (1)$$

2. Suppose that $(B_t)_{t \geq 0}$ has f.d.d. (1). Show that all the following processes have the same f.d.d. (1).

- a) $(-B_t)_{t \geq 0}$.
- b) $(B_t^\lambda)_{t \geq 0} := (\frac{1}{\lambda} B_{\lambda^2 t})_{t \geq 0}$. (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 $0 \cdot B_{1/0} = 0$).

Hint: You can find some basic properties of Gaussian vectors in Section 2.1. This exercise is basically about covariance computation.

It is believable that a stochastic process is more or less determined by all its f.d.d. (which is done by Komolgorov's Extension Theorem, see for example [Shi96, Chap. II.3, Theorem 4]). With the definition of stochastic processes at hand, the next question is what makes a “continuous” stochastic process. To discuss continuity we now take T to be an interval of \mathbb{R} ($T = [a, b]$, $[0, \infty)$, etc). Then, a “continuous” process requires additionally that the map

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

is *continuous* for \mathbb{P} -a.e. ω .

Remark 1.5 For a generic stochastic process $(X_t)_{t \in \mathbb{R}}$, the sets

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

and (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,

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

Although for any fixed sequence (t_n) , the set

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

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

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

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

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 (its usage summarized in **Theorem 1.2**), the other one is to directly build up probability measures on the desired functional spaces (**Section 1.2**).

But assuming that this can be done, we are ready to rigorously define what a Brownian motion is. One last thing to do is to specify how we distinguish between different stochastic processes.

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

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

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

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 1.6 It is tempting to write $P(X_t = Y_t, \forall t \in T) = 1$. 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 it means is that it is true on an event $\tilde{\Omega}$ with $P(\tilde{\Omega}) = 1$. 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 (Ω, \mathcal{F}, P) is assumed to be *complete*, in which case all subsets of zero-probability sets are measurable.

Homework Let $X = (X_t)_{t \geq 0}$ be a stochastic process on (Ω, \mathcal{F}, P) such that $t \mapsto X_t(\omega)$ is continuous for almost every $\omega \in \Omega$. Let τ be a continuous r.v. on (Ω, \mathcal{F}, P) and $Y = (Y_t)_{t \geq 0}$ be defined as

$$Y_t(\omega) = \begin{cases} X_t(\omega), & t \neq \tau(\omega), \\ X_t(\omega) + 1, & t = \tau(\omega). \end{cases}$$

Show that Y is a stochastic process which is a modification of X , but $t \mapsto Y_t(\omega)$ 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

$$\mathcal{L}(B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_m} - B_{t_{m-1}}) = \mathcal{N}(0, \text{diag}\{t_{i+1} - t_i\}_{0 \leq i \leq m-1}), \quad 0 = t_0 \leq t_1 < \dots < t_m. \quad (2)$$

In particular, $P(B_0 = 0) = 1$.

The information of f.d.d. of Brownian motion 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 (2), 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

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

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 *consistent* f.d.d. is guaranteed by Kolmogorov's Extension Theorem, although later in this note we will exploit the Gaussian f.d.d. more to give another more explicit construction of Brownian motion (Section 2.2). Then, using the above theorem we obtain 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 (Ω, \mathcal{F}, P) if $X : \Omega \rightarrow \mathbb{R}$ is $\mathcal{B}(\mathbb{R})/\mathcal{F}$ -measurable. The distribution of X is a measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, given by

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

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

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 $d : M \times M \rightarrow \mathbb{R}_+$ 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 1.7 1. $M = \mathbb{Z}$, $d(x, y) = |x - y|$.

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

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

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

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

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

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

$$(\mathbb{P} \circ X^{-1})(A) = \mathbb{P}(X \in A), \quad A \in \mathcal{B}(M). \quad (3)$$

The measure in (3) is determined its value on all open balls $B_r(x_0)$.

Example 1.8 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 1.9 (Coordinate process) Let μ be a measure on $(\mathcal{C}(\mathbb{R}_+), \mathcal{B}(\mathcal{C}(\mathbb{R}_+)))$. Define

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

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

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

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

$$\lim_{n \rightarrow \infty} \mathbb{E}^{(n)} F(X^{(n)}) = \mathbb{E} F(X).$$

Remark 1.10 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 $\mu_n = \mathbb{P}^{(n)} \circ [X^{(n)}]^{-1}$ and $\mu = \mathbb{P} \circ X^{-1}$. 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 $\mu_n \rightarrow \mu$ in the weak-* topology (since measures on metric spaces form the dual space of bounded continuous functions). In probability it is conventional to call it weak convergence.

The Brownian motion gives rise 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 information of f.d.d. 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 other than Definition 2.4. 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 Brownian motion as random continuous functions. We conclude by mentioning the Hölder-continuity property of Brownian motion.

