1

Brownian Motion

The initial sections of this chapter are devoted to the definition of Brownian motion (the mathematical object, not the physical phenomenon) and a compilation of its basic properties. The properties in question are quite deep, and readers will be referred elsewhere for proofs. Later sections are devoted to the derivation of further properties and to calculation of several interesting distributions associated with Brownian motion.

Before proceeding, readers are advised to at least look through Appendices A and B, which enunciate some standing assumptions (in particular, joint measurability and right-continuity of stochastic processes) and explain several important conventions regarding notation and terminology. As noted there, and in the Guide to Notation and Terminology, the value of a stochastic process X at time t may be written either as X_t or as $X(t)$, depending on convenience. The former notation is generally preferred, but the latter is used when necessary to avoid clumsy typography like subscripts on subscripts.

1.1 Wiener's theorem

A stochastic process X is said to have *independent increments* if the random variables $X(t_0), X(t_1) - X(t_0), \ldots, X(t_n) - X(t_{n-1})$ are independent for any $n \geq 1$ and $0 \leq t_0 < \cdots < t_n < \infty$. It is said to have *stationary* independent increments if moreover the distribution of $X(t) - X(s)$ depends only on $t - s$. Finally, we write $Z \sim \mathcal{N}(\mu, \sigma^2)$ to mean that the random variable Z has the normal distribution with mean μ and variance σ^2. A *standard Brownian motion*, or *Wiener process*, is then defined as a stochastic process X having continuous sample paths, stationary independent increments, and $X(t) \sim \mathcal{N}(0, t)$. Thus, in our terminology, a standard Brownian motion starts at level zero almost surely. A process Y will be called a (μ, σ) Brownian motion if it has the form $Y(t) = Y(0) + \mu t + \sigma X(t)$, where X is a Wiener process and $Y(0)$ is independent of X. It follows that $Y(t + s) - Y(t) \sim$

$\mathcal{N}(\mu s, \sigma^2 s)$. We call μ and σ^2 the *drift* and *variance* of Y, respectively. The term *Brownian motion*, without modifier, will be used to embrace all such processes Y.

There remains the question of whether standard Brownian motion exists and whether it is in any sense unique. That is the subject of Wiener's theorem. For its statement, let $\mathcal{C}$ be the Borel σ-algebra on $C := C[0,\infty)$ as in Section A.2, and let X be the coordinate process on C as in Section A.3. The following is proved in the setting of $C[0,1]$ in Section 8 of Billingsley (1999); the extension to $C[0,\infty)$ is essentially trivial.

Theorem 1.1 (Wiener's Theorem) *There exists a unique probability measure P on $(C, \mathcal{C})$ such that the coordinate process X on $(C, \mathcal{C}, P)$ is a standard Brownian motion.*

This P will be referred to hereafter as the *Wiener measure*. It is left as an exercise to show that a continuous process is a standard Brownian motion if and only if its distribution (see Section A.2) is the Wiener measure. When combined with Theorem 1.1, this shows that standard Brownian motion exists and is unique *in distribution*. No stronger form of uniqueness can be hoped for, because the definitive properties of standard Brownian motion refer only to the distribution of the process.

Before concluding this section we record one more important result. See Chapter 12 of Breiman (1968) for a proof.

Theorem 1.2 *If Y is a continuous process with stationary independent increments, then Y is a Brownian motion.*

This beautiful theorem shows that Brownian motion can actually be defined by stationary independent increments and path continuity alone, with normality following as a consequence of these assumptions. This may do more than any other characterization to explain the significance of Brownian motion for probabilistic modeling.

With an eye toward future requirements, we now introduce the idea of a Brownian motion *with respect to a given filtration*. Let $(\Omega, \mathcal{F}, \mathbb{F}, P)$ be a filtered probability space in the sense of Section A.1, and let X be a continuous process on this space. We say that X is a (μ, σ) Brownian motion with respect to $\mathbb{F}$, or simply a (μ, σ) Brownian motion on $(\Omega, \mathcal{F}, \mathbb{F}, P)$, if

(1.1) X is adapted,

(1.2) $X_t - X_s$ is independent of $\mathcal{F}_s$, $0 \le s \le t$, and

(1.3) X is a (μ, σ) Brownian motion in the sense of Section 1.1.

