BROWNIAN MOTION AS THE LIMITING DISTRIBUTION OF RANDOM WALKS

CHRISTIAN LIU

Abstract. The goal of this expository paper is to provide an accessible introduction to Brownian motion and to prove that Brownian motion is the limiting distribution of scaled and linearly interpolated random walks. On the way to proving this result, which is known either as Donsker's invariance principle or the functional central limit theorem, we examine the Markov and martingale properties of Brownian motion, some results about the convergence of random variables, and the Skorokhod embedding theorem.

Contents

1. Introduction

The classical central limit theorem says that the Normal distribution is the limiting distribution of scaled means of random samples from any distribution with finite variance. This makes the Normal distribution powerful in at least two ways: first, it is a continuous *scaling limit* for discrete models, and second, it is a *universal* limit for many different types of models.

A continuous approximation to a discrete model is a useful result because it is usually easier to work with a continuous function than a discrete function. Integrals can be computed more easily (and more frequently) than sums can, and a function

Date: August 28, 2021.

defined on a continuum can be rescaled in ways that don't work for a function defined on a discrete set. Having a limit universal to many underlying models is also desirable for the practical reason that it allows results deduced from the easiest underlying models to be generalized to more complicated cases.

Donsker's invariance principle, also known as the functional central limit theorem, extends the central limit theorem from random variables to random functions. It states that one-dimensional Brownian motion is the limiting distribution of properly scaled and linearly interpolated random walks defined from any distribution with zero mean and finite variance, and it imbues Brownian motion with these same two properties that make the Normal distribution compelling.

Donsker's invariance principle also helps connect the physical description of Brownian motion to its mathematical representation. Brownian motion was a term first introduced to describe erratic random motion, such as the motion of particles suspended in a fluid. The term is named for botanist Robert Brown, who in 1827 observed that bits of pollen dance around when suspended in water, a phenomenon that physicists now know is caused by the constant bombardment of the suspended particles by neighboring atoms. The mathematical object that formalizes Brownian motion is a specific stochastic process introduced by Norbert Weiner in 1918, and it is sometimes called the Weiner process (although we will refer to it as Brownian motion outside of this introduction). Donsker's invariance principle bridges Brownian motion and the Weiner process: a natural approximation of a particle's movement is its path on a lattice model of discrete space and time, and our theorem says that as the spacing of this lattice goes to zero, the approximation becomes the Weiner process, an exact description of Brownian motion.

This paper closely follows Chapters 2, 5, and 12 of [2], with added remarks and detail. Section 2 will introduce Brownian motion and other necessary tools. Section 3 will discuss properties of Brownian motion, and Section 4 will define random walks. The proof of Donsker's invariance principle is the core of Sections 5 and 6. Familiarity with basic definitions in measure-theoretic probability and the various types of convergence of random objects may be helpful.

2. Preliminaries

The theorem we aim to prove is a statement about convergence in distribution, so we begin by defining what it means for random objects to converge in this way. Then, this section introduces Brownian motion, filtrations, and stopping times, all of which are needed to discuss fundamental properties of Brownian motion in the next section. The definitions here are accompanied by heavy commentary meant to clarify measure theory considerations and provide intuition, but which can be skipped if desired.

2.1. **Convergence in Distribution.** Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, and S a topological space. A *random object* X is a measurable map from Ω to S, and its *distribution* is the probability measure on S defined by $\mathbb{P}(X \in A)$ for, at minimum, any Borel set $A \subset S$.

We say that a sequence of measures $\mu_1, \mu_2, \dots$ on some space *converges weakly* to the measure μ if for all bounded, continuous functions $f : S \to \mathbb{R}$,

$$\int f d\mu_n \to \int f d\mu.$$

A sequence of random objects $X_1, X_2, \ldots$ taking values in some space S converges in distribution to the random object X in S if the distributions of $X_1, X_2, \ldots$ converge weakly to the distribution of X.

Definition 2.1. Let X and $X_1, X_2, \ldots$ be random objects taking values in a space S. We say X_n *converges in distribution* to X if for all bounded, continuous functions $f : S \to \mathbb{R}$,

$$\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)].$$

Convergence in distribution is notated by $X_n \xrightarrow{d} X$.

When S is a *Polish space*, which is a separable and completely metrizable topological space, several conditions are equivalent to convergence in distribution.

Theorem 2.2. *(Portmanteau theorem) Let X and $X_1, X_2, \ldots$ be random objects taking values in a Polish space S. Then, the following are equivalent:*

(1) $X_n \xrightarrow{d} X$
(2) For all closed sets $F \subset S$, $\limsup_{n\to\infty} \mathbb{P}\{X_n \in F\} \leq \mathbb{P}\{X \in F\}$
(3) For all open sets $G \subset S$, $\liminf_{n\to\infty} \mathbb{P}\{X_n \in G\} \geq \mathbb{P}\{X \in G\}$
(4) For all Borel sets $A \subset S$ with $\mathbb{P}(X \in \partial A) = 0$, we have $\lim_{n\to\infty} \mathbb{P}\{X_n \in A\} = \mathbb{P}\{X \in A\}$
(5) For all bounded measurable functions $f : S \to \mathbb{R}$ with f almost surely continuous at X, we have $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$.

See Chapter 12 of [2] for a proof.

2.2. **Brownian Motion.** A random object that maps from a probability space into a function space is called a *random function*. If the function space consists of functions with a one-dimensional parameter, the random function is called a *stochastic process*, and the parameter is usually interpreted as time.

Brownian motion is a particular stochastic process whose parameter takes values in $[0, \infty)$. In the one-dimensional case, Brownian motion is a real-valued function, so it is natural to define one-dimensional Brownian motion by its particulars as a random object in $\mathbb{R}^{[0,\infty)}$.

Definition 2.3. A *linear (one-dimensional) Brownian motion* $\{B(t) : t \geq 0\}$ started at x is a real-valued stochastic process that satisfies the following conditions:

(1) $B(0) = x$
(2) the process has independent increments: for all finite sets of times $0 \leq t_1 \leq t_2 \leq \cdots \leq t_n$, the random variables $B(t_n) - B(t_{n-1}), \ldots, B(t_2) - B(t_1)$ are independent
(3) the process has stationary, Normally distributed increments: for all $t \leq 0$ and $h > 0$, we have $B(t + h) - B(t) \sim \mathrm{N}(0, h)$
(4) almost surely, $t \mapsto B(t)$ is continuous.

A *standard Brownian motion* is a Brownian motion started at the origin.

Remark 2.4. When we describe a linear Brownian motion $\{B(t) : t \geq 0\}$ as a map from some probability space into $\mathbb{R}^{[0,\infty)}$, by default we assume the product topology on $\mathbb{R}^{[0,\infty)}$. Since

$$\mathbb{P}\big(\{B(t) : t \geq 0\} \in \mathbf{C}[0, \infty)\big) = 1$$

by definition, Brownian motion's distribution is a probability measure on $\mathbb{R}^{[0,\infty)}$ defined for, at minimum, each set in

$$\mathcal{A}^* := \sigma\big(\{\mathfrak{B} : \mathfrak{B} \text{ is Borel in } \mathbb{R}^{[0,\infty)} \text{ with the product topology}\} \cup \{\mathbf{C}[0,\infty)\}\big).$$

We now assert that the requirements in Definition 2.3 are consistent with a distribution on $\mathbb{R}^{[0,\infty)}$ and in fact determine a unique distribution.

Theorem 2.5. *(Construction of Brownian motion) Brownian motion exists, i.e. there is a probability measure on* $(\mathbb{R}^{[0,\infty)}, \mathcal{A}^*)$ *that satisfies the conditions in Definition* 2.3.

It is perhaps not surprising that a distribution satisfying conditions (1)-(3) of Definition 2.3 should exist, since the sum of Normally distributed random variables is also Normally distributed. The key result to show for existence is that the continuity requirement of Brownian motion is compatible with the first three properties. This is not obvious; for example, no almost surely continuous stochastic process with Gamma-distributed increments exists, although the sum of Gamma-distributed random variables also stays within the distribution family.

