BROWNIAN MOTION

1. INTRODUCTION

1.1. Wiener Process: Definition.

Definition 1. A *standard (one-dimensional) Wiener process* (also called *Brownian motion*) is a stochastic process $\{W_t\}_{t\geq 0+}$ indexed by nonnegative real numbers t with the following properties:

(1) $W_0 = 0$.
(2) With probability 1, the function $t \to W_t$ is continuous in t.
(3) The process $\{W_t\}_{t\geq 0}$ has *stationary, independent increments*.
(4) The increment $W_{t+s} - W_s$ has the NORMAL$(0, t)$ distribution.

A *standard $d-$dimensional Wiener process* is a vector-valued stochastic process

$$W_t = (W_t^{(1)}, W_t^{(2)}, \dots, W_t^{(d)})$$

whose components $W_t^{(i)}$ are independent, standard one-dimensional Wiener processes. A Wiener process with initial value $W_0 = x$ is gotten by adding x to a standard Wiener process. As is customary in the land of Markov processes, the initial value x is indicated (when appropriate) by putting a superscript x on the probability and expectation operators. The term *independent increments* means that for every choice of nonnegative real numbers $0 \leq s_1 < t_1 \leq s_2 < t_2 \leq \dots \leq s_n < t_n < \infty$, the *increment* random variables

$$W_{t_1} - W_{s_1}, W_{t_2} - W_{s_2}, \dots, W_{t_n} - W_{s_n}$$

are jointly independent; the term *stationary increments* means that for any $0 < s, t < \infty$ the distribution of the increment $W_{t+s} - W_s$ has the same distribution as $W_t - W_0 = W_t$. In general, a stochastic process with stationary, independent increments is called a *Lévy process*; more on these later. The Wiener process is the intersection of the class of *Gaussian processes* with the *Lévy processes.*

It should not be obvious that properties (1)–(4) in the definition of a standard Brownian motion are mutually consistent, so it is not *a priori* clear that a standard Brownian motion exists. (The main issue is to show that properties (3)–(4) do not preclude the possibility of continuous paths.) That it *does* exist was first proved by N. WIENER in about 1920. See the notes on *Gaussian Processes* for more on this.

1.2. Brownian Motion as a Limit of Random Walks.

One of the many reasons that Brownian motion is important in probability theory is that it is, in a certain sense, a limit of rescaled simple random walks. Let $\xi_1, \xi_2, \dots$ be a sequence of independent, identically distributed random variables with mean 0 and variance 1. For each $n \geq 1$ define a continuous–time stochastic process $\{W_n(t)\}_{t\geq 0}$ by

$$W_n(t) = \frac{1}{\sqrt{n}} \sum_{1\leq j\leq \lfloor nt \rfloor} \xi_j \tag{1}$$

This is a random step function with jumps of size $\pm 1/\sqrt{n}$ at times k/n, where $k \in \mathbb{Z}_+$. Since the random variables ξ_j are independent, the increments of $W_n(t)$ are independent. Moreover, for large n the distribution of $W_n(t+s) - W_n(s)$ is close to the $\mathrm{NORMAL}(0,t)$ distribution, by the Central Limit theorem. Thus, it requires only a small leap of faith to believe that, as $n \to \infty$, the distribution of the random function $W_n(t)$ approaches (in a sense made precise below) that of a standard Brownian motion.

Why is this important? First, it explains, at least in part, why the Wiener process arises so commonly in nature. Many stochastic processes behave, at least for long stretches of time, like random walks with small but frequent jumps. The argument above suggests that such processes will look, at least approximately, and on the appropriate time scale, like Brownian motion.

Second, it suggests that many important "statistics" of the random walk will have limiting distributions, and that the limiting distributions will be the distributions of the corresponding statistics of Brownian motion. The simplest instance of this principle is the central limit theorem: the distribution of $W_n(1)$ is, for large n close to that of $W(1)$ (the gaussian distribution with mean 0 and variance 1). Other important instances do not follow so easily from the central limit theorem. For example, the distribution of

$$M_n(t) := \max_{0 \le s \le t} W_n(t) = \max_{0 \le k \le nt} \frac{1}{\sqrt{n}} \sum_{1 \le j \le k} \xi_j \tag{2}$$

converges, as $n \to \infty$, to that of

$$M(t) := \max_{0 \le s \le t} W(t). \tag{3}$$

The distribution of $M(t)$ will be calculated explicitly below, along with the distributions of several related random variables connected with the Brownian path.

1.3. **Transition Probabilities.** The mathematical study of Brownian motion arose out of the recognition by Einstein that the random motion of molecules was responsible for the macroscopic phenomenon of *diffusion*. Thus, it should be no surprise that there are deep connections between the theory of Brownian motion and parabolic partial differential equations such as the heat and diffusion equations. At the root of the connection is the *Gauss kernel*, which is the transition probability function for Brownian motion:

$$P(W_{t+s} \in dy \,|\, W_s = x) \stackrel{\Delta}{=} p_t(x,y)dy = \frac{1}{\sqrt{2\pi t}} \exp\{-(y-x)^2/2t\}dy. \tag{4}$$

This equation follows directly from properties (3)–(4) in the definition of a standard Brownian motion, and the definition of the normal distribution. The function $p_t(y|x) = p_t(x,y)$ is called the *Gauss kernel*, or sometimes the *heat kernel*. (In the parlance of the PDE folks, it is the *fundamental solution* of the heat equation). Here is why:

Theorem 1. *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous, bounded function. Then the unique (continuous) solution $u_t(x)$ to the initial value problem*

$$\frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial^2 u}{\partial x^2} \tag{5}$$

$$u_0(x) = f(x) \tag{6}$$

is given by

$$u_t(x) = Ef(W_t^x) = \int_{y=-\infty}^{\infty} p_t(x,y)f(y)\,dy. \tag{7}$$

Here W_t^x is a Brownian motion started at x.

The equation (5) is called the *heat equation*. That the PDE (5) has only one solution that satisfies the initial condition (6) follows from the *maximum principle*: see a PDE text if you are interested. The more important thing is that the solution is given by the expectation formula (7). To see that the right side of (7) actually does solve (5), take the partial derivatives in the PDE (5) under the integral in (7). You then see that the issue boils down to showing that

$$\frac{\partial p_t(x,y)}{\partial t} = \frac{1}{2}\frac{\partial^2 p_t(x,y)}{\partial x^2}. \tag{8}$$

Exercise: Verify this.

1.4. **Symmetries and Scaling Laws.**

Proposition 1. *Let $\{W(t)\}_{t\geq 0}$ be a standard Brownian motion. Then each of the following processes is also a standard Brownian motion:*

$$\{-W(t)\}_{t\geq 0} \tag{9}$$

$$\{W(t+s) - W(s)\}_{t\geq 0} \tag{10}$$

$$\{aW(t/a^2)\}_{t\geq 0} \tag{11}$$

$$\{tW(1/t)\}_{t\geq 0}. \tag{12}$$

Exercise: Prove this.

The scaling law, in particular, has all manner of important ramifications. It is advisable, when confronted with a problem about Wiener processes, to begin by reflecting on how scaling might affect the answer. Consider,as a first example, the *maximum* and *minimum* random variables

$$M(t) := \max\{W(s) \,:\, 0 \leq s \leq t\} \qquad \text{and} \tag{13}$$

$$M^-(t) := \min\{W(s) \,:\, 0 \leq s \leq t\}. \tag{14}$$

These are well-defined, because the Wiener process has continuous paths, and continuous functions always attain their maximal and minimal values on compact intervals. Now observe that if the path $W(s)$ is replaced by its reflection $-W(s)$ then the maximum and the minimum are interchanged and negated. But since $-W(s)$ is again a Wiener process, it follows that $M(t)$ and $-M^-(t)$ have the same distribution:

$$M(t) \stackrel{\mathcal{D}}{=} -M^-(t). \tag{15}$$

Next, consider the implications of Brownian scaling. Fix $a > 0$, and define

$$\begin{aligned} W^*(t) &= aW(t/a^2) \qquad \text{and} \\ M^*(t) &= \max_{0\leq s\leq t} W^*(s) \\ &= \max_{0\leq s\leq t} aW(s/a^2) \\ &= aM(t/a^2). \end{aligned}$$

By the Brownian scaling property, $W^*(s)$ is a standard Brownian motion, and so the random variable $M^*(t)$ has the same distribution as $M(t)$. Therefore,

$$M(t) \stackrel{\mathcal{D}}{=} aM(t/a^2). \tag{16}$$

On first sight, this relation appears rather harmless. However, as we shall see in section **??**, it implies that the sample paths $W(s)$ of the Wiener process are, with probability one, nondifferentiable at $s = 0$.

