Lecture Note for MAT7093: Stochastic Analysis

LI Liying

March 12, 2024

1 Introduction

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

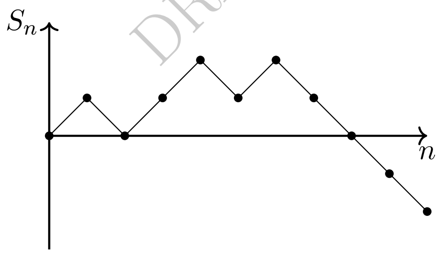
1.1 Stochastic processes

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

Write the partial sum as $S_n = X_1 + X_2 + \cdots + X_n$. We can plot the 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)

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

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. 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 $t \geq 0$,

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

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 $\mathcal{N}(0, t_1), \mathcal{N}(0, t_2 - t_1), \cdots, \mathcal{N}(0, t_m - t_{m-1})$ r.v.'s.

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

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 $0 = t_0 \leq t_1 < \cdots < t_m$,

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

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 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,

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

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 sequences (t_n) such that $t_n \to t_0$.

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

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

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 $(\Omega, \mathcal{F}, \mathsf{P})$, 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.

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

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

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

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

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

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

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 $(\Omega, \mathcal{F}, \mathsf{P})$ 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 1.7 1. $M = \mathbb{Z}$, $d(x,y) = |x-y|$.

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

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

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

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

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

Definition 1.6 *Let (M,d) be a metric space. An M-value random element (r.e.) 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*

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

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 $\big(\mathcal{C}(\mathbb{R}_+),\mathcal{B}(\mathcal{C}(\mathbb{R}_+))\big)$. Define

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

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 (or converge in distribution/law) to X, denoted by $X^{(n)} \Rightarrow_d X$, if for all bounded and continuous $F : \mathcal{C}[0,1] \to \mathbb{R}$,*

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

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 $\mu_n = \mathsf{P}^{(n)}\circ[X^{(n)}]^{-1}$ 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 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 P^W is supported on α-Hölder continuous functions, that is,*

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

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 $\mathcal{B}\big(\mathcal{C}[0,1]\big)$, 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

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

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*:

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

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

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

Moreover, due to stationarity, it only makes sense to have σ^2_{t-s} to be linear: $\sigma^2_{t-s} = K \cdot (t - s)$ 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 $\eta(t) = \dfrac{dB_t}{dt}$?*

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

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

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 (5), this issue could be circumvented by considering the equivalent integral equation

$$x(t) = x(0) + \int_0^t b\big(x(s)\big)\, ds + B(t). \tag{6}$$

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

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

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*

$$\int_0^t f(s)\eta(s)\ ds := \int_0^t f(s)\, dB(s) \tag{7}$$

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*

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

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 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)

$$\sum_{k=1}^{n} u_k(v_{k+1} - v_k) = u_{n+1}v_{n+1} - u_1 v_1 - \sum_{k=1}^{n} v_{k+1}(u_{k+1} - u_k)$$

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 are now consider the SDE

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

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

$$\int_0^t \sigma\big(s, x(s)\big)\, dB_s. \tag{10}$$

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 (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* $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

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* $\mathsf{P}(X = 0) = 1$.

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

Proof: The proof is elementary.

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

$\square$

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

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

$$\mu = \lim_{m \to \infty} \mu_m, \quad \sigma = \lim_{m \to \infty} \sigma_m. \tag{11}$$

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

Proof: The L^2-convergence of $X_m \to X$ implies the existence of both limits in (11). Hence, for each $\xi \in \mathbb{R}$, we have $\varphi_{X_m}(\xi) \to \exp(i\mu\xi - \frac{\sigma^2\xi^2}{2})$, 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. so $\varphi_{X_m}(\xi) \to \varphi_X(\xi)$. 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:

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

By choosing $q > p$, we see that $|X_m - X|^p$ is uniformly integrable. Since $|X_m - X| \to 0$ in probability, this and uniform integrability imply (see [Dur07, Sec. 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 $(B_t)_{t \geq 0}$ be Brownian motion. For any $0 \leq t_1 < t_2 < \cdots < t_m$, 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* 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* $Q : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ *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)* $\{b_1, \dots, b_d\}$, *and* $\varepsilon_1 \geq \varepsilon_2 \geq \dots \geq \varepsilon_r > 0 = \varepsilon_{r+1} = \dots = \varepsilon_d$ *such that*
$$X \overset{d}{=} Y = \mu_X + \sum_{i=1}^r \varepsilon \eta_i \cdot b_i, \quad \eta_i \overset{i.i.d.}{\sim} \mathcal{N}(0,1). \tag{12}$$

Proof: **From Definition 2.2 to Item 1.** Since $\langle \xi, X\rangle$ is Gaussian for every $\xi \in \mathbb{R}^d$, we have
$$\varphi_X(\xi) = \mathsf{E}e^{i\langle \xi, X\rangle} = e^{i\mathsf{E}\langle \xi, X\rangle - \frac{1}{2}\operatorname{Var}(\langle \xi, X\rangle)}.$$
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 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 $\Sigma = \operatorname{diag}\{\varepsilon_1^2, \dots, \varepsilon_d^2\}$.) 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 (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
$$\mathsf{E}\langle v, Y\rangle = \langle v, \mu_X\rangle, \quad \operatorname{Var}(\langle v, Y\rangle) = \sum_{i=1}^r \varepsilon_i^2 \langle v, b_i\rangle^2 = Q(v,v).$$
So X and Y have the same ch.f., and hence $\mathcal{L}(X) = \mathcal{L}(Y)$ as desired.

From Item 2 to Definition 2.2. 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. 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 $Q = (Q_{ij}) = \big(\mathrm{Cov}(X_i, X_j)\big)$ 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

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

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

Homework Let X and Y be i.i.d. with $\mathsf{E}X = \mathsf{E}Y = 0$ and $\mathsf{E}X^2 = \mathsf{E}Y^2 = 1$. 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

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

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 $\lambda(dt) = d\big(-f(t)\big)$. 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 $\mathsf{E}B_tB_s = t \wedge s$. 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 $G(f) = \displaystyle\int_0^1 f(t)dB_t$, 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 $0 = t_0 < t_1 < \cdots < t_m = 1$ and $f(s) = \displaystyle\sum_{i=0}^{m-1} f_i \mathbb{1}_{[t_i,t_{i+1})}(s)$. Then in light of the Riemann–Stieltjes integral, it only makes sense to define $G(f)$ as

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

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

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

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

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

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, 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 $L^2(\Omega, \mathcal{F}, \mathsf{P})$ with values being (centered) Gaussian r.v.'s. The isometry is given by*

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

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

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

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

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

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

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

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

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

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 $t_1 < \cdots < t_m$ and $v_1, \ldots, v_m$,

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

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

$$\mathsf{E}\Big|\int_0^\infty f(t)\,dB_t\Big|^2 = \mathsf{E}\big|G(f)\big|^2 = \int_0^\infty f^2(t)\,dt. \tag{15}$$

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*

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

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 $\mathsf{E}B_tB_s = t\wedge s$. Then $B_t - B_s \sim \mathcal{N}(0, t-s)$, and hence $\mathsf{E}|B_t - B_s|^n \leq K_n(t-s)^{n/2}$ for all $n \geq 1$. Since $\dfrac{n/2-1}{n}$ 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, $\mathsf{P}(X_t = Y_t) = 1$ for all $t \in \mathbb{R}$. Since the set of rational numbers $\mathbb{Q}$ is countable, we have $\mathsf{P}(X_t = Y_t,\ \forall t \in \mathbb{Q}) = 1$. That is, there is a set $\mathcal{N}$ with probability $\mathsf{P}(\mathcal{N}) = 0$, such that for all $\omega \in \mathcal{N}$,

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

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

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

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

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

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

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. □

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

By Markov inequality,

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

By a union bound,

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

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

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

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

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

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 $t \in [\frac{k}{2^n}, \frac{k+1}{2^n}] \cap D$, $0 \le k \le 2^n - 1$, $n \ge n_0$. 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

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

By triangle inequality and (18),

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

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

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

and such n necessarily satisfies

$$\frac{1}{2^{n+1}} \le |t-s| \le \frac{1}{2^{n-1}}. \tag{20}$$

There are two cases.

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

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

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

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

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

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

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

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

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

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

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 $n \ge n_0$,

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

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 $\xi_n \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$ on $(\Omega, \mathcal{F}, \mathsf{P})$. Then by Theorem 2.5,

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

is a Gaussian process with the f.d.d. of a Brownian motion; moreover, the infinite sum converges in $L^2(\Omega, \mathcal{F}, \mathsf{P})$. 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 P-a.e. ω, if the Cauchy criterion holds:

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

then $(B_t^N(\omega))_{t \in [0,1]}$ converges uniformly to some (random) continuous function $\big(\tilde{B}_t(\omega)\big)_{t \in [0,1]}$. 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:

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

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

$$|\xi_n| \le \ln n, \quad \forall n \ge n_0(\omega).$$

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

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

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

There are two fixes. The first one 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

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

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

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

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,

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

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

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

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 $\mathsf{E}X_n = 0$ and $\sum_{n=1}^{\infty} \mathsf{E}X_n^2 < \infty$. 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:

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

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 Filtration and Markov property

3.1 Filtration and stopping times

Definition 3.1 *Let $(X_t)_{t\geq 0}$ be a stochastic process defined on $(\Omega, \mathcal{F}, \mathsf{P})$.*

1. *A filtration $(\mathcal{F}_t)_{t\geq 0}$ is a family of increasing sub–σ-field of $\mathcal{F}_t$, namely,*

$$\mathcal{F}_{t_1} \subset \mathcal{F}_{t_2} \subset \mathcal{F}, \quad \forall 0 \leq t_1 < t_2.$$

2. *X_t is said to be* adapted *to $(\mathcal{F}_t)_{t\geq 0}$, if X_t is measurable w.r.t. $\mathcal{F}_t$ for all $t \geq 0$.*

Example 3.1 (Natural filtration) Let $(X_t)_{t\geq 0}$ be a stochastic process on $(\Omega, \mathcal{F}, \mathsf{P})$. The *natural filtration* is

$$\mathcal{F}_t^X := \sigma\big(X_s : 0 \leq s \leq t\big).$$

Roughly speaking, $\mathcal{F}_t^X$ is the information contained by the process X up to time t. By definition, X_t is $\mathcal{F}_t^X$-measurable, so X is $(\mathcal{F}_t^X)$-adapted.

