Chapter XI
Strong Markov Property, Skorokhod Embedding, and Donsker's Invariance Principle

This chapter ties together a number of the topics introduced in the text via applications to the further analysis of Brownian motion, a fundamentally important stochastic process whose existence was established in Chapter VII and, independently, in Chapter IX.

The discrete-parameter random walk was introduced in Chapter II, where it was shown to have the Markov property. Markov processes on a general state space S with a given transition probability $p(x, dy)$ were introduced in Chapter IX (see Example 1 and Remark 9.4 in Chapter IX). Generalizing from this example, a sequence of random variables $\{X_n : n \geq 0\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$ with values in a measurable space $(S, \mathcal{S})$ has the **Markov property** if for every $m \geq 0$, the conditional distribution of X_{m+1} given $\mathcal{F}_m := \sigma(X_j, 0 \leq j \leq m)$ is the same as its conditional distribution given $\sigma(X_m)$. In particular, the conditional distribution is a function of X_m, denoted by $p_m(X_m, dy)$, where $p_m(x, dy)$, $x \in S$ is referred to as the (one-step) **transition probability** at time m and satisfies the following:

1. For $x \in S$, $p_m(x, dy)$ is a probability on $(S, \mathcal{S})$.
2. For $B \in \mathcal{S}$, the function $x \to p_m(x, B)$ is a real-valued measurable function on S.

In the special case that $p_m(x, dy) = p(x, dy)$, for every $m \geq 0$, the transition probabilities are said to be **homogeneous** or **stationary**. Unless stated otherwise, Markov processes considered in this book are homogenous.

With the random walk example as background, let us recall some basic definitions. Let P_z denote the **distribution** of a discrete-parameter stochastic process $X = \{X_n : n \geq 0\}$, i.e., a probability on the product space $(S^\infty, \mathcal{S}^{\otimes\infty})$, with transition probability $p(x, dy)$ and initial distribution $P(X_0 = z) = 1$. The notation $\mathbb{E}_z$ is used to denote expectations with respect to the probability P_z.

Definition 11.1 Fix $m \geq 0$. The **after-m** (future) process is defined by $X_m^+ := \{X_{n+m} : n \geq 0\}$.

R. Bhattacharya and E.C. Waymire, *A Basic Course in Probability Theory*, Universitext, DOI 10.1007/978-3-319-47974-3_XI

It follows from the definition of a Markov process $\{X_n : n = 0, 1, 2, \dots\}$ with a stationary transition probability given above that for every $n \geq 0$ the conditional distribution of $(X_m, X_{m+1}, \dots, X_{m+n})$, given $\sigma(X_0, \dots, X_m)$ is the same as the P_x-distribution of $(X_0, \dots, X_n)$, evaluated at $x = X_m$. To see this, let f be a bounded measurable function on $(S^{n+1}, \mathcal{S}^{\otimes(n+1)})$. Then the claim is that

$$\mathbb{E}\big(f(X_m, X_{m+1}, \dots, X_{m+n})|\sigma(X_0, \dots, X_m)\big) = g_0(X_m), \tag{11.1}$$

where given $X_0 = x$,

$$g_0(x) := \mathbb{E}_x f(X_0, X_1, \dots, X_n). \tag{11.2}$$

For $n = 0$ this is trivial. For $n \geq 1$, first take the conditional expectation of $f(X_m, X_{m+1}, \dots, X_{m+n})$, given $\sigma(X_0, \dots, X_m, \dots, X_{m+n-1})$ to get, by the Markov property, that

$$\begin{aligned} &\mathbb{E}\big(f(X_m, X_{m+1}, \dots, X_{m+n})\ |\sigma(X_0, \dots, X_m, \dots, X_{m+n-1})\big) \\ &= \mathbb{E}\big(f(x_m, \dots, x_{m+n-1}, X_{m+n})|\sigma(X_{m+n-1})\big)|_{x_m=X_m,\dots,x_{m+n-1}=X_{m+n-1}} \\ &= \int_S f(X_m, \dots, X_{m+n-1}, x_{m+n})p(X_{m+n-1}, dx_{m+n}) \\ &= g_{n-1}(X_m, \dots, X_{m+n-1}), \quad \text{say.} \end{aligned} \tag{11.3}$$

Next take the conditional expectation of the above with respect to $\sigma(X_0, \dots, X_{m+n-2})$ to get

$$\begin{aligned} &\mathbb{E}\big(f(X_m, X_{m+1}, \dots, X_{m+n})\ |\sigma(X_0, \dots, X_m, \dots, X_{m+n-2})\big) \\ &= \mathbb{E}\big(g_{n-1}(X_m, \dots, X_{m+n-1})|\sigma(X_0, \dots, X_{m+n-2})\big) \\ &= \mathbb{E}\big(g_{n-1}(x_m, \dots, x_{m+n-2}, X_{m+n-1})|\sigma(X_{m+n-2})\big)|_{x_m=X_m,\dots,x_{m+n-2}=X_{m+n-2}} \\ &= \mathbb{E}\int_S g_{n-1}(X_m, \dots, X_{m+n-2}, x_{m+n-1})p(X_{m+n-2}, dx_{m+n-1}) \\ &= g_{n-2}(X_m, \dots, X_{m+n-2}), \quad \text{say.} \end{aligned} \tag{11.4}$$

Continuing in this manner one finally arrives at

$$\begin{aligned} &\mathbb{E}\big(f(X_m, X_{m+1}, \dots, X_{m+n})\ |\sigma(X_0, \dots, \dots, X_m)\big) \\ &= \mathbb{E}\big(g_1(X_m, X_{m+1})|\sigma(X_0, \dots, \dots, X_m)\big) \\ &= \int_S g_1(X_m, x_{m+1})p(X_m, dx_{m+1}) = g_0(X_m), \quad \text{say.} \end{aligned} \tag{11.5}$$

Now, on the other hand, let us compute $\mathbb{E}_x f(X_0, X_1, \dots, X_n)$. For this, one follows the same steps as above, but with $m = 0$. That is, first take the conditional expectation of $f(X_0, X_1, \dots, X_n)$, given $\sigma(X_0, X_1, \dots, X_{n-1})$, arriving at $g_{n-1}(X_0, X_1, \ldots, X_{n-1})$. Then take the conditional expectation of this given $\sigma(X_0, X_1, \ldots, X_{n-2})$, arriving at $g_{n-2}(X_0, \ldots, X_{n-2})$, and so on. In this way one again arrives at $g_0(X_0)$, which is (11.1) with $m = 0$, or (11.2) with $x = X_m$.

Since finite-dimensional cylinders $C = B \times S^\infty$, $B \in \mathcal{S}^{\otimes(n+1)}$ $(n = 0, 1, 2, \ldots)$ constitute a π-system, and taking $f = \mathbf{1}_B$ in (11.1), (11.2), one has, for every $A \in \sigma(X_0, \ldots, X_m)$,

$$\mathbb{E}\big(\mathbf{1}_A \mathbf{1}_{[X_m^+ \in C]}\big) = \mathbb{E}\big(\mathbf{1}_A \mathbf{1}_{[(X_m, X_{m+1}, \ldots, X_{m+n}) \in B]}\big) = \mathbb{E}\big(\mathbf{1}_A P_x(C)|_{x=X_m}\big), \tag{11.6}$$

it follows from the π-λ theorem that

$$\mathbb{E}\big(\mathbf{1}_A \mathbf{1}_{[X_m^+ \in C]}\big) = \mathbb{E}\big(\mathbf{1}_A P_x(C)|_{x=X_m}\big), \tag{11.7}$$

for all $C \in \mathcal{S}^\infty$; here $P_x(C)|_{x=X_m}$ denotes the (composite) evaluation of the function $x \mapsto P_x(C)$ at $x = X_m$. Thus, we have arrived at the following equivalent, but seemingly stronger, definition of the Markov property.

Definition 11.2 (*Markov Property*) We say that $X = \{X_n : n \geq 0\}$ has the **(homogeneous) Markov Property** if for every $m \geq 0$, the conditional distribution of X_m^+, given the σ-field $\mathcal{F}_m = \sigma(X_0, \ldots, X_m)$, is P_{X_m}, i.e., equals P_y on the set $[X_m = y]$.

This notion may be significantly strengthened by considering the future evolution given its history up to and including a random stopping time. Let us recall that given a stopping time τ, the **pre-τ σ-field** $\mathcal{F}_\tau$ is defined by

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap [\tau = m] \in \mathcal{F}_m, \forall m \geq 0\}. \tag{11.8}$$

Definition 11.3 The **after-τ process** $X_\tau^+ = \{X_\tau, X_{\tau+1}, X_{\tau+2}, \ldots\}$ is well defined on the set $[\tau < \infty]$ by $X_\tau^+ = X_m^+$ on $[\tau = m]$.

