Brownian Motion

1 Brownian motion: existence and first properties

1.1 Definition of the Wiener process

According to the *De Moivre-Laplace* theorem (the first and simplest case of the central limit theorem), the standard normal distribution arises as the limit of scaled and centered Binomial distributions, in the following sense. Let $\xi_1, \xi_2, \ldots$ be independent, identically distributed *Rademacher* random variables, that is, independent random variables with distribution

$$P\{\xi_i = +1\} = P\{\xi_i = -1\} = \frac{1}{2}, \tag{1}$$

and for each $n = 0, 1, 2, \ldots$ let $S_n = \sum_{i=1}^{n} \xi_i$. The discrete-time stochastic process $\{S_n\}_{n \geq 0}$ is the *simple random walk* on the integers. The De Moivre-Laplace theorem states that for each $x \in \mathbb{R}$,

$$\lim_{n \to \infty} P\{S_n/\sqrt{n} \leq x\} = \Phi(x) := \int_{-\infty}^{x} \varphi_1(y)\, dy \quad \text{where} \tag{2}$$

$$\varphi_t(y) = \frac{1}{\sqrt{2\pi t}} e^{-y^2/2t} \tag{3}$$

is the normal (Gaussian) density with mean 0 and variance t.

The De Moivre-Laplace theorem has an important corollary: the family $\{\varphi_t\}_{t \geq 0}$ of normal densities is closed under convolution. To see this, observe that for any $0 < t < 1$ the sum $S_n = S_{[nt]} + (S_n - S_{[nt]})$ is obtained by adding two independent Rademacher sums; the De Moivre-Laplace theorem applies to each sum separately, and so by an elementary scaling we must have $\varphi_1 = \varphi_t * \varphi_{1-t}$. More generally, for any $s, t \geq 0$,

$$\varphi_s * \varphi_t = \varphi_{s+t}. \tag{4}$$

This law can, of course, be proved without reference to the central limit theorem, either by direct calculation ("completing the square") or by Fourier transform.

However, our argument suggests a "dynamical" interpretation of the equation (4) that the more direct proofs obscure. For any finite set of times $0 = t_0 < t_1 < \cdots < t_m < \infty$ there exist (on some probability space) independent, mean-zero Gaussian random variables $W_{t_{i+1}} - W_{t_i}$ with variances $t_{i+1} - t_i$. The De Moivre-Laplace theorem implies that as $n \to \infty$,

$$\frac{1}{\sqrt{n}}(S_{[nt_1]}, S_{[nt_2]}, \ldots, S_{[nt_m]}) \xrightarrow{\mathcal{D}} (W_{t_1}, W_{t_2}, \ldots, W_{t_m}). \tag{5}$$

The convolution law (4) guarantees that the joint distributions of these limiting random vectors are mutually consistent, that is, if the set of times $\{t_i\}_{i \leq m}$ is enlarged by adding more time points, the joint distribution of $W_{t_1}, W_{t_2}, \ldots, W_{t_m}$ will not be changed. This suggests the possibility of defining a continuous-time stochastic process $\{W_t\}_{t \geq 0}$ in which all of the random vectors $(W_{t_1}, W_{t_2}, \ldots, W_{t_m})$ are embedded.

Definition 1. A *standard (one-dimensional) Wiener process* (also called *Brownian motion*) is a continuous-time stochastic process $\{W_t\}_{t \geq 0}$ (i.e., a family of real random variables indexed by the set of nonnegative real numbers t) with the following properties:

 (A) $W_0 = 0$.
 (B) With probability 1, the function $t \to W_t$ is continuous in t.
 (C) The process $\{W_t\}_{t \geq 0}$ has *stationary, independent increments*.
 (D) For each t the random variable W_t has the NORMAL$(0, t)$ distribution.

A continuous-time stochastic process $\{X_t\}_{t \geq 0}$ is said to have *independent increments* if for all $0 \leq t_0 < t_1 < \cdots < t_m$ the random variables $(X_{t_{i+1}} - X_{t_i})$ are mutually independent; it is said to have *stationary increments* if for any $s, t \geq 0$ the distribution of $X_{t+s} - X_s$ is the same as that of $X_t - X_0$. Processes with stationary, independent increments are known as *Lévy processes*.

Properties (C) and (D) are mutually consistent, by the convolution law (4), but it is by no means clear that there exists a stochastic process satisfying (C) and (D) that has continuous sample paths. That such a process *does* exist was first proved by N. WIENER in about 1920.

Theorem 1. (Wiener) *On any probability space $(\Omega, \mathcal{F}, P)$ that supports an infinite sequence of independent, identically distributed Normal$-(0, 1)$ random variables there exists a standard Brownian motion.*

We will give at least one proof, due to P. LÉVY, later in the course. Lévy's proof shows that exhibits the Wiener process as the sum of an almost surely uniformly convergent series. Lévy's construction, like that of Wiener, gives some insight into the mathematical structure of theorem Wiener process but obscures the connection with random walk. There is another approach, however, that makes direct use

of the central limit theorem (5). Unfortunately, this approach requires a bit more technology (specifically, weak convergence theory for measures on metric space; see he book *Weak Convergence of Probability Measures* by P. Billingsley for details) than we have time for in this course. However, it has the advantage that it also explains, at least in part, why the Wiener process is useful in the modeling of natural processes. Many stochastic processes behave, at least for long stretches of time, like random walks with small but frequent jumps. The weak convergence theory shows that such processes will look, at least approximately, and on the appropriate time scale, like Brownian motion.

Notation and Terminology. A *Brownian motion with initial point x* is a stochastic process $\{W_t\}_{t\geq0}$ such that $\{W_t - x\}_{t\geq0}$ is a standard Brownian motion. Unless otherwise specified, *Brownian motion* means *standard Brownian motion*. To ease eyestrain, we will adopt the convention that whenever convenient the index t will be written as a functional argument instead of as a subscript, that is, $W(t) = W_t$.

1.2 Brownian motion and diffusion

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:

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

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 2. *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*

$$\frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial^2 u}{\partial x^2} \tag{7}$$

$$u_0(x) = f(x) \tag{8}$$

is given by

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

3

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 for details. More important (for us) 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

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

Exercise: Verify this.

1.3 Brownian motion in higher dimensions

Definition 2. A *standard $d-$dimensional Brownian motion* is an $\mathbb{R}^d-$valued continuous-time stochastic process $\{W_t\}_{t\geq 0}$ (i.e., a family of $d-$dimensional random vectors W_t indexed by the set of nonnegative real numbers t) with the following properties.

(A)' $W_0 = 0$.
(B)' With probability 1, the function $t \to W_t$ is continuous in t.
(C)' The process $\{W_t\}_{t\geq 0}$ has *stationary, independent increments*.
(D)' The increment $W_{t+s} - W_s$ has the $d-$dimensional normal distribution with mean vector 0 and covariance matrix tI.

The $d-$dimensional normal distribution with mean vector 0 and (positive definite) covariance matrix Σ is the Borel probability measure on $\mathbb{R}^d$ with density

$$\varphi_\Sigma(x) = ((2\pi)^d \det(\Sigma))^{-1/2} \exp\{-x^T \Sigma^{-1} x/2\};$$

if $\Sigma = tI$ then this is just the product of d one-dimensional Gaussian distributions with mean 0 and variance t. Thus, the existence of $d-$dimensional Brownian motion follows directly from the existence of $1-$dimensional Brownian motion: if $\{W^{(i)}\}_{t\geq 0}$ are independent $1-$dimensional Brownian motions then

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

is a d-dimensional Brownian motion. One of the important properties of the $d-$dimensional normal distribution with mean zero and covariance matrix tI proportional to the identity is its *invariance under orthogonal transformations*. This implies that if $\{W_t\}_{t\geq 0}$ is a $d-$dimensional Brownian motion then for any orthogonal transformation U of $\mathbb{R}^d$ the process $\{UW_t\}_{t\geq 0}$ is also a $d-$dimensional Brownian motion.

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:*

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

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

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

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

Exercise: Prove this.

