Chapter 9

A Taste of Stochastic Integration

As these notes started with measure theory and integration, it is only appropriate that they end with stochastic integration. Although it is one of the highlights of the theory of continuous-time stochastic processes, its analysis, and more generally the analysis of continuous-time processes, has inherent technical difficulties that are insurmountable in the amount of time that is left to us. Therefore, I will conclude these lectures with a very incomplete, somewhat nonrigorous, but hopefully coherent introduction to aspects of stochastic integration. You can learn much more about this topic by reading more specialized texts.\(^1\)

1 The Indefinite Itô Integral

Rather than present a general theory of stochastic integration, I will discuss a special case that is: (i) Broad enough to be applicable for our needs. (ii) Concrete enough so as to make the main ideas clear. You can find a much more in-depth treatment in Dellacherie and Meyer (1982).

If \( H := \{H(s); s \geq 0\} \) is a “nice” stochastic process, we follow more-or-less Itô (1944),\(^2\) and construct the stochastic integral \( \int H \, dW := \int_0^\infty H(s) \, W(ds) \) although with probability one \( W \) is nowhere differentiable (Theorem 8.16). Now let us go ahead and officially redefine what we mean by a stochastic process in continuous-time.\(^3\)

**Definition 9.1** A stochastic process (or process for brevity) \( X := \{X(t); t \geq 0\} \) is a product-measurable function \( X : [0,\infty) \times \Omega \to \mathbb{R} \). We often write \( X(t) \) in place of \( X(t,\omega) \); this is similar to what we did in discrete-time.

In particular, the Brownian motion of the previous section is indeed a stochastic process (check!).

---

\(^1\)Some of my personal favorites are Karatzas and Shreve (1991), McKean (1969), and Revuz and Yor (1999),

\(^2\)See also Bru and Yor (2002) to learn about the recently-discovered work of W. Doeblin on stochastic integrals that paralleled a portion of Itô’s work. This work was developed, independently and at around the same time, as the work of Itô. However, it was interrupted, and later forgotten until quite recently due to the untimely death of Doeblin during the second world war.

\(^3\)This is a nonstandard definition, but reduces technical difficulties without endangering the essence of the theory. One can often show that our notion of a stochastic process is a consequence of much weaker technical assumptions on \( X \); cf. Doob (1953, Chapter II) and Khoshnevisan (2002, Chapter 5.2) under the general heading of “separability.”
Now if \( H \) is nicely-behaved, then it stands to reason that we should define \( \int H \, dW \) as \( \lim_{n \to \infty} \mathcal{I}_n(H) \), where “\( \lim_{n \to \infty} \)” implies an as-yet-unspecified form of a limit, and
\[
\mathcal{I}_n(H) := \sum_{k=0}^{\infty} H \left( \frac{k}{2^n} \right) \left[ W \left( \frac{k+1}{2^n} \right) - W \left( \frac{k}{2^n} \right) \right].
\] (9.1)

It is abundantly clear that \( \mathcal{I}_n(H) \) is a well-defined random variable if, for instance, \( H \) has compact support; i.e., \( H(s) = 0 \) for all \( s \) sufficiently large. The following performs some of the requisite book-keeping about \( n \mapsto \mathcal{I}_n(H) \).

**Lemma 9.2** Suppose there exists a (random or nonrandom) \( T > 0 \) such that with probability one, \( H(s) = 0 \) for all \( s \geq T \). Then, \( \mathcal{I}_n(H) \) is a.s. a finite sum, and
\[
\mathcal{I}_{n+1}(H) - \mathcal{I}_n(H) = \sum_{j=0}^{\infty} \left[ H \left( \frac{2j+1}{2^{n+1}} \right) - H \left( \frac{j}{2^n} \right) \right] \left[ W \left( \frac{j+1}{2^n} \right) - W \left( \frac{2j+1}{2^{n+1}} \right) \right].
\] (9.2)

**Proof (Optional)** The fact that the sum is finite is obvious. Let us derive the stated identity for \( \mathcal{I}_{n+1}(H) - \mathcal{I}_n(H) \).

We consider \( \mathcal{I}_{n+1}(H) = \sum_{k=0}^{2^{n+1}-1} H(k2^{-n-1})[W((k+1)2^{-n-1}) - W(k2^{-n-1})] \), and split this sum according to whether \( k = 2j \) or \( k = 2j+1 \):
\[
\mathcal{I}_{n+1}(H) = \sum_{j=0}^{\infty} H \left( \frac{j}{2^n} \right) \left[ W \left( \frac{2j+1}{2^{n+1}} \right) - W \left( \frac{j}{2^n} \right) \right]
+ \sum_{j=0}^{\infty} H \left( \frac{2j+1}{2^{n+1}} \right) \left[ W \left( \frac{j+1}{2^n} \right) - W \left( \frac{2j+1}{2^{n+1}} \right) \right]
= \sum_{j=0}^{\infty} H \left( \frac{j}{2^n} \right) \left[ W \left( \frac{2j+1}{2^{n+1}} \right) - W \left( \frac{j}{2^n} \right) \right]
+ \sum_{j=0}^{\infty} H \left( \frac{2j+1}{2^{n+1}} \right) \left[ W \left( \frac{j+1}{2^n} \right) - W \left( \frac{2j+1}{2^{n+1}} \right) \right]
- \sum_{j=0}^{\infty} \left[ H \left( \frac{2j+1}{2^{n+1}} \right) - H \left( \frac{j}{2^n} \right) \right] \left[ W \left( \frac{j+1}{2^n} \right) - W \left( \frac{2j+1}{2^{n+1}} \right) \right]
= \mathcal{I}_n(H) - \sum_{j=0}^{\infty} \left[ H \left( \frac{2j+1}{2^{n+1}} \right) - H \left( \frac{j}{2^n} \right) \right] \left[ W \left( \frac{j+1}{2^n} \right) - W \left( \frac{2j+1}{2^{n+1}} \right) \right],
\] (9.3)
as claimed.

\( \square \)

**Definition 9.3** A stochastic process \( H := \{ H(s); t \geq 0 \} \) is said to be adapted to the Brownian filtration \( \mathcal{F} \) if for each \( s \geq 0 \), \( H(s) \) is \( \mathcal{F}_s \)-measurable. It is a compact-support process if there exists a nonrandom \( T \geq 0 \) such that

\( \int H \, dW \) is properly defined. This is the hallmark of Itô's theory of stochastic integration, and unlike the Riemann theory, it cannot be replaced by other rules such as the midpoint- or the right-hand-rule without changing the resulting stochastic integral.

\( \footnote{Notice the left-hand-rule approximation is being used here. This is the hallmark of Ito's theory of stochastic integration, and unlike the Riemann theory, it cannot be replaced by other rules such as the midpoint- or the right-hand-rule without changing the resulting stochastic integral.} \)
that with probability one, \(H(s) = 0\) for all \(s \geq T\). Finally, given \(p \geq 1\), \(H\) is said to be \emph{Dini-continuous in \(L^p(P)\)} if \(H(s) \in L^p(P)\) for all \(s \geq 0\), and 
\[
\int_0^t \psi_p(r)r^{p-1}\,dr < +\infty, \quad \text{where } \psi_p(r) := \sup_{s,t:|s-t| \leq r} \|H(s) - H(t)\|_p
\]
denotes the \emph{modulus of continuity of \(H\) in \(L^p(P)\)}.

\(\Box\)

**Example 9.4** (a) Note that whenever \(H\) is compact-support, continuous, and a.s.-bounded by a nonrandom quantity, then it is a.s. uniformly continuous in \(L^p(P)\) for any \(p \geq 1\). In particular, \(\psi_p(t) \to 0\) as \(t \to 0\). The extra assumption of Dini-continuity in \(L^p(P)\) states that in fact \(\psi_p\) has to converge to zero at some minimum rate. Here are some examples:

(b) Suppose \(H\) is (a.s.) differentiable with a derivative that satisfies \(K := \sup_1 \|H'(t)\|_p < +\infty\).\(^5\) By the fundamental theorem of calculus, if \(t \geq s \geq 0\), then 
\[
\|H(s) - H(s)\|_p \leq \int_s^t \|H'(r)\|_p \,dr \leq K|t - s|.
\]
Therefore, \(\psi_p(t) \leq Kt\), and \(H\) is easily seen to be Dini-continuous in \(L^p(P)\).

(c) Another class of important examples is found by considering processes \(H\) of the form \(H(s) := f(W(s))\), where \(f\) is a nonrandom differentiable function with \(L := \sup_x |f'(x)| < +\infty\). In such a case, \(\|H(s) - H(t)\| \leq L\sqrt{|s-t|}\), and we have \(\psi_p(t) \leq c_p\sqrt{|s-t|}\), where \(c_p = \|Z\|_p\) where \(Z\) is a standard normal variable (why?). This yields the Dini-continuity of \(H\) in \(L^p(P)\) for any \(p \geq 1\).

