1 Definition of the Wiener Process

1.1 The central limit theorem and 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

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

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 $x \in \mathbb{R}$,

$$\lim_{n \to \infty} P\left\{ \frac{S_n}{\sqrt{n}} \leq x \right\} = \Phi(x) := \int_{-\infty}^{x} \varphi_1(y) \, dy$$

where

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

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 $\varphi_1 = \varphi_t \ast \varphi_{1-t}$. More generally, for any $s, t \geq 0$,

$$\varphi_s \ast \varphi_t = \varphi_{s+t}.$$  \hfill (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 $n \to \infty$,

$$\frac{1}{\sqrt{n}} (S_{[nt_1]}, S_{[nt_2]}, \ldots, S_{[nt_m]}) \overset{D}{\to} (W_{t_1}, W_{t_2}, \ldots, W_{t_m}).$$  \hfill (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 \( \text{NORMAL}(0, t) \) 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; we will give a different proof, due to P. Lévy, later in the course.

The convergence (5) shows that, in a certain sense, the Wiener process is an infinitesimal form of the simple random walk. This convergence 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 argument above suggests 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 \( \{ W_t \}_{t \geq 0} \) such that \( \{ W_t - x \}_{t \geq 0} \) 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:

\[ P(W_{t+s} \in dy \mid W_s = x) \overset{\Delta}{=} p_t(x, y)dy = \frac{1}{\sqrt{2\pi t}} \exp\{-\frac{(y - x)^2}{2t}\}dy. \] (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 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

\[
\frac{\partial u}{\partial t} = \frac{1}{2} \frac{\partial^2 u}{\partial x^2} \quad \text{with} \quad u_0(x) = f(x)
\] (7)

is given by

\[
u_t(x) = E f(W_t^x) = \int_{y=-\infty}^{\infty} p_t(x, y) f(y) dy. \] (9)

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

\[
\frac{\partial p_t(x, y)}{\partial t} = \frac{1}{2} \frac{\partial^2 p_t(x, y)}{\partial x^2}. \] (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 \( t \rightarrow W_t \) 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 $\Sigma$ 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
\[
W_t = (W^{(1)}_t, W^{(2)}_t, \ldots, W^{(d)}_t)
\]
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 $\{U W_t\}_{t \geq 0}$ 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:
\[
\begin{align*}
\{-W(t)\}_{t \geq 0} & \quad \text{(11)} \\
\{W(t + s) - W(s)\}_{t \geq 0} & \quad \text{(12)} \\
\{a W(t/a^2)\}_{t \geq 0} & \quad \text{(13)} \\
\{t W(1/t)\}_{t \geq 0} & \quad \text{(14)}
\end{align*}
\]

**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
\[
\begin{align*}
M(t) & := \max\{W(s) : 0 \leq s \leq t\} \quad \text{and} \\
M^-(t) & := \min\{W(s) : 0 \leq s \leq t\}.
\end{align*}
\]
These are well-defined, because the Wiener process has continuous paths, and continuous functions always attain their maximal and minimal values on compact intervals. Now observe that if the path $W(s)$ is replaced by its reflection $-W(s)$ then the maximum and the minimum are interchanged and negated. But since $-W(s)$ is again a Wiener process, it follows that $M(t)$ and $-M^-(t)$ have the same distribution:
\[
M(t) \overset{\mathcal{D}}{=} -M^-(t).
\]
Next, consider the implications of Brownian scaling. Fix \( a > 0 \), and define

\[
W^*(t) = aW(t/a^2) \quad \text{and} \quad M^*(t) = \max_{0 \leq s \leq t} W^*(s)
\]

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

\[
= aM(t/a^2).
\]

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,

\[
M(t) \overset{D}{=} aM(t/a^2). \tag{18}
\]

**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 \( a \).

## 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

\[
\mathcal{F}_t^W := \sigma(\{W_s\}_{0 \leq s \leq t}) \tag{20}
\]

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,

\[
\{M(t) > a\} = \bigcup_{s \in \mathbb{Q}, 0 \leq s \leq t} \{W(s) > a\}. \tag{21}
\]

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 \( \sigma \)–*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 $\{G_t\}_{t \geq 0}$ indexed by time $t$. A stochastic process $X(t)$ is said to be *adapted* to a filtration $\{G_t\}_{t \geq 0}$ if the random variable $X(t)$ is measurable relative to $G_t$ 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 $\{G_t\}_{t \geq 0}$ a filtration of the probability space on which the Wiener process is defined. The filtration is said to be *admissible* for the Wiener process if (a) the Wiener process is adapted to the filtration, and (b) for every $t \geq 0$, the post-$t$ process $\{W_t+s - W_t\}_{s \geq 0}$ is independent of the $\sigma-$algebra $G_t$.