The following theorem shows that for discrete-parameter Markov processes, this stronger (Markov) property that "conditionally given the past and the present the future starts afresh at the present state" holds more generally for a stopping time τ in place of a constant "present time" m.

Theorem 11.1 *(Strong Markov Property)* Let τ be a stopping time for the process $\{X_n : n \geq 0\}$. If this process has the Markov property of Definition 11.2, then on $[\tau < \infty]$ the conditional distribution of the after $-\tau$ process X_τ^+, given the pre-τ σ-field $\mathcal{F}_\tau$, is P_{X_τ}.

Proof Let f be a real-valued bounded measurable function on $(S^\infty, \mathcal{S}^{\otimes\infty})$, and let $A \in \mathcal{F}_\tau$. Then

$$\begin{aligned}\mathbb{E}(\mathbf{1}_{[\tau<\infty]}\mathbf{1}_A f(X_\tau^+)) &= \sum_{m=0}^{\infty}\mathbb{E}(\mathbf{1}_{[\tau=m]}\mathbf{1}_A f(X_m^+))\\ &= \sum_{m=0}^{\infty}\mathbb{E}(\mathbf{1}_{[\tau=m]\cap A}\mathbb{E}_{X_m} f)\\ &= \sum_{m=0}^{\infty}\mathbb{E}(\mathbf{1}_{[\tau=m]\cap A}\mathbb{E}_{X_\tau} f) = \mathbb{E}(\mathbf{1}_{[\tau<\infty]}\mathbf{1}_A\mathbb{E}_{X_\tau} f).\end{aligned}$$

The second equality follows from the Markov property in Definition 11.2 since $A \cap [\tau = m] \in \mathcal{F}_m$. ■

Let us now consider the continuous-parameter Brownian motion process along similar lines. It is technically convenient to consider the canonical model of standard Brownian motion $\{B_t : t \geq 0\}$ started at 0 on $\Omega = C[0,\infty)$ with $\mathcal{B}$ the Borel σ-field on $C[0,\infty)$, P_0, referred to as Wiener measure, and $B_t(\omega) := \omega(t)$, $t \geq 0, \omega \in \Omega$, the coordinate projections. However, for continuous-parameter processes it is often useful to make sure that all events that have probability zero are included in the σ-field for Ω. For example, in the analysis of fine-scale structure of Brownian motion certain sets D may arise that *imply* events $E \in \mathcal{B}$, D ⊂ E, for which one is able to compute $P(E) = 0$. In particular, then, one would want to conclude that D is measurable (and hence assigned $P(D) = 0$ too). For this it may be necessary to replace $\mathcal{B}$ by its σ-field completion $\mathcal{F} = \overline{\mathcal{B}}$. We have seen that this can always be achieved, and there is no loss in generality in assuming that the underlying probability space $(\Omega, \mathcal{F}, P)$ is **complete** from the outset (see Appendix A).

Although the focus is on Brownian motion, just as for the above discussion of random walk, some of the definitions apply more generally and will be so stated in terms of a generic continuous-parameter stochastic process $\{Z_t : t \geq 0\}$, having continuous sample paths (outside a P-null set).

Definition 11.4 For fixed $s > 0$ the **after-s** process is defined by $Z_s^+ := \{Z_{s+t} : t \geq 0\}$.

Definition 11.5 A continuous-parameter stochastic process $\{Z_t : t \geq 0\}$, with a.s. continuous sample paths, such that for each $s > 0$, the conditional distribution of the after-s process Z_s^+ given $\sigma(Z_t, t \leq s)$ coincides with its conditional distribution given $\sigma(Z_s)$ is said to have the **Markov property**.

As will become evident from the calculations in the proof below, the Markov property of a Brownian motion $\{B_t : t \geq 0\}$ follows from the fact that it has independent increments.

Proposition 11.2 *(Markov Property of Brownian Motion)* Let P_x denote the distribution on $C[0,\infty)$ of standard Brownian motion $B^x = \{B_t^x = x + B_t : t \geq 0\}$ started at x. For every $s \geq 0$, the conditional distribution of $(B_s^x)^+ := \{B_{s+t}^x : t \geq 0\}$ given $\sigma(B_u^x : 0 \leq u \leq s)$ is $P_{B_s^x}$.

Proof Write $\mathcal{G} := \sigma(B_u^x : 0 \le u \le s)$. Let f be a real-valued bounded measurable function on $C[0,\infty)$. Then $\mathbb{E}\big(f((B_s^x)^+)|\mathcal{G}\big) = \mathbb{E}\big(\psi(U,V)|\mathcal{G}\big)$, where $U = B_s^x$, $V = \{B_{s+t}^x - B_s^x : t \ge 0\}$, $\psi(y,\omega) := f(\omega^y)$, $y \in \mathbb{R}$, $\omega \in C[0,\infty)$, and $\omega^y \in C[0,\infty)$ is given by $\omega^y(t) = \omega(t) + y$. By the substitution property for conditional expectation (Theorem 2.10), one has

$$\mathbb{E}\big(\psi(U,V)|\mathcal{G}\big) = h(U) = h(B_s^x),$$

where simplifying notation by writing $B_t = B_t^0$ and, in turn, $\{B_t : t \ge 0\}$ for a standard Brownian motion starting at 0,

$$h(y) = \mathbb{E}\psi(y,V) = \mathbb{E}\psi(y,\{B_t : t \ge 0\}) = \mathbb{E}f(B^y) = \int_{C[0,\infty)} f\,dP_y.$$

■

It is sometimes useful to extend the definition of standard Brownian motion as follows.

Definition 11.6 Let $(\Omega,\mathcal{F},P)$ be a probability space and $\mathcal{F}_t, t \ge 0$, a filtration. The **k-dimensional standard Brownian motion with respect to this filtration** is a stochastic process $\{B_t : t \ge 0\}$ on $(\Omega,\mathcal{F},P)$ having (i) stationary, Gaussian increments $B_{t+s} - B_s$ with mean zero and covariance matrix tI_k; (ii) a.s. continuous sample paths $t \mapsto B_t$ on $[0,\infty) \to \mathbb{R}^k$; and (iii) for each $t \ge 0$, B_t is $\mathcal{F}_t$-measurable and $B_t - B_s$ is independent of $\mathcal{F}_s, 0 \le s < t$. Taking $B_0 = 0$ a.s., then $B^x := \{x + B_t : t \ge 0\}$, is referred to as the **standard Brownian motion started at** $x \in \mathbb{R}^k$ (with respect to the given filtration). The stochastic process $X_t = x + \mu t + \sigma B_t, t \ge 0$, where $x, \mu \in \mathbb{R}^k$, and σ is a $k \times k$ matrix defines the k-dimensional Brownian motion started at x and having *drift coefficient* μ and *diffusion coefficient* $D = \sigma^t\sigma$.

For example, one may take the completion $\mathcal{F}_t = \overline{\sigma}(B_s : s \le t), t \ge 0$, of the σ-field generated by the coordinate projections $t \mapsto \omega(t), \omega \in C[0,\infty)$. Alternatively, one may have occasion to use $\mathcal{F}_t = \sigma(B_s, s \le t) \vee \mathcal{G}$, where $\mathcal{G}$ is some σ-field independent of $\mathcal{F}$. The definition of the Markov property can be modified accordingly as follows.

Proposition 11.3 The Markov property of Brownian motions B^x on $\mathbb{R}^k$ defined on $(\Omega,\mathcal{F},P)$ holds with respect to (i) the right-continuous filtration defined by

$$\mathcal{F}_{t+} := \bigcap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon} \qquad (t \ge 0), \tag{11.9}$$

where $\mathcal{F}_t = \mathcal{G}_t := \sigma(B_u : 0 \le u \le t)$, or (ii) $\mathcal{F}_t$ is the P-completion of $\mathcal{G}_t$, or (iii) $\mathcal{F}_t = \mathcal{G}_t \vee \mathcal{G}$ $(t \ge 0)$, where $\mathcal{G}$ is independent of $\mathcal{F}$.

