

BROWNIAN MOTION

1. BROWNIAN MOTION: DEFINITION

Definition 1. A standard Brownian (or a standard Wiener process) is a stochastic process $\{W_t\}_{t \geq 0}$ (that is, a family of random variables W_t , indexed by nonnegative real numbers t , defined on a common probability space (Ω, \mathcal{F}, P)) with the following properties:

- (1) $W_0 = 0$.
- (2) With probability 1, the function $t \rightarrow W_t$ 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 $\text{NORMAL}(0, t)$ distribution.

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 $W_t - W_0 = W_t$.

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. His proof was simplified by P. LÉVY; we shall outline Lévy's construction in section 10 below. But notice that properties (3) and (4) are compatible. This follows from the following elementary property of the normal distributions: If X, Y are independent, normally distributed random variables with means μ_X, μ_Y and variances σ_X^2, σ_Y^2 , then the random variable $X + Y$ is normally distributed with mean $\mu_X + \mu_Y$ and variance $\sigma_X^2 + \sigma_Y^2$.

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 ξ_1, ξ_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

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

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 $\text{NORMAL}(0, t)$ distribution, by the Central Limit theorem. Thus, it requires only a small leap of faith to believe that, as $n \rightarrow \infty$, the

distribution of the random function $W_n(t)$ approaches (in a certain sense)¹ 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

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

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

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

Similarly, for any $a > 0$ the distribution of

$$(4) \quad \tau_n(a) := \min\{t \geq 0 : W_n(t) \geq a\}$$

approaches, as $n \rightarrow \infty$, that of

$$(5) \quad \tau(a) := \min\{t \geq 0 : W(t) = a\}.$$

The distributions of $M(t)$ and $\tau(a)$ will be calculated below.

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:

$$(6) \quad P(W_{t+s} \in dy | W_s = x) \triangleq p_t(x, y) dy = \frac{1}{\sqrt{2\pi t}} \exp\{-(y-x)^2/2t\} dy.$$

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

$$(7) \quad \frac{\partial u}{\partial t} = \frac{1}{2} \frac{\partial^2 u}{\partial x^2}$$

$$(8) \quad u_0(x) = f(x)$$

¹For a formal definition of convergence in distribution of random functions, together with detailed proofs and many examples and statistical applications, see BILLINGSLEY, *Weak Convergence*.

is given by

$$(9) \quad u_t(x) = Ef(W_t^x) = \int_{y=-\infty}^{\infty} p_t(x, y) f(y) dy.$$

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

The equation (7) is called the *heat equation*. That the PDE (7) has only one solution that satisfies the initial condition (8) 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 (9). To see that the right side of (9) actually does solve (7), take the partial derivatives in the PDE (7) under the integral in (9). You then see that the issue boils down to showing that

$$(10) \quad \frac{\partial p_t(x, y)}{\partial t} = \frac{1}{2} \frac{\partial^2 p_t(x, y)}{\partial x^2}.$$

Exercise: Verify this.

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:

$$(11) \quad \{-W(t)\}_{t \geq 0}$$

$$(12) \quad \{W(t+s) - W(s)\}_{t \geq 0}$$

$$(13) \quad \{aW(t/a^2)\}_{t \geq 0}$$

$$(14) \quad \{tW(1/t)\}_{t \geq 0}.$$

Exercise: Prove this.

These properties have important ramifications. The most basic of these involve the *maximum* and *minimum* random variables $M(t)$ and $M^-(t)$ defined by

$$(15) \quad M(t) := \max\{W(s) : 0 \leq s \leq t\} \quad \text{and}$$

$$(16) \quad M^-(t) := \min\{W(s) : 0 \leq s \leq t\}$$

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:

$$(17) \quad M(t) \stackrel{\mathcal{D}}{=} -M^-(t).$$

Property (13) is called the *Brownian scaling* property. It is perhaps the most useful elementary tool in the study of the Wiener process. As a first example, consider its implications for the distributions of the maximum random variables $M(t)$. Fix $a > 0$, and define