The scaling law (13) is especially important. It is often 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

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

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

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 $-M^-(t)$ <mark>have the same distribution</mark>:

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

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

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

$$M^*(t) = \max_{0 \leq s \leq t} W^*(s)$$

$$= \max_{0 \leq s \leq t} aW(s/a^2)$$

$$= aM(t/a^2).$$

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,

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

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

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

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

1.5 The Wiener Isometry

The existence theorem (Theorem 1) was first proved by N. Wiener around 1920. Simpler proofs have since been found, but Wiener's argument contains the germ of an extremely useful insight, which is now known as the *Wiener isometry* (or, in some of the older literature, the *Wiener integral*). Following is an account of Wiener's line of thought.

Suppose that Brownian motion exists, that is, suppose that on some probability space $(\Omega, \mathcal{F}, P)$ there is a centered Gaussian process $\{W_t\}_{t \in [0,1]}$ with covariance $EW_tW_s = \min(s,t)$. The random variables W_t are all elements of the space $L^2(P)$ consisting of the real random variables defined on $(\Omega, \mathcal{F}, P)$ with finite second moments. This space is a *Hilbert space* with inner product $\langle X, Y \rangle = (X, Y)$.

Now consider the Hilbert space $L^2[0,1]$, consisting of all real-valued square-integrable functions on the unit interval, with inner product

$$\langle f, g \rangle = \int_0^1 f(x)g(x)\,dx.$$

Indicator functions $\mathbf{1}_{[0,t]}$ of intervals $[0,t]$ are elements of $L^2[0,1]$, and obviously

$$\langle \mathbf{1}_{[0,t]}, \mathbf{1}_{[0,s]} \rangle = \min(s,t).$$

Thus, the indicators $\mathbf{1}_{[0,t]}$ have exactly the same inner products as do the random variables W_t in the Hilbert space $L^2(P)$. Wiener's key insight was that this identity between inner products implies that there is a linear isometry I_W from $L^2[0,1]$ into $L^2(P)$ mapping each indicator $\mathbf{1}_{[0,t]}$ to the corresponding random variable W_t.

Theorem 3. *(Wiener's Isometry) Let $\{W_t\}_{t \geq 0}$ be a standard Wiener process defined on a probability space $(\Omega, \mathcal{F}, P)$. Then for any nonempty interval $J \subseteq \mathbb{R}_+$ the mapping $\mathbf{1}_{(s,t]} \mapsto W_t - W_s$ extends to a linear isometry $I_W : L^2(J) \to L^2(\Omega, \mathcal{F}, P)$. For every function $\varphi \in L^2(J)$, the random variable $I_W(\varphi)$ is mean-zero Gaussian.*

Proof. Given the identity $\langle \mathbf{1}_{[0,t]}, \mathbf{1}_{[0,s]} \rangle = \langle W_s, W_t \rangle$, the theorem is a straightforward use of standard results in Hilbert space theory. Let H_0 be the set of all finite linear combinations of interval indicator functions $\mathbf{1}_A$. Then H_0 is a dense, linear subspace of $L^2(J)$, that is, every function $f \in L^2(J)$ can be approximated arbitrarily closely in the L^2-metric by elements of H_0. Since I_W is a linear isometry of H_0, it extends uniquely to a linear isometry of $L^2(J)$, by standard results in Hilbert space theory.

We claim that for any $\varphi \in L^2(J)$, the random variable $I_W(\varphi)$ must be (mean-zero) Gaussian (or identically zero, in case $\varphi = 0$ a.e.). This can be seen as follows: Since H_0 is dense in L^2, there exists a sequence φ_n in H_0 such that $\varphi_n \to \varphi$ in L^2. Since I_W is an isometry, $I_W(\varphi_n) \to I_W(\varphi)$ in L^2. Now convergence in L^2 implies

convergence in measure, which in turn implies convergence in distribution: in particular, for any real θ,

$$\lim_{n\to\infty} E \exp i\theta I_W(\varphi_n) = E \exp i\theta I_W(\varphi).$$

Each of the random variables $I_W(\varphi_n)$ has a centered Gaussian distribution, and consequently has a characteristic function (Fourier transform)

$$E \exp i\theta I_W(\varphi_n) = \exp\{-\theta^2 \sigma_n^2/2\}.$$

The weak convergence of these random variables now implies that the variances σ_n^2 converge to a nonnegative limit σ^2, and that the limit random variable $I_W(\varphi)$ has characteristic function

$$E \exp i\theta I_W(\varphi) = \exp\{-\theta^2 \sigma^2/2\}.$$

But this implies that $I_W(\varphi)$ has a (possibly degenerate) Gaussian distribution with mean 0 and variance ^{2}. $\qquad\square$

The Hilbert space isometry I_W suggests a natural approach to explicit representations of the Wiener process, via *orthonormal bases*. The idea is this: if $\{\psi_n\}_{n\in\mathbb{N}}$ is an orthonormal basis of $L^2[0,1]$, then $\{I_W(\psi_n)\}_{n\in\mathbb{N}}$ must be an orthonormal set in $L^2(P)$. Since uncorrelated Gaussian random variables are necessarily independent, it follows that the random variables $\xi_n := I_W(\psi_n)$ must be i.i.d. standard normals. Finally, since I_W is a linear isometry, it must map the L^2-series expansion of $\mathbf{1}_{[0,t]}$ in the basis ψ_n to the series expansion of W_t in the basis ξ_n. Thus, with almost no further work we conclude the following.

Corollary 1. *Assume that the probability space $(\Omega, \mathcal{F}, P)$ supports an infinite sequence ξ_n of independent, identically distributed $N(0,1)$ random variables, and let $\{\psi_n\}_{n\in\mathbb{N}}$ be any orthonormal basis of $L^2[0,1]$. Then for every $t \in [0,1]$ the infinite series*

$$W_t := \sum_{n=1}^{\infty} \xi_n \langle \mathbf{1}_{[0,t]}, \psi_n \rangle \tag{20}$$

converges in the L^2-metric, and the resulting stochastic process $\{W_t\}_{t\in[0,1]}$ is a mean-zero Gaussian process with covariance function

$$EW_t W_s = \min(s,t).$$

Remark 1. Theorem 3 implies not only that the random variables W_t individually have Gaussian distributions, but that for any finite collection $0 < t_1 < t_2 < \cdots < t_m \leq 1$ the random *vector* $(W_{t_i})_{1\leq i\leq m}$ has a (multivariate) Gaussian distribution. To prove this, first observe that it suffices to show that the joint characteristic function

$$E \exp\{i \sum_{j=1}^{m} \theta_j W_{t_j}\}$$

7

is the characteristic function of the appropriate multivariate Gaussian distribution. But this follows from Theorem 3, because for each choice of real constants θ_j,

$$\sum_{j=1}^{m} \theta_j W_{t_j} = I_W \left(\sum_{j=1}^{m} \theta_j \mathbf{1}_{[0,t_j]} \right).$$

Because the convergence in (20) is in the L^2-metric, rather than the sup-norm, there is no way to conclude directly that the process so constructed has a version with continuous paths. Wiener was able to show by brute force that for the particular basis

$$\psi_n(x) = \sqrt{2} \cos \pi n x$$

the series (20) converges (along an appropriate subsequence) not only in L^2 but also uniformly in t, and therefore gives a version of the Wiener process with continuous paths:

$$W_t = \xi_0 t + \sum_{k=1}^{\infty} \sum_{n=2^{k-1}}^{2^k-1} n^{-1} \xi_n \sqrt{2} \sin \pi n t. \tag{21}$$

The proof of this uniform convergence is somewhat technical, though, and moreover, it is in many ways unnatural. Thus, rather than following Wiener's construction, we will describe a different construction, due to P. Lévy.