Definition 3.2 *On the space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathsf{P})$,*

1. *a r.v. T is called a* stopping time *if $\{T \leq t\} \in \mathcal{F}_t$, $\forall t \geq 0$;*
2. *a r.v. T is called an* optional time *if $\{T < t\} \in \mathcal{F}_t$, $\forall t \geq 0$.*

There is a small difference between optional times and stopping times, but under mild assumptions they will be the same. We will see these assumptions by the end of this section. Nevertheless, the next two propositions give some relations between them.

Proposition 3.1 *If T is a stopping time, then T is also optional.*

Proof: We have

$$\{T < t\} = \bigcup_{n=1}^{\infty} \{T \leq t - \frac{1}{n}\} \in \sigma\big(\mathcal{F}_{t-\frac{1}{n}}, n \geq 1\big) \subset \mathcal{F}_t.$$

So T is optional. □

Let $\mathcal{F}_{t+} := \bigcap_{n=1}^{\infty} \mathcal{F}_{t+\frac{1}{n}} = \bigcap_{s>t} \mathcal{F}_s$. The two intersections are equivalent since $\mathcal{F}_t$ is a increasing in t.

Proposition 3.2 *If T is an optional time for $(\mathcal{F}_t)$, then it is a stopping time for $(\mathcal{F}_{t+})$.*

Proof: We have

$$\{T \le t\} = \bigcap_{n=1}^{\infty} \{T < t + \frac{1}{n}\} \in \bigcap_{n=1}^{\infty} \mathcal{F}_{t+\frac{1}{n}} = \mathcal{F}_{t+}.$$

$\square$

Example 3.2 The most common examples of stopping times and optional times are the hitting time of a set. Let $\Gamma \subset \mathbb{R}$ and $(X_t)_{t\ge 0}$ be a $(\mathcal{F}_t)$-adapted process. Then

$$T_\Gamma = \inf\{s \ge 0 : X_s \in \Gamma\}.$$

Proposition 3.3

1. *If Γ is open and X has right-continuous sample paths, then T_Γ is optional.*
2. *If Γ is closed and X has continuous sample paths, then T_Γ is stopping.*

Proof:

1. For $t \ge 0$, we have

   $$\{T_\Gamma < t\} = \{\exists s < t : \ X_s \in \Gamma\} = \{\exists q < t, q \in \mathbb{Q} : X_s \in \Gamma\} = \bigcup_{q\in\mathbb{Q}, q<t} \{X_q \in \Gamma\} \in \mathcal{F}_t,$$

   where the first equality is due to the definition of infimum,and the second equality due to right-continuity of paths and openness of Γ.

2. For $t \ge 0$, we have

   $$\{T_\Gamma > t\} = \{\{X_s\}_{s\in[0,t]} \cap \Gamma = \varnothing\} = \bigcup_{n=1}^{\infty} \{\mathrm{dist}(\{X_s\}_{s\in[0,t]}, \Gamma) \ge \frac{1}{n}\} = \bigcup_{n=1}^{\infty} \bigcap_{q\in[0,t]\cap\mathbb{Q}} \{\mathrm{dist}(X_q, \Gamma) \ge \frac{1}{n}\} \in \mathcal{F}_t.$$

   The continuity of X implies that $\{X_s\}_{s\in[0,t]}$ is a compact set, and hence if it does not intersect a closed set Γ, it must have positive distance to Γ; this gives the second equality.

$\square$

Definition 3.3 *A filtration $(\mathcal{F}_t)_{t\ge 0}$ is* right-continuous *if $\mathcal{F}_{t+} = \mathcal{F}_t$ for all $t \ge 0$.*

For a right-continuous filtration, stopping times and optional times are the same. An effortless way to get right-continuous filtration is just to replace $\mathcal{F}_t$ by $\mathcal{F}_{t+}$. Noting that since $\mathcal{F}_t \subset \mathcal{F}_{t+}$, if X_t is $(\mathcal{F}_t)$-adapted, then it is also $(\mathcal{F}_{t+})$-adapted.

Proposition 3.4 *Let $\mathcal{G}_t = \mathcal{F}_{t+}$. Then $(\mathcal{G}_t)_{t\ge 0}$ is right-continuous.*

Proof: We have

$$\mathcal{G}_{t+} = \bigcap_{n=1}^{\infty} \mathcal{G}_{t+\frac{1}{n}} = \bigcap_{n=1}^{\infty} \mathcal{F}_{(t+\frac{1}{n})+} \subset \bigcap_{n=1}^{\infty} \mathcal{F}_{t+\frac{2}{n}} = \mathcal{F}_{t+} = \mathcal{G}_t.$$

$\square$

It is still a valid question to ask how much $\mathcal{F}_t$ is different from $\mathcal{F}_{t+}$. If the filtration is generated by a nice process like the Brownian motion, then the answer is that $\mathcal{F}_t$ and $\mathcal{F}_{t+}$ only differ by null sets. In the case $t = 0$, this can be formulated by the following zero-one law.

Theorem 3.5 (Blumenthal's 0-1 law) *Let $B = (B_t)_{\geq 0}$ be the standard Brownian motion and $\mathcal{F}_t^B$ be its natural filtration. Then $\mathcal{F}_{0+}^B$ is trivial, i.e., $\mathsf{P}(A) = 0$ or 1 for all $A \in \mathcal{F}_{0+}^B$.*

Remark 3.3 Since $B_0 = 0$ for all ω, $\mathcal{F}_0^B = \{\varnothing, \Omega\}$.

Proof: For any $A \in \mathcal{F}_{0+}^B$, $0 < t_1 < \cdots < t_m$ and bounded continuous $g : \mathbb{R}^m \to \mathbb{R}$, we have

$$\begin{aligned}\mathsf{E}\mathbb{1}_A g(B_{t_1}, \cdots, B_{t_m}) &= \lim_{n\to\infty} \mathsf{E}\mathbb{1}_A g(B_{t_1} - B_{1/n}, \cdots, B_{t_m} - B_{1/n}) \\ &= \mathsf{E}\mathbb{1}_A \lim_{n\to\infty} \mathsf{E}\mathbb{1}_A g(B_{t_1} - B_{1/n}, \cdots, B_{t_m} - B_{1/n}) \\ &= \mathsf{P}(A) \cdot \mathsf{E}g(B_{t_1}, \cdots, B_{t_m}),\end{aligned}$$

where in the first and last equalities, we use the (right-)continuity of $t \mapsto B_t$ at $t = 0$ and the continuity of g, and the Bounded Convergence Theorem, and in the second equality, we use the independence of $B_{t_k} - B_{1/n}$ with $A \in \mathcal{F}_{1/n}$. Then, this implies that $\mathcal{F}_{0+}^B$ is independent of $\sigma(B_t, t > 0)$.

On the other hand, $\mathcal{F}_0^B = \{\varnothing, \Omega\}$, so $\sigma(B_t, t > 0) = \sigma(B_t, t \geq 0)$. Since $\mathcal{F}_{0+}^B \subset \sigma(B_t, t \geq 0)$, we see that $\mathcal{F}_{0+}^B$ is independent of itself. Any such σ-algebra has to be trivial, and this completes the proof. □

Using the zero-one law we can get some surprising results about the sample path of the Brownian motion.

Proposition 3.6 *With probability one,*

$$\forall \varepsilon > 0, \qquad \sup_{0\leq t\leq\varepsilon} B_t > 0 > \inf_{0\leq t\leq\varepsilon} B_t. \tag{25}$$

Proof: Consider the event

$$A = \bigcap_{n=1}^{\infty} \{ \sup_{0\leq t\leq 1/n} B_t > 0\}.$$

Then since A is the intersection of decreasing events, we have

$$\mathsf{P}(A) = \lim_{n\to\infty} \mathsf{P}(\sup_{0\leq t\leq 1/n} B_t > 0) \geq \liminf_{n\to\infty} \mathsf{P}(B_{1/n} > 0) = 1/2.$$

On the other hand, $A \in \mathcal{F}_{0+}^B$, so by Theorem 3.5, $\mathsf{P}(A) = 1$. Hence,

$$\mathsf{P}\big(\sup_{0\leq t\leq 1/n} B_t > 0\big) = 1, \quad \forall n \geq 1.$$

This implies that with probability one, $\sup_{0\leq t\leq\varepsilon} B_t > 0$ for all $\varepsilon > 0$. The other statement for the infimum can be proven similarly. □

We can say something about the zero set of Brownian motion.