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 $\{F_t^W\}_{t \geq 0}$ 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

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

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, this is not standard terminology. Some authors (for instance, Karatzas and Shreve) call such filtrations Brownian filtrations.
\( \mathcal{G} \) 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:

\[
E(1_H \mid \mathcal{G}) = E1_H = P(H) \quad \text{almost surely.} \tag{22}
\]

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 (22) 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

\[
\lim_{n \to \infty} g_n(X_{t_1}, X_{t_2}, \ldots, X_{t_m}) = 1_A \quad \text{almost surely};
\]

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

\[
E(g(X_{t_1}, X_{t_2}, \ldots, X_{t_m}) \mid \mathcal{G}) = Eg(X_{t_1}, X_{t_2}, \ldots, X_{t_m}) \quad \text{almost surely.} \tag{23}
\]

Finally, since any bounded, continuous function \( g \) can be arbitrarily well approximated by finite trig polynomials, to prove (23) 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,

\[
E(\exp\{i \sum_{j=1}^{m} \theta_j X_{t_j}\} \mid \mathcal{G}) = E \exp\{i \sum_{j=1}^{m} \theta_j X_{t_j}\} \quad \text{almost surely.} \tag{24}
\]

**Exercise 1.** 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 (23) and (24) have an easy consequence that is sometimes useful.

**Lemma 1.** Let \( \{X^n_t\}_{n \geq 1, t \in J} \) 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 \( \{X^n_t\}_{t \in J} \) is independent of \( \mathcal{G} \). If

\[
\lim_{n \to \infty} X^n_t = X_t \quad \text{a.s. for each } t \in J \tag{25}
\]

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

**Proof.** It suffices to establish the identity (23) 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 \( \{X^n_t\}_{t \in J} \) is independent of \( \mathcal{G} \), equation (23) holds when \( X \) is replaced by \( X^n \). Since \( g \) is continuous, the hypothesis of the lemma implies that

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

Consequently, by the dominated convergence theorem (applied twice, once for ordinary expectations and once for conditional expectations), equation (23) holds for the limit process \( \{X_t\}_{t \in J} \). \( \square \)
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 $t$,

$$\mathcal{F}_{t^+} = \cap_{s > t} \mathcal{F}_s$$  \hspace{1cm} (26)

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 $\mathcal{F}_{t^+}$ 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 $\mathcal{F}^W_0$ is trivial: it contains only the events $\emptyset$ and $\Omega = \emptyset'$. The augmentation $\mathcal{F}^W_{s^+}$, however, contains interesting events, for instance, the events

$$A_+ = \cap_{n \geq 1} \cup_{q \in \mathbb{Q}; q \leq 2^{-n}} \{W_q > 0\} \quad \text{and} \quad A_- = \cap_{n \geq 1} \cup_{q \in \mathbb{Q}; q \leq 2^{-n}} \{W_q < 0\}.$$  \hspace{1cm} (27)

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 $\mathcal{F}^W_{s^+}$, because the inner union in each definition is an event in $\mathcal{F}^W_{s^+}$.

**Proposition 3.** If $\{W_t\}_{t \geq 0}$ is a standard Brownian motion and $\{\mathcal{F}^W_t\}_{t \geq 0}$ is the standard filtration then the augmented filtration $\{\mathcal{F}^W_{t^+}\}_{t \geq 0}$ 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 $\mathcal{F}^W_s$. 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 $\mathcal{F}^W_{s+2^{-m}}$, and since $\mathcal{F}^W_{s+2^{-m}}$ is contained in $\mathcal{F}^W_{s+2^{-m}}$, it follows that the process $\{\Delta W(s + 2^{-m}, t)\}_{t \geq 0}$ is independent of $\mathcal{F}^W_{s^+}$. 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 1 implies that the process $\{\Delta W(s, t)\}_{t \geq 0}$ is independent of $\mathcal{F}^W_{s^+}$. \hspace{1cm} \Box