Exercise: Use Brownian scaling to deduce a scaling law for the *first-passage time* random variables $\tau(a)$ defined as follows:

$$\tau(a) = \min\{t \,:\, W(t) = a\} \tag{17}$$

or $\tau(a) = \infty$ on the event that the process $W(t)$ never attains the value a.

2. The Strong Markov Property

2.1. **Filtrations and adapted processes.** Property (10) is a rudimentary form of the *Markov property* of Brownian motion. The Markov property asserts something more: not only is the process $\{W(t+s) - W(s)\}_{t\geq 0}$ a standard Brownian motion, but it is independent of the path $\{W(r)\}_{0\leq r\leq s}$ up to time s. This may be stated more precisely using the language of $\sigma-$*algebras*. (Recall that a $\sigma-$*algebra* is a family of events including the empty set that is closed under complementation and countable unions.)

Definition 2. A *filtration* $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ of a probability space $(\Omega, \mathcal{F}, P)$ is a family of $\sigma-$algebras $\mathcal{F}_t$ indexed by $t \in [0, \infty]$, all contained in $\mathcal{F}$, satisfying

(a) if $s \leq t$ then $\mathcal{F}_s \subset \mathcal{F}_t$, and
(b) $\mathcal{F}_\infty = \sigma(\cup_{t\geq 0}\mathcal{F}_t)$.

A stochastic process $(X_t)_{t\geq 0}$ defined on $(\Omega, \mathcal{F}, P)$ is said to be *adapted* to the filtration $\mathbb{F}$ if for each $t \geq 0$ the random variable (or random vector, if the stochastic process is vector-valued) X_t is measurable relative to $\mathcal{F}_t$. (A random vector Y is measurable with respect to a $\sigma-$algebra if for every open set U the event $Y \in U$ is in the $\sigma-$algebra.) The *natural filtration* for a stochastic process $(X_t)_{t\geq 0}$ is the filtration consisting of the smallest $\sigma-$algebras $\mathcal{F}_t^X$ such that the process X_t is adapted.

Example 1. Let W_t be standard Brownian motion and let $M(t)$ be the maximum up to time t. Then for each $t > 0$ and for every $a \in \mathbb{R}$, the event $\{M(t) > a\}$ is an element of $\mathcal{F}_t^W$. To see this, observe that by path-continuity,

$$\{M(t) > a\} = \bigcup_{s\in\mathbb{Q}:0\leq s\leq t} \{W(s) > a\}. \tag{18}$$

Here $\mathbb{Q}$ denotes the set of rational numbers. Because $\mathbb{Q}$ is a countable set, the union in (18) is a countable union. Since each of the events $\{W(s) > a\}$ in the union is an element of the $\sigma-$algebra $\mathcal{F}_t^W$, the event $\{M(t) > a\}$ must also be an element of $\mathcal{F}_t^W$. $\square$

The natural filtration isn't always the only filtration of interest, even if one is only interested in studying just a single process X_t. This is because it is often advantageous to introduce auxiliary random variables and/or stochastic processes Y_t, Z_t, etc. for use in comparison or coupling arguments. In dealing with Wiener processes (or more generally, Lévy processes), we will often want our filtrations to satisfy the following additional property.

Definition 3. Let $\{W_t\}_{t\geq 0}$ be a Wiener process and $\{\mathcal{G}_t\}_{t\geq 0}$ a filtration of the probability space on which the Wiener process is defined. The filtration is said to be *admissible* for the Wiener process if (a) the Wiener process is adapted to the filtration, and (b) for every $t \geq 0$, the post-t process $\{W_{t+s} - W_t\}_{s\geq 0}$ is independent of the $\sigma-$algebra $\mathcal{G}_t$.

Note that the standard filtration is embedded in any admissible filtration $\{\mathcal{G}_t\}_{t\geq 0}$, that is, for each t,

$$\mathcal{F}_t^W \subset \mathcal{G}_{t.}$$

2.2. **Markov and Strong Markov Properties.**

Proposition 2. *(Markov Property) If $\{W(t)\}_{t\geq 0}$ is a standard Brownian motion, then the standard filtration $\{\mathcal{F}_t^W\}_{t\geq 0}$ is admissible.*

Proof. This is nothing more than a sophisticated restatement of the independent increments property of Brownian motion. Fix $s \geq 0$, and consider two events of the form

$$\begin{aligned} A &= \cap_{j=1}^n \{W(s_j) - W(s_{j-1}) \leq x_j\} \in \mathcal{F}_s \qquad \text{and} \\ B &= \cap_{j=1}^m \{W(t_j + s) - W(t_{j-1} + s) \leq y_j\}. \end{aligned}$$

By the independent increments hypothesis, events A and B are independent. Events of type A generate the $\sigma-$algebra $\mathcal{F}_s$, and events of type B generate the smallest $\sigma-$algebra with respect to which the post-s Brownian motion $W(t+s) - W(s)$ is measurable. Consequently, the post-s Brownian motion is independent of the $\sigma-$algebra $\mathcal{F}_s$. $\square$

The Markov property has an important generalization called the *Strong Markov Property*. This generalization involves the notion of a *stopping time* for a filtration $\{\mathcal{F}_t\}_{t\geq 0}$.

Definition 4. A nonnegative random variable τ (possibly taking the value $+\infty$) is a *stopping time* with respect to a filtration $(\mathcal{F}_t)_{t\geq 0}$ if for every $t \geq 0$ the event $\{\tau \leq t\} \in \mathcal{F}_t$. The stopping time τ is *proper* if $\tau < \infty$ on Ω. The *stopping field* $\mathcal{F}_\tau$ associated with a stopping time τ is the $\sigma-$algebra consisting of all events $B \subset \mathcal{F}_\infty$ such that $B \cap \{\tau \leq t\} \in \mathcal{F}_t$ for every $t \geq 0$.

Example 2. $\tau(a) := \inf\{t : W(t) = a\}$ is a stopping time relative to the standard filtration. To see this, observe that, because the paths of the Wiener process are continuous, the event $\{\tau(a) \leq t\}$ is identical to the event $\{M(t) \geq a\}$. We have already shown that this event is an element of $\mathcal{F}_t$. Now let B be the event that the Brownian path $W(t)$ hits b before it hits a. Then $B \in \mathcal{F}_\tau$.

Exercise 1. Fix a filtration $\mathbb{F}$. Prove the following facts:

(a) Every constant $t \geq 0$ is a stopping time with respect to $\mathbb{F}$.

(b) If τ and ν are stopping times then so are $\tau \wedge \nu$ and $\tau \vee \nu$.
(c) If τ and ν are stopping times and $\tau \leq \nu$ then $\mathcal{F}_\tau \subset \mathcal{F}_\nu$.
(d) If τ is a stopping time and τ_n is the smallest dyadic rational $k/2^n$ larger than τ then τ_n is a stopping time.

Theorem 2. *(Strong Markov Property) Let $\{W(t)\}_{t\geq 0}$ be a standard Brownian motion, and let τ be a stopping time relative to the standard filtration, with associated stopping $\sigma-$algebra $\mathcal{F}_\tau$. For $t \geq 0$, define the* post-τ process

$$W^*(t) = W(t+\tau) - W(\tau), \tag{19}$$

and let $\{\mathcal{F}^*_t\}_{t\geq 0}$ be the standard filtration for this process. Then
(a) $\{W^*(t)\}_{t\geq 0}$ is a standard Brownian motion; and
(b) For each $t > 0$, the $\sigma-$algebra $\mathcal{F}^*_t$ is independent of $\mathcal{F}_\tau$.

The Strong Markov property holds more generally for arbitrary Lévy processes — see the notes on Lévy processes for a formal statement and complete proof, and also a proof of the following useful corollary:

Corollary 1. *Let $\{W_t\}_{t\geq 0}$ be a Brownian motion with admissible filtration $\{\mathcal{F}\}_t$, and let τ be a stopping time for this filtration. Let $\{W^*_s\}_{s\geq 0}$ be a second Brownian motion on the same probability space that is independent of the stopping field $\mathcal{F}_\tau$. Then the spliced process*

$$\begin{aligned} \tilde{W}_t &= W_t && \text{fort} \leq \tau, \\ &= W_\tau + W^*_{t-\tau} && \text{fort} \geq \tau \end{aligned} \tag{20}$$

is also a Brownian motion.