Proposition 3.7 *With probability one, there exists a decreasing sequence $t_1(\omega) > t_2(\omega) > \cdots > 0$ such that $B_{t_i} = 0$, i.e., 0 is the limit point of the zero set of B_t.*

Proof: We will construct the sequence (t_i) inductively. By Theorem 3.5, assume (25) holds with probability one.

Take $\varepsilon = 1$ in (25). Then there exists $s_1, s_1' \in (0, 1]$ such that $B_{s_1} > 0 > B_{s_1'}$. Since $t \mapsto B_t$ is continuous, there exists t_1 between s_1 and s_1' such that $B_{t_1} = 0$.

Now suppose that $t_1, t_2, \ldots, t_n$ have been constructed. Then in (25) taking $\varepsilon = t_n$, there exist $s_{n+1}, s_{n+1}' \in (0, t_n]$ such that $B_{s_{n+1}} > 0 > B_{s_{n+1}'}$. Hence there exists t_{n+1} between these two numbers such that $B_{t_{n+1}} = 0$. Clearly $t_{n+1} < t_n$ by this construction. □

Remark 3.4 Suppose that our Brownian motion is constructed on $(\mathcal{C}[0,1], \mathcal{B}(\mathcal{C}[0,1]), \mathsf{P})$. Then clearly, the continuous function f defined by $f(t) = 0$ is not in the set A, so $A \neq \Omega = \mathcal{C}[0,1]$. This means that $\mathcal{F}_0^B \subsetneq \mathcal{F}_{0+}^B$.

Homework For $M > 0$, define $A_M = \bigcap_{n\geq 1} \{ \sup_{0<t\leq 1/n} \frac{B_t}{\sqrt{t}} > M\}$.

1. Show that $\mathsf{P}(A_M) \geq \mathsf{P}(\mathcal{N}(0,1) \geq M)$.
2. Use the zero-one law to deduce that $\mathsf{P}(A_M) = 1$.
3. For every $M > 0$, show that with probability one,
$$\sup_{0<t\leq \frac{1}{n}} \frac{B_t}{\sqrt{t}} > M, \quad \forall n \geq 1.$$
4. Show that with probability one,
$$\sup_{0<t\leq \frac{1}{n}} \frac{B_t}{\sqrt{t}} = +\infty, \quad \forall n \geq 1.$$

3.2 Markov property

We begin with the definition of a Markov process. If the range of t below is restricted to $t = n \in \mathbb{N}$, then one obtains a discrete-time Markov process.

Definition 3.4 *A stochastic process $X = (X_t)_{t\geq 0}$ is Markov if $\forall t, s > 0$,*

$$\mathsf{P}(X_{t+s} \in A \mid \mathcal{F}_t^X) = \mathsf{P}(X_{t+s} \in A \mid X_t), \quad \forall A \in \mathcal{B}(\mathbb{R}), \tag{26}$$

or equivalent,

$$\mathsf{E}\Big[F(X_{t+s}) \mid \mathcal{F}_t^X\Big] = \mathsf{E}\Big[F(X_{t+s}) \mid X_t\Big], \quad \forall F \text{ bounded and measurable.} \tag{27}$$

The intuitive meaning of Markov properties is that, conditioned on the past $(\mathcal{F}_t^X)$ is the same as conditioned at the present (X_t), or in other words, knowing the present state X_t, the future X_{t+s}, $s > 0$ is independent of the past $\mathcal{F}_t^X$.

Remark 3.5 With some more efforts, (26) or (27) are equivalent to their multidimensional versions: for any $t, s_1, \ldots, s_m > 0$,

$$\mathsf{P}((X_{t+s_1}, \cdots, X_{t+s_m}) \in A \mid \mathcal{F}_t^X) = \mathsf{P}((X_{t+s_1}, \cdots, X_{t+s_m}) \in A \mid X_t), \quad \forall A \in \mathcal{B}(\mathbb{R}^m) \tag{28}$$

and

$$\mathsf{E}\Big[F(X_{t+s_1}, \cdots, X_{t+s_m}) \mid \mathcal{F}_t^X\Big] = \mathsf{E}\Big[F(X_{t+s_1}, \cdots, X_{t+s_m}) \mid X_t\Big], \quad \forall F \text{ bounded and measurable.} \tag{29}$$

Since we will deal with conditional expectation very often, it is useful to collect some basic facts about conditional expection here.

Definition 3.5 *Let $X \in L^1(\Omega, \mathcal{F}, \mathsf{P})$ and $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-field. Then $\mathsf{E}[X \mid \mathcal{G}]$ is the unique $\mathcal{G}$-measurable r.v. (up to modification on a zero-probability set) such that for all $A \in \mathcal{G}$,*

$$\mathsf{E}\Big(\mathsf{E}[X \mid \mathcal{G}]\mathbb{1}_A\Big) = \mathsf{E}X\mathbb{1}_A.$$

Conditional expectation has the following properties. Their proofs can be found in any standard graduate probability textbook, say [Dur07, Shi96], etc.

Proposition 3.8 *The following identities are valid as long as the (conditional) expectations involved make sense.*

1. *If $X \in \mathcal{G}$, then $\mathsf{E}[XY \mid \mathcal{G}] = X\mathsf{E}[Y \mid \mathcal{G}]$.*
2. *If X is independent of $\mathcal{G}$, then $\mathsf{E}[X \mid \mathcal{G}] = \mathsf{E}X$ (that is, an almost sure constant).*
3. *If $\mathcal{G}_1 \subset \mathcal{G}_2$, then $\mathsf{E}\big[\mathsf{E}[X \mid \mathcal{G}_1] \mid \mathcal{G}_2\big] = \mathsf{E}\big[\mathsf{E}[X \mid \mathcal{G}_2] \mid \mathcal{G}_1\big] = \mathsf{E}[X \mid \mathcal{G}_1]$.*
   In particular, if $\mathsf{E}[X \mid \mathcal{G}_2]$ is $\mathcal{G}_1$-measurable, then $\mathsf{E}[X \mid \mathcal{G}_1] = \mathsf{E}[X \mid \mathcal{G}_2]$.

Besides, all the well-known limit theorems (Fatou, Monotone/Dominated/Bounded Convergence Theorems, etc) and inequalities (Jensen's equality) also a version for conditional expectation.

A key lemma we will use a lot in the context of Markov processes is the following.

Lemma 3.9 *If $X \in \mathcal{G}$ and Y is independent of $\mathcal{G}$, then for any bounded measurable function $F : \mathbb{R}^2 \to \mathbb{R}$, we have*

$$\mathsf{E}[F(X,Y) \mid \mathcal{G}] = \varphi(X),$$

where φ is a deterministic (Borel measurable) function given by

$$\varphi(x) = \mathsf{E}F(x,Y).$$

The above can also be written in short as

$$\mathsf{E}[F(X,Y) \mid \mathcal{G}] = \Big(\mathsf{E}[F(x,Y) \mid \mathcal{G}]\Big)\Big|_{x=X}. \tag{30}$$

Remark 3.6 We stress that the substitution of $x = X$ into a deterministic function φ makes the right-hand side of (30) $\sigma(X)$-measurable and hence $\mathcal{G}$-measurable.

Proof: Consider the class of functions

$$\mathcal{S} = \{F \text{ bounded measurable } : \mathbb{R}^2 \to \mathbb{R} \text{ such that (30) holds}\}.$$

Then $\mathcal{S}$ forms a monotone class, that is, if $F_n \in \mathcal{S}$ and $F_n \wedge F$, then $F \in \mathcal{S}$ as well. Therefore, to show that $\mathcal{S}$ contains all the bounded measurable functions, by standard measure-theoretical argument, it suffices to show that $F(x,y) = \mathbb{1}_A(x)\mathbb{1}_B(y) \in \mathcal{S}$ for all $A, B \in \mathcal{B}(\mathbb{R})$.

Indeed, since $\mathbb{1}_A(X) \in \mathcal{G}$ and $\mathbb{1}_B(Y)$ is independent of $\mathcal{G}$, we have

$$\mathsf{E}\big[\mathbb{1}_A(X)\mathbb{1}_B(Y) \mid \mathcal{G}\big] = \mathbb{1}_A(X)\mathsf{E}\big[\mathbb{1}_B(Y) \mid \mathcal{G}\big] = \mathbb{1}_A(X)\mathsf{P}(Y \in B) = \varphi(X)$$

where

$$\varphi(x) = \mathsf{E}\mathbb{1}_A(x)\mathbb{1}_B(Y) = \mathbb{1}_A(x)\mathsf{P}(Y \in B).$$

This proves the proposition. □

Example 3.7 The Brownian motion is a Markov process.

In fact, $B_{t+s} - B_t$ is independent of $(B_{t_1}, \cdots, B_{t_m})$ for all $t_1, \cdots, t_m \in [0,t]$, so $B_{t+s} - B_t$ is independent of $\mathcal{F}_t^X$. Hence, for all F bounded measurable, applying Lemma 3.9 to $G(x,y) = F(x+y)$, we have

$$\mathsf{E}\Big[F(B_{t+s}) \mid \mathcal{F}_t^X\Big] = \mathsf{E}\Big[G(B_{t+s} - B_t, B_t) \mid \mathcal{F}_t^X\Big] = \Big[\mathsf{E}G(B_{t+s} - B_t, y)\Big]\Big|_{y=B_t},$$

which is a function of B_t and hence $\sigma(B_t)$-measurable. Then Markov property follows from Item 3 in Proposition 3.8.