**Corollary 1.** For every $t \geq 0$ and every integrable random variable $Y$ that is measurable with respect to the $\sigma$–algebra $\mathcal{F}^W_t$ generated by a standard Wiener process,

$$E(Y | \mathcal{F}^W_{t^+}) = E(Y | \mathcal{F}^W_t).$$  \hspace{1cm} (28)

Consequently, for any event $F \in \mathcal{F}^W_t$ there is an event $F' \in \mathcal{F}^W_t$ such that $P(F \Delta F') = 0$.

**Proof.** By the usual approximation arguments, to prove (28) it suffices to consider random variables of the form

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

8
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 (28). The second assertion of the corollary follows by applying (28) to \( Y = 1_F \); this shows that

\[
1_F = E(1_F | \mathcal{F}_{t^+}^W) = E(1_F | \mathcal{F}_t^W) \quad \text{a.s.,}
\]

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

**Corollary 2.** (Blumenthal 0-1 Law) Every event in \( \mathcal{F}_0^W \) has probability 0 or 1.

**Proof.** The \( \sigma \)-algebra \( \mathcal{F}_0^W \) is the trivial \( \sigma \)-algebra \( \{\emptyset, \Omega\} \), because it is generated by the constant random variable \( W_0 \). Therefore, by Corollary 1, each event \( F \in \mathcal{F}_0^W \) differs from either the empty set or \( \Omega \) by an event of probability 0, and so \( P(F) = 0 \) or 1. \( \square \)

Recall that the events \( A_+ \) and \( A_- \) defined by equations (27) are elements of the \( \sigma \)-algebra \( \mathcal{F}_0^W \). 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 \( \bigcup_{q \in \mathbb{Q} : q \leq 2^{-n}} \{W_q > 0\} \) 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,

\[
P(A_+ \cap A_-) = 1. \quad (29)
\]

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

\(^2\)A stopping time that takes only finite values will be called a **finite** stopping time.
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 \( F_t \).

Exercise 2. (a) Prove that if \( \tau, \sigma \) are stopping times then so are \( \sigma \land \tau \) and \( \sigma \lor \tau \).
(b) Prove that if \( \tau \) 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 \( \tau \) 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 \( \tau \).

Exercise 3. (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 4. Let \( \tau \) be a finite stopping time with respect to the standard filtration of a Brownian motion \( \{ W_t \}_{t \geq 0} \) and let \( \mathcal{F}_W^\tau \) be the associated stopping field. Show that each of the random variables \( \tau, W_\tau, \) and \( M_\tau \) is measurable relative to \( \mathcal{F}_W^\tau \).

2.5 Strong Markov Property

Theorem 2. (Strong Markov Property) Let \( \{ W(t) \}_{t \geq 0} \) be a standard Brownian motion, and let \( \tau \) 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-\( \tau \) process
\[
W^*(t) = W(t + \tau) - W(\tau),
\]
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 3. (Splicing) Let \( \{ W_t \}_{t \geq 0} \) be a Brownian motion with admissible filtration \( \{ \mathcal{F}_t \}_{t \geq 0} \), and let \( \tau \) be a stopping time for this filtration. Let \( \{ W_t^* \}_{t \geq 0} \) be a second Brownian motion on the same probability space that is independent of the stopping field \( \mathcal{F}_\tau \). Then the spliced process
\[
\hat{W}_t = \begin{cases} 
W_t & \text{for } t \leq \tau, \\
W_\tau + W_{t-\tau}^* & \text{for } t \geq \tau
\end{cases}
\]
is also a Brownian motion.
The hypothesis that \( \tau \) 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 2 if \( \tau \) is a finite stopping time then so is \( \tau_m = \min\{k/2^m \geq \tau\} \). Clearly, the sequence \( \tau_m \) is non-increasing in \( m \), and \( \tau_m \to \tau \) as \( m \to \infty \). Hence, by Exercise 1, the stopping fields \( \mathcal{F}_{\tau_m} \) are reverse-ordered, and \( \mathcal{F}_{\tau} \subseteq \mathcal{F}_{\tau_m} \) for every \( m \geq 1 \).