The hypothesis that τ be a stopping time is essential for the truth of the Strong Markov Property. Mistaken application of the Strong Markov Property may lead to intricate and sometimes subtle contradictions. Here is an example: Let T be the first time that the Wiener path reaches its maximum value up to time 1, that is,

$$T = \min\{t : W(t) = M(1)\}.$$

Observe that T is well-defined, by path-continuity, which assures that the set of times $t \leq 1$ such that $W(t) = M(1)$ is closed and nonempty. Since $M(1)$ is the maximum value attained by the Wiener path up to time 1, the post-T path $W^*(s) = W(T+s) - W(T)$ cannot enter the positive half-line $(0, \infty)$ for $s \leq 1 - T$. Later we will show that $T < 1$ almost surely; thus, almost surely, $W^*(s)$ does not immediately enter $(0, \infty)$. Now if the Strong Markov Property were true for the random time T, then it would follow that, almost surely, $W(s)$ does not immediately enter $(0, \infty)$. Since $-W(s)$ is also a Wiener process, we may infer that, almost surely, $W(s)$ does not immediately enter $(-\infty, 0)$, and so $W(s) = 0$ for all s in a (random) time interval of positive duration beginning at 0. But this is impossible, because with probability one,

$$W(s) \neq 0 \qquad \text{for all rational times } s > 0.$$

2.3. **Embedded Simple Random Walks.** Inside every one-dimensional Wiener process W_t are simple random walks. These fit together in a coherent way to form a sort of "skeleton" for the Wiener process that, in certain senses, completely determine everything that the path W_t does. To prove that these embedded simple random walks exist we need the following simple lemma.

Lemma 1. *Define $\tau = \min\{t > 0 : |W_t| = 1\}$. Then with probability 1, $\tau < \infty$, and so τ is a proper stopping time. Furthermore, τ is dominated by a random variable N with a geometric distribution.*

Proof. That τ is a stopping time follows by a routine argument. Thus, the problem is to show that the Wiener process must exit the interval $(-1, 1)$ in finite time, with probability one. This should be a familiar argument: I'll define a sequence of independent Bernoulli trials G_n in such a way that if any of them results in a success, then the path W_t must escape from the interval $[-1, 1]$. Set $G_n = \{W_{n+1} - W_n > 2\}$. These events are independent, and each has probability $p := 1 - \Phi(2) > 0$. Since $p > 0$, infinitely many of the events G_n will occur (and in fact the number N of trials until the first success will have the geometric distribution with parameter p). Clearly, if G_n occurs, then $\tau \leq n + 1$. □

The lemma guarantees that there will be a first time $\tau_1 = \tau$ when the Wiener process has traveled ± 1 from its initial point. Since this time is a stopping time, the post-τ process $W_{t+\tau} - W_\tau$ is an *independent* Wiener process, by the strong Markov property, and so there will be a first time when it has traveled ± 1 from its starting point, and so on. Because the post-τ process is independent of the stopping field $\mathcal{F}_\tau$, it is in particular independent of $W(\tau) = \pm 1$, and so the sequence of future ± 1 jumps is independent of the first. By an easy induction argument, the sequence of ± 1 jumps made in this sequence are independent and identically distributed. Similarly, the sequence of elapsed times are i.i.d. copies of τ. Formally, define $\tau_0 = 0$ and

$$\tau_{n+1} := \min\{t > \tau_n : |W_{t+\tau_n} - W_{\tau_n}| = 1\}. \tag{21}$$

The arguments above imply the following.

Proposition 3. *The sequence $Y_n := W(\tau_n)$ is a simple random walk started at $Y_0 = W_0 = 0$. Furthermore, the sequence of random vectors*

$$(W(\tau_{n+1}) - W(\tau_n), \tau_{n+1} - \tau_n)$$

is independent and identically distributed.

Corollary 2. *With probability one, the Wiener process visits every real number.*

Proof. The recurrence of simple random walk implies that W_t must visit every *integer*, in fact infinitely many times. Path-continuity and the intermediate value theorem therefore imply that the path must travel through every real number. □

There isn't anything special about the values ± 1 for the Wiener process — in fact, Brownian scaling implies that there is an embedded simple random walk on each discrete lattice (i.e., discrete additive subgroup) of $\mathbb{R}$. It isn't hard to see (or to prove, for that matter) that the embedded simple random walks on the lattices $m^{-1}\mathbb{Z}$ "fill out" the Brownian path in such a way that as $m \to \infty$ the polygonal paths gotten by connecting the dots in the embedded simple random walks converge uniformly (on compact time intervals) to the path W_t. This can be used to provide a precise meaning for the assertion made earlier that Brownian motion is, in some sense, a continuum limit of random walks. We'll come back to this later in the course (maybe).

The embedding of simple random walks in Brownian motion has other, more subtle ramifications that have to do with *Brownian local time*. We'll discuss this when we have a few more tools (in particular, the Itô formula) available. For now I'll just remark that the key is the way that the embedded simple random walks on the nested lattices $2^{-k}\mathbb{Z}$ fit together. It is clear that the embedded SRW on $2^{-k-1}\mathbb{Z}$ is a subsequence of the embedded SRW on $2^{-k}\mathbb{Z}$. Furthermore, the *way* that it fits in as a subsequence is exactly the same (statistically speaking) as the way that the embedded SRW on $2^{-1}\mathbb{Z}$ fits into the embedded SRW on $\mathbb{Z}$, by Brownian scaling. Thus, there is an infinite sequence of nested simple random walks on the lattices $2^{-k}\mathbb{Z}$, for $k \in \mathbb{Z}$, that fill out (and hence, by path-continuity, *determine*) the Wiener path. OK, enough for now.

One last remark in connection with Proposition 3: There is a more general — and less obvious — theorem of Skorohod to the effect that *every* mean zero, finite variance random walk on $\mathbb{R}$ is embedded in standard Brownian motion. See sec. **??** below for more.

2.4. **The Reflection Principle.** Denote by $M_t = M(t)$ the maximum of the Wiener process up to time t, and by $\tau_a = \tau(a)$ the first passage time to the value a.

Proposition 4.

$$P\{M(t) \geq a\} = P\{\tau_a \leq t\} = 2P\{W(t) > a\} = 2 - 2\Phi(a/\sqrt{t}). \tag{22}$$

The argument will be based on a symmetry principle that may be traced back to the French mathematician D. ANDRÉ. This is often referred to as the *reflection principle*. The essential point of the argument is this: if $\tau(a) < t$, then $W(t)$ is just as likely to be *above* the level a as to be *below* the level a. Justification of this claim requires the use of the Strong Markov Property. Write $\tau = \tau(a)$. By Corollary 2 above, $\tau < \infty$ almost surely. Since τ is a stopping time, the post-τ process

$$W^*(t) := W(\tau + t) - W(\tau) \tag{23}$$

is a Wiener process, and is independent of the stopping field $\mathcal{F}_\tau$. Consequently, the reflection $\{-W^*(t)\}_{t\geq 0}$ is also a Wiener process, and is independent of the stopping field $\mathcal{F}_\tau$. Thus, if we were to run the original Wiener process $W(s)$ until the time τ of first passage to the value a and then attach not W^* but instead its reflection $-W^*$, we would again obtain a Wiener process. This new process is formally defined as follows:

$$\begin{aligned} \tilde{W}(s) &= W(s) && \text{for } s \leq \tau, \\ &= 2a - W(s) && \text{for } s \geq \tau. \end{aligned} \tag{24}$$

Proposition 5. *(Reflection Principle) If $\{W(t)\}_{t\geq 0}$ is a Wiener process, then so is $\{\tilde{W}(t)\}_{t\geq 0}$.*

Proof. This is just a special case of Corollary 1. □

Proof of Proposition 4. The reflected process $\tilde{W}$ is a Brownian motion that agrees with the original Brownian motion W up until the first time $\tau = \tau(a)$ that the path(s) reaches the level a. In particular, τ is the first passage time to the level a for the Brownian motion $\tilde{W}$. Hence,

$$P\{\tau < t \ \text{ and } \ W(t) < a\} = P\{\tau < t \ \text{ and } \ \tilde{W}(t) < a\}.$$

After time τ, the path $\tilde{W}$ is gotten by reflecting the path W in the line $w = a$. Consequently, on the event $\tau < t$, $W(t) < a$ if and only if $\tilde{W}(t) > a$, and so

$$P\{\tau < t \ \text{ and } \ \tilde{W}(t) < a\} = P\{\tau < t \ \text{ and } \ W(t) > a\}.$$

Combining the last two displayed equalities, and using the fact that $P\{W(t) = a\} = 0$, we obtain

$$P\{\tau < a\} = 2P\{\tau < t \ \text{ and } \ W(t) > a\} = 2P\{W(t) > a\}.$$

□

Corollary 3. *The first-passage time random variable $\tau(a)$ is almost surely finite, and has the* one-sided stable *probability density function of index $1/2$:*

$$f(t) = \frac{ae^{-a^2/2t}}{\sqrt{2\pi t^3}}. \tag{25}$$

Essentially the same arguments prove the following.