Levy's construction of Brownian motion, covered in Chapter 1 of [2], establishes continuity by letting Brownian motion be the limit of a sequence of continuous random functions that converge uniformly, almost surely. A random function $B^{(n)}(\omega)$ in this sequence is defined by setting its values at the times $\{t = \frac{k}{2^n} : k = 0, \dots, 2^n\}$ to be Normal random variables, and then *linearly interpolating* (connecting with straight lines) the values at these times to determine the function for all other t. Using the Borel-Cantelli lemma, it is possible to show that

$$\mathbb{P}\big(B^{(n)}(\omega) \text{ converges uniformly}\big) = 1,$$

so the limit of the $B^{(n)}(\omega)$ is continuous and satisfies the conditions in Definition 2.3.

Theorem 2.6. *(Uniqueness of Brownian motion) Brownian motion is unique, i.e. there is a unique probability measure on* $(\mathbb{R}^{[0,\infty)}, \mathcal{A}^*)$ *satisfying the conditions in Definition* 2.3.

See Chapter 2 of [3] for a proof of uniqueness.

Remark 2.7. It is also natural to view Brownian motion as a random object in the spaces $\mathbf{C}[0,\infty)$ and $\mathbf{C}[0,1]$ instead of $\mathbb{R}^{[0,\infty)}$. We need to investigate which topologies must be applied to these spaces for Brownian motion to define a distribution on the space.

The natural topology for $\mathbf{C}[0,\infty)$ is the subspace topology induced by the product topology on $\mathbb{R}^{[0,\infty)}$. The Borel sets of $\mathbf{C}[0,\infty)$ with this topology include all sets in $\mathcal{A}^*$ assigned positive measure by Brownian motion's distribution on $(\mathbb{R}^{[0,\infty)}, \mathcal{A}^*)$. So, simply restricting our original distribution on $\mathbb{R}^{[0,\infty)}$ to the Borel sets of $\mathbf{C}[0,\infty)$ gives a valid probability measure that characterizes Brownian motion as a random function in $\mathbf{C}[0,\infty)$.

Since Brownian motion's increments are independent and identically distributed, its distribution on $\mathbf{C}[0,\infty)$ must be consistent with viewing $\mathbf{C}[0,\infty)$ as the concatenation of independent copies of $\mathbf{C}[0,1]$. In other words, Brownian motion is also defined by its distribution on $\mathbf{C}[0,1]$ with the subspace topology induced by the product topology. This is useful because proving results for a measure on $\mathbf{C}[0,1]$ can be easier than proving results for a measure on $\mathbf{C}[0,\infty)$.

The classical sup-norm topology on $\mathbf{C}[0,1]$ is strictly finer than the product topology on $\mathbf{C}[0,1]$, since the former is associated with the uniform convergence of functions and the latter with pointwise convergence. However, the two topologies give rise to the same Borel sets. To show this, it suffices to show that a basis of the sup-norm topology is a subset of the Borel sets of the product topology. The collection of neighborhoods of the form

$$\{g \in \mathbf{C}[0,1] : ||f-g||_\infty < \epsilon\} = \bigcup_{n\in\mathbb{N}} \bigcap_{q\in[0,1]\cap\mathbb{Q}} \{g \in \mathbf{C}[0,1] : |f(q)-g(q)| < \epsilon - \frac{1}{n}\}$$

for all $f \in \mathbf{C}[0,1]$ and $\epsilon > 0$ forms such a basis. Thus, Brownian motion can be understood as a random object in $\mathbf{C}[0,1]$ with either topology.

We make this remark because in our proof of Donsker's Invariance Principle, we will show convergence in distribution of random walks to Brownian motion on the space $\mathbf{C}[0,1]$ with the sup-norm topology.

2.3. **Filtrations and Stopping Times.** Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. The *atoms* of $\mathcal{A}$ are the sets in $\mathcal{A}$ that cannot be written as a union of other sets; these indivisible sets form a partition of the state space. A random object X defined on $(\Omega, \mathcal{A}, \mathbb{P})$ must map every outcome in an atom of $\mathcal{A}$ to the same value in its range space in order to be $\mathcal{A}$-measurable. In this way, the σ-algebra $\mathcal{A}$ determines the "clarity" of information that can be known about an outcome from Ω by observing the value of X— more sets in $\mathcal{A}$ correspond to a finer a partition by the atoms and greater clarity.

A filtration is a sequence of σ-algebras that represents gaining clarity over time.

Definition 2.8. A *filtration* on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is a non-decreasing family of sub-σ-algebras of $\mathcal{A}$. A filtration with index set $[0,\infty)$ is notated by $(\mathcal{F}(t) : t \geq 0)$ and satisfies $\mathcal{F}(s) \subset \mathcal{F}(t) \subset \mathcal{A}$ for all $s \leq t$.

A stochastic process $\{X(t) : t \geq 0\}$ is *adapted* to a filtration $(\mathcal{F}(t) : t \geq 0)$ on its probability space if $X(t)$ is $\mathcal{F}(t)$-measurable for all $t \geq 0$.

The *natural filtration* of a stochastic process $\{X(t) : t \geq 0\}$ is the smallest filtration that it is adapted to and is defined by

$$\mathcal{F}^X(t) = \sigma\big(\{X(s) : s \in [0,t]\}\big).$$

A stochastic process adapted to a filtration is analogous to a random variable measurable with respect to a σ-algebra: the filtration determines the clarity of information that can be known about the stochastic process over time. In particular, the natural filtration represents knowing at each time t the path of the stochastic process up to time t. We will only use the natural filtration, but other filtrations exist and can be more powerful; see Remark 3.10.

One immediate use of filtrations is to define stopping times.

Definition 2.9. A random variable T taking values in $[0,\infty]$ is a *stopping time* with respect to a filtration $(\mathcal{F}(t) : t \geq 0)$ on its probability space if $\{T \leq t\} \in \mathcal{F}(t)$ for every $t \geq 0$.

Stopping times let us index a stochastic process by a time that is itself random, because it depends on the realization of the stochastic process. This allows us to do much more than if we were restricted to only using deterministic times.

For example, the first time that a Brownian motion hits some value is a stopping time with respect to its natural filtration. This hitting time is not known in advance, but at each time t, the natural filtration provides enough clarity of information to determine whether the hitting event has occurred and to qualify the hitting time as a stopping time with respect to the filtration.

Remark 2.10. We will refer back to the following facts about stopping times:

(1) Every deterministic time t is a stopping time with respect to every filtration.
(2) If $T_1, T_2, \ldots$ are increasing stopping times with respect to some filtration $(\mathcal{F}(t) : t \geq 0)$, their limit T is a stopping time with respect to the same filtration, since
$$\{T \leq t\} = \bigcap_{n=1}^{\infty} \{T_n \leq t\},$$
and the right-hand side is a countable intersection of elements in $\mathcal{F}(t)$, and thus in $\mathcal{F}(t)$.
(3) Fix n. If T is a stopping time with respect to some filtration $(\mathcal{F}(t) : t \geq 0)$, then the first time of the form $\frac{k}{2^n}$ that is strictly greater than T is a stopping time with respect to the same filtration. We can write the new stopping time as
$$T_n = \frac{\lfloor 2^n T \rfloor + 1}{2^n},$$
and notice that
$$\{T_n \leq t\} = \{\lfloor 2^n T \rfloor \leq 2^n t - 1\} = \bigcup_{m=1}^{\infty} \{2^n T \leq 2^n t - \frac{1}{m}\}.$$

Definition 2.11. A filtration $(\mathcal{F}(t) : t \geq 0)$ indexed by a stopping time T with respect to the filtration is defined as the σ-algebra
$$\mathcal{F}(T) = \{A \in \mathcal{F} : A \cap \{T \leq t\} \in \mathcal{F}(t) \text{ for all } t \geq 0\}.$$

This definition ensures that $\{X(s) : s \in [0, T]\}$ is $\mathcal{F}(T)$-measurable and heuristically gives the events in $\mathcal{F}$ that occur before T.