(d) More generally, suppose we have \(H(s) := f(W(s), s)\), where \(f(x, t)\) is nonrandom, differentiable in each variable, and satisfies: (i) There exists a nonrandom \(T \geq 0\) such that \(f(x, s) = 0\) for all \(s \geq T\); and (ii) \(M := \sup_{x,t} |\partial_x f(x, t) + \partial_t f(x, t)| < +\infty\).\(^6\) Then, 
\[
\|H(s) - H(t)\| \leq |f(W(s), s) - f(W(t), s)| + |f(W(t), s) - f(W(t), t)|.
\]
Applying the fundamental theorem of calculus, we arrive at the following (why?):
\[
\|H(s) - H(t)\| \leq M (|W(s) - W(t)| + |t - s|).
\]
By Minkowski’s inequality (Theorem 2.24), for any \(p \geq 1\), 
\[
\|H(s) - H(s)\|_p \leq M (\|W(s) - W(t)\|_p + |t - s|) = M(c_p|t - s|^{1/2} + |t - s|),
\]
where \(c_p = \|Z\|_p\) (why?). In particular, whenever \(r \in [0,1]\), we have 
\(\psi_p(t) \leq M(c_p + 1)\sqrt{r}\), from which the Dini-continuity of \(H\) follows in any \(L^p(P)(p \geq 1)\).

\(\Box\)

**Remark 9.5** (Cauchy Summability test) Dini-continuity in \(L^p(P)\) is equivalent to the summability of \(\psi_p(2^{-n})\). Indeed, we can write 
\[
\int_0^1 \psi_p(t)t^{-1}\,dt = \sum_{n=0}^\infty \int_{2^{-n-1}}^{2^{-n}} \psi_p(t)\,dt.
\]
Because \(\psi_p\) is nondecreasing, for the \(t\) in this integral, \(\psi_p(2^{-n-1}) \leq \psi_p(t) \leq \psi_p(2^{-n})\), and \(2^n \leq t^{-1} \leq 2^{n+1}\). Therefore, 
\[
\frac{1}{2}\sum_{n=1}^\infty \psi_p(2^{-n}) = \frac{1}{2}\sum_{n=0}^\infty \psi_p(2^{-n-1}) \leq \int_0^1 \frac{\psi_p(t)}{t}\,dt \leq \sum_{n=0}^\infty \psi_p(2^{-n}) .
\]
In particular, \(H\) is Dini continuous in \(L^p(P)\) if and only if \(\sum_n \psi_p(2^{-n}) < +\infty\). This method is generally ascribed to A. L. Cauchy.

\(\Box\)

We can now define \(\int H\,dW\) for adapted compact-support processes that are Dini-continuous in \(L^2(P)\).\(^7\) I will then show how one can improve the assumptions on \(H\).

---

\(^5\)Since \((s, \omega) \mapsto H(s, \omega)\) is product-measurable, \(\int |H'(r)|^p\,dr\) is a random variable, and hence \(\|H'(t)\|_p\) are well-defined (check this).

\(^6\)As is customary, \(\partial g(x, y, z, w, \cdots)\) denotes the derivative of \(g\) with respect to the variable \(x\).

\(^7\)In fact, one needs much less than Dini continuity in \(L^2(P)\) to proceed, but this is all that we need.
Theorem 9.6 (Itô 1944) Suppose $H$ is an adapted compact-support stochastic process that is Dini-continuous in $L^2(\mathbb{P})$. Then, $\int H \, dW := \lim_{n \to \infty} \mathcal{I}_n(H)$ exists in $L^2(\mathbb{P})$, and $\int H \, dW$ has mean zero and variance

$$E \left\{ \left( \int H \, dW \right)^2 \right\} = E \left\{ \int_0^\infty H^2(s) \, ds \right\}. \quad (9.6)$$

Finally, if $a, b \in \mathbb{R}$, and $V$ is another adapted compact-support stochastic process that is Dini-continuous in $L^2(\mathbb{P})$, then with probability one,

$$\int (aH + bV) \, dW = a \int H \, dW + b \int V \, dW. \quad (9.7)$$

Definition 9.7 Equation (9.6) is fundamental to Itô’s theory of stochastic integration, and is called the Itô isometry.

Proof We employ Lemma 9.2, square both sides of the equation therein, and take expectations, and obtain

$$\|\mathcal{I}_{n+1}(H) - \mathcal{I}_n(H)\|_2^2 = \sum_{0 \leq j \leq 2^n T - 1} \left\| H \left( \frac{2j + 1}{2^n} \right) - H \left( \frac{j}{2^n} \right) \right\|_2^2 \times \left\| W \left( \frac{j + 1}{2^n} \right) - W \left( \frac{2j + 1}{2^n} \right) \right\|_2^2. \quad (9.8)$$

To obtain this, we only need the facts that: (i) For $t > s$, $W(t) - W(s)$ is independent of $\mathcal{F}_s$ (Theorem 8.27); and (ii) $H(u)$ is adapted to $\mathcal{F}_s$ for $u \leq s$. But $\|W(s) - W(t)\|_2^2 = t - s$. Hence, in the preceding display, the expectation involving Brownian motion is equal to $2^{-n} \cdot 2^{-n-1} = 2^{-n-1}$, whereas the first expectation there is no more than $\psi_2(2^{-n-1})$. Consequently, $\|\mathcal{I}_{n+1}(H) - \mathcal{I}_n(H)\|_2 \leq \sqrt{T} \psi_2(2^{-n-1})$. In particular, we can use the monotonicity of $\psi_2$ to see that for any nonrandom $N, M \geq 1$,

$$\|\mathcal{I}_{N+M}(H) - \mathcal{I}_N(H)\|_2 \leq \sum_{n=N+1}^{N+M-1} \|\mathcal{I}_{n+1}(H) - \mathcal{I}_n(H)\|_2 \leq \sqrt{T} \sum_{n=N+1}^{N+M-1} \psi_2(2^{-n-1}). \quad (9.9)$$

Thanks to this and Dini continuity (Remark 9.5) in $L^2(\mathbb{P})$, the above goes to zero as $N, M \to \infty$; i.e., $n \mapsto \mathcal{I}_n(H)$ is a Cauchy sequence in $L^2(\mathbb{P})$, which proves the assertion about $L^2(\mathbb{P})$-convergence.

To compute the expectation of $\int H \, dW$, I need only to note that $E\{\mathcal{I}_n(H)\} = 0$ (why?); this is a consequence of the fact that for $t > s$, $W(t) - W(s)$ has mean zero, and is independent of $\mathcal{F}_s$, whereas $H(u)$ is $\mathcal{F}_s$-measurable for each $u \leq s$. Similarly, we can prove (9.6):

$$E \left\{ \left( \int H \, dW \right)^2 \right\} = \lim_{n \to \infty} \|\mathcal{I}_n(H)\|_2^2 = \lim_{n \to \infty} \sum_{k=0}^\infty E \left\{ H^2(k2^{-n}) \right\} 2^{-n} = E \left\{ \int_0^\infty H^2(s) \, ds \right\}, \quad (9.10)$$

where the many exchanges of limits and integrals are all justified by the compact-support assumption on $H$, together with the continuity of the function $t \mapsto \|H(t)\|_2$ (check this!).

Finally, I need to verify (9.7); but this follows from the linearity of $H \mapsto \mathcal{I}_n(H)$ and the existence of $L^2(\mathbb{P})$-limits.

We now drop many of the technical assumptions in Theorem 9.6.
Theorem 9.8 (Itô 1944) Suppose that $H$ is an adapted stochastic process, and $\mathbb{E}\{\int_0^\infty H^2(s) \, ds\} < +\infty$. Then one can define a stochastic integral $\int H \, dW$ that has mean zero and variance $\mathbb{E}\{\int_0^\infty H^2(s) \, ds\}$. Moreover, if $V$ is another such integrand-process, then for all $a,b \in \mathbb{R}$

\[
\int (aH + bV) \, dW = a \int H \, dW + b \int V \, dW, \quad \text{a.s.}
\]

(9.11)

\[
\mathbb{E}\left\{ \int H \, dW \cdot \int V \, dW \right\} = \mathbb{E}\left\{ \int_0^\infty H(s)V(s) \, ds \right\}.
\]

Throughout, let $m$ denote the Lebesgue measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, and let $L^2(m \times \mathbb{P})$ denote the corresponding product $L^2$-space. In particular, note that

\[
\mathbb{E}\left\{ \int_0^\infty H^2(s) \, ds \right\} = \int_{\Omega \times [0,\infty)} H^2 \, dm \, d\mathbb{P} = ||H||^2_{L^2(m \times \mathbb{P})},
\]

(9.12)

and $\mathbb{E}\{\int H(s) V(s) \, ds\}$ is the inner product—in $L^2(m \times \mathbb{P})$—between $H$ and $V$.

The following technical result is the main step in constructing the general stochastic integral.