Roughly speaking, (1.1) and (1.2) say that $\mathcal{F}_t$ contains complete information about the history of X up to time t, but no information at all about the evolution of X after t. For a specific example, one may take the canonical space of Section A.3 with P the Wiener measure. In that case, X is a *standard* Brownian motion on $(\Omega, \mathcal{F}, \mathbb{F}, P)$.

1.2 Quadratic variation and local time

One of the best known properties of Brownian motion is that almost all its sample paths have infinite variation over any time interval of positive length. Thus Brownian sample paths are emphatically *not* VF functions (see Section B.2). In contrast to this negative result, a sharp positive statement can be made about the so-called quadratic variation of Brownian paths. To introduce this important concept we need a few definitions. First, a *partition* of the interval $[0, t]$ is a set of points $\Pi_t = \{t_0, t_1, \ldots, t_n\}$ with $0 = t_0 < \cdots < t_n = t$, and the *mesh* of such a partition is

$$\|\Pi_t\| := \max_{1 \leq k \leq n} (t_k - t_{k-1}).$$

Let $f : [0, \infty) \to \mathbb{R}$ be fixed and define

$$q_t(\Pi_t) := \sum_{k=1}^{n} \left[f(t_k) - f(t_{k-1}) \right]^2 . \tag{1.4}$$

If there exists a number q_t such that $q_t(\Pi_t) \to q_t$ as $\|\Pi_t\| \to 0$, then we call q_t the *quadratic variation* of f over $[0, t]$. The proof of the following proposition is left as an exercise.

Proposition 1.3 *If f is a continuous VF function, then $q_t = 0$ for all $t \geq 0$.*

Let X be a (μ, σ) Brownian motion on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, P)$, and define $Q_t(\omega)$ as q_t is defined above, but with $X(\omega)$ in place of f, assuming for the moment that the limit exists. The following proposition is proved in most standard texts.

Proposition 1.4 *For almost every $\omega \in \Omega$ we have $Q_t(\omega) = \sigma^2 t$ for all $t \geq 0$.*

Three increasingly surprising implications of Proposition 1.4 are as follows. First, the quadratic variation Q_t exists for almost all Brownian paths and all $t \geq 0$. Second, it is not zero if $t > 0$, and hence X almost surely has infinite ordinary variation over $[0, t]$ by Proposition 1.3. Finally, the quadratic variation of X does not depend on ω!

It would be difficult to overstate the significance of Proposition 1.4. We shall see later that it contains the essence of Itô's formula, and that Itô's formula is the key tool for analysis of Brownian motion and related processes. Although a complete proof of Proposition 1.4 would carry us too far afield, there are some easy calculations which at least help to make this critical result plausible. If f is replaced by X in (1.4), then the *expected value* of the sum on the right side is

$$\sum_{k=1}^{n} E\left\{[X(t_k) - X(t_{k-1})]^2\right\} = \sum_{k=1}^{n} \left[\mu^2(t_k - t_{k-1})^2 + \sigma^2(t_k - t_{k-1})\right] \longrightarrow \sigma^2 t \qquad \text{as } \|\Pi_t\| \to 0. \tag{1.5}$$

Similarly, using the independent increments of X, one may calculate explicitly the variance of the sum. (This calculation is left as an exercise.) The variance is found to vanish as $\|\Pi_t\| \to 0$, proving that the sums converge to $\sigma^2 t$ in the L^2 sense as $n \to \infty$. Proposition 1.4 says that they also converge almost surely.

Another nice feature of Brownian paths arises in conjunction with the *occupancy measure* of the process. For each $\omega \in \Omega$ and $A \in \mathcal{B}$ (the Borel σ-algebra on $\mathbb{R}$) let

$$\nu(t, A, \omega) := \int_0^t 1_A\left(X_s(\omega)\right) ds, \qquad t \geq 0,$$

with the integral defined in the Lebesgue sense. Thus $\nu(t, A, \cdot)$ is a random variable representing the amount of time spent by X in the set A up to time t, and $\nu(t, \cdot, \omega)$ is a positive measure on $(\mathbb{R}, \mathcal{B})$ having total mass t; this is the occupancy measure alluded to above. The following theorem, one of the deepest of all results relating to Brownian motion, says that the occupancy measure is absolutely continuous with respect to Lebesgue measure and has a smooth density. See Section 7.2 of Chung and Williams (1990) for a proof.