1.6 Lévy's Construction

Lévy discovered that a more natural orthonormal basis for the construction of the Wiener process is the *Haar wavelet basis*. The Haar basis is defined as follows: first, let $\psi : \mathbb{R} \to \{-1, 1\}$ be the "mother wavelet" function

$$\psi(t) = \begin{cases} 1 & \text{if } 0 \leq t \leq \frac{1}{2}; \\ -1 & \text{if } \frac{1}{2} < t \leq 1; \\ 0 & \text{otherwise.} \end{cases} \tag{22}$$

Then for any integers $n \geq 0$ and $0 \leq k < 2^n$ define the (n,k)th Haar function by

$$\psi_{n,k}(t) = 2^{n/2} \psi(2^n t - k) \tag{23}$$

The function $\psi_{n,k}$ has support $[k/2^n, (k+1)/2^n]$, and has absolute value equal to $2^{n/2}$ on this interval, so its L^2-norm is 1. Note that $\psi_{0,0} = 1$ on $[0,1]$. Moreover, the functions $\psi_{n,k}$ are mutually orthogonal:

$$\langle \psi_{n,k}, \psi_{m,l} \rangle = \begin{cases} 1 & \text{if } n = m \text{ and } k = l; \\ 0 & \text{otherwise.} \end{cases} \tag{24}$$

Exercise 1. Prove that the Haar functions $\{\psi_{n,k}\}_{n\geq 0,0\leq k<2^n}$ form a complete orthonormal basis of $L^2[0,1]$. HINT: Linear combinations of the indicator functions of dyadic intervals $[k/2^n,(k+1)/2^n]$ are dense in $L^2[0,1]$.

In certain senses the Haar basis is better suited to Brownian motion than the Fourier basis, in part because the functions $\psi_{n,k}$ are "localized" (which fits with the independent increments property), and in part because the normalization of the functions forces the scale factor $2^{n/2}$ in (23) (which fits with the Brownian scaling law).

The series expansion (20) involves the inner products of the basis functions with the indicators $\mathbf{1}_{0,t}$. For the Haar basis, these inner products define the *Schauder functions* $G_{n,k}$, which are defined as the indefinite integrals of the Haar functions:

$$G_{n,k}(t) = \langle \mathbf{1}_{[0,t]}, \psi_{n,k}\rangle = \int_0^t \psi_{n,k}(s)\,ds \tag{25}$$

The graphs of these functions are steeply peaked "hats" sitting on the dyadic intervals $[k/2^n,(k+1)/2^n]$, with heights $2^{-n/2}$ and slopes $\pm 2^{n/2}$. Note that $G_{0,0}(t) = t$.

Theorem 4. *(Lévy) If the random variables $\xi_{m,k}$ are independent, identically distributed with common distribution $N(0,1)$, then with probability one, the infinite series*

$$W(t) := \xi_{0,1}t + \sum_{m=1}^{\infty}\sum_{k=0}^{2^m-1}\xi_{m,k}G_{m,k}(t) \tag{26}$$

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

Lemma 1. *If Z is a standard normal random variable then for every $x > 0$,*

$$P\{Z > x\} \leq \frac{2e^{-x^2/2}}{\sqrt{2\pi}x}. \tag{27}$$

Proof.

$$P\{Z > x\} = \frac{1}{\sqrt{2\pi}}\int_x^{\infty}e^{-y^2/2}\,dy \leq \frac{1}{\sqrt{2\pi}}\int_x^{\infty}e^{-xy/2}\,dy = \frac{2e^{-x^2/2}}{\sqrt{2\pi}x}.$$

□

Proof of Theorem 4. By definition of the Schauder functions $G_{n,k}$, the series (26) is a particular case of (20), so the random variables $W(t)$ defined by (26) are centered Gaussian with covariances that agree with the covariances of a Wiener process.

9

Hence, to prove that (26) defines a Brownian motion, it suffices to prove that with probability one the series converges *uniformly* for $t \in [0, 1]$.

The Schauder function $G_{m,k}$ has maximum value $2^{-m/2}$, so to prove that the series (26) converges uniformly it is enough to show that

$$\sum_{m=1}^{\infty} \sum_{k=1}^{2^m} |\xi_{m,k}|/2^{m/2} < \infty$$

with probability 1. To do this we will use the Borel-Cantelli Lemma and the tail estimate of Lemma 1 for the normal distribution to show that with probability one there is a (possibly random) m_* such that

$$\max_k |\xi_{m,k}| \leq 2^{m/4} \quad \text{for all } m \geq m_*. \tag{28}$$

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 (28) holds eventually, it suffices (by Borel-Cantelli) to show that the probabilities of the complementary events are summable. By Lemma 1,

$$P\{|\xi_{m,k}| \geq 2^{m/4}\} \leq \frac{4}{2^{m/4}\sqrt{2\pi}} e^{-2^{m/2}}.$$

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

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

Since this bound is summable in m, Borel-Cantelli implies that with probability 1, eventually (28) must hold. This proves that w.p.1 the series (26) converges uniformly, and therefore $W(t)$ is continuous.

$\square$

2 The Markov and Strong Markov Properties

2.1 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 $\sigma-algebras$. Define

$$\mathcal{F}_t^W := \sigma(\{W_s\}_{0 \leq s \leq t}) \tag{29}$$

to be the smallest $\sigma-$algebra containing all events of the form $\{W_s \in B\}$ where $0 \leq s \leq t$ and $B \subset \mathbb{R}$ is a Borel set. The indexed collection of $\sigma-$algebra $\{\mathcal{F}_t^W\}_{t\geq 0}$ is called the *standard filtration* associated with the Brownian motion.

Example: 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,

$$\{M(t) > a\} = \bigcup_{s\in\mathbb{Q}:0\leq s\leq t} \{W(s) > a\}. \tag{30}$$

Here $\mathbb{Q}$ denotes the set of rational numbers. Because $\mathbb{Q}$ is a countable set, the union in (30) 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 $\mathcal{F}_t^W$.

In general, a *filtration* is an increasing family of $\sigma-$algebras $\{\mathcal{G}_t\}_{t\geq 0}$ indexed by time t. A stochastic process $X(t)$ is said to be *adapted* to a filtration $\{\mathcal{G}_t\}_{t\geq 0}$ if the random variable $X(t)$ is measurable relative to $\mathcal{G}_t$ for every $t \geq 0$. It is sometimes necessary to consider filtrations other than the standard one, because in some situations there are several sources of randomness (additional independent Brownian motions, Poisson processes, coin tosses, etc.).

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*[1] 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 $\mathcal{G}_t$.

Of course, we have not yet defined what it means for a stochastic process $\{X_t\}_{t\in J}$ to be independent of a $\sigma-$algebra $\mathcal{G}$, but the correct definition is easy to guess: the process $\{X_t\}_{t\in J}$ is independent of $\mathcal{G}$ if the $\sigma-$algebra $\sigma(X_t)_{t\in J}$ generated by the random variables X_t (i.e., the smallest $\sigma-$algebra containing all events of the form $X_t \in B$, where $t \in J$ and B is a Borel set) is independent of $\mathcal{G}$.

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 of the Markov Property. 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

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

---

[1]This is not standard terminology. Some authors (for instance, Karatzas and Shreve) call such filtrations *Brownian filtrations*.

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 $\mathcal{F}_s$.

2.2 Sufficient conditions for independence

The Markov property is a straightforward consequence of the independent increments property of the Wiener process, and the proof extends without change to an arbitrary Lévy process, as neither the continuity of paths nor the Gaussian distribution of the increments played any role in the argument. Establishing the independence of a stochastic process from a $\sigma-$algebra is not always so easy, however.

In general, two $\sigma-$algebras $\mathcal{G}$ and $\mathcal{H}$ are independent if for every $G \in \mathcal{G}$ and every $H \in \mathcal{H}$ the events G, H are independent, that is, if $P(G \cap H) = P(G)P(H)$. Equivalently, $\mathcal{G}$ and $\mathcal{H}$ are independent if for every $H \in \mathcal{H}$ the conditional probability of H given $\mathcal{G}$ coincides with the unconditional probability:

$$E(\mathbf{1}_H \,|\, \mathcal{G}) = E\mathbf{1}_H = P(H) \quad \text{almost surely.} \tag{31}$$

When $\mathcal{H} = \sigma(X_t)_{t \in J}$ is the $\sigma-$algebra generated by a (real) stochastic process X_t, other sufficient conditions for independence from $\mathcal{G}$ are sometimes useful. Since any event in $\mathcal{H}$ is the almost sure limit of a sequence of cylinder events (events of the form $\cap_{i=1}^m \{X_{t_i} \in B_i\}$), to prove (31) it suffices to consider cylinder events. Similarly, for any cylinder event $A = \cap_{i=1}^m \{X_{t_i} \in B_i\}$ there is a sequence of bounded, continuous functions $g_n : \mathbb{R}^m \to [0,1]$ such that

$$\lim_{n \to \varepsilon} g_n(X_{t_1}, X_{t_2}, \ldots, X_{t_m}) = \mathbf{1}_A \quad \text{almost surely;}$$

thus, $\mathcal{H}$ is independent of $\mathcal{G}$ if and only if for every finite subset $\{t_i\}_{i \le m} \subset J$ and every bounded, continuous function $g : \mathbb{R}^m \to \mathbb{R}$,

$$E(g(X_{t_1}, X_{t_2}, \ldots, X_{t_m}) \,|\, \mathcal{G}) = Eg(X_{t_1}, X_{t_2}, \ldots, X_{t_m}) \quad \text{almost surely.} \tag{32}$$

