BROWNIAN MOTION

KISHAN JANI

Abstract. Brownian Motion is a continuous time stochastic process $\{B(t) : t \geq 0\}$, defined as having independent increments similar to a random walk. For Brownian motion, the continuous analog of the discrete probability distribution of Random walks is taken to be the normal distribution, with zero mean and variance based on increment size.
The goal of this expository paper is to study the Brownian motion that is resultant from these properties, and explore its wide variety of interesting properties. We start by proving some basic but fundamental properties of Brownian motion like nowhere differentiability and scaling invariance. Then, we make the analogy to the random walk more concrete, proving Donsker's invariance principle theorem that directly relates the two in a weak sup-norm convergence. Lastly, we explore the formulation of a stochastic calculus based on the properties of Brownian motion, along with some applications in physics and finance given by Ito's formula.
For this paper, we assume a background knowledge in probability theory and that of some rudimentary analysis and measure theory.

1. An Introduction and Preliminaries

The random walk, a familiar object in probability theory, is the stochastic process $\{S_t\}_{t\in T}$ over the index set $T = \mathbb{Z}_{\geq 0}$ starting from $S_0 = 0$ defined as $S_t = \sum_{k=0}^{t} X_k$, where $\{X_k\}_{k\in T}$ is a sequence of independent random variables. In a mathematical sense, Brownian motion can be thought of being the continuous time analog of the random walk, with the index set T being generally chosen as $\mathbb{R}_{\geq 0}$. However, it is quite different from the random walk in that the increments are normally distributed. This stochastic process permits some really interesting mathematics, which will be the central focus of this paper. First we introduce some preliminaries necessary for the understanding of the material here.

1.1. Multivariate Gaussian Distribution.

The primary reason behind introducing the multivariate Gaussian distribution is that it becomes applicable in better analyzing the Gaussian nature of Brownian Motion, which will then be used to prove several key theorems.

Definition 1.1.1. A random vector $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$ is called a multivariate d-dimensional standard Gaussian distribution if

(1) $X_1, X_2, \dots, X_d \sim \mathbf{N}(0, 1)$, where $\mathbf{N}(\mu, \sigma)$ is the normal distribution given by the probability density function $f_{\mathbf{N}}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$.
(2) $X_1, X_2, \dots, X_d$ are independent random variables.

The independence property really helps characterize the multivariate distribution: firstly, the joint distribution function will be

$$f_{\mathbf{X}}(\mathbf{t}) = \frac{1}{(\sqrt{2\pi})^d} e^{-\frac{1}{2}(t_1^2+t_2^2+\cdots+t_n^2)} = \frac{1}{(\sqrt{2\pi})^d} e^{-\frac{1}{2}\langle \mathbf{t},\mathbf{t}\rangle}$$

for $\mathbf{t} = (t_1, t_2, \dots t_n) \in \mathbb{R}^d$. For this distribution, the mean and covariance would be

$$\mathbb{E}[\mathbf{X}] = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathrm{Cov}(\mathbf{X}) = \mathbb{1}_d$$

where $\mathbb{1}_d$ is the $d \times d$ identity matrix. Th expectation is zero since all individual components are standard normally distributed with mean zero. The covariance gives the identity matrix since along the diagonal, the variance of each X_i is 1 while the $\mathrm{Cov}(X_i, X_j) = 0$ for $i \neq j$ since all components are independent random variables.

Next, we can scale and shift this definition of standard multivariate normal distribution to create a more generalized normal distribution. For scaling, we start with the following lemma related to invariance under orthogonal scaling.

Lemma 1. *If $\mathbf{X}$ is a d-dimensional standard Gaussian random vector and A an orthogonal $d \times d$ matrix, then $A\mathbf{X}$ is also standard Gaussian.*

Proof. Recall that orthogonality implies that A is invertible and that $A^{\mathrm{T}} = A^{-1}$ with $\det(A) = \pm 1$. Then we can apply a transformation on the density random vector AX as follows:

$$f_{A\mathbf{X}}(\mathbf{t}) = f_{\mathbf{X}}(A^{-1}\mathbf{t}) \mid \det(A^{-1}) \mid = f_{\mathbf{X}}(A^{-1}\mathbf{t}) = \frac{1}{(\sqrt{2\pi})^d} e^{-\frac{1}{2}\langle A^{-1}\mathbf{t}, A^{-1}\mathbf{t}\rangle}$$

Now we can exploit the orthogonality of A to get that $\langle A^{-1}\mathbf{t}, A^{-1}\mathbf{t}\rangle = \langle (A^{-1})^{\mathrm{T}} A^{-1}\mathbf{t}, \mathbf{t}\rangle = \langle (AA^{\mathrm{T}})^{-1}\mathbf{t}, \mathbf{t}\rangle = \langle \mathbf{t}, \mathbf{t}\rangle$. But then that implies

$$f_{A\mathbf{X}}(\mathbf{t}) = \frac{1}{(\sqrt{2\pi})^d} e^{-\frac{1}{2}\langle A^{-1}\mathbf{t}, A^{-1}\mathbf{t}\rangle} = \frac{1}{(\sqrt{2\pi})^d} e^{-\frac{1}{2}\langle \mathbf{t}, \mathbf{t}\rangle} = f_{\mathbf{X}}(\mathbf{t})$$

□

Now for the generalized Gaussian Random Variable, we have the following definition:

Definition 1.1.2. A random vector $\mathbf{Y}$ is d-dimensional Gaussian if we can express it as

$$\mathbf{Y} = A\mathbf{X} + \boldsymbol{\mu}$$

for some m-dimensional standard Gaussian vector $\mathbf{X}$, $d \times m$ matrix A, and some $\boldsymbol{\mu} \in \mathbb{R}^d$. Notice that from this definition, it is also clear that if $Y_1, Y_2, \dots, Y_n$ are independent Gaussian random variables, then $\mathbf{Y} = (Y_1, Y_2, \dots Y_n)^{\mathrm{T}}$ is a Gaussian random vector.

Proposition 2. *For the normal distribution $\mathbf{Y}$ expressed as $\mathbf{Y} := A\mathbf{X} + \mu$, we have expectation $\mathbb{E}[\mathbf{Y}] = \mu$ and covariance $Cov(\mathbf{Y}) = \mathbf{\Sigma} := AA^T$.*

Proof. Notice that for this matrix, $\mathbb{E}[\mathbf{Y}] = \mathbb{E}[A\mathbf{X} + \boldsymbol{\mu}] = A\mathbb{E}[\mathbf{X}] + \boldsymbol{\mu}$, which is equal to $\boldsymbol{\mu}$ since $\mathbf{X}$ is standard Gaussian. Additionally, since we can reinterpret the covariance matrix $\mathrm{Cov}(\mathbf{X}) = \mathbb{E}[(\mathbf{X} - \boldsymbol{\mu}) \cdot (\mathbf{X} - \boldsymbol{\mu})^{\mathrm{T}}]$, we then have for our covariance the following

$$\begin{aligned}
\mathrm{Cov}(\mathbf{Y}) &= \mathbb{E}\big[(A\mathbf{X} + \boldsymbol{\mu} - \mathbb{E}[A\mathbf{X} + \boldsymbol{\mu}]) \cdot (A\mathbf{X} + \boldsymbol{\mu} - \mathbb{E}[A\mathbf{X} + \boldsymbol{\mu}])^{\mathrm{T}}\big] \\
&= \mathbb{E}\big[A(\mathbf{X} - \mathbb{E}[\mathbf{X}]) \cdot (A(\mathbf{X} - \mathbb{E}[\mathbf{X}]))^{\mathrm{T}}\big] \\
&= A\mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}]) \cdot (\mathbf{X} - \mathbb{E}[\mathbf{X}])^{\mathrm{T}}]A^{\mathrm{T}} \\
&= A\mathrm{Cov}(\mathbf{X})A^{\mathrm{T}} = AA^{\mathrm{T}}
\end{aligned}$$

since $\mathrm{Cov}(\mathbf{X}) = \mathbb{1}$. □

We also have the following important result about Gaussian vectors under affine transformations.

Corollary 2.1. *Affine transformations preserve the Gaussian property.*

Proof. To see this, for $\mathbf{Y} \sim \mathbf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, we look at an affine transformation $\mathbf{Y} \to A'\mathbf{Y} + \boldsymbol{\mu}'$ where A' is a $d' \times d$ dimensional matrix and $\boldsymbol{\mu}' \in \mathbb{R}^{d'}$. In this case,

$$A'\mathbf{Y} + \boldsymbol{\mu}' = A'(A\mathbf{X} + \boldsymbol{\mu}) + \boldsymbol{\mu}' = (A'A)\mathbf{X} + (A'\boldsymbol{\mu} + \boldsymbol{\mu}')$$

which makes the transformation multivariate Gaussian by Proposition 2. □

1.2. **Measure Theory and Probability.** We provide a brief background on the measure-theoretic probability concepts that would be needed in this paper, especially in the later sections that deals with convergence. However, this definitely is not a proper introduction to measure theory; for a more comprehensive study of the subject and its use in probability, we refer the reader to [ARDDC00].

Definition 1.2.1 (σ-algebra)**.** For a set X, a σ algebra on X is the set $\mathcal{C}$ of subsets of X such that

(1) $\emptyset \in \mathcal{C}$
(2) If $S \in \mathcal{C}$ then $S^c \in \mathcal{C}$ as well
(3) For a finite or countable collection of sets $S_1, \dots S_n \in \mathcal{C}$, both the union $\bigcup_n S_n$ and the intersection $\bigcap_n S_n$ are in $\mathcal{C}$.

