Chapter 1

One-Dimensional Brownian Motion

1.1. Some motivation

The biologist Robert Brown noticed almost two hundred years ago that bits of pollen suspended in water undergo chaotic behavior. The bits of pollen are much more massive than the molecules of water, but of course there are many more of these molecules than there are bits of pollen. The chaotic motion of the pollen is the result of many infinitesimal jolts by the water molecules. By the central limit theorem (CLT), the law of the motion of the pollen should be closely related to the normal distribution. We now call this law Brownian motion.

During the past half century or so, Brownian motion has turned out to be a very versatile tool for both theory and applications. As we will see in Chapter 6, it provides a very elegant and general treatment of the Dirichlet problem, which asks for harmonic functions on a domain with prescribed boundary values. It is also the main building block for the theory of stochastic calculus, which is the subject of Chapter 5. Via stochastic calculus, it has played an important role in the development of financial mathematics.

As we will see later in this chapter, Brownian paths are quite rough — they are of unbounded variation in every time interval. Therefore, integrals with respect to them cannot be defined in the Stieltjes sense. A new type of integral must be defined, which carries the name of K. Itô, and more recently, of W. Doeblin. This new integral has some unexpected properties. Here is an example: If $B(t)$ is standard Brownian motion at time t with

$B(0) = 0$, then

$$\int_0^t B(s)\,dB(s) = \frac{1}{2}\big[B^2(t) - t\big]. \tag{1.1}$$

Of course, if $B(t)$ could be used as an integrator in the Stieltjes sense, and this were the Stieltjes integral, the right side would not contain the term $-t$.

There are also many important applications of Brownian motion connected with the classical limit theorems of probability theory. If $\xi_1, \xi_2, \ldots$ are i.i.d. random variables with mean zero and variance one, and

$$S_n = \xi_1 + \cdots + \xi_n$$

are their partial sums, the CLT says that $S_n/\sqrt{n}$ converges in distribution to the standard normal law. How can one embed the CLT into a more general theory that includes as one of its consequences the fact that $\max\{0, S_1, \ldots, S_n\}/\sqrt{n}$ converges in distribution to the absolute value of a standard normal? The answer involves Brownian motion in a crucial way, as we will see later in this chapter. Here is an early hint: For $t \geq 0$ and $n \geq 1$, let

$$X_n(t) = \frac{S_{[nt]}}{\sqrt{n}}, \tag{1.2}$$

where $[\cdot]$ is the integer part function. Then $X_n(1) = S_n/\sqrt{n}$, and

$$\max_{0 \leq t \leq 1} X_n(t) = \frac{\max\{0, S_1, \ldots, S_n\}}{\sqrt{n}}.$$

So, we have written both functionals of the partial sums in terms of the stochastic process $X_n(t)$. Once we show that X_n converges in an appropriate sense to Brownian motion, we will have a limit theorem for

$$\max\{0, S_1, \ldots, S_n\},$$

as well as for many other functions of the partial sums.

This chapter represents but a very small introduction to a huge field. For further reading, see [**35**] and [**40**].

1.2. The multivariate Gaussian distribution

Before defining Brownian motion, we will need to review the multivariate Gaussian distribution. Recall that a random variable ξ has the standard Gaussian (or normal) distribution $N(0, 1)$ if it has density

$$\frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty.$$

It is said to be univariate Gaussian if it can be written in the form $\xi = a\zeta + b$, where ζ is standard Gaussian and a, b are real. Note that this definition allows ξ to have zero variance. The Gaussian distribution with mean m and variance σ^2 (obtained above if $b = m$ and $a^2 = \sigma^2$) is denoted by $N(m, \sigma^2)$.

Definition 1.1. The real random vector $(\xi_1, \dots, \xi_n)$ is said to be multivariate Gaussian if all linear combinations

$$\sum_{k=1}^{n} a_k \xi_k$$

of the components have univariate Gaussian distributions.

Remark 1.2. (a) If $\xi_1, \dots, \xi_n$ are independent Gaussians, then $(\xi_1, \dots, \xi_n)$ is multivariate Gaussian.