Finally, since any bounded, continuous function g can be arbitrarily well approximated by finite trig polynomials, to prove (32) it suffices to show that the *conditional* joint characteristic function (Fourier transform) of $X_{t_1}, X_{t_2}, \ldots, X_{t_m}$ given $\mathcal{G}$ coincides (almost surely) with the *unconditional* joint characteristic function, that is,

$$E(\exp\{i \sum_{j=1}^m \theta_j X_{t_j}\} \,|\, \mathcal{G}) = E \exp\{i \sum_{j=1}^m \theta_j X_{t_j}\} \quad \text{almost surely.} \tag{33}$$

Exercise 2. Prove the approximation statements used in the arguments above. Specifically, (a) prove that any event in $\mathcal{H}$ is the almost sure limit of cylinder events; and (b) the indicator function of any cylinder event is the almost sure limit of continuous functions.

The sufficient conditions (32) and (33) have an easy consequence that is sometimes useful.

Lemma 2. *Let $\{X_t^n\}_{n \geq 1, t \in J}$ be a sequence of stochastic processes indexed by $t \in J$, all defined on the same underlying probability space $(\Omega, \mathcal{F}, P)$. Let $\mathcal{G}$ be a $\sigma-$algebra contained in $\mathcal{F}$, and suppose that each process $\{X_t^n\}_{t \in J}$ is independent of $\mathcal{G}$. If*

$$\lim_{n \to \infty} X_t^n = X_t \quad \text{a.s. for each } t \in J \tag{34}$$

then the stochastic process $\{X_t\}_{t \in J}$ is independent of $\mathcal{G}$.

Proof. It suffices to establish the identity (32) for every finite subset $\{t_i\}_{i \leq m}$ of the index set J and every bounded continuous function g on $\mathbb{R}^m$. Because each process $\{X_t^n\}_{t \in J}$ is independent of $\mathcal{G}$, equation (32) holds when X is replaced by X^n. Since g is continuous, the hypothesis of the lemma implies that

$$\lim_{n \to \infty} g(X_{t_1}^n, X_{t_2}^n, \ldots, X_{t_m}^n) = g(X_{t_1}, X_{t_2}, \ldots, X_{t_m}) \quad \text{a.s.}$$

Consequently, by the dominated convergence theorem (applied twice, once for ordinary expectations and once for conditional expectations), equation (32) holds for the limit process $\{X_t\}_{t \in J}$ $\qquad\square$

2.3 Blumenthal's 0-1 Law

Recall that the intersection of any collection of $\sigma-$algebras is a $\sigma-$algebra. Thus, if $\mathbb{F} = \{\mathcal{F}_t\}_{t \geq 0}$ is a filtration, then so is the collection $\mathbb{F}_+ = \{\mathcal{F}_{t+}\}_{t \geq 0}$, where for each t,

$$\mathcal{F}_{t+} = \cap_{s > t} \mathcal{F}_s \tag{35}$$

Call this the *right-continuous augmentation* of the filtration $\mathbb{F}$. The term *right-continuous* is used because the enlarged filtration has the property that $\mathcal{F}_{t+} = \cap_{s > t} \mathcal{F}_{s+}$.

Clearly, $\mathcal{F}_t \subset \mathcal{F}_{t+}$, because the $\sigma-$algebra $\mathcal{F}_t$ is contained in $\mathcal{F}_{t+\varepsilon}$ for every $\varepsilon > 0$. One might wonder what (if any) additional "information" there is in $\mathcal{F}_{t+}$ that is not already contained in $\mathcal{F}_t$.

Example 1. Since the random variable $W_0 = 0$ is constant, the $\sigma-$algebra $\mathcal{F}_0^W$ is trivial: it contains only the events $\emptyset$ and $\Omega = \emptyset^c$. The augmentation $\mathcal{F}_{0+}^W$, however, contains interesting events, for instance, the events

$$A_+ = \cap_{n \geq 1} \cup_{q \in \mathbb{Q}; q \leq 2^{-n}} \{W_q > 0\} \quad \text{and} \tag{36}$$
$$A_- = \cap_{n \geq 1} \cup_{q \in \mathbb{Q}; q \leq 2^{-n}} \{W_q < 0\}.$$

Because Brownian paths are continuous, the event A_+ (respectively, A_-) is the event that the path enters the positive (respectively, negative) halfline at arbitrarily small times. These events are elements of $\mathcal{F}_{0+}^W$, because the inner union in each definition is an event in $\mathcal{F}_{2^{-n}}^W$.

Proposition 3. *If $\{W_t\}_{t\geq 0}$ is a standard Brownian motion and $\{\mathcal{F}_t^W\}_{t\geq 0}$ is the standard filtration then the augmented filtration $\{\mathcal{F}_{t+}^W\}_{t\geq 0}$ is admissible.*

Proof. We must show that for each $s \geq 0$, the post-s process $\{\Delta W(s,t) := W_{t+s} - W_s\}_{t\geq 0}$ is independent of the $\sigma-$algebra $\mathcal{F}_{s+}^W$. By the Markov property (Proposition 2), for any $m \geq 1$ the post-$(s + 2^{-m})$ process $\{\Delta W(s + 2^m, t)\}_{t\geq 0}$ is independent of $\mathcal{F}_{s+2^{-n}}^W$, and since $\mathcal{F}_{s+}^W$ is contained in $\mathcal{F}_{s+2^{-n}}^W$, it follows that the process $\{\Delta W(s + 2^m, t)\}_{t\geq 0}$ is independent of $\mathcal{F}_{s+}^W$. But the path-continuity of the Wiener process guarantees that for each $t \geq 0$,

$$\lim_{m\to\infty} \Delta W(s + 2^{-m}, t) = \Delta W(s, t),$$

and so Lemma 2 implies that the process $\{\Delta W(s,t)\}_{t\geq 0}$ is independent of $\mathcal{F}_{s+}^W$. $\square$

Corollary 2. *For every $t \geq 0$ and every integrable random variable Y that is measurable with respect to the $\sigma-$algebra $\mathcal{F}_\infty^W$ generated by a standard Wiener process,*

$$E(Y \,|\, \mathcal{F}_{t+}^W) = E(Y \,|\, \mathcal{F}_t^W). \tag{37}$$

Consequently, for any event $F \in \mathcal{F}_{t+}^W$ there is an event $F' \in \mathcal{F}_t^W$ such that $P(F \Delta F') = 0$.

Proof. By the usual approximation arguments, to prove (37) it suffices to consider random variables of the form

$$Y = \exp\{i \sum_{j=1}^{m} \theta_j (W_{t_j} - W_{t_{j-1}})\}$$

where $0 \leq t_1 < t_2 < \cdots < t_k = t < t_{k+1} < \cdots < t_m$. For such t_j, the random variable $\sum_{j\leq k} \theta_j (W_{t_j} - W_{t_{j-1}})$ is measurable with respect to $\mathcal{F}_t^W$, and hence also with respect to $\mathcal{F}_{t+}^W$, while by Proposition 3 the random variable $\sum_{k<j\leq m} \theta_j (W_{t_j} - W_{t_{j-1}})$ is independent of $\mathcal{F}_{t+}^W$. Consequently, by the independence and stability properties of conditional expectation (see the notes on conditional expectation posted on the web page),

$$E(Y \,|\, \mathcal{F}_{t+}^W) = \exp\{i \sum_{j\leq k} \theta_j (W_{t_j} - W_{t_{j-1}})\} E \exp\{i \sum_{k<j\leq m} \theta_j (W_{t_j} - W_{t_{j-1}})\} \quad \text{a.s.}$$