$$W^*(t) = aW(t/a^2) \quad \text{and}$$

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

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,

$$(18) \quad M(t) \stackrel{\mathcal{D}}{=} aM(t/a^2).$$

On first sight, this relation appears rather harmless. However, as we shall see in section 7, 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:

$$(19) \quad \tau(a) = \min\{t : W(t) = a\}$$

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

5. THE BROWNIAN FILTRATION AND THE MARKOV PROPERTY

Property (12) 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 σ -algebras. A σ -algebra is by definition a collection of events that includes the empty event \emptyset and is closed under complements and countable unions. Define

$$(20) \quad \mathcal{F}_t := \sigma(\{W(s)\}_{0 \leq s \leq t})$$

to be the σ -algebra consisting of all events that are “observable” by time t . Formally, \mathcal{F}_t is defined to be the smallest σ -algebra containing all events of the form $\{W(r) \leq a\}$, where $a \in \mathbb{R}$ and $r \leq s$. The indexed collection of σ -algebra $\{\mathcal{F}_t\}_{t \geq 0}$ is called the *Brownian filtration*, or the *standard filtration*.

Example: For each $t > 0$ and for every $a \in \mathbb{R}$, the event $\{M(t) > a\}$ is an element of \mathcal{F}_t . To see this, observe that by path-continuity,

$$(21) \quad \{M(t) > a\} = \bigcup_{s \in \mathbb{Q}; 0 \leq s \leq t} \{W(s) > a\}.$$

Here \mathbb{Q} denotes the set of rational numbers. Because \mathbb{Q} is a countable set, the union in (21) is a countable union. Since each of the events $\{W(s) > a\}$ in the union is an element of the σ -algebra \mathcal{F}_t , the event $\{M(t) > a\}$ must also be an element of \mathcal{F}_t .

Proposition 2. (*Markov Property*) *Let $\{W(t)\}_{t \geq 0}$ be a standard Brownian motion, $\{\mathcal{F}_t\}_{t \geq 0}$ the standard filtration, and for $s > 0$, $t \geq 0$ define $W^*(t) = W(t+s) - W(s)$ and let $\{\mathcal{F}_t^*\}_{t \geq 0}$ be its filtration. Then for each $t > 0$ the σ -algebras $\mathcal{F}(s)$ and $\mathcal{F}^*(t)$ are independent.*

Corollary 1. *The random variables $M(s)$ and $M^*(t) = \max_{0 \leq r \leq t} W^*(r)$ are independent.*

Proof of the Markov Property. The Markov Property is nothing more than a sophisticated restatement of the independent increments property of Brownian motion. One first uses independent increments to show that certain types of events in the two σ -algebras are independent: in particular,

$$A = \bigcap_{j=1}^n \{W(s_j) - W(s_{j-1}) \leq x_j\} \in \mathcal{F}_s \quad \text{and} \\ B = \bigcap_{j=1}^m \{W(t_j + s) - W(t_{j-1} + s) \leq y_j\} \in \mathcal{F}_t^*$$

are independent. Events of type A generate the σ -algebra \mathcal{F}_s , and events of type B generate the σ -algebra \mathcal{F}_t^* (by definition!). A standard approximation procedure in measure theory (based on the so-called “ $\pi - \lambda$ ” theorem — see BILLINGSLEY, *Probability and Measure*) now allows one to conclude that the σ -algebras $\mathcal{F}(s)$ and $\mathcal{F}^*(t)$ are independent.

The Markov property has an important generalization called the *Strong Markov Property*. This generalization involves the notion of a *stopping time* for the Brownian filtration. A stopping time is defined to be a nonnegative random variable τ such that for each (nonrandom) $t \geq 0$ the event $\{\tau \leq t\}$ is an element of the σ -algebra \mathcal{F}_t (that is, if the event $\{\tau \leq t\}$ is “determined by” the path $\{W(s)\}_{s \leq t}$ up to time t).

Example: $\tau(a) := \min\{t : W(t) = a\}$ is a stopping time. 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 .