Example 3.8 Let $f \in L^2_{\mathrm{loc}}[0,\infty) = \{g : g\mathbb{1}_{[0,t]} \in L^2[0,t],\ \forall t > 0\}$. Consider the stochastic integral define via the Gaussian white noise:

$$X_t = \int_0^t f(s)\, dB_s := G\Big(f\mathbb{1}_{[0,t]}\Big).$$

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

In fact, the previous analysis for Brownian motion only uses the fact "independent increment" property. To see that such property also holds for X_t, we have from the definition of Gaussian white noise isometry, if $[t_1, t_2] \cap [t_3, t_4] = \varnothing$, then

$$\mathsf{E}(X_{t_4} - X_{t_3})(X_{t_2} - X_{t_1}) = \mathsf{E}G(f\mathbb{1}_{[t_3,t_4]})G(f\mathbb{1}_{[t_1,t_2]}) = \int_0^\infty f^2(s)\mathbb{1}_{[t_1,t_2]}(s)\mathbb{1}_{[t_3,t_4]}(s)\, ds = 0.$$

Since the increments are centered Gaussian, if their covariance is zero, then they are independent.

Homework Let $(B_t)_{t\in[0,1]}$ be the Brownian motion and define $X_t = B_t - tB_1$, $t \in [0,1]$. The process $X = (X_t)_{t\in[0,1]}$ is called the "Brownian Bridge".

1. Show that $(X_t)_{t\geq 0}$ is a centered Gaussian process with covariance

$$\mathsf{E}X_tX_s = s(1-t), \quad \forall 0 \leq s < t \leq 1.$$

2. Let $t > s > s_1 > s_2 > \cdots > s_n \geq 0$. Show that

$$\mathsf{E}\Big(X_t - \frac{1-t}{1-s}X_s\Big)X_{s_i} = 0, \quad 1 \leq i \leq n.$$

   Deduce that $X_t - \dfrac{1-t}{1-s}X_s$ is independent of $(X_{s_1}, \ldots, X_{s_n})$.

3. Let $t > s$. Show that $X_t - \dfrac{1-t}{1-s}X_s$ is independent of $\mathcal{F}^X_s$.
4. Show that $(X_t)_{t\in[0,1]}$ is Markov.

Next we will introduce the strong Markov property. While the usual Markov property states that future and past are conditionally independent if knowing the present, the strong Markov property allows the "present" to occur at a random stopping time. But first we need to understand how to condition on the information before a stopping time. Recall that a stopping time is a r.v. $T \in [0,\infty]$ such that $\{T \leq t\} \in \mathcal{F}^X_t$, $\forall t \geq 0$. In what follows, unless otherwise stated, $\mathcal{F}_t = \mathcal{F}^X_t$ and $\mathcal{F}_\infty = \sigma(\mathcal{F}_t, t \geq 0)$.

Definition 3.6 *The stopping σ-algebra is*

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty : \forall t \geq 0,\ A \cap \{T \leq t\} \in \mathcal{F}_t\}.$$

Intuitively, $\mathcal{F}_T$ contains the information before a stopping time T.

Example 3.9 Let $a \geq 0$ and consider $T = a$ (a constant r.v.). Then T is a stopping time since

$$\{T \leq t\} = \begin{cases} \Omega, & a \leq t, \\ \varnothing, & a > t \end{cases} \in \mathcal{F}_t, \quad \forall t \geq 0.$$

Moreover, $\mathcal{F}_T = \mathcal{F}_a$.

We can compare the stopping σ-algebras for different stopping time, or extract information from the stopping σ-algebra.

Proposition 3.10 *If $S \leq T$ are two stopping times, then $\mathcal{F}_S \subset \mathcal{F}_T$.*

Remark 3.10 Since $S \le T$, "information before S" is less than "information before T".

Proof: If $A \subset \mathcal{F}_S$, then for every $t \ge 0$,

$$A \cap \{T \le t\} = \Big(A \cap \{S \le t\}\Big) \cap \{T \le t\} \in \mathcal{F}_t.$$

So $A \subset \mathcal{F}_T$. This completes the proof. $\square$

Proposition 3.11 *If T is a stopping time and $S \ge T$ is random time such that S is $\mathcal{F}_T$-measurable, then S is also a stopping time.*

Proof: For each $t \ge 0$, since $\{S \le t\} \in \mathcal{F}_T$,

$$\{S \le t\} = \{S \le t\} \cap \{T \le t\} \in \mathcal{F}_t.$$

This completes the proof. $\square$

Remark 3.11 The stopping time S will take the form $S = f(T)$ for some measurable function f with $f(x) \ge x$.

We also need to impose more measurability constraint on our process $X = (X_t)_{t \ge 0}$.

Definition 3.7 *Let $X = (X_t)_{t\ge 0}$ be a stochastic process on $(\Omega, \mathcal{F}, \mathsf{P})$. We say that X is* measurable *if the map*

$$(t,\omega) \mapsto X_t(\omega) : \Big([0,\infty) \times \Omega, \mathcal{B}\big([0,\infty)\big) \otimes \mathcal{F}\Big) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

is measurable.

Proposition 3.12 *Let $X = (X_t)_{t\ge 0}$ be measurable and T be a (finite) r.v., then $X_T(\omega) := X_{T(\omega)}(\omega)$ is a r.v.*

Proof: The map $\omega \mapsto X_{T(\omega)}(\omega)$ is the composition of the following two measurable maps:

$$\omega \mapsto (t', \omega') = (T(\omega), \omega'), \quad (t', \omega') \mapsto X_{t'}(\omega').$$

The first map is measurable since T is a r.v., and the second map is measurable since X is measurable. This proves the proposition. $\square$

For adapted process, we introduce the notion of progressive measurability.

Definition 3.8 *Let $X = (X_t)_{t\ge 0}$ be an adapted process on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge 0}, \mathsf{P})$. We say that X is* progressively measurable *if for every fixed $t \ge 0$, the map*

$$(t,\omega) \mapsto X_t(\omega) : \Big([0,t] \times \Omega, \mathcal{B}\big([0,t]\big) \otimes \mathcal{F}_t\Big) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

is measurable.

Proposition 3.13 *Let $X = (X_t)_{t\ge 0}$ be an adapted process on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge 0}, \mathsf{P})$ which is progressively measurable and let T be a (finite) stopping time. Then $X_T := X_{T(\omega)}(\omega)$ is a $\mathcal{F}_T$-measurable r.v.*

Proof: Let $A \in \mathcal{B}(\mathbb{R})$. We have

$$\{X_T \in A\} \cap \{T \le t\} = \{X_{T\wedge t} \in A\} \cap \{T \le t\}.$$

It suffices to check that $\{X_{T\wedge t} \in A\} \in \mathcal{F}_t$.

In fact, the map $\omega \mapsto X_{T(\omega)\wedge t}(\omega)$ can be written as the composition of the two maps:

$$\omega \mapsto (t', \omega') = (T(\omega) \wedge t, \omega), \quad (t', \omega') \mapsto X_{t'}(\omega').$$

The first map is measurable from $(\Omega, \mathcal{F}_t)$ to $([0,t] \times \Omega, \mathcal{B}([0,t]) \times \mathcal{F}_t)$ by the definition of stopping times, while the second is measurable since X is progressively measurable. Hence, their composition is also measurable. This proves the proposition. $\square$

Proposition 3.14 *If X is $(\mathcal{F}_t)$-adapted and has right-continuous path, then X is also progressively measurable w.r.t. $(\mathcal{F}_t)$.*

Proof: Fix $t > 0$. For $n \geq 1$ and $0 \leq k \leq 2^n - 1$, define

$$X_s^{(n)}(\omega) = X_{(k+1)/2^n}(\omega), \quad s \in \Big(\frac{kt}{2^n}, \frac{(k+1)t}{2^n}\Big].$$

and $X_0^{(n)}(\omega) = X_0(\omega)$. Then for each n, since X is $(\mathcal{F}_t)$-adapted, it is easy to check that $(s, \omega) \mapsto X_s^{(n)}(\omega)$ is $\mathcal{B}([0,t]) \times \mathcal{F}_t$-measurable. Since for every ω, the sample path $s \mapsto X_s(\omega)$ is right-continuous, we have $\lim\limits_{n\to\infty} X_s^{(n)}(\omega) = X_s(\omega)$ for any $(s, \omega) \in [0,t] \times \Omega$. Therefore, the limit map $(s, \omega) \mapsto X_s(\omega)$ is also $\mathcal{B}([0,t]) \times \mathcal{F}_t$-measurable. This proves the proposition. □

We are ready to state the strong Markov property.

Definition 3.9 *A progressively measurable Markov process $X = (X_t)_{t\geq 0}$ has the strong Markov property if for each a.s. finite stopping time S,*

$$\mathsf{P}(X_{T+t} \in A \mid \mathcal{F}_T) = \mathsf{P}(X_{T+t} \mid X_T). \tag{31}$$

Remark 3.12 The strong Markov property can be stated including stopping time T with $\mathsf{P}(T = \infty) > 0$. In that case X_{T+t} makes no sense when $\{T = \infty\}$, so (31) only needs to hold on the set $\{T < \infty\}$. For simplicity, we always assume $T < \infty$ a.s. in the sequel.

