Stochastic Processes, Spring 2020

随机过程论笔记

何憾[1]

June 9, 2020

[1]访问我的个人主页, 可查看并下载其他笔记.

Contents

0.1 Preliminary

Thoughout this note, we denote by $\mathbb{N}_0$ all the non-negetive integers, by $\mathbb{N}$ all the positive integers, and by $\mathbb{R}_+$ all the non-negative real numbers, $\mathbb{R}_+ := [0, \infty)$.

A Stochastic Processes

A stochastic process is a mathematical model for the occurrence, at each moment after the initial time, of a random phenomenon. The randomness is captured by the introduction of a measurable space $(\Omega, \mathcal{F})$, called the sample space. Further, if $\mathbb{P}$ is a probability measure on it, we write "$\mathbb{E}$" the corresponding expectation operator, and "$\mathbb{E}(\cdot\,|\mathcal{G})$" the corresponding conditional expectation opetator, where $\mathcal{G} \subset \mathcal{F}$ is a σ-algebra.

For any map X defined on Ω valued in a Polish space, we write $X \in \mathcal{F}$ if and only if $\sigma(X) \subset \mathcal{F}$. We introduce some further terms. We write $\mathcal{L}[X]$ or $\mathbb{P}_X$ for the distribution of X. For $\mathcal{G} \subset \mathcal{F}$ a sub-σ-algebra or Y random element , we write $\mathcal{L}[X|\mathcal{G}]$, $\mathcal{L}[X|Y]$ for the regular conditional distribution of X given $\mathcal{G}$, Y, respectively.

Throughout this note, let E be a Polish space with Borel algebra $\mathcal{B}(E)$. $(E, \mathcal{B}(E))$ is regarded as the state space. To define a process, we need a index set I interpretated as time. We are mostly interested in the cases $I = \mathbb{N}_0$, $I = [0, \infty)$ (sometimes $I = \mathbb{Z}$, $I = \mathbb{R}$) and without special statements we always assume that I is either $\mathbb{N}_0$ or $\mathbb{R}_+$.

Definition 0.1. A family of random variables $X = \{X_t\}_{t \in I}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $(E, \mathcal{B}(E))$ is called a **stochastic process**.

Remark 0.1. Sometimes families of random variables with more general index sets are called stochastic processes. We come back to this with the point process.

For a fixed sample point $\omega \in \Omega$, the function

$$t \mapsto X_t(\omega) \; ; I \to E$$

is the sample path (realization) of the process X associated with ω. It provides the mathematical model for a random experiment whose outcome can be observed continuously in time (e.g., the number of customers in a queue observed and recorded over a period of time, the trajectory of a molecule subjected to the random disturbances of its neighbors, the output of a communications channel operating in noise).

Let us consider two stochastic processes $X = (X_t)_{t \in I}$ and $Y = (Y_t)_{t \in I}$ defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. When they are regarded as functions of t and ω we would say X and Y were the same if and only if $X_t(\omega) = Y_t(\omega)$ for all t and all ω. However, in the presence of the probability measure $\mathbb{P}$, we could weaken this requirement in at least three different ways to obtain three related concepts of "sameness" between two processes. We list them here.

- X and Y are called *indistinguishable* if almost all their sample paths agree:

$$X_t = Y_t \;\text{ for all } t \quad \text{a.s.}.$$

- X and Y is called a *modification* of each other if, for every $t \in I$, we have $\mathbb{P}\left(X_t = Y_t\right) = 1$.

- X and Y have the same *finite-dimensional distributions* if, for any positive integer $n \geq 1$, and times $t_1 < t_2 < \cdots < t_n$, we have:

$$\mathcal{L}\left(X_{t_1}, \ldots, X_{t_n}\right) = \mathcal{L}\left(Y_{t_1}, \ldots, Y_{t_n}\right) \; .$$

The first property is the strongest; it implies trivially the second one, which in turn yields the third. Clearly, in the case of discrete-time ($\mathrm{I} = \mathbb{N}_0$), the second implies the first. But in the case of continuous-time($\mathrm{I} = \mathbb{R}_+$), two processes can be modifications of one another and yet have completely different sample paths. Here is a standard example:

Example 0.1. Consider a positive random variable U with a continuous distribution. For all $t \geq 0$, put $X_t \equiv 0$ and let

$$Y_t = \begin{cases} 0, & t \neq U \\ 1, & t = U \end{cases} .$$

Y is a modification of X, since for every $t \geq 0$ we have

$$\mathbb{P}\left(X_t = Y_t\right) = \mathbb{P}\left(U \neq t\right) = 1\,.$$

But on the other hand:

$$\mathbb{P}\left(X_t = Y_t \text{ for all } t \geq 0\right) = \mathbb{P}\left(U \neq t \text{ for all } t \geq 0\right) = 0\,.$$

A positive result in this direction is the following. Let $Y = (Y_t)_{t\geq 0}$ be a modification of $X = (X_t)_{t\geq 0}$ and suppose that both processes have a.s. right-continuous sample paths. Then X and Y are indistinguishable.

It does not make sense to ask whether Y is a modification of X, or whether Y and X are indistinguishable, unless X and Y are defined on the same probability space and have the same state space. However, if X and Y have the same state space but are defined on different probability spaces, we can ask whether they have the same finite-dimensional distributions.

For technical reasons in the theory of Lebesgue integration, probability measures are defined on σ-fields and random variables are assumed to be measurable with respect to these σ-fields. Thus, implicit in the statement that a random process $X = \{X_t\}_{t\geq 0}$ is a collection of $\big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$-valued random variables. However, X is really a function of the pair of variables (t, ω), and so, for technical reasons, it is often convenient to have some joint measurability properties.

Definition 0.2. A stochastic process $X = \{X_t\}_{t\geq 0}$ is called **measurable** if, the mapping

$$(t,\omega) \mapsto X_t(\omega) : ([0,\infty)\times\Omega, \mathcal{B}([0,\infty))\times\mathcal{F}) \to (E, \mathcal{B}(E))$$

is measurable.

When X takes values in $\left(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\right)$, it is an immediate consequence of Fubini's theorem that the sample paths of such a process are Borel-measurable functions of $t \in [0,\infty)$, and provided that the components of X have defined expectations, then the same is true for the function $m(t) = \mathbb{E}X_t$. Moreover, if X takes values in $\mathbb{R}$ and A is a subinterval of $[0,\infty)$ such that $\int_A E\,|X_t|\,dt < \infty$, then

$$\int_A |X_t|\,\mathrm{d}t < \infty\mathbb{P}\text{-a.s.}\ , \qquad \text{and} \qquad \int_A \mathbb{E}X_t\,\mathrm{d}t = \mathbb{E}\int_I X_t\,\mathrm{d}t\,.$$

An E-valued stochastic process $X = (X_t)_{t\in I}$ is called

- *real-valued* if $E = \mathbb{R}$,
- a process with *independent increments* if X is real-valued and for all $n \in \mathbb{N}$ and all $t_0, \ldots, t_n \in I$ with $t_0 < t_1 < \ldots < t_n$, we have that

$$\left(X_{t_i} - X_{t_{i-1}}\right)_{i=1,\ldots,n} \qquad \text{is independent}\ ,$$

- a *Gaussian process* if X is real-valued and for $n \in \mathbb{N}$, $t_1, \ldots, t_n \in I$,

$$(X_{t_1}, \ldots, X_{t_n}) \ \text{ is } n\text{-dimensional normally distributed, and}$$

- *integrable* (respectively *square integrable*) if X is real-valued and $\mathbb{E}\,|X_t| < \infty$ (respectively $\mathbb{E}\,|X_t|^2 < \infty$) for all $t \in I$.
- *stationary* if $\mathcal{L}\left[(X_{t+s})_{t\in I}\right] = \mathcal{L}\left[(X_t)_{t\in I}\right]$ for all $s \in I$, and
- a process with *stationary increments* if X is real-valued and

$$\mathcal{L}\left[X_{s+t} - X_t\right] = \mathcal{L}\left[X_s - X_0\right] \qquad \text{for all } s,t \in I\,.$$

A remarkable result show that a Polish space with Borel σ-algebra is a Borel space, i.e., there exists a Borel set $B \in \mathcal{B}(\mathbb{R})$ such that $(E, \mathcal{B}(E))$ and $(B, \mathcal{B}(B))$ are isomorphic. Thus Kolmogorov's extension theorem guarantees the existence of these stochastic process.

Example 0.2. We give some examples with the properties above.

- The Poisson process with intensity λ and the random walk on $\mathbb{Z}$ are processes with stationary independent increments.
- If $X_t, t \in I$, are i.i.d. random variables, then $(X_t)_{t\in I}$ is stationary.
- Let $(X_n)_{n\in\mathbb{Z}}$ be real-valued and stationary and let $k \in \mathbb{N}$ and $c_0, \ldots, c_k \in \mathbb{R}$. Define
$$Y_n := \sum_{i=0}^{k} c_i X_{n-i}$$
Then $Y = (Y_n)_{n\in\mathbb{Z}}$ is a stationary process. If $c_0, \ldots, c_k \geq 0$ and $c_0 + \ldots + c_k = 1$, then Y is called the *moving average* of X (with weights $c_0, \ldots, c_k$).

B Filtrations

There is a very important, nontechnical reason to include σ-fields in the study of stochastic processes, and that is to keep track of information. The temporal feature of a stochastic process suggests a flow of time, in which, at every moment $t \in I$, we can talk about a past, present, and future and can ask how much an observer of the process knows about it at present, as compared to how much he knew at some point in the past or will know at some point in the future. We equip our sample space $(\Omega, \mathcal{F})$ with a *filtration*:

Definition 0.3. Let $\mathfrak{F} = (\mathcal{F}_t)_{t\in I}$ be a family of σ-algebras with $\mathcal{F}_t \subset \mathcal{F}$ for all $t \in I$. $\mathfrak{F}$ is called a **filtration** if

$$\mathcal{F}_s \subset \mathcal{F}_t \text{ for all } s, t \in I \text{ with } s \leq t\,.$$

We set $\mathcal{F}_\infty = \sigma(\cup_t \mathcal{F}_t)$.

The concept of measurability for a stochastic process, introduced in Definition 0.2 is a rather weak one. The introduction of a filtration $\{\mathcal{F}_t\}$ opens up the possibility of more interesting and useful concepts.

Definition 0.4. A stochastic process $X = (X_t)_{t\in I}$ is called **adapted** to the filtration $\mathfrak{F}$ if, for each $t \in I$, X_t is $\mathcal{F}_t$-measurable.

If $X = \{X_t\}_{t\in I}$ is adapted to $\mathfrak{F} = \{\mathcal{F}_t\}_{t\in I}$ and $Y = \{Y_t\}_{t\in I}$ is a modification of X, then Y is also adapted to $\mathfrak{F}$ provided that $\mathcal{F}_0$ contains all the $\mathbb{P}$-negligible sets in $\mathcal{F}$. Note that this requirement is NOT the same as saying that $\mathcal{F}_0$ is complete, since some of the $\mathbb{P}$-negligible sets in $\mathcal{F}$ may not be in the completion of $\mathcal{F}_0$.

Given a stochastic process X, the simplest choice of a filtration is that generated by the process itself: Define $\mathfrak{F}^X = (\mathcal{F}_t^X)_{t\in I}$ by letting

$$\mathcal{F}_t^X := \sigma\left(X_s, s \le t, s \in I\right) .$$

$\mathfrak{F}^X$ is the smallest filtration to which the process X is adapted, called the filtration generated by X. We interpret $A \in \mathcal{F}_t^X$ to mean that by time t, an observer of X knows whether or not A has occurred. The next two exercises illustrate this point.

Example 0.3. Let $X = (X_t)_{t\ge 0}$ be a continuous-time process, *every* sample path of which is RCLL (i.e., right-continuous on $[0,\infty)$ with finite left-hand limits on $(0,\infty)$). Let $s \ge 0$ and

$$A = \{X \text{ is continuous on } [0,s)\} .$$

Note that

$$A = \bigcap_{n=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{\substack{p,q\in[0,s)\cap\mathbb{Q} \\ |p-q|<1/m}} \left\{|X_p - X_q| \le \frac{1}{n}\right\} ,$$

we have $A \in \mathcal{F}_s^X$.

Example 0.4. Let $X = (X_t)_{t\geq 0}$ be a process whose sample paths are RCLL *almost surely.* Let $s \geq 0$ and

$$A = \{X \text{ is continuous on } [0,s)\}.$$

In this case, A can fail to be in $\mathcal{F}_s^X$, but if $\mathfrak{F} = (\mathcal{F}_t)_{t\geq 0}$ is a filtration satisfying $\mathcal{F}_t^X \subset \mathcal{F}_t$ for all $t \geq 0$, and $\mathcal{F}_s$ is complete under $\mathbb{P}$, then $A \in \mathcal{F}_s$.

Note that, we now have

$$A = \bigcap_{n=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{\substack{p,q\in[0,s)\cap\mathbb{Q} \\ |p-q|<1/m}} \left\{ |X_p - X_q| \leq \frac{1}{n} \right\} \quad \mathbb{P}\text{-a.s.},$$

where $A = B$ $\mathbb{P}$ a.s. means that $1_A = 1_B$ $\mathbb{P}$-a.s., i.e., there exists $N \in \mathcal{F}$ so that $A\Delta B \subset N$ and $\mathbb{P}(N) = 0$. Since right-hand site is in $\mathcal{F}_s^X$ and $\mathcal{F}_s$ is complete under $\mathbb{P}$, we have $A \in \mathcal{F}_s$.

We finally construct an example with $A \notin \mathcal{F}_s^X$. Choose $\Omega = [0,2), \mathcal{F} = \mathcal{B}([0,2))$, and $\mathbb{P}(A) = \lambda(A \cap [0,1])$ for $A \in \mathcal{F}$, where λ is Lebesgue mesure. Define, for $\omega \in [0,1]$, $X(t,\omega) = 0$ for all $t \geq 0$; for $\omega \in (1,2)$, $X(t,\omega) = 1_{\{t=\omega\}}$ for all $t \geq 0$. Let $s = 2$. Clearly, in this case $A = [0,1]$. If $A \in \mathcal{F}_s^X$, there exists $B \in \mathcal{B}(\mathbb{R})^{[0,\infty)}$ and sequence $(t_n)_{n\geq 1}$ so that $A = \{(X_{t_n})_{n\geq 1} \in B\}$. Pick $\omega_1 \in (1,2)$ and $\omega_1 \neq t_n$, we have $(X(t_n,\omega))_{n\geq 1} = (0,0,\cdots) \notin B$. On the other hand, Pick $\omega_2 \in [0,1]$, we have $(0,0,\cdots) \in B$. This is a contradiction!

In the case of continuous-time, there is something else to deal with. Let $\mathfrak{F} = \{\mathcal{F}_t\}_{t\geq 0}$ be a filtration. We define $\mathcal{F}_t^- := \sigma\left(\cup_{s<t}\mathcal{F}_s\right)$ to be the σ-field of events strictly prior to $t > 0$ and $\mathcal{F}_t^+ := \cap_{s>t}\mathcal{F}_s$ to be the σ-field of events immediately after $t \geq 0$. We decree $\mathcal{F}_0^- := \mathcal{F}_0$ and say that the filtration $\mathfrak{F}$ is *right-(left-)continuous* if $\mathcal{F}_t = \mathcal{F}_t^+$ (resp., $\mathcal{F}_t = \mathcal{F}_t^-$) holds for every $t \geq 0$.

Exercise 0.1. Let X be a process with every sample path LCRL (i.e., leftcontinuous on $(0,\infty)$ with finite right-hand limits on $[0,\infty)$), and let A be the event that X is continuous on $[0,s]$. Let X be adapted to a right-continuous filtration $\{\mathcal{F}_t\}$. Show that $A \in \mathcal{F}_s$.

Definition 0.5. The stochastic process $X = \{X_t\}_{t\geq 0}$ is called **progressively measurable** with respect to $\mathfrak{F} = \{\mathcal{F}_t\}_{t\geq 0}$ if the mapping

$$(s, \omega) \mapsto X(s, \omega) : ([0, t] \times \Omega, \mathcal{B}([0, t]) \times \mathcal{F}_t) \to (E, \mathcal{B}(E))$$

is measurable, for each $t \geq 0$.

The terminology here comes from [1] . Evidently, any progressively measurable process is measurable and adapted; the following theorem of [1] provides the extent to which the converse is true.

Theorem. *If the stochastic process $X = \{X_t\}_{t\geq 0}$ is measurable and adapted to the filtration $\{\mathcal{F}_t\}$, then it has a progressively measurable modification.*

Nearly all processes of interest are either right- or left- continuous, and for them the proof of a stronger result is easier and will now be given.

Theorem 0.1. *If the stochastic process $X = \{X_t\}_{t\geq 0}$ is adapted to the filtration $\{\mathcal{F}_t\}_{t\geq 0}$ and every sample path is right-continuous or else every sample path is left-continuous, then X is also progressively measurable with respect to $\{\mathcal{F}_t\}$.*

Proof. We treat the case of right-continuity. With $t > 0$, $n \geq 1$, $1 \leq k \leq 2^n$ and $s \in [0, t]$, we define:

$$X^{(n)}(s, \omega) := X\Big(\frac{k}{2^n}, \omega\Big) \quad \text{for} \quad \frac{k-1}{2^n}t < s \leq \frac{k}{2^n}t$$

as well as $X^{(n)}(0, \omega) = X(0, \omega)$. The so-constructed map $(s, \omega) \mapsto X^{(n)}(s, \omega)$ from $[0, t] \times \Omega$ into E is demonstrably $\mathcal{B}[0, t] \times \mathcal{F}_t$-measurable. Besides, by right-continuity we have: $\lim^{(n)}_{n\to\infty}(s, \omega) = X(s, \omega), \forall (s, \omega) \in [0, t] \times \Omega$. Therefore, the (limit) map $(s, \omega) \mapsto X(s, \omega)$ is also $\mathcal{B}[0, t] \times \mathcal{F}_t$ -measurable. $\square$

Remark 0.2. If the stochastic process $X = \{X_t\}_{t\geq 0}$ is right- or left-continuous, but not necessarily adapted to $\{\mathcal{F}_t\}$, then the same argument shows that X is measurable.

A *random time* T is an $\mathcal{F}$-measurable random variable, with values in $I \cup \{\infty\}$. If $X = \{X_t\}_{t \in I}$ is a stochastic process and T is a random time, we define the function X_T on the event $\{T < \infty\}$ by

$$X_T(\omega) := X_{T(\omega)}(\omega)\,,$$

Clearly, if the process $X = \{X_t\}_{t \geq 0}$ is measurable and the random time T with values in $[0, \infty]$ then X_T is a random variable defined on $\{T < \infty\}$. We set the σ-field generated by X_T as

$$\{\{X_T \in A\} : A \in \mathcal{B}(\mathbb{R})\} \cup \{T = \infty\}\,.$$

which is the smallest σ-field on Ω so that X_T is measurable.

Remark 0.3. Sometiones, X_∞ is a well-defined random variable making sense, then X_T can also be defined on Ω, by setting $X_T(\omega) := X_\infty(\omega)$ on $\{T = \infty\}$.

We shall devote our next subsection to a very special and extremely useful class of random times, called stopping times. These are of fundamental importance in the study of stochastic processes, since they constitute our most effective tool in the effort to "tame the continuum of time," as Chung puts it.

C Stopping Times

Let us keep in mind the interpretation of the parameter t as time, and of the σ-field $\mathcal{F}_t$ as the accumulated information up to t. Let us also imagine that we are interested in the occurrence of a certain phenomenon: an earthquake with intensity above a certain level, a number of customers exceeding the safety requirements of our facility, and so on. We are thus forced to pay particular attention to the instant $T(\omega)$ at which the phenomenon manifests itself for the first time. It is quite intuitive then that the event $\{\omega; T(\omega) \leq t\}$, which occurs if and only if the phenomenon has appeared prior to (or at) time t, should be part of the information accumulated by that time.

We can now formulate these heuristic considerations as follows : Let us consider a measurable space $(\Omega, \mathcal{F})$ equipped with a filtration $\mathfrak{F} = \{\mathcal{F}_t\}_{t\in I}$.

Definition 0.6. A random time T is called a **stopping time** of the filtration $\mathfrak{F}$, if the event $\{T \le t\} \in \mathcal{F}_t$, for every $t \in I$.

Clearly, every random time equal to a nonnegative constant is a stopping time.

In the case of discrete time, T is a stopping time of $\mathfrak{F} = \{\mathcal{F}_n\}_{n\ge 0}$ of iff $\{T = n\} \in \mathcal{F}_n$ for all n. In the case of continuous time, if T is a stoppping time of $\{\mathcal{F}_t\}_{t\ge 0}$,

$$\{T < t\} = \bigcup_{n\ge 1} \{T \le t - \frac{1}{n}\} \in \mathcal{F}_t\,.$$

for each $t > 0$, so $\{T = t\} \in \mathcal{F}_t$. However, the converse is NOT ture. Moreover, if a random time T satisfying that $\{T < t\} \in \mathcal{F}_t$ for each $t \ge 0$, then

$$\{T \le t\} = \bigcap_{n\ge 1} \{T \le t + \frac{1}{n}\} \in \mathcal{F}_t^+\,.$$

So such random time T is a stopping time of the $\{\mathcal{F}_t^+\}$. We introduce the following defition.

Definition 0.7. In the case of continuous time, a random time T is called a **optional time** of the filtration $\mathfrak{F} = \{\mathcal{F}_t\}_{t\ge 0}$, if the event $\{T < t\} \in \mathcal{F}_t$, for every $t \ge 0$.

By the arguement above, we get:

Theorem 0.2. *In the case of continuous time, T is an optional time of the filtration $\{\mathcal{F}_t\}_{t\ge 0}$ if and only if it is a stopping time of the (right-continuous) filtration $\{\mathcal{F}_t^+\}_{t\ge 0}$. Particularly, every stopping time is optional, and the two concepts coincide if the filtration is right-continuous.*

Example 0.5. Let $I = \mathbb{N}_0$ and $\mathfrak{F} = \{\mathcal{F}_n\}_{n\geq 0}$ be a filtration. Let $A \in \mathcal{B}(E)$ be measurable. Let $X = (X_n)_{n\geq 0}$ be an adapted E-valued stochastic process. Consider the first time that X hits A :

$$T_A := \inf\{n \geq 0 : X_n \in A\}$$

It is intuitively clear that τ_A should be a stopping time since we can determine by observation up to time n whether $\{\tau_A \leq n\}$ occurs. Consider now the random time of the last visit of X to A:

$$L_A := \sup\{n \geq 0 : X_n \in A\} \ .$$

For a fixed time n, on the basis of previous observations, we cannot determine whether X is already in A for the last time. For this we would have to rely on "prophecy". Hence, in general, L_A is not a stopping time.

Example 0.6. Consider a continuous-time E-valued stochastic process $X = \{X_t\}_{t\geq 0}$ with right-continuous paths, which is adapted to a filtration $\mathfrak{F} = \{\mathcal{F}_t\}_{t\geq 0}$. Consider a subset $A \in \mathcal{B}(E)$ of the state space of the process, and define the hitting time

$$T_A = \inf\{t \geq 0; X_t \in A\} \ .$$

Here we employ the standard convention that the infimum of the empty set is infinity. We will show that,

(i) if A is closed and *every* sample paths of the process X are continuous, then T_A is a $\mathfrak{F}$-stopping time;

(ii) if A is open, then T_A is a $\mathfrak{F}$-optional time.

To show (i), suppose that $\{a_i : i \in \mathbb{N}\}$ is a dense subset of A (such $\{a_i\}$ exists since E is Polish space). Note that for $t \geq 0$,

$$\begin{aligned}\{T_A \leq t\} &= \{\omega : \mathrm{cl}\left(\{X_r(\omega) : r \in [0,t] \cap \mathbb{Q}\}\right) \cap A \neq \emptyset\} \\ &= \bigcap_{n=1}^{\infty} \bigcup_{r\in\mathbb{Q}\cap[0,t]} \bigcup_{i\in\mathbb{N}} \{X_r \in U\left(a_i, 1/n\right)\} \in \mathcal{F}_t\end{aligned}$$

where $U(x,\epsilon) := \{y \in E : d(x,y) < \epsilon\}$. To show (ii), for $t > 0$, note that every sample path of X is right-continuous and A is open, we have

$$\{T_A < t\} = \bigcup_{r \in [0,t) \cap \mathbb{Q}} \{X_r \in A\} \in \mathcal{F}_t \,.$$

If in addition on (ii) X has continuous sample paths, we still can not deduce that T_A is a stoppting time of $\mathfrak{F}$ (of course we assume that $\mathfrak{F}$ is not right-continuous). We give a counterexample.

Suppose $E = \mathbb{R}^d$ and A is a bounded open set. Let B be a d-dimensional Brownian motion starting outside of $\overline{A}$. We may fix a path $\gamma : [0,t] \to \mathbb{R}^d$ with $\gamma[0,t) \cap \overline{A} = \emptyset$ and $\gamma(t) \in \partial A$. Then the σ-algebra $\mathcal{F}_t^B$ contains no nontrivial subset of $\{B(s) = \gamma(s) \ \text{ for all } 0 \leq s \leq t\}$, i.e. no subset other than the empty set and the set itself. If we had $\{T_A \leq t\} \in \mathcal{F}_t^B$, the set

$$\{B(s) = \gamma(s) \text{ for all } 0 \leq s \leq t, T = t\}$$

would be in $\mathcal{F}_t^B$ and (as indicated in Figure 1) a nontrivial subset of this set, which is a contradiction.

Remark 0.4. Because the first hitting times of open or closed sets play an important role, the right-continuous property of the filtration is needed to guarantee that the first hitting times are stopping times.

Let us establish some simple properties of stopping times and optional times.

Lemma 0.3. *Let T and S and be two stopping times of $\mathfrak{F} = \{\mathcal{F}_t\}_{t \in I}$. Then:*

(i) $S \vee T$ *and* $S \wedge T$ *are* $\mathfrak{F}$*-stopping times.*

(ii) $S + T$ *is also a* $\mathfrak{F}$*-stopping time.*

(iii) *For* $s \in I$, $T + s$ *is a* $\mathfrak{F}$*-stopping time. However, in general,* $T - s$ *is not.*

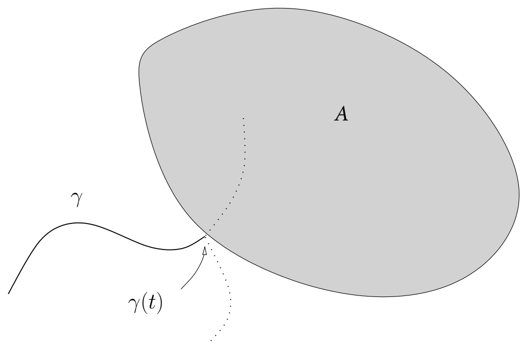


Figure 1: At time t the path γ hits the boundary of A, see the arrow. The two possible dotted continuations indicate that the path may or may not satisfy $T_A = t$.

Before we present the (simple) formal proof, we state that in particular (i) and (iii) are properties we would expect of stopping times. With (i), the interpretation is clear. For (iii), note that $T - s$ peeks into the future by s time units (in fact, $\{T - s \leq t\} \in \mathcal{F}_{t+s}$, while $T + s$ looks back s time units. For stopping times, however, only retrospection is allowed.

Proof. (i). For $t \in I$, we have $\{S \vee T \leq t\} = \{S \leq t\} \cap \{T \leq t\} \in \mathcal{F}_t$ and $\{S \wedge T \leq t\} = \{S \leq t\} \cup \{T \leq t\} \in \mathcal{F}_t$.

(ii). Let $t \in I$. By (i), $T \wedge t$ and $S \wedge t$ are stopping times for any $t \in I$. Hence $T' := (T \wedge t) + 1_{\{T>t\}}$ and $S' := (S \wedge t) + 1_{\{S>t\}}$ are $\mathcal{F}_t$-measurable, and thus $T' + S'$. We conclude $\{T + S \leq t\} = \{T' + S' \leq t\} \in \mathcal{F}_t$.

(iii). For $T + s$, this is a consequence of (ii) (with the stopping time $S \equiv s$). For $T - s$, since T is a stopping time, we have $\{T - s \leq t\} = \{T \leq t + s\} \in \mathcal{F}_{t+s}$. However, in general, $\mathcal{F}_{t+s}$ is a strict superset of $\mathcal{F}_t$; hence $T - s$ is not a stopping time. $\square$

Lemma 0.4. *Let* $\{T_n\}_{n\geq 1}$ *be a sequence of stopping times of* $\mathfrak{F} = \{\mathcal{F}_n\}_{n\geq 0}$ *then the random times*

$$\sup_{n\geq 1} T_n, \quad \inf_{n\geq 1} T_n, \quad \limsup_{n\to\infty} T_n, \quad \liminf_{n\to\infty} T_n$$

are all stopping times of $\mathfrak{F}$*.*

Proof. Not that for each $k \geq 0$,

$$\left\{\sup_{n\geq 1} T_n \leq k\right\} = \bigcap_{n=1}^{\infty} \{T_n \leq k\} \ , \ \left\{\inf_{n\geq 1} T_n \leq k\right\} = \bigcup_{n=1}^{\infty} \{T_n \leq k\} \ ;$$

and

$$\limsup_{n\to\infty} T_n = \inf_{n\geq 1} \sup_{m\geq n} T_m \quad \text{and} \quad \liminf_{n\to\infty} T_n = \sup_{n\geq 1} \inf_{m\geq n} T_m \ ,$$

the desired result follows. □

Lemma 0.5. *Let* $\{T_n\}_{n\geq 1}$ *be a sequence of optional times of* $\mathfrak{F} = \{\mathcal{F}_t\}_{t\geq 0}$ *then the random times*

$$\sup_{n\geq 1} T_n, \quad \inf_{n\geq 1} T_n, \quad \overline{\lim_{n\to\infty}}\, T_n, \quad \underline{\lim_{n\to\infty}}\, T_n$$

are all optional. Furthermore, if the T_n*'s are stopping times, then so is* $\sup_n T_n$*.*

Proof. Note Theorem 0.2 and the identities

$$\left\{\sup_{n\geq 1} T_n \leq t\right\} = \bigcap_{n=1}^{\infty} \{T_n \leq t\} \quad \text{and} \quad \left\{\inf_{n\geq 1} T_n < t\right\} = \bigcup_{n=1}^{\infty} \{T_n < t\} \ .$$

and the proof of Lemma 0.4, the desired result follows. □

Suppose we have a filtration $\{\mathcal{F}_t\}_{t\ inI}$, then how can we measure the information accumulated up to a stopping time T? In order to broach this question, let us suppose that an event A is part of this information, i.e., that the occurrence or nonoccurrence of A has been decided by time T. Now if by time t

one observes the value of T, which can happen only if $T \leq t$, then one must also be able to tell whether A has occurred. In other words, $A \cap \{T \leq t\}$ and $A^c \cap \{T \leq t\}$ must both be $\mathcal{F}_t$-measurable, and this must be the case for any $t \geq 0$. since

$$A^c \cap \{T \leq t\} = \{T \leq t\} \cap (A \cap \{T \leq t\})^c$$

it is enough to check only that $A \cap \{T \leq t\} \in \mathcal{F}_t$ for all $t \in I$.

Definition 0.8. Let T be a stopping time of the filtration $\{\mathcal{F}_t\}$. The *σ-field $\mathcal{F}_T$ of events determined prior to the stopping time T* consists of those events $A \in \mathcal{F}$ for which $A \cap \{T \leq t\} \in \mathcal{F}_t$ for every $t \geq 0$.

It's not hard to verify that $\mathcal{F}_T$ is actually a σ -field and T is $\mathcal{F}_T$ -measurable. Besides, if $T \equiv t$ for some constant $t \in I$, then $\mathcal{F}_T = \mathcal{F}_t$.

Example 0.7. Let $I = \mathbb{N}_0$ and let X be an adapted real-valued stochastic process. Let $a \in \mathbb{R}$ and let

$$T = \inf\{n \geq 0 : X_n \geq a\}$$

be the stopping time of first entrance in $[a, \infty)$. Consider the events

$$A = \{\sup\{X_n : n \in \mathbb{N}_0\} > a - 5\} \ ,$$
$$B = \{\sup\{X_n : n \in \mathbb{N}_0\} > a + 5\} \ .$$

Clearly, for all $n \in \mathbb{N}_0$, $\{T \leq n\} \subset A$, so

$$A \cap \{T \leq n\} = \{T \leq n\} \in \mathcal{F}_n$$

Thus $A \in \mathcal{F}_T$. However, in general, $B \notin \mathcal{F}_T$ since up to time T, we cannot decide whether X will ever exceed $a + 5$.

Lemma 0.6. *If S and T are stopping times of $\mathfrak{F} = (\mathcal{F}_t)_{t \in I}$. Then the following propositions hold.*

(i) *If* $S \le T$*, then* $\mathcal{F}_S \subset \mathcal{F}_T$*.*

(ii) $\mathcal{F}_{S\wedge T} = \mathcal{F}_S \cap \mathcal{F}_T$*, and each of the events* $\{T < S\}, \{S < T\}, \{T \le S\}, \{S \le T\}, \{T = S\}$ *in* $\mathcal{F}_{S\wedge T}$*.*

(iii) *If* $A \in \mathcal{F}_{S\vee T}$ *then* $A \cap \{S \le T\} \in \mathcal{F}_T$*.*

(iv) $\mathcal{F}_{S\vee T} = \sigma(\mathcal{F}_S, \mathcal{F}_T)$*.*

Proof. (i). Take any $A \in \mathcal{F}_S$. For each and $t \in I$, since $S \le T$, we thus get

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

hence $A \in \mathcal{F}_T$.

(ii). By (i), $\mathcal{F}_{S\wedge T} \subset \mathcal{F}_S \cap \mathcal{F}_T$. On the other hand, take any $A \in \mathcal{F}_S \cap \mathcal{F}_T$, then

$$A \cap \{S \wedge T \le t\} = (A \cap \{S \le t\}) \cup (A \cap \{T \le t\}) \in \mathcal{F}_t\,,$$

hence $A \in \mathcal{F}_{S\wedge T}$ and then $\mathcal{F}_{S\wedge T} = \mathcal{F}_S \cap \mathcal{F}_T$. It suffices to show that $\{S < T\} \in \mathcal{F}_S \cap \mathcal{F}_T$, then by symmetry, $\{T < S\} \in \mathcal{F}_{S\wedge T}$, so the same for $\{S \le T\} = \{T < S\}^c$. We firstly show that $\{S < T\} \in \mathcal{F}_T$. When $I = \mathbb{N}_0$, for any $n \ge 0$,

$$\{S < T\} \cap \{T \le n\} = \bigcup_{m=0}^{n} \{S < m\} \cap \{T = m\} \in \mathcal{F}_n\,;$$

when $I = \mathbb{R}_+$, for any $t \ge 0$,

$$\{S < T\} \cap \{T \le t\} = \bigcup_{r\in\mathbb{Q}_+, r\le t} \{S < r\} \cap \{r < T \le t\} \in \mathcal{F}_t\,.$$

To show $\{S < T\} \in \mathcal{F}_S$, note that when $I = \mathbb{N}_0$, for anr $n \ge 0$,

$$\{S < T\} \cap \{S \le n\} = \bigcup_{m=0}^{n} \{T > m\} \cap \{S = m\} \in \mathcal{F}_n\,;$$

when $I = \mathbb{R}_+$ and $t \ge 0$,

$$\{S < T\} \cap \{S \le t\} = \bigcup_{r\in\{t\}\cup(\mathbb{Q}\cap[0,t])} \{S \le r\} \cap \{T > r\} \in \mathcal{F}_t\,,$$

so the desired result follows.

(iii). For any $t \in I$, since $\{S \leq T\} \in \mathcal{F}_{S \wedge T} \subset \mathcal{F}_T$,

$$\begin{aligned} & A \cap \{S \leq T\} \cap \{T \leq t\} \\ & = (A \cap \{S \vee T \leq t\}) \cap (\{S \leq T\} \cap \{T \leq t\}) \in \mathcal{F}_t \,. \end{aligned}$$

(iv). Evidently, $\sigma(\mathcal{F}_S, \mathcal{F}_T) \subset \mathcal{F}_{S \vee T}$. On the other hand, it follows from (iii) that

$$A = (A \cap \{S \leq T\}) \cup (A \cap \{T \leq S\}) \in \sigma(\mathcal{F}_S, \mathcal{F}_T) \,.$$

□

Exercise 0.2. Let T, S be stopping times and Z an integrable random variable. We have (i) $\mathbb{E}[Z|\mathcal{F}_T] = E[Z|\mathcal{F}_{S \wedge T}]$, $\mathbb{P}$ -a.s. on $\{T \leq S\}$ (ii) $\mathbb{E}[\mathbb{E}(Z|\mathcal{F}_T)|\mathcal{F}_S] = E[Z|\mathcal{F}_{S \wedge T}]$, $\mathbb{P}$-a.s..

Now we can start to appreciate the usefulness of the concept of stopping time in the study of stochastic processes.

Theorem 0.7. *In the case of discrete-time, let $X = (X_n)_{n \geq 0}$ adapted to $\mathfrak{F} = (\mathcal{F}_n)_{n \geq 0}$ and let $T < \infty$ be a stopping time of $\mathfrak{F}$. Then X_T is measurable with respect to $\mathcal{F}_T$.*

Proof. Let A be measurable and $n \geq 0$. Hence for all $m \leq n$. $\{T = m\} \cap X_m^{-1}(A) \in \mathcal{F}_m \subset \mathcal{F}_n$ Thus

$$X_T^{-1}(A) \cap \{T \leq m\} = \bigcup_{m=0}^{n} \left(\{T = m\} \cap X_m^{-1}(A)\right) \in \mathcal{F}_n \,.$$ □

Exercise 0.3. Let T, S are stopping times of $\{\mathcal{F}_n\}_{n \geq 0}$. Assume $T < \infty$, then

$$X_T 1_{\{T \leq S\}} \in \mathcal{F}_S \,.$$

In the case of continuous time, for fixed ω, in general, the sample path

$$[0, \infty) \to E \; ; \; t \mapsto X_t(\omega)$$

can not be measurable, hence neither is the composition X_T always measurable. Here one needs assumptions on the regularity of the *paths* $t \mapsto X_t(\omega)$; for example, *right continuity.*

Theorem 0.8. *Let $X = \{X_t\}_{t\geq 0}$ be a progressively measurable process adapted to $\mathfrak{F} = \{\mathcal{F}_t\}_{t\geq 0}$, and let T be a stopping time of $\mathfrak{F}$. Then the random variable X_T, defined on the set $\{T < \infty\} \in \mathcal{F}_T$, is $\mathcal{F}_T$-measurable, and the "stopped process" $X^T = \{X_{T\wedge t}\}_{t\geq 0}$ is progressively measurable with respect to $\mathfrak{F}$.*

Proof. For the first claim, one has to show that for any $A \in \mathcal{B}(E)$ and $t \geq 0$, the event $\{X_T \in A\} \cap \{T \leq t\} \in \mathcal{F}_t$; but this event can also be written in the form $\{X_{T\wedge t} \in A\} \cap \{T \leq t\}$, and so it is sufficient to prove the progressive measurability of the stopped process.

To this end, one observes that the mapping

$$(s, \omega) \mapsto (T(\omega) \wedge s, \omega)$$

of $[0, t] \times \Omega$ into itself is $\mathcal{B}([0, t]) \times \mathcal{F}_t$ -measurable. Besides, by the assumption of progressive measurability, the mapping

$$(s, \omega) \mapsto X(s, \omega) : ([0, t] \times \Omega, \mathcal{B}([0, t]) \times \mathcal{F}_t) \to (E, \mathcal{B}(E))$$

is measurable, and therefore the same is true for the composite mapping

$$(s, \omega) \mapsto X(T(\omega) \wedge s, \omega) : ([0, t] \times \Omega, \mathcal{B}([0, t]) \times \mathcal{F}_t) \to (E, \mathcal{B}(E)) \; .$$

We now complete the proof. $\square$

In the case of continuous time, let T be a optional time of the filtration $\mathfrak{F} = \{\mathcal{F}_t\}_{t\geq 0}$. Then, we can define a σ-field as

$$\{A \in \mathcal{F} : A \cap \{T < t\} \in \mathcal{F}_t \text{ for all } t \geq 0\}$$

On the other hand, T is a stopping time for the right-continuous filtration $\{\mathcal{F}_t^+\}_{t\geq 0}$, so we have *the σ-field $\mathcal{F}_T^+$ of events determined immediately after the optional time T*, given by

$$\mathcal{F}_T^+ = \{A \in \mathcal{F} : A \cap \{T \leq t\} \in \mathcal{F}_t^+ \text{ for all } t \geq 0\},$$

It's easy to check that the two σ-fields coinside. If T is a stopping time, so that both $\mathcal{F}_T, \mathcal{F}_T^+$ are defined, and $\mathcal{F}_T \subset \mathcal{F}_T^+$.

Lemma 0.9. *Given an optional time T of the filtration $\mathfrak{F} = \{\mathcal{F}_t\}_{t\geq 0}$, consider the sequence $\{T_n\}_{n=1}^\infty$ of random times given by*

$$T_n = \sum_{k=1}^\infty \frac{k}{2^n} 1_{\{\frac{k-1}{2^n} \leq T < \frac{k}{2^n}\}} + \infty\, 1_{\{T=\infty\}} \qquad \text{for } n \geq 1 .$$

Then for each n, T_n is a positive stopping time of $\mathfrak{F}$, and $T_n \downarrow T$.

Proof. Evidently, $T_n \downarrow T$. We next show that T_n is a stoppting time. To see this, note that for $t \geq 0$,

$$\{T_n \leq t\} = \bigcup_{1\leq k\leq 2^n t} \{T_n = \frac{k}{2^n}\} = \bigcup_{1\leq k\leq 2^n t} \{\frac{k-1}{2^n} \leq T < \frac{k}{2^n}\} \in \mathcal{F}_t .$$

So T_n is a stopping time of $\mathfrak{F}$. $\square$

Lemma 0.10. $\{T_n\}_{n=1}^\infty$ *is a sequence of optional times and $T = \inf_{n\geq 1} T_n$, then*

$$\mathcal{F}_T^+ = \bigcap_{n=1}^\infty \mathcal{F}_{T_n}^+ .$$

Besides, if each T_n is a positive stopping time and $T < T_n$ on $\{T < \infty\}$ (as in Lemma 0.9), then we have

$$\mathcal{F}_T^+ = \bigcap_{n=1}^\infty \mathcal{F}_{T_n} .$$

Proof. Since $T \leq T_n$ we have $\mathcal{F}_T^+ \subset \mathcal{F}_{T_n}^+$. Thus $\mathcal{F}_T^+ \subset \cap_n \mathcal{F}_{T_n}^+$. On the other hand, if $A \in \mathcal{F}_{T_n}^+$ for all n, then for $t \geq 0$,

$$A \cap \{T < t\} = \bigcup_{n \geq 1} (A \cap \{T_n < t\}) \in \mathcal{F}_t \,.$$

Thus $A \in \mathcal{F}_T^+$. So the first equation holds. To prove the second one, it suffices to show $\mathcal{F}_T^+ \subset \mathcal{F}_{T_n}$ for each n. Then from $\mathcal{F}_T^+ \subset \cap_n \mathcal{F}_{T_n} \subset \cap_n \mathcal{F}_{T_n}^+$ we get the desired result. To this ned, for any $A \in \mathcal{F}_T^+$ and $t \geq 0$, we have

$$A \cap \{T_n \leq t\} = A \cap \{T < t\} \cap \{T_n \leq t\} \in \mathcal{F}_t$$

Thus $A \in \mathcal{F}_{T_n}$. □

We close this section with a statement about the set of jumps for a stochastic process whose sample paths do not admit discontinuities of the second kind.

Definition 0.9. In the case of continuous time, a filtration $\{\mathcal{F}_t\}_{t \geq 0}$ is said to satisfy the **usual conditions** if it is right-continuous and $\mathcal{F}_0$ contains all the $\mathbb{P}$-negligible events in $\mathcal{F}$.

Theorem. *If the process $X = \{X_t\}_{t \geq 0}$ has RCLL paths and is adapted to the filtration $\mathfrak{F} = \{\mathcal{F}_t\}$ which satisfies the usual conditions, then there exists a sequence $\{T_n\}_{n \geq 1}$ of stopping times of $\mathfrak{F}$ which exhausts the jumps of X, i.e.*

$$\{(t,\omega) \in (0,\infty) \times \Omega : X(t,\omega) \neq X(t-,\omega)\} \subset \bigcup_{n=1}^{\infty} \{(t,\omega) \in [0,\infty) \times \Omega : T_n(\omega) = t\}$$

Chapter 1

Discrete-Time Martingales

The study of the dependence between random variables arises in various ways in probability theory. In the theory of Gaussian processes, the basic indicator of dependence is the covariance function, and the inferences made in this theory are determined by the properties of that function. In the theory of Markov chains, the basic dependence is supplied by the transition function, which completely determines the development of the random variables involved in Markov dependence. In this chapter we single out a rather wide class of sequences of random variables (martingales and their generalizations) for which dependence can be studied by methods based on the properties of *conditional expectations*.

1.1 Definitions and Elementary Properties

Although in this chapter we only foucus on discrete-time martingales, we shall introduce martingales indexed by an arbitary index set $I \subset \mathbb{R}$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with a filtration $\mathfrak{F} = (\mathcal{F}_t)_{t \in I}$.

Definition 1.1. Let $X = (X_t)_{t \in I}$ be a real-valued, integrable stochastic process adapted to $\mathfrak{F}$. We say $\{X, \mathfrak{F}, \mathbb{P}\}$ is a

- **martingale**, if $\mathbb{E}\left[X_t|\mathcal{F}_s\right] = X_s$ for all $s,t \in I$ with $t > s$;
- **submartingale** , if $\mathbb{E}\left[X_t|\mathcal{F}_s\right] \geq X_s$ for all $s,t \in I$ with $t > s$;
- **supermartingale**, if $\mathbb{E}\left[X_t|\mathcal{F}_s\right] \leq X_s$ for all $s,t \in I$ with $t > s$.

Remark 1.1. Most of the time, the probability $\mathbb{P}$ is clear, so we shall say X is a (sub-,super-) martingale with respect to $\mathfrak{F}$ or X is a $\mathfrak{F}$-(sub-,super-) martingale for short. We also write $(X, \mathfrak{F})$ or $(X_t, \mathcal{F}_t)_{t\in I}$ is a (sub-,super-) martingale. If we do not explicitly mention the filtration $\mathfrak{F}$, we tacitly assume that $\mathfrak{F}^X = \mathcal{F}^X$ is generated by X, i.e., $\mathcal{F}_t = \sigma\left(X_s, s \leq t\right)$ for all t.

Remark 1.2. If $I = \mathbb{N}$, $\mathbb{N}_0$ or $\mathbb{Z}$, then it is enough to consider at each instant s only $t = s + 1$. In fact, by the tower property of the conditional expectation, we get

$$\mathbb{E}\left[X_{s+2}|\mathcal{F}_s\right] = \mathbb{E}\left[\mathbb{E}\left(X_{s+2}|\mathcal{F}_{s+1}\right)|\mathcal{F}_s\right] .$$

Thus, if the defining equality (or inequality) holds for any time step of size one, by induction it holds for all times.

Remark 1.3. Clearly, for a martingale, the map $t \mapsto \mathbb{E}X_t$ is constant, for submartingales it is monotone increasing and for supermartingales it is monotone decreasing. The name comes from the fact that if f is superharmonic (on $\mathbb{R}^d$), i.e., f has continuous derivatives of order ≤ 2 and $\Delta f \leq 0$, then

$$f(x) \geq \frac{1}{|B(x,r)|}\int_{B(x,r)} f(y)dy$$

where $B(x,r) = \{y : |x-y| \leq r\}$ is the ball of radius r, and $|B(x,r)|$ is the volume of the ball.

Remark 1.4. Let $\mathfrak{F}$ and $\mathfrak{F}'$ be filtrations with $\mathcal{F}_t \subset \mathcal{F}'_t$ for all t, and let X be an $\mathfrak{F}'$ – (sub–, super-) martingale that is adapted to $\mathfrak{F}$. Then X is also a (sub-, super-) martingale with respect to the smaller filtration $\mathfrak{F}$. Indeed, for $s < t$ and for the case of a submartingale,

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

In particular, if $(X, \mathfrak{F})$ is a (sub-, super-) martingale, then $(X, \mathfrak{F}^X)$ must be a (sub-, super-) martingale.

A Examples

We now give some examples, however we will encounter many other examples in this note. We begin by a trivial example.

Example 1.1. Let ξ be an integrable random variable. Define $X = (X_t)_{t\in I}$ by

$$X_t := \xi \ \ , \text{ for all } t \in I\ .$$

Then given any filtration $\mathfrak{F} = (\mathcal{F}_t)_{t\in I}$ satisfying $\sigma(\xi) \subset \mathcal{F}_t$ for each t, X is a martingale with respect to $\mathfrak{F}$.

Example 1.2 (Levy Martingale)**.** Let ξ in L^1 and let $\mathfrak{F} = (\mathcal{F}_t)_{t\in I}$ be a filtration. Define $M = (M_t)_{t\in I}$ by

$$M_t := \mathbb{E}(\xi|\mathcal{F}_t) \text{ for all } t \in I\ .$$

Then $(M_t, \mathcal{F}_t)_{t\in I}$ is a martingale as a consequence of the tower property of conditional expectation. We call it a *Levy martingale.* The key property of M is that it is uniformly integrable.

To see this, without loss of generality, suppose that $\xi \geq 0$. (If not, consider $|\xi| \geq 0$ and note that $|M_t| \leq \mathbb{E}(|\xi||\mathcal{F}_t) =: \tilde{M}_t$ for any t.) Then for any $t \in I$ and $\lambda > 0$,

$$\mathbb{E}\,|M_t|1_{\{|M_t|>\lambda\}} = \mathbb{E}\,M_t 1_{\{M_t>\lambda\}} = \mathbb{E}\,\xi 1_{\{M_t>\lambda\}}$$

and by Markov inequality,

$$\mathbb{P}(M_t > \lambda) \leq \frac{\mathbb{E}M_t}{\lambda} = \frac{\mathbb{E}\xi}{\lambda}\ .$$

Hence

$$\sup_{t \in I} \mathbb{E} |M_t| 1_{\{|M_t| > \lambda\}} = \sup_{t \in I} \mathbb{E}\, \xi 1_{\{M_t > \lambda\}} \to 0 \quad \text{as } \lambda \to \infty$$

follows from the absolute continuity of the integral.

Later, we will see that, in Theorem 1.28, surprisingly, any uniformly integrable martingale is Levy's martingale, in discrete-time case.

Example 1.3. Let $\{\xi_n\}_{n \in \mathbb{N}}$ be independent r.v.'s on $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{E}\, \xi_n = 1$ for each n. Let $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$ is the filtration generated by $\{\xi_n\}$, and set $\mathcal{F}_0 = \{\varnothing, \Omega\}$. Then $M_0 = 1$ and

$$M_n = \prod_{m \leq n} \xi_m \ , \text{ for } n \in \mathbb{N}$$

defines a martingale $\{M_n, \mathcal{F}_n\}_{n \geq 0}$. To prove this, note that

$$\mathbb{E}\left(M_{n+1} | \mathcal{F}_n\right) = M_n \mathbb{E}\left(\xi_{n+1} | \mathcal{F}_n\right) = M_n \quad \text{for } n \in \mathbb{N}_0 \, .$$

We will describe three examples related to *random walk*. In the next threes examples, we always suppose that $\xi_1, \xi_2, \ldots$ are i.i.d. r.v.'s with mean μ. Let $S_n = \xi_1 + \cdots + \xi_n$ and $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$ for $n \geq 1$. Take $S_0 = 0$ and $\mathcal{F}_0 = \{\emptyset, \Omega\}$, denote $\mathfrak{F} = \{\mathcal{F}_n\}_{n \in \mathbb{N}_0}$.

Example 1.4 (Linear martingale)**.** If $\mu = 0$ then $\{S_n, \mathcal{F}_n\}_{n \in \mathbb{N}_0}$ is a martingale.

To prove this, we observe that $S_n \in \mathcal{F}_n$, $\mathbb{E} |S_n| < \infty$, and ξ_{n+1} is independent of $\mathcal{F}_n$, so

$$\mathbb{E}\left(S_{n+1} | \mathcal{F}_n\right) = \mathbb{E}\left(S_n | \mathcal{F}_n\right) + \mathbb{E}\left(\xi_{n+1} | \mathcal{F}_n\right) = S_n + \mu = S_n$$

If $\mu \leq 0$ then the computation just completed shows $\mathbb{E}\left(S_{n+1} | \mathcal{F}_n\right) \leq S_n$, i.e., $\{S_n, \mathcal{F}_n\}$ is a supermartingale. In this case, S_n corresponds to betting on an unfavorable game. If $\mu \geq 0$ then $\{S_n, \mathcal{F}_n\}$ is a submartingale. In this case, S_n corresponds to betting on a favorable game.

Applying the first result to $\xi_i' = \xi_i - \mu$ we see that $\{S_n - n\mu, \mathcal{F}_n\}$ is a martingale.

Example 1.5 (Quadratic Martingale)**.** Suppose now that $\mu = 0$ and $\sigma^2 = \operatorname{var}(\xi_1) < \infty$. In this case $\{S_n^2 - n\sigma^2, \mathcal{F}_n\}_{n\geq 0}$ is a martingale. To show this, note that ξ_{n+1} is independent of $\mathcal{F}_n$,

$$\begin{aligned}\mathbb{E}\left[S_{n+1}^2 - (n+1)\sigma^2|\mathcal{F}_n\right] &= S_n^2 + 2S_n\mathbb{E}\left(\xi_{n+1}|\mathcal{F}_n\right) + \mathbb{E}\left(\xi_{n+1}^2|\mathcal{F}_n\right) - (n+1)\sigma^2 \\ &= S_n^2 + 0 + \sigma^2 - (n+1)\sigma^2 = S_n^2 - n\sigma^2\,.\end{aligned}$$

Example 1.6 (Exponential Martingale)**.** Let $\phi(\theta) = \mathbb{E}\, e^{\theta\xi_i}$ and $Y_i = e^{\theta\xi_i}/\phi(\theta)$ has mean 1. Then let $M_0 = 1$ and

$$M_n = \prod_{i=1}^{n} Y_i = \frac{e^{\theta S_n}}{\phi(\theta)^n}\,,\ \text{for } n \in \mathbb{N}\,.$$

Then $\{M_n, \mathcal{F}_n\}_{n\geq 0}$ is a martingale by Example 1.3.

B Elementary Properties

We turn now to deriving properties of martingales. The following propositttion is elementary, so we omit the proof.

Proposition 1.1. *Let $X = (X_t)_{t\in I}$ and $Y = (Y_t)_{t\in I}$ be real-valued processes adapted to the filtraton $\mathfrak{F} = (\mathcal{F}_t)_{t\in I}$.*

(i) X is a $\mathfrak{F}$-supermartingale if and only if $(-X)$ is a $\mathfrak{F}$-submartingale.

(ii) Let X and Y be $\mathfrak{F}$-martingales and let $a, b \in \mathbb{R}$. Then $(aX + bY)$ is a $\mathfrak{F}$-martingale.

(iii) Let X and Y be $\mathfrak{F}$-supermartingales and $a, b \geq 0$. Then $(aX + bY)$ is a $\mathfrak{F}$-supermartingale.

(iv) Let X and Y be $\mathfrak{F}$-supermartingales. Then $X \wedge Y := (X_t \wedge Y_t)_{t\in I}$ is a $\mathfrak{F}$-supermartingale.

Remark 1.5. It follows from (i) that many statements about supermartingales hold *mutatis mutandis* for submartingales . For example, claim (iii) and (iv) holds for submartingales : if X and Y are $\mathfrak{F}$-submartingales, $(aX + bY)$ is a submartingale, for $a, b \geq 0$ nad $X \vee Y := (X_t \vee Y_t)_{t \in I}$ is a $\mathfrak{F}$-submartingale.

We often do not give the statements both for submartingales and for supermartingales. Instead, we choose representatively one case. Note, however, that those statements that we make explicitly about martingales usually cannot be adapted easily to sub- or super- martingales (such as (ii) in the preceding proposition).

Theorem 1.2. *Let $(X_t, \mathcal{F}_t)_{t \in I}$ be a martingale and let φ be a convex function on $\mathbb{R}$. If*

$$\mathbb{E}\left[\left|\varphi\left(X_t\right)\right|\right] < \infty \text{ for all } t \in I\,,$$

then $\left(\varphi\left(X_t\right), \mathcal{F}_t\right)_{t \in I}$ is a submartingale.

Proof. By assumption, $\left(\varphi\left(X_t\right)\right)_{t \in I}$ is integrable. Jensen's inequality yeilds that, for any $t, s \in I$ with $t > s$,

$$\mathbb{E}\left[\varphi\left(X_t\right)|\mathcal{F}_s\right] \geq \varphi\left(\mathbb{E}\left[X_t|\mathcal{F}_s\right]\right) = \varphi\left(X_s\right)\,.$$

So the desired result holds. ☐

Corollary 1.3. *Let $(X_t, \mathcal{F}_t)_{t \in I}$ be a martingale. Suppose that $p \geq 1$ and $\mathbb{E}\left|X_t\right|^p < \infty$ for all $t \in I$, then $\left(\left|X_t\right|^p, \mathcal{F}_t\right)_{t \in I}$ is a submartingale.*

Corollary 1.4. *Let $(X_t, \mathcal{F}_t)_{t \in I}$ be a submartingale and let φ be a increasing convex function on $\mathbb{R}$. If $\mathbb{E}|\varphi\left(X_t\right)| < \infty$ for all t, then $\left(\varphi\left(X_t\right), \mathcal{F}_t\right)_{t \in I}$ is a submartingale.*

Exercise 1.1. If $\left(X_n, \mathcal{F}_n\right)_{n \geq 0}$ is a supermartingale and $\mathbb{E}\left[X_T\right] \geq \mathbb{E}\left[X_0\right]$ for some $T \in \mathbb{N}$ then $\left(X_n, \mathcal{F}_n, 0 \leq n \leq T\right)$ is a martingale. If there exists a sequence $T_N \to \infty$ with $\mathbb{E}\left[X_{T_N}\right] \geq \mathbb{E}\left[X_0\right]$, then $\left(X_n, \mathcal{F}_n\right)_{n \geq 0}$ is a martingale.

C Discrete Stochastic Integral

From now on, we will foucus on the discrete-time case.

Definition 1.2. A discrete-time process $H = (H_n)_{n\in\mathbb{N}}$ is called **predictable** (or **previsible**) with respect to the filtration $\mathfrak{F} = (\mathcal{F}_n)_{n\in\mathbb{N}_0}$ if

$$H_n \in \mathcal{F}_{n-1} \text{ for every } n \in \mathbb{N}\,.$$

We say that $(H_n, \mathcal{F}_{n-1})_{n\geq 1}$ is predictable for short. In other words, the value of H_n may be predicted (with certainty) from the information available at time $n-1$.

Example 1.7. Let $\xi_1, \xi_2, \ldots$ be i.i.d. random variables, with

$$\mathbb{P}(\xi_1 = 1) = \mathbb{P}(\xi_1 = -1) = \frac{1}{2}\,, \text{ for all } i \in \mathbb{N}\,.$$

Let $\xi = (\xi_i)_{i\in\mathbb{N}}$ and let $\mathfrak{F} = \mathfrak{F}^\xi$ be the filtration generated by ξ.

We interpret ξ_i as the result of a bet that gives a gain or loss of one euro for every euro we put at stake. Just before each gamble we decide how much money we bet. Let H_n be the number of euros to bet in the n th gamble. Clearly, H_n may only depend on the results of the gambles that happened earlier, but not on ξ_m for any $m \geq n$. To put it differently, there must be a function $\phi_n : \{-1,1\}^{n-1} \to \mathbb{R}_+$ such that

$$H_n = \phi_n(\xi_1, \ldots, \xi_{n-1})\,.$$

Hence H is predictable. On the other hand, any predictable H has the form $H_n = \phi_n(\xi_1, \ldots, \xi_{n-1}), n \in \mathbb{N}_0$ for certain functions $\phi_n : \{-1,1\}^{n-1} \to \mathbb{R}_+$. Hence any predictable H is an admissible gambling strategy.

In this subsection, we will be thinking of H_n as the amount of money a gambler will bet at time n. This can be based on the outcomes at times $1, \ldots, n-1$ but not on the outcome at time n.

Once we start thinking of H_n as a gambling system, it is natural to ask how much money we would make if we used it. Let $X = (X_n)_{n\in\mathbb{N}_0}$ be a (possibly unfair) game where X_n is the net amount of money you would have won at time n if you had bet one dollar each time. If you bet according to a *gambling strategy* H then your winnings at time n would be

$$\sum_{k=1}^{n} H_k \left(X_k - X_{k-1}\right) ,$$

since if at time k you have wagered \$3 the change in your fortune would be 3 time that of a person who wagered \$1. Alternatively you can think of X_m is the value of a stock and H_k the number of shares you hold from time $k-1$ to time k.

Definition 1.3. Let $\{X_n, \mathcal{F}_n\}_{n\geq 0}$ be an real-valued adapted process, and let $\{H_n, \mathcal{F}_{n-1}\}_{n\geq 0}$ be a real-valued predictable process. The **discrete stochastic integral** of H with respect to X is the stochastic process $H \cdot X = \{H \cdot X_n\}_{n\geq 0}$ defined by $(H \cdot X)_0 = 0$ and

$$(H \cdot X)_n := \sum_{k=1}^{n} H_k \Delta X_k \qquad \text{for } n \in \mathbb{N}\,.$$

where $\Delta X_k := X_k - X_{k-1}$ for all $k \in \mathbb{N}$. In addition, if X is a martingale, $H \cdot X$ is also called the **martingale transform** of X.

Remark 1.6. An integrabe adapted process $\xi = (\xi_n)_{n\in\mathbb{N}_0}$ is called a *martingale difference sequence* with respect to the filtration $\mathfrak{F} = (\mathcal{F}_n)_{n\in\mathbb{N}_0}$, if for each $n \geq 1$,

$$\mathbb{E}\left(\xi_n | \mathcal{F}_{n-1}\right) = 0\,, \text{a.s.}.$$

We also say that $(\xi_n, \mathcal{F}_n)_{n\geq 1}$ is a martingale difference sequence. The connection between discrete-time martingales and martingale difference sequences is clear: If $(X_n, \mathcal{F}_n)_{n\geq 0}$ is a martingale, then $(\xi_n, \mathcal{F}_n)_{n\geq 0}$ given by $\xi_0 = X_0$, $\xi_n = \Delta X_n, n \geq 1$ is a martingale difference. In turn, if $(\xi_n, \mathcal{F}_n)$ is a martingale difference, then $(X_n, \mathcal{F}_n)_{n\geq 0}$ with $X_n = \xi_0 + \cdots + \xi_n, n \geq 0$ is a martingale.

Example 1.8 (Martingale betting strategy)**.** Let $\xi_1, \xi_2, \ldots$ be independent random variables with

$$\mathbb{P}(\xi_j = 1) = \mathbb{P}(\xi_j = -1) = \frac{1}{2}\,.$$

Let $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and let $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$ for $n \geq 1$. We will refer to such random variables as "coin-tossing" random variables where 1 corresponds to heads and -1 corresponds to tails. Let $M_0 = 0$, and $M_n = \xi_1 + \cdots + \xi_n$ for $n \in \mathbb{N}$. In other words, $(M_n)_{n\geq 0}$ is a simple random walk. We have seen that $(M_n, \mathcal{F}_n)_{n\geq 0}$ is a martingale in Example 1.4. We will consider the following betting strategy.

We start by betting \$1. If we win, we quit; otherwise, we bet \$2 on the next game. If we win the second game, we quit; otherwise we double our bet to \$4 and play. Each time we lose, we double our bet. At the time that we win, we will be ahead \$1. With probability one, we will eventually win the game, so this strategy is a way to beat a fair game. The winnings in this game can be written as

$$(H \cdot M)_n = \sum_{j=1}^{n} H_j \Delta M_j = \sum_{j=1}^{n} H_j \xi_j$$

where the bet $H_1 = 1$ and for $j > 1$

$$H_j = 2^{j-1} 1_{\{\xi_1 = \cdots = \xi_j = -1\}}\,.$$

This is an example of a discrete stochastic integral as in the previous example. By noting that $(H \cdot M)_n = 1$, unless $\xi_1 = \xi_2 = \cdots = \xi_n = -1$ in which case

$$(H \cdot M)_n = -1 - 2^1 - 2^2 - \cdots - 2^{n-1} = -\left(2^n - 1\right)\,.$$

This last event happens with probability $(1/2)^n$, and hence

$$\mathbb{E}\left[(H \cdot M)_n\right] = 1 \cdot \left(1 - 2^{-n}\right) - \left(2^n - 1\right) \cdot 2^{-n} = 0\,, \text{ for each } n \in \mathbb{N}\,.$$

However, we will eventually win which means that with probability one

$$(H \cdot M)_\infty = \lim_{n\to\infty} (H \cdot M)_n = 1$$

and

$$1 = \mathbb{E}\left[(H \cdot M)_\infty\right] > \mathbb{E}\left[(H \cdot M)_n\right] = 0\,.$$

This show that, we will beat the game if we have an infinite amount of time. However, in a finite amount of time, we can't beat it (since $\mathbb{E}(H \cdot M)_n = 0$ for all n). Infact, $((H \cdot M)_n, \mathcal{F}_n)$ is a martingale, i.e., *a fair game.*

The following theorem says, in particular, that we cannot find any locally bounded gambling strategy that transforms a martingale (or, if we are bound to nonnegative gambling strategies, as we are in real life, a supermartingale) into a submartingale. Quite the contrary is suggested by the many invitations to play all kinds of "sure winning systems" in lotteries.

Theorem 1.5 (Stability Theorem)**.** *Let $X = (X_n)_{n\geq 0}$ be an real-valued process adapted to the filtration $\mathfrak{F} = (\mathcal{F}_n)_{n\geq 0}$ with $\mathbb{E}\,|X_0| < \infty$.*

(i) X is a $\mathfrak{F}$-martingale if and only if for any locally bounded [1]$\mathfrak{F}$-predictable process H, $H \cdot X$ is a $\mathfrak{F}$-martingale.

(ii) X is a $\mathfrak{F}$-sub-(super-)martingale if and only if for any non-negative locally bounded $\mathfrak{F}$-predictable process H, $H \cdot X$ is a $\mathfrak{F}$-sub-(super-)martingale.

Proof. We only give the proof of (i), which is similar to the proof of (ii).

Suppose X is a $\mathfrak{F}$-martingale. Clearly, $H \cdot X$ is adapted to $\mathfrak{F}$. Since H is locally bounded, $H \cdot X$ is integrable. Note that for $n \in \mathbb{N}_0$,

$$\begin{aligned}\mathbb{E}\left[(H \cdot X)_{n+1}|\mathcal{F}_n\right] &= (H \cdot X)_n + \mathbb{E}\left[H_{n+1}\left(X_{n+1} - X_n\right)|\mathcal{F}_n\right] \\ &= (H \cdot X)_n + H_{n+1}\mathbb{E}\left(X_{n+1} - X_n|\mathcal{F}_n\right) = (H \cdot X)_n\,.\end{aligned}$$

Thus $H \cdot X$ is a $\mathfrak{F}$-martingale.

On the other hand, fix an $m \in \mathbb{N}$, and let $H_n = 1_{\{n=m\}}$ for all $n \in \mathbb{N}$. Note that $(H \cdot X)_m = (X_m - X_{m-1})$, hence $X_m - X_{m-1}$ is integrable. Since m is

[1] it means that for each n, H_n is bounded.

arbitary and X_0 is integrable, X is integrable. Obeserve that $(H \cdot X)_m = 0$

$$\mathbb{E}\left[(H \cdot X)_m | \mathcal{F}_{m-1}\right] = \mathbb{E}\left(X_m | \mathcal{F}_{m-1}\right) - X_{m-1} = (H \cdot X)_{m-1} = 0\,,$$

so $\mathbb{E}\left(X_m | \mathcal{F}_{m-1}\right) = X_{m-1}$ for all $m \in \mathbb{N}$. Hence X is a $\mathfrak{F}$-martingale. $\square$

Remark 1.7. Although Theorem 1.5 implies that we cannot make money with gambling systems, we can prove many theorems with them.

Recall that a random variable τ taking values in $\mathbb{N}_0 \cup \{\infty\}$, is said to be a *stopping time* if

$$\{\tau = n\} \in \mathcal{F}_n \quad \text{for all } n \in \mathbb{N}_0\,.$$

i.e., the decision to stop at time n must be measurable with respect to the information known at that time.

We will now consider two very special gambling system: bet \$1 at each time $n \le \tau$ then stop playing, or conversely, enter the game until time $n > \tau$ then bet \$1 at each time. Let

$$H_n = 1_{\{n \le \tau\}} \quad \text{for all } n \in \mathbb{N}\,.$$

Clearly, $\{H_n, \mathcal{F}_{n-1}\}$ and $\{1 - H_n, \mathcal{F}_{n-1}\}$ both are predictable, and for each $n \in \mathbb{N}$,

$$(H \cdot X)_n = \sum_{k=1}^{n} 1_{\{k \le \tau\}} \Delta X_k = X_{\tau \wedge n} - X_0\,,$$

$$((1 - H) \cdot X)_n = \sum_{k=1}^{n} 1_{\{k > \tau\}} \Delta X_k = X_n - X_{\tau \wedge n}\,.$$

It follows from Theorem 1.5 that

Theorem 1.6. *Let τ be a stopping time of $\mathfrak{F} = \{\mathcal{F}_n\}_{n \in \mathbb{N}_0}$. Let $(X_n)_{n \in \mathbb{N}_0}$ is a (sub-, super-) martingale with respect to $\mathfrak{F}$. Then*

$$X^\tau := (X_{\tau \wedge n})_{n \in \mathbb{N}_0}\ ,\ X - X^\tau = (X_n - X_{\tau \wedge n})_{n \in \mathbb{N}_0}$$

both are (sub-, super-) martingale with respect to $\mathfrak{F}$. We call X^τ the stopped process.

Exercise 1.2. Show Theorem 1.6 directly.

D Doob's Decomposition for Submartingales

The main result in this subsection is that a submartingale can be decomposed into a sum sonsisting of a *martingale* and a *increasing predictable process*. However, we will consider the decomposition for more stochastic processes.

Let $X = (X_n)_{n\in\mathbb{N}_0}$ be an integrable process adapted to the filtration $\mathfrak{F} = (\mathcal{F}_n)_{n\in\mathbb{N}_0}$. We will decompose X into a sum consisting of a $\mathfrak{F}$-martingale and a $\mathfrak{F}$-predictable process. In other words, we want for all $n \in \mathbb{N}_0$,

$$X_n = M_n + A_n\,,\ \mathbb{E}\left(M_{n+1}|\mathcal{F}_n\right) = M_n\,,\ \text{and}\ \ A_{n+1} \in \mathcal{F}_n\,.$$

So we must have

$$\begin{aligned}\mathbb{E}\left(X_n|\mathcal{F}_{n-1}\right) &= \mathbb{E}\left(M_n|\mathcal{F}_{n-1}\right) + \mathbb{E}\left(A_n|\mathcal{F}_{n-1}\right)\\ &= M_{n-1} + A_n = X_{n-1} - A_{n-1} + A_n\ ,\quad \text{for } n \geq 1\,,\end{aligned}$$

and it follows that

$$A_n - A_{n-1} = \mathbb{E}\left(X_n|\mathcal{F}_{n-1}\right) - X_{n-1}\ ,\quad \text{for } n \geq 1\,. \tag{1.1}$$

Let $A_0 = 0$, then we get

$$A_n = \sum_{m=1}^{n} \mathbb{E}\left(X_m|\mathcal{F}_{m-1}\right) - X_{m-1} = \sum_{m=1}^{n} \mathbb{E}\left(\Delta X_m|\mathcal{F}_{m-1}\right)\ ,\ \text{for } n \geq 1\,.$$

To check that our recipe works, we observe that $A_n \in \mathcal{F}_{n-1}$ for all $n \geq 1$. To prove that $M_n = X_n - A_n$ is a martingale, we note that using (1.1)

$$\begin{aligned}\mathbb{E}\left(M_n|\mathcal{F}_{n-1}\right) &= \mathbb{E}\left(X_n - A_n|\mathcal{F}_{n-1}\right)\\ &= \mathbb{E}\left(X_n|\mathcal{F}_{n-1}\right) - A_n = X_{n-1} - A_{n-1} = M_{n-1}\,,\end{aligned}$$

for all $n \geq 1$, which deduces that $\{M_n, \mathcal{F}_n\}_{n\geq 0}$ is a martingale.

Theorem 1.7 (Doob's Decomposition). *Let $X = (X_n)_{n\geq 0}$ be an integrable process adapted to the filtration $\mathfrak{F} = (\mathcal{F}_n)_{n\geq 0}$. Then there exists a unique decomposition*

$$X = M + A \,,$$

where A is $\mathfrak{F}$-predictable with $A_0 = 0$ and M is a $\mathfrak{F}$-martingale.

Proof. Although the construction above implies the uniqueness, we give a another point to show uniqueness of the decomposition. Let $X = M + A = M' + A'$ be two such decompositions. Then $M - M' = A' - A$ is a predictable martingale, hence $M_n - M'_n = M_0 - M'_0 = 0$ for all n. □

Remark 1.8. Evidently, it follows from (1.1) that X is a submartingale if and only if A is monotone increasing, and for this case, A is called the *increasing process* associated with X, or the *compensator* of X.

Square-integrable martingale The Doob decomposition plays a key role in the study of square-integrable martingales.

Lemma 1.8 (Orthogonality, Conditional Variance). *Let $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ be a square-integrable martingale. Then for $n \geq m \geq 0$,*

$$X_n - X_m \perp L^2(\mathcal{F}_m, \mathbb{P}) \,,$$

and

$$\mathrm{Var}(X_n|\mathcal{F}_m) = \mathbb{E}\left[(X_n - X_m)^2 \,|\mathcal{F}_m\right] = \mathbb{E}\left(X_n^2|\mathcal{F}_m\right) - X_m^2 \quad a.s..$$

Remark 1.9. This is the conditional analogue of

$$\mathrm{Var}(X) = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2$$

and is proved in exactly the same way.

Let $X = \{X_n\}_{n\in\mathbb{N}_0}$ be a square-integrable martingale with respect to the filtration $\mathfrak{F} = \{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$. The application of Doob's decomposition depends on the observation that $\{X_n^2\}_{n\in\mathbb{N}_0}$ is a $\mathfrak{F}$-submartingale. Denote by $\langle X\rangle$ the increasing process associated to X^2, which is called the **quadratic variation process** of X, then for each $n \in \mathbb{N}$,

$$\begin{aligned}\langle X\rangle_n &= \sum_{m=1}^{n} \mathbb{E}\left[\Delta X_m^2 | \mathcal{F}_{m-1}\right] = \sum_{m=1}^{n} \mathbb{E}\left(X_m^2 | \mathcal{F}_{m-1}\right) - X_{m-1}^2 \\ &= \sum_{m=1}^{n} \mathrm{Var}\left(X_m | \mathcal{F}_{m-1}\right) = \sum_{m=1}^{n} \mathbb{E}\left[\left(\Delta X_m\right)^2 | \mathcal{F}_{m-1}\right] ,\end{aligned}$$

and one can see that

$$\mathbb{E}\langle X\rangle_n = \mathbb{E}\left|X_n - X_0\right|^2 = \mathbb{E}X_n^2 - \mathbb{E}X_0^2 .$$

Example 1.9. It is useful to observe that if $S_0 = 0$ and $S_n = \xi_1 + \cdots + \xi_n$, where (ξ_n) is a sequence of independent r.v.'s with $\mathbb{E}\xi_i = 0$ and $\mathbb{E}\xi_i^2 < \infty$, then $S = (S_n)_{n\in\mathbb{N}_0}$ is a square-integrable martingale. The the quadratic variation process of S is given by

$$\langle S\rangle_n = \mathbb{E}S_n^2 = \mathrm{Var}(S_n) , \quad \text{for all } n \geq 1 ,$$

is not random and, indeed, coincides with the variance.

Example 1.10. Let $\xi_1, \xi_2, \ldots$ be independent, square integrable random variables with $\mathbb{E}\xi_n = 1$ for all $n \in \mathbb{N}$. Let $X_n = \Pi_{i=1}^n \xi_i$ for $n \in \mathbb{N}$ and $X_0 = 1$. Then $X = (X_n)_{n\in\mathbb{N}_0}$ is a square integrable martingale with respect to the filtration $\mathfrak{F}$ generated by (ξ_n). For $n \geq 1$,

$$\mathbb{E}\left[\left(\Delta X_n\right)^2 | \mathcal{F}_{n-1}\right] = \mathbb{E}\left[\left(\xi_n - 1\right)^2 X_{n-1}^2 | \mathcal{F}_{n-1}\right] = \mathrm{Var}(\xi_n) X_{n-1}^2 .$$

Hence

$$\langle X\rangle_n = \sum_{m=1}^{n} \mathrm{Var}(\xi_m) X_{m-1}^2 .$$

We see that the square variation process can indeed be a truly random process.

Supplementary materials* If $X = (X_n)_{n\in\mathbb{N}_0}$ and $Y = (Y_n)_{n\in\mathbb{N}_0}$ are square-integrable martingales relative to the filtration $\mathfrak{F} = (\mathcal{F}_n)_{n\in\mathbb{N}_0}$, we put $\langle X, Y\rangle_0 = 0$ and

$$\langle X, Y\rangle_n = \sum_{i=1}^{n} \mathbb{E}\left(\Delta X_m \Delta Y_m | \mathcal{F}_{m-1}\right), \text{ for } n \in \mathbb{N}.$$

One can see that

$$\langle X, Y\rangle_n = \frac{\langle X+Y\rangle_n - \langle X-Y\rangle_n}{4}.$$

It is easily verified that $XY - \langle X, Y\rangle = (X_nY_n - \langle X, Y\rangle_n)_{n\geq 0}$ is a $\mathfrak{F}$-martingale, and hence

$$XY = \big(XY - \langle X, Y\rangle\big) + \langle X, Y\rangle$$

gives the Doob's decomposition of XY.