Associated with any stopping time τ is a σ -algebra \mathcal{F}_τ , defined to be the collection of all events B such that $B \cap \{\tau \leq t\} \in \mathcal{F}_t$. Informally, \mathcal{F}_τ consists of all events that are “observable” by time τ .

Example: Let $\tau = \tau_a$ as above, and let B be the event that the Brownian path $W(t)$ hits b before it hits a . Then $B \in \mathcal{F}_\tau$.

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 σ -algebra \mathcal{F}_τ . For $t \geq 0$, define*

$$(22) \quad W^*(t) = W(t + \tau) - W(\tau),$$

and let $\{\mathcal{F}_t^*\}_{t \geq 0}$ be the filtration of the process $\{W^*(t)\}_{t \geq 0}$. Then

- (a) $\{W^*(t)\}_{t \geq 0}$ is a standard Brownian motion; and
- (b) For each $t > 0$, the σ -algebra \mathcal{F}_t^* is independent of \mathcal{F}_τ .

Details of the proof are omitted (see, for example, KARATZAS & SHREVE, pp. 79ff). Let’s discuss briefly the meaning of the statement. In essence, the theorem says that the post- τ process $W^*(t)$ is itself a standard Brownian motion, and that it is independent of everything that happened up to time τ . Thus, in effect, the Brownian motion “begins anew” at time τ , paying no further attention to what it did before τ . Note that simple random walk (the discrete-time process in which, at each time n one tosses a fair coin to decide whether to move up or down one unit) has an analogous property. If, for instance, one waits until first arriving at some integer k and then continues tossing the coin, the coin tosses after the first arrival at k are independent of the tosses prior to the first arrival. This is not difficult to show:

Exercise: Formulate and prove a Strong Markov Property for simple random walk.

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,

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

6. THE REFLECTION PRINCIPLE AND FIRST-PASSAGE TIMES

Proposition 3.

$$(23) \quad P\{M(t) \geq a\} = P\{\tau_a < t\} = 2P\{W(t) > a\} = 2 - 2\Phi(a/\sqrt{t}).$$

Proof. The argument is based on a symmetry principle that may be traced back to the French mathematician D. ANDRÉ, and 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 claim² requires the use of the Strong Markov Property. First, observe that $\tau_a \wedge t$ is a stopping time. Thus, by the strong Markov property, the post- $\tau_a \wedge t$ process is a standard Brownian motion independent of the path up to time $\tau_a \wedge t$ (and, therefore, independent of $\tau_a \wedge t$). It follows that, for any $s < t$, the *conditional* distribution of $W(t) - W(s)$ given that $\tau_a = s$ is Gaussian with mean 0 and variance $t - s > 0$, and so, by the symmetry of the Gaussian distribution,

$$P(W(t) - W(\tau_a) > 0 | \tau_a = s) = P(W(t) - W(\tau_a) < 0 | \tau_a = s) = 1/2.$$

Integration over $0 < s < t$ against the distribution of $\tau(a)$ gives

$$P\{W(t) - W(\tau_a) > 0 \text{ and } \tau_a < t\} = \frac{1}{2}P\{\tau_a < t\}.$$

But the event $\{W(t) - W(\tau_a) > 0\} \cap \{\tau_a < t\}$ coincides with the event $\{W(t) > a\}$, because (i) if $\tau_a < t$ then, since $W(\tau_a) = a$, the inequality $W(t) - W(\tau_a) > 0$ implies $W(t) > a$; and (ii) if $W(t) > a$ then the Intermediate Value Theorem of calculus and the path-continuity of $W(s)$ implies that $\tau_a < t$. This proves that

$$P\{\tau_a < t\} = 2P\{W(t) > a\} = 2(1 - \Phi(a/\sqrt{t})).$$

Because the expression on the right side of this equation is a continuous function of t , it follows that $P\{\tau(a) < t\} = P\{\tau(a) \leq t\}$. Finally, since the events $\{\tau(a) \leq t\}$ and $\{M(t) \geq a\}$ are the same, we have $P\{\tau(a) \leq t\} = P\{M(t) \geq a\}$. \square

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

$$(24) \quad f(t) = \frac{ae^{-a^2/2t}}{\sqrt{2\pi t^3}}.$$

²Many writers of elementary textbooks, including ROSS, seem to think that no justification is needed; or that readers of their books need not be troubled by the need to justify it.