The Brownian motion has the strong Markov property. We know more about the conditioned process after the any stopping time.

Theorem 3.15 *Let T be a stopping time and define $B_t^{(T)} = B_{T+t} - B_T$. Then $(B_t^{(T)})_{t\geq 0}$ is a standard Brownian motion independent of $\mathcal{F}_T$.*

In particular, Brownian motion has the strong Markov property.

We now use the theorem to check that $(B_t)_{t\geq 0}$ is strongly Markov. The proof of Theorem 3.15 will be postpone to the end of this section.

Derivation of the strong Markov property for $(B_t)_{t\geq 0}$ from Theorem 3.15: Since B is progressively measurable, B_T is $\mathcal{F}_T$-measurable. By Lemma 3.9 and the assumption that $(B_t^{(T)})_{t\geq 0}$ is independent of $\mathcal{F}_T$, for any bounded measurable function F,

$$\mathsf{E}\Big(F(B_{T+t}) \mid \mathcal{F}_T\Big) = \mathsf{E}\Big(F(B_T + B_t^{(T)}) \mid \mathcal{F}_T\Big) = \mathsf{E}\Big(F(B_t^{(T)} + x)\Big)|_{x = B_T} \in \sigma(B_T).$$

So by Item 3 of Proposition 3.8, the strong Markov property holds. □

An important consequence of the strong Markov property is the reflection principle. Consider the maximal process $B_t^* = \sup\limits_{0\leq s\leq t} B_t$ and the hitting time $T_a = \inf\{t \geq 0 : B_t = a\}$ for $a > 0$.

Theorem 3.16 (Reflection Principle) *For $a \geq b$,*

$$\mathsf{P}(B_t^* \geq a,\ B_t < b) = \mathsf{P}(B_t > 2a - b).$$

Proof: Clearly, $\{B_t^* \geq a\} = \{T_a \leq t\} \in \mathcal{F}_{T_a}$ and we have

$$\{B_t^* \geq a,\ B_t < b\} = \{T_a \leq t,\ B_{t-T_a}^{(T_a)} < b - a\}.$$

By Theorem Theorem 3.15, $(B_s^{(T_a)})_{s\ge 0}$ is independent of $\mathcal{F}_T$. Since $T_a \in \mathcal{F}_T$ and Brownian motion is symmetric, we see that in distribution,

$$(T_a, (B_s^{(T_a)})_{s\ge 0}) \stackrel{d}{=} (T_a, (-B_s^{(T_a)})_{s\ge 0}).$$

Therefore,

$$\mathsf{P}(T_a \le t,\ B_{t-T_a}^{(T_a)} < b-a) = \mathsf{P}(T_a \le t,\ -B_{t-T_a}^{(T_a)} < b-a) = \mathsf{P}(T_a \le t, B_{t-T_a}^{(T_a)} > a-b).$$

But on the event on the right-hand side, $B_t = B_{T_a} + B_{t-T_a}^{(T_a)} > 2a-b \ge a$, and by continuity, $B_t \ge a$ implies that $T_a \le t$. So we have

$$\mathsf{P}(T_a \le t, B_{t-T_a}^{(T_a)} > a-b) = \mathsf{P}(B_{t-T_a}^{(T_a)} > a-b) = \mathsf{P}(X_t > 2a-b),$$

where we use strong Markov property in the last equality. This proves the theorem. □

As a corollary, we have the distribution of the hitting time.

Proposition 3.17 *For $a > 0$,*

$$\mathsf{P}(T_a \le t) = \mathsf{P}(B_t^* \ge a) = 2\mathsf{P}(B_t \ge a).$$

Proof: Using Theorem 3.16 for $b = a$, we have

$$\mathsf{P}(B_t^* \ge a) = \mathsf{P}(B_t^* \ge a,\ B_t < a) + \mathsf{P}(B_t^* \ge a, B_t \ge a) = \mathsf{P}(B_t > 2a-a) + \mathsf{P}(B_t \ge a) = 2\mathsf{P}(B_t \ge a).$$

□

Proof of Theorem 3.15: Denote by $W = (W_t)_{t\ge 0}$ be a Brownian motion independent of $B = (B_t)_{t\ge 0}$. By the definition of conditional probability, it suffices to show that for all $0 \le t_1 < t_2 < \cdots < t_m$, all $A \in \mathcal{F}_T$ and all bounded continuous function F on $\mathbb{R}^m$, we have

$$\mathsf{E}F\big(B_{t_1}^{(T)}, B_{t_2}^{(T)}, \cdots, B_{t_m}^{(T)}\big)\mathbb{1}_A = \Big[\mathsf{E}F(W_{t_1}, W_{t_2}, \cdots, W_{t_m})\Big]\mathsf{P}(A). \tag{32}$$

Suppose T takes countably many values. Let $T \in \{s_1, s_2, \cdots\}$. Then the LHS of (32) is equal to

$$\begin{aligned}
&\sum_{n=1}^{\infty} \mathsf{E}F\big(B_{t_1}^{(T)}, B_{t_2}^{(T)}, \cdots, B_{t_m}^{(T)}\big)\mathbb{1}_A\mathbb{1}_{\{T=s_n\}} \\
&= \sum_{n=1}^{\infty} \mathsf{E}F\big(B_{s_n+t_1} - B_{s_n}, \cdots, B_{s_n+t_m} - B_{s_n}\big)\mathbb{1}_{A\cap\{T=s_n\}} \\
&= \sum_{n=1}^{\infty} \mathsf{E}\Big[\mathsf{E}\Big[F\big(B_{s_n+t_1} - B_{s_n}, \cdots, B_{s_n+t_m} - B_{s_n}\big)\mathbb{1}_{A\cap\{T=s_n\}} \mid \mathcal{F}_{s_n}\Big]\Big] \\
&= \sum_{n=1}^{\infty} \mathsf{E}\Big[\mathbb{1}_{A\cap\{T=s_n\}}\mathsf{E}\Big[F\big(B_{s_n+t_1} - B_{s_n}, \cdots, B_{s_n+t_m} - B_{s_n}\big) \mid \mathcal{F}_{s_n}\Big]\Big] \\
&= \sum_{n=1}^{\infty} \Big(\mathsf{E}\mathbb{1}_{A\cap\{T=s_n\}}\Big)\mathsf{E}F(W_{t_1}, \cdots, W_{t_m}) \\
&= \mathsf{P}(A)\cdot\mathsf{E}F(W_{t_1}, \cdots, W_{t_m}).
\end{aligned}$$

There are two crucial steps: in the third equality we use that $A \cap \{T = s_n\} \in \mathcal{F}_{s_n}$, which holds since T is a stopping time; in the fourth equality we use the simple Markov property for B.

General case. We approximate T be a sequence discrete stopping times:

$$T_k(\omega) = \frac{[2^k T] + 1}{2^k} = \sum_{n=0}^{\infty} \frac{n+1}{2^k} \mathbb{1}_{[\frac{n}{2^k}, \frac{n+1}{2^k})}(T(\omega)). \tag{33}$$

Indeed, T_k is stopping since for $t \in [n_0 2^{-k}, (n_0+1)2^{-k})$,

$$\{T_k(\omega) \le t\} = \{T \le \frac{n_0}{2^k}\} \in \mathcal{F}_{\frac{n_0}{2^k}} \subset \mathcal{F}_t,$$

or by Proposition 3.11. Also, since $|T_k - T| \le 2^{-k}$ and $T_k \ge T$, we have $T_k(\omega) \downarrow T(\omega)$ for every ω. Then by the right continuity of $t \mapsto B_t$, $B_t^{(T_k)} \to B_t^{(T)}$ as $k \to \infty$.

Now the left-hand side of (32) is equal to

$$\begin{aligned}
&\lim_{k\to\infty} \mathsf{E} \mathbb{1}_A F\big(B_{t_1+T_k} - B_{T_k}, \cdots, B_{t_m+T_k} - B_{T_k}\big) \\
&= \lim_{k\to\infty} \sum_{n=0}^{\infty} \mathsf{E} \mathbb{1}_{A \cap \{T \in [n2^{-k}, (n+1)2^{-k})\}} F\big(B_{t_1}^{(T_k)}, \cdots, B_{t_m}^{(T_k)}\big) \\
&= \lim_{k\to\infty} \sum_{n=0}^{\infty} \mathsf{E}\Big[\mathsf{E}\Big[\mathbb{1}_{A \cap \{T_k = (n+1)2^k\}} F\big(B_{t_1}^{(T_k)}, \cdots, B_{t_m}^{(T_k)}\big) \mid \mathcal{F}_{T_k}\Big]\Big] \\
&= \lim_{k\to\infty} \sum_{n=0}^{\infty} \mathsf{E}\Big[\mathbb{1}_{A \cap \{T_k = (n+1)2^k\}} \mathsf{E}\Big[F\big(B_{t_1}^{(T_k)}, \cdots, B_{t_m}^{(T_k)}\big) \mid \mathcal{F}_{T_k}\Big]\Big] \\
&= \lim_{k\to\infty} \sum_{n=0}^{\infty} \Big(\mathsf{E} \mathbb{1}_{A \cap \{T_k = (n+1)2^k\}}\Big) \cdot \mathsf{E} F(W_{t_1}, \cdots, W_{t_m}) \\
&= \mathsf{P}(A) \mathsf{E} F(W_{t_1}, \cdots, W_{t_m}).
\end{aligned}$$

In the third equality we use $A \in \mathcal{F}_T \subset \mathcal{F}_{T_k}$ (Proposition 3.10), and in the fourth equality we use the strong Markov property for T_k. □

Remark 3.13 The proof only relies on the simple Markov property (which guarantees strong Markov property for discrete stopping times) and the right-continuity of sample path (which is used for approximation argument).