Example 1.11. In the case when $X_n = \xi_1 + \cdots + \xi_n, Y_n = \eta_1 + \cdots + \eta_n$, where (ξ_n) and (η_n) are sequences of independent random variables with $\mathrm{E}\xi_i = \mathrm{E}\eta_i = 0$ and $\mathrm{E}\xi_i^2 < \infty, \mathrm{E}\eta_i^2 < \infty$, the variable $\langle X, Y\rangle_n$ is given by

$$\langle X, Y\rangle_n = \sum_{i=1}^{n} \mathrm{Cov}\left(\xi_i, \eta_i\right)$$

1.2 Optional Sampling Theorem

In this section, we will dsicuss preservation of martingale property under a random time change. If $X = (X_n)_{n\geq 0}$ is a martingale or a submartingale with respect to the filtration $\mathfrak{F} = (\mathcal{F}_n)_{n\geq 0}$, then we have

$$\mathbb{E}X_0 = \mathbb{E}X_n \quad \text{or} \quad \mathbb{E}X_0 \leq \mathbb{E}X_n$$

for every n. Is this property preserved if the time n is replaced by a finite stopping time ? Unfortunately, in general case, the answer is no, see Example 1.13, 1.14. The following basic theorem describes the "typical" situation, in which, in particular,

$$\mathbb{E}X_0 = \mathbb{E}X_T \quad \text{or} \quad \mathbb{E}X_0 \leq \mathbb{E}X_T. \tag{1.2}$$

A Bounded stopping times

Lemma 1.9. *Suppose T is a stopping time of the filtration $\mathfrak{F} = (\mathcal{F}_n)_{n\in\mathbb{N}_0}$, and T is bounded; i.e., $T \le k$ for smoe $k \in \mathbb{N}$.*

(i) If (X_n) is a $\mathfrak{F}$-martingale, then $\mathbb{E}X_0 = \mathbb{E}X_T = \mathbb{E}X_k$. Moreover, martingale property is preserved at T:

$$\mathbb{E}(X_k|\mathcal{F}_T) = X_T \quad a.s..$$

(ii) If (X_n) is a $\mathfrak{F}$-submartingale, then $\mathbb{E}X_0 \le \mathbb{E}X_T \le \mathbb{E}X_k$. Moreover, submartingale property is preserved at T:

$$\mathbb{E}(X_k|\mathcal{F}_T) \ge X_T \quad a.s.. \tag{1.3}$$

Proof. Evidently, (ii) implies (i), so we only show (ii). By Theorem 1.6, X^T and $X - X^T$ both are $\mathfrak{F}$-submartingale, so

$$\begin{aligned}&\mathbb{E}X_0 = \mathbb{E}X_{T\wedge 0} \le \mathbb{E}X_{T\wedge k} = \mathbb{E}X_T \quad \text{and,}\\ &\mathbb{E}X_k - \mathbb{E}X_T = \mathbb{E}X_k - \mathbb{E}X_{T\wedge k} \ge \mathbb{E}(X_0 - X_{T\wedge 0}) = 0\,.\end{aligned}$$

Thus $\mathbb{E}X_0 \le \mathbb{E}X_T \le \mathbb{E}X_T$. To prove (1.3), we have to show that for any $A \in \mathcal{F}_T$, $\mathbb{E}X_T 1_A \le \mathbb{E}X_k 1_A$, and hence it suffices to show that

$$\mathbb{E}X_T 1_A + \mathbb{E}X_k 1_{A^c} \le \mathbb{E}X_k\,.$$

Define a random time S by $S \coloneqq T1_A + k1_{A^c}$. Clearly, S is a stopping time of $\mathfrak{F}$ and bounded by k, thus

$$\mathbb{E}X_S = \mathbb{E}X_T 1_A + \mathbb{E}X_k 1_{A^c} \le \mathbb{E}X_k\,. \qquad \square$$

Remark 1.10. In fact there is a direct way to show (1.3). Take any $A \in \mathcal{F}_T$. Observe that for any $0 \le i \le k$, $A \cap \{T = i\} \in \mathcal{F}_i$, hence

$$\mathbb{E}X_T 1_A 1_{\{T=i\}} = \mathbb{E}X_i 1_A 1_{\{T=i\}} \le \mathbb{E}X_k 1_A 1_{\{T=i\}}\,.$$

Then we get

$$\mathbb{E}\left[X_T 1_A\right] \leq \mathbb{E}\left[X_k 1_A\right] ,$$

and it follows that $\mathbb{E}(X_k|\mathcal{F}_T) \geq X_T$ a.s.. However, if we want to show $\mathbb{E}(X_T|\mathcal{F}_0) \geq X_0$ a.s., this arguement does not work. But the method used in the proof of Lemma 1.9 works.

Exercise 1.3. Let $X = (X_n)_{n\geq 0}$ be a integrable process adapted to the filtration $\mathfrak{F} = (\mathcal{F}_n)_{n\geq 0}$. Then X is a $\mathfrak{F}$-martingale if and only if for any bounded stopping time T of $\mathfrak{F}$, $\mathbb{E}X_T = \mathbb{E}X_0$.

Lemma 1.10. *Suppose T, S are two bounded stopping times of $\mathfrak{F} = (\mathcal{F}_n)_{n\in\mathbb{N}_0}$ and $S \leq T \leq k$, where $k \in \mathbb{N}$ is a constant.*

(i) If (X_n) is a $\mathfrak{F}$-martingale, then $\mathbb{E}X_0 = \mathbb{E}X_S = \mathbb{E}X_T = \mathbb{E}X_k$. Moreover, martingale property is preserved at time T and S:

$$\mathbb{E}(X_T|\mathcal{F}_S) = X_S \quad a.s..$$

(ii) If (X_n) is a $\mathfrak{F}$-submartingale, then $\mathbb{E}X_0 \leq \mathbb{E}X_S \leq \mathbb{E}X_T \leq \mathbb{E}X_k$. Moreover, submartingale property is preserved at time T and S:

$$\mathbb{E}\left(X_T|\mathcal{F}_S\right) \geq X_S \quad a.s.. \tag{1.4}$$

Proof. We have only to show (ii). In order to prove that $\mathbb{E}X_S \leq \mathbb{E}X_T$, define a $\mathfrak{F}$-predictable process H by letting $H_n := 1_{\{n\leq T\}} - 1_{\{n\leq S\}}$ for $n \geq 1$. Clearly H is non-negative and bounded, so

$$H \cdot X = X^T - X^S$$

is a $\mathfrak{F}$-submartingale by Theorem 1.5. Thus $\mathbb{E}X_S \leq \mathbb{E}X_T$. To show (1.4), we have to show that for any $A \in \mathcal{F}_S$, $\mathbb{E}X_S 1_A \leq \mathbb{E}X_T 1_A$. Then it suffices to prove that

$$\mathbb{E}X_S 1_A + \mathbb{E}X_T 1_{A^c} \leq \mathbb{E}X_T \, .$$

Let N defined by $N := S1_A + T1_{A^c}$. Clearly, N is a stopping time of $\mathfrak{F}$ and bounded by T. In fact, for any $n \in \mathbb{N}_0$,

$$\begin{aligned}\{N = n\} &= \big(\{S = n\} \cap A\big) \cup \big(\{T = n\} \cap A^c\big) \\ &= \big(\{S = n\} \cap A\big) \cup \big(\{T = n\} \cap \{S \le n\} \cap A^c\big) \in \mathcal{F}_n \,.\end{aligned}$$

Then the desired result follows from $\mathbb{E}X_N \le \mathbb{E}X_T$. □

Example 1.12. Let $X = (X_n)$ be a $\mathfrak{F}$- (sub-, super-)martingale , and assume (T_n) is a monotone increasing sequence of bounded stopping times. Then (X_{T_n}) is a (sub-, super-) martingale with respect to the filtration $(\mathcal{F}_{T_n})$.

In particular, for any stopping time T of $\mathfrak{F}$, the stopped process X^T is a (sub-, super-)martingale with respect to to both $\mathfrak{F}$ and $\mathfrak{F}^T := (\mathcal{F}_{n\wedge T})_{n\in\mathbb{N}_0}$.

Theorem 1.11 (Optinal Sampling Theorem I)**.** *Let $X = (X_n)$ be a (sub-)martingale with respect to the filtration $\mathfrak{F} = (\mathcal{F}_n)_{n\in\mathbb{N}_0}$, let T, S be two bounded stopping times of $\mathfrak{F}$. Then*

$$\mathbb{E}(X_T|\mathcal{F}_S) \ (\ge) = \ X_{T\wedge S} \quad a.s.,$$

in other words, $\mathbb{E}(X_T|\mathcal{F}_S)$ $(\ge) =$ X_S a.s. on $\{T \ge S\}$: The (sub-)martingale property is preserved.

Proof. It suffices to show the submartingale case. Take any $A \in \mathcal{F}_S$. We need to show that $\mathbb{E}X_{S\wedge T}1_A \le \mathbb{E}X_T1_A$, so it suffices to prove that

$$\mathbb{E}X_{S\wedge T}1_A + \mathbb{E}X_T1_{A^c} \le \mathbb{E}X_T \,.$$

Let N defined by $N := (S \wedge T)1_A + T1_{A^c} = S1_{A\cap\{S\le T\}} + T1_{(A\cap\{S\le T\})^c}$. We claim that, N is a stopping time of the filtration $\mathfrak{F}$ and bounded by T. In fact, for any positive integer n,

$$\begin{aligned}\{N = n\} &= \big(\{S = n\} \cap A \cap \{S \le T\}\big) \cup \big(\{T = n\} \cap (A \cap \{S \le T\})^c\big) \\ &= \big(\{S = n\} \cap A \cap \{S \le n\}\big) \cup \big(\{T = n\} \cap (A \cap \{S \le n\})^c\big) \in \mathcal{F}_n \,.\end{aligned}$$

Then the desired result follows from $\mathbb{E}X_N \le \mathbb{E}X_T$ by Theorem 1.10. □

Counterexamples The first inequality in Theorem 1.9 need not hold for unbounded stopping times. We will give two important counterexamples, which should be keep in mind as you read the rest of this chapter.

Example 1.13. Let $\{S_n\}$ be a simple random walk with $S_0 = 1$ and let $T = \inf\{n \geq 0 : S_n = 0\}$. As we konw, the simple random walk is ecurrent, hence $\mathbb{P}(T < \infty) = 1$. Then

$$\mathbb{E}S_T = 0 < 1 = \mathbb{E}\, S_0 \,.$$

Later, we will give conditions that guarantee $\mathbb{E}X_0 \leq \mathbb{E}X_T$ for unbounded T.

Example 1.14. Consider the martingale betting strategy in Example 1.8. Let

$$T \coloneqq \inf\{n \geq 1 : \xi_n = 1\} = \inf\{n \geq 1 : (H \cdot X)_n = 1\}\,,$$

and B-C lemma implies $\mathbb{P}(T < \infty) = 1$. But

$$\mathbb{E}(H \cdot X)_T = 1 > 0 = \mathbb{E}\,(H \cdot X)_0 \,.$$

B Finite Stopping Times

Often one does want to conclude the (1.2) for unbounded stopping times, so it is useful to give conditions under which it holds. Let's try to derive the equality and see what conditions we need to impose. Suppose that $X = (X_n)_{n \in \mathbb{N}_0}$ is a martingale and T are stopping times, with respect to the filtration $\mathfrak{F} = (\mathcal{F}_n)_{n \in \mathbb{N}_0}$. In order that X_T makes sense, we will firstly assume that we stop, i.e. $\mathbb{P}(T < \infty) = 1$. Note that for each $n \in \mathbb{N}$,

$$\mathbb{E}X_0 = \mathbb{E}X_{n \wedge T} = \mathbb{E}\, X_T 1_{\{T \leq n\}} + \mathbb{E}\, X_n 1_{\{T > n\}} \,.$$

Secondly, we suppose that $\mathbb{E}\,|X_T| < \infty$, then by domainted convergence theorem,

$$\lim_{n \to \infty} \mathbb{E}X_T 1_{\{T \leq n\}} = \mathbb{E}X_T \,.$$

In Example 1.13 and Example 1.14, this condition did not cause a problem since $S_T = 0$ and $(H \cdot X)_T = 1$. Thus, if we can show that

$$\lim_{n\to\infty} \mathbb{E}\,|X_n|\,1_{\{T>n\}} = \lim_{n\to\infty} \mathbb{E}\,|X_{T\wedge n} - X_T| = 0\,,$$

i.e., $X_{T\wedge n} \to X_T$ in L^1, then we have

$$\mathbb{E}X_T = \mathbb{E}X_0\,.$$

In fact, we can deduce a more stronger result.

Theorem 1.12 (Optional Sampling Theorem II)**.** *Let $X = (X_n)_{n\in\mathbb{N}_0}$ be a (sub-)martingale with respect to the filtration $\mathfrak{F} = (\mathcal{F}_n)_{n\in\mathbb{N}_0}$. Let T be an a.s. finite stopping time of $\mathfrak{F}$ for which $\mathbb{E}|X_T| < \infty$. If $X_{T\wedge n} \to X_T$ in L^1, i.e.,*

$$\lim_{n\to\infty} \mathbb{E}\,|X_n|\,1_{\{T>n\}} = 0\,, \tag{1.5}$$

then for any a.s. finite stopping time S of $\mathfrak{F}$ so that $\mathbb{E}|X_S| < \infty$,

$$\mathbb{E}(X_T|\mathcal{F}_S)\ (\geq) =\ X_{T\wedge S} \quad a.s.,$$

in other words, on $\{T \geq S\}$, $\mathbb{E}(X_T|\mathcal{F}_S)\ (\geq) =\ X_S$ a.s.: The (sub-)martingale property is preserved.

Remark 1.11. Observe that $X_{T\wedge n} \to X_T$ a.s., so $X_{T\wedge n} \to X_T$ in L^1 if and only if the stopped process $(X_{T\wedge n})$ is uniformly integrable. We assume that $\mathbb{E}|X_S| < \infty$ to guatantee $\mathbb{E}|X_{T\wedge S}| < \infty$, since we have $|X_{T\wedge S}| \leq |X_T| + |X_S|$. In fact it suffices ensure that $\mathbb{E}|X_{T\wedge S}| < \infty$. As we will see later, $(X_{T\wedge n})$ is uniformly integrable implies that $\mathbb{E}|X_{T\wedge S}| < \infty$ for any stopping times S of $\mathfrak{F}$.

Proof. We only prove the submartingale case. It suffices to show that for any $A \in \mathcal{F}_S$, $\mathbb{E}\,X_T 1_A \geq \mathbb{E}\,X_{T\wedge S}1_A$, and this is equivalent to

$$\mathbb{E}\left(X_T 1_{\{T>S\}}1_A\right) \geq \mathbb{E}\left(X_S 1_{\{T>S\}}1_A\right).$$

Since $\{S < T\} \subset \{S < \infty\}$, we only need to show for each $n \in \mathbb{N}_0$,

$$\mathbb{E}\left(X_T 1_{\{T>S\}} 1_A 1_{\{S=n\}}\right) \geq \mathbb{E}\left(X_S 1_{\{T>S\}} 1_{\{S=n\}} 1_A\right).$$

Let $B = A \cap \{S = n\} \in \mathcal{F}_n$, then the previous equality is

$$\mathbb{E}\left(X_T 1_{\{T>n\}} 1_B\right) \geq \mathbb{E}\left(X_n 1_{\{T>n\}} 1_B\right).$$

Observe that this equality holds if we suppose T is a bounded, since $E(X_T|\mathcal{F}_n) \geq X_n$ on $\{T > n\}$ by Theorem 1.11. Thus we shall use a sequence of bounded stopping times, $\{T \wedge m\}_{m \geq 1}$ to approximate the unbounded finite stopping times T. Trivially for any $m \geq n$,

$$\mathbb{E}\left(X_{T\wedge m} 1_{\{T>n\}} 1_B\right) = \mathbb{E}\left(X_{T\wedge m} 1_{\{T\wedge m>n\}} 1_B\right) \geq \mathbb{E}\left(X_n 1_{\{T>n\}} 1_B\right). \tag{1.6}$$

Since $X_{T\wedge m} \to X_T$ in L^1, the desired result follows. □

Remark 1.12. In fact, in the case of submartingale, we can ask a weaker condition. Observe that it follows from (1.6) that

$$\mathbb{E}\left(X_T 1_{\{n<T\leq m\}} 1_B\right) + \mathbb{E}\left(X_m 1_{\{T>m\}} 1_B\right) \geq \mathbb{E}\left(X_n 1_{\{T>n\}} 1_B\right). \tag{1.7}$$

Note that $\mathbb{E}\left(X_m 1_{\{T>m\}} 1_B\right) \leq \mathbb{E}\, X_m^+ 1_{\{T>m\}}$. Letting $m \to \infty$, by domainted convergence theorem, we have

$$\mathbb{E}\left(X_T 1_{\{T>n\}} 1_B\right) + \liminf_{m\to\infty} \mathbb{E}\, X_m^+ 1_{\{T>m\}} \geq \mathbb{E}\left(X_n 1_{\{T>n\}} 1_B\right).$$

So we only need to ask

$$\liminf_{m\to\infty} \mathbb{E}\, X_m^+ 1_{\{T>m\}} = 0\,.$$

Corollary 1.13. *Suppose that $X = (X_n)_{n\in\mathbb{N}_0}$ is a $\mathfrak{F}$-supermartingale, T is a a.s. finite stopping time so that $\mathbb{E}|X_T| < \infty$. If*

$$\liminf_{n\to\infty} \mathbb{E}X_n^- 1_{\{T>n\}} = 0\,.$$

then for any a.s. finite stopping time S with respect to $\mathfrak{F}$ and $\mathbb{E}|X_S| < \infty$,

$$\mathbb{E}(X_T|\mathcal{F}_S) \le X_{T\wedge S} \quad a.s.,$$

in other words, on $\{T \ge S\}$, $\mathbb{E}(X_T|\mathcal{F}_S) \le X_S$ a.s.: The supermartingale property is preserved.

C Uniform Integrability and Optional Sampling*

In this subsection, some theorems in Section 1.4 and Section 1.5 are need. Recall that in Theorem 1.12, we suppose X a $\mathfrak{F}$-(sub-)martingale, T is a finite stopping time, and $X_{T\wedge n} \to X_T$ in L^1. In fact, since $X^T = (X_{T\wedge n})_{n\ge 0}$ is also a $\mathfrak{F}$-(sub-)martingale, it follows from Lemma 1.27 that we only need to ensure that $X^T = (X_{n\wedge T})$ is uniformly integrable. Moreover, the condition that T is finite is not needed. If $(X_{n\wedge T})$ is uiformly integrable $\mathfrak{F}$-(sub-)martingale, define

$$X_T(\omega) := \limsup_{n\to\infty} X_{n\wedge T(\omega)}(\omega) \quad \text{for every } \omega \in \Omega\,. \tag{1.8}$$

Then X_T makes sense. Besides, on $\{T < \infty\}$, $X_T(\omega) = X_{T(\omega)}(\omega)$, the definition is the same as before; X_T is integrable and $X_{n\wedge T} \to X_T$ a.s. and in L^1 by Lemma 1.27. The following lemma implies that $(X_{n\wedge T\wedge S})_{n\ge 0}$ is uniformly integrable for any $\mathfrak{F}$-stopping times S. Also, $X_{T\wedge S}$ makes sense and $\mathbb{E}|X_{T\wedge S}| < \infty$.

Lemma 1.14. *Let $X = (X_n)_{n\in\mathbb{N}_0}$ be a uniformly integrable $\mathfrak{F}$-submartingale. Let T be any stopping time of $\mathfrak{F}$. Then $X^T = (X_{n\wedge T})_{n\in\mathbb{N}_0}$ is a uniformly integrable $\mathfrak{F}$-submartingale.*

Proof. Note that for any $n \in \mathbb{N}_0$,

$$\mathbb{E}\,|X_{n\wedge T}| = 2\mathbb{E}X^+_{n\wedge T} - \mathbb{E}X_{n\wedge T} \le 2\mathbb{E}X^+_{n\wedge T} - \mathbb{E}X_0\,.$$

Since (X^+_n) is a submartingale and $n \wedge T$ is a bounded topping time, so

$$\mathbb{E}X^+_{n\wedge T} \le \mathbb{E}X^+_n \le \sup_{n\in\mathbb{N}_0} \mathbb{E}|X_n|\,.$$

Thus

$$\sup_{n\in\mathbb{N}_0} \mathbb{E}\,|X_{n\wedge T}| \le 2 \sup_{n\in\mathbb{N}_0} \mathbb{E}|X_n| - \mathbb{E}X_0 < \infty\,.$$

The convergence theorem for submartingales gives $X_{T\wedge n} \to X_T$ (here X_T is defined by (1.8)) a.s. and $\mathbb{E}|X_T| < \infty$. With this established, the rest is easy. For any $\lambda > 0$, we have

$$\begin{aligned}
&\mathbb{E}|X_{T\wedge n}|1_{\{|X_{T\wedge n}|>\lambda\}} \\
&= \mathbb{E}|X_T|1_{\{|X_T|>\lambda, T\le n\}} + \mathbb{E}|X_n|1_{\{|X_n|>\lambda, T>n\}} \\
&\le \mathbb{E}|X_T|1_{\{|X_T|>\lambda\}} + \sup_{n\in\mathbb{N}_0} \mathbb{E}|X_n|1_{\{|X_n|>\lambda\}}\,.
\end{aligned}$$

Since $\mathbb{E}\,|X_T| < \infty$ and $(X_n)_{n\ge 0}$ is uniformly integrable, the desired result follows. ☐

Therefore, we have the following theorems :

Theorem 1.15 (Optional Sampling Theorem III). *Suppose $X = (X_n)_{n\in\mathbb{N}_0}$ is a (sub-)martingale with respect to the filtration $\mathfrak{F} = (\mathcal{F}_n)_{n\in\mathbb{N}_0}$. Let T be a stopping time of $\mathfrak{F}$. If the stopped process X^T is uniformly integrable, then, for any stopping time S of $\mathfrak{F}$, we have $\mathbb{E}|X_{T\wedge S}| < \infty$, and*

$$\mathbb{E}(X_T|\mathcal{F}_S)\ (\ge) =\ X_{T\wedge S} \quad a.s..$$

in other words, $\mathbb{E}(X_T|\mathcal{F}_S)\ (\ge) =\ X_S$ a.s. on $\{T \ge S\}$: The (sub-)martingale property is preserved.

Altough we have discussed the sufficiency for this theorem at the begining of this subsection, we shall give a rigorous proof for the submartingale case.

Proof. Since X^T is uniformly bounded, X_T makes sense and $\mathbb{E}|X_T| < \infty$. By Lemma 1.14, $\mathbb{E}|X_{S\wedge T}| < \infty$. Now, it suffices to show that for any $A \in \mathcal{F}_S$, denoted by B the intersection $A \cap \{S = n\} \in \mathcal{F}_n$,

$$\mathbb{E}\left(X_T 1_{\{T>n\}} 1_B\right) \ge \mathbb{E}\left(X_n 1_{\{T>n\}} 1_B\right).$$

Note that for any $m \geq n$, we have

$$\mathbb{E}\left(X_{T\wedge m}1_{\{T>n\}}1_B\right) = \mathbb{E}\left(X_{T\wedge m}1_{\{T\wedge m>n\}}1_B\right) \geq \mathbb{E}\left(X_n 1_{\{T>n\}}1_B\right).$$

Since $X_{T\wedge m} \to X_T$ in L^1, the desired result follows. □

For uniformly integrable $\mathfrak{F}$-submartingale $X = (X_n)_{n\in\mathbb{N}_0}$, define X_∞ by $X_\infty := \limsup_n X_n$, then $X_n \to X_\infty$ a.s. and in L^1. Thus for any $\mathfrak{F}$-stopping time T, X_T is well-defined. Besides, this definition here and in (1.8) coincide.

Corollary 1.16. *Let $X = (X_n)$ be a uniformly integrable (sub-)martingale with respect to $\mathfrak{F} = (\mathcal{F}_n)_{n\in\mathbb{N}_0}$. Let T and S are two $\mathfrak{F}$-stopping times. Then X_T, X_S are integrable, and*

$$\mathbb{E}(X_T|\mathcal{F}_S)\ (\geq) =\ X_{T\wedge S} \quad a.s.,$$

in other words, $\mathbb{E}(X_T|\mathcal{F}_S)\ (\geq) =\ X_S$ a.s. on $\{T \geq S\}$: The (sub-)martingale property is preserved.

Corollary 1.17. *Let $X = (X_n)_{n\in\mathbb{N}_0}$ be a uniformly integrable (sub-,super-) martingale with respect to $\mathfrak{F} = (\mathcal{F}_n)$ and let $T_1 \leq T_2 \leq \ldots$ be stopping times of $\mathfrak{F}$. Then $(X_{T_n})_{n\in\mathbb{N}}$ is a uniformly integrable (sub-,super-) martingale with respect to to the filtration $(\mathcal{F}_{T_n})$.*

Theorem 1.18 (Optional Sampling Theorem IV)**.** *Let $X = (X_n)_{n\in\mathbb{N}_0}$ be a (sub-) martingale with respect to the filtration $\mathfrak{F} = (\mathcal{F}_n)_{n\in\mathbb{N}_0}$ with*

$$\mathbb{E}\left(|\Delta X_n| \mid \mathcal{F}_{n-1}\right) \leq C \quad \text{for all } n \geq 1\,.$$

If T is a stopping time of $\mathfrak{F}$ with $\mathbb{E}\,T < \infty$, then the stopped process X^T is uniformly integrable, and hence for any stopping times S, we have $\mathbb{E}|X_{T\wedge S}| < \infty$, and

$$\mathbb{E}(X_T|\mathcal{F}_S)\ (\geq) =\ X_{T\wedge S} \quad a.s.,$$

in other words, $\mathbb{E}(X_T|\mathcal{F}_S)\ (\geq) =\ X_S$ a.s. on $\{T \geq S\}$: The (sub-)martingale property is preserved.

Proof. It suffices to show that X^T is uniformly integrable. Firstly, $\mathbb{E}\,T < \infty$ implies $T < \infty$ a.s.. Observe that

$$\sup_{n\in\mathbb{N}_0} |X_{T\wedge n}| \leq |X_0| + \sum_{k=1}^{\infty} |\Delta X_k|\, 1_{\{k\leq T\}}\,.$$

So, to prove uniform integrability, it suffices to show that the RHS has finite expectation. Now, $\{k \leq T\} \in \mathcal{F}_{k-1}$, so

$$\mathbb{E}\,|\Delta X_k|\, 1_{\{k\leq T\}} = \mathbb{E}\left[\mathbb{E}(|\Delta X_k|\, 1_{\{k\leq T\}}|\mathcal{F}_{k-1})\right] \leq C\,\mathbb{P}(T\geq k)\,,$$

and

$$\mathbb{E}\sum_{k=1}^{\infty} |\Delta X_k|\, 1_{\{k\leq T\}} \leq C\sum_{k=0}^{\infty}\mathbb{P}(T\geq k) = C\,\mathbb{E}T < \infty\,.$$

We have completed the proof. □

Exercise 1.4. Suppose that $X = (X_n)_{n\in\mathbb{N}_0}$ is a non-negative $\mathfrak{F}$-supermartingale. Use Fatou's lemma to show that, for any stopping time T , we have

$$\mathbb{E}X_T \leq \mathbb{E}X_0\,,$$

where $X_T = X_\infty := \limsup_n X_n$ a.s. on $\{T = \infty\}$.

The following example is a direct application of the optional sampling theorem reading.

Example 1.15. Let $X = (X_n)_{n\in\mathbb{N}_0}$ be a uniformly integrable submartingale with respect to $\mathfrak{F}$. Then the family

$$\{X_T : T \text{ is stopping time of } \mathfrak{F}\}$$

is uniformly integrable.

To see this, since the martingale $(X_n)_{n\in\mathbb{N}_0}$ is uniformly integrable, it follows from Corollary 1.16 that for any stopping time T,

$$\mathbb{E}(X_\infty|\mathcal{F}_T) = X_T \quad \text{a.s..}$$

As we have learned in the course of measure theory, since $\mathbb{E}|X_\infty| \le \infty$,

$$\{\mathbb{E}(X_\infty|\mathcal{G}) : \mathcal{G} \text{ is a sub-sigma-field of } \mathcal{F}\}$$

is an uniformly integrable family. Then the desired result follows.

1.3 Applications(I) : Random walks

Let $(\xi_n)_{n\ge1}$ be a sequence of i.i.d. r.v.'s. Let $S_n = S_0 + \xi_1 + \cdots + \xi_n$ for $n \in \mathbb{N}$, where S_0 is some constant x. Let $\mathfrak{F} = (\mathcal{F}_n)_{n\ge0}$ defined by $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$ for $n \ge 1$ and $\mathcal{F}_0 = \{\emptyset, \Omega\}$. We call the process $(S_n)_{n\ge0}$ a random walk starting at x. *Symmetric simple random walk* refers to the special case in which

$$\mathbb{P}(\xi_1 = 1) = \mathbb{P}(\xi_1 = -1) = \frac{1}{2}\,.$$

Asymmetric simple random walk refers to the special case in which

$$\mathbb{P}(\xi_1 = 1) = p\ ,\ \mathbb{P}(\xi_1 = -1) = q$$

with $p, q \in (0,1)$, $p \ne q$ and $p + q = 1$. We will now derive some result by using the three martingales from Example 1.4, 1.5, 1.6.

Theorem 1.19. *Let $(S_n)_{n\ge0}$ be symmetric simple random walk with $S_0 = x \in \mathbb{Z}$. For $j \in \mathbb{Z}$, set $T_j = \inf\{n \ge 0 : S_n = j\}$. Let $\tau_{(a,b)} = T_a \wedge T_b = \inf\{n \ge 0 : S_n \notin (a,b)\}$, then*

(i) For the hitting probability, we have

$$\mathbb{P}_x\left(S_{\tau_{(a,b)}} = a\right) = \frac{b-x}{b-a}\ ,\ \mathbb{P}_x\left(S_{\tau_{(a,b)}} = b\right) = \frac{x-a}{b-a}\,.$$

(ii) For the mean hitting time, we have

$$\mathbb{E}_x\tau_{(a,b)} = (b-x)(x-a)\,.$$

Remark 1.13. This theorem can be showed by the techniques in Markov chian, however, we will give a proof using martingale method.

Proof. In this proof we write $\tau = \tau_{(a,b)}$ for short. To see that $\mathbb{P}_x(\tau < \infty) = 1$, note that if we have $(b-a)$ consecutive steps of size $+1$ we will exit the interval. From this, by induction, it follows that

$$\mathbb{P}_x(\tau > k(b-a)) \leq \left(1 - \frac{1}{2^{b-a}}\right)^k , \text{ for } k \in \mathbb{N} .$$

which implies that $\mathbb{E}_x\tau < \infty$.

(i). Clearly $\mathbb{E}_x |S_\tau| < \infty$ and $(S_{n\wedge\tau})$ are uniformly integrable,

$$\mathbb{E}_x S_\tau = a\,\mathbb{P}_x(S_\tau = a) + b\,\mathbb{P}_x(S_\tau = b) = \mathbb{E}_x S_0 = x .$$

Note that

$$\mathbb{P}_x(S_\tau = a) + \mathbb{P}_x(S_\tau = b) = 1 ,$$

we can slove the equations systerm.

(ii). The second result is an immediate consequence of the first. Note that $(S^2_{\tau\wedge n} - (\tau\wedge n))_{n\geq 0}$ is domainted by an integrable r.v., thus is uniformly integrable. So

$$\mathbb{E}_x(S^2_\tau - \tau) = \mathbb{E}S^2_0 = x^2.$$

Using the result of (i),

$$\mathbb{E}_x\tau = \mathbb{E}_x S^2_\tau - x^2 = a^2\frac{b-x}{b-a} + b^2\frac{x-a}{b-a} = (b-x)(x-a) ,$$

which completes the proof. □

Remark 1.14. Given $n \in \mathbb{N}$, we have $\mathbb{P}_0(T_n < T_{-m}) = m/(m+n)$ for all $m \in \mathbb{N}$. Note that

$$\{T_n < \infty\} = \bigcup_{m=1}^{\infty} \{T_n < T_{-m}\}$$

Thus we get $\mathbb{P}_0(T_n < \infty) = 1$.

As we can see, we deduce Theorem 1.19 from

$$\mathbb{E}_x S_{\tau_{(a,b)}} = x \quad \text{and} \quad \mathbb{E}_x S^2_{\tau_{(a,b)}} = \mathbb{E}_x \tau_{(a,b)} \times 1 = \mathbb{E}_x \tau_{(a,b)}\, \mathbb{E}\xi_1^2\,.$$

In fact, these property holds for more general random walks and stopping times.

Theorem 1.20 (Wald's Identities). *Let $(S_n)_{n\geq 0}$ be a random walk starting at x. Suppose that $\mathbb{E}|\xi_1| < \infty$ and let $\mu = \mathbb{E}\xi_1$. Then, for any stopping times τ with respect to $\mathfrak{F}$ with $\mathbb{E}\,\tau < \infty$,*

$$\mathbb{E}S_\tau = x + \mu\,\mathbb{E}\tau\,. \tag{1.9}$$

If in addition $\mathbb{E}\,\xi_1^2 < \infty$, let $\sigma^2 = \mathrm{Var}(\xi_1)$, then

$$\mathbb{E}(S_\tau - \mu\tau)^2 = x^2 + \sigma^2\,\mathbb{E}\tau\,. \tag{1.10}$$

Proof. Firstly, since $(S_n - n\mu, \mathcal{F}_n)_{n\geq 0}$ is a martingale, note that

$$\mathbb{E}(|\Delta S_n||\mathcal{F}_{n-1}) = \mathbb{E}(|\xi_n||\mathcal{F}_{n-1}) = \mathbb{E}|\xi_1| < \infty\,,$$

by Theorem 1.18 we have

$$\mathbb{E}(S_\tau - \mu\tau) = \mathbb{E}S_0 = x\,.$$

The (1.9) follows. On the one hand, to show (1.10), without loss of generality, we assume that $\mu = 0$. Then

$$(S_n^2 - n\sigma^2, \mathcal{F}_n)_{n\geq 0}$$

is a martingale, so

$$\mathbb{E}\,S^2_{\tau\wedge n} = x^2 + \sigma^2\,\mathbb{E}(\tau\wedge n)\,, \text{ for all } n\,. \tag{1.11}$$

Letting $n\to\infty$, by Fatou's lemma we get $\mathbb{E}S_\tau^2 \leq \sigma^2\,\mathbb{E}\tau$. On the orther hand, it suffices to show that

$$\mathbb{E}S^2_{\tau\wedge n} \leq \mathbb{E}S^2_\tau \quad \text{for all } n\,.$$

Observe form (1.11) that

$$\sup_{n\in\mathbb{N}_0} \mathbb{E}S_{\tau\wedge n}^2 < \infty\,,$$

so $(|S_{\tau\wedge n}|)_{n\geq 0}$ is uniformly integrable. By Theorem 1.12, we have ,

$$\mathbb{E}(|S_\tau||\mathcal{F}_n) \geq |S_{\tau\wedge n}| \quad \text{for all } n\,.$$

Thanks to Jensen's inequality, we have

$$\mathbb{E}(S_\tau^2|\mathcal{F}_n) \geq \mathbb{E}(|S_\tau||\mathcal{F}_n)^2 \geq S_{\tau\wedge n}^2\,,$$

so the desired result follows. □

Remark 1.15. Indeed, since $(S_{\tau\wedge n})$ is L^2 bounded, by L^p convergence theorem (Theorem 1.35), $S_{\tau\wedge n} \to S_\tau$ in L^2. Thus

$$\mathbb{E}S_\tau^2 = \lim_{n\to\infty} \mathbb{E}\, S_{\tau\wedge n}^2 = x^2 + \sigma^2\, \mathbb{E}\,\tau\,.$$

We should study the technique in this proof of (ii), because it is useful in a number of situations. We used that martingale property preserved at bounded stopping times $\tau \wedge n$, then let $n \to \infty$, and use an appropriate convergence theorem.

Theorem 1.21. *Let $(S_n)_{n\geq 0}$ be symmetric random walk with $S_0 = 0$ and let $T_1 = \inf\{n \geq 0 : S_n = 1\}$. Then*

$$\mathbb{E}_0\, s^{T_1} = \frac{1-\sqrt{1-s^2}}{s} \quad \textit{for } s > 0\,.$$

Inverting the generating function we find that

$$\mathbb{P}_0\,(T_1 = 2n-1) = \frac{1}{2n-1}\cdot\frac{(2n)!}{n!n!}\,2^{-2n}\,.$$

Proof. By Remark 1.14, $\mathbb{P}_0(T_1 < \infty) = 1$. We will use the exponential martingale $X_n = e^{\theta S_n}/\phi(\theta)^n$ with $\theta > 0$ and

$$\phi(\theta) = \mathbb{E}\, e^{\theta \xi_1} = \frac{e^\theta + e^{-\theta}}{2}\,.$$

Clearly, $\phi(\theta) \geq 1$. Thus for each n, $|X_{n\wedge T_1}| \leq e^\theta$, and it follows form the bounded convergence theorem that

$$\mathbb{E} X_{T_1} = 1 \quad \text{and hence} \quad \mathbb{E}\left[\phi(\theta)^{-T_1}\right] = e^{-\theta}\,.$$

To convert this into the formula for the generating function we set

$$\phi(\theta) = \frac{e^\theta + e^{-\theta}}{2} = \frac{1}{s}\,.$$

Letting $x = e^{-\theta}$ and doing some algebra we want $x + x^{-1} = 2/s$ or

$$sx^2 - 2x + s = 0\,.$$

The quadratic equation implies

$$x = \frac{2 \pm \sqrt{4 - 4s^2}}{2s} = \frac{1 \pm \sqrt{1 - s^2}}{s}\,,$$

since $\mathbb{E} s^{T_1} = \sum_{k=1}^\infty s^k P(T_1 = k)$ we want the solution that is 0 when $s = 0$, which is $(1 - \sqrt{1 - s^2})/s$.

To invert the generating function we use Newton's binomial formula

$$(1+t)^a = 1 + \binom{a}{1} t + \binom{a}{2} t^2 + \binom{a}{3} t^3 + \dots$$

Taking $t = -s^2$ and $a = 1/2$ we have

$$\sqrt{1 - s^2} = 1 - \binom{1/2}{1} s^2 + \binom{1/2}{2} s^4 - \binom{1/2}{3} s^6 + \dots$$

$$\frac{1 - \sqrt{1 - s^2}}{s} = \binom{1/2}{1} s - \binom{1/2}{2} s^3 + \binom{1/2}{3} s^5 + \dots$$

The coefficient of s^{2n-1} is

$$(-1)^{n-1}\frac{(1/2)(-1/2)\cdots(3-2n)/2}{n!} = \frac{1\cdot 3\cdots(2n-3)}{n!}\cdot 2^{-n} = \frac{1}{2n-1}\frac{(2n)!}{n!n!}2^{-2n}$$

which completes the proof. $\square$

Theorem 1.22. *Let $(S_n)_{n\geq 0}$ be a asymmetric simple random walk starting at $x \in \mathbb{Z}$. Set $\mathbb{P}(\xi_i = 1) = p$ and $\mathbb{P}(\xi_i = -1) = q \equiv 1 - p$ with $p \neq q$.*

(i) Let $\varphi(k) = \left(\frac{q}{p}\right)^k$ for $k \in \mathbb{Z}$, then $(\varphi(S_n))_{n\geq 0}$ is a martingale.

(ii) If we let $T_j = \inf\{n : S_n = j\}$ then for $a < x < b$,

$$\mathbb{P}_x(T_a < T_b) = \frac{\varphi(b) - \varphi(x)}{\varphi(b) - \varphi(a)}, \ \mathbb{P}_x(T_b < T_a) = \frac{\varphi(x) - \varphi(a)}{\varphi(b) - \varphi(a)}.$$

(iii) If $1/2 < p < 1$, $a < 0 < b$ then

$$\mathbb{P}_0(T_a < \infty) = \frac{1}{\varphi(a)}, \ \mathbb{P}_0(T_b < \infty) = 1, \ \mathbb{E}_0 T_b = \frac{b}{p-q}.$$

Proof. (i). Clearly $\varphi(S_n)$ is integrable. For $n \geq 0$,

$$\mathbb{E}_x(\varphi(S_{n+1})|\mathcal{F}_n) = \left(\frac{q}{p}\right)^{S_n}\mathbb{E}_x\left[\left(\frac{q}{p}\right)^{\xi_{n+1}}\middle|\mathcal{F}_n\right] = \left(\frac{q}{p}\right)^{S_n}(p+q) = \varphi(S_n).$$

(ii). Let $\tau \equiv \tau_{(a,b)} := T_a \wedge T_b$, Since $(\varphi(S_{\tau\wedge n}))_{n\geq 0}$ is bounded, so

$$\mathbb{E}_x\varphi(S_\tau) = \mathbb{P}_x(T_a < T_b)\varphi(a) + \mathbb{P}_x(T_b < T_a)\varphi(b) = \mathbb{E}_x\varphi(x).$$

Note that

$$\mathbb{P}_x(T_a < T_b) + \mathbb{P}_x(T_b < T_a) = 1,$$

then we get (ii).

(iii). Note that $T_a < \infty$ if and only if $T_a < T_b$ for some (random) b. Therefore,

$$\mathbb{P}_0(T_a < \infty) = \lim_{b\to\infty} \mathbb{P}_0(T_a < T_b) = \lim_{b\to\infty} \frac{\varphi(b) - \varphi(0)}{\varphi(b) - \varphi(a)} = \frac{1}{\varphi(a)}\,.$$

For the same reason,

$$\mathbb{P}_0\left(T_b < \infty\right) = \lim_{a\to-\infty} \mathbb{P}_x\left(T_b < T_a\right) = \lim_{a\to-\infty} \frac{\varphi(0) - \varphi(a)}{\varphi(b) - \varphi(a)} = 1\,.$$

To show the last conclusion, we note that $(S_n - (p-q)n)_{n\geq 0}$ is a martingale. By optional sampling theorem,

$$\mathbb{E}_0\, S_{T_b\wedge n} = (p-q)\,\mathbb{E}_0\left(T_b \wedge n\right)\,.$$

Note that Now $b \geq S_{T_b\wedge n} \geq \inf_n S_n$ and

$$\mathbb{E}_0\left(-\inf_n S_n\right) = \sum_{j=1}^{\infty} \mathbb{P}_0\left(-\inf_n S_n \geq j\right) = \sum_{j=1}^{\infty} \mathbb{P}_0\left(\tau_{-j} < \infty\right) < \infty\,,$$

then letting $n \to \infty$, the desired result follows from domainted convergencetheorem and monotone convergence theorem. $\square$

1.4 Almost sure martingale convergence

In this section, we present the usual martingale convergence theorems and give some counterexamples. All the result is contained in the following theorem.

Theorem 1.23 (Martingale Convergence Theorem). *Let $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ be a (sub-,super-) martingale. If (X_n) is L^1-bounded, then there exists an integrable r.v. $X_\infty \in \mathcal{F}_\infty$ so that $X_n \to X_\infty$ almost surely.*

The assumptions of Theorem 1.23 do not guarantee convergence in L^1. In other words, a (sub-,super-) martingale is L^1 bounded doesn't ensure it is uniformly integrable.

Example 1.16. Consider the martingale transform $H \cdot M$ in Example 1.8. Since H is a locally bounded predictable process and M is a martingale, $H \cdot M$ is a martingale. Clearly $H \cdot M$ is L^1 bounded:

$$\mathbb{E}|(H \cdot M)_n| \leq 1 + \frac{1}{2^n} \times (1 + 2 + \cdots + 2^{n-1}) \leq 2\,.$$

However, as we have shown, $(H \cdot M)_n \to (H \cdot M)_\infty$ a.s. with

$$1 = \mathbb{E}\left[(H \cdot M)_\infty\right] > \mathbb{E}\left[(H \cdot M)_n\right] = 0\,.$$

So the convergence is not in L^1.

Example 1.17. Let $(\xi_j)_{j\geq 1}$ be i.i.d. with $\mathbb{P}(\xi_j = 0) = \mathbb{P}(\xi_j = 2) = 1/2$, for all $j \geq 1$. and define $M_n = \prod_{j=1}^n \xi_j$ for $n \geq 1$, and $M_0 = 1$. Let $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_n = \sigma(\xi_1, \cdots, \xi_n)$. Then $(M_n, \mathcal{F}_n)_{n\geq 0}$ is a non-negative martingale with $\mathbb{E}(M_n) = 1$ for all n (see Example 1.3). However, by Borel-Cantelli lemma,

$$M_n \to M_\infty = 0 \quad \text{a.s.}.$$

Thus

$$\mathbb{E}(M_n) = 1 > \mathbb{E}(M_\infty) = 0\,,$$

and the convergence is not in L^1.

We shall give another important counterexample.

Example 1.18. Let $\{S_n\}$ be a symmetric simple random walk with $S_0 = 1$. Let

$$T_0 = \inf\{n \geq 0 : S_n = 0\}$$

and let $X_n = S_{T_0 \wedge n}$ for all $n \in \mathbb{N}_0$. Theorem 1.6 implies that $\{X_n\}$ is a nonnegative martingale. Theorem 1.23 implies $\{X_n\}$ converges to a limit $X_\infty <$

∞ that must be $\equiv 0$, since convergence to $k > 0$ is impossible. (If $X_n = k > 0$ then $X_{n+1} = k \pm 1$.) Since $\mathbb{E}X_n = \mathbb{E}X_0 = 1$ for all n and $X_\infty = 0$, convergence cannot occur in L^1.

The road to martingale convergence theorem

We start with the core of the martingale convergence theorems, the so-called *upcrossing inequality.* Let $\mathfrak{F} = (\mathcal{F}_n)_{n\in\mathbb{N}_0}$ be a filtration. Recall that $\mathcal{F}_\infty := \sigma(\cup_n \mathcal{F}_n)$. Suppose $X = (X_n)_{n\in\mathbb{N}_0}$ is a real-valued process adapted to the filtration $\mathfrak{F}$. Let $a, b \in \mathbb{R}$ with $a < b$.

If we think of X as a stock price, it would be a sensible trading strategy to buy the stock when its price has fallen below a and to sell it when it exceeds b at least if we knew for sure that the price would always rise above the level b again. Each time the price makes such an upcrossing from a to b, we make a profit of at least $b - a$. If we get a bound on the maximal profit we can make, dividing it by $b - a$ gives a bound on the maximal number of such upcrossings. If this number is finite for all $a < b$, then the price has to converge.

Let us get into the details. Define stopping times $\sigma_0 \equiv 0$ and

$$\begin{aligned} \tau_k &:= \inf\{n \geq \sigma_{k-1} : X_n \leq a\} \qquad \text{for } k \in \mathbb{N}\,, \\ \sigma_k &:= \inf\{n \geq \tau_k : X_n \geq b\} \qquad \text{for } k \in \mathbb{N}\,. \end{aligned}$$

Note that $\tau_k = \infty$ if $\sigma_{k-1} = \infty$, and $\sigma_k = \infty$ if $\tau_k = \infty$. We say that X has its k th upcrossing over $[a, b]$ between τ_k and σ_k if $\sigma_k < \infty$. For $n \in \mathbb{N}$, define

$$U_n^{a,b}(X) := \sum_{k=0}^{\infty} 1_{\{\sigma_k \leq n\}}$$

as the number of upcrossings over $[a, b]$ until time n.

Lemma 1.24 (Upcrossing Inequality). *Let $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ be a submartingale. Then*

$$\mathbb{E}\, U_n^{a,b}(X) \leq \frac{\mathbb{E}\left[(X_n - a)^+\right] - \mathbb{E}\left[(X_0 - a)^+\right]}{b - a} \quad \textit{for } n \in \mathbb{N}\,.$$

Proof. Formally, the intimated trading strategy H is described for $n \in \mathbb{N}$ by

$$H_n := \begin{cases} 1, & \text{if } n \in \{\tau_k + 1, \ldots, \sigma_k\} \text{ for some } k \in \mathbb{N} \\ 0, & \text{else.} \end{cases}$$

H is nonnegative and predictable since, for all $n \in \mathbb{N}$

$$\{H_n = 1\} = \bigcup_{k=1}^{\infty} \{\tau_k \le n-1\} \cap \{\sigma_k > n-1\}$$

and each of the events is in $\mathcal{F}_{n-1}$.

Define $Y = X \vee a$, then Y is a $\mathfrak{F}$-submartingale. Intuitively,

$$(H \cdot Y)_n \ge (b-a)\, U_n^{a,b}(X) \quad \text{for all } n \in \mathbb{N}\,, \tag{1.12}$$

since each upcrossing results in a profit $\ge (b—a)$ and a final incomplete upcrossing (if there is one) makes a nonnegative contribution to the right-hand side. It is for this reason we had to replace X by Y. Then it follows from Theorem 1.6 that $H \cdot Y$ and $(1-H) \cdot Y$ are submartingales. Now note that

$$Y_n - Y_0 = (H \cdot Y)_n + ((1-H) \cdot Y)_n\,,$$

hence

$$\mathbb{E}\,(Y_n - Y_0) \ge \mathbb{E}\,[(H \cdot Y)_n] + \mathbb{E}[((1-H) \cdot Y)_n] \ge (b-a)\mathbb{E}\, U_n^{a,b}(X)\,,$$

then the desired result follows. Finally, we shall show (1.12) rigorously. Note that, on $\{U_n^{a,b}(X) = k\}$ for some $k \in \mathbb{N}$,

$$Y_{\sigma_i} - Y_{\tau_i} = Y_{\sigma_i} - a \ge b - a \quad \text{for all } i = 1, \cdots, k\,,$$

and hence

$$(H \cdot Y)_{\sigma_k} = \sum_{i=1}^{k} \sum_{j=\tau_i+1}^{\sigma_i} (Y_j - Y_{j-1}) = \sum_{i=1}^{k} (Y_{\sigma_i} - Y_{\tau_i}) \ge k(b-a)\,.$$

For $n \in \{\sigma_k, \ldots, \tau_{k+1}\}$, we have

$$(H \cdot Y)_n = (H \cdot Y)_{\sigma_k} \ge k(b-a).$$

On the other hand, for $n \in \{\tau_{k+1}, \dots, \}$, we have

$$(H \cdot Y)_n \geq (H \cdot Y)_{\tau_{k+1}} = (H \cdot Y)_{\sigma_k} \geq k(b-a)\,.$$

Hence (1.12) holds. ☐

Remark 1.16. The key fact is that $(1-H)\cdot Y$ is a submartingale, In other words, no matter how hard you try you can't lose money betting on a submartingale.

From the upcrossing inequality, we get easily the following theorem.

Theorem 1.25 (Submartingale Convergence). *Let $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ be a submartingale. If (X_n) is L^1 bounded, i.e.,*

$$\sup_{n\in\mathbb{N}_0} \mathbb{E}|X_n| < \infty \qquad \left(\Leftrightarrow \sup_{n\in\mathbb{N}_0} \mathbb{E}X_n^+ < \infty \right),$$

then there exists an integrable r.v. $X_\infty \in \mathcal{F}_\infty$, so that $X_n \to X_\infty$ almost surely.

Remark 1.17. Using the equality $|x| = 2x^+ - x$ and the submartingle property of (X_n), we have

$$\mathbb{E}|X_n| = 2\mathbb{E}X_n^+ - \mathbb{E}X_n \leq 2\mathbb{E}X_n^+ - \mathbb{E}X_0\,.$$

So $\sup_n \mathbb{E}|X_n| < \infty$ is equivalent to $\sup_n \mathbb{E}X_n^+ < \infty$.

Proof. Note that $(x-a)^+ \leq x^+ + |a|$, by upcrossing inequality

$$\mathbb{E}\, U_n^{a,b} \leq \frac{(|a| + \mathbb{E}X_n^+)}{b-a}$$

As $n \uparrow \infty, U_n^{a,b} \uparrow U^{a,b}$ the number of upcrossings of $[a,b]$ by the whole sequence, so if $\sup \mathbb{E}X_n^+ < \infty$ then $\mathbb{E}\, U^{a,b} < \infty$ and hence $U^{a,b} < \infty$ a.s. Since the last conclusion holds for all rational a and b,

$$\bigcup_{a,b\in\mathbb{Q}} \{\liminf X_n < a < b < \limsup X_n\}$$

has probability 0. Thus, $\limsup X_n = \liminf X_n$ a.s., and

$$X_\infty := \lim_{n\to\infty} X_n \,.$$

Fatou's lemma guarantees $\mathbb{E}|X_\infty| \le \liminf \mathbb{E}|X_n| < \infty$, so $|X_\infty| < \infty$ a.s. and hence X_∞ is a integrable random variable. $\square$

Remark 1.18. To prepare for the proof of Theorem 1.56, we should emphasize that we have shown that if the number of upcrossings of (a, b) by $X = (X_n)$ is almost surely finite for all $a, b \in \mathbb{Q}$, then the limit of (X_n) almost surely exists in $[-\infty, \infty]$.

Corollary 1.26 (Supermartingale Convergence)**.** *Let $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ be a supermartingale. If (X_n) is L^1 bounded, i.e.,*

$$\sup_{n\in\mathbb{N}_0} \mathbb{E}|X_n| < \infty \qquad \left(\Leftrightarrow \sup_{n\in\mathbb{N}_0} \mathbb{E}X_n^- < \infty\right),$$

Then there exists an integrable r.v. $X_\infty \in \mathcal{F}_\infty$ so that $X_n \to X_\infty$ almost surely.

Exercise 1.5. Let $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ be a non-negative supermartingale with respect to $\mathfrak{F}$, then there exists an integrable r.v. $X_\infty \in \mathcal{F}_\infty$ so that $X_n \to X_\infty$ almost surely. Moreover, $(X_n, \mathcal{F}_n; 0 \le n \le \infty)$ is a supermartingale.

Example 1.19 (Polya's urn)**.** Suppose we have an urn with red and green balls. At time $n = 0$, we start with r red ball and g green ball. At each positive integer time we choose a ball at random from the urn (with each ball equally likely to be chosen), look at the color of the ball, and then put the ball back in with another c balls of the same color. Let R_n, G_n denote the number of red and green balls in the urn after the draw at time n so that

$$R_0 = r,\ G_0 = g,\ R_n + G_n = r + g + nc \,.$$

and let

$$X_n = \frac{G_n}{R_n + G_n} = \frac{G_n}{r + g + nc}\,, \text{ for } n \in \mathbb{N}_0 \,.$$

Clearly, $X = \{X_n\}$ is integrable. We will show that $\{X_n\}$ is a martingale with respect to $\mathfrak{F}^X$. Note that for all $n \in \mathbb{N}_0$,

$$\begin{aligned}\mathbb{E}\left[X_{n+1}|\mathcal{F}_n\right] &= X_n \frac{G_n+1}{r+g+(n+1)c} + [1-X_n]\frac{G_n}{r+g+(n+1)c} \\ &= \frac{G_n + cX_n}{r+g+nc} = \frac{X_n(r+g+(n+1)c)+cX_n}{r+g+(n+1)c} = X_n\,.\end{aligned}$$

The martingale convergence theorem implies that $X_n \to X_\infty$ a.s. To compute the distribution of the limit, we observe

(i) the probability of getting green on the first m draws then red on the next $\ell = n-m$ draws is

$$\frac{g}{g+r}\cdot\frac{g+c}{g+r+c}\cdots\frac{g+(m-1)c}{g+r+(m-1)c}\cdot\frac{r}{g+r+mc}\cdots\frac{r+(\ell-1)c}{g+r+(n-1)c}$$

(ii) any other outcome of the first n draws with m green balls drawn and ℓ red balls drawn has the same probability since the denominator remains the same and the numerator is permuted.

Consider the special case $c=1, g=1, r=1$. It follows from (i) and (ii) that

$$\mathbb{P}\left(G_n = m+1\right) = \binom{n}{m}\frac{m!(n-m)!}{(n+1)!} = \frac{1}{n+1} \text{ for } m = 0, 2, \cdots, n\,.$$

so we can see that X_∞ has a uniform distribution on $(0,1)$.

If we suppose that $c=1, g=2$, and $r=1$, then

$$P\left(G_n = m+2\right) = \frac{n!}{m!(n-m)!}\frac{(m+1)!(n-m)!}{(n+2)!/2} \to 2x$$

if $n \to \infty$ and $m/n \to x$. In general, the distribution of X_∞ has density

$$\frac{\Gamma((g+r)/c)}{\Gamma(g/c)\Gamma(r/c)} x^{(g/c)-1}(1-x)^{(r/c)-1}$$

This is the beta distribution with parameters g/c and r/c. Later, we will see that the limit behavior changes drastically if, in addition to the c balls of the color chosen, we always add one ball of the opposite color.

1.5 Convergnece in L^1

We are now state the main theorems of this section. Since we have already done all the work in the course of measure theory, the proofs are trivial.

Lemma 1.27 (L^1 Convergence). *Let $(X_n, \mathcal{F}_n)_{n\geq 0}$ be a (sub-, super-) martingale. Then the following are equivalent. (i) $(X_n)_{n\geq 0}$ is uniformly integrable. (ii) $(X_n)_{n\geq 0}$ converges almost surely and in L^1. (iii) $(X_n)_{n\geq 0}$ converges in L^1.*

Proof. We only show that (i) imples (ii). Uniform integrability implies X is L^1 bounded, so the martingale convergence theorem implies that (X_n) converges almost surely to an integrable r.v. $X_\infty \in \mathcal{F}_\infty$, and convergence in L^1 follows from the uniform integrability of X. □

Theorem 1.28 (Characterization of UI Martingale). *Let $(X_n, \mathcal{F}_n)_{n\geq 0}$ be a (sub-, super-)martingale. Then the followings are equivalent.*

(i) $(X_n)_{n\geq 0}$ *is is uniformly integrable.*

(ii) $(X_n)_{n\geq 0}$ *converges to some* $X_\infty \in \mathcal{F}_\infty$ *a.s. and in* L^1.

(iii) There exists an integrable r.v. $X_\infty \in \mathcal{F}_\infty$ *such that* $(X_n, \mathcal{F}_n; 0 \leq n \leq \infty)$ *is a (sub-, super-)martingale with* $\mathbb{E}X_n \to \mathbb{E}X_\infty$.

Remark 1.19. Note that a martingale $(X_n, \mathcal{F}_n)_{n\geq 0}$ satisfying (iii) if and only if it's a Levy meartingale. So this theorem implies that *uniformly integrable martingale and Levy martingale are the same thing* !

Remark 1.20. We should emphasize that although we use the notation X_∞ in (ii) and (iii), it doesn't mean that they are the same random variable. Specifically, if $(X_n, \mathcal{F}_n)_{n\geq 0}$ is a (sub-, super-)martingale with $X_n \to X_\infty$ a.s. and in L^1, then $(X_n, \mathcal{F}_n; 0 \leq n \leq \infty)$ is a (sub-, super-)martingale with $\mathbb{E}X_n \to \mathbb{E}X_\infty$. However, if $(X_n, \mathcal{F}_n; 0 \leq n \leq \infty)$ is a sub-(super-)martingale with $\mathbb{E}X_n \to \mathbb{E}X_\infty$. In general, we don NOT konw if $X_n \to X_\infty$ a.s. and in L^1, unless $(X_n, \mathcal{F}_n; 0 \leq n \leq \infty)$ is a martingale (see Theorem 1.29).

Proof. We have shown that (i) implies (ii) in the preceding lemma.

To show (ii) impleies (iii), as metioned in Remark 1.20, we shall prove that $(X_n, \mathcal{F}_n; 0 \le n \le \infty)$ is a (sub-, super-)martingale. Fix $n \in \mathbb{N}_0$. It is sufficient to show that for all $A \in \mathcal{F}_n$,

$$\mathbb{E}(X_\infty 1_A) \ (\ge, \le) = \ \mathbb{E}(X_n 1_A) \,.$$

Observe that for any $k \ge 0$, we have $\mathbb{E}(X_{n+k} 1_A) \ (\ge, \le) = \ \mathbb{E}(X_n 1_A)$. Letting $k \to \infty$, since $\{X_n\}$ converges to X_∞ in L^1, so

$$\mathbb{E}(X_{n+k} 1_A) \to \mathbb{E}(X_\infty 1_A) \,,$$

and the desired result follows.

We now prove (iii) implies (i). Without loss of generality, we assume that $(X_n, \mathcal{F}_n)_{n\ge 0}$ be a submartingale. (Indeed, for the case of martingale, we have already proved this in Example 1.2.) Firstly we show that $(X_n)_{n\ge 0}$ is L^1-bounded. Since $(X_n^+, \mathcal{F}_n; 0 \le n \le \infty)$ is also a submartingale, we have

$$\sup_{n\ge 0} \mathbb{E} X_n^+ \le \mathbb{E} X_\infty^+ \,.$$

It follows from Remark 1.17 that $(X_n)_{n\ge 0}$ is L^1-bounded. Next, we shall show respectively that $(X_n^+)_{n\ge 0}$ and $(X_n^-)_{n\ge 0}$ are uniformly integrable.

We now prove that $(X_n^+)_{n\ge 0}$ is uniformly integrable. For any $\lambda > 0$, $(X_n^+, \mathcal{F}_n; 0 \le n \le \infty)$ is also a submartingale, we have

$$\mathbb{E} X_n^+ 1_{\{X_n^+ > \lambda\}} = \mathbb{E} X_n 1_{\{X_n > \lambda\}} \le \mathbb{E} X_\infty 1_{\{X_n > \lambda\}} \,.$$

On the other hand, by Markov inequality,

$$\mathbb{P}(X_n > \lambda) \le \frac{\mathbb{E} X_n^+}{\lambda} \le \frac{\mathbb{E} X_\infty^+}{\lambda} \,,$$

so $\sup_{n\ge 0} \mathbb{P}\left(|X_n| > \lambda\right)$ converges to zero as $\lambda \to \infty$. Hence

$$\sup_{n\ge 0} \mathbb{E} X_n^+ 1_{\{X_n^+ > \lambda\}} \le \sup_{n\ge 0} \mathbb{E}\, X_\infty 1_{\{X_n > \lambda\}} \to 0 \ \text{ as } \lambda \to \infty$$

follows from the absolute continuity of the integral.

We now prove that $(X_n^-)_{n\geq 0}$ is uniformly integrable. Observe that for each $m < n$, we have

$$\begin{aligned}\mathbb{E}X_n^- 1_{\{X_n^- > \lambda\}} &= -\mathbb{E}X_n 1_{\{X_n < -\lambda\}} = \mathbb{E}X_n 1_{\{X_n > -\lambda\}} - \mathbb{E}X_n \\ &\leq \mathbb{E}X_\infty 1_{\{X_n > -\lambda\}} - \mathbb{E}X_n \\ &= \mathbb{E}X_\infty - \mathbb{E}X_n - \mathbb{E}X_\infty 1_{\{X_n < -\lambda\}}\,.\end{aligned}$$

Since $\mathbb{E}X_n \to \mathbb{E}X_\infty$, given any $\epsilon > 0$, we can certainly choose $m = m_\epsilon$ so large that $|\mathbb{E}X_\infty - \mathbb{E}X_n| \leq \epsilon$ holds for every $n > m_\epsilon$, and by the absolute continuity of the integral, there exists $\lambda_\epsilon > 0$ so that for any $\lambda > \lambda_\epsilon$,

$$\sup_{n\geq 1} \mathbb{E}X_\infty 1_{\{X_n < -\lambda\}} < \epsilon\ ; \ \sup_{n \leq m_\epsilon} \mathbb{E}X_n^- 1_{\{X_n^- > \lambda\}} < \epsilon\,.$$

Consequently, for any $\lambda > \lambda_\epsilon$ we have:

$$\sup_{n\geq 1} \mathbb{E}X_n^- 1_{\{X_n^- > \lambda\}} < 2\epsilon$$

and thus $(X_n^-)_{n\geq 0}$ is also uniformly integrable. □

Warning: the condition $\mathbb{E}X_n \to \mathbb{E}X_\infty$ in (iii) is necessary! To find a conterexample, it suffices to find a a supermartingle $(X_n, \mathcal{F}_n; 0 \leq n \leq \infty)$ that is not uniformly integrable. Note that for any non-negative martingale $(M_n, \mathcal{F}_n)_{n\geq 0}$, letting $M_\infty = 0$, then $(M_n, \mathcal{F}_n; 0 \leq n \leq \infty)$ is a supermartingale (but may not a martingale). In addition, suppose $M_n \to M_\infty$ a.s., then $(M_n, \mathcal{F}_n; 0 \leq n \leq \infty)$ is uniformly integrable if and only if $M_n \to M_\infty$ in L^1. We already give conterexamples about this, see Example 1.17 and Example 1.18.

This following theorem gives the representation of the limit for a Levy's martingale.

Theorem 1.29 (Limit of Levy Martingale)**.** *Let ξ in L^1 and let $\mathfrak{F} = (\mathcal{F}_n)_{n\in\mathbb{N}_0}$ be a filtration. Then $\{\mathbb{E}(\xi|\mathcal{F}_n)\}_{n\in\mathbb{N}_0}$ is a uiformly integrable martingale with respect to $\mathfrak{F}$, and*

$$\mathbb{E}(\xi|\mathcal{F}_n) \to \mathbb{E}(\xi|\mathcal{F}_\infty) \ \textit{a.s. and in } L^1\,.$$

Proof. We have shown that $\{\mathbb{E}(\xi|\mathcal{F}_n)\}$ is a uiformly integrable martingale. So there exists $X_\infty \in \mathcal{F}_\infty$, $\mathbb{E}(\xi|\mathcal{F}_n) \to X_\infty$ in L^1. It suffices to show that

$$X_\infty = \mathbb{E}(\xi|\mathcal{F}_\infty) \ \text{ a.s..}$$

Recall $\mathcal{F}_\infty = \sigma(\cup_n \mathcal{F}_n)$. Note that $A \in \cup_n \mathcal{F}_n$, by convergence in L^1,

$$\mathbb{E}[\mathbb{E}(\xi|\mathcal{F}_n)1_A] \to \mathbb{E}(X_\infty 1_A)\,,$$

and for large n, $A \in \mathcal{F}_n$ so $\mathbb{E}[\mathbb{E}(\xi|\mathcal{F}_n)1_A] = \mathbb{E}(\xi 1_A)$. Hence

$$\mathbb{E}(X_\infty 1_A) = \mathbb{E}(\xi 1_A)\,, \quad \text{for all } A \in \cup_n \mathcal{F}_n\,.$$

Applying the $\pi - \lambda$ theorem we get

$$\mathbb{E}(X_\infty 1_A) = \mathbb{E}(\xi 1_A)\,, \ \text{for all } A \in \mathcal{F}_\infty\,. \qquad \square$$

An immediate consequence of Theorem 1.29 is:

Corollary 1.30 (Lévy's 0-1 Law)**.** *Let $\mathfrak{F} = (\mathcal{F}_n)_{n\in\mathbb{N}_0}$ be a filtration. For $A \in \mathcal{F}_\infty$, we have*

$$\mathbb{P}(A|\mathcal{F}_n) \to 1_A \ \textit{a.s.}$$

To steal a line from Chung: *The reader is urged to ponder over the meaning of this result and judge for himself whether it is obvious or incredible.* We will now argue for the two points of view.

"It is obvious." $1_A \in \mathcal{F}_\infty$, and $\mathcal{F}_n \uparrow \mathcal{F}_\infty$, so our best guess of 1_A given the information in $\mathcal{F}_n$ should approach 1_A (the best guess given $\mathcal{F}_\infty$).

"It is incredible." Let $X_1, X_2, \ldots$ be independent and suppose $A \in \mathcal{T}$, the tail σ-field. For each n, A is independent of $\mathcal{F}_n$, so $\mathbb{P}(A|\mathcal{F}_n) = \mathbb{P}(A)$. As $n \to \infty$, the left-hand side converges to 1_A a.s., so $\mathbb{P}(A) = 1_A$ a.s., and it follows that $\mathbb{P}(A) \in \{0, 1\}$, i.e., we have proved Kolmogorov's 0-1 law.

The last argument may not show that Lévy's 0-1 law is "too unusual or improbable to be possible," but this and other applications below show that it is a very useful result.

Corollary 1.31 (Dominated Convergence Theorem)**.** *Let $\mathfrak{F} = \{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$ be a filtration. Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a sequence of r.v.'s NOT need to be adapted to $\mathfrak{F}$ and $X_n \to X_\infty$ a.s. If $\{X_n\}$ is domainted by some $Y \in L^1$, i.e., $|X_n| \le Y$ for each n, then*

$$\mathbb{E}\left(X_n|\mathcal{F}_n\right) \to \mathbb{E}\left(X_\infty|\mathcal{F}_\infty\right) \qquad a.s..$$

Proof. Since we have shown that $\mathbb{E}(X_\infty|\mathcal{F}_n) - \mathbb{E}(X_\infty|\mathcal{F}_\infty) \to 0$ a.s., it suffices to show that

$$\mathbb{E}(X_n|\mathcal{F}_n) - \mathbb{E}(X_\infty|\mathcal{F}_n) \to 0 \text{ a.s..}$$

Let $Y_m \coloneqq \sup_{n\ge n} |X_n - X_\infty| \le 2Y$ for $m \ge 1$. Then, clearly, for any fixed m and $n \ge m$,

$$\mathbb{E}(|X_n - X_\infty| \mid \mathcal{F}_n) \le \mathbb{E}(Y_m|\mathcal{F}_n)\,.$$

Letting $n \to \infty$, by Theorem 1.29 we get

$$\limsup_{n\to\infty} \mathbb{E}(|X_n - X_\infty| \mid \mathcal{F}_n) \le \mathbb{E}(Y_m|\mathcal{F}_\infty)\,.$$

Leting $m \to \infty$ now, the desired result follows from the domainted convergence theorem for conditional expectation. ∎

Example 1.20. Let $Z_1, Z_2, \ldots$ be i.i.d. with $\mathbb{E}\,|Z_i| < \infty$, let θ be an independent r.v. with finite mean, and let $Y_i = Z_i + \theta$. If Z_i is normal$(0,1)$ then in statistical terms we have a sample from a normal population with variance 1 and unknown mean. The distribution of θ is called the *prior distribution*, and $P\left(\theta \in \cdot|Y_1, \ldots, Y_n\right)$ is called the *posterior distribution* after n observations. Let $\mathfrak{F}$ be the filtration generated by (Y_n). By SLLN we can see that $\theta \in \mathcal{F}_\infty$, hence it follows from Theorem 1.29 that

$$\mathbb{E}\left(\theta|Y_1, \ldots, Y_n\right) \to \theta \quad \text{a.s..}$$

Example 1.21 (Radon - Nikodym Theorem)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, μ be a finite measure on $\mathcal{F}$ and absolutely continuous to $\mathbb{P}$. Suppose that $\mathcal{F}$ is *countably generated*, i.e., there is a sequence of sets $\{A_n\}$ so that $\mathcal{F} = \sigma(\{A_n\})$. For a concrete example, consider $\Omega = [0,1)$, $\mathcal{F} = \mathcal{B}(\mathbb{R})$ and $\{A_n\}$= all the open intervals in $(0,1)$ with rational endpoints, $\mathbb{P}$=Lebesgue measure. We construct a filtration $\mathfrak{F} = (\mathcal{F}_n)_{n\in\mathbb{N}}$ by letting $\mathcal{F}_n := \sigma(\{A_1,\dots,A_n\})$. Evidently, $\mathcal{F}_n$ is a finite family, for all $n \in \mathbb{N}$. More precisely, there exists a finite subset $\mathcal{C}_n \subset \mathcal{F}_n$ is pairwise disjoint satisfying that, for any $A \in \mathcal{F}_n$,

$$A = \bigcup_{C\in\mathcal{C}_n, C\subset A} C\,.$$

In other words, $\mathcal{C}_n$ decomposes $\mathcal{F}_n$ into its "atoms". Finally, define a stochastic process $X = (X_n)_{n\in\mathbb{N}}$ by

$$X_n := \sum_{C\in\mathcal{C}_n, \mathbb{P}(C)>0} \frac{\mu(C)}{\mathbb{P}(C)}\, 1_C\,.$$

One can easily check that X_n is exactly $\mathrm{d}\mu_n/\,\mathrm{d}\mathbb{P}_n$, and $(X_n, \mathcal{F}_n)$ is a uniformly integrable martingale. Hence, by the same arguement as Theorem 1.29,

$$\frac{\mathrm{d}\mu}{\mathrm{d}\mathbb{P}} \text{ exists, and } = \lim_{n\to\infty} X_n \quad \text{a.s..}$$

Note that for this proof we did not presume the existence of conditional expectations (rather we constructed them explicitly for finite σ-algebras); that is, we did not resort to the Radon–Nikodym theorem in a hidden way.

1.6 Convergence in L^p $(p > 1)$

A Doob's Inequalities

With Kolmogorov's maximal inequality, we became acquainted with an inequality that bounds the probability of large values of the maximum of a square integrable process with independent centered increments. Here we want to improve this inequality in two directions. On the one hand, we replace the independent increments by the assumption that the process of partial sums is a martingale. On the other hand, we can manage with less than second moments; alternatively, we can get better bounds if we have higher moments.

Let $X = (X_n)_{n\in\mathbb{N}_0}$ be a real-valued stochastic process. Then the *running maximum* of the process X, denoted by X^*, is given by

$$X_n^* = \max_{0\le k\le n} X_k \quad \text{for each } n \ge 0\,.$$

Then the running maximum of $|X|$, is

$$|X|_n^* = \max_{0\le k\le n} |X_k| \quad \text{for each } n \ge 0\,.$$

Theorem 1.32 (Doob's Maximal Inequality)**.** *Let $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ be a submartingale. Then for any $n \in \mathbb{N}_0$ and $\lambda > 0$,*

$$\lambda\mathbb{P}\left(X_n^* \ge \lambda\right) \le \mathbb{E}\, X_n 1_{\{X_n^*\ge\lambda\}} \le \mathbb{E}X_n^+\,.$$

Proof. The second inequality is trivial. For the first one, let

$$\tau = \inf\{k \in \mathbb{N}_0 : X_k \ge \lambda\}\,.$$

Then for each $n \in \mathbb{N}$, $\{X_n^* \ge \lambda\} = \{\tau \le n\}$. Evidently, $\tau \wedge n$ is a bounded stopping time, by Theorem 1.9, and note that on $\{\tau \le n\}$, $X_\tau \ge \lambda$,

$$\begin{aligned}\mathbb{E}X_n \ge \mathbb{E}X_{\tau\wedge n} &= \mathbb{E}X_\tau 1_{\{\tau\le n\}} + \mathbb{E}X_n 1_{\{\tau>n\}}\\ &\ge \lambda\,\mathbb{P}(\tau \le n) + \mathbb{E}X_n 1_{\{\tau>n\}}\,.\end{aligned}$$

Thus we have

$$\lambda\,\mathbb{P}(\tau \le n) \le \mathbb{E}X_n - \mathbb{E}X_n 1_{\{\tau>n\}} = \mathbb{E}\,X_n 1_{\{\tau\le n\}}\,.$$

We have completed the proof. $\square$

Example 1.22 (Random Walks)**.** Let $S_n = \xi_1 + \cdots + \xi_n$ where the $\{\xi_m\}$ are independent and have $\mathbb{E}\xi_m = 0$, and $\mathbb{E}\xi_m^2 < \infty$. $\{S_n\}$ is a martingale and $\{S_n^2\}$ is a submartingale. If we let $\lambda = a^2$ and apply Doob's inequality to $\{S_n^2\}$, we get Kolmogorov's maximal inequality:

$$\mathbb{P}\left(\max_{1\le m\le n} |S_m| \ge a\right) \le \frac{\mathbb{E}\, S_n^2}{a^2} \,.$$

Exercise 1.6 (Doob's inequality for supermartingale). Let $(Y_n, \mathcal{F}_n)_{n\ge 0}$ be a super-martingale. Then for each $n \in \mathbb{N}_0$ and $\lambda > 0$,

$$\lambda \mathbb{P}(Y_n^* \ge \lambda) \le \mathbb{E}Y_0 - \mathbb{E}Y_n 1_{\{Y_n^* < \lambda\}} \le \mathbb{E}Y_n^- + \mathbb{E}Y_0 \,.$$

Exercise 1.7 (Doob's Inequality II). Let $X = (X_n)_{n\ge 0}$ be a submartingale with respect to $(\mathcal{F}_n)_{n\ge 0}$. Then for all $\lambda > 0$,

$$\lambda \mathbb{P}\left((-X)_n^* \ge \lambda\right) = \lambda \mathbb{P}\left(\min_{k\le n} X_k \le -\lambda\right) \le \mathbb{E}X_n^+ - \mathbb{E}X_0 \,,$$

and thus

$$\lambda \mathbb{P}\left(|X|_n^* \ge \lambda\right) \le 2\mathbb{E}X_n^+ - \mathbb{E}X_0 \le 2\mathbb{E}|X_n| + \mathbb{E}|X_0| \,.$$

Theorem 1.33 (Doob's L^p Maximal Inequality)**.** *Let $\{X_n, \mathcal{F}_n\}_{n\in\mathbb{N}_0}$ be a non-negative submartingale. Then for any $n \in \mathbb{N}_0$ and $p > 1$,*

$$\|X_n^*\|_p \le \frac{p}{p-1} \|X_n\|_p \,.$$

Remark 1.21. We suppose that X is non-negetive to guarantee that X_n^* is non-negetive, in order to use the following results. As we all know, for any non-negative random variable ξ, we have

$$\mathbb{E}\,\xi^p = p\int_0^\infty t^{p-1}\, \mathbb{P}(\xi \ge t)\, \mathrm{d}t \,.$$

Proof. **Step 1**. By Doob's inequality and Funibi's theorem,

$$\begin{aligned}\mathbb{E}\,|X_n^*|^p &= p\int_0^\infty t^{p-1}\mathbb{P}\left(X_n^* \geq t\right)\mathrm{d}t \leq p\int_0^\infty t^{p-2}\,\mathrm{d}t\int_\Omega X_n 1_{\{X_n^*\geq t\}}\,\mathrm{d}\mathbb{P} \\ &= p\int_\Omega X_n\,\mathrm{d}\mathbb{P}\int_0^{X_n^*} t^{p-2}\,\mathrm{d}t = \frac{p}{p-1}\mathbb{E}\left[X_n\left(X_n^*\right)^{p-1}\right].\end{aligned}$$

Hence, by Hölder's inequality,

$$\mathbb{E}\,|X_n^*|^p \leq \frac{p}{p-1}\left\|X_n\right\|_p\left[\mathbb{E}\left(X_n^*\right)^p\right]^{1/q},$$

where $\frac{1}{p}+\frac{1}{q}=1$. Suppose now $\|X_n^*\|_p<\infty$, then we get the desired result:

$$\|X_n^*\|_p \leq \frac{p}{p-1}\left\|X_n\right\|_p.$$

Step 2. However, if $\|X_n^*\|_p<\infty$ is not satisfied, we consider $X_n^*\wedge K$, where K is a positive integer. Obeserve that

$$\mathbb{P}\left(X_n^*\wedge K \geq t\right) = \mathbb{P}\left(X_n^* \geq t\right),\ \text{for } t\in[0,K],$$

using the same arguement in Step 1, then we get

$$\|X_n^*\wedge K\|_p \leq \frac{p}{p-1}\left\|X_n\right\|_p.$$

Since K is arbitary, letting $K\uparrow\infty$ we have

$$\|X_n^*\|_p \leq \frac{p}{p-1}\left\|X_n\right\|_p. \qquad \square$$

Remark 1.22. If $\{X_n\}$ is a martingale, then $\{|X_n|\}$ is a non-negative submartingale. Applying L^p maximum inequality, we get :

$$\|\,|X|_n^*\,\|_p \leq \frac{p}{p-1}\left\|X_n\right\|_p.$$

If $\{X_n\}$ is a submartingale , $\{X_n^+\}$ is a non-negative submartingale, so

$$\left\|\left(X^+\right)_n^*\right\|_p \leq \frac{p}{p-1}\left\|X_n^+\right\|_p.$$

Example 1.23. There is no L^1 maximal inequality.

