BROWNIAN MOTION: FROM MEASURE THEORY TO QUANTITATIVE BIOLOGY

EMMA ZHANG

Abstract. This paper provides an accessible overview of Brownian motion, a popular stochastic process. It was first observed in 1827 via pollen suspended in water, and has been used widely used ever since, throughout physics, chemistry, economics, and biology. Starting in measure and probability theory, we construct Brownian motion from random walks, prove Markov and Martingale Properties (including the Optional Stopping Theorem), and end with an exploration of Brownian motion's role in comparative quantitative biology, more specifically in modelling random processes of evolution.

Introduction

Brownian motion was first observed in 1827 by biologist Robert Brown while viewing pollen in water under a microscope. He noticed that particles of the pollen grains moved erratically in random directions, as if alive. Then, Brown also found this quality quality in nonliving matter like dust, as Roman poet and philosopher Lucretius describes in his scientific poem "On the Nature of Things" (50 B.C.E): "Observe what happens when sunbeams are admitted into a building and shed light on its shadowy places. You will see a multitude of tiny particles mingling in a multitude of ways... their dancing is an actual indication of underlying movements of matter that are hidden from our sight." Although he studied it for years, Brown never explained the motion, although it started a centuries-long quest to understand and quantify the physical process that became named after him.

Einstein, in his 1905 paper "On the Movement of Small Particles Suspended in Stationary Liquids Required by the Molecular-Kinetic of Heat," finally solved this mystery, attributing the movement to the bombardment of particles in the liquid. He also connected this distribution to the partial differential heat equation with the solution of normal distribution with expectation 0 and variance t. Later, it was officially formalized by Norbert Wiener as a time-continuous stochastic process, and has since became one of the most important processes across disciplines, spanning from physics to finance to biology (which makes its appearance at the end of this paper).

Thus, this paper of Brownian motion, just like it's discovery, spans from the abstract to the applied. To deeply understand Brownian motion, we must first examine measure theory, which seeks to define and analyze notions of size, length, and volume in abstract mathematical spaces. This contains the probability space, defined by the probability measure. From there, random variables are functions that map possible outcomes onto the real number line $\mathbb{R}$. We will construct Brownian motion as a limit of these interpolated random walks, and from there, meditate on various properties by equating Brownian motion to Markov and Martingale processes. With the help of filtrations, we can more closely look at stopping times and hitting times. This includes figuring out if and when a certain value will be reached, as well as the proof of the Optional Stopping

Date: July 15, 2024.

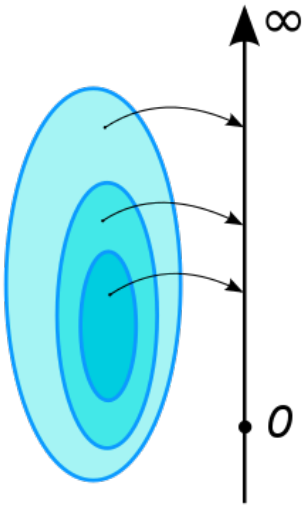

Figure 1. Measures map sets, such as the sample space for all outcomes in a probabilistic experiment, onto $[0, \infty]$
, each subset a σ-algebra Credit: [[Wik24]]

Theorem, which shows that a stopping time has the same expected value as that of the starting time. Lastly, a sample of Brownian motion's powers in modeling quantitative genetics lets us bridge the micro with the macro, math with evolution, and show matrices in which covariance is applied to the branches shared between species.

The fact that a stochastic process constructed from symmetrical random walks can encapsulate so much says a lot about the nature of movement. From subsets of possibilities to heat diffusion to the fairness of games, the study of Brownian motion, like every mathematical theory, helps us not only further the understanding of our world, but also the way we've chosen to conceptualize it. We've defined measures to help us measure the seemingly uncountable, created "flavors" of processes (Markov, Martingale, etc) detailing trajectories across time, and even imagined realizations of Brownian motion as many, maybe infinite, functions all inside a sample space. Although mainly a math paper, this paper seeks to cover the wonder, applications, and embedded ways of thinking in Brownian motion just as eagerly as the math.

Basic Brownian motion sections, and much of the Markov chain and Martingale section, follows Chapter 8 of [[Dob16]], while the biological modeling section references [[Har19]]. Upon finishing this paper, Liu's expository paper [[LIU21]] stems from many of the same ideas covered here, and very eager readers can look towards [[MP10]] for a more in depth textbook on Brownian motion. In addition, although most relevant preliminary concepts will be outlined in this paper, further reading on measure theory and probability will serve as a more solid basis to understand technical definitions and proofs.

1. Measure Theory and Probability

Measure theory assigns abstract mathematical spaces generalized sizes, lengths, and volumes, oftentimes by creating sets and subsets that can be described in terms of their intersections and relationships. Probability relies on measure theory to make sense of its outcomes, abstractly and translated onto the real number line.

1.1. **Measure Spaces.**

Definition 1.1. Let a measurable space be (X, A, μ) with X being a set (in probability, this is the sample space of all outcomes, Ω) and A a special collection of subsets of X, also known as a σ-algebra. A map μ is defined on A and maps the sample space onto $[0, \infty]$. Such a map is a **measure**, as shown in Figure 1. We want a measure to, well, measure the set X, which would mean giving it a generalized length (2d) or generalized volume (3d). This length must be positive or 0, and specific restrictions might hold depending on the measure. Well known measures include the counting measure, the probability measure, and the Lebesgue measure. A measure must follow these conditions:

(1) $\mu(\varnothing) = 0$ (the volume of an empty set is 0).
(2) $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$ (volume of the union of σ-algebras is equal to the volume of individual sigma algebras added together). This property is known as σ-additivity. Outer measures and inner measures have different requirements for this property, so often, they are not measures! That is beyond the scope of this paper, but is interesting to look into.

For Brownian motion, we care about the probability space, which is measured by the probability measure.

1.2. **Probability Spaces.**

Definition 1.2. In probability theory, a **probability space** is a measure space used to define random processes. It looks like $(\Omega, A, \mathbb{P})$, and consists of three elements:

(1) Ω: the **sample space** of all possible outcomes.
(2) A: a **sigma-algebra** (we will define this later), which is a collection of subsets of Ω. Each set is called an event.
(3) $\mathbb{P}$: **probability measure** that maps events onto their probability values from 0 to 1, with 0 being impossible [1] and 1 being certain. Because all events are "cut out" of Ω, $\mathbb{P}(\Omega) = 1$.

We need to divide up Ω into sets because when sample space is uncountably infinite, such as when drawing from continuous outcome spaces, the probability of any discrete point is zero since there are infinitely many discrete points. Instead, we must find the probability of an interval, such as, results from 2 to 5, instead of a discrete point like 3.82.

There is where **sigma-algebras (σ-algebras)** come in, to make the set of all outcomes more measurable by partitioning Ω into intervals, creating the new domains on which to calculate probability.