Proposition 9.9 Given any stochastic process $H \in L^2(m \times \mathbb{P})$ we can find stochastic processes $H_1, H_2, \ldots$, all compact-support and Dini-continuous in $L^2(\mathbb{P})$, such that $\lim_n H_n = H$ in $L^2(m \times \mathbb{P})$.

Theorem 9.8 follows from immediately from this.

Proof of Theorem 9.8 Thanks to Proposition 9.9 we can find adapted stochastic processes $H_n$ that are compact-support Dini-continuous in $L^2(\mathbb{P})$, and converge to $H$ in $L^2(m \times \mathbb{P})$. Thanks to the Itô isometry (equation 9.6), $\int H_n \, dW$ is a Cauchy sequence in $L^2(\mathbb{P})$, since $H_n$ is a Cauchy sequence in $L^2(m \times \mathbb{P})$. Consequently, $\int H \, dW := \lim_n \int H_n \, dW$ exists in $L^2(\mathbb{P})$. The properties of $\int H \, dW$ follow readily from those of $\int H_n \, dW$, and the $L^2(\mathbb{P})$-convergence that we proved earlier.

I will conclude this section by proving the one remaining proposition.

Proof of Proposition 9.9 (Optional) I will proceed in three steps. Each step reduces the problem to a more restrictive class of processes $H$.

Step 1. Reduction to the Compact-Support Case. Define $H_n(t) := H(t)1_{[0,n]}(t)$, and note that $H_n$ is an adapted compact-support stochastic process. Moreover,

\[
\lim_{n \to \infty} ||H - H_n||^2_{L^2(m \times \mathbb{P})} = \lim_{n \to \infty} \mathbb{E}\left\{ \int_0^\infty H^2(s) \, ds \right\} = 0.
\]

(9.13)

In other words, for the remainder of the proof, we can and will assume without loss of generality that $H$ is also compact-support.

Step 2. Reduction to the $L^2$-Bounded Case. I first extend the definition of $H$ to all $\mathbb{R}$ by assigning $H(t) = 0$ if $t < 0$. Next, define for all $n \geq 1$,

\[
H_n(t) := n \int_{t-(1/n)}^t H(s) \, ds, \quad \forall t \geq 0.
\]

(9.14)

133
You should check that $H$ is an adapted stochastic process. Moreover, for all $t \geq T + 1$, $H_n(t) = 0$, so that $H_n$ is also compact-support. Next, I claim that $H_n$ is bounded in $L^2(P)$; i.e., $\sup_n \|H_n(t)\|^2 < +\infty$. Indeed, by the Cauchy–Bunyakovsky–Schwarz inequality (Corollary 2.25), and by the Fubini–Tonelli theorem (Theorem 3.6),

$$
\sup_{t \geq 0} \|H_n(t)\|^2 \leq \sup_{t \geq 0} \int_{t-1/n}^t \|H(s)\|^2 \, ds \leq n \int_0^\infty \|H(s)\|^2 \, ds = n\|H\|^2_{L^2(m \times P)}. \tag{9.15}
$$

(Why?) It remains to prove that $H_n$ converges in $L^2(m \times P)$ to $H$.

Since $\int_0^\infty H^2(s) \, ds < +\infty$ a.s., then thanks to the Lebesgue differentiation theorem (Theorem 7.49), with probability one, $H_n(t) \to H(t)$ for almost every $t \geq 0$. Therefore, by Fubini–Tonelli (Theorem 3.6), $\lim_n H_n = H$, $(m \times P)$-almost surely (why?). According to the dominated convergence theorem (Theorem 2.22), to finish this step we need to only prove that $\sup_n \|H_n\| \in L^2(m \times P)$. In fact, I will prove (9.17) below which is slightly stronger still.

Now note that $\sup_n \|H_n\| \leq \mathcal{M}H$, where the latter is the “maximal function,”

$$
\mathcal{M}H(t) := \sup_{n \geq 1} \left( n \int_{t-1/n}^t |H(s)| \, ds \right), \quad \forall t \geq 0. \tag{9.16}
$$

For each $\omega$, $\mathcal{M}H(t + n^{-1})$ our good-old Hardy–Littlewood maximal function of $H$, and you should check that $\mathcal{M}H$ is an adapted stochastic process, and by applying Corollary 7.51 with $p := 2$, we obtain: $\int_0^\infty |\mathcal{M}H(s)|^2 \, ds \leq 64 \int_0^\infty H^2(s) \, ds$. This is useful only if the right-hand side is finite. But since $H(t) = 0$ for all $t \geq T$ and $\sup_n \|H(t)\|_2 < \infty$, the right-hand side of the preceding inequality is finite for almost-all $\omega$. In particular, we can appeal to Fubini–Tonelli (Theorem 3.6) to take expectations, and then square-roots to deduce that

$$
\|\mathcal{M}H\|_{L^2(m \times P)} \leq 8\|H\|_{L^2(m \times P)}, \tag{9.17}
$$

which is finite. This is the desired inequality, and reduces the problem to $H$’s that are bounded in $L^2(P)$ and compact-support.

Step 3. The Conclusion.

Finally, if $H$ is bounded in $L^2(P)$ and compact-support, then we define $H_n$ by (9.14) and note that $H_n$ is differentiable, and $H_n(t) = n\{H(t) - H(t - n^{-1})\}$. Therefore, $\sup_{t} \|H_n(t)\|_2 \leq 2n \sup_{t} \|H(t)\|_2 < +\infty$, and part (b) of Example 9.4 proves the asserted Dini-continuity of $H_n$. On the other hand, the argument developed in Step 2 proves that $H_n \to H$ in $L^2(m \times P)$, and this concludes the proof. \[\square\]

2 Continuous Martingales in $L^2(P)$

The theories of continuous-time martingale and stochastic integration are intimately connected. Thus, before proceeding further, we take a side-step, and have a quick look at martingale-theory in continuous-time. To avoid unnecessary abstraction, $\mathcal{F}$ will denote the Brownian filtration throughout.\[^8\]

**Definition 9.10** A process $M := \{M(t); t \geq 0\}$ is a (continuous-time) **martingale** if:

1. For all $t \geq 0$, $M(t) \in L^1(P)$.

[^8]: This section’s proofs are optional reading.
2. If \( t \geq s \geq 0 \), then \( \mathbb{E}\{M(t) \mid \mathcal{F}(s)\} = M(s) \), a.s.

\( M \) is said to be a \textit{continuous} \( L^2 \)-\textit{martingale} if \( t \mapsto M(t) \) is almost-surely continuous, and \( M(t) \in L^2(\mathbb{P}) \) for all \( t \geq 0 \).

Essentially all of the theory of martingales in discrete-time has continuous-time translations for continuous \( L^2(\mathbb{P}) \)-martingales. Here is a first sampler.

\textbf{Theorem 9.11 (Optional Stopping Theorem)} \textit{If} \( M \) \textit{is a continuous} \( L^2 \)-\textit{martingale} and \( S \leq T \) \textit{are bounded} \( \mathcal{F} \)-\textit{stopping times}, \textit{then} \( \mathbb{E}\{M(T) \mid \mathcal{F}_S\} = M(S) \), a.s.

\textbf{Proof (Optional)} \textit{Throughout, choose and fix some nonrandom} \( K > 0 \) \textit{such that almost-surely,} \( T \leq K \).

This is a consequence of Theorem 7.33 if \( S \) and \( T \) are simple stopping times. In general, let \( S_n \downarrow S \) and \( T_n \downarrow T \) be the simple stopping times of Lemma 8.26, and note that the condition \( S \leq T \) imposes \( S_m \leq T_n \) \textit{for all} \( n \leq m \) \textit{(why?)}. \textit{We have already seen that for all} \( n \geq m \),

\[ \mathbb{E}\left\{ M(T_n) \mid \mathcal{F}_{S_m} \right\} = M(S_m), \text{ a.s.} \tag{9.18} \]

Moreover, \textit{this very argument implies that} \( M(T_1), M(T_2), \ldots \) \textit{is a (discrete-time) martingale in its own filtration. Since} \( T_n \leq T + 2^{-n} \leq K + 2^{-n} \), \textit{by Exercise 7.9},

\[ \mathbb{E}\left\{ \sup_{n \geq 1} M^2(T_n) \right\} \leq 4 \sup_{n \geq 1} \mathbb{E}\left\{ M^2(T_n) \right\} \leq 4 \mathbb{E}\left\{ M^2 \left( K + \frac{1}{2} \right) \right\} < +\infty. \tag{9.19} \]

Since \( M \) \textit{is continuous, a.s.,} \( M(T_n) \to M(T) \), a.s. \textit{Therefore, by the dominated convergence theorem (Theorem 2.22), we also have} \( M(T_n) \to M(T) \) \textit{in} \( L^2(\mathbb{P}) \). \textit{This} \textit{and conditional Fatou’s lemma (Theorem 7.6) together imply that}

\[ \lim_{n \to \infty} \left\| \mathbb{E}\left\{ M(T_n) \mid \mathcal{F}_{S_m} \right\} - \mathbb{E}\left\{ M(T) \mid \mathcal{F}_{S_m} \right\} \right\|_2 \leq \lim_{n \to \infty} \| M(T) - M(T_n) \|_2 = 0. \tag{9.20} \]

Therefore, \textit{by (9.18),} \( M(S_m) = \mathbb{E}\{M(T) \mid \mathcal{F}_{S_m}\}, \text{ a.s.} \) \textit{To} \textit{finish the proof,} \textit{we simply let} \( m \to \infty \), \textit{and appeal to the time-reversed martingale convergence theorem (in discrete time; Theorem 7.47).} \hfill \Box

The following is a related result whose proof is relegated to the exercises.

