MAT6081A Topics in Analysis I

2^{nd} term, 2016-17

Teacher: Professor Ka-Sing Lau

Schedule: Wednesday, 2.30-5.00 pm

Venue: LSB 222

Topics: Introduction to Stochastic Calculus

In the past thirty years, there has been an increasing demand of stochastic calculus in mathematics as well as in various disciplines such as mathematical finance, pde, physics and biology. The course is a rigorous introduction to this topic. The material include conditional expectation, Markov property, martingales, stochastic processes, Brownian motions, Ito's calculus, and stochastic differential equations.

Prerequisites

Students are expected to have good background in real analysis, probability theory and some basic knowledge of stochastic processes.

References: There will be lecture notes. The other references include

1. A Course in Probability Theory, K.L. Chung, (1974).
2. Measure and Probability, P. Billingsley, (1986).
3. Introduction to Stochastic Integration, H.H. Kuo, (2006).
4. Intro. to Stochastic Calculus with Application, F. Klebaner, (2001).
5. Brownian Motion and Stoch. Cal., I. Karatzas and S. Shreve, (1998).
6. Stoch. Cal. for Finance II– Continuous time model, S. Shreve, (2004).

Everyone knows calculus deals with deterministic objects. On the other hand *stochastic calculus* deals with random phenomena. The theory was introduced by Kiyosi Ito in the 40's, and therefore stochastic calculus is also called *Ito calculus*. Besides its interest in mathematics, it has been used extensively in statistical mechanics in physics, the filter and control theory in engineering. Nowadays it is very popular in the option price and hedging in finance. For example the well-known Black-Scholes model is

$$dS(t) = rS(t)dt \ + \ \sigma S(t)dB(t)$$

where $S(t)$ is the stock price, σ is the volatility, and r is the interest rate, and $B(t)$ is the Brownian motion. The most important notion for us is the Brownian motion. As is known the botanist R. Brown (1828) discovered certain zigzag random movement of pollens suspended in liquid. A. Einstein (1915) argued that the movement is due to bombardment of particle by the molecules of the fluid. He set up some basic equations of Brownian motion and use them to study diffusion. It was N. Wiener (1923) who made a rigorous study of the Brownian motion using the then new theory of Lebesgue measure. Because of that a Brownian motion is also frequently called a Wiener process.

Just like calculus is based on the *fundamental theorem of calculus*, the Ito calculus is based on the *Ito Formula*: Let f be a twice differentiable function on $\mathbb{R}$, then

$$f(B(t)) - f(B(0)) = \int_0^T f'(B(t))dB(t) + \frac{1}{2}\int_0^T f''(B(t))dt$$

where $B(0) = 0$ to denote the motion starts at 0. There are formula for integration, for example, we have

$$\int_0^T B(t)dB(t) = \frac{1}{2}B(t)^2 - \frac{1}{2}T; \qquad \int_0^T tdB(t) = TB(T) - \int_0^T B(t)dt.$$

In this course, the prerequisite is real analysis and basic probability theory. In real analysis, one needs to know σ-fields, measurable functions, measures and integration theory, various convergence theorems, Fubini theorem and the Radon-Nikodym theorem. We will go through some of the probability theory on conditional expectation, optional r.v. (stopping time), Markov property, martingales ([1], [2]). Then we will go onto study the Brownian motion ([2], [3], [5]), the stochastic integration and the Ito calculus ([3], [4], [5]).

Chapter 1

Basic Probability Theory

1.1 Preliminaries

Let Ω be a set and let $\mathcal{F}$ be a family of subsets of Ω, $\mathcal{F}$ is called a *field* if it satisfies

(i) $\emptyset,\ \Omega \in \mathcal{F}$;

(ii) for any $A \in \mathcal{F},\ A^c \in \mathcal{F}$;

(iii) for any $A,\ B \in \mathcal{F},\ A \cup B \in \mathcal{F}$ (hence $A \cap B \in \mathcal{F}$).

It is called a *σ-field* if (iii) is replaced by

(iii)′ for any $\{A_n\}_{n=1}^{\infty} \subset \mathcal{F},\ \cup_{n=1}^{\infty} A_n \in \mathcal{F}$ (hence $\cap_{n=1}^{\infty} A_n \in \mathcal{F}$).

If $\Omega = \mathbb{R}$ and $\mathcal{F}$ is the smallest σ-field generated by the open sets, then we call it the Borel field and denote by $\mathcal{B}$.

A *probability space* is a triple $(\Omega, \mathcal{F}, P)$ such that $\mathcal{F}$ is a σ-field in Ω, and $P : \mathcal{F} \to [0, 1]$ satisfies

(i) $P(\Omega) = 1$

(ii) countable additivity : if $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$ is a disjoint family, then

$$P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n).$$

We call Ω a sample space, $A \in \mathcal{F}$ an event (or measurable set) and P a probability measure on Ω; an element $\omega \in \Omega$ is called an outcome.

Theorem 1.1.1. *(Caratheodory Extension Theorem) Let $\mathcal{F}_0$ be a field of subsets in Ω and let $\mathcal{F}$ be the $\sigma-$ field generated by $\mathcal{F}_0$. Let $P: \mathcal{F}_0 \to [0,1]$ satisfies (i) and (ii) (on $\mathcal{F}_0$). Then P can be extended uniquely to $\mathcal{F}$, and $(\Omega, \mathcal{F}, P)$ is a probability space.*

The proof of the theorem is to use the outer measure argument.

Example 1. Let $\Omega = [0,1]$, let $\mathcal{F}_0$ be the family of set consisting of finite disjoint unions of half open intervals $(a,b]$ and $[0,b]$, Let $P((a,b]) = |b-a|$. Then $\mathcal{F}$ is the Borel field and P is the Lebesgue measure on $[0,1]$.

Example 2. Let $\{(\Omega_n, \mathcal{F}_n, P_n)\}_n$ be a sequence of probability spaces. Let $\Omega = \prod_{n=1}^{\infty} \Omega_n$ be the product space and let $\mathcal{F}_0$ be the family of subsets of the form $E = \prod_{n=1}^{\infty} E_n$, where $E_n \in \mathcal{F}_n$, $E_n = \Omega_n$ except for finitely many n. Define

$$P(E) = \prod_{n=1}^{\infty} P_n(E_n)$$

Let $\mathcal{F}$ be the σ-field generated $\mathcal{F}_0$, then $(\Omega, \mathcal{F}, P)$ is the standard infinite product measure space.

Example 3. (Kolmogorov Extension Theorem) Let P_n be probability measures on $(\prod_{k=1}^{n} \Omega_k, \mathcal{F}_n)$ satisfying the following consistency condition: for $m \leq n$

$$P_n \circ {\pi_{nm}}^{-1} = P_m$$

where $\pi_{nm}(x_1 \cdots x_n) = (x_1 \cdots x_m)$. On $\Omega = \prod_{k=1}^{\infty} \Omega_k$, we let $\mathcal{F}_0$ be the field of sets $F = E \times \prod_{k=n+1}^{\infty} \Omega_k$, $E \in \mathcal{F}_n$ and let

$$P(F) = P_n(E).$$

Then this defines a probability spaces $(\Omega, \ \mathcal{F}, P)$, where $\mathcal{F}$ is the σ-field generated by $\mathcal{F}_0$.

Remark: The probability space in Example 2 is the underlying space for a sequence of independent random variables. Example 3 is for more general sequence of random variables (with the consistency condition).

A random variable (r.v.) X on $(\Omega, \mathcal{F})$ is an (extended) real valued function $X : (\Omega, \mathcal{F}) \to \mathbb{R}$ such that for any Borel subset B of $\mathbb{R}$,

$$X^{-1}(B) = \{\omega : X(\omega) \in B\} \in \mathcal{F}.$$

(i.e. X is $\mathcal{F}$-measurable). We denote this by $X \in \mathcal{F}$. It is well known that

– For $X \in \mathcal{F}$, X is either a simple function (i.e., $\sum_{k=1}^{n} a_k \chi_{A_k}(\omega)$ where $A_k \in \mathcal{F}$), or is the pointwise limit of a sequence of simple functions.

– Let $X \in \mathcal{F}$ and g is a Borel measurable function, then $g(X) \in \mathcal{F}$.

– If $\{X_n\} \subseteq \mathcal{F}$ and $\lim_{n\to\infty} X_n = X$, then $X \in \mathcal{F}$.

– Let $\mathcal{F}_X$ be the σ-field generated by X, i.e., the sub-σ-field $\{X^{-1}(B) : B \in \mathcal{B}\}$. Then for any $Y \in \mathcal{F}_X$, $Y = \varphi(X)$ for some extended-valued Borel function φ on $\mathbb{R}$.

Sketch of proof ([1, p.299]): First prove this for simple r.v. Y so that $Y = \varphi(X)$ for some simple function φ. For a bounded r.v. $Y \geq 0$, we can find a sequence of increasing simple functions $\{Y_n\}$ such that $Y_n = \varphi_n(X)$ and

$Y_n \nearrow Y$. Let $\varphi(x) = \overline{\lim}_n \varphi_n(x)$, hence $Y = \varphi(X)$. Then prove Y for the general case.

A r.v. $X : (\Omega, \mathcal{F}) \to \mathbb{R}$ induces a *distribution (function)* on $\mathbb{R}$:

$$F(x) = F_X(x) = P(X \leq x).$$

It is a non-decreasing, right continuous function with $\lim_{n\to-\infty} F(x) = 0$, $\lim_{n\to\infty} F(x) = 1$. The distribution defines a measure μ

$$\mu((a, b]) = F(b) - F(a)$$

(use the Caratheodory Extension Theorem here). More directly, we can define μ by

$$\mu(B) = P(X^{-1}(B)) \ , \quad B \in \mathcal{B}.$$

The jump of F at x is $F(x) - F(x-) = P(X = x)$. A r.v. X is called a *discrete* if F is a jump function; X is called a *continuous* r.v. if F is continuous, i.e., $P(X = x) = 0$ for each $x \in \mathbb{R}$, and X is said to have a density function $f(x)$ if F is absolutely continuous with the Lebesgue measure and $f(x) = F'(x)$ a.e., equivalently $F(x) = \int_{-\infty}^{x} f(y)dy$.

For two random variables X, Y on $(\Omega, \mathcal{F})$, the random vector (X, Y) : $(\Omega, \mathcal{F}) \to \mathbb{R}^2$ induces a distribution F on $\mathbb{R}^2$

$$F(x, y) = P(X \leq x, \ Y \leq y)$$

and F is called the joint distribution of (X, Y), the corresponding measure μ is given by

$$\mu((a, b] \times (c, d]) = F(b, d) - F(a, d) - F(b, c) + F(a, c),$$

Similarly we can define the joint distribution $F(x_1 \cdots x_n)$ and the corresponding measure.

For a sequence of r.v., $\{X_n\}_{n=1}^{\infty}$, there are various notions of convergence.

(a) $X_n \to X$ a.e. (or a.s.) if $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ (pointwise) for $\omega \in \Omega \setminus E$ where $P(E) = 0$.

(b) $X_n \to X$ in probability if for any $\epsilon > 0$, $\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$.

(c) $X_n \to X$ in distribution if $F_n(x) \to F(x)$ at every continuity point x of F. It is equivalent to $\mu_n \to \mu$ vaguely i.e., $\mu_n(f) \to \mu(f)$ for all $f \in C_0(\mathbb{R})$, the space of continuous functions vanish at ∞ (detail in [1]).

The following relationships are basic ([1] or Royden): $(a) \Rightarrow (b) \Rightarrow (c)$; $(b) \Rightarrow (a)$ on some subsequence. On the other hand we cannot expect (c) to imply (b) as the distribution does not determine X. For example consider the interval $[0,1]$ with the Lebesgue measure, the r.v.'s $X_1 = \chi_{[0,\frac{1}{2}]}$, $X_2 = \chi_{[\frac{1}{2},1]}$, $X_3 = \chi_{[0,\frac{1}{4}]} + \chi_{[\frac{3}{4},1]}$ all have the same distribution.

The expectation of a random variable is defined as

$$E(X) = \int_{\Omega} X(\omega)dP(\omega) = \int_{-\infty}^{\infty} xdF(x) \ (= \int_{-\infty}^{\infty} xd\mu(x))$$

and for a Borel measurable h, we have

$$E(h(X)) = \int_{\Omega} h(X(\omega))dP(\omega) = \int_{-\infty}^{\infty} h(x)dF(x).$$

The most basic convergence theorems are:

(a) Fatou lemma:

$$X_n \geq 0, \quad \text{then} \ \ E(\underline{\lim}_{n\to\infty} X_n) \ \leq \ \underline{\lim}_{n\to\infty} E(X_n).$$

(b) Monotone convergence theorem:

$$X_n \geq 0, \ X_n \nearrow X, \quad \text{then} \ \ \lim_{n\to\infty} E(X_n) = E(X).$$

(c) Dominated convergence theorem:

$$|X_n| \leq Y,\ E(Y) < \infty \ \text{ and } \ X_n \to X \ a.e., \ \text{ then } \lim_{n\to\infty} E(X_n) = E(X).$$

We say that $X_n \to X$ in $L^p, p > 0$ if $E(|X|^p) < \infty$ and $E(|X_n - X|^p) \to 0$ as $n \to \infty$. It is known that L^p convergence implies convergence in probability. The converse also holds if we assume further $E(|X_n|^p) \to E(|X|^p) < \infty$ ([1], p.97).

Two events $A, B \in \mathcal{F}$ are said to be *independent* if

$$P(A \cap B) = P(A)P(B).$$

Similarly we say that the events $A_1, \cdots A_n \in \mathcal{F}$ are independent if for any subsets $A_{j_1}, \cdots, A_{j_k}$,

$$P(\bigcap_{i=1}^{k} A_{j_i}) = \prod_{i=1}^{k} P(A_{j_i}).$$

Two sub-σ-fields $\mathcal{F}_1$ and $\mathcal{F}_2$ are said to be independent if any choice of two sets from each of these σ-fields are independent. Two r.v.'s X, Y are independent if the σ-fields $\mathcal{F}_X$ and $\mathcal{F}_Y$ they generated are independent. Equivalently we have

$$P(X \leq x,\ Y \leq y) = P(X \leq x)\ P(Y \leq y),$$

(i.e., the joint distribution equals the product of their marginal distributions). We say that $X_1 \cdots X_n$ are independent if for any $X_{i_1} \cdots X_{i_k}$, their joint distribution is a product of their marginal distributions.

Proposition 1.1.2. *Let X, Y be independent, then $f(X)$ and $g(Y)$ are independent for any Borel measurable functions f and g.*

Exercises

1. Can you identify the interval $[0,1]$ with the Lebesgue measure to the probability space for tossing a fair coin repeatedly?

2. Prove Proposition 1.1.2.

3. Suppose that $\sup_n |X_n| \leq Y$ and $E(Y) < \infty$. Show that

$$E(\overline{\lim}_{n\to\infty} X_n) \ \geq\ \overline{\lim}_{n\to\infty} E(X_n)$$

4. If $p > 0$ and $E(|X|^p) < \infty$, then $\lim_{n\to\infty} x^p P(|X| > x) = 0$. Conversely, if $\lim_{n\to\infty} x^p P(|X| > x) = 0$, then $E(|X|^{p-\epsilon}) < \infty$ for $0 < \epsilon < p$.

5. For any d.f. F and any $a \geq 0$, we have

$$\int_{-\infty}^{\infty} (F(x+a) - F(x))dx = a$$

6. Let X be a positive r.v. with a distribution F, then

$$\int_0^\infty (1 - F(x))\ dx = \int_0^\infty x\ dF(x).$$

and

$$E(X) = \int_0^\infty P(X > x)\ dx = \int_0^\infty P(X \geq x)\ dx$$

7. Let $\{X_n\}$ be a sequence of identically distributed r.v. with finite mean, then

$$\lim_n \frac{1}{n} E(\max_{1\leq j\leq n} |X_j|) = 0.$$

(Hint: use Ex.6 to express the mean of the maximum)

8. If X_1, X_2 are independent r.v.'s each takes values $+1$ and -1 with probability $\frac{1}{2}$, then the three r.v.'s $\{X_1, X_2, X_1X_2\}$ are pairwise independent but not independent.

9. A r.v. is independent of itself if and only if it is constant with probability one. Can X and $f(X)$ be independent when $f \in \mathcal{B}$?

10 . Let $\{X_j\}_{j=1}^n$ be independent with distributions $\{F_j\}_{j=1}^n$. Find the distribution for $\max_j X_j$ and $\min_j X_j$.

11. If X and Y are independent and $E(|X+Y|^p) < \infty$ for some $p > 0$, then $E(|X|^p) < \infty$ and $E(|Y|^p) < \infty$.

12. If X and Y are independent, $E(|X|^p) < \infty$ for some $p \geq 1$, and $E(Y) = 0$, then $E(|X+Y|^p) \geq E(|X|^p)$.

1.2 Conditional Expectation

Let $\Lambda \in \mathcal{F}$ with $P(\Lambda) > 0$, we define

$$P(E|\Lambda) = \frac{P(\Lambda \cap E)}{P(\Lambda)} \qquad \text{where} \;\; P(\Lambda) > 0.$$

It follow that for a discrete random vector (X, Y),

$$P(Y = y|X = x) = \begin{cases} \dfrac{P(Y = y,\; X = x)}{P(X = x)} \;, & \text{if} \;\; P(X = x) > 0 \;, \\ 0 \;, & \text{otherwise.} \end{cases}$$

Moreover if (X, Y) is a continuous random variable with joint density $f(x, y)$, the conditional density of Y given $X = x$ is

$$f(y|x) = \begin{cases} \dfrac{f(x, y)}{f_X(x)} \;, & \text{if} \;\; f_X(x) > 0 \;, \\ 0 \;, & \text{otherwise} \;. \end{cases}$$

where $f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy$ is the marginal density. The conditional expectation of Y given $X = x$ is

$$E(Y|X = x) = \int_{-\infty}^{\infty} y f(y|x)dy.$$

Note that

$$g(x) := E(Y|X = x) \quad \text{is a function on} \;\; x \;,$$

and hence

$$g(X(\cdot)) := E(Y|X(\cdot)) \quad \text{is a r.v. on} \;\; \Omega \;. \tag{1.2.1}$$

In the following we have a more general consideration for the conditional expectation (and also the conditional probability): $E(Y|\mathcal{G})$ where $\mathcal{G}$ is a sub-σ-field of $\mathcal{F}$.

First let us look at a special case where $\mathcal{G}$ is generated by a measurable partition $\{\Lambda_n\}_n$ of Ω (each member in $\mathcal{G}$ is a union of $\{\Lambda_n\}_n$). Let Y be an integrable r.v., then

$$E(Y|\Lambda_n) = \int_\Omega Y(\omega)dP_{\Lambda_n}(\omega) = \frac{1}{P(\Lambda_n)}\int_{\Lambda_n} Y(\omega)dP(\omega). \qquad (1.2.2)$$

(Here $P_{\Lambda_n}(\cdot) = \frac{P(\,\cdot\,\cap\,\Lambda_n)}{P(\Lambda_n)}$ is a probability measure for $P(\Lambda_n) > 0$). Consider the random variable (as in (1.2.1))

$$Z(\cdot) = E(Y|\mathcal{G})(\cdot) := \sum_n E(Y|\Lambda_n)\chi_{\Lambda_n}(\cdot) \in \mathcal{G}.$$

It is easy to see that if $\omega \in \Lambda_n$, then $Z(\omega) = E(Y|\Lambda_n)$, and moreover

$$\int_\Omega E(Y|\mathcal{G})dP = \sum_n \int_{\Lambda_n} E(Y|\mathcal{G})dP = \sum_n E(Y|\Lambda_n)P(\Lambda_n) = \int_\Omega YdP\ .$$

We can also replace Ω by $\Lambda \in \mathcal{G}$ and obtain

$$\int_\Lambda E(Y|\mathcal{G})dP = \int_\Lambda YdP \qquad \forall\ \Lambda \in \mathcal{G}.$$

Recall that for μ, ν two σ-finite measures on $(\Omega, \mathcal{F})$ and $\mu \geq 0$, ν is called *absolutely continuous* with respect to μ $(\nu \ll \mu)$ if for any $\Lambda \in \mathcal{F}$ and $\mu(\Lambda) = 0$, then $\nu(\Lambda) = 0$. The Radon-Nikodym theorem says that there exists $g = \dfrac{d\nu}{d\mu}$ such that

$$\nu(\Lambda) = \int_\Lambda g d\mu \qquad \forall\ \ \Lambda \in \mathcal{F}.$$

Theorem 1.2.1. *If $E(|Y|) < \infty$ and $\mathcal{G}$ is a sub-σ-field of $\mathcal{F}$, t hen there exists a unique $\mathcal{G}$-measurable r.v., denote by $E(Y|\mathcal{G}) \in \mathcal{G}$, such that*

$$\int_\Lambda YdP = \int_\Lambda E(Y|\mathcal{G})\ dP \qquad \forall\ \Lambda \in \mathcal{G}.$$

Proof. Consider the set-valued function

$$\nu(\Lambda) = \int_\Lambda YdP \qquad \Lambda \in \mathcal{G}.$$

Then ν is a "signed measure" on $\mathcal{G}$. It satisfies

$$P(\Lambda) = 0 \implies \nu(\Lambda) = 0.$$

Hence ν is absolutely continuous with respect to P. By the Radon-Nikodym theorem, the derivative $g = \frac{d\nu}{dP} \in \mathcal{G}$ and

$$\int_\Lambda Y dP = v(\Lambda) = \int_\Lambda g dP \qquad \forall \ \Lambda \in \mathcal{G}.$$

This g is unique: for if we have $g_1 \in \mathcal{G}$ satisfies the same identity,

$$\int_\Lambda Y dP = v(\Lambda) = \int_\Lambda g_1 dP \qquad \forall \ \Lambda \in \mathcal{G}.$$

Let $\Lambda = \{g > g_1\} \in \mathcal{G}$, then $\int_\Lambda (g - g_1) dP = 0$ implies that $P(\Lambda) = 0$. We can reverse g and g_1 and hence we have $P(g \neq g_1) = 0$. It follows that $g = g_1$ $\mathcal{G}$-a.e.