But the same argument shows that

$$E(Y \,|\, \mathcal{F}_t^W) = \exp\{i \sum_{j\leq k} \theta_j (W_{t_j} - W_{t_{j-1}})\} E \exp\{i \sum_{k<j\leq m} \theta_j (W_{t_j} - W_{t_{j-1}})\} \quad \text{a.s.}$$

This proves (37). The second assertion of the corollary follows by applying (37) to $Y = \mathbf{1}_F$: this shows that

$$\mathbf{1}_F = E(\mathbf{1}_F \,|\, \mathcal{F}^W_{t+}) = E(\mathbf{1}_F \,|\, \mathcal{F}^W_t) \quad \text{a.s.,}$$

and so the indicator of F is almost surely equal to an $\mathcal{F}^W_t$ measurable random variable. Since $\mathbf{1}_F$ takes only the values 0 and 1, it follows that its projection on $\mathcal{F}^W_t$ takes only the values 0 and 1, and hence is $\mathbf{1}_{F'}$ for some $F' \in \mathcal{F}^W_t$. $\square$

Corollary 3. *(Blumenthal 0-1 Law) Every event in $\mathcal{F}^W_{0+}$ has probability 0 or 1.*

Proof. The $\sigma-$algebra $\mathcal{F}^W_0$ is the trivial $\sigma-$algebra $\{\emptyset, \Omega\}$, because it is generated by the constant random variable W_0. Therefore, by Corollary 2, each event $F \in \mathcal{F}^W_{0+}$ differs from either $\emptyset$ or Ω by an event of probability 0, and so $P(F) = 0$ or 1. $\square$

Recall that the events A_+ and A_- defined by equations (36) are elements of the $\sigma-$algebra $\mathcal{F}^W_{0+}$. Hence, Blumenthal's Law implies that both A_+ and A_- have probabilities 0 or 1. Which is it? For each $n \geq 1$, the event $\cup_{q \in \mathbb{Q}; q \leq 2^{-n}} \{W_q > 0\}$ has probability at least $1/2$, because the chance that $W_{2^{-n}} > 0$ is $1/2$. Consequently, the intersection must have probability at least $1/2$, and so by Blumenthal's Law, $P(A_+) = 1$. A similar argument shows that $P(A_-) = 1$. Hence,

$$P(A_+ \cap A_-) = 1. \tag{38}$$

On the event A_+, the path W_t takes positive values at indefinitely small positive times; similarly, on A_- the path W_t takes negative values at indefinitely small positive time. Therefore, on the event $A_+ \cap A_-$, the path W_t must cross and recross the origin 0 infinitely often in each time interval $(0, 2^{-n})$. Blumenthal's 0-1 Law implies that this must be the case *with probability one*.

2.4 Stopping times and stopping fields

A *stopping time* for a filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t \geq 0}$ is defined to be a nonnegative (possibly infinite)[2] random variable τ such that for each (nonrandom) $t \in [0, \infty]$ the event $\{\tau \leq t\}$ is in the $\sigma-$algebra $\mathcal{F}_t$.

Example: $\tau(a) := \min\{t \,:\, W(t) = a\}$ 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 $\mathcal{F}_t$.

---

[2]A stopping time that takes only finite values will be called a *finite* stopping time.

Exercise 3. (a) Prove that if τ, σ are stopping times then so are $\sigma \wedge \tau$ and $\sigma \vee \tau$.
(b) Prove that if τ is a *finite* stopping time then so is

$$\tau_n = \min\{k/2^n \geq \tau\}.$$

Thus, any finite stopping time is a decreasing limit of stopping times each of which takes values in a discrete set.

The *stopping field* $\mathcal{F}_\tau$ (more accurately, the *stopping $\sigma-$algebra*) associated with a stopping time τ is the collection of all events B such that $B \cap \{\tau \leq t\} \in \mathcal{F}_t$ for every nonrandom t. Informally, $\mathcal{F}_\tau$ consists of all events that are "observable" by time τ.

Exercise 4. (a) Show that $\mathcal{F}_\tau$ is a $\sigma-$algebra. (b) Show that if $\sigma \leq \tau$ are both stopping times then $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$.

Exercise 5. Let τ be a finite stopping time with respect to the standard filtration of a Brownian motion $\{W_t\}_{t \geq 0}$ and let $\mathcal{F}_\tau^W$ be the associated stopping field. Show that each of the random variables τ, W_τ, and M_τ is measurable relative to $\mathcal{F}_\tau^W$.

2.5 Strong Markov Property

Theorem 5. *(Strong Markov Property) Let $\{W(t)\}_{t \geq 0}$ be a standard Brownian motion, and let τ be a* finite *stopping time relative to the standard filtration, with associated stopping $\sigma-$algebra $\mathcal{F}_\tau$. For $t \geq 0$, define the* post-τ *process*

$$W^*(t) = W(t + \tau) - W(\tau), \tag{39}$$

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 following variant of the strong Markov property is also useful.

Theorem 6. *(Splicing) 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*

$$\begin{aligned}\tilde{W}_t &= W_t && \text{for } t \leq \tau, \\ &= W_\tau + W_{t-\tau}^* && \text{for } t \geq \tau\end{aligned} \tag{40}$$

is also a Brownian motion.

The hypothesis that τ is a stopping time is essential for the truth of the strong Markov Property. Consider the following example. Let T be the first time that the Wiener path reaches its maximum value up to time 1, that is,

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

The random variable T is well-defined, by path-continuity, as this 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 we have already shown, using Blumenthal's Law, that with probability 1, Brownian paths immediately enter both the positive and negative halflines. This is a contradiction, so we conclude that the strong Markov property fails for the random variable T.

Proof of the strong Markov property. First, we will show that it suffices to prove the theorem for stopping times that take values in the discrete set $D_m = \{k/2^m\}_{k \geq 0}$, for some $m \geq 1$; then we will prove that the strong Markov property holds for stopping times that take values in D_m.

Step 1. Suppose that the theorem is true for all stopping times that take values in one of the sets D_m; we will show that it is then true for *any* finite stopping time. By Exercise 3, if τ is a finite stopping time then so is $\tau_m = \min\{k/2^m \geq \tau\}$. Clearly, the sequence τ_m is non-increasing in m, and $\tau_m \to \tau$ as $m \to \infty$. Hence, by Exercise 4, the stopping fields $\mathcal{F}_{\tau_m}$ are reverse-ordered, and $\mathcal{F}_\tau \subset \mathcal{F}_{\tau_m}$ for every $m \geq 1$.

By hypothesis, for each $m \geq 1$ the post-τ_m process $\Delta W(\tau_m, t) = W(t + \tau_m) - W(\tau_m)$ is a standard Brownian motion independent of $\mathcal{F}_{\tau_m}$. Since $\mathcal{F}_\tau \subset \mathcal{F}_{\tau_m}$, the process $\Delta W(\tau_{m,t})$ is also independent of $\mathcal{F}_\tau$. But Brownian paths are right-continuous, so the convergence $\tau_m \downarrow \tau$ implies that for each $t \geq 0$,

$$\lim_{m \to \infty} \Delta W(\tau_m, t) = \Delta W(\tau, t) = W^*(t). \tag{41}$$

Consequently, by Lemma 2, the post-τ process $W^*(t)$ is independent of $\mathcal{F}_\tau$.

To complete the first step of the proof we must show that the process $W^*(t)$ is a Brownian motion. The sample paths are obviously continuous, since the paths of the original Brownian motion $W(t)$ are continuous, so we need only verify

properties (C) and (D) of Definition 1. For this is suffices to show that for any $0 = t_0 < t_1 < \cdots < t_k$,

$$E \exp \left\{ i \sum_{j=1}^{k} \theta_j (W^*_{t_j} - W^*_{t_{j-1}}) \right\} = E \exp \left\{ i \sum_{j=1}^{k} \theta_j (W_{t_j} - W_{t_{j-1}}) \right\} \tag{42}$$

But by hypothesis, each of the processes $\{\Delta W(\tau_m, t)\}_{t \geq 0}$ is a Brownian motion, so for each m the equation (42) holds when W^*_t is replaced by $\Delta W(\tau_m, t)$ in the expectation on the left side. Consequently, the equality (42) follows by (41) and the dominated convergence theorem.