Definition 1.2.2 (Measure)**.** Given a set X and a σ-algebra on X, a measure on $(X, \mathcal{C})$ is a function $\mu : \mathcal{C} \to [0, \infty]$ such that the following properties hold:

(1) $\mu(\emptyset) = 0$
(2) For a finite or countable collection of pairwise disjoint sets $S_1, \dots S_n \in \mathcal{C}$,

$$\mu\Big(\bigcup_n S_n) = \sum_n \mu(S_n)$$

The triple $(X, \mathcal{C}, \mu)$ is called a measure space. Probability spaces, a special kind of measure space, are defined by the triple $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 1.2.3. For an index set I, a filtration on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a family $(\mathcal{F}(i) : i \in I)$ of σ-algebras such that $\mathcal{F}(s) \subset \mathcal{F}(t) \subset \mathcal{F}$ for $s < t$. A probability space with a filtration is called a filtered probability space. The natural index set for Brownian motion is $I = \mathbb{R}_{\geq 0}$.

Example 1.1. A right-continuous filtration is the family $(\mathcal{F}_i^+)_{i \in I}$ such that $\mathcal{F}_i^+ = \bigcap_{z > i} \mathcal{F}_z$.

Definition 1.2.4 (Nonanticipating Processes)**.** A stochastic process $\{X(t)\}_{t \geq 0}$ defined by the filtration $(\mathcal{F}(t))_{t \geq 0}$ is an adaptive or non-anticipating process if $X(t)$ is measurable for any $\mathcal{F}(t)$ for $t \geq 0$.

One last probability theoretic notion that we need is that of martingales, which have a strong connection to Brownian motion, and due to their rich and developed theory, will be used to prove certain results about the latter.

Definition 1.2.5 (Martingales)**.** A discrete time martingale is the sequence of random variables $\{X_n\}_{n \in \mathbb{N}}$ such that for all n, we have that $\mathbb{E}[|X_n|] < \infty$ and $\mathbb{E}[X_{n+1} \mid X_1, X_2, \dots X_n] = X_n$.

One of the very interesting properties of Brownian motion is its deep interconnectedness with Random walks, especially in a limiting sense, wherein they tend to exhibit the same distribution. To see this, however, we need to have a formal notion of convergence of probability measures, which we now provide in the $\mathcal{L}^2$ sense. Note that most results stated here in the $\mathcal{L}^2$ usually extend to more generalized spaces; however, proving these extensions is beyond the scope of this paper. For a much more rigorous and comprehensive treatment of convergence, we refer the reader to [Bil71]. We start by defining the $\mathcal{L}^p$ space.

Definition 1.2.6. The $\mathcal{L}^p$ space is a functional space that is a generalization of the ℓ^p norm of vector spaces. The $\mathcal{L}^p$ space is the vector space of all measurable functions for which

$$\Big(\int_S |f|^p d\mu\Big)^{1/p} < \infty$$

Of particular interest is the $\mathcal{L}^2$ space, or the space of all square-integrable measurable functions, since it can also be thought of as containing all random variables with a finite second moment $\mathbb{E}[X^2] < \infty$. It is also very important since most convergences we will observe, partly due to the quadratic variation of Brownian motion, will be in an $\mathcal{L}^2$ sense. In general, $\mathcal{L}^2$ convergence is a norm commonly seen throughout throughout probability theory.

A significantly important theorem is about the weak convergence of measures, commonly referred to as the Portmanteau theorem. This theorem equates several definitions of weak convergence, some of which we use in proving results. This theorem, along with its proof, can be found in any good measure theory book, take for example [Bil71]. We also have another really important theorem about martingales that we are able to use for proving an *embedding* result about Brownian motion.

Theorem 3 (Levy's Upward Theorem)**.** *Suppose X is an integrable random variable and that $X_n = \mathbb{E}[X \mid \mathcal{F}_n]$. Then the sequence $\{X_n\}_{n\geq 0}$ is a uniformly integrable martingale and $\lim_{n\to\infty} X_n = \mathbb{E}[X \mid \mathcal{F}_\infty]$, where $\mathcal{F}_\infty = \cup_{n=1}^\infty \mathcal{F}_n$, as the union of elements of the filtration, is the smallest σ-algebra with the entire filtration.*

2. Brownian Motion

2.1. Introduction.

Definition 2.1.1. A real-valued stochastic process $\{B(t) : t \in \mathbb{R}_{\geq 0}\}$ with $B(0) = x$ for $x \in \mathbb{R}$ is called a (one-dimensional) Brownian Motion or Wiener Process if

(1) The process consists of independent increments, so that for times $0 \leq t_1 \leq t_2 \leq t_3 \leq \cdots \leq t_n$, the sequence of steps $B(t_n) - B(t_{n-1}), B(t_{n-1}) - B(t_{n-2}). \dots, B(t_2) - B(t_1)$ forms a sequence of independent random variables.
(2) $B(t+s) - B(t) \sim \mathbf{N}(0, s)$
(3) Almost surely, $t \mapsto B(t)$ is continuous.

Lastly, we say that $\{B(t) : t \geq 0\}$ is a standard Brownian motion if $B(0) = 0$.

It is natural to ask if such a stochastic process could actually exist. Fortunately, due to a non-trivial theorem by Wiener, we can affirm the existence of Brownian motion.

Theorem 4 (Wiener 1923)**.** *Standard Brownian Motion exists.*

Clearly, this theorem is fundamental to the study of this subject. However, the Brownian motion constructed by Wiener in his proof is quite complex and lengthy, and is beyond the scope of this paper. This proof, with appropriate motivation and rigor, can be found in [MP10], or any other text that gives a detailed and rigorous treatment of this topic. Another construction due to Levy can be found in [SP14].

Lemma 5. *Brownian Motion is a Gaussian process*

Proof. To prove this lemma, we will show that for $0 \leq t_1 \leq t_2 \leq \cdots \leq t_n$, $(B(t_1), B(t_2), \dots, B(t_n))$ is a Gaussian vector. We first define the random vector $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)^{\mathrm{T}}$ based on increments with $\delta_i = B(t_i) - B(t_{i-1})$. Since the increments of Brownian motion are independent Gaussian random variables, δ_i would be the same, and then it follows that $\boldsymbol{\delta}$ is a Gaussian random vector. Now notice that we can represent $(B(t_1), B(t_2), \dots B(t_n))$ by an affine transformation of $\boldsymbol{\delta}$ given by $B(t_i) = \delta_1 + \delta_2 + \cdots + \delta_i$. So then by Corollary 2.1, $(B(t_1), B(t_2), \dots B(t_n))$ must also be a Gaussian random vector. Note that for this random vector $\mathbb{E}[B(t)] = 0$ and $\mathrm{Cov}(B(t_i), B(t_j)) = \min\{t_i, t_j\}$, which follows from the definition of Brownian motion. □

With these initial formulations, we can see some interesting geometrical invariances that will become fundamental to Brownian motion and further extremely helpful in proving other properties.

Proposition 6. *Suppose $\{B(t) : t \geq 0\}$ is a standard Brownian motion. Then the following processes are also standard Brownian motions:*

(1) *Shifting invariance: For $s \geq 0$, the process defined as $\{W_1(t) := B(t+s) - B(s) : t \geq 0\}$, which is also independent of $\{B(t) : 0 \leq t \leq s\}$, or the Brownian motion before s.*
(2) *Scaling invariance: For $a > 0$, the process $\{W_2(t) : t \geq 0\}$ defined as $W_2(t) := \frac{1}{a}B(a^2 t)$.*
(3) *Inversion invariance: The process $\{W_3(t) : t \geq 0\}$ defined as*

$$W_3(t) := \begin{cases} 0 & t = 0 \\ tB(1/t) & t > 0 \end{cases}$$

Proof. We need to check the properties in Definition 2.1.1 to see that the process is still a Brownian motion.

(1) Note that for any sequential increment of $W_1(t)$, we have

$$W_1(t_{k+1}) - W_1(t_k) = B(t_{k+1} + s) - B(s) - [B(t_k + s) - B(s)] = B(t_{k+1}) - B(t_k),$$

and thus the sequence of steps is independent. The map is trivially continuous, and $W_1(0) = B(0+s) - B(s) = 0$. Finally, for $h > 0$,

$$\begin{aligned} W_1(t+h) - W_1(t) &= B(t+h+s) - B(s) - [B(t+s) - B(s)] \\ &= B(t+s+h) - B(t+s) \sim \mathbf{N}(0, h) \end{aligned}$$

since $\{B(t) : t \geq 0\}$ is a standard Brownian Motion. The existence of all these properties makes $W_1(t)$ a standard Brownian motion as well. To verify its independence from $B(t)$, simply note that since Brownian motion has independent increments, for $t_1, \ldots t_n \geq 0$ and $s \geq s_1, \ldots s_n \geq 0$, $(W(t_1), W(t_2), \ldots, W(t_n)) = (B(t_1 + s) - B(s), \ldots, B(t_n + s) - B(s))$ and $(B(s_1), B(s_2), \ldots B(s_n))$ are independent.
(2) Continuity of the map $W_2(t) : t \to 1/aB(a^2 t)$ follows since $W_2(t)$ is just a composition of continuous functions; furthermore, independence is not affected by scaling. Clearly we also have $W_2(0) = 0$. It remains to see that the random variable is normally distributed, that too with the same mean and variance as $B(t)$. First note that for some $h > 0$, $B(t+h) - B(t) \sim \mathbf{N}(0, h)$. Then

$$\begin{aligned} \mathbb{E}[W_2(t+h) - W_2(t)] &= \mathbb{E}[\frac{1}{a}B(a^2(t+h)) - \frac{1}{a}B(a^2 t)] \\ &= \frac{1}{a}\mathbb{E}[B(a^2 t + a^2 h) - B(a^2 t)] = 0 \end{aligned}$$