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

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

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 \mathbf{P}^W is supported on α -Hölder continuous functions, that is,

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

Remark 1.11 One can show that for every $\alpha \in (0, 1]$, the set of α -Hölder continuous function in $\mathcal{C}[0, 1]$ is in $\mathcal{B}(\mathcal{C}[0, 1])$, using that fact that a continuous function can be determined by its values on rational points.

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

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

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 $-(\nabla U)(x(t))$ 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

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

are independent.

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

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

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*:

$$\dot{x}(t) = -(\nabla u)(x(t)) + \eta(t) \quad (4)$$

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

Free motion case. Let us set $U \equiv 0$ in (4). This means that no external potential (such as the gravity) is taking effect. We can simply integrate (4) to obtain (assuming $x(0) = 0$)

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

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 $0 = t_0 \leq t_1 < \dots < t_m$, $\{x(t_{i+1}) - x(t_i)\}_{1 \leq i \leq m}$ are independent.
- The increments are centered Gaussian: $x(t) - x(s) \sim \mathcal{N}(0, \sigma_{t-s}^2)$. This is because any increment can be written as i.i.d. sums of small r.v.'s:

$$x(t) - x(s) = \sum_{i=0}^{N-1} x(t_{i+1}) - x(t_i), \quad t_i = s + \frac{i(t-s)}{N}.$$

Moreover, due to stationarity, it only makes sense to have σ_{t-s}^2 to be linear: $\sigma_{t-s}^2 = K \cdot (t-s)$ for some constant $K > 0$.

Up to a constant, the only process that satisfies all these conditions is Brownian motion. This means the write noise $\eta(t)$ should be interpreted as the “derivative” of 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 $\eta(t) = \frac{dB_t}{dt}$?*

The $U \neq 0$ case. Let us consider a more general form

$$\dot{x}(t) = b(x(t)) + \eta(t), \quad (5)$$

where $b : \mathbb{R} \rightarrow \mathbb{R}$ 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 (5), this issue could be circumvented by considering the equivalent integral equation

$$x(t) = x(0) + \int_0^t b(x(s)) ds + B(t). \quad (6)$$

Now the noise enters the equation as a Brownian motion $B(t)$, which is a random continuous function. All terms in (6) make sense as long as $x(t)$ 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 (5): the magnitude of the noise is time-dependent.

Let us consider

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

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*

$$\int_0^t f(s)\eta(s) ds := \int_0^t f(s) dB(s) \quad (7)$$

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 Riemann–Stieltjes integral $\int f dg$ is defined as

$$\int_a^b f(s) dg(s) := \lim_{|\Delta| \rightarrow 0} \sum_{i=1}^N f(\xi_i)(g(t_{i+1}) - g(t_i)), \quad (8)$$

where $\Delta : a = t_0 < t_1 < \dots < t_N = b$ 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 or (ξ_i) that are chosen.

Example 1.12 When $g(t) = t$, the Riemann–Stieltjes integral is just the 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 Use the Abel transformation (summation by parts)

$$\sum_{k=1}^n u_k(v_{k+1} - v_k) = u_{n+1}v_{n+1} - u_1v_1 - \sum_{k=1}^n v_{k+1}(u_{k+1} - u_k)$$

to show that integration by parts holds for Riemann–Stieltjes integrals for functions f and g of bounded variation.

Of course, 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 (7) 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 (8) 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 (5): the magnitude of the noise is both time- and space-dependent.

We now consider the SDE

$$\ddot{x}(t) = b(x(t)) + \sigma(t, x(t))\eta(t), \quad (9)$$

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

$$\int_0^t \sigma(s, x(s)) dB_s. \quad (10)$$

We already know that $t \mapsto B_t$ is \mathcal{C}^α 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 $t \mapsto \sigma(t, x(t))$ is at most \mathcal{C}^β with $\beta < 1/2$. One such simple example is $\int_0^t B_s dB_s$. Therefore, it is hopeless to define (10) 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 (10) 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 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 $Y = (Y_t)_{t \geq 0}$.

Again, we emphasize that if $Y \in \mathcal{C}^\beta$, $\beta > 1/2$, then the stochastic integral can be defined for every fixed realization of Brownian motion, but such treatment cannot cover even the simple case where $Y_t = B_t$ itself.

2. Develop a good solution theory for the SDE (9).

2 Construction and properties of Brownian motion

2.1 Gaussian r.v.’s and vectors

Gaussianity is crucial in the study of Brownian motion. In many ways, Brownian motion can be seen as a generalization of Gaussian vectors. In this section, we review some basic facts about Gaussian r.v.’s and vectors.

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

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

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

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

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

Proposition 2.1

1. If X is a Gaussian r.v. on $(\Omega, \mathcal{F}, \mathbb{P})$, then $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$, $\forall p \in (0, \infty)$. In particular, for $X \sim \mathcal{N}(\mu, \sigma^2)$, $\mathbb{E}X = 0$ and $\text{Var}(X) = \sigma^2$.
2. If $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ and X_i are independent, then $X_1 + X_2 + \dots + X_n \sim \mathcal{N}(\mu_1 + \dots + \mu_n, \sigma_1^2 + \dots + \sigma_n^2)$.