\textbf{Theorem 9.12 (Doob’s Maximal Inequalities)} \textit{If} \( M \) \textit{is a continuous} \( L^2 \)-\textit{martingale, then for all} \( \lambda, t > 0 \),

\[ \mathbb{P}\left\{ \sup_{0 \leq s \leq t} |M(s)| \geq \lambda \right\} \leq \frac{1}{\lambda} \mathbb{E}\left\{ |M(t)| \mid \mathbb{P}\left\{ \sup_{0 \leq s \leq t} |M(s)| \geq \lambda \right\}. \tag{9.21} \]

\textit{In} \textit{particular, for all} \( p > 1 \),

\[ \mathbb{E}\left\{ \sup_{0 \leq s \leq t} |M(s)|^p \right\} \leq \left( \frac{p}{p - 1} \right)^p \mathbb{E}\{ |M(t)|^p \}. \tag{9.22} \]
3 The Definite Itô Integral

It is a good time to also mention the definite Itô integral, which is simply defined as \( \int_0^t H \, dW := \int H_1 \, dW \) for all adapted processes \( H \) such that \( \mathbb{E}\{\int_0^t H^2(s) \, ds\} < +\infty \) for all \( t \geq 0 \). This defines a collection of random variables \( \int_0^t H \, dW \) — one for each \( t \geq 0 \). The following is of paramount importance, since it says something about the properties of the random function \( t \mapsto \int_0^t H \, dW \).

**Theorem 9.13** If \( H \) is an adapted stochastic process such that \( \mathbb{E}\{\int_0^t H^2(s) \, ds\} < +\infty \), then we can construct the process \( t \mapsto \int_0^t H \, dW \) such that it is a continuous \( L^2 \)-martingale.

**Proof (Optional)** According to Theorem 9.8, \( \int_0^t H \, dW \) exists, so we can proceed by verifying the assertions of the theorem. This will be done in three steps.

**Step 1. Reduction to \( H \) that is Dini-Continuous in \( L^2(P) \).**

Suppose we have proved the theorem for all processes \( H \) that are adapted and Dini-continuous in \( L^2(P) \). In this first step we prove that this implies the remaining assertions of the theorem.

Let \( H \) be an adapted process such that for all \( t \geq 0 \), \( \mathbb{E}\{\int_0^t H^2(s) \, ds\} < +\infty \). We can find adapted processes \( H_n \) that are Dini-continuous in \( L^2(P) \) and

\[
\lim_{n \to \infty} \mathbb{E}\left\{ \int_0^t [H_n(s) - H(s)]^2 \, ds \right\} = 0.
\] (9.23)

Indeed, we can apply Proposition 9.9 to \( H_1 \), and use the recipe of the said proposition for \( H_n \). Then apply the proposition to \( H_n \). This shows that in fact \( H_n \) can even be chosen independently of \( t \) as well.

By the Itô isometry (equation 9.6),

\[
\lim_{n \to \infty} \mathbb{E}\left\{ \int_0^t H \, dW - \int_0^t H_n \, dW \right\}^2 = \lim_{n \to \infty} \mathbb{E}\left\{ \int_0^t (H(s) - H_n(s))^2 \, ds \right\} = 0.
\] (9.24)

But \( \int_0^t H_n \, dW - \int_0^t H_{n+k} \, dW = \int_0^t (H_n - H_{n+k}) \, dW \), a.s., and is a continuous \( L^2 \)-martingale. Therefore, by Doob’s maximal inequality (Theorem 9.12), for any nonrandom but fixed \( T > 0 \),

\[
\lim_{n \to \infty} \mathbb{E}\left\{ \sup_{0 \leq t \leq T} \left( \int_0^t H_n \, dW - \int_0^t H_{n+k} \, dW \right)^2 \right\} = 0.
\] (9.25)

In particular, for each \( T > 0 \), there exists a stochastic process \( X := \{X(t); t \geq 0\} \) and a (random) subsequence \( n' \to \infty \) such that with probability one, \( \lim_{n' \to \infty} \sup_{0 \leq t \leq T} |\int_0^t H_{n'} \, dW - X(t)| = 0 \). Moreover, the same uniform convergence holds in \( L^2(P) \), and along the original subsequence \( n \to \infty \). Consequently, this and (9.24) together show that \( X \) is a particular construction of \( t \mapsto \int_0^t H \, dW \) that is a.s.-continuous and adapted. In other words, \( X \) is (obviously) adapted and a.s.-continuous, but it also satisfies

\[
P\left\{X(t) = \int_0^t H \, dW\right\} = 1, \quad \forall t \geq 0.
\] (9.26)

(Why?) Finally, \( X(t) \in L^2(P) \) for all \( t \geq 0 \), so it remains to prove that \( X \) is a martingale. But remember that we are assuming that \( t \mapsto \int_0^t H_n \, dW \) is a martingale.
By the conditional Jensen inequality (7.6), and by $L^2(\mathbb{P})$-convergence,
\[
\left\| \mathbb{E}\{X(t + s) \mid \mathcal{F}_s\} - \mathbb{E}\left\{ \int_0^{t+s} H_n dW \right\} \right\|_2 = \left\| \mathbb{E}\{X(t + s) - \int_0^{t+s} H_n dW\} \right\|_2 
\leq \left\| X(t + s) - \int_0^{t+s} H_n dW \right\|_2,
\]
which goes to zero as $n \to \infty$. On the other hand, since $t \to \int_0^t H_n dW$ is a martingale, this also shows that $\int_0^t H_n dW \to \mathbb{E}\{X(t + s) \mid \mathcal{F}_s\}$ in $L^2(\mathbb{P})$. But we have already seen that $\int_0^t H_n dW \to X(t)$ in $L^2(\mathbb{P})$. Therefore, with probability one, $\mathbb{E}\{X(t + s) \mid \mathcal{F}_s\} = X(s)$; i.e., $X$ is martingale as claimed.

**Step 2. A Continuous Martingale in the Dini-Continuous Case.**

Now we suppose that $H$ is in addition Dini-continuous in $L^2(\mathbb{P})$, and prove the theorem in this special case. Together with Step 1, this completes the proof. The argument is based on a trick. Define,
\[
J_n(H)(t) := \sum_{k \leq 2^n, t-1} H\left(\frac{k}{2^n}\right) \times \left[ W\left(\frac{k+1}{2^n}\right) - W\left(\frac{k}{2^n}\right) \right]
\]
\[
+ H\left(\frac{2^n t - 1}{2^n}\right) \times \left[ W(t) - W\left(\frac{2^n t - 1}{2^n}\right) \right].
\]

This is a minor variant of $\mathcal{I}_n(H1_{[0, t]}).$ Indeed, you should check that
\[
J_n(H)(t) - \mathcal{I}_n (H1_{[0,t]}) = H\left(\frac{2^n t - 1}{2^n}\right) \times \left[ W(t) - W\left(\frac{2^n t - 1}{2^n}\right) \right],
\]
whose $L^2(\mathbb{P})$-norm goes to zero as $n \to \infty$. But $J_n(H)$ is also a stochastic process that is (a) adapted, and (b) continuous in $t$. In fact, it is also a martingale. Here is why: Suppose $t \geq s \geq 0$. Then there exist integers $0 \leq k \leq K \leq 2^n s - 1$ such that $s \in D(k; n) := [k2^{-n}, (k + 1)2^{-n})$ and $t \in D(K; n)$. Then,
\[
J_n(H)(t) - J_n(H)(s)
= \sum_{k \leq j < K} H\left(\frac{j}{2^n}\right) \times \left[ W\left(\frac{j+1}{2^n}\right) - W\left(\frac{j}{2^n}\right) \right]
\]
\[
+ H\left(\frac{K}{2^n}\right) \times \left[ W(t) - W\left(\frac{K}{2^n}\right) \right] - H\left(\frac{k}{2^n}\right) \times \left[ W(s) - W\left(\frac{k}{2^n}\right) \right],
\]

where $\sum_{k \leq j < K} (\cdots) := 0$ (in the case that $k = K$). Since $W$ has independent increments, $\mathbb{E}\{[\cdots] \mid \mathcal{F}_s\} = 0$, a.s. where $[\cdots]$ is any of the terms of the preceding display in the square-brackets. The adaptedness of $H$ and Corollary 8.10 together show that $\mathbb{E}\{J_n(H)(t) - J_n(H)(s) \mid \mathcal{F}_s\} = 0$, a.s., which then proves the martingale property.