(b) Definition 1.1 is much stronger than the statement that each ξ_k is Gaussian. For example, suppose ζ is standard Gaussian, and

$$\xi = \begin{cases} +\zeta & \text{if } |\zeta| \le 1; \\ -\zeta & \text{if } |\zeta| > 1. \end{cases}$$

Then ξ is also standard Gaussian. However, since $|\zeta + \xi| \le 2$ and $\zeta + \xi$ is not constant, $\zeta + \xi$ is not Gaussian, so (ζ, ξ) is not bivariate Gaussian.

Remark 1.3. Definition 1.1 has a number of advantages over the alternative, in which one specifies the joint density of $(\xi_1, \dots, \xi_n)$:

(a) It does not require that $(\xi_1, \dots, \xi_n)$ have a density. For example, (ξ, ξ) is bivariate Gaussian if ξ is Gaussian.

(b) It makes the next result immediate.

Proposition 1.4. *Suppose $\xi = (\xi_1, \dots, \xi_n)$ is Gaussian and A is an $m \times n$ matrix. Then the random vector $\zeta = A\xi$ is also Gaussian.*

Proof. Any linear combination of $\zeta_1, \dots, \zeta_m$ is some other linear combination of $\xi_1, \dots, \xi_n$. □

An important property of a multivariate Gaussian vector ξ is that its distribution is determined by the mean vector $E\xi$ and the covariance matrix, whose (i, j) entry is $\mathrm{Cov}(\xi_i, \xi_j)$. To check this statement, we use characteristic functions. Recall that the characteristic function of a random variable with the $N(m, \sigma^2)$ distribution is

$$\exp\left\{ itm - \frac{1}{2} t^2 \sigma^2 \right\}.$$

Therefore, if $\xi = (\xi_1, \dots, \xi_n)$ is multivariate Gaussian, its joint characteristic function is given by

$$\phi(t_1, \dots, t_n) = E \exp\left\{ i \sum_{j=1}^{n} t_j \xi_j \right\} = \exp\left\{ im - \frac{1}{2} \sigma^2 \right\},$$

where m and σ^2 are the mean and variance of $\sum_{j=1}^n t_j \xi_j$:

$$m = \sum_{j=1}^n t_j E\xi_j \quad \text{and} \quad \sigma^2 = \sum_{j,k=1}^n t_j t_k \operatorname{Cov}(\xi_j, \xi_k).$$

Since $\phi(t_1, \ldots, t_n)$ depends on ξ only through its mean vector and covariance matrix, these determine the characteristic function of ξ, and hence its distribution by Proposition A.24. This observation has the following consequence:

Proposition 1.5. *If $\xi = (\xi_1, \ldots, \xi_n)$ is multivariate Gaussian, then the random variables $\xi_1, \ldots, \xi_n$ are independent if and only if they are uncorrelated.*

Proof. That independence implies uncorrelatedness is always true for random variables with finite second moments. For the converse, suppose that $\xi_1, \ldots, \xi_n$ are uncorrelated, i.e., that $\operatorname{Cov}(\xi_j, \xi_k) = 0$ for $j \neq k$. Take $\zeta_1, \ldots, \zeta_n$ to be independent, with ζ_i having the same distribution as ξ_i. Then ξ and $\zeta = (\zeta_1, \ldots, \zeta_n)$ have the same characteristic function, and hence the same distribution, by Proposition A.24. It follows that $\xi_1, \ldots, \xi_n$ are independent. $\square$

The next exercise will be useful in Chapter 6 — see Proposition 6.12.

Exercise 1.6. Show that if $\xi = (\xi_1, \ldots, \xi_n)$, where $\xi_1, \ldots, \xi_n$ are i.i.d. standard Gaussian random variables, and O is an $n \times n$ orthogonal matrix, then $O\xi$ has the same distribution as ξ.

For the next exercise, recall that $\Rightarrow$ denotes convergence in distribution — see Definition A.18.

Exercise 1.7. (a) Suppose that $\xi_k \Rightarrow \xi$ and that ξ_k has the $N(m_k, \sigma_k^2)$ distribution for each k. Prove that ξ is $N(m, \sigma^2)$ for some m and σ^2, and that $m_k \to m$ and $\sigma_k^2 \to \sigma^2$. (Suggestion: First reduce the problem to the mean zero case by symmetrization, i.e., consider $\xi_k - \overline{\xi}_k$, where $\overline{\xi}_k$ is independent of ξ_k and has the same distribution.)