where the last equality follows from the definition of Brownian motion since $a^2 t \in \mathbb{R}_{\geq 0}$ is a point on the process. We employ a similar process for the variance, where

$$\begin{aligned} \mathrm{Var}[W_2(t+h) - W_2(t)] &= \mathrm{Var}[\frac{1}{a}B(a^2(t+h)) - \frac{1}{a}B(a^2 t)] \\ &= \frac{1}{a^2}\mathrm{Var}[B(a^2 t + a^2 h) - B(a^2 t)] = \frac{1}{a^2} \cdot a^2 h = h \end{aligned}$$

and thus the Brownian nature follows.
(3) A proof of the third part requires a little more analysis. Clearly, $W_3(t)$ for $t \geq 0$ is a Gaussian process as well, and it also follows that the random vector $\mathbf{W}_3(t)$ of the motion

has expectation zero. The covariance between two components for $t > 0$ and $h \geq 0$ is given by

$$\text{Cov}(W_3(t+h), W_3(t)) = t(t+h)\text{Cov}(B(1/(t+h)), B(1/t)) = t(t+h) \cdot \frac{1}{t+h} = t$$

which are same as the results one would obtain for $B(t)$. Since Gaussian processes are uniquely determined by their mean and variance, it follows that $W_3(t)$ is similar in distribution to $B(t)$. It remains to show that the paths are continuous, specifically at $t = 0$, since the $t > 0$ case is trivial. To show for $t = 0$, note that since $\mathbb{Q}$ is countable and $\{X(t) : t \geq 0, t \in \mathbb{Q}\}$ has the same distribution as a Brownian motion,

$$\lim_{\substack{t \downarrow 0 \\ t \in \mathbb{Q}}} X(t) = 0$$

Next, since $\mathbb{Q}_{>0}$ is dense in $\mathbb{R}_{>0}$ and since $X(t)$ is a.s. continuous for $t > 0$, the limiting behaviour can be extended to the $\mathbb{R}_{>0}$ and thus $\lim\limits_{t \downarrow 0} X(t) = 0$ as desired.

□

These scaling and inversion properties of Brownian Motion lead to some interesting results.

Example 2.1. We can establish a bound on the Brownian motion since almost surely, we must have

$$\lim_{t \to \infty} \frac{B(t)}{t} = 0$$

This is easy to see with the time inversion invariance. Defining $u := 1/t$,

$$\lim_{t \to \infty} \frac{B(t)}{t} = \lim_{u \to 0^+} uB(1/u) = \lim_{u \to 0^+} W_3(u) = W_3(0) = 0$$

Example 2.2. We can analyse the first exit time of the Brownian motion, defined for the interval $[a, b]$ where $a < 0 < b$ as $\tau(a, b) = \inf\{t \geq 0 : B(t) = a \text{ or } B(t) = b\}$ primarily due to a scaling property. Notice that if we define $W(t) := 1/aB(a^2 t)$,

$$\begin{aligned}\mathbb{E}[\tau(a,b)] &= \inf\left\{t \geq 0 : \frac{1}{a}B(t) = 1 \text{ or } \frac{1}{a}B(t) = b/a\right\} \\ &= a^2 \inf\left\{t \geq 0 : W(t) = 1 \text{ or } W(t) = b/a\right\} = a^2\mathbb{E}[\tau(1, b/a)]\end{aligned}$$

Firstly, this implies that $\mathbb{E}[\tau(-b, b)]$ is a constant multiple of b^2. Secondly, this shows how we can extend this scaling invariance to have probabilistic and statistical interpretations; in this case:

$$\mathbb{P}\big(t < \infty : B(t) = a, B(t) \text{ exists } [a, b]\big) = \mathbb{P}\big(t < \infty : W(t) = 1, W(t) \text{ exists } [1, b/a]\big)$$

2.2. **Nowhere differentiability of Brownian motion.** From the results of Proposition 6, an extremely interesting feature of Brownian motion emerges: despite being continuous, it is extremely pathological in that it is almost surely nowhere differentiable. As one might expect, this theorem raises several problems in analyzing the Brownian motion and its subsequent calculus as the standard definition of the derivative is rendered meaningless. This problem is addressed with the stochastic integral, which is covered in Section 4.

Theorem 7 (Nondifferentiability of Brownian Motion)**.** *Almost surely, Brownian motion is nowhere differentiable. Furthermore, for all t,*

$$D^* B(t) = +\infty \text{ and } D_* B(t) = -\infty$$

almost surely, where

$$D^* f(x) := \limsup_{h \to 0} \frac{f(x+h) - f(x)}{h}, \quad D_* f(x) = \liminf_{h \to 0} \inf \frac{f(x+h) - f(x)}{h}$$

Proof. Due to the scaling invariance of Brownian motion it is sufficient to look at the $[0,1]$. Aiming for a contradiction, suppose there exists $t_0 \in [0,1]$ such that $-\infty < D_*B(t_0) \leq D^*B(t_0) < \infty$. Since Brownian motion is bounded in $[0,1]$, there must exist M such that for t_0, we have

$$\sup_{h\in[0,1]} \frac{|B(t_0+h)-B(t_0)|}{h} \leq M$$

It suffices to show that this event has probability zero regardless of M. We partition the unit interval into 2^n sub-intervals, taking the limit as $n\to\infty$. Let t_0 be contained within the interval $[(k-1)/2^n, k/2^n]$ for $n \geq 3$ and $1 \leq k \leq 2^n-3$.Then if we look at some $j \in [1, 2^n-k]$, we can use the triangle inequality to get that the following must also necessarily hold

$$\begin{aligned} \left|B\Big(\frac{k+j}{2^n}\Big) - B\Big(\frac{k+j-1}{2^n}\Big)\right| &\leq \left|B\Big(\frac{k+j}{2^n}\Big) - B(t_0)\right| + \left|B(t_0) - B\Big(\frac{k+j-1}{2^n}\Big)\right| \\ &\leq \frac{M(2j+1)}{2^n} \end{aligned}$$

For convenience of summability for $n \geq 3$ while avoiding overlap, we must set $j = 1,2,3$. Next, we examine the probability of the latter inequality actually being true. For this, first we define the event

$$E(n,k) = \left\{\left|B\Big(\frac{k+j}{2^n}\Big) - B\Big(\frac{k+j-1}{2^n}\Big)\right| \leq \frac{M(2j+1)}{2^n} \text{ for } j=1,2,3\right\}$$

We ultimately wish to show that $\lim_{n\to\infty} \mathbb{P}(\bigcup_k E(n,k)) = 0$. Now notice that for $1 \leq k \leq 2^n-3$, the invariance properties of Brownian motion can be used as follows to bound the probability:

$$\begin{aligned} \mathbb{P}(E(n,k)) &= \mathbb{P}\left(\left\{\left|B\Big(\frac{k+j}{2^n}\Big) - B\Big(\frac{k+j-1}{2^n}\Big)\right| \leq \frac{M(2j+1)}{2^n} \text{ for } j=1,2,3\right\}\right) \\ &\leq \prod_{j=1}^{3} \mathbb{P}\left(\left\{\left|B\Big(\frac{k+j}{2^n}\Big) - B\Big(\frac{k+j-1}{2^n}\Big)\right| \leq \frac{M(2j+1)}{2^n}\right\}\right) \\ &= \prod_{j=1}^{3} \mathbb{P}\left(\left\{\left|B\Big(\frac{1}{2^n}\Big) - B(0)\right| \leq \frac{M(2j+1)}{2^n}\right\}\right) \end{aligned}$$

where the last equality follows by translational invariance. Bounding above by the $j=3$ case, we get

$$\begin{aligned} \mathbb{P}(E(n,k)) &\leq \mathbb{P}\left(\left|B\Big(\frac{1}{2^n}\Big) - B(0)\right| \leq \frac{7M}{2^n}\right)^3 \\ &= \mathbb{P}\left(\left|2^{n/2}B(2^{-n})\right| \leq \frac{7M2^{n/2}}{2^n}\right)^3 = \mathbb{P}\Big(|B(1)| \leq \frac{7M}{2^{n/2}}\Big)^3 \leq \Big(\frac{7M}{2^{n/2}}\Big)^3 \end{aligned}$$

The last equality follows from scaling invariance, and the last inequality follows from the fact that the density of the Gaussian is bounded by $1/\sqrt{2\pi} < 1/2$. Then over all possible k,

$$\mathbb{P}\Big(\bigcup_{k=1}^{2^n-3} E(n,k)\Big) \leq \sum_{k=1}^{2^n-3} \mathbb{P}(E(n,k)) \leq 2^n \mathbb{P}(E(n,k)) \leq 2^n\Big(\frac{7M}{2^{n/2}}\Big)^3 = \frac{(7M)^3}{2^{n/2}}$$

Now, observe that

$$\sum_{n=1}^{\infty} \mathbb{P}\Big(\bigcup_{k=1}^{2^n-3} E(n,k)\Big) \leq \sum_{n=1}^{\infty} \frac{(7M)^3}{2^{n/2}} = \frac{(7M)^3}{\sqrt{2}} \cdot \frac{1}{1-\frac{1}{\sqrt{2}}} = \frac{(7M)^3}{\sqrt{2}-1} < \infty,$$

so then we can apply the Borel-Cantelli Lemma to see that for infinitely many n,

$$\mathbb{P}(B(t) \text{ is differentiable}) = \mathbb{P}\Big(\text{there exists } t_0 \in [0,1] \text{ such that } \sup_{h\in[0,1]} \frac{|B(t_0+h)-B(t_0)|}{h} \leq M < \infty\Big)$$

$$\leq \mathbb{P}\Big(\limsup_{n\to\infty} \bigcup_{k=1}^{2^n-3} E(n,k)\Big) = 0$$

proving that $B(t)$ is almost surely nowhere differentiable. □

Due to this nowhere differentiability, we see that a significant problem emerges in the formulation of a proper calculus for such continuous-time stochastic processes. This was an area of interest in the study of Brownian motion, resolved by Ito through stochastic calculus, which we will explore in the final section.