Theorem 1.5 *There exists* $l : [0, \infty) \times \mathbb{R} \times \Omega \to \mathbb{R}$ *such that, for almost every* ω, $l(t, x, \omega)$ *is jointly continuous in* t *and* x *and*

$$\nu(t, A, \omega) = \int_A l(t, x, \omega)\, dx \qquad \textit{for all } t \geq 0 \textit{ and } A \in \mathcal{B}.$$

The most difficult and surprising part of this result is the continuity of l in x, a smoothness property that testifies to the erratic behavior of Brownian paths. (Consider the occupancy measure corresponding to a continuously differentiable sample path. You will see that it does not have a continuous density at points x that are achieved as local maxima or minima of the path.) From Theorem 1.5 it follows that, for almost all ω,

$$l(t,x,\omega) = \lim_{\epsilon \downarrow 0} \frac{1}{2\epsilon} \int_0^t 1_{[x-\epsilon, x+\epsilon]}(X_s(\omega))\, ds \tag{1.6}$$

for all $t \geq 0$ and $x \in \mathbb{R}$. Consequently $l(\cdot, x, \omega)$ is a continuous increasing function *that increases only at time points t where $X(t,\omega) = x$*. The stochastic process $l(\cdot, x, \cdot)$ is called the *local time* of X at level x.

Proposition 1.6 *If $u : \mathbb{R} \to \mathbb{R}$ is bounded and measurable, then for almost all ω we have*

$$\int_0^t u(X_s(\omega))\, ds = \int_{\mathbb{R}} u(x) l(t,x,\omega)\, dx, \qquad t \geq 0. \tag{1.7}$$

Proof If u is the indicator 1_A for some $A \in \mathcal{B}$, then (1.7) follows from Theorem 1.5. Thus (1.7) holds for all simple functions u (finite linear combinations of indicators). For any positive, bounded, measurable u we can construct simple functions $\{u_n\}$ such that $u_n(x) \uparrow u(x)$ for almost every x (Lebesgue measure). Because (1.7) is valid for each u_n, it is also valid for u by the monotone convergence theorem. Moreover, the right side of (1.7) is finite because $l(t,\cdot,\omega)$ has compact support. The proof is concluded by the observation that every bounded, measurable function is the difference of two positive, bounded measurable functions. □

1.3 Strong Markov property

Again let X be a (μ, σ) Brownian motion on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, P)$. When we speak of stopping times (see Section A.1), implicit reference is being made to the filtration $\mathbb{F}$. Here and later we write $T < \infty$ as shorthand for the more precise statement $P\{T < \infty\} = 1$.

Theorem 1.7 *Let $T < \infty$ be a stopping time, and define $X_t^* = X_{T+t} - X_T$ for $t \geq 0$. Then X^* is a (μ, σ) Brownian motion with starting state zero and X^* is independent of $\mathcal{F}_T$.*

Let $\mathcal{F}^*$ be the smallest σ-algebra with respect to which all the random variables $\{X_t^*, t \geq 0\}$ are measurable. The last phrase of the theorem means that $\mathcal{F}_T$ and $\mathcal{F}^*$ are independent σ-algebras. Theorem 1.7 is proved in Section 37 of Billingsley (1995). This result articulates the strong Markov property in a form unique to Brownian motion. See Chapter 3 for an equivalent statement that suggests more clearly what is meant by a strong Markov process in general.