**Step 3. The Conclusion.**

To finish the proof, suppose $H$ is an adapted process that is Dini-continuous in $L^2(\mathbb{P})$. A calculation similar to that of Lemma 9.2 reveals that for any nonrandom $T > 0$,
\[
\lim_{n \to \infty} \sup_{0 \leq t \leq T} \mathbb{E}\left\{ \left( J_{n+1}(H)(t) - J_n(H)(t) \right)^2 \right\} = 0.
\]

137
Therefore, by Doob’s maximal inequality (Theorem 9.12),

\[
\lim_{n \to \infty} E \left\{ \sup_{0 \leq t \leq T} \left( J_{n+1}(H)(t) - J_n(H)(t) \right)^2 \right\} = 0.
\] (9.32)

This implies that a subsequence of \( J_n(H) \) converges a.s. and uniformly for all \( t \in [0,T] \) to some process \( X \). Since \( J_n(H) \) is a continuous process, \( X \) is necessarily continuous a.s. Furthermore, the argument applied in Step 1 shows that there too \( X \) is a martingale. Finally, we have already seen that for any fixed \( t \geq 0 \), \( J_n(H)(t) - \mathcal{I}_n(H1_{[0,t]}) \to 0 \) in \( L^2(P) \). Since \( \mathcal{I}_n(H1_{[0,t]}) \to \int_0^t H \, dW \) in \( L^2(P) \), this shows that \( P\{X(t) = \int_0^t H \, dW\} = 1 \), which proves the result.

\[ \square \]

4 Quadratic Variation

We now elaborate a little on quadratic variation; cf. Theorem 8.9. Quadratic variation is a central theme in continuous-time martingale theory, but it requires too much time to study properly. Therefore, we will only develop the portions for which we have immediate use.

Throughout, we define the second-order analogue of \( \mathcal{I}_n \) (9.1);

\[
\mathcal{Q}_n(H) := \sum_{k=0}^{\infty} H \left( \frac{k}{2^n} \right) \times \left[ W \left( \frac{k+1}{2^n} \right) - W \left( \frac{k}{2^n} \right) \right]^2.
\] (9.33)

**Theorem 9.14** Suppose \( H \) is an adapted compact-support process that is uniformly continuous in \( L^2(P) \); i.e., \( \lim_{r \to 0} \psi_2(r) = 0 \). Then,

\[
\lim_{n \to \infty} \mathcal{Q}_n(H) = \int_0^\infty H(s) \, ds, \quad \text{in } L^2(P).
\] (9.34)

**Proof** To simplify the notation, write

\[
d_{k,n} := W((k+1)2^{-n}) - W(k2^{-n}), \quad \forall k = 0,1, \ldots, n = 1,2, \ldots.
\] (9.35)

Recall next that we can find a nonrandom \( T > 0 \) such that for all \( s \geq T \), \( H(s) = 0 \), a.s. Throughout, we keep such a \( T \) fixed.

*Step 1. Approximating the Lebesgue Integral.*

I begin by proving that the Riemann–integral approximation of the Lebesgue integral \( \int_0^\infty H(s) \, ds \) converges in \( L^2(P) \). Namely, note that

\[
\left| \sum_{k=0}^{\infty} H \left( \frac{k}{2^n} \right) 2^{-n} - \int_0^\infty H(s) \, ds \right| \leq \sum_{0 \leq k \leq 2^n T-1} \int_{k2^{-n}}^{(k+1)2^{-n}} \left| H \left( \frac{k}{2^n} \right) - H(s) \right| \, ds,
\] (9.36)

since if we remove the absolute values, the preceding becomes an identity. In particular, apply Minkowski’s inequality (Theorem 2.24) to deduce that

\[
\left\| \sum_{k=0}^{\infty} H \left( \frac{k}{2^n} \right) 2^{-n} - \int_0^\infty H(s) \, ds \right\|_2 \leq \sum_{0 \leq k \leq 2^n T-1} 2^{-n} \psi_2(2^{-n}) \leq T \psi_2(2^{-n}).
\] (9.37)
As $n \to \infty$, the above converges to zero, and Step 1 follows.

**Step 2. Completing the Proof.**

Note that $H(k2^{-n})$ is independent of $d_{k,n}$, and the latter has mean zero and variance $2^{-n}$. Therefore,

$$Q_n(t) - \sum_{k=0}^{\infty} H\left(\frac{k}{2^n}\right)2^{-n} = \sum_{k=0}^{\infty} H\left(\frac{k}{2^n}\right)\left[d_{k,n}^2 - 2^{-n}\right]. \quad (9.38)$$

Next, we square the above and take expectations.

$$\left\|Q_n(t) - \sum_{k=0}^{\infty} H\left(\frac{k}{2^n}\right)2^{-n}\right\|^2_2 = \sum_{0 \leq k \leq 2^n T-1} E\left\{H^2\left(\frac{k}{2^n}\right)\right\}E\left\{[d_{k,n}^2 - 2^{-n}]^2\right\}$$

$$+ 2 \sum_{0 \leq j < k \leq 2^n T-1} E\left\{H\left(\frac{k}{2^n}\right)H\left(\frac{j}{2^n}\right)[d_{k,n}^2 - 2^{-n}] [d_{j,n}^2 - 2^{-n}]\right\} \quad (9.39)$$

$$= \sum_{0 \leq k \leq 2^n T-1} E\left\{H^2\left(\frac{k}{2^n}\right)\right\}E\left\{[d_{k,n}^2 - 2^{-n}]^2\right\}$$

$$+ 2 \sum_{0 \leq j < k \leq 2^n T-1} E\left\{H\left(\frac{k}{2^n}\right)H\left(\frac{j}{2^n}\right)[d_{k,n}^2 - 2^{-n}]\right\} \times E\left\{d_{j,n}^2 - 2^{-n}\right\}$$

$$= \sum_{0 \leq k \leq 2^n T-1} E\left\{H^2\left(\frac{k}{2^n}\right)\right\}E\left\{[d_{k,n}^2 - 2^{-n}]^2\right\}. $$

But $d_{k,n}$ is normal with mean zero and variance $2^{-n}$; so $2^{-n/2}d_{k,n}$ is standard normal, and so, $[d_{k,n}^2 - 2^{-n}]$ has the same distribution as $2^{-n}[Z^2 - 1]$ where $Z$ is standard normal. But $E([Z^2 - 1]^2) = E(Z^4) - 1 = 2$, thanks to (8.24). Therefore, $\|Q_n(t) - \sum_k H(k2^{-n})2^{-n}\|^2_2 = 2 \cdot 4^{-n} \sum_{0 \leq k \leq 2^n T-1} \|H(k2^{-n})\|^2_2$. But Dini-continuity insures that $t \mapsto \|H(t)\|_2$ is continuous, and hence bounded on $[0,T]$ by some constant $K_T$. Thus, $\|Q_n(t) - \sum_k H(k2^{-n})2^{-n}\|^2_2 \leq TK_T^2 2^{-n+1} \to 0$. This and Step 1 together prove the result. \qed

## 5 Itô's Formula and Three Applications

The chain rule of calculus states that given two continuously-differentiable functions $f$ and $g$, $(f \circ g)' = f'(g)g'$. In its integrated form, this is integration-by-parts, and states that for all $t \geq s \geq 0$ (say),

$$f(g(t)) - f(g(s)) = \int_s^t f'(g(u))g'(u) \, du. \quad (9.40)$$

Itô's formula tells us what happens if we replace $g$ by the nowhere-differentiable function $W$.

**Theorem 9.15 (Itô's Formula; Itô 1944)** If $f : \mathbb{R} \to \mathbb{R}$ has two continuous derivatives, then for all $t \geq s \geq 0$, the following holds with probability one:

$$f(W(t)) - f(W(s)) = \int_s^t f'(W(r))W(dr) + \frac{1}{2} \int_s^t f''(W(r)) \, dr. \quad (9.41)$$
provided that there exist constants $A,B > 0$ such that $|f'(x)| \leq Ae^{B|x|}$.

Remark 9.16 (a) In other words, the nowhere-differentiability of $W$ forces us to replace the right-hand side of (9.40) with a stochastic integral plus a second-derivative term.

(b) Itô’s formula continues to hold even if we assume only that $f''$ exists almost-everywhere, and that $\int_{0}^{t} (f'(W(r))^{2} \, dr < +\infty$, a.s. Of course, then we have to make sense of the stochastic integral, etc. Rather than prove such refinement here, I urge you to have a look at Dellacherie and Meyer (1982) for a definitive account.