Definition 1.2.2. *Given an integrable r.v. Y and a sub-σ-field $\mathcal{G}$, we say that $E(Y|\mathcal{G})$ is the* conditional expectation *of Y with respect to $\mathcal{G}$ (also denote by $E_{\mathcal{G}}(Y)$) if it satisfies*

(a) $E(Y|\mathcal{G}) \in \mathcal{G}$;

(b) $\int_\Lambda Y dP = \int_\Lambda E(Y|\mathcal{G}) dP \quad \forall \ \Lambda \in \mathcal{G}$.

If $Y = \chi_\Delta \in \mathcal{F}$, we define $P(\Delta|\mathcal{G}) = E(\chi_\Delta|\mathcal{G})$ and call this the conditional probability with respect to $\mathcal{G}$.

Note that the *conditional probability* can be put in the following way:

(a)' $\quad P(\Delta|\mathcal{G}) \in \mathcal{G}$;

(b)' $\quad P(\Delta \cap \Lambda) = \int_\Lambda P(\Delta|\mathcal{G}) dP \quad \forall \ \ \Lambda \in \mathcal{G}$.

It is a simple exercise to show that the original definition of $P(\Delta|\Lambda)$ agrees with this new definition by taking $\mathcal{G} = \{\emptyset, \ \Lambda, \ \Lambda^c, \ \Omega\}$.

Note that $E(Y|\mathcal{G})$ is "almost everywhere" defined, and we call one such function as a "version" of the conditional expectation. For brevity we will not mention the "a.e." in the conditional expectation unless necessary. If $\mathcal{G}$ is the sub-σ-field $\mathcal{F}_X$ generated by a r.v. X, we write $E(Y|X)$ instead of $E(Y|\mathcal{F}_X)$. Similarly we can define $E(Y|X_1, \cdots, X_n)$.

Proposition 1.2.3. *For $E(Y|X) \in \mathcal{F}_X$, there exists an extended-valued Borel measurable φ such that $E(Y|X) = \varphi(X)$, and φ is given by*

$$\varphi = \frac{d\lambda}{d\mu} \ ,$$

where $\lambda(B) = \int_{X^{-1}(B)} Y dP, \quad B \in \mathscr{B}$, and μ is the associated probability of the r.v. X on $\mathbb{R}$.

Proof. Since $E(Y|X) \in \mathcal{F}_X$, we can write $E(Y|X) = \varphi(X)$ for some Borel measurable φ (see §1). For $\Lambda \in \mathcal{F}$, there exists $B \in \mathcal{B}$ such that $\Lambda = X^{-1}(B)$. Hence

$$\int_\Lambda E(Y|X)dP = \int_\Omega \chi_B(X)\varphi(X)dP = \int_{\mathbb{R}} \chi_B(X)\varphi(X)d\mu = \int_B \varphi(x)d\mu$$

On the other hand by the definition of conditional probability,

$$\int_\Lambda E(Y|X)dP = \int_{X^{-1}(B)} Y dP = \lambda(B).$$

It follows that $\lambda(B) = \int_B \varphi(x)d\mu$ for all $B \in \mathcal{B}$. Hence $\varphi = \dfrac{d\lambda}{d\mu}$. $\square$

The following are some simple facts of the conditional expectation:

- If $\mathcal{G} = \{\phi, \Omega\}$, then $E(Y|\mathcal{G})$ is a constant function and equals $E(Y)$.

- If $\mathcal{G} = \{\phi, \Lambda, \Lambda^c, \Omega\}$, then $E(Y|\mathcal{G})$ is a simple function which equals $E(Y|\Lambda)$ on Λ, and equals $E(Y|\Lambda^c)$ on Λ^c,

– If $\mathcal{G} = \mathcal{F}$ or $Y \in \mathcal{G}$, then $E(Y|\mathcal{G}) = Y$.

– If (X, Y) has a joint density function, then $E(Y|X)$ coincides with the expression in (1.2.1).

Using the defining relationship of conditional expectation, we can show that the linearity, the basic inequalities and the convergence theorems for $E(\cdot)$ also hold for $E(\cdot\ |\mathcal{G})$. For example we have

Proposition 1.2.4. *(Jensen inequality) If φ is a convex function on $\mathbb{R}$, and Y and $\varphi(Y)$ are integrable r.v., then for each sub-σ-algebra $\mathcal{G}$,*

$$\varphi\big(E(Y|\mathcal{G})\big) \leq E\big(\varphi(Y)|\mathcal{G}\big)$$

Proof. If Y is a simple r.v., then $Y = \sum_{j=1}^{n} y_j \chi_{\Lambda_j}$ with $\Lambda \in \mathcal{F}$. It follows that

$$E(Y|\mathcal{G}) = \sum_{j=1}^{n} y_j E(\chi_{\Lambda_j}|\mathcal{G}) = \sum_{j=1}^{n} y_j P(Y_{\Lambda_j}|\mathcal{G})$$

and

$$E(\varphi(Y)|\mathcal{G}) = \sum_{j=1}^{n} \varphi(y_j) P(Y_{\Lambda_j}|\mathcal{G}).$$

Since $\sum_{j=1}^{n} P(\Lambda_j|\mathcal{G}) = 1$, the inequality holds by the convexity of φ.

In general we can find a sequence of simple r.v. $\{Y_m\}$ with $|Y_m| \leq |Y|$ and $Y_m \to Y$, then apply the above together with the dominated convergence theorem. $\square$

Proposition 1.2.5. *Let Y and YZ be integrable r.v. and $Z \in \mathcal{G}$, then we have*

$$E(YZ|\mathcal{G}) = ZE(Y|\mathcal{G}).$$

Proof. It suffices to show that for $Y, Z \geq 0$

$$\int_\Lambda ZE(Y|\mathcal{G})dP = \int_\Lambda ZY dP \qquad \forall\, \Lambda \in \mathcal{G}.$$

Obviously, this is true for $Z = \chi_\Delta,\ \Delta \in \mathcal{G}$. We can pass it to the simple r.v. Then use the monotone convergence theorem to show that it hold for all $Z \geq 0$, and then the general integrable r.v. □

Proposition 1.2.6. *Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be sub-σ-fields of $\mathcal{F}$ and $\mathcal{G}_1 \subseteq \mathcal{G}_2$. Then for Y integrable r.v.*

$$E(E(Y|\mathcal{G}_2)|\mathcal{G}_1) = E(Y|\mathcal{G}_1) = E(E(Y|\mathcal{G}_1)|\mathcal{G}_2). \tag{1.2.3}$$

Moreover

$$E(Y|\mathcal{G}_1) = E(Y|\mathcal{G}_2) \quad \textit{iff} \quad E(Y|\mathcal{G}_2) \in \mathcal{G}_1. \tag{1.2.4}$$

Proof. Let $\Lambda \in \mathcal{G}_1$, then $\Lambda \in \mathcal{G}_2$. Hence

$$\int_\Lambda E(E(X|\mathcal{G}_2)|\mathcal{G}_1)dP = \int_\Lambda E(Y|\mathcal{G}_2)dP = \int_\Lambda Y dP = \int_\Lambda E(Y|\mathcal{G}_1)dP,$$

and the first identity in (1.2.3) follows. The second identity is by $E(Y|\mathcal{G}_1) \in \mathcal{G}_2$ (recall that $Z \in \mathcal{G}$ implies $E(Z|\mathcal{G}) = Z$).

For the last part, the necessity is trivial, and the sufficiency follows from the first identity. □

As a simple consequence, we have

Corollary 1.2.7. $E(E(Y|X_1, X_2)|X_1) = E(Y|X_1) = E(E(Y|X_1)|X_1, X_2)$.

Exercises

1. (Bayes' rule) Let $\{\Lambda_n\}$ be a $\mathcal{F}$-measurable partition of Ω and let $E \in \mathcal{F}$ with $P(E) > 0$. Then

$$P(\Lambda_n|E) = \frac{P(\Lambda_n)\ P(E|\Lambda_n)}{\sum_n P(\Lambda_n)P(E|\Lambda_n)}\ .$$

2. If the random vector (X, Y) has probability density $p(x, y)$ and X is integrable, then one version of $E(X|X + Y = z)$ is given by

$$\int xp(x, z - x)dx\ /\ \int p(x, z - x)dx\ .$$

3. Let X be a r.v. such that $P(X > t) = e^{-t},\ t > 0$. Compute $E(X|X \vee t)$ and $E(X|X \wedge t)$ for $t > 0$. (Here $\vee$ and $\wedge$ mean maximum and minimum respectively.

4. If X is an integrable r.v., Y is a bounded r.v., and $\mathcal{G}$ is a sub-σ-field, then

$$E\big(E(X|\mathcal{G})Y\big) = E\big(XE(Y|\mathcal{G})\big).$$

5. Prove that $\mathrm{var}(E(Y|\mathcal{G})) \le \mathrm{var}(Y)$.

6. Let X, Y be two r.v., and let $\mathcal{G}$ be a sub-σ-field. Suppose

$$E(Y^2|\mathcal{G}) = X^2, \quad E(Y|\mathcal{G}) = X,$$

then $Y = X$ a.e.

7. Give an example that $E(E(Y|X_1)|X_2) \neq E(E(Y|X_2)|X_1)$. (*Hint: it suffices to find an example* $E(X|Y) \neq E(E(X|Y)|X)$ *for* Ω *to have three points*).

1.3 Markov Property

Let A be an index set and let $\{\mathcal{F}_\alpha : \alpha \in A\}$ be family of sub-σ-fields of $\mathcal{F}$. We say that the family of $\mathcal{F}_\alpha$'s are *conditionally independent* relative to $\mathcal{G}$ if for any $\Lambda_i \in \mathcal{F}_{\alpha_i}$ $i = 1, \cdots, n$,

$$P(\bigcap_{j=1}^{n} \Lambda_j|\mathcal{G}) \ = \ \prod_{j=1}^{n} P(\Lambda_i|\mathcal{G}). \tag{1.3.1}$$

Proposition 1.3.1. *For $\alpha \in A$, let $\mathcal{F}^{(\alpha)}$ denote the sub-σ-field generated by $\mathcal{F}_\beta$, $\beta \in A\backslash\{\alpha\}$. Then the family $\{\mathcal{F}_\alpha\}_\alpha$ are conditionally independent relative to $\mathcal{G}$ if and only if*

$$P(\Lambda|\ \mathcal{F}^{(\alpha)} \vee \mathcal{G}) = P(\Lambda|\ \mathcal{G}), \quad \Lambda \in \mathcal{F}_\alpha$$

where $\mathcal{F}^{(\alpha)} \vee \mathcal{G}$ is the sub-σ-field generated by $\mathcal{F}^{(\alpha)}$ and $\mathcal{G}$.

Proof. We only prove the case $A = \{1, 2\}$, i.e.,

$$P(\Lambda|\ \mathcal{F}_2 \vee \mathcal{G}) = P(\Lambda|\ \mathcal{G}), \quad \Lambda \in \mathcal{F}_1. \tag{1.3.2}$$

The general case follows from the same argument. To prove the sufficiency, we assume (1.3.2). To check (1.3.1), let $\Lambda \in \mathcal{F}_1$, then for $M \in \mathcal{F}_2$,

$$\begin{aligned} P(\Lambda \cap M\ |\mathcal{G}) &= E(P(\Lambda \cap M\ |\mathcal{F}_2 \vee \mathcal{G})|\mathcal{G}) \\ &= E(P(\Lambda|\ \mathcal{F}_2 \vee \mathcal{G})\chi_M|\mathcal{G}) \\ &= E(P(\Lambda|\ \mathcal{G})\chi_M|\mathcal{G}) \qquad \text{(by (1.3.2))} \\ &= P(\Lambda|\ \mathcal{G})P(M|\ \mathcal{G}). \end{aligned}$$

Hence $\mathcal{F}_1$ and $\mathcal{F}_2$ are $\mathcal{G}$-independent.

To prove the necessity, suppose (1.3.1) holds, we claim that for $\Delta \in \mathcal{G}$, $\Lambda \in \mathcal{F}_1$ and $M \in \mathcal{F}_2$,

$$\int_{M\cap\Delta} P(\Lambda|\mathcal{G})\ dP = \int_{M\cap\Delta} P(\Lambda|\ \mathcal{F}_2 \vee \mathcal{G})\ dP$$

Since the sets of the form $M \cap \Delta$ generate $\mathcal{G} \vee \mathcal{F}_2$, we have $P(\Lambda|\mathcal{G}) = P(\Lambda|\ \mathcal{F}_2 \vee \mathcal{G})$. i.e., (1.3.2) holds.

The claim follows from the following: let $\Lambda \in \mathcal{F}_1$, $M \in \mathcal{F}_2$, then

$$\begin{aligned} E(P(\Lambda|\mathcal{G})\chi_M|\mathcal{G}) &= P(\Lambda|\mathcal{G})P(M|\mathcal{G}) \\ &= P(\Lambda \cap M\ |\mathcal{G}) \qquad \text{(by (1.3.1))} \\ &= E(P(\Lambda|\mathcal{F}_2 \vee\ \mathcal{G})\chi_M|\mathcal{G}) \qquad \Box \end{aligned}$$

Corollary 1.3.2. *Let $\{X_\alpha\}_{\alpha \in A}$ be a family of r.v. and let $\mathcal{F}_\alpha$ be the sub-σ-field generated by X_α. Then the X_α's are independent if and only if for any Borel set B,*

$$P(X_\alpha \in B|\mathcal{F}^{(\alpha)}) = P(X_\alpha \in B).$$

Moreover the above condition can be replaced by: for any integrable $Y \in \mathcal{F}_\alpha$,

$$E(Y|\mathcal{F}^{(\alpha)}) \ = \ E(Y).$$

Proof. The first identity follows from Proposition 1.3.1 by taking $\mathcal{G}$ as the trivial σ-field. The second one follows from an approximation by simple function and use the first identity. $\Box$

To consider the Markov property, we first consider an important basic case.

Theorem 1.3.3. *Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of independent r.v. and each X_n has a distribution μ_n on $\mathbb{R}$. Let $S_n = \sum_{j=1}^{n} X_j$. Then for $B \in \mathscr{B}$,*

$$P(S_n \in B \mid S_1, \cdots, S_{n-1}) = P(S_n \in B \mid S_{n-1}) = \mu_n(B - S_{n-1})$$

(Hence S_n is independent of $S_1, \cdots, S_{n-2}$ given S_{n-1}.)

Proof. We divide the proof into two steps.

Step 1. We show that

$$P(X_1 + X_2 \in B \mid X_1) = \mu_2(B - X_1)$$

First observe that $\mu_2(B - X_1)$ is in $\mathcal{F}_{X_1}$. Let $\Lambda \in \mathscr{F}_{X_1}$, then $\Lambda = X_1^{-1}(A)$ for some $A \in \mathscr{B}$, and

$$\begin{aligned}
\int_\Lambda \mu_2(B - X_1)\ dP &= \int_A \mu_2(B - x_1)\ d\mu_1(x_1) \\
&= \int_A \Big(\int_{x_1+x_2 \in B} d\mu_2(x_2) \Big)\ d\mu_1(x_1) \\
&= \iint_{x_1 \in A,\ x_1+x_2 \in B} d(\mu_1 \times \mu_2)(x_1, x_2) \\
&= P(X_1 \in A,\ X_1 + X_2 \in B) \\
&= \int_\Lambda P(X_1 + X_2 \in B \mid \mathscr{F}_{X_1})\ dP
\end{aligned}$$

This implies that $\mu_2(B - X_1) = P(X_1 + X_2 \in B \mid X_1)$.

Step 2. The second equality in the proposition follows from Step 1 by applying to S_{n-1} and X_n. To prove the first identity, we let $\mu^n = \mu_1 \times \cdots \times \mu_n = \mu^{n-1} \times \mu_n$. Let $B_j \in \mathcal{B}$, $1 \le j \le n-1$, and let $\Lambda = \bigcap_{j=1}^{n-1} S_j^{-1}(B_j) \in \mathcal{F}(S_1, \cdots, S_{n-1})$. We show as in Step 1,

$$\int_\Lambda \mu_n(B - S_{n-1})\ dP = \int_\Lambda P(S_n \in B | S_1, \cdots, S_{n-1})\ dP$$

and the identity $\mu_n(B - S_{n-1}) = P(S_n \in B | S_1, \cdots, S_{n-1})$ follows. $\square$

Definition 1.3.4. *We call a sequence of random variables $\{X_n\}_{n=0}^\infty$ a (discrete time) stochastic process. It is called a* Markov process *(*Markov chain *if the state space is countable or finite) if for any n and $B \in \mathcal{B}$,*

$$P(X_{n+1} \in B | X_0, \cdots, X_n) = P(X_{n+1} \in B | X_n).$$

Let $I \subseteq \mathbb{N}_0 := \mathbb{N} \cup \{0\}$ and let $\mathcal{F}_I$ denote the sub-σ-field generated by $\mathcal{F}_n,\ n \in I$. Typically, $I = \{n\}$, or $[0, n]$, or (n, ∞); $\mathcal{F}_{\{n\}}$ denotes the events at the present, $\mathcal{F}_{[0,n]}$ denotes the events from the past up to the present, and $\mathcal{F}_{(n,\infty)}$ denotes the events in the future. The above Markov property means the future depends on the present and is independent of the past.

One of the most important examples of Markov process is the sequence $\{S_n\}_{n=0}^{\infty}$ in Theorem 1.2.3.

Theorem 1.3.5. *Let $\{X_n\}_{n=0}^{\infty}$ be a stochastic process, then the following are equivalent:*

(a) $\{X_n\}_{n=0}^{\infty}$ has the Markov property;

(b) $P(M|\mathcal{F}_{[0,n]}) = P(M|X_n)$ for all $n \in \mathbb{N}$ and $M \in \mathcal{F}_{(n,\infty)}$;

(c) $P(M_1 \cap M_2\ |X_n) = P(M_1|X_n)\ P(M_2|X_n)$ for all $M_1 \in \mathcal{F}_{[0,n]},\ M_2 \in \mathcal{F}_{(n,\infty)}$ and $n \in \mathbb{N}$.

The conditions remain true if $\mathcal{F}_{(n,\infty)}$ is replaced by $\mathcal{F}_{[n,\infty)}$ (Exercise). Condition (c) can be interpreted as conditioning on the present, the past and the future are independent.