Definition 1.3. Each σ-algebra is closed under

(1) Complements (A^C) , or everything else in the sample space except the set.
(2) Unions ($A \cup B$), "A or "B" both happen
(3) Intersections ($A \cap B$), "A and B happens"

This allows us to create sets within Ω that hold up to basic properties we can work with: if one event happens, there will also be a defined probability where it doesn't happen (even if that's 0); if two events happen in the same space, we can also think about either event happening or both events happening.

The Borel σ-algebra is a specific type of σ-algebra:

[1] *3Blue1Brown has a video on why probability 0 doesn't necessarily mean impossible, check it out here

Definition 1.4. A **Borel σ-algebra** is a σ-algebra generated on open sets (topologies or distances) of $\mathbb{R}$. An open set is defined by: A is an open set if for every $x \in A$, a ball $B(x, r)$ with center x and radius r is contained in A.

Borel σ-algebras aid in the translation of a σ-algebra (set of probabilities) to a random variable on the real line (figure 2). For a random variable X to be measurable, the pre-image of every Borel set must be an event in the underlying probability space σ-algebra corresponding to the probabilities in $\mathbb{P}$.

1.3. **Generating Sigma Algebras.** To generate σ-algebras, first we must recognize that the intersection of sigma algebras are sigma algebras:

$$\{A_i\}_{i\in I} \to \bigcup_{i\in I} A_i \text{ is a } \sigma\text{-algebra}$$

because if the σ-algebra and their complements are present in all the intersecting sets, they also are in the intersection. Similarly, intersections and unions hold as well. Then, we can also find the "smallest" σ-algebra generated by a collection of sets:

Definition 1.5. The smallest σ-algebra **generated** by a collection of sets is created by taking the intersection of every σ-algebra containing that collection.

1.4. **Random Variables.**

Definition 1.6. A **random variable** on the probability space $(\Omega, A, \mathbb{P})$ is an σ-measurable function from the set of all possible outcomes to the set of real numbers: $X : \Omega \to \mathbb{R}$. This is shown in Figure **??**. For every Borel set $B \in \mathcal{B}(\mathbb{R})$, the pre-image of B refers to the function ω in Ω. The outcomes of ω is a σ-algebra of Ω, and hence, it is σ-measurable:

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

Since Borel sets occur on real open intervals, this just shows the mapping of a random function from a probability sample space (Ω) to the real line and how it still holds σ-measurable properties as it is made from σ-algebras.

1.5. **Probability Preliminaries.**

Definition 1.7. An **expectation** in probability is a weighted mean denoted by $E(X)$. For a list of numbers $X_1, X_2, X_3 \ldots$, the expectation is defined by

$$\mathbb{E}(X) = \sum_{i=1}^{n} X_i \mathbb{P}_i \tag{1.1}$$

For continuous random variables like $W(t)$ in Equation 2.4,

$$\mathbb{E}(X) = \int_{\Omega} X d\mathbb{P} \tag{1.2}$$

The bounds Ω being the set of all functions in Ω from $(-\infty, \infty)$.

Definition 1.8. The **variance** is a measure of how much the random variable varies around the expectation:

$$Var(X) = E((X - E(X))^2), \tag{1.3}$$

denoted as σ^2.

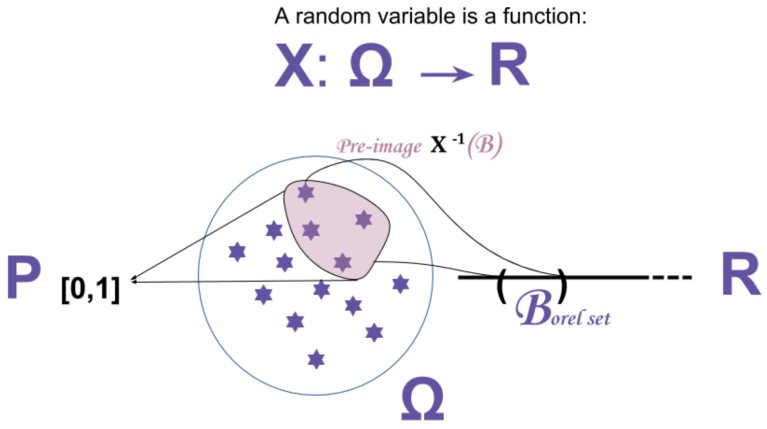

Figure 2. shows random variables as functions that map $\Omega \to \mathbb{R}$ (Credit: rinterested)

Definition 1.9. The **covariance** is the measure of the relationship between two random variables and the extent to which they change together:

$$Cov(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]. \tag{1.4}$$

Definition 1.10. A **normal distribution** or **Gaussian distribution** is a continuous probability for a real valued random variable, typically shown a bell curve. The probability density function is as follows:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{1.5}$$

2. Brownian Motion

Definition 2.1. A stochastic process X_t is a collection of random variables indexed by time. In measure theory terms, it is a function of $t \in [0, \infty) \times \Omega \to \mathbb{R}$, random variables that "realize" the probability space into Borel sets on the real line.

Definition 2.2. A continuous time stochastic process $(X_t)_{t \geq 0} = \{X_t, 0t < \infty\}$ is an uncountable collection of random variables indexed by the non-negative real numbers.

Brownian Motion is a continuous-time, continuous-state stochastic process defined on $\mathbb{R}^{(0,\infty)}$, where random values can fall between $[0, \infty)$. Brownian motion can exist on multiple dimensions, but we will focus on the one-dimensional version.

The most common form of Brownian motion is Standard Brownian motion.

Definition 2.3. A stochastic process $(B_t)_{t \geq 0}$ is a Standard Brownian Motion if it satisfies the following properties:

(1) $B_0 = 0$ (with probability 1). Brownian Motion starts at 0 when $t = 0$.
(2) **Independent Increments:** If $0 \leq q < r \leq s < t$, then $B_t - B_s$ is independent of $B_r - B_q$.

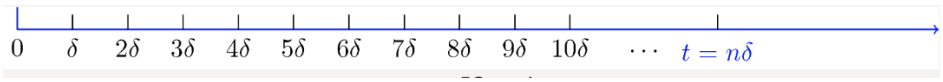

Figure 3. cite or recreate

(3) **Stationary Increments of Normal Distribution:** Each interval of $B_{(t+s)} - B_t$, given that $s < t$, is normally distributed with expectation 0 and variance s shown by $\sim N(0, s)$, and independent of starting time t.
(4) **Continuous paths**: $t \mapsto B(t)$ is continuous, with probability 1.

Brownian motion is often seen as the "functional version" of the **central limit theorem** (which states that the distribution of sample means approximates normal distribution as the sample size gets larger, regardless of the population's distribution) as it is a simple representation of random variables converging to normal distribution, giving us distribution of the third property.