3. Properties of Brownian Motion

Brownian motion possesses many interesting properties both as a random function and as the set of points in its path. For example, it is everywhere continuous and nowhere differentiable, and questions about the area of its path and the likelihood of its path to self-intersect have been well-studied. Brownian motion also produces related stochastic processes inviting study; some, such as its running maximum, are easier to understand than others, such as loop-erased Brownian motion.

In this section, we present just the properties of Brownian motion immediately relevant to proving our theorem. These are its scaling invariance and behavior as a Markov process and martingale.

3.1. Scaling Invariance.

An easy invariance property of Brownian motion is that a scaled Brownian motion remains a Brownian motion, if the function's time parameter is also changed appropriately.

Proposition 3.1. *(Scaling invariance) Let $\{B(t) : t \geq 0\}$ be a standard Brownian motion. For any constant $a \neq 0$, the process $\{B^*(t) : t \geq 0\}$ defined by*
$$B^*(t) = \frac{1}{a} B(a^2 t)$$
is also a Brownian motion.

Proof. Scaling and changing the time parameter of $\{B(t) : t \geq 0\}$ do not affect its continuity or the independence of its increments. Also, $B^*(s)-B^*(t) = \frac{1}{a}(B(a^2s)-B(a^2t))$ has distribution $\frac{1}{a}\mathrm{N}(0, a^2(s-t)) \overset{d.}{=} \mathrm{N}(0, s-t)$, so the increments of $\{B^*(t) : t \geq 0\}$ have the correct distribution to be a Brownian motion. □

3.2. **As a Markov Process.** A Markov process with a transition kernel is the continuous time and space extension of a Markov chain with a transition matrix, which lives in discrete time and space.

Definition 3.2. A *Markov transition kernel* is a function $p : [0, \infty) \times \mathbb{R}^d \times \mathfrak{B} \to \mathbb{R}$, where $\mathfrak{B}$ is the Borel σ-algebra on $\mathbb{R}^d$, such that

(1) $p(t, x, \cdot)$ is a Borel probability measure on $\mathbb{R}^d$ for all $t \in [0, \infty)$ and starting points $x \in \mathbb{R}^d$. When we integrate a function $f : \mathbb{R}^d \to \mathbb{R}$ with respect to this measure, we use the notation $\int f(y)p(t, x, dy)$
(2) $p(\cdot, \cdot, A)$ is a measurable function in (t, x). In particular, this means $p(t, \cdot, A)$ is a measurable function in x that we can integrate with respect to any measure on $\mathbb{R}^d$. Thus, it is well-defined to require that
(3) for all $s, r \geq 0$, $x \in \mathbb{R}^d$, and $A \in \mathfrak{B}$, we have

$$p(s + r, x, A) = \int_{\mathbb{R}^d} p(r, y, A)p(s, x, dy). \tag{3.3}$$

The idea is to use p to represent, for a given stochastic process $\{X(t) : t \geq 0\}$, times $s, r \geq 0$, starting position $x \in \mathbb{R}^d$, and measurable set $A \in \mathfrak{B}$,

$$p(r, x, A) = \mathbb{P}\big(X(s + r) \in A | X(s) = x\big).$$

This interpretation of p makes sense only if the stochastic process's value at a time r increment in the future depends only on its current position x and the choice of r, i.e. it is Markov.

Definition 3.4. A stochastic process $\{X(t) : t \geq 0\}$ is a *(time-homogeneous) Markov process* with respect to a filtration $(\mathcal{F}(t) : t \geq 0)$ and with transition kernel p if

(1) it is adapted to $(\mathcal{F}(t) : t \geq 0)$ and
(2) it has the *Markov property*, which means that for all $s, r \geq 0$ and $A \in \mathfrak{B}$, almost surely,
$$\mathbb{P}\big(X(s + r) \in A|\mathcal{F}(s)\big) = \mathbb{P}\big(X(s + r) \in A|X(s)\big) = p(r, X(s), A).$$

The future evolution of a partially realized Markov process depends only on the last realized state of the process, and not on any other history of the path.

Theorem 3.5. *(Markov property) Linear Brownian motion is a Markov process with respect to its natural filtration and with transition kernel p defined by $p(t, x, .) = \mathrm{N}(x, t)$.*

For a Brownian motion $\{B(t) : t \geq 0\}$ started at $x \in \mathbb{R}$, the Markov property says that for any $s \geq 0$, the process $\{B(t+s)-B(s) : t \geq 0\}$ is a standard Brownian motion that is independent of the process $\{B(t) : t \in [0, s]\}$.

Proof. From the independence of Brownian motion's increments,

$$\mathbb{P}\big(B(s + r) \in A|\mathcal{F}(s)\big) = \mathbb{P}\big(B(s + r) \in A|B(s)\big) = \mathbb{P}(Z \in A),$$

where Z is a random variable with distribution $\mathrm{N}(B(s), r)$. It suffices to show that p is a valid transition kernel.

We know that p satisfies the first two requirements of Definition 3.2 because the Normal distribution is a Borel probability measure and varies continuously in its parameters. The last requirement is satisfied because the convolution of two Normally distributed random variables, which is the integral on the right-hand side of (3.3), has the Normal distribution on the left-hand side of (3.3). ☐

We can generalize the Markov property to any almost surely finite stopping time; this is known as the strong Markov property.

Definition 3.6. A Markov process is a *(time-homogeneous) strong Markov process* if it has the *strong Markov property*, which means that for any stopping time T satisfying $\mathbb{P}(T < \infty) = 1$, for all $r \geq 0$ and $A \in \mathfrak{B}$, almost surely,

$$\mathbb{P}\big(X(T+r) \in A | \mathcal{F}(T)\big) = \mathbb{P}\big(X(T+r) \in A | X(T)\big) = p(r, X(T), A). \tag{3.7}$$

Theorem 3.8. *(Strong Markov property) Linear Brownian motion is a strong Markov process.*

For a standard Brownian motion $\{B(t) : t \geq 0\}$, the strong Markov property says that for any almost surely finite stopping time T, the process $\{B(T+t) - B(T) : t \geq 0\}$ is a standard Brownian motion that is independent of $\mathcal{F}(T)$.

Proof. First, we show that the strong Markov property holds for stopping times T with only countably many possible values, by properties of σ-algebras. Approximation of an arbitrary stopping time by stopping times of this form proves the full result.

Fix n. Let T_n be the first time of the form $\frac{k}{2^n}$ that occurs after T, which is a stopping time by Remark 2.10. Notate by $B^{(s)}$ the Brownian motion restarted and standardized at time s, i.e. $\{B(s+t) - B(s) : t \geq 0\}$, and further abbreviate $B^* = B^{(T_n)}$. We want to show that $\sigma(B^*)$ is independent of $\mathcal{F}(T_n)$.

Choose $\{B^* \in A\} \in \sigma(B^*)$ and $E \in \mathcal{F}(T_n)$. Since $\mathbb{P}(T_n < \infty) = 1$, by countable additivity over disjoint events,

$$\mathbb{P}\big(\{B^* \in A\} \cap E\big) = \sum_{k=0}^{\infty} \mathbb{P}\big(\{B^{(k/2^n)} \in A\} \cap E \cap \{T_n = \frac{k}{2^n}\}\big).$$

Notice that $E \cap \{T_n = \frac{k}{2^n}\}$ is an element of $\mathcal{F}(\frac{k}{2^n})$ since T_n is a stopping time with respect to $(\mathcal{F}(t) : t \geq 0)$, so by the Markov property of Brownian motion, $E \cap \{T_n = \frac{k}{2^n}\}$ is independent of $\{B^{(k/2^n)} \in A\}$. Also by the Markov property, for each k, the stochastic process $B^{(k/2^n)}$ is a Brownian motion, so

$$\begin{aligned}
\mathbb{P}\big(\{B^* \in A\} \cap E\big) &= \sum_{k=0}^{\infty} \mathbb{P}(B^{(k/2^n)} \in A)\mathbb{P}\big(E \cap \{T_n = \frac{k}{2^n}\}\big) \\
&= \mathbb{P}(B^{(0)} \in A) \sum_{k=0}^{\infty} \mathbb{P}\big(E \cap \{T_n = \frac{k}{2^n}\}\big) \\
&= \mathbb{P}(B^{(0)} \in A)\mathbb{P}(E),
\end{aligned}$$

and we can conclude both that B^* is independent of $\mathcal{F}(T_n)$ and that B^* is a Brownian motion by taking $E = \Omega$. This proves (3.7) for stopping times taking values of form $\frac{k}{2^n}$.