1.4 Brownian martingales

Here again we denote by X a (μ, σ) Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, P)$. Thus $X_t - X_s$ is independent of $\mathcal{F}_s$ for $s \le t$ by (1.2). If $\mu = 0$, then we have

$$E(X_t - X_s | \mathcal{F}_s) = E(X_t - X_s) = 0 \tag{1.8}$$

and

$$E\left[(X_t - X_s)^2 | \mathcal{F}_s\right] = E\left[(X_t - X_s)^2\right] = \sigma^2 (t - s). \tag{1.9}$$

Obviously (1.8) can be restated as

$$E(X_t | \mathcal{F}_s) = X_s \tag{1.10}$$

and then the left side of (1.9) reduces to

$$\begin{aligned} E\left[(X_t - X_s)^2 | \mathcal{F}_s\right] &= E(X_t^2 | \mathcal{F}_s) - 2E(X_t X_s | \mathcal{F}_s) + X_s^2 \\ &= E(X_t^2 | \mathcal{F}_s) - 2X_s E(X_t | \mathcal{F}_s) + X_s^2 \\ &= E(X_t^2 | \mathcal{F}_s) - X_s^2. \end{aligned} \tag{1.11}$$

Substituting (1.11) into (1.9) and rearranging terms gives

$$E(X_t^2 - \sigma^2 t | \mathcal{F}_s) = X_s^2 - \sigma^2 s. \tag{1.12}$$

Now (1.10) and (1.12) can be restated as follows.

Proposition 1.8 *If $\mu = 0$, then X and $\{X_t^2 - \sigma^2 t, t \ge 0\}$ are martingales on* $(\Omega, \mathcal{F}, \mathbb{F}, P)$.

From (1.2) and (1.3) we know that the conditional distribution of $X_t - X_s$ given $\mathcal{F}_s$ is $\mathcal{N}(\mu(t-s), \sigma^2(t-s))$. From this it follows that

$$E\left[\exp\{\beta(X_t - X_s)\} | \mathcal{F}_s\right] = \exp\left\{\mu\beta(t-s) + \tfrac{1}{2}\sigma^2\beta^2(t-s)\right\} \tag{1.13}$$

for any $\beta \in \mathbb{R}$ and $s < t$. Now let

$$q(\beta) := \mu\beta + \tfrac{1}{2}\sigma^2\beta^2, \qquad \beta \in \mathbb{R}, \tag{1.14}$$

(the letter q is mnemonic for *quadratic*) and note that (1.13) can be rewritten as

$$E\left[\exp\{\beta(X_t - X_s) - q(\beta)(t-s)\} | \mathcal{F}_s\right] = 1. \tag{1.15}$$

From (1.15) it is immediate that $E[V_\beta(t) | \mathcal{F}_s] = V_\beta(s)$, where

$$V_\beta(t) := \exp\{\beta X_t - q(\beta)t\}, \qquad t \ge 0. \tag{1.16}$$

Thus we arrive at the following.

Proposition 1.9 *V_β is a martingale on $(\Omega, \mathcal{F}, \mathbb{F}, P)$ for each $\beta \in \mathbb{R}$.*

Hereafter we call V_β the *Wald martingale* with dummy variable β. It plays a central role in the calculations of Chapter 3.

1.5 Two characterizations of Brownian motion

In the calculations leading up to Proposition 1.9 we used the following: if a random variable ξ is distributed $\mathcal{N}(\mu, \sigma^2)$ then

$$E\left(e^{\beta\xi}\right) = e^{q(\beta)} \qquad \text{for all } \beta \in \mathbb{R},$$

where $q(\cdot)$ is the quadratic function (1.14). Moreover, the converse of that statement is also true: see Curtiss (1942). Combining that with the definitions in Section 1.1, one easily obtains the following converse of Proposition 1.9.

Proposition 1.10 *Let X be a continuous adapted process on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, P)$. If process V_β defined by (1.16) is a martingale for any $\beta \in \mathbb{R}$, then X is a (μ, σ) Brownian motion with respect to $\mathbb{F}$.*

The following more difficult converse is broadly useful. Its proof is beyond the scope of this book but can be found in many advanced texts; see, for example, Section 6.1 of Chung and Williams (1990).