Definition 2.4. (Central Limit Theorem) Let $X_1, X_2, X_3, \ldots X_n$ be a sample of n independent random variables with expectation μ and variance σ^2. As limit $n \to \infty$, the limit of the distribution converges to the normal distribution with expectation 0 and variance 1:

$$(\overline{X} - \mu)/\sigma_n \to Z$$

or

$$(\overline{X} - \mu)/\sigma_n \sim \mathcal{N}(0, 1).$$

More on the central limit theorem, including proofs, can be found here.

2.1. **Brownian Motion as a Limit of Random Walks.** Brownian motion is, at its essence, a process constructed from random walks (or at least the limit of random walks). A random walk is a stochastic process formed by successive summation of independent, identically distributed random variables (i.i.d.s).
Examples of random walks include flipping a coin or the drunkard's walk. We will prove Brownian motion's construction using random walks by following Chapter 11.4.1 of [[PN14]].

Theorem 2.5. *Brownian motion exists, and satisfies the above conditions.*

Proof. Let us construct Brownian Motion from a symmetric random walk. We can divide the real line $[0, \infty)$ into tiny intervals of length δ. Each sub-interval is a time slot of length δ. These intervals are $[0, \delta), [\delta, 2\delta), [2\delta, 3\delta) \ldots [(k-1\delta), k\delta)$ for $k < \infty$. We define: in each time slot, we toss a fair coin. We define the random variables X_i as follows:

$$X_i = \begin{cases} +\sqrt{\delta}, & \text{with probability } 1/2, \\ -\sqrt{\delta}, & \text{with probability } 1/2, \end{cases} \tag{2.1}$$

where X_is are independent (i.i.d's). Square roots are used instead of 1 and -1 as some constructions do so we do not have to scale the random walks later. Note that:

$$E(X_i) = 0 \text{ due to being symmetrical} \tag{2.2}$$

$$Var(X_i) = \delta \tag{2.3}$$

Then, we can define the process $W(t)$ where $W(0) = 0$. This satisfies the first condition of Brownian motion, $B(0) = 0$. We define t as the sum of the n amount of δ's up to that point:

$\sum_{i=1}^{n} \delta_i$ shortened as $t = n\delta$. $W(t)$ is given by:

$$W(t) = W(n\delta) = \sum_{i=1}^{n} X_i. \tag{2.4}$$

Since $W(t)$ is the sum of n i.i.d. variables, the expectation and variance can be found by:

$$E(W(t)) = \sum_{i=1}^{n} E(X_i) \tag{2.5}$$

$$= 0, \tag{2.6}$$

$$Var(W(t)) = \sum_{i=1}^{n} Var(X_i) \tag{2.7}$$

$$= nVar(X_1) \tag{2.8}$$

$$= n\delta \tag{2.9}$$

$$= t. \tag{2.10}$$

Since coin tosses are independent, $W(t)$ has independent increments such that for $0 \leq t_1 < t_2 < t_3 \ldots < t_n$,

$$W(t_2) - W(t_1), W(t_3) - W(t_2), \ldots W(t_n) - W(t_{n-1}) \tag{2.11}$$

are independent, proving condition two. For $t \in (0, \infty)$, as n goes to infinity, δ goes to zero, meaning each interval is condensing. By the Central Limit Theorem (Definition 2.5), $W(t)$ will become a normal random variable:

$$W(t) \sim \mathcal{N}(0, t). \tag{2.12}$$

Additionally, stationary increments means that each interval of $W(t)$ must only depend on the length of the interval, and not the exact location of the interval on the line: the distribution of $W(t_2) - W(t_1)$ must be equal to $W(t_2 + s) - W(t_1 + s)$. We will prove this below. For $0 \leq t_1 < t_2$, let $t_1 = n_1\delta$ and $t_2 = n_2\delta$, we see that

$$W(t_1) = \sum_{i=1}^{n_1} X_i, \tag{2.13}$$

$$W(t_2) = \sum_{i=1}^{n_2} X_i. \tag{2.14}$$

Then, to find the interval in between,

$$W(t_1) - W(t_2) = \sum_{i=1+n_1}^{n_2} X_i. \tag{2.15}$$

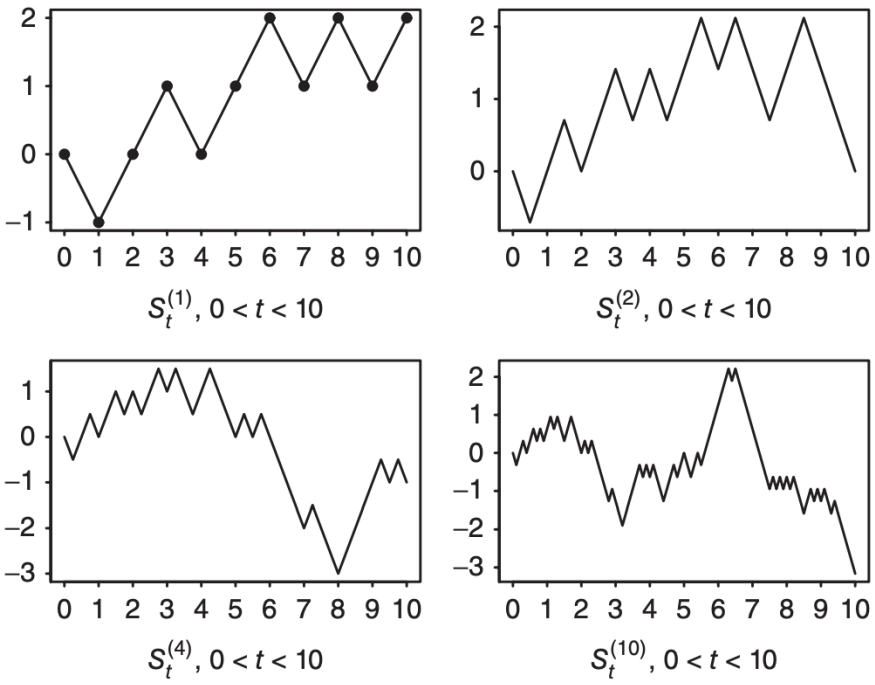

Figure 4. Brownian motion is constructed as n approaches infinity

Following the same process as before, we can naturally find the expectation and variance of this interval:

$$E(W(t_1) - W(t_2)) = E(\sum_{1+n_1}^{n_2} X_i) \tag{2.16}$$

$$= 0 \tag{2.17}$$

$$Var(W(t_1) - W(t_2)) = Var(\sum_{1+n_1}^{n_2} X_i) \tag{2.18}$$

$$= (n_2 - n_1)Var(X_1) \tag{2.19}$$

$$= (n_2 - n_1)\delta \tag{2.20}$$

$$= t_2 - t_1 \tag{2.21}$$

Hence, $W(t_2) - W(t_1)$ converges to $\mathcal{N}(0, t_2 - t_1)$. This satisfies the third condition of Brownian motion, that each interval is stationary and normally distributed with expectation 0 and variance t, and the distribution is only affected by the length, $(n_2 - n_1)\delta$ or $t_2 - t_1$, of the interval. See Remark 2.7 for notes on continuity.