Again, the counterexample is provided by Example 1.18. Let $\{S_n\}$ be a simple random walk starting from $S_0 = 1$. For each $j \in \mathbb{N}_0$, let

$$\tau_j = \inf\{n : S_n = j\},$$

be the first hitting time to state j. Define $X_n = S_{\tau_0 \wedge n}$. $\{X_n\}$ is a non-negative martingale implies $\mathbb{E}X_n = \mathbb{E}X_0 = \mathbb{E}S_0 = 1$ for all n. Using hitting probabilities for simple random walk, for $m \geq 1$,

$$\mathbb{P}(\tau_m < \tau_0) = \mathbb{P}\left(\max_{n\in\mathbb{N}_0} X_n \geq m\right) = \frac{1}{m}.$$

So

$$\mathbb{E}\left(\max_n X_n\right) = \sum_{m=1}^{\infty} \mathbb{P}\left(\max_n X_n \geq m\right) = \infty.$$

Since $X_n^* \uparrow \max_n X_n$, the monotone convergence theorem implies that

$$\|X_n^*\|_1 = \mathbb{E}X_n^* \uparrow \infty.$$

B L^p Convergnece Theorem

As we konw, in general, for a stochastic process $X = (|X_n|^p)_{n\geq 0}$ to be uniformly integrable it is not enough that $(X_n)_{n\geq 0}$ be L^p-bounded, where $p > 1$. However, if X is a martingale or a nonnegative submartingale, then Doob's inequality implies that the statements are equivalent. In particular, in this case, almost sure convergence implies convergence in L^p.

Lemma 1.34. *Let $\{X_n, \mathcal{F}_n\}_{n\in\mathbb{N}_0}$ be a non-negative submartingale, let $p > 1$. If $\{X_n\}$ is L^p-bounded, i.e.,*

$$\sup_{n\in\mathbb{N}_0} \|X_n\|_p < \infty,$$

then $\{X_n\}$ converges to a random variable $X_\infty \in L^p(\mathcal{F}_\infty)$ a.s. and in L^p.

Proof. Since $\{X_n\}$ is L^p bounded, it is L^1 bounded, by martingale convergence theorem, there exists an integrable $X_\infty \in \mathcal{F}_\infty$ so that $X_n \to X_\infty$ a.s.. It suffices to show that $\{X_n^p\}$ is uniformly integrable. By L^p maximum inequality, we have

$$\|X_n^*\|_p \le \frac{p}{p-1}\|X_n\|_p$$

Letting $n \to \infty$, since $X_n^* \uparrow \sup_n X_n$, we have

$$\mathbb{E} \sup_{n \in \mathbb{N}_0} X_n^p < \infty.$$

Observe that

$$X_n^p \le \sup_{n \in \mathbb{N}_0} X_n^p \quad \text{for all } n\,.$$

Thus $\{X_n^p\}$ is uniformly integrable, which implies the conclusion. □

Theorem 1.35 (L^p Convergence)**.** *Let $\{X_n, \mathcal{F}_n\}_{n\in\mathbb{N}_0}$ be a martingale, let $p > 1$. Then the following statements are equivalent.*

(i) $\{X_n\}$ *is L^p bounded, i.e.,* $\sup_{n\in\mathbb{N}_0} \|X_n\|_p < \infty$.

(ii) $\{X_n\}$ *converges a random variable $X_\infty \in L^p(\mathcal{F}_\infty)$ a.s. and in L^p.*

(iii) *There exists a random variable $X_\infty \in L^p(\mathcal{F}_\infty)$ such that $(X_n, \mathcal{F}_n; 0 \le n \le \infty)$ is a martingale.*

Proof. (i) ⇒ (ii). Since $\{X_n\}$ is L^p bounded, it is L^1 bounded, by martingale convergence theorem, there exists an integrable $X_\infty \in \mathcal{F}_\infty$ so that $X_n \to X_\infty$ a.s.. It fuffices to show that $\{|X_n|^p\}$ is uniformly integrable. Note that $\{|X_n|\}$ is a non-negative submartingale, by Lemma 1.34 , the desired result follows.

(ii) ⇒ (iii). Fix $n \in \mathbb{N}_0$. It is sufficient to show that for all $A \in \mathcal{F}_n$,

$$\mathbb{E}(X_n 1_A) = \mathbb{E}(X_\infty 1_A)\,.$$

However, for any $k \ge 0$, $\mathbb{E}(X_n 1_A) = \mathbb{E}(X_{n+k} 1_A)$. Letting $k \to \infty$, since $\{X_n\}$ converges to X_∞ in L^p, of course in L^1, so

$$\mathbb{E}(X_{n+k} 1_A) \to \mathbb{E}(X_\infty 1_A)\,,$$

and the desired result follows.

(iii) $\Rightarrow$ (i). Since $(X_n, \mathcal{F}_n; 0 \leq n \leq \infty)$ is a martingale, by Jensen's inequality we have

$$|X_n|^p \leq \mathbb{E}\left(|X_\infty|^p|\mathcal{F}_n\right) \quad \text{ for all } n\,.$$

Thus $\sup_{n\in\mathbb{N}_0} \mathbb{E}|X_n^p| \leq \mathbb{E}|X_\infty^p| < \infty$. □

Corollary 1.36. *Let $\xi \in L^p$ for $p > 1$. Let $\mathfrak{F} = (\mathcal{F}_n)_{n\geq 0}$ be a filtration. Then*

$$\mathbb{E}\left(\xi|\mathcal{F}_n\right) \to \mathbb{E}\left(\xi|\mathcal{F}_\infty\right), \quad \textit{a.s. and in } L^p\,.$$

Example 1.24. Let $(\xi_n)_{n\in\mathbb{N}}$ be non-negative independent random variables with $\mathbb{E}\xi_n = 1$ for all n. Set $X_0 = 1$, $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and

$$X_n = \prod_{j=1}^n \xi_j\,,\ \mathcal{F}_n = \sigma(\xi_1, \cdots, \xi_n) \ \text{ for } n \in \mathbb{N}\,.$$

Then $(X_n, \mathcal{F}_n)$ is a non-negative martingale. By martingale convergence theorem, $\{X_n\}$ converges a.s. to an integrable r.v. $X_\infty \in \mathcal{F}_\infty$. Note that

$$\sum_{n=1}^\infty \log \xi_n > -\infty$$

is a tail event, so the Kolmogorov 0-1 law implies $\mathbb{P}(X_\infty = 0) \in \{0, 1\}$. The next result gives a concrete criterion for which of the two alternatives occurs.

Let $a_j = \mathbb{E}\sqrt{\xi_j}$ for $j \geq 1$. Clearly, $a_j > 0$, and by Hölder inequality,

$$a_j = \mathbb{E}\sqrt{\xi_j} \leq \sqrt{\mathbb{E}\xi_j} = 1\,.$$

Then we have the following criterion.

(i) If $\Pi_j a_j = 0$, then $X_\infty = 0$ a.s.

(ii) If $\Pi_j a_j > 0$, then X_n converges to X_∞ in L^1, paticularly, $\mathbb{E}\left[X_\infty\right] = 1$, hence we have $X_\infty > 0$ a.s..

Proof. Define

$$Y_0 = 1, \quad Y_n = \prod_{j=1}^{n} \frac{\sqrt{\xi_j}}{a_j}, \text{ for } n \in \mathbb{N}.$$

Then $Y = (Y_n)$ is a non-negative martingale with respect to $\{\mathcal{F}_n\}$, and the relation with X is

$$Y_n^2 = \frac{X_n}{\Pi_{j=1}^n a_j}, \ \mathbb{E}\left[Y_n^2\right] = \frac{1}{\Pi_{j=1}^n a_j}, \text{ for } n \in \mathbb{N}.$$

If $\Pi_j a_j = 0$, since (Y_n) is a non-negative martingale, it converges a.s. to some limit $Y_\infty \in L^1$. Therefore, almost surely,

$$X_n = Y_n^2 \times \Pi_{j=1}^n a_j \to Y_\infty^2 \times 0 = 0 = X_\infty .$$

If $\Pi_j a_j > 0$, note that

$$\sup_n \mathbb{E}\, Y_n^2 \leq \frac{1}{\Pi_{j=1}^\infty a_j} < \infty .$$

By Lemma 1.34 , we deduce that $\sup_{n\in\mathbb{N}} Y_n^2 \in L^1$. Note that for all n,

$$X_n = Y_n^2 \Pi_{j=1}^n a_j \leq \sup_{n\in\mathbb{N}} Y_n^2 .$$

So $\{X_n\}$ is uniformly integrable. □

Futher reading* We will state the famous law of the iterated logarithm (LIT). The essential tool of the proof is Borel-Cantelli lemma, based on an estimation of probability using Doob's inequality. Indeed, the proof basically the same as the LIT for Brownian motion.

Theorem 1.37 (Law of the Iterated Logarithm)**.** *Let $(\xi_n)_{n\geq 1}$ be i.i.d. Gaussian random variable with zero mean and unit variance. Define $S_n = \sum_{j=1}^n \xi_j$. Then, the law of the iterated logarithm claim that*

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n \log\log n}} = 1 \quad a.s..$$

Clearly $S_n \sim N(0,n)$. We use the esitimation of the tail normal probability: For $x>0$,

$$\left(x^{-1}-x^{-3}\right)\exp\left(-x^2/2\right) \le \int_x^\infty \exp\left(-y^2/2\right)dy \le x^{-1}\exp\left(-x^2/2\right) .$$

So, for any positive function f,

$$\mathbb{P}\left(S_n > \sqrt{nf(n)}\right) \sim (2\pi)^{-1/2} f(n)^{-1/2}\exp(-f(n)/2) .$$

The last result implies that if $\epsilon>0$, then

$$\sum_{n=1}^\infty \mathbb{P}\left(S_n > \sqrt{nf(n)}\right) \begin{cases} <\infty & \text{when } f(n)=(2+\epsilon)\log n \\ =\infty & \text{when } f(n)=(2-\epsilon)\log n \end{cases}$$

and hence by the Borel-Cantelli lemma that

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\log n}} \le 1 \quad \text{a.s..}$$

Next, we will try to imporve this method.

Proof of Theorem 1.37. Define $\varphi(n)=\sqrt{2n\log\log n}$. It is sufficient to show that, for any $\epsilon>0$, we have almost surely,

$$\limsup_{n\to\infty}\frac{S_n}{\varphi(n)} \le 1+\epsilon, \quad \limsup_{n\to\infty}\frac{S_n}{\varphi(n)} \ge 1-\epsilon .$$

Upper bound. Since $(S_n)_{n\ge 0}$ is a martingale, and for $\theta>0$, the function $x\mapsto e^{\theta x}$ is convex, we have $\left(e^{\theta S_n}\right)$ is a submartingale. By Doob's inequality, we have that, for any $c>0$

$$\begin{aligned}\mathbb{P}\left(\max_{0\le k\le n} S_k \ge c\right) &= \mathbb{P}\left(\max_{0\le k\le n} e^{\theta S_k} \ge e^{\theta c}\right) \\ &\le e^{-\theta c}\mathbb{E}\left[e^{\theta S_n}\right] = e^{-\theta c+\theta^2\frac{n}{2}} .\end{aligned}$$

Pick $\theta=c/n$, we have that, for any $c>0$

$$\mathbb{P}\left(\max_{0\le k\le n} S_k \ge c\right) \le e^{-\frac{c^2}{2n}} .$$

Thus,

$$\mathbb{P}\left(\max_{0\le k\le n} S_k \ge (1+\epsilon)\varphi(n)\right) \le (\log n)^{-(1+\epsilon)^2}.$$

In order to use the Borel-Cantelli lemma, fix some $q > 1$, then we have

$$\mathbb{P}\left(\max_{0\le k\le q^m} S_k \ge (1+\epsilon)\varphi\left(q^m\right)\right) \le (m\log q)^{-(1+\epsilon)^2}.$$

Thus

$$\sum_{m=1}^{\infty} \mathbb{P}\left(\max_{k\le q^m} S_k \ge (1+\epsilon)\varphi\left(q^m\right)\right) < \infty.$$

By Borel-Cantelli lemma, we have almost surely,

$$\max_{k\le q^m} S_k \le (1+\epsilon)\varphi\left(q^m\right), \qquad \text{for } m \text{ large enough.} \tag{1.13}$$

On this event, for $q^m \le n \le q^{m+1}$, we have

$$S_n \le \max_{k\le q^{m+1}} S_k \le (1+\epsilon)\varphi\left(q^{m+1}\right) \le (1+\epsilon)\varphi(qn).$$

Therefore, almost surely,

$$\limsup_{n\to\infty} \frac{S_n}{\varphi(n)} \le (1+\epsilon)\lim_{n\to\infty}\frac{\varphi(qn)}{\varphi(n)} = (1+\epsilon)\sqrt{q}.$$

Let $q \downarrow 1$, we have almost surely,

$$\limsup_{n\to\infty} \frac{S_n}{\varphi(n)} \le 1+\epsilon.$$

Lower bound. Note that S_n is Gaussian with mean zero and variance n, thus

$$\begin{aligned}\mathbb{P}\left(S_n \ge (1-\epsilon)\varphi(n)\right) &= \int_{(1-\epsilon)\sqrt{2\log\log n}} \frac{1}{2\pi} e^{-y^2/2}\,\mathrm{d}y. \\ &\sim (2\pi)^{-1/2}(1-\epsilon)^{-1}(2\log\log n)^{-1/2}(\log n)^{-(1-\epsilon)^2}\end{aligned}$$

Fix some $q > 1$, but this time q will be (large) integer. Since to get independent events, we will we look at

$$\begin{aligned}&\mathbb{P}\left(S\left(q^{m+1}\right) - S\left(q^m\right) \ge (1-\epsilon)\varphi\left(q^{m+1}-q^m\right)\right) \\ &\sim (2\pi)^{-1/2}(1-\epsilon)^{-1}(2\log(m\log q))^{-1/2}(m\log q)^{-(1-\epsilon)^2}.\end{aligned}$$

Therefore,

$$\sum_{m=1}^{\infty} \mathbb{P}\left(S\left(q^{m+1}\right)-S\left(q^{m}\right) \geq (1-\epsilon)\varphi\left(q^{m+1}-q^{m}\right)\right)=\infty\,.$$

By Borel-Cantelli Lemma, we have almost surely

$$S\left(q^{m+1}\right)-S\left(q^{m}\right) \geq (1-\epsilon)\varphi\left(q^{m+1}-q^{m}\right), \qquad \text{i.o..}$$

By (1.13), we have almost surely

$$S\left(q^{m}\right) \geq -(1+\epsilon)\varphi\left(q^{m}\right), \qquad \text{for } m \text{ large.}$$

Combining these two, we have almost surely,

$$S\left(q^{m+1}\right) \geq (1-\epsilon)\varphi\left(q^{m+1}-q^{m}\right)-(1+\epsilon)\varphi\left(q^{m}\right), \qquad \text{i.o..}$$

Therefore, almost surely

$$\limsup_{n\to\infty} \frac{S\left(q^{m}\right)}{\varphi\left(q^{m}\right)} \geq \sqrt{\frac{q-1}{q}}(1-\epsilon)-\frac{1+\epsilon}{\sqrt{q}}\,.$$

Let $q \to \infty$, we have almost surely,

$$\limsup_{n\to\infty} \frac{S\left(q^{m}\right)}{\varphi\left(q^{m}\right)} \geq 1-\epsilon\,. \qquad \square$$

1.7 Sets of Convergence

Let $(X_n, \mathcal{F}_n)_{n\geq 0}$ be a adapted real-valued stochastic process. Denote by

$$\{X_n \to\}$$

the set of sample points for which $\lim_n X_n$ exists and is finite. Let us write $A \subset B$ a.s. when $\mathbb{P}\left(1_A \leq 1_B\right)=1$, and $A=B$ a.s. if $\mathbb{P}(A\Delta B)=0$.

If $(X_n, \mathcal{F}_n)_{n\geq 0}$ is a submartingale and $\sup_n \mathbb{E}X_n^+ < \infty$, then according to Martingale convergence theorem, we have

$$\{X_n \to\} = \Omega \quad \text{a.s..}$$

In this section, let us consider the structure of sets $\{X_n \to\}$ of convergence for submartingales when the hypothesis $\sup_n \mathbb{E}X_n^+ < \infty$ is not satisfied.

However, we need some other consitions. In the first resule for martingale, we will assume that the martingale difference sequence $(\Delta X_n)_{n\geq 1}$ is domainted by a integrable r.v., in other words,

$$\mathbb{E} \sup_{n\in\mathbb{N}} |\Delta X_n| < \infty\,, \tag{1.14}$$

It's evdently that (1.14) is satisfied if there exists a constant C so that

$$|\Delta X_n| \leq C < \infty\,,\ \text{for all } n \geq 1 \quad \text{a.s..}$$

Our first result shows that martingales with bounded increments either converge or oscillate between ∞ and $-\infty$.

Theorem 1.38. *Let $\{X_n, \mathcal{F}_n\}_{n\geq 0}$ be a martingale. If $\{\Delta X_n\}_{n\geq 1}$ is domainted by an integrable r.v., i.e., (1.14) holds, then*

$$\left\{\inf_n X_n = -\infty, \sup X_n = +\infty\right\} \cup \{X_n \to\} = \Omega \quad a.s..$$

Proof. Since $\{-X_n, \mathcal{F}_n\}_{n\geq 0}$ is also a martingale, it suffices to show

$$\{\sup X_n < \infty\} = \{X_n \to\} \quad \text{a.s..}$$

Then we apply the conclution to $\{-X_n\}$, and get

$$\{\inf X_n > -\infty\} = \{X_n \to\} \quad \text{a.s..}$$

which implies the desired result.

The inclusion $\{X_n \to\} \subset \{\sup X_n < \infty\}$ is evident. To establish the opposite inclusion, take some $k \in \mathbb{N}$, let

$$\tau_k = \inf\{n \in \mathbb{N}_0 : X_n \geq k\}\ . \tag{1.15}$$

we consider the stopped process $X^{\tau_k} = \{X_{\tau_k \wedge n}\}$, which is a $\mathfrak{F}$-submartingale. Note that

$$X_{\tau_k \wedge n} \leq |X_0| + k + \Delta X_{\tau_k} 1_{\{1 \leq \tau_k < \infty\}} \quad \text{for all } n \text{ a.s..} \tag{1.16}$$

Since X satisfies (1.14), $\sup_n \mathbb{E} X^+_{\tau_k \wedge n}$ is finite. Applying Theorem 1.25, the $\mathfrak{F}$-submartingale $X^{\tau_k} = \{X_{\tau_k \wedge n}\}$ converges. Note that on $\{\tau_k = \infty\}$ we have $X_{\tau_k \wedge n} = X_n$, so

$$\{\tau_k = \infty\} \subset \{X_n \to\} \quad \text{a.s..}$$

Since k is arbitary, we have

$$\bigcup_{k \in \mathbb{N}} \{\tau_k = \infty\} = \{\sup X_n < \infty\} \subset \{X_n \to\} \quad \text{a.s..}$$

We have completed the proof. □

A For Submartingales

From the proof above, we can see that, the main result

$$\{\sup X_n < \infty\} = \{X_n \to\} \quad \text{a.s.}$$

holds since the stopped process X^{τ_k} converges for each $k \in \mathbb{N}$. To guarantee this, we made the conditions of convergence theorem for submartingales satisfied, by contorlling the expectation of the RHS in (1.16).

Firstly, we only need the condition that X is a submartingale to guarantee that $X^{\tau_k}, k \in \mathbb{N}$ are submartingales. Secondly, to ensure the convergence for these submartingles,

$$\mathbb{E}\left[\Delta X_{\tau_k} 1_{\{1 \leq \tau_k < \infty\}}\right] < \infty\,, \quad \text{for any } k \in \mathbb{N}\,. \tag{1.17}$$

is enough. Clearly, (1.14) implies (1.17). Thus we have,

Lemma 1.39. *If $\{X_n, \mathcal{F}_n\}_{n \in \mathbb{N}_0}$ is a submartingale satisfying (1.17), then*

$$\{\sup X_n < \infty\} = \{X_n \to\} \quad \textit{a.s..}$$

To give a more comprehensive description of the set of convergence of submartingales, we shall study the increasing process associated to the submartingales.

Theorem 1.40. *Let $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ be a submartingale and*

$$X_n = M_n + A_n \qquad \text{for } n \geq 0$$

its Doob's decomposition. Since A is increasing, $A_\infty := \lim_n A_n$ is well-defined. Then the following propositions holds.

(i) If X is a nonnegative submartingale, then

$$\{A_\infty < \infty\} \subset \{X_n \to\} \subset \{\sup X_n < \infty\} \quad a.s..$$

(ii) If X satisfies (1.17) then

$$\{\sup X_n < \infty\} = \{X_n \to\} \subset \{A_\infty < \infty\} \quad a.s..$$

(iii) If X is a nonnegative submartingale and X satisfies (1.17) then

$$\{\sup X_n < \infty\} = \{X_n \to\} = \{A_\infty < \infty\} \quad a.s..$$

Proof. (i). The second inclusion is obvious, to establish the first inclusion, we introduce the times

$$\sigma_k = \inf\{n \in \mathbb{N}_0 : A_{n+1} > k\}$$

where $k \in \mathbb{N}$. Then $A_{\sigma_k \wedge n} \leq k$ for each $n \in \mathbb{N}_0$. Since A is predictable, σ_k is a stopping time to $\mathfrak{F}$. Therefore, the stopped process $X^{\sigma_k} = \{X_{\sigma_k\wedge n}\}$ is a non-negative $\mathfrak{F}$-submartingale, and $M^{\sigma_k} = \{M_{\sigma_k\wedge n}\}$ is a $\mathfrak{F}$-martingale,

$$\mathbb{E}X_{\sigma_k\wedge n} = \mathbb{E}M_0 + \mathbb{E}A_{\sigma_k\wedge n} \leq \mathbb{E}X_0 + k\,, \text{ for all } n\,. \tag{1.18}$$

In other words, X^{σ_k} is L^1 bounded. By the martingale convergence theorem, X^{σ_k} converges almost surely. On $\{\sigma_k = \infty\}$, we have $X_n = X_{\sigma_k\wedge n}$ for all n, so

$$\{A_\infty \leq k\} = \{\sigma_k = \infty\} \subset \{X_n \to\} \quad \text{a.s..}$$

Therefore

$$\{A_\infty < \infty\} = \bigcup_{k\in\mathbb{N}} \{A_\infty \le k\} \subset \{X_n \to\} .$$

(ii). The first equation follows from Lemma 1.39. To show the second, note that, for τ_k defined in (1.15) and fixed $n \in \mathbb{N}$,

$$\mathbb{E}A_{\tau_k\wedge n} = \mathbb{E}X_{\tau_k\wedge n} - \mathbb{E}X_0 \le 2\mathbb{E}|X_0| + k + \mathbb{E}\left[\Delta X_{\tau_k} 1_{\{1\le\tau_k<\infty\}}\right] ,$$

and letting $n \to \infty$, by monotone convergence theorem,

$$\mathbb{E}A_{\tau_k} < \infty .$$

Hence $\{\tau_k = \infty\} \subset \{A_\infty < \infty\}$, a.s. and we obtain the required conclusion since $\cup_k \{\tau_k = \infty\} = \{\sup X_n < \infty\}$.

(iii). This is an immediate consequence of (i) and (ii). $\square$

Remark 1.23. The hypothesis in (i) that X is non-negative can be replaced by the hypothesis $\mathbb{E}\left[\sup_n X_n^-\right] < \infty$, since we only use non-negativity in (1.18) to deduce that $X^{\sigma_k} = \{X_{\sigma_k\wedge n}\}$ is convergent.

Example 1.25. Let $(\xi_n)_{n\ge1}$ be a non-negative integrable process adapted to $(\mathcal{F}_n)_{n\in\mathbb{N}}$. Set $S_0 = 0$ and $\mathcal{F}_0 = \{\varnothing, \Omega\}$, $S_n = \xi_1 + \cdots + \xi_n$ for $n \in \mathbb{N}$. Clearly, $(S_n)_{n\in\mathbb{N}_0}$ is a non-negative $\mathfrak{F}$-submartingale. Then by Doob's decomposition and Theorem 1.40,

$$\left\{\sum_{n=1}^{\infty} \mathbb{E}\left(\xi_n|\mathcal{F}_{n-1}\right) < \infty\right\} \subset \left\{\sum_{n=1}^{\infty} \xi_n < \infty\right\} \quad \text{a.s..}$$

and if, in addition, $\mathbb{E}\left[\sup_n \xi_n\right] < \infty$, then

$$\left\{\sum_{n=1}^{\infty} \mathbb{E}\left(\xi_n|\mathcal{F}_{n-1}\right) < \infty\right\} = \left\{\sum_{n=1}^{\infty} \xi_n < \infty\right\} \quad \text{a.s..}$$

Now let $\xi_n = 1_{B_n}$, we get the following theorem, which genelize the B-C lemma.

Corollary 1.41 (Second Borel-Cantelli Lemma II). *Let $\mathfrak{F} = \{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$ be a filtration. $\{B_n\}_{n\in\mathbb{N}}$ is a sequence of events with $B_n \in \mathcal{F}_n$. Then*

$$\left\{\sum_{n=1}^{\infty} \mathbb{P}\left(B_n|\mathcal{F}_{n-1}\right) = \infty\right\} = \{B_n \ i.o.\ \} \quad a.s..$$

B For Square-Integrable Martingales

Let's see an example first.

Example 1.26 (Kolmogorov's two-series theorem). Let $\xi_1, \xi_2, \ldots$ be a sequence of independent r.v.'s with $\mathbb{E}\xi_n = 0$ and $\mathbb{E}\xi_n^2 < \infty$. By the Kolmogorov's two-series theorem, which we learned in the courese of probability theory, the series $\sum_i \xi_i$ converges a.s. if and only if $\sum_i \mathbb{E}\xi_i^2 < \infty$. Note that the sequence $S = (S_n)_{n\geq 0}$ with $S_n = \sum_{i=1}^n \xi_i$ is a square-integrable martingale with respect to the filtration generated by (ξ_n). The quadratic variation process is given by $\langle S\rangle_n = \sum_{i=1}^n \mathbb{E}\xi_i^2$, and the theorem just stated can be interpreted as follows:

$$\{\langle S\rangle_\infty < \infty\} = \{S_n \to\} \qquad \text{a.s.},$$

where $\langle S\rangle_\infty = \lim_n \langle S\rangle_n$.

Thus we try to genelize the result to all the square-integrable martingales.

Theorem 1.42. *Let $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ be a square-integrable martingale. Let $\langle X\rangle$ be the quadratic variation process of X. Then*

$$\{\langle X\rangle_\infty < \infty\} \subset \{X_n \to\} \qquad a.s..$$

If, in addition, $\mathbb{E}\left(\sup_n |\Delta X_n|^2\right) < \infty$, then

$$\{\langle X\rangle_\infty < \infty\} = \{X_n \to\} \qquad a.s..$$

Proof. The proof of the first proposition depends on the observation

$$\{X_n \to\} = \left\{X_n^2 \to,\ (X_n + 1)^2 \to\right\}.$$

Thus, consider the following two $\mathfrak{F}$-submartingales, $X^2 = \left(X_n^2\right)_{n\geq 0}$ and $(X+1)^2 = \left((X_n+1)^2\right)_{n\geq 0}$. We notice that the quadratic variation process of $(X+1)^2$, coincides with X^2. Applying Theorem 1.40, we have

$$\{\langle X\rangle_\infty < \infty\} \subset \left\{X_n^2 \to,\ (X_n+1)^2 \to\right\} = \{X_n \to\}\ .$$

If in addition, $\mathbb{E}\sup_n |\Delta X_n|^2 < \infty$, by Theorem 1.40 (iii), it is sufficient to show that X^2 satisfies (1.17). Let $\tau_k = \inf\left\{n \geq 0 : X_n^2 > k\right\}$. Then, on the set $\{1 \leq \tau_k < \infty\}$,

$$\begin{aligned}\left|\Delta X_{\tau_k}^2\right| &= \left|X_{\tau_k}^2 - X_{\tau_k-1}^2\right| = |X_{\tau_k} + X_{\tau_k-1}| \cdot |X_{\tau_k} - X_{\tau_k-1}| \\ &\leq |X_{\tau_k} - X_{\tau_k-1}|^2 + 2\,|X_{\tau_k-1}| \cdot |X_{\tau_k} - X_{\tau_k-1}| \\ &\leq (\Delta X_{\tau_k})^2 + 2k^{1/2}\,|\Delta X_{\tau_k}|\ ,\end{aligned}$$

whence

$$\begin{aligned}\mathbb{E}\left|\Delta X_{\tau_k}^2\right| 1_{\{\tau_k<\infty\}} &\leq \mathbb{E}\sup|\Delta X_n|^2 + 2k^{1/2}\mathbb{E}\sup|\Delta X_n| \\ &\leq \mathbb{E}\sup|\Delta X_n|^2 + 2\sqrt{k\mathbb{E}\sup|\Delta X_n|^2} < \infty\,.\end{aligned}$$

This completes the proof of the theorem. □

Exercise 1.8. Suppose $X = (X_n)_{n\in\mathbb{N}_0}$ is a square-integrable $\mathfrak{F}$-martingale.

(i) Show that $\mathbb{E}\left(\sup|X_n|^2\right) \leq 4\mathbb{E}\langle X\rangle_\infty$ by using Doob's L^p maximum inequality.

(ii) By using (i) and stopping X to show that $\{\langle X\rangle_\infty < \infty\} \subset \{X_n \to\}$.

(iii) Show that $\mathbb{E}\left(\sup_n |X_n|\right) \leq 3\mathbb{E}\langle X\rangle_\infty^{1/2}$ by using Doob's inequality and stopping X. This ia a slightly better result than (i).

As an illustration of Theorem 1.42, we present the following result, which can be considered as a kind of the *law of large numbers* for square-integrable martingales.

Theorem 1.43. *Let $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ be a square-integrable martingale. Let $f \geq 1$ be a increasing function on $(0,\infty)$ with $\int_0^\infty f(t)^{-2}\,\mathrm{dt} < \infty$. Then*

$$\frac{X_n}{f\left(\langle X\rangle_n\right)} \to 0 \quad on \quad \{\langle X\rangle_\infty = \infty\} \quad a.s..$$

Proof. Evidently, $f(t) \to \infty$ as $t \to \infty$, so

$$f\left(\langle X\rangle_n\right) \to \infty \quad \text{on} \quad \{\langle X\rangle_\infty = \infty\} \quad \text{a.s..}$$

By Kronecker's lemma, it suffices to show that

$$\sum_{n=1}^\infty \frac{\Delta X_n}{f\left(\langle X\rangle_n\right)} < \infty \quad \text{on} \quad \{\langle X\rangle_\infty = \infty\} \quad \text{a.s..}$$

Let $H_n = f\left(\langle X\rangle_n\right)^{-1}$ for $n \geq 1$. Then H is locally bounded and predictable, so $H \cdot X$ is a martingale, given by

$$(H\cdot X)_n = \sum_{i=1}^n \frac{\Delta X_i}{f\left(\langle X\rangle_i\right)}\,, \text{ for } n \geq 1\,.$$

Clearly, $H \cdot X$ is a square-integrable $\mathfrak{F}$-martingale, and let $\langle H\cdot X\rangle$ be it's the quadratic process, then

$$\begin{aligned}
\langle H\cdot X\rangle_n &= \sum_{i=1}^n E[|\Delta(H\cdot X)_i|^2|\mathcal{F}_{i-1}] \\
&= \sum_{i=1}^n \frac{E(|\Delta X_i|^2|\mathcal{F}_{i-1})}{f\left(\langle X\rangle_i\right)^2} \\
&= \sum_{i=1}^n \frac{\Delta\langle X\rangle_i}{f\left(\langle X\rangle_i\right)^2}\,.
\end{aligned}$$

Our hypotheses on f imply that

$$\langle H\cdot X\rangle_\infty = \sum_{n=1}^\infty \frac{\Delta\langle X\rangle_n}{f\left(\langle X\rangle_n\right)^2} \leq \sum_{n=1}^\infty \int_{[\langle X\rangle_{n-1},\langle X\rangle_n)} f(t)^{-2}dt < \infty \text{ a.s..}$$

By Theorem 1.42, we have result:

$$\sum_{n=1}^\infty \frac{\Delta X_n}{f\left(\langle X\rangle_n\right)} < \infty \quad \text{a.s..}$$

Since only on $\{\langle X\rangle_\infty = \infty\}$, we have $f(\langle X\rangle_n) \uparrow \infty$, in which case can we use the Kronecker's lemma, so we get

$$\frac{X_n}{f(\langle X\rangle_n)} \to 0 \quad \text{on} \quad \{\langle X\rangle_\infty = \infty\} \quad \text{a.s..}$$

□

Example 1.27 (Rate of Convergence in SLLN)**.** Let $\xi_1, \xi_2, \ldots$ be i.i.d with $E\xi_1 = \mu$ and $E\xi_1^2 = \sigma^2 > 0$. Let $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$ be the filtration generated by $\xi_1, \xi_2, \cdots$, let $S_n = \sum_{i=1}^n \xi_i$ for $n \geq 1$ and $S_0 = 0$. Then $(S_n - n\mu, \mathcal{F}_n)_{n\geq 0}$ is a square integrable $\mathfrak{F}$-martingale with quadrativ variation process $(n\sigma^2)_{n\geq 0}$.

Given $\epsilon > 0$, set $f(t) = \left(t \log^{1+2\epsilon} t\right)^{1/2} \vee 1$ for $t > 0$. Then f satisfies the hypotheses of Theorem 1.43. Clearly, $n\sigma^2 \to \infty$,

$$\frac{S_n - n\mu}{n^{\frac{1}{2}} \log^{\frac{1}{2}+\epsilon} n} \to 0 \quad \text{a.s..}$$

Corollary 1.44 (Second Borel-Cantelli Lemma, III)**.** *Suppose that $(B_n)_{n\geq 1}$ is adapted to $\mathfrak{F} = (\mathcal{F}_n)_{n\geq 0}$. Let $p_n \coloneqq \mathbb{P}(B_n|\mathcal{F}_{n-1})$ for all $n \geq 1$, then*

$$\frac{\sum_{m=1}^n 1_{B_m}}{\sum_{m=1}^n p_m} \to 1 \qquad a.s. \ \ on \ \Big\{\sum_m p_m = \infty\Big\}.$$

Proof. Observe that

$$\frac{\sum_{m=1}^n 1_{B(m)}}{\sum_{m=1}^n p_m} - 1 = \frac{\sum_{m=1}^n 1_{B(m)} - p_m}{\sum_{m=1}^n p_m},$$

and $(1_{B_m} - p_m)_{m\geq 0}$ is a martingale difference with respect to $\mathfrak{F}$. Thus we can define a martingale $X = (X_n)_{n\in\mathbb{N}_0}$, so that

$$\frac{\sum_{m=1}^n 1_{B(m)}}{\sum_{m=1}^n p_m} - 1 = \frac{X_n}{\sum_{m=1}^n p_m}.$$

Evidently, X is square integrable with quadratic variation

$$\begin{aligned}\langle X\rangle_n &= \sum_{m=1}^n E\left[(\Delta X_m)^2 | \mathcal{F}_{m-1}\right] \\ &= \sum_{m=1}^n E\left[(1_{B_m} - p_m)^2 | \mathcal{F}_{m-1}\right] = \sum_{m=1}^n p_m - p_m^2.\end{aligned}$$

On $\{\langle X\rangle_\infty < \infty\}$, (X_n) converges almost surely by Theorem 1.42, so

$$\frac{X_n}{\sum_{m=1}^n p_m} \to 0 \quad \text{a.s. on}\, \{\langle X\rangle_\infty < \infty\} \cap \Big\{ \sum_m p_m = \infty \Big\}\,.$$

Apply Theorem 1.43 with $f(t) = t \vee 1$, we get

$$\frac{X_n}{\sum_{m=1}^n p_m} \to 0 \quad \text{a.s. on} \;\; \{\langle X\rangle_\infty = \infty\}\;.$$

and the desired conclusion follows from $\{\langle X\rangle_\infty = \infty\} \subset \{\sum_m p_m = \infty\}$. □

Example 1.28 (Bernard Friedman's Urn)**.** Consider a variant of Polya's urn in which we add a balls of the color drawn and b balls of the opposite color where $a \geq 0$ and $b > 0$. We will show that if we start with g green balls and r red balls, where $g, r > 0$, then the fraction of green balls

$$g_n \to \frac{1}{2}\,.$$

Let G_n and R_n be the number of green and red balls after the n th draw is completed. Let B_n be the event that the n th ball drawn is green, and let D_n be the number of green balls drawn in the first n draws. It follows from Theorem 1.44 that

$$D_n / \sum_{m=1}^n g_{m-1} \to 1 \qquad \text{a.s. on} \qquad \sum_{m=1}^\infty g_{m-1} = \infty \tag{$\star$}$$

which always holds since $g_m \geq g/(g+r+(a+b)m)$. At this point, the argument breaks into three cases.

Case 1, $a = b = c$. In this case, the result is trivial since we always add c balls of each color.

Case 2, $a > b$. We begin with the observation

$$g_{n+1} = \frac{G_{n+1}}{G_{n+1} + R_{n+1}} = \frac{g + aD_n + b\,(n - D_n)}{g + r + n(a+b)} \tag{$*$}$$

If $\limsup g_n \le x$ then $(\star)$ implies $\limsup D_n/n \le x$ and (since $a > b$)

$$\limsup_{n\to\infty} g_{n+1} \le \frac{ax + b(1-x)}{a+b} = \frac{b + (a-b)x}{a+b}$$

The right-hand side is a linear function with slope < 1 and fixed point at $1/2$ so starting with the trivial upper bound $x = 1$ and iterating we conclude that limsup $g_n \le 1/2$. Interchanging the roles of red and green shows $\liminf g_n \ge 1/2$ and the result follows.

Case 3, $a < b$. The result is easier to believe in this case since we are adding more balls of the type not drawn but is a little harder to prove. The trouble is that when $b > a$ and $D_n \le xn$, the right-hand side of $(*)$ is maximized by taking $D_n = 0$, so we need to also use the fact that if r_n is fraction of red balls, then

$$r_{n+1} = \frac{R_{n+1}}{G_{n+1} + R_{n+1}} = \frac{r + bD_n + a\left(n - D_n\right)}{g + r + n(a+b)}$$

Combining this with the formula for g_{n+1}, it follows that if $\limsup g_n \le x$ and $\limsup r_n \le y$ then

$$\limsup_{n\to\infty} g_n \le \frac{a(1-y) + by}{a+b} = \frac{a + (b-a)y}{a+b}$$
$$\limsup_{n\to\infty} r_n \le \frac{bx + a(1-x)}{a+b} = \frac{a + (b-a)x}{a+b}$$

Starting with the trivial bounds $x = 1, y = 1$ and iterating (observe the two upper bounds are always the same), we conclude as in Case 2 that both limsups are $\le 1/2$.

C For Square-Integrable Submartingales with Bounded Increments

We shall let $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ is a square-integrable submartingale with bounded increments. Clearly, under the condition (ΔX_n) is bounded, $X_0 \in L^2$ implies (X_n) is square-integrable. Without loss of generality, we assume that $X_0 = 0$.

Theorem 1.45. *Let $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}_0}$ be a submartingale with $X_0 = 0$ and bounded increments, i.e., $|\Delta X_n| \le C$. Let $X = M + A$ be its Doob decomposition. Then*

$$\{\langle M\rangle_\infty + A_\infty < \infty\} = \{X_n \to\} \quad a.s.,$$

or, equivalently,

$$\left\{\sum_{n=1}^{\infty} \mathbb{E}\left[\Delta X_n + (\Delta X_n)^2 \,|\mathcal{F}_{n-1}\right] < \infty\right\} = \{X_n \to\} \quad a.s..$$

Proof. For $n \ge 1$, we have

$$A_n = \sum_{k=1}^{n} \mathbb{E}\left(\Delta X_k|\mathcal{F}_{k-1}\right)\,,\ M_n = \sum_{k=1}^{n} \Delta X_k - \mathbb{E}\left(\Delta X_k|\mathcal{F}_{k-1}\right)\,.$$

and

$$\begin{aligned}\langle M\rangle_n &= \sum_{k=1}^{n} \mathbb{E}\left[\{\Delta X_k - \mathbb{E}\left(\Delta X_k|\mathcal{F}_{k-1}\right)\}^2|\mathcal{F}_{k-1}\right] \\ &= \sum_{k=1}^{n} \mathbb{E}\left[(\Delta X_k)^2|\mathcal{F}_{k-1}\right] - \mathbb{E}(\Delta X_k|\mathcal{F}_{k-1})^2\,.\end{aligned}$$

Note that M_n is a square-integrable martingale with bounded increments, by Theorem 1.42 we have

$$\{\langle M\rangle_\infty < \infty\} = \{M_n \to\} \quad \text{a.s..}$$

Since A_n is a increasing process, by Theorem 1.40, we have

$$\{X_n \to\} \subset \{A_\infty < \infty\} = \{A_n \to\} \quad \text{a.s..}$$

Thus

$$\{\langle M\rangle_\infty + A_\infty < \infty\} = \{X_n \to\} \quad \text{a.s..}$$

Moreover, the convergence of A_n imples the convergences of the series

$$\sum_k \mathbb{E}(\Delta X_k|\mathcal{F}_{k-1})^2\,.$$

and the desired result follows. ∎

Kolmogorov's three-series theorem gives a necessary and sufficient condition for the convergence, with probability 1, of a series $\sum \xi_n$ of independent random variables. The following theorem describes sets of convergence of $\sum \xi_n$ without the assumption that the random variables $\xi_1, \xi_2, \ldots$ are independent.

Exercise 1.9 (Kolmogorov's Three-Series Theorem). Let $(\xi_n, \mathcal{F}_n)_{n\geq 1}$ be an adapted stochastic sequence. Let $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Take any $C > 0$, then

$$\Big\{ \sum_n \xi_n \quad \text{converges} \Big\} = A \quad \text{a.s.},$$

where A is the set of sample points, for which the following three series

$$\sum \mathbb{P}\left(|\xi_n| \geq c | \mathcal{F}_{n-1}\right), \quad \sum \mathbb{E}\left(\bar{\xi}_n | \mathcal{F}_{n-1}\right), \quad \sum \operatorname{Var}\left(\bar{\xi}_n | \mathcal{F}_{n-1}\right)$$

converge, where $\bar{\xi}_n := \xi_n 1_{\{|\xi_n| \leq C\}}$.

1.8 Applications(II): Locally Absoulute Continuity

Let $(\Omega, \mathcal{F})$ be a probability space with a filtration $\mathfrak{F} = (\mathcal{F}_n)_{n\in\mathbb{N}_0}$ so that that $\mathcal{F} = \mathcal{F}_\infty := \sigma(\cup_n \mathcal{F}_n)$. Let's suppose that two probability measures $\mathbb{P}$ and $\tilde{\mathbb{P}}$ are given on $(\Omega, \mathcal{F})$. Let's write

$$\mathbb{P}_n = \mathbb{P}|_{\mathcal{F}_n}, \quad \tilde{\mathbb{P}}_n = \tilde{\mathbb{P}}|_{\mathcal{F}_n}$$

for the restrictions of these measures to $\mathcal{F}_n$.

Definition 1.4. We say that $\tilde{\mathbb{P}}$ is **locally absolutely continuous** with respect to $\mathbb{P}$ and write $\tilde{\mathbb{P}} \overset{\text{loc}}{\ll} \mathbb{P}$, if

$$\tilde{\mathbb{P}}_n \ll \mathbb{P}_n$$

for every $n \in \mathbb{N}_0$.

The fundamental question that we shall consider in this subsection is the determination of conditions under which local absolute continuity $\tilde{\mathbb{P}} \overset{\text{loc}}{\ll} \mathbb{P}$ implies one of the properties $\tilde{\mathbb{P}} \ll \mathbb{P}, \tilde{\mathbb{P}} \perp \mathbb{P}$. It will become clear that martingale theory is the mathematical apparatus that lets us give definitive answers to these questions.

Theorem 1.46. *Suppose* $\tilde{\mathbb{P}} \overset{loc}{\ll} \mathbb{P}$. *Let*

$$X_n := \frac{\mathrm{d}\tilde{\mathbb{P}}_n}{\mathrm{d}\mathbb{P}_n} \text{ for all } n\,, \text{ and } X_\infty := \limsup_{n\to\infty} X_n\,.$$

Then $(\{X_n, \mathcal{F}_n\}_{n\in\mathbb{N}_0}, \mathbb{P})$ *is a non-negative martingale and hence* $X_n \to X_\infty$ $\mathbb{P}$-*a.s.*[2] *with* $\mathbb{E}|X_\infty| < \infty$. *The key thing is that for any* $A \in \mathcal{F}$, *we have*

$$\tilde{\mathbb{P}}(A) = \int_A X_\infty \,\mathrm{d}\mathbb{P} + \tilde{\mathbb{P}}(A \cap \{X_\infty = \infty\})\,. \tag{1.19}$$

Remark 1.24. Define $\tilde{\mathbb{P}}_c$ and $\tilde{\mathbb{P}}_s$ by letting $\tilde{\mathbb{P}}_c(A) := X_\infty.\mathbb{P}$ and $\tilde{\mathbb{P}}_s(A) := \tilde{\mathbb{P}}(A \cap \{X_\infty = \infty\})$ for all $A \in \mathcal{F}$. Then $\tilde{\mathbb{P}} = \tilde{\mathbb{P}}_c + \tilde{\mathbb{P}}_s$ gives the *Lebesgue decomposition* of $\tilde{\mathbb{P}}$ with respect to $\mathbb{P}$.

Proof. It is clear that X_n is $\mathcal{F}_n$ measurable; and if $A \in \mathcal{F}_n$, then

$$\begin{aligned}\int_A X_{n+1}\,\mathrm{d}\mathbb{P} &= \int_A \frac{\mathrm{d}\tilde{\mathbb{P}}_{n+1}}{\mathrm{d}\mathbb{P}_{n+1}}\,\mathrm{d}\mathbb{P} = \tilde{\mathbb{P}}_{n+1}(A) = \tilde{\mathbb{P}}_n(A) \\ &= \int_A \frac{\mathrm{d}\tilde{\mathbb{P}}_n}{\mathrm{d}\mathbb{P}_n}\,\mathrm{d}\mathbb{P} = \int_A X_n\,\mathrm{d}\mathbb{P}\,.\end{aligned}$$

It follows that, with respect to $\mathbb{P}$, the stochastic sequence $(X_n, \mathcal{F}_n)_{n\geq 0}$ is a non-negaive martingale. Applying the martingale convergence theorem, we get $X_n \to X_\infty$ $\mathbb{P}$-a.s. and $\mathbb{E}|X_\infty| < \infty$. We want to check that the equality (1.19) in the theorem holds.

[2] Since there appers different probability measures, we need to keep track of the measure to which the a.s. refers.

Step 1. If $\tilde{\mathbb{P}} \ll \mathbb{P}$. Then esists $\xi \in L^1$ so that $\tilde{\mathbb{P}} = \xi.\mathbb{P}$ Then

$$X_n = \frac{\mathrm{d}\tilde{\mathbb{P}}_n}{\mathrm{d}\mathbb{P}_n} = \frac{\mathrm{d}(\xi.\mathbb{P})|_{\mathcal{F}_n}}{\mathrm{d}\mathbb{P}|_{\mathcal{F}_n}} = \mathbb{E}(\xi|\mathcal{F}_n)\,, \text{ for all } n \ \mathbb{P}\text{-a.s..}$$

By Theorem 1.29,

$$X_\infty = \mathbb{E}(\xi|\mathcal{F}) = \xi \quad \mathbb{P}\text{-a.s..}$$

Step 2. To find a probability measure ρ on $(\Omega, \mathcal{F})$ so that $\tilde{\mathbb{P}} \ll \rho$, let

$$\rho = (\tilde{\mathbb{P}} + \mathbb{P})/2,\ \rho_n = \left(\tilde{\mathbb{P}}_n + \mathbb{P}_n\right)/2\,, \text{ for all } n\,.$$

Clearly, $\tilde{\mathbb{P}}_n \ll \mathbb{P}_n \sim \rho_n$ for all n and $\tilde{\mathbb{P}} \ll \rho$, $\mathbb{P} \ll \rho$. Let

$$Y_n = \frac{\mathrm{d}\tilde{\mathbb{P}}_n}{\mathrm{d}\rho_n}\,,\ Z_n = \frac{\mathrm{d}\mathbb{P}_n}{\mathrm{d}\rho_n}\,,\ Y = \frac{\mathrm{d}\tilde{\mathbb{P}}}{\mathrm{d}\rho}\,,\ Z = \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\rho}\,.$$

Note that $Y_n + Z_n = 2$, so by Step 1, $\{Y_n\}$ and $\{Z_n\}$ are non-negative bounded $\mathfrak{F}$-martingales on $(\Omega, \mathcal{F}, \rho)$, with ρ-a.s. limits Y and Z. By $\mathbb{P}_n \sim \rho_n$,

$$Z_n > 0 \ \text{ for all } n \quad \rho\text{-a.s.,}$$

and as we konw, $X_n Z_n = Y_n$ ρ-a.s for all n. Letting $n \to \infty$, on $\{Z > 0\}$ we have

$$X_\infty Z = Y\,, \quad \rho\text{-a.s..} \tag{1.20}$$

and note that $Y + Z = 2$, ρ-a.s. we have $\rho(Z = 0, Y = 0) = 0$, hence on $\{Z = 0\}$ we have

$$X_\infty = \infty\,, \quad \rho\text{-a.s..}$$

Given $A \in \mathcal{F}_\infty$ we have

$$\begin{aligned}
\tilde{\mathbb{P}}(A) = \int_A Y \,\mathrm{d}\rho &= \int_A X_\infty Z 1_{\{Z>0\}} \,\mathrm{d}\rho + \int_A Y 1_{\{Z=0\}} \,\mathrm{d}\rho \\
&= \int_A X_\infty 1_{\{Z>0\}} \,\mathrm{d}\mathbb{P} + \int_A Y 1_{\{X_\infty=\infty\}} \,\mathrm{d}\rho \\
&= \int_A X_\infty \,\mathrm{d}\mathbb{P} + \tilde{\mathbb{P}}(A \cap \{X_\infty = \infty\})\,.
\end{aligned}$$

In the last equality we used that $\mathbb{P}(Z = 0) = \mathbb{P}(X_\infty = \infty) = 0$. $\square$

Remark 1.25. Firstly, to get (1.20) we must assume $\{Z > 0\}$. In other words,

$$X_\infty Z = Y \ \rho\text{-a.s.}$$

is NOT ture. Secondly, by martingale convergence we konw that the limit of $\{X_n\}$ is $\mathbb{P}$-almost surely unique. Generaly speaking, let X'_∞ be one of the limit, then the theorem becomes false because (1.20) may not hold, we only have, that on $\{Z > 0\}$

$$X_\infty Z = Y\,, \quad \mathbb{P}\text{-a.s.}.$$

The Lebesgue decomposition implies the following useful tests for absolute continuity or singularity of locally absolutely continuous probability measures.

Corollary 1.47. *Let $\tilde{\mathbb{P}} \overset{loc}{\ll} \mathbb{P}$. Let X_∞ be same as in Theorem 1.46, then*

$$\tilde{\mathbb{P}} \ll \mathbb{P} \Leftrightarrow \mathbb{E}X_\infty = 1 \Leftrightarrow X_\infty < \infty \ \tilde{\mathbb{P}}\text{-}a.s.\,,$$

and

$$\tilde{\mathbb{P}} \perp \mathbb{P} \Leftrightarrow X_\infty = 0 \ \mathbb{P}\text{-}a.s. \Leftrightarrow X_\infty = \infty \ \tilde{\mathbb{P}}\text{-}a.s..$$

Example 1.29 (Kakutani Dichotomy for Infinite Product Measures)**.** Let $\tilde{\mathbb{P}}$ and $\mathbb{P}$ be two product probability measures on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}_{\mathbb{R}}^{\mathbb{N}})$. Let π_n be the n-th projection, i.e., $\pi_n(\omega) = \omega_n$, $\omega = (\omega_n) \in \mathbb{R}^{\mathbb{N}}$. Let $\tilde{\mathbb{P}}_n$ and $\mathbb{P}_n$ be the restrictions of $\tilde{\mathbb{P}}$ and $\mathbb{P}$ to $\mathcal{F}_n = \sigma\left(\pi_m : m \le n\right)$, and let

$$X_n = \frac{\mathrm{d}\tilde{\mathbb{P}}_n}{\mathrm{d}\mathbb{P}_n}\,.$$

Let $F_n(x) = \tilde{\mathbb{P}}\left(\pi_n \le x\right), G_n(x) = \mathbb{P}\left(\pi_n \le x\right)$. Suppose $F_n \ll G_n$ and let

$$\xi_n = \frac{\mathrm{d}F_n}{\mathrm{d}G_n}\,.$$

Evidently, random variables $\{\xi_n(\pi_n)\}$, defined on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}_{\mathbb{R}}^{\mathbb{N}}, \mathbb{P})$, is independent with

$$\mathbb{E}_{\mathbb{P}}\left[\xi_n(\pi_n)\right] = \int_{\mathbb{R}} \xi_n(x) G_n(\mathrm{d}x) = \int_{\mathbb{R}} \mathrm{d}F_n = 1\,, \text{ for each} n\,.$$

On the other hand, we have

$$X_n = \frac{\mathrm{d}\tilde{\mathbb{P}}_n}{\mathrm{d}\mathbb{P}_n} = \prod_{m=1}^{n} \xi_m(\pi_m)\,.$$

Thus, by Example 1.24, $\{X_n\}$ is a martingale with respect to $\{\mathcal{F}_n\}$ and defined on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}_{\mathbb{R}}^{\mathbb{N}}, \mathbb{P})$ so that $X_n \to X_\infty$, $\mathbb{P}$ -a.s., and

$$\mathbb{P}(X_\infty = 0) \in \{0, 1\}$$

and it follows from Theorem 1.47 that either $\tilde{\mathbb{P}} \ll \mathbb{P}$ or $\tilde{\mathbb{P}} \perp \mathbb{P}$. And we have a concrete criterion for which of the two alternatives occurs.

Theorem. $\tilde{\mathbb{P}} \ll \mathbb{P}$ *or* $\tilde{\mathbb{P}} \perp \mathbb{P}$, *according as*

$$\prod_n \int \sqrt{\xi_n}\,\mathrm{d}G_n \;\; > 0 \quad \textit{or} \;\; = 0\,.$$

1.9 Applications(III): Galton-Watson Trees

This section is taken from [7]. We start by studying a few basic properties of supercritical Galton-Watson trees. The main aim of this section is to introduce the notion of size-biased trees. In particular, we see in Subsection 1.3 how this allows us to prove the well-known Kesten - Stigum theorem. This notion of size-biased trees can be developed to study more complicated models.

A Galton-Watson Trees and Extinction Probabilities

We are interested in processes involving (rooted) trees. The simplest rooted tree is the regular rooted tree, where each vertex has a fixed number (say m, with $m > 1$) of offspring. For example, here is a rooted binary tree:

Let Z_n denote the number of vertices (also called particles or individuals) in the n-th generation, then $Z_n = m^n, \forall n \geq 0$.

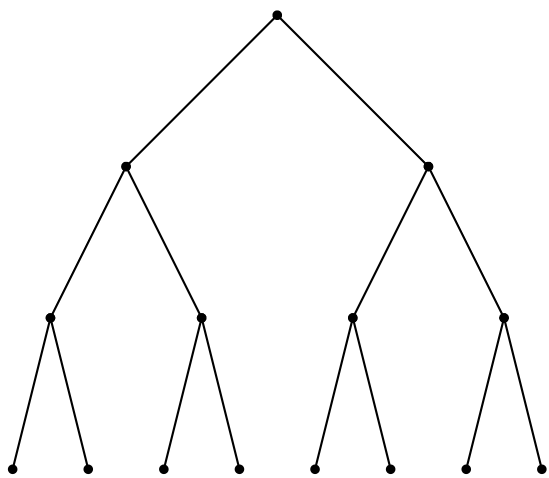


Figure 1.1: A rooted binary tree, i.e., $m = 2$.

In probability theory, we often encounter trees where the number of offspring of a vertex is *random*. The easiest case is when these random numbers are i.i.d., which leads to a Galton-Watson tree. A Galton–Watson tree starts with one initial ancestor (sometimes, it is possible to have several or even a random number of initial ancestors, in which case it will be explicitly stated). It produces a certain number of offspring according to a given probability distribution. The new particles form the first generation. Each of the new particles produces offspring according to the same probability distribution, independently of each other and of everything else in the generation. And the system regenerates.

We write p_i for the probability that a given particle has i children, $i \geq 0$; thus $\sum_{i=0}^{\infty} p_i = 1$. In the case of a regular m-ary tree, $p_i = \delta_{i,m}$. To avoid trivial discussions, we assume throughout that $p_0 + p_1 < 1$. As before, we write Z_n for the number of particles in the n -th generation.

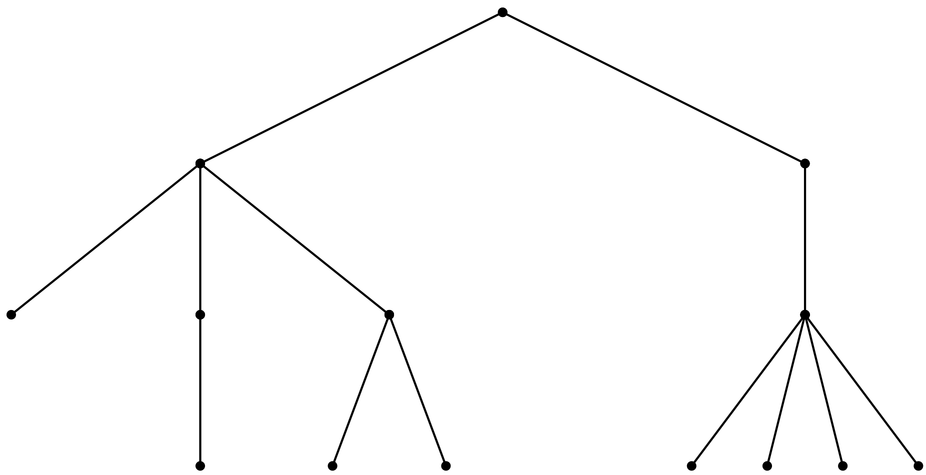


Figure 1.2: A Galton–Watson tree with $Z_0 = 1, Z_1 = 2, Z_2 = 4, Z_3 = 7$.

We will give a more precise description of the model. Let

$$\xi_i^{(n)}\ ,\ i \geq 1,\, n \geq 0$$

be i.i.d. nonnegative integer-valued random variables with distribution $(p_i)_{i\geq 0}$. The tree starts with one ancestor: $Z_0 = 1$. The ancestor has $Z_1 = \xi_1^{(0)}$ children which forms the 1st generation. For the particles in the 1st generation, they have $\xi_j^{(1)}$ children for $j = 1, \ldots, Z_1$. The number of particles in 2nd generation is then

$$Z_2 = \sum_{j=1}^{Z_1} \xi_j^{(1)}$$

Generally, given Z_n, the particles in n-th generation have $\xi_j^{(n)}$ children for $j =$

$1, \dots, Z_n$. The number of particles in $(n+1)$-th generation is

$$Z_{n+1} = \sum_{j=1}^{Z_n} \xi_j^{(n)}$$

Clearly, $(Z_n)_{n \geq 0}$ is a *Markov chain* on $\mathbb{N}_0$, with an *absorbing* state zero, i.e., if $Z_n = 0$ for a certain n, then $Z_j = 0$ for all $j \geq n$.

Extinction Probability One of the first questions we ask is about the *extinction probability*

$$q := \mathbb{P}\left(Z_n = 0 \text{ eventually }\right).$$

Since the event $\{Z_n = 0\}$ being non-decreasing in n, we have

$$q = \mathbb{P}\big(\bigcup_{n=1}^{\infty} \{Z_n = 0\}\big) = \lim_{n \to \infty} \mathbb{P}\left(Z_n = 0\right).$$

To compute the distribution of Z_n, we introduce it's genetating function f_n,

$$f_n(s) := \mathbb{E}s^{Z_n} = \sum_{i=0}^{\infty} \mathbb{P}(Z_n = i)s^i, \quad s \in [0,1],$$

with the convention that $0^0 = 1$. $f \equiv f_1$ is exactly the generating function of offspring fistribution :

$$f(s) = \mathbb{E}s^{Z_1} = \sum_{i=0}^{\infty} p_i s^i, \quad s \in [0,1].$$

Note that

$$f(0) = p_0, \quad f(1) = 1, \quad f'(1) = m := \mathbb{E}Z_1 \in (0, \infty].$$

Lemma 1.48. *$f_n(s) = f^{(n)}(s)$ for all $s \in [0,1]$, where $f^{(n)}$ denotes the n-th fold composition of f.*

Proof. For fixed $n \geq 1$, conditioning on Z_{n-1}, Z_n is the sum of Z_{n-1} i.i.d. random variables having the common distribution which is that of Z_1, thus

$$\mathbb{E}\left(s^{Z_n} | Z_{n-1}\right) = f(s)^{Z_{n-1}},$$

which implies

$$f_n(s) = \mathbb{E}\, s^{Z_n} = \mathbb{E}\, f(s)^{Z_{n-1}} = f_{n-1}(f(s)).$$

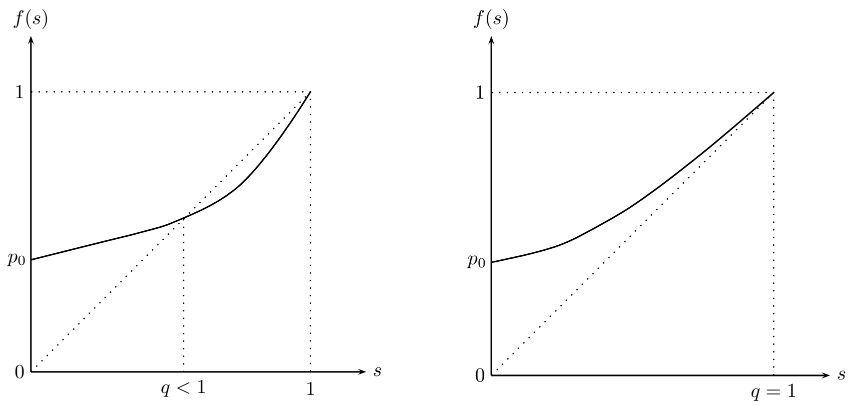
By induction, the desired result follows. $\square$

Theorem 1.49. *The extinction probability q is the smallest root of the equation $f(s) = s$ for $s \in [0,1]$,*

Proof. By Lemma 1.48, we have

$$q = \lim_{n\to\infty} \mathbb{P}(Z_n = 0) = \lim_{n\to\infty} f_n(0).$$

Let us look at the graph of the function f on $[0,1]$. The function is (strictly) increasing and strictly convex, with $f(0) = p_0 \geq 0$ and $f(1) = 1$. In particular, it has at most two fixed points.



(a) $m > 1$ (b) $m \leq 1$

Figure 1.3: Graph of f

If $m \leq 1$, then $p_0 > 0$, and $f(s) > s$ for all $s \in [0, 1)$, which implies $f_n(0) \to 1$. In other words, $q = 1$ is the unique root of $f(s) = s$.

Assume now $m \in (1, \infty]$. This time, $f_n(0)$ converges increasingly to the unique root of $f(s) = s, s \in [0, 1)$ In particular, $q < 1$. □

Remark 1.26. Theorem 1.49 tells us that in the subcritical case (i.e., $m < 1$) and in the critical case ($m = 1$), the Galton-Watson process dies out with probability 1, whereas in the supercritical case ($m > 1$), the Galton-Watson process survives with (strictly) positive probability. Of course, we will be mainly interested in the supercritical case $m > 1$.

Martingales Suppose now $m \in (0, \infty)$. Let us introduce

$$W_n := \frac{Z_n}{m^n}, \quad n \geq 0$$

and $\mathcal{F}_0 = \{\emptyset, \Omega\}$,

$$\mathcal{F}_n = \sigma\left(\xi_i^m : i \geq 1, 0 \leq m \leq n-1\right) \quad n \geq 1\,.$$

It is clear that $\{W_n\}$ is a martingale with respect to $\{\mathcal{F}_n\}$. In fact, $Z_n \in \mathcal{F}_n$ and for $n \geq 0$,

$$\begin{aligned}\mathbb{E}(Z_{n+1}|\mathcal{F}_n) &= \mathbb{E}\left(\sum_{k=0}^{\infty}\sum_{i=1}^{k} \xi_i^{(n)} 1_{\{Z_n = k\}} \middle| \mathcal{F}_n\right) \\ &= \sum_{k=0}^{\infty} k\, m 1_{\{Z_n = k\}} = m Z_n\,.\end{aligned}$$

Since $\{W_n\}$ is non-negative, we have

$$W_n \to W_\infty \quad \mathbb{P}\text{-a.s.}$$

Where $W_\infty = \limsup W_n$ is an integrable random variable. It is, however, possible that $W_\infty = 0$. So it is important to know when W_∞ is non-degenerate.

We make the trivial remark that $W_\infty = 0$ if the system dies out. In particular, we have $W_\infty = 0$ a.s. if $m \leq 1$. What happens if $m > 1$?

We start with two simple observations. The first says that in general, $\mathbb{P}(W_\infty = 0\ )$ equals q or 1, whereas the second tells us that W_∞ is non-degenerate if the offspring distribution admits a finite second moment.

Proposition 1.50. *Assume $m < \infty$. Then $\mathbb{P}(W_\infty = 0)$ equals either q or 1, and hence*

$$\{W_\infty = 0\} = \{Z_n = 0 \quad \textit{eventually}\ \} \quad \textit{a.s.}.$$

Proof. There is nothing to prove if $m \leq 1$. So let us assume $m \in (1, \infty)$. By markov property, for $n \geq 1$,

$$Z_n = \sum_{i=1}^{\xi} Z_{n-1}^{(i)}$$

where $\{(Z_n^{(i)})_{n\geq 0}, i \geq 1\}$ are i.i.d. copies of (Z_n) and ξ is independent of $\{(Z_n^{(i)})_{n\geq 0}, i \geq 1\}$ with offspring distribution.

Dividing on both sides by m^n and letting $n \to \infty$, it follows that mW_∞ is distributed as $\sum_{i=1}^{\xi} W_\infty^{(i)}$, where $\{W_\infty^{(i)}, i \geq 1\}$ are i.i.d. copies of W_∞, independent of Z_1. In particular,

$$\mathbb{P}(W_\infty = 0) = \mathbb{E}[\mathbb{P}(W_\infty = 0)^\xi] = f(\mathbb{P}(W_\infty = 0)).$$

$\mathbb{P}(W_\infty = 0)$ is a root of $f(s) = s$ for $s \in [0, 1]$. So, $\mathbb{P}(W_\infty = 0)$ is q or 1. □

Theorem 1.51. *If $\mathbb{E}Z_1^2 < \infty$ and $m > 1$, then $\{W_n\}$ is L^2 bounded and $W_n \to W_\infty$ almost surely and in L^2. In particular, $\mathbb{E}(W_\infty) = 1$, and $\mathbb{P}(W_\infty = 0) = q$.*

Proof. By conditional variation formular,

$$\mathbb{E}\left(W_n^2 | \mathcal{F}_{n-1}\right) = W_{n-1}^2 + \mathbb{E}\left[(W_n - W_{n-1})^2 | \mathcal{F}_{n-1}\right].$$

To compute the second term, we observe

$$\mathbb{E}\left[(W_n - W_{n-1})^2 \,|\mathcal{F}_{n-1}\right] = m^{-2n}\mathbb{E}\left((Z_n - mZ_{n-1})^2 \,|\mathcal{F}_{n-1}\right)\,,$$

and

$$\mathbb{E}\left[(Z_n - mZ_{n-1})^2 \,|\mathcal{F}_{n-1}\right] = \mathbb{E}\left[\Big(\sum_{i=1}^{Z_{n-1}} \xi_i^{(n-1)} - mk\Big)^2 \Big| \mathcal{F}_{n-1}\right] = Z_{n-1}\sigma^2\,.$$

Combining the last three equations gives

$$\mathbb{E}W_n^2 = \mathbb{E}W_{n-1}^2 + \frac{\sigma^2}{m^{2n}}\mathbb{E}\left(Z_{n-1}\right) = \mathbb{E}W_{n-1}^2 + \frac{\sigma^2}{m^{n+1}}\,.$$

So the desired result follows. $\square$

It turns out that the second moment condition in the preceding theorem can be weakened to an $X \log X$-type integrability condition.

Theorem 1.52 (Kesten and Stigum)**.** *Assume* $1 < m < \infty$. *Then*

$$\mathbb{E}(W_\infty) = 1 \Leftrightarrow \mathbb{P}(W_\infty > 0|\textit{non-extinction}) = 1 \Leftrightarrow \mathbb{E}\left(Z_1 \log^+ Z_1\right) < \infty\,.$$

Remark 1.27. The conclusion in the Kesten-Stigum theorem can also be stated as

$$\mathbb{E}(W_\infty) = 1 \Leftrightarrow \mathbb{P}(W_\infty = 0) = q \Leftrightarrow \sum_{i=1}^{\infty} p_i\, i \log i < \infty\,.$$

The condition $\mathbb{E}\left(Z_1 \log^+ Z_1\right) < \infty$ may look technical. We will see in the next paragraph why this is a natural condition.

B Size-Biased Galton-Watson Trees*

In order to introduce size-biased Galton-Watson processes, we need to view the tree as a random element in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. To do this, let

$$\mathcal{U} := \{\varnothing\} \cup \bigcup_{k=1}^{\infty} \mathbb{N}^k\,,$$

If $u, v \in \mathcal{U}$, we denote by uv the concatenated element, with $u\varnothing = \varnothing u = u$.

Definition 1.5. A **tree** ω is a subset of $\mathcal{U}$ satisfying:

(i) $\varnothing \in \omega$;

(ii) if $uj \in \omega$ for some $j \in \mathbb{N}$, then $u \in \omega$;

(iii) if $u \in \omega$ then $uj \in \omega$ if and only if $1 \leq j \leq N_u(\omega)$ for some non-negative integer $N_u(\omega)$.

In the language of trees, if $u \in \mathcal{U}$ is an element of the tree ω, u is a *vertex* of the tree, and $N_u(\omega)$ the number of children. Vertices of ω are labeled by their line of descent: if $u = i_1 \cdots i_n \in \mathcal{U}$, then u is the i_n-th child of the i_{n-1}-th child of ... of the i_1-th child of the initial ancestor $\varnothing$.

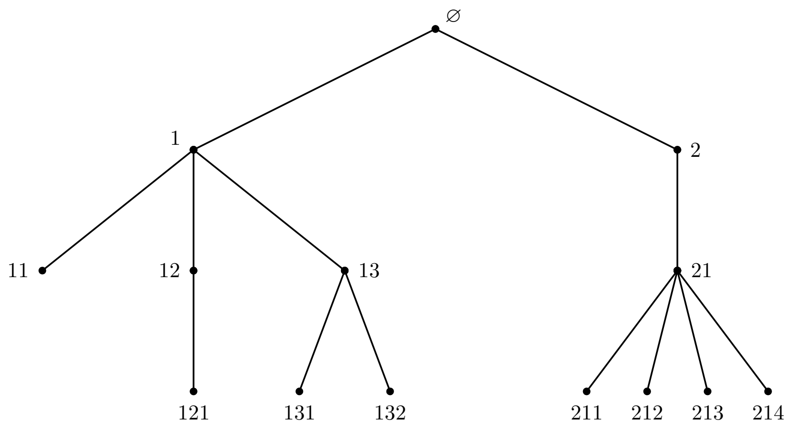


Figure 1.4: Vertices of a tree as elements of $\mathcal{U}$

Let Ω be the space of all trees. We now endow it with a sigma-algebra. For any $u \in \mathcal{U}$, let $\Omega_u := \{\omega \in \Omega : u \in \omega\}$ denote the subspace of Ω consisting of all the trees containing u as a vertex. (In particular, $\Omega_\varnothing = \Omega$ because all the trees contain the root as a vertex, according to part (i) of the definition.) The promised sigma-algebra associated with Ω is defined by

$$\mathcal{F} := \sigma\left(\Omega_u : u \in \mathcal{U}\right) ,$$

and for each $n \geq 0$, let

$$\mathcal{F}_n := \sigma\left(\Omega_u : u \in \mathcal{U}\, , |u| \leq n\right),$$

where $|u|$ is the length of u (representing the generation of the vertex u in the language of trees). Clearly, $(\mathcal{F}_n)_{n\geq 0}$ is a filtration so that $\mathcal{F} = \sigma(\cup_n \mathcal{F}_n)$. For any tree $\omega \in \Omega$, let $Z_n(\omega)$ be the number of individuals in the n-th generation, i.e.,

$$Z_n(\omega) := \sum_{u\in\omega} 1_{\{|u|=n\}} = \sum_{\substack{u\in\omega \\ |u|=n-1}} N_u(\omega)\, .$$

It is easily checked that for any n, $Z_n \in \mathcal{F}_n$ is a random variable with non-negative integer values.

Let $\mathbb{T} : \Omega \to \Omega$ be the identity application. Let $(p_i)_{i\geq 0}$ be a distuibution. According to Neveu [6], there exists a probability $\mathbb{P}$ on Ω such that the law of $\mathbb{T}$ under $\mathbb{P}$ is the law of the Galton-Watson tree with reproduction distribution (p_k).

Let $\tilde{\mathbb{P}}$ be the probability on $(\Omega, \mathcal{F})$ such that for any n.

$$\tilde{\mathbb{P}}|_{\mathcal{F}_n} = W_n.\mathbb{P}|_{\mathcal{F}_n}\, .$$

Since (W_n) is a martingale with respect to $(\mathcal{F}_n)$, the existence of $\tilde{\mathbb{P}}$ is guaranteed by Kolmogorov's extension theorem. Note that for any n,

$$\tilde{\mathbb{P}}\left(Z_n > 0\right) = \mathbb{E}\left(1_{(Z_n>0\}}W_n\right) = \mathbb{E}W_n = 1\, .$$

Therefore,

$$\tilde{\mathbb{P}}\left(Z_n > 0,\ \text{for all } n\right) = 1\, .$$

In other words, there is almost surely non-extinction of the Galton-Watson tree T under the new probability $\tilde{\mathbb{P}}$. The Galton-Watson tree T under $\tilde{\mathbb{P}}$ is called a **size-biased** Galton-Watson tree.

Let us give a description of its paths. Let $N := N_\varnothing$. If $N \geq 1$, then there are N individuals in the first generation. We write $\mathbb{T}_1, \mathbb{T}_2, \cdots, \mathbb{T}_N$ for the N subtrees rooted at each of the N individual in the first generation, i.e., for $\omega \in \Omega_u$, $\mathbb{T}_u(\omega) := \{v \in \mathcal{U} : uv \in \omega\}$.

Theorem 1.53. *Let $k \geq 1$. If $A_1, A_2, \cdots, A_k$ are elements of $\mathcal{F}$, then*

$$\begin{aligned}&\tilde{\mathbb{P}}\left(N = k, \mathbb{T}_1 \in A_1, \cdots, \mathbb{T}_k \in A_k\right)\\ =&\frac{kp_k}{m}\frac{1}{k}\sum_{i=1}^{k}\mathbb{P}\left(A_1\right)\cdots\mathbb{P}\left(A_{i-1}\right)\tilde{\mathbb{P}}\left(A_i\right)\mathbb{P}\left(A_{i+1}\right)\cdots\mathbb{P}\left(A_k\right) .\end{aligned}$$

Proof. if we can prove the desired identity for all n and all $A_1, A_2, \cdots, A_k \in \mathcal{F}_n$. By the monotone class theorem, the identity holds true for any $A_1 \in \mathcal{F}$ and all n and all $A_2, A_3, \cdots, A_k \in \mathcal{F}_n$, By the monotone class theorem again, it holds for any $A_1 \in \mathcal{F}, A_2 \in \mathcal{F}$, and all n and all $A_3, A_4, \cdots, A_k \in \mathcal{F}_n$. Iterating the procedure n times completes the argument. Hence, we may assume that $A_1, A_2, \cdots, A_k$ are elements of $\mathcal{F}_n$, for some n. Then

$$\tilde{\mathbb{P}}\left(N = k, \mathbb{T}_1 \in A_1, \cdots, \mathbb{T}_k \in A_k\right) = \mathbb{E}\left(\frac{Z_n}{m^n}1_{\{N=k,\mathbb{T}_1\in A_1,\cdots,\mathbb{T}_k\in A_k\}}\right) ,$$

On the event $\{N = k\}$, we can write $Z_n = \sum_{i=1}^{k} Z_{n-1}^{(i)}$, where $Z_{n-1}^{(i)}$ denotes the number of individuals in the $(n-1)$-th generation of the subtree rooted at the i-th individual in the first generation. Accordingly,

$$\tilde{\mathbb{P}}\left(N = k, \mathbb{T}_1 \in A_1, \cdots, \mathbb{T}_k \in A_k\right) = \frac{p_k}{m^n}\sum_{i=1}^{k}\mathbb{E}\left(Z_{n-1}^{(i)}1_{\{\mathbb{T}_1\in A_1,\cdots,\mathbb{T}_k\in A_k\}}\right) .$$

Since

$$\begin{aligned}\mathbb{E}\left(Z_{n-1}^{(i)}1_{\{\mathbb{T}_1\in A_1,\ldots,\mathbb{T}_k\in A_k\}}\right) &= \mathbb{E}\left(Z_{n-1}1_{\{\mathbb{T}\in A_i\}}\right)\prod_{j\neq i}\mathbb{P}\left(A_j\right)\\ &= m^{n-1}\tilde{\mathbb{P}}\left(A_i\right)\prod_{j\neq i}\mathbb{P}\left(A_j\right) ,\end{aligned}$$

the desired identity follows. ◻

Theorem 1.53 tells us the following fact about the size-biased Galton-Watson tree: The root has the biased distribution, i.e., having k children with probability $\frac{kp_k}{m}$, among the individuals in the first generation, one of them is chosen randomly (according to the uniform distribution) such that the subtree rooted at this vertex is a size-biased Galton-Watson tree, whereas the subtrees rooted at all other vertices in the first generation are independent copies of the usual Galton-Watson tree.

