Discrete Applications of Brownian Motion

Dmitriy Kunisky, Spring 2013 Junior Paper

Advised by Michael Damron

May 7, 2013

1 Preliminaries

1.1 Notation

- Two random variables X, Y with the same law will be said to satisfy the equivalence relation $X \sim Y$.
- $\mathcal{N}(\mu, \sigma^2)$ will always denote a random variable that is normally distributed with mean μ and variance σ^2. Some useful elementary bounds on the standard normal distribution that will be used throughout are collected in Section 8.1.
- The minimum of two numbers x and y is denoted $x \wedge y$, and the maximum $x \vee y$.
- The weak convergence or convergence in distribution of a sequence of random variable will be denoted $X_n \Rightarrow X$. The *portmanteau theorem* gives many useful equivalent characterizations of this convergence [2, Theorem 3.25].

1.2 Overview of Stochastic Processes

Definition. A ($\mathbb{R}^d$-valued) *stochastic process* X is a collection of random variables $\{X_t\}_{t \in T}$ taking values in $\mathbb{R}^d$, defined on an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$. T is a totally ordered index set usually interpreted as time; we typically deal with continuous one-sided time, $T = [0, \infty)$. We will use one of two equivalent notations for the random variable the process takes at each time, either X_t or $X(t)$; the interpretation of each is given below.

In the most basic sense, X is then just a function $X : T \times \Omega \to \mathbb{R}^d$. The slice of this function for a fixed $t \in T$ is, as described above, a random variable denoted $X_t = X(t, \cdot) : \Omega \to \mathbb{R}^d$. Alternatively, the slice for a fixed $\omega \in \Omega$ is a *sample path*, a function $X(\cdot, \omega) : T \to \mathbb{R}^d$ representing a single possible path "traced out" by the process in $\mathbb{R}^d$ over all time. When we use the notation $X(t)$, we are suggesting that X be thought of as its sample path. With the probability measure $\mathbb{P}$ on Ω, the stochastic process then represents from this perspective a probability measure on the space of functions $f : T \to \mathbb{R}^d$. Thus, a stochastic process has two rather different interpretations:

1. X is a time-evolving family of $\mathbb{R}^d$-valued random variables X_t, or the induced time-evolving probability measure on $\mathbb{R}^d$. Here, X is similar to a diffusion in physics: often we impose a condition such as $X_0 = 0$ deterministically (so that the measure induced by X_0 is the Dirac measure concentrated at the origin), and as t increases, the measure "relaxes" into $\mathbb{R}^d$.

2. X is a random function $T \to \mathbb{R}^d$ (or even more time-agnostically, a random *curve* in $\mathbb{R}^d$ when time is continuous), or the induced probability measure on the space of such functions. This perspective is more amenable to questions about the global geometry of the sample path.

We will usually be more interested in the latter interpretation to study global properties (we can define amounts of time spent in domains, hitting and exit times, integrals against the sample path, etc). When we refer to the "distribution of X" or "law of X", we mean the probability measure on this function space. We define several notions of equivalence on processes.

Definition. Processes X, Y are *versions* of each other if $\mathbb{P}\{X_t = Y_t\} = 1$ for all t. They are *almost versions* if this holds for almost all t (in the Lebesgue sense).

Many of the interesting properties of a stochastic process are properties of its samples at finitely many time points, which we formalize in the following way.

Definition. The *finite-dimensional distributions* of a process X are the distributions of the finite vectors $(X_{t_1}, \ldots, X_{t_n})$ for any collection of $t_i \in T$.

By the following theorem, these distributions completely specify the distribution of a stochastic process (a proof of this technical fact can be found in [2, p. 25]).

Theorem 1.1. Two $\mathbb{R}^d$-valued stochastic processes X, Y have the same distribution if and only if they have the same finite-dimensional distributions.

This justifies the following finitary definition of independence of processes.

Definition. Processes X, Y are *independent* if any finite vectors $(X_{t_1}, \ldots, X_{t_n})$ and $(Y_{s_1}, \ldots, Y_{s_m})$ are independent for $t_i, s_j \in T$.

One property of a process that is not definable in this way is continuity, where we impose the requirement of continuity on the entire sample path at once.

Definition. A process X is *continuous* if its sample paths are almost surely (with respect to the measure on functions $T \to \mathbb{R}^d$ induced by the process) continuous.

We usually informally assume that an almost-sure property like this holds with certainty, just by restricting the underlying probability space to the almost sure set. The following is a useful basic theorem on continuity, a proof can be found in [2, p. 35].

Theorem 1.2 (Kolmogorov continuity criterion). Let X be a random process, and suppose that for any time $T > 0$, there are constants $\alpha, \beta, K > 0$ so that

$$\mathbb{E}[|X_t - X_s|^\alpha] \leq K|t - s|^{1+\beta}$$

for all $0 \leq s, t \leq T$. Then, there is a continuous version of X.

These properties and the theorems describing them will be useful for seeing why the axioms for Brownian motion specify a unique process (in terms of distribution).

A last relevant class of stochastic processes is defined below, the direct generalization to the continuous setting of the definition of Gaussian random vectors.

Definition. A process X is *Gaussian* if for any times $t_1, \ldots, t_n$ the vector $(X_{t_1}, \ldots, X_{t_n})$ is a Gaussian random vector (a property which has many explicit characterizations, the most useful for us will be that for any $c_i \in \mathbb{R}$, the random variable $c_1 X_{t_1} + \cdots + c_n X_{t_n}$ is Gaussian). A standard theorem [2, Lemma 11.1] gives that the distribution of a Gaussian process is completely determined by the means $\mathbb{E}[X_s]$ and covariances $\mathrm{Cov}[X_s, X_t]$.

1.3 Filtrations and Stopping Times

We also want an abstract means of defining “making decisions” on the basis of what has been seen of a stochastic process up to a certain time. For instance, in the classical example of the process representing a gambler's wealth while playing repeated rounds of a game, we want to delineate the information available to a gambler after N rounds of play, from which he decides whether to continue playing or to stop.

Definition. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a *filtration* of the space is a time-indexed family $\mathcal{F}(t)$ of σ-algebras on Ω so that $\mathcal{F}(s) \subseteq \mathcal{F}(t) \subseteq \mathcal{F}$ for $s < t$. A stochastic process $X(t)$ defined on a space with a filtration $\mathcal{F}(t)$ is called *adapted* to the filtration if $X(t)$ is $\mathcal{F}(t)$-measurable for each t.

Note that given a stochastic process $X(t)$, letting $\mathcal{F}(t)$ be the σ-algebra generated by the random variables $X(s)$ for $0 \le s \le t$ gives a filtration to which $X(t)$ is immediately adapted, which we denote $\mathcal{F}_X^0(t)$, omitting the subscript if the process is clear from context. This filtration represents the events whose occurrence or non-occurrence can be determined by observing the process up to time t. We can also define a slightly larger filtration,

$$\mathcal{F}^+(t) = \bigcap_{s>t} \mathcal{F}^0(t) = \bigcap_{s>t} \mathcal{F}^+(t),$$

allowing events that are determined by observing the process to any $t + \epsilon$ where $\epsilon > 0$ (the latter expression above is equivalent, and represents the important property of *right continuity* that the $\mathcal{F}^+$ filtration possesses but $\mathcal{F}^0$ does not). Some events belonging to $\mathcal{F}^+(t)$ but not $\mathcal{F}^0(t)$ are continuity, differentiability, and monotonicity of the sample path at time t, because all of these can be determined by looking at the sample path on an arbitrarily small open interval around t.

We additionally consider the notion of a time-valued random variable T whose distribution does not "run ahead" of $\mathcal{F}(t)$: whether or not $T \le t$ can be determined by the information in $\mathcal{F}(t)$.

Definition. A random variable T taking values in $[0, \infty]$ defined on a probability space filtered by $\mathcal{F}(t)$ is called a *stopping time* with respect to $\mathcal{F}(t)$ if $\{T \le t\} \in \mathcal{F}(t)$ for each $t \ge 0$.

In the example of the gambler playing a repeated game, a stopping time corresponds to a gambler's strategy: the only decision he makes at every round is whether to stop or continue, and he only has the information of the outcomes up to that round in order to make this decision. Thus, proving statements about arbitrary stopping times can usually be interpreted as making certain guarantees about the efficacy of any such strategy.

Example. It is easy to check some initial intuitive examples of stopping times, taking care to distinguish between which are adapted to $\mathcal{F}^0$ versus $\mathcal{F}^+$.

1. Any deterministic time $T = t_0$ is a stopping time with respect to any filtration.

2. Any stopping time with respect to $\mathcal{F}_X^0(t)$ is a stopping time with respect to $\mathcal{F}_X^+(t)$, since $\mathcal{F}_X^0(t) \subset \mathcal{F}_X^+(t)$.

3. If $X(t)$ is a continuous stochastic process and C is a closed set, then the *hitting time* of C, given by $H_C = \inf\{t \ge 0 : X(t) \in H\}$, is a stopping time with respect to $\mathcal{F}_X^0(t)$: $X(H_C) \in C$, because the path $B(t)$ is continuous and C is a closed set, so $X(H_C)$ is a limit point of points in C, and thus is in C. Thus, up to a set of measure zero, the event $\{H_C \le t\}$ is just the event $\{X(s) \in C \text{ for some } s \in [0, t]\} \in \mathcal{F}_X^0(t)$.