Proof It is enough to prove that $B_{t+s} - B_s$ is independent of $\mathcal{F}_{s+}$ for every $t > 0$. Let $G \in \mathcal{F}_{s+}$ and $t > 0$. For each $\varepsilon > 0$ such that $t > \varepsilon$, $G \in \mathcal{F}_{s+\varepsilon}$, so that if $f \in C_b(\mathbb{R}^k)$, one has

$$\mathbb{E}(\mathbf{1}_G f(B_{t+s} - B_{s+\varepsilon})) = P(G) \cdot \mathbb{E} f(B_{t+s} - B_{s+\varepsilon}).$$

Letting $\varepsilon \downarrow 0$ on both sides,

$$\mathbb{E}(\mathbf{1}_G f(B_{t+s} - B_s)) = P(G)\mathbb{E} f(B_{t+s} - B_s).$$

Since the indicator of every closed subset of $\mathbb{R}^k$ is a decreasing limit of continuous functions bounded by 1 (see the proof of Alexandrov's theorem in Chapter VII), the last equality also holds for indicator functions f of closed sets. Since the class of closed sets is a π-system, and the class of Borel sets whose indicator functions f satisfy the equality is a σ-field, one can use the π-λ theorem to obtain the equality for all $B \in \mathcal{B}(\mathbb{R}^k)$. The proofs of (ii) and (iii) are left to Exercise 2. ■

One may define the σ-field governing the "past up to time τ" as the σ-field of events $\mathcal{F}_\tau$ given by

$$\mathcal{F}_\tau := \sigma(Z_{t\wedge\tau} : t \geq 0). \tag{11.10}$$

The stochastic process $\{\tilde{Z}_t : t \geq 0\} := \{Z_{t\wedge\tau} : t \geq 0\}$ is referred to as the **process stopped at** τ. Events in $\mathcal{F}_\tau$ depend only on the process stopped at τ. The stopped process contains no further information about the process $\{Z_t : t \geq 0\}$ beyond the time τ. Alternatively, in analogy with the discrete-parameter case, a description of the past up to time τ that is often more useful for checking whether a particular event belongs to it may be formulated as follows.

Definition 11.7 Let τ be a stopping time with respect to a filtration $\mathcal{F}_t$, $t \geq 0$. The **pre-τ σ-field** is

$$\mathcal{F}_\tau = \{F \in \mathcal{F} : F \cap [\tau \leq t] \in \mathcal{F}_t \text{ for all } t \geq 0\}.$$

For example, using this definition it is simple to check that

$$[\tau \leq t] \in \mathcal{F}_\tau, \forall t \geq 0, \qquad [\tau < \infty] \in \mathcal{F}_\tau. \tag{11.11}$$

Remark 11.1 We will always use[1] Definition 11.7, and not (11.10). Note, however, that $t \wedge \tau \leq t$ for all t, so that $\sigma(X_{t\wedge\tau} : t \geq 0\}$ is contained in $\mathcal{F}_\tau$ (see Exercise 1).

The future relative to τ is the **after-τ process** $Z_\tau^+ = \{(Z_\tau^+)_t : t \geq 0\}$ obtained by viewing $\{Z_t : t \geq 0\}$ from time $t = \tau$ onwards, for $\tau < \infty$. This is

$$(Z_\tau^+)_t(\omega) = Z_{\tau(\omega)+t}(\omega), \qquad t \geq 0, \qquad \text{on } [\tau < \infty]. \tag{11.12}$$

[1]The proof of the equivalence of (11.10) and that of Definition 11.7 for processes with continuous sample paths may be found in Stroock and Varadahn (1980), p. 30.

Theorem 11.4 *(Strong Markov Property for Brownian Motion)* Let $\{B_t : t \geq 0\}$ be a k-dimensional Brownian motion with respect to a filtration $\{\mathcal{F}_t : t \geq 0\}$ starting at 0 and let P_0 denote its distribution (Wiener measure) on $C[0, \infty)$. For $x \in \mathbb{R}^k$ let P_x denote the distribution of the Brownian motion process $B_t^x := x + B_t, t \geq 0$, started at x. Let τ be a stopping time. On $[\tau < \infty]$, the conditional distribution of B_τ^+ given $\mathcal{F}_\tau$ is the same as the distribution of $\{B_t^y : t \geq 0\}$ starting at $y = B_\tau$. In other words, this conditional distribution is P_{B_τ} on $[\tau < \infty]$.

Proof First assume that τ has countably many values ordered as $0 \leq s_1 < s_2 < \cdots$. Consider a finite-dimensional function of the after-τ process of the form

$$h(B_{\tau+t_1'}, B_{\tau+t_2'}, \ldots, B_{\tau+t_r'}), \qquad [\tau < \infty], \tag{11.13}$$

where h is a bounded continuous real-valued function on $(\mathbb{R}^k)^r$ and $0 \leq t_1' < t_2' < \cdots < t_r'$. It is enough to prove

$$\mathbb{E}\left[h(B_{\tau+t_1'}, \ldots, B_{\tau+t_r'})\mathbf{1}_{[\tau<\infty]} \mid \mathcal{F}_\tau\right] = [\mathbb{E}h(B_{t_1'}^y, \ldots, B_{t_r'}^y)]_{y=B_\tau}\mathbf{1}_{[\tau<\infty]}. \tag{11.14}$$

That is, for every $A \in \mathcal{F}_\tau$ we need to show that

$$\mathbb{E}(\mathbf{1}_A h(B_{\tau+t_1'}, \ldots, B_{\tau+t_r'})\mathbf{1}_{[\tau<\infty]}) = \mathbb{E}\left(\mathbf{1}_A \left[\mathbb{E}h(B_{t_1'}^y, \ldots, B_{t_r'}^y)\right]_{y=B_\tau} \mathbf{1}_{[\tau<\infty]}\right). \tag{11.15}$$

Now

$$[\tau = s_j] = [\tau \leq s_j] \cap [\tau \leq s_{j-1}]^c \in \mathcal{F}_{s_j},$$

so that $A \cap [\tau = s_j] \in \mathcal{F}_{s_j}$. Express the left side of (11.15) as

$$\sum_{j=1}^{\infty} \mathbb{E}\big(\mathbf{1}_{A\cap[\tau=s_j]} h(B_{s_j+t_1'}, \ldots, B_{s_j+t_r'})\big). \tag{11.16}$$

By the Markov property, the jth summand in (11.16) equals

$$\mathbb{E}(\mathbf{1}_A\mathbf{1}_{[\tau=s_j]}[\mathbb{E}h(B_{t_1'}^y, \ldots, B_{t_r'}^y)]_{y=B_{s_j}}) = \mathbb{E}(\mathbf{1}_A\mathbf{1}_{[\tau=s_j]}[\mathbb{E}h(B_{t_1'}^y, \ldots, B_{t_r'}^y)]_{y=B_\tau}).$$

Summing this over j, one obtains the desired relation (11.15). This completes the proof in the case that τ has countably many values $0 \leq s_1 < s_2 < \cdots$.

The case of more general τ may be dealt with by approximating it by stopping times assuming countably many values. Specifically, for each positive integer n define

$$\tau_n = \begin{cases} \frac{j}{2n} & \text{if } \frac{j-1}{2^n} < \tau \leq \frac{j}{2^n}, \quad j = 0, 1, 2, \ldots \\ \infty & \text{if } \tau = \infty. \end{cases} \tag{11.17}$$

Since

$$\left[\tau_n = \frac{j}{2^n}\right] = \left[\frac{j-1}{2^n} < \tau \leq \frac{j}{2^n}\right] = \left[\tau \leq \frac{j}{2^n}\right] \setminus \left[\tau \leq \frac{j-1}{2^n}\right] \in \mathcal{F}_{j/2^n},$$

it follows that

$$[\tau_n \leq t] = \bigcup_{j:j/2^n \leq t} \left[\tau_n = \frac{j}{2^n}\right] \in \mathcal{F}_t \qquad \text{for all } t \geq 0.$$

Therefore, τ_n is a stopping time for each n and $\tau_n(\omega) \downarrow \tau(\omega)$ as $n \uparrow \infty$ for each $\omega \in \Omega$. Also one may easily check that $\mathcal{F}_\tau \subset \mathcal{F}_{\tau_n}$ from the definition (see Exercise 1). Let h be a bounded continuous function on $(\mathbb{R}^k)^r$. Define

$$\varphi(y) \equiv \mathbb{E}h(B^y_{t'_1}, \ldots, B^y_{t'_r}). \tag{11.18}$$

One may also check that φ is continuous using the continuity of $y \to (B^y_{t'_1}, \ldots, B^y_{t'_r})$. Let $A \in \mathcal{F}_\tau (\subset \mathcal{F}_{\tau_n})$. Applying (11.15) to $\tau = \tau_n$ one has

$$\mathbb{E}(\mathbf{1}_A h(B_{\tau_n+t'_1}, \ldots, B_{\tau_n+t'_r})\mathbf{1}_{[\tau_n<\infty]}) = \mathbb{E}(\mathbf{1}_A \varphi(B_{\tau_n})\mathbf{1}_{[\tau_n<\infty]}). \tag{11.19}$$

Since h, φ are continuous, $\{B_t : t \geq 0\}$ has continuous sample paths, and $\tau_n \downarrow \tau$ as $n \to \infty$, Lebesgue's dominated convergence theorem may be used on both sides of (11.19) to get

$$\mathbb{E}(\mathbf{1}_A h(B_{\tau+t'_1}, \ldots, B_{\tau+t'_r})\mathbf{1}_{[\tau<\infty]}) = \mathbb{E}(\mathbf{1}_A \varphi(B_\tau)\mathbf{1}_{[\tau<\infty]}). \tag{11.20}$$

This establishes (11.15). Since finite-dimensional distributions determine a probability on $C[0, \infty)$, the proof is complete. ■

Remark 11.2 Note that the proofs of the Markov property (Proposition 11.3) and the strong Markov property (Theorem 11.1) hold for $\mathbb{R}^k$-valued Brownian motions on $\mathbb{R}^k$ with arbitrary drift and positive definite diffusion matrix (Exercise 2).