Corollary 4.

$$P\{M(t) \in da \text{ and } W(t) \in a - db\} = \frac{2(a+b)\exp\{-(a+b)^2/2t\}}{(2\pi)^{1/2}t^{3/2}}\, dadb \tag{26}$$

It follows, by an easy calculation, that for every t the random variables $|W_t|$ and $M_t - W_t$ have the same distribution. In fact, the *processes* $|W_t|$ and $M_t - W_t$ have the same joint distributions:

Proposition 6. *(P. Lévy) The processes $\{M_t - W_t\}_{t\geq 0}$ and $\{|W_t|\}_{t\geq 0}$ have the same distributions.*

Exercise 2. Prove this. Hints: (A) It is enough to show that the two processes have the same *finite-dimensional distributions*, that is, that for any finite set of time points $t_1, t_2, \dots, t_k$ the joint distributions of the two processes at the time points t_i are the same. (B) By the Markov property for the Wiener process, to prove equality of finite-dimensional distributions it is enough to show that the two-dimensional distributions are the same. (C) For this, use the Reflection Principle.

Remark 1. The reflection principle and its use in determining the distributions of the max M_t and the first-passage time $\tau(a)$ are really no different from their analogues for simple random walks, about which you learned in Stat 312. In fact, we could have obtained the results for Brownian motion directly from the corresponding results for simple random walk, by using embedding.

Exercise 3. Brownian motion with absorption.

(A) Define Brownian motion with absorption at 0 by $Y_t = W_{t\wedge\tau(0)}$, that is, Y_t is the process that follows the Brownian path until the first visit to 0, then sticks at 0 forever after. Calculate the transition probability densities $p_t^0(x, y)$ of Y_t.

(B) Define Brownian motion with absorption on $[0, 1]$ by $Z_t = W_{t\wedge T}$, where $T = \min\{t : W_t = 0 \,\text{or}\, 1\}$. Calculate the transition probability densities $q_t(x, y)$ for $x, y \in (0, 1)$.

3. Wald Identities for Brownian Motion

3.1. **Wald Identities; Examples.**

Proposition 7. *Let $\{W(t)\}_{t\geq 0}$ be a standard Wiener process and let $\mathbb{G} = \{\mathcal{G}_t\}_{t\geq 0}$ be an admissible filtration. Then each of the following is a continuous martingale relative to $\mathbb{G}$ (with $\theta \in \mathbb{R}$):*

(a) $\{W_t\}_{t\geq 0}$
(b) $\{W_t^2 - t\}_{t\geq 0}$
(c) $\{\exp\{\theta W_t - \theta^2 t/2\}\}_{t\geq 0}$
(d) $\{\exp\{i\theta W_t + \theta^2 t/2\}\}_{t\geq 0}$

Consequently, for any bounded stopping time τ, each of the following holds:

$$EW(\tau) = 0; \tag{27}$$

$$EW(\tau)^2 = E\tau; \tag{28}$$

$$E\exp\{\theta W(\tau) - \theta^2\tau/2\} = 1 \qquad \forall \theta \in \mathbb{R}; \text{ and} \tag{29}$$

$$E\exp\{i\theta W(\tau) + \theta^2\tau/2\} = 1 \qquad \forall \theta \in \mathbb{R}. \tag{30}$$

The definition of martingales in continuous time is the same as in discrete time: $\{M_t\}_{t\geq 0}$ is said to be a martingale if for all $t, s \geq 0$,

$$E(M_{t+s} \mid \mathcal{F}_t) = M_t \quad a.s.$$

In fact we won't use anything about martingales here except for the Wald identities. The proofs will be given later. Observe that for *nonrandom* times $\tau = t$, these identities follow from elementary properties of the normal distribution. Also, if τ is an *unbounded* stopping time, then the identities may fail to be true: for example, if $\tau = \tau(1)$ is the first passage time to the value 1, then $W(\tau) = 1$, and so $EW(\tau) \neq 0$. Finally, it is crucial that τ should be a stopping time: if, for instance, $\tau = \min\{t \leq 1 : W(t) = M(1)\}$, then $EW(\tau) = EM(1) > 0$.

Example 3. Fix constants $a, b > 0$, and define $T = T_{-a,b}$ to be the first time t such that $W(t) = -a$ or $+b$. The random variable T is a finite, but unbounded, stopping time, and so the Wald identities may not be applied directly. However, for each integer $n \geq 1$, the random variable $T \wedge n$ *is* a bounded stopping time. Consequently,

$$EW(T \wedge n) = 0 \qquad \text{and} \qquad EW(T \wedge n)^2 = ET \wedge n.$$

Now until time T, the Wiener path remains between the values $-a$ and $+b$, so the random variables $|W(T \wedge n)|$ are uniformly bounded by $a + b$. Furthermore, by path-continuity, $W(T \wedge n) \to W(T)$ as $n \to \infty$. Therefore, by the dominated convergence theorem,

$$EW(T) = -aP\{W(T) = -a\} + bP\{W(T) = b\} = 0.$$

Since $P\{W(T) = -a\} + P\{W(T) = b\} = 1$, it follows that

$$\boxed{P\{W(T) = b\} = \frac{a}{a+b}.} \tag{31}$$

The dominated convergence theorem also guarantees that $EW(T \wedge n)^2 \to EW(T)^2$, and the monotone convergence theorem that $ET \wedge n \uparrow ET$. Thus,

$$EW(T)^2 = ET.$$

Using (31), one may now easily obtain

$$\boxed{ET = ab.} \tag{32}$$

Example 4. Let $\tau = \tau(a)$ be the first passage time to the value $a > 0$ by the Wiener path $W(t)$. As we have seen, τ is a stopping time and $\tau < \infty$ with probability one, but τ is not bounded. Nevertheless, for any $n < \infty$, the truncation $\tau \wedge n$ is a bounded stopping time, and so by the third Wald identity, for any $\theta > 0$,

$$E\exp\{\theta W(\tau \wedge n) - \theta^2(\tau \wedge n)\} = 1. \tag{33}$$

Because the path $W(t)$ does not assume a value larger than a until after time τ, the random variables $W(\tau \wedge n)$ are uniformly bounded by a, and so the random variables in equation (33) are dominated by the constant $\exp\{\theta a\}$. Since $\tau < \infty$ with probability one, $\tau \wedge n \to \tau$

as $n \to \infty$, and by path-continuity, the random variables $W(\tau \wedge n)$ converge to a as $n \to \infty$. Therefore, by the dominated convergence theorem,

$$E \exp\{\theta a - \theta^2(\tau)\} = 1.$$

Thus, setting $\lambda = \theta^2/2$, we have

$$\boxed{E \exp\{-\lambda \tau_a\} = \exp\{-\sqrt{2\lambda} a\}.} \tag{34}$$

The only density with this Laplace transform[1] is the one–sided stable density given in equation (25). Thus, the Optional Sampling Formula gives us a second proof of (22).

Exercise 4. First Passage to a Tilted Line. Let W_t be a standard Wiener process, and define $\tau = \min\{t > 0 : W(t) = a - bt\}$ where $a, b > 0$ are positive constants. Find the Laplace transform and/or the probability density function of τ.

Exercise 5. Two-dimensional Brownian motion: First-passage distribution. Let $Z_t = (X_t, Y_t)$ be a two-dimensional Brownian motion started at the origin $(0, 0)$ (that is, the coordinate processes X_t and Y_t are independent standard one-dimensional Wiener processes).

(A) Prove that for each real θ, the process $\exp\{\theta X_t + i\theta Y_t\}$ is a martingale relative to an admissible filtration.

(B) Deduce the corresponding Wald identity for the stopping time $\tau(a) = \min\{t : W_t = a\}$, for $a > 0$.

(C) What does this tell you about the distribution of $Y_{\tau(a)}$?