3. Reflection Principles and Applications

3.1. **The strong Markov property and reflection.** Similar to what we see for the Random Walk on $\mathbb{Z}$, Brownian motion also permits many interesting reflection principles that help better analyze certain properties of the overall motion.

Definition 3.1.1. The maximum value of a Brownian motion is given by $M(t) := \max_{0\leq s\leq t}\{B(s)\}$, while the minimum can similarly be expressed as $m(t) := \min_{0\leq s\leq t}\{B(s)\}$

Proposition 8 (Reflection Principles)**.** *The following reflection principles are associated with a Brownian motion $\{B(t) : t \geq 0\}$:*

(1) *For τ a stopping time, $a \in \mathbb{R}$ such that $B(\tau) = a$, the standard Brownian motion $\{B(t) : t \geq 0\}$ reflected at τ, defined as*

$$W(t) = \begin{cases} B(t) & t < \tau \\ B(\tau) - \big[B(t) - B(\tau)\big] & t \geq \tau \end{cases}$$

is also a standard Brownian motion.

(2) *$M(t)$ has the same distribution as $|B(t)|$, that is for $a \in \mathbb{R}$*

$$\mathbb{P}(M(t) > a) = \mathbb{P}(\tau > t) = 2\mathbb{P}(B(t) > a) = \mathbb{P}(|B(t)| > a) = 2 - 2\Phi(a/\sqrt{t})$$

A rigorous proof of both these principles is based on the strong Markov property that Brownian motion obeys:

Lemma 9 (Strong Markov Property)**.** *Brownian motion obeys the strong Markov property, namely that for a standard motion $\{B(t) : t \geq 0\}$ and a stopping time τ defined with respect to it, $\Delta(\tau) = B(t+\tau) - B(\tau)$ for $t \geq 0$ is also a standard Brownian motion. Furthermore, for each $t > 0$, $\{\Delta(s) : 0 \leq s < t\}$ is independent of $\mathcal{F}^+(\tau)$.*

Proof. Equipped with this property,

(1) For $t > \tau$, first note that $B(t) = B(\tau) + B(t) - B(\tau)$. Additionally, by the strong Markov Property, for $0 \leq s \leq \tau$, we have $B(s+\tau) - B(\tau)$ also a Brownian Motion independent of $B(s)$. But now since negation also preserves the Brownian property, we know that $B(\tau) - B(s+\tau)$ for $s \geq 0$, or equivalently for $t > \tau$, must also be a Brownian motion. We can then use this on the first relationship as follows:

$$\{B(t), t > \tau\} = \{B(\tau) + B(t) - B(\tau), t > \tau\} \sim \{B(\tau) - \big[B(t) - B(\tau)\big], t > \tau\} = \{2a - B(t), t > \tau\}$$

which proves the first part of the theorem.

(2) We define the stopping time $\tau = \inf\{t \geq 0 : B(t) = a\}$ and let $W(t)$ be the Brownian motion reflected at τ. Notice that we can divide the desired probability into two cases:

$$\{M(t) > a\} = \{B(t) > a\} \cup \{B(t) \leq a, M(t) > a\}$$

The union between these events is disjoint so the probabilities can be summed directly. The key detail to notice next is that the second event $\{B(t) \leq a, M(t) > a\}$ is the same as saying $W(t) \geq a$. Since the latter set in the union implies that $t < \tau$, we will have $W(t) = B(t)$ for the interval, and thus $\mathbb{P}(M(t) > a) = \mathbb{P}(B(t) > a) + \mathbb{P}(W(t) > a) = 2\mathbb{P}(B(t) > a)$. Note that the other statements follow quite trivially, since $2\mathbb{P}(B(t) > a) = \mathbb{P}(|B(t) > a)$, and $\mathbb{P}(B(t) > a) = 2(1 - \mathbb{P}(B(t) \leq a)) = 2 - 2\Phi(a/\sqrt{t})$.

□

Example 3.1 (A Notion of Absorption)**.** We define the absorbed Brownian motion $A(t)$ to be

$$A(t) = \begin{cases} B(t) & M(t) < a \\ a & M(t) \geq a \end{cases}$$

We can derive the cumulative distribution function as

$$\mathbb{P}(A(t) \leq x) = \Phi\Big(\frac{x}{\sqrt{t}}\Big) - 1 + \Phi\Big(\frac{2a - x}{\sqrt{t}}\Big)$$

For $x < a$, we have the cumulative distribution probability as

$$\begin{aligned} \mathbb{P}(A(t) \leq x) &= \mathbb{P}(B(t) \leq x, M(t) < a) \\ &= \mathbb{P}(B(t) \leq x) - \mathbb{P}(B(t) \leq x, M(t) \geq a) \\ &= \mathbb{P}(B(t) \leq x) - \mathbb{P}(B(t) \leq x, \tau \leq t) \end{aligned}$$

where the last equality is a simple consequence of the Intermediate Value theorem: if the maximum has been hit above a, then the stopping time giving $B(\tau) = a$ must have been hit sometime before that. Notice that by the reflection principle, since $t \geq \tau$, for the second probability we have

$$\begin{aligned} \mathbb{P}(B(t) \leq x, \tau \leq t) &= \mathbb{P}(2a - B(t) \leq x, \tau \leq t) \\ &= \mathbb{P}(B(t) \geq 2a - x, \tau \leq t) = \mathbb{P}(B(t) \geq 2a - x) \end{aligned}$$

The last equality here follows because if $B(t) \geq 2a - x > a$, then it is directly implied that $a > x$ and so $\tau \leq t$ for the reflection. Then we can combine to get

$$\begin{aligned} \mathbb{P}(A(t) \leq x) &= \mathbb{P}(B(t) \leq x) - \mathbb{P}(B(t) \geq 2a - x) \\ &= \mathbb{P}(B(t) \leq x) + \mathbb{P}(B(t) \leq 2a - x) - 1 \\ &= \Phi\Big(\frac{x}{\sqrt{t}}\Big) - 1 + \Phi\Big(\frac{2a - x}{\sqrt{t}}\Big) \text{ as desired.} \end{aligned}$$

Example 3.2 (Arcsine Law)**.** For a standard Brownian motion $\{B(t) : t \geq 0\}$ restricted into $t \in [0, 1]$, the time at which the maximum value is achieved, defined by the random variable τ_M where $B(\tau_M) = M(1)$, is arcsine distributed.

Note that due to the scaling invariance of Brownain motion, we can extend the restricted Brownian motion in $[0, 1]$ to infinity to apply these properties in the general scenario. We define a maximum variable $M(t, X(t))$, which gives the maximum process for $X(s)$ over the interval $0 \leq s \leq t$. For $s \in [0, 1]$ in our interval,

$$\begin{aligned} \mathbb{P}(\tau_M < s) &= \mathbb{P}(M(s) > \max_{s \leq u \leq 1} B(u)) \\ &= \mathbb{P}(M(s) - B(s) > \max_{s \leq u \leq 1} B(u) - B(s)) \\ &= \mathbb{P}\Big(M(s, B(s - t) - B(s)) > M(1 - s, B(s + t) - B(s))\Big) \end{aligned}$$

By the reflection principle in the second part of Proposition 9, we can take the absolute values, which gives

$$\mathbb{P}\Big(M(s, B(s-t)-B(s)) > M(1-s, B(s+t)-B(s))\Big) = \mathbb{P}(|B_1(s)| \leq B_2(1-s))$$

where $B_1(s) := B(s-t) - B(s)$ and $B_2(1-s) := B(s+t) - B(s)$. Using the scaling invariance of Brownian Motion, we normalize B_1 and B_2 and introduce in place normalized gaussian random variables Z_1 and Z_2:

$$\begin{aligned}\mathbb{P}(|B_1(s)| \leq B_2(1-s)) &= \mathbb{P}\big(\sqrt{s}|Z_1| > \sqrt{1-s}|Z_2|\big) \\ &= \mathbb{P}\Big(\frac{|Z_2|}{|Z_1|} > \frac{\sqrt{s}}{\sqrt{1-s}}\Big) = \mathbb{P}\Big(\frac{|Z_2|}{\sqrt{|Z_1|^2+|Z_2|^2}} > \sqrt{s}\Big)\end{aligned}$$

The natural substitution then seems to be $(Z_1, Z_2) = (r\sin\theta, r\cos\theta)$ for a random variable $\theta \in [0, 2\pi]$. So then we have

$$\mathbb{P}\Big(\frac{|Z_2|}{\sqrt{|Z_1|^2+|Z_2|^2}} > \sqrt{s}\Big) = \mathbb{P}(|\sin\theta| < \sqrt{s}) = 4\mathbb{P}(\theta < \sin^{-1}(\sqrt{s})) = \frac{2}{\pi}\sin^{-1}(\sqrt{s})$$

which is the arcsine distribution.