(b) State an analogue of (a) for Gaussian random vectors, and prove it using part (a). (Recall the Cramér-Wold device, Theorem A.26.)

The main topic of this book is a class of stochastic processes; in this chapter, they are Gaussian. We conclude this section with formal definitions of these concepts.

Definition 1.8. A stochastic process is a collection of random variables indexed by time. It is a discrete time process if the index set is a subset of the integers, and a continuous time process if the index set is $[0, \infty)$ (or sometimes, $(-\infty, \infty)$).

Definition 1.9. A stochastic process $X(t)$ is Gaussian if for any $n \geq 1$ and any choice of times $t_1, \ldots, t_n$, the random vector $(X(t_1), \ldots, X(t_n))$ has a multivariate Gaussian distribution. Its mean and covariance functions are $EX(t)$ and $\text{Cov}(X(s), X(t))$ respectively.

1.3. Processes with stationary independent increments

As we will see shortly, Brownian motion is not only a Gaussian process, but is a process with two other important properties — stationarity and independence of its increments. Here is the relevant definition.

Definition 1.10. A stochastic process $(X(t), t \geq 0)$ has stationary increments if the distribution of $X(t) - X(s)$ depends only on $t - s$ for any $0 \leq s \leq t$. It has independent increments if the random variables $\{X(t_{j+1}) - X(t_j), 1 \leq j < n\}$ are independent whenever $0 \leq t_1 < t_2 < \cdots < t_n$ and $n \geq 1$.

The simplest process with stationary independent increments is the Poisson process $N(t)$ with parameter $\lambda > 0$. It has the following properties:

(i) $N(t, \omega)$ is an increasing right continuous step function in t with jumps of size 1,

and

(ii) $N(t) - N(s)$ is Poisson distributed with parameter $\lambda(t - s)$ for $0 \leq s < t$.

It can be constructed in the following way: Let $\tau_1, \tau_2, \ldots$ be independent and identically distributed random variables that are exponentially distributed with parameter λ. Then let

$$N(t) = \#\{k \geq 1 : \tau_1 + \cdots + \tau_k \leq t\}. \tag{1.3}$$

Exercise 1.11. With $N(t)$ defined as in (1.3), show that if $0 < s < t$, then $N(s)$ and $N(t) - N(s)$ are independent Poisson distributed random variables with parameters λs and $\lambda(t - s)$ respectively.

1.4. Definition of Brownian motion

To see that the properties introduced in the previous two sections are likely to have a bearing on the definition of Brownian motion, note that the process $X_n(t)$ defined in (1.2) has independent increments, and that except for the effect of time discretization, it has stationary increments. Therefore, any limit $X(t)$ of $X_n(t)$ as $n \to \infty$, if it exists in any reasonable sense, will have stationary independent increments. Also, by the central limit theorem, $X(t)$ will have the $N(0, t)$ distribution. Thus, we would expect Brownian motion to be Gaussian and have stationary independent increments. The following result relates these properties.

Proposition 1.12. *The following two statements are equivalent for a stochastic process* $(X(t), t \ge 0)$*:*

(a) $X(t)$ *has stationary independent increments, and* $X(t)$ *is* $N(0,t)$ *for each* $t \ge 0$.

(b) $X(t)$ *is a Gaussian process with* $EX(t) = 0$ *and*

$$\mathrm{Cov}(X(s), X(t)) = s \wedge t.$$

Proof. Suppose (a) holds. To show that the process is Gaussian, take a_k's and t_k's as required in Definitions 1.1 and 1.9. Without loss of generality, we may assume that $0 = t_0 < t_1 < \cdots < t_n$. Summing by parts, and using $X(0) = 0$, we see that there are b_k's so that

$$\sum_{k=1}^{n} a_k X(t_k) = \sum_{k=1}^{n} b_k \big[X(t_k) - X(t_{k-1})\big].$$

The right side is a sum of independent Gaussians, and hence is Gaussian. To check the covariance statement, take $s < t$ and write

$$\begin{aligned}\mathrm{Cov}(X(s), X(t)) &= EX(s)X(t)\\ &= EX(s)\big[X(t) - X(s)\big] + EX^2(s)\\ &= s = s \wedge t.\end{aligned}$$

For the converse, assume (b). Then for $s < t$, $X(t) - X(s)$ is Gaussian with mean zero and

$$\mathrm{Var}(X(t) - X(s)) = t - 2(s \wedge t) + s = t - s,$$

so the process has stationary increments and has the correct marginal distributions.