4. If U is an open set, then H_U is a stopping time with respect to $\mathcal{F}_X^+(t)$ but not $\mathcal{F}_X^0(t)$. The first claim follows by using the continuity of $X(t)$ and the right continuity of $\mathcal{F}^+$:

$$\mathcal{F}^+(t) = \bigcap_{\epsilon>0} \mathcal{F}^+(t+\epsilon).$$

   A counterexample for the second claim can be easily derived from the Brownian motion we will define, but is intuitively clear: $X(H_U)$ is a limit point of U but may not belong to U, and the sample path can with positive probability either enter U or "turn away from" U.

Since hitting times are one of our basic objects of study (as are open sets), we will typically use $\mathcal{F}_X^+(t)$ as the main filtration for stopping times, rather than $\mathcal{F}_X^0(t)$. Once we define the standard Brownian motion $B(t)$, all filtrations unless explicitly defined otherwise will be $\mathcal{F}_B^+(t)$, and a process will simply be called *adapted* if it is adapted to this filtration.

This allows us to use a slightly weaker definition of a stopping time.

Proposition 1.3. If T is a time-valued random variable with $\{T < t\} \in \mathcal{F}(t)$ for each $t \geq 0$, and $\mathcal{F}(t)$ is right-continuous, then T is a stopping time.

Proof. We have

$$\{T \leq t\} = \bigcap_{n=1}^{\infty} \left\{T < t + \frac{1}{n}\right\} \in \bigcap_{n=1}^{\infty} \mathcal{F}\left(t + \frac{1}{n}\right) = \mathcal{F}(t).$$

□

A last useful property of stopping times is that they satisfy a simple convergence theorem.

Proposition 1.4. If $T_n \nearrow T$ is an increasing sequence of stopping times with respect to a filtration, then T is a stopping time with respect to the same filtration.

Proof. We have

$$\{T \leq t\} = \bigcap_{n=1}^{\infty} \{T_n \leq t\}$$

where each term in the intersection belongs to the filtration. □

The main intuitive guideline we will follow about stopping times is that when we can show some property for a family of processes for deterministic times, the same property usually also holds (with more work) for stopping times in place of deterministic times (see the weak/strong Markov properties, and the optional stopping theorem).

2 Introduction

2.1 Motivation: A Functional Central Limit Theorem

In probability theory, a basic object of study is the discrete random walk defined by $S_n = \sum_{i=1}^n X_i$, where the X_i are i.i.d random variables with (after normalization) mean 0 and variance 1. The global or limiting behavior of such walks is of particular importance; the central limit theorem, for instance, is the statement that $\frac{1}{\sqrt{n}} S_n \Rightarrow \mathcal{N}(0,1)$, the remarkable aspect of the theorem being that the limiting distribution does not depend on the initial distribution of the X_i (we call this sort of behavior *universality* or *universal convergence*.

However, it is easy to empirically verify that such universal limiting behavior is exhibited for many different properties of S_n. For instance, some numerical experimentation readily yields all of the following observations.

1. The central limit theorem: $\frac{1}{\sqrt{n}}S_n$ universally converges in distribution to $\mathcal{N}(0,1)$.

2. The maximum $M_n = \max_{1\le i\le n} S_i$ converges universally when normalized to $\frac{M_n}{\sqrt{n}}$. By considering just the simple random walk and performing the necessary combinatorics [1, Section 3.10], it is straightforward to see that this limiting distribution is that of $|\mathcal{N}(0,1)|$.

3. The proportion of time spent on positive numbers, $P_n = \frac{1}{n}\#\{i : 1 \le i \le n, S_i > 0\}$, converges universally to an arcsine distribution (for the simple random walk, this is the classical first arcsine law).

To see more clearly what the connection between these limit theorems is, we must shift to continuous-time representations of S_n, so that we can treat them as random functions on a real interval. To that end, we first interpolate linearly between the discrete points $S(n)$ to construct a random path that is given by a continuous stochastic process:

$$S(t) = S_{\lfloor t\rfloor} + (t - \lfloor t\rfloor)(S_{\lfloor t\rfloor+1} - S_{\lfloor t\rfloor}).$$

Then, we scale larger and larger portions of the trajectory of $S(t)$ into the interval $[0,1]$, defining $S_n^*(t) = \frac{1}{\sqrt{n}}S(nt)$ on $t \in [0,1]$. So, we have S_n^* as a random function on $[0,1]$ is continuous, and hence belongs to $C([0,1])$ (which we make into a metric space by giving it the supremum or L^∞ norm). The central limit theorem can then be stated as $S_n^*(1) \Rightarrow \mathcal{N}(0,1)$, the maximum operator is then the continuous supremum operator $M_n = \sup_{t\in[0,1]} S_n^*(t)$, and the proportion of time spent above the x-axis is a Lebesgue measure, $m(\{t : t \in [0,1], S_n^*(t) > 0\})$.

In general, most of the interesting properties of the random walks that can be encoded into a single real number can be realized as *functionals*, or mappings $\Lambda : C([0,1]) \to \mathbb{R}$, applied to S_n^*. Note that those mentioned above, and most functionals that arise from combinatorial considerations on the underlying discrete random walks, are also continuous with respect to the supremum norm metric given to $C([0,1])$ above.

Now, suppose that the S_n^* had a limit in distribution; that is, that there existed some stochastic process B so that, in terms of their distributions on $C([0,1])$, we have $S_n^* \Rightarrow B$. This would imply by the portmanteau theorem that for any bounded continuous functional $\Lambda : C([0,1]) \to \mathbb{R}$, we have

$$\mathbb{E}[\Lambda(S_n^*)] \to \mathbb{E}[\Lambda(B)].$$

Now, since for any bounded continuous $h : \mathbb{R} \to \mathbb{R}$ we have that $h \circ \Lambda$ is also continuous whenever Λ is just continuous, we have

$$\mathbb{E}[h(\Lambda(S_n^*))] \to \mathbb{E}[h(\Lambda(B))],$$

and thus $\Lambda(S_n^*) \Rightarrow \Lambda(B)$ as $\mathbb{R}$-valued random variables for any continuous functional Λ.

If the limiting process B were independent of the distributions of the original X_i, then we would have a unified explanation of the many limit theorems applying to random walks: any continuous functional of the (linear interpolation of) the discrete random walk would converge universally in distribution to the same functional of B. We will later prove that this is indeed the case in Donsker's invariance principle (also called the functional central limit theorem), but for now we can first easily isolate some of the basic properties of this hypothetical limit process.

Proposition 2.1. If $S_n^* \Rightarrow B$, then the process $\{B_t\}$ satisfies the following:

1. $B_0 = 0$.
2. The differences $B_t - B_s$, called the *increments* of B, satisfy that for any $t_0 < \cdots < t_n$, the n random variables $B_{t_k} - B_{t_{k-1}}$ with $1 \le k \le n$ are jointly independent.
3. The increments are normally distributed: $B_t - B_s \sim \mathcal{N}(0, t - s)$ when $t \ge s$.
4. B is a continuous process.

Proof. Property (1) is immediate, since $S_0 = 0$. For (2) and (3), note that $B(t) - B(s)$ is converged on in distribution by

$$\frac{1}{\sqrt{n}}\Big(S_{\lfloor nt\rfloor} + (nt - \lfloor nt\rfloor)(S_{\lfloor nt\rfloor+1} - S_{\lfloor nt\rfloor}) - S_{\lfloor ns\rfloor} - (ns - \lfloor ns\rfloor)(S_{\lfloor ns\rfloor+1} - S_{\lfloor ns\rfloor})\Big),$$

And since the inner terms other than $S_{\lfloor nt\rfloor} - S_{\lfloor ns\rfloor}$ are bounded in magnitude, their contribution tends to zero (because of the $\frac{1}{\sqrt{n}}$ factor), so $B(t) - B(s)$ is converged on in distribution by

$$\frac{1}{\sqrt{n}}(S_{\lfloor nt\rfloor} - S_{\lfloor ns\rfloor}) = \frac{1}{\sqrt{n}}\sum_{ns<i\le nt} X_i.$$

By the central limit theorem, this converges in law to $\mathcal{N}(0, t - s)$, giving (3). For (2), we note that X_i in the sums as above for disjoint intervals (s, t) are disjoint, so since the X_i are independent, the increments are jointly independent. Property (4) follows from the Kolmogorov continuity criterion, taking $\alpha = 4$, $\beta = 1$, and $K = 2$. □

In particular, we have $B_t \sim \mathcal{N}(0, t)$ for all t, by taking the increment starting at $B_0 = 0$. We can refine this description slightly by characterizing Brownian motion as a Gaussian process.

Proposition 2.2. B is a Gaussian process, with means and covariances given by

$$\mathbb{E}[B_{t_i}] = 0 \text{ and } \mathrm{Cov}[B_{t_i}, B_{t_j}] = \min(t_i, t_j).$$

Proof. By the characterization given earlier, to show that B is Gaussian it suffices to show that any linear combination of the components $B_{t_1}, \ldots, B_{t_n}$ is Gaussian in distribution. To do this, we take advantage of the independent increments: let $a_i \in \mathbb{R}$, and define $t_0 = 0$ so that $B_{t_0} = 0$, then by a simple summation-by-parts type manipulation we find:

$$\sum_{k=1}^{n} a_k B_{t_k} = \sum_{k=1}^{n}\left(\sum_{i=k}^{n} a_n\right)(B_{t_k} - B_{t_{k-1}}),$$

which by independent increments is a sum of independent Gaussians, and hence is itself Gaussian.