(c) The strange-looking condition on $f'$ ensures that it does not grow too fast. This condition can be removed nearly altogether but this development rests on introducing a new family of processes that are called local martingales. To see how this exponential-growth condition comes up in the context of Theorem 9.8, note that for $\int_{0}^{t} f'(W(s)) W(ds)$ to be well defined, we need $\mathbb{E}\{\int_{0}^{t} [f'(W(s))]^{2} \, ds\}$ to be finite for all $t \geq 0$. But then the said condition on $f'$ implies:

$$
\mathbb{E}\left\{\int_{0}^{t} [f'(W(s))]^{2} \, ds\right\} = \int_{-\infty}^{\infty} \int_{0}^{t} [f'(x)]^{2} e^{-x^{2}/(2t)} \, ds \, dx \\
\leq \int_{-\infty}^{\infty} \int_{0}^{t} [f'(x)]^{2} e^{-x^{2}/(2t)} \, ds \, dx \leq A \int_{-\infty}^{\infty} \int_{0}^{t} e^{2B|x|} e^{-x^{2}/(2t)} \, ds \, dx,
$$

which is finite (in fact less than or equal to $4A \exp(2B^{2}t)$.)

□

Proof in the Case that $f'''$ is Bounded and Continuous

Without loss of generality, $s := 0$ (why?). The proof of Itô’s formula starts out in the same manner as that of (9.40). Namely, we apply Taylor’s expansion with remainder, and write

$$
f(W(2^{-n} [2^{n}t - 1])) - f(0) = \sum_{0 \leq k \leq 2^{n}t-1} \left[ f\left(W\left(\frac{k+1}{2^{n}}\right)\right) - f\left(W\left(\frac{k}{2^{n}}\right)\right)\right] \\
= \sum_{0 \leq k \leq 2^{n}t-1} f'(W\left(\frac{k}{2^{n}}\right)) \times \left[ W\left(\frac{k+1}{2^{n}}\right) - W\left(\frac{k}{2^{n}}\right)\right] \\
+ \frac{1}{2} \sum_{0 \leq k \leq 2^{n}t-1} f''(W\left(\frac{k}{2^{n}}\right)) \times \left[ W\left(\frac{k+1}{2^{n}}\right) - W\left(\frac{k}{2^{n}}\right)\right]^{2} \\
+ \sum_{0 \leq k \leq 2^{n}t-1} R_{k,n} \times \left[ W\left(\frac{k+1}{2^{n}}\right) - W\left(\frac{k}{2^{n}}\right)\right]^{3},
$$

where $|R_{k,n}| \leq \sup_{x} |f'''(x)| := M < \infty$, uniformly for all $k,n$. In the last identity, the first term converges in $L^{2}(P)$ to $\int_{0}^{t} f'(W(s)) W(ds)$ by Theorem 9.6; see also Example 9.4. The second term, on the other hand, converges in $L^{2}(P)$ to $\frac{1}{2} \int_{0}^{t} f''(W(s)) \, ds$; cf. Theorem 9.14. In addition, $f(W(2^{-n} [2^{n}t - 1])) \rightarrow f(W(t))$ a.s. and in $L^{2}(P)$ by continuity, and thanks to the dominated convergence theorem (Theorem 2.22). It, therefore, suffices to prove that as $n \rightarrow \infty$,

$$
\sum_{0 \leq k \leq 2^{n}t-1} R_{k,n} \times \left[ W\left(\frac{k+1}{2^{n}}\right) - W\left(\frac{k}{2^{n}}\right)\right]^{3} \overset{L^{1}(P)}{\longrightarrow} 0.
$$

(9.44)
On the other hand,

$$\begin{align*}
E \left\{ \sum_{0 \leq k \leq 2^n t-1} |R_{k,n}| \times \left| W \left( \frac{k+1}{2^n} \right) - W \left( \frac{k}{2^n} \right) \right|^3 \right\} \\
\leq M \sum_{0 \leq k \leq 2^n t-1} E \left\{ \left| W \left( \frac{k+1}{2^n} \right) - W \left( \frac{k}{2^n} \right) \right|^3 \right\} \leq M2^n t \cdot 2^{-3n/2} E\{|Z|^3\} \to 0, \quad \text{as } n \to \infty,
\end{align*}$$

(9.45)

where \(Z\) is standard normal. Equation (9.44) follows, whence the proof. \(\square\)

Ito’s formula is particularly useful because it identifies various martingales. This in turn leads to explicit calculations. The following is a brief sampler; it is proved by applying the Itô formula with \(f(x) = x\), and \(f(x) = x^2\), respectively.

**Corollary 9.17** The Brownian motion \(W\) is a continuous \(L^2\)-martingale. The process \(t \mapsto W^2(t) - t\) is another continuous \(L^2\)-martingale. In fact, \(W^2(t) - t = \frac{1}{2} \int_0^t W \, dW\), a.s.

Next is an interesting refinement; it is proved by similar arguments involving Taylor series expansions that were used to derive Theorem 9.15.

**Theorem 9.18** Suppose \(W\) is Brownian motion started at a given point \(W(0) = x_0 \in \mathbb{R}\). If \(f = f(x,t)\) is twice continuously-differentiable in \(x\) and continuously-differentiable in \(t\), and if for all \(t \geq 0\),

$$E\{\int_0^t |\partial_x f(W(s),s)|^2 \, ds\} < +\infty,$$

then with probability one,

$$f(W(t),t) - f(x_0,0) = \int_0^t \partial_x f(W(s),s) \, W(ds) + \int_0^t \left[ \frac{1}{2} \partial_{xx}^2 f(W(s),s) + \partial_x f(W(s),s) \right] ds,$$

(9.46)

where \(\partial_{xx}^2 f(x,t) := \partial_x \partial_x f(x,t)\). This theorem is valid even if \(f\) is complex-valued.

Of course, I owe you the definition of \(\int H \, dW\) when \(H\) is complex-valued, but this is easy: Whenever possible, \(\int H \, dW := \int \text{Re}(H) \, dW + i \int \text{Im}(H) \, dW\).

Rather than prove this theorem, I merely apply it to make three interesting computations.\(^9\)

### 5.1 A Look at Exit Distributions

If \(W\) denotes the Brownian motion started at some fixed point \(x_0 \in (-1, 1)\), one might wish to know when it leaves a given interval \((-1, 1)\) (say). The following remarkable formula of Paul Lévy answers a generalization of this question.\(^{10}\)

\(^9\)The treatment of the present notes is very far from being complete. Perhaps its greatest omission is William Feller’s characterization of one-dimensional diffusions—one of the crowning achievements of its day—and its connections to the problems below; cf. Feller (1955b, 1955a, 1956). For a pedagogic account that includes some of the most recent progress in this area see Bass (1998).

\(^{10}\)For this and much more, see Knight (1981, Chapter 4).
Theorem 9.19 (Lévy 1951) Choose and fix \( a, b > 0 \), and define \( T_{-b,a} := \inf \{ s > 0 : W(s) = a \ or \ -b \} \), where \( \inf \emptyset := +\infty \). If \( W(0) = x_0 \in (-b, a) \) if also fixed, then the characteristic function of \( T_{-b,a} \) is given by the following: For all real numbers \( \lambda \neq 0 \),

\[
E \left\{ e^{i\lambda T_{-b,a}} \right\} = \frac{e^{(1+i)x_0\sqrt{\lambda}}}{e^{(1+i)a\sqrt{\lambda}} + e^{-(1+i)b\sqrt{\lambda}}} + \frac{e^{-(1+i)x_0\sqrt{\lambda}}}{e^{(1+i)b\sqrt{\lambda}} + e^{-(1+i)a\sqrt{\lambda}}}. \tag{9.47}
\]

**Proof** Let us apply Itô’s formula (Theorem 9.18) with \( f(x, t) := \psi(x)e^{it\lambda} \), where \( \lambda \neq 0 \) is fixed, and the function \( \psi \) satisfies the following (complex) eigenvalue problem:

\[
\frac{1}{2}\psi''(x) = i\lambda\psi(x), \quad \psi(a) = \psi(-b) = 1. \tag{9.48}
\]

You can directly check that the solution is

\[
\psi(x) := \frac{e^{(1+i)x\sqrt{\lambda}}}{e^{(1+i)a\sqrt{\lambda}} + e^{-(1+i)b\sqrt{\lambda}}} + \frac{e^{-(1+i)x\sqrt{\lambda}}}{e^{(1+i)b\sqrt{\lambda}} + e^{-(1+i)a\sqrt{\lambda}}}. \tag{9.49}
\]

Clearly, \( |\partial_x f(x, t)| \) is bounded in \((x, t)\) so that \( E \left\{ \int_0^t |\partial_x f(W(s), s)|^2 ds \right\} < +\infty \). Moreover, the eigenvalue problem for \( \psi \) implies that \( \frac{1}{2}\partial_x^2 f(x, t) + \partial_t f(x, t) = 0 \). Therefore, Theorem 9.18 tells us that \( f(W(t), t) - f(x_0, 0) \) is a mean-zero (complex) martingale. By the optional stopping theorem (Theorem 9.11),

\[
E \left\{ f(W(T_{-b,a} \land t), T_{-b,a} \land t) \right\} = f(x_0, 0), \tag{9.50}
\]

which equals \( \psi(x_0) \). Thanks to the dominated convergence theorem (Theorem 2.22), and by the a.s. continuity of \( W \), we can let \( t \to \infty \) to deduce that

\[
E \left\{ f(W(T_{-b,a}), T_{-b,a}) \right\} = \psi(x_0). \tag{9.51}
\]