■

Proof. (Levy's Construction): The most well known construction of Brownian motion is Levy's construction, which proves the existence of Brownian motion by linearly-interpolating (connecting by straight lines) values of a function $B^{(n)}(\omega)$ of random variables at times $t = \frac{k}{2n} : 0, ..., 2n$. Using Borel-Cantelli lemma, it is shown that

$$\mathbb{P}(B^{(n)}(\omega) \text{ converges uniformly}) = 1 \tag{2.22}$$

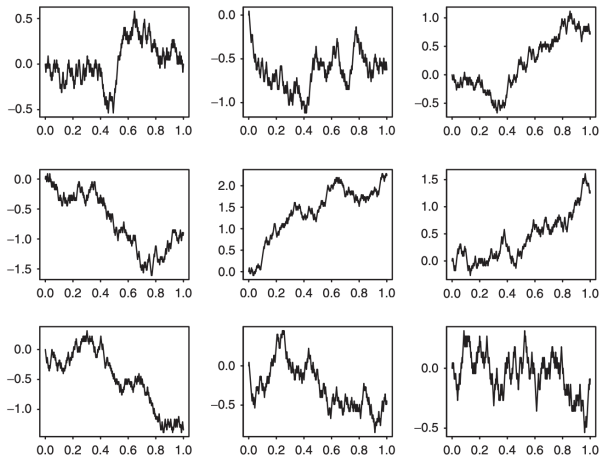

Figure 8.2 Sample paths of Brownian motion on [0, 1].

Figure 5. Realizations of Brownian Motion (Dobrow 2016)

Therefore, the process is continuous, and increments are independent and stationary. A more rigorous proof can be found (include citation). ■

2.2. **Sample Space of Brownian Motion.** Let Brownian motion be defined on the probability space $(\Omega, A, \mathbb{P})$.

Remark 2.6. The sample set Ω is the set of all continuous functions $[0, \infty)$. The random variable is a function of ω in Ω where each ω is a continuous function: $X = X_\omega$. In a stochastic process, X is also a function of time t, where $X_t = X_t(\omega)$. (However, ω is usually suppressed for simplicity.) While $t \geq 0$, X is a random variable. As shown in Proof 2.1, Brownian motion is the limit of such random variables.
Letting $\omega \in \Omega$ vary generates different values of Brownian motion at a fixed time t. On the other hand, fixing ω creates a realization or random function $f(x) = B_\omega(t)$ as t varies. Therefore, the continuity requirement can be more precisely stated as:

$$\mathbb{P}(\omega \in \Omega : B_\omega(t) \text{ is a continuous function of t}) = 1$$

Figure 5 shows different such random paths of Brownian Motion.

3. Filtrations and Stopping Times

Next, we introduce a new way to approach the flow of information of Brownian motion. Filtrations are important because they help define a specific subset of information to focus on up to time t. Many theorems, including those concerning Markov and Martingale processes, do not apply unless we can partition the Brownian motion ω function up to a distinct time. For example, how are we going to know there is a time greater than 0 where the expected value at that time is equal to the starting expected value if we don't know where to stop to check (Theorem 5.3)?

In an intuitive way, we can imagine that we are viewing Brownian motion unfold in real time—we know the events of the past and not the events of the future, so we can only use what we know right now (which includes what we learned previously) to see if a certain condition is met (this condition is a stopping time, as we will discuss later in this section). To connect this back to evolutionary biology, a filtration can be thought of as going from

one generation to the next. If we are at our generation, we know the traits we have, and the traits our parents, grandparents, etc have, but we do not know what traits our children or grandchildren will have, even if it must all be contained in the set of alleles and their possibilities (Ω).

3.0.1. *Filtrations.* In measure theory terms, a filtration is a sequence of sub-σ-algebra inside A. As t increases, the sub-σ-algebras "expand," allowing us to know an increasing amount of information. It's like creating more and more space to be able to ask: "Has this happened yet? Can we know this from the information leading up to here?" "What about now?"

Definition 3.1. Let $(\Omega, A, \mathbb{P})$ be a probability space, and I as an index set. For every $i \in I$, Let $\mathcal{F}_i$ be a sub-σ-algebra of A. Then a filtration is

$$F := (\mathcal{F}_i)_{i \in I}$$

where $\mathcal{F}_k \subseteq \mathcal{F}_l$ given that $k \leq l$. For example, if the index is the natural numbers, we can view the sequence of σ-algebras as

$$\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \mathcal{F}_3 \subseteq \mathcal{F}_4 \ldots \subseteq \mathcal{F}_i.$$

We know more and more about the process as t increases.

In a natural filtration, this same process is done but with random variables:

Definition 3.2. Let $X(t)_{(0,\infty]}$ be a stochastic process on the probability space $(\Omega, A, \mathbb{P})$ where $\sigma(X_s | s \leq t)$ denotes the σ-algebra created by the random variables X_i. Following this,

$$(\mathcal{F}_t^X)_{t \in [0,\infty)} := \sigma(X_s | s \leq t) \tag{3.1}$$

is a natural filtration of A.

Remark 3.3. This can also be defined as the σ-algebra

$$\mathcal{F}_t^X = A \in \mathcal{F} : A \cup \{T \leq t\} \in \{(\mathcal{F}_t^X)_{t \in [0,\infty)} \text{ for all } t \geq 0\} \tag{3.2}$$

which ensures all X_s is $\mathcal{F}(T)$-measurable.

3.0.2. *Hitting Times and Stopping Times.* Continuing our metaphor, we can think of a hitting time as the first time a certain characteristic a will appear in a certain lineage in the subset of all realizations of an evolutionary process $\omega \in \Omega$.

Definition 3.4. Let $a \in \mathbb{R}$ be a certain condition to be reached. Let any time $T_a = \{t : B_t = a\}$ be called a **hitting time**. Therefore, $T_a = min\{t : B_t = a\}$ is the **first hitting time**.

Definition 3.5. A random variable $T : \Omega \to [0, \infty) \cup \{\infty\}$ is called an $\mathcal{F}_T$**-stopping time** if $(T \leq t : t \geq 0) \in (\mathcal{F}_t^X)_{t \in [0,\infty)}$ for some $t \in [0, \infty)$. This can be written as:

$$\{T \leq t\} = \{\omega \in \Omega | T(\omega) \leq t\}$$

If $T \leq t$, it means that the realization of a function ω in the set of all possible paths Ω happens before t, making it a stopping time because whether or not it is a hitting time can be determined in the information available up to time t.

4. Markov and Martingale Properties

We can use these ideas of filtrations and stopping times to look at some interesting properties and characteristics of Brownian motion, helping us better understand the patterns of this symmetrical random walk through time. In particular, we can explore how how future states relate to the current state (oftentimes, moreso than the past), whether or not and how we can expect certain conditions or characteristics to be met, and even Brownian motion's relationship to other random walks (some of which can be embedded straight into Brownian motion!).