Homework Let $B = (B_t)_{t\ge0}$ and $\tilde{B} = (\tilde{B}_t)_{t\ge0}$ be two independent $(\mathcal{F}_t)$-adapted Brownian motion on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge0}, \mathsf{P})$. Let T be an a.s. finite stopping time. Define

$$W_t(\omega) = \begin{cases} B_t(\omega), & t \le T(\omega), \\ B_{T(\omega)} + \Big(\tilde{B}_t(\omega) - \tilde{B}_{T(\omega)}(\omega)\Big), & t > T(\omega). \end{cases}$$

1. Show that $(W_t)_{t\ge0}$ is a continuous, $(\mathcal{F}_t)$-adapted stochastic process.
2. Prove that $W = (W_t)_{t\ge0}$ is a standard Brownian motion by showing that W and B have the same finite dimensional distribution, namely, for all $0 \le t_1 < t_2 < \cdots < t_m$ and all Borel sets $A_1, A_2, \ldots, A_m$,

$$\mathsf{P}(W_{t_1} \in A_1, \cdots, W_{t_m} \in A_m) = \mathsf{P}(B_{t_1} \in A_1, \cdots, B_{t_m} \in A_m).$$

Homework Let $B = (B_t)_{t\ge0}$ be the standard Brownian motion. For $a > 0$, let $T_a = \inf\{t \ge 0 : B_t = a\}$ be the first hitting time of a. For $\lambda > 0$, define the Laplace transform of T_a: $e^{-\varphi(\lambda,a)} = \mathsf{E} e^{-\lambda T_a}$. It is not hard to show that φ is a continuous function and we will assume that.

1. Use the strong Markov property to show that $T_a, T_{2a} - T_a, T_{3a} - T_{2a}, \ldots$ are i.i.d. random variables.
2. Show that
$$\varphi(\lambda, na) = n\varphi(\lambda, a), \quad n \geq 1,$$
and use continuity to conclude that $\varphi(\lambda, a) = \varphi(\lambda, 1)a$, $a > 0$.
3. Use the fact that $(\lambda B_{\lambda^{-2}t})_{t\geq 0}$ is also a standard Brownian motion for every $\lambda > 0$ to show that $T_{a\lambda}$ and $\lambda^2 T_a$ have the same distribution.
4. Show that $\varphi(\lambda^2, a) = \varphi(1, \lambda a)$ and conclude that there is a constant $c > 0$ such that
$$\mathsf{E}e^{-\lambda T_a} = e^{-c\sqrt{\lambda}a}.$$

3.3 Augmentation and usual condition

Definition 3.10 *We say that a filtration $(\mathcal{F}_t)$ satisfies the "usual condition" if*

1. *$\mathcal{F}_t = \mathcal{F}_{t+}$, i.e., it is right-continuous,*
2. *$\mathcal{F}_t$ is a complete σ-field.*

We recall the definition of a complete σ-field.

Definition 3.11 *We say that $\mathcal{G}$ is complete under the probability measure P if $N_1 \subset N_2$ where $N_2 \in \mathcal{G}$ and $\mathsf{P}(N_2) = 0$, then $N_1 \in \mathcal{G}$.*

We have seen that if a filtration is right-continuous, then optional times and stopping times are the same. In general, it is just simpler to work with complete probability space. We can always complete a σ-field by adding all the subsets of null sets. The completion of $\mathcal{G}$ under the probability measure P is

$$\begin{aligned}\bar{\mathcal{G}} &= \{G : \exists F \subset \mathcal{G} \text{ and } \mathsf{P}\text{-null set } N \in \mathcal{G} \text{ s.t. } F\Delta G \subset N\} \\ &= \{G : \exists F_1, F_2 \in \mathcal{G}, F_1 \subset F_2, \mathsf{P}(F_1) = \mathsf{P}(F_2) \text{ s.t. } F_1 \subset G \subset F_2\}.\end{aligned}$$

The completed measure on $\bar{\mathcal{G}}$ is defined by $\mathsf{P}(G) = \mathsf{P}(F)$.

With a $(\mathcal{F}_t)$-adapted process X, define the following collections of null sets

$$\begin{aligned}\mathcal{N}_t &= \{N : \exists F \subset \mathcal{F}_t^X : \mathsf{P}(F) = 0, \ N \subset F\} \\ \mathcal{N}_\infty &= \{N : \exists F \subset \mathcal{F}_\infty^X : \mathsf{P}(F) = 0, \ N \subset F\}.\end{aligned}$$

There are two ways to complete a filtration.

- **Completion**
$$\overline{\mathcal{F}}_t = \sigma(\mathcal{F}_t^X \cup \mathcal{N}_t) = \{G : \exists F \in \mathcal{F}_t^X \text{ s.t. } F\Delta G \in \mathcal{N}_t\}.$$
- **Augmentation**
$$\widetilde{\mathcal{F}}_t = \sigma(\mathcal{F}_t^X \cup \mathcal{N}_\infty) = \{G : \exists F \in \mathcal{F}_t^X \text{ s.t. } F\Delta G \in \mathcal{N}_\infty\}. \tag{34}$$

As we seen in Section 3.1, $\overline{\mathcal{F}}_t$ may not be right continuous: using the set A in the proof of Proposition 3.6, we see

$$\{\varnothing, \Omega\} = \mathcal{F}_0 = \overline{\mathcal{F}}_0 \subsetneq \mathcal{F}_{0+} \subset \overline{\mathcal{F}}_{0+}.$$

Indeed, from the zero-one law Theorem 3.5, even though $\mathcal{F}_{0+}$ is trivial, it still contains information strictly after time $t = 0$. This tells us just doing completion by adding null sets up to time t cannot lead to right-continuous filtration. However, if we do the augmentation, then the resulting filtration will be right-continuous, and thus satisfies the "usual condition".

Theorem 3.18 *Let X be the standard Brownian motion. Then the augmented filtration $(\widetilde{\mathcal{F}}_t)_{t\geq 0}$ is right-continuous.*

Proof: The first step is to show that for every bounded $\mathcal{F}^B_\infty$-measurable r.v. Y and $t \geq 0$,

$$\mathsf{E}\Big[Y \mid \mathcal{F}^B_{t+}\Big] = \mathsf{E}\Big[Y \mid B_t\Big]. \tag{35}$$

To prove (35), it suffices consider Y taking the form

$$Y = f(B_{t_1}, \cdots, B_{t_n}), \quad 0 \leq t_1 < \ldots < t_{m-1} < t \leq t_m < \cdots < t_n,$$

where f is a bounded continuous function. For $t = 0$, this is the main step in the proof of Theorem 3.5. For $t > 0$, the proof is similar; in order to get $\mathcal{F}^B_{t+}$ instead of $\mathcal{F}^B_t$, one needs to use the right-continuity of $t \mapsto B_t$: for $A \in \mathcal{F}^B_{t+}$,

$$\mathsf{E}\mathbb{1}_A f(B_{t_1}, \cdots, B_{t_n}) = \lim_{\varepsilon \downarrow 0} \mathsf{E}\mathbb{1}_A f(B_{t_1}, \cdots, B_{t_{m-1}}, B_{t_m+\varepsilon}, \cdots, B_{t_n+\varepsilon}).$$

Suppose that (35) is proved. Let $F \in \mathcal{F}^B_{t+} \subset \mathcal{F}^B_\infty$. Then by (35), $\mathsf{E}[\mathbb{1}_F \mid \mathcal{F}^B_{t+}]$ has a $\sigma(B_t)$-measurable version Z. On the other hand, $\mathsf{E}[\mathbb{1}_F \mid \mathcal{F}^B_{t+}] = \mathbb{1}_F$ a.s. Hence, for $A = \{Z = 1\} \in \mathcal{F}^B_t$, we have $F\Delta A \in \mathcal{N}_\infty$. This implies $F \in \widetilde{\mathcal{F}}_t$. Since F is arbitrary, $\mathcal{F}^B_{t+} \subset \widetilde{\mathcal{F}}_t$.

Next, let $F \subset \widetilde{\mathcal{F}}_{t+} = \bigcap_{n\geq 1} \widetilde{\mathcal{F}}_{t+\frac{1}{n}}$. Then by definition, there exist $G_n \in \mathcal{F}^B_{t+\frac{1}{n}}$ such that $F\Delta G_n \in \mathcal{N}_\infty$. We have

$$\begin{aligned}
& F\Delta G_n \in \mathcal{N}_\infty \\
\Leftrightarrow\ & \mathbb{1}_F + \mathbb{1}_{G_n} = 0 \mod 2 \text{ a.s.}, \quad \forall n \geq 1, \\
\Leftrightarrow\ & \mathbb{1}_F + \limsup_{n\to\infty} \mathbb{1}_{G_n} = 0 \mod 2 \text{ a.s.} \\
\Leftrightarrow\ & F\Delta(\limsup_{n\to\infty} G_n) \in \mathcal{N}_\infty,
\end{aligned}$$

where

$$\limsup_{n\to\infty} G_n = \bigcap_{k=1}^\infty \bigcup_{m=k}^\infty G_m \in \bigcap_{k=1}^\infty \mathcal{F}^B_{t+\frac{1}{k}} \subset \mathcal{F}^B_{t+}.$$

Since $\mathcal{F}^B_{t+} \subset \widetilde{\mathcal{F}}_t$ and $\widetilde{\mathcal{F}}_t$ is complete, we have $F \subset \widetilde{\mathcal{F}}_t$. This shows $\widetilde{\mathcal{F}}_{t+} \subset \widetilde{\mathcal{F}}_t$ and hence the right-continuity of $(\widetilde{\mathcal{F}}_t)_{t\geq 0}$. □

Remark 3.14 We only use the simply Markov property and the right-continuity of the Brownian motion.