Proof: The proof is elementary.

1. Direct computation using the Gaussian density.
2. Use the ch.f. of Gaussian r.v.'s.

□

Gaussian r.v.'s have nice properties as elements in $L^2(\Omega, \mathcal{F}, \mathbb{P})$.

Proposition 2.2 If $X_m \sim \mathcal{N}(\mu_m, \sigma_m^2)$ and $X_m \rightarrow X$ in $L^2(\Omega, \mathcal{F}, \mathbb{P})$, then $X \sim \mathcal{N}(\mu, \sigma^2)$ with

$$\mu = \lim_{m \rightarrow \infty} \mu_m, \quad \sigma = \lim_{m \rightarrow \infty} \sigma_m. \quad (11)$$

Moreover, $X_m \rightarrow X$ in $L^p(\Omega, \mathcal{F}, \mathbb{P})$ for any $p > 0$.

Proof: The L^2 -convergence of $X_m \rightarrow X$ implies the existence of both limits in (11). Hence, for each $\xi \in \mathbb{R}$, we have $\varphi_{X_m}(\xi) \rightarrow \exp(i\mu\xi - \frac{\sigma^2\xi^2}{2})$, which is the ch.f. of $\mathcal{N}(\mu, \sigma^2)$ -Gaussian. On the other hand, the L^2 -convergence of $X_m \rightarrow X$ also implies that $X_m \rightarrow X$ in probability, and thus in distribution. so $\varphi_{X_m}(\xi) \rightarrow \varphi_X(\xi)$. Therefore, $\varphi_X(\xi) = \exp(i\mu\xi - \frac{\sigma^2\xi^2}{2})$, and X indeed has $\mathcal{N}(\mu, \sigma^2)$ distribution, with μ, σ given by (11).

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

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

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

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. $X = (X_1, \dots, X_d)$ 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 $(B_t)_{t \geq 0}$ be Brownian motion. For any $0 \leq t_1 < t_2 < \dots < t_m$, both random vectors

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

are Gaussian.

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

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

Theorem 2.3 Each of the following is an equivalent definition for a random vector $X \in \mathbb{R}^d$ to be Gaussian.

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

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

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

$$X \stackrel{d}{=} Y = \mu_X + \sum_{i=1}^r \varepsilon_i \eta_i \cdot b_i, \quad \eta_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1). \quad (12)$$

Proof: From **Definition 2.2** to **Item 1**. Since $\langle \xi, X \rangle$ is Gaussian for every $\xi \in \mathbb{R}^d$, we have

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

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

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

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

From Item 1 to Item 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 $\Sigma = \text{diag}\{\varepsilon_1^2, \dots, \varepsilon_d^2\}$.) 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 (12). 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

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

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

From Item 2 to Definition 2.2. It is already done above. \square

A Gaussian vector is non-degenerate if the quadratic form Q is non-degenerate, i.e., all eigenvalues are strictly positive. A non-degenerate Gaussian vector has a density, which is more familiar to most people.

Proposition 2.4 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 $Q = (Q_{ij}) = (\text{Cov}(X_i, X_j))$ is the covariance matrix.

Remark 2.3 Since the distribution of a Gaussian vector is determined by its covariance matrix, the f.d.d. of a centered Gaussian process $X = (X_t)_{t \in T}$ is completely determined by its covariance function

$$\Gamma(s, t) := \text{Cov}(X_s, X_t) = \mathbb{E}X_s X_t, \quad s, t \in T.$$

For Brownian motion, $\Gamma(s, t) = s \wedge t$.

Homework Let X and Y be i.i.d. with $\mathbb{E}X = \mathbb{E}Y = 0$ and $\mathbb{E}X^2 = \mathbb{E}Y^2 = 1$. Suppose that the distribution of (X, Y) is rotational invariant, i.e.,

$$\mathcal{L}(X, Y) = \mathcal{L}(X \cos \theta + Y \sin \theta, -X \sin \theta + Y \cos \theta), \quad \forall \theta \in \mathbb{R}.$$

Show that $\mathcal{L}(X) = \mathcal{L}(Y) = \mathcal{N}(0, 1)$.

Hint: rotational invariance implies that the ch.f. takes the form $\varphi_{X,Y}(\xi, \eta) = F(\xi^2 + \eta^2)$.

A Banach space is an infinite-dimensional vector space. The generalization of Gaussian vectors to the infinite dimension is *Gaussian measures on Banach spaces*.

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 $\langle \lambda, X \rangle$ is a Gaussian r.v. for any linear functional $\lambda \in E^*$.

Example 2.4 For Gaussian vectors in \mathbb{R}^d , $E = \mathbb{R}^d = E^*$, that is, any linear functional is the inner product with a fixed vector v . This is exactly **Definition 2.2**.