But \( f(W(T_{-b,a})) = \psi(W(T_{-b,a}))e^{i\lambda T_{-b,a}} = e^{i\lambda T_{-b,a}} \), since \( \psi(a) = \psi(-b) = 1 \). This proves the theorem. \( \square \)

### 5.2 A Second Look at Exit Distributions

Let us have a second look at Theorem 9.19 in the simplest setting where \( x_0 := 0 \) and \( a = b = 1 \). In this case, (9.47) somewhat simplifies to the following elegant form:

\[
E \left\{ e^{i\lambda T_{-1,1}} \right\} = \frac{2}{e^{\sqrt{2}\lambda} + e^{-\sqrt{2}\lambda}} = \frac{1}{\cosh \left( \sqrt{2\lambda} \right)}, \quad \forall \lambda \in \mathbb{R} \setminus \{0\}. \tag{9.52}
\]

This requires only the elementary fact that \((1 + i) = \sqrt{2} \) (check!). So in principle, this determines the distribution of \( T := T_{-1,1} \) although reality may present a different picture. For instance, I note that if it were not for (9.52), then we could not even easily prove that \([\cosh(\sqrt{2\lambda})]^{-1}\) is a characteristic function of a probability measure. Or for that matter, can you see from (9.52) that \( T \) has finite moments of all orders?

The following contains a different representation of the distribution of \( T \) that answers the question about the moments of the stopping time \( T \).

---

11See Ciesielski and Taylor (1962) for an interesting multidimensional analogue.
Theorem 9.20 (Chung 1947) For any $t > 0$,
\[
P\{T > t\} = 4 \pi \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \exp\left(\frac{-n^2 \pi^2 t}{8}\right). \tag{9.53}
\]
In particular, $P\{T > t\} = \frac{4}{\pi} \exp(-\frac{1}{8} \pi^2 t) \times (1 + \Theta(t))$, where $\Theta(t) \to 0$ as $t \to \infty$.

In particular, when $t$ is large, $P\{T > t\} \leq 2 \exp(-\frac{1}{4} \pi^2 t)$. In lieu of Lemma 5.9, for any $p \geq 1$,
\[
E\{T^p\} = p \int_0^{\infty} t^{p-1} P\{T > t\} dt < +\infty, \tag{9.54}
\]
which shows that $T$ has moments of all orders, as asserted earlier.

Theorem 9.20 implies also the following unusual formula:

Corollary 9.21 (Chung 1947) The distribution function of $\sup_{0 \leq s \leq 1} |W(s)|$ satisfies the following:
\[
P\left\{ \sup_{0 \leq s \leq 1} |W(s)| \leq x \right\} = 4 \pi \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \exp\left(\frac{-n^2 \pi^2 x^2}{8}\right). \tag{9.55}
\]
In particular, $P\{\sup_{0 \leq s \leq 1} |W(s)| \leq x\} = \frac{4}{\pi} \exp(-\frac{1}{8} \pi^2 x^2) \times (1 + \Theta(x^2))$, where $\Theta(x^2) \to 0$ as $x \to 0$.

I will prove only Theorem 9.20. While it is possible to write a very quick proof, I will sketch an argument that I believe shows the motivation behind the proof. [It is a sketch only because I will not develop some of the details of the PDE and Fourier-analysis arguments.]

Proof of Theorem 9.20 (Sketch) This proof is presented in three quick steps.

Step 1. Itô’s Formula and a Sturm–Liouville Problem.
Itô’s formula (Theorem 9.18) tells us that modulo technical conditions, $f(W(t), t) - f(0, 0)$ is a mean-zero martingale provided that $f$ satisfies the partial differential equation, \( \frac{1}{2} \partial^2_{xx} f + \partial_t f = 0 \). It is known that $f$ must have the form \( f(x, t) = \psi(x)e^{\lambda t} \); this is called separation of variables. Rather than prove this fact, I will merely be guided by it, and seek functions of the type $f(x, t) := \psi(x)e^{\lambda t}$ that satisfy the said PDE. But this is an easy task, for in terms of $\psi$ the PDE is: $\psi'' + 2\lambda \psi = 0$ (check!). Any $\psi$ that satisfies this ODE yields a function $f$ for which we then have (modulo technical integrability), $E\{f(W(T \wedge t), T \wedge t) = f(0, 0)\)$. In the last part I used the optional stopping theorem (Theorem 9.11). Equivalently, \[
E\{f(W(T), T); T \leq t\} + E\{f(W(t), t); T > t\} = f(0, 0). \tag{9.56}
\]
Of course, $W(T) \in \{-1, 1\}$. Therefore, if we add to the ODE the conditions that $\psi(\pm 1) = 0$, then we obtain \[
E\{f(W(t), t); T > t\} = f(0, 0). \tag{9.57}
\]
Equivalently, suppose $\psi$ solves the Sturm–Liouville problem: $\psi'' = -2\lambda \psi$ and $\psi(\pm 1) = 0$. Then, \[
E\{\psi(W(t)); T > t\} = e^{-\lambda t} \psi(0). \tag{9.58}
\]
The typical solution to the said Sturm–Liouville problem is \( \psi(x) = \cos\left(\frac{n\pi x}{2}\right) \) where \( n = 1, 2, \ldots \) (check!). This function solves \( \psi'' = -2\lambda\psi, \psi(\pm 1) = 0 \) with \( \lambda = \frac{1}{8}n^2\pi^2 \). Since \( \psi(0) = 1 \), we have

\[
E \left\{ \cos \left( \frac{n\pi W(t)}{2} \right); T > t \right\} = \exp \left( -\frac{n^2\pi^2 t}{8} \right). \tag{9.59}
\]

**Step 2. Fourier Series.**

Let \( L^2(-2, 2) \) denote the collection of all measurable functions \( g : [-2, 2] \to \mathbb{R} \) such that \( \int_{-2}^{2} g^2(x) \, dx < +\infty \).

Then, Theorem 8.31 and a little fiddling with the variables shows us that \( L^2(-2, 2) \) has the following complete orthonormal basis:

\[
\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \sin\left(\frac{1}{2}n\pi x\right), \frac{1}{\sqrt{2}} \cos\left(\frac{1}{2}n\pi x\right) \quad (n, m = 1, 2, \ldots). \]

In particular, any \( \phi \in L^2(-2, 2) \) has the representation,

\[
\phi(x) = \frac{A_0}{2} + \sum_{n=1}^{\infty} A_n \cos \left( \frac{n\pi x}{2} \right) + \frac{1}{\sqrt{2}} \sum_{n=1}^{\infty} B_n \sin \left( \frac{n\pi x}{2} \right), \tag{9.60}
\]

where the sums converge in \( L^2(-2, 2) \), \( A_0 := \int_{-2}^{2} \phi(x) \, dx, A_n := \frac{1}{\sqrt{2}} \int_{-2}^{2} \phi(x) \cos\left(\frac{1}{2}n\pi x\right) \, dx \), and \( B_n := \frac{1}{\sqrt{2}} \int_{-2}^{2} \phi(x) \sin\left(\frac{1}{2}n\pi x\right) \, dx \) \((n, m = 1, 2, \ldots)\).

**Step 3. Putting it Together.**

Apply Step 2 to the function \( \phi(x) := 1_{(-1,1)}(x) \) to obtain \( A_0 = 1 \), \( A_n = 2^{3/2}(n\pi)^{-1}(-1)^{n+1} \), and \( B_n = 0 \) \((n = 1, 2, \ldots)\). Thus,

\[
1_{(-1,1)}(x) = \frac{1}{2} + \sum_{n=1}^{\infty} \left( -\frac{1}{n} \right)^{n+1} \cos \left( \frac{n\pi x}{2} \right). \tag{9.61}
\]

Plug in \( x := W(t, \omega) \), multiply by \( 1_{\{T(\omega) > t\}} \), and integrate \( [P(d\omega)] \) to obtain

\[
P \{ W(t) \in (-1, 1), T > t \} = \frac{1}{2} P\{T > t\} + \sum_{n=1}^{\infty} \left( -\frac{1}{n} \right)^{n+1} E \left\{ \cos \left( \frac{n\pi W(t)}{2} \right); T > t \right\}. \tag{9.62}
\]

[A word of caution: In a fully-rigorous treatment, we need to justify this exchange of \( L^2(P) \) limit and expectation.] Two quick observations are in order: The left-hand side equals \( P\{T > t\} \); this follows from the fact that \( \{T > t\} = \{\sup_{s \leq t} |W(s)| < 1\} \). The second observation is that the right-hand side is computable via (9.59). After a little algebra, this completes our proof. \( \Box \)

The following is a corollary of the proof. It turns out to be the starting-point of some of the many deep connections between killed Markov processes and the boundary-theory of second-order differential equations.