4. Brownian Motion as the limit of a Random Walk

4.1. **Limiting behavior.** The analogy between random walks and Brownian motion previously mentioned actually extends further than the similar Markov nature. In particular, Brownian motion can be interpreted as the limit of a Random Walk through a functional extension of the Central Limit theorem, for Brownian Motion. To set up this limiting behaviour, let $X_1, X_2, \ldots$ be a sequence of i.i.d random variables with mean 0 and variance 1 and let $S_n = \sum_{i=1}^{n} X_n$ represent the corresponding random walk. Furthermore, we linearly interpolate the integral time index so that

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

which is a random function in $\mathcal{C}[0, \infty)$. Then if we consider the normalized sequence of functions $\mathscr{Z}_n(t)$ in $[0, 1]$

$$\mathscr{Z}_n(t) := \frac{S(nt)}{\sqrt{n}} \text{ for all } t \in [0, 1]$$

we have the following result.

Theorem 10 (Donsker's Invariance Principle)**.** *Let $\{B(t)\}_{t\geq 0}$ be a standard Brownian motion. Then on $\mathcal{C}[0, 1]$, the space of continuous functions on $[0, 1]$, with the metric induced by the sup-norm, the normalized sequence $\{\mathscr{Z}_n(t) : n > 0\}$ converges in distribution to $B(t)$ for $t \in [0, 1]$.*

The *invariance* here can be associated with the invariance of the distribution from discretized to continuous time steps, along with an invariance along variation, since any process with mean 0 and finite variance would converge to a Brownian motion. Furthermore, notice that Donsker's invariance principle lends further insight into the ubiquity of Brownian phenomena. For instance, consider the motion of microscopic particles, which under the influence of a quantum phenomena and inter-particle interactions, can be categorized as random. Then since these random influences lead to a random small displacement, interpreting their motion as Brownian is helpful since their displacement is essentially the limit of a random walk under small displacements and small time intervals.

4.2. **Embedding Theorems.** The proof of this theorem relies on significant convergence theorems over metric spaces, and thus involve some degree of measure theory. The idea behind this principle is that we can embed random variables in the behaviour of the stopping times of a Brownian Motion, which will follow the same law and have the same first-moment. This is the statement of the Skorokhod embedding theorem that we will see below. Then we choose a sequence of such stopping times with iid increments and interpret it as a random walk, whose limiting behaviour would converge to a Brownian motion. First however, we state the necessary conditions that govern Skorokhod's embedding theorem, which are fortunately addressed completely by Wald's lemmas.

Lemma 11 (Wald's lemmas)**.** *The following properties hold for a Brownian motion:*

(1) *Let $\{B(t)\}_{t\geq 0}$ be a standard linear Brownian motion and τ a stopping time for which either $\mathbb{E}[\tau] < \infty$ or $\{B(\min\{t,\tau\} : t \geq 0\}$ is dominated by an integrable random variable. Then $\mathbb{E}[B(\tau)] = 0$.*
(2) *Let $B(t)$ be a Brownian motion and τ a stopping time for which $\mathbb{E}[\tau] < \infty$. Then*

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

Theorem 12 (Skorokhod Embedding theorem)**.** *Let X be a real-valued random variable with $\mathbb{E}[X] = 0$ and finite second moment $\mathbb{E}[X^2] < \infty$. Then there exists a stopping time τ of the Brownian motion $\{B(t)\}_{t\geq 0}$ with respect to the natural filtration $(\mathcal{F}(t) : t \geq 0)$ such that X can be embedded onto $B(\tau)$, that is to say that $B(\tau)$ has the same law as X and $\mathbb{E}[X^2] = \mathbb{E}[\tau]$.*

Notice that by Wald's Identities, we have that $\mathbb{E}[B^2(\tau)] = \mathbb{E}[\tau]$, which makes the *embedding* further explicit. Before proving the theorem, we first look at an example embedding.

Example 4.1. We consider the simple case where X can take values a and b. For $\mathbb{E}[X] = 0$, we need $a < 0 < b$ and $\mathbb{P}(X = a) = b/(b-a)$ and $\mathbb{P}(X = b) = -a/(b-a)$. It is easy to see that for the stopping time defined on the basis of this random variable $\tau = \inf\{t : B(t) \notin (a,b)\}$, we have $\mathbb{E}[\tau] = -ab$, and that $B(\tau)$ has the same law as X.

For the proving the existence of such an embedding, we first construct the following intermediate embedding on martingales. Specifically, we look at the binary martingale, which is defined as follows.

Definition 4.2.1. A *binary splitting martingale* is a martingale in which for each $n \geq 1$, the random variable X_{n+1} is characterized by two values whenever the event $E(x_0, x_1, \ldots, x_n) = \{X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n\}$ has a positive probability. In a sense, there is a binary split between these two values at X_{n+1} for a given condition on the values of the previous event.

Lemma 13 (Dubins' Embedding theorem)**.** *Let Y be a real-valued random variable with finite second moment $\mathbb{E}[X]^2 < \infty$. Then there exists a martingale $\{X_n\}_{n\in\mathbb{N}}$ such that X_n converges to X almost surely in $\mathcal{L}^2$.*

Proof. To construct the aforementioned embedding, we define the following ± 1 elementary binary splitting martingale over $\{\Omega, \mathcal{F}, \mathbb{P}\}$. Take $X_0 = \mathbb{E}[X]$ and the basic trivial filtration $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Then, we define a sequence of auxillary Bernoulli variables as follows for $n > 0$

$$\xi_n = \begin{cases} 1 \text{ if } X \geq X_{n-1} \\ -1 \text{ if } X < X_{n-1} \end{cases}$$

Next, let $\mathcal{F}_n = \sigma(\xi_0, \xi_1, \ldots \xi_{n-1})$ be the σ-algebra generated by these random variables, and let $X_n = \mathbb{E}[X \mid \mathcal{F}_n]$. With this definition of the martingale (the fact that it is a martingale trivially follows) and its associated filtration $\{\mathcal{F}_n\}_{n\geq 0}$, it is evident that the martingale is binary splitting. Notice that the filtration $\mathcal{F}_n$ partitions the probability space into 2^n different sets of the form $E(x_0, x_1, \ldots, x_n) = \{X_0 = x_0, \ldots, X_n = x_n$ where $x_i = \pm 1$ for $1 \leq i \leq n\}$, each of which has a positive probability. Each random variable X_{n+1} takes values ± 1 based on $\mathbb{E}(x_0, x_1, \ldots, x_n)$, and thus the binary splitting nature is also evident. Now since $\mathbb{E}[X^2] < \infty$, X is bounded in $\mathcal{L}^2$, and thus by an application of Jensen's inequality, it follows that

$$\mathbb{E}[X_n^2] = \mathbb{E}\big[(\mathbb{E}[X \mid \mathcal{F}_n])^2\big] \leq \mathbb{E}\big[\mathbb{E}[X^2 \mid \mathcal{F}_n]\big] = \mathbb{E}[X^2] < \infty$$

Jensen's inequality can be applied since $\varphi : x \to x^2$ is a convex function from $\mathbb{R}$ to $\mathbb{R}$, and the last equality follows by the law of total expectation. This gives us that $\mathbb{E}[X_n^2]$ is bounded in $\mathcal{L}^2$, and so we can apply Levy's upward theorem for martingales and the convergence of $\mathcal{L}^2$ bounded martingales to say that $\lim\limits_{n\to\infty} X_n = X_\infty = \mathbb{E}[X \mid \mathcal{F}_\infty]$ for the filtration $\mathcal{F}_\infty = \bigcup_{i=1}^\infty \mathcal{F}_i$. Then it only remains to show that $X = X_\infty$ almost surely.

Claim 13.1. *Almost surely we have*

$$\lim_{n\to\infty} \xi_n(X - X_{n+1}) = |X - X_\infty|$$

With this claim, we note that

$$\mathbb{E}[\xi_n(X - X_{n+1})] = \mathbb{E}\big[\xi_n \mathbb{E}[X - X_{n+1} \mid \mathcal{F}_{n+1}]\big] = 0$$

and since $\xi_n(X - X_{n+1})$ is bounded in $\mathcal{L}^2$ and uniformly integrable, by the dominated convergence theorem,

$$\mathbb{E}[X - X_\infty] = \mathbb{E}[\lim_{n\to\infty} \xi_n(X - X_{n+1})] = \lim_{n\to\infty} \mathbb{E}[\xi(X - X_{n+1})] = 0$$

from the claim. But since $\mathbb{E}[X - X_\infty] = 0$, we have $X = X_\infty$ almost surely. The final step is to prove the claim. If $X(\omega) = X_\infty(\omega)$ for some $\omega \in \Omega$, then the claim naturally holds. In the case that $X(\omega) > X_\infty(\omega)$, then for $n > N$ for sufficiently large N we must have $X(\omega) > X_n(\omega)$, so we set $\xi_n = 1$ to get the claim. The same argument holds if $X(\omega) < X_\infty(\omega)$, except we choose $\xi_n = -1$, and with the claim, the embedding theorem has been proved. □