Example 2.5 For Brownian motion, $X = (B_t)_{t \in [0,1]}$, $E = \mathcal{C}[0, 1]$, and E^* is the space of all finite signed measures on $[0, 1]$. Then for $\lambda = \lambda(dt) \in E^*$, $\langle \lambda, X \rangle$ is a centered Gaussian with variance

$$\text{Var}(\langle \lambda, X \rangle) = \mathbb{E} \int_0^1 \int_0^1 B_s \lambda(ds) B_t \lambda(dt) = \int_0^1 [\mathbb{E} B_s B_t] \lambda(ds) \lambda(dt) \int_0^1 \int_0^1 (s \wedge t) \lambda(ds) \lambda(dt),$$

where in the last equality the exchange of integration and expectation needs justification.

For the construction of Brownian motion, the variance of $\langle \lambda, X \rangle$, $\lambda \in E^*$, will be given first, and then some general theory will guarantee the existence of a corresponding (centered) Gaussian measure as long as the variance functional induces a positive definite quadratic form, similar to Gaussian vectors.

Homework Let $f(t) = \lambda((t, 1])$.

1. Suppose that $\lambda(dt) = \rho(t) dt$ for some $\rho \in \mathcal{C}[0, 1]$. Show that

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

Hint: use integration by parts.

2. (Optional) Prove the same identity for an arbitrary signed measure $\lambda(dt)$.

Hint: if $\lambda(dt)$ is a signed measure, then f defined as above has bounded variation and $\lambda(dt) = d(-f(t))$. Use integration by parts for Riemann–Stieltjes integrals.

2.2 Gaussian white noise

The goal of this section is to construct a centered Gaussian process $(B_t)_{t \in [0,1]}$ with covariance $\mathbb{E}B_t B_s = t \wedge s$. After the construction, the resulting process (called “pre-Brownian motion” in [LeG16]) may not be a.s. continuous; we will discuss how to get continuity in Sections 2.2 and 2.3.

The Kolmogorov’s Extension Theorem ([Shi96, Chap. II.3, Theorem 4]) already guarantees the existence of a stochastic process with any prescribed *consistent* f.d.d. However, in the special case of Brownian motion, it is advantageous to have a more explicit construction using the Gaussian white noise.

Surprisingly, it is more convenient to first define a more general stochastic integral $G(f) = \int_0^1 f(t)dB_t$, and then define Brownian motion as a special stochastic integral

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

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. This will also motivate the introduction of Gaussian white noise, and the definition of Itô integrals later.

First: f piecewise constant

Suppose that $[0,1]$ is partitioned into $0 = t_0 < t_1 < \dots < t_m = 1$ 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

$$G(f) := \sum_{i=0}^{m-1} f_i \cdot (B_{t_{i+1}} - B_{t_i}). \quad (13)$$

We did not specify $f(1)$, but it does not enter the definition of (13) anyway, so it is safe to ignore it. The r.v. in (13) is a sum of i.i.d. Gaussian r.v.’s, so it is also Gaussian. It has zero mean, and a variance

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

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

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

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 $f_n \rightarrow f$ in $L^2[0,1]$ and f_n are all piecewise constant. Note that

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

Since $f_n \rightarrow f$, (f_n) is a Cauchy sequence in $L^2[0,1]$, and hence $(G(f_n))$ is a Cauchy sequence in $L^2(\Omega, \mathcal{F}, \mathbb{P})$. But $L^2(\Omega, \mathcal{F}, \mathbb{P})$ is a complete metric space, which means every Cauchy sequence has a limit; let us denote the limit of $G_N(f)$ by $G(f)$. Note that all $G(f_n)$ are Gaussian, so by Proposition 2.2, 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 $L^2(\Omega, \mathcal{F}, \mathbb{P})$ with values being (centered) Gaussian r.v.'s. The isometry is given by

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

Theorem 2.5 If the Hilbert space $H = L^2(E, \mathcal{E}, \mu)$ is separable, then there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that the Gaussian white noise $G : H \rightarrow L^2(\Omega, \mathcal{F}, \mathbb{P})$ exists.

Remark 2.6 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 standard real analysis. “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 an ONB.

Proposition 2.6 If H is a separable Hilbert space, then there exist $(e_n)_{n \geq 1} \subset H$, such that

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

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

where the infinite sum is converging in H .

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

Proof of Theorem 2.5: Pick an ONB $(e_n)_{n \geq 1}$ for $H = L^2(E, \mathcal{E}, \mu)$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ 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

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

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

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

Since $f \in H = L^2(E, \mathcal{E}, \mu)$ and $|f|_H^2 = \sum_{n=1}^{\infty} |\langle e_n, f \rangle|^2 < \infty$, $\{G_N(f)\}_{N \geq 1}$ is Cauchy in $L^2(\Omega, \mathcal{F}, \mathbb{P})$.