There is a more sophisticated form of the reflection principle that is useful in certain calculations. Set $\tau = \tau(a)$, where $\tau(a)$ is the first passage time to the value a . The random variable τ is a stopping time, and we have now shown that it is finite with probability one. By the Strong Markov Property, the post- τ process

$$(25) \quad W^*(t) := W(\tau + t) - W(\tau)$$

is a Wiener process, and is independent of the stopping field \mathcal{F}_τ . Consequently, $\{-W^*(t)\}_{t \geq 0}$ is also a Wiener process, and is also independent of the stopping field \mathcal{F}_τ . 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:

$$(26) \quad \begin{aligned} \tilde{W}(s) &= W(s) && \text{for } s \leq \tau, \\ &= 2a - W(s) && \text{for } s \geq \tau. \end{aligned}$$

Proposition 4. (*Reflection Principle*) *If $\{W(t)\}_{t \geq 0}$ is a Wiener process, then so is $\{\tilde{W}(t)\}_{t \geq 0}$.*

Proof. An exercise for the interested reader. □

The analogous property for the simple random walk is worth noting. Simple random walk is performed by repeatedly tossing a fair coin, moving one step to the right on every H, and one step to the left on every T. The simple random walk with reflection in the level a , for an integer value a , is gotten by changing the law of motion at the time of first passage to a : after this time, one moves one step to the *left* on every H, and one step to the *right* on every T. It is fairly obvious (and easy to prove) that this modification does not change the “statistics” of the random walk.

The Reflection Principle for Brownian motion, as formalized in Proposition 4, allows one to calculate the *joint* distribution of $M(t)$ and $W(t)$:

Corollary 3.

$$(27) \quad P\{M(t) \geq a \text{ and } W(t) \leq a - b\} = P\{W(t) \geq a + b\} \quad \forall a, b > 0.$$

Proof. Because the paths $W(s)$ and $\tilde{W}(s)$ coincide up to time τ , the event that $M(t) \geq a$ coincides with the event $\tilde{M}(t) \geq a$, where $\tilde{M}(t)$ is defined to be the maximum of the path $\tilde{W}(s)$ for $0 \leq s \leq t$. Thus, by (26),

$$\{M(t) \geq a \text{ and } W(t) \leq a - b\} = \{\tilde{M}(t) \geq a \text{ and } \tilde{W}(t) \geq a + b\} = \{\tilde{W}(t) \geq a + b\}.$$

The Reflection Principle implies that the events $\{\tilde{W}(t) \geq a + b\}$ and $\{W(t) \geq a + b\}$ have the same probability, and so the corollary follows. □

Corollary 4.

$$(28) \quad P\{M(t) \in da \text{ and } W(t) \in a - db\} = \frac{2(a+b) \exp\{-(a+b)^2/2t\}}{(2\pi)^{1/2} t} da db$$

7. BEHAVIOR OF 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 (more on this later). 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. Complete proofs of these facts are beyond the scope of these notes. However, some closely related facts may be established using only the formula (23) and elementary arguments.

7.1. Zero Set of a Brownian Path. The zero set is

$$(29) \quad \mathcal{Z} = \{t \geq 0 : W(t) = 0\}.$$

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,

$$(30) \quad P\{\mathbb{Q}_+ \cap \mathcal{Z} \neq \emptyset\} = 0.$$

Proposition 5. *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 (11)) that the distribution of $M^-(t)$, the minimum up to time t , is the same as that of $-M(t)$. By formula (23), 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 6. *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 $\lim_{n \rightarrow \infty} t_n = t$.*