The examples below illustrate the usefulness of Theorem 11.4 in typical computations. In all these examples $B = \{B_t : t \geq 0\}$ is a one-dimensional standard Brownian motion starting at zero. For $\omega \in C([0, \infty) : \mathbb{R})$ define, for every $a \in \mathbb{R}$,

$$\tau^{(1)}_a(\omega) \equiv \tau_a(\omega) := \inf\{t \geq 0 : \omega(t) = a\}, \tag{11.21}$$

and, recursively,

$$\tau^{(r+1)}_a(\omega) := \inf\{t > \tau^{(r)}_a : \omega(t) = a\}, \quad r \geq 1, \tag{11.22}$$

with the usual convention that the infimum of an empty set of numbers is ∞.

Similarly, in the context of the simple random walk, put $\Omega = \mathbb{Z}^\infty = \{\omega = (\omega_0, \omega_1, \dots) : \omega_n \in \mathbb{Z}, \forall n \geq 1\}$, and define

$$\overline{\tau}_a^{(1)}(\omega) \equiv \overline{\tau}_a(\omega) := \inf\{n \geq 0 : \omega_n = a\}, \tag{11.23}$$

and, recursively,

$$\overline{\tau}_a^{(r+1)}(\omega) := \inf\{n > \overline{\tau}_a^{(r)} : \omega_n = a\}, \quad r \geq 1. \tag{11.24}$$

Example 1 *(Recurrence of Simple Symmetric Random Walk)* Consider the simple symmetric random walk $S^x := \{S_n^x = x + S_n^0 : n \geq 0\}$ on $\mathbb{Z}$ started at x. Suppose one wishes to prove that $P_x(\overline{\tau}_y < \infty) = 1$ for $y \in \mathbb{Z}$. This may be obtained from the (ordinary) Markov property applied to $\varphi(x) := P_x(\overline{\tau}_y < \overline{\tau}_a)$, $a \leq x \leq y$. For $a < x < y$, conditioning on S_1^x, and writing $S_1^{x+} = \{S_{1+n}^x : n \geq 0\}$, we have

$$\begin{aligned}\varphi(x) &= P_x(\overline{\tau}_y < \overline{\tau}_a) = P(\overline{\tau}_y \circ S^x < \overline{\tau}_a \circ S^x) \\ &= P(\overline{\tau}_y \circ S_1^{x+} < \overline{\tau}_a \circ S_1^{x+}) \\ &= \mathbb{E}_x P_{S_1^x}(\overline{\tau}_y < \overline{\tau}_a) = \mathbb{E}\varphi(S_1^x) \\ &= \mathbb{E}(\mathbf{1}_{[S_1^x = x+1]}\varphi(x+1) + \mathbf{1}_{[S_1^x = x-1]}\varphi(x-1)) \\ &= \frac{1}{2}\varphi(x+1) + \frac{1}{2}\varphi(x-1), a < x < y,\end{aligned} \tag{11.25}$$

with boundary values $\varphi(y) = 1$, $\varphi(a) = 0$. Solving, one obtains $\varphi(x) = (x - a)/(y - a)$. Thus $P_x(\tau_y < \infty) = 1$ follows by letting $a \to -\infty$ using basic "continuity properties" of probability measures. Similarly, letting $y \to \infty$, one gets $P_x(\overline{\tau}_a < \infty) = 1$. Write $\overline{\eta}_a := \inf\{n \geq 1 : \omega_n = a\}$ for the **first return time to** a. Then $\overline{\eta}_a = \overline{\tau}_a$ on $\{\omega : \omega_0 \neq a\}$, and $\overline{\eta}_a > \overline{\tau}_a = 0$ on $\{\omega : \omega_0 = a\}$. By conditioning on S_1^x again, one has $P_x(\overline{\eta}_x < \infty) = \frac{1}{2}P_{x-1}(\overline{\tau}_x < \infty) + \frac{1}{2}P_{x+1}(\overline{\tau}_x < \infty) = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 1 = 1$. While this calculation required only the Markov property, next consider the problem of showing that the process will return to y infinitely often. One would like to argue that, conditioning on the process up to its return to y, it merely starts over. This of course is the strong Markov property. So let us examine carefully the calculation to show that under P_x, the rth passage time to y, $\overline{\tau}_y^{(r)}$, is a.s. finite for every $r = 1, 2, \dots$. First note that by the (ordinary) Markov property, $P_x(\tau_y < \infty) = 1\ \forall x$. To simplify notation, write $\tau_y^{(r)} = \overline{\tau}_y^{(r)} \circ S^x$, and $S_{\tau_y^{(r)}}^{x+} = \{S_{\tau_y^{(r)}+n}^{x+} : n \geq 0\}$ is then the after-$\tau_y^{(r)}$ process (for the random walk S^x). Applying the strong Markov property with respect to the stopping time $\tau_y^{(r)}$ one has, remembering that $S_{\tau_y^{(r)}}^x = y$,

$$\begin{aligned}
P(\overline{\tau}_y^{(r+1)} < \infty) &= P(\tau_y^{(r)} < \infty, \overline{\eta} \circ S^{x+}_{\tau_y^{(r)}} < \infty) \\
&= \mathbb{E}\big(\mathbf{1}_{[\tau_y^{(r)} < \infty]} P_y(\overline{\eta} < \infty)\big) \\
&= \mathbb{E}\big(\mathbf{1}_{[\tau_y^{(r)} < \infty]}\big) \cdot 1 \\
&= P(\overline{\tau}_y^{(r)} < \infty) = 1 \quad (r = 1, 2, \ldots),
\end{aligned} \tag{11.26}$$

by induction on r. If $x = y$, then $\overline{\tau}_x^{(1)}$ is replaced by $\overline{\eta}_x$. Otherwise, the proof remains the same. This is equivalent to the **recurrence** of the state y in the sense that

$$P(S_n^x = y \text{ for infinitely many } n) = P(\cap_{r=1}^{\infty}[\tau_y^{(r)} < \infty]) = 1.$$

Example 2 *(Boundary Value Distribution of Brownian Motion)* Let $B^x = \{B_t^x := x + B_t : t \geq 0\}$ be a one-dimensional standard Brownian motion started at $x \in [c, d]$ for $c < d$, and let $\tau_y = \overline{\tau}_y \circ B^x$. The stopping time $\tau_c \wedge \tau_d$ denotes the first time for B^x to reach the "boundary" states $\{c, d\}$, referred to as a **hitting time** for B^x. Define

$$\psi(x) := P(B^x_{\tau_c \wedge \tau_d} = c) \equiv P(\{B_t^x : t \geq 0\} \text{ reaches } c \text{ before } d), \qquad (c \leq x \leq d). \tag{11.27}$$

Fix $x \in (c, d)$ and $h > 0$ such that $[x - h, x + h] \subset (c, d)$. In contrast to the discrete-parameter case there is no "first step" to consider. It will be convenient to consider $\tau = \tau_{x-h} \wedge \tau_{x+h}$, i.e., τ is the first time $\{B_t^x : t \geq 0\}$ reaches $x - h$ or $x + h$. Then $P(\tau < \infty) = 1$, by the law of the iterated logarithm (see Exercise 6 for an alternative argument). Now, by the strong Markov property (Theorem 11.4), applied with respect to τ,

$$\begin{aligned}
\psi(x) &= P(\{B_t^x : t \geq 0\} \text{ reaches } c \text{ before } d) \\
&= P(\{(B_\tau^{x+})_t : t \geq 0\} \text{ reaches } c \text{ before } d) \\
&= \mathbb{E}(P(\{(B_\tau^{x+})_t : t \geq 0\} \text{ reaches } c \text{ before } d \mid \mathcal{F}_\tau)).
\end{aligned} \tag{11.28}$$