**Corollary 9.22** Suppose \( \phi \in L^2(-2, 2) \) has the Fourier series representation (9.60), and \( \phi(\pm 1) = 0 \); i.e., for all \( n \geq 1 \), \( B_n = 0 \). Then for any \( t \geq 0 \),

\[
E \{ \phi(W(t)); T > t \} = \frac{2A_0}{\pi} \sum_{n=1}^{\infty} \left( -\frac{1}{n} \right)^{n+1} e^{-\frac{1}{4}n^2\pi^2 t} + \frac{1}{\sqrt{2}} \sum_{n=1}^{\infty} A_n e^{-\frac{1}{4}n^2\pi^2 t}. \tag{9.63}
\]
5.3 Brownian Motion with Positive Drift

The final application of this section is the next result is about Brownian motion with drift $\beta > 0$; this is the process $W_\beta(t) := W(t) + \beta t$. To motivate the next result, let me first mention that thanks to the law of the iterated logarithm (Exercise 8.4) with probability one, $t^{-1} W(t) \to 0$ as $t \to \infty$. This implies that for large values of $t$, $W_\beta(t)$ behaves like the nonrandom function $\beta t$, and in particular, $\inf_{t \geq 0} W_\beta(t) > -\infty$, a.s. On the other hand, by the law of the iterated logarithm of $W$ near $t = 0$ (Exercise 8.4), when $t^{-1} W_\beta(t) \to +\infty$. This shows that $0 \geq \inf_{t \geq 0} W_\beta(t) > -\infty$, a.s. (why?). In other words, $\inf_{t \geq 0} W_\beta(t)$ is a nontrivial negative random variable. Our next theorem finds its distribution.\footnote{There is a remarkable analysis of Brownian motion with drift that produces this result as a trivial by-product; cf. Williams (1974), and in particular Theorem 2.4 there.}

**Theorem 9.23** For any $a, b > 0$, define $S_{a,b} := \inf \{ s > 0 : W_\beta(s) = a or -b \}$. Then,

$$P\{W_\beta(S_{a,b}) = -b\} = \frac{1 - e^{-2\beta a}}{e^{2\beta b} - e^{-2\beta a}}.$$  (9.64)

In particular, let $a \to \infty$ to see that for any $b > 0$,

$$P\left\{ \inf_{t \geq 0} W_\beta(t) \leq -b \right\} = e^{-2\beta b}.$$  (9.65)

In other words, $-\inf_{t \geq 0} W_\beta(t)$ is an exponential random variable with mean $(2\beta)^{-1}$.

**Proof** For any $\alpha \in \mathbb{R}$, $M_\alpha(t) := \exp(\alpha W(t) - \frac{1}{2} \alpha^2 t)$ is a mean-one martingale. [To prove this, apply Theorem 9.18 with $f(x, t) := \exp(\alpha x - \frac{1}{2} \alpha^2 t)$.] Choose $\alpha = 2\beta$ and rewrite this statement to see that $t \to \exp(-2\beta W_{\beta}(t))$ is a mean-one martingale. Now $S_{a,b}$ is a stopping time (why?). Therefore, by the optional stopping theorem (Theorem 9.11), for any $t > 0$, $E\{ \exp(-2\beta W_{\beta}(S_{a,b} \wedge t)) \} = 1$. Since $\sup_{s \leq S_{a,b}} |W_{\beta}(t)| \leq \max(a, b)$ is finite, by the dominated convergence theorem (Theorem 2.22), we can let $t \to \infty$ and obtain: $E\{ \exp(-2\beta W_{\beta}(S_{a,b})) \} = 1$. But this expectation is easily computed, since $W_{\beta}(S_{a,b}) = a$ with probability $p$ and $W_{\beta}(S_{a,b}) = -b$ with probability $1 - p$, where $p := P\{W_{\beta}(S_{a,b}) = a\}$. This proves the first assertion. The remainder of this theorem follows suit from the first assertion. \qed

6 Exercises

**Exercise 9.1** In this exercise, we construct a Dini-continuous process in $L^p(\mathbb{P})$ that is not a.s. continuous.

1. If $t > s > 0$, prove that

$$P\{W(s) > 0, W(t) < 0\} = \int_0^\infty \frac{e^{-\frac{y^2}{2(s-a)}}}{\sqrt{2\pi s}} P\{W(t-s) > y\} \, dy.$$  (9.66)

2. Prove that $P\{W(t-s) > y\} \leq e^{-\frac{y^2}{2(t-s)^2}}$.

3. Conclude that $P\{W(s) > 0, W(t) < 0\} \leq \frac{1}{2} \sqrt{\frac{e}{t-s}}$.

4. Use this to prove that given any $\eta > 0$, $H(s) := 1_{(0,\infty)}(W(s))$ is Dini-continuous in $L^2(\mathbb{P})$ as $s$ varies in $[\eta, 1]$, but $H$ is not a.s. continuous.
Exercise 9.2 In this exercise, you are asked to construct a rather general abstract integral that is due to Young (1970).

A function $f : [0, 1] \to \mathbb{R}$ is said to be Hölder continuous of order $\alpha > 0$ if there exists a finite constant $K$ such that for all $s, t \in [0, 1]$ $|f(s) - f(t)| \leq K|t - s|^{\alpha}$. Let $C^\alpha$ denote the collection of all such functions. When $\alpha = 0$, we define $C^0$ to be the collection of all continuous real functions on $[0, 1]$.

1. Prove that when $\alpha > 1$, $C^\alpha$ contains only constants, whereas $C^1$ includes but is not limited to all continuously-differentiable functions.

2. If $0 < \alpha < 1$, then prove that $C^\alpha$ is a complete normed linear space that is normed by

$$
\|f\|_{C^\alpha} := \sup_{s, t \in [0, 1], s \neq t} \frac{|f(s) - f(t)|}{|s - t|^{\alpha}}.
$$

3. Given two functions $f$ and $g$, define

$$
\int_0^1 f \delta_n g := \sum_{k=0}^{2^n - 1} f \left( \frac{k}{2^n} \right) \times \left[ g \left( \frac{k + 1}{2^n} \right) - g \left( \frac{k}{2^n} \right) \right], \quad \forall n \geq 1.
$$

Suppose for some $\alpha, \beta \leq 1$, $f \in C^\alpha$ and $g \in C^\beta$. Prove that whenever $\alpha + \beta > 1$, then $\int_0^1 f \delta g := \lim_n \int_0^1 f \delta_n g$ exists, and extends the Riemann integral (i.e., the case where $g(x) = x$). This is called the Young integral.

(Hint: Lemma 9.2.)

Exercise 9.3 In this exercise you are asked to prove Theorem 9.12 and its variants in steps. We say that $M$ is a submartingale if it is defined as a martingale, except whenever $t \geq s \geq 0$, we have $\mathbb{E}(M(t) | \mathcal{F}_s) \geq M(s)$, a.s. $M$ is a supermartingale if $-M$ is a submartingale. A process $M$ is said to be a continuous $L^2$-submartingale (respectively, continuous $L^2$-supermartingale) if it is a submartingale (respectively supermartingale), $t \mapsto M(t)$ is a.s.-continuous, and for all $t \geq 0$, $M(t) \in L^2(\mathcal{P})$.

1. If $Y$ is in $L^2(\mathcal{P})$, then prove that $M(t) := \mathbb{E}(Y | \mathcal{F}_t)$ is a martingale. (This is the Doob martingale in continuous time.)

2. If $M$ is a martingale and $\psi$ is convex, then $\psi(M)$ is a submartingale provided that $\psi(M(t)) \in L^1(\mathcal{P})$ for each $t \geq 0$.

3. If $M$ is a submartingale and $\psi$ is a nondecreasing convex function, and if $\psi(M(t)) \in L^1(\mathcal{P})$ for all $t \geq 0$, then $\psi(M)$ is a submartingale.

4. Prove that the first inequality in Theorem 9.12 holds if $|M|$ is replaced by any a.s.-continuous submartingale.

(Hint: In the last part, you will need to prove that $\sup_{0 \leq s \leq t} M(s)$ is measurable!)

Exercise 9.4 (Gambler’s Ruin Formula) If $W$ denotes a Brownian motion, then for any $a \in \mathbb{R}$, define $T_a := \inf\{s \geq 0 : W(s) = a\}$ where $\inf \emptyset := \infty$. Recall that $T_a$ is an $\mathcal{F}$-stopping time (Proposition 8.18). If $a, b > 0$, then carefully prove that $\mathbb{P}(T_a < T_b) = b \div (a + b)$. Finally, compute $\mathbb{E}(T_a)$.

(Hint: Use Corollaries 8.10, 9.17, and Theorem 9.11.)
Exercise 9.5 Prove Corollary 9.21.
(Hint: Use Theorem 9.20 and the scaling property of Brownian motion; cf. Theorem 8.9.)