4.1. As A Markov Process.

A Markov process is a continuous-state stochastic process where the probability of an event is only dependent on the state attained in the previous event.

Definition 4.1. A Markov process $(X_t)_{t\geq 0}$ can be mathematically represented if

$$P((X_{t+s} \leq y)|X_u, 0 \leq u \leq s) = P((X_{t+s} \leq y)|X_s) \tag{4.1}$$

for all $s, t > 0$ and real y.

This means the probability of state X time t after s only depends on the state at X_s, not anything in between.

Additionally,

Definition 4.2. A Markov process $(X_t)_{t\geq 0}$ is time-homogeneous if the probability in Equation (3.1) does not depend on s, but instead, only the starting value, X_0:

$$P((X_{t+s} \leq y)|X_s) = P((X_t \leq y)|X_0) \tag{4.2}$$

Every stochastic process generates a natural filtration $(\mathcal{F}_t^x)_{t\in[0,\infty)}$, as discussed in the previous section. This filtration can be used to show that Brownian Motion is a Markov process that has the Markov property. But first, we must define a Markov process as a transition kernel, a continuous time and space extension of the discrete Markov chain.

Definition 4.3. A Markov transition kernel is a function $p : [0, \infty) \times \mathbb{R}^d \times \mathfrak{B} \to \mathbb{R}$, where $\mathfrak{B}$ is a Borel σ-algebra on $\mathbb{R}^d$ such that:

1. $p(t, x, \cdot)$ is a probability measure that takes in a σ-algebra $A \in \mathfrak{B}$ for all $t \in [0, \infty)$ and starting points $x \in \mathbb{R}^d$. We can integrate a function $f : \mathbb{R}^d \to \mathbb{R}$ with respect to this probability measure $\int f(y)p(t, x, dy)$. Integrating a function brings it onto the real number line, the measure serving as a benchmark to scale higher volumes into lengths.
2. $p(\cdot, \cdot, A)$ represents a measurable function in (t, x). This also means $p(t, \cdot, A)$ is a function in x that can be integrated with respect to any measure on $\mathbb{R}^d$.
3. Thus, for all $s, r \geq 0, x \in \mathbb{R}^d$, and $A \in \mathfrak{B}$,

$$p(s + r, x, A) = \int p(r, y, A)p(s, x, dy)$$

which combines statements 1 and 2. This represents a stochastic process $(X_t)_{t\geq 0}$ where

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

meaning the process at increment r in the future only depends on current, starting event x, and the chosen increment r, which satisfies the Markov property and the stationary increments condition of Brownian motion.

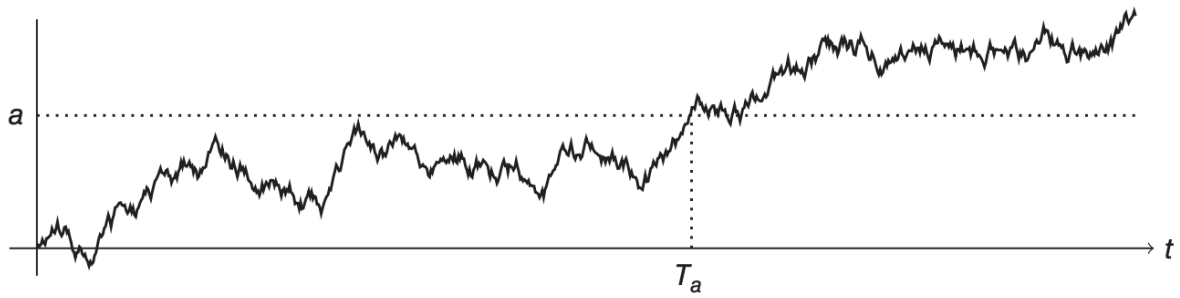

Figure 6. T_a is the first time Brownian motion hits a.

Definition 4.4. A stochastic process $(X_t)_{t\geq 0}$ is a time-continuous Markov process with respect to the filtration $(\mathcal{F}_t^x)_{t\geq 0}$ if

(1) it is adapted to $(\mathcal{F}_t^x)_{t\geq 0}$.
(2) satisfies the Markov property:

$$\mathbb{P}((X(s+r) : \mathcal{F}_t^s) = \mathbb{P}((X(s+r) : X(s)) = p(r, X(s), A)$$

showing how a partially realized process (adapted to filtration) depends on it's last state.

Theorem 4.5. *(Markov Property) Linear Brownian motion is a Markov process with respect to its natural filtration, shown as the transition kernel p defined by $p(t, x, \cdot) = N(x, t)$. It possesses the Markov property, which shows that $B_{t+s} - B_s : t \geq 0$ which starts at $x \in \mathbb{R}$ is independent of $B_t : t \in [0, s)$.*

Proof. From the independence of Brownian Motion's increments:

$$\mathbb{P}(B_{s+r} \in A : \mathcal{F}_s^x) = \mathbb{P}(B_{s+r} \in A : B_s) = \mathbb{P}(Z \in A)$$

where Z is a random variable with the distribution $N(B(s), r)$, which is "anchored" at the previous event with the variance of the length of the segment, adjacent to both the Markov property and the distribution of Brownian motion as $N(0, t)$ since the starting point of Brownian motion is $B_0 = 0$, and by the martingale property, the expectation is constant. Therefore, p is a valid transition kernel. We know p satisfies 1 and 2 of [def whatever] because the Normal distribution is a Borel probability measure since it is defined on the Borel σ-algebra (can assign probability to all Borel sets), measuring no outcomes is null, measuring all outcomes adds up to probability 1, and it is sigma additive. The third requirement is satisfied because the integral of convolution of the two normally distributed variables on the right of Definition 3.5 has the same normal distribution of the left. ■

Theorem 4.6. *A Brownian Motion $(B_t)_{t\geq 0}$ is a Markov process with regards to its natural filtration since the Markov property states $B_s - B_t$ is independent of $\mathcal{F}_t^x$ for all t.*

Proof. A detailed proof can be found in Theorem 3.8 of [LIU21] ■

The strong Markov property can be used to find the distribution of T_a, the hitting times.

The natural filtration $\mathcal{F}_t^x$ reveals information about the stochastic process up to t. As random variable T must be less than t, we can check if T_a occurred before T or not.