So, it remains to calculate the covariances. Let $s \le t$, then $\mathbb{E}[B_s] = \mathbb{E}[B_t] = 0$, so

$$\begin{aligned}
\mathrm{Cov}[B_s, B_t] &= \mathbb{E}[B_s B_t] \\
&= \mathbb{E}[B_s \cdot (B_t - B_s + B_s)] \\
&= \mathbb{E}[B_s \cdot (B_t - B_s)] + \mathbb{E}[B_s^2] \\
&= \mathbb{E}[B_s] \cdot \mathbb{E}[B_t - B_s] + \mathrm{Var}[B_s] \\
&= \mathrm{Var}(B_s) \\
&= s,
\end{aligned}$$

where we have used that B_s and $B_t - B_s$ are independent increments. □

Note that this statement is thus equivalent to the first three properties in the previous proposition. We will use either of these equivalent characterizations to define a Brownian motion.

Definition. Any real-valued stochastic process $\{B_t\}$ that is a continuous and Gaussian with means of $\mathbb{E}[B_t] = 0$ and covariances of $\mathrm{Cov}[B_s, B_t] = \min(s, t)$ is called a *(one-dimensional) standard Brownian motion*, or a *Brownian motion started at x* if the means are instead identically some other $x \in \mathbb{R}$.

Of course, this leaves open the two key issues of existence and uniqueness: we have yet to exhibit an actual Brownian motion, nor have we shown that the distribution the Brownian motion induces (the measure on the space $C([0, 1])$) is unique.

Uniqueness is straightforward: the Gaussian process characterization of Brownian motion specifies the finite-dimensional distributions of B, so by Theorem 1.1, it follows that all standard Brownian motions have the same distribution.

Existence is more difficult: it is possible to take a mostly abstract approach, first proving the existence of a suitable Gaussian process by the Kolmogorov consistency theorem, and then applying the Kolmogorov continuity criterion to show that this has a continuous version; this is the approach taken in [2, Theorem 11.5]. It is also possible to construct this process directly, essentially constructing the scaling limit of some particular discrete random walk as described above; this is presented in [3, Theorem 1.3]. We will for our purposes assume the existence of a standard Brownian motion.

2.2 Brownian Motion in $\mathbb{R}^d$

We define the Brownian motion on $\mathbb{R}^d$ by $B(t) = (B_1(t), \dots, B_d(t))$, where the coordinates $B_i(t)$ are independent one-dimensional standard Brownian motions. It is an easy exercise to verify that the higher-dimensional analog of the independence and distribution of increments (they will now be Gaussian vectors instead of just one-dimensional Gaussian distributions) conditions are equivalent to this characterization.

3 Basic Symmetries and Bounds

The simplest symmetry of Brownian motion is analogous to the scaling symmetry of the normal distribution, $\frac{1}{a}\mathcal{N}(0, a^2) \sim \mathcal{N}(0, 1)$.

Proposition 3.1 (Scaling Invariance). Suppose $B(t)$ is a standard Brownian motion and let $a > 0$. Then, the process $X(t) = \frac{1}{a}B(a^2t)$ is also a standard Brownian motion.

Proof. All properties but the distribution of the increments are not affected by the transformation. So, it suffices to check that $X(t) - X(s) = \frac{1}{a}(B(a^2t) - B(a^2s))$ is normally distributed with mean 0 and variance $t - s$, which is immediate by linearity of expectation and because $B(a^2t) - B(a^2s)$ has variance $a^2t - a^2s$. □

This can be seen as a statement of the self-similarity of Brownian motion: the distribution of $B(t)$ on any time interval $t \in [0, x)$ for $x > 0$ is sufficent to determine the distribution over arbitrarily large spans of time.

We can in passing flesh out the analogy between standard Brownian motion and the standard normal distribution: if we define the transformation $\mathcal{F} : S(t) \mapsto \frac{1}{2}S(4t)$ on some vector space of stochastic processes containing $B(t)$, we have that $B(t)$ is a fixed point of $\mathcal{F}$. The convergence under scaling of discrete random walks to $B(t)$ under Donsker's invariance principle can then be interpreted as a convergence result for iteration of this map. This is the analog for processes of the transformation $X \mapsto \frac{1}{2}(X^{(1)} + X^{(2)} + X^{(3)} + X^{(4)})$ (where $X^{(i)}$ are independent copies of X) on the Hilbert space of L^2 random variables converging to $\mathcal{N}(0, 1)$ by the central limit theorem.

Another basic invariance property is unique to the time-dependent setting of processes, and is correspondingly more powerful, giving that the behavior of $B(t)$ on $[0, x)$ actually is enough to determine the behavior of $B(t)$ on $(\frac{1}{x}, \infty)$, and in particular the asymptotic behavior at infinity.

Proposition 3.2 (Time Inversion). Suppose $B(t)$ is a standard Brownian motion, then the process

$$X(t) = \begin{cases} 0 & t = 0, \\ tB(t^{-1}) & t > 0 \end{cases}$$

is also a standard Brownian motion.

Proof. It is immediate that $X(t)$ is a Gaussian process, and that for a finite vector $(X(t_1), \ldots, X(t_n))$, the coordinate means are zero. The covariances are

$$\mathrm{Cov}[X(t+h), X(t)] = t(t+h)\mathrm{Cov}\left[B\left(\frac{1}{t+h}\right), B\left(\frac{1}{t}\right)\right] = t(t+h)\frac{1}{t+h} = t,$$

so $X(t)$ is a Gaussian process with the same statistics as $B(t)$.

It remains to show that $X(t)$ is continuous. Continuity of $X(t)$ is immediate when $t > 0$ by continuity of $B(t)$, so it suffices to consider $t = 0$. Continuity at $t = 0$ is equivalent to the limit $\lim_{t \to 0} X(t) = 0$. Equivalently, we want $\lim_{t\to 0} tB(\frac{1}{t}) = \lim_{t\to\infty} \frac{1}{t}B(t) = 0$.

Fix $\epsilon > 0$, and consider the $n \in \mathbb{Z}_+$ for which $|\frac{1}{n}B(n)| > \epsilon$, or $|\frac{1}{\sqrt{n}}B(n)| > \epsilon\sqrt{n}$. Let $A_{n,\epsilon}$ denote this event. Since $\frac{1}{\sqrt{n}}B(n) \sim \mathcal{N}(0, 1)$, we have

$$\mathbb{P}(A_{n,\epsilon}) = 2\mathbb{P}(\mathcal{N} > \epsilon\sqrt{n}) < 2e^{-\epsilon^2 n/2},$$

and in particular for fixed ϵ the $\mathbb{P}(A_{n,\epsilon})$ over all $n \in \mathbb{Z}_+$ are summable, so by the Borel-Cantelli lemma only finitely many $A_{n,\epsilon}$ occur. This means that for all but finitely many $n \in \mathbb{Z}_+$, we have

$|\frac{1}{n}B(n)| \le \epsilon$. Also, for $q \in \mathbb{Q} \cap (n, n+1)$, we have

$$\mathbb{P}\left(\left|\frac{1}{n}(B(q) - B(n))\right| > \epsilon\right) \le \mathbb{P}\left(\left|\frac{1}{\sqrt{q-n}}(B(q) - B(n))\right| > \epsilon\sqrt{n}\right),$$

and since $B(q) - B(n) \sim \mathcal{N}(0, q-n)$ we then have that the same argument from above applies, so that for all but finitely many q we have $|\frac{1}{n}(B(q) - B(n))| \le \epsilon$, and so $|\frac{1}{n}B(q)| \le 2\epsilon$, using the above result for sufficiently large n. Since the q for which this holds are dense in $(n, n+1)$, by continuity we obtain that for all $t \in (n, n+1)$, $|\frac{1}{t}B(t)| \le 2\epsilon$, for sufficiently large n, which gives the result. □

Again, the main use of this transformation is that it connects the behavior of $B(t)$ near zero to its behavior at infinity. For instance, we can obtain some basic growth bounds.

Corollary 3.3 (Simple Asymptotics). Almost surely,

$$\begin{aligned}
\lim_{t\to\infty} \frac{B(t)}{t} &= 0, \\
\limsup_{t\to\infty} \frac{B(t)}{\sqrt{t}} = \limsup_{t\to 0} \frac{B(t)}{\sqrt{t}} &= +\infty, \\
\liminf_{t\to\infty} \frac{B(t)}{\sqrt{t}} = \liminf_{t\to 0} \frac{B(t)}{\sqrt{t}} &= -\infty.
\end{aligned}$$

Proof. Let $X(t)$ be the time inversion of $B(t)$. Then, we have $\lim_{t\to\infty} \frac{B(t)}{t} = \lim_{t\to\infty} X(\frac{1}{t})$, which by continuity is almost surely equal to $X(0) = 0$, giving the first claim.