Iterating the procedure, we obtain a decomposition of the size-biased Galton-Watson tree with an (infinite) *spine* and with i.i.d. copies of the usual Galton-Watson tree: The root $\varnothing =: w_0$ has the biased distribution, i.e., having k children with probability $\frac{kp_k}{m}$. Among the children of the root, one of them is chosen randomly (according to the uniform distribution) as the element of the spine in the first generation (denoted by w_1). We attach subtrees rooted at all other children; these subtrees are independent copies of the usual Galton-Watson tree. The vertex w_1 has the biased distribution. Among the children of w_1, we choose at random one of them as the element of the spine in the second generation (denoted by w_2). Independent copies of the usual Galton-Watson tree are attached as subtrees rooted at all other children of w_1, whereas w_2 has the biased distribution. And so on.

From technical point of view, it is more convenient to connect size-biased Galton-Watson trees with Galton-Watson branching processes with immigration, described as follows.

A Galton-Watson branching processes with *immigration* starts with no individual (say), and is characterized by a reproduction law and an immigration law. At generation n (for $n \geq 1$), Y_n new individuals immigrate into the system, while all individuals regenerate independently and following the same reproduction law; we assume that $(Y_n, n \geq 1)$ is a collection of i.i.d. random variables following the same immigration law, and independent of everything else in that generation.

Our description of the size-biased Galton-Watson tree can be reformulated

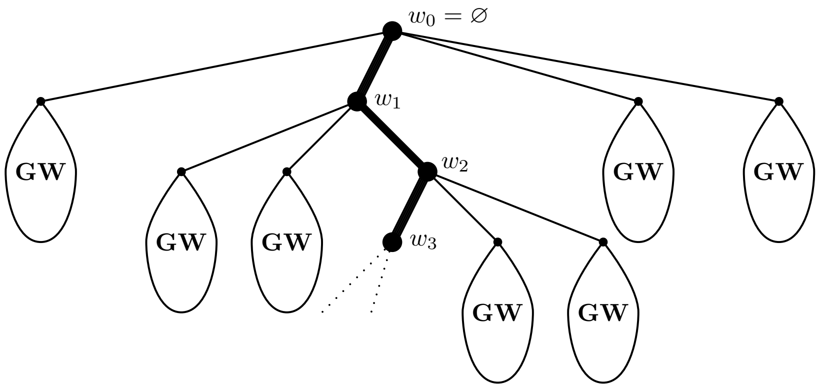


Figure 1.5: A size-baised Galton–Watson tree

in the following way: $(Z_n - 1)_{n\geq 0}$ under $\tilde{\mathbb{P}}$ is a Galton-Watson branching process with immigration, whose immigration law is that of $\tilde{N} - 1$, so that $\{\tilde{N} = k\}$ has probability $\frac{kp_k}{m}$ for $k \geq 1$.

More precisely, there exists another probability space $(E, \mathcal{E}, \mathbf{P})$ and a processes $(X_n)_{n\geq 0}$ defined on it, is a Galton-Watson branching processes with immigration starts with no individual. The reproduction law is $(p_i)_{i\geq 0}$ and immigration law is that of $\tilde{N} - 1$ so that

$$\mathbf{P}(\tilde{N} = k) = \frac{kp_k}{m}\,,\ \text{for } k \geq 1\,.$$

Then $(Z_n - 1)_{n\geq 0}$ under $\tilde{\mathbb{P}}$ concides with $(X_n)_{n\geq 0}$ under $\mathbf{P}$.

C Proof of the Kesten–Stigum Theorem*

We prove the Kesten-Stigum theorem, by means of size-biased Galton-Watson trees. Let us start with an elementary result, which is a classical exercise in the course of probebility theory.

Lemma 1.54. *Let $X, X_1, X_2, \cdots$ be i.i.d. non-negative random variables. Then*

$$\limsup_{n\to\infty} \frac{X_n}{n} = \begin{cases} 0\,, & \mathbb{E}X < \infty\,; \\ \infty\,, & \mathbb{E}X = \infty\,. \end{cases} \qquad a.s..$$

[4] presented a concept of conditional (sub-) martingales given a sigma-field, which is a natural generalization of martingales: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\{\mathcal{F}_n\}$ a filtration and $\mathcal{G} \subset \mathcal{F}_1$ a σ -field. Then the following definition of conditional submartingale (martingale) can be introduced: for an adapted sequence $\{X_n\}$, we say it is a *conditional submartingale* with respect to $\{F_n\}$ given $\mathcal{G}$, if $\mathbb{E}(|X_n||\mathcal{G}) < \infty$ a.s. and $\mathbb{E}(X_{n+1}|\mathcal{G}) \geq X_n$ a.s. for all n.

Theorem (Convergence of Conditional Submartingale, [4])**.** *Let $\{X_n\}$ be a conditional supermartingale given σ-field $\mathcal{G}$ with respect to $\{\mathcal{F}_n\}$. If*

$$\sup_n \mathbb{E}(|X_n|\,|\mathcal{G}) < \infty \qquad a.s.,$$

then $\{X_n\}$ converges to some r.v. X such that $\mathbb{E}(X|\mathcal{G}) < \infty$ a.s..

The Kesten–Stigum theorem will be a consequence of Seneta's theorem for branching processes with immigration.

Theorem 1.55 (Seneta's Theorem)**.** *Let $(Z_n)_{n\geq 0}$ denote the number of individuals of a Galton-Waston branching process with immigration $(Y_n)_{n\geq 1}$ starting with no individual. Assume that $m \in (1, \infty)$, where m denotes the expectation of the reproduction law.*

(i) If $\mathbb{E}\left(\log^+ Y_1\right) < \infty$, then $\lim \frac{Z_n}{m^n}$ exists and is finite a.s.

(ii) If $\mathbb{E}\left(\log^+ Y_1\right) = \infty$, then $\limsup \frac{Z_n}{m^n} = \infty, a.s.$

Proof. Assume $\mathbb{E}\left(\log^+ Y_1\right) < \infty$. Lemma 1.54 tells us in this case that for any $\sum_k \frac{Y_k}{m^k} < \infty$ a.s. Let $\mathcal{Y}$ be the sigma-algebra generated by (Y_n). Clearly,

$$\mathbb{E}\left(Z_{n+1}|\mathcal{F}_n, \mathcal{Y}\right) = mZ_n + Y_{n+1} \geq mZ_n$$

thus $\left(\frac{Z_n}{m^n}\right)$ is a sub-martingale given $\mathcal{Y}$, and

$$\mathbb{E}\left(\frac{Z_n}{m^n} \mid \mathcal{Y}\right) = \sum_{k=0}^{n} \frac{Y_k}{m^k} \quad \text{a.s..}$$

Thus

$$\sup_n \mathbb{E}\left(\frac{Z_n}{m^n} \mid \mathcal{Y}\right) < \infty \quad \text{a.s..}$$

As a consequence, almost surely, $\lim \frac{Z_n}{m^n}$ exists and is finite.

Assume now $\mathbb{E}\left(\log^+ Y_1\right) = \infty$. Then by Lemma 1.54,

$$\limsup_{n\to\infty} \frac{\log Y_n}{n} = \infty \quad \text{a.s.,}$$

since $Z_n \geq Y_n$ it follows that for any $\limsup \frac{Z_n}{m^n} = \infty$ a.s.. □

Proof of the Kesten-Stigum theorem.

Let $m \in (1, \infty)$. Note that $\mathbf{E}[\log^+(\tilde{N} - 1)]$ is finite if and only if $\mathbf{E}(\log^+ \tilde{N})$ is , and

$$m\, \mathbf{E}(\log^+ \tilde{N}) = \sum_{i=1}^{\infty} i p_i \, \log i = \mathbb{E}(Z_1 \log^+ Z_1)\,.$$

If $\sum_i p_i \, i \log i < \infty$, then $\mathbf{E} \log^+(\tilde{N} - 1) < \infty$. By Seneta's theorem,

$$\lim_{n\to\infty} \frac{Z_n - 1}{m^n} = W_\infty < \infty \quad \tilde{\mathbb{P}}\text{-a.s..}$$

By Corollary 1.47, $\mathbb{E}(W_\infty) = 1$; in particular, $\mathbb{P}(W_\infty = 0) < 1$ and thus $\mathbb{P}(W_\infty = 0) = q$.

If $\sum_{i=1}^{\infty} p_i i \log i = \infty$, then $\mathbf{E} \log^+(\tilde{N} - 1) = \infty$. By Seneta's theorem, $\tilde{\mathbb{P}}$-almost surely,

$$\limsup_{n\to\infty} \frac{Z_n - 1}{m^n} = W_\infty = \infty \quad \tilde{\mathbb{P}}\text{-a.s..}$$

By Corollary 1.47, $W_\infty = 0$, $\mathbb{P}$-a.s. We complete the proof. □

1.10 Backwards Martingales

The concepts of filtration and martingale don't require the index set (often interpreted as time) to be a subset of $\mathbb{R}_+$. Hence we can consider the case the martingale indexed by non-positive integers.

Definition 1.6. Let $X = (X_{-n})_{n\geq 0}$ be a martingale with respect to $\mathfrak{F} = (\mathcal{F}_{-n})_{n\geq 0}$. Then X is called a **backwards martingale**.

Remark 1.28. A backwards martingale is always uniformly integrable. This follows from Example 1.2 and the fact that

$$\mathbb{E}\left(X_0|\mathcal{F}_{-n}\right) = X_{-n} \quad \text{a.s. for each } n \geq 0\,.$$

Theorem 1.56 (Convergence Theorem for Backwards Martingales)**.** *Let $X = (X_{-n})_{n\geq 0}$ be a martingale with respect to $\mathfrak{F} = (\mathcal{F}_{-n})_{n\geq 0}$. Let $\mathcal{F}_{-\infty} := \cap_{n=1}^{\infty}\mathcal{F}_{-n}$. Then there exists a r.v. $X_{-\infty} \in \mathcal{F}_{-\infty}$ such that $X_{-n} \to X_{-\infty}$ a.s. and in L^1. Moreover,*

$$X_{-\infty} = \mathbb{E}\left(X_0|\mathcal{F}_{-\infty}\right) \quad a.s..$$

Proof. Let $a < b$ and fix $n \in \mathbb{N}$. Let $U^{a,b}_{-n}$ be the number of upcrossings of X over $[a, b]$ between times $-n$ and 0. Further, let $U^{a,b} = \lim_{n\to\infty} U^{a,b}_{-n}$.

The upcrossing inequality for submartinglaes, Lemma 1.24 implies that

$$\mathbb{E}U^{a,b}_{-n} \leq \frac{\mathbb{E}\left(X_0 - a\right)^+ - \mathbb{E}\left(X_{-n} - a\right)^+}{b-a} \leq \frac{\mathbb{E}\left(X_0 - a\right)^+}{b-a}\,.$$

Letting $n \to \infty$ and using the monotone convergence theorem, we have

$$\mathbb{E}\,U^{a,b} < \infty\,.$$

By Remark 1.18, almost surely, the limit $X_{-\infty}$ exists. Clearly $X_{-\infty} \in \mathcal{F}_{-\infty}$. Since $\{X_{-n}\}$ is uniformly integrable, by Fatou's lemma,

$$\mathbb{E}|X_{-\infty}| \leq \liminf_{n\to\infty} \mathbb{E}|X_{-n}| < \infty\,,$$

so $X_{-\infty}$ is integrable, and hence X_{-n} converges to X in L^1.

Now, it suffices to show that for any $A \in \mathcal{F}_{-\infty}$, $\mathbb{E}(X_{-\infty}1_A) = \mathbb{E}(X_0 1_A)$. Since $\mathbb{E}\left(X_0|\mathcal{F}_{-n}\right) = X_{-n}$, we have

$$\mathbb{E}(X_{-n}1_A) = \mathbb{E}(X_0 1_A)\,, \text{ for all } n \in \mathbb{N}\,.$$

Letting $n \to \infty$, the desired result follows from that $(X_{-n}) \to X_{-\infty}$ in L^1. $\square$

Let $(\mathcal{G}_n)_{n\geq 0}$ be a decreasing sequence of sub-σ-fields on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$; i.e., $\mathcal{G}_{n+1} \subset \mathcal{G}_n \subset \mathcal{F}$ for all $n \geq 0$. Let $\mathcal{G}_\infty = \cap_n \mathcal{G}_n$. Let $(\xi_n)_{n\geq 0}$ be a sequence of integrable random variables adapted to $(\mathcal{G}_n)_{n\geq 0}$; i.e., $\xi_n \in \mathcal{G}_n$ for all $n \geq 0$. Suppose that for each $n \geq 0$,

$$\xi_{n+1} = \mathbb{E}(\xi_n|\mathcal{G}_{n+1}) \qquad \text{a.s..} \tag{1.21}$$

(Notice that this is equivalent to $\xi_n = \mathbb{E}(\xi_0|\mathcal{G}_n)$ a.s. for all $n \geq 0$.) Then let

$$X_{-n} = \xi_n \;\text{ and }\; \mathcal{F}_{-n} = \mathcal{G}_n\,, \qquad \text{for all } n \geq 0\,. \tag{1.22}$$

Clearly $(X_{-n})_{n\geq 0}$ is a backwards martingale with respect to $(\mathcal{F}_{-n})$. Conversely, for any backwards martingale (X_{-n}) with respect to $(\mathcal{F}_{-n})$, define $(\xi_n, \mathcal{G}_n)$ by (1.22). Then $(\xi_n, \mathcal{G}_n)$ is adapted, integrable and satisfies (1.21). Terefore, we also say that $(\xi_n)_{n\geq 0}$ is a baskwards martingle with respect to $(\mathcal{G}_n)_{n\geq 0}$. As we can see, the backwards martingale is indeed the conditional expections of an integrable random variable given a decreasing sequence of σ-fields.

It follows directly from the convergence theorem of backwards martingales that

Corollary 1.57. *For any integrable random variable ξ,*

$$\mathbb{E}\left(\xi|\mathcal{G}_n\right) \to \mathbb{E}\left(\xi|\mathcal{G}_\infty\right) \qquad \textit{a.s. and in } L^1\,.$$

We can enhance the result above, as following.

Proposition 1.58. *Let $(\mathcal{G}_n)_{n\geq 0}$ be a decreasing sequence of σ-fields. Let ξ_n, ξ_∞ be random variables. (Note that we did NOT assume (ξ_n) is adapted.) Suppose that $\xi_n \to \xi_\infty$ a.s. and $|\xi_n| \leq \eta$ for some $\eta \in L^1$. Then*

$$\mathbb{E}\left(\xi_n|\mathcal{G}_n\right) \to \mathbb{E}\left(\xi_\infty|\mathcal{G}_\infty\right) \qquad a.s.\ and\ in\ L^1\,.$$

Proof. By Corollary 1.57, it suffices to show that

$$\mathbb{E}\left(|\xi_n - \xi_\infty||\mathcal{G}_n\right) \to 0 \qquad \text{a.s. and in } L^1\,.$$

The L^1 convergence trivially follows from the almost sure convergence and the Lebesgue domainted convergence theorem. To show the almost sure convergence, let $\eta_m \coloneqq \sup_{n\geq m}|\xi_n - \xi_\infty| \leq 2\eta$, then for fixed m, by Corollary 1.57 we have

$$\limsup_{n\to\infty} \mathbb{E}\left(|\xi_n - \xi_\infty||\mathcal{G}_n\right) \leq \lim_{n\to\infty} \mathbb{E}(\eta_m|\mathcal{G}_n) = \mathbb{E}(\eta_m|\mathcal{G}_\infty) \quad \text{a.s.}.$$

Letting $m \to \infty$, by the domainted convergence for conditional expectation, the desired result follows. $\square$

Example 1.30 (Ballot Theorem)**.** Let $\{\xi_j, 1 \leq j \leq n\}$ be i.i.d. nonnegative integervalued r.v.'s, let $S_k = \xi_1 + \cdots + \xi_k$, and let $G = \{S_j < j \text{ for } 1 \leq j \leq n\}$. Then

$$\mathbb{P}\left(G|S_n\right) = \left(1 - \frac{S_n}{n}\right)^+ .$$

The result is trivial when $S_n \geq n$, so suppose $S_n < n$. Let $X_{-j} = S_j/j$ and $\mathcal{F}_{-j} = \sigma\left(S_j, \ldots, S_n\right)$ for $1 \leq j \leq n$. It's easy to see that (X_{-j}) is a backwards martingale with respect to $(\mathcal{F}_{-j})$. Let $T = \inf\{k \geq -n : X_k \geq 1\}$ and set $T = -1$ if the set is $\emptyset$. We claim that $X_T = 1$ on G^c. To check this, note that if $S_{j+1} < j+1$ then the fact that the ξ_i are nonnegative integer values implies $S_j \leq S_{j+1} \leq j$. since $G \subset \{T = -1\}$ and $S_1 < 1$ implies $S_1 = 0$, we have $X_T = 0$ on G. Noting $\mathcal{F}_{-n} = \sigma\left(S_n\right)$ and using optional stopping theorem, we see that on $\{S_n < n\}$

$$P\left(G^c|S_n\right) = E\left(X_T|\mathcal{F}_{-n}\right) = X_{-n} = S_n/n$$

Subtracting from 1 and recalling that this computation has been done under the assumption $S_n < n$ gives the desired result.

Remark 1.29. To explain the name, consider an election in which candidate B gets β votes and A gets $\alpha > \beta$ votes. Let $\xi_1, \xi_2, \ldots, \xi_n$ be i.i.d. and take values 0 or 2 with probability 1/2 each. Interpreting 0 's and 2 's as votes for candidates A and B, we see that $G = \{A \text{ leads } B \text{ throughout the counting }\}$ so if $n = \alpha + \beta$

$$P(G|B \text{ gets } \beta \text{ votes }) = \left(1 - \frac{2\beta}{n}\right)^+ = \frac{\alpha - \beta}{\alpha + \beta}$$

Convergence of Backwards Submartingales Let $\{\mathcal{F}_n\}_{n\geq 0}$ be a decreasing sequence of sub-σ-fields on $(\Omega, \mathcal{F}, \mathbb{P})$, and let the $\{X_n\}_{n\geq 0}$ be a sequence of random variables. We say that $\{X_n\}_{n\geq 0}$ is a *backwards submartingale* with respect to $\{\mathcal{F}_n\}$, if $\mathbb{E}\,|X_n| < \infty$, X_n is $\mathcal{F}_n$-measurable, and $\mathbb{E}\,(X_n|\mathcal{F}_{n+1}) \geq X_{n+1}$ a.s., for every $n \geq 0$.

Theorem 1.59. *Let $\{X_n\}_{n\geq 0}$ be a backwards submartingale with respect to $\{\mathcal{F}_n\}$. Then there exists a $\overline{\mathbb{R}}$-valued random variable $X_\infty \in \mathcal{F}_\infty$ so that $X_n \to X_\infty$ a.s.. If in addition,*

$$\sup_{n\in\mathbb{N}_0} \mathbb{E}|X_n| < \infty \quad \left(\Leftrightarrow \lim_{n\to\infty} \mathbb{E}X_n > -\infty\right),$$

then $\{X_n\}$ is uniformly integrable and hence X_∞ is an integrale random variable and $\{X_n\}$ converges to X_∞ in L^1.

Proof. As in the proof of Theorem 1.56, the upcrossing inequality implies the existence of such a X_∞. It suffices to show that if $\{X_n\}$ is L^1 bounded, then it's uniformly integrable. But firstly, let's deal with the equivallence condition of L^1-boundedness. Since $\{X_n^+\}$ is also a backwards martingale with respect to $(\mathcal{F}_n)$, we have

$$E|X_n| = 2\mathbb{E}X_n^+ - \mathbb{E}X_n \leq 2\mathbb{E}X_0^+ - \mathbb{E}X_n\,.$$

Since $(\mathbb{E}X_n)$ is decreasing, so $\lim_n \mathbb{E}X_n > -\infty$ implies that $\{X_n\}$ is L^1 bounded.

We now prove that $\{X_n^+\}$ is uniformly integrable. For any $\lambda > 0$, since $\{X_n^+\}$ is also a submartingale with respect to $\{\mathcal{F}_n\}$,

$$\mathbb{E}X_n^+ 1_{\{X_n^+ > \lambda\}} = \mathbb{E}X_n 1_{\{X_n > \lambda\}} \leq \mathbb{E}X_0 1_{\{X_n > \lambda\}} .$$

On the other hand, by Markov inequality,

$$\mathbb{P}(X_n > \lambda) \leq \frac{\mathbb{E}X_n^+}{\lambda} \leq \frac{\mathbb{E}X_0^+}{\lambda} ,$$

so $\sup_{n\geq 0} \mathbb{P}(|X_n| > \lambda)$ converges to zero as $\lambda \to \infty$. Hence

$$\sup_{n\geq 0} \mathbb{E}X_n^+ 1_{\{X_n^+ > \lambda\}} \leq \sup_{n\geq 0} \mathbb{E}\, X_0 1_{\{X_n > \lambda\}} \to 0 \quad \text{as } \lambda \to \infty$$

follows from the absolute continuity of the integral.

We now prove that $\{X_n^-\}$ is uniformly integrable. Observe that for each $m < n$, we have

$$\begin{aligned}\mathbb{E}X_n^- 1_{\{X_n^- > \lambda\}} &= -\mathbb{E}X_n 1_{\{X_n < -\lambda\}} = \mathbb{E}X_n 1_{\{X_n > -\lambda\}} - \mathbb{E}X_n \\ &\leq \mathbb{E}X_m 1_{\{X_n > -\lambda\}} - \mathbb{E}X_n \\ &= \mathbb{E}X_m - \mathbb{E}X_n - \mathbb{E}X_m 1_{\{X_n < -\lambda\}} .\end{aligned}$$

So given $\epsilon > 0$, we can certainly choose $m = m_\epsilon$ so large that $0 \leq \mathbb{E}X_m - \mathbb{E}X_n \leq \epsilon$ holds for every $n > m_,$ and for that $m = m_\epsilon$, by the absolute continuity of the integral, there exists $\lambda_\epsilon = \lambda(m_\epsilon, \epsilon) > 0$ so that for any $\lambda > \lambda_\epsilon$,

$$\sup_{n>m} \mathbb{E}X_m 1_{\{X_n < -\lambda\}} < \epsilon \ ; \ \sup_{n\leq m} \mathbb{E}X_n^- 1_{\{X_n^- > \lambda\}} < \epsilon .$$

Consequently, for any $\lambda > \lambda_\epsilon$ we have:

$$\sup_{n\geq 1} \mathbb{E}X_n^- 1_{\{X_n^- > \lambda\}} < 2\epsilon$$

and thus $\{X_n^-\}$ is also uniformly integrable. $\square$

1.11 Applacation(IV): Exchangeability

With many data acquisitions, such as telephone surveys, the order in which the data come does not matter. Mathematically, we say that a family of random variables is *exchangeable* if the joint distribution does not change under finite permutations. De Finetti's structural theorem says that a countable family of E-valued exchangeable random variables can be described by a two-stage experiment. At the first stage, a probability distribution Ξ on E is drawn at random. At the second stage, i.i.d. random variables with distribution Ξ are implemented.

Recall that a finite permutation is a bijection $\varrho : I \to I$ that leaves all but finitely many points unchanged.

Definition 1.7. Let I be an arbitrary index set and let E be a Polish space. A family $(X_i)_{i\in I}$ of random variables with values in E is called **exchangeable** if

$$\mathcal{L}\left[\left(X_{\varrho(i)}\right)_{i\in I}\right] = \mathcal{L}\left[(X_i)_{i\in I}\right]$$

for any finite permutation $\varrho : I \to I$.

Remark 1.30. Clearly, the following are equivalent. $(X_i)_{i\in I}$ is exchangeable if and only if for each $n \in \mathbb{N}$, $i_1, \ldots, i_n \in I$ are pairwise distinct and $j_1, \ldots, j_n \in I$ are pairwise distinct, there holds

$$\mathcal{L}\left[(X_{i_1}, \ldots, X_{i_n})\right] = \mathcal{L}\left[(X_{j_1}, \ldots, X_{j_n})\right]$$

In particular, exchangeable random variables are identically distributed.

Example 1.31. If $(X_i)_{i\in I}$ is i.i.d., then $(X_i)_{i\in I}$ is exchangeable.

Example 1.32. Let Y be a random variable with values in $[0, 1]$. Assume that, given Y, the random variables $(X_i)_{i\in I}$ are independent and Bernouli(Y) distributed, i.e., for any two disjoint nonempty finite subset of I, namely J_1 and J_0,

$$\mathbb{P}\left(X_j = 1, j \in J_1\,;\ X_j = 0, j \in J_2 | Y\right) = Y^{|J_0|}(1-Y)^{|J_2|}\,.$$

Then $(X_i)_{i\in I}$ is exchangeable.

A Exchangeable Sequences

Although we defined exchangeability for arbitary family of random variables, we only deal with the case that the family is countable. Now, let $X = (X_n)_{n\in\mathbb{N}}$ be a stochastic process with values in E, a Polish space. The following settings were discussed in the course of probability theory when talking *Hewitt-Sage 0-1 law.*

For each $n \in \mathbb{N}$, let $S(n)$ be the set of permutations of $\mathbb{N}$ so that $\varrho(k) = k$ for all $k > n$. For $\varrho \in S(n)$ and $x \in E^{\mathbb{N}}$, denote $x^{\varrho} = \left(x_{\varrho(1)}, x_{\varrho(2)}, \ldots\right) \in E^{\mathbb{N}}$. If $A \in \sigma(X_n, n \in \mathbb{N})$ is an event, then there is a measurable $B \in \mathcal{B}(E)^{\mathbb{N}}$ with $A = \{X \in B\}$. We denote $A^{\varrho} = \{X^{\varrho} \in B\}$ for $\varrho \in S(n)$. Then we define, for each $n \in \mathbb{N}$,

$$\mathcal{E}_n := \{A : A^{\varrho} = A \text{ for all } \varrho \in S(n)\} \ .$$

Clearly, $\mathcal{E}_{n+1} \subset \mathcal{E}_n$. Then define

$$\mathcal{E} := \bigcap_{n\in\mathbb{N}} \mathcal{E}_n = \{A : A^{\varrho} = A \text{ for all finite permutation } \varrho \text{ of } \mathbb{N}\} \ ,$$

is called the **exchangeable σ-algebra** (for X). Each $A \in \mathcal{E}$ is called a exchangeable event.

Remark 1.31. Denote by $\mathcal{T}$ the tail σ-algebra (for X), i.e.,

$$\mathcal{T} = \bigcap_{n\in\mathbb{N}} \sigma(X_{n+1}, X_{n+2}, \ldots) \ .$$

Then $\mathcal{T} \subset \mathcal{E}$, since $\sigma(X_{n+1}, X_{n+2}, \ldots) \subset \mathcal{E}_n$ for $n \in \mathbb{N}$. Moreover, strict inclusion is possible. Indeed, let $E = \{0, 1\}$ and let $X_1, X_2, \ldots$ be independent random variables with $\mathbb{P}(X_n = 1) = 2^{-n}$. Then $\sum_{n=1}^{\infty} X_n$ is measurable with respect to $\mathcal{E}$ but not with respect to $\mathcal{T}$.

Lemma 1.60. *Let $X = (X_n)_{n\in\mathbb{N}}$ be exchangeable. Let $\varphi : E^{\mathbb{N}} \to \mathbb{R}$ is measurable. If $\mathbb{E}|\varphi(X)| < \infty$, then for all $n \in \mathbb{N}$ and all $\varrho \in S(n)$*

$$\mathbb{E}\left[\varphi(X)|\mathcal{E}_n\right] = \mathbb{E}\left[\varphi\left(X^\varrho\right)|\mathcal{E}_n\right] ,$$

Moreover,

$$\mathbb{E}\left[\varphi(X)|\mathcal{E}_n\right] = A_n(\varphi) := \frac{1}{n!}\sum_{\varrho\in S(n)} \varphi\left(X^\varrho\right) .$$

Proof. Let $\{X \in B\} \in \mathcal{E}_n$, where $B \in \mathcal{B}(E)^{\mathbb{N}}$. We will show that

$$\mathbb{E}\left[\varphi(X)1_{\{X\in B\}}\right] = \mathbb{E}\left[\varphi(X^\varrho)1_{\{X\in B\}}\right] .$$

Since X and X^ϱ have the same distribution,

$$\mathbb{E}[\varphi(X)1_{\{X\in B\}}] = \mathbb{E}\left[\varphi\left(X^\varrho\right)1_{\{X^\varrho\in B\}}\right] .$$

Note that $\{X \in B\} \in \mathcal{E}_n$, $\varrho \in S(n)$, we have $\{X \in B\} = \{X^\varrho \in B\}$ and

$$\mathbb{E}\left[\varphi(X)1_{\{X\in B\}}\right] = \mathbb{E}\left[\varphi(X^\varrho)1_{\{X\in B\}}\right] .$$

From this follows , note that $A_n(\varphi)$ is $\mathcal{E}_n$ -measurable, so

$$\mathbb{E}\left[\varphi(X)|\mathcal{E}_n\right] = \mathbb{E}\left[\frac{1}{n!}\sum_{\varrho\in S(n)}\varphi\left(X^\varrho\right)|\mathcal{E}_n\right] = \frac{1}{n!}\sum_{\varrho\in S(n)}\varphi\left(X^\varrho\right) .$$

We complete the proof. □

Example 1.33. Let $X = (X_n)_{n\in\mathbb{N}}$ be exchangeable real random variables and $\mathbb{E}|X_1| < \infty$. Then

$$\mathbb{E}(X_1|\mathcal{E}_n) = \frac{1}{n!}\sum_{\varrho\in S(n)} X_{\varrho(1)} = \frac{1}{n}\sum_{i=1}^{n} X_i .$$

By Theorem 1.57,

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}\left(X_1|\mathcal{E}\right) \qquad \text{a.s. and in } L^1$$

In fact $\mathbb{E}\left(X_1|\mathcal{E}\right)$ is $\mathcal{T}$-measurable, since it's easy to see that

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}X_i \quad \text{and} \quad \liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}X_i$$

are $\mathcal{T}$-measurable. Hence

$$\mathbb{E}\left(X_1|\mathcal{E}\right)=\mathbb{E}\left[\mathbb{E}\left(X_1|\mathcal{E}\right)|\mathcal{T}\right]=\mathbb{E}\left[X_1|\mathcal{T}\right].$$

Example 1.34 (SLLN)**.** If $X_1, X_2, \ldots$ are real and i.i.d. with $\mathbb{E}\left|X_1\right|<\infty$, then

$$\frac{1}{n}\sum_{i=1}^{n}X_i\to\mathbb{E}\left[X_1\right] \qquad \text{a.s..}$$

By Kolmogorov 0-1 law (Hewitt-Sage 0-1 law, respectively) the tail σ-algebra $\mathcal{T}$ (the exchangebale σ-algebra $\mathcal{E}$) is trivial, hence

$$\begin{gathered}\mathbb{E}\left[X_1|\mathcal{T}\right]=\mathbb{E}X_1 \quad \text{a.s.}\\ \left(\mathbb{E}\left[X_1|\mathcal{E}\right]=\mathbb{E}X_1 \quad \text{a.s. }\right)\end{gathered}$$

By Example 1.33, SLLN follows.

We close this subsection with a generalization of Example 1.33. This conclusion from the convergence theorem for backwards martingales will be used in an essential way in the next section.

Theorem 1.61. *Let $X=\left(X_n\right)_{n\in\mathbb{N}}$ be an exchangeable family of random variables with values in E. Let $\varphi: E^{\mathbb{N}}\to\mathbb{R}$ be measurable so that $\mathbb{E}\left|\varphi\left(X\right)\right|<\infty$. Then*

$$\mathbb{E}[\varphi(X)|\mathcal{E}]=\lim_{n\to\infty}A_n(\varphi) \qquad \textit{a.s. and in } L^1.$$

If in addition, $\varphi(x)$ can be regarded as a function of $(x_1,\cdots,x_k)$, $x\in E^{\mathbb{N}}$, that is $\varphi(x)=\phi(x_1,\cdots,x_k)$ for some $\phi: E^k\to\mathbb{R}$, then

$$\mathbb{E}[\varphi(X)|\mathcal{E}]=\mathbb{E}[\varphi(X)|\mathcal{T}]$$

Proof. Since $\mathbb{E}\left[\varphi(X)|\mathcal{E}_n\right] = A_n(\varphi)$, by Theorem 1.57,

$$A_n(\varphi) \to \mathbb{E}[\varphi(X)|\mathcal{E}] \quad \text{a.s. and in } L^1$$

As for Example 1.33, we can argue that $\mathbb{E}[\varphi(X)|\mathcal{E}] = \lim_n A_n(\varphi)$ is $\mathcal{T}$-measurable. Indeed, for all $l \in \mathbb{N}$,

$$\limsup_{n\to\infty} \frac{\left|\left\{\varrho \in S(n) : \varrho^{-1}(i) \le l \text{ for some } i \in \{1,\dots,k\}\right\}\right|}{n!} = 0\,.$$

Thus, for large n, the dependence of $A_n(\varphi)$ on the first l coordinates is negligible. Hence

$$\mathbb{E}\left(\varphi(X)|\mathcal{E}\right) = \mathbb{E}\left[\mathbb{E}\left(\varphi(X)|\mathcal{E}\right)|\mathcal{T}\right] = \mathbb{E}\left[\varphi(X)|\mathcal{T}\right]\,. \qquad \square$$

Corollary 1.62. *Let $X = (X_n)_{n\in\mathbb{N}}$ be exchangeable. Then, for any $A \in \mathcal{E}$ there exists a $C \in \mathcal{T}$ with $\mathbb{P}(A\triangle C) = 0$.*

Proof. Since $\mathcal{E} \subset \sigma\left(X_1, X_2, \dots\right)$, by the approximation theorem for measures, there exists a sequence $(A_n)_{n\in\mathbb{N}}$ with $A_n \in \sigma\left(X_1, \dots, X_n\right)$ and such that $\mathbb{P}\left(A\Delta A_n\right) \to 0$. Let $B_n \in \mathcal{B}(E)^n$ be measurable with

$$A_n = \{(X_1, \dots, X_n) \in B_n\}$$

for all $n \in \mathbb{N}$. Letting $\varphi_n(x) := 1_{B_n}(x_1, \cdots, x_n)$, for $x \in E^{\mathbb{N}}$. Theorem 1.61 implies that

$$\begin{aligned} 1_A = \mathbb{E}[1_A|\mathcal{E}] &= \mathbb{E}\left[\lim_{n\to\infty} 1_{A_n}|\mathcal{E}\right] = \mathbb{E}\left[\lim_{n\to\infty}\varphi_n(X)|\mathcal{E}\right] \\ &= \lim_{n\to\infty}\mathbb{E}\left[\varphi_n(X)|\mathcal{E}\right] = \lim_{n\to\infty}\mathbb{E}\left[\varphi_n(X)|\mathcal{T}\right] =: \psi \quad \text{a.s.}. \end{aligned}$$

Hence there is a $\mathcal{T}$-measurable function ψ with $\psi = 1_A$ almost surely. We can assume that $\psi = 1_C$ for some $C \in \mathcal{T}$. $\square$

Corollary 1.63 (Hewitt-Savage 0-1 Law)**.** *Let $X = (X_n)_{n\in\mathbb{N}}$ be i.i.d. random variables. Then the exchangeable σ-algebra $\mathcal{E}$ for X is trivial; that is, $\mathbb{P}(A) \in \{0,1\}$ for all $A \in \mathcal{E}$.*

B De Finetti's Theorem

Hence we shall show that a countably infinite exchangeable family of random variables is an i.i.d. family given the exchangeable σ-algebra $\mathcal{E}$. Furthermore, we compute the conditional distribution of the individual random variables. As a first step, we define conditional independence formally.

Definition 1.8. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathcal{G} \subset \mathcal{F}$ be a σ-algebra and let $(\mathcal{A}_i)_{i \in I}$ be an arbitrary family of subsets of $\mathcal{F}$. Assume that for any finite $J \subset I$, any choice of $A_j \in \mathcal{A}_j$, $j \in J$

$$\mathbb{P}\left(\cap_{j \in J} A_j | \mathcal{G}\right) = \prod_{j \in J} \mathbb{P}\left(A_j | \mathcal{G}\right) \quad \text{a.s.}$$

Then the family $(\mathcal{A}_i)_{i \in I}$ is called **independent given** $\mathcal{G}$.

Remark 1.32. A family $(X_i)_{i \in I}$ of random variables is called independent (and identically distributed) given $\mathcal{G}$ if $(\sigma(X_i))_{i \in I}$ are independent given $\mathcal{G}$ (and the conditional distributions $\mathbb{P}(X_i \in \cdot | \mathcal{G})$ are equal).

Let $X = (X_n)_{n \in \mathbb{N}}$ be a stochastic process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in a Polish space E. Let $\mathcal{E}$ be the exchangeable σ-algebra and let $\mathcal{T}$ be the tail σ-algebra.

Theorem 1.64 (de Finetti)**.** *The sequence $X = (X_n)_{n \in \mathbb{N}}$ is exchangeable iff there exists a σ-algebra $\mathcal{G} \subset \mathcal{F}$ such that (X_n) is i.i.d. given $\mathcal{G}$. In this case, $\mathcal{G}$ can be chosen to equal the exchangeable σ-algebra $\mathcal{E}$ or the tail-σ-algebra $\mathcal{T}$.*

Proof. Necessity. Let X be exchangeable and let $\mathcal{G} = \mathcal{E}$ or $\mathcal{G} = \mathcal{T}$.It fuffices to show that, for any given $k \in \mathbb{N}$ and bounded measurable maps $f_i : E \to \mathbb{R}$, $i = 1, 2, \cdots k$, we have

$$\mathbb{E}\left[\prod_{i=1}^{k} f_i(X_i) | \mathcal{G}\right] = \prod_{i=1}^{k} \mathbb{E}\left[f_i(X_i) | \mathcal{G}\right] \quad \text{a.s., and}$$

$$\mathbb{E}\left[f_i(X_i) | \mathcal{G}\right] = \mathbb{E}\left[f_i(X_1) | \mathcal{G}\right] \quad \text{a.s. for each } i\,.$$

The second indentity is evident, since by Theorem 1.61,

$$\mathbb{E}\left[f_i\left(X_i\right)|\mathcal{G}\right] = \lim_{n\to\infty}\frac{f_i(X_1)+\cdots+f_i(X_n)}{n} = \mathbb{E}\left[f_i\left(X_i\right)|\mathcal{G}\right] \quad \text{a.s..}$$

Now, define $\hat{f}_i : E^{\mathbb{N}} \to \mathbb{R}$ by letting $\hat{f}_i(x) = f_i(x_i)$ for $x \in \mathbb{E}^{\mathbb{N}}$. Let $\varphi_j = \Pi_{i=1}^{j}\hat{f}_i$ for $j \leq k$. Then it sufficcies to show that

$$\mathbb{E}\left[\varphi_k(X)|\mathcal{G}\right] = \prod_{i=1}^{k}\mathbb{E}\left[f_i\left(X_i\right)|\mathcal{G}\right] \quad \text{a.s..}$$

which is equivalent to

$$\lim_{n\to\infty} A_n(\varphi_k) = \lim_{n\to\infty}\prod_{i=1}^{k} A_n(\hat{f}_i) \quad \text{a.s..}$$

By induction, we only need to show that

$$\lim_{n\to\infty} A_n(\varphi_k) = \lim_{n\to\infty} A_n(\varphi_{k-1})A_n(\hat{f}_k) \quad \text{a.s..}$$

Note that

$$\begin{aligned}
&A_n\left(\varphi_{k-1}\right)A_n\left(\hat{f}_k\right) = \frac{1}{n!}\sum_{\varrho\in S(n)}\varphi_{k-1}\left(X^{\varrho}\right)\frac{1}{n}\sum_{i=1}^{n} f_k\left(X_i\right)\\
&= \frac{1}{n\cdot n!}\sum_{\substack{\varrho\in S(n)\\ k\leq i\leq n}}\varphi_{k-1}\left(X^{\varrho}\right)f_k\left(X_{\rho(i)}\right) + \frac{1}{n\cdot n!}\sum_{\substack{\varrho\in S(n)\\ 1\leq i<k}}\varphi_{k-1}\left(X^{\varrho}\right)f_k\left(X_{\rho(i)}\right)\\
&= \frac{n-k+1}{n\cdot n!}\sum_{\varrho\in S(n)}\varphi_k\left(X^{\varrho}\right) + \frac{1}{n\cdot n!}\sum_{\varrho\in S(n)}\varphi_{k-1}\left(X^{\varrho}\right)\sum_{i=1}^{k-1} f_k\left(X_{\rho(i)}\right)\\
&= \frac{n-k+1}{n}A_n\left(\varphi_k\right) + R_{n,k}
\end{aligned}$$

where

$$\begin{aligned}
\left|R_{n,k}\right| &\leq 2\left\|\varphi_{k-1}\right\|_\infty\cdot\left\|f_k\right\|_\infty\cdot\frac{1}{n\cdot n!}\sum_{\varrho\in S(n)}(k-1)\\
&= 2\left\|\varphi_{k-1}\right\|_\infty\cdot\left\|f_k\right\|_\infty\cdot\frac{k-1}{n}.
\end{aligned}$$

Leeting $n \to \infty$, the desired result follows.

Sufficiency. Now let X be i.i.d. given $\mathcal{G}$ for a suitable σ -algebra $\mathcal{G} \subset \mathcal{F}$. For any bounded measurable function $\varphi : E^{\mathbb{N}} \to \mathbb{R}$ can be regraded as a function of the first n components, and for any $\varrho \in S(n)$, we have

$$\mathbb{E}[\varphi(X)|\mathcal{G}] = \mathbb{E}\left[\varphi\left(X^{\varrho}\right)|\mathcal{G}\right] .$$

Hence

$$\mathbb{E}[\varphi(X)] = \mathbb{E}[\mathbb{E}[\varphi(X)|\mathcal{G}]] = \mathbb{E}\left[\mathbb{E}\left[\varphi\left(X^{\varrho}\right)|\mathcal{G}\right]\right] = \mathbb{E}\left[\varphi\left(X^{\varrho}\right)\right] ,$$

whence X is exchangeable. □

Futher results* Denote by $\mathcal{M}_1(E)$ the set of probability measures on E equipped with the topology of weak convergence. That is, a sequence $(\mu_n)_{n\in\mathbb{N}}$ in $\mathcal{M}_1(E)$ converges weakly to a $\mu \in \mathcal{M}_1(E)$ if and only if $\int f\,d\mu_n \to \int f\,d\mu$ for any bounded continuous function $f : E \to \mathbb{R}$. At this point, we use the topology only to make $\mathcal{M}_1(E)$ a measurable space, namely with the Borel σ-algebra $\mathcal{B}\left(\mathcal{M}_1(E)\right)$. Now we can study random variables with values in $\mathcal{M}_1(E)$, so-called random measures .

For $x \in E^{\mathbb{N}}$, $\xi_n(x) = \frac{1}{n}\sum_{i=1}^n \delta_{x_i} \in \mathcal{M}_1(E)$. Then the random measure

$$\Xi_n := \xi_n(X) := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$$

is called the *empirical distribution* of $X_1, \ldots, X_n$.

Theorem (de Finetti Representation Theorem)**.** *The family $X = (X_n)_{n\in\mathbb{N}}$ is exchangeable if and only if there is a σ-algebra $\mathcal{G} \subset \mathcal{F}$ and an $\mathcal{G}$-measurable random measure $\Xi_\infty : \Omega \to \mathcal{M}_1(E)$ with the property that given Ξ_∞, $(X_n)_{n\in\mathbb{N}}$ is i.i.d. with $\mathcal{L}\left[X_1|\Xi_\infty\right] = \Xi_\infty$. In this case, we can choose $\mathcal{G} = \mathcal{E}$ or $\mathcal{G} = \mathcal{T}$.*

Proof. Let X be exchangeable. Then, by Theorem 1.64, there exists a σ -algebra $\mathcal{G} \subset \mathcal{F}$ such that $(X_n)_{n\in\mathbb{N}}$ is i.i.d. given $\mathcal{G}$. As E is Polish, there exists a regular conditional distribution, $\Xi_\infty := \mathcal{L}[X_1|\mathcal{G}]$. For measurable $A_1, \ldots, A_n \subset E$, we have $\mathbb{P}(X_i \in A_i|\mathcal{G}) = \Xi_\infty(A_i)$ for all $i = 1, \ldots, n$; hence

$$\begin{aligned}
&\mathbb{P}(X_i \in A_i, 1 \le i \le n|\Xi_\infty) \\
&= \mathbb{E}\big[\,\mathbb{P}(X_i \in A_i, 1 \le i \le n|\mathcal{G})\,|\,\Xi_\infty\big] \\
&= \mathbb{E}\left[\prod_{i=1}^{n} \Xi_\infty(A_i)\,|\Xi_\infty\right] = \prod_{i=1}^{n} \Xi_\infty(A_i)
\end{aligned}$$

Therefore, $\mathcal{L}[X|\Xi_\infty] = \Xi_\infty^{\otimes\mathbb{N}}$, the desired result follows. $\square$

Example 1.35. Let $(X_n)_{n\in\mathbb{N}}$ be exchangeable and assume $X_n \in \{0, 1\}$. Then there exists a random variable $Y : \Omega \to [0, 1]$ such that,i.e., for any two disjoint nonempty finite subset of $\mathbb{N}$, namely J_1 and J_0,

$$\mathbb{P}(X_j = 1, j \in J_1\,;\; X_j = 0, j \in J_2|Y) = Y^{|J_0|}(1 - Y)^{|J_2|}\,.$$

In other words, $(X_n)_{n\in\mathbb{N}}$ is independent given Y and Bernoulli(Y)-distributed. Compare Example 1.32.

This result is useful for people concerned about the foundations of statistics, since from the palatable assumption of symmetry one gets the powerful conclusion that the sequence is a mixture of i.i.d. sequences.

Chapter 2

Ergodic Theorems

Laws of large numbers, e.g., for i.i.d. integrable random variables $X_1, X_2, \ldots$, state that the sequence of averages converges a.s. to the expected value,

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}\, X_1 \quad \text{a.s.}.$$

Hence averaging over one realization of many random variables is equivalent to averaging over all possible realizations of one random variable. In the terminology of statistical physics this means that the *time average*, or *path* (Greek: *odos*) *average*, equals the *space average*. The "space" in "space average" is the probability space in mathematical terminology, and in physics it is considered the space of admissible states with a certain energy (Greek: *ergon*). Combining the Greek words gives rise to the name *ergodic theory*, which studies laws of large numbers for possibly dependent, but stationary, random variables.

In fract, the theory of stationary stochastic processes can be stated outside the framework of probability theory as the theory of one-parameter groups of transformations of a measure space that preserve the measure; this theory is very close to the general theory of dynamical systems and to ergodic theory.

2.1 Definitions and Examples

A Stationarty Sequence

Let $I \subset \mathbb{R}$ be a set that is closed under addition, for us the important examples are $\mathbb{N}_0$, $\mathbb{N}$, $\mathbb{Z}$, $[0, \infty)$ and $\mathbb{R}$.

Definition 2.1. A stochastic process $X = (X_t)_{t \in I}$ taking values in a Poish space $(E, \mathcal{B}(E))$ is called **stationary** (in the strict sense) if

$$\mathcal{L}\left[(X_{t+s})_{t \in I}\right] = \mathcal{L}\left[(X_t)_{t \in I}\right] \text{ for all } s \in I\,.$$

Remark 2.1. Although we give a general definition, we will focus on the case that X is a seqnence, i.e., $I = \mathbb{N}_0$. Moreover, one can see that, by induction, X is stationary if and only if

$$\mathcal{L}\left[(X_n)_{n \in \mathbb{N}_0}\right] = \mathcal{L}\left[(X_{n+1})_{n \in \mathbb{N}_0}\right]\,,$$

which is equivalent to $\mathcal{L}(X_0, \cdots, X_n) = \mathcal{L}(X_1, \cdots, X_{n+1})$ for all $n \in \mathbb{N}$.

We begin by giving four examples that will be our constant companions.

Example 2.1 (IID Sequence)**.** If $X = (X_n)_{n \geq 0}$ is i.i.d. random variables, then X is stationary. If only $X = (X_n)_{n \geq 0}$ is identically distributed (without the independence), then in general X is not stationary. For example, consider $X_1 = X_2 = X_3 = \cdots$ but $X_0 \neq X_1$. Then X is not stationary.

Example 2.2 (Markov Chain)**.** Let $X = (X_n)_{n \geq 0}$ be a Markov chain on a Poish space $(E, \mathcal{B}(E))$ with transition probability $P(x, A)$ and stationary distribution π, i.e., $\pi(A) = \int \pi(dx)P(x, A)$. If X_0 has distribution π then $X_0, X_1, \ldots$ is a stationary sequence. A special case to keep in mind for counterexamples is the chain with state space $E = \{0, 1\}$ and transition probability $P(x, \{1-x\}) = 1$. In this case, the stationary distribution has $\pi(0) = \pi(1) = 1/2$ and $(X_0, X_1, \ldots) = (0, 1, 0, 1, \ldots)$ or $(1, 0, 1, 0, \ldots)$ with probability $1/2$ each.

Example 2.3 (Rotation of the Circle)**.** Let $\Omega = [0,1), \mathcal{F} =$ Borel subsets, $\mathbb{P} =$ Lebesgue measure. Let $\theta \in (0,1)$, and for $n \geq 0$, let

$$X_n(\omega) = (\omega + n\theta) \bmod 1^{1}, \text{ for } \omega \in [0,1)\,.$$

This example is a special case of Example 2.2: Let $p(x, \{y\}) = 1$ if $y = (x + \theta) \bmod 1$.

B Measure-Preserving Dynamical Systerm

In the following, assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $\varphi : \Omega \to \Omega$ is a measurable map.

Definition 2.2. φ is called a **measure-preserving** transformation, if

$$\mathbb{P}\left(\varphi^{-1}(A)\right) = \mathbb{P}(A) \quad \text{for all } A \in \mathcal{F}\,. \tag{2.1}$$

In this case, $(\Omega, \mathcal{F}, \mathbb{P}, \varphi)$ is called a **measure-preserving dynamical system**. Evidently, the condition (2.1) can be changed by

$$\mathbb{P}\left(\varphi^{-k}(A)\right) = \mathbb{P}(A) \quad \text{for all } k \in \mathbb{N},\ A \in \mathcal{F}\,.$$

Remark 2.2. In fact, measure-preserving dynamical system have a close relationship with stationary sequence. Roughly speaking, we can regard them as the same thing and the reason can be found in Example 2.6. In my opinion, the measure-preserving dynamical system is more easy to deal with at most times.

Let us consider the physical hypotheses that lead to the consideration of measure preserving transformations. Suppose that Ω is the phase space of a system that evolves (in discrete time) according to a given law of motion. If ω is the state at instant $n = 1$, then $\varphi^n(\omega)$ [2] where φ is the translation operator

[1] $x \bmod 1 := x - [x]$.

[2] Let φ^n be the nth iterate of φ defined inductively by $\varphi^n = \varphi\left(\varphi^{n-1}\right)$, for $n \geq 1$, where $\varphi^0 =$ identity.

induced by the given law of motion, is the state attained by the system after n steps. Moreover, if A is some set of states ω, then $\varphi^{-1}(A) = \{\omega : \varphi\omega \in A\}$ is, by definition, the set of states ω that lead to A in one step. Therefore, if we interpret Ω as an incompressible fluid, the condition $\mathbb{P}\left(\varphi^{-1}(A)\right) = \mathbb{P}(A)$ can be thought of as the rather natural condition of conservation of volume.

Example 2.4. Let $\Omega = \{\omega_1, \ldots, \omega_n\}$ consist of $n \geq 2$ points. let $\mathcal{F} = 2^\Omega$ be the collection of its subsets, and let $\varphi(\omega_i) = \omega_{i+1}, 1 \leq i \leq n-1$, and $\varphi(\omega_n) = \omega_1$. If $\mathbb{P}(\omega_i) = 1/n$, then the transformation φ is measure-preserving.

Example 2.5 (Rotation of the Circle)**.** Let $\Omega = [0,1), \mathcal{F} = \mathcal{B}([0,1))$, $\mathbb{P}$ is the Lebesgue measure, $\theta \in [0,1)$ then $\varphi(x) = (x+\theta) \bmod 1$ for $x \in [0,1)$ is a measure-preserving transformation. To see the reason for the name, map $[0,1)$ into $\mathbb{C}$ by $x \to \exp(2\pi i x)$.

Example 2.6. $(\Omega, \mathcal{F}, \mathbb{P}, \varphi)$ is called a measure-preserving dynamical system. Given any random variable X_0 taking values in a Polish space $(E, \mathcal{B}(E))$, then we define $X = (X_n)_{n\geq 0}$ by setting

$$X_n = X_0 \circ \varphi^n\,, \text{ for } n \geq 0\,.$$

We claim that $X = (X_n)_{n\geq 0}$ is stationary. To check this, let $n \geq 1$, $B \in \mathcal{B}(\mathbb{R})^{n+1}$, and $k \geq 1$, then

$$\begin{aligned}
\mathbb{P}\left((X_k, \ldots, X_{k+n}) \in B\right) &= \mathbb{P}\left((X_0, \ldots, X_n) \circ \varphi^k \in B\right) \\
&= \mathbb{P}\left(\varphi^{-k}\{(X_0, \ldots, X_n) \in B\}\right) \\
&= \mathbb{P}\left((X_0, \ldots, X_n) \circ \in B\right)\,.
\end{aligned}$$

Thus X is stationary. This example of a stationary sequence is more than an important example. In fact, it is the only example!

Let $X = (X_n)_{n\geq 0}$ be a stationary sequence taking values in a Polish space $(E, \mathcal{B}(E))$. Consider the distribution of X, i.e., the probability measure $\mathbb{P}_X$

on sequence space $(E^{\mathbb{N}_0}, \mathcal{B}(E)^{\mathbb{N}_0})$. As we konw, projections $(\pi_n)_{n\geq 0}$ (that is $\pi_n(\omega) = \omega_n$ for all $\omega \in E^{\mathbb{N}_0}$) has the same distribution as $(X_n)_{n\geq 0}$, in other words, the identity mapping has the same distribution with X. If we let φ be the *shift operator*, i.e.,

$$\varphi(\omega_0, \omega_1, \ldots) = (\omega_1, \omega_2, \ldots) \text{ for } \omega \in E^{\mathbb{N}_0}.$$

Then $\pi_n(\omega) = \pi_0(\varphi^n \omega)$ for all $n \geq 0$ and φ is measure preserving since $(X_n)_{n\geq 0}$ is stationary.

We will now give some important definitions. Here and in what follows we assume that φ is a measure-preserving transformation on $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 2.3. An event $A \in \mathcal{F}$ is called **invariant** (under φ) if $\varphi^{-1}(A) = A$ and **quasi-invariant** if $\varphi^{-1}(A) = A$ a.s. (i.e., $1_{\varphi^{-1}(A)} = 1_A$ a.s.). Denote the σ-algebra of invariant events and quasi-invariant events, respectively, by

$$\mathcal{I} = \left\{A \in \mathcal{F} : \varphi^{-1}(A) = A\right\} \text{ and } \mathcal{I}^* = \left\{A \in \mathcal{F} : \varphi^{-1}(A) = A \text{ a.s. }\right\}.$$

It is easily verified that the classes $\mathcal{I}$ and $\mathcal{I}^*$ of invariant or quasi-invariant events, respectively, are σ-algebras. The following lemma establishes a connection between invariant and quasi-invariant sets.

Lemma 2.1. *If A is an quasi-invariant set, then there is an invariant set B such that $A = B$ a.s.*

Proof. Let $B = \limsup_n \varphi^{-n}(A)$. Then $\varphi^{-1}(B) = \limsup_n \varphi^{-(n+1)}A = B$, i.e., $B \in \mathcal{I}$. It is easily seen that $A\Delta B \subset \cup_{k=0}^{\infty}\left(\varphi^{-k}(A)\Delta\varphi^{-(k+1)}(A)\right)$. But

$$\mathbb{P}\left(\varphi^{-k}A\Delta\varphi^{-(k+1)}A\right) = \mathbb{P}\left(A\Delta\varphi^{-1}A\right) = 0$$

Hence $\mathbb{P}(A\Delta B) = 0$. □

Definition 2.4. The measure-preserving dynamical systerm $(\Omega, \mathcal{F}, \mathbb{P}, \varphi)$ is called **ergodic** if $\mathcal{I}$ is trivial, i.e., every $A \in \mathcal{I}$ has measure either zero or one. Sometimes, we say the measure-preserving transformation φ is ergodic for short.

Remark 2.3. Firstly, by Lemma 2.1, it's easy to see that φ is ergodic iff $\mathcal{I}^*$ is trivial. Secondly, if φ is not ergodic then the space can be split into two sets A and A^c, each having positive probability, so that $\varphi(A) = A$ and $\varphi(A^c) = A^c$. In other words, φ is not "irreducible."

Lemma 2.2. *Let* $(\Omega, \mathcal{F}, \mathbb{P}, \varphi)$ *be a measure-preserving dynamical systerm.*

(i) *A random variable* ξ *is* $\mathcal{I}$*-(*$\mathcal{I}^*$*-) measurable if and only if* X *is invariant (quasi-inavriant), i.e.,* $\xi \circ \varphi = \xi$ *(almost surely).*

(ii) $(\Omega, \mathcal{F}, \mathbb{P}, \varphi)$ *is ergodic if and only if any (quasi-)invariant random variable is almost surely constant.*

Proof. (i). The statement is obvious if $X = 1_A$ is an indicator function. The general case, can be inferred by the usual approximation arguments.

(ii). Assume that $(\Omega, \mathcal{F}, \mathbb{P}, \varphi)$ is ergodic. Then, for any $c \in \mathbb{R}$, we have $\{\xi > c\} \in \mathcal{I}$ and thus $\mathbb{P}(\xi > c) \in \{0, 1\}$. We conclude that

$$\xi = \inf\{c \in \mathbb{R} : \mathbb{P}(\xi > c) = 0\} \quad \text{a.s..}$$

Assume any $\mathcal{I}$-measurable map is a.s. constant. If $A \in \mathcal{I}$, then 1_A is $\mathcal{I}$-measurable and hence a.s. equals either 0 or 1. Thus $\mathbb{P}(A) \in \{0, 1\}$. □

To investigate further the meaning of ergodicity, we return to our examples, renumbering them because the new focus is on checking ergodicity.

Example 2.7 (Rotation of the Circle)**.** Example 2.5 is not ergodic if $\theta = m/n$ where $m < n$ are positive integers. Since if B is a Borel subset of $[0, 1/n)$ and

$$A = \bigcup_{k=0}^{n-1} (B + \frac{k}{n})$$

then A is invariant.

Conversely, if θ is irrational, then φ is ergodic. To prove this, we need a fact from Fourier analysis. If f is a measurable function on [0,1) with $\int_0^1 f^2(x)dx < \infty$, then f can be written as $f(x) = \sum_{k\in\mathbb{Z}} c_k e^{2\pi ikx}$ where the equality is in the sense that as $K \to \infty$,

$$\sum_{k=-K}^{K} c_k e^{2\pi ikx} \to f(x) \text{ in } L^2[0,1)\,.$$

and this is possible for only one choice of the coefficients $c_k = \int_0^1 f(x)e^{-2\pi ikx}dx$. Now

$$f(\varphi(x)) = \sum_{k\in\mathbb{Z}} c_k e^{2\pi ik(x+\theta)} = \sum_{k\in\mathbb{Z}} \left(c_k e^{2\pi ik\theta}\right) e^{2\pi ikx}$$

The uniqueness of the coefficients c_k implies that $f(\varphi(x)) = f(x)$ if and only if $c_k\left(e^{2\pi ik\theta} - 1\right) = 0$. If θ is irrational, this implies $c_k = 0$ for $k \neq 0$, so f is constant. Applying the last result to $f = 1_A$ with $A \in \mathcal{I}$ shows that $A = \emptyset$ or [0,1) a.s.

Definition 2.5. The stationary sequence $X = (X_n)_{n\geq 0}$ taking values in a Polish space $(E, \mathcal{B}(E))$ is called **ergodic**, if the measure-preserving dynamical systerm $(\mathbb{E}^{\mathbb{N}_0}, \mathcal{B}(E)^{\mathbb{N}_0}, \mathbb{P}_X, \varphi)$ (from Example 2.6) is ergodic.

Remark 2.4. We can define the ergodic sequence in another way. Let $X = (X_n)_{n\geq 0}$ be a stationary sequence taking values in a Polish space E. A set $A \in \mathcal{F}$ is called invariant with respect to X, if there is a set $B \in \mathcal{B}(E)^{\mathbb{N}_0}$ such that $B = \varphi^{-1}(B)$ (recall that φ is the shift operator on $E^{\mathbb{N}_0}$) and

$$A = \{(X_n)_{n\geq 0} \in B\}\,.$$

The collection of all such invariant sets is a σ-algebra, denoted by $\mathcal{I}_X$. Similarly, we can define $\mathcal{I}_X^*$. Indeed, let $\mathcal{I}, \mathcal{I}^*$ be the invariant and quasi-invariant σ-field of $(\mathbb{E}^{\mathbb{N}_0}, \mathcal{B}(E)^{\mathbb{N}_0}, \mathbb{P}_X, \varphi)$, respectively, then $\mathcal{I}_X = X^{-1}(\mathcal{I})$ and $\mathcal{I}_X^* = X^{-1}(\mathcal{I}^*)$. Thus X is ergodic if and only if $\mathcal{I}_X$ or $\mathcal{I}_X^*$ is trivial.

Proposition 2.3. *If $X = (X_n)_{n\geq 0}$ is a stationary sequence taking values in a Polish space E, and $g : E^{\mathbb{N}_0} \to \mathbb{R}$ is measurable. Define $Y = (Y_n)_{n\geq 0}$ by letting $Y_n = g(X_n, X_{n+1}, \ldots)$. Then Y is stationary and if X is ergodic then so is Y.*

Proof. Denote by φ the shift operator on $E^{\mathbb{N}_0}$. Then $Y_n = g(\varphi^n(X))$. Define $\psi : \mathbb{E}^{\mathbb{N}_0} \to \mathbb{E}^{\mathbb{N}_0}$,

$$x \mapsto (g(\varphi^n(x)))_{n\geq 0}\,.$$

Then $Y = \psi(X)$ and $\varphi(Y) = \psi(\varphi(X))$ since $\varphi \circ \psi = \psi \circ \varphi$. Since $\mathcal{L}(X) = \mathcal{L}(\varphi(X))$, we have $\mathcal{L}(Y) = \mathcal{L}(\varphi(Y))$, i.e., Y is stationary.

We show that $\mathcal{I}_Y \subset \mathcal{I}_X$. Take any $B \in \mathcal{I}$, then $\{Y \in B\} = \{X \in \psi^{-1}(B)\}$. It suffices to show that $\psi^{-1}(B) \in \mathcal{I}$, which is evident:

$$\varphi^{-1} \circ \psi^{-1}(B) = \psi^{-1} \circ \varphi^{-1}(B) = \psi^{-1}(B)\,.$$

So the desired result follows. □

Example 2.8 (IID Sequence)**.** Let $X = (X_n)_{n\in\mathbb{N}_0}$ be i.i.d. random variables valued in a Polish space E. Then for any invariant set $A \in \mathcal{I}_X$, and $k \in \mathbb{N}$,

$$A = \{(X_n)_{n\geq 0} \in B\} = \{(X_{n+k})_{n\geq 0} \in B\} \in \sigma(X_k, X_{k+1}, \ldots)\,.$$

Thus

$$A \in \bigcap_{n=1}^{\infty} \sigma(X_n, X_{n+1}, \ldots) = \mathcal{T}, \qquad \text{the tail sigma-field,}$$

so $\mathcal{I}_X \subset \mathcal{T}$. For an i.i.d. sequence, Kolmogorov's 0-1 law implies $\mathcal{T}$ is trivial, so $\mathcal{I}_X$ is trivial and the sequence is ergodic.

Example 2.9 (Markov Chains)**.** Suppose that $X = (X_n)_{n\geq 0}$ is a Markov chain on the countable state space S with initial stationary distribution π so that $\pi(x) > 0$ for all $x \in S$. As we konw, all states are (positive) recurrent, and we can write

$$S = \cup_i R_i\,,$$

where the R_i are disjoint irreducible closed sets. If $X_0 \in R_i$ then with probability one, $X_n \in R_i$ for all $n \geq 1$ almost surely, so

$$\{X_0 \in R_i\} \in \mathcal{I}_X^*.$$

The last observation shows that if the Markov chain is reducible then the sequence is not ergodic.

To prove the converse, observe that if $A = \{X \in B\} \in \mathcal{I}_X$, where $B \in (2^S)^{\mathbb{N}_0}$ so that $B = \varphi^{-1}(B)$, where φ is the shift operator on $S^{\mathbb{N}_0}$. Then $1_A \circ \varphi^n = 1_A$. So if we let $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$, the shift invariance of 1_A and the Markov property imply

$$\mathbb{E}_\pi(1_A|\mathcal{F}_n) = \mathbb{E}_\pi\left(1_{\{X\in B\}}|\mathcal{F}_n\right) = \mathbb{E}_\pi\left(1_{\{\varphi^n(X)\in B\}}|\mathcal{F}_n\right) = h(X_n)$$

where $h(x) = \mathbb{E}_x 1_A$ for $x \in S$. Lévy's 0-1 law implies that the left-hand side converges to 1_A as $n \to \infty$. Since $X = (X_n)_{n\geq 0}$ is irreducible and recurrent, then for any $x \in S$, the righthand side $= h(x)$ i.o., so either $h(x) = 0$ or $h(x) = 1$ and $\mathbb{P}_\pi(A) \in \{0, 1\}$ almost surely.

This example also shows that $\mathcal{I}$ and $\mathcal{T}$ may be different. When the transition probability p is irreducible $\mathcal{I}$ is trivial, but if all the states have period $d > 1$, $\mathcal{T}$ is not. Since if $S_0, \ldots, S_{d-1}$ is the cyclic decomposition of S, then

$$\mathcal{T} = \sigma(\{X_0 \in S_r\} : 0 \leq r < d) .$$

The proof can be found in [2], Theorem 5.7.3.

2.2 Ergodic Theorems

Our basic set up consists of

$(\Omega, \mathcal{F}, \mathbb{P}, \varphi)$	a measure-preserving dynamical system,
$X_n(\omega) = X_0(\varphi^n\omega)$	where X_0 is a random variable.

Hence $X = (X_n)_{n\geq 0}$ is a stationary real-valued stochastic process. Let $S_0 = 0$ and

$$S_n = \sum_{k=0}^{n-1} X_k$$

denote the nth partial sum for $n \geq 1$. Ergodic theorems are laws of large numbers for (S_n).

Theorem 2.4 (Individual Ergodic Theorem)**.** *Let $X_0 \in L^1$. Then*

$$\frac{S_n}{n} = \frac{1}{n}\sum_{k=0}^{n-1} X_0 \circ \varphi^k \to \mathbb{E}\left(X_0|\mathcal{I}\right) \quad a.s.$$

In particular, if φ is ergodic, then $S_n/n \to \mathbb{E}\, X_0$ a.s..

The result due to Birkhoff is sometimes called the pointwise or individual ergodic theorem. When the sequence is ergodic, and we take $X_0 = 1_A$, it follows that the asymptotic fraction of time $\varphi^m \in A$ is $\mathbb{P}(A)$.

The proof we give is based on an odd integration inequality due to Yosida and Kakutani. We follow Garsia. The proof is not intuitive, but none of the steps are difficult. We start with a preliminary lemma.

Lemma 2.5 (Maximal Ergodic Lemma)**.** *Suppose that $X_0 \in L^1$. Define $M_n = \max\{0, S_1, \ldots, S_n\}, n \in \mathbb{N}$. Then*

$$\mathbb{E}\left(X_0 1_{\{M_n>0\}}\right) \geq 0 \quad \textit{for every } n \in \mathbb{N}\,.$$

Proof. Note tha $M_n(\varphi(\omega)) \geq S_k(\varphi(\omega))$, for $1 \leq k \leq n$, we have

$$X_0 + M_n \circ \varphi \geq X_0 + S_k \circ \varphi = S_{k+1}\,.$$

Thus $X_0 \geq S_{k+1} - M_n \circ \varphi$ for $1 \leq k \leq n$. Manifestly, $S_1 = X_0$ and $M_n \circ \varphi \geq 0$ and hence also (for $k = 0$) $X_0 \geq S_1 - M_n \circ \varphi$. Therefore,

$$X_0 \geq \max\{S_1, \ldots, S_n\} - M_n \circ \varphi\,.$$

Furthermore, we have

$$\{M_n = 0\} \subset \{M_n = 0\} \cap \{M_n \circ \varphi \geq 0\} \subset \{M_n - M_n \circ \varphi \leq 0\} \ .$$

Since φ is measure-preserving, we conclude that

$$\begin{aligned}\mathbb{E}\left(X_0 1_{\{M_n>0\}}\right) &\geq \mathbb{E}\left[\left(\max\{S_1,\ldots,S_n\} - M_n \circ \varphi\right) 1_{\{M_n>0\}}\right] \\ &= \mathbb{E}\left[\left(M_n - M_n \circ \varphi\right) 1_{\{M_n>0\}}\right] \\ &\geq \mathbb{E}\left(M_n - M_n \circ \varphi\right) = \mathbb{E}\left(M_n\right) - \mathbb{E}\left(M_n\right) = 0\,.\end{aligned}$$

We complete the proof. □

Proof of Theorem 2.4. By Lemma 2.2, we have

$$\mathbb{E}\left(X_0|\mathcal{I}\right) \circ \varphi = \mathbb{E}\left(X_0|\mathcal{I}\right) \quad \text{a.s..}$$

Hence, by passing to $\tilde{X}_n := X_n - \mathbb{E}\left(X_0|\mathcal{I}\right)$, without loss of geneality, we can assume $\mathbb{E}\left(X_0|\mathcal{I}\right) = 0$. Define

$$Z := \limsup_{n\to\infty} \frac{1}{n} S_n\,.$$

Let $\epsilon > 0$, we shall show that $\mathbb{P}(Z > \epsilon) = 0$. From this we infer $\mathbb{P}(Z > 0) = 0$ and similarly (with $-X$ instead of X) also $\liminf_n \frac{1}{n} S_n \geq 0$ a.s.. Then we get the desired result: $\frac{1}{n} S_n \to 0$ a.s..

Evidently, $Z \circ \varphi = Z$; hence $\{Z > \epsilon\} \in \mathcal{I}$. Define

$$\begin{aligned}X_n^\epsilon &:= \left(X_n - \epsilon\right) 1_{\{Z>\epsilon\}}, \\ S_n^\epsilon &:= X_0^\epsilon + \ldots + X_{n-1}^\epsilon = \left(S_n - \epsilon\right) 1_{\{Z>\epsilon\}}\,, \\ M_n^\epsilon &:= \max\{0, S_1^\epsilon, \ldots, S_n^\epsilon\} \ .\end{aligned}$$

Then $\{M_n^\epsilon > 0\} \uparrow$ and

$$\bigcup_{n=1}^{\infty} \{M_n^\epsilon > 0\} = \{Z > \epsilon\}\,.$$

Dominated convergence yields $\mathbb{E}\left(X_0^\epsilon 1_{\{M_n^\epsilon>0\}}\right) \to \mathbb{E}\left(X_0^\epsilon 1_{\{Z>\epsilon\}}\right) = \mathbb{E}\,X_0^\epsilon$. By the maximal-ergodic lemma (applied to X^ϵ), we have

$$\mathbb{E}\,X_0^\epsilon = \lim_{n\to\infty} \mathbb{E}\left(X_0^\epsilon 1_{\{M_n^\epsilon>0\}}\right) \geq 0\,.$$

However,

$$\begin{aligned}\mathbb{E}\left(X_0^\epsilon\right) &= \mathbb{E}\left[\left(X_0-\epsilon\right)1_{\{Z>\epsilon\}}\right] \\ &= \mathbb{E}\left[\,\mathbb{E}\left(X_0|\mathcal{I}\right)1_{\{Z>\epsilon\}}\right] - \epsilon\,\mathbb{P}(Z>\epsilon) \\ &= 0-\epsilon\,\mathbb{P}(Z>\epsilon)\,.\end{aligned}$$

We conclude that $\mathbb{P}(Z>\epsilon)=0$. □

As a consequence, we obtain the *statistical ergodic theorem*, or *L^p-ergodic theorem*, that was found by von Neumann in 1931 right before Birkhoff proved his ergodic theorem, but was published only later. Before we formulate it, we state one more lemma.

Lemma 2.6. *Let $p \geq 1$ and let $X_0, X_1, \ldots.$ be identically distributed, real random variables with $X_0 \in L^p$. Define $Y_n := \left|\frac{1}{n}\sum_{k=0}^{n-1} X_k\right|^p$ for $n \in \mathbb{N}$. Then $(Y_n)_{n\in\mathbb{N}}$ is uniformly integrable.*

Proof. Evidently, the singleton $\{|X_0|^p\}$ is uniformly integrable. Hence, there exists a monotone increasing convex map $f:[0,\infty)\to[0,\infty)$ with $\frac{f(x)}{x}\to\infty$ for $x\to\infty$ and $\mathbb{E}\left[f\left(|X_0|^p\right)\right]<\infty$. It is enough to show that

$$\sup_{n\in\mathbb{N}} \mathbb{E}\left[f\left(Y_n\right)\right] < \infty\,.$$

By Jensen's inequality (for $x \mapsto |x|^p$), we have

$$Y_n \leq \frac{1}{n}\sum_{k=0}^{n-1} |X_k|^p$$

Again, by Jensen's inequality (now applied to f), we get that

$$f\left(Y_n\right) \leq f\left(\frac{1}{n}\sum_{k=0}^{n-1}|X_k|^p\right) \leq \frac{1}{n}\sum_{k=0}^{n-1} f\left(|X_k|^p\right)$$

Hence $\mathbb{E}\left[f\left(Y_n\right)\right] \leq \frac{1}{n}\sum_{k=0}^{n-1}\mathbb{E}\left[f\left(|X_k|^p\right)\right] = \mathbb{E}\left[f\left(|X_0|^p\right)\right]$ for all $n\in\mathbb{N}$. □

Corollary 2.7 (L^p-Ergodic Theorem)**.** *Let $p \geq 1$ and $X_0 \in L^p$, then*

$$\frac{S_n}{n} = \frac{1}{n} \sum_{k=0}^{n-1} X_k \to \mathbb{E}\left(X_0 | \mathcal{I}\right) \qquad \textit{in } L^p\,.$$

In particular, if φ is ergodic, then $\frac{S_n}{n} \to \mathbb{E}\, X_0$ in L^p.

Proof. Although this follows trivially from Lemma 2.6, we shall give a direct proof. We have to show that

$$\left\| \frac{1}{n} \sum_{k=0}^{n-1} X_k - \mathbb{E}\left(X_0|\mathcal{I}\right) \right\|_p \to 0\,.$$

Let $M > 0$ be a large real numbers. We will truncate X_n at level M: Define $X_n' = X_n 1_{\{|X_n| \leq M\}}$ for all n. Clearly we have $X_n' = X_0' \circ \varphi^n$. Let $X_n'' = X_n 1_{\{|X_n| > M\}}$. Using the inequality $|x + y|^p \leq 2^{p-1}|x|^p + |y|^p$ for $x, y \in \mathbb{R}$, we have

$$\begin{aligned} &\left\| \frac{1}{n} \sum_{k=0}^{n-1} X_k - \mathbb{E}\left(X_0|\mathcal{I}\right) \right\|_p \\ &\leq \left\| \frac{1}{n} \sum_{k=0}^{n-1} X_k' - \mathbb{E}\left(X_0'|\mathcal{I}\right) \right\|_p + \frac{1}{n} \sum_{k=0}^{n-1} \left\| X_0'' \circ \varphi^k \right\|_p + \left\| \mathbb{E}\left(X_0''|\mathcal{I}\right) \right\|_p\,. \end{aligned}$$

It follows from the individual ergodic theorem and the bounded convergence theorem that

$$\left\| \frac{1}{n} \sum_{k=0}^{n-1} X_k' - \mathbb{E}\left(X_0'|\mathcal{I}\right) \right\|_p \to 0\,.$$

On the other hand, for each k,

$$\left\| X_0'' \circ \varphi^k \right\|_p = \left\| X_0'' \right\|_p = \left\| X_0 1_{\{|X_0| > M\}} \right\|_p\,,$$

and by Jensen's inequality $|\mathbb{E}(X_0''|\mathcal{I})|^p \leq \mathbb{E}(|X_0''|^p|\mathcal{I})$, so

$$\left\| \mathbb{E}\left(X_0''|\mathcal{I}\right) \right\|_p \leq \left\| X_0'' \right\|_p = \left\| X_0 1_{\{|X_0| > M\}} \right\|_p\,.$$

Thus,

$$\limsup_{n\to\infty}\left\|\frac{1}{n}\sum_{k=0}^{n-1}X_k-\mathbb{E}\left(X_0|\mathcal{I}\right)\right\|_p\leq 2\left\|X_0 1_{\{|X_0|>M\}}\right\|_p .$$

Letting $M\to\infty$, the desired result follows. □

Corollary 2.8 (Ergodic Theorem for Stationary Sequence)**.** *Let $\xi=(\xi_n)_{n\geq 0}$ be a stationary random sequence with $\xi_0\in L^p$ where $p\geq 1$. Then*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\xi_k=\mathbb{E}\left(\xi_0|\mathcal{I}_\xi\right)\qquad a.s.\ \ and\ in\ L^p .$$

Proof. It follows evidently from the individual ergodic theorem that $\frac{1}{n}\sum_{k=0}^{n-1}\xi_k$ converges a.s. and in L^p. We have only to show that if the random variable η is the limit (a.s. and in L^p) of $\frac{1}{n}\sum_{k=1}^{n}\xi_k$, then it can be taken equal to $\mathbb{E}\left(\xi_0|\mathcal{I}_\xi\right)$. To this end, notice that we can set

$$\eta(\omega)=\limsup_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\xi_k(\omega)\,,\ \text{for } \omega\in\Omega\,.$$

It follows from the definition of limsup that for the random variable $\eta(\omega)$ so defined, the sets $\{\eta<y\}\in\mathcal{I}_\xi$, for $y\in\mathbb{R}$, and therefore η is $\mathcal{I}_\xi$-measurable. We shall show that for any $A\in\mathcal{I}_\xi$,

$$\mathbb{E}(\eta 1_A)=\mathbb{E}(\xi_0 1_A)\,.$$

Then the desired result follows.