Proof of Skorokhod Embedding Theorem. Notice that if we take the binary splitting martingale from Dubins embedding, then since X_n is supported on at most two values based on $E(x_1, x_2, \ldots x_n)$, we can use the stopping time seen in Example 4.1. Next, we construct a sequence of times $\tau_0 \leq \tau_1 \leq \tau_2 \leq \cdots$ with $\lim_{n\to\infty} \tau_n = \tau$ defined as

$$\tau_n = \inf\Big\{t \geq \tau_{n-1} : B(t) \notin \big(f_n(B(\tau_1), \ldots, B(\tau_{n-1}), \xi_n) \text{ for } \xi_n = \pm 1\big)\Big\}$$

such that $\mathbb{E}[X_n^2] = \mathbb{E}[\tau_n]$ and $B(\tau_n) \sim X_n$. Notice that by the strong Markov property, $\{B(\tau_{n-1} + t) - B(\tau_{n-1})\}_{t\geq 0}$ is independent of the Brownian filtration $\mathcal{F}^{(B)}_{\tau_{n-1}}$, and so we can confirm that the distribution of each $B(\tau_n)$ would indeed match that of X_n as planned. To verify that we indeed have $\mathbb{E}[X_n^2] = \mathbb{E}[\tau_n]$, we use the law of total expectation

$$\begin{aligned}\mathbb{E}[\tau_n - \tau_{n-1}] &= \mathbb{E}\big[\mathbb{E}[\tau_n - \tau_{n-1} \mid \mathcal{F}^{(B)}_{\tau_{n-1}}]\big] \\ &= \mathbb{E}\big[\mathbb{E}[(B(\tau_n) - B(\tau_{n-1}))^2 \mid \mathcal{F}^{(B)}_{\tau_{n-1}}]\big] \\ &= \mathbb{E}\big[\mathbb{E}[(B(\tau_n) - B(\tau_{n-1})^2]\big] = \mathbb{E}[(X_n - X_{n-1})^2]\end{aligned}$$

where the last line follows by the strong Markov property since it asserts that $(B(\tau_n) - B(\tau_{n-1}))^2$ is independent of the filtration $\mathcal{F}^{(B)}_{\tau_{n-1}}$. The desired result then follows by properties of the martingale,

$$\mathbb{E}[\tau_n] = \sum_{j=1}^n \mathbb{E}[\tau_n - \tau_{n-1}] = \sum_{j=1}^n \mathbb{E}[(X_n - X_{n-1})^2] = \mathbb{E}[X_n^2]$$

Now, we use the dominated convergence theorem and $\mathcal{L}^2$ convergence

$$\mathbb{E}[\tau] = \lim_{n\to\infty} \mathbb{E}[\tau_n] = \lim_{n\to\infty} \mathbb{E}[X_n^2] = \mathbb{E}[X^2] < \infty$$

Furthermore, by the continuity of Brownian motion, $\lim_{n\to\infty} B(\tau_n) = B(\tau)$, which converges in distribution to $\lim_{n\to\infty} X_n = X$ as shown. ☐

Finally, we can prove Donsker's invariance principle after the introduction of a lemma.

Lemma 14. *Let $B(t)$ be a linear Brownian motion. Then for any random variable X with mean zero and variance one, there exists a sequence of stopping times*

$$0 = \tau_0 \leq \tau_1 \leq \tau_2 \leq \cdots$$

such that

(1) *The sequence $\{B(\tau_n)\}_{n\geq 0}$ has the distribution of a random walk and its increments have the law of X.*
(2) *For the sequence of $\mathcal{C}[0,1]$ random functions $\{\mathcal{Z}_n(t) : n > 0\}$ now constructed from the random walk $\{B(\tau_n)\}_{n\geq 0}$, so that*

$$\mathcal{Z}_n(t) = \frac{\mathcal{B}(nt)}{\sqrt{n}} \text{ where } \mathcal{B}(u) = B(\tau_{\lfloor u \rfloor}) + (u - \lfloor u \rfloor)(B(\tau_{\lfloor u \rfloor + 1}) - B(\tau_{\lfloor u \rfloor}))$$

we have the following sup*-convergence*

$$\lim_{n\to\infty} \mathbb{P}\Big(\sup_{0\leq t\leq 1} \Big| \mathcal{Z}_n(t) - \frac{B(nt)}{\sqrt{n}} \Big| > \varepsilon \Big) = 0$$

Proof. To prove the first part, we develop an inductive argument using the Skorokhod embedding theorem to claim that the following construction is valid.

Claim 14.1. *There exists a sequence of stopping times $0 = \tau_0 \leq \tau_1 \leq \tau_2 \leq \cdots$ such that $B(\tau_n)$ is the embedded random walk with $\mathbb{E}[\tau_n] = n$.*

By the embedding theorem, we first define a stopping time τ_1 such that $\mathbb{E}[\tau_1] = 1$ and $B(\tau_1) \overset{d}{=} X$, which sets the base case. For the inductive step, suppose we have τ_k a stopping time such that $\mathbb{E}[\tau_k] = k$ and $B(\tau_k) \overset{d}{=} X$. Then by the strong Markov property of Brownian motion, since

$$\{B'(t) := B(\tau_k + t) - B(\tau_k)\}_{t\geq 0}$$

is independent of the filtration $\mathcal{F}^+(\tau_k)$, we can consider another stopping time $\tau_{k\to k+1}$ such that $\mathbb{E}[\tau_{k\to k+1}] = 1$ and $B'(\tau_{k\to k+1}) \overset{d}{=} X$. Then notice that if we define $\tau_{k+1} := \tau_k + \tau_{k\to k+1}$, we get $\mathbb{E}[\tau_{k+1}] = 1 + \mathbb{E}[\tau_k] = k+1$ and that $B(\tau_{k+1})$ is the $(k+1)$th entry of the sequence which has increments given by the law of X due to our construction of $\tau_{i\to i+1}$.

For the second part we define the event E_n that the distributions $\mathcal{Z}_n(t)$ and $W_n(t) := \frac{B(nt)}{\sqrt{n}}$ do not converge, which is to say that there exists $t \in [0,1]$ such taht $|\mathcal{Z}_n(t) - W(t)| > \varepsilon$. Then our theorem will be proven if we show that $\lim_{n\to\infty} \mathbb{P}(E_n) = 0$. In the partitioning of the unit interval, let k be the integer that contains t in the sense that $(k-1)/n \leq t \leq k/n$. Since this interval is between integers, $\mathcal{Z}_n(t)$ is defined by its linear interpolation, so we have

$$E_n \subset \left\{ \exists\, t \in [0,1] \text{ such that } \left| \frac{\mathcal{B}(\tau_k)}{\sqrt{n}} - W_n(t) \right| > \epsilon \right\} \bigcup \left\{ \exists\, t \in [0,1] \text{ such that } \left| \frac{\mathcal{B}(\tau_{k-1})}{\sqrt{n}} - W_n(t) \right| > \epsilon \right\}$$

Next, by scaling invariance, we can substitute $B(\tau_k) = \sqrt{n} W_n(\tau_k/n)$ to get

$$E_n = E_n^* \subset \big\{ \exists\, t \in [0,1] \text{ such that } \big| W_n(\tau_k/n) - W_n(t) \big| > \epsilon \big\}$$
$$\cup \big\{ \exists\, t \in [0,1] \text{ such that } \big| W_n(\tau_{k-1}/n) - W_n(t) \big| > \epsilon \big\}$$

And then for $0 < \delta < 1$, we can further generate the inclusion by splitting based on stopping times:

$$E_n = E_n^* \subset \big\{ \exists\, s,t \in [0,2] \text{ such that } |s-t| < \delta,\ |W_n(s) - W_n(t)| > \varepsilon \big\}$$
$$\cup \big\{ \exists\, t \in [0,1] \text{ such that } |\tau_k/n - t| \vee |\tau_{k-1}/n - t| \geq \delta \big\}$$

Notice that by this redfinition, the first event is independent of n, so then since Brownian motion is continuous, we can choose arbitrarily small $\delta > 0$ to make the probability as small as possible, hence can send it to zero. Then all we need to show is that the second converges to zero, for which since originally $t \in [(k-1)/n, k/n]$, we obtain

$$\begin{aligned}
&\mathbb{P}\left\{\exists\, t \in [0,1] \text{ such that } |\tau_k/n - t| \vee |\tau_{k-1}/n - t| \geq \delta\right\} \\
&\quad \leq \mathbb{P}\left\{\sup_{1\leq k\leq n}\Big(\frac{\tau_k - (k-1)}{n}\Big)\vee\Big(\frac{k-\tau_{k-1}}{n}\Big) \geq \delta\right\} \\
&\quad \leq \mathbb{P}\left\{\sup_{1\leq k\leq n}\Big(\frac{\tau_k - (k-1)}{n}\Big) \geq \delta/2\right\} + \left\{\sup_{1\leq k\leq n}\Big(\frac{k-\tau_{k-1}}{n}\Big) \geq \delta/2\right\} = 0
\end{aligned}$$

As the last step towards our proof, we will show that each of the summands indeed converges to zero as $n \to \infty$. To see this, first note that

$$\lim_{n\to\infty}\frac{\tau_n}{n} = \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}(\tau_k - \tau_{k-1}) = \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\tau_{k\to k+1} = 1$$

where the convergence to 1 follows by the Komlmogorov law of large numbers since $\mathbb{E}[\tau_{k\to k+1}] = 1$ for all k by construction. Next, it remains to notice that since $\lim_{n\to\infty}\tau_n/n = 1$, we would also have that

$$\lim_{n\to\infty}\sup_{0\leq k\leq n}\frac{|\tau_k - k|}{n} = 0$$

Hence, the probabilities being added converge to zero, and so does $\lim_{n\to\infty}\mathbb{P}(E_n)$, which was our original objective. □

Proof of Donsker's Invariance Princple. With the lemma, we can prove Donsker's theorem simply by validating one of the conditions of the Portmanteau theorem for the sequence $\mathcal{Z}_n$ and $W_n(t)$. With definitions consistent with the lemma, let $K \subset \mathcal{C}[0,1]$ be closed. We define

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

From this definition, we have the following probabilistic inequality

$$\mathbb{P}(\mathcal{Z}_n \in K) \leq \mathbb{P}(W_n \in K[\epsilon]) + \mathbb{P}(\|\mathcal{Z}_n - W_n\|_{\text{sup}} > \epsilon$$

From the previous lemma, we know that

$$\lim_{n\to\infty}\mathbb{P}\Big(\sup_{0\leq t\leq 1}|\mathcal{Z}_n - W_n| > \varepsilon\Big) = 0$$

To treat the second term, notice that since we can choose epsilon arbitrarily, we have

$$\lim_{n\to\infty}\lim_{\epsilon\to 0+}\mathbb{P}(W_n \in K(\varepsilon)) = \lim_{\epsilon\to 0+}\mathbb{P}(B \in K(\epsilon)) = \mathbb{P}(B \in \cap_{\epsilon>0}K(\epsilon) = \mathbb{P}(B \in K)$$

So then the inequality becomes

$$\lim_{n\to\infty}\mathbb{P}(\mathcal{Z}_n \in K) \leq \mathbb{P}(B \in K)$$

and by an application of the Portmanteau theorem, it follows that $\mathcal{Z}_n(t) \overset{d}{=} B(t)$ as desired. □

5. Stochastic Calculus

5.1. **Setting up the stochastic integral.** As mentioned before, the nowhere differentiability of Brownian motion poses a significant problem with regards to how one might formulate a proper and intuitive calculus for it, especially in the absence of a limiting behavior. An extremely successful approach turns out to be to start with the integral, and after demonstrating convergence in a mean-square sense, further the theory of differentiation and differential equations. We mark the foundations of stochastic calculus in this section, as well as look at some key results like Ito's formula and explore some of its applications.