For the second claim, fix $c > 0$ and define the event

$$E_c = \{B(n) > c\sqrt{n} \text{ for infinitely many } n \in \mathbb{Z}_+\},$$

then by Fatou's lemma and the scaling property,

$$\mathbb{P}E_c \ge \limsup_{n\to\infty} \mathbb{P}\{B(n) > c\sqrt{n}\} = \mathbb{P}\{B(1) > c\},$$

which is positive (since $B(1) = B(1) - B(0)$ is normally distributed). Also, E_c is a tail event (since $B(n) = \sum_{k=1}^{n} (B(k) - B(k-1))$, where the summands are independent), and thus has probability 0 or 1 by Kolmogorov's 0-1 law, so $\mathbb{P}E_c = 1$, and taking the intersection of E_c over all $c \in \mathbb{Z}_+$ shows that $\frac{B(n)}{\sqrt{n}}$ is unbounded from above almost surely. Time inversion then immediately gives the second claim: $\frac{B(t)}{\sqrt{t}} = \sqrt{t}X(\frac{1}{t})$, which as $t \to \infty$ is the same as $\frac{X(t)}{\sqrt{t}}$ as $t \to 0$. A symmetric argument gives the two claims for the infimum. □

As we see in the proofs of the second two claims above, time inversion is better adapted to working with processes of the form $\tilde{B}(t) = \frac{B(t)}{\sqrt{t}}$, which we will identify in its own class.

Definition. A real-valued stochastic process $X(t)$ is called a *normalized standard Brownian motion* if $\sqrt{t}X(t)$ is a standard Brownian motion. We denote some fixed instance of such a process by $\tilde{B}(t)$ as above, corresponding to our fixed instance $B(t)$ of a standard Brownian motion.

The statement of time inversion for these processes is simply that if $\tilde{B}(t)$ is a normalized standard Brownian motion, then so is $\tilde{B}(1/t)$.

We use the term normalized because the variance of $\tilde{B}(t)$ is identically 1 regardless of t; thus, unlike $B(t)$, $\tilde{B}(t)$ has the same distribution at each time point. This is a very useful feature, enabling us to use bounds on just the standard normal distribution to get relatively strong results for Brownian motion. An additional interesting observation, which we will partly clarify with the law of the iterated logarithm, is that though for any fixed t we have $\tilde{B}(t) \sim \mathcal{N}(0,1)$, the above shows that $\tilde{B}(t)$ is still almost surely unbounded. Lastly, while when $s < t$ we had $\mathrm{Cov}[B(s), B(t)] = s$, here we have

$$\mathrm{Cov}[\tilde{B}(s), \tilde{B}(t)] = \frac{1}{\sqrt{st}}\mathrm{Cov}[B(s), B(t)] = \frac{s}{\sqrt{st}} = \sqrt{\frac{s}{t}},$$

which tends to zero as $\frac{s}{t} \to 0$, which is a weak statement of eventual almost-independence of points sampled along the path of $\tilde{B}(t)$. This heuristic idea will be used in the proof of the law of the iterated logarithm for Brownian motion, where this near-independence will be used in the converse of the Borel-Cantelli lemma.

4 The Markov Property

As in the finite setting with Markov chains, the Markov property for stochastic processes formalizes the idea of a process having "no memory", in that the process starts over at any fixed time.

Theorem 4.1 (Weak Markov Property). Suppose that $B(t)$ is a Brownian motion in $\mathbb{R}^d$. Then for any $s > 0$, the process $X(t) = B(t+s) - B(s)$ is a Brownian motion started at the origin and finitely independent of the process $\{B(t) : t \in [0, s]\}$.

Proof. That $B(t+s)-B(s)$ satisfies the properties of a Brownian motion is immediate (the increments are of the form $(B(t+a+s) - B(s)) - (B(t+s) - B(s)) = B(t+s+a) - B(t+s)$ and thus are also increments of the original Brownian motion). The independence statement follows from the independence of increments of $B(t)$. □

This combined with time inversion allows us to show relatively easily that Brownian motion is almost surely not differentiable at any fixed point.

Corollary 4.2. For any fixed $t \geq 0$, the sample path of B is almost surely not differentiable at t. In particular, the upper and lower derivatives diverge as follows:

$$D^*B(t) = \limsup_{h \to 0} \frac{B(t+h) - B(t)}{h} = +\infty,$$
$$D_*B(t) = \liminf_{h \to 0} \frac{B(t+h) - B(t)}{h} = -\infty.$$

Proof. First, suppose $t = 0$. Let X be the time inversion of B. We then have

$$D^*B(0) \geq \limsup_{n \to \infty} \frac{B(\frac{1}{n}) - B(0)}{\frac{1}{n}} \geq \limsup_{n \to \infty} \sqrt{n} B(\frac{1}{n}) = \limsup_{n \to \infty} \tilde{X}(n),$$

which we have seen is infinite. By the same argument, $D_*B(0) = -\infty$.

Now, suppose $t > 0$, then we just apply the above to the walk defined by $B'(s) = B(t+s) - B(t)$, which is a standard Brownian motion by the above. □

Note that this is *not* the statement that the sample path is almost surely nowhere differentiable, we will prove this later as a consequence of the asymptotic growth bound in the law of the iterated logarithm. This does however imply that on a countable dense subset of times, such as $\mathbb{Q}_+$, the Brownian motion is again almost surely not differentiable at any of the points in that subset.

Introducing filtrations (let $\mathcal{F}^0$ and $\mathcal{F}^+$ be as defined in the preliminaries, for the process $B(t)$), we can obtain more general results.

Theorem 4.3 (Intermediate Markov Property). The process $B(s+t)-B(s)$ is independent of $\mathcal{F}^+(s)$.

Proof. By continuity, if $s_n \searrow s$, then $B(t+s)-B(s) = \lim_{n\to\infty}(B(s_n+t)-B(s_n))$, and each $B(s_n+t)-B(s_n)$ is independent of $\mathcal{F}^+(s)$, hence $B(t+s)-B(s)$ is independent of $\mathcal{F}^+(s)$, and the statement for finite vectors of increments follows similarly. □

A useful corollary is that if we define the *germ σ-algebra* $\mathcal{F}^+(0)$, then we obtain a 0-1 law for elements of this σ-algebra.

Corollary 4.4 (Blumenthal 0-1 Law). If $A \in \mathcal{F}^+(0)$, then $\mathbb{P}(A) \in \{0,1\}$.

Proof. By the above proposition, A is independent of itself. □

Defining the *tail σ-algebra* to be $\mathcal{T} = \mathcal{F}^\infty = \bigcap_{t\geq 0}\sigma(B(s) : s \geq t)$, then by time inversion on the above corollary, we obtain another such law.

Corollary 4.5 (Tail 0-1 Law). If $A \in \mathcal{F}^\infty$, then $\mathbb{P}(A) \in \{0,1\}$.

Proof. Time inversion maps the tail σ-algebra onto the germ σ-algebra, and the result then follows by the Blumenthal 0-1 law. □

This gives an easy proof that Brownian motion changes sign infinitely often both near the origin and near infinity.

Corollary 4.6. Define the random variables

$$\begin{aligned}
\tau_+ &= \inf_{t>0}\{B(t) > 0\},\\
\tau_0 &= \inf_{t>0}\{B(t) = 0\},\\
\tau_- &= \inf_{t>0}\{B(t) < 0\},\\
\sigma_+ &= \sup_{t>0}\{B(t) > 0\},\\
\sigma_0 &= \sup_{t>0}\{B(t) = 0\},\\
\sigma_- &= \sup_{t>0}\{B(t) < 0\},
\end{aligned}$$

then $\mathbb{P}\{\tau_\bullet = 0\} = \mathbb{P}\{\sigma_\bullet = \infty\} = 1$.

Proof. By time inversion, it suffices to prove the claim for the first three random variables. By symmetry, the claim for τ_- will follow from that for τ_+, and from both of these by the intermediate value theorem the claim will follow for τ_0, so it suffices to prove the claim for τ_+. We have

$$\{\tau_+ = 0\} = \bigcap_{n=1}^{\infty} \left\{ B(\epsilon) > 0 \text{ for some } \epsilon \in (0, 1/n) \right\} \in \mathcal{F}^+(0),$$

so it has probability either 0 or 1. And, we have $\mathbb{P}\{\tau_+ \le t\} \ge \mathbb{P}\{B(t) > 0\} = \frac{1}{2} > 0$, so the probability $\mathbb{P}\{\tau_+ = 0\}$ is positive and thus 1. □

Now, we apply our principle of pushing results for deterministic times through to all stopping times by the following extension of these results.

Theorem 4.7 (Strong Markov Property). For every almost surely finite stopping time T, the process $B(T+t) - B(T)$ is a standard Brownian motion independent of $\mathcal{F}^+(T)$.

Proof of Strong Markov Property. Let us define a sequence of decreasing discrete approximations $T_n \searrow T$ by $T_n = (m+1)2^{-n}$ whenever $T \in [m2^{-n}, (m+1)2^{-n})$, where we are just "rounding T up" to lie in a discrete dyadic set. We will first show that the statement holds for the T_n. After this, the general result follows easily since we have

$$B(s+t+T) - B(t+T) = \lim_{n\to\infty} B(s+t+T_n) - B(t+T_n),$$

which gives the distribution and independence of the increments, as well as independence from $\mathcal{F}^+(T)$, since $\mathcal{F}^+(T) \subset \mathcal{F}^+(T_n)$ for each n.