By the L^p convergence, we have

$$\frac{1}{n}\sum_{k=0}^{n-1}\mathbb{E}(\xi_k 1_A)\to\mathbb{E}(\eta 1_A)\,.$$

Assume $B\in\mathcal{B}\left(\mathbb{R}\right)^{\mathbb{N}_0}$ is invariant (under the shift operator φ) such that $A=\{(\xi_n)_{n\geq 0}\in B\}=\{(\xi_{n+k})_{n\geq 0}\in B\}$ for all $k\geq 0$. Then since ξ is stationary,

$$\mathbb{E}\left(\xi_k 1_A\right)=\mathbb{E}\left(\xi_k 1_{\left\{(\xi_{n+k})_{n\geq 0}\in B\right\}}\right)=\mathbb{E}\left(\xi_0 1_{\left\{(\xi_n)_{n\geq 0}\in B\right\}}\right)=\mathbb{E}\left(\xi_0 1_A\right) .$$

Hence it follows that $\eta=\mathbb{E}(\xi_0|\mathcal{I}_\xi)$. Moreover, as we can see, $\mathbb{E}(\xi_k|\mathcal{I}_\xi)=\mathbb{E}(\xi_0|\mathcal{I}_\xi)$ for all $k\geq 1$. □

A Examples

Our next step is to see what ergodic theorems says about our examples.

Example 2.10 (IID Sequence)**.** Since $\mathcal{I}$ is trivial, the ergodic theorem implies that

$$\frac{1}{n}\sum_{m=0}^{n-1} X_m \to \mathbb{E}X_0 \qquad \text{a.s. and in } L^1$$

The a.s. convergence is the strong law of large numbers.

Remark 2.5. We can prove the L^1 convergence in the law of large numbers without invoking the ergodic theorem. By SLLN and monotone convergence theorem,

$$\mathbb{E}\,\frac{1}{n}\sum_{m=0}^{n-1} X_m^+ \to \mathbb{E}X_0^+ \ ; \ \mathbb{E}\,\frac{1}{n}\sum_{m=0}^{n-1} X_m^- \to \mathbb{E}X_0^- \, .$$

Convergence of the L^1-norm and almost surely convergence imply the convergence in L^1.

Example 2.11 (Markov Chain)**.** Let $X = (X_n)_{n\geq 0}$ be an irreducible Markov chain on a countable state space S with initial stationary distribution π. Let f be a function on S with

$$\sum_{x\in S} |f(x)|\pi(x) < \infty$$

In Example 2.9 we showed that $\mathcal{I}_X$ is trivial, note that $\mathbb{E}_\pi|f(X_0)| < \infty$, so applying the ergodic theorem gives

$$\frac{1}{n}\sum_{m=0}^{n-1} f\left(X_m\right) \to \sum_{x\in S} f(x)\pi(x) \quad \mathbb{P}_\pi\text{-a.s. and in } L^1(\mathbb{P}_\pi)\, .$$

Recall that in the course of Markov chain, we have shown that the ergodic theorem holds for X under $\mathbb{P}_x$, given any $x \in S$, although X can be not stationary under $\mathbb{P}_x$.

Example 2.12. Let $\mathbb{P}$ and $\tilde{\mathbb{P}}$ be probability measures on the measurable space $(\Omega, \mathcal{F})$, and let $(\Omega, \mathcal{F}, \mathbb{P}, \varphi)$ and $(\Omega, \mathcal{F}, \tilde{\mathbb{P}}, \varphi)$ be ergodic. Then either $\mathbb{P} = \tilde{\mathbb{P}}$ or $\mathbb{P} \perp \tilde{\mathbb{P}}$.

Indeed, if $\mathbb{P} \neq \tilde{\mathbb{P}}$, then there exists an random variable X_0 with $|X_0| \leq 1$ and $\int X_0 \, \mathrm{d}\mathbb{P} \neq \int X_0 \, \mathrm{d}\tilde{\mathbb{P}}$. However, by individual ergodic theorem,

$$\frac{1}{n} \sum_{k=0}^{n-1} X_0 \circ \varphi^k \overset{n\to\infty}{\longrightarrow} \begin{cases} \int X_0 \ \mathrm{d}\mathbb{P} & \mathbb{P} \text{ -a.s.} \\ \int X_0 \ \mathrm{d}\tilde{\mathbb{P}} & \tilde{\mathbb{P}} \text{ -a.s.} \end{cases}$$

If we define

$$A = \left\{ \frac{1}{n} \sum_{k=0}^{n-1} X_0 \circ \varphi^k \to \int X_0 \ \mathrm{d}\mathbb{P} \right\},$$

then $\mathbb{P}(A) = 1$ and $\tilde{\mathbb{P}}(A) = 0$, Thus $\mathbb{P} \perp \tilde{\mathbb{P}}$.

Example 2.13 (Rotation of the Circle)**.** Suppose that $\theta \in (0, 1)$ is irrational, so that by a result in Example 2.7 $\mathcal{I}$ is trivial. If we set $X_0(\omega) = 1_A(\omega)$, with $A \in \mathcal{B}([0, 1))$, then the ergodic theorem implies

$$\frac{1}{n} \sum_{m=0}^{n-1} 1_A(\varphi^m \omega) \to \lambda(A) \,, \quad \lambda\text{-a.s. } \omega \,,$$

where λ denotes the Lebesgue measure. The last result for $\omega = 0$ is usually called *Weyl's equidistribution theorem*, although Bohl and Sierpinski should also get credit.

To recover the number theoretic result, we will now show that:

Theorem. *If $A = [a, b)$ then the exceptional set is $\emptyset$.*

To see this, let $A_k = [a + 1/k, b - 1/k)$. If $b - a > 2/k$, the ergodic theorem implies

$$\frac{1}{n} \sum_{m=0}^{n-1} 1_{A_k} (\varphi^m \omega) \to b - a - \frac{2}{k}$$

for $\omega \in \Omega_k$ with $\lambda(\Omega_k) = 1$. Let $G = \cap_k \Omega_k$, where the intersection is over integers k with $b - a > 2/k$. $\lambda(G) = 1$ so G is dense in $[0, 1)$. If $x \in [0, 1)$ and $\omega_k \in G$ with $|\omega_k - x| < 1/k$, then $\varphi^m \omega_k \in A_k$ implies $\varphi^m x \in A$, so

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} 1_A(\varphi^m x) \geq b - a - \frac{2}{k}$$

for all large enough k. Noting that k is arbitrary and applying similar reasoning to A^c shows

$$\frac{1}{n} \sum_{m=0}^{n-1} 1_A(\varphi^m x) \to b - a\,.$$

Example 2.14 (Benford's Law)**.** As Gelfand first observed, the equidistribution theorem says something interesting about 2^m. Let $\theta = \log_{10} 2$,and $A_k = [\log_{10} k, \log_{10}(k+1))$ where $1 \leq k \leq 9$. Taking $x = 0$ in the last result, we have

$$\frac{1}{n} \sum_{m=0}^{n-1} 1_{A_k}(m\theta \bmod 1) \to \log_{10}\left(\frac{k+1}{k}\right).$$

A little thought reveals that the first digit of $2^m = 10^{m\theta}$ is k if and only if $m\theta$ mod $1 \in A_k$, thus

$$\frac{1}{n} \sum_{m=0}^{n-1} 1_{\{\text{the first digit of } 2^m \text{ is } k\}} \to \log_{10}\left(\frac{k+1}{k}\right).$$

The numerical values of the limiting probabilities are

1	2	3	4	5	6	7	8	9
.3010	.1761	.1249	.0969	.0792	.0669	.0580	.0512	.0458

The limit distribution on $\{1, \ldots, 9\}$ is called Benford's (1938) law, although it was discovered by Newcomb (1881). As Raimi (1976) explains, in many tables the observed frequency with which k appears as a first digit is approximately $\log_{10}((k+1)/k)$. Some of the many examples that are supposed to follow Benford's law are: census populations of 3259 counties, 308 numbers from Reader's

Digest, areas of 335 rivers, 342 addresses of American Men of Science. The next table compares the percentages of the observations in the first five categories to Benford's law:

	1	2	3	4	5
Census	33.9	20.4	14.2	8.1	7.2
Reader's Digest	33.4	18.5	12.4	7.5	7.1
Rivers	31.0	16.4	10.7	11.3	7.2
Benford's Law	30.1	17.6	12.5	9.7	7.9
Addresses	28.9	19.2	12.6	8.8	8.5

The fits are far from perfect, but in each case Benford's law matches the general shape of the observed distribution. The IRS and other government agencies use Benford's law to detect fraud. When records are made up the first digit distribution does not match Benford's law.

2.3 Applications

A Recurrence of Random Walks

In this section, we will study the recurrence properties of stationary sequences. Our first result is an application of the ergodic theorem. Let $X = (X_n)_{n\in\mathbb{N}}$ be a stationary sequence taking values in $\mathbb{R}^d$. Let $S_n = X_1 + \cdots + X_n$, and $S_0 = 0$. let $T^0 = 0$ and $T^j = \{n > T^{j-1} : S_n = 0\}$. Then

$$\{T^1 = \infty\} = \{S_n \neq 0 \text{ for all } n \geq 1\}$$

be the event of an "escape" from 0. Let

$$R_n = |\{S_1, \dots, S_n\}|$$

denote the range of $S = (S_n)_{n\geq 1}$; that is, the number of points visited from time 1 to time n.

Lemma 2.9. *As $n \to \infty$, $R_n/n \to \mathbb{P}\left(T^1 = \infty | \mathcal{I}_X\right)$ a.s.*

Proof. Suppose $X = (X_n)_{n\in\mathbb{N}}$ be the *canonical process*, that is $(\Omega, \mathcal{F}, \mathbb{P}) = \left(\left(\mathbb{R}^d\right)^{\mathbb{N}}, \mathcal{B}\left(\mathbb{R}^d\right)^{\mathbb{N}}, \mathbb{P}\right)$ with $X_n(\omega) = \omega_n$ for $\omega \in \left(\mathbb{R}^d\right)^{\mathbb{N}}$. Denote by φ the shift operator on $(\mathbb{R}^d)^{\mathbb{N}}$. Now, $(\Omega, \mathcal{F}, \mathbb{P}, \varphi)$ is a typically measure-preserving dynamical systerm. Firstly, note that

$$\begin{aligned} R_n &= \sum_{k=1}^{n} 1_{\{S_k \neq S_l \text{ for all } k<l\leq n\}} \geq \sum_{k=1}^{n} 1_{\{S_k \neq S_l \text{ for all } l>k\}} \\ &= \sum_{k=1}^{n} 1_{\{S_l - S_k \neq 0 \text{ for all } l>k\}} = \sum_{k=1}^{n} 1_{\{T^1=\infty\}} \circ \varphi^k . \end{aligned}$$

Birkhoff's ergodic theorem yields

$$\liminf_{n\to\infty} \frac{R_n}{n} \geq \mathbb{P}(T^1 = \infty | \mathcal{I}) \quad \text{a.s.}$$

For the converse inequality, note that for every $n \geq m$,

$$\begin{aligned} R_n &= \sum_{k=1}^{n} 1_{\{S_k \neq S_l \text{ for all } k<l\leq n\}} \leq \sum_{k=1}^{n-m} 1_{\{S_l \neq S_k \text{ for all } k<l\leq n\}} + m \\ &\leq \sum_{k=1}^{n-m} 1_{\{S_l - S_k \neq 0 \text{ for all } k<l\leq k+m\}} + m = \sum_{k=1}^{n-m} 1_{\{T^1>m\}} \circ \varphi^k + m . \end{aligned}$$

Again, by the individual ergodic theorem,

$$\limsup_{n\to\infty} \frac{R_n}{n} \leq \mathbb{P}\left(T^1 > m | \mathcal{I}\right) \quad \text{a.s..}$$

Since $\left\{T^1 > m\right\} \downarrow \left\{T^1 = \infty\right\}$, by domainted convergence theorem for conditional expection, we have $\mathbb{P}\left(T^1 > m | \mathcal{I}\right) \to \mathbb{P}\left(T^1 = \infty | \mathcal{I}\right)$. Thus

$$\frac{R_n}{n} \to \mathbb{P}\left(T^1 = \infty | \mathcal{I}\right) \quad \text{a.s..}$$

Finally, suppose X is not the canonical process. by the proof above, R_n/n converges almost surely. Let

$$\xi(\omega) \coloneqq \limsup_{n\to\infty} \frac{R_n(\omega)}{n}, \text{ for } \omega \in \Omega .$$

Then $R_n/n \to \xi$ a.s. and in L^1 (since $R_n/n \leq 1$), it suffices to show that $\xi = \mathbb{P}(T^1 = \infty|\mathcal{I}_X)$ a.s., i.e.,

$$\mathbb{E}\,\xi 1_A = \mathbb{P}(A \cap \{T^1 = \infty\})\,,\ \text{for } A \in \mathcal{I}_X\,.$$

Consider the canonical process: $\tilde{X} = (\tilde{X}_n)$ on measure-preserving dynamical systerm $\left(\left(\mathbb{R}^d\right)^{\mathbb{N}}, \mathcal{B}\left(\mathbb{R}^d\right)^{\mathbb{N}}, \mathbb{P}_X, \varphi\right)$, where φ is the shift operator. Let $\tilde{R}_n$ and $\tilde{T}^1$ be defined as R_n and T^1, respectively. Then

$$\{T^1 = \infty\} = \{X \in \{\tilde{T}^1 = \infty\}\}\,.$$

Take any $A = \{X \in \tilde{A}\} \in \mathcal{I}_X$, where $\tilde{A} \in \mathcal{I}$, then

$$\begin{aligned}\mathbb{E}\,\xi 1_A &= \lim_{n\to\infty} \mathbb{E}\big(\frac{R_n}{n} 1_{\{X\in\tilde{A}\}}\big) = \lim_{n\to\infty} \int_{\tilde{A}} \frac{\tilde{R}_n}{n}\,\mathrm{d}\mathbb{P}_X = \int_{\tilde{A}} \mathbb{P}_X(\tilde{T}^1 = \infty|\mathcal{I})\,\mathrm{d}\mathbb{P}_X \\ &= \mathbb{P}_X(\{\tilde{T}^1 = \infty\} \cap \tilde{A}) = \mathbb{P}(A \cap \{T_1 = \infty\})\,.\end{aligned}$$

We complete the proof. $\square$

From Lemma 2.9, we get a result about the recurrence of random walks with stationary increments that is (for integer valued random walks) a generalization of the *Chung-Fuchs theorem* : a random walk on $\mathbb{Z}$ with centered increments is recurrent.

Theorem 2.10. *Let $X = (X_n)_{n\geq 1}$ be an integer-valued, integrable, stationary sequence. Let $S_n = X_1 + \cdots + X_n$ and $S_0 = 0$. If $\mathbb{E}\,(X_1|\mathcal{I}_X) = 0$, then $\mathbb{P}(T^1 < \infty) = 1$. Moreover, $\mathbb{P}\,(S_n = 0$ i.o.$) = 1$.*

Remark 2.6. In other words, mean zero implies recurrence. The condition

$$\mathbb{E}\,(X_1|\mathcal{I}_X) = 0$$

is needed to rule out trivial examples that have mean 0 but are a combination of a sequence with positive and negative means, e.g., $\mathbb{P}\,(X_n = 1 \text{ for all } n) = \mathbb{P}\,(X_n = -1 \text{ for all } n) = 1/2$.

Proof. Step 1. If $\mathbb{E}(X_1|\mathcal{I}) = 0$ then the ergodic theorem implies $S_n/n \to 0$ a.s. Thus

$$\lim_{n\to\infty} \frac{1}{n} \max_{1\le k\le n} |S_k| = 0 .$$

Note that

$$R_n \le \max_{1\le k\le n} S_k - \min_{1\le k\le n} S_k + 1 \le 2 \max_{1\le k\le n} |S_k| + 1 ,$$

we have $R_n/n \to 0$ and Lemma 2.9 implies $\mathbb{P}(T^1 = \infty) = 0$.

Step 2. For $n \ge 0$, define

$$\zeta_n := \inf\{m \ge 1 : S_{m+n} - S_n = 0\} , \; B_n := \{\zeta_n < \infty\} \text{ and } B = \bigcap_{n=0}^{\infty} B_n .$$

Observe that $\zeta_0 = T^1$, we have $\mathbb{P}(\zeta_0 < \infty) = 1$. By stationarity, $\mathbb{P}(\zeta_n < \infty) = 1$ for every $n \ge 0$. Hence $\mathbb{P}(B) = 1$.

Note that $T^n = T^{n-1} + \zeta_{T^{n-1}}$ for $n \ge 1$, where T^n is the time of the nth return of $(S_n)_{n\ge 0}$ to the origin. On B we have $T^n < \infty$ for every $n \ge 0$ and hence

$$\mathbb{P}(S_n = 0 \text{ i.o. }) = \mathbb{P}(T^n < \infty \text{ for all } n \ge 1) \ge \mathbb{P}(B) = 1 . \qquad \square$$

If in Theorem 2.10 the random variables X_n are not integer-valued, then there is no hope that $S_n = 0$ for any $n \in \mathbb{N}$ with positive probability. On the other hand, in this case, there is also some kind of recurrence property, namely $S_n/n \to 0$ almost surely by the ergodic theorem. Note, however, that this does not exclude the possibility that $S_n \to \infty$ with positive probability; for instance, if S_n grows like $\sqrt{n}$. The next theorem shows that if the $(X_n)_{n\ge 1}$ are integrable, then the process of partial sums can go to infinity only with a linear speed.

Theorem 2.11. *Let $(X_n)_{n\in\mathbb{N}}$ be an integrable stationary ergodic process and define $S_n = X_1 + \ldots + X_n$ for $n \in \mathbb{N}_0$. Then the following statements are equivalent.*

$$\textit{(i) } \mathbb{P}(S_n \to \infty) > 0 , \quad \textit{(ii) } S_n \to \infty \ \textit{a.s.}, \quad \textit{(iii) } \frac{S_n}{n} \to \mathbb{E}\, X_1 > 0 \ \textit{a.s.}.$$

Proof. Trivially, (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i).

(i) $\Rightarrow$ (ii). Note that $\{S_n \to \infty\}$ is an invariant event and thus has probability either 0 or 1.

(ii) $\Rightarrow$ (iii). The convergence follows by the individual ergodic theorem. Hence, it is enough to show that

$$\lim_{n\to\infty} \frac{S_n}{n} > 0 \quad \text{a.s.}.$$

Let $S^- := \inf\{S_n : n \in \mathbb{N}_0\}$. By assumption (ii), we have $S^- > -\infty$ almost surely and $\tau := \sup\{n \in \mathbb{N}_0 : S_n = S^-\}$ is finite almost surely. Hence there is an constant $N \in \mathbb{N}$ with $\mathbb{P}(\tau \le N) \ge 1/2$. For $n \ge 0$ and $\varepsilon \ge 0$, let

$$A_n^\varepsilon := \{S_m > S_n + \varepsilon \text{ for all } m > n\} \ ,$$

Since $(X_n)_{n\ge 1}$ is stationary, we have $\mathbb{P}(A_n^\varepsilon) = A_0^\varepsilon$. Observe that

$$\mathbb{P}(\tau < N) = \mathbb{P}\left(\bigcup_{n=0}^{N-1} A_n^0\right) \ge \frac{1}{2} \, .$$

Since $A_n^\varepsilon \uparrow A_n^0$ for fixed n as $\varepsilon \downarrow 0$, there is an $\epsilon > 0$ with

$$p := \mathbb{P}\left(A_0^\epsilon\right) \ge \frac{1}{4N} > 0.$$

By Proposition 2.3, as $(X_n)_{n\ge 1}$ is ergodic, $\left(1_{A_n^\epsilon}\right)_{n\ge 0}$ is also ergodic. By the individual ergodic theorem, we conclude that

$$\frac{1}{n}\sum_{i=0}^{n-1} 1_{A_n^\epsilon} \to p \quad \text{a.s.}.$$

Hence there exists an $n_0 = n_0(\omega)$ such that $\sum_{i=0}^{n-1} 1_{A_n^\epsilon} \ge \frac{pn}{2}$ for all $n \ge n_0$. This implies $S_n \ge \frac{pn\epsilon}{2}$ for $n \ge n_0$ and hence

$$\mathbb{E}X_1 = \lim_{n\to\infty} \frac{S_n}{n} \ge \frac{p}{2}\epsilon > 0 \, . \qquad \square$$

B Recurrence of Stationary Sequence

Extending the reasoning in the proof of Theorem 2.10 gives a result about recurrence of stationary sequence. Let $X_0, X_1, \ldots$ be a stationary sequence taking values in $(E, \mathcal{B}(E))$. Let $A \in \mathcal{B}(E)$, let $T^0 = 0$, and for $n \geq 1$, let $T^n = \inf\left\{m > T^{n-1} : X_m \in A\right\}$ be the time of the n th return to A. Let $\sigma^n = T^n - T^{n-1}$ length of the nth excursion to A.

However, we give a useful lemma first, which can be proved easily by Kolmogorov's consistency theorem.

Lemma 2.12. *Let $X = (X_n)_{n\geq 0}$ be a stationary sequence takeing values in $(E, \mathcal{B}(E))$. Then X can be extended to a stationary process $\tilde{X} = \left(\tilde{X}_n\right)_{n\in\mathbb{Z}}$.*

Theorem 2.13 (Recurrence, mean return time)**.** *If $\mathbb{P}\left(\sigma^1 < \infty \mid X_0 \in A\right) = 1$. Then*

$$(\sigma^n)_{n\geq 1} \quad \textit{conditional on} \quad \{X_0 \in A\}$$

is a stationary sequence. Particularly, $\mathbb{P}(X_n \in A \text{ i.o. } | X_0 \in A) = 1$. If, in addition, $\mathbb{P}(\sigma^1 < \infty) = 1$, then

$$\mathbb{E}\left(\sigma^1 | X_0 \in A\right) = \frac{1}{\mathbb{P}\left(X_0 \in A\right)}.$$

Remark 2.7. If $(X_n)_{n\geq 0}$ is an irreducible Markov chain on a countable state space S starting from its stationary distribution π, and $A = \{x\}$, then Theorem 2.13 says $\mathbb{E}_x T_x = 1/\pi(x)$. Theorem 2.13 extends that result to an arbitrary $A \subset S$ and drops the assumption that $(X_n)_{n\geq 0}$ is a Markov chain.

Proof. We first show that under $\mathbb{P}\left(\cdot | X_0 \in A\right)$, $(\sigma^n)_{n\geq 1}$ is stationary. It suffices to show that for any $n \geq 1$ and $m_1, \cdots m_n \in \mathbb{N}$,

$$\mathbb{P}\left(\sigma^1 = m_1, \cdots, \sigma^n = m_n | X_0 \in A\right) = \mathbb{P}\left(\sigma^2 = m_1, \cdots, \sigma^{n+1} = m_n | X_0 \in A\right)$$

Our first step is to extend $(X_n)_{n\geq 0}$ to a two-sided stationary sequence $(X_n)_{n\in\mathbb{Z}}$. Then observe that

$$\begin{aligned}
&\mathbb{P}\left(X_0 \in A, \sigma^2 = m_1, \cdots, \sigma^{n+1} = m_n\right) \\
&= \sum_{k=1}^{\infty} \mathbb{P}\left(X_0 \in A, \sigma^1 = k, \sigma^2 = m_1, \cdots, \sigma^{n+1} = m_n\right) \\
&= \sum_{k=1}^{\infty} \mathbb{P}\left(\hat{\sigma}_1 = -k, X_0 \in A, \sigma^1 = m_1, \cdots, \sigma^n = m_n\right) ,
\end{aligned}$$

where $\hat{\sigma}_1 \coloneqq \sup\{j < 0 : X_j \in A\}$. Note that for any $k \in \mathbb{N}$,

$$\begin{aligned}
\mathbb{P}(\hat{\sigma}_1 = -k, X_0 \in A) &= \mathbb{P}(X_{-k} \in A, X_{-k+1} \notin A, \cdots, X_{-1} \notin A, X_0 \in A) \\
&= \mathbb{P}(X_0 \in A, X_1 \notin A, \cdots X_{k-1} \notin A, X_k \in A) \\
&= \mathbb{P}(X_0 \in A, \sigma^1 = k) .
\end{aligned}$$

So $\mathbb{P}(\hat{\sigma}_1 > -\infty | X_0 \in A) = 1$ and hence

$$\begin{aligned}
&\mathbb{P}\left(X_0 \in A, \sigma^2 = m_1, \cdots, \sigma^{n+1} = m_n\right) \\
&= \mathbb{P}\left(X_0 \in A, \sigma^1 = m_1, \cdots, \sigma^n = m_n\right) .
\end{aligned}$$

Firstly, we can see that $\mathbb{P}(\sigma^n < \infty | X_0 \in A) = 1$ for all $n \geq 1$. Secondly, the desired stationarity property holds.

To complete the proof, we compute

$$\begin{aligned}
&\mathbb{E}\left(\sigma^1 | X_0 \in A\right) = \sum_{k=1}^{\infty} \mathbb{P}\left(\sigma^1 \geq k | X_0 \in A\right) \\
&= \frac{1}{\mathbb{P}(X_0 \in A)} \sum_{k=1}^{\infty} \mathbb{P}\left(\sigma^1 \geq k, X_0 \in A\right) \\
&= \frac{1}{\mathbb{P}(X_0 \in A)} \sum_{k=1}^{\infty} \mathbb{P}(X_0 \in A, X_1 \notin A, \cdots, X_{k-1} \notin A) \\
&= \frac{1}{\mathbb{P}(X_0 \in A)} \sum_{k=1}^{\infty} \mathbb{P}(X_{-k} \in A, X_{-k+1} \notin A, \cdots, X_{-1} \notin A) \\
&= \frac{1}{\mathbb{P}(X_0 \in A)} \sum_{k=1}^{\infty} \mathbb{P}(\hat{\sigma}_1 = -k) .
\end{aligned}$$

We show that if $\mathbb{P}(\sigma^1 < \infty) = 1$, then $\mathbb{P}(\hat{\sigma}_1 > -\infty) = 1$ and the desired result follows. To see this, note that

$$\begin{aligned}\mathbb{P}(\hat{\sigma}_1 < -k) &= \mathbb{P}(X_{-k} \notin A, \cdots, X_{-1} \notin A) \\ &= \mathbb{P}(X_1 \notin A, \cdots, X_k \notin A) = \mathbb{P}(\sigma^1 > k)\,.\end{aligned}$$

We complete the proof. □

Remark 2.8. In fact we have a stronger conclusion: if $\mathbb{P}\left(\sigma^1 < \infty\right) = 1$ and $B \subset E$ so that $A \cap B = \emptyset$, then

$$\mathbb{E}\left(\sum_{1 \le m \le T_1} 1_{\{X_m \in B\}} \mid X_0 \in A\right) = \frac{\mathbb{P}\left(X_0 \in B\right)}{\mathbb{P}\left(X_0 \in A\right)}\,.$$

When $A = \{x\}$ and X_n is a Markov chain, this is the "cycle trick" for defining a stationary measure.

2.4 A Subadditive Ergodic Theorem

In this section we will prove Liggett's version of Kingman's subadditive ergodic theorem.

Theorem 2.14 (Subadditive Ergodic Theorem)**.** *Suppose* $\{X_{m,n} : 0 \le m < n\}$ *satisfies:*

(i) $X_{0,n} \le X_{0,m} + X_{m,n}$ *for all* $0 < m < n$.

(ii) *For each* $k \in \mathbb{N}$, $\left(X_{nk,(n+1)k}\right)_{n \ge 1}$ *is a stationary sequence.*

(iii) *The distribution of* $(X_{m,m+k})_{k \ge 1}$ *does not depend on* m.

(iv) $\mathbb{E}X_{0,1}^+ < \infty$ *and for each* n, $\mathbb{E}X_{0,n} \ge Cn$, *where* C *is a constant.*

Then there exists $\Gamma \in L^1$ *so that*

$$\frac{X_{0,n}}{n} \to \Gamma \quad \textit{a.s. and in } L^1 .$$

Moreover,

$$\mathbb{E}X = \lim_{n\to\infty} \mathbb{E}\,\frac{X_{0,n}}{n} = \gamma := \inf_{n\geq 1} \mathbb{E}\frac{X_{0,n}}{n} .$$

If all the stationary sequences in (ii) are ergodic then $X = \gamma$ *a.s.*

Remark 2.9. Kingman assumed (iv), but instead of (i)-(iii) he assumed that $X_{\ell,n} \leq X_{\ell,m} + X_{m,n}$ for all $\ell < m < n$ and that the distribution of

$$\{X_{m+k,n+k}, 0 \leq m < n\}$$

does not depend on k. In two of the four applications in the next, these stronger conditions do not hold.

Before giving the proof, which is somewhat lengthy, we will consider several examples for motivation. The first example shows that Theorem 2.14 contains Birkhoff's ergodic theorem as a special case.

Example 2.15 (Stationary Sequence)**.** Suppose $\xi_1, \xi_2, \ldots$ is a integrable stationary sequence and let $X_{m,n} := \xi_{m+1} + \cdots + \xi_n$ for all $0 \leq m < n$. Then by Theorem 2.14, we get the individual ergodic theorem.

Example 2.16 (Range of Random Walk)**.** Suppose $\xi_1, \xi_2, \ldots$ is a stationary sequence and set $S_n = \xi_1 + \cdots + \xi_n$. Let $X_{m,n} = |\{S_{m+1}, \ldots, S_n\}|$ for all $0 \leq m < n$. Applying Theorem 2.14 now gives $X_{0,n}/n \to \Gamma$ a.s. and in L^1, but it does not tell us what the limit is.

Example 2.17 (Longest Common Subsequences)**.** Given two ergodic stationary sequences $(\xi_n)_{n\geq 1}$ and $(\eta_n)_{n\geq 1}$. For $0 \leq m < n$, let $L_{m,n} = \max\{K : \xi_{i_k} = \eta_{j_k}$ for $1 \leq k \leq K,$ where $m < i_1 < \ldots < i_K \leq n$ and $m < j_1 < \ldots < j_K \leq n\}$. It is clear that

$$L_{0,m} + L_{m,n} \leq L_{0,n} .$$

So $X_{m,n} = -L_{m,n}$ is subadditive. Applying Theorem 2.14 to $\{X_{m,n} : 0 \le m < n\}$ now, we conclude that

$$\frac{L_{0,n}}{n} \to \gamma = \sup_{m \ge 1} \mathbb{E} \frac{L_{0,m}}{m} .$$

The examples above should provide enough motivation for now. In the next subsection, we will give four more applications of Theorem 2.14.

Proof of Theorem 2.14. Let

$$\overline{X} = \limsup_{n\to\infty} \frac{X_{0,n}}{n} , \ \underline{X} = \liminf_{n\to\infty} \frac{X_{0,n}}{n} .$$

The proof will be broken up into four steps:

1. We check that $\mathbb{E}|X_{0,n}| \le C'n$ for some constant $C' > 0$. Let $\gamma_n = \mathbb{E}X_{0,n}$, then
$$\lim_{n\to\infty} \frac{\gamma_n}{n} = \gamma := \inf_{n\ge 1} \frac{\gamma_n}{n}$$
2. $\mathbb{E}\overline{X} \le \gamma$, and if for each $k \in \mathbb{N}$, $\left(X_{nk,(n+1)k}\right)_{n\ge 1}$ is are stationary ergodic sequence, then $\overline{X} \le \gamma$ a.s.
3. $\mathbb{E}\underline{X} \ge \gamma$. This is the “difficult half” according to Kingman.
4. $X_{0,n}/n$ converges to $\overline{X} = \Gamma$ a.s. and in L^1.

Step 1. The first thing to check is that $\mathbb{E}|X_{0,n}| \le Cn$. To do this, we note that (i) implies

$$X^+_{0,n} \le X^+_{0,m} + X^+_{m,n} .$$

Repeatedly using the last inequality and invoking (iii) gives $\mathbb{E}X^+_{0,n} \le n\mathbb{E}X^+_{0,1}$. Since $|x| = 2x^+ - x$, it follows from (iv) that

$$\mathbb{E}|X_{0,n}| \le 2\mathbb{E}X^+_{0,n} - \mathbb{E}X_{0,n} \le C'n < \infty .$$

Note that (i) and (iii) imply that

$$\gamma_{m+n} \leq \gamma_m + \gamma_n \,, \text{ for all } m, n \geq 1 \,.$$

Define γ by

$$\gamma = \inf_{n \geq 1} \frac{\gamma_n}{n} \,,$$

which is finite by (iv). Fix an $m \geq 1$ and write $n = km + l$, where $0 \leq l < m$. Thus we have

$$\gamma_n \leq k\gamma_m + \gamma_l \,.$$

As $n \to \infty, n/k \to m$, so that

$$\limsup_{n \to \infty} \frac{\gamma_n}{n} \leq \frac{\gamma_m}{m} \,.$$

since m is arbitrary, we conclude that

$$\gamma \leq \liminf_{n \to \infty} \frac{1}{n} \gamma_n \leq \limsup_{n \to \infty} \frac{1}{n} \gamma_n \leq \gamma \,.$$

Step 2. Fix an $m \geq 1$. Dividing by $n = km + \ell \, (0 \leq \ell < m)$, making repeated use of (i), we get

$$X_{0,n} \leq \sum_{j=1}^{k} X_{(j-1)m,jm} + X_{km,n} \,.$$

Thus

$$\frac{X_{0,n}}{n} \leq \frac{k}{n} \cdot \frac{1}{k} \sum_{j=1}^{k} X_{(j-1)m,jm} + \frac{X_{km,n}}{n} \,. \tag{2.2}$$

On the one hand, using (ii) and the ergodic theorem now gives that, as $n \to \infty$,

$$\frac{1}{k} \sum_{j=1}^{k} X_{(j-1)m,jm} \to \Gamma_m \quad \text{ a.s. and in } L^1 \,,$$

with $\mathbb{E}\,\Gamma_m = \gamma_m$. If in addition $\left(X_{nk,(n+1)k}\right)_{n \geq 1}$ is ergodic, then we can choose $\Gamma_m = \gamma_m$. On the other hand, we shall show that

$$\frac{X_{km,n}}{n} \to 0 \quad \text{a.s.}.$$

It suffices to show that for any given $\epsilon > 0$,

$$\mathbb{P}(|X_{km,n}| > n\epsilon \text{ i.o.}) = 0\,.$$

We compute

$$\begin{aligned}\sum_{n=1}^{\infty} \mathbb{P}(|X_{km,n}| > n\epsilon) &= \sum_{n=1}^{\infty} \mathbb{P}(|X_{0,\ell}| > n\epsilon) \\ &\le \sum_{n=1}^{\infty} \mathbb{P}\big(\max_{1\le \ell < m} |X_{0,\ell}| > n\epsilon\big) < \infty\,,\end{aligned}$$

since $\mathbb{E} \max_{1\le\ell<m} |X_{0,\ell}| < \infty$. Letting n to infty in (2.2), we get

$$\overline{X} \le \frac{\Gamma_m}{m} \quad \text{and hence } \mathbb{E}\overline{X} \le \frac{\gamma_m}{m}\,.$$

Taking the infimum over m, we have $\mathbb{E}\overline{X} \le \gamma$. Note that if all the stationary sequences in (ii) are ergodic, we have $\overline{X} \le \gamma$.

Remark 2.10. If (i)-(iii) hold, $\mathbb{E}X_{0,1}^+ < \infty$, and $\inf \mathbb{E}X_{0,m}/m = -\infty$, then it follows from the last argument that as $X_{0,n}/n \to -\infty$ a.s..

Step 3. For each $m \ge 1$, let

$$\underline{X}_m := \liminf_{n\to\infty} \frac{X_{m,m+n}}{n}\,.$$

(i) implies

$$X_{0,m+n} \le X_{0,m} + X_{m,m+n}$$

Dividing both sides by n and letting $n \to \infty$ gives $\underline{X} \le \underline{X}_m$ a.s. However, (iii) implies that $\underline{X}_m$ and $\underline{X}$ have the same distribution so

$$\underline{X} = \underline{X}_m \quad \text{a.s..}$$

Let $\epsilon > 0$ and let $Z = \epsilon + (\underline{X} \vee -M)$. Clearly, $\mathbb{E}|Z| < \infty$. Let

$$Y_{m,n} = X_{m,n} - (n-m)Z \quad \text{ for all } 0 \le m < n\,.$$

Then $\{Y_{m,n}\}$ satisfies (i)$-$(iv), since $Z_{m,n} = -(n-m)Z$ does, and has

$$\underline{Y} \equiv \liminf_{n\to\infty} \frac{Y_{0,n}}{n} = \underline{X} - Z \leq -\epsilon\,. \tag{2.3}$$

Let $T_m = \inf\{n \geq 1 : Y_{m,m+n} \leq 0\}$ [3], (iii) implies $T_m \overset{law}{=} T_0$ and

$$\mathbb{E}\left(Y_{m,m+1}1_{\{T_m>N\}}\right) = \mathbb{E}\left(Y_{0,1}1_{\{T_0>N\}}\right)\,.$$

(2.3) implies that $\mathbb{P}(T_0 < \infty) = 1$, thus we can pick N large enough so that

$$\mathbb{E}\left(Y_{0,1}1_{\{T_0>N\}}\right) \leq \epsilon\,.$$

We truncate T_m by N: Let

$$S_m = T_m 1_{\{T_m\leq N\}} + 1_{\{T_m>N\}}\,.$$

Note that $Y(m, m+T_m) \leq 0$, and on $\{T_m > N\}$, $Y(m, m+S_m) = Y_{m,m+1} > 0$, we have

$$Y(m, m+S_m) \leq Y_{m,m+1}1_{\{T_m>N\}}\,.$$

Let $R_0 = 0$, and for $k \geq 1$, let $R_k = R_{k-1} + S(R_{k-1})$. For given n, let $K = \max\{k : R_k \leq n\}$. From (i), it follows that

$$Y(0,n) \leq Y(R_0, R_1) + \cdots + Y(R_{K-1}, R_K) + Y(R_K, n)$$

since $n - R_K \leq S(R_K) \leq N$, we have

$$Y_{0,n} \leq \sum_{m=0}^{n-1} Y_{m,m+1}1_{\{T_m>N\}} + \sum_{j=1}^{N} |Y_{n-j,n-j+1}|\,,$$

here we have used (i) on $Y(R_K, n)$. Dividing both sides by n, taking expected values, and letting $n \to \infty$ gives

$$\limsup_{n\to\infty} \frac{\mathbb{E}Y_{0,n}}{n} \leq \mathbb{E}\left(Y_{0,1}1_{T_0>N}\right) \leq \epsilon\,.$$

[3] This is not a stopping time but there is nothing special about stopping times for a stationary sequence!

It follows from (a) and the definition of $Y_{0,n}$ that

$$\gamma = \lim_{n\to\infty} \frac{\mathbb{E}X_{0,n}}{n} \le 2\epsilon + \mathbb{E}(\underline{X} \vee -M)\,.$$

Since $\epsilon \ge 0$ and M are arbitrary, it follows that $\mathbb{E}\underline{X} \ge \gamma$.

Step 4. It only remains to prove convergence in L^1. Let

$$\Gamma = \inf_{m\ge 1} \frac{\Gamma_m}{m}\,.$$

By step 2 and Step 1, we have $\overline{X} \le \Gamma$ and $\mathbb{E}\Gamma \le \gamma$. By Step 3,

$$\gamma \le \mathbb{E}\underline{X} \le \mathbb{E}\overline{X} \le \mathbb{E}\Gamma \le \gamma\,.$$

Thus we have $X_{0,n}/n \to \Gamma$ almost surely. To show the convergence in L^1, observing that $|z| = 2z^+ - z$, we can write

$$\begin{aligned}\mathbb{E}\left|\frac{X_{0,n}}{n} - \Gamma\right| &= 2\mathbb{E}\left(\frac{X_{0,n}}{n} - \Gamma\right)^+ - \mathbb{E}\left(\frac{X_{0,n}}{n} - \Gamma\right)\\ &\le 2\mathbb{E}\left(X_{0,n}/n - \Gamma\right)^+ \ ,\end{aligned}$$

since $\mathbb{E}\left(X_{0,n}/n\right) = \mathbb{E}\left(\Gamma_n/n\right) \ge \mathbb{E}\Gamma$. Using the trivial inequality $(x+y)^+ \le x^+ + y^+$ and noticing $\Gamma_m \ge \Gamma$ now gives

$$\mathbb{E}\left(\frac{X_{0,n}}{n} - \Gamma\right)^+ \le \mathbb{E}\left(\frac{X_{0,n}}{n} - \frac{\Gamma_m}{m}\right)^+ + \mathbb{E}\left(\frac{\Gamma_m}{m} - \Gamma\right)$$

As we can see the second term is small if m is large. To bound the other term, observe that (i) implies

$$\mathbb{E}\left(\frac{X_{0,n}}{n} - \frac{\Gamma_m}{m}\right)^+ \le \mathbb{E}\left(\frac{1}{n}\sum_{j=1}^{k} X_{(j-1)m,jm} - \frac{\Gamma_m}{m}\right)^+ + \mathbb{E}\left(\frac{X(km,n)}{n}\right)^+ \ ,$$

where $n = km + \ell$ with $0 \le \ell < m$. The second term tends to 0 as $n \to \infty$. For the first, the ergodic theorem implies

$$\mathbb{E}\left|\frac{1}{n}\sum_{j=1}^{k} X_{(j-1)m,jm} - \frac{\Gamma_m}{m}\right| \to 0\,,$$

so the proof of Theorem 2.14 is complete. □

A Applications

In this subsection, we will give three applications of our subadditive ergodic theorem. These examples are independent of each other and can be read in any order. In Example 2.19 we encounter situations to which Liggett's version applies but Kingman's version does not.

Example 2.18 (First Passage Percolation)**.** Consider $\mathbb{Z}^d$ as a graph with edges connecting each $x, y \in \mathbb{Z}^d$ with $|x-y|=1$. Assign an independent nonnegative random variable $\tau(e)$ to each edge that represents the time required to traverse the edge going in either direction. If e is the edge connecting x and y, let $\tau(x,y)=\tau(y,x)=\tau(e)$. If $x=x_0, x_1, \cdots, x_n=y$ is a path from x to y, i.e., a sequence with $|x_m-x_{m-1}|=1$ for $1 \leq m \leq n$ we define the *travel time* for the path to be $\tau\left(x_0, x_1\right)+\cdots+\tau\left(x_{n-1}, x_n\right)$ Define the *passage time* from x to y, $T(x,y)=$ the infimum of the travel times over all paths from x to y. Let $X_{m,n}=T(mu, nu)$ where $u=(1,0,\ldots,0) \in \mathbb{Z}^d$.

Clearly $X_{0,n} \leq X_{0,m}+X_{m,n}$ and $X_{0,n} \geq 0$, so if $\mathbb{E}\tau(x,y)<\infty$ then (iv) holds, and Theorem 2.14 implies that $X_{0,n}/n \to \Gamma$ a.s. To see that the limit is constant, enumerate the edges in some order $e_1, e_2, \ldots$ and observe that Γ is measurable with respect to the tail σ-field of the i.i.d. sequence $\left(\tau\left(e_i\right)\right)$.

Example 2.19 (Age-Dependent Branching Processes)**.** This is a variation of the branching process introduced in which each individual lives for an amount of time with distribution F before producing k off spring with probability p_k. The description of the process is completed by supposing that the process starts with one individual in generation 0 who is born at time 0, and when this particle dies, its offspring start independent copies of the original process.

Suppose $p_0=0$, let $X_{0,m}$ be the birth time of the first member of generation m, and let $X_{m,n}$ be the time lag necessary for that individual to have an offspring in generation n. In case of ties, pick an individual at random from those in generation m born at time $X_{0,m}$. It is clear that $X_{0,n} \leq X_{0,m}+X_{m,n}$. since

$X_{0,n} \geq 0$, (iv) holds if we assume F has finite mean. Applying 2.14 now, it follows that

$$\frac{X_{0,n}}{n} \to \gamma \quad \text{a.s..}$$

The limit is constant because the sequences $\{X_{nk,(n+1)k}, n \geq 0\}$ are i.i.d. hence ergodic. As usual, one has to use other methods to identify the constant.

Remark 2.11. The inequality $X_{\ell,n} \leq X_{\ell,m} + X_{m,n}$ is false when $\ell > 0$, because if we call i_m the individual that determines the value of $X_{m,n}$ for $n > m$ then i_m may not be a descendant of i_ℓ.

Chapter 3

Brownian Motion

Brownian movement is the name given to the irregular movement of pollen, suspended in water, observed by the botanist Robert Brown in 1828. This random movement, now attributed to the buffeting of the pollen by water molecules, results in a dispersal or diffusion of the pollen in the water. The range of application of Brownian motion as defined here goes far beyond a study of microscopic particles in suspension and includes modeling of stock prices, of thermal noise in electrical circuits, of certain limiting behavior in queueing and inventory systems, and of random perturbations in a variety of other physical, biological, economic, and management systems.

Why we study it? There are many answers to this question, but to us there seem to be four main ones:

- Virtually every interesting class of processes contains Brownian motion : Brownian motion is a *continuous martingale*, a *Markov process*, a *diffusion*, a *Gaussian process*, a *Levy process*,...;
- Brownian motion is sufficiently concrete that one can do *explicit calculations*, which are impossible for more general objects;

- Brownian motion can be used as a *building block* for other processes (indeed, a number of the most important results on Brownian motion state that the most general process in a certain class can be obtained from Brownian motion by some sequence of transformations);
- last but not least, Brownian motion is a rich and beautiful mathematical object in its own right.

3.1 Fundamentals

Brownian motion is a process of tremendous practical and theoretical significance. The first thing is to define Brownian motion.

Definition 3.1. A real-valued process $\{B_t\}_{t\geq 0}$ defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called a **standard, one-dimensional Brownian motion** if the following holds.

(i) Almost surely, $B_0 = 0$.

(ii) The process has *independent increments*, that is, for any given times $0 \leq t_1 < t_2 < \cdots < t_n < \infty$ the increments

$$B\left(t_n\right) - B\left(t_{n-1}\right), B\left(t_{n-1}\right) - B\left(t_{n-2}\right), \cdots, B\left(t_2\right) - B\left(t_1\right)$$

are independent random variables.

(iii) The process has *stationary increments*, that is, for all $t \geq 0$ and $h > 0$, the increments $B(t+h) - B(t)$ are normally distributed with expectation 0 and variance h.

(iv) For every $\omega \in \Omega$, the sample path $t \mapsto B(t, \omega)$ is continuous.

If we change condition (i) by (i'): $\mathbb{P}(B \in A) = \mu(A)$ for all $A \in \mathcal{B}(\mathbb{R})$, where μ is a probability measure on $\mathbb{R}$, then $\{B_t\}_{t\geq 0}$ is called a Brownian motion with initial distribution μ. Therefore, there may exists a family of probability measures $\{\mathbb{P}_\mu : \mu \text{ is a p.m. on } (\mathbb{R}, \mathcal{B}(\mathbb{R}))\}$ on sample space $(\Omega, \mathcal{F})$ and B is a Brownian motion with initial distribution μ under $\mathbb{P}_\mu$. For $x \in \mathbb{R}$ and $\mu = \delta_x$, we say that B is a Brownian motion starting at x, and denote the corresponding probability measure by $\mathbb{P}_x$. Trivially, under $\mathbb{P}_0$, $\{B_t+x\}_{t\geq 0}$ is a Brownian motion starting at x.

We will address the nontrivial question of the existence of a Brownian motion later, for the moment let us step back and look at some technical points.

Remark 3.1. We can make the definition of Brownian motion in another way. A real-valued process $B = \{B_t\}_{t\geq 0}$ adapted to the filtration $\mathfrak{F} = \{\mathcal{F}_t\}_{t\geq 0}$ is called a standard Brownian motion with respect to $\mathfrak{F}$, if $B_0 = 0$ a.s.; for $0 \leq s < t$, the increment $B_t - B_s$ is independent of $\mathcal{F}_s$ and is normally distributed with mean 0 and variance $t - s$; and every sample path of B is continuous. Thus the Brownian motion $\{B_t\}_{t\geq 0}$ defined in Definition 3.1 without using filtration, is indeed a Brownian motion with respect to $\mathfrak{F}^B$, the filtration generated by B.

Moreover, if B is a $\mathfrak{F}$-standard Brownian motion, since B is adapted to $\mathfrak{F}$, we deduce that $\mathfrak{F}$ must be a "larger" filtration than $\mathfrak{F}^B$: for all $t \geq 0$, $\mathcal{F}_t^B \subset \mathcal{F}_t$. Then clearly B is a $\mathfrak{F}^B$-standard Brownian motion. It is often interesting, and necessary, to work with a filtration $\mathfrak{F}$ which is larger than $\mathfrak{F}^B$. For instance, we shall see in [3] Example 5.3.5 that the stochastic differential equation (5.3.1) does not have a solution, unless we take the driving process W to be a Brownian motion with respect to a filtration which is strictly larger than $\left\{\mathcal{F}_t^W\right\}$.

Remark 3.2. For two continuous-time stochastic process X and Y, we can think of them as equivalent if they are *indistinguishable*. We therefore appropriately relax the definition of Brownian motion. A process $B = \{B_t\}_{t\geq 0}$ satisfying (i), (ii), (iii) with *almost surely* continuous sample paths is also called a standard Brownian motion.

The "canonical" space for Brownian motion, the one most convenient for many future developments, is $C[0, \infty)$ the space of all continuous, real-valued functions on $[0,\infty)$ with metric

$$d(\phi_1,\phi_2) := \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{p_n(\phi_1-\phi_2)}{1+p_n(\phi_1-\phi_2)}, \qquad \text{where}$$

$$p_n(\phi_1-\phi_2) := \max_{t\in[0,n]} |\phi_1(t)-\phi_2(t)| \qquad \text{for all } n.$$

Define π_t the *evaluation mapping* (*coordinate mapping*) $\phi \mapsto \phi(t)$ for all $\phi \in C[0,\infty)$. Then one can show that (see Lemma 4.1, 4.2)

(i) $C[0,\infty)$ equipped tith the topology induced by d defined above, is a complete, separable metric space. In fact, the topology is a locally convex vector topology incuded by a countable family of seminorms.

(ii) The Borel σ-algebra $\mathcal{B}(C[0,\infty))$ coinsides with the σ-algebra generated by the *evaluation* (*coordinate*) mappings $\{\pi_t\}_{t\geq 0}$, i.e.,

$$\sigma(\pi_t : t \geq 0) = \mathcal{B}(C[0,\infty)).$$

So, if $X = \{X_t\}_{t\geq 0}$ is a *continuous* stochastic process on $(\Omega,\mathcal{F},\mathbb{P})$. We can regard X as a random variable on $(\Omega,\mathcal{F},\mathbb{P})$ with values in $(C[0,\infty),\mathcal{B}(C[0,\infty))$, and the *law* of X, is determined by its finite-dimensional distributions.

Later, we show how to construct a measure $\mathbb{P}$, called *Wiener measure*, on $(C[0,\infty),\mathcal{B}(C[0,\infty)))$, so that the coordinate mapping process is Brownian motion. That is, for any $n \geq 1$ and any real numbers $0 = t_0 < t_1 < \cdots < t_n$,

$$\begin{aligned}&\mathbb{P}(\pi_{t_k} \in A_j, 0 \leq k \leq n)\\ &= \int_{A_0} \delta_0(\mathrm{d}x_0) \int_{A_1} p(t_1,0,x_1)\,\mathrm{d}x_1 \cdots \int_{A_n} p(t_n-t_{n-1},x_{n-1},x_n)\,\mathrm{d}x_n\end{aligned}$$

where δ_0 is the point measure with suppoer $\{0\}$ and

$$p(t,x,y) := \frac{1}{\sqrt{2\pi t}} \exp\left\{-\frac{|x-y|^2}{2t}\right\} \qquad \text{for } t>0, x,y \in \mathbb{R}.$$

In the following example, we will see that, if we are interested in the sample path properties of a stochastic process, we may need to specify more than just its finite-dimensional distributions.

Example 3.1. Suppose that $\{B_t\}_{t\geq 0}$ is a standard Brownian motion and U is uniformly distributed on $[0,1]$ and independent with $\{B_t\}_{t\geq 0}$. Then the process $\{\tilde{B}_t : t \geq 0\}$ defined by

$$\tilde{B}_t = \begin{cases} B_t & \text{if } t \neq U \\ 0 & \text{if } t = U \end{cases} , \quad t \geq 0$$

is a modification of B, but is discontinuous if $B(U) \neq 0$, i.e. with probability one, and hence this process is not a Brownian motion.

A Brownian motions as Gaussian processes

Firstly, recall that a random vector $X = (X_1, \cdots, X_n)$ is called a *Gaussian random vector*, if there exists an $n \times m$ matrix A, and vector $b \in \mathbb{R}^n$ such that $X^T = AY + b$, where Y is an m-dimensional vector with independent standard normal entries. As we konw, a Gaussian random vector X has independent entries if and only if its covariance matrix is diagonal. In other words, the entries in a Gaussian vector are uncorrelated if and only if they are independent.

In complete generality, a (real- valued) process $(X_t)_{t\in T}$ indexed by some set T is said to be a *Gaussian process* if, for any given $t_1, \cdots, t_n$, the vector

$$\big(X(t_1), \cdots, X(t_n)\big)$$

is a Gaussian random vector. Thus the law of the process X is specified by the functions

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

Suppose now $\{B_t\}_{t\geq 0}$ is a standard Brownian motion and $0 = t_0 < t_1 < t_2 < \cdots < t_n$, then we can write $B(t_1), \cdots, B(t_n)$ as linear combinations of the independent standard normal random variables

$$\frac{B(t_j) - B(t_{j-1})}{\sqrt{t_j - t_{j-1}}}, \quad j = 1, \ldots, n\,.$$

Hence $\{B_t\}_{t\geq 0}$ is a Gaussian process with mean zero. For $s < t$,

$$\begin{aligned}\mathbb{E}\left[B_s B_t\right] &= \mathbb{E}B_s^2 + \mathbb{E}\left[B_s\left(B_t - B_s\right)\right] \\ &= s + \mathbb{E}\left[B_s\right]\mathbb{E}\left[B_t - B_s\right] = s\,,\end{aligned}$$

which gives the general rule

$$\operatorname{Cov}\left(B_s, B_t\right) = s \wedge t\,.$$

As we have pointed, one cannot tell from the finite-dimensional distributions alone whether or not the paths are continuous. Thus, We have the following equivalent definition of standard Brownian motion.

Theorem 3.1. *$\{B_t\}_{t\geq 0}$ is a standard Brownian motion if and only if*

(i) *$\{B_t\}$ is a Gaussian process.*

(ii) $\mathbb{E}B_t = 0$ *and* $\operatorname{Cov}(B_s, B_t) = s \wedge t$ *for all t, s.*

(iii) *For every $\omega \in \Omega$, the function $t \mapsto B(t, \omega)$ is continuous.*

This simple fact turns out to be an extremely efficient means of checking when a process is a Brownian motion, and the following four simple but extremely important examples serve to illustrate this. We will give some transformations on the space of functions that changes the individual Brownian random functions without changing the distribution.

Theorem 3.2 (Invariance)**.** *$\{B_t\}_{t\geq 0}$ is a standard Brownian motion, then we have the following invaiant properties.*

(i) *(Symmetry) $\{-B_t\}_{t\geq 0}$ is a standard Brownian motion.*

(ii) *(Increments reversal) For any given $t > 0$, $\{B(t-s) - B(t)\}_{0\leq s\leq t}$ is a standard Brownian motion.*

(iii) (Scaling invariance) For any given $a > 0$, the process $\{W_t\}_{t\geq 0}$ defined by

$$W(t) = \frac{1}{a} B\left(a^2 t\right)$$

is a standard Brownian motion.

(iv) (Time inversion) The process $\{X_t\}_{t\geq 0}$ defined by

$$X(t) = \begin{cases} 0 & \text{for } t = 0 \\ tB\left(\frac{1}{t}\right) & \text{for } t > 0 \end{cases}$$

is a standard Brownian motion.

Proof. We only show (iv). Obviously, $\{X_t\}_{t\geq 0}$ is also a Gaussian process and the Gaussian random vectors $(X(t_1), \ldots, X(t_n))$ have expectation zero. The covariances, for $t > 0, h \geq 0$, are given by

$$\operatorname{Cov}(X(t+h), X(t)) = (t+h)\, t \ \operatorname{Cov}[B\big(\frac{1}{t+h}\big), B\big(\frac{1}{t}\big)] = t\,.$$

Hence the law of all the finite-dimensional distributions

$$(X(t_1), X(t_2), \cdots, X(t_n)), \quad \text{for } 0 \leq t_1 \leq \cdots \leq t_n\,,$$

are the same as for Brownian motion. The paths of $t \mapsto X(t)$ are clearly continuous for all $t > 0$ and in $t = 0$ we use the following two facts: First, as the set $\mathbb{Q}$ of rationals is countable, the distribution of $\{X_t : t > 0, t \in \mathbb{Q}\}$ is the same as for a Brownian motion, and hence

$$\lim_{\substack{t\downarrow 0 \\ t\in\mathbb{Q}}} X_t = 0 \text{ a.s..}$$

Second, $\mathbb{Q} \cap (0, \infty)$ is dense in $(0, \infty)$ and $\{X_t : t \geq 0\}$ is almost surely continuous on $(0, \infty)$, so that

$$\lim_{t\downarrow 0} X_t = 0 \ \text{ a.s..}$$

So $\{X_t\}$ has almost surely continuous paths. By Theorem 3.1 , it's a standard Brownian motion (in the indistinguishable sence). $\square$

Remark 3.3. Scaling invariance implies that the self-similarity of the sample paths of Brownian motions, and we roughly say that Brownian motions are in some sense random fractals .

Remark 3.4. The symmetry inherent in the time inversion property becomes more apparent if one considers the Ornstein-Uhlenbeck diffusion $\{X_t\}_{t\in\mathbb{R}}$, which is given by

$$X(t) = e^{-t}B\left(e^{2t}\right) \text{ for all } t \in \mathbb{R}$$

This is a Markov process (this will be explained properly later), such that $X(t)$ is standard normally distributed for all t. It is a *diffusion* with a drift towards the origin proportional to the distance from the origin. Unlike Brownian motion, the Ornstein-Uhlenbeck diffusion is time reversible: The time inversion formula gives that $\{X(t)\}_{t\geq 0}$ and $\{X(-t)\}_{t\geq 0}$ have the same law. For t near $-\infty$, $X(t)$ relates to the Brownian motion near time 0, and for t near ∞, $X(t)$ relates to the Brownian motion near ∞.

Example 3.2. As an example for the use of scaling invariance, let $a < 0 < b$, and look at

$$\tau_{(a,b)} = \inf\left\{t \geq 0 : B_t \in \{a, b\}\right\},$$

the first exit time of a standard Brownian motion from the interval $[a, b]$. Then, with $W(t) = \frac{1}{b}B\left(b^2 t\right)$ we have

$$\mathbb{E}_0\, \tau_{(a,b)} = b^2\ \mathbb{E}_0\ \inf\{t \geq 0 : W_t = a/b \text{ or } W_t = 1\} = b^2\ \mathbb{E}_0\ \tau_{(\frac{a}{b},1)}$$

which implies that $\mathbb{E}\,\tau_{(-b,b)}$ is a constant multiple of b^2. Also

$$\mathbb{P}_0\left(B_t \text{ exits } [a, b] \text{ at } a\right) = \mathbb{P}_0\left(W_t \text{ exits } [\frac{a}{b}, 1] \text{ at } 1\right)$$

is only a function of the ratio a/b. The scaling invariance property will be used extensively in all the following sections, and we shall often use the phrase that a fact holds ‘by Brownian scaling’ to indicate this.

Example 3.3. We have

$$\mathbb{P}_0\left(\sup_{t\geq 0} B_t = +\infty,\ \inf_{t\geq 0} B_t = -\infty\right) = 1\,.$$

To see this, let $Z := \sup_t B_t$. By Brownian scaling, for any $c > 0$, we have

$$cZ \overset{law}{=} Z\,,$$

so the law of Z is concentrated on $\{0, +\infty\}$. Then it suffices to show that $\mathbb{P}_0(Z = 0)$. Note that

$$\mathbb{P}_0(Z = 0) \leq \mathbb{P}_0\left(B(1) \leq 0\,,\ \sup_{t\geq 0} B(t+1) - B(1) = 0\right) = \frac{1}{2}\mathbb{P}_0(Z = 0)\,.$$

Thus $\mathbb{P}_0(Z = 0) = 0$, and the desired result follows.

Time inversion is a useful tool to relate the properties of Brownian motion in a neighbourhood of time $t = 0$ to properties at infinity. To illustrate the use of time inversion we exploit Theorem 3.1 (iv) to get an interesting statement about the long-term behaviour from an easy statement at the origin.

Example 3.4 (Law of large numbers)**.** Almost surely,

$$\lim_{t\to\infty} \frac{B_t}{t} = 0\,.$$

To see this, let $\{X_t\}_{t\geq 0}$ be the time inversion of $\{B_t\}$. Thus, it's easy to see that

$$\lim_{t\to\infty} \frac{B_t}{t} = \lim_{t\to\infty} X\big(\frac{1}{t}\big) = 0 \quad \text{a.s.}.$$

B Brownian motions as Lévy processes

Let $\{X_t\}_{t\geq 0}$ be a continuous-time processes. Recall that we say it has *independent increments* if, for each $n \geq 1$ and $0 \leq t_1 < t_2 < \cdots < t_n < \infty$, the random variables

$$X(t_2) - X(t_1) \cdots, X(t_n) - X(t_{n-1})$$

are independent; it has *stationary increments* iff, for any t, $s \geq 0$,

$$X(t+s) - X(s) \overset{\text{law}}{=} X(t) - X(0)\,.$$

We give the following definition: $\{X_t\}_{t\geq 0}$ is called a *Lévy process* if, (i) $X_0 = 0$ a.s., (ii) $\{X_t\}_{t\geq 0}$ has independent, stationary increments, and (iii) $\{X_t\}_{t\geq 0}$ is *continuous in probability* : for all $t \geq 0$, $X(t+h) \xrightarrow{\mathbb{P}} X(t)$ as $h \to 0\,(t+h \geq 0)$, i.e., for any $\epsilon > 0$,

$$\lim_{h\to 0, t+h\geq 0} \mathbb{P}\left(|X(t+h) - X(t)| > \epsilon\right) = 0\,.$$

Remark 3.5. Note that in the presence of (i) and (ii), (iii) is equivalent to the condition $\lim_{h\downarrow 0} \mathbb{P}(|X_h| > \epsilon) = 0$.

A Lévy process may thus be viewed as the continuous-time analog of a random walk. We have already discussed an example of a Lévy process with right-continuous paths, the Poisson process. As we can see, Brownian motion $\{B_t\}$ with drift μ and variance σ^2 starting at origin, is a Lévy process with continuous paths : (i),(ii) are clear, and $B(t+h) \to B(t)$ a.s., since the sample paths continuous almost surely. However, surprisingly, the converse is ture.

Theorem 3.3. *Suppose $\{X_t\}_{t\geq 0}$ is a Lévy process with continuous paths. Then there exists μ and σ^2 such that $\{X_t\}_{t\geq 0}$ is a Brownian motion with drift μ and variance σ^2, starting in* 0.

Proof. All we need to show is that $X(1)$ has a normal distribution. Let

$$Y_{n,j} = X\left(\frac{j}{n}\right) - X\left(\frac{j-1}{n}\right), \text{ for } j = 1, 2, \cdots, n\,.$$

then

$$X_1 = Y_{n,1} + \cdots + Y_{n,n}\,.$$

We shall use Lindeberg-Feller's theorem to show the desired result.

Step 1. We claim that all the moments of X_t are finite. To see this, let

$$M = \max_{0\le t\le 1} |X_t| \ ,$$

since $t \mapsto X_t$ is continuous, we have $M < \infty$ a.s. , so there exists some positive integer k such that

$$\mathbb{P}(M \ge k) \le \frac{1}{2} \, .$$

Then using continuity of the paths, by stopping at the first time t that $|X_t| = nk$, we can see that

$$\mathbb{P}(M \ge (n+1)k \mid M \ge nk) \le \frac{1}{2} \, .$$

and hence

$$\mathbb{P}(M \ge nk) \le \frac{1}{2^n} \, .$$

Thus all the moments of M are finite. Of course all the moments of X_t are finite for a given t.

Step 2. Let $\mathbb{E}X_1 = \mu$, and $\mathrm{Var}\,(X_1) = \sigma^2$. Since the increments are independent and stationary,

$$\mathbb{E}Y_{n,j} = \frac{\mu}{n} \, , \ \mathrm{Var}\,(Y_{n,j}) = \frac{\sigma^2}{n} \, , \ \text{for all } j = 1, \cdots, n$$

Let $\xi_{n,j} = Y_{n,j} - \mathbb{E}\,Y_{n,j}$, then $\mathbb{E}\xi_{n,j} = 0$ and

$$\sum_{j=1}^{n} \mathbb{E}\,\xi_{n,j}^2 = \sigma^2$$

It suffices to check the Lindeberg's condition : for any $\epsilon > 0$,

$$\sum_{j=1}^{n} \mathbb{E}\,\xi_{n,j}^2 1_{\{|\xi_{n,j}|>\epsilon\}} \to 0 \, .$$

Define $N_n = \sum_{j=1}^{n} 1_{\{|\xi_{n,j}|>\epsilon\}}$, note that $|\xi_{n,j}| \le |Y_{n,j}| + |\mathbb{E}Y_{n,j}|$ we have

$$\sum_{j=1}^{n} \xi_{n,j}^2 1_{\{|\xi_{n,j}|>\epsilon\}} \le \big(2M + \frac{|\mu|}{n}\big)^2 N_n$$

By C-B-S inequality, Lindeberg's condition follows from

$$\mathbb{E}N_n^2 \to 0\,. \tag{3.1}$$

Step 3. To show (3.1), put

$$Z_n = \max_{1\le j\le n} |Y_{n,j}|\,.$$

Continuity of the paths implies that $Z_n \to 0$ a.s., and hence for every $\epsilon > 0$,

$$\mathbb{P}\left(Z_n \le \epsilon\right) \to 1\,.$$

Since the increments are independent and stationary, we have

$$\mathbb{P}\left(Z_n \le \epsilon\right) = \left[1 - \mathbb{P}\left(|Y_{n,1}| > \epsilon\right)\right]^n \le \mathrm{e}^{-n\mathbb{P}(|Y_{n,1}|>\epsilon)}$$

Therefore, for every $\epsilon > 0$, $n\,\mathbb{P}\left(|Y_{n,1}| > \epsilon\right) \to 0$, which deduce that

$$n\,\mathbb{P}\left(|\xi_{n,1}| > \epsilon\right) \to 0\,.$$

So $\mathbb{E}N_n = n\mathbb{P}\left(|\xi_{n,1}| \ge \epsilon\right) \to 0$,

$$\mathrm{Var}(N_n) = \sum_{j=1}^n \mathrm{Var}(1_{\{|\xi_{n,j}|>\epsilon\}}) \le \sum_{j=1}^n \mathbb{E}1_{\{|\xi_{n,j}|>\epsilon\}} = \mathbb{E}N_n \to 0,$$

and then (3.1) follows. □

3.2 Existence of Brownian motion

It is a substantial issue whether the conditions imposed on the finite-dimensional distributions in the definition of Brownian motion allow the process to have continuous sample paths, or whether there is a contradiction. Now we show that there is no contradiction and, fortunately, Brownian motion exists.

Theorem 3.4 (Wiener's theorem)**.** *Standard Brownian motion exists.*

Sketch : we construct Brownian motion as a uniform limit of continuous functions, to ensure that it automatically has continuous paths. We first construct Brownian motion on the interval $[0,1]$ as a random element on the space $\mathrm{C}[0,1]$ of continuous functions on $[0,1]$. The idea is to construct the right joint distribution of Brownian motion step by step on the finite sets

$$\mathcal{D}_n = \left\{ \frac{k}{2^n} : 0 \le k \le 2^n \right\}$$

of dyadic points. We then interpolate the values on $\mathcal{D}_n$ linearly and check that the uniform limit of these continuous functions exists and is a Brownian motion.

Proof. **Step 1**. Let $\mathcal{D} = \bigcup_{n=0}^{\infty} \mathcal{D}_n$ and let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which a collection $\{Z_t : t \in \mathcal{D}\}$ of independent, standard normally distributed r.v.'s can be defined. Let $B(0) := 0$ and $B(1) := Z_1$. For each $n \in \mathbb{N}$ we define the random variables $B(d), d \in \mathcal{D}_n$ such that

(i) for all $r < s < t$ in $\mathcal{D}_n$ the random variable $B(t) - B(s)$ is normally distributed with mean zero and variance $t - s$, and is independent of $B(s) - B(r)$

(ii) the vectors $(B(d) : d \in \mathcal{D}_n)$ and $(Z_t : t \in \mathcal{D} \backslash \mathcal{D}_n)$ are independent.

Note that we have already done this for $\mathcal{D}_0 = \{0, 1\}$. Proceeding inductively we may assume that we have succeeded in doing it for some $n-1$. We then define $B(d)$ for $d \in \mathcal{D}_n \backslash \mathcal{D}_{n-1}$ by

$$B(d) = \frac{1}{2}\left[B\big(d - \frac{1}{2^n}\big) + B\big(d + \frac{1}{2^n}\big) \right] + \frac{Z_d}{2^{\frac{n+1}{2}}}$$

Note that the first summand is the linear interpolation of the values of B at the neighbouring points of d in $\mathcal{D}_{n-1}$. Therefore $B(d)$ is independent of $(Z_t : t \in \mathcal{D} \backslash \mathcal{D}_n)$ and the second property is fulfilled.

Moreover, as

$$\frac{1}{2}\left[B\big(d - \frac{1}{2^n}\big) + B\big(d + \frac{1}{2^n}\big) \right]$$

depends only on $(Z_t : t \in \mathcal{D}_{n-1})$, it is independent of Z_d. By our induction assumptions both terms are normally distributed. Then their sum, and their difference

$$B(d) - B\big(d - \frac{1}{2^n}\big)\,,\ B\big(d + \frac{1}{2^n}\big) - B(d)$$

are independent and normally distributed with mean zero and variance $\frac{1}{2^n}$.

Indeed, all increments

$$B(d) - B\big(d - \frac{1}{2^n}\big)\,,\ d \in \mathcal{D}_n \backslash \{0\},$$

are independent. To see this, it suffices to show that they are pairwise independent, as the vector of these increments is Gaussian. We have seen in the previous paragraph that pairs

$$B(d) - B\big(d - \frac{1}{2^n}\big)\,,\ B\big(d + \frac{1}{2^n}\big) - B(d)$$

with $d \in \mathcal{D}_n \backslash \mathcal{D}_{n-1}$ are independent. The other possibility is that the increments are over intervals separated by some $d \in \mathcal{D}_{n-1}$. Choose $d \in \mathcal{D}_j$ with this property and minimal j, so that the two intervals are contained in $\left[d - 2^{-j}, d\right]$, respectively $\left[d, d + 2^{-j}\right]$. By induction the increments over these two intervals of length 2^{-j} are independent, and the increments over the intervals of length 2^{-n} are constructed from the independent increments $B(d) - B\left(d - 2^{-j}\right)$, respectively $B\left(d + 2^{-j}\right) - B(d)$ using a disjoint set of variables $(Z_t : t \in \mathcal{D}_n)$. Hence they are independent and this implies the first property, and completes the induction step.

Step 2. Having thus chosen the values of the process on all dyadic points, we interpolate between them. Formally, for each n denote by $(B_n(t))_{t \geq 0}$ the continuous process obtained by linear interpolation from $\{B(d) : d \in D_n\}$.

$$B_n(t) = \begin{cases} B(d) \text{ for } t = d \in \mathcal{D}_n \\ \text{linear in between} \end{cases}$$

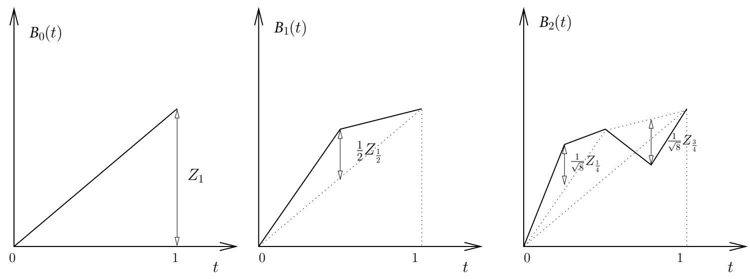


Figure 3.1: The first three steps in the construction of Brownian motion

Then, set $F_n(t) = B_n(t) - B_{n-1}(t)$, so

$$F_n(t) = \begin{cases} \dfrac{Z_d}{2^{(n+1)/2}}\,, \text{ for } t = d \in \mathcal{D}_n \backslash \mathcal{D}_{n-1} \\ 0\,, \text{ for } t = d \in \mathcal{D}_{n-1} \\ \text{linear between consecutive points in } \mathcal{D}_n \end{cases} . \tag{3.2}$$

Clearly,

$$\|F_n\|_\infty = \sup\left\{ \frac{|Z_d|}{2^{(n+1)/2}} : d \in \mathcal{D}_n \backslash \mathcal{D}_{n-1} \right\} .$$

So for large n,

$$\mathbb{P}\left(\|F_n| \geq \frac{c\sqrt{n}}{2^{(n+1)/2}} \right) \leq 2^n \mathbb{P}\left(|Z_1| \geq c\sqrt{n}\right) \leq 2^n \exp\left(\frac{-c^2 n}{2}\right) .$$

Take $c > \sqrt{2\log 2}$, by the Borel-Cantelli lemma there exists a random (but almost surely finite) N such that

$$\|F_n\|_\infty < c\sqrt{n}2^{-\frac{n}{2}}\,, \quad \text{for all } n \geq N\,.$$

This upper bound implies that, almost surely, the sequence $\{B_n\}$ is uniformly convergent on $[0,1]$. We denote the continuous limit by

$$\{B(t) : t \in [0,1]\}\,.$$

Step 3. We check that the increments of this process have the right finite-dimensional distributions, namely *Brownian distribution.* This follows directly from the properties of B on the dense set $\mathcal{D} \subset [0,1]$ and the continuity of the paths. Indeed, suppose that $t_1 < t_2 < \cdots < t_n$ are in $[0,1]$. We find $\{t_{i,k}\}$ in $\mathcal{D}$ so that $t_{i,k} \uparrow t_k$ as $i \to \infty$, and infer from the continuity of B that, for $1 \leq i \leq n-1$,

$$B(t_{i+1}) - B(t_i) = \lim_{k \uparrow \infty} B(t_{i+1,k}) - B(t_{i,k}) .$$

Using the characteristic function of Gsussian r.v., we have that the increments $B(t_{i+1}) - B(t_i)$ are independent Gaussian r.v's with mean 0 and variance $t_{i+1} - t_i$. We have thus constructed a continuous process $B : [0,1] \to \mathbb{R}$ with the same finite-dimensional distributions as Brownian motion. Take a sequence $B_0, B_1, \cdots$ of independent C$[0,1]$-valued random variables with the distribution of this process, and define $\{B(t)\}_{t\geq 0}$ by gluing together the parts, more precisely by

$$B(t) = B_{[t]}(t - [t]) + \sum_{i=0}^{[t]-1} B_i(1), \text{ for all } t \geq 0 .$$

This defines a continuous random function $B : [0,\infty) \to \mathbb{R}$ and one can see easily from what we have shown so far that it is a standard Brownian motion. ☐

Remark 3.6. If Brownian motion is constructed as a family $\{B_t\}_{t\geq 0}$ of random variables on some probability space Ω, it is sometimes useful to know that the mapping :

$$(t,\omega) \mapsto B(t,\omega)$$

is measurable on the product space $[0,\infty) \times \Omega$.

We point that this can be achieved by Lévy's construction. In fact, the Brownian motion is defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which a collection $\{Z_t : t \in \mathcal{D}\}$ of independent, standard normally distributed random variables are defined. It is easy to see from the construction that, for any $n \in \mathbb{N}$, the functions B_n are jointly measurable as a function of $Z_d, d \in \mathcal{D}_n$ and $t \in [0,1]$. Therefore it is also jointly measurable as a function of $\omega \in \Omega$ and $t \in [0,1]$ and this carries over to $(\omega, t) \mapsto B(\omega, t)$ by summation and taking a limit.

3.3 Nondifferentiability of Brownian motion

In this section, we always suppose that $B = \{B_t\}_{t\geq 0}$ is a standard Brownian motion. One manifestation is that the paths of Brownian motion have no intervals of monotonicity.

Theorem 3.5. *Almost surely, Brownian motion $\{B_t\}_{t\geq 0}$ doesn't have a monotone interval.*

Proof. Given an interval $[a,b] \subset [0,\infty]$, if it is an interval of monotonicity, we pick numbers

$$a = a_0 < a_1 < \cdots < a_n = b$$

and divide $[a,b]$ into n sub-intervals $[a_i, a_{i+1}]$. Each increment

$$B(a_i) - B(a_{i-1}),\, i = 1 \cdots, n$$

has to have the same sign. As the increments are independent, this has probability

$$2 \cdot \frac{1}{2^n}$$

and taking $n \to \infty$ shows that the probability that $[a,b]$ is an interval of monotonicity must be zero.

Taking a countable union gives that, almost surely, there is no nondegenerate interval of monotonicity with rational endpoints, but each nondegenerate interval would have a nondegenerate rational sub-interval. $\square$

Proposition 3.6. *Almost surely,*

$$\limsup_{n\to\infty} \frac{B_n}{\sqrt{n}} = +\infty\,, \quad \textit{and} \quad \liminf_{n\to\infty} \frac{B_n}{\sqrt{n}} = -\infty\,.$$

Proof. We clearly have, by Fatou's lemma,

$$\mathbb{P}\left(B_n > C\sqrt{n} \text{ i.o.}\right) \geq \limsup_{n\to\infty} \mathbb{P}\left(B_n > C\sqrt{n}\right)$$

By the scaling property, the expression in the limsup equals $\mathbb{P}(B_1 > C)$, which is positive. Denote $X_n = B_n - B_{n-1}$ for all n then

$$\left\{\frac{B_n}{\sqrt{n}} > C \quad \text{i.o.} \right\} = \left\{\frac{1}{\sqrt{n}} \sum_{j=1}^{n} X_j > C \quad \text{i.o.} \right\}$$

is an tail event with respect to $\{X_n\}$. Hence the Kolmogorov 0-1 law gives that,

$$\mathbb{P}\left(B_n > C\sqrt{n} \quad \text{i.o.}\right) = 1\,.$$

Taking the intersection over all positive integers C gives the first part of the statement and the second part is proved analogously. $\square$

Remark 3.7. Let $\{B_t\}$ be (one-dimensional) Brownian motion and let

$$A = \cap_n \{B_t = 0 \text{ for some } t \geq n\}\,.$$

Then $\mathbb{P}_x(A) = 1$ for all $x \in \mathbb{R}$. In other words, one-dimensional Brownian motion is recurrent. For any starting point x, it will return to 0 "infinitely often," i.e., there is a sequence of times $t_n \uparrow \infty$ so that $B(t_n) = 0$. We have to be careful with the interpretation of the phrase in quotes since starting from 0, B_t will hit 0 infinitely many times by time $\epsilon > 0$. See Theorem 3.26.