The strong Markov property now gives that

$$\psi(x) = \mathbb{E}(\psi(B_\tau^x)), \tag{11.29}$$

so that by symmetry of Brownian motion, i.e., B^0 and $-B^0$ have the same distribution,

$$\begin{aligned}
\psi(x) &= \psi(x-h)P(B_\tau^x = x - h) + \psi(x+h)P(B_\tau^x = x + h) \\
&= \psi(x-h)\frac{1}{2} + \psi(x+h)\frac{1}{2},
\end{aligned} \tag{11.30}$$

where, by (11.27), $\psi(x)$ satisfies the boundary conditions $\psi(c) = 1$, $\psi(d) = 0$. Therefore,

$$\psi(x) = \frac{d - x}{d - c}. \tag{11.31}$$

Now, by (11.31) (see also Exercise 6),

$$P(\{B_t^x : t \geq 0\} \text{ reaches } d \text{ before } c) = 1 - \psi(x) = \frac{x - c}{d - c} \tag{11.32}$$

for $c \leq x \leq d$. It follows, on letting $d \uparrow \infty$ in (11.31), and $c \downarrow -\infty$ in (11.32) that

$$P(\overline{\tau}_y < \infty) = 1 \qquad \text{for all } x, y. \tag{11.33}$$

As another illustrative application of the strong Markov property one may derive a Cantor-like structure of the random set of zeros of Brownian motion as follows.

Example 3.

Proposition 11.5 With probability one, the set $\mathcal{Z} := \{t \geq 0 : B_t = 0\}$ of zeros of the sample path of a one dimensional standard Brownian motion, starting at 0, is uncountable, closed, unbounded, and has no isolated point. Moreover, $\mathcal{Z}$ a.s. has Lebesgue measure zero.

Proof The law of iterated logarithm (LIL) may be applied as $t \downarrow 0$ to show that with probability one, $B_t = 0$ for infinitely many t in every interval $[0, \varepsilon]$. Since $t \mapsto B_t(\omega)$ is continuous, $\mathcal{Z}(\omega)$ is closed. Applying the LIL as $t \uparrow \infty$, it follows that $\mathcal{Z}(\omega)$ is unbounded a.s.

We will now show that for $0 < c < d$, the probability is zero of the event $A(c, d)$, say, that B has a single zero in $[c, d]$. For this consider the stopping time $\tau := \inf\{t \geq c : B_t = 0\}$. By the strong Markov property, B_τ^+ is a standard Brownian motion, starting at zero. In particular, τ is a point of accumulation of zeros from the right (a.s.). Also, $P(B_d = 0) = 0$. This implies $P(A(c, d)) = 0$. Considering all pairs of rationals c, d with $c < d$, it follows that $\mathcal{Z}$ has no isolated point outside a set of probability zero (see Exercise 4 for an alternate argument).

Finally, for each $T > 0$ let $H_T = \{(t, \omega) : 0 \leq t \leq T, B_t(\omega) = 0\} \subset [0, T] \times \Omega$. By the Fubini–Tonelli theorem, denoting the Lebesgue measure on $[0, \infty)$ by m, one has

$$(m \times P)(H_T) = \int_0^T \left\{ \int_\Omega \mathbf{1}_{[B_t = 0]}(\omega) P(d\omega) \right\} dt = \int_0^T P(B_t = 0) dt = 0, \tag{11.34}$$

so that $m(\{t \in [0, T] : B_t(\omega) = 0\}) = 0$ for P-almost all ω. ■

The following general consequence of the Markov property can also be useful in the analysis of the (infinitesimal) fine-scale structure of Brownian motion and may be viewed as a corollary to Proposition 11.3. As a consequence, for example, one sees that for any given function $\varphi(t)$, $t > 0$, the event

$$D_\varphi := [B_t < \varphi(t) \text{ for all sufficiently small } t] \tag{11.35}$$

will certainly occur or is certain not to occur. Functions φ for which $P(D_\varphi) = 1$ are said to belong to the **upper class**. Thus $\varphi(t) = \sqrt{2t \log \log t}$ belongs to the upper class by the law of the iterated logarithm for Brownian motion (Theorem 10.9).

Proposition 11.6 *(Blumenthal's Zero–One Law)* With the notation of Proposition 11.3,

$$P(A) = 0 \text{ or } 1 \qquad \forall\, A \in \mathcal{F}_{0+}. \tag{11.36}$$

Proof It follows from (the proof of) Proposition 11.3 that $\mathcal{F}_{s+}$ is independent of $\sigma\{B_{t+s} - B_s : t \geq 0\}\ \forall\, s \geq 0$. Set $s = 0$ to conclude that $\mathcal{F}_{0+}$ is independent of $\sigma(B_t : t \geq 0) \supset \mathcal{F}_{0+}$. Thus $\mathcal{F}_{0+}$ is independent of $\mathcal{F}_{0+}$, so that $\forall\, A \in \mathcal{F}_{0+}$ one has $P(A) \equiv P(A \cap A) = P(A) \cdot P(A)$. ■

In addition to the strong Markov property, another powerful tool for the analysis of Brownian motion is made available by observing that both the processes $\{B_t : t \geq 0\}$ and $\{B_t^2 - t : t \geq 0\}$ are martingales. Thus one has available the optional sampling theory.

Example 4 *(Hitting by BM of a Two-Point Boundary)* Let $\{B_t^x : t \geq 0\}$ be a one-dimensional standard Brownian motion starting at x, and let $c < x < d$. Let τ denote the stopping time, $\tau = \inf\{t \geq 0 : B_t^x = c \text{ or } d\}$. Then writing $\psi(x) := P(\{B_t^x\}_{t\geq 0}$ reaches d before c), one has (see (11.31))

$$\psi(x) = \frac{x-c}{d-c} \qquad c < x < d. \tag{11.37}$$

Applying the optional sampling theorem to the martingale $X_t := (B_t^x - x)^2 - t$, one gets $\mathbb{E}X_\tau = 0$, or $(d-x)^2\psi(x) + (x-c)^2(1-\psi(x)) = \mathbb{E}\tau$, so that $\mathbb{E}\tau = [(d-x)^2 - (x-c)^2]\psi(x) + (x-c)^2$, or

$$\mathbb{E}\tau = (d-x)(x-c). \tag{11.38}$$

Consider now a Brownian motion $\{Y_t^x : t \geq 0\}$ with nonzero drift coefficient μ and diffusion coefficient $\sigma^2 > 0$, starting at x. Then $\{Y_t^x - t\mu : t \geq 0\}$ is a martingale, so that (see Exercise 6) $\mathbb{E}(Y_\tau^x - \mu\tau) = x$, i.e., $d\psi_1(x) + c(1-\psi_1(x)) - \mu\mathbb{E}\tau = x$, or

$$(d-c)\psi_1(x) - \mu\mathbb{E}\tau = x - c, \tag{11.39}$$

where $\psi_1(x) = P(Y_\tau^x = d)$, i.e., $\{Y_t^x : t \geq 0\}$ reaches d before c. There are two unknowns, ψ_1 and $\mathbb{E}\tau$ in (11.39), so we need one more relation to solve for them. Consider the exponential martingale $Z_t := \exp\left\{\xi(Y_t^x - t\mu) - \frac{\xi^2\sigma^2}{2}t\right\}$ $(t \geq 1)$. Then $Z_0 = e^{\xi x}$, so that $e^{\xi x} = \mathbb{E}Z_\tau = \mathbb{E}\exp\{\xi(d-\tau\mu) - \xi^2\sigma^2\,\tau/2\}\mathbf{1}_{[Y_\tau^x = d]} + \mathbb{E}[\exp\{\xi(c - \tau\mu) - \xi^2\sigma^2\,\tau/2\}\mathbf{1}_{[Y_\tau^x = c]}]$. Take $\xi \neq 0$ such that the coefficient of τ in the exponent is zero, i.e., $\xi\mu + \xi^2\,\sigma^2/2 = 0$, or $\xi = -2\mu/\sigma^2$. Then optional stopping yields

$$
\begin{aligned}
e^{-2\mu x/\sigma^2} &= \exp\{\xi d\}\psi_1(x) + \exp\{\xi c\}(1 - \psi_1(x)), \\
&= \psi_1(x)\left[\exp\left\{-\frac{2\mu d}{\sigma^2}\right\} - \exp\left\{-\frac{2\mu c}{\sigma^2}\right\}\right] + \exp\left\{-\frac{2\mu c}{\sigma^2}\right\},
\end{aligned}
$$

or

$$
\psi_1(x) = \frac{\exp\{-2\mu x/\sigma^2\} - \exp\{-2\mu c/\sigma^2\}}{\exp\{-\frac{2\mu d}{\sigma^2}\} - \exp\{-\frac{2\mu c}{\sigma^2}\}}. \tag{11.40}
$$

One may use this to compute $\mathbb{E}\tau$:

$$
\mathbb{E}\tau = \frac{(d-c)\psi_1(x) - (x-c)}{\mu}. \tag{11.41}
$$

Checking the hypothesis of the optional sampling theorem for the validity of the relations (11.37)–(11.41) is left to Exercise 6.