To prove the statement for the stopping time T_n, define two more Brownian motions:

$$\begin{aligned} B^k(t) &= B(t + \frac{k}{2^n}) - B(\frac{k}{2^n}), \\ B^*(t) &= B(t + T_n) - B(T_n). \end{aligned}$$

Note that $B^* = B^k$ if and only if $T_n = \frac{k}{2^n}$, and so the B^k are in this sense piecewise components of B^*. Now, suppose we have some event $E \in \mathcal{F}^+(T_n)$. Then, for any event $\{B^* \in A\}$, we have

$$\begin{aligned} \mathbb{P}(\{B^* \in A\} \cap E) &= \sum_{k=0}^{\infty} \mathbb{P}\left(\{B^k \in A\} \cap E \cap \{T_n = k2^{-n}\}\right) \\ &= \sum_{k=0}^{\infty} \mathbb{P}\left(\{B^k \in A\}\right) \cdot \mathbb{P}\left(E \cap \{T_n = k2^{-n}\}\right), \end{aligned}$$

since by the intermediate Markov property, $\{B^k \in A\}$ is independent of $E \cap \{T_n = k2^{-n}\} \in \mathcal{F}^+(k2^{-n})$. By the weak Markov property, $\mathbb{P}(\{B^k \in A\})$ does not depend on k since all B^k are standard Brownian motions, so we have $\mathbb{P}\{B^k \in A\} = \mathbb{P}\{B \in A\}$ and thus continuing the above,

$$\mathbb{P}(\{B^* \in A\} \cap E) = \mathbb{P}\{B \in A\} \sum_{k=0}^{\infty} \mathbb{P}(E \cap \{T_n = k2^{-n}\}) = \mathbb{P}\{B \in A\}\mathbb{P}(E),$$

hence B^* is a standard Brownian motion and is independent of E, for all $E \in \mathcal{F}^+(T_n)$, which shows the claim for T_n and completes the proof. □

One important application is the reflection principle, which allows us not only to restart but also to reflect a Brownian motion at any stopping time.

Theorem 4.8 (Reflection Principle). If T is an almost surely finite stopping time and $B(t)$ a standard brownian motion, then the process

$$B^*(t) = \begin{cases} B(t) & t \leq T, \\ 2B(T) - B(t) & t > T \end{cases}$$

is also a standard Brownian motion.

Proof. By the strong Markov property, both $B(T+t) - B(T)$ and $-(B(T+t) - B(t))$ are standard Brownian motions. Concatenating either of them to the standard Brownian motion $B(t)$ on the interval $[0, T]$ then gives two stochastic processes with the same distributions. But, concatenating the first gives $B(t)$, and concatenating the second gives $B^*(t)$, thus $B^*(t)$ is a standard Brownian motion. □

This gives a direct attack on the question of the distribution of the maximum of Brownian motion (one of our motivating limit functionals from the introduction), the process

$$M(t) = \max_{s \in [0,t]} B(t).$$

The elegant description is proven easily by the reflection principle.

Corollary 4.9. For each fixed t, $M(t)$ has the same law as $|B(t)|$.

Proof. Let $a > 0$, then we want to show that $\mathbb{P}\{M(t) > a\} = \mathbb{P}\{|B(t)| > a\}$. Let $T = H_a$ be the hitting time of a, and let $B^*(t)$ be Brownian motion reflected at T. Then,

$$\{M(t) > a\} = \{B(t) > a\} \cup \{M(t) > a \text{ and } B(t) \leq a\}.$$

This union is disjoint, and the second event $\{B^*(t) \geq a\}$. So,

$$\mathbb{P}\{M(t) > a\} = 2\mathbb{P}\{B(t) > a\} = \mathbb{P}\{|B(t)| > a\}.$$ □

5 The Martingale Property

Another general property of stochastic processes that Brownian motion enjoys is the martingale property, which instead of describes not the time invariance of a process but a certain stationarity of its conditional distributions at points in time.

Definition. A process $X(t)$ adapted to a filtration $\mathcal{F}(t)$ is called a *martingale* with respect to the filtration if $\mathbb{E}[|X(t)|] < \infty$ for each $t \geq 0$, and for any $0 \leq s \leq t$ we have $\mathbb{E}[X(t) \mid \mathcal{F}(s)] = X(s)$ almost surely. It is called a *local martingale* on $t \in [0, T]$ if there exist stopping times T_n increasing almost surely to T so that $X(t \wedge T_n)$ is a martingale for each n.

By the Markov property, standard Brownian motion is a martingale with respect to $\mathcal{F}^+(t)$:

$$\begin{aligned}\mathbb{E}[B(t) \mid \mathcal{F}^+(s)] &= \mathbb{E}[B(t) - B(s) \mid \mathcal{F}^+(s)] + \mathbb{E}[B(s) \mid \mathcal{F}^+(s)] \\ &= \mathbb{E}[B(t) - B(s)] + B(s) \\ &= B(s).\end{aligned}$$

On the other hand, normalized standard Brownian motion, though it is stationary in distribution, is not a martingale:

$$\mathbb{E}\left[\frac{1}{\sqrt{t}}B(t)\,\middle|\,\mathcal{F}^+(s)\right] = \frac{1}{\sqrt{t}}\mathbb{E}\left[B(t)\,\middle|\,\mathcal{F}^+(s)\right] = \frac{1}{\sqrt{t}}B(s),$$

so on the average we predict the process to drift towards the origin as time goes on from any given point; a martingale is not allowed to have any such bias. We will use a standard fact about martingales extensively, which applies our principle of extending deterministic time results to stopping times to the defining martingale property. A proof can be found in [2, Theorem 6.12].

Theorem 5.1 (Optional Stopping). Let $X(t)$ be a continuous martingale, and $0 \le S \le T$ be stopping times. If there is a random variable Z so that the process $X(t \wedge T)$ is dominated by Z almost surely at every time, then almost surely we have

$$\mathbb{E}[X(T) \mid \mathcal{F}(S)] = X(S).$$

The optional stopping theorem applied to Brownian motion gives Wald's lemmas, which identify the first two moments of $B(T)$ for a stopping time T. Applying the optional stopping theorem to Brownian motion with $S = 0$ gives us the first moment of $B(T)$ for an integrable stopping time T directly.

Corollary 5.2 (Wald's First Lemma). Let T be a stopping time so that $B(t \wedge T)$ is dominated by an integrable random variable, then $\mathbb{E}[B(T)] = 0$.

A technical result gives that in fact the condition here holds whenever $\mathbb{E}[T] < \infty$, a more applicable version of the result. One application is an orthogonality (or weak independence) statement for $B(S)$ and $B(T) - B(S)$.

Corollary 5.3. Let $S \le T$ be stopping times with $\mathbb{E}[T] < \infty$. Then,

$$\mathbb{E}[B(T)^2] = \mathbb{E}[B(S)^2] + \mathbb{E}[(B(T) - B(S))^2].$$

Proof. Note that in the L^2 space of random variables with the expectation inner product, this is the statement that $B(S)$ and $B(T) - B(S)$ are orthogonal. By the tower property of conditional expectation, we have

$$\mathbb{E}[B(T)]^2 = \mathbb{E}[B(S)^2] + 2\mathbb{E}\Big[B(S)\mathbb{E}(B(T) - B(S) \mid \mathcal{F}(S))\Big] + \mathbb{E}[(B(T) - B(S))^2].$$

By linearity of expectation and the martingale property, the inner expectation here is zero, and the result follows. □

To identify the second moment of $B(T)$, we identify another martingale besides $B(t)$ itself.

Lemma 5.4. $B(t)^2 - t$ is a martingale.

Proof. That the process is adapted to the filtration corresponding to $B(t)$ is immediate, and we have

$$\begin{aligned}\mathbb{E}[B(t)^2 - t \mid \mathcal{F}^+(s)] &= \mathbb{E}[(B(t) - B(s))^2 \mid \mathcal{F}^+(s)] + 2\mathbb{E}[B(t)B(s) \mid \mathcal{F}^+(s)] - B(s)^2 - t \\ &= (t - s) + 2B(s)^2 - B(s)^2 - t \\ &= B(s)^2 - s.\end{aligned}$$

□

Next, we obtain the second moment of $B(T)$.

Theorem 5.5 (Wald's Second Lemma). Let T be a stopping time with $\mathbb{E}[T] < \infty$. Then, we have $\mathbb{E}[B(T)^2] = \mathbb{E}[T]$.

Proof. Define the stopping times $T_n = \inf\{t \geq 0 : |B(t)| = n\}$, the hitting times of positive integer n by $|B(t)|$. Now, we have $B(t \wedge T \wedge T_n)^2 - t \wedge T \wedge T_n$ dominated by $n^2 + T$, an integrable random variable, and by the optional stopping theorem $\mathbb{E}[B(T \wedge T_n)^2] = \mathbb{E}[T \wedge T_n]$. Using monotone converge and Fatou's lemma to bound $\mathbb{E}[B(T)^2]$ by the limits of the $\mathbb{E}[B(T \wedge T_n)^2]$ from below and above respectively, the result follows. □

For us, these results are only of instrumental use in preparing for the Skorokhod embedding in the next section. However, in general the relationship between martingales and the universality of Brownian motion is far deeper: firstly, by a theorem of Lévy, $B(t)$ is unique in being a martingale and having $B(t)^2 - t$ be a martingale as well; secondly, by a theorem of Dambis, Dubins, and Schwarz, *any* continuous (local) martingale is actually Brownian motion, after a time change.

6 Discrete Random Walks: Skorokhod Embedding

We are now equipped to see the application of Brownian motion to discrete random walks and vice-versa, as outlined in the introduction. The next two sections will yield two tools for "embedding" discrete random walks in Brownian motion in different ways, and will give applications of these tools to various limit theorems regarding discrete random walks, whose proof is eased by going through continuous Brownian motion.