To check independence of the increments, take $0 \le t_1 < t_2 < \cdots < t_n$, and write the vector of increments in the form

$$(X(t_2) - X(t_1), \ldots, X(t_n) - X(t_{n-1})) = A(X(t_1), \ldots, X(t_n))$$

for an appropriately chosen matrix A. Therefore, by Proposition 1.4, the vector of increments is Gaussian. So, in order to check the independence of the increments, it is enough by Proposition 1.5 to show that the increments are uncorrelated. To do so, take $u < v \le s < t$. Then

$$\begin{aligned}\mathrm{Cov}(X(v) - X(u), X(t) - X(s)) &= v \wedge t - v \wedge s - u \wedge t + u \wedge s\\ &= v - v - u + u = 0.\end{aligned}$$

□

Exercise 1.13. Suppose $X(t)$ is a stochastic process with stationary independent increments that satisfies $EX(1) = 0, EX^2(1) = 1$, and $X(t)$ has the same distribution as $\sqrt{t}X(1)$ for $t \geq 0$. Use the CLT to show that $X(t) - X(s)$ is $N(0, |t-s|)$ for $s, t \geq 0$.

Definition 1.14. A stochastic process $(X(t), t \geq 0)$ is said to have continuous paths if

$$P(\{\omega : X(t, \omega) \text{ is continuous in } t\}) = 1. \tag{1.4}$$

Definition 1.15. Standard Brownian motion $B(t)$ is a stochastic process with continuous paths that satisfies the equivalent properties (a) and (b) in Proposition 1.12.

Of course, it is not at all obvious that there exists a probability space on which one can construct a standard Brownian motion. Showing that this is the case is the objective of the next section. Taking this for granted for the time being, here are two exercises that provide some practice with the definition. In both cases, and throughout this book, $B(t)$ denotes standard Brownian motion.

Exercise 1.16. Let

$$X(t) = \int_0^t B(s)\, ds.$$

(a) Explain why $X(t)$ is a Gaussian process.

(b) Compute the mean and covariance functions of X.

(c) Compute $E(X(t) - X(s))^2$, and compare its rate of decay as $t \downarrow s$ with that of $E(B(t) - B(s))^2$.

Exercise 1.17. Compute

$$P(B(s) > 0, B(t) > 0), \quad 0 < s < t.$$

(Suggestion: Write the event in terms of two i.i.d. standard Gaussians, and change to polar coordinates.)

If it were not for the path continuity requirement in Definition 1.15, the existence of standard Brownian motion on some probability space would follow from Kolmogorov's extension theorem — see Theorem A.1. Since every event in this probability space is determined by the process at only countably many times, and continuity of a path is not determined by its values at countably many times, this approach would lead to the awkward situation in which the set in question,

$$C = \{\omega : B(t, \omega) \text{ is continuous in } t\},$$

is not an event. Therefore, we would not even be able to discuss the issue of whether its probability is 1.

The situation is even more serious than this. Even if C were measurable, it would not be possible to prove that $P(C) = 1$ follows from properties (a) and (b). To see this, take a process B on some probability space satisfying properties (a) and (b), and let τ be a continuous random variable that is independent of B. Define a new process by

$$X(t,\omega) = \begin{cases} B(t,\omega) & \text{if } t \neq \tau(\omega); \\ B(t,\omega) + 1 & \text{if } t = \tau(\omega). \end{cases}$$

Then not both B and X can have continuous paths. However, since

$$P(X(t) = B(t)) = P(\tau \neq t) = 1$$

for every t, it follows that X also satisfies properties (a) and (b) in Proposition 1.12. Therefore, there exist processes that satisfy properties (a) and (b) but do not have continuous paths. Note that with this construction,

$$P(X(t) = B(t) \text{ for all } t) = 0.$$

The fact that this can happen even though $P(X(t) = B(t)) = 1$ for every t is an early indication of how different things can be in discrete and continuous time. In the next section, we will see how to resolve these issues.