Theorem 4.7. *For standard Brownian motion, let T_a be the stopping time where process B_t hits level a. The density function of T_a is*

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

Proof. Consider standard Brownian motion, such as the one we constructed from random walks in section 2.1. There is a 1/2 chance B_t will be above and a 1/2 chance it will be below the line $y = 0$ at any given t, because each interval has a 1/2 chance of going up and a 1/2 chance of going down and $B_0 = 0$. Similarly, if we start at a, B_t is also equally likely to be above or below a. This gives:

$$\mathbb{P}(B_t > a | T_a < t) = \mathbb{P}(B_t > 0) = \frac{1}{2}$$

and therefore,

$$\mathbb{P}(B_t > a | T_a < t) = \frac{\mathbb{P}(B_t > a, T_a < t)}{\mathbb{P}(T_a < t)} = \frac{\mathbb{P}(B_t > a)}{\mathbb{P}(T_a < t)}$$

the second equality by conditional probability and the third because of continuity of simple paths. Namely, if $B_t > a$, T_a has to be reached before t since a continuous path cannot "skipped over" T_a. Then, we have

$$\mathbb{P}(T_a < t) = 2\mathbb{P}(B_t > a)$$

We multiply the probability by two because of Brownian motion's symmetric property: $B_t < -a$ would be the same thing as $B_t > a$. Next,

(4.3)

$$= 2 \int_a^\infty \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x^2)}{2t}} dx \tag{4.4}$$

$$= 2 \int_{a/\sqrt{t}}^\infty \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x^2)}{2}} dx \tag{4.5}$$

$$= 2 \int_{|a|/\sqrt{t}}^\infty \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x^2)}{2}} dx. \tag{4.6}$$

the absolute value again for for the symmetry. Integrating with respect to t will give the probability density function of the first hitting time:

$$f_{T_a}(t) = \frac{|a|}{\sqrt{2\pi t^3}} e^{-\frac{(a^2)}{2}}. \tag{4.7}$$

■

Following this, we can begin to look at some very strange properties of Brownian motion hitting times.

Theorem 4.8. *For a level y=a, $a < \infty$, $\mathbb{P}(T_a < \infty) = 1$.*

Proof.

$$\mathbb{P}(T_a < \infty) = \lim_{t \to \infty} (T_a < t) = \lim 2 \int_{|a|/\sqrt{t}}^\infty \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x^2)}{2}} dt = 2(\frac{1}{2}) = 1. \tag{4.8}$$

■

This means Brownian motion is guaranteed to hit level a, for all a, no matter how large. However,

Theorem 4.9. *For any point $y = a$, $\mathbb{E}(T_a) = +\infty$.*

Proof.

$$\mathbb{E}(T_a) = \int_0^\infty \frac{t|a|}{\sqrt{2\pi t^3}} e^{-\frac{(a^2)}{2t}} dt = +\infty \tag{4.9}$$

■

Together, these two findings mean that although there is guaranteed to be a time T_a where a is reached in Brownian motion, the average (expected) time of this happening is infinite. Looking at the integral, t has the possibility to be very large, and these very long hitting times contribute to the infinite expectation since the integral does not converge fast enough to offset the linear t. Hence, while almost every value can be guaranteed, it is impossible to predict when they will hit.

4.2. **As A Martingale.** Just like how a Markov chain does not rely on past events, a martingale is also a stochastic process that ties Brownian motion's future state to its current state.

Definition 4.10. A stochastic process $(Y_t)_{t\geq 0}$ is a martingale if:

(1) $E(Y_t|Y_r, 0 \leq r \leq s) = Y_s$ for all $0 \leq s \leq t$.
(2) $E(|Y_t|) < \infty$.

The first point means the expectation of future events is equal to the last event (previously referred to as the current event). Since this "chain" consistently adheres to the last event, the expectation of any event is also the expectation of the very first event:

$$E(E(Y_t|Y_{s+r}) = E(Y_s) = E(Y_0) \tag{4.10}$$

for all t since

$$E(Y_0) = E(E(Y_t|Y_{0+s}) = E(Y_s).$$

The existence of a stopping time where $\mathbb{E}(Y_t) = \mathbb{E}(Y_0)$ is explored more thoroughly in 5.3.

Theorem 4.11. *A symmetric random walk is a martingale.*

Proof. Let there be a random walk S consisting of random events X.

$$X_i = \begin{cases} +1, \text{ with probability } 1/2, \\ -1, \text{ with probability } 1/2, \end{cases}$$

for $i = 1, 2 \ldots$ For $n \geq 1$, let $S_n = X_1 + \cdots + X_n$ with $S_0 = 0$. Then, the first condition of a martingale is satisfied as

$$\begin{aligned} E(S_{n+1}|S_0, \ldots, S_n) &= E(S_n + X_{n+1}|S_0, \ldots, S_n) \\ &= E(S_n|S_0, \ldots, S_n) + E(X_{n+1}|S_0, \ldots, S_n) \\ &= S_n + E(X_{n+1}) = S_n \end{aligned}$$

since each new X is not dependent on the previous walks (each new interval has a $1/2$ probability of up or down regardless of current position) and $E(X_{n+1}) = 0$ because the next

walk has an equal chance in each direction. The second condition is satisfied intuitively as each X_i is bounded $+1$ or -1,

$$E(|S_n|) = E(\sum_{i=1}^{n} X_i) \leq E(|\sum_{i=1}^{n} X_i|) = \sum_{i=1}^{n} E(|X_i|) < \infty. \tag{4.11}$$

■

And since Brownian motion is constructed out of random walks (theorem something), Brownian Motion is also a martingale.

Theorem 4.12. *Standard Brownian motion $(B_t)_{t\geq 0}$ is a martingale.*

Proof.

$$E(B_t|B_r, 0 \leq r \leq s) = E(B_t - B_s + B_s|B_r, 0 \leq r \leq s) \tag{4.12}$$

$$= E(B_t - B_s|B_r, 0 \leq r \leq s) - E(B_s|B_r, 0 \leq r \leq s) \tag{4.13}$$

$$= E(B_t - B_s) + B_s = B_s \tag{4.14}$$

This proves the first condition, that the expectation of future events only depends on the current event. As for the second condition, which is fulfilled through independent increments,

$$E(|B_t|) = \int_{-\infty}^{\infty} |x| \frac{1}{\sqrt{2\pi t}} e^{\frac{-x^2}{2t}} dx = \int_0^{\infty} x\sqrt{\frac{2}{\pi t}} e^{\frac{-x^2}{2t}} dx = \sqrt{\frac{2t}{\pi}} < \infty. \tag{4.15}$$

■

5. Optional Stopping Theorem

The optional stopping theorem is a really interesting result that investigates where the martingale property [def whatever this ends up being] continues to apply in a stochastic process at random stopping times. It formalizes the idea that fair games can remain fair even when stopped at random times because it can hold the same properties. These results are essential to proving Wald's Lemmas and later Shorokhod's Embedding Theorem. This proof we will be following is referenced in [[LaL13]].

First, let's introduce some notation that will guide us in the proof.

Let $X = X_n$ be a stochastic process, and let T be a stopping time. For any positive integer n and $\omega \in \Omega$, we define

$$T \wedge n(\omega) = \min\{T(\omega), n\},$$

which takes either $T(\omega)$ or n, depending on which is smaller. Hence, as $n \to \infty$, $T \wedge n(\omega) \to T(\omega)$. Then we can define a stopped process:

Definition 5.1. A stopped process $X^T = X_n^T$ is given by

$$X_n^T = X_{T\wedge n(\omega)}(\omega),$$

the value of ω at this minimum time.