$\square$

Step 2: We must prove that if τ is a stopping time that takes values in D_m, for some $m \geq 0$, then the post-τ process $\{W^*_t\}_{t \geq 0}$ is a Brownian motion and is independent of $\mathcal{F}_\tau$. For ease of notation, we will assume that $m = 0$; the general case can be proved by replacing each integer n in the following argument by $n/2^m$.

It will suffice to show that if B is any event in $\mathcal{F}_\tau$ then for any $0 = t_0 < t_1 < \cdots < t_k$ and $\theta_1, \theta_2, \ldots, \theta_k$,

$$E \mathbf{1}_B \exp \left\{ i \sum_{j=1}^{k} \theta_j (W^*_{t_j} - W^*_{t_{j-1}}) \right\} = P(B) E \exp \left\{ i \sum_{j=1}^{k} \theta_j (W_{t_j} - W_{t_{j-1}}) \right\}. \tag{43}$$

Since τ takes only nonnegative integer values, the event B can be partitioned as $B = \cup_{n \geq 0}(B \cap \{\tau = n\})$. Since τ is a stopping time, the event $B \cap \{\tau = n\}$ is in the $\sigma-$algebra $\mathcal{F}_n$; moreover, on $\{\tau = n\}$, the post-τ process coincides with the post-n process $\Delta W(n, t) = W_{t+n} - W_n$. Now the Markov property implies that the post-n process $\Delta W(n, t)$ is a Brownian motion independent of $\mathcal{F}_n$, and hence independent of $B \cap \{\tau = n\}$. Consequently,

$$E \mathbf{1}_B \mathbf{1}_{\{\tau=n\}} \exp \left\{ i \sum_{j=1}^{k} \theta_j (W^*_{t_j} - W^*_{t_{j-1}}) \right.$$

$$= E \mathbf{1}_B \mathbf{1}_{\{\tau=n\}} \exp \left\{ i \sum_{j=1}^{k} \theta_j (\Delta W(n, t_j) - \Delta W(n, t_{k-1})) \right\}$$

$$= P(B \cap \{\tau = n\}) E \exp \left\{ i \sum_{j=1}^{k} \theta_j (W_{t_j} - W_{t_{j-1}}) \right\}.$$

Summing over $n = 0, 1, 2, \ldots$ (using the dominated convergence theorem on the left and the monotone convergence theorem on the right), we obtain (43). $\square$

Proof of the Splicing Theorem 6. This is done using the same strategy as used in proving the strong Markov property: first, one proves that it suffices to consider stopping times τ that take values in D_m, for some m; then one proves the result for this

18

restricted class of stopping times. Both steps are very similar to the corresponding steps in the proof of the strong Markov property, so we omit the details. □

3 Embedded Simple Random Walks

Lemma 3. *Define* $\tau = \min\{t > 0 : |W_t| = 1\}$. *Then with probability* 1, $\tau < \infty$.

Proof. 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 the random variable $W(\tau)$, 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

$$\tau_{n+1} := \min\{t > \tau_n : |W_{t+\tau_n} - W_{\tau_n}| = 1\}. \tag{44}$$

The arguments above imply the following.

Proposition 4. *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 4. *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 $2^{-m}\mathbb{Z}$ "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 4: 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.

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 5.

$$P\{M(t) \geq a\} = P\{\tau_a \leq t\} = 2P\{W(t) > a\} = 2 - 2\Phi(a/\sqrt{t}). \tag{45}$$

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 4

above, $\tau < \infty$ almost surely. Since τ is a stopping time, the post-τ process

$$W^*(t) := W(\tau + t) - W(\tau) \tag{46}$$

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:

$$\begin{aligned} \tilde{W}(s) &= W(s) && \text{for } s \leq \tau, \\ &= 2a - W(s) && \text{for } s \geq \tau. \end{aligned} \tag{47}$$

Proposition 6. *(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 the Splicing Theorem 6. $\qquad\square$

Proof of Proposition 5. 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

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

$\qquad\square$

Corollary 5. *The first-passage time random variable $\tau(a)$ is almost surely finite, and has the* one-sided stable *probability density function of index $1/2$:*

$$f(t) = \frac{ae^{-a^2/2t}}{\sqrt{2\pi t^3}}. \tag{48}$$

Proof. Differentiate the equation (45) with respect to the variable t. $\qquad\square$

Essentially the same arguments prove the following.

Corollary 6.

$$P\{M(t) \in da \text{ and } W(t) \in a - db\} = \frac{2(a+b)\exp\{-(a+b)^2/2t\}}{(2\pi)^{1/2}t^{3/2}} \, dadb \qquad (49)$$

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

Exercise 6. 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 $t_1, t_2, \ldots, t_k$ 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 2. 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 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 7. 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 $x, y \in (0, 1)$.

5 Wald Identities for Brownian Motion

5.1 Doob's optional sampling formula

A *continuous-time martingale* with respect to a filtration $\{\mathcal{F}_t\}_{t \geq 0}$ is an adapted process $\{X_t\}_{t \geq 0}$ such that each X_t is integrable, and such that for any $0 \leq s \leq t < \infty$,

$$E(X_t \mid \mathcal{F}_s) = X_s \quad \text{almost surely.} \qquad (50)$$

Equivalently, for every finite set of times $0 \leq t_1 < t_2 < \cdots < t_n$, the finite sequence $\{X_{t_i}\}_{i \leq n}$ is a discrete-time martingale with respect to the discrete filtration $\{\mathcal{F}_{t_i}\}_{i \leq n}$. Observe that the definition does not require that the sample paths $t \mapsto X_t$ be continuous, or even measurable.

Theorem 7. *(Doob) If $\{X_t\}_{t \geq 0}$ is a martingale with respect to the filtration $\{\mathcal{F}_t\}_{t \geq 0}$ such that $\{X_t\}_{t \geq 0}$ has right-continuous paths and if τ is a* bounded *stopping time for this filtration, then X_τ is an integrable random variable and*

$$EX_\tau = EX_0. \tag{51}$$

Proof. By hypothesis there is a finite constant C such that $\tau \leq C$. By Exercise 3, the random variables $\tau_m := \min\{k/2^m \geq \tau\}$ are stopping times. Clearly, each τ_m takes values in a finite set ,

$$C + 1 \geq \tau_1 \geq \tau_2 \geq \cdots \geq \cdots , \quad \text{and}$$
$$\tau = \lim_{m \to \infty} \tau_m.$$

Since τ_m takes values in a discrete set, Doob's identity for discrete-time martingales implies that for each m,

$$EX_{\tau_m} = EX_0.$$

Furthermore, because the sample paths of $\{X_t\}_{t \geq 0}$ are right-continuous, $X_\tau = \lim X_{\tau_m}$ pointwise. Hence, by the dominated convergence theorem, to prove the identity (51) it suffices to show that the sequence X_{τ_m} is *uniformly integrable*. For this, observe that the ordering $C + 1 \geq \tau_1 \geq \tau_2 \geq \cdots$ of the stopping times implies that the sequence $\{X_{\tau_m}\}_{m \geq 1}$ is a *reverse* martingale with respect to the backward filtration $\{\mathcal{F}_{\tau_m}\}$. Any reverse martingale is uniformly integrable. $\qquad \square$

Remark 3. (a) If we knew *a priori* that $\sup_{t \leq C+1} |X_t|$ were integrable then there would be no need to show that the sequence $\{X_{\tau_m}\}_{m \geq 1}$ is uniformly integrable, and so the use of results concerning reverse martingales could be avoided. All of the martingales that will figure into the results of section 5.2 will have this property. (b) It can be shown that every continuous-time martingale has a version with right-continuous paths. (In fact, this can be done by using only the upcrossings and maximal inequalities for discrete-time martingales, together with the fact that the dyadic rationals are dense in $[0, \infty)$.) We will not need this result, though.

5.2 The Wald identities

You should recall (see the notes on discrete-time martingales) that there are several discrete-time martingales associated with the simple random walk on $\mathbb{Z}$ that are quite useful in first-passage problems. If $S_n = \sum_{i=1}^n \xi_i$ is a simple random walk

on $\mathbb{Z}$ (that is, the random variables $\{\xi_i\}_{i\geq 1}$ are independent, identically distributed Rademacher, as in section 1) and $\mathcal{F}_n = \sigma(\xi_i)_{i\leq n}$ is the $\sigma-$algebra generated by the first n steps of the random walk, then each of the following sequences is a martingale relative to the discrete filtration $(\mathcal{F}_n)_{n\geq 0}$:

(a) S_n;
(b) $S_n^2 - n$;
(c) $\exp\{\theta S_n\}/(\cosh\theta)^n$.