Therefore, the following limit in $L^2(\Omega, \mathcal{F}, \mathbb{P})$

$$G(f) = \lim_{N \rightarrow \infty} G_N(f) = \sum_{n=1}^{\infty} \xi_n \langle e_n, f \rangle \tag{14}$$

exists. Since $G(f)$ is the L^2 -limit of Gaussians, it is also Gaussian; moreover, by **Proposition 2.1**, it has distribution $\mathcal{N}(0, |f|_H^2)$. \square

Example 2.7 A Gaussian vector in \mathbb{R}^d is also associated with a Gaussian white noise expansion, with $H = (\mathbb{R}^d, |\cdot|_H)$, and

$$|v|_H^2 = v^T Q v = \sum_{i=1}^r \varepsilon_i^2 |\langle v, b_i \rangle|^2.$$

Compare with **Item 2** in **Theorem 2.3**.

Example 2.8 $H = L^2(\mathbb{R}_{\geq 0}, \mathcal{B}(\mathbb{R}_{\geq 0}), dt)$. Then $B_t = G(\mathbb{1}_{[0,t]})$ is a centered Gaussian process, with covariance

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

That is, $(B_t)_{t \geq 0}$ has the same f.d.d. as 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 do a little bit more work. This can be also derived from the definition of Gaussian white noise. In fact, any isometry between Hilbert spaces must be linear, so for any $t_1 < \dots < t_m$ and v_1, \dots, v_m ,

$$v_1 B_{t_1} + \dots + v_m B_{t_m} = G\left(\sum_{i=1}^m v_i \mathbb{1}_{[0,t_i]}\right)$$

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

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

Remark 2.9 Use the GWN construction of BM, for $f \in L^2[0, \infty)$,

$$\mathbb{E} \left| \int_0^\infty f(t) dB_t \right|^2 = \mathbb{E} |G(f)|^2 = \int_0^\infty f^2(t) dt. \quad (15)$$

This is the simplest form of the celebrated “Itô’s Isometry”.

2.3 Continuity of Brownian motion via Kolmogorov’s Continuity Theorem

A powerful tool to get continuous modification of a stochastic process is the celebrated Komolgorov Continuity Theorem. It extracts information of path regularity from the f.d.d.

Theorem 2.7 Let $(X_t)_{t \in [0, T]}$ be a stochastic process that satisfies

$$\mathbb{E} |X_t - X_s|^\alpha \leq K |t - s|^{1+\beta}, \quad \forall 0 \leq s, t \leq T.$$

Then X has a modification \tilde{X} which is γ -Hölder continuous for all $\gamma < \beta/\alpha$.

Example 2.10 Let $(B_t)_{t \in [0, 1]}$ be a Gaussian process with $\mathbb{E} B_t B_s = t \wedge s$. Then $B_t - B_s \sim \mathcal{N}(0, t - s)$, and hence $\mathbb{E} |B_t - B_s|^n \leq K_n (t - s)^{n/2}$ for all $n \geq 1$. Since $\frac{n/2 - 1}{n}$ can be arbitrarily close to $1/2$, (B_t) has a modification which is γ -Hölder for all $\gamma < 1/2$.

We first reduce **Theorem 2.7** to the case of a fixed γ .

Lemma 2.8 If X and Y are continuous stochastic processes on \mathbb{R} , and Y is a modification of X , then Y is a version of X .

Proof: By the definition of modifications, $P(X_t = Y_t) = 1$ for all $t \in \mathbb{R}$. Since the set of rational numbers \mathbb{Q} is countable, we have $P(X_t = Y_t, \forall t \in \mathbb{Q}) = 1$. That is, there is a set \mathcal{N} with probability $P(\mathcal{N}) = 0$, such that for all $\omega \in \mathcal{N}$,

$$X_t(\omega) = Y_t(\omega), \quad \forall t \in \mathbb{Q}. \quad (16)$$

Noting that $t \mapsto X_t(\omega)$ and $t \mapsto Y_t(\omega)$ are always continuous. Hence, if for any ω the condition (16) holds, then it follows that

$$X_t(\omega) = Y_t(\omega), \quad \forall t \in \mathbb{R}. \quad (17)$$

So (17) holds except on a null-set \mathcal{N} ; this means that Y is a version of X . \square

Lemma 2.9 For Theorem 2.7, it suffices to prove it for any fixed $\gamma < \alpha/\beta$.