These comments suggest the importance of the following definition.

Definition 1.18. Two stochastic processes $X(t)$ and $Y(t)$ are versions of one another if $P(X(t) = Y(t)) = 1$ for all t.

The following exercise introduces an important variant of Brownian motion, which is known as the Brownian bridge, or tied down Brownian motion. It arises in the study of empirical distribution functions — see Section 7.8 of [**18**].

Exercise 1.19. Define $X(t) = B(t) - tB(1)$ for $0 \leq t \leq 1$, where B is standard Brownian motion.

(a) Show that X is a Gaussian process, and compute its covariance function $\text{Cov}(X(s), X(t))$.

(b) Show that for $0 < t_1 < \cdots < t_n < 1$, the (joint) distribution of $\big((B(t_1), \ldots, B(t_n)) \mid |B(1)| \leq \epsilon\big)$ converges to the (joint) distribution of $(X(t_1), \ldots, X(t_n))$ as $\epsilon \downarrow 0$.

The next exercise gives a limit law for the occupation time of A by Brownian motion up to time t — take the f below to be 1_A. This approach is computationally intensive, but provides good practice in working with Gaussian integrals and basic properties of Brownian motion. The proof becomes much easier and neater once we have developed some theory — see Exercise 5.59.

Exercise 1.20. Let B be standard Brownian motion, and put

$$X(t) = \frac{1}{\sqrt{t}} \int_0^t f(B(s))\, ds,$$

where $f \in L_1(R^1)$ and $\int f(x)\, dx = 1$.

(a) Show that

$$\lim_{t\to\infty} EX(t) = \sqrt{2/\pi} \quad \text{and} \quad \lim_{t\to\infty} EX^2(t) = 1.$$

(b) Use the method of moments — see Theorem A.31 — to prove that $X(t) \Rightarrow |Z|$ as $t \to \infty$, where Z is standard normal, following the outline below.

(i) For $\alpha_1, \ldots, \alpha_k > 0$, let $I(\alpha_1, \ldots, \alpha_k) =$

$$\int \cdots \int_{0<r_1<\cdots<r_k<1} r_1^{\alpha_1 - 1}(r_2 - r_1)^{\alpha_2 - 1} \cdots (r_k - r_{k-1})^{\alpha_k - 1}\, dr_1 \cdots dr_k.$$

Show that

$$\lim_{t\to\infty} EX^k(t) = \frac{k!}{(2\pi)^{k/2}} I\left(\frac{1}{2}, \ldots, \frac{1}{2}\right).$$

(ii) Using the Beta integral

$$\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\, du = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)},$$

check that

$$I(\alpha_1, \ldots, \alpha_k) = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)} I(\alpha_1+\alpha_2, \alpha_3, \ldots, \alpha_k),$$

so that

$$I\left(\frac{1}{2}, \ldots, \frac{1}{2}\right) = \frac{[\Gamma(1/2)]^k}{\Gamma(k/2)(k/2)}.$$

(iii) Check the recursion

$$E|Z|^k = (k-1)E|Z|^{k-2}, \quad k \geq 2,$$

and use it to compute $E|Z|^k$.

1.5. The construction

In this section, we will give one construction of Brownian motion. Another construction is outlined in the exercises.

Theorem 1.21. *There exists a probability space* $(\Omega, \mathcal{F}, P)$ *on which standard Brownian motion* B *exists.*

Proof. The finite-dimensional distributions of B are determined by properties (a) and (b) in Proposition 1.12. By Kolmogorov's extension theorem, Theorem A.1, there exists a probability space on which random variables $B(t)$ are defined with these joint distributions for $t \in Q^+$, the set of positive rationals. We will prove that for every $N \geq 1$,

(1.5) $\qquad B(t, \omega)$ is uniformly continuous in t for $t \in Q \cap [0, N]$ a.s.

Once this is done, $B(t)$ can be extended to all $t \geq 0$ by continuity. Note that the uniformity is important here. If $b \notin Q$, the function

$$f(t) = \begin{cases} 1 & \text{if } t \geq b; \\ 0 & \text{if } t < b \end{cases}$$

is continuous on Q, but not uniformly continuous on Q. It cannot be extended from Q to R^1 by continuity.