Proof. $(b) \Rightarrow (c)$. Let $Y_i = \chi_{M_i}$ with $M_1 \in \mathcal{F}_{[0,n]},\ M_2 \in \mathcal{F}_{(n,\infty)}$, then

$$\begin{aligned} P(M_1|X_n)\ P(M_2|X_n) &= E(Y_1|X_n)\ E(Y_2|X_n) = E(Y_1 E(Y_2|X_n)|X_n) \\ &= E(Y_1 E(Y_2|\mathcal{F}_{[0,n]})|X_n) = E(E(Y_1 Y_2|\mathcal{F}_{[0,n]})|X_n) \\ &= E(Y_1 Y_2|X_n) = P(M_1 \cap M_2\ |X_n). \end{aligned}$$

$(c) \Rightarrow (b)$. Let $\Lambda \in \mathcal{F}_{[0,n]}$ be the test set, and let $Y_1 = \chi_\Lambda,\ Y_2 = \chi_M \in$

$\mathcal{F}_{(0,\infty)}$. Then

$$\begin{aligned}\int_\Lambda P(M|X_n)\,dP &= E\big(Y_1E(Y_2|X_n)\big) = E\big(E(Y_1E(Y_2|X_n))|X_n\big)\\ &= E\big(E(Y_1|X_n)E(Y_2|X_n)\big) = E\big(E(Y_1Y_2|X_n)\big)\\ &= \int_\Omega P(\Lambda\cap M|X_n)\,dP = P(\Lambda\cap M).\end{aligned}$$

This implies $P(M|X_n) = P(M|\mathcal{F}_{[0,n]})$.

$(b) \Rightarrow (a)$ is trivial.

$(a) \Rightarrow (b)$. We claim that for each n,

$$E(Y|\mathcal{F}_{[0,n]}) = E(Y|\ X_n) \qquad \forall\ Y \in \mathcal{F}_{[n+1,n+k]},\quad k = 1, 2, \cdots . \tag{1.3.3}$$

This will establish (b) for $M \in \bigcup_{k=1}^\infty \mathcal{F}_{(n,n+k)}$; this family of M generates $\mathcal{F}_{(0,\infty)}$.

Note that the Markov property implies (1.3.3) is true for $k = 1$. Suppose the statement is true for k, we consider $Y = Y_1Y_2 \in \mathcal{F}_{[n+1,n+k+1]}$, where $Y_1 \in \mathcal{F}_{[n+1,n+k]}$ and $Y_2 \in \mathcal{F}_{n+k+1}$. Then

$$\begin{aligned}E(Y|\mathcal{F}_{[0,n]}) &= E(\ E(Y|\mathcal{F}_{[0,n+k]})\ |\ \mathcal{F}_{[0,n]})\\ &= E(\ Y_1E(Y_2|\mathcal{F}_{[0,n+k]})\ |\ \mathcal{F}_{[0,n]})\\ &= E(\ Y_1E(Y_2|\mathcal{F}_{n+k})\ |\ \mathcal{F}_{[0,n]}) \qquad \text{(by Markov)}\\ &= E(\ Y_1E(Y_2|\mathcal{F}_{n+k})\ |\ \mathcal{F}_{n}) \qquad \text{(by induction)}\\ &= E(\ Y_1E(Y_2|\mathcal{F}_{[n,n+k]})\ |\ \mathcal{F}_{[0,n]}) \qquad \text{(by Markov)}\\ &= E(\ E(Y_1Y_2|\mathcal{F}_{[n,n+k]})\ |\ \mathcal{F}_{[0,n]})\\ &= E(Y_1Y_2|\mathcal{F}_n)\\ &= E(Y|\mathcal{F}_n).\end{aligned}$$

This implies the inductive step for $Y = \chi_{M_1\cap M_2} = \chi_{M_1}\chi_{M_2}$ with $M_1 \in \mathcal{F}_{[n+1,n+k]}$ and $M_2 \in \mathcal{F}_{n+k+1}$. But the class of all such Y generates $\mathcal{F}_{[n+1,n+k]}$. This implies the claim and completes the proof of the theorem. $\square$

The following random variable plays a central role in stochastic process.

Definition 1.3.6. *A r.v.* $\alpha : \Omega \to \mathbb{N}_0 \cup \{\infty\}$ *is called a* stopping time *(or Markov time or optional r.v.) with respect to* $\{X_n\}_{n=0}^{\infty}$ *if*

$$\{\omega : \alpha(\omega) = n\} \in \mathcal{F}_{[0,n]} \qquad \textit{for each} \quad n \in \mathbb{N}_0 \cup \{\infty\}.$$

It is easy to see the definition can be replaced by $\{\omega : \alpha(\omega) \leq n\} \in \mathcal{F}_{[0,n]}$. In practice, the most important example is: for a given $A \in \mathcal{B}$, let

$$\alpha_A(\omega) = \min\{n \geq 0 : \ X_n(\omega) \in A\}.$$

($\alpha_A(\omega) = \infty$ if $X_n(\omega) \notin A$ for all n.) This is the r.v. of the first time the process $\{X_n\}_{n=0}^{\infty}$ enters A. It is clear that

$$\{\omega : \ \alpha_A(\omega) = n\} = \bigcap_{j=0}^{n-1} \{\omega : \ X_j(\omega) \in A^c, \ X_n(\omega) \in A\} \in \mathcal{F}_{[0,n]},$$

and similarly for $n = \infty$. Hence α_A is a stopping time.

Very often α represents the random time that a specific event happens, and $\{X_{\alpha+n}\}_{n=1}^{\infty}$ is the process after the event has occurred. We will use the following terminologies:

- The pre-α field $\mathcal{F}_\alpha$ is the sets $\Lambda \in \mathcal{F}_{[0,\infty)}$ of the form

$$\Lambda = \bigcup_{0 \leq n \leq \infty} \{\{\alpha = n\} \cap \Lambda_n\}, \qquad \Lambda_n \in \mathcal{F}_{[0,n]}. \tag{1.3.4}$$

  It follows that $\Lambda \in \mathcal{F}_\alpha$ if and only if $\{\alpha = n\} \cap \Lambda \in \mathcal{F}_n$ for each n.

- The post α-process is $\{X_{\alpha+n}\}_{n=1}^{\infty}$ where $X_{\alpha+n}(\omega) = X_{\alpha(\omega)+n}(\omega)$. The post-$\alpha$ field $\mathcal{F}'_\alpha$ is the sub-σ-field generated by the post-α process.

Proposition 1.3.7. *Let $\{X_n\}_{n=0}^{\infty}$ be a stochastic process and let α be a stopping time. Then $\alpha \in \mathcal{F}_\alpha$ and $X_\alpha \in \mathcal{F}_\alpha$.*

Proof. For α to be $\mathcal{F}_\alpha$-measurable, we need to show that $\{\alpha = k\} \in \mathcal{F}_\alpha$. This follows from (1.3.4) by taking $\Lambda_n = \emptyset$ for $n \neq k$ and $\Lambda_k = \Omega$.

That $X_\alpha \in \mathcal{F}_\alpha$ follows from

$$\{\omega : \ X_\alpha(\omega) \in B\} = \bigcup_n \{\omega : \alpha(\omega) = n, \ X_n(\omega) \in B\} \in \mathcal{F}_\alpha$$

for any Borel set $B \in \mathcal{B}$. $\square$

Theorem 1.3.8. *Let $\{X_n\}_{n=0}^{\infty}$ be a Markov-process and α is an a.e. finite stopping time, then for each $M \in \mathcal{F}'_\alpha$,*

$$P(M|\mathcal{F}_\alpha) = P(M| \ \alpha, X_\alpha). \tag{1.3.5}$$

We call this property the *strong* Markov-property.

Proof. Note that the generating sets of $\mathcal{F}'_\alpha$ are $M = \bigcap_{j=1}^{l} X_{\alpha+j}^{-1}(B_j), \ B_j \in \mathcal{B}$. Let $M_n = \bigcap_{j=1}^{l} X_{n+j}^{-1}(B_j) \in \mathcal{F}_{(n,\infty)}$, We claim that

$$P(M| \ \alpha, X_\alpha) = \sum_{n=1}^{\infty} P(M_n|X_n)\chi_{\{\alpha=n\}}. \tag{1.3.6}$$

Indeed if we consider $P(M_n|X_n) = \varphi_n(X_n)$, then it is clear $\sum_{n=1}^{\infty} \varphi_n(X_n)\chi_{\{\alpha=n\}}$ is measurable with respect to the σ-field generated by α and X_α. By making use of Theorem 1.3.5(b), we have

$$\begin{aligned}
\int_{\{\alpha=m, \ X_\alpha \in B\}} \sum_{n=1}^{\infty} P(M_n|X_n)\chi_{\{\alpha=n\}} \, dP &= \int_{\{\alpha=m, \ X_m \in B\}} P(M_m|X_m) \, dP \\
&= \int_{\{\alpha=m, \ X_m \in B\}} P(M_m|\mathcal{F}_{[0,m]}) \, dP \\
&= P(\{\alpha = m, \ X_m \in B\} \cap M_m) \\
&= P(\{\alpha = m, X_\alpha \in B\} \cap M).
\end{aligned}$$

(The last equality is due to $M_m \cap \{\alpha = m\} = M \cap \{\alpha = m\}$). Hence the claim follows.

Now to prove the theorem, let $\Lambda \in \mathcal{F}_\alpha,\ \Lambda = \bigcup_{n=0}^{\infty}(\{\alpha = n\} \cap \Lambda_n)$, then

$$\begin{aligned} P(\Lambda \cap M) &= \sum_{n=0}^{\infty} P(\{\alpha = n, \Lambda_n\} \cap M_n) \\ &= \sum_{n=0}^{\infty} \int_{\{\alpha=n\}\cap\Lambda_n} P(M_n|\mathcal{F}_{[0,n]})\ dP \\ &= \sum_{n=0}^{\infty} \int_{\Lambda} P(M_n|X_n)\chi_{\{\alpha=n\}}\ dP \quad \text{(by Theorem 1.3.5(b))} \\ &= \int_{\Lambda} P(M_n|\ \alpha, X_\alpha)\ dP \qquad \text{(by (1.3.6))}. \end{aligned}$$

The theorem follows from this. $\square$

We remark that when α is the constant n, then we can omit the α in (1.3.5) and it reduces to the Markov property as in Theorem 1.3.5. Also if the process is homogeneous (i.e., invariant on the time n), then we can omit the α there. It is because in (1.3.6), the right side can be represented as $\sum_{n=1}^{\infty} \varphi(X_n)\chi_{\{\alpha=n\}}$ (instead of $\varphi_n(X_n)$) which is $\mathcal{F}_\alpha$-measurable. In this case we can rewrite (1.3.5) as

$$P(X_{\alpha+1} \in B|\mathcal{F}_\alpha) = P(X_{\alpha+1} \in B|X_\alpha) \qquad \forall\ B \in \mathcal{B},$$

a direct analog of the definition of Markov property.

There is a constructive way to obtain Markov processes. For a Markov chain $\{X_n\}_{n=0}^{\infty}$, we mean a stochastic process that has a state space $S = \{a_1, a_2, \cdots, a_N\}$ (finite or countable) and a transition matrix

$$P = \begin{pmatrix} p_{11} & \cdots & p_{1N} \\ \vdots & \cdots & \vdots \\ p_{N1} & \cdots & p_{NN} \end{pmatrix}$$

where $p_{ij} \geq 0$ and the row sum is 1; the p_{ij} is the probability from i to j. Suppose the process starts at X_0 with initial distribution $\mu = (\mu_1, \cdots, \mu_N)$, let X_n denote the location of the chain at the n-th time according to the transition matrix P, then $\{X_n\}_{n=0}^{\infty}$ satisfies the Markov property:

$$P(X_{n+1} = x_{n+1}|X_0 = x_0, \cdots, X_n = x_n) = P(X_{n+1} = x_{n+1}|X_n = x_n) = p_{ij}.$$

Also it follows that

$$\begin{aligned} & P(X_0 = x_0,\ X_1 = x_1, \cdots,\ X_n = x_n) \\ = & P(X_0 = x_0)P(X_1 = x_1|X_0 = x_0) \cdots P(X_n = x_n|X_{n-1} = x_{n-1}) \\ = & \mu_{x_0} p_{x_0 x_1} \cdots p_{x_{n-1} x_n}. \end{aligned}$$

More generally, we consider the state space to be $\mathbb{R}$. Let $\mu : \mathbb{R} \times \mathcal{B} \to [0, 1]$ satisfies

(a) for each x, $\mu(x, \cdot)$ is a probability measure;

(b) for each B, $\mu(\cdot, B)$ is a Borel measurable function.

Let $\{X_n\}_{n=0}^{\infty}$ be a sequence of r.v. with finite dimensional joint distributions $\mu^{(n)}$ for $X_0, \cdots, X_n$ given by

$$\begin{aligned} P(\bigcap_{j=0}^{n} \{X_j \in B_j\}) &= \mu^{(n)}(B_0 \times \cdots \times B_n) \\ &:= \int \cdots \int_{B_0 \times \cdots \times B_n} \mu_0(dx_0)\mu(x_0, dx_1) \cdots \mu(x_{n-1}, dx_n). \end{aligned}$$

where μ_0 is the distribution function of X_0.

It is direct to check from definition that

$$P(X_{n+1} \in B|X_n) \ = \ \mu(X_n, B),$$

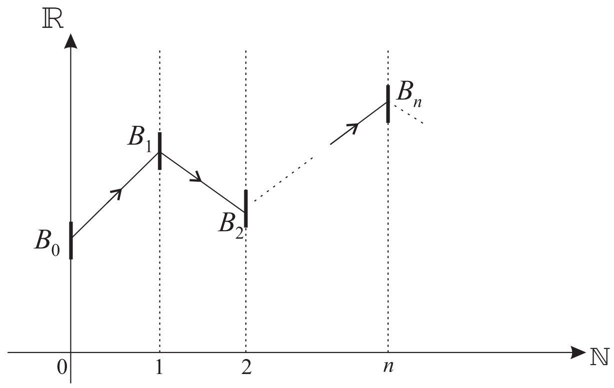

Figure 1.1:

i.e.,

$$P(X_{n+1} \in B | X_n = x) \; = \; \mu(x, B).$$

Hence $\mu(x, B)$ represents the probability that in the $(n+1)$-step the chain is in B, starting at x in the n-th step. To see that $\{X_n\}_{n=0}^{\infty}$ satisfies the Markov property, we let $\Lambda = \bigcap_{j=0}^{n} \{X_j \in B_j\}$, then

$$\begin{aligned}
\int_{\Lambda} P(X_{n+1} \in B | X_n) \, dP \; &= \; \int \cdots \int_{B_0 \times \cdots \times B_n} \mu(x_n, B) \, d\mu^{(n)}(x_0, \cdots, x_n) \\
&= \; \int \cdots \int_{B_0 \times \cdots B_n \times B} \mu_0(dx_0) \prod_{j=1}^{n+1} \mu(x_{j-1}, dx_j) \\
&= \; P(\Lambda \cap \{X_{n+1} \in B\}).
\end{aligned}$$

This implies

$$P(X_{n+1} \in B | X_n) = P(X_{n+1} \in B | X_1, \cdots, X_n)$$

and the Markov property follows.

We call the above $\{X_n\}_{n=0}^{\infty}$ a *stationary* (or *homogeneous*) Markov process and $\mu(x, B)$ the transition probability.

Exercises

1. Let $\{X_n\}_{n=0}^{\infty}$ be a Markov process. Let f be a one-to-one Borel measurable function on $\mathbb{R}$ and let $Y_n = f(X_n)$. Show that $\{Y_n\}_{n=0}^{\infty}$ is also a Markov process (with respect to the fields generated by $f(X_n)$); but the conclusion does not hold if we do not assume f is one-to-one.

2. Prove the strong Markov property in the form of Theorem 1.3.5(c).

3. If α_1 and α_2 are both stopping times, so are $\alpha_1 \wedge \alpha_2$, $\alpha_1 \vee \alpha_2$ and $\alpha_1 + \alpha_2$. However $\alpha_1 - \alpha_2$ is not necessary a stopping time.

4. Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of i.i.d.r.v. Let $\{\alpha_k\}_{k=1}^{\infty}$ be a sequence of strictly increasing finite stopping times. Then $\{X_{\alpha_k+1}\}_{k=1}^{\infty}$ is also a sequence of i.i.d.r.v. (This is the gambling-system theorem given by Doob).

5. A sequence $\{X_n\}_{n=0}^{\infty}$ is a Markov chain of *second order* if

$$P(X_{n+1} = j | X_0 = i_0, \cdots, X_n = i_n) = P(X_{n+1} = j | X_{n-1} = i_{n-1}, X_n = i_n).$$

Show that nothing really new is involved because the sequence (X_n, X_{n+1}) is a Markov chain.

6. Let $\mu^{(n)}(x, B)$ be the n-step transition probability in the stationary Markov process. Prove the Chapman-Kolmogorov equation

$$\mu^{(m+n)}(x, B) = \int_{\mathbb{R}} \mu^{(m)}(x, dy)\mu^{(n)}(y, B) \qquad \forall\ m, n \in \mathbb{N}.$$

1.4 Martingales

We first consider a simple example in analysis. Let f be an integrable function on $[0,1]$, let $\mathcal{P}_n = \{0 = \frac{1}{2^n} \leq \cdots \leq \frac{k}{2^n} \cdots \leq 1\}$ be a partition of $[0,1]$ and let $I_{n,k} = [\frac{k}{2^n}, \frac{k+1}{2^n})$. We define the average function f_n of f on the partition $\mathcal{P}_n$:

$$f_n(x) = \sum_{k=0}^{2^{n-1}} a_{n,k}\ \chi_{I_{n,k}}\ , \qquad x \in I_{n,k}. \tag{1.4.1}$$

where $a_{n,k} = \frac{1}{|I_{n,k}|}\int_{I_{n,k}} f(x)dx$. Then $\{f_n\}_n$ converges to f in L^1. Moreover $\{f_n\}_n$ has the following consistency property: for $m > n$

$$f_n(x) = \frac{1}{|I_{n,k}|}\int_{I_{n,k}} f_m(y)dy \qquad x \in I_{n,k}. \tag{1.4.2}$$

This property has been reformulated by Doob in the more general probability setting.

Definition 1.4.1. *Let $\{(X_n, \mathcal{F}_n)\}_{n=1}^{\infty}$ be a sequence of r.v. such that $X_n \in \mathcal{F}_n$. It is called a martingale if*

(a) $\mathcal{F}_n \subset \mathcal{F}_{n+1}$;

(b) $E(|X_n|) < \infty$;

(c) $X_n = E(X_{n+1}|\mathcal{F})$.

It is called a supermartingale (or submartingale) if $\geq$ (or $\leq$ respectively) in (c) holds. We will call $\{X_n\}_n$ a s-martingale if it is any one of the three cases.

Condition (c) can be strengthened as $X_n = E(X_m|\mathcal{F}_n)$ for $m > n$. It follows from

$$E(X_m|\mathcal{F}_n) = E(E(X_m|\mathcal{F}_{m-1})|\mathcal{F}_n) = E(X_{m-1}|\mathcal{F}_n) = \cdots = E(X_n|\mathcal{F}_n) = X_n\ .$$

Martingale has its intuitive background in gambling. If X_n is interpreted as the gambler's capital at time n, then the defining property says that his expected capital after next game, played with the knowledge of the entire past and present, is exactly equal to his current capital. In other words, his expected gain is zero, and is in this sense the game is said to be "fair". The supermartingale and submartingale can be interpreted similarly.

Example 1. As a direct analog of the above function case, we let X be an integrable r.v. and let $\{\mathcal{F}_n\}_{n=1}^{\infty}$ be an increasing sequence of sub-σ-fields (e.g., take $\mathcal{F}_n$ to be a partition). Let $X_n = E(X|\mathcal{F}_n)$. Then $\{X_n\}_{n=1}^{\infty}$ is a martingale. Indeed we see that

$$E(|X_n|) = E(|E(X|\mathcal{F}_n)|) \leq E(E(|X||\mathcal{F}_n)) = E(|X|) < \infty$$

and (b) follows. For (c), we observe that

$$E(X_{n+1}|\mathcal{F}_n) = E(E(X|\mathcal{F}_{n+1})|\mathcal{F}_n) = E(X|\mathcal{F}_n) = X_n.$$

Example 2. Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of independent integrable r.v. with mean zero. Let $S_n = \sum_{j=1}^{n} X_n$ and $\mathcal{F}_n = \mathcal{F}(X_1, \cdots, X_n)$. Then

$$\begin{aligned} E(S_{n+1}|\mathcal{F}_n) &= E(S_n + X_{n+1}|\mathcal{F}_n) \\ &= S_n + E(X_{n+1}|\mathcal{F}_n) \\ &= S_n + E(X_{n+1}) \\ &= S_n. \end{aligned}$$

Hence $\{(S_n, \mathcal{F}_n)\}$ is a martingale.