For arbitrary T, the sequence $T_1, T_2, \ldots$ with T_n defined as above converges to T from above. We proved that $\sigma(B^{(T_n)}) \perp\!\!\!\perp \mathcal{F}(T_n)$, and $\mathcal{F}(T) \subset \mathcal{F}(T_n)$, so $\sigma(B^{(T_n)}) \perp\!\!\!\perp \mathcal{F}(T)$.

Although $\sigma(B^{(T)})$ is a superset of $\sigma(B^{(T_n)})$, we can show that every increment of $B^{(T)}$ is independent from $\mathcal{F}(T_n)$ by exploiting the continuity of Brownian motion. Let X be the random variable denoting the increment $B^{(T)}(s+t)-B^{(T)}(t)$ of $B^{(T)}$. Let X_n be the random variables denoting the increments $B^{(T_n)}(s+t)-B^{(T_n)}(t)$ of $B^{(T_n)}$. Then, almost surely,

$$X = B(s+t+T) - B(t+T) = \lim_{n\to\infty} \big(B(s+t+T_n) - B(t+T_n)\big) = \lim_{n\to\infty} X_n. \tag{3.9}$$

For any event $E \in \mathcal{F}(T_n)$ and $a < b$, by the dominated convergence theorem,

$$\begin{aligned} \mathbb{P}\big(\{X \in (a,b)\} \cap E\big) &= \int 1_{\{X\in(a,b)\}\cap E} d\mathbb{P} \\ &= \lim_{n\to\infty} \int 1_{\{X_n\in(a,b)\}\cap E} d\mathbb{P} \\ &= \lim_{n\to\infty} \mathbb{P}(X_n \in (a,b))\mathbb{P}(E) \\ &= \mathbb{P}(X \in (a,b))\mathbb{P}(E). \end{aligned}$$

This shows that every increment of $B^{(T)}$ is independent from $\mathcal{F}(T_n)$, and since $B^{(T)}$ is defined by its increments, we conclude that $B^{(T)}$ is independent from $\mathcal{F}(T_n)$.

By a similar argument, we can use the limit in (3.9) to show that the increments of $B^{(T)}$ are independent from each other and normally distributed. Since $B^{(T)}$ is also almost surely continuous, it satisfies the requirements to be a Brownian motion, completing the proof for arbitrary T. □

Remark 3.10. We can also prove that Brownian motion has the Markov and strong Markov properties with respect to its *right-continuous natural filtration*

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

where $\mathcal{F}^0$ is Brownian motion's natural filtration.

The filtration $\mathcal{F}^+$ is larger than $\mathcal{F}^0$— heuristically, $\mathcal{F}^+$ uses Brownian motion's continuity to know Brownian motion's path infinitesimally further into the future than what has been realized at time t. Therefore, the Markov property with respect to $\mathcal{F}^+$ is a stronger statement than what we have shown, since it says that a restarted Brownian motion $\{B(s+t) - B(s) : t \geq 0\}$ is independent of a larger set of events than just those in $\mathcal{F}^0$. Studying Brownian motion with the filtration $\mathcal{F}^+$ is useful because $\mathcal{F}^+$ admits more stopping times, and Chapter 2 of [2] explores this in more detail.

Remark 3.11. The Markov properties can also be defined with conditional expectations instead of probabilities; see pages 156-158 of [4] for details.

3.3. As a Martingale.

Definition 3.12. A real-valued stochastic process $\{X(t) : t \geq 0\}$ is a *martingale* with respect to the filtration $(\mathcal{F}(t) : t \geq 0)$ if

1. it is adapted to $(\mathcal{F}(t) : t \geq 0)$,
2. $\mathbb{E}[|X(t)|] < \infty$ for all $t \geq 0$, and
3. for all $s \leq t$,

$$\mathbb{E}[X(t)|\mathcal{F}(s)] = X(s), \text{ almost surely.} \tag{3.13}$$

A martingale is a stochastic process whose expected change from some last known value is always 0. If we know information about a martingale up until time t, then the best guess for the value that the process will take for any time $s > t$ is the value of the process at t.

The optional stopping theorem says that a martingale, defined to satisfy (3.13) for all deterministic times, will also satisfy (3.13) for certain stopping times.

Definition 3.14. A stochastic process $\{X(t) : t \geq 0\}$ is *dominated* by an integrable random variable if there exists a random variable D such that $\mathbb{E}[|D|] < \infty$ and for all $t \geq 0$,

$$|X(t)| \leq D, \text{ almost surely.}$$

Theorem 3.15. *(Optional stopping theorem) Let $S \leq T$ be stopping times for a martingale $\{X(t) : t \geq 0\}$ with respect to a filtration $(\mathcal{F}(t) : t \geq 0)$. If $\{X(t \wedge T) : t \geq 0\}$ is dominated by an integrable random variable, then*

$$\mathbb{E}[X(T)|\mathcal{F}(S)] = X(S), \text{ almost surely.}$$

See Chapter 2 of [2] for a proof. The optional stopping theorem is needed to prove Wald's lemmas for Brownian motion, which we will state momentarily.

Proposition 3.16. *(Martingale property) Linear Brownian motion is a martingale.*

Proof. Let $\{B(t) : t \geq 0\}$ be a linear Brownian motion. We have, for $s \leq t$,

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

by the Markov property and definition of conditional expectation in the second step. $\square$

Wald's lemmas for Brownian motion give the expectation and variance of a Brownian motion at a stopping time meeting certain criteria. We again refer the reader to Chapter 2 of [2] for the full proofs.

Theorem 3.17. *(Wald's first lemma for Brownian motion) Let T be a stopping time for a standard linear Brownian motion $\{B(t) : t \geq 0\}$ satisfying either $\mathbb{E}[T] < \infty$ or that $\{B(t \wedge T) : t \geq 0\}$ is dominated by an integrable random variable. Then,*

$$\mathbb{E}[B(T)] = 0.$$

The crux of the proof is to show that $\mathbb{E}[T] < \infty$ implies that $\{B(t \wedge T) : t \geq 0\}$ is dominated by an integrable random variable. Then, letting $S = 0$ in the optional stopping theorem yields the result.

Theorem 3.18. *(Wald's second lemma for Brownian motion) Let T be a stopping time for a standard linear Brownian motion $\{B(t) : t \geq 0\}$ satisfying $\mathbb{E}[T] < \infty$. Then,*

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

The proof uses the optional stopping theorem for the martingale $\{B(t)^2 - t : t \geq 0\}$ and the stopping times $S = 0$ and $R = T \wedge T_n$, where $T_n = \inf\{t : |B(t)| = n\}$. The optional stopping theorem is valid here because $\{B(t \wedge R)^2 - t \wedge R : t \geq 0\}$ is dominated by $n^2 + T$.