There are corresponding continuous-time martingales associated with the Wiener process that can be used to obtain analogues of the Wald identities for simple random walk.

Proposition 8. *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 fo the following holds:

$$EW(\tau) = 0; \tag{52}$$

$$EW(\tau)^2 = E\tau; \tag{53}$$

$$E\exp\{\theta W(\tau) - \theta^2\tau/2\} = 1 \qquad \forall\theta \in \mathbb{R};\ and \tag{54}$$

$$E\exp\{i\theta W(\tau) + \theta^2\tau/2\} = 1 \qquad \forall\theta \in \mathbb{R}. \tag{55}$$

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. The martingale property follows immediately from the independent increments property. In particular, the definition of an admissible filtration implies that $W(t+s) - W(s)$ is independent of $\mathcal{G}_s$; hence (for instance),

$$E(\exp\{\theta W_{t+s} - \theta^2(t+s)/2\}\,|\,\mathcal{G}_s) = \exp\{\theta W_s - \theta^2 s/2\}E(\exp\{\theta W_{t+s} - W_s - \theta^2 t/2\}\,|\,\mathcal{G}_s)$$
$$= \exp\{\theta W_s - \theta^2 s/2\}E\exp\{\theta W_{t+s} - W_s - \theta^2 t/2\}$$
$$= \exp\{\theta W_s - \theta^2 s/2\}.$$

The Wald identities follow from Doob's optional sampling formula. $\qquad\square$

24

5.3 Wald identities and first-passage problems for Brownian motion

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,

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

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

$$\boxed{P\{W(T) = b\} = \frac{a}{a+b}.} \tag{56}$$

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 (56), one may now easily obtain

$$\boxed{ET = ab.} \tag{57}$$

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 $\theta > 0$,

$$E \exp\{\theta W(\tau \wedge n) - \theta^2 (\tau \wedge n)\} = 1. \tag{58}$$

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 (58) 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

$$\boxed{E \exp\{-\lambda\tau_a\} = \exp\{-\sqrt{2\lambda}a\}.} \tag{59}$$

The only density with this Laplace transform[3] is the one–sided stable density given in equation (48). Thus, the Optional Sampling Formula gives us a second proof of (45).

Exercise 8. 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 9. 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 any admissible filtration.

(B) Deduce the corresponding Wald identity for the first passage 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 10. 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 $\{\sqrt{2}\sin k\pi x\}_{k \geq 1}$ (together with) 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,

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

---

[3]Check a table of Laplace transforms.

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

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

6 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 $Z = \{t \le 1 : W(t) = 0\}$) is a homeomorphic image of the Cantor set. These facts may be established using only the formula (45), Brownian scaling, the strong Markov property, and elementary arguments.

6.1 Zero Set of a Brownian Path

The zero set is
$$\mathcal{Z} = \{t \ge 0 \, : \, W(t) = 0\}. \tag{60}$$
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 (non-random) $t > 0$, 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,

$$P\{\mathbb{Q}_+ \cap \mathcal{Z} \neq \emptyset\} = 0. \tag{61}$$

Proposition 9. *With probability one, the Brownian path $W(t)$ has infinitely many zeros in every time interval $(0, \varepsilon)$, where $\varepsilon > 0$.*

Proof. We have already seen that this is a consequence of the Blumenthal 0-1 Law, but we will now give a different proof using what we have learned about the distribution of $M(t)$. 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 (45), 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. $\qquad\square$

Proposition 10. *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 9, 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

28

$(t - \varepsilon, t)$ 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 $t_n \to t$. □

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.

6.2   Nondifferentiability of Paths

Proposition 11. *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

$$|W_t - W_{t_*}| \le C|t - t_*| \quad \text{for all } t \in (t_* - \varepsilon, t_* + \varepsilon), \tag{62}$$

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 $0 \le m \le 4^k$ such that[4]

$$|W((m + i + 1)/4^k) - W((m + i)/4^k)| \le 16C/4^k \quad \text{for each } i = 0, 1, 2. \tag{63}$$

I'll show that the probability of this event is 0. Let $B_{m,k} = B_{k,m}(C)$ be the event that (63) holds, and set $B_k = \cup_{m \le 4^k} B_{m,k}$; then by the Borel-Cantelli lemma it is enough to show that (for each $C < \infty$)

$$\sum_{k=1}^{\infty} P(B_m) < \infty. \tag{64}$$

The trick is *Brownian scaling*: in particular, for all $s, t \ge 0$ the increment $W_{t+s} - W_t$ is Gaussian with mean 0 and standard deviation $\sqrt{s}$. Consequently, since the three increments in (63) are independent, each with standard deviation 2^{-k}, and since the standard normal density is bounded above by $1/\sqrt{2\pi}$,

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

---

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

where Z is standard normal. Since B_k is the union of 4^k such events $B_{m,k}$, it follows 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 11. 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

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

7 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

$$QV(X; \Pi) = \sum_{j=1}^{n} (X(t_j) - X(t_{j-1}))^2. \tag{65}$$

Theorem 8. *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,*

$$\lim_{n \to \infty} QV(W; \Pi_n) = 1. \tag{66}$$

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 8, I'll discuss a much simpler special case[5], 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 12. *Let $\{W(t)\}_{t \geq 0}$ be a standard Brownian motion. For each $t > 0$, with probability one,*

$$\lim_{n \to \infty} QV(W; \mathcal{D}_n[0,t]) = t. \tag{67}$$

Proof. Proof of Proposition 12. First let's prove convergence in probability. To simplify things, assume that $t = 1$. Then for each $n \geq 1$, the random variables

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

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

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

---

[5]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 (67) holds a.s.

Exercise 12. Prove that if (67) holds a.s. for each dyadic rational in the unit interval, then with probability one it holds for all t.

$\square$

In general, when the partition Π is not a dyadic partition, the summands in the formula (66) 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 $\mathrm{sgn}(\zeta)$ and $\mathrm{sgn}(\zeta)$ are independent of $\mathcal{G}$. Then*

$$E((\xi + \zeta)^2 \mid \mathcal{G}) = \xi^2 + \zeta^2. \tag{68}$$

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

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

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

Proof of Theorem 8. 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,

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

By the reverse martingale convergence theorem,

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

Exercise 13. Prove that the limit is constant a.s., and that the constant is 1.

$\Box$

8 Skorohod's Theorem

In section 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, in a certain sense.

Theorem 9. *(Skorohod Embedding I) Let F be any probability distribution on the real line with mean 0 and variance $\sigma^2 < \infty$. Then on some probability space there exist (i) a standard Wiener process $\{W_t\}_{t \geq 0}$; (ii) an admissible filtration $\{\mathcal{F}_t\}_{t \geq 0}$; and (iii) a sequence of finite stopping times $0 = T_0 \leq T_1 \leq T_2 \leq \cdots$ such that*

- *(A) the random vectors $(T_{n+1} - T_n, W_{T_{n+1}} - W_{T_n})$ are independent, identically distributed;*
- *(B) each random variable , $W_{T_{n+1}} - W_{T_n}$ has distribution F; and*
- *(C) $ET_{n+1} - ET_n = \sigma^2$.*

Thus, in particular, the sequence $\{W_{T_n}\}_{n \geq 0}$ has the same joint distribution as a random walk with step distribution F, and $T_n/n \to \sigma^2$ as $n \to \infty$. For most applications of the theorem, this is all that is needed. However, it is natural to wonder about the choice of filtration: is it true that there are stopping times $0 = T_0 \leq T_1 \leq T_2 \leq \cdots$ with respect to the *standard* filtration such that (A), (B), (C) hold? The answer is yes, but the proof is more subtle.