For a function f, we define the upper and lower right derivatives

$$D^* f(t) = \limsup_{h\downarrow 0} \frac{f(t+h) - f(t)}{h}$$

and

$$D_* f(t) = \liminf_{h\downarrow 0} \frac{f(t+h) - f(t)}{h}\,.$$

We now show that for any fixed time t, almost surely, Brownian motion is not differentiable at t. For this we use Proposition 3.6 and the invariance under time inversion.

Proposition 3.7. *Fix $t \geq 0$, then, almost surely, Brownian motion is not differentiable at t. Moreover,*

$$D^* B(t) = +\infty \quad \textit{and} \quad D_* B(t) = -\infty\,.$$

Proof. Without loss of generality, we assume $t = 0$. Or we can consider $\{\widetilde{B}_s\}$, defined by $\widetilde{B}(s) = B(t+s) - B(t)$.

For a standard Brownian motion $\{B_t\}$, we construct a further Brownian motion $\{X_t\}$ by time inversion. Then

$$D^* X(0) \geq \limsup_{n\to\infty} \frac{X\left(\frac{1}{n}\right) - X(0)}{\frac{1}{n}} = \limsup_{n\to\infty} nX\left(\frac{1}{n}\right) = \limsup_{n\to\infty} B(n)\,.$$

By Proposition 3.6, the we find that $D^* X(0)$ is infinite. Similarly, one can see that $D_* X(0) = -\infty$, showing that X is not differentiable at 0. □

While the previous proof shows that every t is almost surely a point of nondifferentiability for the Brownian motion, this does not imply that almost surely every t is a point of non-differentiability for the Brownian motion!

Remark 3.8. The behaviour of Brownian motion at a fixed time $t > 0$ reflects the behaviour at typical times in the following sense. Suppose $\mathcal{X} = \{f \in \mathrm{C}[0,\infty) : f \text{ is differentable at } 0\}$. Then we have shown that for any $t \geq 0$,

$$\{B(t+s) - B(t) : s \geq 0\} \in \mathcal{X} \quad \text{a.s.} \tag{3.3}$$

Denote the (random) set $D = \{t : \{B(t+s) - B(t) : s \geq 0\} \notin \mathcal{X}\}$, then we shall show that almost surely, D has Lebesgue measure zero. To see this, using the joint measurability mentioned in the remark of Theorem 3.4 and Fubini's theorem,

$$\mathbb{E} \int_0^\infty 1_D \,\mathrm{d}t = \int_0^\infty \mathbb{P}\left(\{B(t+s) - B(t) : s \geq 0\} \notin \mathcal{X}\right) \mathrm{d}t = 0\,.$$

Thus $\lambda(D) = 0$ a.s., where λ is the Lebesgue measure.

Theorem 3.8 (Paley, Wiener and Zygmund)**.** *Almost surely, Brownian motion is nowhere differentiable. Moreover, almost surely,*

$$\text{either } D^*B(t) = +\infty \text{ or } D_*B(t) = -\infty \text{ or both, for each } t \geq 0\,.$$

Proof. Suppose that there is a $t_0 \in [0,1]$ such that $-\infty < D_*B(t_0) \leq D^*B(t_0) < \infty$, then

$$\limsup_{h \downarrow 0} \frac{|B(t_0+h) - B(t_0)|}{h} < \infty\,.$$

Using the continuity of paths, there exists some M so that

$$\sup_{h\in[0,1]} \frac{|B(t_0+h) - B(t_0)|}{h} \leq M\,.$$

Thus,

$$\begin{aligned} &\{\exists t_0 \in [0,1], -\infty < D_*B(t_0) \leq D^*B(t_0) < \infty\} \\ &\subset \bigcup_{M=1}^{\infty} \left\{\exists t_0 \in [0,1] \text{ s.t. } \sup_{h\in[0,1]} \frac{|B(t_0+h) - B(t_0)|}{h} \leq M\right\} \end{aligned}$$

It suffices to show that, for any fixed M, we have

$$\mathbb{P}\left(\exists t_0 \in [0,1] \text{ s.t. } \sup_{h\in[0,1]} \frac{|B(t_0+h) - B(t_0)|}{h} \leq M\right) = 0\,.$$

If t_0 is contained in the binary interval $\left[\frac{k-1}{2^n}, \frac{k}{2^n}\right]$ for some $n \geq 2$ and $k \geq 1$, then for all $1 \leq j \leq 2^n - 1$ the triangle inequality gives

$$\begin{aligned} &\left|B\left(\frac{k+j}{2^n}\right) - B\left(\frac{k+j-1}{2^n}\right)\right| \\ &\leq \left|B\left(\frac{k+j}{2^n}\right) - B(t_0)\right| + \left|B(t_0) - B\left(\frac{k+j-1}{2^n}\right)\right| \leq \frac{2j+1}{2^n}M\,. \end{aligned}$$

Define events

$$A_{n,k} \coloneqq \left\{ \left|B\left(\frac{k+j}{2^n}\right) - B\left(\frac{k+j-1}{2^n}\right)\right| \leq \frac{2j+1}{2^n}M \ \text{ for } j = 1,2,3 \right\},$$

Then, clearly

$$\left\{\exists\ t_0 \in [0,1] \text{ s.t. } \sup_{h\in[0,1]} \frac{|B(t_0+h)-B(t_0)|}{h} \leq M\right\} \subset \bigcap_{n=2}^{\infty} \bigcup_{k=1}^{2^n} A_{n,k}\,.$$

By independence of the increments and the scaling property, for $1 \leq k \leq 2^n$,

$$\begin{aligned}\mathbb{P}(A_{n,k}) &\leq \prod_{j=1}^{3} \mathbb{P}\left(\left|B\left(\frac{k+j}{2^n}\right) - B\left(\frac{k+j-1}{2^n}\right)\right| \leq \frac{2j+1}{2^n}M\right)\\ &\leq \mathbb{P}\left(|B(1)| \leq \frac{7M}{\sqrt{2^n}}\right)^3 \leq \left(\frac{7M}{\sqrt{2^n}}\right)^3.\end{aligned}$$

Hence

$$\mathbb{P}\left(\bigcup_{k=1}^{2^n} A_{n,k}\right) \leq 2^n\left(\frac{7M}{\sqrt{2^n}}\right)^3 = \frac{(7M)^3}{\sqrt{2^n}}\,,$$

Thus

$$\mathbb{P}\left(\exists\ t_0 \in [0,1] \text{ s.t. } \sup_{h\in[0,1]} \frac{|B(t_0+h)-B(t_0)|}{h} \leq M\right) \leq \frac{(7M)^3}{\sqrt{2^n}}\,.$$

Letting $n \to \infty$, the desired result follows. □

Remark 3.9. The proof of Theorem 3.8 can be tightened to prove that, for any $\alpha > 1/2$, almost surely, the sample paths of Brownian motion are nowhere locally α-Hölder continuous.

Another important regularity property, which Brownian motion does NOT possess is to be of bounded variation. Recall that a function $f : [a,b] \to \mathbb{R}$ is a function of *bounded variation* if

$$V_f := \sup_{P\in\mathcal{P}} \sum_{j=1}^{k} |f(x_j) - f(x_{j-1})| < \infty$$

where the supremum is taken over the set $\mathcal{P} = \{P : a = x_0 < x_1 \cdots < x_n\}$ all the partitions of $[a,b]$. If the supremum is infinite f is said to be of *unbounded variation*. As we konw, f is of bounded variation if and only if it can be written as the difference of two increasing functions.

Theorem 3.9. *Suppose that the sequence of partitions*

$$0 = t_0^{(n)} \leq t_1^{(n)} \leq \cdots \leq t_{k(n)-1}^{(n)} \leq t_{k(n)}^{(n)} = t$$

is nested, i.e. at each step one or more partition points are added, and the mesh

$$\Delta(n) := \sup_{1 \leq j \leq k(n)} \left| t_j^{(n)} - t_{j-1}^{(n)} \right|$$

converges to zero. Then, almost surely,

$$\lim_{n\to\infty} \sum_{j=1}^{k(n)} \left| B\left(t_j^{(n)}\right) - B\left(t_{j-1}^{(n)}\right) \right|^2 = t \,,$$

and therefore Brownian motion is of unbounded variation.

Remark 3.10. For a sequence of partitions as above, we call

$$\lim_{n\to\infty} \sum_{j=1}^{k(n)} \left(B\left(t_j^{(n)}\right) - B\left(t_{j-1}^{(n)}\right) \right)^2$$

the quadratic variation of Brownian motion. The fact that Brownian motion has finite quadratic variation will be of crucial importance in Stochastic integrals and applications. However, the analogy to the notion of bounded variation of a function is not perfect : there exists a sequence of partitions

$$0 = t_0^{(n)} \leq t_1^{(n)} \leq \cdots \leq t_{k(n)-1}^{(n)} \leq t_{k(n)}^{(n)} = t$$

with mesh converging to zero, such that almost surely

$$\limsup_{n\to\infty} \sum_{j=1}^{k(n)} \left(B\left(t_j^{(n)}\right) - B\left(t_{j-1}^{(n)}\right) \right)^2 = \infty$$

In particular, the condition that the partitions in Theorem 1.35 are nested cannot be dropped entirely. See Exercise 1.15 and Exercise 1.16, [5].

Lemma 3.10. *If* Y, Z *are independent, symmetric random variables in* L^2, *then*

$$\mathbb{E}\left[(Y+Z)^2 | Y^2 + Z^2\right] = Y^2 + Z^2 \quad a.s..$$

Proof. By symmetry of Z we have

$$\mathbb{E}\left[(Y+Z)^2|Y^2+Z^2\right] = \mathbb{E}\left[(X-Z)^2|Y^2+Z^2\right]$$

Both sides of the equation are finite, so that we can take the difference and obtain

$$\mathbb{E}\left[YZ|X^2+Z^2\right] = 0$$

and the result follows immediately. □

Proof of Theorem 3.9. **Step 1**. For each $n \geq 1$, let

$$X_n := \sum_{j=1}^{k(n)} \left(B\left(t_j^{(n)}\right) - B\left(t_{j-1}^{(n)}\right)\right)^2 .$$

Clearly, for all n we have $\mathbb{E}X_n = t$ and

$$\mathbb{E}\left|X_n - t\right|^2 = \sum_{j=1}^{k(n)} \left|t_j^{(n)} - t_{j-1}^{(n)}\right|^2 \mathrm{Var}(B_1^2) \leq \mathrm{Var}(B_1^2)\Delta(n)\, t$$

Thus $X_n \to t$ in L^2. To see that $(X_n)_{n\geq 1}$ converges almost surely, we use the theory of martingales in discrete time. We denote by $\mathcal{G}_n$ the σ-algebra generated by the random variables $X_n, X_{n+1}, \cdots$. Then

$$\mathcal{G}_\infty := \bigcap_{k=1}^{\infty} \mathcal{G}_k \subset \cdots \subset \mathcal{G}_{n+1} \subset \mathcal{G}_n \subset \cdots \subset \mathcal{G}_1 .$$

We show that $(X_n)_{n\geq 1}$ is a backward martingale, i.e.,

$$X_{n+1} = \mathbb{E}\left(X_n|\mathcal{G}_{n+1}\right) \quad \text{a.s..} \tag{3.4}$$

Then, by Theorem 1.57, we get $X_n \to \mathbb{E}(X_1|\mathcal{G}_\infty)$ a.s. and in L^1. Thus, $\mathbb{E}(X_1|\mathcal{G}_\infty) = t$ almost surely and the desired result follows.

Step 2. By inserting elements in the sequence, if necessary, we may assume that at each step exactly one point is added to the partition. The proof of (3.4) is easy with the help of Lemma 3.10. Indeed, if $s \in \left(t_{j-1}^{(n)}, t_j^{(n)}\right)$ is the inserted point we apply it to the symmetric, independent random variables $B(s) - B\left(t_{j-1}^{(n)}\right), B\left(t_j^{(n)}\right) - B(s)$. Then using the independence of the increments of Brownian motion,

$$\begin{aligned}
&\mathbb{E}\left[\left(B\left(t_j^{(n)}\right) - B\left(t_{j-1}^{(n)}\right)\right)^2 |\mathcal{G}_{n+1}\right] \\
&= \mathbb{E}\left[\left(B\left(t_j^{(n)}\right) - B\left(t_{j-1}^{(n)}\right)\right)^2 \middle| \left(B\left(t_j^{(n)}\right) - B(s)\right)^2 + \left(B(s) - B\left(t_{j-1}^{(n)}\right)\right)^2\right] \\
&= \left(B\left(t_j^{(n)}\right) - B(s)\right)^2 + \left(B(s) - B\left(t_{j-1}^{(n)}\right)\right)^2 \quad \text{a.s..}
\end{aligned}$$

Hence

$$X_{n+1} = \mathbb{E}\left(X_n | \mathcal{G}_{n+1}\right) \quad \text{a.s..}$$

Step 3. By the Hölder continuous property, we can find, for any $\alpha \in (0, 1/2)$, a (random) $n \in \mathbb{N}$ such that $|B(a) - B(b)| \leq |a - b|^\alpha$ for all $a, b \in [0, t]$ with $|a - b| \leq \Delta(n)$. Hence

$$X_n \leq \sum_{j=1}^{k(n)} \Delta(n)^\alpha \left|B\left(t_j^{(n)}\right) - B\left(t_{j-1}^{(n)}\right)\right|$$

Therefore, once we show that the (X_n) converge almost surely to a positive random variable, it follows immediately that Brownian motion is almost surely of unbounded variation. □

3.4 Continuity properties of Brownian motion

The definition of Brownian motion already requires that the sample functions are continuous almost surely. This implies that on the interval $[0, 1]$ (or any other compact interval) the sample functions are uniformly continuous, i.e., there exists some (random) function φ with $\lim_{h \downarrow 0} \varphi(h) = 0$ called a *modulus of* *continuity* of the function $B : [0,1] \to \mathbb{R}$ such that

$$\limsup_{h \downarrow 0} \sup_{t \in [0,1-h]} \frac{|B(t+h) - B(t)|}{\varphi(h)} \leq 1\,.$$

Can we achieve such a bound with a deterministic function φ, i.e., is there a nonrandom modulus of continuity for the Brownian motion ? Surprisingly, the answer is yes, as the following theorem shows.

Theorem 3.11. *For standard Brownian motion $\{B(t)\}$, there exists a constant $C > 0$ such that, almost surely, for every sufficiently small $h > 0$,*

$$|B(t+h) - B(t)| \leq C\sqrt{h \log \frac{1}{h}} \quad \textit{for all} \quad t \in [0, 1-h]\,. \tag{3.5}$$

First proof. This follows quite elegantly from Lévy's construction of standard Brownian motion. Recall the notation introduced in (3.2) and that we have represented Brownian motion as a series

$$B(t) = \sum_{n=0}^{\infty} F_n(t)\,,$$

where each F_n is a piecewise linear function. We have shown that for any $c > \sqrt{2 \log 2}$ there exists a (random) $N \in \mathbb{N}$, such that,

$$\|F_n\|_\infty < c\sqrt{n} 2^{-\frac{n}{2}}\,, \quad \text{for all } n \geq N\,.$$

Now for each $t, t+h \in [0,1]$,

$$|B(t+h) - B(t)| \leq \sum_{n=0}^{\infty} |F_n(t+h) - F_n(t)|\,,$$

Suppose that h is (again random and) small enough, then the positive integer ℓ, defined by $h \in (\frac{1}{2^\ell}, \frac{1}{2^{\ell-1}}]$, exceeds N. Hence

$$\sum_{n=\ell+1}^{\infty} |F_n(t+h) - F_n(t)| \leq 2 \sum_{n=\ell+1}^{\infty} \|F_n\|_\infty \leq 2c \sum_{n=\ell+1}^{\infty} \sqrt{n} 2^{-\frac{n}{2}}$$

and there exists constants C_1, C_2 such that

$$\sum_{n=\ell+1}^{\infty} \sqrt{n} 2^{-\frac{n}{2}} \le C_1 \sqrt{\ell}\, 2^{-\frac{\ell}{2}} \le C_2 \sqrt{h \log \frac{1}{h}} \, .$$

Hence, using the mean-value theorem, we get for all $\ell > N$ that $|B(t+h) - B(t)|$ is bounded by

$$h \sum_{n=0}^{\ell} \|F_n'\|_\infty + C_2 \sqrt{h \log \frac{1}{h}}.$$

Note that

$$\|F_n'\|_\infty \le \frac{2\,\|F_n\|_\infty}{2^{-n}} \le 2c\sqrt{n} 2^{n/2} \qquad \text{for } n \ge N \, ,$$

Thus

$$h \sum_{n=0}^{\ell} \|F_n'\|_\infty \le h \sum_{n=0}^{N} \|F_n'\|_\infty + 2ch \sum_{n=N}^{\ell} \sqrt{n} 2^{n/2}$$

We now suppose that h is (again random and) small enough that the first summand is smaller than $\sqrt{h \log \frac{1}{h}}$, and note that there exists constants C_3, C_4 so that

$$2ch \sum_{n=N}^{\ell} \sqrt{n} 2^{n/2} \le C_3 h \sqrt{\ell}\, 2^{\ell/2} \le C_4 \sqrt{h \log \frac{1}{h}} \, ,$$

we get (3.5). □

This upper bound is pretty close to the optimal result. The following lower bound confirms that the only missing bit is the precise value of the constant.

Theorem 3.12. *For every constant $C < \sqrt{2}$, almost surely, for every $\epsilon > 0$ there exist $0 < h < \epsilon$ and $t \in [0, 1-h]$ with*

$$|B(t+h) - B(t)| \ge C \sqrt{h \log \frac{1}{h}} \, .$$

Proof. Let $C < \sqrt{2}$ and define, for integers $k, n \ge 0$, the events

$$A_{n,k} = \left\{ B\left(\frac{k+1}{e^n}\right) - B\left(\frac{k}{e^n}\right) > C\sqrt{n} e^{-n/2} \right\} \, .$$

Then for any $k \geq 0$

$$\mathbb{P}\left(A_{n,k}\right) = \mathbb{P}\left(B\left(\frac{1}{e^n}\right) > C\sqrt{n}e^{-n/2}\right) = \mathbb{P}\left(B(1) > C\sqrt{n}\right) \geq \frac{C\sqrt{n}}{C^2 n + 1}\frac{1}{\sqrt{2\pi}}e^{-C^2 n/2}$$

By our assumption on C, we have $e^n \mathbb{P}\left(A_{n,k}\right) \to \infty$ as $n \uparrow \infty$. Therefore, using $1 - x \leq e^{-x}$ for all x

$$\mathbb{P}\left(\bigcap_{k=0}^{[e^n - 1]} A_{n,k}^c\right) = \left(1 - \mathbb{P}\left(A_{n,0}\right)\right)^{e^n} \leq \exp\left(-e^n \mathbb{P}\left(A_{n,0}\right)\right) \to 0\,.$$

Thus

$$\mathbb{P}\left(\bigcup_{k=0}^{[e^n - 1]} A_{n,k} \quad \text{i.o.}\right) \geq \limsup_{n\to\infty} \mathbb{P}\left(\bigcup_{k=0}^{[e^n - 1]} A_{n,k}\right) = 1\,.$$

The desired result follows. □

Remark 3.11. One can determine the constant c in the best possible modulus of continuity $\varphi(h) = c\sqrt{h\log(1/h)}$ precisely. Indeed, our proof of the lower bound yields a value of $c = \sqrt{2}$ which turns out to be optimal. This striking result is due to Paul Lévy: Almost surely,

$$\limsup_{h\downarrow 0} \sup_{0\leq t\leq 1-h} \frac{|B(t+h) - B(t)|}{\sqrt{2h\log(1/h)}} = 1\,.$$

Let I be an interval of $\mathbb{R}$ and $f : I \to \mathbb{R}$. Recall that f is said to be *locally α-Hölder continuous*, if there exists $\delta > 0$ and $C > 0$ such that

$$|f(x) - f(y)| \leq C|x-y|^\alpha, \qquad \text{for all } x, y \in I \text{ with } |y-x| < \delta\,. \tag{3.6}$$

We say f is said to be *locally α-Hölder continuous* at $x \in I$, if there exists $\delta, C > 0$ such that (3.6) holds for any $y \in I$ with $|y - x| < \delta$. Note that locally α-Hölder continuous is uniformly locally α-Hölder continuous at every point. We refer to $\alpha > 0$ as the *Hölder exponent* and to $C > 0$ as the *Hölder constant*. Clearly, α-Holder continuity gets stronger, as the exponent α gets larger.

The results of this section so far indicate that, for Brownian motion, the transition between paths which are α-Hölder continuous and paths which are not happens at $\alpha = \frac{1}{2}$.

Corollary 3.13. *Let $\alpha < \frac{1}{2}$. Then, almost surely, Brownian motion is locally α-Hölder continuous.*

Remark 3.12. This result is optimal in the sense that, for $\alpha > 1/2$, almost surely, at every point, Brownian motion fails to be locally α-Hölder continuous. See Remark 3.9. Points where Brownian motion is locally $1/2$-Hölder continuous exist almost surely, but they are very 'rare'. See Chapter10, [5].

Second proof of Theorem 3.11* We will give another proof of Theorem 3.11, assuming that the distributuion of the maximum of Brownian motion is known.

Proof. For any $h > 0$, let

$$\mathrm{Osc}(h) := \{|B(t) - B(s)| : t, s \in [0,1], |t - s| \leq h\}.$$

It's easy to see that Theorem 3.5 hols if and only if there exists a constant $C > 0$ and almost surely,

$$\lim_{h \to 0} \frac{\mathrm{Osc}(h)}{\sqrt{h \log \frac{1}{h}}} \leq C. \tag{3.7}$$

Firstly, take a positive integer n and let

$$I_{n,m} = \left[\frac{m-1}{2^n}, \frac{m}{2^n}\right]$$

for $m = 1, 2, \cdots, 2^n$ and denote

$$\Delta_{n,m} = \sup_{t \in I_{n,m}} \left| B(t) - B(\frac{m-1}{2^n}) \right|.$$

By Brownian scaling, for a standard Brownian motion $\{W(t)\}$, $\Delta_{n,m}$ is indetically distributed to

$$\frac{M \vee \widetilde{M}}{\sqrt{2^n}}$$

where

$$M := \max_{0\le t\le 1} W(t) \text{ and } \widetilde{M} := \max_{0\le t\le 1} -W(t)\,.$$

Thus for given $x > 0$,

$$\mathbb{P}\left(\Delta_{n,m} > \frac{2x}{\sqrt{2^n}}\right) = \mathbb{P}\left(M \vee \widetilde{M} > 2x\right)$$

and note that

$$\{M \vee \widetilde{M} > 2x\} \subset \{M > x\} \cup \{\widetilde{M} > x\}\,,$$

we have

$$\begin{aligned}\mathbb{P}\left(\Delta_{n,m} > \frac{2x}{\sqrt{2^n}}\right) &\le 2\mathbb{P}\left(M > x\right) = 2\ \mathbb{P}\left(|B(1)| > x\right)\\ &= 4\int_x^\infty (2\pi)^{-\frac{1}{2}} \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t \le \exp\left(-\frac{x^2}{2}\right)\,.\end{aligned}$$

So

$$\mathbb{P}\left(\text{ exists } m \text{ s.t. } \Delta_{n,m} > \frac{2x}{\sqrt{2^n}}\right) \le 2^n \exp\left(-\frac{x^2}{2}\right) = \exp\left(n\log 2 - \frac{x^2}{2}\right)\,.$$

In order that

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\text{ exists } m \text{ s.t. } \Delta_{n,m} > \frac{2x_n}{\sqrt{2^n}}\right) < \infty\,,$$

we let $x_n = c\sqrt{n}$ where $c > \sqrt{2\log 2}$, by B-C lemma we have

$$\mathbb{P}\left(\text{ exists } m \text{ s.t. } \Delta_{n,m} > 2c\sqrt{\frac{n}{2^n}} \text{ i.o. }\right) = 0\,.$$

That is,

$$\mathbb{P}\left(\liminf_{n\to\infty}\left\{\sup_{1\le m\le 2^n} \Delta_{n,m} \le 2c\sqrt{\frac{n}{2^n}}\right\}\right) = 1\,. \tag{3.8}$$

One can show that (3.7) follows from .(3.8). $\square$

3.5 The Markov Property

Let d be a positive integer. In this section we define the notion of a d-dimensional Markov process and cite d-dimensional Brownian motion as an example. There are several equivalent statements of the Markov property, and we spend some time developing them.

A Brownian Motion in Several Dimensions

Definition 3.2. Let μ be a probability measure on $\mathbb{R}^d$. Let $B = \{B_t\}_{t\geq 0}$ be a $\mathbb{R}^d$-valued process, defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and adapted to the filtration $\mathfrak{F} = \{\mathcal{F}_t\}_{t\geq 0}$. B is called a d**-dimensional Brownian motion** with initial distribution μ, with respect to $\mathfrak{F}$, if

(i) $\mathbb{P}(B_0 \in A) = \mu(A)$, for all $A \in \mathcal{B}\left(\mathbb{R}^d\right)$;

(ii) for $0 \leq s < t$, the increment $B_t - B_s$ is independent of $\mathcal{F}_s$ and is normally distributed with mean zero and covariance matrix equal to $(t-s)I_d$, where I_d is the $d \times d$ identity matrix;

(iii) *every* sample path of B is continuous.

If the filtration $\mathfrak{F}$ is exactly $\mathfrak{F}^B$, the filtration generated by B, we say B is a Brownian motion with initial distribution μ for short. If μ assigns measure one to some singleton $\{x\}$, we say that B is a d-dimensional Brownian motion starting at x. Besides, a one-dimensional Brownian motion is called *linear*.

Here is one way to construct a d-dimensional Brownian motion with initial distribution μ. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space so that $X, B^{(1)}, \cdots, B^{(d)}$ defined on it, where X taking values in $\mathbb{R}^d$ has distribution μ, $B^{(i)}$ is standard brownian motion for $1 \leq i \leq d$, and $X, B^{(1)}, \cdots, B^{(d)}$ are independent. Then let

$$B_t = X + (B_t^{(1)}, \cdots, B_t^{(d)}) \quad \text{for } t \geq 0\,.$$

Then B is the desired object.

There is a second construction of d-dimensional Brownian motion with initial distribution μ, a construction which motivates the concept of Markov family, to be introduced in this section. Let $P^{(i)}, i = 1, \ldots, d$ be d copies of Wiener measure on $(C[0,\infty), \mathcal{B}(C[0,\infty)))$. Then $\mathbb{P}_0 := P^{(1)} \times \cdots \times P^{(d)}$ is a measure, called *d-dimensional Wiener measure*, on $\left(C[0,\infty)^d, \mathcal{B}\left(C[0,\infty)\right)^d\right)$. Under $\mathbb{P}_0$, the coordinate mapping process $B(t,\omega) := \omega(t)$ is a d-dimensional Brownian motion starting at the origin. For $x \in \mathbb{R}^d$, we define the probability measure $\mathbb{P}_x$ on $\mathcal{B}\left(C[0,\infty)^d\right)$ by

$$\mathbb{P}_x(A) = \mathbb{P}_0(A - x) \quad \text{for all} \quad A \in \mathcal{B}\left(C[0,\infty)\right)^d ,$$

where $A - x := \left\{\omega \in C[0,\infty)^d : \omega(\cdot) + x \in A\right\}$. Under $\mathbb{P}_x$, the coordinate mapping process $B = \{B_t\}_{t\geq 0}$ is a d-dimensional Brownian motion starting at x. Finally, for a probability measure μ on $\mathbb{R}^d$, we define $\mathbb{P}^\mu$ on $\mathcal{B}\left(C[0,\infty)^d\right)$ by

$$\mathbb{P}_\mu(A) = \int_{\mathbb{R}^d} \mathbb{P}_x(A)\mu(\mathrm{d}x), \quad \text{for all } \in \mathcal{B}\left(C[0,\infty)\right)^d .$$

Infact, one can show that the mapping $x \mapsto \mathbb{P}_x(A) ; \mathbb{R}^d \to [0,1]$ is measurable by using Dynkin's π-λ theorem. Then it's not hard to check that under $\mathbb{P}_\mu$, the coordinate mapping process $B = \{B_t\}_{t\geq 0}$ is a d-dimensional Brownian motion with initial distribution μ.

We giving a straight formulation of the facts for a Brownian motion.

Theorem 3.14 (Markov property)**.** *Let $\{B_t\}_{t\geq 0}$ be a d-dimensional Brownian motion with initial distribution μ. Then, for any $s > 0$, $\{B(t+s) - B(s)\}_{t\geq 0}$ is again a Brownian motion started in the origin and it is independent of the process $\{B_t : 0 \leq t \leq s\}$.*

Proof. It is easy to check that $\{B(t+s) - B(s)\}_{t\geq 0}$ satisfies the definition of a d-dimensional Brownian motion. The independence statement follows directly from the independence of the increments of a Brownian motion. $\square$

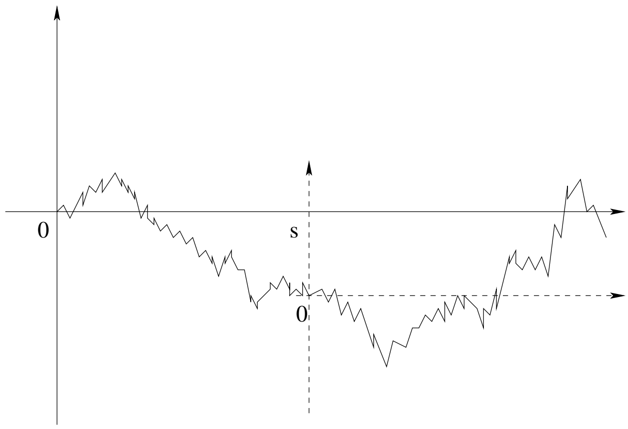


Figure 3.2: One-dimensional Brownian motion starts afresh at time s.

Definition 3.3. Given a Polish space E, we denote by $\overline{\mathcal{B}(E)}^{\mu}$ the completion of the Borel σ-field $\mathcal{B}(E)$ with respect to the probability measure μ on $(E, \mathcal{B}(E))$. The **universal σ-field** is $\mathcal{U}(E) := \cap_{\mu} \overline{\mathcal{B}(E)}^{\mu}$, where the intersection is over all probability measures μ. Besides, a real-valued $\mathcal{U}(E)/\mathcal{B}(\mathbb{R})$-measurable function is said to be **universally measurable**.

Proposition 3.15. *$\mathcal{U}(E)$ is the universal σ-field on E. Then*

(i) Let $A \subset E$. Then $A \in \mathcal{U}(E)$ iff for every p.m. μ on $(E, \mathcal{B}(E))$, there exists $B_{\mu} \in \mathcal{B}(E))$ so that $A = B_{\mu}$ μ-a.e..

(ii) Let $f : E \to \mathbb{R}$. Then f is universally measurable iff for each p.m. μ on $(E, \mathcal{B}(E))$, there is a Borel-measurable function g_{μ} so that $f = g_{\mu}$ μ-a.e..

Proof. Note that

$$\overline{\mathcal{B}(E)}^{\mu} = \{A \subset E : \exists\, B \in \mathcal{B}(E) \ \text{ s.t. } A = B \ \mu\text{-a.e.}\},$$

we get (i). (ii) is trivially follows from (i). $\square$

Definition 3.4. A **d-dimensional Brownian family** is a adapted $\mathbb{R}^d$-valued process $B = \{B_t\}_{t\geq 0}$ on a filtrated measurable space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0})$, and a family of probability measures $\{\mathbb{P}_x\}_{x\in\mathbb{R}^d}$ such that

(i) for each $A \in \mathcal{F}$, the mapping $x \mapsto \mathbb{P}_x(A)$ is universally measurable;

(ii) under each $\mathbb{P}_x$, B is a d-dimensional Brownian motion with respect to $\{\mathcal{F}_t\}_{t\geq 0}$ and starting at x.

We have already seen how to construct a family of probability measures $\{\mathbb{P}_x\}_{x\in\mathbb{R}^d}$ on the canonical space $\left(C[0,\infty)^d, \mathcal{B}\left(C[0,\infty)\right)^d\right)$ so that the coordinate mapping process, relative to the filtration it generates, is a Brownian motion starting at x under any $\mathbb{P}_x$. With $\mathcal{F} = \mathcal{B}\left(C[0,\infty)^d\right)$, indeed, for this canonical example of a d-dimensional Brownian family, the mapping $x \mapsto \mathbb{P}_x(A)$ is actually Borel-measurable for each $A \in \mathcal{F}$. The reason we formulate Definition 3.4 with the weaker measurability condition is to allow expansion of $\mathcal{F}$ to a larger σ-field (see 待补充).

B Markov peocesses and Markov families

Suppose now that $\{X_t\}_{t\geq 0}$ is a $\mathbb{R}^d$-valued stochastic process. Intuitively, the *Markov property* says that if we know the process $\{X_t\}_{t\geq 0}$ on the interval $[0,s]$, for the prediction of the future $\{X_{s+t}\}_{t\geq 0}$ this is as useful as just knowing the endpoint X_s. Moreover, a process is called a *time-homogeneous* Markov process if it starts *afresh* at any fixed time s. We shall make make a rigorous definition.

Definition 3.5. Let μ be a p.m. on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Let $X = \{X_t\}_{t\geq 0}$ be a $\mathbb{R}^d$-valued adapted process defined on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$. X is called a **Markov process** with respect to $\{\mathcal{F}_t\}_{t\geq 0}$, having initial distribution μ, if

(i) $\mathbb{P}(X_0 \in A) = \mu(A)$ for all $A \in \mathcal{B}(\mathbb{R}^d)$;

(ii) for any $t, s \geq 0$ and $A \in \mathcal{B}(\mathbb{R}^d)$, we have

$$\mathbb{P}(X_{s+t} \in A \mid \mathcal{F}_s) = \mathbb{P}(X_{s+t} \in A \mid X_s) \quad \text{a.s.}.$$

If a stochastic process X is called a Markov process without any reference to a filtration, then the minimal filtration of X is implied.

Clearly, the regular conditional distribution of X_{s+t} given X_s exists, and we denote it by $P_{s,t+s}$. That is, $P_{s,t+s}$ is a probability kernel from $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ to itself, statisfying

$$\mathbb{P}(X_{s+t} \in A | X_s) = P_{s,t+s}(X_s, A) \quad \text{a.s. for all } A \in \mathcal{B}(\mathbb{R}^d)\,.$$

If for any fixed $t \geq 0$, the transition kernels $P_{s,t+s}$ don't depend on s, in other words, $P_{s,t+s} = P_{0,t}$ for all $s \geq 0$, then we say the Markov process is **time-homogeneous**, and we denote P_t for $P_{s,t+s}$. In this note, we only discuss time-homogeneous Markov processes. Clearly, for any $x \in \mathbb{R}^d$, we have

$$P_0(x, A) = \delta_x(A) \;\text{ for every } A \in \mathcal{B}(\mathbb{R}^d)\,. \tag{3.9}$$

Moreover, We shoud emphsize that $\{P_t\}_{t\geq 0}$ satisies the *Chapman-Kolmogorov (C-K) equation*: For $t, s \geq 0$, $x \in \mathbb{R}^d$ and $A \in \mathcal{B}(\mathbb{R}^d)$,

$$P_{t+s}(x, A) = \int_{\mathbb{R}^d} P_t(y, A) P_s(x, \mathrm{d}y)\,. \tag{3.10}$$

We say a family of probability kernels $\{P_t\}_{t\geq 0}$ is a family of *Markov transition kernels* if it satisfies (3.9) and (3.10). Recall the course of discrete-time *Markov chains* can we recognise the pattern behind this definition: The Markov transition kernels $\{P_t\}_{t\geq 0}$ plays the role of the transition matrix P in this setup: it determinate the Markov process uniquely (up to the initial distribution).

Theorem 3.16. *Let μ be a Borel probability measure on $\mathbb{R}^d$. Let $\{P_t\}_{t\geq 0}$ be Markov transition kernels. Then X is a Markov process having Markov transition kernels $\{P_t\}_{t\geq 0}$ and initial distribution μ if and only if, for any $n \geq 1$, $A_j \in \mathcal{B}(\mathbb{R}^d), 0 \leq j \leq n$, and $0 = t_0 < t_1 < \cdots < t_n < \infty$ we have*

$$\begin{aligned} &\mathbb{P}(X(t_0) \in A_0, \cdots, X(t_n) \in A_n) \\ &= \int_{A_0} \mu(\mathrm{d}x_0) \int_{A_1} P_{t_1}(x_0, \mathrm{d}x_1) \cdots \int_{A_n} P_{t_n - t_{n-1}}(x_{n-1}, \mathrm{d}x_n)\,. \end{aligned} \tag{3.11}$$

Proof. The necessity is trivial, we only show the sufficiency, i.e.,

$$\mathbb{P}(X_{s+t} \in A \mid \mathcal{F}_s) = P_t(X_s, A) \quad \text{a.s..}$$

Let $C = \{X(t_0) \in A_0, \cdots, X(t_n) \in A_n\} \in \mathcal{F}_s^X$, where $A_0, \cdots, A_n \in \mathcal{B}(\mathbb{R}^d)$ and $0 = t_0 < t_1 < \cdots < t_n = s$. We firstly show that

$$\begin{aligned}
&\mathbb{E}\left[P_t(X_s, A)1_C\right] = \mathbb{P}(C, X_{s+t} \in A) \\
&= \int_{A_0} \mu(\mathrm{d}x_0) \int_{A_1} P_{t_1}(x_0, \mathrm{d}x_1) \cdots \int_{A_n} P_{t_n - t_{n-1}}(x_{n-1}, \mathrm{d}x_n) P_t(x_n, A) \,.
\end{aligned}$$

where the second equals sgin follows from (3.11). To see this, observe that for $D \in \mathcal{B}(\mathbb{R}^d)$,

$$\begin{aligned}
&\mathbb{E}\left[1_{\{X(s) \in D\}} 1_C\right] \\
&= \int_{A_0} \mu(\mathrm{d}x_0) \int_{A_1} P_{t_1}(x_0, \mathrm{d}x_1) \cdots \int_{A_n} P_{t_n - t_{n-1}}(x_{n-1}, \mathrm{d}x_n) 1_{\{x_n \in D\}} \,.
\end{aligned}$$

Linearity implies that for simple functions f,

$$\begin{aligned}
&\mathbb{E}\left[f(X(s))1_C\right] \\
&= \int_{A_0} \mu(\mathrm{d}x_0) \int_{A_1} P_{t_1}(x_0, \mathrm{d}x_1) \cdots \int_{A_n} P_{t_n - t_{n-1}}(x_{n-1}, \mathrm{d}x_n) f(x_n) \,,
\end{aligned}$$

and the bounded convergence theorem implies that it is valid for bounded measurable f, e.g., $f(x) = P_t(x, A)$. Hence

$$\mathbb{E}\left[P_t(X_s, A)1_C\right] = \mathbb{P}(C, X_{t+s} \in A) \,.$$

Using $\pi - \lambda$ theorem, we get the desired result. $\square$

Our experience with d-dimensional Brownian motion indicates that it is notationally and conceptually helpful to have a whole family of probability measures, rather than just one. Toward this end, we define the concept of a Markov family.

Definition 3.6. Let $\{P_t\}_{t\geq 0}$ be Markov transition kernels. A d-dimensional **Markov family** is an adapted process $X = \{X_t\}_{t\geq 0}$ defined on some filtrated measurable space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0})$, together with a family of probability measures $\{\mathbb{P}_x\}_{x\in\mathbb{R}^d}$ such that

(i) for each $A \in \mathcal{F}$, the mapping $x \mapsto \mathbb{P}_x(A)$ is universally measurable on $\mathbb{R}^d$;

(ii) under probability measure $\mathbb{P}_x$, X is a Markov process with respect to $\{\mathcal{F}_t\}_{t\geq 0}$, having transition kernels $\{P_t\}_{t\geq 0}$ and starting at x. That is, $\mathbb{P}_x(X_0 = x) = 1$ and for every $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\mathbb{P}_x\left(X_{s+t} \in A | \mathcal{F}_s\right) = P_t(X_s, A) \quad \mathbb{P}_x\text{-a.s..} \tag{3.12}$$

Firstly, let $s = 0$ in (3.12) we can see that $P_t(x, A) = \mathbb{P}_x(X_t \in A)$. This gives an intuitive explanation of Markov property. Secondly, for any Borel probability measure μ on $\mathbb{R}^d$. define $\mathbb{P}_\mu$ on $(\Omega, \mathcal{F})$ by letting

$$\mathbb{P}_\mu(A) = \int_{\mathbb{R}^d} \mathbb{P}_x(A)\mu(\mathrm{d}x) \qquad \text{for all } A \in \mathcal{F}\,,$$

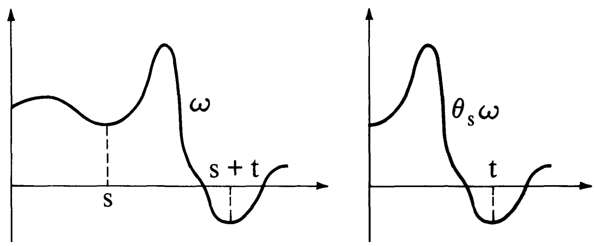
then X is a Markov process with respect to $\mathfrak{F}$, having transition kernels $\{P_t\}_{t\geq 0}$ and initial distribution μ.

It happens sometimes, for a given adapted process $X = \{X_t\}_{t\geq 0}$ on a measurable space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0})$, that one can construct a family of so-called *shift operators* $\theta_s : \Omega \to \Omega, s \geq 0$, such that each θ_s is measurable and

$$X_t\left(\theta_s\omega\right) = X_{t+s}(\omega) \quad \forall\, \omega \in \Omega, \quad s, t \geq 0\,.$$

The most obvious examples occur when Ω is either $\left(\mathbb{R}^d\right)^{[0,\infty)}$ or $C[0,\infty)^d$, $\mathcal{F}$ is the smallest σ-field containing all finite-dimensional cylinder sets, and X is the coordinate mapping process $X(t,\omega) = \omega(t)$. We can then define

$$\left(\theta_s\omega\right)(t) = \omega(s+t), \quad t \geq 0\,.$$



Theorem 3.17. *Let $X = \{X_t\}_{t\geq 0}$ be a Markov family having transition kernels $\{P_t\}_{t\geq 0}$ defined on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \{\mathbb{P}_x\}_{x\in\mathbb{R}^d})$. Let $\{\theta_s\}_{s\geq 0}$ be a family of shift-operators. Then, for any bounded (or non-negative) random variavle $Y \in \mathcal{F}_\infty^X$ and $x \in \mathbb{R}^d$,*

$$\mathbb{E}_x\left[Y \circ \theta_s \mid \mathcal{F}_s\right] = \mathbb{E}_{X_s} Y \quad \mathbb{P}_x\text{-a.s.},$$

where the right-hand side is the function $x \mapsto \mathbb{E}_x Y$ evaluated at X_s.

Proof. Firstly, we show that the theorem holds, if $Y = 1_{\{X\in C\}}$ where C is a finite-dimensional rectangular cylinder set in $(\mathbb{R}^d)^{[0,\infty)}$. To see this, let

$$C = \left\{\phi : [0, \infty) \to \mathbb{R}^d : \phi(t_0) \in A_0, \phi(t_1) \in A_1, \cdots, \phi(t_n) \in A_n\right\},$$

where $0 = t_0 < t_1 < \cdots < t_n < \infty$, $A_j \in \mathcal{B}(\mathbb{R}^d), 0 \leq j \leq n$ and $n \geq 1$. Then what we need to show is that

$$\begin{aligned}
&\mathbb{P}_x\left(X(s) \in A_0, X(s+t_1) \in A_1, \cdots, X(s+t_n) \in A_n | \mathcal{F}_s\right) \\
&= 1_{\{X(s)\in A_0\}} \int_{A_1} P_{t_1}(X(s), \mathrm{d}x_1) \cdots \int_{A_n} P_{t_n - t_{n-1}}(x_{n-1}, \mathrm{d}x_n),
\end{aligned}$$

which can be proved by induction on n and by Markov property.

Secondly, using $\pi - \lambda$ theorem, we get that the theorem holds if $Y = 1_{\{X\in A\}}$, where $A \in \mathcal{B}(\mathbb{R}^d)^{[0,\infty)}$. Finally, recalling that any bounded (or non-negative) measureable $Y \in \sigma(X)$, there exists a bounded (or non-negative) measurable funciton f on $\left((\mathbb{R}^d)^{[0,\infty)}, \mathcal{B}(\mathbb{R}^d)\right)$ so that $Y = f(X)$. So Y can be approximated by simple functions $\sum_i c_i 1_{\{X\in A_i\}}$, then we complete the proof. $\square$

C Brownian motion as a Markov process

The following lemma is useful when computing conditional probability or conditional expectation, which can be shown by the canonical method in measure theory and so we omit the proof.

Lemma 3.18. *If ξ and η are E-valued random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, $\mathcal{G}$ is a sub-σ-field of $\mathcal{F}$. Suppose ξ is independent of $\mathcal{G}$ and $\eta \in \mathcal{G}$, then*

$$\mathcal{L}[(\xi, \eta)|\mathcal{G}] = \mathcal{L}[(\xi, \eta)|\eta] \text{ and } \mathcal{L}[(\xi, \eta)|\eta = y] = \mathcal{L}[(\xi, y)] .$$

Suppose $\{B_t\}_{t\geq 0}$is a d-dimensional Brownian, we denote $\mathfrak{F}^B = \left\{\mathcal{F}_t^B\right\}_{t\geq 0}$ the filtration generated by B. By Theorem 3.14 and Lemma 3.18, or directly by Theorem 3.16, we deduce that :

Theorem 3.19. *$\{B_t\}_{t\geq 0}$ is a Markov process with respect to $\{\mathcal{F}_t^B\}_{t\geq 0}$, and the probability kernel $P_t(x, \cdot)$ is a Gaussian distribution with mean x and covariance matrix $t\,\mathrm{I}_d$.*

Note that for $t > 0$, $P_t(x, \cdot)$ has probability density, denoted by $p_t(x, y)$ pr $p(t, x, y)$, given by :

$$p_t(x, y) = (2\pi t)^{-\frac{d}{2}} \exp\left(-\frac{\|x - y\|^2}{2t}\right) \qquad \text{for } x, y \in \mathbb{R}^d . \tag{3.13}$$

so for $A \in \mathcal{B}(\mathbb{R})^d$ we have, $P_t(x, A) = \int_A p(t, x, y)\,\mathrm{d}y$. Besides, the C-K equation is just the fact that the sum of two independent Gaussian random vectors is a Gaussian random vector with the sum of the covariance matrices, can be rewritten as

$$p_{t+s}(x, y) = \int_{\mathbb{R}^d} p_s(x, \mathrm{d}z) p_t(z, y), \text{ for } z \in \mathbb{R}^d .$$

We give another example of the Markov process obtained from Brownian motion.

Example 3.5. The one-dimensional *refected Brownian motion* $\{X_t\}_{t\geq 0}$ defined by

$$X_t = |B_t|\,, \quad \text{for all } t \geq 0\,,$$

is a Markov process. It's not hard to show this by using Theorem 3.16. Moreover, its transition kernel $P_t(x, \cdot)$ is the law of $|Z|$ for Z normally distributed with mean x and variance t, which we call *the modulus normal distribution with parameters x and t.*

We intend to improve Markov property of Brownian motion slightly. Suppose that $\{B_t\}_{t\geq 0}$ is a d-dimensional Brownian motion. In a first step, we improve this and allow a slightly larger (augmented) σ-algebra $\mathcal{F}_t^+$ defined by

$$\mathcal{F}_t^+ = \bigcap_{s>t} \mathcal{F}_s^B \tag{3.14}$$

Clearly, the family $\left\{\mathcal{F}_t^+\right\}_{t\geq 0}$ is again a filtration and $\mathcal{F}_t^B \subset \mathcal{F}_t^+$. The fields $\{\mathcal{F}_t^+\}$ are nicer because they are *right continuous*:

$$\bigcap_{t>s} \mathcal{F}_t^+ = \bigcap_{t>s} \left(\bigcap_{u>t} \mathcal{F}_u^B \right) = \bigcap_{u>s} \mathcal{F}_u^B = \mathcal{F}_s^+\,.$$

In other words, the $\mathcal{F}_t^+$ allow us an "infinitesimal peek at the future". As we will see later, this property is very important in developing the theory. If $f(h) > 0$ for all $h > 0$, then in $d = 1$ the random variable

$$\limsup_{h\downarrow 0} \frac{B(t+h) - B(t)}{f(h)}$$

is measurable with respect to $\mathcal{F}_t^+$ but not $\mathcal{F}^B(t)$. We will see below that there are no interesting examples, i.e., $\mathcal{F}_t^+$ and $\mathcal{F}^B(t)$ are the same (up to null sets).

Theorem 3.20. *$B = \{B_t\}_{t\geq 0}$ is a Markov processes with respect to $\{\mathcal{F}_t^+\}_{t\geq 0}$ having the same transition kernals as in Theorem 3.19.*

Proof. It suffices to show that, for $t > s$ and $A \in \mathcal{B}(\mathbb{R})^d$ we have, almost surely,

$$\mathbb{P}(B_t \in A \mid \mathcal{F}_s^+) = P_{t-s}(B_s A)\,.$$

where the transition kernel $P_t(x,\cdot)$ has density (3.13). To see this, take any $h>0$, note that

$$\begin{aligned}\mathbb{P}\left(B_t \in A \mid \mathcal{F}_s^+\right) &= \mathbb{E}\left(\mathbb{P}\left(B_t \in A \mid \mathcal{F}_{s+h}\right) \mid \mathcal{F}_s^+\right) \\ &= \mathbb{E}\left(P_{t-s-h}(B(s+h), A) \mid \mathcal{F}_s^+\right) .\end{aligned}$$

Note that $P_\cdot(\cdot, A)$ is bounded and continuous on $(0,\infty)\times\mathbb{R}$, letting $h\downarrow 0$, by domainted convergence theorem, we have

$$\mathbb{P}\left(B_t \in A \mid \mathcal{F}_s^+\right) = \mathbb{E}\left(\lim_{h\downarrow 0} P_{t-s-h}(B_{s+h}, A) \mid \mathcal{F}_s^+\right) = P_{t-s}(B_s, A)\,.$$

We now complete the proof. $\square$

By Theorem 3.17, 3.19, 3.20, we have that for any bounded random variable $Y \in \mathcal{F}_\infty^B$ and any $s \geq 0$,

$$\mathbb{E}\left[Y\circ\theta_s|\mathcal{F}_s^+\right] = \mathbb{E}_{B(s)}Y = \mathbb{E}\left[Y\circ\theta_s|\mathcal{F}_s^B\right] \quad \text{a.s..}$$

Following this, we can show that :

Corollary 3.21. *For any bounded random variable $Y \in \mathcal{F}_\infty^B$,*

$$\mathbb{E}\left[Y|\mathcal{F}_s^+\right] = \mathbb{E}\left[Y|\mathcal{F}_s^B\right] \quad \textit{a.s..} \tag{3.15}$$

Proof. Indeed, it suffices to show (3.15) holds when $Y = 1_{\{B\in C\}}$ where C is a finite-dimensional measurable rectangular cylinder. In this case, Y can be written as $Y_1 \cdot Y_2\circ\theta_s$, where $Y_1 \in \mathcal{F}_s^B$, $Y_2 \in \mathcal{F}_\infty^B$, and they are bounded. So

$$\mathbb{E}\left[Y|\mathcal{F}_s^+\right] = Y_1\mathbb{E}\left[Y_2\circ\theta_s|\mathcal{F}_s^+\right] = \mathbb{E}\left[Y|\mathcal{F}_s^B\right] \quad \text{a.s..}$$ $\square$

Form (3.15), we get the following theorem.

Theorem 3.22. *For any $s \geq 0$ and $A \in \mathcal{F}_s^+$, we can find $C \in \mathcal{F}_s$ so that $\mathbb{P}(A\Delta C) = 0$.*

An alternative form of the (imporved) Markov property is the following one. In fact, by Lemma 3.18, it implies Theorem 3.20.

Theorem 3.23. *For every $s \geq 0$ the process $\{B(t+s)-B(s)\}_{t\geq 0}$ is independent of the σ-algebra $\mathcal{F}_s^+$. In other words, conditional on $\mathcal{F}_s^+$, the process $\{B(t+s)-B(s)\}_{t\geq 0}$ is a standard d-dimensional Brownian motion.*

Evidently, this theorem follows from Theorem 3.22 directly. However, we will give another heuristic and important proof.

Lemma 3.24. *Let ξ and $\xi_n, n \in \mathbb{N}$ be E-valued random varibles on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\xi_n \to \xi$ almost surely. Let $\mathcal{G} \subset \mathcal{F}$ be a σ-field. If for each n, ξ_n is independent of $\mathcal{G}$, then ξ is independent of $\mathcal{G}$.*

Proof. For any $f \in C_b(E)$ and bounded random variable $\eta \in \mathcal{G}$, we have

$$\mathbb{E}[f(\xi_n)\eta] = \mathbb{E}f(\xi_n)\ \mathbb{E}\eta\,.$$

Letting $n \to \infty$, by the bounded convergence theorem, $\mathbb{E}[f(\xi)\eta] = \mathbb{E}f(\xi)\mathbb{E}\eta\,$. By Lusin's theorem, for any bounded Borel measurable funciton f on E, this equation still holds. Then the desired result follows. □

Proof of Theorem 3.23. Pick a strictly decreasing sequence $\{s_n : n \in \mathbb{N}\}$ converging to s. For any $t_1, \cdots, t_m \geq 0$, by Contiuty of sample paths, Markov property and Lemma 3.24, the vector

$$\begin{aligned}&(B(t_1+s)-B(s), \cdots, B(t_m+s)-B(s)) \\ &= \lim_{j\uparrow\infty}(B(t_1+s_j)-B(s_j), \cdots, B(t_m+s_j)-B(s_j)) \quad \text{a.s.}\end{aligned}$$

is independent of $\mathcal{F}_s^+$, and so is the process $\{B(t+s)-B(s) : t \geq 0\}$. □

We now look at the **germ σ-algebra** $\mathcal{F}^+(0)$, which heuristically comprises all events defined in terms of Brownian motion on an infinitesimal small interval to the right of the origin.

Theorem 3.25 (Blumenthal's 0-1 law)**.** *Let $\{B_t\}_{t\geq 0}$ be d-dimensional Brownian motion starting form $x \in \mathbb{R}^d$. Then $\mathcal{F}_0^+$ is trivial, i.e., for each $A \in \mathcal{F}_0^+$,*

$$\mathbb{P}_x(A) \in \{0, 1\}\,. \tag{3.16}$$

Proof. This 0-1 law follows from Corollary 3.22 directly. Or using Theorem 3.23 for $s = 0$, we see that $\mathcal{F}^B(\infty) := \sigma(B(t) : t \geq 0)$ is independent of $\mathcal{F}^+(0)$. This applies in particular to $A \in \mathcal{F}^+(0)$, which therefore is independent of itself, hence has probability zero or one. □

This result is very useful in studying the local behavior of Brownian paths. As a first application we show that a standard linear Brownian motion has positive and negative values and zeros in every small interval to the right of 0.

Theorem 3.26. *Let $\{B_t\}_{t\geq 0}$ be a standard linear Brownian motion. Define*

$$\tau = \inf\{t > 0 : B_t > 0\} \quad \text{and} \quad \sigma_0 = \inf\{t > 0 : B_t = 0\}\,,$$

then

$$\mathbb{P}_0(\tau = 0) = \mathbb{P}_0(\sigma_0 = 0) = 1\,.$$

Proof. The event

$$\{\tau = 0\} = \bigcap_{n=1}^{\infty} \{\text{ there is } r \in \mathbb{Q} \cap (0, 1/n) \text{ such that } B(r) > 0\}$$

is clearly in $\mathcal{F}_0^+$. Hence we just have to show that this event has positive probability. This follows, as $\mathbb{P}_0(\tau \leq t) \geq \mathbb{P}_0(B_t > 0) = 1/2$ for $t > 0$. Hence

$$\mathbb{P}_0(\tau = 0) \geq \frac{1}{2}$$

and we have shown the first part.

The same argument works replacing $B_t > 0$ by $B_t < 0$ and from these two facts $\mathbb{P}_0(\sigma_0 = 0) = 1$ follows, using the intermediate value property of continuous functions. □

A further application is a 0-1 law for the tail σ-algebra of a d-dimensional Brownian motion. Define $\mathcal{F}'_t = \sigma(B_s : s \geq t)$. Let

$$\mathcal{T} = \bigcap_{t \geq 0} \mathcal{F}'_t \tag{3.17}$$

be the **tail σ-algebra** of all tail events.

Theorem 3.27 (Zero-one law for tail events)**.** *Let $\{B_t\}_{t\geq 0}$ be a d-dimensional Brownian motion starting at $x \in \mathbb{R}^d$. Suppose $A \in \mathcal{T}$ is a tail event. Then*

$$\mathbb{P}_x(A) \in \{0, 1\}.$$

Proof. It suffices to look at the case $x = 0$. Under the time inversion of Brownian motion, the tail σ-algebra is mapped on the germ σ-algebra, which contains only sets of probability zero or one, by Blumenthal's $0 - 1$ law. $\square$

3.6 The strong Markov property

Heuristically, the Markov property states that Brownian motion is started anew at each deterministic time instance. It is a crucial property of Brownian motion that this holds also for stopping times but become false for random times.

Theorem 3.28 (Strong Markov property)**.** *Let $B = \{B_t\}_{t\geq 0}$ be a d-dimensional Brownian motion with initial distribution μ defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then, for every finite stopping time T of $\{\mathcal{F}^+(t)\}_{t\geq 0}$, the process*

$$\{B(T + t) - B(T) : t \geq 0\}$$

is a standard d-dimensional Brownian motion independent of $\mathcal{F}^+_T$.

Proof. We first show our statement for the stopping times T_n which discretely approximate T from above,

$$T_n = \sum_{k=1}^{\infty} \frac{k}{2^n} 1_{\left\{\frac{k-1}{2^n} \leq T < \frac{k}{2^n}\right\}} \qquad \text{for } n \geq 1\,,$$

see Lemma 0.9. Write $B_{n,k} = \{B_{n,k}(t)\}_{t\geq 0}$ for the Brownian motion defined by $B_{n,k}(t) = B\left(t + k/2^n\right) - B\left(k/2^n\right),$ and $B_n = \{B_n(t)\}_{t\geq 0}$ for the process defined by $B_n(t) = B\left(t + T_n\right) - B\left(T_n\right)$. Suppose that $C \in \mathcal{F}^+_{T_n}$. Then, for every event $\{B_n \in A\},$ where $A \in \mathcal{B}(C[0,\infty))$, we have

$$\begin{aligned}\mathbb{P}\left(\{B_n \in A\} \cap C\right) &= \sum_{k=0}^{\infty} \mathbb{P}\left(\{B_{n,k} \in A\} \cap C \cap \{T_n = k/2^n\}\right) \\ &= \sum_{k=0}^{\infty} \mathbb{P}\left(B_{n,k} \in A\right) \mathbb{P}\left(C \cap \{T_n = k/2^n\}\right)\end{aligned}$$

using that $\{B_{n,k} \in A\}$ is independent of $C \cap \{T_n = k/2^n\} \in \mathcal{F}^+\left(k/2^n\right)$. Now, by Markov property of Brownian motion, $\mathbb{P}\left(B_{n,k} \in A\right) = \mathbb{P}\left(B \in A\right)$ does not depend on n, k and hence we get

$$\begin{aligned}\mathbb{P}\left(\{B_n \in A\} \cap C\right) &= \sum_{k=0}^{\infty} \mathbb{P}\left(B_{n,k} \in A\right) \mathbb{P}\left(C \cap \{T_n = k/2^n\}\right) \\ &= \mathbb{P}\left(B \in A\right) \sum_{k=0}^{\infty} \mathbb{P}\left(C \cap \{T_n = k/2^n\}\right) = \mathbb{P}(B \in A)\mathbb{P}(C)\end{aligned}$$

which shows that B_n is a standrad Brownian motion and independent of $\mathcal{F}^+_{T_n} \supset \mathcal{F}^+_T$. Take any $0 \leq t_1 < \cdots < t_j < \infty$, then

$$\begin{aligned}&(B(t_1 + T) - B(T), \cdots, B(t_j + T) - B(T)) \\ &= \lim_{n\uparrow\infty} \left(B\left(t_1 + T_n\right) - B\left(T_n\right), \cdots, B\left(t_j + T_n\right) - B\left(T_n\right)\right)\end{aligned}$$

is independent of $\mathcal{F}^+_T$, by Lemma 3.24 and has the same distribution as $(B(t_1 + T_n) - B(T_n), \cdots, B(t_j + T_n) - B(T_n))$. Thus $\{B(t+T) - B(T)\}_{t\geq 0}$ is a stardand d-dimensional Brownian motion and is independent of $\mathcal{F}^+_T$. $\square$

We given an exanmple that the strong Markov property become false for random times.

Example 3.6. Let

$$\tau = \inf\left\{t \geq 0 : B_t = \max_{0\leq s\leq 1} B_s\right\}.$$

It is intuitively clear that τ is a finite random time, but not a stopping time of $\{\mathcal{F}^+(t)\}_{t\geq 0}$. From Theorem 3.26, we can see that

$$\mathbb{P}(0 < \tau < 1) = 1\,.$$

Observe that the increment $B(\tau + t) - B(\tau)$ is non-positive in a small neighbourhood to the right of 0, which contradicts the strong Markov property and Theorem 3.26.

A Strong Markov Processes and Families

Definition 3.7. Let μ a Borel probability measure on $\mathbb{R}^d$. A progressively measurable, d-dimensional process $X = \{X_t\}_{t\geq 0}$ on some $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ is said to be a **strong Markov process** with transition kernel $\{P_t\}_{t\geq 0}$ and initial distribution μ, if

(i) $\mathbb{P}\left(X_0 \in A\right) = \mu(A)$, for all $A \in \mathcal{B}\left(\mathbb{R}^d\right)$;

(ii) for any optional time S of $\{\mathcal{F}_t\}_{t\geq 0}$, $t \geq 0$ and $A \in \mathcal{B}\left(\mathbb{R}^d\right)$,

$$\begin{aligned}\mathbb{P}\left(X_{S+t} \in A|\mathcal{F}_S^+\right) &= \mathbb{P}\left(X_{S+t} \in A|X_S\right)\\ &= P_t(X_S, A), \quad \text{a.s. on } \{S < \infty\}\,.\end{aligned}$$

Sometimes we rewrite (ii) as the following form

$$\mathbb{P}\left(X_{S+t} \in A, S < \infty|\mathcal{F}_S^+\right) = P_t(X_S, A)1_{\{S<\infty\}} \quad \text{a.s.}.$$

Definition 3.8. Let $\{P_t\}_{t\geq 0}$ be Markov transition kernels on $(\mathbb{R}^d, (\mathbb{R}^d))$. A d-dimensional **strong Markov family** is a progressively measurable process $X = \{X_t\}_{t\geq 0}$ on some filtrated measurable space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0})$ together with a family of probability measure $\{\mathbb{P}_x\}_{x\in \mathrm{R}^d}$ such that:

(i) for each $A \in \mathcal{F}$, the mapping $x \mapsto \mathbb{P}_x(A)$ is universally measurable;

(ii) Under $\mathbb{P}_x$, X is a strong Markov process with respect to $\{\mathcal{F}_t\}_{t\geq 0}$ starting at x. That is, $\mathbb{P}_x(X(0) = x) = 1$ and for every optional time S of $\{\mathcal{F}_t\}_{t\geq 0}$, $A \in \mathcal{B}(\mathbb{R}^d)$, and $t \geq 0$,

$$\mathbb{P}_x\left(X_{S+t} \in A | \mathcal{F}_S^+\right) = P_t(X_S, A) \quad \mathbb{P}_x\text{-a.s. on } \{S < \infty\}.$$

Remark 3.13. An optional time of $\{\mathcal{F}_t\}_{t\geq 0}$ is a stopping time of $\left\{\mathcal{F}_t^+\right\}_{t\geq 0}$. Because of the assumption of progressive measurability, the random variable X_S is $\mathcal{F}_S^+$-measurable. Moreover, if S is a stopping time of $\{\mathcal{F}_t\}_{t\geq 0}$, then X_S is $\mathcal{F}_S$ measurable. In this case, we can take conditional expectations with respect to $\mathcal{F}_S$ on both sides of (ii) to obtain

$$\mathbb{P}_x\left(X_{S+t} \in A | \mathcal{F}_S\right) = P_t(X_S, A) \quad \mathbb{P}_x\text{-a.s. on } \{S < \infty\}.$$

Setting S equal to a constant $s \geq 0$, we obtain that, every strong Markov family is a Markov family. Likewise, every strong Markov process is a Markov process. However, not every Markov family enjoys the strong Markov property; a counterexample to this effect, involving a progressively measurable process X, appears in Wentzell (1981), p. 161.

Remark 3.14. By Theorem 3.28 and Lemma 3.18, we have : A d-dimensional Brownian family is a strong Markov family. A d-dimensional Brownian motion is a strong Markov process.

Whenever S is an optional time of $\{\mathcal{F}_t\}$ and $u > 0$, then $S + u$ is a stopping time of $\{\mathcal{F}_t\}$.This fact can be used to replace the constant s in the proof of Theorem 3.17 by the optional time S, thereby obtaining the following result.

Theorem 3.29. *Let* $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \{X_t\}_{t\geq 0}, \{P_t\}_{t\geq 0}, \{\mathbb{P}_x\}_{x\in\mathbb{R}^d})$ *be a strong Markov family. Let* $\{\theta_s\}_{s\geq 0}$ *be a family of shift-operators. Then, for any bounded (or non-negative) random variavle* $Y \in \mathcal{F}_\infty^X$ *any optional time* S *of* $\{\mathcal{F}_t\}_{t\geq 0}$*, and* $x \in \mathbb{R}^d$*,*

$$\mathbb{E}_x[Y \circ \theta_S | \mathcal{F}_S^+] = \mathbb{E}_{X_S} Y \quad \mathbb{P}_x\text{-a.s. on } \{S < \infty\}.$$

Proof. Firstly, we show that the theorem holds, if $Y = 1_{\{X\in C\}}$, where C is a finite-dimensional rectangular cylinder set in $(\mathbb{R}^d)^{[0,\infty)}$. To see this, let

$$C = \{\phi : [0,\infty) \to \mathbb{R}^d : \phi(t_0) \in A_0, \phi(t_1) \in A_1, \cdots, \phi(t_n) \in A_n\},$$

where $0 = t_0 < t_1 < \cdots < t_n < \infty$, $A_j \in \mathcal{B}(\mathbb{R}^d), 0 \le j \le n$ and $n \ge 1$. What we need to show is that

$$\begin{aligned}&\mathbb{P}_x\left(X(S) \in A_0, X(S+t_1) \in A_1, \cdots, X(S+t_n) \in A_n | \mathcal{F}_S^+\right)\\ &= 1_{\{X_S\in A_0\}} \int_{A_1} P_{t_1}(X_S, \mathrm{d}x_1) \cdots \int_{A_n} P_{t_n - t_{n-1}}(x_{n-1}, \mathrm{d}x_n),\end{aligned}$$

which can be proved by induction on n and by strong Markov property.

Secondly, using $\pi - \lambda$ theorem, we get that the theorem holds if $Y = 1_{\{X\in A\}}$, where $A \in \mathcal{B}(\mathbb{R}^d)^{[0,\infty)}$. Finally, recalling that any bounded (or non-negative) measureable $Y \in \mathcal{F}_\infty^X$, there exists a bounded (or non-negative) measurable funciton f on $\left((\mathbb{R}^d)^{[0,\infty)}, \mathcal{B}(\mathbb{R}^d)\right)$ so that $Y = f(X)$. So Y can be approximated by simple functions $\sum_i c_i 1_{\{X\in A_i\}}$, then we complete the proof. □

For right-continuous processes a slight extension of the strong Markov property is given.

Proposition 3.30. *Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\ge 0}, \{X_t\}_{t\ge 0}, \{P_t\}_{t\ge 0}, \{\mathbb{P}_x\}_{x\in\mathrm{R}^d})$ be a strong Markov family, and the process X be right-continuous. Let S be an optional time of $\{\mathcal{F}_t\}$. Let T an $\mathcal{F}_S^+$-measurable random time satisfying $S(\omega) \le T(\omega)$ for all ω. Then, for any $x \in \mathbb{R}^d$ and $A \in \mathcal{B}(\mathbb{R}^d)$,*

$$\mathbb{P}_x\left(X_T \in A | \mathcal{F}_S^+\right) = P_{T-S}(X_S, A) \quad \mathbb{P}_x\text{-a.s. on } \{T < \infty\}.$$

Proof. It suffices to show that for $f \in C_0(\mathbb{R}^d)$,

$$\mathbb{E}_x\left(f(X_T)|\mathcal{F}_S^+\right) = \int f(y) P_{T-S}(X_S, \mathrm{d}y) \quad \mathbb{P}_x\text{-a.s.on } \{T < \infty\}.$$

We first show our statement for the stopping times T_n which discretely approximate T from above,

$$T_n = \sum_{k=1}^{\infty}(S + \frac{k}{2^n})\, 1_{\{\frac{k-1}{2^n} \leq T - S < \frac{k}{2^n}\}} + \infty_{T=\infty} \qquad \text{for } n \geq 1\,,$$

It's easy to see that

$$\mathbb{E}_x \left(f(X_{T_n})|\mathcal{F}_S^+\right) = \int f(y) P_{T_n - S}(X_S, \mathrm{d}y) \quad \mathbb{P}_x\text{-a.s. on } \{T < \infty\}\,.$$

The bounded convergence theorem for conditional expectations and the right-continuity of X imply that the left-hand side converges to The bounded convergence theorem for conditional expectations and the rightcontinuity of X imply that the left-hand side converges to $\mathbb{E}_x \left(f(X_T)|\mathcal{F}_S^+\right)$ as $n \to \infty$. Since $E_y f(X_t)$ is right-continuous in t for every $y \in \mathbb{R}^d$, the right-hand side converges to the desired integral. □

B Feller property

Suppose $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \{X_t\}_{t\geq 0}, \{P_t\}_{t\geq 0}, \{\mathbb{P}_x\}_{x\in\mathbb{R}^d})$ is a Markov family. If for any $f \in C_0(\mathbb{R}^d)$ and $t \geq 0$, the function

$$x \mapsto \mathbb{E}_x f(X_t) = \int f(y) P_t(x, \mathrm{d}y)$$

is continuous on $\mathbb{R}^d$, then we say that $\{X_t\}_{t\geq 0}$ has *Feller property.*

As an example, for a d-dimensional Brownian family, it's obviously a Markov family. Besides, for any $f \in C_0(\mathbb{R}^d)$ and $t \geq 0$, the function

$$x \mapsto \mathbb{E}_x f(B_t) = \int f(y) p_t(x, y)\, \mathrm{d}y\ ,$$
$$\text{where}\ \ p_t(x, y) = (2\pi t)^{-d/2} \exp\{-\frac{\|x - y\|^2}{2t}\}\,,$$

is obviously continuous on $\mathbb{R}^d$, so it has Feller property.

Infact, every right-contionuous Markov family with Feller property has the strong Markov property:

Theorem 3.31. $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \{X_t\}_{t\geq 0}, \{P_t\}_{t\geq 0}, \{\mathbb{P}_x\}_{x\in\mathbb{R}^d})$ *is a right continuous Markov family with Feller property. Let* $\{\theta_s\}_{s\geq 0}$ *be a family of shift-operators. Let* $(s,\omega)\mapsto Y_s(\omega)\,;\,[0,\infty)\times\Omega\to\mathbb{R}$ *be bounded (or non-negative) and* $\mathcal{B}([0,\infty))\times\mathcal{F}^X_\infty$ *measurable. Then for any optional time* S *of* $\{\mathcal{F}_t\}_{t\geq 0}$ *and* $x\in\mathbb{R}^d$*, we have*

$$\mathbb{E}_x\left[Y_S\circ\theta_S|\mathcal{F}_S^+\right] = \mathbb{E}_{X_S}Y_S \quad \mathbb{P}_x\text{-a.s. on } \{S<\infty\}\,.$$

where $\mathbb{E}_{X_S}Y_S$ *is the function* $\varphi(s,x)=\mathbb{E}_xY_s$ *evaluated at* (S,X_S)*.*

Remark 3.15. Clearly, $Y:\left([0,\infty)\times\Omega,\mathcal{B}([0,\infty))\times\mathcal{F}^X_\infty\right)\to(\mathbb{R},\mathcal{B}(\mathbb{R}))$ is measurable and bounded (or non-negative) *if and only if* there exists

$$f:\left([0,\infty)\times(\mathbb{R}^d)^{[0,\infty)},\mathcal{B}([0,\infty))\times\mathcal{B}(\mathbb{R}^d)^{[0,\infty)}\right)\to(\mathbb{R},\mathcal{B}(\mathbb{R}))$$

is measurable and bounded (or non-negative) so that $Y_s(\omega)=f(s,X(\omega))$. Therefore, Theorem 3.31 is equivalent to say for the function f above,

$$\mathbb{E}_x\left[f(S,X\circ\theta_S)|\mathcal{F}_S^+\right]=\mathbb{E}_{X_S}f(S,X) \quad \mathbb{P}_x\text{-a.s. on } \{S<\infty\}\,. \tag{3.18}$$

where $\mathbb{E}_{X_S}f(S,X)$ is the function $\varphi(s,x)=\mathbb{E}_xf(s,X)$ evaluated at (S,X_S).

Proof. **Step 1**. We first prove the (3.18) for the stopping times S_n defined by

$$S_n=\sum_{k=1}^{\infty}\frac{k}{2^n}1_{\left\{\frac{k-1}{2^n}\leq S<\frac{k}{2^n}\right\}}+\infty\,1_{\{S=\infty\}} \qquad \text{for } n\geq 1\,.$$

We showed that $\{S_n\}$ are stopping times of $\{\mathcal{F}_t\}_{t\geq 0}$ and $\mathcal{F}_S^+=\cap_n\mathcal{F}(S_n)$ in Lemma 0.9, 0.10. We break things down according to the value of S_n apply the Markov property, and put the pieces back together. For $A\in\mathcal{F}_{S_n}$,

$$\mathbb{E}_x\left[f(S_n,X\circ\theta_{S_n})1_{\{S_n<\infty\}}1_A\right]=\sum_{k=1}^{\infty}\mathbb{E}_x\left[f(k/2^n,X\circ\theta_{k/2^n})1_{\{S_n=k/2^n\}}1_A\right]\,.$$

Now since $A \in \mathcal{F}_{S_n}$, $A \cap \{S_n = k/2^n\} \in \mathcal{F}(k/2^n)$, it follows from the Markov property that the above sum is

$$\sum_{n=1}^{\infty} \mathbb{E}_x \left[\mathbb{E}_{X(k/2^n)} f(k/2^n, X) 1_{\{S_n = k/2^n\}} 1_A \right] = \mathbb{E}_x \left[\varphi(S_n, X(S_n)) 1_{\{S_n < \infty\}} 1_A \right] .$$

Thus we have

$$\mathbb{E}_x \left[f(S_n, X\theta_{S_n}) | \mathcal{F}(S_n) \right] = \varphi(S_n, X(S_n)) \quad \mathbb{P}_x\text{-a.s. on } \{S_n < \infty\} . \tag{3.19}$$

Step 2. To be able to taking limits as $n \to \infty$ in (3.19), we restrict our attention to f's of the form

$$f(s, \phi) = h(s) \prod_{j=0}^{m} f_j \left(\phi \left(t_j \right) \right) \text{ for } s \geq 0, \; \phi \in (\mathbb{R}^d)^{[0,\infty)} ,$$

where $0 = t_0 < t_1 < \ldots < t_m$ and $f_j \in C_0(\mathbb{R}^d)$ and $h \in C_0(\mathbb{R})$. The Feller property implies that $x \to \int f_j(y) P_t(x, \mathrm{d}y)$ is continuous. From this, it follows that

$$\begin{aligned} \varphi(s, x) &= \mathbb{E}_x Y_s = \mathbb{E}_x f(s, X) \\ &= h(s) f_0(x) \int f_1(y_1) P_{t_1}(x, \mathrm{d}y_1) \cdots \int f_m(y_m) P_{t_m - t_{m-1}}(y_{m-1}, \mathrm{d}y_m) \end{aligned}$$

is bounded and continuous on $[0, \infty) \times \mathbb{R}^d$. By (3.19) and $\{S_n < \infty\} = \{S < \infty\}$ we have

$$\mathbb{E}_x \left[Y_{S_n} \circ \theta_{S_n} | \mathcal{F}_S^+ \right] = \mathbb{E}_x \left[\varphi(S_n, X(S_n)) | \mathcal{F}_S^+ \right] \quad \mathbb{P}_x\text{-a.s. on} \{S < \infty\}$$

Now, as $n \to \infty$, $S_n \downarrow S$, by the right-continuity of sample paths, $X\left(S_n\right) \to X_S, \varphi\left(S_n, X\left(S_n\right), \right) \to \varphi(S, X_S)$ and

$$\begin{aligned} f(S_n, X \circ \theta_{S_n}) &= h(S_n) \prod_{j=0}^{m} f_j \left(X(S_n + t_j) \right) \\ &\to h(S) \prod_{j=0}^{m} f_j \left(X(S + t_j) \right) = f(S, X \circ \theta_S) , \end{aligned}$$

so the bounded convergence theorem for conditional expectation implies

$$\mathbb{E}_x\left[f(S, X\circ\theta_S)|\mathcal{F}_S^+\right] = \mathbb{E}_x\left[\varphi(S, X(S))|\mathcal{F}_S^+\right] \quad \mathbb{P}_x\text{-a.s. on } \{S<\infty\}\,.$$

Step 3. To complete the proof now, we will apply the canonical method. We let $\mathcal{H}$ be the collection of f for which (3.18) holds. Firstly, we show that (3.18) holds for $f = 1_C$ that C has the form

$$C = H\times\{\phi\in(\mathbb{R}^d)^{[0,\infty)} : \phi(t_0)\in G_0, \phi(t_1)\in G_1, \cdots, \phi(t_m)\in G_m\}$$

where $0 = t_0 < t_1 < \cdots < t_n < \infty$, and H, G_j are bounded open sets in $[0,\infty)$ and $\mathbb{R}^d$, sespectively. Pick $\{h^{(n)}\}$, $\{f_j^{(n)}\}$ in $C_0(\mathbb{R})$ and $C_0(\mathbb{R}^d)$ respectively, that $h^n\uparrow 1_H$, and $f_j^n\uparrow 1_{G_j}$ as $n\uparrow\infty$. The facts that (3.18) holds for

$$f^{(n)}(s,\phi) = h^{(n)}(s)\prod_{j=1}^m f_j^{(n)}\left(\phi\left(t_j\right)\right)$$

and bounded convergence theorem implies that $f = 1_C\in\mathcal{H}$. Secondly, using $\pi-\lambda$ theorem, we get that the theorem holds if $f = 1_A$ is the incicator function of any $A\in\mathcal{B}([0,\infty))\times\mathcal{B}(\mathbb{R}^d)^{[0,\infty)}$. Finally, recalling that any bounded (or non-negative) measureable f can be approximated by simple functions, we complete the proof. □

3.7 Path properties

Part of the appeal of Brownian motion lies in the fact that the distribution of certain of its functionals can be obtained in closed form. In this section, if not specifically stated, we always suppose $\{B_t\}_{t\geq 0}$ a standard one-dimensional Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P}_0)$. Perhaps the most fundamental of these functionals is the *passage time* T_a to a level $a\in\mathbb{R}$, defined by

$$T_a \coloneqq \inf\{t\geq 0 : B_t = a\}\ .$$

We recall that a passage time for a continuous process is a stopping time (Example 0.6). This application of the strong Markov property (Theorem 3.31) shows why we want to allow the function f that we apply to the shifted path to depend on the stopping time.