Let

$$\Delta_n = \sup_{\substack{s,t \in Q \cap [0,N] \\ |s-t| \leq \frac{1}{n}}} |B(t) - B(s)|.$$

We need to prove that $\Delta_n \to 0$ a.s. Since Δ_n is decreasing in n, it is enough to prove convergence in probability. To see this, recall that convergence in probability implies that a subsequence converges a.s. By the monotonicity, convergence along a subsequence implies convergence along the full sequence.

The next step is to reduce the number of arguments of B that need to be considered in the supremum. To do so, let

$$Y_{k,n} = \sup_{\substack{t \in Q \\ \frac{k-1}{n} \leq t \leq \frac{k}{n}}} \left| B(t) - B\left(\frac{k-1}{n}\right) \right|.$$

Then

$$\Delta_n \leq 3 \max_{1 \leq k \leq nN} Y_{k,n}. \tag{1.6}$$

The factor of 3 above arises in the following way. For a given $t \in Q \cap [0, N]$, choose k so that $\frac{k-1}{n} \leq t \leq \frac{k}{n}$. If $|s - t| \leq \frac{1}{n}$, then $\frac{k-2}{n} \leq s \leq \frac{k+1}{n}$. If, for example, $\frac{k}{n} \leq s \leq \frac{k+1}{n}$, bound $|B(t) - B(s)|$ by

$$\left| B(t) - B\left(\frac{k-1}{n}\right) \right| + \left| B\left(\frac{k}{n}\right) - B\left(\frac{k-1}{n}\right) \right| + \left| B(s) - B\left(\frac{k}{n}\right) \right|,$$

which is at most the right side of (1.6). A similar argument applies if $\frac{k-2}{n} \leq s \leq \frac{k-1}{n}$ or $\frac{k-1}{n} \leq s \leq \frac{k}{n}$.

Noting that the distribution of $Y_{k,n}$ does not depend on k, write for $\epsilon > 0$

$$P\left(\max_{1 \leq k \leq nN} Y_{k,n} > \epsilon \right) \leq \sum_{k=1}^{nN} P(Y_{k,n} > \epsilon) = nNP(Y_{1,n} > \epsilon). \tag{1.7}$$

To check that the right side above tends to 0 as $n \to \infty$, we will apply Doob's inequality for discrete time submartingales — see Theorem A.42. If $0 < t_1 < \cdots < t_m$ are rational, $(B(t_1), \ldots, B(t_m))$ is a martingale, since the successive differences are independent and have mean 0. Therefore, $(B^4(t_1), \ldots, B^4(t_m))$ is a (nonnegative) submartingale by Propostion A.37. Doob's inequality, Theorem A.42, gives

$$P\left(\max_{1\le k\le m} |B(t_k)| > \epsilon\right) \le \frac{1}{\epsilon^4} EB^4(t_m).$$

Note that the bound on the right side depends on t_m, but not on m — this is very important in the next step. Applying this to a sequence of subsets that exhausts $Q \cap [0, \frac{1}{n}]$, and using the fact that $EB^4(t)$ is increasing in t, we see that

$$P(Y_{1,n} > \epsilon) \le \frac{1}{\epsilon^4} EB^4\left(\frac{1}{n}\right).$$

Since $B(t)$ has the same distribution as $\sqrt{t}B(1)$ (the $N(0, t)$ distribution), we conclude that

$$P(Y_{1,n} > \epsilon) \le \frac{EB^4(1)}{n^2\epsilon^4},$$

so that the right side of (1.7) tends to 0 as $n \to \infty$. Therefore, $\Delta_n \to 0$ in probability by (1.6) as required. □

Exercise 1.22. Use Doob's inequality, Theorem A.42, applied to even powers larger than the fourth in the proof of Theorem 1.21 to obtain the following improvement: for every $N > 0$ and every $0 < \alpha < \frac{1}{2}$ there is a random variable C so that

$$|B(t) - B(s)| \le C|t - s|^\alpha, \quad 0 \le s, t \le N,$$

i.e., the paths are locally Hölder continuous with exponent α. (It follows from Exercise 1.30 below that this is not true if $\alpha = \frac{1}{2}$.)

The following exercises develop another approach to the construction problem.