Proposition 1.4.2. *If $\{(X_n, \mathcal{F}_n)\}_{n=1}^{\infty}$ is a submartingale, and φ is increasing and convex in $\mathbb{R}$. If $\{\varphi(X_n)\}$ is integrable, then $\{(\varphi(X_n), \mathcal{F}_n)\}$ is also a submartingale.*

Proof. Since $X_n \leq E(X_{n+1}|\mathcal{F}_n)$, by the property of φ, we have

$$\varphi(X_n) \leq \varphi(E(X_{n+1}|\mathcal{F}_n)) \leq E(\varphi(X_{n+1})|\mathcal{F}_n) \qquad \Box$$

It follows that if $\{X_n\}_{n=0}^{\infty}$ is a martingale (or submartingale), then $\{|X_n|^p\}_{n=0}^{\infty}$, $p \geq 1$ (provided that $X_n \in L^p$) and $\{X_n^+\}_{n=0}^{\infty}$ are submartingales. Also if $\{X_n\}$ is a supermartingale, so does $\{X_n \wedge a\}_n$ for any $a \in \mathbb{R}$.

Theorem 1.4.3. *(Doob's decomposition Theorem) For any submartingale $\{(X_n, \mathcal{F}_n)\}_{n=1}^{\infty}$, X_n can be decomposed as*

$$X_n = Y_n + Z_n$$

where $\{(Y_n, \mathcal{F}_n)\}_{n=1}^{\infty}$ is a martingale and $\{Z_n\}$ is a non-negative increasing process.

Proof. We define the difference r.v.

$$D_1 = X_1, \quad D_j = X_j - X_{j-1}, \quad j \geq 2.$$

Then $X_n = \sum_{j=1}^{n} D_j$, and the defining relation of submartingale yields

$$E(D_j|\mathcal{F}_{j-1}) \geq 0, \quad j \geq 2. \tag{1.4.3}$$

We consider yet another difference

$$S_1 = D_1, \quad S_j = D_j - E(D_j|\mathcal{F}_{j-1}),$$

and let

$$Y_n = \sum_{j=1}^{n} S_j, \qquad Z_n = \sum_{j=1}^{n} E(D_j|\mathcal{F}_{j-1}).$$

It is clear that $X_n = Y_n + Z_n$, $X_1 = Y_1$, $Z_1 = 0$ and $\{Z_n\}_{n=1}^{\infty}$ is a non-negative increasing process (by (1.4.3)). On the other hand, note that $E(S_j|\mathcal{F}_{j-1}) = 0$, it follows that

$$E(Y_n|\mathcal{F}_{n-1}) = \sum_{j=1}^{n-1} S_j = Y_{n-1}$$

and hence a martingale. $\square$

For an increasing family of sub-σ-fields $\{\mathcal{F}_n\}_{n=1}^{\infty}$, let $\mathcal{F}_\infty = \bigcup_{n=1}^{\infty} \mathcal{F}_n$ and let α be a stoping time with respect to $\{\mathcal{F}_n\}_{n=1}^{\infty}$, i.e.,

$$\alpha : \Omega \to \mathbb{N} \cup \{\infty\} \quad \text{such that} \quad \{\alpha = n\} \in \mathcal{F}_n$$

As in last section, the pre-α field $\mathcal{F}_\alpha$ is the family of sets

$$\Lambda = \bigcup_n (\{\alpha = n\} \cap \Lambda_n), \quad \Lambda_n \in \mathcal{F}_n.$$

The following theorems aim at replacing the constant time of a martingale by a stoping time.

Theorem 1.4.4. *Let Y be integrable r.v. and let $X_n = E(Y|\mathcal{F}_n)$ where $\mathcal{F}_n$ is an increasing family of sub-σ-fields (it is a martingale). Then for any stopping time α, we have $X_\alpha = E(Y|\mathcal{F}_\alpha)$.*

Moreover if β is also a stopping time and $\alpha \le \beta$, then $\{(X_\alpha, \mathcal{F}_\alpha),\ (X_\beta, \mathcal{F}_\beta)\}$ is a two term martingale (i.e., $X_\alpha = E(X_\beta|\mathcal{F}_\alpha)$).

Proof. Note that $X_\alpha \in \mathcal{F}_\alpha$. We claim that it is also integrable. Indeed as

$$|X_n| = |E(Y|\mathcal{F}_n)| \le E(|Y||\mathcal{F}_n),$$

we have

$$\int_\Omega |X_\alpha| dP = \sum_n \int_{\{\alpha=n\}} |X_n| dP \le \sum_n \int_{\{\alpha=n\}} |Y| dP = \int_\Omega |Y| dP < \infty.$$

Now if $\Lambda \in \mathcal{F}_\alpha$, let $\Lambda_n = \Lambda \cap \{\alpha = n\}$, then

$$\int_\Lambda X_\alpha dP = \sum_n \int_{\Lambda_n} X_n dP = \sum_n \int_{\Lambda_n} Y dP = \int_\Lambda Y dP\ .$$

Hence $X_\alpha = E(Y|\mathcal{F}_\alpha)$.

For the last statement, note that $\mathcal{F}_\alpha \subset \mathcal{F}_\beta$, hence

$$E(X_\beta|\mathcal{F}_\alpha) = E(E(Y|\mathcal{F}_\beta)|\mathcal{F}_\alpha) = X_\alpha. \qquad \square$$

Corollary 1.4.5. *Under the above assumption and suppose $\{\alpha_i\}_{i=1}^{\infty}$ is an increasing sequence of stopping times. If $\{(X_n, \mathcal{F}_n)\}_n$ is an s-martingale, then $\{(X_{\alpha_i}, \mathcal{F}_{\alpha_i})\}_i$ is an s-martingale.*

Unlike Theorem 1.4.4, in the following theorem, we do not assume that the $\{X_n\}$ is the conditional expectation of an integrable Y.

Theorem 1.4.6. *Let $\{(X_n, \mathcal{F}_n)\}_n$ be a s-martingale. Let $\alpha \le \beta$ be two bounded stopping times, then $\{(X_\alpha, \mathcal{F}_\alpha), (X_\beta, \mathcal{F}_\beta)\}$ is also an s-martingale (of the same type).*

Proof. We prove the theorem for supermartingale. For submartigale, we consider $\{-X_n\}$ instead.

Let $\Lambda \in \mathcal{F}_\alpha$, and let $\Lambda_j = \Lambda \cap \{\alpha = j\}$ $(\in \mathcal{F}_j)$. Then for $k \ge j$, $\Lambda_j \cap \{\beta > k\} \in \mathcal{F}_k$, hence

$$\begin{aligned}\int_{\Lambda_j \cap \{\beta \ge k\}} X_k dP &= \int_{\Lambda_j \cap \{\beta > k\}} X_k dP + \int_{\Lambda_j \cap \{\beta = k\}} X_k dP \\ &\ge \int_{\Lambda_j \cap \{\beta > k\}} X_{k+1} dP + \int_{\Lambda_j \cap \{\beta = k\}} X_k dP\end{aligned}$$

i.e.,

$$\int_{\Lambda_j \cap \{\beta \ge k\}} X_k dP - \int_{\Lambda_j \cap \{\beta \ge k+1\}} X_{k+1} dP \ge \int_{\Lambda_j \cap \{\beta = k\}} X_\beta dP$$

Summing over k, $j \le k \le m$, where m is the upper bound of β, then

$$\int_{\Lambda_j \cap \{\beta \ge j\}} X_\alpha dP - \int_{\Lambda_j \cap \{\beta \ge m+1\}} X_{m+1} dP \ge \int_{\Lambda_j \cap \{j \le \beta \le m\}} X_\beta dP$$

Hence

$$\int_{\Lambda_j} X_\alpha dP \ge \int_{\Lambda_j} X_\beta dP$$

Summing over $1 \le j \le m$,we have

$$\int_{\Lambda} X_\alpha dP \ge \int_{\Lambda} X_\beta dP \qquad \forall\ \Lambda \in \mathcal{F}_\alpha. \qquad \Box$$

Corollary 1.4.7. *If* $\{(X_n, \mathcal{F}_n)\}$ *is a martingale or a supermartingale, then the same is for* $\{(X_{\alpha\wedge n}, \mathcal{F}_{\alpha\wedge n})\}$ *for any stopping time* α.

The theorem still holds if α, β are unbounded. For this we need to associate a random variable X_∞ at ∞.

Theorem 1.4.8. *Assume* $\lim_{n\to\infty} X_n = X_\infty$ *exists and is integrable. Let* α, β *be two arbitrary stopping times. Then Theorem 1.4.6 still hold if* $\{(X_n, \mathcal{F}_n)\}_{n\in N_\infty}$ *is a supermartingale.*

Proof. We first assume that $X_n \geq 0$ and $X_\infty = 0$. Then $X_\alpha \leq \liminf_{n\to\infty} X_{\alpha\wedge n}$, and hence X_α is integrable by Fatou's lemma. The same is for X_β.

From the proof of Theorem 1.4.6, we can conclude that for any m

$$\int_{\Lambda\cap\{\alpha=j\}} X_\alpha dP \geq \int_{\Lambda\cap\{\alpha=j\}\cap\{\beta\leq m\}} X_\beta dP\ .$$

By letting $m \to \infty$ and summing over all j, we have

$$\int_{\Lambda\cap\{\alpha<\infty\}} X_\alpha dP \geq \int_{\Lambda\cap\{\beta<\infty\}} X_\beta dP\ .$$

In addition we have $X_\alpha = X_\infty = 0$ on $\{\alpha = \infty\}$, and $X_\beta = X_\infty = 0$ on $\{\beta = \infty\}$, We conclude that

$$\int_\Lambda X_\alpha dP = \int_\Lambda X_\beta dP$$

and hence $\{(X_\alpha, \mathcal{F}_\alpha),\ (X_\beta, \mathcal{F}_\beta)\}$ is a supermartingale.

For the general case we let

$$X'_n = E(X_\infty|\mathcal{F}_n), \quad X''_n = X_n - X'_n.$$

Then $\{X'_n\}$ is a martingale, and $X_n \geq X'_n$ by the defining property of supermartingale apply to X_n and X_∞. We can apply the above proved case to X''_n, and conclude that $\{X_n\}$ is a supermartingale. $\square$

The above theorems are referred as Doob's *optional sampling theorems.* In terms of gambling, one would hope to devise a strategy to gain advantage of the outcome, but the theorems say that such a strategy does not exist, at least mathematically. The reader can refer to [1, p.327](and the exercises there) for a discussion of the *gambler's ruin problem.*

We use the above stopping time consideration to prove a useful inequality for sub-martingales.

Theorem 1.4.9. *If* $\{(X_j, \mathcal{F}_j)\}_{j=1}^n$ *is a submartingale, then for any real* λ*, we have*

$$\lambda \, P(\max_{1\le j\le n} X_j > \lambda) \;\le\; \int_{\{\max_{1\le j\le n} X_j > \lambda\}} X_n dP \;\le\; E(X_n^+).$$

Proof. Let α be the first j such that $X_j \ge \lambda$ if such $1 \le j \le n$ exists, otherwise let $\alpha = n$. It is clear that α is a stopping time, and hence $\{X_\alpha, X_n\}$ is a submartingale (Theorem 1.4.6). If we write

$$M = \{\max_{1\le j\le n} X_j \ge \lambda\},$$

then $M \in \mathcal{F}_\alpha$ and $X_\alpha \ge \lambda$ on M, hence the first inequality follows from

$$\lambda P(M) \le \int_M X_\alpha dP \le \int_M X_n dP.$$

The second inequality is clear. $\square$.

Corollary 1.4.10. *If* $\{(X_n, \mathcal{F}_n)\}_{n=1}^\infty$ *is a martingale, then for any* $\lambda > 0$*, we have*

$$P(\max_{1\le j\le n} |X_j| > \lambda) \le \int_{\{\max_{1\le j\le n} |X_j| > \lambda\}} |X_n| dP \le \frac{1}{\lambda} E(|X_n|) \ .$$

In addition if $E(|X_n|^2) < \infty$*, then we also have*

$$P(\max_{1\le j\le n} |X_j| > \lambda) \;\le\; \frac{1}{\lambda^2} E(|X_n|^2) \ .$$

For a sequence of independent r.v. $\{X_n\}_{n=0}^{\infty}$ with zero mean and finite variance, we let $S_n = \sum_{j=1}^{n} X_j$. It is well known (Kolmogorov's inequality [1, p. 116]) that for any $\lambda > 0$,

$$P(\max_{1\le j\le n} |S_j| > \lambda) \ \le \ \frac{1}{\lambda^2} E(|S_n|^2).$$

We see that the inequality follows directly from the above corollary.

To conclude this section, we prove a deep theorem on the convergence of the $\{X_n\}_n$, which is also due to Doob. It involves an ingenious method in the proof.

Theorem 1.4.11. *If $\{(X_n, \mathcal{F}_n)\}_{n=0}^{\infty}$ is an L^1-bounded submartingale, then $\{X_n\}_{n=0}^{\infty}$ converges a.e. to a finite limit.*

Proof. First we define, for any pair of rationals a, b, let

$$\Lambda_{[a,b]} = \{\omega : \ \liminf_{n\to\infty} X_n(\omega) < a < b < \limsup_{n\to\infty} X_n(\omega)\} \tag{1.4.4}$$

We show that $\Lambda_{[a,b]}$ is a zero set for any $a, b \in \mathbb{Q}$. It follows that

$$\{\omega : \ \liminf_{n\to\infty} X_n(\omega) < \limsup_{n\to\infty} X_n(\omega)\} = \bigcup_{a,b\in Q,\ a<b} \Lambda_{[a,b]}$$

is a zero set. Note that $\liminf_{n\to\infty} X_n$ is finite almost everywhere (by Fatou lemma and the L^1-boundedness assumption, $E(\liminf |X_n|) \le \liminf E(|X_n|) < \infty$), hence the theorem follows.

It remains to prove (1.4.4). We first introduce some notations. Let $\{x_1, \cdots, x_n\}$ be a numerical sequence, for $a < b$, let

$$\alpha_1 = \min\{j : \ 1 \le j \le n, \ x_j \le a\},$$

$$\alpha_2 = \min\{j : \ \alpha_1 < j \le n, \ x_j \ge b\}.$$

Inductively we define

$$\alpha_{2k-1} = \min\{j : \ \alpha_{2k-2} < j \le n, \ x_j \le a\},$$

$$\alpha_{2k} = \min\{j : \ \alpha_{2k-1} < j \le n, \ x_j \ge b\}.$$

Let α_l be the last one defined. We can think of connecting the consecutive x_i by line segments, Let ν be the number of times the line segments comes from $\le a$ to $\ge b$, i.e., the number of upcrossing through the interval $[a, b]$, it is seen that $\nu = [l/2]$.

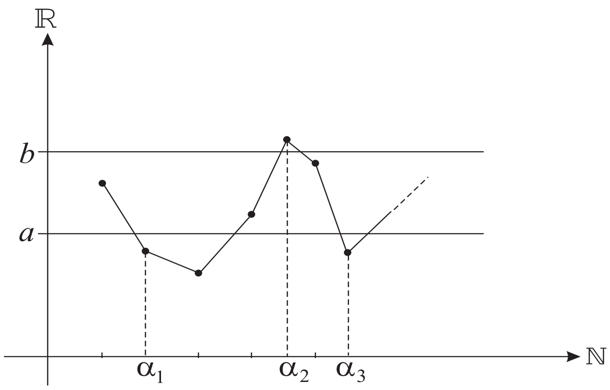

Figure 1.2:

Lemma 1.4.12. *Let* $\{(X_j, \mathcal{F}_j)\}_{j=1}^n$ *be a submartingale and assume that* $X_j \ge 0$. *Let* $\nu^{(n)}_{[0,b]}$ *be the r.v. of the number of upcrossing of* $[0, b]$ *by the sample sequence* $\{X_j(\omega) : \ 1 \le j \le n\}$. *Then*

$$E(\nu^{(n)}_{[0,b]}) \ \le \ \frac{E(X_n - X_1)}{b} \ .$$

Proof. For convenience, we let $\alpha_0 = 1$ and $\alpha_{l+1} = \alpha_{l+2} = \cdots = \alpha_n = n$. Then we have a sequence of stopping times with

$$1 = \alpha_0 \le \alpha_1 < \cdots < \alpha_l \le \alpha_{l+1} \cdots \le \alpha_n = n.$$

We write

$$X_n - X_1 = X_{\alpha_n} - X_{\alpha_0} = \sum_{j=1}^{n-1} (X_{\alpha_{j+1}} - X_{\alpha_j}) = \Big(\sum_{j \ odd} + \sum_{j \ even} \Big)(X_{\alpha_{j+1}} - X_{\alpha_j}).$$

It follows that

$$\sum_{j\ odd} \big(X_{\alpha_{j+1}}(\omega) - X_{\alpha_j}(\omega)\big) \ \geq\ [l(\omega)/2]\cdot b \ =\ \nu^{(n)}_{[0,b]}(\omega)\cdot b.$$

On the other hand by Theorem 1.4.6, $\{X_{\alpha_j} : 0 \leq j \leq n\}$ is a submartingale, so that for each $0 \leq j \leq n-1$, $E(X_{\alpha_{j+1}} - X_{\alpha_j}) \geq 0$. Consequently

$$E(\sum_{j\ even} (X_{\alpha_{j+1}} - X_{\alpha_j})) \geq 0.$$

Therefore $E(X_n - X_1) \geq E(v^{(n)}_{[0,b]})\cdot b$ which yields the lemma. $\square$.

Now to complete the proof of (1.4.4), we consider the upcrossing on any [a,b]. We replace the r.v. in the lemma by $(X_n - a)^+$. The sequence $\{(X_n - a)^+\}_n$ is still submartingale and by the lemma,

$$E(\nu^{(n)}_{[a,b]}) \leq \frac{E(X_n - a)^+ - E(X_1 - a)^+}{b-a} \leq \frac{E({X_n}^+) + |a|}{b-a}.$$

Let $\nu_{[a,b]} = \lim_{n\to\infty} \nu^{(n)}_{[a,b]}$. The L^1-boundedness of $\{X_n\}_n$ implies that $E(\nu_{[a,b]}) < \infty$. Hence $\nu_{[a,b]}$ is finite with probability 1. Note that

$$\begin{aligned}\Lambda_{[a,b]} &= \{\omega :\ \liminf_{n\to\infty} X_n(\omega) \leq a < b \leq \limsup_{n\to\infty} X_n(\omega)\}\\ &\subseteq \{\omega :\ \nu_{[a,b]}(\omega) = \infty\},\end{aligned}$$

hence $\Lambda_{[a,b]}$ is a zero set and (1.4.4) follows. This completes the proof of the theorem. $\square$

Corollary 1.4.13. *Every uniformly bounded s-martingale converges a.e. Also every positive supermartingale and every negative submartingale converges a.e.*

Proof. The first statement follows directly from Theorem 1.4.8 and that $\{X_n\}$ is a submartingale if and only $\{-X_n\}$ is a supermartingale.

For the second part we use Doob's decomposition theorem (Theorem 1.4.3. Let $\{X_n\}_n$ be a positive supermartingale, then $X_n = Y_n - Z_n$ where $\{Y_n\}$ is a martingale and $Z_n \geq 0$, $\{Z_n\} \nearrow$. Since $X_n \geq 0$, it follows that $0 \leq Z_n \leq Y_n$. Let $Z_\infty = \lim_{n\to\infty} Z_n$. It is finite a.e. because

$$E(Z_\infty) = \lim_{n\to\infty} E(Z_n) \leq E(Y_1) < \infty.$$

Also since $\{X_n\}_n$ is a supermartingale,

$$E(Y_n) = E(X_n) + E(Z_n) \leq E(X_1) + E(Z_\infty).$$

This implies $\{Y_n\}_n$ is L^1-uniformly bounded and $\{Y_n\}_n$ converges to a finite limit a.e. (Theorem 1.4.8). The same holds for $\{X_n\}_n$. $\square$

Recall that a sequence of r.v. $\{X_n\}_{n=1}^\infty$ is called uniformly integrable if

$$\lim_{k\to\infty} \int_{|X_n|\geq k} |X_n| dP = 0 \qquad \text{uniformly on } n \ .$$

It is clear that it implies that $\{X_n\}_{n=1}^\infty$ is L^1- bounded. Also, if $X_n \to X$ a.e., then the uniformly boundedness implies that $X_n \to X$ in L^1 ([1, p.96-97]).