Proof: Suppose that there are modifications $X^{(n)}$ of X which is $\gamma_n = (\alpha/\beta - 1/n)$ -Hölder continuous. Then by Lemma 2.8, $X^{(n)}$, $n \geq 1$, are all versions of each other. In particular, there exist null-sets $\mathcal{N}^{(n)}$ such that

$$\forall \omega \in (\mathcal{N}^{(n)})^c : \quad X_t^{(1)} = X_t^{(n)}, \quad t \in [0, T].$$

Let $\mathcal{N} = \bigcup_{n \geq 2} \mathcal{N}^{(n)}$. Then \mathcal{N} is also a null-set, and for all $\omega \in \mathcal{N}^c$, $X_t^{(1)} = X_t^{(n)}$, $\forall n, t$. Hence, $X^{(1)}$ is γ_n -Hölder for all $n \geq 1$ on the set \mathcal{N} . Since γ_n is arbitrarily close to α/β , $X^{(1)}$ is γ -Hölder for any $\gamma < \alpha/\beta$ on \mathcal{N} . The proof is complete. \square

Proof of Theorem 2.7: Without loss of generality set $T = 1$. Let $\gamma < \beta/\alpha$.

By Markov inequality,

$$P(|X_{k/2^n} - X_{(k-1)/2^n}| > 2^{-\gamma n}) \leq K \frac{(1/2^n)^{1+\beta}}{2^{-\gamma n \alpha}} = K 2^{-n(1+\beta-\alpha\gamma)}.$$

By a union bound,

$$P\left(\sup_{1 \leq k \leq 2^n} |X_{k/2^n} - X_{(k-1)/2^n}| > 2^{-\gamma n}\right) \leq K \cdot 2^{-(\beta-\alpha\gamma)n}.$$

Since $\sum_{n=1}^{\infty} 2^{-(\beta-\alpha\gamma)n} < \infty$, by Borel–Cantelli, there exists $n_0 = n_0(\omega)$ such that for $n \geq n_0$,

$$|X_{k/2^n} - X_{(k-1)/2^n}| \leq 2^{-\gamma n}, \quad \forall 1 \leq k \leq 2^n. \quad (18)$$

Claim: for a.e. ω , X is uniformly γ -Hölder continuous on $D = \bigcup D_n = \bigcup (\mathbb{Z}/2^n \cap [0, 1])$, that is, there exists $M = M(\omega) > 0$ such that

$$|X_s - X_t| < M|t - s|^\gamma, \quad \forall t, s \in D.$$

Assume that the claim is proved. Noting that D is dense in $[0, 1]$, we can define

$$\tilde{X}_t = \begin{cases} X_t, & t \in D, \\ \lim_{D \ni t_m \rightarrow t} X_{t_m}, & t \notin D. \end{cases}$$

By the uniform γ -Hölder continuity, the limit is independent of (t_m) , and the resulting \tilde{X}_t is γ -Hölder continuous with the same constant $C(\omega)$.

Now we turn to the proof of the claim.

Let $t \in [\frac{k}{2^n}, \frac{k+1}{2^n}] \cap D$, $0 \leq k \leq 2^n - 1$, $n \geq n_0$. Then there exist a sequence $k/2^n = p_n/2^n$, $p_{n+1}/2^{n+1}, \dots, p_N/2^N = t$ such that

$$\left| \frac{p_m}{2^m} - \frac{p_{m+1}}{2^{m+1}} \right| = \frac{1}{2^{m+1}}, \quad n \leq m < N.$$

By triangle inequality and (18),

$$|X_t - X_{k/2^n}| \leq \sum_{m=n}^{N-1} |X_{p_m/2^m} - X_{p_{m+1}/2^{m+1}}| \leq \sum_{m=n}^{\infty} 2^{-\gamma m} = \frac{2^{-\gamma n}}{1 - 2^{-\gamma}}. \quad (19)$$

In particular, this and triangle inequality imply that X_t is bounded on $t \in D$. Let $M_0(\omega) = \sup_D X_t$.

For every $s < t$ in D , we can find the biggest n such that

$$\frac{k-1}{2^n} \leq s < \frac{k}{2^n} \leq t < \frac{k+1}{2^n},$$

and such n necessarily satisfies

$$\frac{1}{2^{n+1}} \leq |t - s| \leq \frac{1}{2^{n-1}}. \quad (20)$$

There are two cases.

Case 1: $n < n_0$. Since $|t - s| \geq 2^{-n_0}$, we have

$$\frac{|X_t - X_s|}{|t - s|^\gamma} \leq \frac{2M_0}{(2^{-n_0})^\gamma} := M_1(\omega).$$

Case 2: $n \geq n_0$. By triangle inequality, (19) and (20), we have

$$|X_s - X_t| \leq |X_s - X_{k/2^n}| + |X_{k/2^n} - X_t| \leq \frac{2^{-\gamma n+1}}{1 - 2^{-\gamma}} \leq \frac{2}{1 - 2^{-\gamma}} (2|t - s|)^\gamma := M_2|t - s|^\gamma.$$

Let $M = \max(M_1(\omega), M_2)$. Then $|X_t - X_s| \leq M|t - s|^\gamma$ for all $t, s \in D$. The claim is proved. \square

Homework The *Brown sheet* $(\mathbb{B}_{s,t})_{s,t \in [0,1]}$ is a centered Gaussian process with covariance

$$\mathbb{E}\mathbb{B}_{s,t}\mathbb{B}_{s',t'} = (s \wedge s')(t \wedge t'), \quad s, t, s', t' \in [0, 1].$$

It can be constructed via GWN with $H = L^2([0, 1]^2, \mathcal{B}([0, 1]^2), ds \times dt)$ and $\mathbb{B}_{s,t} = G(\mathbb{1}_{[0,s] \times [0,t]})$.