Proof. That \mathcal{Z} is closed follows from path-continuity, as noted earlier. Fix a rational number $q > 0$ (nonrandom), and define $v = v_q$ to be the first time $t \geq q$ such that $W(t) = 0$. Because $W(q) \neq 0$ almost surely, the random variable v_q is well-defined and is almost surely strictly greater than q . By the Strong Markov Property, the post- v_q process $W(v_q + t) - W(v_q)$ is, conditional on the stopping field \mathcal{F}_{v_q} , a Wiener process, and consequently, by Proposition 5, it has infinitely many zeros in every time interval $(0, \varepsilon)$, with probability one. Since $W(v_q) = 0$, 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 $(v_q, v_q + \varepsilon)$, 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 $t_n \rightarrow t$, 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 $t = v_q$. In this case, by the preceding paragraph, there must be a *decreasing* sequence t_n of zeros such that $t_n \rightarrow t$. \square

It may be shown that every compact perfect set of Lebesgue measure zero is homeomorphic to the Cantor set. (This is not especially difficult – consult your local mathematician for assistance.) 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.

7.2. Nondifferentiability of Paths.

Proposition 7. *For each $t \geq 0$, the Brownian path is almost surely not differentiable at t .*

Note: It follows that, with probability one, the Brownian path is not differentiable at any *rational* time $t \geq 0$. DVORETSKY, ERDÖS, and KAKUTANI proved an even stronger statement: with probability one, the Brownian path is *nowhere* differentiable. This, like Proposition 7, is ultimately a consequence of Brownian scaling.

Proof. By the Markov property, it suffices to prove that the Brownian path is almost surely not differentiable at $t = 0$. Suppose that it were: then the difference quotients $W(t)/t$ would be bounded for t in some neighborhood of 0, that is, for some $A < \infty$ and $\varepsilon > 0$ it would be the case that

$$W(t) < At \quad \text{for all } 0 \leq t < \varepsilon.$$

Fix an integer $A > 0$. By formula (23),

$$P\{M(t) < At\} = 2 - 2\Phi(At/\sqrt{t}) = 2 - 2\Phi(A\sqrt{t}) \rightarrow 0$$

as $t \rightarrow 0$. Consequently, for each fixed $A > 0$, the probability that $W(t) < At$ for all $t \leq \varepsilon$ is *zero*. Since the union of countably many events of probability zero has probability zero, it follows that for any $\varepsilon > 0$ the event that $W(t)/t$ remains bounded for $0 \leq t < \varepsilon$ has probability zero. \square

8. WALD IDENTITIES FOR BROWNIAN MOTION

Proposition 8. *Let $\{W(t)\}_{t \geq 0}$ be a standard Wiener process, and let τ be a bounded stopping time. Then*

$$(31) \quad EW(\tau) = 0;$$

$$(32) \quad EW(\tau)^2 = E\tau;$$

$$(33) \quad E \exp\{\theta W(\tau) - \theta^2 \tau / 2\} = 1 \quad \forall \theta \in \mathbb{R}; \text{ and}$$

$$(34) \quad E \exp\{i\theta W(\tau) + \theta^2 \tau / 2\} = 1 \quad \forall \theta \in \mathbb{R}.$$

Observe that for *nonrandom* times $\tau = t$, these identities follow from elementary properties of the normal distribution. Notice also that 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 $EW(\tau) = EM(1) > 0$.

Proof. By hypothesis, τ is a bounded stopping time, so there is a nonrandom $N < \infty$ such that $\tau < N$ almost surely. By the Strong Markov Property, the post- τ process $W(\tau + t) - W(\tau)$ is a standard Wiener process that is independent of the stopping field \mathcal{F}_τ , and, in particular, independent of the random vector $(\tau, W(\tau))$. Hence, the conditional distribution of $W(N) - W(\tau)$ given that $\tau = s$ is the normal distribution with mean 0 and variance $N - s$. It follows that

$$E(\exp\{\theta(W(N) - W(\tau)) - \theta^2(N - \tau)\} | W(\tau), \tau) = 1$$

Therefore,

$$\begin{aligned} Ee^{\theta W(\tau) - \theta^2 \tau / 2} &= E \exp\{\theta W(\tau) - \theta^2 \tau / 2\} \cdot 1 \\ &= E \exp\{\theta W(\tau) - \theta^2 \tau / 2\} E(\exp\{\theta(W(N) - W(\tau)) - \theta^2(N - \tau)\} | W(\tau), \tau) \\ &= EE(\exp\{\theta W(\tau) - \theta^2 \tau / 2\} \exp\{\theta(W(N) - W(\tau)) - \theta^2(N - \tau)\} | W(\tau), \tau) \\ &= EE(\exp\{\theta W(N) - \theta^2 N / 2\} | W(\tau), \tau) \\ &= E \exp\{\theta W(N) - \theta^2 N / 2\} \\ &= 1. \end{aligned}$$