3.4 Sample path properties of Brownian motion

In this section we mention some interesting sample path properties of Brownian motion.

Proposition 3.19 (Nowhere monotone) *With probability one, there is no interval $[a, b]$ such that*

$$B_{t_1} \leq B_{t_2} \leq B_{t_3}, \quad \forall a \leq t_1 < t_2 < t_3 \leq b,$$

or

$$B_{t_1} \geq B_{t_2} \geq B_{t_3}, \quad \forall a \leq t_1 < t_2 < t_3 \leq b,$$

Proof: For any $q_1 < q_2$, by Proposition 3.6, with probability one,

$$\sup_{q_1 \le s \le q_2} B_s > B_{q_1} > \inf_{q_1 \le s \le q_2} B_s.$$

Hence, with probability one, Brownian motion is non-monotone in any given interval. By a union bound, Brownian motion is non-monotone simultaneously in all intervals $[q_1, q_2]$, $q_1, q_2 \in \mathbb{Q}$. Since any monotone interval $[a, b]$, if existing, will contain a monotone sub-interval with rational endpoints, the desired conclusion follows. □

Proposition 3.20 (Nowhere differentiable) *With probability one, for every $t \ge 0$, either*

$$D^+ B_t = \limsup_{h \to 0+} \frac{B_{t+h} - B_t}{h} = \infty,$$

or

$$D_+ B_t = \liminf_{h \to 0+} \frac{B_{t+h} - B_t}{h} = -\infty,$$

Proof: See [KS98, pp. 110, Chap. 2, Theorem 9.18]. □

Proposition 3.21 *With probability one, all local maxima of $t \mapsto B_t$ is strict.*

Proof: For $t_1 < t_2 < t_3 < t_4$, let

$$A_{t_1,t_2,t_3,t_4} = \{\omega : \sup_{s \in [t_3,t_4]} B_s - \sup_{s \in [t_1,t_2]} B_s \ne 0\}.$$

Then on $\bigcap_{t_i \in \mathbb{Q}} A_{t_1,\dots,t_4}$, all local maxima are strict. It suffices to show that $\mathsf{P}(A_{t_1,\dots,t_4}) = 1$ for all t_i. Indeed, we have

$$\sup_{s \in [t_3,t_4]} B_s - \sup_{s \in [t_1,t_2]} B_s = (B_{t_3} - B_{t_2}) + \inf_{s \in [t_1,t_2]} (B_{t_2} - B_s) + \sup_{s \in [t_3,t_4]} (B_s - B_{t_3})$$

which are sum of three independent, continuous random variables. Hence $\mathsf{P}(A_{t_1,\dots,t_4}) = 1$. □

Proposition 3.22 *With probability one, the zero set*

$$N(\omega) = \{t \ge 0 : B_t = 0\}$$

is a perfect set (a closed, measure-zero set with no isolated point, like the Cantor set).

Proof: We have

$$\{\omega : N(\omega) \text{ has an isolated point}\} = \bigcup_{a,b \in \mathbb{Q}} \{\omega : \text{ there is exactly one } s \in (a, b) \text{ such that } B_s(\omega) = 0\}.$$

For $t \ge 0$, let $\beta_t = \inf\{s > t : B_s = 0\}$. Then β_t are stopping times. By Proposition 3.7, $\beta_0 = 0$. By the strong Markov properties, $B_{\beta(t)+h} - B_{\beta(t)}$ is a standard Brownian motion, so $\beta_{\beta(t)} = \beta_t$ almost surely. Hence,

$$\{\omega : \text{ there is exactly one } s \in (a, b) \text{ such that } B_s(\omega) = 0\} \subset \{\omega : \beta_a(\omega) < b \text{ and } \beta_{\beta_a(\omega)}(\omega) > b\}$$

has zero probability. This completes the proof. □

4 Martingales

4.1 Basic martingale theory

Definition 4.1 *An adapted stochastic process $(M_t)_{t\geq 0}$ on a probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathsf{P})$ is called a martingale if $M_t \in L^1(\Omega, \mathcal{F}, \mathsf{P})$ for all $t \geq 0$, and for all $s, t \geq 0$,*

$$\mathsf{E}\Big[M_{t+s} \mid \mathcal{F}_t\Big] = M_t.$$

If t only takes discrete values (like $\mathbb{Z}$), then we call (M_t) a discrete martingale.

Remark 4.1 If the filtration is not specified, we take the natural filtration $\mathcal{F}_t = \mathcal{F}_t^X$.

Example 4.2 Let X_i be independent random variables with $\mathsf{E}X_i = 0$. Then the partial sum $S_n = X_1 + \cdots + X_n$ forms a martingale, since by independence,

$$\mathsf{E}\Big[S_{n+m} \mid X_1, \cdots, X_n\Big] = X_1 + \cdots + X_n + \mathsf{E}[X_{n+1} + \cdots + X_m] = S_n.$$

Proposition 4.1 *Let $(X_t)_{t\geq 0}$ be a stochastic process with mean zero independent increments. Then*

1. *$(X_t)_{t\geq 0}$ is a martingale.*
2. *If $X_t \in L^2$ for all $t \geq 0$, then $(X_t^2 - \mathsf{E}X_t^2)_{t\geq 0}$ is a martingale.*
3. *If for some $\lambda \in \mathbb{R}$, $\mathsf{E}e^{\lambda X_t} < \infty$ for all $t \geq 0$, then $\left(\dfrac{e^{\lambda X_t}}{\mathsf{E}e^{\lambda X - t}}\right)_{t\geq 0}$ is a martingale.*

Proof:

1. This is obvious.
2. We have for $t > s$,

$$\begin{aligned}
&\mathsf{E}\Big[X_t^2 - X_s^2 \mid \mathcal{F}_s\Big] \\
&= \mathsf{E}\Big[\big(X_t - X_s + X_s\big)^2 - X_s^2 \mid \mathcal{F}_s\Big] \\
&= \mathsf{E}\Big[(X_t - X_s)^2 \mid \mathcal{F}_s\Big] + 2X_s\mathsf{E}\big[X_t - X_s \mid \mathcal{F}_s\big] \\
&= \mathsf{E}(X_t - X_s)^2 = \mathsf{E}(X_t - X_s)(X_t + X_s) - 2\mathsf{E}X_s(X_t - X_s) \\
&= \mathsf{E}X_t^2 - \mathsf{E}X_s^2.
\end{aligned}$$

3. We have for $t > s$,

$$\begin{aligned}
\mathsf{E}\Big[e^{\lambda X_t} \mid \mathcal{F}_s\Big] &= e^{\lambda X_s}\mathsf{E}\Big[e^{\lambda(X_t - X_s)} \mid \mathcal{F}_s\Big] \\
&= e^{\lambda X_s}\mathsf{E}e^{\lambda(X_t - X_s)} \\
&= e^{\lambda X_s}\frac{\mathsf{E}e^{\lambda X_t}}{\mathsf{E}e^{\lambda X_s}}.
\end{aligned}$$

□

Example 4.3 Let $(B_t)_{t\geq 0}$ be Brownian motion. Then $(B_t)_{t\geq 0}$, $(B_t^2 - t)_{t\geq 0}$, $(e^{\lambda B_t - \frac{1}{2}\lambda^2 t^2})_{t\geq 0}$ are all martingales.

Example 4.4 Let $f \in L^2_{\mathrm{loc}}[0,\infty)$ and consider the stochastic integral defined via Gaussian white noise

$$M_t = \int_0^\infty \mathbb{1}_{[0,t]}(s) f(s)\, dB_s = G\Big(\mathbb{1}_{[0,t]} f\Big).$$

Then (M_t) has mean zero independent increments, and the processes

$$(M_t)_{t\geq 0}, \quad \left(M_t^2 - \int_0^t f^2(s)\, ds\right)_{t\geq 0}, \quad \left(e^{\lambda M_t - \frac{1}{2}\lambda^2 \int_0^t f^2(s)\, ds}\right)_{t\geq 0}$$

are all martingales.

Example 4.5 Let $(N_t)_{t\geq 0}$ be a Poisson process with intensity λ, i.e.,

$$N_t = \max\{n \geq 0 : \xi_1 + \cdots + \xi_n \leq t\}$$

where $(\xi_i)_{i\geq 1}$ are i.i.d. $\mathrm{Exp}(\lambda)$ random variables. Then $(N_t - \lambda t)_{t\geq 0}$ has mean zero independent increments.

Definition 4.2 *Let $(M_t)_{t\geq 0}$ be an adapted process and assume that $M_t \in L^1$ for all $t \geq 0$. We say that $(M_t)_{t\geq 0}$ is a* super-martingale *if*

$$\mathsf{E}[X_t \mid \mathcal{F}_s] \leq X_s, \quad \forall 0 \leq s < t,$$

and say that $(M_t)_{t\geq 0}$ is a sub-martingale *if*

$$\mathsf{E}[X_t \mid \mathcal{F}_s] \geq X_s, \quad \forall 0 \leq s < t.$$

One can use convex/concave functions to generate super- or sub-martingale from martingales.