Our main goal for this chapter is to derive a beautiful result of Skorokhod (1965) representing a general random walk (partial sum process) as values of a Brownian motion at a sequence of successive stopping times (with respect to an enlarged filtration). This will be followed by a proof of the functional central limit theorem (invariance principle) based on the Skorokhod embedding representation. Recall that for $c < x < d$,

$$
P(\tau_d^x < \tau_c^x) = \frac{x-c}{d-c}, \tag{11.42}
$$

where $\tau_a^x := \overline{\tau}_a(B^x) \equiv \inf\{t \geq 0 : B_t^x = a\}$. Also,

$$
\mathbb{E}(\tau_c^x \wedge \tau_d^x) = (d-x)(x-c). \tag{11.43}
$$

Write $\tau_a = \tau_a^0$, $B^0 = B = \{B_t : t \geq 0\}$. Consider now a two-point distribution $F_{u,v}$ with support $\{u, v\}$, $u < 0 < v$, having mean zero. That is, $F_{u,v}(\{u\}) = v/(v-u)$ and $F_{u,v}(\{v\}) = -u/(v-u)$. It follows from (11.42) that with $\tau_{u,v} = \tau_u \wedge \tau_v$, $B_{\tau_{u,v}}$ has distribution $F_{u,v}$ and, in view of (11.43),

$$
\mathbb{E}\tau_{u,v} = -uv = |uv|. \tag{11.44}
$$

In particular, the random variable $Z := B_{\tau_{u,v}}$ with distribution $F_{u,v}$ is naturally **embedded** in the Brownian motion. We will see by the theorem below that any given non-degenerate distribution F with mean zero may be similarly embedded by randomizing over such pairs (u, v) to get a random pair (U, V) such that $B_{\tau_{U,V}}$ has distribution F, and $\mathbb{E}\tau_{U,V} = \int_{(-\infty,\infty)} x^2 F(dx)$, the variance of F. Indeed, this is achieved by the distribution γ of (U, V) on $(-\infty, 0) \times (0, \infty)$ given by

$$
\gamma(du\, dv) = \theta(v-u)F_-(du)F_+(dv), \tag{11.45}
$$

where F_+ and F_- are the restrictions of F to $(0, \infty)$ and $(-\infty, 0)$, respectively. Here θ is the normalizing constant given by

$$1 = \theta\left[\left(\int_{(0,\infty)} vF_+(dv)\right) F_-((-\infty, 0)) + \left(\int_{(-\infty,0)} (-u)F_-(du)\right) F_+(0, \infty)\right],$$

or, noting that the two integrals are each equal to $\frac{1}{2}\int_{-\infty}^{\infty} |x|F(dx)$ since the mean of F is zero, one has

$$1/\theta = \left(\frac{1}{2}\int_{-\infty}^{\infty} |x|F(dx)\right)[1 - F(\{0\})]. \tag{11.46}$$

Let $(\Omega, \mathcal{F}, P)$ be a probability space on which are defined (1) a standard Brownian motion $B \equiv B^0 = \{B_t : t \geq 0\}$, and (2) a sequence of i.i.d. pairs (U_i, V_i) independent of B, with the common distribution γ above. Let $\mathcal{F}_t := \sigma\{B_s : 0 \leq s \leq t\} \vee \sigma\{(U_i, V_i) : i \geq 1\}, t \geq 0$. Define the $\{\mathcal{F}_t : t \geq 0\}$-stopping times (Exercise 13)

$$T_0 \equiv 0, \quad T_1 := \inf\{t \geq 0 : B_t = U_1 \text{ or } V_1\},$$
$$T_{i+1} := \inf\{t > T_i : B_t = B_{T_i} + U_{i+1} \text{ or } B_{T_i} + V_{i+1}\}\ (i \geq 1).$$

Theorem 11.7 *(Skorokhod Embedding)* Assume that F has mean zero and finite variance. Then (a) B_{T_1} has distribution F, and $B_{T_{i+1}} - B_{T_i}$ $(i \geq 0)$ are i.i.d. with common distribution F, and (b) $T_{i+1} - T_i$ $(i \geq 0)$ are i.i.d. with

$$\mathbb{E}\,(T_{i+1} - T_i) = \int_{(-\infty,\infty)} x^2 F(dx). \tag{11.47}$$

Proof (a) Given (U_1, V_1), the conditional probability that $B_{T_1} = V_1$ is $\frac{-U_1}{V_1-U_1}$. Therefore, for all $x > 0$,

$$\begin{aligned} P\left(B_{T_1} > x\right) &= \theta\int_{\{v>x\}}\int_{(-\infty,0)} \frac{-u}{v-u}\cdot(v-u)F_-(du)F_+(dv) \\ &= \theta\int_{\{v>x\}}\left\{\int_{(-\infty,0)} (-u)F_-(du)\right\} F_+(dv) \\ &= \int_{\{v>x\}} F_+(dv), \end{aligned} \tag{11.48}$$

since $\int_{(-\infty,0)}(-u)F_-(du) = \frac{1}{2}\int |x|F(dx) = 1/\theta$. Thus the restriction of the distribution of B_{T_1} on $(0, \infty)$ is F_+. Similarly, the restriction of the distribution of B_{T_1} on $(-\infty, 0)$ is F_-. It follows that $P(B_{T_1} = 0) = F(\{0\})$. This shows that B_{T_1} has distribution F. Next, by the strong Markov property, the conditional distribution of $B_{T_i}^+ \equiv \{B_{T_i+t} : t \geq 0\}$, given $\mathcal{F}_{T_i}$, is $P_{B_{T_i}}$ (where P_x is the distribution of B^x).

Therefore, the conditional distribution of $B^+_{T_i} - B_{T_i} \equiv \{B_{T_i+t} - B_{T_i}; t \geq 0\}$, given $\mathcal{F}_{T_i}$, is P_0. In particular, $Y_i := \{(T_j, B_{T_j}) : 1 \leq j \leq i\}$ and $X^i := B^+_{T_i} - B_{T_i}$ are independent. Since Y_i and X^i are functions of $B \equiv \{B_t : t \geq 0\}$ and $\{(U_j, V_j); 1 \leq j \leq i\}$, they are both independent of (U_{i+1}, V_{i+1}). Since $\tau^{(i+1)} := T_{i+1} - T_i$ is the first hitting time of $\{U_{i+1}, V_{i+1}\}$ by X^i, it now follows that (1) $(T_{i+1} - T_i \equiv \tau^{(i+1)}, B_{T_{i+1}} - B_{T_i} \equiv X^i_{\tau^{(i+1)}})$ is independent of $\{(T_j, B_{T_j}) : 1 \leq j \leq i\}$, and (2) $(T_{i+1} - T_i, B_{T_{i+1}} - B_{T_i})$ has the same distribution as (T_1, B_{T_1}).

(b) It remains to prove (11.47). But this follows from (11.44):

$$\begin{aligned}
\mathbb{E}T_1 &= \theta \int_{(0,\infty)} \int_{(-\infty,0)} (-uv)(v-u) F_-(du) F_+(dv) \\
&= \theta \left[\int_{(0,\infty)} v^2 F_+(dv) \cdot \int_{(-\infty,0)} (-u) F_-(du) + \int_{(-\infty,0)} u^2 F_-(du) \cdot \int_{(0,\infty)} v F_+(dv) \right] \\
&= \int_{(0,\infty)} v^2 F_+(dv) + \int_{(-\infty,0)} u^2 F_-(du) = \int_{(-\infty,\infty)} x^2 F(dx).
\end{aligned}$$

■

We now present an elegant proof of **Donsker's invariance principle**, or **functional central limit theorem**, using Theorem 11.7. Consider a sequence of i.i.d. random variables Z_i $(i \geq 1)$ with common distribution having mean zero and variance 1. Let $S_k = Z_1 + \cdots + Z_k$ $(k \geq 1)$, $S_0 = 0$, and define the polygonal random function $S^{(n)}$ on [0, 1] as follows:

$$S_t^{(n)} := \frac{S_{k-1}}{\sqrt{n}} + n\left(n - \frac{k-1}{n}\right) \frac{S_k - S_{k-1}}{\sqrt{n}} \quad \text{for } t \in \left[\frac{k-1}{n}, \frac{k}{n}\right], 1 \leq k \leq n. \tag{11.49}$$

That is, $S_t^{(n)} = \frac{S_k}{\sqrt{n}}$ at points $t = \frac{k}{n}$ $(0 \leq k \leq n)$, and $t \mapsto S_t^{(n)}$ is linearly interpolated between the endpoints of each interval $\left[\frac{k-1}{n}, \frac{k}{n}\right]$.