Corollary 1.4.14. *If $\{(X_n, \mathcal{F}_n)\}_{n=1}^\infty$ is a submartingale and is uniformly integrable, then $X_\infty = \lim_{n\to\infty} X_n$ a.e. and in L^1.*

Remark. Theorem 1.4.11 and Corollary 1.4.14 are more or less that the converse of Example 1. However for Example 2, the sum $\{S_n\}_{n=1}^\infty$ of i.i.d.r.v. $\{X_n\}_{n=0}^\infty$ with zero mean forms a martingale, but does not converge; it is because the L^1-bounded condition is not satisfies. In fact, we can show that

$$\lim_{n\to\infty} E\big(\frac{|S_n|}{\sqrt{n}}\big) = \sqrt{\frac{2}{\pi}}\sigma$$

where σ is the variance of X_n. For more detail, the reader can refer to [1, Chapter 5, 6] for the law of large number and the central limit theorem for $\{S_n\}_{n=1}^\infty$.

Exercises

1. Suppose $\{(X_n^{(k)}, \mathcal{F}_n)\}_n,\ k = 1, 2$ are two martingales, α is a finite stopping time and $X_\alpha^{(1)} = X_\alpha^{(2)}$. Define $X_n = X_n^{(1)} \chi_{\{n \le \alpha\}} + X_n^{(2)} \chi_{\{n \le \alpha\}}$. Show that $\{(X_n, \mathcal{F}_n)\}_n$ is a martingale.

2 If $\{(X_n, \mathcal{F}_n)\}_n$, $\{(Y_n, \mathcal{F}_n\}_n$ are martingales, then $\{(X_n + Y_n, \mathcal{F}_n)\}$ is again a martingale. However it may happen that $\{X_n\}_n$, $\{Y_n\}_n$ are martingales, but $\{X_n + Y_n\}_n$ is not a martingale. (Note the the σ-field generated by $X_n + Y_n$ may not have the same σ-field $\mathcal{F}_n$.)

3 Prove that for any L^1-bounded s-martingale $\{(X_n, \mathcal{F}_n)\}_n$, and for any α stopping time, then $E(|X_\alpha|) < \infty$.

4. If X is an integrable r.v., then the collection of r.v., $D(X|\mathcal{G})$ with $\mathcal{G}$ ranging over all Borel subfields of $\mathcal{F}$, is uniformly integrable.

5. Find an example of a positive martingale that is not uniformly integrable.

6. Find an example of a martingale $\{X_n\}_n$ such that $X_n \to -\infty$. This implies that in a fair game one player may lose an arbitrary large amount if he stays on long enough. (Hint: Try sums of independent but not identically distributed r.v. with mean 0.)

7. If $\{X_n\}_n$ is a uniformly integrable submartingale, then for any stopping time α, $\{X_{\alpha \wedge n}\}_n$ is again a uniformly integrable submartingale and

$$E(X_1) \le E(X_\alpha) \le \sup_n E(X_n).$$

8 Prove that for any s-martingale, we have for each $\lambda > 0$,

$$\lambda P(\sup_n |X_n| \ge \lambda) \ \le 3 \sup_n E(|X_n|)\ .$$

For a martingale or a positive or nonnegative s-martingale the constant 3 may be replaced by 1.

9. Let $\{X_n\}_n$ be a positive supermartingale. Then for almost every ω, $X_k(\omega) = 0$ implies $X_n(\omega) = 0$ for all $n \geq k$.

10. Every L^1-martingale is the difference of two positive L^1-bounded martingales. (Hint, take one of them to be $\lim_{k\to\infty} E(X_k^+|\mathcal{F}_n)$).

11. If $\{X_n\}$ is a martingale or positive submartingale such that $\sup_n E(X_n^2) \leq \infty$, then $\{X_n\}_n$ converges in L^2 as well as a.e.

12. Show that if $\{(X_n, \mathcal{F}_n)\}_n$ is a submartingate, $X_n \geq 0$, then for $p > 1$,

$$|| \max_{\{1\leq k\leq n\}} ||_p \ \leq \ \frac{p}{p-1} ||X_n||_p \ .$$

(Hint: Show that for $Y \geq 0$, $E(Y^p) = p\int_0^\infty \lambda^{p-1} P(Y \geq \lambda) d\lambda$.)

Chapter 2

Brownian Motion

2.1 Continuous time stochastic processes

We call a family of random variables $\{X_t\}_{t\geq 0}$ on $(\Omega, \mathcal{F}, P)$ a *continuous time* stochastic process. For each $\omega \in \Omega$, $X(\cdot, \omega) = X_{(\cdot)}(\omega)$ is called a sample path. Usually we treat $X(\cdot, \omega) = \omega(t)$ (this can be justified).

There are two most important classes of continuous time stochastic processes. The first one is the **Poisson process** $\{N_t\}_{t\geq 0}$, the number of arrivals in time $[0, t]$ according to an arrival rate λ per unit time.

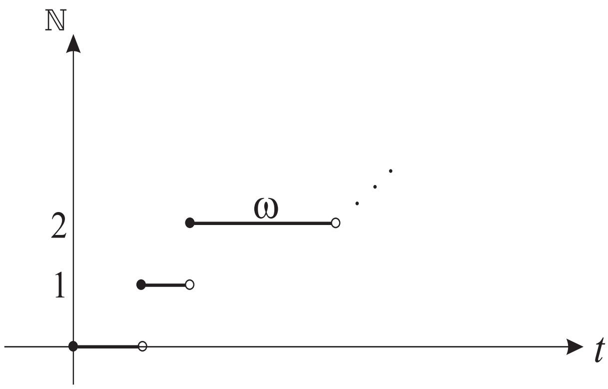

Figure 2.1:

Recall that a Poisson random variable X with rate λ has distribution

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k = 0,\ 1,\ 2\ \cdots$$

Hence N_t has distribution

$$P(N_t = k) = e^{-\lambda t}\frac{(\lambda t)^k}{k!}, \qquad k = 0,\ 1,\ 2\ \cdots$$

A Poisson process is characterized by

1. $N_0 = 0$;
2. Independent increment: for $0 < t_1 < t_2 < \cdots < t_n$,

   $$N_{t_1},\ N_{t_2} - N_{t_1},\ N_{t_3} - N_{t_2},\ \cdots,\ N_{t_n} - N_{t_{n-1}}$$

   are independent.
3. Poisson increment: for $t > s$, $N_t - N_s \sim N_{(t-s)}$, i.e., it has a Poisson distribution with rate $\lambda(t-s)$.

The next one is the **Brownian motion** $\{B_t\}_{t\geq 0}$. It is also called a Wiener process due to the pioneer work of Wiener in the 20's. Recall that a one dimension normal distribution $N(\mu, \sigma^2)$ has density function

$$\frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad x \in \mathbb{R}$$

and $N(0,1)$ is called the standard normal distribution. The Brownian motion is defined by

1. $B_0 = 0$;
2. Independent increment: for $0 < t_1 < t_2 < \cdots < t_n$,

   $$B_{t_1},\ B_{t_2} - B_{t_1},\ B_{t_3} - B_{t_2}, \cdots, B_{t_n} - B_{t_{n-1}}$$

   are independent;

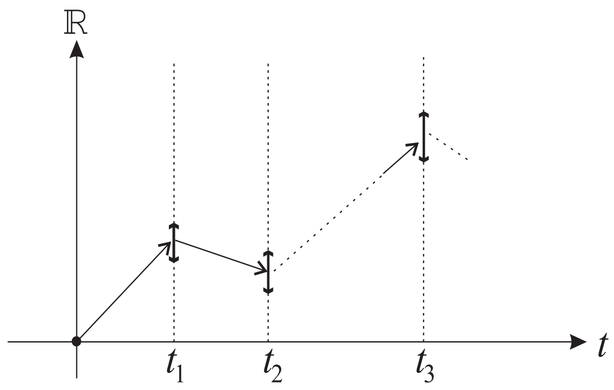

Figure 2.2:

3. Normal increment: for $t > s$, $B_t - B_s$, has normal distribution $N(0, t-s)$.

We will see in the next section that almost all sample paths are continuous, but not differentiable anywhere. We can also define in the same way the higher dimensional Brownian motion, i.e., $\{B_t\}_{t\geq 0}$ has range in R^d; the corresponding density function in (3) is

$$\frac{1}{(2\pi(t-s))^{d/2}}\, e^{-\frac{|x|^2}{2(t-s)}}, \quad x \in \mathbb{R}^d.$$

The Brownian motion was first formulated by Einstein to study diffusion. Heuristically we can realize it as the following: it is direct to check that $p(t,x) = (2\pi t)^{-\frac{d}{2}} e^{-|x|^2/2t}$ satisfies

$$\frac{\partial p(t,x)}{\partial t} = \frac{1}{2}\Delta p(t,x)$$

where $\Delta = \sum_{i=1}^d \frac{\partial^2}{\partial x^2}$ is the Laplacian. Hence it satisfies the *heat equation*

$$\frac{\partial u}{\partial t} = \frac{1}{2}\Delta u \qquad \text{on } \mathbb{R}^d.$$

If we are given an initial condition $u(x,0) = f(x)$, it is known that the solution is given by

$$u(x,t) = \int_{R^d} f(y)p(t, x-y)dy = (f * p_t)(x) = E_x(f(B_t)).$$

Equivalently, we can put it in terms of the Brownian motion $u(x,t) = E(f(x - B_t))$. The study of heat equation can be put entirely into a probabilistic setting.

In view of the two definitions above, there is another more general type of stochastic processes called *Lévy processes*. They are $\{X_t\}_{t\geq 0}$ defined by replacing (3) with a *stationary increment* condition, i.e., for $t > s$, $X_t - X_s$ has the same distribution as $X_{(t-s)}$. The reader can refer to [Ito, Stochastic Process, Springer, 2004] for detail.

In the following we outline the theoretical existence of a probability space $(\Omega, \mathcal{F}, P)$ for a stochastic process $\{X_t\}_{t\geq 0}$, and the measurability problem arised. The space and the σ-field are constructed by the family of finite dimensional distributions as for the discrete time case $\{X_n\}_{n=1}^{\infty}$.

Let $T = [0, \infty)$ and let $\mathbb{R}^T$ denote all functions $\omega : T \to \mathbb{R}$. For $t_1 < \cdots < t_n$, the n-variate r.v. $(X_{t_1}, \cdots, X_{t_n})$ induces a distribution $\mu_{t_1 \cdots t_n}$ on $\mathbb{R}^d$. Let $\mathcal{F}$ be the σ-field generated by $(X_{t_1}, \cdots, X_{t_n})$, i.e., by sets (cylinder sets) of the form

$$E_{t_1 \cdots t_n} = \{\omega : \omega(t_i) \in E_i\} \quad \text{with } \ E_i \ \text{ Borel sets}, \qquad 0 \leq t_1 \cdots < t_n.$$

It can be checked that the family $\{\mu_{t_1 \cdots t_n}\}_{t_1 < \cdots < t_n}$ satisfies the consistency condition:

$$\mu_{t_1 \cdots t_{i-1} t_{i+1} \cdots t_n} = \mu_{t_1 \ldots t_n} \circ \varphi_i^{-1}$$

where $\varphi_i : \mathbb{R}^n \to \mathbb{R}^{n-1}$, $(x_1 \ldots x_n) \to (x_1 \ldots x_{i-1}, x_{i+1} \ldots x_n)$ is the projection map. In this case $\mu_{t_1 \ldots t_{i-1} t_{i+1} \ldots t_n}$ is the marginal distribution of $\mu_{t_1 \ldots t_n}$. By the Kolmogorov extension theorem, there exists a probability P on $(\Omega, \mathcal{F})$ and $\{X_t\}_{t\geq 0}$ is the stochastic process with respect to $(\Omega, \mathcal{F}, P)$.

The probability space defined in this way is, however, still needed to be refined. One of the problems we often encounter is the measurability of union of uncountably many sets with indices from $T = [0, \infty)$. Another problem is that the σ-field $\mathcal{F}$ thus defined does not impose any condition on the continuity of the sample paths on $[0, \infty)$ as is seen in the following example.

Example. Consider $(\Omega, \mathcal{F}, P)$ on which there is a continuous random variable τ with values in $[0, T)$ (i.e., $P(\tau = t) = 0$ for all $t \geq 0$). Define $X_t(\omega) \equiv 0$ for all $t \geq 0$, and

$$Y_t(\omega) = \begin{cases} 1 & \text{if } \ \tau(\omega) = t \\ 0 & \text{if } \ \tau(\omega) \neq t \end{cases}$$

Then the only sample path of $X(\cdot, \omega)$ is 0, but each sample path of $Y(\cdot, \omega)$ has a jump at $\tau(\omega) = t$. On the other hand, it follows from the assumption on τ that $P(Y_t = 1) = P(\tau = t) = 0$ for each t, hence $P(X_t = Y_t) = 1$ for each t. Therefore $\{X_t\}_{t \geq 0}$, $\{Y_t\}_{t \geq 0}$ have the same finite dimensional distribution, they equals the point mass with probability 1 at the path $\omega \equiv 0$.

We will resolve the problem as follows:

Definition 2.1.1. *Two stochastic processes* $\{X_t\}_{t>0}$, $\{Y_t\}_{t \geq 0}$ *on* $(\Omega, \mathcal{F}, P)$ *is called a* version *of each other if* $P(X_t = Y_t) = 1$ *for all* $t \geq 0$.

Note that if we let $N_t = \{X_t \neq Y_t\}$, they are zero set with respect to P. We would like to have $\bigcup_{t \geq 0} N_t$ to be a zero set, however, it is not necessary measurable from the construction of probability space. We will use the following theoretical device to overcome this dilemma. Let D be a countable subset of $T = [0, \infty)$, a function $x : T \to \mathbb{R}$ is called *separable* if for any $t \in T$, there exists a sequence $\{t_n\} \subseteq D$, $t_n \to t$ and $x(t_n) \to x(t)$. For example continuous functions or right continuous functions are separable with respect to the rationals.

Definition 2.1.2. *A stochastic process* $\{X_t\}_{t \geq 0}$ *on* $(\Omega, \mathcal{F}, P)$ *is separable with respect to* D *if there exists an* $\mathcal{F}$*-null set* N *such that* $X(\cdot, \omega)$ *is separable with* *respect to D for all $\omega \notin N$.*

The process $\{Y_t\}_{t\geq 0}$ in the Example is not separable. For if otherwise let D be a countable set in the definition. For any $\tilde{t} \notin D$, let ω be such that $\tau(\omega) = \tilde{t}$, then $Y(\tilde{t}, \omega) = 1$ and $Y(t, \omega) = 0$ for all $t \neq \tilde{t}$. Hence for $\{t_n\} \subseteq D$ and $t_n \to t$, $Y(t_n, \omega) \nrightarrow Y(\tilde{t}, \omega)$.

The following is the main theorem

Theorem 2.1.3. . *Let $\{X_t\}_{t\geq 0}$ be a process on $(\Omega, \mathcal{F}, P)$, then there exists on the same space a separable process $\{X_t'\}_{t>0}$ such that $P(X_t' = X_t) = 1$ for every $t > 0$.*

Sketch of proof ([2, p.555-559]). Note that for any fixed t and for any countable set $D(\subset [0, T))$, the set of ω for which $X(\cdot, \omega)$ is separable with respect to D at t can be written as

$$\bigcap_{n=1}^{\infty} \bigcup_{\substack{|s-t|<\frac{1}{n} \\ s\in D}} \{\omega : \ |X(s,\omega) - X(t,\omega)| < \frac{1}{n}\}.$$

The main task is to construct D (independent of t) so the above set has probability 1. To prove this, we take any interval $I \subseteq T$ and $J \subseteq \mathbb{R}$, and let

$$p(C) = P(\bigcap_{s\in C} (X_s \notin J))$$

for any countable set $C \subset I$. Observe that as C increases, $p(C)$ decreases. we can choose C_n such that $p(C_n) \to \inf_C p(C)$, and let $U_{(I,J)} = \bigcup_n C_n$. Then

$$P\big(\{X_t \in J\} \cap \bigcap_{s\in C_{I,J}} (X_s \notin J)\big) = 0$$

(otherwise, we consider $C_{I,J} \cup \{t\}$ and obtain a contradiction). Let $D = \bigcup C_{(I,J)}$ where (I, J) runs through all intervals I and J with rational end points. If we let

$$N(t) = \bigcup_{I,J} \Big(\{X_t \in J\} \cap \bigcap_{s\in C_{I,J}} (X_s \notin J) \Big)$$

Then we have $P(N(t)) = 0$. It is direct to check that D has the property we want. Now we define the separable version of $\{X_t\}_{t\geq 0}$ as

$$X'(t,\omega) = \begin{cases} X(t,\omega) & \text{if } \ t \in D \ \text{ or } \ t \in D \ \& \ \omega \notin N(t), \\ \limsup_{n\to\infty} X(s_n(t),\omega) & \text{if } \ t \notin D \ \text{ and } \ \omega \in N(t) \end{cases}$$

where $s_n(t)$ is a fixed sequence converges to t. It follows that for each t, and for $\omega \notin N(t)$, $X'(\cdot,\omega)$ is separable with respect to D at t. $\square$

We introduce the following definitions on a probability space $(\Omega, \mathcal{F}, P)$:

− a family $\mathbb{F} = \{\mathcal{F}_t\}_{t\geq 0}$ of sub-σ-fields in $\mathcal{F}$ is called a *filtration* if $\{\mathcal{F}_t\}$ is an increasing sequence of $\sigma-$fields on t;

− a process $\{X_t\}_{t\geq 0}$ is said to be *adaptable* to $\mathbb{F}$ if $X_t \in \mathcal{F}_t$ for each $t > 0$;

− a filtration $\mathbb{F}$ is called *right continuous* if $\mathcal{F}_{t+} = \mathcal{F}_t$ (by definiton $\mathcal{F}_{t+} = \bigcap_{s>t} \mathcal{F}_s$).

For any filtration $\{\mathcal{F}_t\}$, let $\mathcal{G}_t = \mathcal{F}_{t+}$, then $\mathbb{G} = \{\mathcal{G}_t\}_{t\geq 0}$ is right continuous. It is clear that if $\{X_t\}_{t\geq 0}$ is adaptable to $\mathbb{F}$, then it is also adaptable to $\mathbb{G}$. For reasons that will be obvious later, we assume without loss of generality that $\mathbb{F}$ is right continuous. It is also convenient to enlarge $\mathcal{F}_0$ (hence all $\mathcal{F}_t$) to include all subsets of the zero sets (completion by null sets).

With the filtration $\mathbb{F}$, we can define the necessary terminologies as before:

− Markov property: $P(X_{t+s} \in E \mid \mathcal{F}_t) = P(X_{t+s} \in E \mid X_t)$ for $t, s > 0$;

− Martingale: $X_s = E(X_t \mid \mathcal{F}_s)$ for $t > s$;

− Stopping time $\alpha : \Omega \to [0,\infty)$ such that $\{\tau \leq t\} \in \mathcal{F}_t$.

Exercises

1. Let $\xi : \Omega \to [0, \infty)$ be a random variable which satisfies

$$P(\xi \geq t + s \mid \xi \geq s) \qquad \forall\, t, s \geq 0$$

(*lack of memory property*). Show that this property is equivalent to ξ being an exponential distribution, i.e., $P(\xi \geq t) = e^{-\lambda t},\ t > 0$, the waiting time with arrival rate λ.

2. Let $X(t)$ be a Poisson process, let $S_i = \inf\{t > 0 :\ X(t) = n\}$ and let $\xi_n = S_n - S_{n-1}$ be the waiting time of the interarrivals. Show that the $\{\xi_n\}_{n=1}^{\infty}$ are i.i.d. exponential random variables.

3. Conversely, let $\{\xi_n\}_{n=1}^{\infty}$ be i.i.d. exponential random variables. Let $\tau_n = \xi_1 + \cdots + \xi_n$, and let

$$X(t) = \max\{n : \tau_n \leq t\}, \qquad t > 0\ .$$

Show that $X(t)$ is a Poisson process. (This is an alternative way to define a Poisson process.) Use the picture of a sample path to realize τ_n and $X(t)$ are "inverse" of each other (like the inverse function).