Definition 5.1.1 (Progressively Measurable processes)**.** A process $\{X(t,\omega) : t \geq 0, \omega \in \Omega\}$ is called a *progressively measurable process* if for each $t \geq 0$, the map $X : [0,t] \times \Omega \to \mathbb{R}$ is measurable with respect to the σ-algebra $\mathfrak{B}([0,t]) \otimes \mathcal{F}(t)$, where $\mathfrak{B}$ is a Borel measure and $\mathcal{F}$ is the complete filtration.

Lemma 15. *Any non-anticipating process that is either left or right-continuous is also progressively measurable.*

We do not prove this lemma, but a relatively straightforward proof can be found in [MP10]. Equipped with this and the notion of $\mathcal{L}^2$ convergence, however, we now define the stochastic integral, whose existence will be our main focus for a majority of this section.
Let $\{F(t,\omega) : t \geq 0, \omega \in \Omega\}$ be a progessively measurable process that is of the form

$$F(t,\omega) = \sum_{i=1}^{k} f_i(\omega)\mathbb{1}_{(t_i,t_{i+1}]}(t) \text{ for } 0 \leq t_1 \leq t_2 \leq \cdots \leq t_{k+1} \text{ and } \mathcal{F}(t_i) \text{ measurable } f_i$$

which analogous to the classical case, can define the integral finitely as

$$\int_0^\infty F(s)dB(s) := \sum_{i=1}^{k} f_i(B(t_{i+1}) - B(t_i))$$

Now we must show that every progressively measurable process can be expressed as the limiting behavior of a sequence of progressively measurable processes $\{F_n\}_{n\in\mathbb{N}}$, and for this the definition of the stochastic integral of F would be

$$\int_0^\infty F(s)dB(s) := \lim_{n\to\infty} \int_0^\infty F_n(s)dB(s)$$

Naturally, this raises several questions about existence and convergence, which we address in the next definition.

Definition 5.1.2. Let F be a progressively measurable process such that $\mathbb{E}\int_0^\infty F(s)^2 ds < \infty$. We define the *stochastic integral* to then be

$$\int_0^\infty F(s)dB(s) := \lim_{n\to\infty} \int_0^\infty F_n(s)dB(s)$$

where the convergence will be proven in the $\mathcal{L}^2$ sense, but can be extend to generalized spaces beyond $\mathcal{L}^2$ (see [SP14]). However, we require for the integral the following conditions to be satisfied:

(1) *Existence of limiting behavior*: Every progressively measurable process $\{F(t,\omega) : t \geq 0, \omega \in \Omega\}$ with $\mathbb{E}\int_0^\infty F(s)^2 ds < \infty$ can be approximated by progressively measurable step processes H_n in the ℓ^2 norm sense.
(2) *Monotone convergence for the limiting sequence*: For each of these sequences the above limiting integration relation holds in an $\mathcal{L}^2$ sense.
(3) *Invariance*: The limit is independent of the chosen sequence H_n.

The first lemma we introduce towards forming the stochastic integral will prove (1):

Lemma 16. *For every progressively measurable process F with properties as above, there is a sequence of progressivley measurable processes F_n such that* $\lim_{n\to\infty} \|F - F_n\|_2 = 0$.

Proof. We prove this lemma and form a proper approximation through three different convergence results with measures:

(1) **Approximation by a bounded progressively measurable process:** If we define a form of cut-off at time $n > 0$ with F_n as $F_n(s,\omega) := F(s,\omega)$ for $s \le n$ and $F_n(s,\omega) = 0$ otherwise, then clearly we get $\lim_{n\to\infty} \|F - F_n\|_2 = 0$. Next we consider the truncation $F_n(s,\omega) := \min\{F(s,\omega), n\}$ for large n on a finite interval, for which the ℓ^2 convergence also holds.

(2) **Approximation by a bounded, a.s. continuous, progressively measurable process:** Let $h = 1/n$, and setting $F(s,\omega) = F(0,\omega)$ for negative s, we can define a continuous process through the average

$$F_n(s,\omega) = \frac{1}{h}\int_{s-h}^{s} F(u,\omega)du$$

which taken over past entries, is non-anticipating. It is almost surely continuous, and thus by the aforementioned lemma, F_n is a progressively measurable process. The furthermore, as an average, the following limiting behavior is clear:

$$\lim_{n\to\infty} \frac{1}{h}\int_{s-h}^{s} F(u,\omega)du = \lim_{h\to 0^+} \frac{1}{h}\int_{s-h}^{s} F(u,\omega)du = F(s,\omega)$$

Finally, it remains to show that we indeed have convergence in an ℓ^2 sense. This is a simple consequence of the dominated convergence theorem over $\mathcal{L}^p$ spaces, and so we also have $\lim_{n\to\infty} \|F - F_n\|_2 = 0$.

(3) **Approximation by a step process:** The bounded, almost surely continuous and progressively measurable function defined above can easily be converted into a step process by defining $F_n(s,\omega) := F(k/n,\omega)$ for $j/n \le s \le (j+1)/n$. It is easy to check both progressive measurability and the desired limiting behavior, completing the proof.

□

Next, following a quick lemma, we can move on to proving the last two properties.

Lemma 17. *For F a progressively measurable step process such that $\mathbb{E}\int_0^\infty F(s)^2 ds < \infty$, we have that*

$$\mathbb{E}\left[\left(\int_0^\infty F(s)dB(S)\right)^2\right] = \mathbb{E}\int_0^\infty F(s)^2 ds$$

Proof. Due to the strong Markov property, this lemma becomes relatively easy to prove since the cross terms vanish. We consider the progressively measurable step process $F = \sum_{i=1}^k f_i \mathbb{1}_{(t_i,t_{i+1}]}$

$$\mathbb{E}\left(\int_0^\infty F(s)dB(S)\right)^2 = \mathbb{E}\left[\sum_{i=1}^k\sum_{j=1}^k f_i f_j\big(B(t_{i+1}) - B(t_i)\big)\big(B(t_{j+1}) - B(t_j)\big)\right]$$

$$= \sum_{i=1}^k \mathbb{E}\left[f_i^2\big(B(t_{i+1}) - B(t_i)\big)^2\right] + 2\sum_{i\neq j}\mathbb{E}\Big[f_i f_j\big(B(t_{i+1}) - B(t_i)\big)\big(B(t_{j+1}) - B(t_j) \mid \mathcal{F}(t_j)\big)\Big]$$

$$= \sum_{i=1}^k \mathbb{E}\left[f_i^2\big(B(t_{i+1}) - B(t_i)\big)^2\right] = \sum_{i=1}^k \mathbb{E}[f_i^2](t_{i+1} - t_i) = \int_0^\infty F(s)^2 ds$$

□

Notice that this lemma leads directly to the following corollary:

Corollary 17.1. *Suppose that for a progressively measurable step process $\{F_n(s,\omega)\}_{n\in\mathbb{N}}$ we have*

$$\lim_{m,n\to\infty} \mathbb{E}\int_0^\infty (F_m - F_n)^2 ds = 0$$

then we can make a similar statement mean-squared sense for the stochastic integral, namely it follows that

$$\lim_{m,n\to\infty} \mathbb{E}\left[\left(\int_0^\infty F_m(s) - F_n(s)dB(S)\right)^2\right] = 0$$

This final corollary allows us sufficient background to summarize and develop the results into one comprehensive theorem that addresses (2) and (3) from the original definition, completing the formulation of the stochastic integral for step processes.

Theorem 18 (The stochastic integral for step processes)**.** *Let F be a progressively measurable process. Suppose that $\{F_n\}_{n\in\mathbb{N}}$ is a sequence of progressively measurable step processes such that*

$$\lim_{n\to\infty} \mathbb{E}\int_0^\infty (F - F_n)^2 ds = 0$$

Then the stochastic integral is defined as

$$\lim_{n\to\infty} \int_0^\infty F_n(s)dB(s) =: \int_0^\infty F(s)dB(s)$$

which is convergent in an $\mathcal{L}^2$ sense and is independent of the sequence F_n. Furthermore, we also have that

$$\mathbb{E}\left(\int_0^\infty F(s)dB(S)\right)^2 = \mathbb{E}\int_0^\infty F(s)^2 ds$$

Proof. Applying the triangle inequality to Corollary 17.1 qualifies F_n for its assumption and so it follows that $\int_0^\infty F_n(s)dB(s)$ is a Cauchy sequence in $\mathcal{L}^2$ by virtue of the assertion of Corollary 17.1. Thus the stated limit exists, and independence of choice comes from the Corollary as well. The last statement holds simply because of Lemma 17 applied to F_n under the limit. □

While we have achieved the existence of the stochastic integral, the problem of continuity remains since the theorem only works for progressively measurable *step processes.* Fortunately, there is a natural extension into continuity, which we prove in the following theorem.