By hypothesis, for each \( m \geq 1 \) the post-\( \tau_m \) 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 \( \mathcal{F}_{\tau} \subseteq \mathcal{F}_{\tau_m} \), 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 \( t \geq 0 \),

\[
\lim_{m \to \infty} \Delta W(\tau_m, t) = \Delta W(\tau, t) = W^*(t). \tag{32}
\]

Consequently, by Lemma 1, the post-\( \tau \) 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 \( 0 = t_0 < t_1 < \cdots < t_k \),

\[
E \exp \left\{ i \sum_{j=1}^{k} \theta_j(W_{t_j}^* - W_{t_{j-1}}^*) \right\} = E \exp \left\{ i \sum_{j=1}^{k} \theta_j(W_{t_j} - W_{t_{j-1}}) \right\}, \tag{33}
\]
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 (33) holds when \( W^* \) is replaced by \( \Delta W(\tau_m, t) \) in the expectation on the left side. Consequently, the equality (33) follows by (32) and the dominated convergence theorem.

\[ \square \]

**Step 2:** We must prove that if \( \tau \) is a stopping time that takes values in \( D_m \), for some \( m \geq 0 \), then the post-\( \tau \) process \( \{W^*_t\}_{t \geq 0} \) is a Brownian motion and is independent of \( F_\tau \). 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 \( F_\tau \) then for any \( 0 = t_0 < t_1 < \cdots < t_k \) and \( \theta_1, \theta_2, \ldots, \theta_k \),

\[
E_1 B \exp \left\{ i \sum_{j=1}^k \theta_j (W^*_{t_j} - W^*_{t_{j-1}}) \right\} = P(B)E \exp \left\{ i \sum_{j=1}^k \theta_j (W_{t_j} - W_{t_{j-1}}) \right\}.
\]

(34)

Since \( \tau \) takes only nonnegative integer values, the event \( B \) can be partitioned as \( B = \cup_{n \geq 0} (B \cap \{\tau = n\}) \). Since \( \tau \) is a stopping time, the event \( B \cap \{\tau = n\} \) is in the \( \sigma \)-algebra \( F_n \); moreover, on \( \{\tau = n\} \), the post-\( n \) 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 \( F_n \), and hence independent of \( B \cap \{\tau = n\} \). Consequently,

\[
E_1 B \mathbf{1}_{\{\tau = n\}} \exp \left\{ i \sum_{j=1}^k \theta_j (W^*_{t_j} - W^*_{t_{j-1}}) \right\}
\]

\[
= E_1 B \mathbf{1}_{\{\tau = n\}} \exp \left\{ i \sum_{j=1}^k \theta_j (\Delta W(n, t_j) - \Delta W(n, t_{k-1})) \right\}
\]

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

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 (34).

\[ \square \]

**Proof of the Splicing Theorem**

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 \( \tau \) that take values in \( D_m \), for some \( m \); then one proves the result for this 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.

\[ \square \]

### 3 Embedded Simple Random Walks

**Lemma 2.** 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 \( \tau \leq n + 1 \). \hfill \Box

The lemma guarantees that there will be a first time \( \tau_1 = \tau \) when the Wiener process has traveled \( \pm 1 \) from its initial point. Since this time is a stopping time, the post-\( \tau \) 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 \( \pm 1 \) from its starting point, and so on. Because the post-\( \tau \) process is independent of the stopping field \( F_\tau \), it is, in particular, independent of the random variable \( W(\tau) \), and so the sequence of future \( \pm 1 \) jumps is independent of the first. By an easy induction argument, the sequence of \( \pm 1 \) jumps made in this sequence are independent and identically distributed. Similarly, the sequence of elapsed times are i.i.d. copies of \( \tau \). Formally, define \( \tau_0 = 0 \) and

\[
\tau_{n+1} := \min \{ t > \tau_n : |W_{t+\tau_n} - W_{\tau_n}| = 1 \}
\]

(35)

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 3.** 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. \hfill \Box

There isn’t anything special about the values \( \pm 1 \) 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.

\[
P\{M(t) \geq a\} = P\{\tau_a \leq t\} = 2P\{W(t) > a\} = 2 - 2\Phi(a/\sqrt{t}).
\] (36)