Theorem 10. *(Skorohod Embedding II) 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. Then there exist stopping times $0 = T_0 \leq T_1 \leq T_2 \leq \cdots$ with respect to the standard filtration such that the conclusions (A), (B), (C) of Theorem 9 hold.*

I will only prove Theorem 9. The proof will hinge on a representation of an arbitrary mean-zero probability distribution as a *mixture* of mean-zero two-point

distributions. A *two-point distribution* is, by definition, a probability measure whose support has only two points. For any two points $-a < 0 < b$ there is a unique two-point distribution $F_{a,b}$ with support $\{-a, b\}$ and mean 0, to wit,

$$F_{a,b}(\{b\}) = \frac{a}{a+b} \quad \text{and} \quad F_{a,b}(\{-a\}) = \frac{b}{a+b}.$$

The variance of $F_{a,b}$ is ab.

Proposition 13. *For any Borel probability distribution F on R with mean zero there exists a Borel probability distribution G on $\mathbb{R}_+^2$ such that*

$$F = \int F_{a,b} \, dG(a, b). \tag{69}$$

Hence,

$$\int x^2 \, dF(x) = \int ab \, dG(a, b). \tag{70}$$

Proof for compactly supported F. First we will prove this for distributions F with finite support. To do this, we induct on the number of support points. If F is supported by only two points then $F = F_{a,b}$ for some $a, b > 0$, and so the representation (69) holds with G concentrated at the single point (a, b). Suppose, then, that the result is true for mean-zero distributions supported by fewer than m points, and let F be a mean-zero distribution supported by m points. Then among the m support points there must be two satisfying $-a < 0 < b$, both with positive probabilities $F(-a)$ and $F(b)$. There are three possibilities:

$$-aF(-a) + bF(b) = 0, \quad \text{or}$$
$$-aF(-a) + bF(b) > 0, \quad \text{or}$$
$$-aF(-a) + bF(b) < 0.$$

In the first case, F can obviously be decomposed as a convex combination of $F_{a,b}$ and a probability distribution F' supported by the remaining $m - 2$ points in the support of F In the second case, where $-aF(-a) + bF(b) > 0$, there is some value $0 < \beta < F(b)$ such that $-aF(-a) + \beta F(b) = 0$, and so F can be decomposed as a convex combination of $F_{a,b}$ and a distribution F' supported by b and the remaining $m - 2$ support points. Similarly, in the third case F is a convex combination of $F_{a,b}$ and a distribution F' supported by $-a$ and the remaining $m - 2$ supported points of F. Thus, in all three cases (69) holds by the induction hypothesis. The formula (70) follows from (69) by Fubini's theorem.

Next, let F be a Borel probability distribution on $\mathbb{R}$ with mean zero and compact support $[-A, A]$. Clearly, there is a sequence of probability distributions F_n

converging to F in distribution such that F_n is supported by finitely many points. In particular, if X is a random variable with distribution F, then set

$$X_n = \max\{k/2^n \ : \ k/2^n \le X\}$$

and let F_n be the distribution of X_n. Clearly, $X_n \to X$ pointwise as $n \to \infty$. The distributions F_n need not have mean zero, but by the dominated convergence theorem, $EX_n \to EX = 0$. Thus, if we replace X_n be $X_n - EX_n$ we will have a sequence of mean-zero distributions converging to F, each supported by only finitely many points, all contained in $[-A - 1, A + 1]$. Therefore, since the representation (69) holds for distributions with finite support, there are Borel probability distributions G_n such that

$$F_n = \int F_{a,b} \, dG_n(a, b) \quad \text{and} \quad \sigma_n^2 := \int x^2 \, dF_n(x) = \int ab \, dG_n(a, b).$$

Moreover, since each F_n has supports contained in $[-A - 1, A + 1]$, the mixing distributions G_n have supports contained in $[0, A + 1]^2$. Hence, by Helly's selection principle (i.e., Banach-Alaoglu to those of you from Planet Math), there is a subsequence G_k that converges weakly to a Borel probability distribution G. It follows routinely sthat

$$F = \int F_{a,b} \, dG(a, b) \quad \text{and} \quad \sigma^2 := \int x^2 \, dF(x) = \int ab \, dG(a, b).$$

$\square$

Exercise 14. Finish the proof: Show that the representation (69) holds for all mean-zero, finite variance distributions.

Remark 4. Proposition 13 can also be deduced from CHOQUET'S theorem. The set of mean-zero, probability distributions on $\mathbb{R}$ with variances bounded by C is a convex, weak-∗compact subset of the space of finite Borel measures on $\mathbb{R}$. It is not difficult to see that the extreme points are the two-point distributions $F_{a,b}$. Therefore, the representation (69) is a special case of Choquet's theorem.

Proof of Theorem 9. Consider first the case where F is supported by only two points. Since F has mean zero, the two support points must satisfy $a < 0 < b$, with $p = F(\{b\}) = 1 - F(\{a\})$. Let $\{W_t\}_{t \ge 0}$ be a standard Wiener process, and let T be the first time that the Wiener process visits either a or b. Then $T < \infty$ almost surely, and T is a stopping time with respect to any admissible filtration. Equations (56)–(57) imply that the distribution of W_T is F, and $ET = \sigma^2$. (Exercise: Fill in the details.)

To complete the proof in the case of a two-point distribution, we now use the strong Markov property and an induction argument. Assume that stopping times

$0 = T_0 \leq T_1 \leq T_2 \leq \cdots \leq T_m$ have been defined in such a way that properties (A), (B), (C) hold for $n < m$. Define T_{m+1} to be the first time after T_m that $W_{T_{m+1}} - W_{T_m} = a$ or b; then by the strong Markov property, the random vector $(T_{m+1} - T_m, W_{T_{m+1}} - W_{T_m})$ is independent of the random vectors $(T_{n+1} - T_n, W_{T_{n+1}} - W_{T_n})$ for $n < m$ (since these are all measurable with respect to the stopping field $\mathcal{F}_{T_m}$), and $(T_{m+1} - T_m, W_{T_{m+1}} - W_{T_m})$ has the same distribution as does $(T_1 - T_0, W_{T_1} - W_{T_0})$. This completes the induction for two-point distributions. It should be noted that the random variables T_n are stopping times with respect to any admissible filtration.

Now let F be any mean-zero distribution with finite variance σ^2. By Proposition 13, F has a representation as a mixture of two-point distributions $F_{a,b}$, with mixing measure G. Let $(\Omega, \mathcal{F}, P)$ be a probability space that supports a sequence $\{(A_n, B_n)\}_{n \geq 1}$ of independent, identically distributed random vectors each with distribution G, and an independent Wiener process $\{W_t\}_{t \geq 0}$. (Such a probability space can always be realized as a product space.) Let $\mathcal{F}_0$ be the $\sigma-$algebra generated by the random vectors (A_n, B_n), and for each $t \geq 0$ let $\mathcal{F}_t$ be the $\sigma-$algebra generated by $\mathcal{F}_0$ and the random variables W_s, for $s \leq t$; since the random vectors (A_n, B_n) are independent of the Wiener process, the filtration $\{\mathcal{F}_t\}_{t \geq 0}$ is admissible. Define stopping times T_n inductively as follows: $T_0 = 0$, and

$$T_{n+1} = \min\{t \geq 0 \: : \: W_{T_n+t} - W_{T_n} = -A_n \text{ or } B_n\}.$$

Conditional on $\mathcal{F}_0$, the random vectors $(T_{n+1} - T_n, W_{T_{n+1}} - W_{T_n})$ are independent, by the strong Markov property. Furthermore, conditional on $\mathcal{F}_0$, the random variable $W_{T_{n+1}} - W_{T_n}$ has the two-point distribution $F_{a,b}$ with $a = A_n$ and $b = B_n$. Since the unconditional distribution of (A_n, B_n) is G, it follows that unconditionally the random vectors $(T_{n+1} - T_n, W_{T_{n+1}} - W_{T_n})$ are independent, identically distributed, and $W_{T_{n+1}} - W_{T_n}$ has distribution F. That $ET_{n+1} - ET_n = \sigma^2$ follows from the variance formula in the mixture representation (69).

$\square$

Proof of Theorem 10 for the uniform distribution on (-1,1). The general case is proved by showing *directly* (without using the Choquet representation (69)) that a mean-zero probability distribution F is a limit of finitely supported mean-zero distributions. I will consider only the special case where F is the uniform distribution (normalized Lebesgue measure) 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 $[-1, 1]$. $\qquad \square$

Exercise 15. 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.