Theorem 11.8 *(Invariance Principle)* $S^{(n)}$ converges in distribution to the standard Brownian motion, as $n \to \infty$.

Proof Let T_k, $k \geq 1$, be as in Theorem 11.7, defined with respect to a standard Brownian motion $\{B_t : t \geq 0\}$. Then the random walk $\{S_k : k = 0, 1, 2, \ldots\}$ has the same distribution as $\{\widetilde{S}_k := B_{T_k} : k = 0, 1, 2, \ldots\}$, and therefore, $S^{(n)}$ has the same distribution as $\widetilde{S}^{(n)}$ defined by $\widetilde{S}^{(n)}_{k/n} := n^{-\frac{1}{2}} B_{T_K}$ $(k = 0, 1, \ldots, n)$ and with linear interpolation between k/n and $(k+1)/n$ $(k = 0, 1, \ldots, n-1)$. Also, define, for each $n = 1, 2, \ldots$, the standard Brownian motion $\widetilde{B}^{(n)}_t := n^{-\frac{1}{2}} B_{nt}$, $t \geq 0$. We will show that

$$\max_{0 \leq t \leq 1} \left| \widetilde{S}^{(n)}_t - \widetilde{B}^{(n)}_t \right| \longrightarrow 0 \quad \text{in probability as } n \to \infty, \tag{11.50}$$

which implies the desired weak convergence. Now

$$\max_{0\le t\le 1}\left|\widetilde{S}_t^{(n)}-\widetilde{B}_t^{(n)}\right| \le n^{-\frac{1}{2}}\max_{1\le k\le n}\left|B_{T_k}-B_k\right|$$
$$+\max_{0\le k\le n-1}\left\{\max_{\frac{k}{n}\le t\le\frac{k+1}{n}}\left|\widetilde{S}_t^{(n)}-\widetilde{S}_{k/n}^{(n)}\right|+n^{-\frac{1}{2}}\max_{k\le t\le k+1}|B_t-B_k|\right\}$$
$$= I_n^{(1)}+I_n^{(2)}+I_n^{(3)},\quad \text{say.} \tag{11.51}$$

Now, writing $\tilde{Z}_k=\tilde{S}_k-\tilde{S}_{k-1}$, it is simple to check (Exercise 14) that as $n\to\infty$,

$$I_n^{(2)}\le n^{-\frac{1}{2}}\max\{|\tilde{Z}_k|:1\le k\le n\}\to 0 \quad \text{in probability,}$$
$$I_n^{(3)}\le n^{-\frac{1}{2}}\max_{0\le k\le n-1}\max\{|B_t-B_k|:k\le t\le k+1\}\to 0 \quad \text{in probability.}$$

Hence we need to prove, as $n\to\infty$,

$$I_n^{(1)}:=n^{-\frac{1}{2}}\max_{1\le k\le n}\left|B_{T_k}-B_k\right|\longrightarrow 0 \quad \text{in probability.} \tag{11.52}$$

Since $T_n/n\to 1$ a.s., by SLLN, it follows that (Exercise 14)

$$\varepsilon_n:=\max_{1\le k\le n}\left|\frac{T_k}{n}-\frac{k}{n}\right|\longrightarrow 0 \quad \text{as } n\to\infty \text{ (almost surely).} \tag{11.53}$$

In view of (11.53), there exists for each $\varepsilon>0$ an integer n_ε such that $P(\varepsilon_n<\varepsilon)>1-\varepsilon$ for all $n\ge n_\varepsilon$. Hence with probability greater than $1-\varepsilon$ one has for all $n\ge n_\varepsilon$ the estimate

$$I_n^{(1)}\le \max_{\substack{|s-t|\le n\varepsilon,\\ 0\le s,t\le n+n\varepsilon}} n^{-\frac{1}{2}}|B_s-B_t| = \max_{\substack{|s-t|\le n\varepsilon,\\ 0\le s,t\le n(1+\varepsilon)}}\left|\widetilde{B}_{s/n}^{(n)}-\widetilde{B}_{t/n}^{(n)}\right|$$
$$= \max_{\substack{|s'-t'|\le\varepsilon,\\ 0\le s',t'\le 1+\varepsilon}}\left|\widetilde{B}_{s'}^{(n)}-\widetilde{B}_{t'}^{(n)}\right| \stackrel{d}{=} \max_{\substack{|s'-t'|\le\varepsilon,\\ 0\le s',t'\le 1+\varepsilon}}|B_{s'}-B_{t'}|$$
$$\longrightarrow 0 \quad \text{as } \varepsilon\downarrow 0,$$

by the continuity of $t\to B_t$. Given $\delta>0$ one may then choose $\varepsilon=\varepsilon_\delta$ such that for all $n\ge n(\delta):=n_{\varepsilon_\delta}$, $P(I_n^{(1)}>\delta)<\delta$. Hence $I_n^{(1)}\to 0$ in probability. ■

For another application of Skorokhod embedding let us see how to obtain a **law of the iterated logarithm** (LIL) for sums of i.i.d. random variables using the LIL for Brownian motion.

Theorem 11.9 *(Law of the Iterated Logarithm)* Let $X_1, X_2, \ldots$ be an i.i.d. sequence of random variables with $\mathbb{E}X_1=0$, $0<\sigma^2:=\mathbb{E}X_1^2<\infty$, and let $S_n=X_1+\cdots+X_n$, $n\ge 1$. Then with probability one,

$$\limsup_{n\to\infty}\frac{S_n}{\sqrt{2\sigma^2 n\log\log n}}=1.$$

Proof By rescaling if necessary, one may take $\sigma^2 = 1$ without loss of generality. In view of the Skorokhod embedding one may replace the sequence $\{S_n : n \geq 0\}$ by the embedded random walk $\{\tilde{S}_n = B_{T_n} : n \geq 0\}$. By the SLLN one also has $\frac{T_n}{n} \to 1$ a.s. as $n \to \infty$. In view of the law of the iterated logarithm for Brownian motion, it is then sufficient to check that $\frac{\tilde{S}_{[t]} - B_t}{\sqrt{t \log\log t}} \to 0$ a.s. as $t \to \infty$. From $\frac{T_n}{n} \to 1$ a.s., it follows for given $\varepsilon > 0$ that with probability one, $\frac{1}{1+\varepsilon} < \frac{T_{[t]}}{t} < 1 + \varepsilon$ for all t sufficiently large. Let $t_n = (1+\varepsilon)^n$, $n = 1, 2, \ldots$. Then one has

$$
\begin{aligned}
M_{t_n} &:= \max\left\{|B_s - B_{t_n}| : \frac{t_n}{1+\varepsilon} \leq s \leq t_n(1+\varepsilon)\right\} \\
&\leq \max\left\{|B_s - B_{t_n}| : \frac{t_n}{1+\varepsilon} \leq s \leq t_n\right\} + \max\left\{|B_s - B_{t_n}| : t_n \leq s \leq t_n(1+\varepsilon)\right\} \\
&\leq M_{n,1} + M_{n,2}, say.
\end{aligned}
$$

Since $t_n - \frac{t_n}{1+\varepsilon} = \frac{t_n\varepsilon}{1+\varepsilon} < t_n(1+\varepsilon) - t_n = t_n\varepsilon$, $M_{n,2}$ is stochastically larger than $M_{n,1}$, so that $P(M_{t_n} > 2\sqrt{3\varepsilon t_n \log\log t_n}) \leq 2P(M_{n,2} > \sqrt{3\varepsilon t_n \log\log t_n})$. It follows from the scaling property of Brownian motion, using Lévy's Inequality and Feller's tail probability estimate, that

$$
\begin{aligned}
P\left(M_{t_n} > 2\sqrt{3\varepsilon t_n \log\log t_n}\right) &\leq 2P\left(\max_{0 \leq u \leq 1} |B_u| > \sqrt{3 \log\log t_n}\right) \\
&\leq 8P\left(B_1 \geq \sqrt{3 \log\log(t_n)}\right) \\
&\leq \frac{8}{\sqrt{3\log\log t_n}} \exp\left(-\frac{3}{2}\log\log t_n\right) \\
&\leq cn^{-\frac{3}{2}}
\end{aligned}
$$

for a constant $c = (\log(1+\varepsilon))^{\frac{-3}{2}} > 0$. Summing over n, it follows from the Borel–Cantelli lemma I that with probability one, $M_{t_n} \leq \sqrt{3\varepsilon t_n \log\log t_n}$ for all but finitely many n. Since a.s. $\frac{1}{1+\varepsilon} < \frac{T_{[t]}}{t} < 1 + \varepsilon$ for all t sufficiently large, one has that

$$
\limsup_{t \to \infty} \frac{|\tilde{S}_{[t]} - B_t|}{\sqrt{t \log\log t}} \leq \sqrt{3\varepsilon}.
$$

Letting $\varepsilon \downarrow 0$ one has the desired result. ■

Exercise Set XI

1. (i) If τ_1, τ_2 are stopping times, show that $\tau_1 \vee \tau_2$ and $\tau_1 \wedge \tau_2$ are stopping times.
   (ii) If $\tau_1 \leq \tau_2$ are stopping times, show that $\mathcal{F}_{\tau_1} \subset \mathcal{F}_{\tau_2}$.