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 3 above, \(\tau < \infty\) almost surely. Since \(\tau\) is a stopping time, the post-\(\tau\) process

\[
W^*(t) := W(\tau + t) - W(\tau)
\] (37)

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 \(\tau\) 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:

\[
\tilde{W}(s) = W(s) \quad \text{for } s \leq \tau,
\]

\[
= 2a - W(s) \quad \text{for } s \geq \tau.
\] (38)

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 3.
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, $\tau$ is the first passage time to the level $a$ for the Brownian motion $W$. Hence,

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

After time $\tau$, 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

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

\[\square\]

**Corollary 4.** The first-passage time random variable $\tau(a)$ is almost surely finite, and has the one-sided stable probability density function of index $1/2$:

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

(39)

**Proof.** Differentiate the equation (36) with respect to the variable $t$. \[\square\]

Essentially the same arguments prove the following.

**Corollary 5.**

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

(40)

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

(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 $p^0_t(x, y)$ 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 $\mathbb{F}_{t \geq 0}$ is an adapted process $\{X_t\}_{t \geq 0}$ such that each $X_t$ is integrable, and such that for any $0 \leq s \leq t < \infty$,

$$E(X_t | \mathcal{F}_s) = X_s \quad \text{almost surely.} \quad (41)$$

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 4. (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 $\tau$ is a bounded stopping time for this filtration, then $X_\tau$ is an integrable random variable and

$$EX_\tau = EX_0. \quad (42)$$

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

$$C + 1 \geq \tau_1 \geq \tau_2 \geq \cdots \geq \cdots,$$

and

$$\tau = \lim_{m \to \infty} \tau_m.$$

Since $\tau_m$ 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 (42) it suffices to show that the sequence $X_{\tau_m}$ 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 $\{\mathcal{F}_m\}$. Any reverse martingale is uniformly integrable. \qed
Remark 2. (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 \( F_n = \sigma(\xi_i)_{i \leq n} \) 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 \( (F_n)_{n \geq 0} \):

(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 \( \mathbb{G} = \{G_t\}_{t \geq 0} \) 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 \( \tau \), each of the following holds:

\[
E W(\tau) = 0; \tag{43}
\]
\[
E(W(\tau))^2 = E \tau; \tag{44}
\]
\[
E \exp\{\theta W(\tau) - \theta^2 \tau/2\} = 1 \quad \forall \theta \in \mathbb{R}; \text{ and} \tag{45}
\]
\[
E \exp\{i\theta W(\tau) + \theta^2 \tau/2\} = 1 \quad \forall \theta \in \mathbb{R}. \tag{46}
\]

Observe that for nonrandom times \( \tau = t \), these identities follow from elementary properties of the normal distribution. Notice also that if \( \tau \) 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 \( \tau \) 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 $G_s$; hence (for instance),
\[
E(\exp\{\theta W_{t+s} - \theta^2(t+s)/2\} \mid G_s) = \exp\{\theta W_s - \theta^2 s/2\}E(\exp\{\theta W_{t+s} - W_s - \theta^2 t/2\} \mid G_s)
\]
\[
= \exp\{\theta W_s - \theta^2 s/2\}E \exp\{\theta W_{t+s} - W_s - \theta^2 t/2\}
\]
\[
= \exp\{\theta W_s - \theta^2 s/2\}.
\]
The Wald identities follow from Doob’s optional sampling formula.

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,
\[
EW(T \wedge n) = 0 \quad \text{and} \quad EW(T \wedge n)^2 = ET \wedge n.
\]
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, $W(T \wedge n) \to W(T)$ 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
\[
P\{W(T) = b\} = \frac{a}{a+b}.
\]
(47)
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 (47), one may now easily obtain
\[
ET = ab.
\]
(48)

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, $\tau$ is a stopping time and $\tau < \infty$ with probability one, but $\tau$ 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 $\theta > 0$,
\[
E \exp\{\theta W(\tau \wedge n) - \theta^2(\tau \wedge n)\} = 1.
\]
(49)
Because the path $W(t)$ does not assume a value larger than $a$ until after time $\tau$, the random variables $W(\tau \wedge n)$ are uniformly bounded by $a$, and so the random variables in equation (49) 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

$$E \exp\{-\lambda \tau a\} = \exp\{-\sqrt{2}\lambda a\}. \tag{50}$$

The only density with this Laplace transform is the one-sided stable density given in equation (39). Thus, the Optional Sampling Formula gives us a second proof of (36).