Our first step is to produce a converse to Wald's lemmas, which given a random variable X produces a stopping time T so that $B(T)$ has the probability law of T. By Wald's lemmas, for an integrable stopping time T we also have $\mathbb{E}[B(T)] = 0$ and $\mathbb{E}[B(T)^2] = \mathbb{E}[T] < \infty$, so there is only hope of producing T so that $B(T) \sim X$ if $\mathbb{E}[X] = 0$ and $\mathbb{E}[X^2] < \infty$. Our theorem will give us that these are the only obstructions to the construction of such a T.

Theorem 6.1 (Skorokhod Embedding Theorem). Suppose X is a real-valued random variable with $\mathbb{E}[X] = 0$ and $\mathbb{E}[X^2] < \infty$, then there is a stopping time T so that $B(T) \sim X$ and $\mathbb{E}[T] = \mathbb{E}[X^2]$.

We first prove the theorem for the simple case of X taking on only two values.

Lemma 6.2. If X takes on only the values $a < 0$ and $b > 0$ with $\mathbb{E}[X] = 0$, then there a Skorokhod embedding T, where $\mathbb{E}[T] = -ab$.

Proof. Since $\mathbb{E}[X] = 0$, we must have $\mathbb{P}\{X = a\} = \frac{b}{b-a}$ and $\mathbb{P}\{X = b\} = \frac{-a}{b-a}$. Then, define

$$T = \inf\{t > 0 : B(t) \notin (u, v)\}$$

to be the exit time of (u, v). Clearly this is a stopping time, and by Wald's first lemma we have $\mathbb{E}[B(T)] = 0$. This means that

$$0 = a\mathbb{P}\{B(T) = a\} + b\mathbb{P}\{B(T) = b\},$$

which in combination with $\mathbb{P}\{B(T) = a\} + \mathbb{P}\{B(T) = b\} = 1$ gives after a simple algebraic manipulation that $B(T)$ has the same distribution as X.

To use Wald's second lemma, we must check $\mathbb{E}[T] < \infty$:

$$\mathbb{E}[T] = \int_0^\infty \mathbb{P}\{T > t\}dt = \int_0^\infty \mathbb{P}\{B(s) \in (a, b) \text{ for all } s \in [0, t]\}dt,$$

and we have letting $m = \max_{x\in(a,b)} \mathbb{P}_x\{B(1) \in (a, b)\}$, some constant, that the integrand is at most m^k when $t \geq k$. Thus, the integral converges. So, by Wald's second lemma and the probabilities for $B(T)$, we have

$$\mathbb{E}[T] = \mathbb{E}[B(T)^2] = \frac{a^2b}{b-a} + \frac{-b^2a}{b-a} = -ab.$$

□

So, we have that we can embed any probability measure with support at two points. Now, we will express a general probability measure with the correct mean and variance properties as an average of a continuum of such probability measures. This will allow us to generalize from the above result immediately to the embedding full result.

Lemma 6.3. Let μ be a probability measure on $\mathbb{R}$, with $\int x d\mu = 0$ and $\sigma^2 = \int x^2 d\mu < \infty$. Then, there is a probability measure θ on $(-\infty, 0) \times [0, \infty)$, such that

$$\mu(A) = \int \left(\frac{y}{y-x}\chi_A(x) + \frac{-x}{y-x}\chi_A(y) \right) d\theta,$$

and also $\sigma^2 = \int -xy d\theta$.

Proof. Let $m = \int_{[0,\infty)} y d\mu = -\int_{(-\infty,0)} u d\mu$, then we define

$$\theta(d(x, y)) = \frac{1}{m}(y - x)\mu(dx)\mu(dy),$$

a scaling of the product measure $\mu(du)\mu(dv)$ by $\frac{1}{m}(y - x)$. Then, to check that this is a probability measure, note

$$\begin{aligned}
\int d\theta &= \frac{1}{m}\int_{(-\infty,0)} d\mu(dx) \int_{[0,\infty)} \mu(dy)(y - x) \\
&= \frac{1}{m}\int_{(-\infty,0)} \mu(dx)(m - x\mu([0, \infty))) \\
&= \frac{1}{m}\Big(m\mu((-\infty, 0)) + m\mu([0, \infty))\Big) \\
&= 1
\end{aligned}$$

thus θ is indeed a probability measure. And, we have

$$\begin{aligned}\int \left(\frac{y}{y-x}\chi_A(x) + \frac{-x}{y-x}\chi_A(y) \right) d\theta &= \frac{1}{m}\int_{(-\infty,0)} \mu(dx) \int_{[0,\infty)} \mu(dy)(y\chi_A(x) - x\chi_A(y)) \\ &= \int_{(-\infty,0)} \mu(dx)\chi_A(x) + \int_{[0,\infty)} \mu(dy)\chi_A(y) \\ &= \mu(A),\end{aligned}$$

and the last property is checked in the same way. □

Proof of Skorokhod Embedding. Let X be arbitrary with $\mathbb{E}[X] = 0$ and $\mathbb{E}[X^2] = \sigma^2 < \infty$. Let μ be the probability law of X, and θ as in the second lemma above. And, let $\xi = (\xi_x, \xi_y)$ be a random variable taking values in $(-\infty, 0) \times [0, \infty)$ with distribution given by the measure θ. Denote by $T_{x,y}$ the stopping time constructed in the first lemma above, for the random variable taking values x and y. Let $\mathcal{G}(t)$ be the filtration generated by $\mathcal{F}_0(t)$ together with ξ, and let $T = T_{\xi_x,\xi_y}$.

We have $T \le T_{u,v}$ when $u < \xi_x$ and $v > \xi_y$, and combining this with the continuity of $B(t)$ we obtain

$$\{T \le t\} = \bigcup_{\substack{u,v\in\mathbb{Q} \\ u<0<v}} (\{\xi \in (u,0) \times [0,v)\} \cap \{T_{u,v} \le t\}) \in \mathcal{G}(t),$$

so T is a stopping time with respect to $\mathcal{G}(t)$. We have for $r < 0$ that

$$\begin{aligned}\mathbb{P}(X \le r) &= \int_{(-\infty,r]\times[0,\infty)} \frac{y}{y-x} d\theta \\ &= \int_{(-\infty,r]\times[0,\infty)} \mathbb{P}\{B(T_{u,v}) = u\} \\ &= \mathbb{P}\{B(T) \le x\}\end{aligned}$$

and in the same way for $r \ge 0$ we have $\mathbb{P}\{X > r\} = \mathbb{P}\{B(T) > x\}$, hence $B(T) \sim X$. And, we have

$$\mathbb{E}[T] = -\mathbb{E}[\xi_x\xi_y] = -\int uv\, d\theta = \sigma^2.$$

□

As an intermediate application, we will show the power of this theorem by proving the law of the iterated logarithm for Brownian motion and then inferring it for discrete random walks.

6.1 Application: The Law of the Iterated Logarithm (LIL)

We saw in our elementary asymptotic bounds that the standard Brownian motion $B(t)$ is almost surely dominated by t, in that $\frac{B(t)}{t} \to 0$ almost surely, but is almost surely *not* dominated by $\sqrt{t}$, in that $\frac{B(t)}{\sqrt{t}}$ is of unbounded variation almost surely. We would like to identify a function $\psi(t)$ that is deterministic and gives an asymptotic envelope for $B(t)$, in that $\limsup_{t\to\infty} \frac{B(t)}{\psi(t)} = 1$ and $\liminf_{t\to\infty} \frac{B(t)}{\psi(t)} = -1$ almost surely. The basic bounds give $\sqrt{t} \ll \psi(t) \ll t$, and we will show that in fact the lower bound is very close to the envelope, in that taking $\psi(t) = \sqrt{2t\ell(t)}$ gives the two results above, where $\ell(t) = \log\log(t)$ is the namesake iterated logarithm. First, we give a heuristic reason to believe that this is the correct function.

Recall that we defined $\tilde{B}(t) = \frac{B(t)}{\sqrt{t}}$ earlier, where the distribution of $\tilde{B}(t)$ is standard normal for all $t > 0$. Then, the statements about $B(t)$ and $\psi(t)$ above are equivalent to the same statements for $\tilde{B}(t)$ and $\tilde{\psi}(t) = \sqrt{2\ell(t)}$. Combining the asymptotic in 8.3 and the standard normal distribution of $\tilde{B}(t)$, we find

$$\begin{aligned}\mathbb{P}\{B(t) > L\psi(t)\} &= \mathbb{P}\{\tilde{B}(t) > L\tilde{\psi}(t)\} \\ &\sim \frac{1}{L\tilde{\psi}(t)}\frac{1}{\sqrt{2\pi}}e^{-\frac{L^2\tilde{\psi}(t)^2}{2}} \\ &= \frac{L}{\sqrt{2\ell(t)}}\frac{1}{\sqrt{2\pi}}e^{-L^2\ell(t)} \\ &= \frac{L}{2\sqrt{\pi\ell(t)}}(\log(t))^{-L^2}.\end{aligned}$$

Now, ideally we could choose some collection of t increasing to infinity so that the events whose probability we are estimating above are independent, then the question of whether infinitely many of them occur would reduce to the summability of the series whose terms are the final approximation by the Borel-Cantelli lemma. Unfortunately, these events will not be independent; however, we will see that we can achieve near-independence by taking rapidly increasing t. In particular, if we fix some $a > 1$ and make the change of variables $t = a^u$, then the above becomes

$$\mathbb{P}\{B(t) > L\psi(t)\} \sim \frac{L}{2\sqrt{\pi\ell(a^u)}}(u\log(a))^{-L^2},$$

which is summable as a series over positive integer u if and only if $L > 1$. Were these events to be independent, this would be precisely what we need to apply Borel-Cantelli and obtain our result. The work that remains is to circumvent the fact that these events are not in fact independent, which will ultimately reduce to taking a limit as $a \to \infty$, since as the points a^u become increasingly distant, they "approach independence", and the summability criterion does not depend on a.