4. Show that if X is measurable in the sub-σ-field $\sigma\{X_t : t \in T\}$, then X is measurable in $\sigma\{X_t : t \in S\}$ for some countable subset $S \subset T$.

5. Let $\{X_t\}_{t \geq 0}$ be a stochastic process on $(\Omega, \mathcal{F}, P)$ and $A \in \mathcal{F}$. Show that there is a countable set $S \subset T$ such that $P(A \mid X_t, t \in T) = P(A \mid X_t, t \in S)$.

6. Let $K(s,t)$ be a real function over $T \times T$. Suppose that K is symmetric and nonnegative definite on T. Show that there is a process $\{X_t\}_{t \geq 0}$ for which $(X_{t_1}, \cdots, X_{t_n})$ has the central (zero mean) normal distribution with covariance $\mathrm{cov}(X_{t_i}, X_{t_j}) = K(t_i, t_j),\ \ i, j = 1, \cdots, k\ .$

2.2 Brownian motion and sample paths

For a normal r.v. $X \sim N(0, \sigma^2)$, the symmetry implies $E(X^{2k+1}) = 0$, and the integration by parts yields

$$E(X^{2k}) = 1 \cdot 3 \cdot 5 \cdots (2k-1) \cdot \sigma^{2k} \ . \tag{2.2.1}$$

We also need the following elementary properties of the normal r.v.'s :

– Suppose $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$, and they are independent, then $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

– Suppose $\mathbf{X} = (X_1, \cdots, X_n)$ is a n-variate normal r.v. with distribution $N(\boldsymbol{\mu}, \Sigma)$ where Σ is a symmetric, positive definite $n \times n$-matrix. The density function is given by

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} (\det \Sigma)^{\frac{1}{2}}} \exp\big(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^t \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\big) \ .$$

By a direct calculation, we have $\Sigma = [\mathrm{cov}(X_i, X_j)]$ where $\mathrm{cov}(X_i, X_j) = E((X_i - \mu_i)(X_j - \mu_j))$.

– Suppose $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$. Let $\mathbf{Y} = A\mathbf{X} + \mathbf{c}$ where A is non-singular, then by a change of variable, $\mathbf{Y}$ has density $g(\mathbf{y}) = |\det A|^{-1} f(A^{-1}\mathbf{y} - \mathbf{c})$. A direct substitution yields $\mathbf{Y} \sim N(A\boldsymbol{\mu} + \mathbf{c},\ \Sigma')$ where $\Sigma' = A\Sigma A^t$.

Let $\{B_t\}_{t \geq 0}$ be the Brownian motion defined as in Section 2.1, then the process is *stationary* in the sense that the distribution $B_t - B_s$ depends only on the difference $t - s$. Since B_t has distribution $N(0, t)$, it follows that

$$E(B_t) = 0 \ , \qquad E(B_t^2) = t.$$

Moreover, by using independence, we have

$$E(B_s B_t) = \min(s, t) \tag{2.2.2}$$

This can be checked directly: assume $s < t$, then

$$\begin{aligned} E(B_sB_t) &= E\big(B_s(B_s + (B_t - B_s))\big) \\ &= E(B_s^2) + E(\big(B_s(B_t - B_s)\big) \\ &= E(B_s^2) + E(B_s)E(B_t - B_s) = s. \end{aligned}$$

For $0 < t_1 < t_2 \cdots < t_n$, the joint distribution of $(B_{t_1}, B_{t_2} - B_{t_1}, \cdots, B_{t_n} - B_{t_{n-1}})$ is given by

$$\begin{aligned} f_{t_1\cdots t_n}(x_1, \cdots, x_n) &= \prod_{i=0}^{n} \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}} \exp\Big(- \frac{(x_i - x_{i-1})^2}{2(t_i - t_{i-1})}\Big) \\ &= \frac{1}{\sqrt{2\pi}(\det \Sigma)^{1/2}} \exp\big(- \frac{1}{2}(\mathbf{z}^t\Sigma^{-1}\mathbf{z})\big) \end{aligned}$$

where $\mathbf{z} = (x_1,\ x_2 - x_1,\ \cdots,\ x_n - x_{n-1})$ and

$$\Sigma = \begin{pmatrix} t_1 & 0 & \cdots & 0 \\ 0 & t_2 - t_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & t_n - t_1 \end{pmatrix}.$$

On the other hand the distribution of the $n-variate$ random vector $(B_{t_1},\ B_{t_2}, \cdots B_{t_n})$ is given by

$$g_{t_1,\cdots,t_n}(\mathbf{x}) = \frac{1}{\sqrt{2\pi}(\det \Sigma')^{1/2}} \exp\big(- \frac{1}{2}(\mathbf{x}^t\Sigma'^{-1}\mathbf{x})\big) \tag{2.2.3}$$

where

$$\Sigma' = \begin{pmatrix} t_1 & t_1 & t_1 & \cdots & t_1 \\ t_1 & t_2 & t_2 & \cdots & t_2 \\ t_1 & t_2 & t_3 & \cdots & t_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ t_1 & t_2 & t_3 & \cdots & t_n \end{pmatrix}.$$

This follows from the transformation of the above multivariate normal random vector :

$$\begin{pmatrix} B_{t_1} \\ B_{t_2} \\ \vdots \\ B_{t_n} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & & & \vdots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} B_{t_1} \\ B_{t_2} - B_{t_1} \\ \vdots \\ B_{t_n} - B_{t_{n-1}} \end{pmatrix}$$

and notice that $\operatorname{cov}(B_{t_i}, B_{t_j}) = \min\{t_i,\ t_j\}$. By using this and the construction of the probability space in the last section, we can conclude the existence of a probability space for the Brownian motion.

One of the main properties of the Brownian motion is the continuity of the sample paths.

Theorem 2.2.1. *Let $\{B_t\}_{t\geq 0}$ be a Brownian motion, then there is a version such that with probability 1, the sample paths $B(\cdot, \omega)$ are continuous.*

Proof. Let D denote the set of dyadic rationals in $[0, \infty)$ and let $I_{n,k} = [\frac{k}{2^n}, \frac{k+1}{2^n})$ be the dyadic intervals. Let

$$E_n = \big\{\omega : \max_{0\leq k\leq n2^n} \big(\sup_{r\in I_{n,k}\cap D} |B(r,\omega) - B(\frac{k}{2^n}, \omega)|\big) > \frac{1}{n})\big\}.$$

We divide the proof into three steps.

(i) We claim that $\sum_{n=1}^{\infty} P(E_n) < \infty$. This will be proved in Lemma 2.2.5.

(ii) It follows from (i) and the Borel-Cantelli lemma that

$$E = \overline{\lim_{n\to\infty}}\, E_n = \bigcap_{l=1}^{\infty} \bigcup_{n=l}^{\infty} E_n$$

is a zero set. Observe that for any $\omega \notin E$ there exists ℓ such that for all $n \geq \ell$, $\omega \notin E_n$. It follows that for any $\epsilon > 0$ and $t \in [0, \infty)$, we can find $n_0 > \max\{\ell, t\}$ and $1/n_0 < \epsilon/3$, such that for any $n > n_0$, we have (by $\omega \notin E_n$),

$$|B(r,\omega) - B(\frac{k}{2^n}, \omega)| \leq \frac{1}{n}, \quad \forall\ r \in I_{n,k} \cap D,\ \ 0 \leq \frac{k}{2^n} \leq n\ .$$

This implies that

$$|B(r,\omega) - B(r',\omega)| \leq \varepsilon, \quad \forall\ \ r, r' \in [0,t] \cap D, \ \ |r - r'| \leq \frac{1}{2^n}.$$

We conclude from this that if $\omega \notin E$, then $B(\cdot,\omega)$ is uniformly continuous on the dyadic rationals, and hence $B(\cdot,\omega)$ has a continuous extension $B'(\cdot,\omega)$ on $[0,\infty)$,

$$B'_t(\omega) = B'(t,\omega) = \begin{cases} \lim\limits_{r\to t} B(r,\omega) & \text{if } \ \omega \notin E \\ 0 & \text{if } \ \omega \in E \end{cases}$$

where the r's are the dyadic rationals decrease to t.

(iii) Next we observe that the joint distribution of $(B_{t_1}, \cdots, B_{t_k})$ is the limit of the distributions of $\{B_{r_1(n)}, \cdots, B_{r_k(n)}\}_{n=1}^{\infty}$, where the $r_j(n)$'s are rationals and $r_j(n) \searrow t_j$. (check this by the density functions (2.2.3)). Also note that $(B'_{t_1}, \cdots, B'_{t_k})$ also has the same distribution (see the following Lemma 2.2.2). Therefore $\{B_t\}_t$ and $\{B'_t\}_t$ have the same finite dimensional distributions in the same probability space. In view of $P(B_t \neq B'_t) = P(E) = 0$ for each $t \geq 0$, we conclude that $\{B'_t\}_t$ is a continuous version of $\{B_t\}_t$. $\Box$

The following simple lemma is needed in (iii).

Lemma 2.2.2. *Let $\{X_n\}_n$ and X be k-dimensional r.v. Suppose $X_n \to X$ in probability, and $F_n(x) \to F(x)$ for all x, then F is the distribution function of X.*

Proof. We only prove the 1-dimensional case for simplicity. Let F_X be the distribution function of X. Since $X_n \to X$ in probability, for $\epsilon > 0$, there exists n such that for $k > n$, $P(|X_k - X| \geq \epsilon) \leq \epsilon$. Hence for $k > n$,

$$\begin{aligned} P(X_k \leq x) &\leq P(X \leq x + \epsilon) + P(|X_k - X| \geq \epsilon) \\ &\leq P(X \leq x + \epsilon) + \epsilon \ . \end{aligned}$$

It follows that $\overline{\lim}_n F_n(x) \leq F_X(x)$. By considering $P(X_k > x+h)$, we can use similar technique to show that for $h > 0$, $F_X(x) \leq \underline{\lim}_n F_n(x+h)$. Putting the two inequalities together,

$$F(x) = \overline{\lim}_n F_n(x) \leq F_X(x) \leq \underline{\lim}_n F_n(x+h) = F(x+h)$$

Therefore $F_X(x) = F(x)$ follow by taking $h \to 0$. $\square$

Finally we prove $\sum_{n=1}^{\infty} P(B_n) < \infty$ in (i), which will complete the proof of Theorem 2.2.1. We need a technical lemma.

Lemma 2.2.3. *Suppose $X_1, ..., X_n$ are independent r.v. and are symmetric about 0. Let $S_n = X_1 + ... + X_n$. Then for $\alpha > 0$ and $\epsilon > 0$,*

(i) $P\big(\max_{k\leq n} S_k \geq \alpha\big) \leq 2P\big(S_n \geq \alpha\big)$;

(ii) $P\big(\max_{k\leq n} S_k \geq \alpha\big) \geq 2P\big(S_n \geq \alpha + 2\varepsilon\big) - \sum_{k=1}^{n} P\big(X_k \geq \varepsilon\big)$.

Proof. Note that

$$\begin{aligned} P\big(\max_{k\leq n} S_k \geq \alpha\big) &= P\big(\max_{k\leq n} S_k \geq \alpha,\ S_n \geq \alpha\big) + P\big(\max_{k\leq n} S_k \geq \alpha,\ S_n < \alpha\big) \\ &= P\big(S_n \geq \alpha\big) + P\big(\max_{k\leq n} S_k \geq \alpha, S_n < \alpha\big). \end{aligned}$$

(i) We need only show that the last term is $\leq P(S_n \geq \alpha)$. Let $A_k = \{\max_{i<k} S_i < \alpha \leq S_k\}$ (k is the first time $S_i \geq \alpha$). Then

$$\begin{aligned} P\big(\max_{k\leq n} S_k \geq \alpha,\ S_n < \alpha\big) &= \sum_{k=1}^{n-1} P\big(A_k \cap \{S_n < \alpha\}\big) \\ &\leq \sum_{k=1}^{n-1} P\big(A_k \cap \{S_n - S_k < 0\}\big) \\ &= \sum_{k=1}^{n-1} P\big(A_k \cap \{S_n - S_k > 0\}\big) \\ &\leq \sum_{k=1}^{n-1} P\big(A_k \cap \{S_n > \alpha\}\big) \\ &\leq P\big(S_n \geq \alpha\big). \end{aligned}$$

(Note that the key step is to switch "< 0" to "> 0" in the second equality, because A_k is independent of $\{S_n - S_k\}$ and that $S_n - S_k$ is symmetric about 0.) This proves (i).

(ii) We make use of the following two trivial relations

$$\text{(a)} \qquad S_{k-1} < \alpha,\ X_k < \varepsilon,\ S_n - S_k < -\varepsilon \implies S_n < \alpha,$$

$$\text{(b)} \qquad S_{k-1} < \alpha,\ X_k < \varepsilon,\ S_n \geq \alpha + 2\varepsilon \implies S_n - S_k > \varepsilon.$$

Following the same idea as in (i), we have

$$\begin{aligned}
& \sum_{k=1}^{n-1} P\big(A_k \cap \{S_n < \alpha\}\big) \\
\geq\ & \sum_{k=1}^{n-1} P\big(A_k \cap \{X_k < \varepsilon,\ S_n - S_k < -\varepsilon\}\big) \qquad \text{(by (a))} \\
\geq\ & \sum_{k=1}^{n-1} P\big(A_k \cap \{X_k < \varepsilon,\ S_n - S_k > \varepsilon\}\big) \qquad \text{(by indep. and symm.)} \\
\geq\ & \sum_{k=1}^{n-1} P\big(A_k \cap \{X_k < \varepsilon,\ S_n \geq \alpha + 2\varepsilon\}\big) \qquad \text{(by (b))} \\
\geq\ & \sum_{k=1}^{n-1} P\big(A_k \cap \{S_n \geq \alpha + 2\varepsilon\}\big) - P\big(X_k \geq \varepsilon\big) \\
\geq\ & P\big(S_n \geq \alpha + 2\varepsilon\big) - \sum_{k=1}^{n-1} P\big(X_k \geq \varepsilon\big).
\end{aligned}$$

Combining with the previous part, we have (ii). $\square$

It follows easily from the above that

Corollary 2.2.4. *Under the above assumption*

$$P\big(\max_{k \leq n} |S_k| \geq \alpha\big) \leq 2P\big(\,|\,S_n\,| \geq \alpha\big).$$

Proof. We make use of the symmetry:

$$\begin{aligned}
P\big(\max_{k\leq n}|S_k| \geq \alpha\big) &= P\big(\max_{k\leq n} S_k \geq \alpha\big) + P\big(\max_{k\leq n}(-S_k) \geq \alpha\big) \\
&\leq 2\big(P(S_n \geq \alpha) + P(-S_n \geq \alpha)\big) = 2P\big(|S_n| \geq \alpha\big) \qquad \square.
\end{aligned}$$

Finally we prove the main lemma for Theorem 2.2.1.

Lemma 2.2.5. *With the notations in Theorem 2.2.1, we have* $\sum_n P(B_n) < \infty$.

Proof. We fix δ and t, then by Lemma 2.2.3

$$\begin{aligned} P\big(\max_{i \le 2^m} \big|B(t+\frac{i}{2^m}\delta) - B(t)\big| \ge \alpha\big) &\le 2P\big(|B(t+\delta) - B(t)| \ge \alpha\big) \\ &\le \frac{2}{\alpha^4} E\big(|B(t+\delta) - B(t)|^4\big) \\ &= \frac{6\delta^2}{\alpha^4} \ . \end{aligned}$$

(We have made use of $P(|X| \ge \alpha) \le \alpha^{-4} E(|X|^4)$, and for X normal r.v., $E(X^4) = 3\sigma^4$.) Let $m \to \infty$, we have

$$P(\sup_{0<r<1, r\in D} \big|B(t+r\delta) - B(t)\big| > \alpha) \le \frac{6\delta^2}{\alpha^4}.$$

Therefore for $E_n = \{\omega : \max_{0 \le t \le n2^n} \big(\sup_{r \in I_{n_k} \cap D} |B(r,\omega) - B(k2^{-n},\omega)| > \frac{1}{n}\big)\}$,

$$P(E_n) \le n2^n \big(6 \cdot 2^{-2n}\big)/\big(\frac{1}{n}\big)^4 \ = \ 6n^5 2^{-n} \ .$$

Hence $\sum P(B_n) < \infty$. $\square$

Remark 1. By Theorem 2.2.1, we can assume, in addition to (i)-(iii) in the Brownian motion,

(iv) For each $\omega, B(\cdot, \omega)$ is continuous.

Remark 2. In view of the estimation in Lemma 2.2.5 and the existence of a separable version for any given stochastic process (Theorem 2.1.3), Theorem 2.2.1 can be extended to the more general case:

Theorem 2.2.6. *(*Kolomogorov's continuity theorem*) Let* $\{X_t\}$ *be a stochastic processes. Assume that there exists* $\alpha, \beta > 0$ *such that*

$$E\big(|X(t) - X(s)|^\alpha\big) \le K|t-s|^{1+\beta} \qquad \forall \ \ t, s \ge 0 \ .$$

Then $X(t)$ *has a continuous version.*

The reader can refer to [3, p.31] for the detail.

Definition 2.2.7. *A stochastic process* $\{X_t\}_{t\geq 0}$ *is called a measurable process on* $(\Omega, \mathcal{F}, P)$ *if* $X: \ T \times \Omega \ \longrightarrow \ \mathbb{R}$ *is* $\mathcal{B} \times \mathcal{F}$ *measurable.*

Proposition 2.2.8. *The Brownian motion* $\{B_t\}_{t\geq 0}$ *is a measurable process.*

Proof. Let

$$B^{(n)}(t, \omega) = B(\frac{k}{2^n}, \omega), \qquad \frac{k}{2^n} \leq t < \frac{k+1}{2^n}, \quad k = 0, 1, 2, \cdots$$

Then the map $B^{(n)}(\cdot, \cdot)$ is $\mathcal{B} \times \mathcal{F}$ measurable, as

$$\big\{(t, \omega): \ B^{(n)}(t, \omega) \geq a\big\} \ = \ \bigcup\nolimits_{k,n} \Big(\big[k2^{-n}, \ (k+1)2^{-(n)}\big) \times \{B^{(n)}_{k2^{-n}}(\omega) \geq a\}\Big)$$

By the continuity of the sample path, $B^{(n)}(t, \omega) \to B(t, \omega)$, hence $B(\cdot, \cdot)$ is $\mathcal{B} \times \mathcal{F}$ measurable. $\square$

As a corollary of the estimation in Lemma 2.2.3, we have

Theorem 2.2.9. *For the Brownian motion* $\{B_t\}_{t\geq 0}$*, we have*

$$P\big(\sup_{s \leq t} B_s \geq \alpha\big) \ = \ 2P\big(B_t \geq \alpha\big), \qquad \forall \ \ \alpha \geq 0.$$

Proof. From Lemma 2.2.3(i), we have

$$P\big(\max_{k \leq 2^m} \ B_{k2^{-m}t} \geq \alpha\big) \ \leq \ 2P\big(B_t \geq \alpha\big).$$

Hence as $m \to \infty$, the continuity of $B_{(\cdot)}(\omega)$ implies that

$$P(\sup_{s \leq t} B_s \geq \alpha) \ \leq \ 2P(B_t \geq \alpha)$$

On the other hand, Lemma 2.2.3(ii) implies

$$\begin{aligned} P\big(\sup_{s\le t} B_s \ge \alpha\big) &\ge P\big(\max_{k\le 2^m} B_{k2^{-m}t} \ge \alpha\big) \\ &\ge 2P\big(B_t \ge \alpha + \frac{2}{m}\big) - 2^m P\big(B_{2^{-m}t} \ge \frac{1}{m}\big). \end{aligned}$$

By the Chebychev inequality $(P(|X| \ge \epsilon)) \le \epsilon^{-k} E(|X|^k)$, using $k = 4$), we conclude that the last term $\le 2^m(3(t2^{-m})^2)/m^{-4} = 3m^4t^2 2^{-m}$. This implies that

$$P\big(\sup_{s\le t} B_s \ge \alpha\big) \ge 2P\big(B_t \ge \alpha\big)$$

and the theorem follows. $\square$

We will give a simple proof this theorem again in Section 4 using the strong Markov property and the continuity of the sample paths. In the following we show that the almost all the sample paths are non-differentiable everywhere. First we observe a simple invariant property of the Brownian motion.