**Exercise 7. 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 $\tau$.

**Exercise 8. 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 $\theta$, 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 9. 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 $\sqrt{2} \sin k\pi x$ for $k \geq 1$ (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

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

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

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

\footnote{Check a table of Laplace transforms.}
where \( \hat{u}(k) = \sqrt{2} \int_0^1 u(y) \sin(k \pi y) \, dy \) is the \( k \)th Fourier coefficient of \( u \).

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

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

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 \( Z = \{ t \leq 1 : W(t) = 0 \} \) is a homeomorphic image of the Cantor set. These facts may be established using only the formula (36), Brownian scaling, the strong Markov property, and elementary arguments.

6.1 Zero Set of a Brownian Path

The zero set is

\[
Z = \{ t \geq 0 : W(t) = 0 \}.
\]  \( \text{(51)} \)

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

\[
E|Z| = E \int_0^\infty \mathbf{1}_{\{0\}}(W_t) \, dt
= \int_0^\infty E\mathbf{1}_{\{0\}}(W_t) \, dt
= \int_0^\infty P\{W_t = 0\} \, dt
= 0,
\]

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

\[
P\{\mathbb{Q}_+ \cap Z \neq \emptyset\} = 0.
\]  \( \text{(52)} \)
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 (36), 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. □

Proposition 10. With probability one, the zero set $Z$ of a Brownian path is a perfect set, that is, $Z$ is closed, and for every $t \in Z$ there is a sequence of distinct elements $t_n \in Z$ such that $\lim_{n \to \infty} t_n = t$.

Proof. That $Z$ 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 $t \geq 0$ such that $W(t) = 0$. Because $W(q) \neq 0$ almost surely, the random variable $\nu_q$ is well-defined and is almost surely strictly greater than $q$. By the Strong Markov Property, the post-$\nu_q$ 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 $(t - \varepsilon, t)$ 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, Erdős, & 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 $t_\star \in (0, 1)$ the path were differentiable at $t = t_\star$; then for some $\varepsilon > 0$ and some $C < \infty$ it would be the case that

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

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

$$|W((m + i + 1)/4^k) - W((m + i)/4^k)| \leq 16C/4^k \quad \text{for each} \quad i = 0, 1, 2.$$  \hspace{1cm} (54)

I’ll show that the probability of this event is 0. Let $B_{m,k} = B_{k,m}(C)$ be the event that (54) holds, and set $B_k = \bigcup_{m \leq 4k} B_{m,k}$; then by the Borel-Cantelli lemma it is enough to show that (for each $C < \infty$)

$$\sum_{k=1}^{\infty} P(B_m) < \infty.$$ \hspace{1cm} (55)

The trick is Brownian scaling: in particular, for all $s, t \geq 0$ the increment $W_{t+s} - W_t$ is Gaussian with mean 0 and standard deviation $\sqrt{s}$. Consequently, since the three increments in (54) are independent, each with standard deviation $2^{-k}$, and since the standard normal density is bounded above by $1/\sqrt{2\pi}$,

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

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

$$P(B_k) \leq 4^k (32C/2^k \sqrt{2\pi})^3 \leq (32C/\sqrt{2\pi})^3 / 2^k.$$  \hspace{1cm} (56)

This is obviously summable in $k$. \hfill \Box

**Exercise 10. 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

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

(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

\footnote{The constant 16 might really be 32, possibly even 64.}
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 \( \Pi \) is the length of its longest interval \( t_i - t_{i-1} \). If \( \Pi \) is a partition of \([0, t]\) and if \( 0 < s < t \), then the restriction of \( \Pi \) 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 \( \Pi' \) is a refinement of the partition \( \Pi \) if the sequence of points \( t_i \) that defines \( \Pi \) is a subsequence of the sequence \( t'_{j_i} \) that defines \( \Pi' \). A nested sequence of partitions is a sequence \( \Pi_n \) such that each is a refinement of its predecessor. For any partition \( \Pi \) and any continuous-time stochastic process \( X_t \), define the quadratic variation of \( X \) relative to \( \Pi \) by

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

(56)

**Theorem 5.** Let \( \Pi_n \) 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,

\[
\lim_{n \to \infty} QV(W; \Pi_n) = 1.
\]

(57)

**Note 1.** It can be shown, without too much additional difficulty, that if \( \Pi'_n \) is the restriction of \( \Pi_n \) to \([0, t]\) then with probability one, for all \( t \in [0, 1] \),

\[
\lim_{n \to \infty} QV(W; \Pi'_n) = t.
\]

Before giving the proof of Theorem 5, 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 \( n \)th dyadic partition \( D_n[0, t] \) 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 \( s \).