Wald's lemmas immediately imply the following result:

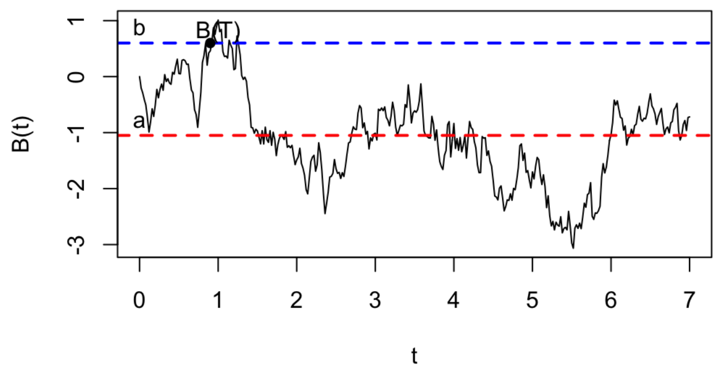

FIGURE 1. The stopping time T discussed in Theorem 3.19 is the first time that a Brownian motion hits $\{a, b\}$.

Theorem 3.19. *Choose $a < 0 < b$. Let $T = \min\{t \geq 0 : B(t) \notin (a, b)\}$ be the stopping time for a Brownian motion $\{B(t) : t \geq 0\}$ that is the first time that the Brownian motion hits a or b. Then,*

$$\mathbb{P}\big(B(T) = a\big) = \frac{b}{|a| + b}$$

and

$$\mathbb{E}[T] = |a|b.$$

Proof. We can use Wald's lemmas because $\mathbb{E}[T] < \infty$. Notice

$$\begin{aligned}\mathbb{E}[T] = \int_0^\infty \mathbb{P}(T > t)dt &= \int_0^\infty \mathbb{P}\big(B(s) \in (a, b) \text{ for all } s \in [0, t]\big)dt \\ &< \sum_{t=1}^\infty \mathbb{P}\big(B(s) \in (a, b) \text{ for all } s \in [0, t]\big) \\ &\leq \sum_{t=1}^\infty \big(\max_{x \in (a,b)} \mathbb{P}\big(B_x(s) \in (a, b) \text{ for all } s \in [0, 1]\big)\big)^t < \infty,\end{aligned}$$

where $\{B_x(t) : t \geq 0\}$ is a Brownian motion started at x.

By Wald's first lemma, $\mathbb{E}[B(T)] = 0$. Let $p = \mathbb{P}(B(T) = a)$. Then, solving

$$\mathbb{E}[B(T)] = ap + b(1 - p) = 0$$

gives the first result.

By Wald's second lemma,

$$\mathbb{E}[T] = \mathbb{E}[B(T)^2] = a^2p + b^2(1 - p) = |a|b.$$

□

4. Random Walk

Definition 4.1. A *random walk* is a sequence of random variables whose increments are independent and identically distributed with finite variance. A random walk $\{S_n : n \in \mathbb{N}\}$ can be written as $S_n = \sum_{k=1}^{n} X_k$, where $X_1, X_2, \dots$ are independent and identically distributed random variables.

A classic example of a random walk is a random walk on the integers, which starts at any integer and at each step moves to the left with probability p or to the right with probability $1 - p$.

Remark 4.2. We can view a random walk not as a sequence of random variables, but as a random sequence. Linearly interpolating the points of a random sequence produces a random continuous function, and we will derive distributions of random walks on the space of continuous functions in this manner.

Proposition 4.3. *A random walk possesses analogous properties to Brownian motion. It is a (time-homogeneous) Markov process, and it is a martingale if its increments are mean zero. A random walk is scale invariant in the sense that if X_t is a random walk then so are*

$$\tilde{X}_t = cX_t \text{ and } \hat{X}_t = X_{nt},$$

for any $c \in \mathbb{R}$ and $n \in \mathbb{N}$.

We leave the details of these proofs to the reader.

5. Skorokhod Embedding Theorem

Our proof of Donsker's invariance principle employs the Skorokhod embedding theorem, which is a result interesting in its own right. The Skorokhod embedding theorem says that we can embed a random variable with almost any distribution into a Brownian motion.

We will prove the Skorokhod embedding theorem using martingales, so we begin this section by defining uniform integrability and stating two martingale convergence theorems.

5.1. Martingale Convergence Theorems.

Definition 5.1. The $\mathbf{L}^p$ *norm* of a random variable X is

$$||X||_{\mathbf{L}^p} = \mathbb{E}[|X_n|^p]^{\frac{1}{p}}.$$

A sequence $\{X_n : n \in \mathbb{N}\}$ is *bounded in* $\mathbf{L}^p$ if $||X_n||_{\mathbf{L}^p}$ is uniformly bounded.

Definition 5.2. A sequence $\{X_n : n \in \mathbb{N}\}$ is *uniformly integrable* if, for every $\epsilon > 0$, there exists $K > 0$ such that for all n,

$$\mathbb{E}[|X_n| 1_{|X_n| \geq K}] < \epsilon.$$

Sufficient criteria for uniform integrability are:

1. $\{X_n : n \in \mathbb{N}\}$ is dominated by an integrable random variable,
2. $\{X_n : n \in \mathbb{N}\}$ is $\mathbf{L}^p$-bounded for some $p > 1$, or
3. $\{X_n : n \in \mathbb{N}\}$ is $\mathbf{L}^1$-convergent.

Uniform integrability is a property of stochastic processes that is especially useful for studying martingales.

Lemma 5.3. *A sequence of random variables that converges in probability also converges in* $\mathbf{L}^1$ *if and only if the sequence is uniformly integrable.*

See Chapter 13 of [5] for proof.

Theorem 5.4. *(Martingale convergence theorem) Let* $\{X_n : n \in \mathbb{N}\}$ *be a martingale bounded in* $\mathbf{L}^1$*. Then, there exists an integrable random variable* X *on the same probability space such that* $X_n \to X$ *almost surely.*

See [6] for proof. The idea is that an $\mathbf{L}^1$-bounded martingale satisfies $\sup \mathbb{E}[|X_n|] < \infty$, so if the martingale does not converge, it must oscillate and have infinitely many *upcrossings* between some a and b, where it dips below a and then rises above b. However, this is incompatible with the definition that the expected value of future changes of a martingale is 0.

We can say something even stronger about $\mathbf{L}^2$-bounded martingales, though it will not be needed for our later proofs.

Theorem 5.5. *(Martingale convergence theorem for* $\mathbf{L}^2$*-bounded martingales) Let* $\{X_n : n \in \mathbb{N}\}$ *be a martingale bounded in* $\mathbf{L}^2$*. Then, there exists an integrable random variable* X *on the same probability space such that* $X_n \to X$ *in* $\mathbf{L}^2$*.*

This result is proved in Chapter 12 of [2].

5.2. Skorokhod Embedding Theorem.

Definition 5.6. We say that we can *embed* a random variable into Brownian motion $\{B(t) : t \geq 0\}$ if there exists a stopping time T such that $\mathbb{E}[T] < \infty$ and $B(T)$ has the distribution of the desired random variable.

We have in fact already embedded one distribution into Brownian motion: the distribution with mean zero that is supported on two values, in Theorem 3.19. Wald's lemmas tell us that any other distribution we wish to embed must also have zero mean and finite variance, since Brownian motion at any stopping time T satisfying $\mathbb{E}[T] < \infty$ will have these properties, but the Skorokhod embedding theorem proves that these are the distribution's only required properties.

Definition 5.7. A random sequence $\{X_n : n \in \mathbb{N}\}$ is *binary splitting* if, for any event $A(x_0, \ldots, x_n) = \{X_0 = x_0, \ldots, X_n = x_n\}$ with positive probability, we have

$$\mathbb{P}\big(\omega \in A(x_0, \ldots, x_n) : X_{n+1}(\omega) \in S(x_0, \ldots, x_n)\big) = 1,$$

where $S(x_0, \ldots, x_n)$ is a set of at most two values.

In other words, the distribution of X_{n+1} conditioned on the event $A(x_0, \ldots, x_n)$ is supported on at most two values.

Lemma 5.8. *Let* X *be a random variable with finite mean and variance. Then, there exists a binary splitting martingale* $\{X_n : n \in \mathbb{N}\}$ *such that* $X_n \to X$ *almost surely and in* $\mathbf{L}^2$*.*

Proof. We proceed with a constructive proof.