Exercise 6. Eigenfunction expansions. These exercises show how to use Wald identities to obtain eigenfunction expansions (in this case, Fourier expansions) of the transition probability densities of Brownian motion with absorption on the unit interval $(0, 1)$. You will need to know that the functions $\{\sqrt{2} \sin k\pi x\}_{k \geq 1}$ constitute an orthonormal basis of $L^2[0, 1]$. Let W_t be a Brownian motion started at $x \in [0, 1]$ under P^x, and let $T = T_{[0,1]}$ be the first time that $W_t = 0$ or 1.

(A) Use the appropriate martingale (Wald) identity to check that

$$E^x \sin(k\pi W_t) e^{k^2\pi^2 t/2} \mathbf{1}\{T > t\} = \sin(k\pi x).$$

(B) Deduce that for every C^∞ function u which vanishes at the endpoints $x = 0, 1$ of the interval,

$$E^x u(W_{t \wedge T}) = \sum_{k=1}^{\infty} e^{-k^2\pi^2 t/2} (\sqrt{2} \sin(k\pi x)) \hat{u}(k)$$

where $\hat{u}(k) = \sqrt{2} \int_0^1 u(y) \sin(k\pi y)\, dy$ is the kth Fourier coefficient of u.

(C) Conclude that the sub-probability measure $P^x\{W_t \in dy\, ; T > t\}$ has density

$$q_t(x, y) = \sum_{k=1}^{\infty} e^{-k^2\pi^2 t/2} 2 \sin(k\pi x) \sin(k\pi y).$$

[1] Check a table of Laplace transforms.

3.2. **Proof of the Wald Identities.** This argument follows a strategy that is useful generally in the study of continuous-time stochastic processes. (The strong Markov property can also be proved this way – see the lecture notes on Lévy processes on my 385 web page.) First, we will consider the case of *elementary* stopping times, that is, stopping times that take values only in a finite set. Then we will show that the general case follows by an approximation argument.

Step 1. Let τ be a stopping time that takes one of the values $0 = t_0 < t_1 < \cdots < t_K$. Then for each j the event $F_j = \{\tau = t_j\}$ is in the $\sigma-$algebra $\mathcal{F}_{t_j}$. (Exercise: Check this, using the definition of a continuous-time stopping time.) Consequently, τ is a stopping time for the discrete-time filtration

$$\mathcal{F}_0 \subset \mathcal{F}_{t_1} \subset \cdots \subset \mathcal{F}_{t_K}.$$

Let $\{M_t\}_{t\geq 0}$ be one of the processes listed in (a)-(d) of Proposition 7, e.g., $M_t = W_t^2 - t$. Then the discrete sequence M_{t_j} is a discrete-time martingale with respect to the discrete-time filtration $\{\mathcal{F}_{t_j}\}_{0\leq j\leq K}$, that is,

$$E(M_{t_{j+1}} \,|\, \mathcal{F}_{t_j}) = M_{t_j}.$$

(This follows from the independent increments property of Brownian motion.) Now Doob's Optional Sampling theorem for discrete time martingales applies, yielding the identities (27)–(30).

Step 2. Next, let τ be an arbitrary bounded stopping time, and let $T < \infty$ be a constant such that $\tau \leq T$. For each $m = 1, 2, \dots$ define

$$\tau_m = k/2^m \quad \text{if} (k-1)/2^m < \tau \leq k/2^m \quad \text{for some} \;\; k = 1, 2, \dots$$

Then (i) each τ_m is a stopping time; (ii) each τ_m takes values in a finite set; and (iii) $\lim_{m\to\infty} \tau_m = \tau$. (Exercise: Check all three of these statements. The important one is (i).) Also, the sequence τ_m is non-increasing, that is, $\tau_1 \geq \tau_2 \geq \cdots \geq \tau$. By Step 1,

$$EM_{\tau_m} = EM_0$$

for each $m = 1, 2, \dots$ and each of the martingales M_t listed in (a)-(d) of Proposition 7. Consider, for instance, the martingale $M_t = \exp\{\theta W_t - \theta^2 t/2\}$ where $\theta > 0$. Since $\tau \leq T$, it follows that $\tau_m \leq T + 1$, and so

$$\exp\{\theta W_{\tau_m}\} \leq \exp\{\theta \max_{t\leq T} W_t\}.$$

By Proposition 4, the random variable on the right has finite expectation. (Exercise: Explain this.) Therefore, all of the random variables $\exp\{\theta W_{\tau_m}\}$ are dominated by an integrable random variable. Since $\tau_m \to \tau$, the path-continuity of Brownian motion implies that

$$\lim_{m\to\infty} \exp\{\theta W_{\tau_m} - \theta^2 \tau_m/2\} = \exp\{\theta W_\tau - \theta^2 \tau/2\}.$$

Consequently, by the Dominated Convergence Theorem,

$$1 = \lim_{m\to\infty} E\exp\{\theta W_{\tau_m} - \theta^2 \tau_m/2\} = E\exp\{\theta W_\tau - \theta^2 \tau/2\}.$$

This proves the third Wald identity. The other cases are similar. (You should work through the details for the second Wald identity as an exercise.) $\square$

4. Brownian Paths

In the latter half of the nineteenth century, mathematicians began to encounter (and invent) some rather strange objects. Weierstrass produced a continuous function that is nowhere differentiable. Cantor constructed a subset C (the "Cantor set") of the unit interval with zero area (Lebesgue measure) that is nevertheless in one-to-one correspondence with the unit interval, and has the further disconcerting property that between any two points of C lies an interval of positive length totally contained in the complement of C. Not all mathematicians were pleased by these new objects. Hermite, for one, remarked that he was "revolted" by this plethora of nondifferentiable functions and bizarre sets.

With Brownian motion, the strange becomes commonplace. With probability one, the sample paths are nowhere differentiable, and the zero set $Z = \{t \leq 1 : W(t) = 0\})$ is a homeomorphic image of the Cantor set. These facts may be established using only the formula (22), Brownian scaling, the strong Markov property, and elementary arguments.

4.1. **Zero Set of a Brownian Path.** The zero set is

$$\mathcal{Z} = \{t \geq 0 \,:\, W(t) = 0\}. \tag{35}$$

Because the path $W(t)$ is continuous in t, the set $\mathcal{Z}$ is closed. Furthermore, with probability one the Lebesgue measure of $\mathcal{Z}$ is 0, because Fubini's theorem implies that the *expected* Lebesgue measure of $\mathcal{Z}$ is 0:

$$\begin{aligned} E|\mathcal{Z}| &= E\int_0^\infty \mathbf{1}_{\{0\}}(W_t)\,dt \\ &= \int_0^\infty E\mathbf{1}_{\{0\}}(W_t)\,dt \\ &= \int_0^\infty P\{W_t = 0\}\,dt \\ &= 0, \end{aligned}$$

where $|\mathcal{Z}|$ denotes the Lebesgue measure of $\mathcal{Z}$. Observe that for any fixed (nonrandom) $t > 0$, the probability that $t \in \mathcal{Z}$ is 0, because $P\{W(t) = 0\} = 0$. Hence, because $\mathbb{Q}_+$ (the set of positive rationals) is countable,

$$P\{\mathbb{Q}_+ \cap \mathcal{Z} \neq \emptyset\} = 0. \tag{36}$$

Proposition 8. *With probability one, the Brownian path $W(t)$ has infinitely many zeros in every time interval $(0, \varepsilon)$, where $\varepsilon > 0$.*

Proof. First we show that for every $\varepsilon > 0$ there is, with probability one, at least one zero in the time interval $(0, \varepsilon)$. Recall (equation (9)) that the distribution of $M^-(t)$, the minimum up to time t, is the same as that of $-M(t)$. By formula (22), the probability that $M(\varepsilon) > 0$ is one; consequently, the probability that $M^-(\varepsilon) < 0$ is also one. Thus, with probability one, $W(t)$ assumes both negative and positive values in the time interval $(0, \varepsilon)$. Since the path $W(t)$ is continuous, it follows, by the Intermediate Value theorem, that it must assume the value 0 at some time between the times it takes on its minimum and maximum values in $(0, \varepsilon]$.