\(^5\)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.
Proposition 12. Let \( \{ W(t) \}_{t \geq 0} \) be a standard Brownian motion. For each \( t > 0 \), with probability one,
\[
\lim_{n \to \infty} QV(W; D_n[0,t]) = t.
\] (58)

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
\[
\xi_{n,k} \overset{d}{=} 2^n (W(k/2^n) - W((k - 1)/2^n))^2, \quad k = 1, 2, \ldots, 2^n
\]
are independent, identically distributed \( \chi^2 \) 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
\[
QV(W; D_n[0,1]) = 2^{-n} \sum_{k=1}^{2^n} \xi_{n,k}.
\]
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 \( \chi^2 \) 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
\[
P\{|QV(W; D_n[0,1]) - 1| \geq \varepsilon\} = P\{|\sum_{k=1}^{2^n} (\xi_{n,k} - 1)| \geq 2^n \varepsilon\} \leq \frac{E\xi_{1,1}^2}{4^n \varepsilon^2}.
\]
Since \( \sum_{n=1}^{\infty} 1/4^n < \infty \), the Borel–Cantelli lemma implies that, with probability one, the event \( |QV(W; D_n[0,1]) - 1| \geq \varepsilon \) occurs for at most finitely many \( n \). Since \( \varepsilon > 0 \) can be chosen arbitrarily small, it follows that \( \lim_{n \to \infty} QV(W; D_n[0,1]) = 1 \) almost surely. The same argument shows that for any dyadic rational \( t \in [0, 1] \), the convergence (58) holds a.s.

Exercise 11. Prove that if (58) holds a.s. for each dyadic rational in the unit interval, then with probability one it holds for all \( t \).
Proof. Expand the square, and use the fact that $\xi^2$ and $\zeta^2$ are $\mathcal{G}$–measurable to extract them from the conditional expectation. What’s left is

$$2E(\xi \zeta | \mathcal{G}) = 2E(\text{sgn}(\xi)\text{sgn}(\zeta) | \xi | \mathcal{G})$$

$$= 2|\xi| |\zeta| E(\text{sgn}(\xi)\text{sgn}(\zeta) | \mathcal{G})$$

$$= 0,$$

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

Proof of Theorem 5. Without loss of generality, we may assume that each partition $\Pi_{n+1}$ is gotten by splitting one interval of $\Pi_n$, and that $\Pi_0$ is the trivial partition of $[0,1]$ (consisting of the single interval $[0,1]$). Thus, $\Pi_n$ consists of $n+1$ nonoverlapping intervals $J^n_k = [t^n_{k-1}, t^n_k]$. Define

$$\xi_{n,k} = W(t^n_k) - W(t^n_{k-1}),$$

and for each $n \geq 0$ let $\mathcal{G}_n$ be the $\sigma$–algebra generated by the random variables $\xi^2_{m,k}$, 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 3 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,

$$\lim_{n \to \infty} QV(W; \Pi_n) = E(QV(W; \Pi_0) | \cap_{n \geq 0} \mathcal{G}_n) = E(W^2_1 | \cap_{n \geq 0} \mathcal{G}_n) \quad \text{almost surely.}$$

Exercise 12. Prove that the limit is constant a.s., and that the constant is 1.

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 an appropriate sense.

Theorem 6. (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 7. (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. There is a stopping time \( T \) (for the standard filtration) with expectation \( ET = \sigma^2 \) such that the random variable \( W(T) \) has distribution \( F \).