Let X be defined on $(\Omega, \mathcal{A}, \mathbb{P})$. Recursively define a random sequence $\{X_n : n \in \mathbb{N}\}$ and filtration $(\mathcal{A}_n : n \in \mathbb{N})$ with the help of another random sequence $\{\xi_n : n \in \mathbb{N}\}$, all on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, as follows:

For $n = 0$,

$$\begin{aligned} \mathcal{F}_0 &= \{\emptyset, \Omega\}, \\ X_0 &= \mathbb{E}[X], \text{ and} \\ \xi_0 &= \begin{cases} 1 & \text{if } X \geq X_0, \\ -1 & \text{if } X < X_0, \end{cases} \end{aligned}$$

and for $n > 0$,

$$\begin{aligned} \mathcal{F}_n &= \sigma(\xi_0, \ldots, \xi_{n-1}), \\ X_n &= \mathbb{E}[X|\mathcal{F}_n], \text{ and} \\ \xi_n &= \begin{cases} 1 & \text{if } X \geq X_n, \\ -1 & \text{if } X < X_n. \end{cases} \end{aligned}$$

The atoms of $\mathcal{F}_n$ are the 2^n many sets

$$\begin{aligned} &\{\omega : X < X_0, X < X_1, \ldots, X < X_{n-1}\}, \\ &\{\omega : X \geq X_0, X < X_1, \ldots, X < X_{n-1}\}, \\ &\ldots, \\ &\{\omega : X \geq X_0, X \geq X_1, \ldots, X \geq X_{n-1}\} \end{aligned}$$

that partition Ω based on how X relates to each of $X_0, \ldots, X_{n-1}$. We deduce that each atom of $\mathcal{F}_n$ is the union of at most two atoms from $\mathcal{F}_{n+1}$.

Because X_n is measurable with respect to $\mathcal{F}_n$, it maps each atom of $\mathcal{F}_n$ to a constant value. An event $A(x_0, \ldots, x_{n-1})$ determines which atom of $\mathcal{F}_{n-1}$ an outcome is drawn from, so given an event $A(x_0, \ldots, x_{n-1})$, the random variable X_n is supported on at most two values. We conclude that $\{X_n : n \in \mathbb{N}\}$ is binary splitting.

To verify that $\{X_n : n \in \mathbb{N}\}$ is a martingale with respect to $(\mathcal{F}_n : n \in \mathbb{N})$, we need to show that for $m \leq n$,

$$\mathbb{E}[X_n|\mathcal{F}_m] = X_m, \text{ almost surely.}$$

This is true because, by the law of iterated expectations, for $m \leq n$,

$$\mathbb{E}[\mathbb{E}[X|\mathcal{F}_n]|\mathcal{F}_m] = \mathbb{E}[X|\mathcal{F}_m], \text{ almost surely.}$$

We move on to showing the convergence results. Notice that for all n,

$$||X_n||^2_{\mathbf{L}^2} + ||X - X_n||^2_{\mathbf{L}^2} = ||X||^2_{\mathbf{L}^2}, \tag{5.9}$$

because $X_n = \mathbb{E}[X|\mathcal{F}_n]$ is the projection of X into the Hilbert space $\mathbf{L}^2(\Omega, \mathcal{F}_n) = \{Y : \Omega \to \mathbb{R} \text{ such that } Y \text{ is } \mathcal{F}_n\text{-measurable and } \mathbb{E}[Y^2] < \infty\}$. Since the right-hand side $||X||^2_{\mathbf{L}^2} = \mathbb{E}[X^2]$ is finite, both $\{X_n : n \in \mathbb{N}\}$ and $\{X - X_n : n \in \mathbb{N}\}$ are $\mathbf{L}^2$-bounded.

The $\mathbf{L}^2$-boundedness of $\{X_n : n \in \mathbb{N}\}$ means that the martingale converges to a limit L almost surely by Theorem 5.4 (and in $\mathbf{L}^2$ by Theorem 5.5). If we can show $X = L$ almost surely, then $X_n \to X$ almost surely (and in $\mathbf{L}^2$).

The $\mathbf{L}^2$-boundedness of $\{\xi_n(X - X_n) : n \in \mathbb{N}\}$ means that it is uniformly integrable, so we can derive a statement about $\mathbf{L}^1$ convergence from the result that

$$\xi_n(X - X_{n+1}) \to |X - L| \text{ almost surely.} \tag{5.10}$$

First, we verify (5.10) by considering all outcomes in Ω. When $X(\omega) = L(\omega)$, statement (5.10) is $L - X_{n+1} \to 0$, which is true. When $X(\omega) < L(\omega)$, we have $X < X_n$ and thus $\xi_n = 1$ for all large n. In the final case, $X(\omega) > L(\omega)$, we have $\xi_n = -1$ for all large n, and again the statement is true.

By the uniform convergence of the sequence, equation (5.10) implies

$$\mathbb{E}\big[\xi_n(X - X_{n+1})\big] \to \mathbb{E}\big[|X - L|\big].$$

This left-hand side is a sequence of 0s by the definition of X_{n+1} and the law of iterated expectation, so $\mathbb{E}[|X-L|] = 0$. This proves, almost surely, that $X = L$. ☐

Because we know how to embed a mean zero random variable supported on two values into a Brownian motion, Lemma 5.8 directly leads to the Skorokhod embedding theorem.

Theorem 5.11. *(Skorokhod embedding theorem) Let X be a random variable with $\mathbb{E}[X] = 0$ and $\mathbb{E}[X^2] < \infty$. Then, there exists a stopping time T with respect to the natural filtration of a standard Brownian motion $\{B(t) : t \geq 0\}$ such that $B(T)$ has the distribution of X and $\mathbb{E}[T] = \mathbb{E}[X^2] < \infty$.*

Proof. Let $\{X_n : n \in \mathbb{N}\}$ be a binary splitting martingale that converges to X. By the binary splitting property of $\{X_n : n \in \mathbb{N}\}$ and Theorem 3.19, we can find stopping times $T_0 \leq T_1 \leq \dots$ such that $B(T_n)$ is distributed as X_n by choosing hitting values that depend on $A(x_0, \dots, x_{n-1})$.

The increasing sequence T_n converges to a limit T almost surely, and T is a stopping time by Remark 2.10. Continuity of Brownian motion and $T_n \to T$ implies $B(T_n) \to B(T)$ almost surely. By construction, $B(T_n) \to X$ almost surely, so $B(T)$ and X have the same distribution.

Wald's second lemma says $\mathbb{E}[T_n] = \mathbb{E}[X_n^2]$. We have $\mathbb{E}[T_n] \to \mathbb{E}[T]$ by the monotone convergence theorem and $\mathbb{E}[X_n^2] \to \mathbb{E}[X^2]$ by the dominated convergence theorem (X_n^2 are dominated by X^2 by (5.9)), so $\mathbb{E}[T] = \mathbb{E}[X^2] < \infty$. It suffices to choose T as our desired stopping time. ☐

6. Donsker's Invariance Principle

We arrive at our main theorem, Donsker's invariance principle. We will prove the theorem by showing that linearly interpolated and scaled random walks are close in norm to a particular sequence of Brownian motions, and that this implies convergence in distribution of the random walks to Brownian motion. We streamline the main proof by presenting three lemmas.

The first lemma says that it is possible to embed an entire random walk into Brownian motion, using the Skorokhod embedding theorem and strong Markov property of Brownian motion.