We now show that, almost surely, $W(t)$ has *infinitely* many zeros in the time interval $(0, \varepsilon)$. By the preceding paragraph, for each $k \in \mathbb{N}$ the probability that there is at least one zero in $(0, 1/k)$ is one, and so with probability one there is at least one zero in every $(0, 1/k)$. This implies that, with probability one, there is an infinite sequence t_n of zeros converging to zero: Take any zero $t_1 \in (0,1)$; choose k so large that $1/k < t_1$; take any zero $t_2 \in (0, 1/k)$; and so on. $\square$

Proposition 9. *With probability one, the zero set $\mathcal{Z}$ of a Brownian path is a* perfect *set, that is, $\mathcal{Z}$ is closed, and for every $t \in \mathcal{Z}$ there is a sequence of* distinct *elements $t_n \in \mathcal{Z}$ such that $\lim_{n\to\infty} t_n = t$.*

Proof. That $\mathcal{Z}$ is closed follows from path-continuity, as noted earlier. Fix a rational number $q > 0$ (nonrandom), and define $\nu = \nu_q$ to be the first time $t \geq$ such that $W(t) = 0$. Because $W(q) \neq 0$ almost surely, the random variable ν_q is well-defined and is almost surely strictly greater than q. By the Strong Markov Property, the post-ν_q process $W(\nu_q + t) - W(\nu_q)$ is, conditional on the stopping field $\mathcal{F}_\nu$, a Wiener process, and consequently, by Proposition 8, it has infinitely many zeros in every time interval $(0, \varepsilon)$, with probability one. Since $W(\nu_q) = 0$, and since the set of rationals is countable, it follows that, almost surely, the Wiener path $W(t)$ has infinitely many zeros in every interval $(\nu_q, \nu_q + \varepsilon)$, where $q \in \mathbb{Q}$ and $\varepsilon > 0$.

Now let t be any zero of the path. Then either there is an increasing sequence t_n of zeros such that $t_n \to t$, or there is a real number $\varepsilon > 0$ such that the interval $(t - \varepsilon, t)$ is free of zeros. In the latter case, there is a rational number $q \in (t - \varepsilon, t)$, and $t = \nu_q$. In this case, by the preceding paragraph, there must be a *decreasing* sequence t_n of zeros such that $t_n \to t$. $\square$

It can be shown (this is not especially difficult) that every compact perfect set of Lebesgue measure zero is homeomorphic to the Cantor set. Thus, with probability one, the set of zeros of the Brownian path $W(t)$ in the unit interval is a homeomorphic image of the Cantor set.

4.2. **Nondifferentiability of Paths.**

Proposition 10. *With probability one, the Brownian path W_t is nowhere differentiable.*

Proof. This is an adaptation of an argument of DVORETSKY, ERDÖS, & KAKUTANI 1961. The theorem itself was first proved by PALEY, WIENER & ZYGMUND in 1931. It suffices to prove that the path W_t is not differentiable at any $t \in (0,1)$ (why?). Suppose to the contrary that for some $t_* \in (0,1)$ the path were differentiable at $t = t_*$; then for some $\varepsilon > 0$ and some $C < \infty$ it would be the case that

$$|W_t - W_{t_*}| \leq C|t - t_*| \quad \text{for all } \ t \in (t_* - \varepsilon, t_* + \varepsilon), \tag{37}$$

that is, the graph of W_t would lie between two intersecting lines of finite slope in some neighborhood of their intersection. This in turn would imply, by the triangle inequality, that for infinitely many $k \in \mathbb{N}$ there would be some $0 \leq m \leq 4^k$ such that[2]

$$|W((m+i+1)/4^k) - W((m+i)/4^k)| \leq 16C/4^k \quad \text{for each } \ i = 0, 1, 2. \tag{38}$$

I'll show that the probability of this event is 0. Let $B_{m,k} = B_{k,m}(C)$ be the event that (38) holds, and set $B_k = \cup_{m \leq 4^k} B_{m,k}$; then by the Borel-Cantelli lemma it is enough to show that (for each $C < \infty$)

$$\sum_{k=1}^{\infty} P(B_m) < \infty. \tag{39}$$

---

[2]The constant 16 might really be 32, possibly even 64.

The trick is *Brownian scaling*: in particular, for all $s, t \geq 0$ the increment $W_{t+s} - W_t$ is Gaussian with mean 0 and standard deviation $\sqrt{s}$. Consequently, since the three increments in (38) are independent, each with standard deviation 2^{-k}, and since the standard normal density is bounded above by $1/\sqrt{2\pi}$,

$$P(B_{m,k}) = P\{|Z| \leq 16C/2^k\}^3 \leq (32C/2^k\sqrt{2\pi})^3.$$

This implies that

$$P(B_k) \leq 4^k(32C/2^k\sqrt{2\pi})^3) \leq (32C/\sqrt{2\pi})^3/2^k.$$

This is obviously summable in k. □

Exercise 7. Local Maxima of the Brownian Path. A continuous function $f(t)$ is said to have a *local maximum* at $t = s$ if there exists $\varepsilon > 0$ such that

$$f(t) \leq f(s) \qquad \text{for all } t \in (s - \varepsilon, s + \varepsilon).$$

(A) Prove that if the Brownian path $W(t)$ has a local maximum w at some time $s > 0$ then, with probability one, it cannot have a local maximum at some later time s^* with the same value w. HINT: Use the Strong Markov Property and the fact that the rational numbers are countable and dense in $[0, \infty)$.

(B) Prove that, with probability one, the times of local maxima of the Brownian path $W(t)$ are dense in $[0, \infty)$

(C) Prove that, with probability one, the set of local maxima of the Brownian path $W(t)$ is *countable*. HINT: Use the result of part (A) to show that for each local maximum (s, W_s) there is an interval $(s - \varepsilon, s + \varepsilon)$ such that

$$W_t < W_s \quad \text{for all } t \in (s - \varepsilon, s + \varepsilon),\ t \neq s.$$

5. LÉVY'S CONSTRUCTION

The existence of Brownian motion – that is to say, the existence of a stochastic process with the properties listed in Definition 1 – was first proved by Wiener in about 1920. His proof was grounded on an important idea, namely that there must be an L^2 isometry between the linear space of Gaussian random variables generated by the Wiener process and the space $L^2[0,1]$ of square-integrable functions on the unit interval; this led him to discover the Fourier expansion of a Wiener path. However, the proof that this series converges uniformly to a continuous function of t is not easy.

Lévy later realized that the sine basis of $L^2[0,1]$ is not the most natural basis of $L^2[0,1]$ for studying Brownian motion. The *Haar basis* (the simplest of the wavelet bases) is much better suited, because it has "self-similarity" properties that mesh with those of the Wiener process (see Proposition 1). It turns out to be relatively simple to prove that the Haar function expansion of a Brownian path converges uniformly, as we will see.

Lévy's strategy was to build the random variables W_t one by one for all times $t \in \mathbb{D}$, the set of dyadic rationals (rationals of the form $k/2^m$). It is enough to consider only values of $t \in [0,1]$, because if we can build a stochastic process $\{W_t\}_{t\in[0,1]}$ that satisfies the requirements of Definition 1 then we can just as easily build countably many independent copies and then piece them together to get a Wiener process defined for all $t \geq 0$. So the plan is this: first draw a standard normal random variable and designate it as W_1; then draw $W_{1/2}$ from the conditional distribution given W_1; then draw $W_{1/4}$ and $W_{3/4}$ from the conditional distribution given $W_{1/2}$ and W_1; etc. Once we have constructed all of the W_t for $t \in \mathbb{D}$ we will show that with probability 1 they fit together to form a process with continuous paths.

The construction is based on a simple property of normal distributions:

Lemma 2. *Let X, Y be independent random variables, each normally distributed, with mean 0 and variances $s > 0$ and $t > 0$, respectively. Then the conditional distribution of X given that $X + Y = z$ is*

$$\mathcal{D}(X \mid X + Y = z) = \text{Normal}(zs/(s+t), st/(s+t)). \tag{40}$$

Proof. Exercise. □

Next is a variation of the Markov property.

Lemma 3. *Let $t_0 = 0 < t_1 < \cdots \leq t_K = 1$ be a finite set of times. Let W_{t_i} be random variables indexed by these times such that $W_0 = 0$ and such that the increments $W_{t_i} - W_{t_i - 1}$ are independent, Gaussian random variables with mean 0 and variances $t_i - t_{i+1}$. Then for each $0 < J < K$, the conditional distribution of W_{t_J} given the values of all the remaining random variables W_{t_i}, where $i \neq J$, is the same as its conditional distribution given only the values of $W_{t_{J-1}}$ and $W_{t_{J+1}}$.*

Proof. Conditioning on the values W_{t_i}, where $i \neq J$, is the same as conditioning on the values $W_{t_{J-1}}, W_{t_{J+1}}$, and the increments $W_{t_{i+1}} - W_{t_i}$ where $i < J - 1$ or $i + 1 \geq J + 2$. But these increments are independent of the random vector $(W_{t_{J-1}}, W_{t_J}, W_{t_{J+1}})$, and so the lemma follows from the next exercise.