Lastly,

Proposition 5.2. *If a stochastic process $X = X_n$ is a martingale, then the stopped process $X^T = X_{T\wedge n}$ is also a martingale. In particular,*

$$\mathbb{E}(X_0) = \mathbb{E}(X_{T\wedge n}) \tag{5.1}$$

A proof for this can be found in page 4 of [LaL13].

Now, we can begin to tackle the Optional Stopping Theorem.

Theorem 5.3. *(Optional Stopping Theorem) Let $(\Omega, A, \mathbb{P})$ be a probability space, $\mathcal{F} = \{F_n\}$ a filtration on Ω, and $X = X_n$ a martingale with respect to $\mathcal{F}$. T is a stopping time. If one of the following conditions hold:*

1. *There is a positive integer N such that $T(\omega) \leq N$ for all $\omega \in \Omega$*
2. *There is a positive real number K such that $|X_n(\omega)| < K$ for all n and all $\omega \in \Omega$, where T is almost surely finite.*
3. *or $\mathbb{E}(T) < \infty$ and there is a positive real number K such that*

$$|X_n(\omega) - X_{n-1}(\omega)| < K,$$

Then X_T is integrable, and

$$\mathbb{E}(X_T) = \mathbb{E}(X_0).$$

Proof. All these conditions require T to be finite, meaning there will be a stopping time before the martingale ends to study. We also know that $X_{T\wedge n} \to X_t$ as $n \to \infty$. We also know that $\mathbb{E}(X_{T\wedge n}) = \mathbb{E}(X_0)$, and that $X_{T\wedge n}$ is integrable. Let's look at the first condition first. For $n > N$, $T \wedge N$ simplifies to T because $T(w) \leq N < n$. Therefore,

$$\mathbb{E}(X_{T\wedge n}) = \mathbb{E}(X_T) = \mathbb{E}(X_0). \tag{5.2}$$

Next, let's consider the second and third conditions. If the second condition holds, X_n is bounded such that

$$|X_{T\wedge n}(\omega)| < K \tag{5.3}$$

for all n and all $\omega \in \Omega$. We can rewrite this as

$$X_{T\wedge n}(\omega) = X_0(\omega) + \sum_{i=1}^{T\wedge n} X_i(\omega) - X_{i-1}(\omega) \tag{5.4}$$

for all ω. If the third condition also holds, we have

$$|X_{T\wedge n}(\omega)| \leq X_0(\omega) + \sum_{i=1}^{T\wedge n} X_i|(\omega) - X_{i-1}(\omega)| \leq X_0 + K \cdot T \wedge n(\omega) \tag{5.5}$$

because K limits the length of an increment of X, multiplied by $T \wedge n(\omega)$. X_0 is integrable, and so $\mathbb{E}(KT)$=$K\mathbb{E}(T) < 0$ because bound by the real number K and T is finite. In this way, both conditions two and three bind $X_{T\wedge n}$ by an integrable random variable, so the Dominated Convergence theorem applies. Therefore, X_T is integrable:

$$\lim_{n\to\infty} \int_\Omega X_{T\wedge n}(\omega)\mathbb{P}d(\omega) = \int_\Omega X_T(\omega)\mathbb{P}d(\omega). \tag{5.6}$$

This then gives

$$\lim_{n\to\infty} \mathbb{E}(X_{T\wedge n}) = \mathbb{E}(X_T). \tag{5.7}$$

Since we've established that $\mathbb{E}(X_{T\wedge n)} = \mathbb{E}(X_0)$ for all n, we have $\mathbb{E}(X_T) = \mathbb{E}(X_0)$ as desired. ■

5.1. **Wald's Lemmas and Skorokhod's Embedding Theorem.** The Optional Stopping Theorem is important in proving Wald's Lemmas and later Skorokhod's Embedding Theorem, which allows any random walk to be embedded into Brownian motion, provided that certain expectation conditions hold. As an introductory on Brownian motion, the formal proofs are beyond the scope of this paper, but these results are very interesting and can be a nice finish to your study on Brownian motion, especially looking deeper into martingales and relations to random walks. More about the Skorokhod's Embedding theorem can be found in [[Obł04]]

6. Brownian Motion in Quantitative Genetics

Now that we've discussed measure theory as a basis for Brownian motion, properties and random walk constructions of Brownian motion, and ways Brownian motion has the martingale and Markov properties which can be applied in conjunction stopping times, a look at Brownian motion in modeling, especially for fundamental as evolutionary biology will serve as a finale to demonstrate the widespread powers of Brownian motion.

6.1. **Genetic Drift.** The easiest way to apply Brownian motion to quantitative biology is to look at genetic drift. Here, we assume character is determined by many genes, each of small effect, and selection does not take place. Mutations are random and have small effects on each character.

We denote the mean value of a trait as $\overline{z}$, in a population with size N_e. Mutations (random variables) are drawn from a distribution with mean 0 and variance σ_m^2. We assume the number of alleles are so large that more than one mutation cannot happen to each allele. The population evolves purely based on this mutation and genetic drift (random chance). The limit of these random walks, the mean value of the trait, is thus a Brownian motion path as time t increases.

As we simulate this model many times, we would have a set of evolved population, each trajectory $\omega \in \Omega$. These populations would have roughly the same mean trait value, but would roughly differ. As a martingale, the expectation of this value would remain constant and equal to the initial value:

$$\mathbb{E}[\overline{z}_t] = \overline{z}_0. \tag{6.1}$$

Next, we must look at phenotypic variance σ_B^2, which represents simultaneously the variance in a trait over time, and the between-population variance, since different populations simply took different paths over time guided by this variance. For, this we look at additive genetic variance (which does not take into account gene to gene interactions like dominance and epistasis), indicated by σ_a^2. This σ_a^2 decreases from genetic drift and increases with mutations. We can model this as

$$\mathbb{E}[\sigma_a^2(t+1)] = 1 - \frac{1}{2N_e}\mathbb{E}[\sigma_a^2(t)] + \sigma_m^2 \tag{6.2}$$

where the first term represents the decrease in genetic variance due to genetic drift and the second the increase with mutations.

After some calculus, finding the equilibrium value of σ_a^2, and some substitution (more detailed explanation [[Har19]]), we can conclude that the variance at time t can be represented as

$$\sigma_B^2(t) = \frac{h^2\sigma_W^2 t}{N_e} \tag{6.3}$$

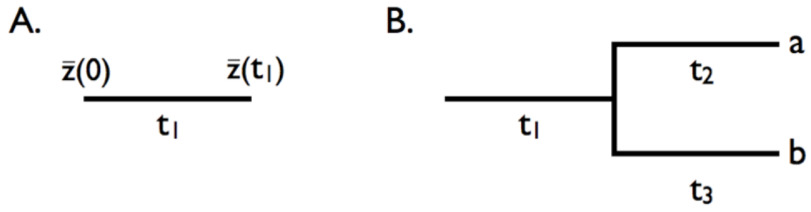

Figure 7. A. Evolution on a single lineage. B. Evolution relating species a and b, across the intervals t_1, t_2, and t_3. Credit to Luke J. Harmon [cite]

where h^2 measures heritability (proportion of phenotypic variance due to additive genetic effects, N_e the effective population size, and σ_W^2 the within-population variance (different from σ_a^2 because it also includes non additive genetic and environmental effects).