Exercise 1.23. Let $\{\phi_n\}$ be a complete orthonormal family in $L_2[0, 1]$, and define

$$\psi_n(t) = \int_0^t \phi_n(s)\, ds = \langle \phi_n, 1_{[0,t]}\rangle,$$

where $\langle\cdot\rangle$ is the usual inner product in $L_2[0, 1]$:

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

Then let $\{\xi_n\}$ be i.i.d. standard normal random variables. Use Corollary A.16 to show that for each $t \in [0,1]$, the series

$$B(t) = \sum_n \xi_n \psi_n(t) \tag{1.8}$$

converges a.s. and in L_2, and that the resulting process B satisfies properties (a) and (b) in Proposition 1.12. In doing so, you will find Parseval's identity

$$\langle f, g\rangle = \sum_n \langle f, \phi_n\rangle\langle g, \phi_n\rangle$$

useful.

Exercise 1.24. Show that if $\{\xi_n\}$ are i.i.d. standard normal random variables, then for every $\epsilon > 0$,

$$\lim_{n\to\infty} \xi_n n^{-\epsilon} = 0 \text{ a.s.}$$

Exercise 1.25. In Exercise 1.23, take the orthonormal family to be the Haar functions, which are defined as follows: $\phi_0 \equiv 1$,

$$\phi_{0,1}(t) = \begin{cases} +1 & \text{if } 0 \le t \le \frac{1}{2}; \\ -1 & \text{if } \frac{1}{2} < t \le 1, \end{cases}$$

$$\phi_{1,1}(t) = \begin{cases} +\sqrt{2} & \text{if } 0 \le t \le \frac{1}{4}; \\ -\sqrt{2} & \text{if } \frac{1}{4} < t \le \frac{1}{2}; \\ 0 & \text{if } \frac{1}{2} < t \le 1, \end{cases} \qquad \phi_{1,2}(t) = \begin{cases} 0 & \text{if } 0 \le t \le \frac{1}{2}; \\ +\sqrt{2} & \text{if } \frac{1}{2} < t \le \frac{3}{4}; \\ -\sqrt{2} & \text{if } \frac{3}{4} < t \le 1, \end{cases}$$

and in general, for $1 \le k \le 2^n$,

$$\phi_{n,k}(t) = \begin{cases} +2^{n/2} & \text{if } \frac{k-1}{2^n} \le t \le \frac{k-(1/2)}{2^n}; \\ -2^{n/2} & \text{if } \frac{k-(1/2)}{2^n} < t \le \frac{k}{2^n}; \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find a good upper bound in terms of the coefficients $a_1, \ldots, a_{2^n}$ for

$$\max_{0\le t\le 1} \left| \sum_{k=1}^{2^n} a_k \phi_{n,k}(t) \right|.$$

(b) Use Exercise 1.24 to show that the series in (1.8) converges uniformly on $[0,1]$, and hence defines a standard Brownian motion there.

Remark 1.26. The uniform convergence of the series (1.8) holds for all complete orthonormal families $\{\phi_n\}$, but the proof is more difficult than it is for the Haar functions.

The Brownian motion property is preserved by several transformations, as we now check.

Theorem 1.27. *Suppose that B is a standard Brownian motion. Then the following processes are also:*

$$X_1(t) = B(t+s) - B(s), \quad s > 0 \text{ fixed.}$$
$$X_2(t) = \frac{B(ct)}{\sqrt{c}}, \quad c > 0 \text{ fixed.}$$
$$X_3(t) = \begin{cases} tB(1/t) & \text{if } t > 0; \\ 0 & \text{if } t = 0. \end{cases}$$

Proof. The first two cases are left as an exercise. For the third, note that X_3 is a mean zero Gaussian process that is continuous, except possibly at 0. To check the covariance, write

$$EX_3(s)X_3(t) = st\min(1/s, 1/t) = s \wedge t.$$

To check continuity at 0, write

$$\{\omega : \lim_{t\downarrow 0} B(t) = 0\} = \bigcap_{m\geq 1}\bigcup_{n\geq 1}\{\omega : |B(t)| \leq 1/m \text{ for all } t \in Q \cap (0, 1/n)\}$$

and

$$\{\omega : \lim_{t\downarrow 0} X_3(t) = 0\} = \bigcap_{m\geq 1}\bigcup_{n\geq 1}\{\omega : |X_3(t)| \leq 1/m \text{ for all } t \in Q \cap (0, 1/n)\}.$$

The right sides of these two identities have the same probability, since the two processes have the same finite-dimensional distributions, and the events depend on the processes at only countably many times. The left side of the first has probability 1, since B has continuous paths. Therefore the left side of the second also has probability 1. □

Exercise 1.28. Check that $X_1(t)$ and $X_2(t)$ in Theorem 1.27 are Brownian motions.