Theorem 19. *Suppose we have a progressively measurable process $\{F(s,\omega) : s \geq 0, \omega \in \Omega\}$ such that $\mathbb{E}\int_0^t F(s,\omega)^2 ds < \infty$ for all t. Then there almost surely exists a continuous modification of $\int_0^t F(s,\omega)dB(s)$ that is a martingale so that $\mathbb{E}\int_0^t F(s,\omega)dB(s) = 0$ for all t.*

Notice that our bounds of $[0,\infty)$ have been replaced by finite bounds. This comes from a simple addition to our definition of the stochastic integral: we define $F^{(t_0)}(s,\omega) := F(s,\omega)\mathbb{1}_{s\leq t_0}$ as a progressively measurable process based on $\{F(s,\omega) : s \geq 0, \omega \in \Omega\}$ such that $\mathbb{E}\int_0^\infty F(s)^2 ds < \infty$. Clearly, all properties are preserved, but due to the updated definition for the process $s > t_0$, we have the stochastic integral up to t_0 as

$$\int_0^t F(s)dB(s) := \int_0^\infty F^{(t_0)}(s)dB(s)$$

Proof. Moving on to the proof, our general goal will be to embed the properties of the stochastic integral for the progressively measurable processes to a continuous martingale $Y(t)$ that is in the question. Let F_n be a sequence of step processes such that we can fix large t_0 to have $\|F_n - F^{(t_0)}\|_2 \longrightarrow 0$, and so by the corollary and the argument in the previous theorem,

$$\mathbb{E}\Big(\int_0^\infty F_n(s) - F^{(t_0)}(s)dB(s)\Big)^2 \to 0$$

Naturally for $s \leq t$, $\int_0^s F_n(u)dB(u)$ is $\mathcal{F}(s)$ measurable, and beyond that $\int_s^t F_n(u)dB(u)$ is independent of the filtration due to a variant of the strong Markov property. This makes

$$\left\{X_n(s) := \int_0^s F_n(u)dB(u) : 0 \leq s \leq t_0\right\}$$

a martingale for all n. Based on the limiting behaviour of $X_n(s)$, we define another martingale

$$\left\{Y(t) := \mathbb{E}\Big[\int_0^{t_0} F(u)\,dB(u) \;\mid\; \mathcal{F}(t)\Big] : 0 \leq t \leq t_0\right\} \text{ such that } Y(t_0) = \int_0^{t_0} F(u)dB(u)$$

But now notice that by Doob's Maximal inequality,

$$\begin{aligned}
\mathbb{E}\Big[\sup_{0\leq t\leq t_0} \Big(X_n(t) - Y(t)\Big)^2\Big] &= \mathbb{E}\Big[\sup_{0\leq t\leq t_0} \Big(\int_0^t F_n(s)\,dB(s) - Y(t)\Big)^2\Big] \\
&\leq 4\mathbb{E}\Big(\int_0^{t_0} F_n(s)\,dB(s) - Y(t_0)\Big)^2 \\
&= 4\mathbb{E}\Big(\int_0^{t_0} F_n(s) - F(s)\,dB(s)\Big)^2 \to 0 \text{ as } n \to \infty
\end{aligned}$$

Thus, $Y(t)$, as the limiting behaviour of the continuous time martingale $X_n(t)$, is a continuous time martingale itself. Taking the limit in an $\mathcal{L}^2$ sense, we can examine properties of $Y(t)$ in the context of $X_n(t)$: for $0 \leq t \leq t_0$, the random variable $\int_0^t H(u)dB(u)$ must now be $\mathcal{F}(t)$ measurable, while the remaining portion of $\int_t^{t_0} H(u)dB(u)$ is independent of the filtration and has zero expectation. But this leads to the following

$$\begin{aligned}
Y(t) = \mathbb{E}[Y(t_0) \mid \mathcal{F}(t)] &= \mathbb{E}\Big[\int_0^{t_0} F(u)\,dB(u) \mid \mathcal{F}(t)\Big] \\
&= \mathbb{E}\Big[\int_0^{t} F(u)\,dB(u) \mid \mathcal{F}(t)\Big] + \mathbb{E}\Big[\int_t^{t_0} F(u)\,dB(u) \mid \mathcal{F}(t)\Big] \\
&= \int_0^t F(u)\,dB(u) + \mathbb{E}\Big[\int_t^{t_0} F(u)\,dB(u)\Big] = \int_0^t F(u)\,dB(u)
\end{aligned}$$

proving that $Y(t)$ is matches $\int_0^t F(u)\,dB(u)$, making the latter a suitable continuous analog for the stochastic integral. □

5.2. **Ito's Formula and Applications.** Apart from formulating the stochastic integral, Ito was also responsible for producing results that extensively characterized the nature of stochastic calculus and furthered the scope of integration of such functions. In this next section, we look at his eponymous formula, stated without proof as a rigorous proof is somewhat involved. One can find this proof, along with a much more rigorous and deeply measure theoretic treatment of the topic, in [SP14]. The book also delves extensively into the many applications of Ito's formula in stochastic differential equations; however, a more application-centric text with arguably less rigor is [Cal12], which lays heavy focus on the resultant Black-Scholes equation and other applications of stochastic calculus to finance. With that, we now provide Ito's formula in two different versions:

Theorem 20 (Ito's Formula I). *Let $f : \mathbb{R} \to \mathbb{R}$ be a twice continuously differentiable function such that $\mathbb{E}\int_0^t f'(B(s))^2 ds < \infty$ for some $t > 0$. Then almost surely for $0 \leq s \leq t$, we have*

$$f(B(s)) - f(B(0)) = \int_0^s f'(B(u))dB(u) + \frac{1}{2}\int_0^s f''(B(u))du$$

Ito's formula can be thought of as the fundamental theorem of calculus for stochastic processes, and plays a large role in the solution of stochastic differential equations as we will see in the following examples. That being said, we now state another version of Ito's formula that is more general.

Theorem 21 (Ito's Formula II)**.** *Let $\{W(s) : s \geq 0\}$ be an increasing, continuous and non-anticipating stochastic process. Let $f : \mathbb{R} \to \mathbb{R}$ be a twice continuously differentiable function in x, and a continuously differentiable function in y. Then if*

$$\mathbb{E}\int_0^t \Big(\partial_x f(B(s), W(s))\Big)^2 ds < \infty$$

almost surely, for $0 \leq s \leq t$ we have

$$f(B(s), W(s) - f(B(0), W(0)) = \int_0^s \partial_x f(B(u), W(u)) dB(u) + \int_0^s \partial_y f(B(u), W(u)) dW(u)$$
$$+\frac{1}{2}\int_0^s \partial_{xx} f(B(u), W(u) du$$

To see the faculty and sheer applicability of Ito's formula, consider the Ornstein Uhlenbeck process.

Example 5.1 (The Ornstein-Uhlenbeck process)**.** The Ornstein-Uhlenbeck process, used to model the velocity of a large Brownian particle under the influence of friction, obeys the stochastic differential equation

$$dX(t) = -\alpha X(t)dt + \beta dB(t)$$

A solution in terms of $X(t)$ can be calculated by virtue of Ito's formula. We start by defining the function $Y(t) = F(t, X(t))$ given by $f(x,t) = xe^{\alpha t}$. Notice that in terms of traditional ODEs, the $e^{\alpha t}$ term serves as our integrating factor. We have $\partial_t f(x,t) = \alpha x e^{\alpha t}$, $\partial_x f(x,t) = e^{\alpha t}$ and finally $\partial_{xx} f(x,t) = 0$. Then we can use Ito's formula on $f(t, X(t))$ to get that

$$Y(t) - Y(0) = \int_0^t \alpha X(s) e^{\alpha s} ds + \int_0^t e^{\alpha s} dX(s) + \underbrace{\frac{1}{2}\int_0^t \partial xx f(x,s) ds}_{=0}$$
$$= \underbrace{\int_0^t \alpha X(s) e^{\alpha s} ds + \int_0^t -\alpha X(s) e^{\alpha s} ds}_{=0} + \int_0^t \beta e^{\alpha s} dB(s)$$

where we used the original SDE $dX(t) = -\alpha X(t)dt + \beta dB(t)$. Then substituting $X(t)$ from our original map, we get

$$X(t)e^{\alpha t} = X(0) + \int_0^t \beta e^{\alpha s} dB(s)$$
$$\text{or } X(t) = X(0)e^{-\alpha t} + \int_0^t \beta e^{\alpha(s-t)} dB(s)$$

References

[ARDDC00] Robert B Ash, B Robert, Catherine A Doleans-Dade, and A Catherine. *Probability and measure theory*. Academic Press, 2000.

[Bil71] Patrick Billingsley. *Weak convergence of measures: Applications in probability*. SIAM, 1971.

[Cal12] Ovidiu Calin. An introduction to stochastic calculus with applications to finance. *Ann Arbor*, 2012.

[MP10] Peter Mörters and Yuval Peres. *Brownian motion*, volume 30. Cambridge University Press, 2010.

[SP14] René L Schilling and Lothar Partzsch. *Brownian motion: an introduction to stochastic processes*. Walter de Gruyter GmbH & Co KG, 2014.