Theorem 3.32 (Reflection principle). *Let $\{B_t\}_{t\geq 0}$ be a standard Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P}_0)$. Then the process $\{B_t^*\}_{t\geq 0}$ called Brownian motion reflected at a and defined by*

$$B_t^* := \begin{cases} B_t \ , & t \leq T_a \, ; \\ 2a - B_t \ , & t > T_a \, . \end{cases}$$

is also a standard Brownian motion.

Proof. Clearly, evert sample path of B^* is continuous, so it suffices to show that B^* has the same finite-distribution as B. Without loss of generality, let $a > 0$. Then we shall show that for any $n \geq 1$, $A_j \in \mathcal{B}(\mathbb{R})$, $1 \leq j \leq n$ and $0 = t_0 < t_1 < \cdots < t_n < \infty$

$$\mathbb{P}_0(B^*(t_j) \in A_j \text{ for all } j) = \mathbb{P}_0(B(t_j) \in A_j \text{ for all } j)\, .$$

Note that, for any $1 \leq m \leq n$,

$$\begin{aligned} &\mathbb{P}\big(B(t_j) \in A_j, j < m; \, t_{m-1} \leq T_a < t_m; \, B(t_j) \in 2a - A_j, j \geq m \mid \mathcal{F}_{T_a}^+\big) \\ &= 1_{\{B(t_j)\in A_j, j<m; \, t_{m-1}\leq T_a\}} \, \mathbb{P}_0\big(2a - B(t_j) \in A_j, j \geq m, T_a < t_m \mid \mathcal{F}_{T_a}^+\big)\, . \end{aligned}$$

By strong Markov property we have,

$$\begin{aligned} &\mathbb{P}_0\big(2B(T_a) - B(t_j) \in A_j, j \geq m, T_a < t_m \mid \mathcal{F}_{T_a}^+\big) \\ &= \mathbb{P}_0\big(2B \circ \theta_{T_a}(0) - B \circ \theta_{T_a}(t_j - T_a) \in A_j, j \geq m, T_a < t_m \mid \mathcal{F}_{T_a}^+\big) \\ &= \mathbb{E}_0\big(f(T_a, B \circ \theta_{T_a}) \mid \mathcal{F}_{T_a}^+\big)\, , \end{aligned}$$

where $f : [0, \infty) \times C[0, \infty) \to \mathbb{R}$ is given by

$$f(t, \phi) = 1_{\{t<t_m\}} 1_{\{2\phi(0)-\phi(t_j-t)\in A_j, j\geq m\}}\, .$$

We omit the proof that f is measurable and compute that

$$\begin{aligned}\phi(t,x) &:= \mathbb{E}_x f(t,B) = 1_{\{t<t_m\}}\mathbb{P}_x(2x - B(t_j - t) \in A_j, j \geq m) \\ &= 1_{\{t<t_m\}}\mathbb{P}_x(B(t_j - t) \in A_j, j \geq m) \\ &= 1_{\{t<t_m\}} \int_{A_m} p_{t_m - t}(x, \mathrm{d}y_m) \cdots \int_{A_n} p_{t_n - t_{n-1}}(a, \mathrm{d}y_n)\end{aligned}$$

the last equals sign is because under $\mathbb{P}_x$, $\{B_t\}_{t\geq 0}$ and $\{2x - B_t\}_{t\geq 0}$ has the same distribution. Thus by strong Markov property (Theorem 3.31),

$$\begin{aligned}&\mathbb{P}_0\big(2a - B(t_j) \in A_j, j \geq m, T_a < t_m \mid \mathcal{F}^+_{T_a}\big) \\ &= 1_{\{T_a<t_m\}} \int_{A_m} p_{t_m - T_a}(a, \mathrm{d}y_m) \cdots \int_{A_n} p_{t_n - t_{n-1}}(a, \mathrm{d}y_n) \\ &= \mathbb{P}_0\big(B(t_j) \in A_j, j \geq m, T_a < t_m \mid \mathcal{F}^+_{T_a}\big)\,.\end{aligned}$$

Then we get

$$\begin{aligned}&\mathbb{P}\big(B(t_j) \in A_j, j < m;\, t_{m-1} \leq T_a < t_m;\, B(t_j) \in 2a - A_j, j \geq m \mid \mathcal{F}^+_{T_a}\big) \\ &= \mathbb{P}\big(B(t_j) \in A_j, j < m;\, t_{m-1} \leq T_a < t_m;\, B(t_j) \in A_j, j \geq m \mid \mathcal{F}^+_{T_a}\big)\,.\end{aligned}$$

Since m is arbitary in $\{0, 1, \cdots, n\}$, we get

$$\mathbb{P}_0(B^*(t_j) \in A_j \text{ for all } j) = \mathbb{P}_0(B(t_j) \in A_j \text{ for all } j)\,. \qquad \square$$

Exercise 3.1. Show the reflection principle by using Proposition 3.30.

There is a more general result which implies the reflection principle.

Theorem (Splicing)**.** *Let $\{B_t\}_{t\geq 0}$ be a standard d-dimensional Brownian motion. Let T be a optional time for $\{\mathcal{F}^B_t\}_{t\geq 0}$. Let $\{W_t\}_{t\geq 0}$ be a standard d-dimensional Brownian motion, which is independent of $\mathcal{F}^+_T$. Then the spliced process*

$$B^*(t) := \begin{cases} B(t)\,, & t \leq T\,; \\ B(T) + W(t-T)\,, & t > T\,. \end{cases} \tag{3.20}$$

is also a standard d-dimensional Brownian motion.

A Hitting times

Our next goal is to compute the distribution of the hitting times T_a.

Theorem 3.33. *Let $\{B_t\}_{t\geq 0}$ be a standard, one-dimensional Brownian motion, and for $a \geq 0$, let T_a be the first passage time to a. Then*

$$\mathbb{P}_0\left(T_a \leq t\right) = 2\ \mathbb{P}_0\left(B_t \geq a\right) = \mathbb{P}_0\left(|B_t| \geq a\right)\ .$$

Moreover, T_a is a continuous random variable with probability density

$$\rho(x) = \frac{a}{\sqrt{2\pi x^3}} e^{-\frac{a^2}{2x}}\, 1_{\{x>0\}} \tag{3.21}$$

Proof. Note that

$$\{T_a \leq t\} = \{T_a \leq t, B_t \geq a\} \cup \{T_a \leq t, B_t < a\}\ .$$

By the continuity of paths,

$$\mathbb{P}_0\left(T_a \leq t, B_t \geq a\right) = \mathbb{P}_0\left(B_t \geq a\right)\ ,$$

and

$$\mathbb{P}_0\left(T_a \leq t, B_t < a\right) = \mathbb{P}_0\left(T_a^* \leq t, B_t^* > a\right) = \mathbb{P}_0\left(B_t^* > a\right)$$

where B^* is defined in (3.20), and T_a^* is the first hitting time of B^*, in fact $T_a^* = T_a$. By Reflection principle, B^* is a standard Brownian motion under $\mathbb{P}_0$, thus we have

$$\mathbb{P}_0\left(T_a \leq t\right) = 2\mathbb{P}(B_t > a) = \mathbb{P}_0(|B_t| > a)\,.$$

Clearly, $\mathbb{P}_0(T_a < \infty) = 1$. Thus T_a is a random variable. On the other hand,

$$\mathbb{P}_0\left(T_a \leq t\right) = 2\mathbb{P}_0\left(B_t > a\right) = 2\int_a^\infty \frac{1}{\sqrt{2\pi t}} e^{-\frac{y^2}{2t}}\,\mathrm{d}y\,.$$

Then we can compute the density of T_a. $\square$

Theorem 3.34. *Under $\mathbb{P}_0$, $\{T_a\}_{a\geq 0}$ has stationary independent increments. is an increasing Markov process with transition kernel given by the densities*

$$p_a(t,s) = \frac{a}{\sqrt{2\pi(s-t)^3}} \exp\left\{-\frac{a^2}{2(s-t)}\right\} 1_{\{s>t\}}, \qquad \textit{for } a>0\,.$$

This process is called the stable subordinator of index $1/2$.

Proof. The first step is to notice that if $0<a<b$ then

$$T_b - T_a = \inf\{t \geq 0 : B(T_a + t) - B(T_a) = b - a\} \quad \mathbb{P}_0\text{-a.s.}.$$

Let $W(t) = B(T_a+t) - B(T_a)$ and $\{T_a^W\}_{a\geq 0}$ the hitting times of W. By strong Markov property (Theorem 3.28),

$$\mathcal{L}[T_b - T_a] = \mathcal{L}[T_{b-a}^W] = \mathcal{L}[T_{b-a}]\,.$$

To show that the increments are independent, let $0 = a_0 < a_1 \ldots < a_n < \infty$, let f_j, $1 \leq j \leq n$ be bounded and measurable, and let $F_j = f_j\left(T_{a_j} - T_{a_{i-1}}\right)$. Conditioning on $\mathcal{F}^+(T_{a_{n-1}})$ and using the preceding calculation we have

$$\mathbb{E}_0\left(\prod_{i=1}^{n} F_j\right) = \mathbb{E}_0\left(\prod_{i=1}^{n-1} F_j \cdot \mathbb{E}_0\left[F_n | \mathcal{F}^+_{T_{a_{n-1}}}\right]\right) = \mathbb{E}_0\left(\prod_{i=1}^{n-1} F_j\right)\mathbb{E}_0 F_n\,.$$

By induction, it follows that $\mathbb{E}_0 \Pi_{j=1}^n F_j = \Pi_{j=1}^n \mathbb{E}_0 F_j$, which implies the desired conclusion. The densities of transition kernels given by 3.21. □

It is easy to determine the Laplace transform

$$\varphi_a(\lambda) = \mathbb{E}_0 \exp\left(-\lambda T_a\right) \qquad \text{for } a \geq 0$$

To do this, we start by observing that Theorem 3.34 implies

$$\varphi_x(\lambda)\varphi_y(\lambda) = \varphi_{x+y}(\lambda)\,.$$

It follows easily from this that $\varphi_a(\lambda) = \exp(-ac(\lambda))$. To identify $c(\lambda)$, recall that $\mathcal{L}(T_a) =$ implies

$$\mathbb{E}\exp\left(-T_a\right) = \mathbb{E}\exp\left(-a^2 T_1\right)$$

so $ac(1) = c\left(a^2\right)$, i.e., $c(\lambda) = c(1)\sqrt{\lambda}$. Since all of our arguments also apply to σB_t we cannot hope to compute $c(1)$. Later we will use martingale method to show that

$$\mathbb{E}_0\left(\exp\left(-\lambda T_a\right)\right) = \exp(-a\sqrt{2\lambda})\,.$$

Also, we can using Theorem 3.33 to compute this directly.

Exercise 3.2. Show that $\mathbb{P}_0\left(\lim_{b\to a} T_b = T_a\right) = 1$ for all $a \in \mathbb{R}$.

B Running Maximum

Set $\{B_t\}_{t\geq 0}$ a one-dimensional Brownian motion on $(\Omega, \mathcal{F}, \{\mathbb{P}_x\}_{x\in\mathbb{R}})$. Define the **running maximum**(or maximum-to-date)

$$M_t = \max_{0\leq s\leq t} B_s\,.$$

A priori it is not at all clear what the distribution of this random variable is, but we can determine it as a consequence of the reflection principle: Note that

$$\{M_t > a\} = \{T_a < t\}\,, \quad \mathbb{P}_0\text{-a.s.}$$

combine this and Theorem 3.33 , we get:

Theorem 3.35. *Under $\mathbb{P}_0$, for each $t > 0$, the distribution of M_t is given by*

$$\mathbb{P}_0(M_t > a) = \mathbb{P}_0(|B_t| > a), \quad \textit{for all } a > 0\,.$$

In other words, M_t as a modulus normal distribution with parameters 0 and t.

Remark 3.16. For $x > y$, using reflection principle to compute

$$\mathbb{P}\left(B_t \leq y < x \leq M_t\right)\,,$$

we can show that for $t > 0$, (M_t, B_t) is a continuous random vector, with density:

$$\frac{2(2x-y)}{\sqrt{2\pi t^3}} \exp\left\{-\frac{(2x-y)^2}{2t}\right\} 1_{\{x\geq y, x\geq 0\}}$$

Example 3.7 (Large law of number)**.** $\{B_t\}_{t\geq 0}$ is a standard Brownian motion, then

$$\frac{B_t}{t} \to 0 \quad \text{a.s.}$$

To see this, by SLLN we have $B(n)/n \to 0$ almost surely. Then let

$$X_n = \sup_{n-1\leq t\leq n} (B_t - B_n), \quad Y_n = \inf_{n-1\leq t\leq n} (B_t - B_n)\,,$$

Then by Markov property, $\{X_n\}$ is a i.i.d. sequence and $X_1 = M_1$, $\mathcal{L}[Y_1] = \mathcal{L}[-M_1]$, thus by SLLN again, we have

$$\frac{X_n}{n} \to 0 \quad \text{and} \quad \frac{Y_n}{n} \to 0 \quad \text{a.s.}.$$

Therefore $B_t/t \to 0$ almost surely.

We now exploit the Markov property to study the local and global maxima of a standard Brownian motion.

Theorem 3.36. *Let $\{B_t\}_{t\geq 0}$ a standard Brownian motion. Then for any given two nonoverlapping closed time intervals $[a_1, b_1]$ and $[a_2, b_2]$, i.e. $b_1 \leq a_2$, the maxima of Brownian motion on them are different almost surely. In other words,*

$$\mathbb{P}_0\left(\max_{a_1\leq t\leq b_1} B_t = \max_{a_2\leq t\leq b_2} B_t\right) = 0\,.$$

Proof. Denote by m_1 and m_2, the maxima of Brownian motion on $[a_1, b_1]$ and $[a_2, b_2]$ respectively. On the one hand, let

$$m_1 - B(b_1) = \max_{a_1\leq t\leq b_1} B(t) - B(b_1) = \max_{0\leq t\leq b_1-a_1} B(b_1 - t) - B(b_1) =: \tilde{m}_1\,.$$

Since $\{B(b_1 - t) - B(b_1)\}_{0\leq t\leq b_1-a_1}$ is a standard Brownian motion (Theorem 3.2 (ii)), $\tilde{m}_1$ has the same distribution as $|B(b_1 - a_1)|$ by Theorem 3.35. On the other hand, let

$$m_2 - B(a_2) = \max_{0\leq t\leq b_2-a_2} (B(a_2 + t) - B(a_2)) =: \tilde{m}_2\,.$$

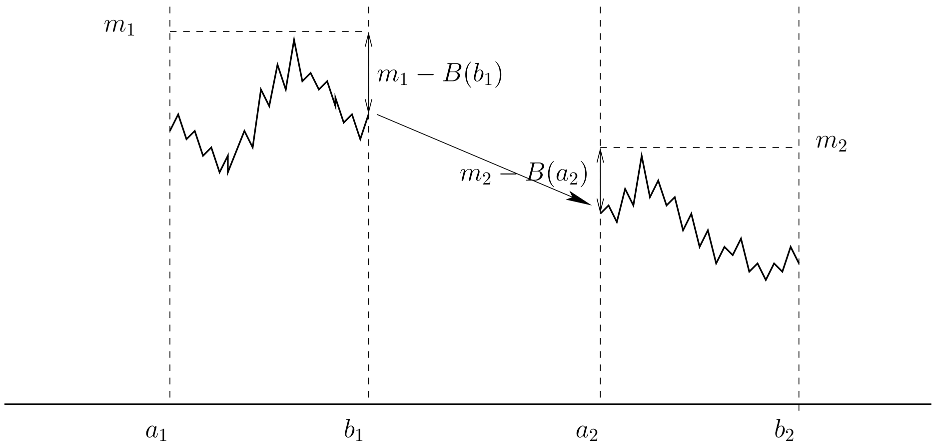


Figure 3.3: $\tilde{m}_1$, $B(a_2) - B(b_1)$ and $\tilde{m}_2$.

Applying Theorem 3.14 and Theorem 3.35, we get that $\tilde{m}_2$ has the same distribution as $|B(b_2 - a_2)|$. Thus

$$\{m_1 = m_2\} = \{\tilde{m}_2 - \tilde{m}_1 = B(a_2) - B(a_1)\} ,$$

By Markov proprty, $\tilde{m}_2$, $\tilde{m}_1$, $B(a_2) - B(a_1)$ are indenpendent continuous random variable, hence this event has probability 0. $\square$

Corollary 3.37. *Let $\{B_s\}_{0\leq s\leq t}$ a standard Brownian motion, then almost surely,*

(i) every local maximum is a strict local maximum.

(ii) the set of times where the local maxima are attained is countable and dense.

(iii) the global maximum is attained at a unique time in $(0, t)$.

Proof. (i). By Theorem 3.36, almost surely, all nonoverlapping pairs of nondegenerate compact intervals with rational endpoints have different maxima. If Brownian motion however has a non-strict local maximum, there are two such intervals where Brownian motion has the same maximum.

(ii) In particular, almost surely, the maximum over any nondegenerate compact interval with rational endpoints is not attained at an endpoint. Hence every such interval contains a local maximum, and the set of times where local maxima are attained is dense. As every local maximum is strict, this set has at most the cardinality of the collection of these intervals.

(iii) Almost surely, for any rational number $q \in [0, 1]$ the maximum in $[0, q]$ and in $[q, 1]$ are different. Note that, if the global maximum is attained for two points $t_1 < t_2$ there exists a rational number $t_1 < q < t_2$ for which the maximum in $[0, q]$ and in $[q, 1]$ agree. $\square$

We now recited a famous theorem of Paul Lévy, which shows that the difference of the maximum process of a Brownian motion and the Brownian motion itself is a reflected Brownian motion. See Figure 3.4.

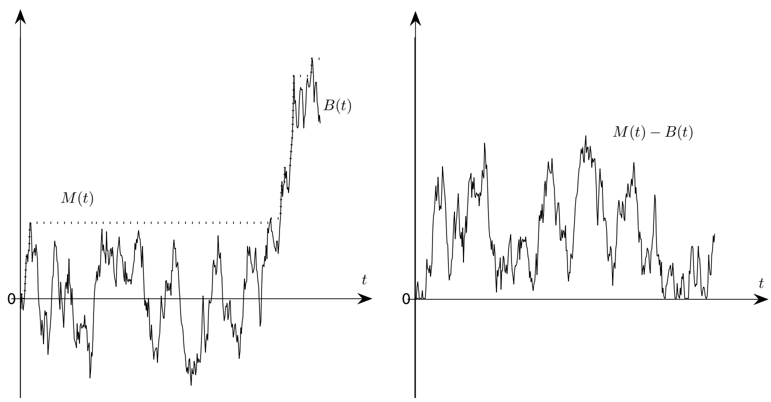


Figure 3.4: B_t and $M_t - B_t$.

Theorem 3.38 (Paul Lévy)**.** *Let $\{M_t\}_{t\geq 0}$ be the running maximum of standard Brownian motion $\{B_t\}_{t\geq 0}$. Then, the process $\{M_t - B_t\}_{t\geq 0}$ is a reflected Brownian motion.*

The main step is to show that the process $\{M_t - B_t\}_{t\geq 0}$ is a Markov process and its Markov transition kernel $P_t(x,\cdot)$ has modulus normal distribution with parameters x and t. Once this is established, it is immediate that the finite-dimensional distributions of this process agree with those of a reflected Brownian motion (Example 3.5). Obviously, $\{M_t - B_t\}_{t\geq 0}$ has almost surely continuous sample paths.

Proof. Fix $s, t \geq 0$. We shall show that

$$\mathcal{L}[M_{t+s} - B_{t+s}|\mathcal{F}_s^{M-B}] = \mathcal{L}[M_{t+s} - B_{t+s}|M_s - B_s]\,.$$

Let $\tilde{B} = \{\tilde{B}_u\}_{u\geq 0}$ defined by $\tilde{B}_u = B_{u+s} - B_s$ for $u \geq 0$ and $\tilde{M}_t = \max_{0\leq u\leq t} \tilde{B}(u)$. Observe that $M_{s+t} = M_s \vee (B_s + \tilde{M}_t)$, and hence

$$\begin{aligned} M_{t+s} - B_{t+s} &= \big(M_s \vee (B_s + \tilde{M}_t)\big) - B_{s+t} \\ &= \big(M_s - B_s\big) \vee \tilde{M}_t - \big(B_{s+t} - B_s\big) \end{aligned}.$$

That is,

$$M_{t+s} - B_{t+s} = \big(M_s - B_s\big) \vee \tilde{M}_t - \tilde{B}(t)\,.$$

Note that, by Markov property, $\tilde{B}_t$ and $\tilde{W}_t$ both are independent of $\mathcal{F}_s^B \supset \mathcal{F}_t^{M-B}$, it follows from Lemma 3.18 that

$$\mathcal{L}\left[\big(M_s - B_s\big) \vee \tilde{M}_t - \tilde{B}_t|\mathcal{F}_s^{M-B}\right] = \mathcal{L}\left[\big(M_s - B_s\big) \vee \tilde{M}_t - \tilde{B}_t|M_s - B_s\right]\,.$$

To finish, by Lemma 3.18, for every $y \geq 0$,

$$\mathcal{L}\left[\big(M_s - B_s\big) \vee \tilde{M}_t - \tilde{B}(t) \mid M_s - B_s = y\right] = \mathcal{L}[y \vee M_t - B_t]\,.$$

Hence it suffices to check that, for $y \geq 0$

$$\mathcal{L}[y \vee M_t - B_t] = \mathcal{L}[\,|y + B_t|\,]\,.$$

Let $W_u = B_{t-u} - B_t$ for $0 \leq u \leq t$, then $W = \{W_u\}_{0 \leq u \leq t}$ is a standard Brownian motion by Theorem 3.2 (ii). Let $M_t^W = \max_{0 \leq u \leq t} W_u$, then almost surely,

$$y \vee M_t - B_t = \big(y - B_t\big) \vee \big(M_t - B_t\big) = \big(y + W_t\big) \vee M_t^W .$$

For any $a > 0$, by reflection priciple,

$$\begin{aligned}
&\mathbb{P}_0\left(\big(y + W_t\big) \vee M_t^W > a\right) \\
&= \mathbb{P}_0\left(y + W_t > a\right) + \mathbb{P}_0\left(W_t < a - y, T_a^W < t\right) \\
&= \mathbb{P}_0\left(y + W_t > a\right) + \mathbb{P}_0\left(W_t > a + y\right) \\
&= \mathbb{P}_0\left(y + W_t > a\right) + \mathbb{P}_0\left(-W_t > a + y\right) = \mathbb{P}_0\left(|y + W_t| > a\right) .
\end{aligned}$$

□

So the desired result follows.

Exercise 3.3. For each $t \geq 0$, let

$$\lambda(t) = \arg\max_{0 \leq s \leq t} B_s,$$

notice that $\lambda(t)$ is well-defined, since almost surely all the local maximum of Brownian motion are distinct. Show that for $s \leq t$,

$$\mathbb{P}(\lambda(t) \leq s) = \frac{2}{\pi} \arcsin \sqrt{\frac{s}{t}} .$$

C Zero points of Brownian Motion

Let $B = \{B_t\}_{t \geq 0}$ be a standard, one-dimensional Brownian motion. In this subsection we will investigate the (random) set

$$\mathcal{Z} = \{t \geq 0 : B_t = 0\} . \tag{3.22}$$

Because the paths of B is continuous, the set $\mathcal{Z}$ is closed. Furthermore, With probability one, the Lebesgue measure of $\mathcal{Z}$ is 0, since Fubini's theorem implies that the expected Lebesgue measure of $\mathcal{Z}$ is 0:

$$\mathbb{E}|\mathcal{Z}| = \mathbb{E} \int_0^\infty 1_{\{B_t = 0\}}\, \mathrm{d}t = \int_0^\infty \mathbb{P}\left(B_t = 0\right) \mathrm{d}t = 0 .$$

where $|\mathcal{Z}|$ denotes the Lebesgue measure of $\mathcal{Z}$. Observe that for any fixed (non-random) $t > 0$, the probability that $t \in \mathcal{Z}$ is 0, because $\mathbb{P}(B_t = 0) = 0$. Hence, because $\mathbb{Q}_+$ is countable,

$$\mathbb{P}\left(\mathbb{Q}_+ \cap \mathcal{Z} \neq \emptyset\right) = 0\,.$$

First, we introduce the famous *arcsine law.* Define the last zero point of B in $[0, t]$ as L_t, that is

$$L_t := \sup\{\mathcal{Z} \cap [0, t]\} = \sup\{s \in [0, t] : B_s = 0\}\,. \tag{3.23}$$

Theorem 3.39 (Arcsin law)**.** *For any $s \in [0, t]$,*

$$\mathbb{P}_0(L_t \leq s) = \mathbb{P}_0(\mathcal{Z} \cap (s, t] = \emptyset) = \frac{2}{\pi} \arcsin \sqrt{\frac{s}{t}}\,. \tag{3.24}$$

Proof. By the continuity of paths, we have $\{L_t \leq s\} = \{\mathcal{Z} \cap (s, t] = \emptyset\} = \{B_{s+u} > 0 \text{ for all } 0 < u \leq t - s\} \cup \{B_{s+u} < 0 \text{ for all } 0 < u \leq t - s\}$. Thus by Markov property,

$$\mathbb{P}_0(L_t \leq s | \mathcal{F}_s^B) = \mathbb{P}_{B_s}(T_0 > t - s)\,.$$

So

$$\mathbb{P}_0\left(L_t \leq s\right) = \int p_s(0, x)\mathbb{P}_x\big(T_0 > t - s\big)\,\mathrm{d}x\,.$$

Note that, by symmetry, $\mathbb{P}_x(T_0 > t - s) = \mathbb{P}_0(T_{|x|} > t - s)$, the integral can be rewritten as

$$\int p_s(0, x)\mathbb{P}_0\big(|B(t - s)| \leq |x|\big)dx = \int p_s(0, x)\mathbb{P}\big(\sqrt{t - s}|Z_1| \leq |x|\big)dx\,,$$

where Z_1 has distribution $N(0, 1)$. Let Z_2 has identical distribution and independent with Z_1, then

$$\int p_s(0, x)\mathbb{P}\big(\sqrt{t - s}|Z_1| \leq |x|\big)dx = \mathbb{P}\left(\sqrt{t - s}|Z_1| \leq \sqrt{s}|Z_2|\right)\,,$$

letting $(Z_1, Z_2) = (R\cos\Theta, R\sin\Theta)$, we konw that Θ uniformly distributed in $[0, 2\pi]$, (and R has a *Rayleigh distribution*), thus

$$\begin{aligned}&\mathbb{P}\left(\sqrt{t - s}|Z_1| \leq \sqrt{s}|Z_2|\right) = \mathbb{P}\left(\sqrt{t - s}|\sin\Theta| \leq \sqrt{s}|\cos\Theta|\right)\\&= \mathbb{P}\left(|\tan\Theta| \leq \sqrt{\frac{s}{t - s}}\right) = \frac{2}{\pi}\arcsin\sqrt{\frac{s}{t}}\,.\end{aligned}$$

We complete the proof. $\square$

We will give a second proof of Theorem 3.26 using arcsin law.

Corollary 3.40. *Let* $\sigma_0 := \inf\{t > 0 : B_t = 0\}$*, then* $\mathbb{P}_0(\sigma_0 = 0) = 1$*.*

Proof. By the arcsine law we have, for any $t > 0$,

$$\mathbb{P}_0(0 < L_t < t) = 1\,.$$

Thus $\sigma_0 \le t$ a.s. Clearly $\sigma_0 = 0$ a.s.. $\square$

Theorem 3.41. *Almost surely, $\mathcal{Z}$ is a perfect set.*[1]

Proof. As noted earlier, $\mathcal{Z}$ is closed from path-continuity. We divided Z into three parts:

$$\mathcal{Z}(\omega) = \{0\} \cup C_-(\omega) \cup C_+(\omega)\,,$$

where

$$C_-(\omega) = \{t > 0 : \exists\, t_n \uparrow t \ \text{ s.t. } B(t_n, \omega) = 0\} \quad \text{and}\,,$$
$$C_+(\omega) = \{t > 0 : \exists\, \delta > 0 \ \text{ s.t. } B(s,\omega) \neq 0, \forall\, s \in (t-\delta, t),\ \text{but } B(t,\omega) = 0\}\,.$$

Clearly, $C_-(\omega) \subset d(\mathcal{Z}(\omega))$, it suffices to show that $C_+(\omega) \subset d(\mathcal{Z}(\omega))$ almost surely. Now, fix a rational number $q > 0$, and define ν_q by

$$\nu_q = \inf\{t \ge q : B_t = 0\}\,,$$

then one can see that $C_+(\omega) \subset \{\nu_q(\omega) : q \in \mathbb{Q}_+\}$. It is enough to show that, almost surely, $\nu_q \in d(\mathcal{Z})$. By the strong Markov Property, the process

$$B(\nu_q + t) - B(\nu_q) = B(\nu_q + t)$$

is a standard Brownian motion. Consequently, by Corollary 3.40,

$$\inf\{t > 0 : B(\nu_q + t) = 0\} = 0 \qquad \text{a.s.}$$

Thus $\nu_q \in d(\mathcal{Z})$ almost surely, as required. $\square$

[1] In general topology, a subset of a topological space is perfect if it is closed and has no isolated points.

Remark 3.17. It can be shown that every compact perfect set of Lebesgue measure zero is homeomorphic to the *Cantor set*. Thus, with probability one, the set of zeros of the Brownian path B_t in the unit interval is a homeomorphic image of the Cantor set.

Fractal dimension* How "big" is the set $\mathcal{Z}$? To discuss this we need to discuss the notion of a dimension of a set. There are two similar notions of dimension, *Hausdorff dimension* and *box dimension*, which can give fractional dimensions to sets. (There is a phrase "fractal dimension" which is used a lot in scientific literature. As a rule, the people who use this phrase are not distinguishing between Hausdorff and box dimension and could mean either one.) The notion of dimension we will discuss here will be that of box dimension, but all the sets we will discuss have Hausdorff dimension equal to their box dimension.

Suppose we have a bounded set A in d -dimensional space $\mathbb{R}^d$. Suppose we cover A with d -dimensional balls of diameter ϵ. How many such balls are needed? If A is a line segment of length 1 (one-dimensional set), then ϵ^{-1} such balls are needed. If A is a two-dimensional square, however, on the order of ϵ^{-2} such balls are needed. One can see that for a standard k -dimensional set, we need ϵ^{-k} such balls. This leads us to define the (box) dimension of the set A to be the number D such that for small ϵ the number of balls of diameter ϵ needed to cover A is on the order of ϵ^{-D}.

Example 3.8. Consider the fractal subset of $[0,1]$, the Cantor set. The Cantor set A can be defined as a limit of approximate Cantor sets A_n. We start with $A_0 = [0,1]$. The next set A_1 is obtained by removing the open middle interval $(1/3, 2/3)$, so that

$$A_1 = \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right]$$

The second set A_2 is obtained by removing the middle thirds of the two intervals in A_1, hence

$$A_2 = \left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right]$$

In general A_{n+1} is obtained from A_n by removing the "middle third" of each interval. The Cantor set A is then the limit of these sets A_n

$$A = \bigcap_{n=1}^{\infty} A_n$$

Note that A_n consists of 2^n intervals each of length 3^{-n}. Suppose we try to cover A by intervals of length 3^{-n}

$$\left[\frac{k-1}{3^n}, \frac{k}{3^n}\right]$$

We need 2^n such intervals. Hence the dimension D of the Cantor set is the number such that $2^n = \left(3^{-n}\right)^{-D}$, i.e.,

$$D = \frac{\ln 2}{\ln 3} \approx 0.631$$

Now consider the set $\mathcal{Z}$ and consider $\mathcal{Z}_1 = \mathcal{Z} \cap [0,1]$. We will try to cover $\mathcal{Z}_1$ by one-dimensional balls (i.e., intervals) of diameter (length) $\epsilon = 1/n$. For ease we will consider the n intervals

$$\left[\frac{k-1}{n}, \frac{k}{n}\right], \quad k = 1, 2, \ldots n$$

How many of these intervals are needed to cover $\mathcal{Z}_1$? Such an interval is needed if $\mathcal{Z}_1 \cap [(k-1)/n, k/n] \neq \emptyset$. Note that, by arcsince law,

$$P(k,n) = \mathbb{P}\left(\mathcal{Z}_1 \cap \left[\frac{k-1}{n}, \frac{k}{n}\right] \neq \emptyset\right) = 1 - \frac{2}{\pi} \arctan \sqrt{k-1}$$

Therefore, the expected number of the intervals needed to cover Z_1 looks like

$$\sum_{k=1}^{n} P(k,n) = \sum_{k=1}^{n} \left[1 - \frac{2}{\pi} \arctan \sqrt{k-1}\right]$$

To estimate the sum, we need to consider the Taylor series for arctan $(1/t)$ at $t = 0$ (which requires remembering the derivative of arctan),

$$\arctan\frac{1}{t} = \frac{\pi}{2} - t + O\left(t^2\right)$$

In other words, for x large,

$$\arctan x \approx \frac{\pi}{2} - \frac{1}{x}$$

Hence

$$\sum_{k=1}^{n} P(k,n) \approx 1 + \sum_{k=2}^{n} \frac{2}{\pi\sqrt{k-1}} \approx \frac{2}{\pi}\int_1^n (x-1)^{-1/2}dx \approx \frac{4}{\pi}\sqrt{n}$$

Hence it takes on the order of $\sqrt{n}$ intervals of length $1/n$ to cover Z_1, or, in other words,

Theorem. *The fractal dimension of the zero set $\mathcal{Z}$ is $1/2$.*

3.8 The law of the iterated logarithm

In this section, we will always suppose $\{B_t\}_{t\geq 0}$ is a standard one-dimensional Brownian motion. Although at any given time t and for any open set $U \subset \mathbb{R}$ the probability of the event $\{B_t \in U\}$ is positive, over a long time Brownian motion cannot grow arbitrarily fast. We have seen in Example 3.4 that,

$$\lim_{t\to\infty} \frac{B(t)}{t} \to 0 \quad \text{a.s.}.$$

Whereas Proposition 3.6 ensures that

$$\limsup_{t\to\infty} \frac{|B(t)|}{\sqrt{t}} = \infty \quad \text{a.s.}.$$

It is therefore natural to ask for the asymptotic smallest upper envelope of the Brownian motion, i.e. for a function $\psi : (1,\infty) \to \mathbb{R}$ such that

$$\limsup_{t\to\infty} \frac{B(t)}{\psi(t)} = 1\,.$$

The *law of the iterated logarithm* (whose name comes from the answer to this question but is by now firmly established for this type of upper-envelope results) provides such a 'gauge' function, which determines the almost-sure *asymptotic growth* of a Brownian motion.

Theorem 3.42 (Law of the Iterated Logarithm for Brownian motion)**.** *Let* $\{B_t\}_{t\geq 0}$ *be a standard one-dimensional Brownian motion. Then, almost surely,*

$$\limsup_{t\to\infty} \frac{B(t)}{\sqrt{2t\log\log t}} = 1\,.$$

Proof. Upper bound. The key to the proof is

$$\begin{aligned}\mathbb{P}_0\left(M_t > a\right) = 2\mathbb{P}_0(B_t > a) = 2\int_{\frac{a}{\sqrt{t}}}^{\infty} \frac{1}{2\pi}\exp\left\{-\frac{y^2}{2}\right\}\mathrm{d}y\,. \\ \leq \frac{\sqrt{t}}{\pi a}\exp\left\{-\frac{a^2}{2t}\right\}\,,\end{aligned}$$

in the last step we used that

$$\int_x^{\infty}\exp\left\{-\frac{y^2}{2}\right\}\mathrm{d}y \leq \frac{1}{x}\exp\left\{-\frac{x^2}{2}\right\}\,.$$

So, taking $a = \sqrt{tf(t)}$, we get

$$\mathbb{P}_0\left(M_t > \sqrt{tf(t)}\right) \leq \frac{1}{\pi\sqrt{f(t)}}\exp\left\{-\frac{f(t)}{2}\right\} \tag{3.25}$$

If we choose $f(t) = 2(1+\epsilon)^2\log t$, then (3.25) becomes

$$\mathbb{P}_0\left(M_t > (1+\epsilon)\sqrt{2t\log t}\right) \leq \frac{1}{\pi(1+\epsilon)\sqrt{2\log t}}\; t^{-(1+\epsilon)^2}\,.$$

By Borel-Cantelli lemma, almose surely,

$$\limsup_{n\to\infty}\frac{M_n}{\sqrt{2n\log n}} \leq 1+\epsilon\,.$$

If we replace log by log log, i.e., $f(t) = 2(1+\epsilon)^2 \log\log t$, then we get that

$$\mathbb{P}_0\left(M_t > (1+\epsilon)\psi(t)\right) \leq \frac{1}{\pi(1+\epsilon)\sqrt{2\log\log t}} \, (\log t)^{-(1+\epsilon)^2} .$$

Let $\alpha > 1$ and take $t = \alpha^n$. By Borel-Cantelli lemma, almose surely,

$$\limsup_{n\to\infty} \frac{M(\alpha^n)}{\psi(\alpha^n)} \leq 1+\epsilon . \tag{3.26}$$

For $t \in [\alpha^n, \alpha^{n+1}]$, we have

$$\frac{B(t)}{\psi(t)} \leq \frac{M(\alpha^{n+1})}{\psi(\alpha^{n+1})} \, \frac{\psi(\alpha^{n+1})}{\psi(\alpha^n)} .$$

Letting $t \to \infty$, then we get

$$\limsup_{t\to\infty} \frac{B(t)}{\sqrt{2t\log\log t}} \leq (1+\epsilon)\sqrt{\alpha} .$$

Since ϵ and α is arbitary, we get the upper bound.

Lower bound. To show the lower bound, firstly we note that

$$\begin{aligned}\mathbb{P}_0(B_t \geq (1-\epsilon)\psi(t)) &= \int_{(1-\epsilon)\sqrt{2\log\log t}} \frac{1}{2\pi} e^{-y^2/2} \mathrm{d}y \\ &\sim \frac{1}{\pi(1-\epsilon)\sqrt{2\log\log t}} \, (\log t)^{-(1-\epsilon)^2}\end{aligned}$$

But we can not use the second Borel-Cantelli lemma, since B_t are not independent. So, to get independent events, we choose a strictly increasing sequence $\{t_n\}$ tending to infinity, and

$$\begin{aligned}&\mathbb{P}_0(B(t_{n+1}) - B(t_n) \geq (1-\epsilon)\psi(t_{n+1}-t_n)) \\ &\sim \frac{1}{\pi(1-\epsilon)\sqrt{2\log\log(t_{n+1}-t_n)}} \, [\log(t_{n+1}-t_n)]^{-(1-\epsilon)^2}\end{aligned}$$

Let $\alpha > 1$ and $t_n = \alpha^n$. It follows from the second Borel-Cantelli lemma that, almost surely,

$$B(\alpha^{n+1}) - B(\alpha^n) \geq (1-\epsilon)\psi(\alpha^{n+1}-\alpha^n) \quad \text{i.o..}$$

Therefore, alomst surely,

$$\limsup_{n\to\infty} \frac{B(\alpha^{n+1})}{\psi(\alpha^{n+1})} \geq (1-\epsilon) \lim_{n\to\infty} \frac{\psi(\alpha^{n+1}-\alpha^n)}{\psi(\alpha^{n+1})} + \limsup_{n\to\infty} \frac{B(\alpha^n)}{\psi(\alpha^n)} \frac{\psi(\alpha^n)}{\psi(\alpha^{n+1})}$$
$$\geq (1-\epsilon)\sqrt{\frac{\alpha-1}{\alpha}} - \frac{1}{\alpha}\,.$$

In the last inequality we used a consequence of (3.26):

$$\limsup_{n\to\infty} \frac{B(\alpha^n)}{\psi(\alpha^n)} \geq \liminf_{n\to\infty} \frac{B(\alpha^n)}{\psi(\alpha^n)} \geq -1\,.$$

Thus we get

$$\limsup_{t\to\infty} \frac{B(t)}{\sqrt{2t\log\log t}} \geq \limsup_{n\to\infty} \frac{B(\alpha^{n+1})}{\psi(\alpha^{n+1})} \geq (1-\epsilon)\sqrt{\frac{\alpha-1}{\alpha}} - \frac{1}{\alpha}\,.$$

Letting $\alpha \to \infty$ and $\epsilon \downarrow 0$, the desired result follows. □

Remark 3.18. By symmetry and time inversion, the following propositions are equibalent: Almost surely,

(i) $\displaystyle\limsup_{t\to\infty} \frac{B(t)}{\sqrt{2t\log\log t}} = 1$, (ii) $\displaystyle\liminf_{t\to\infty} \frac{B(t)}{\sqrt{2t\log\log t}} = -1$,

(iii) $\displaystyle\limsup_{t\downarrow 0} \frac{B(t)}{\sqrt{2t\log\log(1/t)}} = 1$, (iv) $\displaystyle\liminf_{t\downarrow 0} \frac{B(t)}{\sqrt{2t\log\log(1/t)}} = -1$.

Remark 3.19. It's easy to see that Theorem 3.42 is euivalent to Theorem 1.37. (In fact, the proof of these two are biscally the same.) Since, almost surely,

$$\limsup_{t\to\infty} \frac{|B_t - B_{[t]}|}{\sqrt{2t\log\log t}} = 0\,.$$

To see this, note that $|B_t - B_{[t]}| \leq \xi_{[t]}$, where

$$\xi_n \coloneqq \max_{n\leq t\leq n+1} |B_t - B_n| \quad \text{for all } n\,.$$

Clearly, ξ_n are i.i.d., and

$$\limsup_{n\to\infty} \frac{\xi_n}{\sqrt{2n\log\log n}} = 0\,.$$

3.9 The martingale property

In the previous section we have taken a particular feature of Brownian motion, the Markov property, and strong Markov property. In this section we follow a similar plan, taking a different feature of Brownian motion, the martingale property, as a starting point.

Lemma. *Let $\{X_t\}_{t\geq 0}$ be a right continuous martingale adapted to a right continuous filtration $\{\mathcal{F}_t\}_{t\geq 0}$. If T is a bounded stopping time of $\{\mathcal{F}_t\}_{t\geq 0}$, then*

$$\mathbb{E}X_T = \mathbb{E}X_0 \,.$$

Theorem 3.43. *Let $\{B_t\}_{t\geq 0}$ be a one-dimensional Brownian motion starting at x. Then with respect to the σ-fields $\{\mathcal{F}_t^+\}_{t\geq 0}$ defined in (3.14),*

(i) $\{B_t\}_{t\geq 0}$ is a continuous martingale ;

(ii) $\{B_t^2 - t\}_{t\geq 0}$ is a continuous martingale ;

(iii) $\left\{\exp\left(\theta B_t - \frac{\theta^2}{2}t\right)\right\}_{t\geq 0}$ is a martingale (called exponential martingale) for any $\theta \in \mathbb{R}$ (or indeed, for $\theta \in \mathbb{C}$).

Proof. The Markov property implies that for any $t \geq s$,

$$\mathbb{E}_x\left(B_t|\mathcal{F}_s^+\right) = B_s + \mathbb{E}_x(B_t - B_s|\mathcal{F}_s^+) = B_s \,.$$

so (i) follows. To show (ii), writing $B_t^2 = (B_s + B_t - B_s)^2$ we have

$$\begin{aligned}\mathbb{E}_x\left(B_t^2|\mathcal{F}_s^+\right) &= B_s^2 + \mathbb{E}_x\left((B_t - B_s)^2|\mathcal{F}_s^+\right) \\ &= B_s^2 + (t-s)\end{aligned}.$$

To prove (iii), bringing $\exp(\theta B_s)$ outside

$$\begin{aligned}\mathbb{E}_x\left(\exp\{\theta B_t\}|\mathcal{F}_s^+\right) &= \exp\{\theta B_s\}\,\mathbb{E}_x\left(\exp\{\theta(B_t - B_s)\}|\mathcal{F}_s^+\right) \\ &= \exp\{\theta B_s\}\exp\left\{\theta^2(t-s)/2\right\} \,.\end{aligned}$$

We complete the proof. □

Remark 3.20. The simple fact (ii) is a pointer to the development of stochastic integrals; once that theory is developed, we shall be in a position to prove the following startling converse to (ii):

Theorem (Lévy). *Let $\{X_t\}_{t\geq 0}$ be a continuous martingale with respect to $\{\mathcal{F}_t\}_{t\geq 0}$, $X_0 = 0$ and suppose that*

$$\{X_t^2 - t\}_{t\geq 0} \quad \text{is a martingale with respect to } \{\mathcal{F}_t\}_{t\geq 0}\,.$$

Then X is a standard $\{\mathcal{F}_t\}$-Brownian motion.

Remark 3.21. These exponential martingales are extremely useful in many ways: we use them to compute the Brownian first-passage distribution to a level, and we derive the law of the iterated logarithm using them. One small point to note here in connection with the exponential martingales is that if we define the Hermite polynomials $H_n(t,x)$ by

$$\exp\left\{\theta x - \frac{\theta^2}{2}t\right\} := \sum_{n\geq 0} \frac{\theta^n}{n!} H_n(t,x)\,.$$

Then, for $0 \leq s \leq t$, (iii) implies that

$$\sum_{n\geq 0} \frac{\theta^n}{n!}\mathbb{E}_x\left[H_n\left(t, B_t\right)|\mathcal{F}_s^+\right] = \sum_{n\geq 0} \frac{\theta^n}{n!} H_n\left(s, B_s\right)\,.$$

so, by comparing coefficients of θ^n, we deduce that for each n, $\{H_n\left(t, B_s\right)\}_{t\geq 0}$ is a martingale with respect to $\{\mathcal{F}_t^+\}_{t\geq 0}$. It's easy to check that $H_1(t,x) = x$ and $H_2(t,x) = x^2 - t$. So in fact, (iii) implies (i) and (ii).

A Exit probabilities and expected exit times

We now use the martingale property and the optional stopping theorem to obtain exit probabilities and expected exit times for a linear Brownian motion.

Theorem 3.44 (Exit probability). *$\{B_t\}_{t\geq 0}$ is a linear Brownian motion starting at $x \in (a,b)$. Let $\tau \equiv \tau_{(a,b)} := \inf\{t \geq 0 : B_t \notin (a,b)\}$, (clearly almost surely $\tau = T_a \wedge T_b$), then*

$$\begin{aligned} \mathbb{P}_x\big(B_\tau = a\big) = \mathbb{P}_x\big(T_a < T_b\big) = \frac{x-a}{b-a}\,, \\ \mathbb{P}_x\big(B_\tau = b\big) = \mathbb{P}_x\big(T_b < T_a\big) = \frac{b-x}{b-a}\,. \end{aligned} \tag{3.27}$$

Proof. Note that $|B_{\tau\wedge t}| \leq |a| + |b|$ for each t, by optional stopping theorem we have that

$$\mathbb{E}_x(B_{\tau\wedge t}) = \mathbb{E}_x(B_0) = x\,.$$

Letting $t \to \infty$ and using the bounded convergence theorem, it follows that

$$\begin{aligned} a\,\mathbb{P}_x\big(B_\tau = a\big) + b\,\mathbb{P}_x\big(B_\tau = b\big) = x\,, \\ \mathbb{P}_x\big(B_\tau = a\big) + \mathbb{P}_x\big(B_\tau = b\big) = 1\,. \end{aligned}$$

Solving this we obtain (3.27). □

Lemma 3.45. *Under $\mathbb{P}_x$, τ has a moment generating function,i.e., $\mathbb{E}_x\, e^{\theta\tau}$ is finite for some $\theta > 0$. In particular, it has finite moments of all orders.*

Proof. We have the crude inequality

$$\sup_{x\in(a,b)} \mathbb{P}_x(\tau > 1) \leq \sup_{x\in(a,b)} \mathbb{P}_x(B_1 \in (a,b))\,.$$

The right side can be expressed in terms of a normal distribution but it is sufficient to see that it is strictly less than 1, and we denote it by δ.

The next step is a basic argument using the Markovian character of the process. For any $x \in (a,b)$ and $n \geq 1$,

$$\mathbb{P}_x(\tau > n) \leq \mathbb{P}_x\left(B_1, \cdots, B_n \in [a,b]\right) \leq \delta^n\,, \text{ for all } n\,.$$

In fact the argument above yields more. For any ϵ such that $e^\epsilon\delta < 1$, we have

$$\mathbb{E}_x\, e^{\epsilon\tau} \leq \sum_{n=1}^{\infty} e^{\epsilon n}\, \mathbb{P}_x(n-1 < \tau \leq n) \leq \sum_{n=1}^{\infty} e^{\epsilon n}\delta^{n-1} < \infty\,.$$ □

Theorem 3.46 (Mean exit time)**.** *$\{B_t\}_{t\geq 0}$ is a linear Brownian motion starting at $x \in (a, b)$. Let $\tau \equiv \tau_{(a,b)} := \inf\{t \geq 0 : B_t \notin (a, b)\}$, then*

$$\mathbb{E}_x \tau = (x - a)(b - x)\,.$$

Proof. Note that

$$|B^2_{\tau\wedge t} - \tau \wedge t| \leq |a| \vee |b| + \tau\,, \text{ for all } t$$

by optional stopping theorem and domainted convergence theorem, we have that

$$\mathbb{E}_x(B^2_\tau - \tau) = \mathbb{E}_x(B(0)^2) = x^2$$

Thus

$$\mathbb{E}_x \tau = \mathbb{E}_x\, B^2_\tau - x^2 = (x - a)(b - x)\,,$$

as required, and in order to compute $\mathbb{E}_x B^2_\tau$ we used (3.27). □

Theorem 3.47. *Let $a \geq 0$, then $\mathbb{E}_0\, \mathrm{e}^{-\lambda T_a} = \mathrm{e}^{-a\sqrt{2\lambda}}$ for $\lambda \geq 0$.*

Proof. Using exponential martingales and we have

$$\mathbb{E}_0 \exp\left\{\theta B_{T_a\wedge t} - \frac{\theta^2}{2} T_a \wedge t\right\} = 1\,.$$

Taking $\theta = \sqrt{2\lambda}$, letting $t \to \infty$ and using the bounded convergence theorem (since $B_{T_a\wedge t} \leq a$ $\mathbb{P}_0$-a.s.) gives

$$\mathbb{E}_0 \exp\left\{\sqrt{2\lambda}a - \lambda T_a\right\} = 1\,.$$

So the desired result follows. □

Theorem 3.48. *For $x \in (a, b)$, the Laplace transform of τ is given by:*

$$\mathbb{E}_x\, \mathrm{e}^{-\lambda\tau} = \frac{\sinh\sqrt{2\lambda}(b - x) + \sinh\sqrt{2\lambda}(x - a)}{\sinh\sqrt{2\lambda}(b - a)} \quad \text{for } \lambda > 0\,.$$

Here sinh denotes the "hyperbolic sine" function, as cosh denotes the "hyperbolic cosine" function:

$$\sinh\theta = \frac{e^{\theta} - e^{-\theta}}{2}\ , \ \cosh\theta = \frac{e^{\theta} + e^{-\theta}}{2}\ .$$

Proof. Observe that

$$\begin{aligned}
\mathbb{E}_x\, \mathrm{e}^{-\lambda T_a} &= \mathbb{E}_x\, \mathrm{e}^{-\lambda\tau} 1_{\{T_a<T_b\}} + \mathbb{E}_x\, \mathrm{e}^{-\lambda T_a} 1_{\{T_a>T_b\}} \\
&= \mathbb{E}_x\, \mathrm{e}^{-\lambda\tau} 1_{\{T_a<T_b\}} + \mathbb{E}_x \left[\mathrm{e}^{-\lambda\tau}\mathbb{E}_x \left(\mathrm{e}^{-\lambda(T_a-\tau)} 1_{\{T_a>\tau\}} | \mathcal{F}_T\right)\right] \\
&= \mathbb{E}_x\, \mathrm{e}^{-\lambda\tau} 1_{\{T_a<T_b\}} + \mathbb{E}_x\, \mathrm{e}^{-\lambda\tau} 1_{\{T_a>T_b\}}\, \mathbb{E}_b\, \mathrm{e}^{-\lambda T_a}\ .
\end{aligned}$$

In the last equality we used the strong Markov property. Interchange the roles of a and b to get a second equation, use Theorem 3.47, and solve to get

$$\begin{aligned}
\mathbb{E}_x\, \mathrm{e}^{-\lambda\tau} 1_{\{T_a<T_b\}} &= \frac{\sinh\sqrt{2\lambda}(b-x)}{\sinh\sqrt{2\lambda}(b-a)}\ ; \\
\mathbb{E}_x\, \mathrm{e}^{-\lambda\tau} 1_{\{T_b<T_a\}} &= \frac{\sinh\sqrt{2\lambda}(x-a)}{\sinh\sqrt{2\lambda}(b-a)}\ .
\end{aligned}$$

Then the desired result follows. □

B A general method

To obtain comprehensive information regarding the joint distribution of the time and place of exit from (a, b), we use the more powerful exponential martingale. Let us denote the exponential martingale (with parameter θ) by $\{M_t\}_{t\geq 0}$, then for every $x \in (a, b)$:

$$\mathbb{E}_x\, M_{t\wedge\tau} = \mathbb{E}_x\, M_0 = \mathrm{e}^{\theta x}\ .$$

Since $\{|M_{t\wedge\tau}|\}$ is bounded, we obtain by dominated convergence that

$$\mathbb{E}_x M_\tau = \mathbb{E}_x\, \exp\left\{\theta B_\tau - \frac{\theta^2\tau}{2}\right\} = \mathrm{e}^{\theta x}\ .$$

Putting

$$f_a(x) = \mathbb{E}_x\left(\exp\left\{-\frac{\theta^2\tau}{2}\right\}1_{\{B_\tau=a\}}\right),$$
$$f_b(x) = \mathbb{E}_x\left(\exp\left\{-\frac{\theta^2\tau}{2}\right\}1_{\{B_\tau=b\}}\right).$$

We have the equation

$$e^{\theta x} = e^{\theta a}f_a(x) + e^{\theta b}f_b(x), \quad x \in (a,b)\,. \tag{3.28}$$

We have also the equation

$$f_a(x) + f_b(x) = \mathbb{E}_x\left(\exp\left\{-\frac{\theta^2\tau}{2}\right\}\right). \tag{3.29}$$

Unlike the situation in Theorem 3.44, these two equations do not yield the three unknowns involved. There are several ways of circumventing the difficulty. The first one is to uncover a third hidden equation in (3.28):

$$e^{-\theta x} = e^{-\theta a}f_a(x) + e^{-\theta b}f_b(x), \quad x \in (a,b)\,. \tag{3.30}$$

Combine (3.28) and (3.30) can we slove $f_a(x)$ and $f_b(x)$. But this quickie method depends on a lucky quirk. By contrast, the method developed here, though much longer, belongs to the mainstream of probabilistic analysis and is of wide applicability. It is especially charming in the setting of $\mathbb{R}^1$.

We begin with the observation that if x is the midpoint of the interval (a,b), then $f_a(x) = f_b(x)$ by symmetry, so that in this case (3.28) is solvable for $f_a(x)$. Changing the notation we fix x and consider the interval $(x-h, x+h)$. We obtain from (3.28) that

$$f_{x-h}(x) = \frac{e^{\theta x}}{e^{\theta(x-h)} + e^{\theta(x+h)}} = \frac{1}{e^{\theta h} + e^{-\theta h}}$$

and consequently by (3.29),

$$\mathbb{E}_x\left(\exp\left\{-\frac{\theta^2}{2}\tau_{(x-h,x+h)}\right\}\right) = \frac{1}{\cosh(\theta h)} \tag{3.31}$$

With this foot in the door, we will push on to calculate $f_a(x)$. Recall $x \in (a, b)$, hence for sufficiently small $h > 0$ we have $[x - h, x + h] \subset (a, b)$ and so

$$\tau_{(x-h,x+h)} < \tau_{(a,b)}$$

We shall denote $\tau_{(x-h,x+h)}$ by τ_h below. Now starting at x, the path upon its exit from $(x - h, x + h)$ will be at $x - h$ or $x + h$ with probability $1/2$ each. From the instant τ_h onward, the path moves as if it started at these two new positions by the strong Markov property, because τ_h is a stopping time. This verbal description is made symbolic below.

$$\mathbb{E}_x \left(\exp \left\{ -\frac{\theta^2 \tau}{2} \right\} 1_{\{B_\tau = a\}} \right) = \mathbb{E}_x \left[\exp \left\{ -\frac{\theta^2 \tau_h}{2} \right\} \mathbb{E}_{B(\tau_h)} \left(-\frac{\theta^2 \tau}{2} 1_{\{B_\tau = a\}} \right) \right] .$$

Then,

$$f_a(x) = \mathbb{E}_x \left(\exp \left\{ -\frac{\theta^2 \tau_h}{2} \right\} \right) \frac{1}{2} \left[f_a(x - h) + f_a(x + h) \right] . \tag{3.32}$$

Using (3.31) we may rewrite this as follows:

$$\frac{f_a(x + h) - 2 f_a(x) + f_a(x - h)}{h^2} = \frac{2 \cosh(\theta h) - 2}{h^2} f_a(x) . \tag{3.33}$$

Letting $h \downarrow 0$ we see that the left member in (3.33) converges to $\theta^2 f_a(x)$. It is also immediate from (3.32) that

$$f_a(x) < \frac{1}{2} \left[f_a(x - h) + f_a(x + h) \right] \tag{3.34}$$

valid for $a < x - h < x + h < b$. Since f_a is also bounded, (3.34) implies f_a is continuous (in fact convex) in (a, b).

To prove this we write the condition in the form $f_a(x) - f_a(x - h) \leq f_a(x + h) - f_a(x)$ from which we derive the inequalities

$$\begin{aligned} f_a(x - \nu h) - f_a(x - (\nu + 1)h) &\leq f_a(x + h) - f_a(x) \\ &\leq f_a(x + (\nu + 1)h) - f_a(x + \nu h) \end{aligned}$$

valid for every integer $\nu \geq 0$. If we add these for values of ν from $v = 0$ to $\nu = n-1$, we obtain the estimate

$$\frac{f_a(x) - f_a(x-nh)}{n} \leq f_a(x) - f_a(x-h) \leq f_a(x+h) - f_a(x) \leq \frac{f_a(x+nh) - f_a(x)}{n}\,.$$

Letting $n \to \infty$, since f_a is bounded, we get that f is continuous.

Now if f_a has a second derivative f_a'' in (a,b), then an easy exercise in calculus shows that the limit as $h \downarrow 0$ of the left member in (3.33) is equal to $f_a''(x)$. What is less easy is to show that a close converse is also true. This is known as *Schwarz's theorem* on generalized second derivative, a basic lemma in Fourier series. We state it in the form needed below.

Theorem (Schwarz's Theorem)**.** *Let f be continuous in (a,b) and suppose that*

$$\lim_{h \to 0} \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} = \varphi(x) \quad \forall x \in (a,b)$$

where $\varphi \in C(a,b)$. Then $f \in C^2(a,b)$ and $f'' = \varphi$.

Since f_a has been shown to be continuous, Schwarz's theorem applied to (3.33) yields the differential equation:

$$f_a''(x) = \theta^2 f(x), \quad x \in (a,b)\,.$$

The most general solution of this equation is given by

$$f_a(x) = Ae^{\theta x} + Be^{-\theta x} \tag{3.35}$$

where A and B are two arbitrary constants. To determine these we compute the limits of $f_a(x)$ as $x \to a$ and $x \to b$ from inside (a,b). From (3.32) and Theorem 3.44 we infer that

$$\limsup_{x \to b} f_a(x) \leq \lim_{x \to b} \mathbb{P}_x(B_\tau = a) = \lim_{x \to b} \frac{b-x}{b-a} = 0\,,$$
$$\limsup_{x \to b} f_b(x) \leq \lim_{x \to b} \mathbb{P}_x(B_\tau = b) = \lim_{x \to b} \frac{x-a}{b-a} = 1\,.$$

Since $f_a \geq 0$, the first relation above shows that $\lim_{x\to b} f_a(x) = 0$, using (3.33) we see that

$$e^{\theta b} \leq e^{\theta b} \liminf_{x\to b} f_b(x) .$$

so $\lim_{x\to b} f_b(x) = 1$. Similarly we have

$$\lim_{x\to a} f_a(x) = 1, \quad \lim_{x\to a} f_b(x) = 0$$

Thus we obtain from (3.35)

$$0 = Ae^{\theta b} + Be^{-\theta b}, \quad 1 = Ae^{\theta a} + Be^{-\theta a} .$$

Solving for A and B and substituting into (3.35), we obtain

$$f_a(x) = \frac{\operatorname{sh}\theta(b-x)}{\operatorname{sh}\theta(b-a)}, \quad f_b(x) = \frac{\operatorname{sh}\theta(x-a)}{\operatorname{sh}\theta(b-a)} .$$

Exercise 3.4. Show that for $0 \leq \theta < \pi^2/2(b-a)^2$, we have

$$\mathbb{E}_x\ \mathrm{e}^{\theta\tau_{(a,b)}} = \frac{\cos\left(\sqrt{2\theta}\left(x - \frac{a+b}{2}\right)\right)}{\cos\left(\sqrt{2\theta}\left(\frac{b-a}{2}\right)\right)}$$

Prove that $\mathbb{E}_x\ \mathrm{e}^{\theta\tau_{(a,b)}} = +\infty$ for $\theta = \pi^2/2(b-a)^2$.

C More martingales

Given the function f on $\mathbb{R}$, $f(x) = x^2$ we were able to subtract a suitable term from $f(B_t)$ to obtain a martingale. To get a feeling for what we wish to subtract in the case of a general f, we look at the analogous problem for the simple random walk $\{S_n\}_{n\geq 0}$. A straightforward calculation gives, for $f : \mathbb{Z} \to \mathbb{R}$

$$\begin{aligned}
&\mathbb{E}\left[f\left(S_{n+1}\right)|\sigma\left(S_1,\ldots,S_n\right)\right] - f\left(S_n\right) \\
&= \frac{1}{2}\left(f\left(S_n+1\right) - 2f\left(S_n\right) + f\left(S_n-1\right)\right) = \frac{1}{2}\tilde{\Delta} f\left(S_n\right) ,
\end{aligned}$$

where $\tilde{\Delta}$ is the second difference operator $\tilde{\Delta} f(x) := f(x+1) - 2f(x) + f(x-1)$. Hence

$$f(S_n) - \frac{1}{2} \sum_{k=0}^{n-1} \tilde{\Delta} f(S_k)$$

defines a (discrete time) martingale. In the Brownian motion case, one would expect a similar result with $\tilde{\Delta} f$ replaced by its continuous analogue, the Laplacian

$$\Delta f(x) = \sum_{i=1}^{d} \frac{\partial^2 f}{\partial x_i^2}.$$

Theorem 3.49. *Let $f \in C^2(\mathbb{R}^d)$, i.e., $f : \mathbb{R}^d \to \mathbb{R}$ is twice continuously differentiable. Let $\{B_t\}_{t \geq 0}$ be a d-dimensional Brownian motion. Further suppose that, for all $t \geq 0$ and $x \in \mathbb{R}^d$, we have*

$$\mathbb{E}_x |f(B_t)| < \infty \ , \ \mathbb{E}_x \int_0^t |\Delta f(B_s)| ds < \infty .$$

Then the process $\{X_t\}_{t \geq 0}$ defined by

$$X_t = f(B_t) - \frac{1}{2} \int_0^t \Delta f(B_s) ds$$

is a continuous martingale with respect to $\{\mathcal{F}_t^+\}_{t \geq 0}$ under $\mathbb{P}_x$.

Proof. For any $0 \leq s < t$, by Markov property,

$$\begin{aligned} &\mathbb{E}_x[X_t | \mathcal{F}_s^+] \\ &= \mathbb{E}_{B_s}[f(B_{t-s})] - \frac{1}{2} \int_0^s \Delta f(B_u) du - \int_0^{t-s} \mathbb{E}_{B_s} \left[\frac{1}{2} \Delta f(B_u) \right] du \end{aligned}$$

Now, using integration by parts and $\frac{1}{2} \Delta_y p(t,x,y) = \frac{\partial}{\partial t} p(t,x,y)$, where $p(t,x,y)$ is the Brownian transition densities, we find

$$\begin{aligned} &\mathbb{E}_{B_s} \left[\frac{1}{2} \Delta f(B_u) \right] = \frac{1}{2} \int_{\mathbb{R}^d} p(u, B_s, y) \Delta f(y) \, \mathrm{d}y \\ &= \frac{1}{2} \int_{\mathbb{R}^d} \Delta_y p(u, B_s, y) f(y) \, \mathrm{d}y = \int_{\mathbb{R}^d} \frac{\partial}{\partial u} p(u, B_s, y) f(y) \, \mathrm{d}y . \end{aligned}$$

Hence by Fubini's theorem,

$$
\begin{aligned}
&\int_0^{t-s} \mathbb{E}_{B_s}\left[\frac{1}{2}\Delta f(B_u)\right] \mathrm{d}u \\
&= \int_0^{t-s} \mathrm{d}u \int_{\mathbb{R}^d} \frac{\partial}{\partial u} p(u, B_s, y) f(y)\, \mathrm{d}y \\
&= \int_{\mathbb{R}^d} f(y)\, \mathrm{d}y \int_0^{t-s} \frac{\partial}{\partial u} p(u, B_s, y)\, \mathrm{d}u \\
&= \int_{\mathbb{R}^d} f(y)\, \mathrm{d}y \lim_{\epsilon \downarrow 0} \int_\epsilon^{t-s} \frac{\partial}{\partial u} p(u, B_s, y)\, \mathrm{d}u \\
&= \int_{\mathbb{R}^d} p(t-s, B_s, y) f(y)\, \mathrm{d}y - \lim_{\epsilon \downarrow 0} \int_{\mathbb{R}^d} p(\epsilon, B_s, y) f(y)\, \mathrm{d}y \\
&= \mathbb{E}_{B_s}[f(B_{t-s})] - f(B_s)\,.
\end{aligned}
$$

This confirms the martingale property. □

3.10 Dirichlet and Poisson problems in $\mathbb{R}^1$

In classical potential theory (see Kellogg [1]) there are a clutch of famous problems which had their origins in electromagnetism. We begin by stating two of these problems in Euclidean space $\mathbb{R}^d$, where d is the dimension. Let D be a nonempty bounded open set (called a "domain" when it is connected), and let ∂D denote its boundary. Let Δ denote the Laplacian, namely the differential operator

$$
\Delta = \sum_{j=1}^{d} \left(\frac{\partial}{\partial x_j}\right)^2 .
$$

A function f defined in D is called *harmonic* there iff $\Delta f = 0$ in D. This of course requires that f is twice differentiable. If f is *locally integrable* in D, namely has a finite Lebesgue integral over any compact subset of D, then it is harmonic in D if and only if the following "surface averaging property" is true.

Let $B(x,\delta)$ denote the closed ball with center x and radius δ. For each $x \in D$ and $\delta > 0$ such that $B(x,\delta) \subset D$, we have

$$f(x) = \frac{1}{\sigma(\partial B(x,\delta))} \int_{\partial B(x,\delta)} f(y)\sigma(dy) \tag{3.36}$$

where $\sigma(dy)$ is the area measure on $\partial B(x,\delta)$. This alternative characterization of harmonic function is known as Gauss's theorem and plays a basic role in probabilistic potential theory, because probability reasoning integrates better than differentiates.

Dirichlet's problem. Given D and a continuous function f on ∂D, to find a function φ which is continuous in $\bar{D}$ and satisfies:

$$\begin{aligned} \Delta\varphi &= 0 \quad \text{in } D\,, \\ \varphi &= f \quad \text{on } \partial D\,. \end{aligned} \tag{3.37}$$

Poisson's problem. Given D and a bounded continuous function f in D, to find a function φ which is continuous in $\bar{D}$ and satisfies

$$\begin{aligned} \Delta\varphi &= f \quad \text{in } D\,, \\ \varphi &= 0 \quad \text{on } \partial D\,. \end{aligned} \tag{3.38}$$

We have stated these problems in the original forms, of which there are well-known generalizations. As stated, a unique solution to either problem exists provided that the boundary ∂D is not too irregular. Since we shall treat only the one-dimensional case in this section, we need not be concerned with the general difficulties.

In $\mathbb{R}^1$, a domain is just an bounded open nonempty interval $I = (a,b)$. Its boundary ∂I consists of the two points $\{a,b\}$. Since $\Delta f = f''$, a harmonic function is just a linear function. The boundary function f reduces to two arbitrary values assigned to the points a and b, and no question of its continuity arises. Thus in $\mathbb{R}^1$ Dirichlet's problem reads as follows.

Dirichlet's Problem in $\mathbb{R}^1$. Given two arbitrary numbers $f(a)$ and $f(b)$, to find a function φ which is linear in (a,b) and continuous in $[a,b]$, so that $\varphi(a) = f(a), \varphi(b) = f(b)$.

This is a (junior) high school problem of analytic geometry. The solution is given by

$$\frac{b-x}{b-a}f(a)+\frac{x-a}{b-a}f(b)\,. \tag{3.39}$$

Now we will write down the probabilistic solution, as follows

$$\varphi(x)=\mathbb{E}_x[f(B_\tau)]\ ,\ x\in[a,b] \tag{3.40}$$

where $\tau=\tau_{(a,b)}=\inf\{t\geq 0: B_t\notin(a,b)\}$. If we evaluate the right member of (3.40) by Theorem 3.44, we see at once that it is the same as given in (3.39). But we will prove that φ is the sought solution by the general method developed in the last section, because the same pattern of proof works in any dimension. Using the notation τ_h as in "the general method", we obtain

$$\varphi(x)=\mathbb{E}_x\left[\mathbb{E}_{B(\tau_h)}f(B_\tau)\right]=\frac{1}{2}[\varphi(x-h)+\varphi(x+h)]\ , \tag{3.41}$$

for $x\in(a,b)$ and for h so that $(x-h,x+h)\subset(a,b)$. This is the one-dimensional case of Gauss's criterion for harmonicity (3.36). Since φ is bounded (so it is locally integrable), it follows from the criterion that φ is harmonic, namely linear. But we can also invoke Schwarz's Theorem to deduce this result, indeed the generalized second derivative of φ is identically zero by (3.41).

It remains to show that as $x\to a$ or b from inside (a,b), $\varphi(x)$ tends to $\varphi(a)=f(a)$ or $\varphi(b)=f(b)$ respectively. This is a consequence of the probabilistic relations below:

$$\lim_{x\to a}\mathbb{P}_x\left(\tau=T_a\right)=1\ ,\ \lim_{x\to b}\mathbb{P}_x\left(\tau=T_b\right)=1$$

which are immediate by Theorem 3.44. But since no such analogue is available in dimension >1, another proof more in the general spirit is indicated below.

Assuming (5.8), we have

$$\begin{aligned}\varphi(x)&=E^x\left\{f\left(X\left(T_a\right)\right);\tau=T_a\right\}+E^x\left\{f\left(X\left(T_b\right)\right);\tau=T_b\right\}\\&=P^x\left\{\tau=T_a\right\}f(a)+P^x\left\{\tau=T_b\right\}f(b)\end{aligned}$$

and consequently

$$\lim_{x\to a}\varphi(x) = 1\cdot f(a) + 0\cdot f(b) = f(a)\,,$$
$$\lim_{x\to b}\varphi(x) = 0\cdot f(a) + 1\cdot f(b) = f(b)\,.$$

Thus φ is continuous on $[a,b]$.

Poisson's Problem in $\mathbb{R}^1$. Given a bounded continuous function f in (a,b) to find a function φ which is continuous in $[a,b]$ such that

$$\begin{aligned} \frac{1}{2}\varphi''(x) &= -f(x) \quad \text{for } x\in(a,b)\,; \\ \varphi(a) &= \varphi(b) = 0\,. \end{aligned} \tag{3.42}$$

The constants $\frac{1}{2}$ and -1 in the differential equation are chosen for the sake of convenience, as will become apparent below. This is a simple calculus problem which can be solved by setting

$$\begin{aligned} \varphi(x) &= \int_a^x \int_a^y -2f(z)\,\mathrm{d}z\,\mathrm{d}y + cx + d \\ &= \int_a^x 2(y-x)f(y)\,\mathrm{d}y + cx + d \end{aligned}$$

and determining the constants $d=-ca$ and

$$c = \frac{1}{b-a}\int_a^b (b-y)f(y)\,\mathrm{d}y$$

by the boundary conditions $\varphi(a)=0$ and $\varphi(b)=0$. Substituting these values for c and d and rearranging we can write the solution above as

$$\varphi(x) = \int_a^b g(x,y)f(y)dy\,, \tag{3.43}$$

where

$$g(x,y) = \frac{2[(x\wedge y)-a]\,[b-(x\vee y)]}{b-a}\ , \text{ for } x,y\in(a,b)\,. \tag{3.44}$$

Note that $g(x,y)>0$ in (a,b) and $g(x,y)=g(y,x)$. We put $g(x,y)=0$ outside $(a,b)\times(a,b)$. The function g is known as the *Green's function* for (a,b) because representing the solution of (3.42) in the form (3.43) is an example of the classical method of solving differential equations by Green's functions.

Now we will write down the probabilistic solution of Poisson's problem (3.42), as follows:

$$\varphi(x)=\mathbb{E}_x\left(\int_0^{\tau} f(B_t)\,\mathrm{d}t\right) \text{ for } x\in[a,b]\,. \tag{3.45}$$

Note that the integral above may be regarded as over $(0,\tau)$ so that f need be defined in (a,b) only.

Clearly $\varphi(a)=\varphi(b)=0$. To show that φ satisfies the differential equation, we proceed by "the general method". For $x\in(a,b)$, and $(x-h,x+h)\subset(a,b)$, we have

$$\mathbb{E}_x\left(\int_0^{\tau} f(B_t)\,\mathrm{d}t\right)=\mathbb{E}_x\left(\int_0^{\tau_h} f(B_t)\,\mathrm{d}t\right)+\mathbb{E}_x\left(\int_{\tau_h}^{\tau} f(B_t)\,\mathrm{d}t\right),$$

and by strong Markov property

$$\begin{aligned}\mathbb{E}_x\left(\int_{\tau_h}^{\tau} f(B_t)\,\mathrm{d}t\right)&=\mathbb{E}_x\left[\mathbb{E}_x\left(\int_{\tau_h}^{\tau} f(B_t)\,\mathrm{d}t\,\middle|\,\mathcal{F}(\tau_h)\right)\right]\\&=\mathbb{E}_x\left[\mathbb{E}_{B(\tau_h)}\left(\int_0^{\tau} f(B_t)\,\mathrm{d}t\right)\right]\\&=\frac{1}{2}\left[\varphi(x+h)+\varphi(x-h)\right].\end{aligned}$$

Put

$$\varpi(x,h)=\mathbb{E}_x\left(\int_0^{\tau_h} f\left(B_t\right)\mathrm{d}t\right).$$

Thus

$$\varphi(x)=\varpi(x,h)+\frac{1}{2}[\varphi(x+h)+\varphi(x-h)]\,.$$

In order to use Schwartz's Theorem, We rewrite the equality above as

$$\frac{\varphi(x+h)-2\varphi(x)+\varphi(x-h)}{h^2}=-\frac{2\varpi(x,h)}{h^2}\,.$$

Firstly, we show that φ is continuous in (a,b). Without loss of generality we may suppose $f\le 0$; for the general case will follow from this case and

$f = (-f^-) - (-f^+)$. Since $f \geq 0, \varpi \geq 0$, we have

$$\varphi(x) \leq \frac{1}{2}[\varphi(x+h) + \varphi(x-h)]\,;$$

also $\varphi(x) \leq \|f\|_\infty \mathbb{E}_x\tau \leq \|f\|(b-a)^2/4$. Thus φ is continuous (in fact convex).

Next, we show that for each $x \in (a,b)$,

$$\lim_{h\downarrow 0} \frac{\varpi(x,h)}{h^2} = \lim_{h\downarrow 0} \frac{1}{h^2}\,\mathbb{E}_x\left(\int_0^{\tau_h} f\left(B_t\right)\mathrm{d}t\right) = f(x)\,.$$

It follows from Theorem 3.46 that

$$\mathbb{E}_x\left(\int_0^{\tau_h} f\left(B_0\right)\mathrm{d}t\right) = f(x)\,\mathbb{E}_x(\tau_h) = f(x)h^2\,.$$

So it suffices to show that

$$\lim_{h\downarrow 0} \frac{1}{h^2}\,\mathbb{E}_x\left(\int_0^{\tau_h} \left|f\left(B_t\right) - f\left(B_0\right)\right|\mathrm{d}t\right) = 0\,. \tag{3.46}$$

Observe that for each $t \in (0,\tau)$, $B_t \in (x-h, x+h)$, then we have

$$\frac{1}{h^2}\,\mathbb{E}_x\left(\int_0^{\tau_h} \left|f\left(B_t\right) - f\left(B_0\right)\right|\mathrm{d}t\right) \leq \sup_{|y-x|<h} |f(y) - f(x)|\,,$$

which implies (3.46). So using Schwartz's Theorem, $\varphi \in C^2(a,b)$ with

$$\varphi'' = -2f\,.$$

Finally, since

$$|\varphi(x)| \leq \|f\|\mathbb{E}_x\tau = \|f\|(x-a)(b-x)\,,$$

$\varphi(x)$ converges to zero as $x \to a$ or $x \to b$. Thus φ is continuous in $[a,b]$.