Theorem 1.11 *Let X be a continuous martingale on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, P)$, and further suppose that, for each $t > 0$, X has quadratic variation t over the interval $[0, t]$. Then X is a standard Brownian motion with respect to $\mathbb{F}$.*

1.6 The innovation theorem

Let $W = \{W_t,\ t \geq 0\}$ be a standard Brownian motion defined on some probability space $(\Omega, \mathcal{F}, P)$, and let $\xi = \{\xi_t,\ t \geq 0\}$ be a bounded process defined on that same space, independent of W. Now imagine a decision maker who observes the process

$$Y_t := W_t + \int_0^t \xi_s \, ds, \qquad t \geq 0, \tag{1.17}$$

and wishes to estimate, in some sense, the trajectory of ξ given the observed trajectory of Y. In this context it is usual to describe ξ as the "signal" to be estimated and W as the "noise" with which it is confounded. The decision maker seeks to "filter" the observed process Y, extracting from it an estimate of the signal ξ.

A number of such filtering problems (that is, specially structured examples of the general problem just described) will be considered in Chapter 8, where the following theorem plays a central role. In preparation, let $\mathbb{F} = \{\mathcal{F}_t,\ t \geq 0\}$ be the filtration generated by the observed process Y (see Section A.2 for the meaning of that phrase) and define

$$\mu_t := E(\xi_t | \mathcal{F}_t), \qquad t \geq 0, \tag{1.18}$$

and

$$Z_t := Y_t - \int_0^t \mu_s \, ds, \qquad t \geq 0. \tag{1.19}$$

Recall from Section A.2 that all stochastic processes are assumed to be jointly measurable in this book. Because conditional expectations are defined only up to an equivalence, (1.18) does not in itself define a *bona fide* process. That is, (1.18) does not specify a jointly measurable function $\mu_t(\omega)$, and without joint measurability the integral in (1.19) is not well defined. To remedy this problem we can invoke a basic result in stochastic process theory: there exists a (jointly measurable) process $\mu = \{\mu_t,\ t \geq 0\}$ such that $\mu_t = E(\xi_t|\mathcal{F}_t)$ almost surely for all $t \geq 0$. In fact, one can take μ to be what is called an optional process, thereby ensuring that the process Z defined by (1.19) is adapted to $\mathbb{F}$; see Theorem 3.6 and Lemma 3.11 of Chung and Williams (1990), or Section VI.7 of Rogers and Williams (1987). It is this choice of μ to which we refer hereafter.

In filtering theory Z is called the "innovations process," and adopting the terminology of Poor and Hadjiliadis (2008), we call the following result "the innovation theorem."

Theorem 1.12 (Innovation Theorem) *The process* $Z = \{Z_t,\ t \geq 0\}$ *defined by* (1.18) *and* (1.19) *is a standard Brownian motion with respect to the filtration* $\mathbb{F}$ *that is generated by* Y.

Proof From (1.18) and the tower property of conditional expectations we have that

$$E\left(\int_s^t (\xi_u - \mu_u)\, du \Big| \mathcal{F}_s\right) = 0 \qquad \text{for } 0 \leq s \leq t. \tag{1.20}$$

Moreover,

$$E(W_t - W_s | \mathcal{F}_s) = 0 \qquad \text{for } 0 \leq s \leq t, \tag{1.21}$$

because $W_t - W_s$ is independent of both ξ and $\{W_u,\ 0 \le u \le s\}$. Also, Z is adapted to $\mathbb{F}$, so it is a martingale with respect to $\mathbb{F}$ by (1.17), (1.19), (1.20), and (1.21). Finally, from (1.17) and (1.19) it follows that Z is continuous and has the same quadratic variation as W; that is, its quadratic variation is t over each interval $[0, t]$. Thus Z is a standard Brownian motion with respect to $\mathbb{F}$ by Theorem 1.11. □

1.7 A joint distribution (Reflection principle)

Let X be a (μ, σ) Brownian motion *with starting state zero* on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, P)$. Also, let $M_t := \sup\{X_s, 0 \le s \le t\}$ and then define the joint distribution function

$$F_t(x, y) := P\{X_t \le x,\ M_t \le y\}. \tag{1.22}$$

Because $X_0 = 0$ by hypothesis, one need only calculate $F_t(x, y)$ for $y \ge 0$ and $x \le y$; the discussion is hereafter restricted to (x, y) pairs satisfying those two conditions. We shall compute F for *standard* Brownian motion in this section and then extend the calculation to general μ and σ in Section 1.9. Temporarily fixing $\mu = 0$ and $\sigma = 1$, note first that