Proposition 2.2.10. *(Scaling property) For $c > 0$, let*

$$B'_t(\omega) = c^{-1} B_{c^2 t}(\omega)\ .$$

Then $\{B'_t\}_{t\ge 0}$ is again a Brownian motion.

Proof. It is clear that $\{B'_t\}_{t\ge 0}$ has independent increment, we only need to see the increment has a normal distribution with the correct variance. Recall that if $X \sim N(0, \sigma^2)$, then $cX \sim N(0, c^2\sigma^2)$. Hence

$$B'_t(\omega) - B'_s(\omega) = c^{-1}(B_{c^2t}(\omega) - B_{c^2s}(\omega)).$$

It is a normal r.v. with variance $c^{-2}(c^2t - c^2s) = t - s$. This implies that $\{B'_t\}_{t\ge 0}$ is a Brownian motion. $\square$

Theorem 2.2.11. *Except for a set of zero probability, $B(\cdot, \omega)$ is nowhere differentiate.*

Proof. Let

$$X_{n,k} = \max_{i=0,1,2} \{ |B(\frac{k+(i+1)}{2^n}) - B(\frac{k+i}{2^n})| \}$$

be the maximum oscillation of B_t on three consecutive segments. Then by Proposition 2.2.10,

$$B(\frac{k+(i+1)}{2^n}) \ - \ B(\frac{k+i}{2^n}) \ \sim \ B_{2^{-n}} \ \sim \ 2^{-\frac{n}{2}} B_1.$$

Hence for any $n,\ k$ and $\epsilon > 0$, by independence, we have

$$P\big(X_{n,k} < \epsilon\big) = P\big(|B_1| \ \leq \ 2^{n/2}\epsilon\big)^3 = \big(\frac{1}{\sqrt{2\pi}} \int_{|x|<2^{n/2}\epsilon} e^{-\frac{x^2}{2}} dx\big)^3 \leq (2 \cdot 2^{n/2}\epsilon)^3$$

Define

$$Y_n = \min_{k \leq 2^n} X_{n,k}$$

as the smallest oscillation of the $\{X_{n,k}\}_k$, then $P(Y_n < \epsilon) \leq n2^n(2 \cdot 2^{n/2}\epsilon)^3$. In particular,

$$P(Y_n \ < \ n2^{-n}) \leq n2^n\big(2 \cdot 2^{n/2} \cdot n2^{-n}\big)^3 \to 0. \tag{2.2.4}$$

Now consider the upper and lower derivative of $B(\cdot, \omega)$ from the right ,

$$D^+B(t,\omega) = \limsup_{h \to 0^+} \big(B(t+h,\omega) - B(t,\omega)\big)/h$$

$$D_+B(t,\omega) = \liminf_{h \to 0^+} \big(B(t+h,\omega) - B(t,\omega)\big)/h$$

Let $E = \{\omega : \ \exists \ t > 0 \ni \ -\infty < D_+B(t,\omega) \leq \ D^+B(t,\omega) < \infty\}$, we claim that $P(E) = 0$. Hence for $\omega \notin E_s$, $\ D^+B(t,\omega) = \infty$ or $D_+B(t,\omega) = -\infty$, and the theorem follows.

To prove the claim, let $\omega \in E$, then there exists $K > 0$,

$$-K < D_+B(t,\omega) \ \leq \ D^+B(t,\omega) < K \ .$$

This implies that there exists $\delta > 0$ such that for $t < s < t + \delta$,

$$|B(s,\omega) - B(t,\omega)| \ \le\ K|s-t|.$$

Let n_0 be such that $n_0 > \max\{4K, t\}$ and $4/2^{n_0} < \delta$; for $n > n_0$, let k be such that

$$|\frac{k}{2^n} - t| < \delta, \qquad k = 0, 1, 2, 3,$$

then

$$X_{n,k}(\omega) \ \le\ 4K2^{-n} \ <\ n2^{-n}.$$

It follows that $Y_n(\omega) \ \le\ n2^{-n}$. Let $A_n = \{Y_n \le n2^{-n}\}$. Note that $\omega \in E$ implies $\omega \in A_n$ for $n \ge n_0$, i.e., $\omega \in \bigcup_{k=1}^\infty \bigcap_{n=k}^\infty A_n$ $(= \underline{\lim}_{n\to\infty} A_n)$. Therefore by (2.2.4),

$$P(E) \ \le\ P\big(\underline{\lim}_n A_n\big) \ =\ \lim_{n\to\infty} P\big(\bigcap_{k=n}^{\infty} A_k\big) \ \le\ \lim_{n\to\infty} P\big(A_n\big) \to 0$$

This proves the claim and the theorem follows.

Remark. It is well-known that the the regularity of the sample path can be made precise.

Theorem (Law of iterated logarithm). Let $\{B_t\}_{t\ge 0}$ be a Brownian motion. Then

$$P\Big(\underline{\lim}_{s\to 0} \frac{B_{t+s} - B_s}{\sqrt{2t\log\log\frac{1}{t}}} = -1,\ \ \overline{\lim}_{s\to 0} \frac{B_{t+s} - B_s}{\sqrt{2t\log\log\frac{1}{t}}} = 1\Big) = 1.$$

The proof can be found in standard probability books (e.g., Breiman, Probability). There is a nice proof in "Diffusion Processes and Stochastic Calculus, Baudoin, 2014", using Doob's maximal inequality on the exponential martingale $\{e^{\alpha B_t - \frac{\alpha^2}{2}t}\}_{t\ge 0}$, and the Borel-Cantelli lemma. The theorem implies that the sample paths are Hölder continuous for order $\frac{1}{2} - \varepsilon$ for any $\varepsilon > 0$. The reader can refer to Falconer, Fractal Geometry for a direct proof, also for the Hausdorff dimension of the paths.

Exercies

1. Show that the Poisson Process is a measurable process.

2. Let $\{B_t\}_{t\geq 0}$ be Brownian motion. For fixed t and s, find the distribution of $B_t + B_s$.

3. Show that $\lim_{t\to 0} tB(1/t) = 0$ almost surely. Define $B'_t = tB_{1/t}$ for $t > 0$. Prove that $\{B'_t\}_t$ is again a Brownian motion.

4. Show that $\bigcap_{t>0} \sigma\{B_s : \ s \geq t\}$ is a sub-σ-field contains only sets of probability 0 and 1. Do the same for $\bigcap_{\epsilon>0} \sigma\{B_t : \ 0 < t < \epsilon\}$; give non-trivial examples in the σ-field.

5. Let $\{W_t\}_{t\geq 0}$ be a stochastic process having independent, stationary increments and satisfies $E(W_t) = 0$, $E(W_t^2) = t$. Show that if the finite-dimensional distributions are preserved by the scaling transformation $W(t) \sim c^{-1}W_{c^2 t}, c > 0$, then $\{W_t\}_{t\geq 0}$ is a Brownian motion (Hint: use the Lindeberg theorem [2, p. 368]).

6. (Fourier expansion of Brownian motion)

(a) Show that for $s, t \in [-\pi, \pi]$,

$$\min(s,t) = \frac{ts}{\pi} + \frac{2}{\pi}\sum_{n\geq 1} \frac{\sin nt \ \sin ns}{n^2} \ .$$

(b) Let $\{X_0\}_{n=0}^{\infty}$ be i.i.d standard normal random variables, then

$$W_t = \frac{t}{\sqrt{\pi}} \ X_0 + \sqrt{\frac{2}{\pi}} \ \sum_{n\geq 1} \frac{\sin nt}{n} X_n$$

is a Brownian motion on $[0, \pi]$ (see Breiman, Probability, 1968, P. 259-261).

2.3 Some basic properties

Let $f : [0,t] \to \mathbb{R}$ be a real-valued function, we say that f is of *bounded variation* if

$$V(f) \;=\; \sup_{\mathcal{P}} \sum_{i=1}^{n} |f(t_i) - f(t_{i-1})| < \infty \ .$$

where the supremum is taken over all partition $\mathcal{P} = \{0 = t_1 < t_2 < ... < t_n = t\}$ of [0, t]. It is known that if f is of bounded variation, then f is differentiable a.e. A function f is said to have *quadratic variation* if the limit

$$\lim_{||\mathcal{P}|| \to 0} \sum_{i=1}^{n} |f(t_i) - f(t_{i-1})|^2 \qquad \text{exists} \ ,$$

where $||\mathcal{P}|| = \max_i\{|t_i - t_{i-1}|\}$. The following shows that bounded variation and bounded quadratic variation are two non-compatible conditions.

Proposition 2.3.1. *If $f : [0,t] \to \mathbb{R}$ is continuous and is of bounded variation, then f has zero quadratic variation.*

Proof. Observe that for any $\mathcal{P} = \{0 = t_1 < ... < t_n = t\}$,

$$\sum_{i=1}^{n} |f(t_i) - f(t_{i-1})|^2 \le \max_i |f(t_i) - f(t_{i-1})| \cdot V(f)$$

By the uniform continuity of f, the above expression tends to 0 as $||\mathcal{P}||$ tends to 0. $\square$

Theorem 2.3.2. *Let $[B](t)$ denote the quadratic variation of B_t, then $[B](t) = t$ a.e.*

Proof Let $\delta_n = ||\mathcal{P}_n||$ and satisfies $\sum_{n=1}^{\infty} \delta_n < \infty$. For a partition $\mathcal{P}_n$ with $||\mathcal{P}_n|| \le \delta_n$, let

$$T_n = \sum_{i=1}^{n_k} |B(t_i) - B(t_{i-1})|^2.$$

Then the expectation

$$E(T_n) = E\big(\sum_{i=1}^{n_k} |B(t_i) - B(t_{i-1})|^2\big) = \sum_{i=1}^{n_k}(t_i - t_{i-1}) \ = \ t. \qquad (2.3.1)$$

We claim that

$$\sum_{n=1}^{\infty} E(T_n - E(T_n))^2 = \sum_{n=1}^{\infty} \mathrm{Var}(T_n) < \infty \quad a.e.$$

It follows that $E\big(\sum_{n=1}^{\infty}(T_n - E(T_n))^2\big) < \infty$. Hence $\sum_{n=1}^{\infty}(T_n - E(T_n))^2 < \infty$ a.e., and

$$\lim_{n\to\infty} (T_n - E(T_n)) = 0.$$

This together with (2.3.1) implies that $[B](t) = \lim_{n\to\infty} E(T_n) = t$ *a.e.*

The claim follows from

$$\begin{aligned}
\mathrm{Var}(T_n) &= \mathrm{Var}(\sum_{i=1}^{n_k} |B(t_i) - B(t_{i-1})|^2) \\
&= \sum_{i=1}^{n_k} \mathrm{Var}(|B(t_i) - B(t_{i-1})|^2) \\
&\le \sum_{i=1}^{n_k} E((B(t_i) - B(t_{i-1}))^4) \\
&= \sum_{i=1}^{n_k} 3 \cdot (t_i - t_{i-1})^2 \\
&\le 3||\mathcal{P}_n|| \cdot \sum_{i=1}^{n_k}(t_i - t_{i-1}) \le 3t\delta_n
\end{aligned}$$

and $\sum_{n=1}^{\infty} \mathrm{Var}(T_n) \le 3t \sum_{n=1}^{\infty} \delta_n < \infty$.

Recall that $\{X_t\}_{t\ge 0}$ is a martingale if $E(|X_t|) < \infty$ and for any $s > 0$

$$E(X_{t+s} \mid \mathcal{F}_t) = X(t) \quad a.e.$$

Here $\{\mathcal{F}_t\}_t$ is filtration (right continuous sub-σ-field) generated by $\{X_r : \ 0 \le r \le t\}$.

Theorem 2.3.3. *The following processes are martingales:*
(i) $\{B_t\}_{t\geq 0}$ *; (ii)* $\{B_t^2 - t\}_{t\geq 0}$ *; (iii)* $\{e^{\xi B_t - \frac{\xi^2}{2}t}\}_{t\geq 0}$.

Proof. The proof depends on the independence of $B_{t+s}-B_t$ and $B_r,\ 0 \leq r \leq t$, and also

$$E(g(B_{t+s} - B_t) \mid \mathcal{F}_t) \ = \ E(g(B_{t+s} - B_t))$$

where g is a Borel measurable function.

(i) Since $B_t \sim N(0,t)$, $E(|B_t|) < \infty$. By independence,

$$\begin{aligned} E(B_{t+s}|\mathcal{F}_t) &= E(B_t + (B_{t+s} - B_t)|\mathcal{F}_t) \\ &= E(B_t|\mathcal{F}_t) \ + \ E(B_{t+s} - B_t \mid \mathcal{F}_t) \\ &= B_t + B(B_{t+s} - B_t) = B_t \ . \end{aligned}$$

(ii) Note that $E(B_t^2) = t < \infty$ and

$$\begin{aligned} E(B_{t+s}^2) &= (B_t + (B_{t+s} - B_t))^2 \\ &= B_t^2 + 2B_t(B_{t+s} - B_t) + (B_{t+s} - B_t)^2. \end{aligned}$$

Hence $E(B_{t+s}^2 \mid \mathcal{F}_t) = B_t^2 + 0 + s$. It follows that

$$E\big((B_{t+s}^2 - (t+s)) \mid \mathcal{F}_t\big) \ = \ B_t^2 - t.$$

(iii) It is easy to show by using completing square that

$$E(e^{\xi B_t}) \ = \ \int_{\mathcal{R}} e^{\xi x} \cdot \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t} dx \ = \ e^{t\xi^2/2}.$$

We then apply the same proof as in (ii). $\square$

A process $\{X_t\}_{t\geq 0}$ is a Markov process if for any $s, t > 0$,

$$P(X_{t+s} \in E \mid \mathcal{F}_t) \ = \ P(X_{t+s} \in E \mid X_t)$$

where $\mathcal{F}_t$ generated by $X_r,\ 0 \leq r \leq t$.

Theorem 2.3.4. *$\{B_t\}_{t\geq 0}$ is a Markov process.*

Proof. Since $\mathcal{F}_t$ is generated by $B_{t_1},\ B_{t_2},\ ...,\ B_{t_n}$ for any $0 < t_1 < ... < t_n = t$, to suffices to show that

$$P(B_{t+s} \in E \mid B_{t_1}, ..., B_{t_n}) \;=\; P(B_{t+s} \in E \mid B_{t_n}). \tag{2.3.2}$$

Let $X_1 = B_{t_1},\ \ X_i = B_{t_i} - B_{t_{i-1}}, i = 1,\ \cdots,\ n$ and $X_{n+1} = B_{t+s} - B_{t_n}$. Also let $S_i = X_1 + ... + X_i$, the sum of independent random variables. We have proved in Theorem 1.3.3 that

$$P\big(S_{n+1} \in E \mid S_1, ..., S_n\big) \;=\; P(S_{n+1} \in E \mid S_n) \;=\; \mu_{n+1}(E - S_n)$$

This verifies (2.3.2) with $S_i = B_{t_i}$ and μ_{n+1} the density function of B_{t+s}. $\square$

With the $\{B_t\}_{t\geq 0}$ as a Markov process, it has a transition probability $P(y,t;x,s) \;=\; P\big(B_t \leq y \mid B_s = s\big)$. It follows that the density function is

$$f(y,t;x,s) \;=\; \frac{1}{\sqrt{2\pi(t-s)}}e^{-(y-x)^2/2(t-s)}.$$

The transition probability satisfies the stationary property $P(y,t;x,s) = P(y,t-s;x,0)$.

Analogous to the discrete case, a random variable $\tau:\ \Omega \to [0,\infty)$ is called a stopping time if

$$\{\tau \leq t\} \in \mathcal{F}_t \qquad \forall\ t \geq 0 \tag{2.3.3}$$

where $\{\mathcal{F}_t\}_{t\geq 0}$ is a filtration (see Section 2.1). It follows that if τ is a stopping time, then $\{\tau < t\} \in \mathcal{F}_t$; this follows from

$$\{\tau < t\} = \bigcup\nolimits_{n=1}^{\infty}\{\tau \leq t - \frac{1}{n}\} \in F_{t-1/n} \subseteq \mathcal{F}_t.$$

Also by the right continuity of $\{\mathcal{F}_t\}_{t\geq 0}$, it is easy to show that " $\{\tau < t\} \in \mathcal{F}_t$ for all $t \geq 0$" actually equivalent to (2.3.3).

The pre-τ-field $\mathcal{F}_\tau$ is defined as the family $M \in \mathcal{F}$ such that

$$M \cap \{\tau \le t\} \in \mathcal{F}_t \ , \qquad t \ge 0 \ . \tag{2.3.4}$$

The post-τ field $\mathcal{F}'_\tau$ is defined as the sub-σ-field generated by the process $\mathcal{F}_{\tau+t}$.

Example. Let $\mathcal{F}_t = \sigma\{B_s : \ s \le t\}$, then $\tau = \inf\{t : \ B_t = 1\}$ is a stopping time. Indeed let r denote a rational, then

$$\{\tau < t\} = \bigcup_{r<t} \bigcap_{m \ge 0} \{B_r \ge 1 - 1/m\} \ \in \ \mathcal{F}_t$$

The event $M = \{\inf_{s<\alpha} B_s > -1\}$ is the set of paths that hit 1 before hit -1. It is in $\mathcal{F}_\tau$ because (r, s are rationals)

$$M^c \cap \{\tau < t\} = \bigcup_{s<r<t} \bigcap_{m,n \ge 0} \{ \ B_s \le -(1 - 1/n), \ B_t \ge 1 - 1/m\} \ \in \ \mathcal{F}_t.$$

Theorem 2.3.5. *Let τ be a stopping time finite a.e., then*

$$B^*_t = B_{\tau+t} - B_\tau$$

is a Brownian motion. Moreover for $M \in \mathcal{F}_\tau$, and for E any Borel set in $\mathbb{R}^k$,

$$\begin{aligned} P\big(((B^*_{t_1} \cdots B^*_{t_k}) \in E) \cap M\big) \ &= \ P\big((B^*_{t_1}, \cdots, B^*_{t_n}) \in E\big)P(M) \\ &= \ P\big((B_{t_1} \cdots B_{t_n}\big) \in E)P(M) \ . \end{aligned}$$

Proof. We will prove the identities, then by taking $M = \Omega$, $(B^*_{t_1}, \cdots, B^*_{t_k})$ has the same distribution as $(B_{t_1} \cdots B_{t_k})$. Hence $\{B^*_t\}_{t \ge 0}$ is a Brownian motion.

We first prove the case τ has a countable range D. Note that B^*_t is in the post-τ field $\mathcal{F}'_\tau$, and for any Borel set E in $\mathbb{R}$

$$\big\{B^*_t \in E\big\} = \bigcup_{s \in D} \big\{B_{s+t} - B_s \in E, \ \tau = s\big\}.$$

Let $M \in \mathcal{F}_\tau$ and $E \subseteq \mathbb{R}^k$, then by the independence,

$$\begin{aligned} & P(((B^*_{t_1}, \cdots, B^*_{t_n}) \in E) \cap M) \\ = & \sum_{s \in D} P\big((B^*_{t_1}, \cdots, B^*_{t_n}) \in E) \cap M \cap \{\tau = s\}\big) \qquad (2.3.5) \\ = & \sum_{s \in D} P((B^*_{t_1} \cdots B^*_{t_n}) \in E) P(M \cap \{\tau = s\}) \\ = & P((B^*_{t_1}, \cdots, B^*_{t_n}) \in E) P(M). \end{aligned}$$

For the second identity, we take $M = \Omega$, then we can replace the $(B^*_{t_1}, \cdots, B^*_{t_n})$ in (2.3.4) by $(B_{t_1}, \cdots, B_{t_n})$ and follow by the same argument.