Proposition 4.2 *If $(M_t)_{t\geq 0}$ is a martingale, and $\varphi : \mathbb{R} \to \mathbb{R}$ is a convex function, then $\big(\varphi(M_t)\big)_{t\geq 0}$ is a sub-martingale.*

Proof: Using Jensen's inequality for conditional expectation, we have for all $s < t$,

$$\mathsf{E}\big[\varphi(M_t) \mid \mathcal{F}_s\big] \geq \varphi\Big(\mathsf{E}\big[X_t \mid \mathcal{F}_s\big]\Big) = \varphi(X_s). \tag{36}$$

□

Corollary 4.3 *If $(M_t)_{t\geq 0}$ is a sub-martingale and $\varphi : \mathbb{R} \to \mathbb{R}$ is convex and increasing, then $\big(\varphi(M_t)\big)_{t\geq 0}$ is also a sub-martingale.*

Proof: Since φ is increasing and $(M_t)_{t\geq 0}$ is a sub-martingale, the last equality in (36) will become

$$\varphi\Big(\mathsf{E}\big[X_t \mid \mathcal{F}_s\big]\Big) \geq \varphi(X_s),$$

and this completes the proof. □

Example 4.6 The function $|x|^p$ ($p \geq 1$) is convex. The functions $x \vee a$ ($a \in \mathbb{R}$), $x^+ = x \vee 0$ are convex and increasing.

4.2 Convergence of martingales

In this section we discuss the a.s.-limit and L^1-limit of martingales. The main tools are *Doob's Up-crossing Theorem* and uniform integrability.

Let (X_t) be an adapted process (continuous-time or discrete-time) and $a < b$. Consider the following stopping times: $T_b^{(0)} = -\infty$,

$$T_a^{(\ell)} = \inf\{t \geq T_b^{(\ell-1)} : X_t \leq a\}, \quad T_b^{(\ell)} = \inf\{t \geq T_a^{(\ell)} : X_t \geq b\}, \quad \ell \geq 1.$$

In every interval $[T_a^{(\ell)}, T_b^{(\ell)}]$, the process (X_t) completes an up-crossing. The total number of up-crossing in a given interval $[0, n]$ is defined by

$$U_{ab}^X[0, n] = \max\{k : T_b^{(k)} \leq n\}.$$

Theorem 4.4 *Let $(X_n)_{n\geq 1}$ be a sub-martingale, then*

$$\mathsf{E}U_{ab}^X[0, n] \leq \frac{1}{b-a}\mathsf{E}(X_n - a)^+.$$

We have the following corollary about a.s. convergence of martingales.

Proposition 4.5 *If $(X_n)_{n\geq 1}$ is a sub-martingale, and $\sup_n \mathsf{E}X_n^+ < \infty$. Then there exists X such that $X_n \to X$ a.s.*

Proof: The up-crossing number is increasing in n, and hence by assumption,

$$\mathsf{E}U_{ab}^X[0, \infty) = \lim_{n\to\infty} \mathsf{E}U_{ab}^X[0, n] \leq \frac{\sup_n \mathsf{E}X_n^+ + |a|}{b-a} < \infty.$$

This implies $U_{ab}^X[0, \infty) < \infty$ a.s., that is, with probability one, any interval $[a, b]$ is being up-crossed by at most finitely many times. As a consequence, for fixed

$$\liminf_{n\to\infty} X_n < a < b < \limsup_{n\to\infty} X_n$$

cannot happen. Taking a union bound over all $[a, b]$ with $a, b \in \mathbb{Q}$, we see that with probability one,

$$\limsup_{n\to\infty} X_n = \liminf_{n\to\infty} X_n.$$

This proves the statement. □

Example 4.7 If a martingale $(X_n)_{n\geq 0}$ is non-negative, then $\mathsf{E}X_n^+ = \mathsf{E}X_n = \mathsf{E}X_0$, and hence $\lim_{n\to\infty} X_n$ exists.

Next we will discuss the L^1-convergence. Recall the definition of uniform integrability for a family of random variables $\{X_n\}$.

Definition 4.3 *A family of random variables (X_n) is* uniformly integrable, *if*

$$\lim_{M\to\infty} \sup_n \mathsf{E}\mathbb{1}_{\{|X_n|\geq M\}}|X_n| = 0.$$

Uniform integrability is the necessary and sufficient condition for L^1-convergence.

Theorem 4.6 *If $X_n \to X$ a.s., then $X_n \to X$ if and only if (X_n) is uniformly integrable.*

Example 4.8 The following conditions will imply uniform integrability.

1. If there exists $Z \in L^1$ such that $|X_n| \le Z$ for all n, then (X_n) is uniformly integrable. (This is Dominated Convergence Theorem.)
2. If $\sup_n \mathsf{E}|X_n|^p < \infty$ for some $p > 1$, then (X_n) is uniformly integrable.
3. Let $Z \in L^1$. Then the collection of r.v.'s $\{\mathsf{E}[Z \mid \mathcal{G}] : \mathcal{G} \subset \mathcal{F}\}$ is uniformly integrable.

We will prove the last point.

Proposition 4.7 *Let $Z \in L^1(\Omega, \mathcal{F}, \mathsf{P})$. Then the collection of r.v.'s*

$$\{\mathsf{E}[Z \mid \mathcal{G}] : \mathcal{G} \text{ is a sub–}\sigma\text{-field of } \mathcal{F}\}$$

is uniformly integrable.

Proof: Since $Z \in L^1(\Omega, \mathcal{F}, \mathsf{P})$, for every $\varepsilon > 0$, there is $\delta > 0$ such that whenever $\mathsf{P}(A) < \delta$, $\mathsf{E}|Z|\mathbb{1}_A < \varepsilon$.

By Jensen's inequality, for $A = \{|\mathsf{E}[Z \mid \mathcal{G}]| \ge M\} \in \mathcal{G}$, we have

$$\mathsf{E}\mathbb{1}_A|\mathsf{E}[Z \mid \mathcal{G}]| \le \mathsf{E}\mathbb{1}_A\mathsf{E}[|Z| \mid \mathcal{G}] = \mathsf{E}\mathsf{E}[|Z|\mathbb{1}_A \mid \mathcal{G}] = \mathsf{E}|Z|\mathbb{1}_A.$$

When $A = \Omega$, the above inequality gives $\mathsf{E}|\mathsf{E}[Z \mid \mathcal{G}]| \le \mathsf{E}|Z|$. Then by Markov inequality,

$$\mathsf{P}(A) \le \frac{\mathsf{E}|Z|}{M},$$

uniformly for all sub–σ-field $\mathcal{G}$. Combining all these together we prove the statement. □

Proposition 4.8 *A martingale (X_n) is uniformly integrable, if and only if there exists $X_\infty \in L^1$ such that $X_n = \mathsf{E}[X_\infty \mid \mathcal{F}_n]$.*

Proof: **The "⇒" direction.** Uniform integrability implies that $\sup_n \mathsf{E}|X_n| < \infty$, hence Proposition 4.5 implies that there exists X_∞ such that $X_n \to X_\infty$ a.s. But (X_n) is also uniformly integrable, so the limit is also in L^1. Then,

$$\mathsf{E}[X_\infty \mid \mathcal{F}_n] = \lim_{m\to\infty} \mathsf{E}[X_{n+m} \mid \mathcal{F}_n] = X_n$$

as desired.

The "⇐" direction. It follows from Proposition 4.7. □

Adaption to the continuous-time.

The L^1-convergence relies on the uniform integrability, which holds true for continuous-time. The a.s. convergence relies on the up-crossing inequality, and we need extra continuity assumption to take the limit of approximation.

Theorem 4.9 (Continuous-time Doob's Up-crossing Inequality) *Let $(X_t)_{t\ge 0}$ be a right-continuous sub-martingale, then for all $T > 0$*

$$\mathsf{E}U_{ab}^X[0,T] \le \frac{1}{b-a}\mathsf{E}(X_T - a)^+.$$

Proof: We can restrict the definition of up-crossings to $D_n = \mathbb{Z}/2^n$. We denote the number of up-crossing observed on D_n by $U^X_{ab}[0,T] \cap D_n$. Since D_n has few points, the number of up-crossing is smaller (an up-crossing can occur on an interval $(k/2^n, (k+1)/2^n)$ and not "seen" by the set D_n). But since X has right-continuous path,

$$U^X_{ab}[0,T] \cap D_n \uparrow U^X_{ab}[0,T], n \to \infty,$$

almost surely. Now on D_n, $(X_t)_{t\in D_n}$ is just a discrete martingale, and we have

$$\mathsf{E}U^X_{ab}[0,T] \cap D_n \leq \frac{1}{b-a}\mathsf{E}(X_T - a)^+.$$

Taking the limit $n \to \infty$ and using the Monotone Convergence Theorem prove the statement. □

5 Notations

5.1 Abbreviations

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

5.2 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

5.3 Relations

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

5.4 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 $[a,b]$

5.5 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 $\mathcal{X}$
$A\Delta B = (A \setminus B) \cup (B \setminus A)$	the difference set.

5.6 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.

[KS98] Ioannis Karatzas and Steven Shreve. *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics. Springer-Verlag, New York, 2 edition, 1998.

[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.