Once phenotypes are affected by enough mutations, the central limit theorem (Definition 2.4) guarantees the distribution of phenotypes in the population is normal.

6.2. **Selection.** We have now seen how genetic drift can be modeled with Brownian motion. It turns out, so can selection, in a very similar process. Oftentimes, genetic drift and selection are simultaneously acting on a population. Then, the variance σ_B^2 in selection can be shown as:

$$\sigma_B^2(t) = (\frac{h^2\sigma_W^2 t}{N_e} + \sigma_s^2)t. \tag{6.4}$$

Notice our substitution of Equation 6.3. However, when the variation due to selection is much stronger than variation due to drift (as it usually is),

$$\sigma_B^2 \approx \sigma_s^2. \tag{6.5}$$

Brownian motion's role in selection is achieved through the optimum, the optimal value where the strong stabilizing selection for a particular trait converges towards. This change in this model can be modeled in a Brownian motion fashion, assuming that there is at least some stabilizing selection (to the order of $1/t_{ij}$, where t_{ij} is the number of generations separating pairs of populations). This means the population is under strong stabilizing selection in any one generation, but because the optimum changes, the long term trajectory of trait's mean value follows Brownian motion.

6.3. **Phylogenetic Trees.** We can model this Brownian motion-like evolution with a phylogenetic tree. First we can consider this along a single branch (A. in Figure 7), which serves as the interval t_1, and we can denote the starting value as $\overline{z}_0$. We evolve with some rate parameter σ_B^2, giving

$$\mathbb{E}[\overline{z}_t] \sim N(\overline{z}_0, \sigma_B^2 t_1). \tag{6.6}$$

Then, we have figure B (Figure 7), also starting from $\overline{z}_0$ and evolving with rate parameter σ_B^2. The trait mean in both a and b evolves according to Brownian motion, a on the intervals of time $t_1 + t_2$ and b on the intervals of time $t_1 + t_3$. This gives

$$\overline{x}_a \sim N[\overline{z}_0, \sigma_B^2(t_1 + t_2)], \tag{6.7}$$

$$\begin{bmatrix} \sigma^2(t_1 + t_2) & \sigma^2 t_1 \\ \sigma^2 t_1 & \sigma^2(t_1 + t_3) \end{bmatrix} = \sigma^2 \begin{bmatrix} t_1 + t_2 & t_1 \\ t_1 & t_1 + t_3 \end{bmatrix} = \sigma^2 \mathbf{C}$$

Figure 8. Phylogenetic variance-covariance matrix for Figure 7

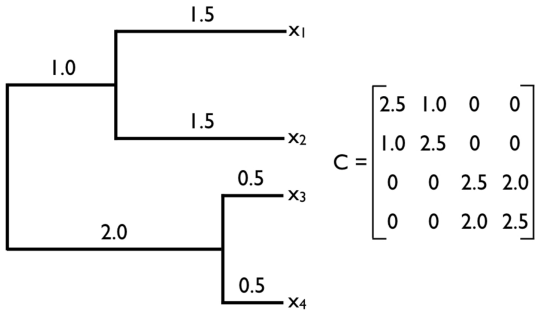

Figure 9. Variance-Covariance phylogenetic matrix for a small tree

and

$$\overline{x}_b \sim N[\overline{z}_0, \sigma_B^2(t_1 + t_3)]. \tag{6.8}$$

Because a and b share a common ancestor, t_1, we can calculate this similarity in the traits by finding the covariance. First, we note that

$$\overline{x}_a = \Delta\overline{x}_1 + \Delta\overline{x}_2 \tag{6.9}$$

$$\overline{x}_b = \Delta\overline{x}_1 + \Delta\overline{x}_3 \tag{6.10}$$

where $\Delta\overline{x}_1, \Delta\overline{x}_2, \Delta\overline{x}_3$ represent evolution over the intervals $t_1, t_2,$ and t_3 respectively. These evolutions are random variables with mean 0 and variances $\sigma_B^2 t_1, \sigma_B^2 t_2,$ and $\sigma_B^2 t_1$. Therefore, $\overline{x}_a$ and $\overline{x}_b$ are the sums of their respective random variables. Unsurprisingly, the covariance between a and b is just their shared term:

$$Cov(\overline{x}_b, \overline{x}_b) = Var(\Delta\overline{x}_1) = \sigma_B^2 t_1. \tag{6.11}$$

We can consider that the trait values for the two species came from a single draw from a multivariate normal distribution. Each trait has expected value $\overline{z}_0$. Then, we can form a variance-covariance matrix as shown in Figure 8. This matrix, C, is commonly encountered in comparative biology, and is called a phylogenetic variance-covariance matrix. This matrix follows a special structure: it is an $n \times n$ matrix, with a row and column corresponding to each of the n taxa. The diagonal represents the distance each taxa is from the root of the tree, while the off-diagonal refers to the shared branch lengths between pairs of taxa. This matrix is symmetrical, because the branches a shares with b is the same as the branches b shares with a. Under Brownian motion, these shared path lengths are proportional to the covariances of trait values. A full phylogenetic variance-covariance matrix is shown in Figure 9, describing the expected evolution of trait values under Brownian motion.

Acknowledgements

This paper was written as a part of the 2024 Independent Research and Paper Writing class of Euler Circle. I would like to thank Thomas Kaminsky for productive and interesting conversations and revisions. I would also like to thank Simon Rubinstein-Salzedo for making Euler Circle possible.

References

[Dob16] Robert P Dobrow. *Introduction to stochastic processes with R*. John Wiley & Sons, 2016.
[Har19] Luke J Harmon. *Phylogenetic comparative methods*. Independent Traverse City, MI, USA, 2019.
[LaL13] Scott M LaLonde. The martingale stopping theorem. *math.dartmouth.edu*, 2013.
[LIU21] CHRISTIAN LIU. Brownian motion as the limiting distribution of random walks. *Uchicago REU*, 2021.
[MP10] Peter Mörters and Yuval Peres. *Brownian motion*, volume 30. Cambridge University Press, 2010.
[Obł04] Jan Obłój. The skorokhod embedding problem and its offspring. *Project Euclid*, 2004.
[PN14] Hossein Pishro-Nik. *Introduction to probability, statistics, and random processes*. Kappa Research, LLC Blue Bell, PA, USA, 2014.
[Wik24] Wikipedia. Measure (mathematics) — Wikipedia, the free encyclopedia. `http://en.wikipedia.org/w/index.php?title=Measure%20(mathematics)&oldid=1234289136`, 2024. [Online; accessed 14-July-2024].