Theorem 6.4 (LIL for Brownian Motion). If $B(t)$ is a standard Brownian motion and $\psi(t) = \sqrt{2t\ell(t)}$, then almost surely

$$\limsup_{t\to\infty}\frac{B(t)}{\psi(t)} = 1 \text{ and } \limsup_{t\to\infty}\frac{B(t)}{\psi(t)} = -1.$$

Proof. Note that by the symmetry of reflection at the origin, the second claim follows immediately from the first, so it suffices to consider the supremum. The easier part is bounding the supremum from above, since there we can use the direction of the Borel-Cantelli lemma not requiring independent events, which gives the result easily.

Letting $a > 1$ and $\epsilon > 0$, define the event

$$A_n = \left\{\max_{t\in[0,a^n]} B(t) \geq (1+\epsilon)\psi(q^n)\right\}.$$

By Corollary 4.9, the maximum of $B(t)$ up to a fixed time t_0 has the same distribution as $|B(t_0)|$, so applying this and transfering to the normalized Brownian motion and corresponding bound,

$$\begin{aligned}\mathbb{P}A_n &= \mathbb{P}\left\{|B(a^n)| \geq (1+\epsilon)\psi(a^n)\right\} \\ &= \mathbb{P}\{|\tilde{B}(a^n)| \geq (1+\epsilon)\tilde{\psi}(a^n)\} \\ &= 2\mathbb{P}\{\tilde{B}(a^n) \geq (1+\epsilon)\tilde{\psi}(a^n)\}\end{aligned}$$

where the last step follows by symmetry. Now, we apply the tail bound for the standard normal variable $\tilde{B}(a^n)$ and discard non-exponential terms to obtain

$$\mathbb{P}A_n \leq 2 \cdot \frac{1}{(1+\epsilon)\tilde{\psi}(a^n)} \frac{1}{\sqrt{2\pi}} e^{-\frac{(1+\epsilon)^2\tilde{\psi}(a^n)^2}{2}} \leq 2e^{-(1+\epsilon)^2 \log\log(a^n)^2} = \frac{2}{(n \log a)^{(1+\epsilon)^2}}.$$

This series is summable in n, so by Borel-Cantelli we have that almost surely, only finitely many of the A_n occur. Now, given a large t, let n be such that $a^{n-1} \leq t < a^n$, then we expand

$$\frac{B(t)}{\psi(t)} = \frac{B(t)}{\psi(a^n)} \cdot \frac{\psi(a^n)}{a^n} \cdot \frac{t}{\psi(t)} \cdot \frac{a^n}{t} = \frac{B(t)}{\psi(a^n)} \cdot \frac{\frac{\psi(a^n)}{a^n}}{\frac{\psi(t)}{t}} \cdot \frac{a^n}{t}.$$

The first term is almost surely eventually bounded by $(1+\epsilon)$ by what we have just shown, the last term is at most a, and the middle term is bounded by 1 since $\frac{\psi(s)}{s}$ is a decreasing function. Thus, $\frac{B(t)}{\psi(t)} \leq a(1+\epsilon)$ almost surely for all large t, and letting $\epsilon \to 0$ and $a \to 1$ gives $\limsup \frac{B(t)}{\psi(t)} \leq 1$ almost surely.

To bound the supremum from below, we need the converse Borel-Cantelli lemma, and thus need to produce a sequence of independent events. Again fixing $a > 1$, we define the events

$$D_n = \{B(a^n) - B(a^{n-1}) \geq \psi(a^n - a^{n-1})\},$$

which are independent since the increments $B(a^n) - B(a^{n-1})$ are independent. The random variable $B(a^n) - B(a^{n-1})$ is normal with mean zero and variance $a^n - a^{n-1}$, so

$$\mathbb{P}D_n = \mathbb{P}\left\{\mathcal{N} \geq \frac{\psi(a^n - a^{n-1})}{\sqrt{a^n - a^{n-1}}}\right\} = \mathbb{P}\{\mathcal{N} \geq \tilde{\psi}(a^n - a^{n-1})\}.$$

Applying the tail lower bound from 8.2, we have for some $c > 0$ that

$$\mathbb{P}D_n \geq \frac{c}{\sqrt{2\log\log(a^n - a^{n-1})}} e^{-\log\log(q^n - q^{n-1})} \geq \frac{ce^{-\log(n \log a)}}{\sqrt{2\log(n \log a)}} > \frac{d}{n \log n}$$

for some other constant $d > 0$, and so in particular the series of $\mathbb{P}D_n$ is not summable, thus by the converse of Borel-Cantelli we have almost surely, infinitely many of the D_n occur. Combining this with the upper bound proved above, for infinitely many n we have

$$B(q^n) \geq B(q^{n-1}) + \psi(q^n - q^{n-1}) \geq -2\psi(q^{n-1}) + \psi(q^n - q^{n-1}).$$

So, almost surely for infinitely many n,

$$\frac{B(q^n)}{\psi(q^n)} \geq \frac{-2\psi(q^{n-1}) + \psi(q^n - q^{n-1})}{\psi(q^n)} \geq -\frac{2}{\sqrt{q}} + \frac{q^n - q^{n-1}}{q^n} = 1 - \frac{2}{\sqrt{q}} - \frac{1}{q},$$

and taking $q \to \infty$ gives the lower bound, completing the proof. □

Using this, we can obtain a general result for discrete random walks with relatively little work (the discrete combinatorial proofs of this theorem are much more technical and lengthy).

Theorem 6.5 (Hartman-Wintner LIL). Let S_n be a random walk with increments $S_n - S_{n-1}$ of mean 0 and variance 1. Then,

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

Proof. Let $T_1 \le T_2 \le \cdots$ be a sequence of stopping times as from the Skorokhod embedding so that $S_n \sim B(T_n)$, and where $T_n - T_{n-1} = \mathrm{Var}(S_2 - S_1) = 1$. By the LIL for Brownian motion, we have

$$\limsup_{t\to\infty} \frac{B(t)}{\sqrt{2t\log\log t}} = 1,$$

almost surely, and so it suffices to show that

$$\limsup_{t\to\infty} \frac{B(t) - B(T_{\lfloor t\rfloor})}{2t\log\log t} = 0$$

almost surely. By the law of large numbers, $\frac{1}{n}T_n \to 1$ almost surely, so letting $\epsilon > 0$ and t_0 large enough such that

$$\frac{1}{1+\epsilon} \le \frac{T_{\lfloor t\rfloor}}{t} \le 1+\epsilon$$

for all $t \ge t_0$. Then, we define

$$M(t) = \sup_{s\in[\frac{t}{1+\epsilon}, t(1+\epsilon)]} |B(s) - B(t)|,$$

and note that it suffices to show that

$$\limsup_{t\to\infty} \frac{M(t)}{\sqrt{2t\log\log t}} = 0$$

almost surely. Let $t_n = (1+\epsilon)^n$, and

$$L_n = \sup_{s\in[t_{n-1}, t_{n+2}]} |B(s) - B(t_{n-1})|.$$

Then, by the triangle inequality we have when $t \in [t_n, t_{n+1}]$, then $M(t) \le 2L_n$. Let $\delta = (1+\epsilon)^3 - 1$, then $t_{n+2} - t_{n-1} = \delta t_{n-1}$. Scaling and the reflection principle then gives

$$\begin{aligned}
\mathbb{P}\left(L_n > \sqrt{3\delta t_{n-1}\log\log t_{n-1}}\right) &= \mathbb{P}\left(\sup_{s\in[0,1]} |B(s)| > \sqrt{3\log\log t_{n-1}}\right) \\
&\le 2\mathbb{P}\left(\sup_{s\in[0,1]} B(s) > \sqrt{3\log\log t_{n-1}}\right) \\
&= 4\mathbb{P}\left(B(1) > \sqrt{3\log\log t_{n-1}}\right) \\
&\le \frac{2}{\sqrt{3\log\log t_{n-1}}} \exp\left(-\frac{3}{2}\log\log t_{n-1}\right) \\
&\le n^{-\frac{3}{2}}
\end{aligned}$$

for n sufficiently large. These probabilities are summable, so by the Borel-Cantelli lemma,

$$\limsup_{t\to\infty} \frac{M(t)}{\sqrt{t\log\log t}} \le \limsup_{n\to\infty} \frac{2L_n}{\sqrt{t_{n-1}\log\log t_{n-1}}} \le 2\sqrt{3\delta},$$

and letting $\epsilon \to 0$ we have $\delta \to 0$, completing the proof. □

7 Discrete Random Walks: Donsker Invariance Principle

We lastly turn to proving the functional central limit theorem described in the introduction. Letting X_i be i.i.d random variables, normalized so that $\mathbb{E}[X_i] = 0$ and $\mathrm{Var}(X_i) = 1$ (this is really just the assumption of finite mean and variance for the X_i), we will consider the random walk $S_n = \sum_{k=1}^n X_k$ and produce a continuous random process $S(t)$ by interpolating linearly between these integer points:

$$S(t) = S_{\lfloor t\rfloor} + \{t\}(S_{\lfloor t\rfloor+1} - S_{\lfloor t\rfloor}).$$

We now define a family S_n^* of derived stochastic processes by

$$S_n^*(t) = \frac{1}{\sqrt{n}}S(nt)$$

for $t \in [0,1]$. Intuitively, S_n^* scales $S(t)$ on $[0,n]$ into the interval $[0,1]$.