For the general τ, we let

$$\tau_n = \begin{cases} \dfrac{k}{2^n} & \text{if } \dfrac{k-1}{2^n} \le \tau < \dfrac{k}{2^n}, \quad k = 1, 2, \cdots, \\ 0\,, & \tau = \infty. \end{cases}$$

It is clear that $\tau_n \searrow \tau$. For $k2^{-n} \le \tau < (k+1)2^{-n}$,

$$\{\tau_n \le t\} = \{\tau \le k2^{-n}\} \in F_{k2^{-n}} \subseteq F_t.$$

This implies τ_n is a stopping time. Let $B^{(n)}_t = B_{t+\tau_n} - B_{\tau_n}$ and $M \in \mathcal{F}_\tau$ $(\subseteq F_{\tau_n})$. Then for H a closed rectangle in $\mathbb{R}^k$, by the above , we have

$$P\big(((B^{(n)}_{t_1}, \cdots, B^{(n)}_{t_k}) \in H) \cap M\big) \;=\; P((B_{t_1}, \cdots, B_{t_n}) \in H\big)P(M),$$

Since $\{\tau_n(\omega)\}$ converges to $\tau(\omega)$ and the sample paths are continuous, we can take limit so that

$$P\big(((B^*_{t_1}, \cdots, B^*_{t_k}) \in H) \cap M\big) \;=\; P((B_{t_1}, \cdots, B_{t_n}) \in H\big)P(M).$$

The rest of the theorem follows readily. $\square$

Theorem 2.3.6. *(*The strong Markov property*)* *Let τ be a stopping time finite a.e., then for any Borel set $E \subseteq \mathbb{R}$,*

$$P\big(B_{t+\tau} \in E \mid \mathcal{F}_\tau\big) \;=\; P(B_{t+\tau} \in E \mid B_\tau)\;.$$

Proof. We write $B_{t+\tau} = (B_{t+\tau} - B_\tau) + B_\tau = B_t^* + B_t$. It follows from Theorem 2.3.5 that $B_\tau \in \mathcal{F}_\tau$ and $B_t^* \in \mathcal{F}'_\tau$, and they are independent. We can use Theorem 1.3.3 : Let X,Y be independent and $Y \in \mathcal{F}$, then $P(X + Y \in E \mid \mathcal{F}) = P(X + Y \in E \mid Y)$, to conclude the theorem. $\square$

Exercises

1. Consider a process with three states $\{a, b, c\}$, and follows the rule that a goes to b, b goes to c and c goes to a. This is a Markov chain. Show that

$$P(X_3 = c \mid X_2 = a \text{ or } b,\ X_1 = c) \neq P(X_3 = c \mid X_2 = a \text{ or } b).$$

Explain this situation in regard to the independence of the future and the past subject to the present.

2. Show that a process $\{X(t)\}_{t \geq 0}$ with stationary and independent increment and right continuous sample paths has the strong Markov property.

3. Let $\{N(t)\}_t$ be a Poisson process with rate λ. Prove the following are martingales: a. $N(t) - \lambda t$; b. $(N(t) - \lambda t)^2 - \lambda t$; c. $e^{\log(1-\xi)N(t)+\xi\lambda t}$, $0 < \xi < 1$.

4. For $T < \infty$, is $X(t) = B(T - t) - B(T)$ a Brownian motion on $[0, T]$?

2.4 The exit time and hitting time

We use $P_x(\cdot)$ to denote the probability of the Brownian motion starting at x. For $a \in \mathbb{R}$, let

$$T_a = \inf\{t > 0 : B(t) = a\}$$

be the first time of the Brownian motion hitting a. Then T_a is a stopping time. We first give two propositions to describe the exit and hitting probability.

Proposition 2.4.1. *Let $a < x < b$ and $\tau = \min\{T_a, T_b\}$ be the exit time, then $P_x(\tau < \infty) = 1$ and the waiting time for exit is $E_x(\tau) < \infty$.*

Proof. Note that

$$\{\tau > 1\} \ = \ \big\{B(r) \in (a,b) \ \ \forall\, 0 < r < 1\big\} \ \subseteq \ \{B(1) \in (a,b)\}.$$

Then

$$\max\nolimits_{z\in(a,b)} P_z(B(1) \in (a,b)) \ \le \ \max\nolimits_{z\in(a,b)} \Big(\frac{1}{\sqrt{2\pi}} \int_a^b e^{-(z-y)^2/2} dy\Big) \ := \ \theta \ < 1 \ .$$

It follows that

$$\begin{aligned}
& P_x\big(\tau > n\big) \\
= \ & P_x\big(\tau > n-1 \text{ and } B(r) \in (a,b)\ \forall\ n-1 < r \le n\big) \\
= \ & P_x(\tau > n-1)\ P_x\big(B'(r) + B(n-1) \in (a,b)\ \forall\ 0 < r < 1 \mid \{\tau > n-1\}\big),
\end{aligned}$$

where $B'(r) = B(r + (n-1)) - B(n-1)$. The last part can be estimated as follows:

$$\begin{aligned}
\le \ & P_x\big(B'(1) + B(n-1) \in (a,b) \mid \{\tau > n-1\}\big) \\
= \ & P_x\big(B'(1) + B(n-1) \in (a,b) \mid \{B_{n-1} \in (a,b)\}\big) \\
= \ & \big(P_x(B_{n-1} \in (a,b))\big)^{-1} \int_a^b P(B'(1) + y \in (a,b)) d\mu_x(y) \le \theta,
\end{aligned}$$

where μ_x is the distribution of B_{n-1} starts at x. Hence

$$P_x(\tau > n) \ \le \ P_x(\tau > n-1)\ \theta \ \le \ \cdots \ \le \ \theta^n.$$

This implies that $P_x(\tau = \infty) = 0$, i.e., $P_x(\tau < \infty) = 1$. For the second part, we make use of $E(X) \le \sum_{n=0}^{\infty} P(X > n)$ for $X \ge 0$:

$$E(\tau) \le \sum_{n=1}^{\infty} \theta^n < \infty \ . \qquad \square$$

Proposition 2.4.2. *For any $a, b \in \mathbb{R}$, $P_a(P_b < \infty) = 1$.*

Remark. It follow that $P_a(T_a < \infty) = 1$ at any a, and the path will return to a again and again. This property is called the *recurrent* property. We will see in Proposition 2.4.4 that, unlike Proposition 2.4.1, the waiting time for return is ∞ .

Proof. We show that $P_0(T_1 < \infty) = 1$. The other cases are similar. Observe that for any $a \ne b$, the symmetry implies $P_{(a+b)/2}(T_a < T_b) = \frac{1}{2}$. Hence

$$P_0\big(T_{-1} < T_1\big) = \frac{1}{2}, \quad P_{-1}\big(T_{-3} < T_1\big) = \frac{1}{2}, \quad P_{-3}\big(T_{-7} < T_1\big) = \frac{1}{2}, \ \cdots$$

By the continuity of the paths, to reach $-(2^n - 1)$, they must pass though $-1, -3, \cdots$. Let $A_n = \big\{T_{-(2^n-1)} < T_1\big\}$, by the strong Markov property,

$$P_0\big(A_n\big) = P_0\big(T_{-1} < T_1)P_{-1}(T_{-3} < T_1\big) \cdots P_{-(2^n-1)}\big(T_{-(2^n-1)} < T_1\big).$$

It implies that $P_0(A_n) = 2^{-n}$, so that $P_0(\bigcap_{n=1}^{\infty} A_n) = 0$. This yields

$$1 = P_0\big(\bigcup_{n=1}^{\infty} A_n^c\big) = \lim_n P_0\big(T_1 \le T_{-(2^n-1)}\big) = P_0\big(T_1 < \infty\big)$$

and the proposition follows. $\square$

Let $M(t) = \max_{0 \le s \le t} B(s)$. It is clear that

$$\{M(t) \ge a\} = \{T_a \le t\}. \tag{2.4.1}$$

Theorem 2.4.3. *For $a \in \mathbb{R}$, $P\big(M(t) \geq a\big) = 2P(B(t) \geq a)$.*

Remark. We have proved this in Theorem 2.2.9. Here we give a simple proof by the hitting time, using the paths are continuous.

Proof. Note that $\{B(t) \geq a\} = \{T_a \leq t, B(t) - B(T_a) \geq 0\}$, and $\{T_a \leq t\} \in \mathcal{F}_{T_a}$, which is independent of $\{B(t) - B(T_a) \geq 0\} \in \mathcal{F}'_{T_a}$. Hence

$$\begin{aligned} P(\{B(t) \geq a\}) &= P(T_a \leq t, \quad B(t) - B(T_a) \geq 0) \\ &= P\big(T_a \leq t\big)P\big(B(t) - B(T_a) \geq 0\big) \\ &= P\big(T_a \leq t\big)P\big(B^*(t - T_a) \geq 0\big) \\ &= P\big(T_a \leq t\big) \cdot \frac{1}{2} \qquad \text{(by symmetry)} \\ &= \frac{1}{2}P(M(t) \geq a). \qquad \square \end{aligned}$$

As an application of Theorem 2.4.3, we have

Proposition 2.4.4. *The r.v. $T_a : (\Omega, \mathcal{F}, P) \to [0, \infty)$ has density*

$$f_{T_a} = \frac{|a|}{\sqrt{2\pi}} t^{-3/2} e^{-|a|^2/2t}, \qquad t > 0 \ , \tag{2.4.2}$$

and $E(T_a) = \infty$.

Proof. Let $a > 0$, we have by Theorem 2.4.2,

$$\begin{aligned} P(T_a \leq t) &= P(M(t) \geq a) \ = \ 2P(B(t) \geq a) \\ &= 2\int_a^\infty e^{-y^2/2t} dy = \sqrt{\frac{\pi}{2}} \int_{\pi/\sqrt{t}}^\infty e^{-u^2} du. \end{aligned}$$

The density function follows from taking derivative of the above. Since the density of T_a is $\approx t^{-3/2}$, it is clear that $E(T_a) = \infty$. $\square$

Corollary 2.4.5. *For any* $0 < a < b$, $T_b - T_a$ *is independent of* $B(t)$, $t \leq T_a$; *the distribution function of* $T_b - T_a$ *is*

$$f_{T_b - T_a}(t) = \frac{b-a}{\sqrt{2\pi}}\ t^{3/2} e^{-(b-a)^2/2t}\ .$$

Proof. From Theorem 2.3.5, $B^*(t) = B(t+T_a) - B(T_a)$ is a Brownian motion, and is independent of $B(s),\ s \leq T_a$. Hence the same is for

$$T_b - T_a \ = \ \inf\{t > 0 :\ B^*(t) = b - a\},$$

and the density is given by Proposition 2.4.4. $\square$

We use the reflection property of the Brownian motion in Theorem 2.4.3. In the following, we formulate it into a theorem.

Theorem 2.4.6. *(*Reflection principle*) Let* τ *be a stopping time. Define*

$$\widehat{B}(t) \ = \ \begin{cases} B(t), & \text{if } t \leq \tau \\ 2B(\tau) - B(t), & \text{if } t > \tau \end{cases}$$

Then $\widehat{B}_t$ *is also a Brownian motion.*

Remark. Note that for $t > \tau$, $\widehat{B}_t(\omega) = -\big(B_t(\omega) - B_\tau(\omega)\big) + B_\tau(\omega)$ is the reflection along $a = B_\tau(\omega)$.

Proof. Let

$$C[0,\infty) \ = \ \big\{f \text{ continuous on } [0,\infty),\ f(0) = 0\big\}$$

be equipped with the σ-field generated by the cylinder sets. (It contains all the continuous sample paths of the Brownian motion.) Define a map Φ : $[0,\infty) \times C[0,\infty) \times C[0,\infty) \to C[0,\infty)$ by

$$\Phi(T, f, g) = \begin{cases} f(t), & \text{if } 0 \leq t \leq T\ , \\ f(t) + g(t-T), & \text{if } t \geq T\ . \end{cases}$$

It is clear that Φ is measurable.

Note that $B_{t\wedge\tau}$, it is $\mathcal{F}_\tau$-measurable. Let $B_s^* = B_{s+\tau} - B_\tau,\ s > 0$. Then both B_s^* and $-B_s^*$ are Brownian motions with the same distribution, and are independent of $\mathcal{F}_\tau$. Hence $(\tau,\ B_{t\wedge\tau},\ B_s^*)$ and $(\tau,\ B_{t\wedge\tau}, -B_s^*)$ have the same distribution (as r.v. on $[0,\infty) \times C[0,\infty) \times C[0,\infty)$). It follows that

$$\Phi\big(\tau, B_{(\cdot)\wedge\tau},\ B^*_{(\cdot)}\big) \quad \text{and} \quad \Phi\big(\tau,\ B_{(\cdot)\wedge\tau},\ -B^*_{(\cdot)}\big)$$

have the same distribution. Note that the first one is just B_t and the second one is $\widehat{B}(t)$. We conclude that $\widehat{B}(t)$ is also a Brownian motion. $\square$

As a corollary we have

Corollary 2.4.7. *The joint distribution of* $(B(t), M(t))$ *has density*

$$f_{B,M}(x,y) = \frac{2}{\sqrt{2\pi}}\,\frac{2y-x}{t^{3/2}}e^{-(2y-x)^2/2t}, \qquad y \geq 0,\ x\ .$$

Proof. Let $y \geq 0, x$, and let $\widehat{B}(t)$ be the reflection of $B(t)$ at T_y. Then

$$\begin{aligned} P(B(t) \leq x,\ M(t) \geq y) &= P(B(t) \leq x,\ T_y \leq t) \\ &= P(\widehat{B}(t) \geq 2y - x,\ T_y \leq t) \\ &= P(\widehat{B}(t) \geq 2y - x) \\ &= 1 - \frac{1}{\sqrt{2\pi t}}\int_{2y-x}^{\infty} e^{-u^2/2t}du\ . \end{aligned}$$

The density function is obtained by taking partial derivatives on x and y . $\square$

In the rest of the section, we consider the zeros of $B(t)$.

Lemma 2.4.8. *Let* $\mathcal{Z}_t = \{B(s) = 0 \ \text{ for some } \ s \in (0,t)\}$. *Then for* $a \neq 0$,

$$P_a(\mathcal{Z}_t) \ = \ \frac{|a|}{\sqrt{2\pi}}\int_0^t u^{-3/2}e^{-a^2/2u}du\ .$$

Proof. Assume $a > 0$. Then

$$\begin{aligned}
P_a(\mathcal{Z}_t) &= P(\min_{0\le s\le t}(B(s)+a) \le 0) \\
&= P(\min_{0\le s\le t} B(s) \le -a) \\
&= P(M(t) \ge a) \qquad \text{(by symmetry)} \\
&= P(T_a \le t) \\
&= \frac{a}{\sqrt{2\pi}} \int_0^t u^{-3/2} e^{-a^2/2u} du \quad \text{(by Proposition 2.4.4).}
\end{aligned}$$

Proposition 2.4.9. *The probability that $\{B(t)\}_t$ has at least one zero in time (r, s) is* $\frac{2}{\pi}\arccos\sqrt{\frac{r}{s}}$.

Proof. Let $A_{r,s} = \{B(t) = 0 \text{ for some } t \in (r, s)\}$, and let

$$h(x) = P(A_{r,s} \mid B_r = x) = P_x(A_{r,s}).$$

Hence by the above lemma,

$$\begin{aligned}
P(A_{r,s}) &= \int_{-\infty}^{\infty} h(x) \frac{1}{\sqrt{2\pi r}} e^{-x^2/2r} dx \\
&= \sqrt{\frac{2}{\pi r}} \int_0^{\infty} \Big(\frac{x}{\sqrt{2\pi}} \int_0^{s-r} u^{-3/2} e^{x^2/2u} du \Big) e^{-x^2/2r} dx \\
&= \cdots \\
&= \frac{2}{\pi} \arctan \sqrt{\frac{s-r}{r}} = \frac{2}{\pi} \arccos \sqrt{\frac{r}{s}} . \qquad \square
\end{aligned}$$

Corollary 2.4.10. *For $0 < r < s$, the probability that no zero in (r, s) is* $\frac{2}{\pi}\arcsin\sqrt{\frac{r}{s}}$

Proof. It follows from the above and

$$1 - \frac{2}{\pi}\arccos\sqrt{\frac{r}{s}} = \frac{2}{\pi}\arcsin\sqrt{\frac{r}{s}} \qquad \square$$

.

To conclude, we prove a special property of the zero sets of the sample paths.

Lemma 2.4.11. $P(\bigcap_{0\le t\le 1} B(t) < 0) = P(\bigcap_{0\le t\le 1} B(t) > 0) = 0.$

Proof. We make use of Theorem 2.4.3 :

$$\begin{aligned} P\big(\bigcap_{0\le t\le 1} B(t) \le 0\big) &= P\big(\max_{0\le t\le 1} B(t) \le 0\big) \\ &= 1 - P(\max_{0\le t\le 1} B(t) > 0) \\ &= 1 - 2P(B(1) > 0) \\ &= 1 - 2\int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 0 . \qquad \square \end{aligned}$$

Lemma 2.4.12. *Let $\mathcal{Z}_1(\omega) = \{t : \ B(t,\omega) = 0, \ 0 \le t \le 1\}$, then for almost all ω, $\mathcal{Z}_1(\omega)$ has 0 as a limit point.*

Proof. From Lemma 2.4.11, we see that

$$P\big(B(t) \text{ crosses } 0 \text{ for some } 0 \le t \le 1\big) = 1.$$

By the scaling property (Proposition 2.2.10), we conclude that

$$P\big(B(t) \text{ crosses } 0 \text{ for some } 0 \le t \le r\big) = 1.$$

In particular, we take $r_n \searrow 0$. Then

$$P\big(\bigcap_{n=1}^{\infty} \{B(t) \text{ crosses } 0 \text{ for some } 0 \le t \le r_n\}\big) = 1.$$

This implies the lemma. $\square$

Theorem 2.4.13. *For almost all ω, $\mathcal{Z}_1(\omega)$ is a perfect set (hence uncountable) and has Lebesgue measure zero .*

Proof. We use $|\mathcal{Z}_1|$ to denote the Lebesgue measure of $\mathcal{Z}_1$, it is a r.v. and

$$\begin{aligned} E(|\mathcal{Z}_1|) &= E\big(\int_0^1 \chi_{\{B(t)=0\}}\ dt\big) \\ &= \int_0^1 E\big(\chi_{\{B(t)=0\}}\big)\ dt \\ &= \int_0^1 P(B(t)=0)\ dt \ = \ 0. \end{aligned}$$

Hence $|\mathcal{Z}_1| = 0$ P-a.e.

Next we note that $B(\cdot,\omega)$ is continuous, hence $\mathcal{Z}_1(\omega)$ is closed. We need to show that it has no isolated point, and it is a perfect set.

To this end for any rational $r \in (0,1)$, let τ_r be the least $t \geq r$ such that $B(t) = 0$, then τ_r is a stopping time. Let

$$A_r = \{\omega : \ \tau_r(\omega) \text{ is the limit point of } \mathcal{Z}_1(\omega)\}.$$

Then by the strong Markov property and Lemma 2.4.12, we have $P(A_r) = 1$. It follows that $P(\bigcap_r A_r) = 1$ (where the intersection is taken over all rationals ≥ 0. Now for any $\omega \in \bigcap_r A_r$ and for $s \in \mathcal{Z}_1(\omega),\ s > 0$, if s is a left limit point of $\mathcal{Z}_1(\omega)$, then it is not an isolated point. If s is not a left limit point of $\mathcal{Z}_1(\omega)$, then $s = \tau_r(\omega)$ for some rational $r < s$. This implies that s is the right limit point of $\mathcal{Z}_1(\omega)$ (by the strong Markov property, and use the lemma). In either case, s is not an isolated point. Hence $\mathcal{Z}_1(\omega)$ has no isolated point and the proof of theorem is complete.

Exercises

1 Show that $M(t) = \sup_{0<s\leq t}\{B(s)\}$, $|B(t)|$ and $M(t) - B(t)$ have the same distribution.

2. For $a, b > 0$, let $\tau = \min\{T_{-a}, T_b\}$ be the first that the Brownian motion hits a or b. What is $P\big(B(\tau) = -a\big)$ and $P\big(B(\tau) = b\big)$?

What are the hitting probabilities if we change a and b to two slant barriers $-a + rt$ and $b + rt$ for some $r > 0$.

3. Suppose $X_1, \cdots, X_n$ are independent and each has density function

$$h_a(t) = \frac{a}{\sqrt{2\pi}}\, t^{-3/2} e^{-a^2/2t}, \quad t > 0$$

(the density function for the first time the B.M. hits a). Show that

(a) $(X_1 + \cdots + X_n)/n^2$ also has the same distribution. Contrast this with the law of large numbers.

(b) $P\big((\max_{k\leq n} X_k)/n^2 \leq x\big) \to e^{-a\sqrt{2/(\pi x)}}$ for $x > 0$.

4. (a) Show that the probability of the last zero preceding time 1 is distribution over $(0, 1)$ with density $\pi^{-1}(t(1-t))^{-1/2}$.

(b) Similarly calculate the distribution of the position of the first zero following time 1.

(c) Calculate the joint distribution of the two zeros in (a) and (b).