2. (i) Extend the Markov property for one-dimensional Brownian motion (Proposition 11.2) to k-dimensional Brownian motion with respect to a given filtration. (ii) Prove parts (ii), (iii) of Proposition 11.3.
3. Suppose that X, Y, Z are three random variables with values in arbitrary measurable spaces $(S_i, \mathcal{S}_i)$, $i = 1, 2, 3$. Assume that regular conditional distributions exist; see Chapter II for general conditions. Show that $\sigma(Z)$ is conditionally independent of $\sigma(X)$ given $\sigma(Y)$ if and only if the conditional distribution of Z given $\sigma(Y)$ a.s. coincides with the conditional distribution of Z given $\sigma(X, Y)$.
4. Prove that the event $A(c, d)$ introduced in the proof of Proposition 11.5 is measurable, i.e., the event $[\tau < d, B_t > 0\ \forall \tau < t \leq d]$ is measurable.
5. Consider a Markov chain $X = \{X_n : n = 0, 1, 2 \dots\}$ on a countable state space. Assume i is (point) recurrent: $P(X_n = i \text{ i.o.}|\mathrm{X}_0 = \mathrm{i}) = 1$. If j is a state such that $p_{ij}^{(n)} > 0$ for some n, prove that (i) the probability that j is reached starting from i is one, and (ii) j is (point) recurrent. [*Hint*: Consider visiting j between successive returns to i as i.i.d. events.]
6. Check the conditions for the application of the optional sampling theorem (Theorem 3.8(b)) for deriving (11.37)–(11.41). [*Hint*: For Brownian motion $\{Y_t^x : t \geq 0\}$ with a drift μ and diffusion coefficient $\sigma^2 > 0$, let $Z_1 = Y_1^x - x$, $Z_k = Y_k^x - Y_{k-1}^x (k \geq 1)$. Then $Z_1, Z_2, \dots$ are i.i.d. and Corollary 3.10 applies with $a = c, b = d$. This proves $P(\tau < \infty) = 1$. The uniform integrability of $\{Y_{t\wedge\tau}^x : t \geq 0\}$ is immediate, since $c \leq Y_{t\wedge\tau}^x \leq d$ for all $t \geq 0$.]
7. Let $u' < 0 < v'$. Show that if $F = F_{u',v'}$ is the mean-zero two-point distribution concentrated at $\{u', v'\}$, then $P((U, V) = (u', v')) = 1$ in the Skorokhod embedding of F defined by $\gamma(du\, dv)$.
8. Given any distribution F on $\mathbb{R}$, let $\tau := \inf\{t \geq 0 : B_t = Z\}$, where Z is independent of $B = \{B_t : t \geq 0\}$ and has distribution F. Then $B_\tau = Z$. One can thus embed a random walk with (a nondegenerate) step distribution F (say, with mean zero) in different ways. However, show that $\mathbb{E}\tau = \infty$. [*Hint*: The stable distribution of $\tau_a := \inf\{t \geq 0 : B_t = a\}$ has infinite mean for every $a \neq 0$. To see this, use Corollary 10.6 to obtain $P(\tau_a > t) \geq 1 - 2P(B_t > a) = P(|B_t| \leq a) = P(|B_1| \leq \frac{a}{\sqrt{t}})$, whose integral over $[0, \infty)$ is divergent.]
9. Prove that $\varphi(\lambda) := \mathbb{E}\exp\{\lambda\tau_{u,v}\} \leq \mathbb{E}\exp\{\lambda\tau_{-a,a}\} < \infty$ for $\lambda < \lambda_0(a)$ for some $\lambda_0(a) > 0$, where $a = \max\{-u, v\}$. Here $\tau_{u,v}$ is the first passage time of standard Brownian motion to $\{u, v\}$, $u < 0 < v$. [*Hint*: Use Corollary 3.10 with $X_n := B_n - B_{n-1}$ $(n \geq 1)$.]
10. (i) Show that for every $\lambda \geq 0$, $X_t := \exp\{\sqrt{2\lambda}B_t - \lambda t\}$, $t \geq 0$, is a martingale.
    (ii) Use the optional sampling theorem to prove $\varphi(\lambda) = 2\left(e^{\sqrt{2\lambda}a} + e^{-\sqrt{2\lambda}a}\right)^{-1}$, where $\varphi(\lambda) = \mathbb{E}\exp(\lambda\tau_{-a,a})$, in the notation of the previous exercise.
11. Refer to the notation of Theorem 11.8.
    (i) Prove that $T_i - T_{i-1}$ $(i \geq 1)$ has a finite moment-generating function in a neighborhood of the origin if F has compact support.

(ii) Prove that $\mathbb{E}T_1^2 < \infty$ if $\int |z|^2 F(dz) < \infty$. [*Hint*: $\tau_{u,v} \le \tau_{-a,a}$ with $a := \max\{-u, v\} \le v - u$ and $\mathbb{E}\tau_{U,V}^2 \le c\theta \int (v-u)^2 F_+(dv) F_-(du)$ for some $c > 0$.]

12. In Theorem 11.7 suppose F is a symmetric distribution. Let X_i $(i \ge 1)$ be i.i.d. with common distribution F and independent of $\{B_t : t \ge 0\}$. Let $\widetilde{T}_1 := \inf\{t \ge 0 : B_t \in \{-X_1, X_1\}$, $\widetilde{T}_i := \widetilde{T}_{i-1} + \inf\{t \ge 0 : B_{\widetilde{T}_{i-1}+t} \in \{-X_i, X_i\}\}$ $(i \ge 1)$, $\widetilde{T}_0 = 0$.

  (i) Show that $B_{\widetilde{T}_i} - B_{\widetilde{T}_{i-1}}$ $(i \ge 1)$ are i.i.d. with common distribution F, and $\widetilde{T}_i - \widetilde{T}_{i-1}$ $(i \ge 1)$ are i.i.d.
  (ii) Prove that $\mathbb{E}\widetilde{T}_1 = \mathbb{E}X_1^2$, and $\mathbb{E}\widetilde{T}_1^2 = c\mathbb{E}X_1^4$, where c is a constant to be computed.
  (iii) Compute $\mathbb{E}e^{-\lambda \widetilde{T}_1}$ for $\lambda \ge 0$.

13. Prove that T_i $(i \ge 0)$ defined by (11.47) are $\{\mathcal{F}_t\}$–stopping times, where $\mathcal{F}_t$ is as defined there.
14. (i) Let Z_k, $k \ge 1$, be i.i.d. with finite variance. Prove that $n^{-\frac{1}{2}} \max\{|Z_k| : 1 \le k \le n\} \to 0$ in probability as $n \to \infty$. [*Hint*: $nP(Z_1 > \sqrt{n}\,\varepsilon) \le \frac{1}{\varepsilon^2}\,\mathbb{E}Z_1^2 \mathbf{1}[|z : 1 \ge \sqrt{n}\,\varepsilon], \forall \varepsilon > 0$].
  (ii) Derive (11.47). [*Hint*: $\varepsilon_n = \max_{1 \le k \le n} |\frac{T_k}{k} - 1| \cdot \frac{k}{n} \le \left\{\max_{1 \le k \le k_0} |\frac{T_k}{k} - 1|\right\} \cdot \frac{k_0}{n} + \max_{k \ge k_0} \left|\frac{T_k}{k} - 1\right| \forall k_0 = 1, 2, \ldots$.]