Exercise 8. Let $(Z, Y_1, Y_2, .., Y_L)$ and $(X_1, X_2, \ldots, X_K)$ be independent random vectors. Show that for any (Borel) set B,

$$P(Z \in B \mid Y_1, Y_2, .., Y_L, X_1, X_2, \ldots, X_K) = P(Z \in B \mid Y_1, Y_2, .., Y_L).$$

□

Thus, to build random variables $W(k/2^m)$ so that the successive increments are independent Normal$-(0, 2^{-m})$, we may first build $W(j/2^{m-1})$ so that successive increments are independent Normal$-(0, 2^{-m+1})$, then, conditional on these values, *independently* draw the random variables $W((2j+1)/2^m)$ from the conditional distributions (40). This equation implies that, given $W(j/2^m)$ and $W((j+1)/2^m)$,

$$W((2j+1)/2^{m+1}) = \frac{1}{2}(W(^m) + W((j+1)/2^m)) + 2^{-(m+2)/2}\xi_{m,j}$$

where the random variables $\xi_{m,j}$ (where $m \geq 1$ and $1 \leq j \leq 2^m$) are independent standard normals. (You should check that you understand how this follows from Lemmas 2 – 3.) The upshot is that we now have an inductive procedure for constructing the random variables W_t at dyadic rational t from countably many independent standard normal random variables.

Now we want to show that these fit together to give a continuous path W_t defined for all $t \in [0, 1]$. To do this, let's use the mth level random variables $W(k/2^m)$ to build a polygonal function $\{W_m(t)\}_{t\in[0,1]}$ by connecting the successive dots $(k/2^m, W(k/2^m))$, where $k = 0, 1, \ldots, 12^m$. Observe that W_{m+1} is obtained from $W_m(t)$ by adding "hat functions" of height $2^{-(m+2)/2}\xi_{m,k}$ in each interval $[k/2^m, (k+1)/2^m]$. Formally,

$$W_0(t) = tW(1) \quad \text{and} \quad W_{m+1}(t) = W_m(t) + \sum_{k=1}^{2^m} \xi_{m,k} G_{m,k}(t)/2^{(m+2)/2}$$

where $G_{m,k}(t)$ are the *Schauder functions* (i.e., hat functions)

$$G_{m,k}(t) = 2^{m+1}t - (k-1) \qquad \text{for } (k-1)/2^{m+1} \leq t \leq k/2^{m+1}; \tag{41}$$
$$= k + 1 - 2^{m+1}t \qquad \text{for } k/2^{m+1} \leq t \leq (k+1)/2^{m+1};$$
$$= 0 \qquad \text{otherwise.}$$

Theorem 3. *(Lévy) If the random variables $Z_{m,k}$ are independent, identically distributed with common distribution $N(0,1)$, then with probability one, the infinite series*

$$W(t) := Z_{0,1}t + \sum_{m=1}^{\infty}\sum_{k=1}^{2^m} Z_{m+1,k}G_{m,k}(t)/2^{(m+2)/2} \tag{42}$$

converges uniformly for $0 \leq t \leq 1$. The limit function $W(t)$ is a standard Wiener process.

Proof. Each hat $G_{m,k}$ has maximum value 1, so to show that the series converges uniformly it would be enough to show that for some constant $C < \infty$ not depending on m,

$$\max_k |\xi_{m,k}| \leq C$$

because then the series $\sum_{m=1}^{\infty}$ would be dominated by the geometric series $C\sum 2^{-(m+2)/2}$. Unfortunately, this isn't true. What we *will* show is that with probability one there is a (possibly random) m_* such that

$$\max_k |\xi_{m,k}| \leq 2^{m/4} \quad \text{for all } \; m \geq m_*. \tag{43}$$

This will imply that almost surely the series is *eventually* dominated by a multiple of the geometric series $\sum 2^{-(m+2)/4}$, and consequently converges uniformly in t.

To prove that (43) holds eventually, we will use the Borel-Cantelli Lemma. First, for all m, k,

$$P\{|\xi_{m,k}| \geq 2^{(m+2)/4}\} \leq \frac{2}{\sqrt{2\pi}}\int_{2^{m/2}}^{\infty} e^{-2^{m/2}x/2}\,dx \leq \sqrt{2/\pi}e^{-2^{m-1}}/2^{m/2}.$$

Hence, by the Bonferroni inequality (i.e., the crude union bound),

$$P\{\max_{1\leq k\leq 2^m} |\xi_{m,k}| \geq 2^{(m+2)/4}\} \leq 2^{m/2}\sqrt{2/\pi}e^{-2^{m-1}}.$$

Since this bound is summable in m, Borel-Cantelli implies that with probability 1, eventually (43) must hold. This proves that w.p.1 the series (42) converges uniformly, and therefore $W(t)$ is continuous.

It remains only to show that properties (3)-(4) of Definition 1 are satisfied, that is, that for all $0 = t_0 < t_1 < \cdots < t_K = 1$ the increments $W(t_{j+1} - W(t_j))$ are independent and normally distributed with mean 0 and variances $t_{j+1} - t_j$. Now this is true for *dyadic rational* times t_j, by construction. But because the paths $t \mapsto W_t$ are continuous, the increments can be arbitrarily well-approximated by increments at dyadic times. Since limits of Gaussians are Gaussian, properties (3)-(4) of Definition 1 follow. □

6. Quadratic Variation

Fix $t > 0$, and let $\Pi = \{t_0, t_1, t_2, \ldots, t_n\}$ be a *partition* of the interval $[0,t]$, that is, an increasing sequence $0 = t_0 < t_1 < t_2 < \cdots < t_n = t$. The *mesh* of a partition Π is the length of its longest interval $t_i - t_{i-1}$. If Π is a partition of $[0,t]$ and if $0 < s < t$, then the *restriction* of Π to $[0,s]$ (or the restriction to $[s,t]$) is defined in the obvious way: just terminate the sequence t_j at the largest entry before s, and append s. Say that a partition Π'

is a *refinement* of the partition Π if the sequence of points t_i that defines Π is a subsequence of the sequence t'_j that defines Π'. A *nested sequence* of partitions is a sequence Π_n such that each is a refinement of its predecessor. For any partition Π and any continuous-time stochastic process X_t, define the *quadratic variation* of X relative to Π by

$$QV(X;\Pi) = \sum_{j=1}^{n}(X(t_j) - X(t_{j-1}))^2. \tag{44}$$

Theorem 4. *Let Π_n be a nested sequence of partitions of the unit interval $[0,1]$ with mesh $\to 0$ as $n \to \infty$. Let W_t be a standard Wiener process. Then with probability one,*

$$\lim_{n\to\infty} QV(W;\Pi_n) = 1. \tag{45}$$

Note 1. It can be shown, without too much additional difficulty, that if Π_n^t is the restriction of Π_n to $[0,t]$ then with probability one, for all $t \in [0,1]$,

$$\lim_{n\to\infty} QV(W;\Pi_n^t) = t.$$

Before giving the proof of Theorem 4, I'll discuss a much simpler special case[3], where the reason for the convergence is more transparent. For each natural number n, define the nth *dyadic partition* $\mathcal{D}_n[0,t]$ to be the partition consisting of the *dyadic rationals* $k/2^n$ of depth n (here k is an integer) that are between 0 and t (with t added if it is not a dyadic rational of depth n). Let $X(s)$ be any process indexed by s.

Proposition 11. *Let $\{W(t)\}_{t\geq 0}$ be a standard Brownian motion. For each $t > 0$, with probability one,*

$$\lim_{n\to\infty} QV(W;\mathcal{D}_n[0,t]) = t. \tag{46}$$

Proof. Proof of Proposition 11. First let's prove convergence in probability. To simplify things, assume that $t = 1$. Then for each $n \geq 1$, the random variables

$$\xi_{n,k} \stackrel{\Delta}{=} 2^n(W(k/2^n) - W((k-1)/2^n))^2, \qquad k = 1, 2, \dots, 2^n$$

are independent, identically distributed χ^2 with one degree of freedom (that is, they are distributed as the square of a standard normal random variable). Observe that $E\xi_{n,k} = 1$. Now

$$QV(W;\mathcal{D}_n[0,1]) = 2^{-n}\sum_{k=1}^{2^n}\xi_{n,k}.$$

The right side of this equation is the average of 2^n independent, identically distributed random variables, and so the Weak Law of Large Numbers implies convergence in probability to the mean of the χ^2 distribution with one degree of freedom, which equals 1.