1. Show that for each $p \geq 1$, there is some constant $K_p > 0$,

$$\mathbb{E}|\mathbb{B}_{s,t} - \mathbb{B}_{s',t'}|^{2p} \leq K_p(|s - s'|^p + |t - t'|^p), \quad s, t, s', t' \in [0, 1].$$

2. Let $0 < \gamma < 1/2$. Show that with probability one, there is a random constant $n_0 = n_0(\omega)$ such that for all $n \geq n_0$,

$$\left| \mathbb{B}_{\frac{k}{2^n}, \frac{\ell}{2^n}} - \mathbb{B}_{\frac{k'}{2^n}, \frac{\ell'}{2^n}} \right| \leq 2^{-\gamma n}, \quad 0 \leq k, \ell, k', \ell' \leq 2^n, \quad |k - k'| + |\ell - \ell'| \leq 1.$$

2.4 Lévy's construction of Brownian motion

Using the proof of [Theorem 2.5](#), we can express Brownian motion explicitly in the form of (14). In fact, let $\{e_n\}$ be an ONB of $L^2([0, 1], dt)$ and $\xi_n \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$ on $(\Omega, \mathcal{F}, \mathbb{P})$. Then by [Theorem 2.5](#),

$$B_t(\omega) = \sum_{n=1}^{\infty} \xi_n(\omega) \langle e_n(x), \mathbb{1}_{[0,t]}(x) \rangle \quad (21)$$

is a Gaussian process with the f.d.d. of a Brownian motion; moreover, the infinite sum converges in $L^2(\Omega, \mathcal{F}, \mathbb{P})$. But we cannot derive continuity of $t \mapsto B_t(\omega)$ for fixed ω .

Let us take a closer look at the infinite series (21). Note that $\beta_n(t) = \langle e_n(x), \mathbb{1}_{[0,t]}(x) \rangle$ is a deterministic, continuous function. Hence, for every fixed N ,

$$B_t^N(\omega) = \sum_{n=1}^N \xi_n(\omega) \beta_n(t)$$

is also continuous in t for every ω . From classical analysis, for \mathbb{P} -a.e. ω , if the Cauchy criterion holds:

$$\sup_{t \in [0,1]} |B_t^N - B_t^{N'}|(\omega) \rightarrow 0, \quad N, N' \rightarrow \infty, \quad (22)$$

then $(B_t^N(\omega))_{t \in [0,1]}$ converges uniformly to some (random) continuous function $(\tilde{B}_t(\omega))_{t \in [0,1]}$. The two processes B and \tilde{B} must have the same f.d.d., since for fixed t , \tilde{B}_t is the a.s.-limit of B_t^N , while B_t is the L^2 -limit of B_t^N ; in other words, \tilde{B} will be a continuous modification of B .

The usual approach to verify the Cauchy criterion is to use *Weierstrass M-test*, which is an estimate for absolute convergence:

$$\sup_{t \in [0,1]} |B_t^N - B_t^{N'}|(\omega) \leq \sum_{N \leq n < N'} |\xi_n| \sup_{t \in [0,1]} |\beta_n(t)|. \quad (23)$$

Since $\xi_n \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$, it is easy to control the growth of ξ_n : by Borel–Cantelli and the Gaussian tail estimate $\mathbb{P}(|\mathcal{N}(0, 1)| \geq a) \leq e^{-a^2/2}$, with probability one, there is a random constant $n_0 = n_0(\omega)$ s.t.

$$|\xi_n| \leq \ln n, \quad \forall n \geq n_0(\omega).$$

Therefore, to apply the M -test, all we need is

$$\sum_{n=1}^{\infty} \ln n \cdot \sup_{t \in [0,1]} |\beta_n(t)| < \infty. \quad (24)$$

Can (24) be true? Let us look at a common choice for ONB on $L^2[0, 1]$ from Fourier series:

$$\{e_n(x)\} = \{1, \sqrt{2} \sin(2\pi n \cdot x), \sqrt{2} \cos(2\pi n \cdot x)\}.$$

For the corresponding $\beta_n(t)$, one has

$$\sup_{t \in [0,1]} |\beta_n(t)| \sim \frac{1}{n}.$$

Since $\sum_{n=1}^{\infty} \frac{\ln n}{n}$ diverges, the M -test cannot apply.

There are two fixes. The first one is to choose $\{e_n(x)\}$ more cleverly, so the Cauchy criterion (22) holds. See Lévy's construction in the exercise below.