Green's function If we equate the two solutions of Poisson's problem given in (3.43) and (3.45), we obtain

$$\mathbb{E}_x\left(\int_0^{\tau} f(B_t)\,\mathrm{d}t\right) = \int_a^b g(x,y)f(y)\,\mathrm{d}y \tag{3.47}$$

for every bounded continuous f on (a, b). Let us put for $x \in \mathbb{R}$ and $A \in \mathcal{B}(\mathbb{R})$:

$$V(x, A) = \mathbb{E}_x \left(\int_0^\tau 1_{\{B_t \in A\}} \, \mathrm{d}t \right) .$$

Clearly, for every x, $V(x, \cdot)$ is a (finite) measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. By canonical method in measure theory, for any bounded Borel measurble function f we have

$$\int f(y) V(x, \mathrm{d}y) = \mathbb{E}_x \left(\int_0^\tau f(B_t) \, \mathrm{d}t \right) .$$

Then it follows from (3.47) that V is a finite transition kernel from $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ to itself, and

$$V(x, A) = \int_A g(x, y) \, \mathrm{d}y \, .$$

In other words, $V(x, \cdot)$ has $g(x, \cdot)$ as its Radon-Nikodym derivative with respect to the Lebesgue measure. The kernel V is sometimes called the potential of the Brownian motion killed at τ. It is an important object for the study of this process, since $V(x, A)$ gives the expected occupation time of A starting form x.

Exercise 3.5. Determine the measure $H(x, \cdot)$ on ∂I so that the solution to Dirichlet's problem may be written as

$$\int_{\partial I} f(y) H(x, dy)$$

The analogue in $\mathbb{R}^d$ is called the *harmonic measure* for I. It is known in the classical theory that this measure may be obtained by taking the "interior normal derivative" of $g(x, y)$ with respect to y. Find out what this means in $\mathbb{R}^1$.

Feynman—Kac functional*

As a final application of "the general method", we will treat a fairly new problem. Reversing the previous order of discussion, let us consider

$$\varphi(x) = \mathbb{E}_x \left(\exp \left\{ \int_0^\tau q(B_t) \, \mathrm{d}t \right\} f(B_\tau) \right) , \quad x \in [a, b] \tag{3.48}$$

where $\tau = \tau_{(a,b)}$ is the exit time of (a, b), q is a bounded continuous function in $[a, b]$, f as in Dirichlet's problem. Clearly, $\varphi(a) = f(a)$, $\varphi(b) = f(b)$. The exponential factor in (3.48) is called the *Feynman-Kac functional*. Let us write $e(u) = \int_0^u q(B_t)\,\mathrm{d}t$ for $u \geq 0$.

An immediate question is whether φ is finite. If $q \equiv$ a constant c, and $f \equiv 1$, then $\varphi \equiv \infty$ for sufficiently large c, by Exercise 3.4.

Chapter 4

Brownian Motion and Random Walk

The key to the developments in the section is that we can embed sums of mean zero inde- pendent random variables (see Section 8.1) and martingales (see Section 8.2) into Brownian motion and derive central limit theorems. These results lead in Section 8.3 to CLTs for stationary sequences and in Section 8.4 to the convergence of rescaled empirical distribu- tions to Brownian Bridge. Finally, in Section 8.5 we use the embedding to prove a law of the iterated logarithm.

4.1 The Invariance Principle

A Space $C[0,\infty)$

The sample spces for the Brownian motions we built in Sections 2 and 3 were, respectively, the space $\mathbb{R}^{[0,\infty)}$ of all real-valued functions on $[0,\infty)$ and a space Ω rich enough to carry a countable collection of independent, standard normal random variables. The "canonical" space for Brownian motion, the one most convenient for many future developments, is $C[0,\infty)$ the space of all continuous, real-valued functions on $[0,\infty)$ with metric

$$\begin{aligned} d(\omega_1,\omega_2) &:= \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{p_n(\omega_1-\omega_2)}{1+p_n(\omega_1-\omega_2)}, \quad \text{where} \\ p_n(\omega_1-\omega_2) &:= \max_{t\in[0,n]} |\omega_1(t)-\omega_2(t)| \quad \text{for all } n\,. \end{aligned} \tag{4.1}$$

In this section, we show how to construct a measure, called Wiener measure, on this space so that the coordinate mapping process is Brownian motion. This construction is given as the proof of Donsker's invariance principle and involves the notion of weak convergence of random walks to Brownian motion.

Lemma 4.1. *$C[0,\infty)$ equipped tith the topology induced by d defined above, is a complete, separable metric space.*

Proof. It's easy to see that d is a complete metric on $C[0,\infty)$. To see that $C[0,\infty)$ is seperable, let

$$P_Q := \{p : [0,\infty) : p \text{ is a ploynomimal with rational cofficients}\}\,.$$

Clearly, P_Q is countable. By the Stone-Weierstrass theorem, P_Q is dense in $C[0,\infty)$, and hence the desired result follows. ☐

Lemma 4.2. *The Borel σ-algebra $\mathcal{B}(C[0,\infty))$ coinsides with the σ-algebra generated by the evaluation mappings $\{\pi_t\}_{t\geq 0}$, i.e.,*

$$\sigma(\pi_t : t\geq 0) = \mathcal{B}(C[0,\infty))\,.$$

Fuermore, for any $s\geq 0$, define $\psi_s : C[0,\infty)\to C[0,\infty)$ by letting $(\psi_s\phi)(t) = \omega(t\wedge s)$, then

$$\sigma(\pi_s : 0\leq t\leq s) = \psi_s^{-1}\mathcal{B}(C[0,\infty))$$

Proof. Firstly, we show that for any $\epsilon > 0$ and $\omega\in C[0,\infty)$

$$\{\omega' : d(\omega',\omega)\leq\epsilon\} \in \sigma(\pi_t : t\geq 0)\,, \tag{4.2}$$

which implies $\mathcal{B}(C[0,\infty)) \subset \sigma(\pi_t : t \geq 0)$. To this end, observe that $\omega' \to p_n(\omega' - \omega)$ is measurable with respect to $\sigma(\pi_t : t \geq 0)$, since for any $\epsilon > 0$,

$$\{\omega' : p_n(\omega' - \omega) \leq \epsilon\} = \bigcap_{r \in [0,n] \cap \mathbb{Q}} \{|\pi_r - \omega_r| \leq \epsilon\}\,.$$

By the definition of the metric d,

$$d(\omega, \omega') = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{p_n(\omega - \omega')}{1 + p_n(\omega - \omega')}\,.$$

Combine these two, then (4.2) follows. On the other hand, for each t, π_t is continuous function on $C[0,\infty)$. So for any open subset O of $\mathbb{R}$, $\{\pi_t \in O\}$ is open, and thus

$$\{\pi_t \in O\} \in \mathcal{B}(C[0,\infty))\,.$$

From this, we get

$$\sigma(\pi_t : t \geq 0) \subset \mathcal{B}(C[0,\infty))\,.$$

We now complete the proof. □

Next, we shall provide a characterization of tightness of the probability measures on $C[0,\infty)$. To do so, we define for each $\omega \in C[0,\infty), T > 0$, and $\delta > 0$ the *modulus of continuity* on $[0,T]$ by

$$m^T(\omega, \delta) := \max_{\substack{|s-t| \leq \delta \\ 0 \leq s,t \leq T}} |\omega(s) - \omega(t)|\,.$$

It's eacy to show that $m^T(\omega,\delta)$ is non-decreasing in δ, and $\lim_{\delta \downarrow 0} m^T(\omega,\delta) = 0$ for each $\omega \in C[0,\infty)$. Besides, $m^T(\omega,\delta)$ is continuous in $\omega \in C[0,\infty)$ under the metric d. In fact, by triangular inequality, for any $\omega, \omega' \in C[0,\infty)$,

$$\begin{aligned} m^T(\omega,\delta) - m^T(\omega',\delta) &\leq \max_{0 \leq s \leq T} |\omega(s) - \omega'(s)| + \max_{0 \leq t \leq T} |\omega(t) - \omega'(t)| \\ &\leq 2p_{\lceil T \rceil}(\omega - \omega')\,. \end{aligned} \tag{4.3}$$

We shall need the following version of the Arzelà-Ascoli theorem.

Theorem 4.3 (Arzelà-Ascoli theorem)**.** *Let $A \subset C[0,\infty)$. Then A is relative compact if and only if the following two conditions hold:*

$$\sup_{\omega\in A}|\omega(0)| < \infty\,, \tag{4.4}$$

$$\lim_{\delta\downarrow 0}\sup_{\omega\in A} m^T(\omega,\delta) = 0 \quad \textit{for every } T > 0\,. \tag{4.5}$$

Remark 4.1. In fact, these two conditions are equivalent to that, for any $T > 0$, $\{\omega|_{[0,T]} : \omega \in A\}$ is *uniformly bounded* and *equicontinuous* on $[0,T]$.

Proof. Assume that the closure of A, denoted by $\bar{A}$, is compact. Since $\bar{A}$ is contained in the union of the open sets

$$G_n = \{\omega : |\omega(0)| < n\}, \quad n = 1, 2, \ldots$$

it must be contained in some particular G_n, and (4.4) follows. Given $T > 0$ and $\epsilon > 0$, since A is totally bounded, there exists a finite ϵ-net A_ϵ of A, that is A_ϵ is finite and

$$A \subset \bigcup_{\omega'\in A_\epsilon} B(\omega',\epsilon)\,.$$

We choose $\delta(\epsilon) > 0$ so that

$$m^T(\omega',\delta) \le \epsilon \quad \text{for all } \omega' \in A_\epsilon,\ \delta \le \delta(\epsilon)\,.$$

Then for any $\omega \in A$, there exists $\omega' \in A_\epsilon$ with $d(\omega,\omega') < \epsilon$. By (4.3) and (4.1),

$$m^T(\omega,\delta) \le \epsilon + C(T)\epsilon \quad \text{for all } \delta \le \delta(\epsilon)$$

where $C(T)$ is a positive constant depended only on T. Thus (4.5) holds.

We now assume(4.4), (4.5) and prove the compactness of $\bar{A}$. Since $C[0,\infty)$ is a metric space, it suffices to prove that every sequence $\{\omega_n\}_{n=1}^{\infty}$ in A has a convergent subsequence in $C[0,\infty)$.

Firstly, we will show that for any given $t \geq 0$, $\pi_t(A) = \{\omega(t) : \omega \in A\}$ is a bounded subset of $\mathbb{R}$. To this end, fix $T > t$ and note that for some $\delta_1 > 0$, we have $m^T(\omega, \delta_1) \leq 1$ for each $\omega \in A$. So, let $k = [\frac{t}{\delta_1}]$ and we have

$$|\omega(t) - \omega(0)| \leq \sum_{l=1}^{k} |\omega(k\delta_1) - \omega((k-1)\delta)| + |\omega(t) - \omega(k\delta_1)| \leq k + 1 .$$

Secondly, it follows that for each $r \in \mathbb{Q}_+$, $\{\omega_n(r)\}$ is bounded. Let $\{r_1, r_2, \cdots\}$ be an enumeration of $\mathbb{Q}_+$. Then choose $\{\omega_n^{(1)}\}$ a subsequence of $\{\omega_n\}$ with $\{\omega_n^{(1)}(r_1)\}$ converging to a limit denoted by $\omega(r_1)$. From $\{\omega_n^{(1)}\}$ choose a further subsequence $\{\omega_n^{(2)}\}$ such that $\omega_n^{(2)}(r_2)$ converges to a limit $\omega(r_2)$. Continue this process, and then let $\{\tilde{\omega}_n\} = \{\omega_n^{(n)}\}$ be the *diagonal sequence*. We have $\tilde{\omega}_n(r) \to \omega(r)$ for each $r \in \mathbb{Q}_+$.

Thirdly, let us note from (4.5) that for each $T > 0$ and $\epsilon > 0$, there exists $\delta_\epsilon > 0$ such that $|\tilde{\omega}_n(s) - \tilde{\omega}_n(t)| \leq \epsilon$ whenever $0 \leq s, t \leq T$, $|s - t| \leq \delta_\epsilon$, and $n \geq 1$. The same inequality, therefore, holds for ω when we impose the additional condition $s, t \in \mathbb{Q}_+$. It follows that ω is uniformly continuous on $[0, T] \cap \mathbb{Q}_+$, and so has an extension to a continuous function, also denoted by ω, on $[0, T]$. Moreover, $|\omega(s) - \omega(t)| \leq \epsilon$ whenever $0 \leq s, t \leq T$ and $|s - t| \leq \delta_\epsilon$. Now it suffices to show that $\{\tilde{\omega}_n\}$ converges to ω uniformly on $[0, T]$. For the above ϵ, δ_ϵ, there exists a large integer $K = K_\epsilon$ so that

$$[0, T] \subset \bigcup_{j=1}^{K_\epsilon} B(r_j, \delta_\epsilon) .$$

Since $\tilde{\omega}_n(r) \to \omega(r)$ for all r, we can find a large integer N_ϵ so that for all $n \geq N_\epsilon$,

$$|\tilde{\omega}_n(r_j) - \omega(r_j)| \leq \epsilon , \text{ for all } 1 \leq j \leq K_\epsilon .$$

Thus for any $t \in [0, T]$, there exists r_j $(1 \leq j \leq K_\epsilon)$ so that $|t - r_j| \leq \delta_\epsilon$, and hence

$$|\tilde{\omega}_n(t) - \omega(t)| \leq |\tilde{\omega}_n(t) - \tilde{\omega}_n(r_j)| + |\tilde{\omega}_n(r_j) - \omega(r_j)| + |\omega(r_j) - \omega(t)| \leq 3\epsilon ,$$

for all $n \geq N_\epsilon$. Then we get

$$\max_{0 \leq t \leq T} |\tilde{\omega}_n(t) - \omega(t)| \leq 3\epsilon, \text{ for all } n \geq N_\epsilon .$$

Since T is arbitary, thus ω is well-defined on $[0, \infty)$ and $\{\tilde{\omega}_n\}_{n=1}^\infty$ converges to ω in $C[0, \infty)$. □

Theorem 4.4. *Let $\{\mathbb{P}_\alpha\}_{\alpha \in \Lambda}$ be a sequence of probability measures on $C[0, \infty)$. Let $\pi = \{\pi_t\}_{t \geq 0}$ be the coordinate mappings (π is identity map). Then $\{\mathbb{P}_\alpha\}$ is tight if and only if*

$$\lim_{x \to \infty} \sup_{\alpha \in \Lambda} \mathbb{P}_\alpha(|\pi_0| > x) = 0\,, \tag{4.6}$$

$$\lim_{\delta \downarrow 0} \sup_{\alpha \in \Lambda} \mathbb{P}_\alpha\left(m^T(\pi, \delta) > \epsilon\right) = 0\,, \qquad \textit{for all } T > 0, \epsilon > 0\,. \tag{4.7}$$

Proof. Suppose first that $\{\mathbb{P}_\alpha\}$ is tight. Given $\eta > 0$, there is a compact set K with $\mathbb{P}_\alpha(K) \geq 1 - \eta$, for every α. According to Arzelà-Ascoli theorem, for sufficiently large $x > 0$, we have $|\omega(0)| \leq x$ for all $\omega \in K$; this proves (4.6). According to the same theorem, if T and ϵ are also given, then there exists δ_0 such that $m^T(\omega, \delta) \leq \epsilon$ for $0 < \delta < \delta_0$ and $\omega \in K$. This gives us (4.7).

Let us now assume (4.6) and (4.6). Given a positive integer T and $\eta > 0$ we choose $x > 0$ so that

$$\sup_{\alpha \in \Lambda} \mathbb{P}_\alpha(|\pi_0| > x) \leq \frac{\eta}{2^{T+1}}\,.$$

We choose $\delta_k > 0, k = 1, 2, \ldots$ such that

$$\sup_{\alpha \in \Lambda} \mathbb{P}_\alpha\left(m^T(\pi, \delta_k) > \frac{1}{k}\right) \leq \frac{\eta}{2^{T+k+1}}\,.$$

Define

$$A_T = \left\{\omega : |\omega(0)| \leq x, m^T(\omega, \delta_k) \leq \frac{1}{k}, \text{ for all } k \geq 1\right\}, \quad A = \bigcap_{T=1}^{\infty} A_T\,.$$

So

$$\mathbb{P}_\alpha(A_T) \geq 1 - \sum_{k=0}^{\infty} \frac{\eta}{2^{T+k+1}} = 1 - \frac{\eta}{2^T}, \quad \mathbb{P}_\alpha(A) \geq 1 - \eta,$$

for every $\alpha \in \Lambda$. By Arzelà-Ascoli theorem, A is compact, so $\{\mathbb{P}_\alpha\}$ is tight. $\square$

Exercise 4.1. Let $\{\mathbb{P}_n\}_{n\geq 1}$ be a sequence of probability measures on $C[0,\infty)$ which converges weakly to a probability measure $\mathbb{P}$. Suppose, in addition, that $\{f_n\}_{n\geq 1}$ is a uniformly bounded sequence of real-valued, continuous functions on $C[0,\infty)$ converging to a continuous function f, the convergence being uniform on compact subsets of $C[0,\infty)$. Then

$$\lim_{n\to\infty} \int_{C[0,\infty)} f_n(\omega)\mathbb{P}_n(\mathrm{d}\omega) = \int_{C[0,\infty)} f(\omega)\mathbb{P}(\mathrm{d}\omega)\,.$$

Remark 4.2. Theorem 4.3, 4.4 and Exercise 4.1 have natural extensions to $C[0,\infty)^d = C([0,\infty),\mathbb{R}^d)$, the space of continuous $\mathbb{R}^d$-valued functions on $[0,\infty)$. The proofs of these extensions are the same as for the one-dimensional case.

B Convergence of finite-fimensional distributions

Suppose that X is a continuous process on some $(\Omega,\mathcal{F},\mathbb{P})$. For each ω the function $t \mapsto X_t(\omega)$ is a member of $C[0,\infty)$, which we denote by $X(\omega)$. Since $\mathcal{B}(C[0,\infty))$ is generated by the one-dimensional cylinder sets and $X_t(\cdot)$ is $\mathcal{F}$-measurable for each fixed t, the random function $X : \Omega \to C[0,\infty)$ is $\mathcal{F}/\mathcal{B}(C[0,\infty))$ -measurable. Thus, if $\{X^{(n)}\}_{n\geq 1}$ is a sequence of continuous processes (with each $X^{(n)}$ defined on a perhaps distinct probability space $(\Omega_n,\mathcal{F}_n,P_n))$, we can ask whether $X^{(n)} \xrightarrow{\text{law}} X$. We can also ask whether the finite-dimensional distributions of $\{X^{(n)}\}_{n\geq 1}$ converge to those of X, i.e., whether

$$\left(X_{t_1}^{(n)}, X_{t_2}^{(n)}, \ldots, X_{t_d}^{(n)}\right) \xrightarrow{\text{law}} (X_{t_1}, X_{t_2}, \ldots, X_{t_d})$$

The latter question is considerably easier to answer than the former, since the convergence in distribution of finite-dimensional random vectors can be resolved by studying characteristic functions.

For any finite subset $\{t_1, \ldots, t_d\}$ of $[0, \infty)$, let us define the projection mapping $\pi_{t_1,\ldots,t_d} : C[0, \infty) \to \mathbb{R}^d$ as

$$\pi_{t_1,\ldots,t_d}(\omega) = (\omega(t_1), \ldots, \omega(t_d)) .$$

If the function $f : \mathbb{R}^d \to \mathbb{R}$ is bounded and continuous, then the composite mapping $f \circ \pi_{t_1,\ldots,t_d} : C[0, \infty) \to \mathbb{R}$ enjoys the same properties; thus, $X^{(n)} \xrightarrow{\text{law}} X$ implies

$$\begin{aligned}\lim_{n\to\infty} \mathbb{E}_n f\left(X^{(n)}_{t_1}, \ldots, X^{(n)}_{t_d}\right) &= \lim_{n\to\infty} \mathbb{E}_n (f \circ \pi_{t_1,\ldots,t_d})\left(X^{(n)}\right) \\ &= \mathbb{E}(f \circ \pi_{t_1,\ldots,t_d})(X) = \mathbb{E}f(X_{t_1}, \ldots, X_{t_d}) .\end{aligned}$$

In other words, if the sequence of processes $\left\{X^{(n)}\right\}_{n\geq 1}$ converges in distribution to the process X, then all finite-dimensional distributions converge as well. The converse holds in the presence of tightness (Theorem 4.5), but not in general; this failure is illustrated by the following example.

Example 4.1. Consider the sequence of (nonrandom) processes

$$X^{(n)}_t = nt \cdot 1_{[0,\frac{1}{2n}]}(t) + (1 - nt) \cdot 1_{(\frac{1}{2n},\frac{1}{n}]}(t) \quad t \geq 0 ,$$

and let $X_t = 0, t \geq 0$. Then, all finite-dimensional distributions of $X^{(n)}$ converge weakly to the corresponding finite-dimensional distributions of X, but the sequence of processes $\{X^{(n)}\}_{n\geq 1}$ does not converge in distribution to the process X, since

$$\max_{0\leq t\leq 1} X^{(n)}_t \equiv 1/2 > 0 = \max_{0\leq t\leq 1} X_t.$$

Theorem 4.5. *Let $\{X^{(n)}\}_{n\geq 1}$ be a tight sequence of continuous processes with the property that, whenever $0 \leq t_1 < \cdots < t_d < \infty$, then the sequence of random vectors $\{(X^{(n)}_{t_1}, \cdots, X^{(n)}_{t_d})\}$ converges in distribution. Let $\mathbb{P}_n$ be the measure induced on $(C[0, \infty), \mathcal{B}(C[0, \infty)))$ by $X^{(n)}$. Then $\{\mathbb{P}_n\}_{n=1}^{\infty}$ converges weakly to a measure $\mathbb{P}$, under which the coordinate mapping process $\pi_t(\omega) := \omega(t)$ on $C[0, \infty)$ satisfies*

$$\left(X^{(n)}_{t_1}, \ldots, X^{(n)}_{t_d}\right) \xrightarrow{\text{law}} (\pi_{t_1}, \ldots, \pi_{t_d}) , \quad 0 \leq t_1 < \cdots < t_d < \infty,\ d \geq 1 .$$

Proof. In order that $\{\mathbb{P}_n\}$ converges weakly, since we have known $\{\mathbb{P}_n\}$ is tight, it suffices to show that each weakly convergent subsequence of $\{\mathbb{P}_n\}$ has the same limit. In other words, let $\{\mathbb{P}_{l_n}\}$, $\{\mathbb{P}_{k_n}\}$ be two weakly convergent subsequence with limit $\tilde{\mathbb{P}}$, $\hat{\mathbb{P}}$ respectively, then $\tilde{\mathbb{P}} = \hat{\mathbb{P}}$. This is trivial since the finite-dimensional distributions of $\tilde{\mathbb{P}}$, $\hat{\mathbb{P}}$ coincide: For $f \in C_b(\mathbb{R}^d)$ and $0 \le t_1 < \cdots < t_d < \infty$,

$$\begin{aligned}
\tilde{\mathbb{E}} f(\pi_{t_1}, \cdots, \pi_{t_d}) &= \int f \circ \pi_{t_1,\dots,t_d} \, \mathrm{d}\tilde{\mathbb{P}} = \lim_{n\to\infty} \int f \circ \pi_{t_1,\dots,t_d} \, \mathrm{d}\mathbb{P}_{l_n} \\
&= \lim_{n\to\infty} \int f \, \mathrm{d}(\mathbb{P}_{l_n} \circ \pi^{-1}_{t_1,\dots,t_d}) = \lim_{n\to\infty} \int f \, \mathrm{d}(\mathbb{P}_{k_n} \circ \pi^{-1}_{t_1,\dots,t_d}) \\
&= \lim_{n\to\infty} \int f \circ \pi_{t_1,\dots,t_d} \, \mathrm{d}\mathbb{P}_{k_n} = \int f \circ \pi_{t_1,\dots,t_d} \, \mathrm{d}\hat{\mathbb{P}} \\
&= \hat{\mathbb{E}} f(\pi_{t_1}, \cdots, \pi_{t_d}) \, .
\end{aligned}$$

We now complete the proof. □

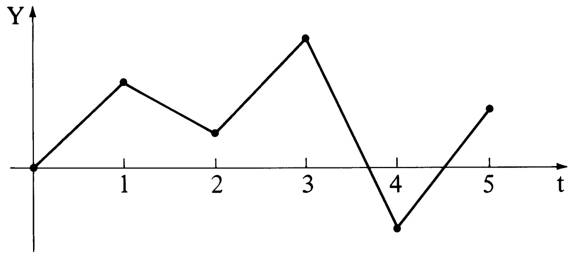
We shall need the following result.

Exercise 4.2. Let $\{X^{(n)}\}$, $\{Y^{(n)}\}$, and X be random variables with values in a Polish space $(E, \mathcal{B}(E))$. Assume that for each $n \ge 1$, $X^{(n)}$ and $Y^{(n)}$ are defined on the same probability space. If $X^{(n)} \xrightarrow{\text{law}} X$ and $\rho\left(X^{(n)}, Y^{(n)}\right) \to 0$ in probability. Then $Y^{(n)} \xrightarrow{\text{law}} X$.

C The invariance principle

Let us consider now a sequence $\{\xi_j\}_{j\ge1}$ of independent, identically distributed random variables with mean zero and positive variance σ^2, as well as the sequence of partial sums $S_0 = 0, S_n = \sum_{j=1}^n \xi_j, n \ge 1$. A continuous-time process $Y = \{Y_t\}_{t\ge0}$ can be obtained from the sequence $\{S_n\}_{n\ge0}$ by *linear interpolation*:

$$Y_t := S_{[t]} + (t - [t])\xi_{[t]+1}, \quad t \ge 0 \, . \tag{4.8}$$



Scaling appropriately both time and space, we obtain from Y a sequence of processes $\left\{X^{(n)}\right\}$:

$$X_t^{(n)} = \frac{1}{\sigma\sqrt{n}} Y_{nt}, \quad t \geq 0 \tag{4.9}$$

Note that with $s = \frac{k}{n}$ and $t = \frac{k+1}{n}$, the increment $X_t^{(n)} - X_s^{(n)} = \frac{1}{\sigma\sqrt{n}}\xi_{k+1}$ is independent of $\mathcal{F}_s^{X^{(n)}} = \sigma\left(\xi_1, \cdots, \xi_k\right)$. Furthermore, $X_t^{(n)} - X_s^{(n)}$ has zero mean and variance $t-s$. This suggests that $\{X^{(n)}\}$ is approximately a Brownian motion. We now show that, even though the random variables ξ_j are not necessarily normal, the central limit theorem dictates that the limiting distributions of the increments of $X^{(n)}$ are normal.

Theorem 4.6. *With $\{X^{(n)}\}$ defined by (4.9) and $0 \leq t_1 < \cdots < t_d < \infty$, we have*

$$\left(X_{t_1}^{(n)}, \ldots, X_{t_d}^{(n)}\right) \xrightarrow{\text{law}} (B_{t_1}, \ldots, B_{t_d}) \qquad \text{as } n \to \infty\,,$$

where $\{B_t, \}_{t \geq 0}$ is a standard, one-dimensional Brownian motion.

Proof. We take the case $d = 2$; the other cases differ from this one only by being notationally more cumbersome. Set $s = t_1, t = t_2$. We wish to show

$$\left(X_s^{(n)}, X_t^{(n)}\right) \xrightarrow{\text{law}} (B_s, B_t)\ .$$

Since

$$\left|X_t^{(n)} - \frac{1}{\sigma\sqrt{n}} S_{[tn]}\right| \leq \frac{1}{\sigma\sqrt{n}} \left|\xi_{[tn]+1}\right|\ ,$$

we have by the Chebyshev inequality,

$$\mathbb{P}\left(\left|X_t^{(n)} - \frac{1}{\sigma\sqrt{n}} S_{[tn]}\right| > \epsilon\right) \le \frac{1}{\epsilon^2 n} \to 0 \, .$$

It is clear then that

$$\left\| \left(X_s^{(n)}, X_t^{(n)}\right) - \frac{1}{\sigma\sqrt{n}} \left(S_{[sn]}, S_{[tn]}\right) \right\| \to 0 \quad \text{in probability}$$

so, by Problem 4.16, it suffices to show

$$\frac{1}{\sigma\sqrt{n}} \left(S_{[sn]}, S_{[tn]}\right) \xrightarrow{\text{law}} (B_s, B_t)$$

From continuity mapping theorem, we see that this is equivalent to proving

$$\frac{1}{\sigma\sqrt{n}} \left(\sum_{j=1}^{[sn]} \xi_j , \sum_{j=[sn]+1}^{[tn]} \xi_j \right) \xrightarrow{\text{law}} (B_s, B_t - B_s)$$

We shall use characteristic function to show this. The independence of the random variables $\{\xi_j\}$ implies

$$\begin{aligned} &\mathbb{E}\left[\exp\left\{ \frac{iu}{\sigma\sqrt{n}} \sum_{j=1}^{[sn]} \xi_j + \frac{iv}{\sigma\sqrt{n}} \sum_{j=[sn]+1}^{[tn]} \xi_j \right\}\right] \\ &= \mathbb{E}\left[\exp\left\{ \frac{iu}{\sigma\sqrt{n}} \sum_{j=1}^{[sn]} \xi_j \right\}\right] \cdot \mathbb{E}\left[\exp\left\{ \frac{iv}{\sigma\sqrt{n}} \sum_{j=[sn]+1}^{[tn]} \xi_j \right\}\right] \, . \end{aligned} \tag{4.10}$$

By the central limit theorem,

$$\frac{1}{\sigma\sqrt{n}} \sum_{j=1}^{[sn]} \xi_j \xrightarrow{\text{law}} N(0, s) \, .$$

We have

$$\lim_{n\to\infty} \mathbb{E}\left[\exp\left\{ \frac{iu}{\sigma\sqrt{n}} \sum_{j=1}^{[sn]} \xi_j \right\}\right] = e^{-u^2 \frac{s}{2}} \, .$$

Similarly,

$$\lim_{n\to\infty} \mathbb{E}\left[\exp\left\{\frac{iv}{\sigma\sqrt{n}}\sum_{j=[sn]+1}^{[tn]}\xi_j\right\}\right] = e^{-v^2\frac{t-s}{2}}\,.$$

Substitution of these last two equations into (4.10) completes the proof. □

Actually, the sequence $\{X^{(n)}\}$ of linearly interpolated and normalized random walks in (4.9) converges to Brownian motion in distribution. For the tightness required to carry out such an extension (recall Theorem 4.5), we shall need two auxiliary results.

We are now in a position to establish the main result of this section, namely the convergence in distribution of the sequence of normalized random walks in (4.9) to Brownian motion. This result is also known as the *invariance principle*.

Theorem 4.7 (Donsker's Invariance Principle)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which is given a sequence $\{\xi_j\}_{j\geq 1}$ of independent, identically distributed random variables with mean zero and positive variance σ^2. Define $X^{(n)}$ by* (4.9) *and let $\mathbb{P}_n$ be the measure induced by $X^{(n)}$ on $(C[0,\infty), \mathcal{B}(C[0,\infty)))$. Then $\{\mathbb{P}_n\}$ converges weakly to a measure $\mathbb{P}_*$ under which the coordinate mapping process $W_t(\omega) := \omega(t)$ on $C[0,\infty)$ is a standard, one-dimensional Brownian motion.*

Proof. In light of Theorem 4.5 and 4.6, it remains to show that $\{X^{(n)}\}_{n=1}^{\infty}$ is tight. For this we use Theorem 4.4, and since $X_0^{(n)} = 0$ a.s. for every n, we need only establish, for arbitrary $\epsilon > 0$ and $T > 0$, the convergence

$$\lim_{\delta\downarrow 0}\sup_{n\geq 1}\mathbb{P}\left(\max_{\substack{|s-t|\leq\delta\\ 0\leq s,t\leq T}}\left|X_s^{(n)} - X_t^{(n)}\right| > \epsilon\right) = 0\,. \tag{4.11}$$

We may replace "$\sup_{n\geq 1}$" in this expression by "$\limsup_{n\to\infty}$", since for a finite number of integers n we can make the probability appearing in (4.11) as small as we choose, by reducing δ. But

$$\mathbb{P}\left(\max_{\substack{|s-t|\leq\delta\\ 0\leq s,t\leq T}}\left|X_s^{(n)} - X_t^{(n)}\right| > \epsilon\right) = \mathbb{P}\left(\max_{\substack{|s-t|\leq n\delta\\ 0\leq s,t\leq nT}}|Y_s - Y_t| > \epsilon\sigma\sqrt{n}\right),$$

and

$$\max_{\substack{|s-t|\le n\delta \\ 0\le s,t\le nT}} |Y_s - Y_t| \le \max_{\substack{|s-t|\le \lceil \delta \rceil \\ 0\le s,t\le \lceil nT\rceil}} |Y_s - Y_t| \le \max_{\substack{0\le k\le \lceil nT\rceil \\ 1\le j\le \lceil n\delta\rceil}} |S_{k+j} - S_k| ,$$

where the last inequality follows from the fact that Y is piecewise linear and changes slope only at integer values. Now (4.11) follows from Lemma 4.8. □

Remark 4.3. A standard, one-dimensional Brownian motion defined on any probability space can be thought of as a random variable with values in $C[0,\infty)$; regarded this way, it induces the *Wiener measure* on $(C[0,\infty), \mathcal{B}(C[0,\infty)))$. For this reason, we call $(C[0,\infty), \mathcal{B}(C[0,\infty)), \mathbb{P}_*)$ where $\mathbb{P}_*$ is Wiener measure, the *canonical probability space* for Brownian motion.

Lemma 4.8. *For any $T > 0$ and $\epsilon > 0$,*

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} \mathbb{P}\left(\max_{\substack{0\le k\le \lceil nT\rceil \\ 1\le j\le \lceil n\delta\rceil}} |S_{k+j} - S_k| > \epsilon\sigma\sqrt{n} \right) = 0\,.$$

Proof. For $0 < \delta \le T$, let $m = m_\delta \ge 2$ be the unique integer satisfying $m - 1 \le \frac{T}{\delta} < m$. Since $\frac{\lceil nT\rceil}{\lceil n\delta\rceil} \to \frac{T}{\delta} < m$, we have $\lceil nT\rceil < m\lceil n\delta\rceil$ for sufficiently large n.

Step 1. Let n be sufficiently large. We shall show that

$$\begin{aligned} &\left\{ \max_{\substack{0\le k\le \lceil nT\rceil \\ 1\le j\le \lceil n\delta\rceil}} |S_{k+j} - S_k| > \epsilon\sigma\sqrt{n} \right\} \\ &\subset \bigcup_{p=0}^{m} \left\{ \max_{1\le j\le \lceil n\delta\rceil} \left|S_{j+p\lceil n\delta\rceil} - S_{p\lceil n\delta\rceil}\right| > \frac{1}{3}\epsilon\sigma\sqrt{n} \right\} . \end{aligned} \tag{4.12}$$

Suppose $|S_{j+k} - S_k| > \epsilon\sigma\sqrt{n}$ for some k in $\{0, 1, \cdots, \lceil nT\rceil\}$ and j in $\{1, \cdots \lceil n\delta\rceil\}$. There exists then a unique integer $p, 0 \le p \le m - 1$, such that

$$p\lceil n\delta\rceil \le k < (p+1)\lceil n\delta\rceil\,.$$

There are two possibilities for $k + j$:

- One possibility is that $p\lceil n\delta\rceil < k+j \le (p+1)\lceil n\delta\rceil$, in which case either $\left|S_k - S_{p\lceil n\delta\rceil}\right| > \frac{1}{3}\epsilon\sigma\sqrt{n}$, or else $\left|S_{k+j} - S_{p\lceil n\delta\rceil}\right| > \frac{1}{3}\epsilon\sigma\sqrt{n}$.
- The second possibility is that $(p+1)\lceil n\delta\rceil < k+j < (p+2)\lceil n\delta\rceil$, in which case either $\left|S_k - S_{p\lceil n\delta\rceil}\right| > \frac{1}{3}\epsilon\sigma\sqrt{n}$, $\left|S_{(p+1)p\lceil n\delta\rceil} - S_{p\lceil n\delta\rceil}\right| > \frac{1}{3}\epsilon\sigma\sqrt{n}$, or else $\left|S_{k+j} - S_{(p+1)\lceil n\delta\rceil}\right| > \frac{1}{3}\epsilon\sigma\sqrt{n}$.

In conclusion, we see (4.12) holds.

Step 2. Observe that for all p,

$$\mathbb{P}\left(\max_{1\le j\le\lceil n\delta\rceil}\left|S_{j+p\lceil n\delta\rceil} - S_{p\lceil n\delta\rceil}\right| > \frac{1}{3}\epsilon\sigma\sqrt{n}\right) = \mathbb{P}\left(\max_{1\le j\le\lceil n\delta\rceil}|S_j| > \frac{1}{3}\epsilon\sigma\sqrt{n}\right),$$

By step 1, we have

$$\begin{aligned}\mathbb{P}\left(\max_{\substack{0\le k\le\lceil nT\rceil\\1\le j\le\lceil n\delta\rceil}}|S_{k+j} - S_k| > \epsilon\sigma\sqrt{n}\right) &\le (m+1)\mathbb{P}\left(\max_{1\le j\le\lceil n\delta\rceil}|S_j| > \frac{1}{3}\epsilon\sigma\sqrt{n}\right).\\ &\le \frac{2T}{\delta}\mathbb{P}\left(\max_{1\le j\le\lceil n\delta\rceil}|S_j| > \frac{1}{3}\epsilon\sigma\sqrt{n}\right).\end{aligned}$$

So it suffices to show

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty}\frac{1}{\delta}\,\mathbb{P}\left(\max_{1\le j\le\lceil n\delta\rceil}|S_j| > \epsilon\sigma\sqrt{n}\right) = 0\,. \tag{4.13}$$

Step 3. We now define

$$\tau = \inf\left\{j\ge 0; |S_j| > \varepsilon\sigma\sqrt{n}\right\}.$$

With $0<\delta<\frac{\varepsilon^2}{2}$, we have (imitating the proof of the Kolmogorov inequality e.g., Chung (1974), p. 116)

$$\begin{aligned}\mathbb{P}\left(\max_{0\le j\le\lceil n\delta\rceil}|S_j| > \varepsilon\sigma\sqrt{n}\right) &\le \mathbb{P}\left(\left|S_{\lceil n\delta\rceil}\right| \ge \sigma\sqrt{n}(\varepsilon - \sqrt{2\delta})\right)\\ &\quad + \sum_{j=1}^{\lceil n\delta\rceil}\mathbb{P}\left(\left|S_{\lceil n\delta\rceil}\right| < \sigma\sqrt{n}(\varepsilon-\sqrt{2\delta})\middle|\tau = j\right)\mathbb{P}(\tau=j)\end{aligned} \tag{4.14}$$

But if $\tau = j$, then $\left|S_{\lceil n\delta \rceil}\right| < \sigma\sqrt{n}(\varepsilon - \sqrt{2\delta})$ implies $\left|S_{\lceil n\delta \rceil} - S_j\right| > \sigma\sqrt{2n\delta}$. By the Chebyshev inequality, the probability of this event is bounded above by

$$\frac{1}{2n\delta\sigma^2}\mathbb{E}\left[\left(S_{\lceil n\delta \rceil} - S_j\right)^2 |\tau = j\right] = \frac{1}{2n\delta\sigma^2}\mathbb{E}\left(\sum_{i=j+1}^{\lceil n\delta \rceil} \xi_i^2\right) \leq \frac{1}{2}\,.$$

Returning to (4.14), we may now write

$$\begin{aligned}
&\mathbb{P}\left(\max_{0\leq j\leq \lceil n\delta \rceil} |S_j| > \varepsilon\sigma\sqrt{n}\right) \\
\leq &\mathbb{P}\left(\left|S_{\lceil n\delta \rceil}\right| \geq \sigma\sqrt{n}(\varepsilon - \sqrt{2\delta})\right) + \frac{1}{2}\mathbb{P}(\tau \leq [n\delta]) \\
\leq &\mathbb{P}\left(\left|S_{\lceil n\delta \rceil}\right| \geq \sigma\sqrt{n}(\varepsilon - \sqrt{2\delta})\right) + \frac{1}{2}\mathbb{P}\left(\max_{0\leq j\leq \lceil n\delta \rceil} |S_j| > \varepsilon\sigma\sqrt{n}\right)\,.
\end{aligned}$$

So, it follows that

$$\mathbb{P}\left(\max_{0\leq j\leq \lceil n\delta \rceil} |S_j| > \varepsilon\sigma\sqrt{n}\right) \leq 2\mathbb{P}\left(\left|S_{\lceil n\delta \rceil}\right| \geq \sigma\sqrt{n}(\varepsilon - \sqrt{2\delta})\right)\,.$$

Since

$$\frac{S_{\lceil n\delta \rceil}}{\sigma\sqrt{n\delta}} \xrightarrow{\text{law}} N(0,1)\,,$$

we have

$$\begin{aligned}
&\lim_{\delta\downarrow 0}\limsup_{n\to\infty}\frac{1}{\delta}\,\mathbb{P}\left(\max_{1\leq j\leq \lceil n\delta \rceil} |S_j| > \epsilon\sigma\sqrt{n}\right) \\
\leq &\lim_{\delta\downarrow 0}\frac{2}{\delta}\limsup_{n\to\infty}\,\mathbb{P}\left(\frac{|S_{\lceil n\delta \rceil}|}{\sigma\sqrt{n\delta}} \geq \frac{\varepsilon - \sqrt{2\delta}}{\sqrt{\delta}}\right) \\
= &\lim_{\delta\downarrow 0}\frac{2}{\delta}\,\mathbb{P}\left(|N(0,1)| \geq \frac{\varepsilon - \sqrt{2\delta}}{\sqrt{\delta}}\right) \\
\leq &\lim_{\delta\downarrow 0}\frac{2\sqrt{\delta}}{(\varepsilon - \sqrt{2\delta})^3}\mathbb{E}|N(0,1)|^3\,.
\end{aligned}$$

Now we complete the proof. ☐

4.2 Skorokhod embedding

We will give a motivation about the *Skorokhod embedding*. Let $\{S_n\}_{n\geq 0}$ be a simple random walk on the line, starting at the origin. For positive integer n, let $X_n = S_n - S_{n-1}$ be the n-step of the walk, as we konw, $\{X_n\}$ is i.i.d. with the distribution

$$\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2\,.$$

Let $\{B(t)\}_{t\geq 0}$ be a standard Brownian motion. We will find a sequence of stopping times $\{\tau_n\}$ of $\{B(t)\}$ so that Let $\tau_1 \coloneqq \inf\{t \geq 0 : |B_t| = 1\}$. As we showed in Theorem 3.44, 3.46, $\mathbb{E}\tau_1 = 1$, and

$$\mathbb{P}(B_{\tau_1} = 1) = \mathbb{P}(B_{\tau_1} = -1) = 1/2\,.$$

Next, we define

$$\begin{aligned}\tau_2 \coloneqq& \inf\{t \geq \tau_1 : |B(t) - B(\tau_1)| = 1\}\\ =& \inf\{t \geq 0 : |B(t+\tau_1) - B(\tau_1)| = 1\} + \tau_1\,,\end{aligned}$$

is a stopping times for Brownian motion. It follows from the strong Markov property that, $\{B(t+\tau_1) - B(\tau_1)\}_{t\geq 0}$ is a standard Brownian motion and independent of $\mathcal{F}^B_{\tau_1}$. So

$$B_{\tau_2} - B_{\tau_1}\ ,\ B_{\tau_1}$$

are independent and identically distributed random variables.

We define stopping times τ_n of the Brownian motion by induction. Let

$$\begin{aligned}\tau_n \coloneqq& \inf\{t \geq \tau_1 : |B(t) - B(\tau_{n-1})| = 1\}\\ =& \inf\{t \geq 0 : |B(t+\tau_{n-1}) - B(\tau_{n-1})| = 1\} + \tau_{n-1}\,,\end{aligned}$$

where we set $\tau_0 = 0$. By Strong Markov property,

$$B_{\tau_n} - B_{\tau_{n-1}}, \cdots, B_{\tau_1} - B_{\tau_0}$$

are independent and identically distributed random variables. Therefore, $\{B_{\tau_n}\}$ is simple random walk, i.e.,

$$\mathcal{L}[\{S_n\}] = \mathcal{L}[\{B_{\tau_n}\}]\,.$$

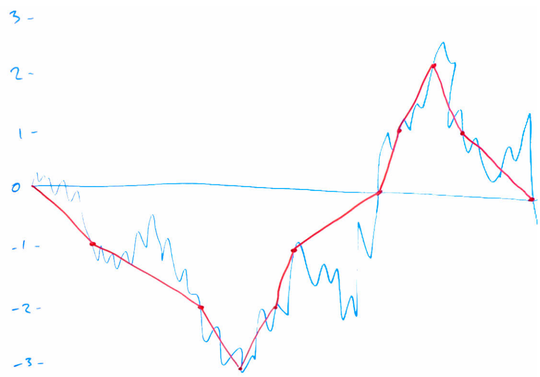


Figure 4.1: Embedding simple random walk in Brownian motion.

For random walks $\{S_n\}_{n\geq 0}$ with general increments, we use a random variable X to represent the distribution of the increment. So, if we want to embed the random walk $\{S_n\}_{n\geq 0}$ in Brownian motion, it's neceaasry to ask that if we can embed a random variable X in Brownian motion. In other words, for a given random variable X representing the increment, does there exist a stopping time τ of Brownian motion with $\mathbb{E}\tau < \infty$, such that B_τ has the law of X. This problem is called the *Skorokhod embedding problem(SEP)*. By Wald's lemmas, for any integrable stopping time τ, we have

$$\mathbb{E}B_\tau = 0 \quad \text{and} \quad \mathbb{E}B_\tau^2 = \mathbb{E}\tau < \infty\,,$$

so that the Skorokhod embedding problem can only be solved for random variables X with mean zero and finite second moment. However, these are the only restrictions, as the following result shows.

Theorem 4.9 (Skorokhod Embedding Theorem)**.** *Let X be a real valued random variable with $\mathbb{E}X = 0$ and $\mathbb{E}X^2 < \infty$, then there exists a stopping time τ, of the natural filtration $\{\mathcal{F}_t^B\}_{t\geq 0}$ of the Brownian motion, such that B_τ has the same law as X and $\mathbb{E}\tau = \mathbb{E}X^2$.*

Example 4.2. Assume that X may take two values $a < b$. In order that $\mathbb{E}X = 0$, we must have $a < 0 < b$ and $\mathbb{P}(X = a) = b/(b-a)$ and $P(X = b) = -a/(b-a)$. We have seen in As we showed in Theorem 3.44, 3.46 that, for the stopping time $\tau = \inf\{t : B(t) \notin (a, b)\}$ the random variable B_T has the same distribution as X, and that $\mathbb{E}\tau = -ab$ is finite.

We now present two proofs of the Skorokhod embedding theorem, which actually represent different constructions of the required stopping times. Both approaches, Dubins'embedding, and the Azéma-Yor embedding are very elegant and have their own merits.

A Solution by Dubins

The first one, due to Dubins, is particularly simple and based on the notion of binary splitting martingales. We say that a martingale $\{X_n\}_{n\geq 0}$ is *binary splitting* if, whenever for some $x_0, \ldots, x_n \in \mathbb{R}$ the event

$$\{X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n\}$$

has positive probability, the distribution of X_{n+1} conditioned on this event,

$$\mathcal{L}[X_{n+1}|X_0 = x_0, \ldots, X_n = x_n]\,,$$

is supported on at most two values.

Lemma 4.10. *Let X be a random variable with $\mathbb{E}X^2 < \infty$. Then there is a binary splitting martingale $\{X_n\}_{n\geq 0}$ such that $X_n \to X$ a.s. and in L^2.*

Proof. We define the martingale $\{X_n\}_{n\geq 0}$ and the associated filtration $\{\mathcal{G}_n\}_{n\geq 0}$ recursively. Let

$$\mathcal{G}_0 := \{\emptyset, \Omega\}\,,\ X_0 := \mathbb{E}X\,,\ \xi_0 = \begin{cases} 1 & \text{if } X \geq X_0 \\ -1 & \text{if } X < X_0 \end{cases}\,.$$

For any $n \geq 1$, let

$$\mathcal{G}_n = \sigma(\xi_0, \cdots, \xi_{n-1}) \ , \ X_n = \mathbb{E}(X|\mathcal{G}_n) \ , \ \xi_n = \begin{cases} 1 & \text{if } X \geq X_n \\ -1 & \text{if } X < X_n \end{cases} .$$

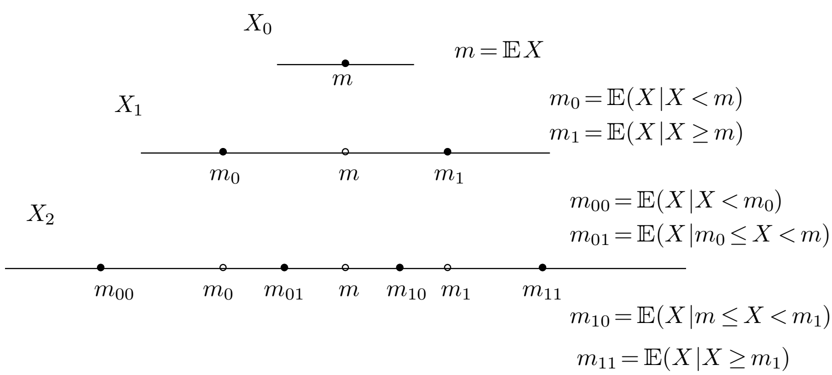


Figure 4.2: The ditribution of X_0, X_1, X_2.

Note that $\mathcal{G}_n$ is generated by a partition $\mathcal{P}_n$ of the underlying probability space into 2^n sets, each of which has the form

$$\{X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n\} \ .$$

As each element of $\mathcal{P}_n$ is a union of two elements of $\mathcal{P}_{n+1}$, the martingale $\{X_n\}_{n \geq 0}$ is binary splitting. By Theorem 1.35 (iii), we get

$$X_n \to X_\infty := \mathbb{E}[X|\mathcal{G}_\infty] \quad \text{a.s. and in } L^2 .$$

where $\mathcal{G}_\infty = \sigma(\cup_i \mathcal{G}_i)$. To conclude the proof we have to show that $X = X_\infty$ almost surely. We claim that, almost surely,

$$\lim_{n \uparrow \infty} \xi_n (X - X_{n+1}) = |X - X_\infty| \tag{4.15}$$

Indeed, if $X(\omega) = X_\infty(\omega)$ this is easy. If $X(\omega) < X_\infty(\omega)$ then for some large enough N we have $X(\omega) < X_n(\omega)$ for any $n > N$, hence $\xi_n = -1$ and (4.15) holds. Similarly, if $X(\omega) > X_\infty(\omega)$ then $\xi_n = 1$ for $n > N$ and so (4.15) holds. Using that ξ_n is $\mathcal{G}_{n+1}$ -measurable, we find that

$$\mathbb{E}\left[\xi_n\left(X - X_{n+1}\right)\right] = \mathbb{E}\left[\xi_n \mathbb{E}\left(X - X_{n+1} | \mathcal{G}_{n+1}\right)\right] = 0$$

Hence, we conclude that $\mathbb{E}\left|X - X_\infty\right| = 0$. □

Proof of Theorem 4.9. From Lemma 4.10, we take the binary splitting martingale $\{X_n\}_{n \geq 0}$ such that $X_n \to X$ almost surely and in L^2. Define $\tau_0 = 0$, and

$$\tau_1 = \inf\{t \geq 0 : B_t \in \operatorname{supp} \mathcal{L}(X_1)\},$$

where $\operatorname{supp}\mathcal{L}(X_1)$ is the support of the distribution of X_1. Then, clearly, B_{τ_n} has the same distribution of X_1. Let

$$\tau_n = \inf\{t \geq \tau_{n-1} : B_t \in \operatorname{supp} \mathcal{L}(X_n)\},$$

It's not hard o check that $\tau_n \in L^1$, B_{tau_n} has the same distribution of X_n and $\mathbb{E}\tau_n = \mathbb{E}X_n^2$. As τ_n is an increasing sequence, we have $\tau_n \uparrow \tau$ almost surely for some stopping time τ. Also, by the monotone convergence theorem

$$\mathbb{E}\tau = \lim_{n \uparrow \infty} \mathbb{E}\tau_n = \lim_{n \uparrow \infty} \mathbb{E}X_n^2 = \mathbb{E}X^2 .$$

As B_{τ_n} converges in distribution to X by construction, and converges almost surely to B_τ by continuity of the Brownian sample paths, we get that B_τ is distributed as X. □

B Solution by Azéma and Yor

In this subsection we discuss a second solution to the Skorokhod embedding problem with a more explicit construction of the stopping times.

Theorem 4.11. *Suppose that X is a real valued random variable with $\mathbb{E}X = 0$ and $\mathbb{E}X^2 < \infty$. Let*

$$\Psi(x) = \begin{cases} \mathbb{E}(X|X \geq x), & \text{if } \mathbb{P}(X \geq x) > 0 \\ 0, & \text{otherwise} \end{cases} .$$

For a Brownian motion $\{B_t\}_{t\geq 0}$ let $\{M_t\}_{t\geq 0}$ be the maximum process and define a stopping time τ by

$$\tau = \inf\{t \geq 0 : M_t \geq \Psi(B_t)\} .$$

Then $\mathbb{E}\tau = \mathbb{E}X^2$ and B_τ has the same law as X.

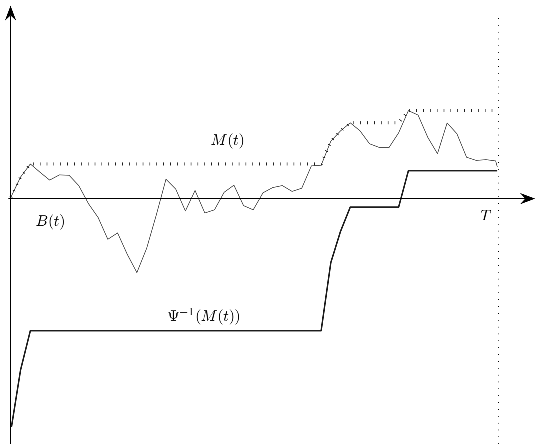


Figure 4.3: The Azéma-Yor embedding: the path is stopped when the Brownian motion hits the level $\Psi^{-1}(M_t)$, where $\Psi^{-1}(x) = \sup\{a : \Psi(a) \leq x\}$

Remark 4.4. It's easy to see that Ψ is increasing and $\Psi(x) > x$ on $(-\infty, x_0)$, where $x_0 \coloneqq \sup\{x : \mathbb{P}(X \geq x) > 0\}$. In fact, for $x_1 < x_2$, with $\mathbb{P}(X > x_2) > 0$, we have

$$\begin{aligned}
&\mathbb{E}(X|X \geq x_1) \leq \mathbb{E}(X|X \geq x_2) \\
\Leftrightarrow\ &\mathbb{P}(X \geq x_2)\mathbb{E}[X1_{\{X \geq x_1\}}] \leq \mathbb{P}(X \geq x_1)\mathbb{E}[X1_{\{X \geq x_2\}}] \\
\Leftrightarrow\ &\mathbb{P}(X \geq x_2)\mathbb{E}[X1_{\{x_1 \leq X < x_2\}}] \leq \mathbb{P}(x_1 \leq X < x_2)\mathbb{E}[X1_{\{X \geq x_2\}}] .
\end{aligned}$$

The last inequality is obvious. Moreover, τ is a stopping time of Brown motion since ψ is left-continuous.

Remark 4.5. Clearly, τ is a stopping time of the natural filtration. In fact, we have

$$\{\tau \leq t\} = \bigcap_{n\in\mathbb{N}} \bigcup_{r\in[0,t]\cap\mathbb{Q}} \left\{M_r \geq \Psi(B_r) - \frac{1}{2^n}\right\} \in \mathcal{F}_t^B .$$

Besides, $\tau < \infty$ almost surely is trivial. In fact, if $\mathbb{P}(X = 0) = 1$, then $\tau = 0$.. Otherwise, take any $x < 0 < y = \Psi(x)$, and let τ_x be the first visit to x following the first visit to y. Then $\tau \leq \tau_x < \infty$ a.s.

We proceed in three steps. In the first step we formulate an embedding for random variables taking only finitely many values.

Lemma 4.12. *Suppose the random variable X with mean zero takes only finitely many values $\{x_1 < x_2 < \cdots < x_n\}$. Define $y_1 < y_2 < \cdots < y_{n-1}$ by $y_i = \Psi(x_{i+1})$, and define stopping times $\tau_0 = 0$,*

$$\tau_i = \inf\{t \geq \tau_{i-1} : B(t) \notin (x_i, y_i)\} \quad \textit{for } i \leq n-1 .$$

Then τ_{n-1} satisfies $\mathbb{E}\tau_{n-1} = \mathbb{E}X^2$ and $B_{\tau_{n-1}}$ has the same law as X.

Proof. First observe that $y_i \geq x_{i+1}$ and this equality holds if and only if $i = n-1$. We have $\mathbb{E}\tau_{n-1} < \infty$, and hence $\mathbb{E}\,\tau_{n-1} = \mathbb{E}B^2_{\tau_{n-1}}$. So it suffices to show that $B_{\tau_{n-1}}$ has the same distribution as X. To this end, we will show a stronger result. For $1 \leq k \leq n_1$, define random variables

$$Y_k = \begin{cases} y_k , & \text{if } X \geq x_{k+1} \\ x_j , & \text{if } X = x_j \leq x_k \end{cases} .$$

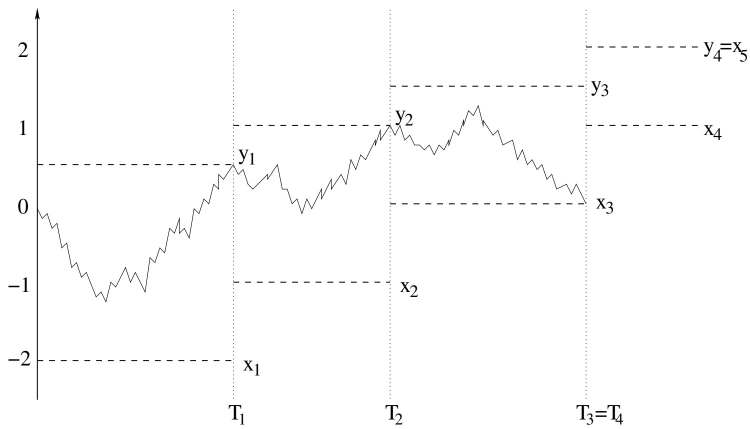


Figure 4.4: The Azéma-Yor embedding for the uniform distribution on the set $\{-2,-1,0,1,2\}$. The drawn path samples the value $B_\tau = 0$ with $\tau = \tau_4$

Clearly $Y_{n-1} = X$. We now argue that

$$\mathcal{L}\left(B_{\tau_1}, \ldots, B_{\tau_{n-1}}\right) = \mathcal{L}\left(Y_1, \ldots, Y_{n-1}\right) .$$

Note that Y_1 and B_{τ_1} has expectation zero and takes on the two values x_1, y_1. For $k \geq 2$, given $Y_{k-1} = y_{k-1}$ $(B_{\tau_{k-1}} = y_{k-1})$, the random variable Y_k (B_{τ_k}) takes the values x_k, y_k and has expectation y_{k-1}. Given $Y_{k-1} = x_j$ $(B_{\tau_{k-1}} = x_j)$, $j \leq k-1$, we have $Y_k = x_k$ $(B_{\tau_k} = x_j)$. Hence the two tuples have the same law and, in particular, $B_{\tau_{n-1}}$ has the same law as X. □

Lemma 4.13. *The stopping time τ_{n-1} constructed in Lemma 4.12 and the stopping time τ in Theorem 4.11 are equal.*

Proof. Firstly, we will show that $\tau \leq \tau_{n-1}$. To this end, it's enough to show that $M_{\tau_{n-1}} \geq \Psi(B_{\tau_{n-1}})$. Suppose that $B_{\tau_{n-1}} = x_k$, then $\Psi\left(B_{\tau_{n-1}}\right) = y_{k-1}$.

(i) If $k \leq n-1$, then $B_{\tau_{k-1}} = y_{k-1} = \Psi(x_k)$, and hence $M_{\tau_k} \geq y_{k-1}$.

(ii) If $k = n$, we also have $M_{\tau_{n-1}} = x_n = y_{n-1}$.

So in any case, we have $\tau \le \tau_{n-1}$.

Secondly, we shall show that $\tau_{n-1} \le \tau$. To see this, suppose that $B_{\tau_{n-1}} = x_k$. For $j \le k$, and $\tau_{j-1} \le t < \tau_j$, we have $B_t \in (x_j, y_j)$ and this implies $M_t < y_j \le \Psi(B_t)$. Hence $\tau_{n-1} \le \tau$.

Finally, we have seen $\tau_{n-1} = \tau$. $\square$

Now we completes the proof of Theorem 4.11 for random variables taking finitely many values. The general case follows from a limiting procedure.

Lemma 4.14. *Given a centred random variable X with finite variance. There exist centred random variables X_n taking only finitely many values, such that X_n converges to X in law and for $\Psi_n(x) = \mathbb{E}\left(X_n | X_n \ge x\right)$, the embedding stopping times*

$$\tau_n = \inf\{t \ge 0 : M_t \ge \Psi_n(B_t)\}$$

converge a.s. to τ. Infer that B_τ has the same law as X, and $\mathbb{E}\tau = \mathbb{E}X^2$.

Proof. For each $n \in \mathbb{N}$, divide the intersection of the support of X with the interval $[-n, n]$ into finitely intervals with mesh $1/n$. If $x_1 < \cdots < x_m$ are the partition points, construct the law of X_n by placing, for any $j \in \{0, \ldots, m\}$, atoms of size $\mathbb{P}\left(X \in [x_j, x_{j+1})\right)$ in position $\mathbb{E}\left(X | x_j \le X < x_{j+1}\right)$, using the convention $x_0 = -\infty$ and $x_{m+1} = \infty$. By construction, X_n takes only finitely many values. Evidently, $\mathbb{E}X_n = 0$ and X_n converges to X in distribution.

Then, one can show that $\tau_n \to \tau$ almost surely. (How?) This implies that $B_{\tau_n} \to B_\tau$ a.s., and therefore also in distribution, which implies that X has the same law as B_τ. Fatou's lemma implies that

$$\mathbb{E}\tau \le \liminf_{n\to\infty} \mathbb{E}\tau_n = \liminf_{n\to\infty} \mathbb{E}X_n^2 < \infty\,.$$

Hence, by Wald's second lemma, $\mathbb{E}X^2 = \mathbb{E}B_\tau^2 = \mathbb{E}\tau$. $\square$

From Theorem 8.1.1, it is only a small step to:

Theorem 4.15. *Let $X_1, X_2, \ldots$ be i.i.d. with a distribution F which has mean 0 and variance 1, and let $S_n = X_1 + \cdots + X_n$. There is a sequence of stopping times $\tau_0 = 0, \tau_1, \tau_2, \ldots$ such that $\mathcal{L}[\{S_n\}] = \mathcal{L}[\{B_{\tau_n}\}]$ and $\tau_n - \tau_{n-1}$ are independent and identically distributed.*

Proof. Let $\tau_0 = 0$, and define

$$\tau_n = \inf\{t \geq \tau_{n-1} : \max_{\tau_{n-1} \leq s \leq t} B_t - B_{\tau_{n-1}} \geq \Psi(B_t - B_{\tau_{n-1}})\}\,.$$

By strong Markov property, it's easy $\square$

Appendix A

Process

A.1 Stochastic Processes and σ-Fields

Thus, a stochastic process is a collection of random variables $X = \{X_t : 0 \leq t < \infty\}$ on $(\Omega, \mathscr{F})$, which take values in a second measurable space $(S, \mathscr{S})$, called the state space. For our purposes, the state space $(S, \mathscr{S})$ will be the d-dimensional Euclidean space equipped with the σ-field of Borel sets, i.e., $S = \mathbb{R}^d, \mathscr{S} = \mathscr{B}\left(\mathbb{R}^d\right)$, The index $t \in [0, \infty)$ of the random variables X_t admits a convenient interpretation as time.

Appendix B

Convergence of Measures

One focus of probability theory is distributions that are the result of an interplay of a large number of random impacts. Often a useful approximation can be obtained by taking a limit of such distributions, for example, a limit where the number of impacts goes to infinity. With the Poisson distribution, we have encountered such a limit distribution that occurs as the number of very rare events when the number of possibilities goes to infinity. In many cases, it is necessary to rescale the original distributions in order to capture the behavior of the essential fluctuations, e.g., in the central limit theorem. While these theorems work with real random variables, we will also see limit theorems where the random variables take values in more general spaces such as the space of continuous functions when we model the path of the random motion of a particle.

In this chapter, we provide the abstract framework for the investigation of convergence of measures. We introduce the notion of weak convergence of probability measures on general (mostly Polish) spaces and derive the fundamental properties. We start with a short overview of some topological definitions and theorems.

B.1 A Topology Primer

In the following, let (E, τ) be a Hausdorff topological space with the Borel σ-algebra $\mathcal{B}(E)$. A measure defined on $\mathcal{B}(E)$ is called a *Borel measure*. For $A \subset E$, we denote by $\bar{A}$ the closure of A, by A° the interior and by ∂A the boundary of A.

(E, τ) is called *metrizable* if there exists a metric d on E such that τ is induced by the open balls $B_\varepsilon(x) := \{y \in E : d(x, y) < \varepsilon\}$. A metric d on E is called *complete* if any Cauchy sequence with respect to d converges in E. (E, τ) is called completely metrizable if there exists a complete metric on E that induces τ. If (E, d) is a metric space and $A, B \subset E$, then we write $d(A, B) = \inf\{d(x, y) : x \in A, y \in B\}$ and $d(x, B) := d(\{x\}, B)$ for $x \in E$.

A metrizable space (E, τ) is called *separable* if there exists a countable dense subset of E. Separability in metrizable spaces is equivalent to the existence of a countable base of the topology; that is, a countable set $\mathcal{U} \subset \tau$ with

$$A = \bigcup_{U \in \mathcal{U} : U \subset A} U \quad \text{for all } A \in \tau.$$

(For example, choose the ε-balls centered at the points of a countable subset and let ε run through the positive rational numbers.) A compact metric space is always separable (simply choose for each $n \in \mathbb{N}$ a finite cover $\mathcal{U}_n \subset \tau$ comprising balls of radius $\frac{1}{n}$ and then take $\mathcal{U} := \bigcup_{n \in \mathbb{N}} \mathcal{U}_n$)

Definition B.1. A topological space (E, τ) is called a **Polish space** if it is separable and if there exists a complete metric that induces the topology τ.

Examples of Polish spaces are countable discrete spaces (however, not $\mathbb{Q}$ with the usual topology), the Euclidean spaces $\mathbb{R}^n$, and the space $C([0, 1])$ of continuous functions $[0, 1] \to \mathbb{R}$, equipped with the supremum norm $\|\cdot\|_\infty$. In practice, all spaces that are of importance in probability theory are Polish spaces, and hence we will always suppose that E is a Polish space.

Definition B.2. A σ-finite measure μ on $(E, \mathcal{B}(E))$ is called

(i) **locally finite** if, for any point $x \in E$, there exists an open neighborhood U of x such that $\mu(U) < \infty$;

(ii) **inner regular** if

$$\mu(A) = \sup\{\mu(K) : K \subset A \text{ is compact}\} \quad \text{for all } A \in \mathcal{B}(E) \ ;$$

(iii) **outer regular** if

$$\mu(A) = \inf\{\mu(U) : U \supset A \text{ is open}\} \quad \text{for all } A \in \mathcal{B}(E)$$

(iv) **regular** if μ is inner and outer regular, and

(v) a **Radon measure** if μ is an locally finite inner regular measure.

Remark B.1. The Lebesgue measure λ on $\mathbb{R}^d$ is a regular Radon measure. However, not all σ-finite measures on $\mathbb{R}^d$ are regular.

Consider the measure $\mu = \sum_{q \in \mathbb{Q}} \delta_q$. Clearly, this measure is σ-finite; however, it is neither locally finite nor outer regular.

We introduce the following spaces of measures on E:

$$\begin{aligned}
\mathcal{M}(E) &:= \{\text{Radon measures on } (E, \mathcal{B}(E))\}\,, \\
\mathcal{M}_f(E) &:= \{\text{finite measures on } (E, \mathcal{B}(E))\}\,, \\
\mathcal{M}_1(E) &:= \{\mu \in \mathcal{M}_f(E) : \mu(E) = 1\} \ , \\
\mathcal{M}_{\le 1}(E) &:= \{\mu \in \mathcal{M}_f(E) : \mu(E) \le 1\} \ .
\end{aligned}$$

The elements of $\mathcal{M}_{\le 1}(E)$ are called *sub-probability measures* on E. Further, we agree on the following notation for spaces of continuous functions:

$$\begin{aligned}
C(E) &:= \{f : E \to \mathbb{R} \text{ is continuous }\}\,, \\
C_b(E) &:= \{f \in C(E) \text{ is bounded }\}\,, \\
C_c(E) &:= \{f \in C(E) \text{ has compact support }\}\,.
\end{aligned}$$

Unless otherwise stated, the vector spaces $C(E), C_b(E)$ and $C_c(E)$ are equipped with the supremum norm.

Lemma B.1. *Let $\mu \in \mathcal{M}_f(E)$, then for any $\varepsilon > 0$, there is a compact set $K \subset E$ with $\mu(E \backslash K) < \varepsilon$.*

Proof. Since E is seperable, there exists a countable dense subset $\{x_1, x_2 \cdots\}$ of E. Thus, for each $n \in \mathbb{N}$, $E = \cup_i B_{1/n}(x_i)$. Note that μ is finite, so for any $\varepsilon > 0$ and $n \in \mathbb{N}$, we can find $N_n = N_{n,\varepsilon} \in \mathbb{N}$ such that

$$\mu\left(E \backslash \bigcup_{i=1}^{N_n} B_{1/n}(x_i)\right) < \frac{\varepsilon}{2^n}.$$

Let

$$A = \bigcap_{n=1}^{\infty} \bigcup_{i=1}^{N_n} B_{1/n}(x_i)$$

By construction, A is totally bounded, so $\overline{A}$ is compact. Furthermore, it follows that

$$\mu(E \backslash \overline{A}) \leq \mu(E \backslash A) < \sum_{n=1}^{\infty} \frac{\varepsilon}{2^n} = \varepsilon.$$

We complete the proof. □

Theorem B.2. *If E is Polish and if $\mu \in \mathcal{M}_f(E)$, then μ is regular. In particular, $\mathcal{M}_f(E) \subset \mathcal{M}(E)$.*

Proof. Outer regularity, Step 1. Let $B \subset E$ be closed and let $\varepsilon > 0$. Let d be a complete metric on E compatible with the topology. For $\delta > 0$, let

$$B_\delta := \{x \in E : d(x, B) < \delta\}$$

be the open δ-neighborhood of B. As B is closed, we have $\bigcap_{\delta > 0} B_\delta = B$. Since μ is upper semicontinuous, there is a $\delta > 0$ such that

$$\mu(B_\delta) \leq \mu(B) + \varepsilon.$$

Step 2. Let $B \in \mathcal{B}(E)$ and $\varepsilon > 0$. Consider the class of sets

$$\mathcal{A} := \{V \cap C : V \text{ is open }, C \text{ is closed }\}$$

Clearly, we have $\mathcal{B}(E) = \sigma(\mathcal{A})$. It is easy to check that $\mathcal{A}$ is a semiring. Hence by the approximation theorem for measures, there are mutually disjoint sets $A_n = V_n \cap C_n \in \mathcal{A}, n \in \mathbb{N}$, such that

$$B \subset A := \bigcup_{n=1}^{\infty} A_n \quad \text{and} \quad \mu(A) \leq \mu(B) + \varepsilon/2.$$

As shown in the first step, for any $n \in \mathbb{N}$, there is an open set $W_n \supset C_n$ such that $\mu(W_n) \leq \mu(C_n) + \varepsilon 2^{-n-1}$. Hence also $U_n := V_n \cap W_n$ is open. Let $B \subset U := \bigcup_{n=1}^{\infty} U_n$. We conclude that $\mu(U) \leq \mu(A) + \sum_{n=1}^{\infty} \varepsilon 2^{-n-1} \leq \mu(B) + \varepsilon$.
Inner regularity. Replacing B by B^c, the outer regularity yields the existence of a closed set $D \subset B$ with $\mu(B \backslash D) < \varepsilon/2$. By Lemma B.1, there exists a compact set K with $\mu\left(K^C\right) < \varepsilon/2$. Define $C = D \cap K$. Then $C \subset B$ is compact and $\mu(B \backslash C) < \varepsilon$. Hence μ is also inner regular. □

Let (E, d_E) and (F, d_F) be metric spaces. A function $f : E \to F$ is called *Lipschitz continuous* if there exists a constant $K < \infty$, the so-called Lipschitz constant, with

$$d_F(f(x), f(y)) \leq K \cdot d_E(x, y) \text{ for all } x, y \in E.$$

Denote by $\operatorname{Lip}_K(E; F)$ the space of Lipschitz continuous functions with constant K and by $\operatorname{Lip}(E; F) = \bigcup_{K>0} \operatorname{Lip}_K(E; F)$ the space of Lipschitz continuous functions on E. We abbreviate $\operatorname{Lip}_K(E) := \operatorname{Lip}_K(E; \mathbb{R})$ and $\operatorname{Lip}(E) := \operatorname{Lip}(E; \mathbb{R})$.

A family $\mathcal{C}$ of measurable functions $E \to \mathbb{R}$ is called a separating family for $\mathcal{M}(E)$ if, for any two measures $\mu, \nu \in \mathcal{M}(E)$, the following holds:

$$\left(\int f \, d\mu = \int f \, dv \text{ for all } f \in \mathcal{C} \cap \mathcal{L}^1(\mu) \cap \mathcal{L}^1(v)\right) \Longrightarrow \mu = v\,.$$

Lemma B.3. *Let (E, d) be a metric space. For any closed set $F \subset E$ and any $\varepsilon > 0$, there is a Lipschitz continuous map $\rho_{F,\varepsilon} : E \to [0, 1]$ with*

$$\rho_{F,\varepsilon}(x) = \begin{cases} 1, & \text{if } x \in F\,, \\ 0, & \text{if } d(x, F) \geq \varepsilon\,. \end{cases}$$

Proof. Let

$$\rho_{F,\varepsilon}(x) := \begin{cases} 1 - \frac{d(x,F)}{\varepsilon}\,, & d(x, F) < \varepsilon\,, \\ 0\,, & d(x, F) \geq \varepsilon\,. \end{cases}$$

Clearly, $\rho_{F,\varepsilon} \in \mathrm{Lip}(E; [0, 1])$. ☐

Theorem B.4. *Let (E, d) be a metric space.*

(i) $\mathrm{Lip}_1(E; [0, 1])$ *is separating for $\mathcal{M}(E)$.*

(ii) *If in addition, E is locally compact, then $C_c(E) \cap \mathrm{Lip}_1(E; [0, 1])$ is separating for $\mathcal{M}(E)$.*

Proof. (i). Assume $\mu_1, \mu_2 \in \mathcal{M}(E)$ are measures with $\int f\, d\mu_1 = \int f\, d\mu_2$ for all $f \in \mathrm{Lip}_1(E; [0, 1])$. Since μ_1, μ_2 are inner regular, it is enough to show that $\mu_1(K) = \mu_2(K)$ for any compact set K.

Now let $K \subset E$ be compact. Since μ_i $(i = 1, 2)$ are locally finite, there is a open neighborhood U of K such that $\mu_i(U) < \infty$. Since $K \subset U$, for any $0 < \varepsilon < d(K, U^c)$ we have $K \subset \{x : d(x, K) < \varepsilon\} \subset U$.

Let $\rho_{K,\varepsilon}$ be the map from Lemma B.3, we have $1_K \leq \rho_{K,\varepsilon} \leq 1_U$. since $\rho_{K,\varepsilon} \stackrel{\varepsilon\to 0}{\longrightarrow} 1_K$, we get by dominated convergence that

$$\mu_i(K) = \lim_{\varepsilon\to 0} \int \rho_{K,\varepsilon} d\mu_i\,.$$

Note that $\epsilon\rho_{K,\varepsilon} \in \mathrm{Lip}_1(E; [0, 1]) \cap L^1(\mu_1) \cap L^1(\mu_2)$, by assumption,

$$\int \rho_{K,\varepsilon} d\mu_1 = \varepsilon^{-1} \int (\varepsilon\rho_{K,\varepsilon})\, d\mu_1 = \varepsilon^{-1} \int (\varepsilon\rho_{K,\varepsilon})\, d\mu_2 = \int \rho_{K,\varepsilon} d\mu_2$$

This implies $\mu_1(K) = \mu_2(K)$; hence $\mu_1 = \mu_2$.

(ii). If E is locally compact, then in (i) we can choose the open neighborhoods U of K to be relatively compact. Thus $\rho_{K,\varepsilon}$ has compact support and is thus in $C_c(E)$. □

B.2 Weak and vague convergence

In the lsat section , we saw that integrals of bounded continuous functions f determine a Radon measure on a metric space (E, d). If E is locally compact, it is enough to consider f with compact support. This suggests that we can use $C_b(E)$ and $C_c(E)$ as classes of test functions in order to define the convergence of measures.

Definition B.3. Let E be a Polish space.

(i) Let $\mu, \mu_1, \mu_2, \ldots \in \mathcal{M}_f(E)$. We say that $(\mu_n)_{n\in\mathbb{N}}$ **converges weakly** to μ, formally $\mu_n \to \mu$ (weakly) or $\mu = \text{w-lim}_n \, \mu_n$, if

$$\int f \, d\mu_n \to \int f \, d\mu \quad \text{for all } \; f \in C_b(E) \, .$$

(ii) Let $\mu, \mu_1, \mu_2, \ldots \in \mathcal{M}(E)$. We say that $(\mu_n)_{n\in\mathbb{N}}$ **converges vaguely** to μ formally $\mu_n \to \mu$ (vaguely) or $\mu = \text{v-lim}_n \, \mu_n$, if

$$\int f \, d\mu_n \to \int f \, d\mu \quad \text{for any } \; f \in C_c(E) \, .$$

Remark B.2. While weak convergence implies convergence of the total masses, since $1 \in C_b(E)$, with vague convergence a mass defect (but not a mass gain) can be experienced in the limit.

Bibliography

[1] K. L. Chung and J. L. Doob. Fields, optionality and measurability. *American Journal of Mathematics*, 87(2):397–424, 1965.

[2] R. Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

[3] S. E. S. Ioannis Karatzas. *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics. Springer, 1 edition, 1987.

[4] D. Majerek and W. Zieba. Conditional martingales. *Acta Math. Vietnam*, 32:41–50, 2007.

[5] P. Mörters and Y. Peres. *Brownian motion*, volume 30. Cambridge University Press, 2010.

[6] J. Neveu. Arbres et processus de galton-watson. In *Annales de l'IHP Probabilités et statistiques*, volume 22, pages 199–207, 1986.

[7] Z. Shi. Random walks and trees. In *Esaim: proceedings*, volume 31, pages 1–39. EDP Sciences, 2011.