Lemma 6.1. *Let $\{S_n : n \in \mathbb{N}\}$ be a random walk defined by $S_n = \sum_{k=1}^n X_k$, where $\{X_k : k \in \mathbb{N}\}$ are independent and identically distributed random variables, distributed as some X with $\mathbb{E}[X] = 0$ and $\mathbb{E}[X^2] = 1$. Then, there exist stopping times $T_1, T_2, \dots$ with respect to the right-continuous natural filtration of a standard Brownian motion $\{B(t) : t \geq 0\}$ such that $\{B(T_n) : n \in \mathbb{N}\}$ has the distribution of the random walk.*

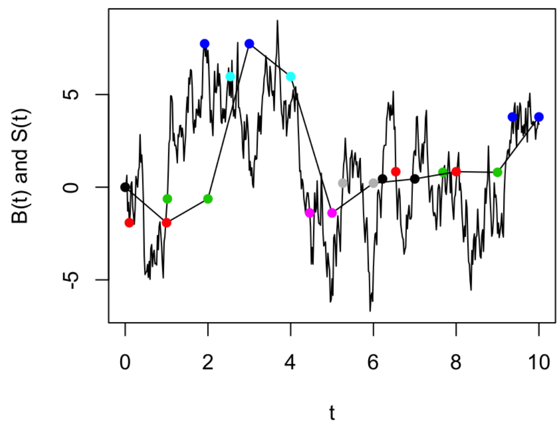

FIGURE 2. Lemma 6.1 says it is possible to embed a random walk into a Brownian motion.

Proof. By Skorokhod embedding, we can find T_1 such that $B(T_1) = X$ in distribution and $\mathbb{E}[T_1] = 1$. Define a new standard Brownian motion $\{B^{(2)}(t) : t \geq 0\}$ as the standardized Brownian motion restarted at T_1, i.e. let

$$B^{(2)}(t) = B(T_1 + t) - B(T_1).$$

The strong Markov property of $\{B(t) : t \geq 0\}$ says that $\{B^{(2)}(t) : t \geq 0\}$ is independent of both $\{B(t) : t \in [0, T_1]\}$ and $(T_1, B(T_1))$, so we can use Skorokhod embedding again to find T_2' such that $B^{(2)}(T_2') = X$ in distribution and $\mathbb{E}[T_2'] = 1$.

We need to show that $T_2 = T_1 + T_2'$ is a stopping time with respect to the right-continuous natural filtration of $\{B(t) : t \geq 0\}$, which we notate as $(\mathcal{F}^+(t) : t \geq 0)$. Write

$$\begin{aligned}\{T_1 + T_2' \leq t\} &= \bigcap_{n \in \mathbb{N}} \{T_1 + T_2' < t + \frac{1}{n}\} \\ &= \bigcap_{n \in \mathbb{N}} \bigcup_{r \in \mathbb{Q} \cap (0, t + \frac{1}{n})} \{T_1 < r\}\{T_2' < t + \frac{1}{n} - r\}.\end{aligned}$$

While $\{T_1 < r\}$ is not necessarily measurable with respect to $\mathcal{F}^0(t)$, where $(\mathcal{F}^0(t) : t \geq 0)$ is $\{B(t) : t \geq 0\}$'s natural filtration, it is measurable with respect to $\mathcal{F}^+(t)$. For $r < t + \frac{1}{n}$, the event $\{T_2' < t + \frac{1}{n} - r\}$ is also $\mathcal{F}^+(t)$-measurable, because the natural right-continuous filtration of $\{B(t) : t \geq 0\}$ contains the natural right-continuous filtration of $\{B^{(2)}(t) : t \geq 0\}$). As a result, $\{T_1 + T_2' \leq t\}$ is $\mathcal{F}^+(t)$-measurable and T_2 is a stopping time.

After finding T_2, define $\{B^{(3)}(t) : t \geq 0\}$ by

$$B^{(3)}(t) = B(T_2 + t) - B(T_2),$$

and use Skorokhod embedding to find $T_3 = T_1 + T_2' + T_3'$. Continuing to repeat this process defines stopping times $T_1, T_2, \ldots$ for which

$$B(T_n) = B(T_1 + \sum_{k=2}^{n} T_k') = B(T_1) + \sum_{k=2}^{n} B^{(k)}(T_k') \stackrel{d.}{=} \sum_{i=1}^{n} X_i,$$

where the last equality is justified by Brownian motion's strong Markov property. Note that

$$\mathbb{E}[T_n] = \mathbb{E}[T_1 + \sum_{k=2}^{n} T_k'] = n < \infty,$$

which completes our embedding. □

The next lemma gives a condition resembling convergence in probability that is sufficient to show that two sequences converge to the same distribution.

Lemma 6.2. *Let $A_1, A_2, \ldots$ and $B_1, B_2, \ldots$ be random objects with each pair (A_n, B_n) defined on the same probability space, and with all A_n and B_n taking values in a Polish normed space $(S, ||\cdot||)$. If*

$$\mathbb{P}(||A_n - B_n|| > \epsilon) \to 0,$$

and B_n is identically distributed with the distribution of B, then A_n converges in distribution to B.

Proof. We can apply the Portmanteau theorem since S is a Polish space. We will show that $A_n \xrightarrow{d.} B$ by checking that for every closed subset $F \in S$,

$$\limsup_{n\to\infty} \mathbb{P}(A_n \in F) \leq \mathbb{P}(B \in F).$$

Choose any closed set $F \in S$ and let $F[\epsilon]$ enlarge F by ϵ distance, so

$$F[\epsilon] = \{x \in S : ||x - y|| \leq \epsilon \text{ for some } y \in F\}.$$

If the random object A_n is in F, then either B_n is in $F[\epsilon]$ or A_n and B_n are greater than ϵ apart, so

$$\mathbb{P}(A_n \in F) \leq \mathbb{P}(B_n \in F[\epsilon]) + \mathbb{P}(||A_n - B_n|| > \epsilon).$$

In the limit, because $\mathbb{P}(B_n \in F[\epsilon])$ does not depend on n,

$$\limsup_{n\to\infty} \mathbb{P}(A_n \in F) \leq \mathbb{P}(B \in F[\epsilon]).$$

Letting $\epsilon \to 0$ gives the desired result by continuity of measures:

$$\lim_{\epsilon\to 0} \mathbb{P}(B \in F[\epsilon]) = \mathbb{P}(B \in \bigcap_{\epsilon>0} F[\epsilon]) = \mathbb{P}(B \in F).$$

□

Finally, we prove an arithmetic statement in our third lemma.

Lemma 6.3. *Let a_n be a sequence in $\mathbb{R}$ satisfying $\frac{a_n}{n} \to 1$. Then, the sequence also satisfies*

$$\sup_{k\in[0,n]} \left| \frac{a_k - k}{n} \right| \underset{n\to\infty}{\to} 0.$$

Proof. We want to find N such that for any fixed $n \geq N$, every $k \in [0, n]$ satisfies

$$\left| \frac{a_k - k}{n} \right| < \epsilon. \tag{6.4}$$

Fix n. Because $\frac{a_k}{k} \to 1$, there exists N' such that

$$\left| \frac{a_k - k}{k} \right| < \epsilon \text{ for all } k \geq N',$$

and (6.4) holds for $n \geq k \geq N'$.

Since there are only finitely many $k < N'$, we can let $N'' = \max_{k<N'} |a_k - k|/\epsilon$, and (6.4) holds for all $n \geq N''$ and $k < N'$. Letting $N = N' \vee N''$ proves the result. □

Theorem 6.5. *(Donsker's invariance principle) Let $\{S_n : n \in \mathbb{N}\}$ be a random walk defined by $S_n = \sum_{k=1}^n X_k$, where $\{X_k : k \in \mathbb{N}\}$ are independent and identically distributed random variables, distributed as some X with $\mathbb{E}[X] = 0$ and $\mathbb{E}[X^2] = 1$.*

Let $S(t) \in \mathbf{C}[0, \infty)$ be the random function that linearly interpolates the points of $\{S_n : n \in \mathbb{N}\}$, so

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

Define a sequence of random functions in $\mathbf{C}[0, 1]$ by

$$S_n^*(t) = \frac{S(nt)}{\sqrt{n}} \text{ for } t \in [0, 1].$$

Then, $S_n^*(t)$ converges in distribution to a standard Brownian motion $\{B(t) : t \in [0, 1]\}$ on the space $\mathbf{C}[0, 1]$ of continuous functions on the unit interval with metric induced by the sup-norm.