I will only prove Theorem 6. The proof will hinge on the following 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 \( \mathbb{R} \) with mean zero there exists a Borel probability distribution \( G \) on \( \mathbb{R}_+^2 \) such that

\[
F = \int F_{a,b} \, dG(a,b).
\]

Hence,

\[
\int x^2 \, dF(x) = \int ab \, dG(a,b).
\]

**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 (60) 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:

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

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 (60) holds by the induction hypothesis. The formula (61) follows from (60) 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

\[X_n = \max\{k/2^n : k/2^n \leq X\}\]

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 → 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 (60) 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 that

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

\[\square\]

**Exercise 13.** Finish the proof: Show that the representation (60) holds for all mean-zero, finite variance distributions.

**Remark 3.** 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 (60) is a special case of Choquet’s theorem.

**Proof of Theorem 3** 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 \(\{W_t\}_{t \geq 0}\) 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 (47)–(48) 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 \( (T_n - T, W_{T_{n+1}} - W_{T_n}) \) for \( n < m \) (since these are all measurable with respect to the stopping field \( F_{T_m} \)), 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.

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

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

\[
d = \frac{pb + pc}{pb + pc}.
\]

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

Exercise 14. (A) Check that the distribution of \( W_T \) is \( F \).
(B) Check that \( E T = \sigma^2 \).
(C) Complete the induction step.

Proof for the uniform distribution on (-1,1). The general case is proved using the special case of finitely supported distributions by taking a limit. Let’s first consider the special case where \( F \) is the uniform distribution (normalized Lebesgue measure) on \([-1,1]\). Define a sequence of stopping times \( \tau_n \) 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 \(n \to \infty\) 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 \(\tau_n\) 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 \(\pm 1\). (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 \([-1, 1]\).

**Exercise 15.** Show that if \(\tau_n\) is an increasing sequence of stopping times such that \(\tau = \lim \tau_n\) is finite with probability one, then \(\tau\) is a stopping time.

**Proof in the general case.** The strategy will be to approximate \(F\) by a sequence \(F_n\) of distributions such that \(F_n\) is supported by \(2^n\) (or fewer) points. The distributions \(F_n\) are constructed inductively. First, decompose \(F\) as a mixture of two probability distributions \(F_+, F_-\), where \(F_+\) is supported by \([0, \infty)\) and \(F_-\) by \((\infty, 0)\); in particular,

\[
F = \alpha F_+ + (1 - \alpha) F_-,
\]

where \(\alpha = F([0, \infty)) = 1 - F(\infty, 0)\). \(F_+\) and \(F_-\) are the normalized restrictions of \(F\) to the half-lines \([0, \infty)\) and \((\infty, 0)\), respectively. Let \(\mu_+\) and \(\mu_-\) be the means of the distributions \(F_+\) and \(F_-\). Since \(F\) has mean \(0\), it must be the case that \(\alpha \mu_+ + (1 - \alpha) \mu_- = 0\). Define \(F_1\) to be the mean-zero two-point distribution supported by \(\{\mu_+, \mu_-\}\):

\[
F_1 = \alpha \delta_{\mu_+} + (1 - \alpha) \delta_{\mu_-}.
\]

Define \(\tau_1\) to be the first time that \(W_{\tau_1} = \mu_+\) or \(\mu_-\); then clearly the distribution of \(W_{\tau_1}\) is \(F_1\).

The distribution \(F_2\) is now constructed repeating the steps above for each of the distributions \(F_-\) and \(F_+\). First, decompose \(F_+\) as a mixture of two distributions \(F_{++}\) and \(F_{++}\) supported by the intervals \([0, \mu_+)\) and \([\mu_+, \infty)\):

\[
F_+ = \alpha_+ F_{++} + (1 - \alpha_+) F_{++};
\]

and similarly, decompose \(F_-\) as a mixture of distributions \(F_{--}\) and \(F_{++}\) supported by the intervals \((\infty, \mu_-)\) and \((\mu_-, \infty)\):

\[
F_- = \alpha_- F_{--} + (1 - \alpha_-) F_{++}.
\]

Thus, \(F_{++}, F_{--}, F_{++}, F_{++}\) are the normalized restrictions of \(F\) to the four intervals with endpoints \(\mu_-, 0, \mu_+\) determined by the first step of the construction, and the constants \(\alpha_\pm\) satisfy

\[
\alpha_\pm = F([\mu_+, \infty)) \quad \text{and} \quad \alpha_- (1 - \alpha_\pm) = F(\infty, \mu_-).
\]

\(\square\)