Homework For $n \geq 0$ and $0 \leq k \leq 2^n - 1$, let

$$e_{n,k}(x) = \begin{cases} 2^{\frac{n}{2}}, & \frac{k}{2^n} \leq x < \frac{2k+1}{2^{n+1}}, \\ -2^{\frac{n}{2}}, & \frac{2k+1}{2^{n+1}} \leq x < \frac{k+1}{2^n}, \\ 0, & \text{otherwise,} \end{cases} \quad \beta_{n,k}(t) = \langle e_{n,k}, \mathbb{1}_{[0,t]} \rangle,$$

and $\xi_{n,k} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$. Define $\Delta B_t^n = \sum_{k=0}^{2^n-1} \xi_{n,k} \beta_{n,k}(t)$ and $B_t^N = \sum_{n=0}^N \Delta B_t^n$.

1. Show that $\{e_{n,k}\}$ is orthonormal, i.e.,

$$\int_0^1 e_{n,k}(x) e_{n',k'}(x) dx = \mathbb{1}_{n=n'} \mathbb{1}_{k=k'}.$$

2. Show that

$$\sup_{t \in [0,1]} |\Delta B_t^n| \leq 2^{-n/2} \cdot \max_{0 \leq k \leq 2^n-1} |\xi_{n,k}|.$$

Hint: note that for fixed n , $e_{n,k}$ has disjoint support for different k .

3. Use $P(|\mathcal{N}(0, 1)| \geq a) \leq e^{-a^2/2}$ and Borel–Cantelli Lemma to show that with probability one, there is a random constant $n_0 = n_0(\omega)$ such that

$$|\xi_{n,k}| \leq n, \quad \forall 0 \leq k \leq 2^n - 1, \quad n \geq n_0.$$

4. Conclude that with probability 1, $\{B_t^N(\omega), t \in [0, 1]\}_{N \geq 1}$ is Cauchy in $\mathcal{C}[0, 1]$, that is,

$$\lim_{N, N' \rightarrow \infty} \sup_{t \in [0,1]} |B_t^N(\omega) - B_t^{N'}(\omega)| = 0, \quad \text{a.e. } \omega.$$

Another convenient description of Lévy's construction is the following. Let X_k be i.i.d. $\mathcal{N}(0, 1)$ and $S_k = X_1 + \dots + X_k$. Define

$$\tilde{S}_t = \begin{cases} S_k, & t = k \in \mathbb{Z}, \\ (t - k)S_{k+1} + (t + 1 - k)S_k, & t \in (k, k + 1). \end{cases}$$

Then

$$B_t^N \stackrel{d}{=} \frac{\tilde{S}_{2^N t}}{2^{N/2}}.$$

In this representation, it is easy to verify that B^N has the same f.d.d. as Brownian motion at $t \in \mathbb{Z}/2^N$. By the Functional CLT, B^N converges to Brownian motion in distribution.

Another fix is to utilize the fluctuation of i.i.d. Gaussian and improve the bound on the right hand side of (23). As a comparison, recall the Kolmogorov's One-Series Theorem.

Theorem 2.10 *Let X_n be independent with $\mathbb{E}X_n = 0$ and $\sum_{n=1}^{\infty} \mathbb{E}X_n^2 < \infty$. Then $\sum_{n=1}^{\infty} X_n$ converges a.s.*

As a consequence of Theorem 2.10, we can put random ± 1 in front of $1/n$ and get a conditionally converging sum $\sum_{n=1}^{\infty} \frac{\pm 1}{n}$ since $\sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$. However, $\sum_{n=1}^{\infty} \frac{1}{n} = \infty$ so absolute convergence bound like (23) will fail.

In infinite dimension, the analogue is $\sum_{n=1}^{\infty} \beta_n^2 < \infty$ in the L^2 -sense:

$$\sum_{n=1}^{\infty} \int_0^1 \beta_n^2(t) dt = \int_0^1 \sum_{n=0}^{\infty} \langle e_n, \mathbb{1}_{[0,t]} \rangle^2 dt = \int_0^1 |\mathbb{1}_{[0,t]}|_{L^2[0,1]}^2 dt = \int_0^1 t dt < \infty.$$

Some general theory about Gaussian measures is develop to guarantee that (21) always converges almost surely, whatever the choice of the ONB $\{e_n\}$, which is a refinement of the construction in [Theorem 2.5](#) (see e.g. [\[PZ14, Part I, Theorem 2.12\]](#)).

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 $\{0, 1, 2, \dots\}$
\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>.
- [LeG16] Jean-François LeGall. *Brownian Motion, Martingales, and Stochastic Calculus*, volume 274 of *Graduate Texts in Mathematics*. Springer International Publishing, Cham, 2016.
- [PZ14] Giuseppe Da Prato and Jerzy Zabczyk. *Stochastic Equations in Infinite Dimensions*. Cambridge University Press, 2014.
- [Shi96] A. N. Shiryaev. *Probability*, volume 95 of *Graduate Texts in Mathematics*. Springer New York, New York, NY, 1996.

DRAFT