$$\begin{aligned} F_t(x, y) &= P\{X_t \le x\} - P\{X_t \le x,\ M_t > y\} \\ &= \Phi\left(xt^{-1/2}\right) - P\{X_t \le x,\ M_t > y\} \end{aligned} \tag{1.23}$$

where $\Phi(\cdot)$ is the $\mathcal{N}(0, 1)$ distribution function. Now the term $P\{X_t \le x, M_t > y\}$ can be calculated heuristically using the so-called reflection principle (note that the restriction $\mu = 0$ is critical here) as follows: for every sample path of X that hits level y before time t but finishes below level x at time t, there is another equally probable path (shown by the dotted line in Figure 1.1) that hits y before time t and then travels *upward* at least $y - x$ units to finish above level $y + (y - x) = 2y - x$ at time t. Thus

$$\begin{aligned} P\{X_t \le x,\ M_t > y\} &= P\{X_t \ge 2y - x\} \\ &= P\{X_t \le x - 2y\} = \Phi\left((x - 2y)t^{-1/2}\right). \end{aligned} \tag{1.24}$$

This argument is not rigorous, of course, but it can be made so using the strong Markov property of Section 1.3, as follows. Let T be the first t at which $X_t = y$, and define X^* as in Theorem 1.7. From Theorem 1.7 it follows that

$$\begin{aligned} P\{X_t \le x,\ M_t > y\} &= P\{T < t,\ X^*(t - T) \le x - y\} \\ &= P\{T < t,\ X^*(t - T) \ge y - x\}. \end{aligned}$$

(The strong Markov property is needed to justify the *second* of these equalities.) By definition $X^*(t-T) = X(t) - y$ and thus we arrive at (1.24). Combining (1.23) and (1.24) gives the following proposition. For the corollary, differentiate with respect to x.

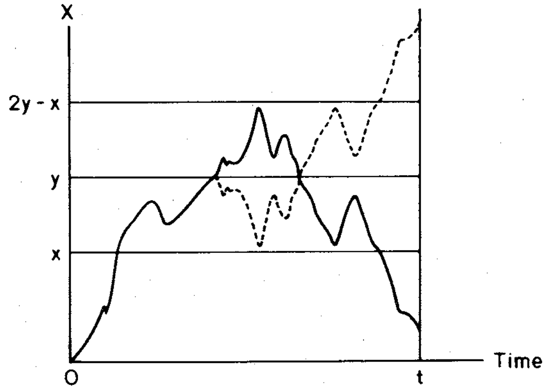

Figure 1.1 The reflection principle.

Proposition 1.13 *If* $\mu = 0$ *and* $\sigma = 1$*, then*

$$P\{X_t \le x,\ M_t \le y\} = \Phi\left(xt^{-1/2}\right) - \Phi\left((x-2y)t^{-1/2}\right). \tag{1.25}$$

Corollary 1.14 $P\{X_t \in dx,\ M_t \le y\} = g_t(x, y)\, dx$, *where*

$$g_t(x, y) := \left[\phi\left(xt^{-1/2}\right) - \phi\left((x-2y)t^{-1/2}\right)\right] t^{-1/2} \tag{1.26}$$

and $\phi(z) := (2\pi)^{-1/2} \exp(-z^2/2)$ *is the* $N(0, 1)$ *density function.*

1.8 Change of drift as change of measure

Continuing the development in the previous section, let $T > 0$ be fixed and deterministic, and restrict X to the time domain $[0, T]$. Starting with the (μ, σ) Brownian motion $X = \{X_t,\ 0 \le t \le T\}$ on $(\Omega, \mathcal{F}, \mathbb{F}, P)$, suppose we want to construct a $(\mu + \theta, \sigma)$ Brownian motion, also with time domain $[0, T]$. One approach is to keep the original space $(\Omega, \mathbb{F}, P)$ and define a new process $Z_t(\omega) = X_t(\omega) + \theta t,\ 0 \le t \le T$. Then Z is a $(\mu + \theta, \sigma)$ Brownian motion on $(\Omega, \mathcal{F}, \mathbb{F}, P)$.

Another approach is to keep the original process X and change the probability measure. The idea is to replace P by some other probability measure