The other three identities can be established in a similar fashion. □

Example 1: 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,

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

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, $W(T \wedge n) \rightarrow W(T)$ as $n \rightarrow \infty$. 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

$$(35) \quad \boxed{P\{W(T) = b\} = \frac{a}{a+b}.$$

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

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

Using (35), one may now easily obtain

$$(36) \quad \boxed{ET = ab.}$$

Example 2: 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 $\theta > 0$,

$$(37) \quad E \exp\{\theta W(\tau \wedge n) - \theta^2(\tau \wedge n)\} = 1.$$

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 (37) are dominated by the constant $\exp\{\theta a\}$. Since $\tau < \infty$ with probability one, $\tau \wedge n \rightarrow \tau$ as $n \rightarrow \infty$, and by path-continuity, the random variables $W(\tau \wedge n)$ converge to a as $n \rightarrow \infty$. Therefore, by the dominated convergence theorem,

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

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

$$(38) \quad \boxed{E \exp\{-\lambda \tau a\} = \exp\{-\sqrt{2\lambda} a\}.}$$

The only density with this Laplace transform³ is the one-sided stable density given in equation (24). Thus, the Optional Sampling Formula gives us a second proof of (23).

9. SKOROHOD'S THEOREM

The calculation in Example 1 of the preceding section shows that there are *Rademacher* random variables⁴ lurking in the Brownian path: if $T = T_{-1,1}$ is the first time that the Wiener path visits one of the values ± 1 then $W(T)$ is a Rademacher random variable. Skorohod discovered that *any* mean zero, finite variance random variable has a replica in the Brownian path.

Theorem 3. *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 with expectation $ET = \sigma^2$ such that the random variable $W(T)$ has distribution F .*

We refrain from giving a complete proof. Instead, we shall discuss the case of the uniform distribution on the interval $[-1, 1]$, as the calculations are simple in this case. Define a sequence of stopping times τ_n as follows:

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

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 $n \rightarrow \infty$ the values fill

³Check a table of Laplace transforms.

⁴A *Rademacher random variable* is a random variable taking the values ± 1 with probabilities $1/2$.

the interval. Consequently, the distribution of $W(\tau_n)$ converges to the uniform distribution on $[-1, 1]$ as $n \rightarrow \infty$.

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, $W(\tau_n) \rightarrow W(\tau)$ almost surely. As we have seen, the distributions of the random variables $W(\tau_n)$ approach the uniform distribution on $[-1, 1]$ as $n \rightarrow \infty$, so it follows that the random variable $W(\tau)$ is uniformly distributed on $[-1, 1]$.

Exercise 1: 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.

Exercise 2: Use one of the Wald identities to show that, for the stopping time τ constructed above,

$$(39) \quad E\tau = \frac{1}{2} \int_{-1}^1 x^2 dx.$$

10. LÉVY'S CONSTRUCTION

Each of the three principal figures in the development of the mathematical theory of Brownian motion – Wiener, Kolmogorov, and Lévy – gave a construction of the process. Of the three, Lévy's is the shortest, and requires least in the way of preparation. It is based on a very simple property of normal distributions:

Lemma 1. *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*

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

Proof. Exercise. □

Lévy's idea was that one could use this fact to build a sequence of successive approximations $W_n(t)$ to the Wiener process $W(t)$, each a (random) polygonal function linear in each of the time intervals $[k/2^n, (k + 1)/2^n]$. This would be done in such a way that the values of $W_n(t)$ at the dyadic rationals $k/2^m$ of level m , where $k = 0, 1, \dots, 2^m$ would remain the same for all $n \geq m$. Thus, the joint distribution of the random variables

$$W_m(0), W_m(1/2^m), W_m(2/2^m), \dots, W_m(1)$$

should be chosen in such a way that their joint distribution is the same as that for the corresponding points on the Wiener path. The trick is that one can arrange for this joint distribution by choosing each $W_m(k/2^m)$, where k is *odd*, from the *conditional* distribution of $W(k/2^m)$ given the values $W((k - 1)/2^m) = W_{m-1}((k - 1)/2^m)$ and $W((k + 1)/2^m) = W_{m-1}((k + 1)/2^m)$. This is where Lemma 1 comes in. It guarantees that the conditional variance does not depend on the values $W_{m-1}((k - 1)/2^m)$ and $W_{m-1}((k + 1)/2^m)$, and that the conditional mean is just the average of these values; in particular, since $W_{m-1}(t)$ is linear between times $(k - 1)/2^m$ and $(k + 1)/2^m$, the conditional mean of $W_m(k/2^m)$ given the values $W_{m-1}((k - 1)/2^m)$ and $W_{m-1}((k + 1)/2^m)$ is $W_{m-1}(k/2^m)$! Hence, to obtain $W_{m+1}(t)$ from $W_m(t)$:

$$(41) \quad W_{m+1}(t) = W_m(t) + \sum_{k=1}^{2^m} Z_{m+1,k} G_{m,k}(t) / 2^{(m+2)/2}$$

where the random variables $Z_{m,k}$ are independent, with the unit normal distribution, and the functions $G_{m,k}(t)$ are the *Schauder functions*:

$$(42) \quad \begin{aligned} G_{m,k}(t) &= 2^{m+1}t - (k-1) && \text{for } (k-1)/2^{m+1} \leq t \leq k/2^{m+1}; \\ &= k+1 - 2^{m+1}t && \text{for } k/2^{m+1} \leq t \leq (k+1)/2^{m+1}; \\ &= 0 && \text{otherwise.} \end{aligned}$$

The construction of the initial approximation $W_0(t)$ is trivial: since the distribution of $W(t)$ is Normal(0, 1), merely set

$$(43) \quad W_0(t) = Z_{0,1}t,$$

where $Z_{0,1}$ is a unit normal.

Theorem 4. (*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*

$$(44) \quad W(t) := Z_{0,1}t + \sum_{m=0}^{\infty} \sum_{k=1}^{2^m} Z_{m+1,k} G_{m,k}(t) / 2^{(m+2)/2}$$

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

11. QUADRATIC VARIATION

Fix $t > 0$, and let $\Pi = \{t_0, t_1, t_2, \dots, t_n\}$ be a *partition* of the interval $[0, t]$, that is, an increasing sequence $0 = t_0 < t_1 < t_2 < \dots < t_n = t$. For each natural number n , define the *n*th *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 . For any partition Π , define the *quadratic variation* of X relative to Π by

$$(45) \quad QV(X; \Pi) = \sum_{j=1}^n (X(t_j) - X(t_{j-1}))^2.$$

Proposition 9. *Let $\{W(t)\}_{t \geq 0}$ be a standard Brownian motion. For each $t > 0$, with probability one,*

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

The primary significance of this result is that it is the key to the *Itô* formula of stochastic calculus, about which we shall have much to say later in the course. It also implies that the Brownian path cannot be of bounded variation in any interval, and so in particular is not differentiable in any interval. (Even more is known: the Brownian path is *nowhere differentiable*.)

Partial Proof. We will only prove the weaker statement that the convergence in (46) holds in probability. To simplify things, assume that $t = 1$. Then for each $n \geq 1$, the random variables

$$\xi_{n,k} \triangleq 2^n (W(k/2^n) - W((k-1)/2^n))^2, \quad k = 1, 2, \dots, 2^n$$

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 implies that

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

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 $\lim_{n \rightarrow \infty} QV(W; \mathcal{D}_n[0, 1]) = 1$ almost surely.