The stronger statement that the convergence holds with probability one can easily be deduced from the Chebyshev inequality and the Borel–Cantelli lemma. The Chebyshev inequality and Brownian scaling implies that

$$P\{|QV(W;\mathcal{D}_n[0,1]) - 1| \geq \varepsilon\} = P\{|\sum_{k=1}^{2^n}(\xi_{n,k} - 1)| \geq 2^n\varepsilon\} \leq \frac{E\xi_{1,1}^2}{4^n\varepsilon^2}.$$

[3]Only the special case will be needed for the Itô calculus. However, it will be of crucial importance — it is, in essence the basis for the *Itô formula*.

Since $\sum_{n=1}^{\infty} 1/4^n < \infty$, the Borel–Cantelli lemma implies that, with probability one, the event $|QV(W; \mathcal{D}_n[0,1]) - 1| \geq \varepsilon$ occurs for at most finitely many n. Since $\varepsilon > 0$ can be chosen arbitrarily small, it follows that $\lim_{n\to\infty} QV(W; \mathcal{D}_n[0,1]) = 1$ almost surely. The same argument shows that for any dyadic rational $t \in [0,1]$, the convergence (46) holds a.s.

Exercise 9. Prove that if (46) holds a.s. for each dyadic rational in the unit interval, then with probability one it holds for all t.

□

In general, when the partition Π is not a dyadic partition, the summands in the formula (45) for the quadratic variation are (when $X = W$ is a Wiener process) still independent χ^2 random variables, but they are no longer identically distributed, and so Chebyshev's inequality by itself won't always be good enough to prove a.s. convergence. The route we'll take is completely different: we'll show that for nested partitions Π_n the sequence $QV(W; \Pi_n)$ is a *reverse martingale* relative to an appropriate filtration $\mathcal{G}_n$.

Lemma 4. *Let ξ, ζ be independent Gaussian random variables with means 0 and variances $\sigma_\xi^2, \sigma_\zeta^2$, respectively. Let $\mathcal{G}$ be any $\sigma-$algebra such that the random variables ξ^2 and ζ^2 are $\mathcal{G}-$measurable, but such that* $\operatorname{sgn}(\zeta)$ *and* $\operatorname{sgn}(\zeta)$ *are independent of $\mathcal{G}$. Then*

$$E((\xi + \zeta)^2 \mid \mathcal{G}) = \xi^2 + \zeta^2. \tag{47}$$

Proof. Expand the square, and use the fact that ξ^2 and ζ^2 are $\mathcal{G}-$measurable to extract them from the conditional expectation. What's left is

$$\begin{aligned} 2E(\xi\zeta \mid \mathcal{G}) &= 2E(\operatorname{sgn}(\xi)\operatorname{sgn}(\zeta)|\xi|\,|\zeta| \mid \mathcal{G}) \\ &= 2|\xi|\,|\zeta| E(\operatorname{sgn}(\xi)\operatorname{sgn}(\zeta) \mid \mathcal{G}) \\ &= 0, \end{aligned}$$

because $\operatorname{sgn}(\xi)$ and $\operatorname{sgn}(\zeta)$ are independent of $\mathcal{G}$. □

Proof of Theorem 4. Without loss of generality, we may assume that each partition Π_{n+1} is gotten by splitting one interval of Π_n, and that Π_0 is the trivial partition of $[0,1]$ (consisting of the single interval $[0,1]$). Thus, Π_n consists of $n+1$ nonoverlapping intervals $J_k^n = [t_{k-1}^n, t_k^n]$. Define

$$\xi_{n,k} = W(t_k^n) - W(t_{k-1}^n),$$

and for each $n \geq 0$ let $\mathcal{G}_n$ be the $\sigma-$algebra generated by the random variables $\xi_{m,k}^2$, where $m \geq n$ and $k \in [m+1]$. The $\sigma-$algebras $\mathcal{G}_n$ are decreasing in n, so they form a *reverse filtration*. By Lemma 4, the random variables $QV(W; \Pi_n)$ form a reverse martingale relative to the reverse filtration $\mathcal{G}_n$, that is, for each n,

$$E(QV(W; \Pi_n) \mid \mathcal{G}_{n+1}) = QV(W; \Pi_{n+1}).$$

By the reverse martingale convergence theorem,

$$\lim_{n\to\infty} QV(W; \Pi_n) = E(QV(W; \Pi_0) \mid \cap_{n\geq 0} \mathcal{G}_n) = E(W_1^2 \mid \cap_{n\geq 0} \mathcal{G}_n) \quad \text{almost surely.}$$

Exercise 10. Prove that the limit is constant a.s., and that the constant is 1.

□

7. SKOROHOD'S THEOREM

In section 2.3 we showed that there are simple random walks embedded in the Wiener path. Skorohod discovered that *any* mean zero, finite variance random walk is also embedded. To prove this it suffices (in view of the strong Markov property) to show that for any mean 0, finite variance distribution F there is a stopping time T such that W_T has distribution F.

Theorem 5. *Let F be any probability distribution on the real line with mean 0 and variance $\sigma^2 < \infty$, and let $W(t)$ be a standard Wiener process. There is a stopping time T (for the standard filtration) with expectation $ET = \sigma^2$ such that the random variable $W(T)$ has distribution F.*

Proof for finitely supported distributions. Suppose first that F is supported by two points $a < 0 < b$, to which it attaches positive probabilities p, q. Since F has mean 0, it must be that $pa + qb = 0$. Define T to be the first time that W reaches either a or b; then Wald's first identity implies that $p = P\{W_T = a\}$ and $q = P\{W_T = b\}$ (Exercise!). Thus, two-point distributions are embedded.

The general case of a probability distribution F with finite support can now be proved by induction on the number of points in the support. To see how the induction step goes, consider the case of a three-point distribution F which puts probabilities p_a, p_b, p_c on the points $a < 0 \leq b < c$. (Note: Since the distribution F has mean 0, there must be at least one point in the support on either side of 0.) Define d to be the weighted average of b and c:

$$d = \frac{bp_b + cp_c}{p_b + p_c}.$$

Now define two stopping times: (A) Let ν be the first time that W reaches either a or d. (B) If $W_\nu = a$, let $T = \nu$; but if $W_\nu = d$, then let T be the first time after ν that W reaches either b or c.

Exercise 11. (A) Check that the distribution of W_T is F.
(B) Check that $ET = \sigma^2$.
(C) Complete the induction step.

$\square$

Proof for the uniform distribution on (-1,1). The general case is proved using the special case of finitely supported distributions by taking a limit. I'll do only the special case of the uniform distribution on $[-1, 1]$. Define a sequence of stopping times τ_n as follows:

$$\tau_1 = \min\{t > 0 \,:\, W(t) = \pm 1/2\}$$
$$\tau_{n+1} = \min\{t > \tau_n \,:\, W(t) - W(\tau_n) = \pm 1/2^{n+1}\}.$$

By symmetry, the random variable $W(\tau_1)$ takes the values $\pm 1/2$ with probabilities $1/2$ each. Similarly, by the Strong Markov Property and induction on n, the random variable $W(\tau_n)$ takes each of the values $k/2^n$, where k is an *odd* number between -2^n and $+2^n$, with probability $1/2^n$. Notice that these values are equally spaced in the interval $[-1, 1]$, and that as $\to \infty$ the values fill the interval. Consequently, the distribution of $W(\tau_n)$ converges to the uniform distribution on $[-1, 1]$ as $n \to \infty$.

The stopping times τ_n are clearly increasing with n. Do they converge to a finite value? Yes, because they are all bounded by $T_{-1,1}$, the first passage time to one of the values ± 1. (Exercise: Why?) Consequently, $\tau := \lim \tau_n = \sup \tau_n$ is finite with probability one. By path-continuity, $W(\tau_n) \to W(\tau)$ almost surely. As we have seen, the distributions of the random variables $W(\tau_n)$ approach the uniform distribution on $[-1, 1]$ as $n \to \infty$, so it follows that the random variable $W(\tau)$ is uniformly distributed on $[-1, 1]$. $\square$

Exercise 12. Show that if τ_n is an increasing sequence of stopping times such that $\tau = \lim \tau_n$ is finite with probability one, then τ is a stopping time.