Theorem 7.1 (Donsker Invariance Principle). On the space $C[0,1]$ of continuous functions on $[0,1]$ with the metric induced by the supremum norm, the sequence S_n^* converges in distribution to a standard Brownian motion $B(t)$ on $t \in [0,1]$.

Lemma 7.2. For any random variable X with mean 0 and variance 1, there is a sequence of stopping times $0 = T_0 \le T_1 \le \cdots$ such that $B(T_n)$ has the distribution of the random walk S_n defined by adding independent variables identical to X, and the sequence of functions S_n^* satisfies

$$\lim_{n\to\infty} \mathbb{P}\left\{\sum_{t\in[0,1)} \left|\frac{B(nt)}{\sqrt{n}} - S_n^*(t)\right| > \epsilon\right\} = 0.$$

Proof. By Skorokhod embedding, choose a stopping time T_1 with $\mathbb{E}[T_1] = 1$ so that $B(T_1) \sim X$. By the strong Markov property, $B_2(t) = B(T_1 + t) - B(T_1)$ is a Brownian motion and independent of $\mathcal{F}^+(T_1)$. In particular, $B_2(t)$ is independent of both T_1 and $B(T_1)$. Thus, we can choose a stopping time T_2' so that $\mathbb{E}[T_2'] = 1$ and $B_2(T_2') \sim X$. Then, $T_2 = T_1 + T_2'$ is a stopping time for $B(t)$ with $\mathbb{E}[T_2] = 2$, and so that $B(T_2) \sim S_2 = X_1 + X_2$. We repeat this inductively to produce a sequence $0 \le T_1 \le T_2 \le \cdots$ where $B(T_n) \sim S_n$ and $\mathbb{E}[T_n] = n$.

It remains to verify the convergence statement. Let us write

$$B_n(t) = \frac{B(nt)}{\sqrt{n}}$$

and let A_n be the event that there exists $t \in [0,1)$ with $|S_n^*(t) - B_n(t)| > \epsilon$, then we must show $\mathbb{P}(A_n) \to 0$. Let k be such that $\frac{k-1}{n} \le t < \frac{k}{n}$, then since S_n^* is linear on this interval, we have that A_n is contained in the union of there being a $t \in [0,1)$ such that $|\frac{S_k}{\sqrt{n}} - B_n(t)| > \epsilon$, and of there being a

t such that $|\frac{S_{k-1}}{\sqrt{n}} - B_n(t)| > \epsilon$. Since $S_k = B(T_k) = \sqrt{n}B_n(\frac{T_k}{n})$, we have

$$\begin{aligned} A_n \subseteq \{&\exists t \in [0,1) : |B_n(T_k/n) - B_n(t)| > \epsilon\} \cup \\ &\{\exists t \in [0,1) : |B_n(T_{k-1}/n) - B_n(t)| > \epsilon\}. \end{aligned}$$

Let A_n^* be the event on the right-hand side. Then, for any $0 < \delta < 1$, we have

$$\begin{aligned} A_n^* \subseteq \{&\exists s,t \in [0,2] : |s-t| < \delta, |B_n(s) - B_n(t)| > \epsilon\} \cup \\ &\{\exists t \in [0,1) : |T_k/n - t| \vee |T_{k-1}/n - t| \geq \delta\}. \end{aligned}$$

Note that the first event here does not depend on n, and its probability tends to zero as $\delta \to 0$ since Brownian motion is uniformly continuous on $[0,2]$. So, it remains to show that for any $\delta > 0$, the probability of the second set above tends to zero as $n \to \infty$. To prove this, we huse that by the law of large numbers, $\lim_{n\to\infty} \frac{T_n}{n} = 1$, and thus $\lim_{n\to\infty} \sup_{0\leq k\leq n} |\frac{T_k}{k} - k|/n = 0$. This gives that $\mathbb{P}(A_n^*) \to 0$, completing the proof. □

Proof of Donsker Invariance Principle. Choose the stopping times T_i as in the lemma, and note that $B_n(t)$ is a standard Brownian motion by scaling invariance. Suppose $K \subseteq C([0,1])$ is closed and let

$$K_\epsilon = \{f \in C([0,1]) : \|f - g\|_\infty \leq \epsilon \text{ for some } g \in K\}.$$

Then, we have $\mathbb{P}\{S_n^* \in K\} \leq \mathbb{P}\{B_n \in K_\epsilon\} + \mathbb{P}\{\|S_n^* - B_n\|_\infty > \epsilon\}$. The second term tends to zero as $n \to \infty$, and the first term is independent of n since the B_n are identical in law, and is equal to $\mathbb{P}\{B \in K_\epsilon\}$ for a fixed standard Brownian motion B. Since K is closed, we have

$$\lim_{\epsilon\to 0} \mathbb{P}\{B \in K_\epsilon\} = \mathbb{P}\left\{B \in \bigcap_{\epsilon>0} K_\epsilon\right\} = \mathbb{P}\{B \in K\},$$

and so $\limsup_{n\to\infty} \mathbb{P}\{S_n^* \in K\} \leq \mathbb{P}\{B \in K\}$. The result then follows as this is one of the equivalent conditions from the portmanteau theorem. □

We recall our corollary from the introduction that motivated the limit theorem.

Corollary 7.3. For any continuous functional $\Lambda : C([0,1]) \to \mathbb{R}$, we have $\Lambda(S_n^*) \Rightarrow \Lambda(B)$.

A wide variety of limit theorems for discrete random walks are now readily available by one of two techniques: firstly, we can just directly compute limiting distributions on $B(t)$ if it is convenient, or alternatively we can compute limiting distributions on a particular well-behaved discrete random walk, say the simple symmetric random walk, since we know that limit laws will not depend on the specific random walk. Two examples mentioned in the introduction illustrate these methods; we first note in passing that since the central limit theorem is an easy corollary of this theorem, we have actually found a new proof of the central limit theorem which, though rather roundabout, makes no reference to characteristic functions or Fourier analysis.

For the maximum of the discrete random walk, if we let $M_n = \max_{1\leq k\leq n} S_k$, then $\frac{1}{\sqrt{n}} M_n \Rightarrow \sup_{t\in[0,1]} B(t)$, which by Corollary 4.9 has the distribution of $|B(1)|$, or $|\mathcal{N}(0,1)|$.

For the proportion of time spent above the x-axis, we have

$$\frac{1}{n}\#\{k : 1 \le k \le n, S_k > 0\} \Rightarrow m(\{s \in [0,1] : B(s) > 0\}),$$

and so in particular the limiting law for a particular discrete random walk is the same as that for all discrete walks. So, the limiting law is the same as that for the simple symmetric random walk, which from classical theory follows the arcsine distribution [1, Section 3.10].

8 Appendix

8.1 Standard Normal Distribution Bounds

Some basic bounds on the tails of the standard normal distribution are useful.

Lemma 8.1 (Tail Bounds). Suppose N is a random variable with standard normal distribution, then for any $x > 0$, we have

$$\frac{1}{x + \frac{1}{x}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \le \mathbb{P}\{\mathcal{N}(0,1) > x\} \le \frac{1}{x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Proof. For the inequality on the right, we estimate

$$\mathbb{P}\{\mathcal{N}(0,1) > x\} \le \frac{1}{\sqrt{2\pi}} \int_x^\infty \frac{u}{x} e^{-\frac{u^2}{2}} \, du = \frac{1}{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

For the inequality on the left, define

$$f(x) = x e^{-\frac{x^2}{2}} - (x^2 + 1) \int_x^\infty e^{-\frac{u^2}{2}} \, du,$$

so that $f(0) < 0$ and $f(x) \to 0$ as $x \to \infty$. We also have

$$f'(x) = -2x \left(\int_x^\infty e^{-\frac{u^2}{2}} \, du - \frac{e^{-\frac{x^2}{2}}}{x} \right),$$

which is positive for $x > 0$ by the other inequality proved above. Thus, $f(x) \le 0$ for all x, completing the proof. $\square$

For technical work, a simplified version is usually enough for us.

Corollary 8.2. There is a universal constant $c > 0$ such that

$$\frac{c}{x} e^{-x^2/2} \le \mathbb{P}\{\mathcal{N}(0,1) > x\} \le e^{-x^2/2}.$$

For heuristics, an asymptotic version is useful.

Corollary 8.3. For large x,

$$\mathbb{P}\{\mathcal{N}(0,1) > x\} \sim \frac{1}{x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

References

[1] Geoffrey Grimmett, David Stirzaker, *Probability and Random Processes*. Oxford University Press, Third Edition, 2001.

[2] Olav Kallenberg, *Foundations of Modern Probability*. Springer, New York, 1997.

[3] Peter Morters, Yuval Peres, *Brownian Motion*. Cambridge University Press, 2010.

[4] Cosma Shalizi, *Almost None of the Theory of Stochastic Processes*. Draft, 2010.