Proof. Using Lemma 6.1, embed the random walk $\{S_n : n \in \mathbb{N}\}$ into a standard Brownian motion $\{B(t) : t \geq 0\}$. Let $W_1, W_2, \ldots$ be Brownian motions defined by rescaling $\{B(t) : t \geq 0\}$ according to

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

By Lemma 6.2, we are done if we show that

$$\mathbb{P}(||S_n^* - W_n(t)||_{sup} > \epsilon) \to 0.$$

This is equivalent to setting A_n as the event that $|S_n^*(t) - W_n(t)| > \epsilon$ for some $t \in [0, 1)$, and showing that $\mathbb{P}(A_n) \to 0$.

Let $k = \lceil nt \rceil$ be the unique integer so that k and $k-1$ are the closest integers to nt and $k - 1 < nt \leq k$. Then, $S(t)$ is on the line connecting S_{k-1} and S_k, and S_n^* is linear between $S_n^*(\frac{k-1}{n}) = S_{k-1}$ and $S_n^*(\frac{k}{n}) = S_k$. As a result, we have

$$\begin{aligned}
A_n \subset A_n^* &\coloneqq \{\exists\, t \in [0, 1) : \left| \frac{S_k}{\sqrt{n}} - W_n(t) \right| > \epsilon\} \\
&\qquad \cup \{\exists\, t \in [0, 1) : \left| \frac{S_{k-1}}{\sqrt{n}} - W_n(t) \right| > \epsilon\} \\
&= \{\exists\, t \in [0, 1) : \left| W_n(\frac{T_k}{n}) - W_n(t) \right| > \epsilon\} \\
&\qquad \cup \{\exists\, t \in [0, 1) : \left| W_n(\frac{T_{k-1}}{n}) - W_n(t) \right| > \epsilon\},
\end{aligned}$$

where the second line follows because $S_k = B(T_k) = \sqrt{n} W_n(\frac{T_k}{n})$. We see that A_n^* is the event that there is some $t \in [0,1)$ for which W_n's value at t is far from its value at both $\frac{T_{k-1}}{n}$ and $\frac{T_k}{n}$.

Next, fix $\delta \in (0,1)$. Either $\frac{T_{k-1}}{n}$ and $\frac{T_k}{n}$ are both at least δ away from t, or they are not, and there is an element less than δ away from t with value far from W_n's value at t. In other words,

$$
\begin{aligned}
A_n^* \subset \{\exists\, s,t \in [0,2] : |s-t| < \delta,\ |W_n(s) - W_n(t)| > \epsilon\} \\
\cup \{\exists\, t \in [0,1) : \big|\frac{T_k}{n} - t\big| \vee \big|\frac{T_{k-1}}{n} - t\big| \geq \delta\}.
\end{aligned} \tag{6.6}
$$

The first event in (6.6) does not depend on n. Because W_n is a Brownian motion and uniformly continuous on $[0,2]$, we can make the probability of this event as small as desired by choosing small δ. (We cannot guarantee that this probability will be 0.)

For the second event, note that by construction, $(T_n - T_{n-1} : n \in \mathbb{N})$ are independent and identically distributed random variables with mean 1. Thus, by the strong law of large numbers,

$$\frac{T_n}{n} = \frac{\sum_{k=1}^n (T_k - T_{k-1})}{n} \to 1 \text{ almost surely.}$$

By Lemma 6.3, this implies

$$
\begin{aligned}
&\sup_{k\in[0,n]} \left|\frac{T_k - k}{n}\right| \to 0 \text{ almost surely} \\
&\Rightarrow\ \mathbb{P}\Big(\bigcup_{N\in\mathbb{N}} \bigcap_{n\geq N} \big\{ \sup_{k\in[0,n]} \left|\frac{T_k - k}{n}\right| < \delta \big\}\Big) = 1 \\
&\Leftrightarrow\ \mathbb{P}\Big(\bigcap_{n\geq N} \big\{ \sup_{k\in[0,n]} \left|\frac{T_k - k}{n}\right| < \delta \big\}\Big) \underset{N\to\infty}{\to} 1.
\end{aligned}
$$

The probability of an event is greater than its intersection with other events, so

$$
\begin{aligned}
&\mathbb{P}\Big(\sup_{k\in[0,n]} \left|\frac{T_k - k}{n}\right| < \delta \Big) \underset{n\to\infty}{\to} 1 \\
&\Leftrightarrow\ \mathbb{P}\Big(\sup_{k\in[0,n]} \left|\frac{T_k - k}{n}\right| \geq \delta \Big) \underset{n\to\infty}{\to} 0.
\end{aligned} \tag{6.7}
$$

Finally, recall $t \in [\frac{k-1}{n}, \frac{k}{n})$ and let $n > 2/\delta$. Then the second event in (6.6) has probability

$$
\begin{aligned}
&\mathbb{P}\big(\exists\, t \in [0,1) : \big|\frac{T_k}{n} - t\big| \vee \big|\frac{T_{k-1}}{n} - t\big| \geq \delta\big) \\
&\quad \leq \mathbb{P}\Big(\sup_{k\in[1,n]} \frac{(T_k - (k-1)) \vee (k - T_{k-1})}{n} \geq \delta \Big) \\
&\quad \leq \mathbb{P}\Big(\sup_{k\in[1,n]} \frac{T_k - k}{n} \geq \delta/2 \Big) + \mathbb{P}\Big(\sup_{k\in[1,n]} \frac{(k-1) - T_{k-1}}{n} \geq \delta/2 \Big),
\end{aligned}
$$

and both terms of the final sum converge to 0 as $n \to \infty$ by (6.7).

Thus, for large n, we can make the probability of A_n arbitrarily small by choosing sufficiently small δ, proving $\mathbb{P}(A_n) \to 0$ and the theorem. □

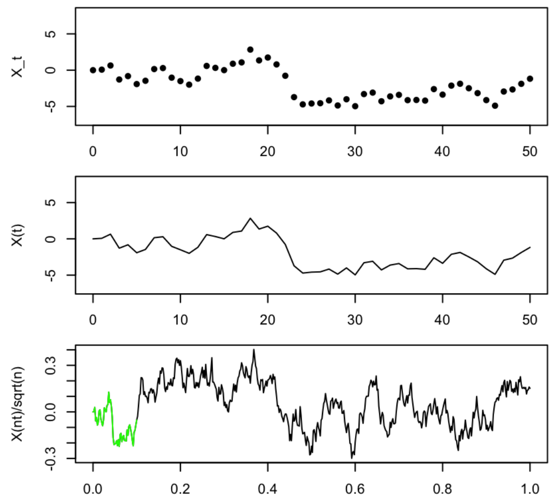

Figure 3. Donsker's invariance principle says that a random walk (first panel) that is linearly interpolated (second panel) and properly rescaled (third panel) has distribution converging to Brownian motion.

Acknowledgments

First and foremost, I would like to thank Jin Woo Sung for being an excellent mentor and for his time and effort in helping me work through the material presented above. I also owe thanks to the REU speakers in the Probability and Analysis subgroups for introducing me to new topics, and to Greg Lawler for holding weekly office hours that were a treat to attend. Finally, I am grateful to Peter May for organizing the REU and supporting the students by attending and reading all of our presentations and papers.

References

[1] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov Chains and Mixing Times.* American Mathematical Society, 2008.
[2] Peter Morters and Yuval Peres. *Brownian Motion.* Cambridge University Press, 2010.
[3] L. C. G. Rogers and David Williams. *Diffusions, Markov Processes and Martingales, Vol 1.* Cambridge University Press, 2000.
[4] Stewart N. Ethier and Thomas G. Kurtz. *Markov Processes: Characterization and Convergence.* Willey Series in Probability and Mathematical Statistics, 1986.
[5] David Williams. *Probability with Martingales.* Cambridge University Press, 1991.
[6] David Gamarnik. "Martingale Convergence Theorem." https://ocw.mit.edu/courses/sloan-school-of-management/15-070j-advanced-stochastic-processes-fall-2013/lecture-notes/MIT15_070JF13_Lec11Add.pdf