The next two exercises show that Theorem 1.27 can be quite useful, in spite of its simplicity. Note that $B(n)$ is a sum of the i.i.d. increments $B(k) - B(k-1)$, so that $B(n)/n \to 0$ a.s. as $n \to \infty$ by the strong law of large numbers, Theorem A.17. A reasonable question is whether one can take the limit along all $t \to \infty$, rather than just along the integers.

Exercise 1.29. Use the fact that $X_3(t)$ from Theorem 1.27 is a Brownian motion to show that $B(t)/t \to 0$ a.s. as $t \to \infty$.

Exercise 1.30. Use the fact that $X_3(t)$ from Theorem 1.27 is a Brownian motion and Corollary A.28 to show that

$$\limsup_{t\downarrow 0} \frac{B(t)}{\sqrt{t}} = +\infty \quad \text{and} \quad \liminf_{t\downarrow 0} \frac{B(t)}{\sqrt{t}} = -\infty \text{ a.s.}$$

In particular,

$$P\big(\forall \epsilon > 0, B(t) \text{ takes both signs in } [0,\epsilon)\big) = 1.$$

The next three exercises are designed to show that the property of path continuity is rather rare among continuous time processes. They also serve to introduce the symmetric stable processes (see Definition A.29 for the definition of a stable law) and compound Poisson processes. Both are examples of Lévy processes, which are introduced in Chapter 3. Note that Brownian motion is a symmetric stable process with $\alpha = 2$.

The acronym "càdlàg" (continue à droite, limites à gauche) below refers to paths that are right continuous and have left limits. The càdlàg property helps to resolve many measurability issues. For example, if $X(t)$ has càdlàg paths, then the set

$$C = \{\omega : X(t,\omega) \text{ is continuous for } t \in [0,N]\}$$

is measurable, since it can be written as

$$C = \Big\{\omega : \lim_{n\to\infty} \sup_{\substack{s,t\in Q\cap[0,N] \\ |s-t|\le \frac{1}{n}}} |X(s) - X(t)| = 0\Big\}.$$

Exercise 1.31. Suppose that $(X(t), t \ge 0)$ is a stochastic process with the following three properties:

(i) It has stationary independent increments.

(ii) For each t, $X(t)$ has a symmetric stable law of index $\alpha \in (0,2]$. Specifically, take its characteristic function to be $Ee^{iuX(t)} = e^{-t|u|^\alpha}$. A consequence if $\alpha < 2$ is that $P(|X(1)| \ge x) \sim cx^{-\alpha}$ for some constant $c > 0$ as $x \to \infty$. (Note that this is false if $\alpha = 2$.)

(iii) Its sample paths are a.s. càdlàg. (That there is a version of the process with this property will be proved in Chapter 3.)

(a) Prove that for each t, $X(\cdot)$ is a.s. continuous at t.

(b) Express

$$P\left(\max_{1\le k\le n} \left| X\left(\frac{k}{n}\right) - X\left(\frac{k-1}{n}\right)\right| \ge \epsilon\right)$$

in terms of the distribution of $X(1)$.

(c) Prove that if $\alpha < 2$, then

$$P(X(\cdot) \text{ is continuous on } [0,1]) = 0.$$

Exercise 1.32. Let $N(t)$ be a rate λ Poisson process, and let Y_j be i.i.d. random variables independent of N. Define a continuous time process

$(Z(t), 0 \le t \le 1)$ by

$$Z(t) = \sum_{j=1}^{N(t)} Y_j.$$

Such processes are called compound Poisson.

(a) Compute the characteristic function of $Z(t)$ in terms of the characteristic function of Y_j.

(b) Show that Z has stationary independent increments.

(c) Prove that for each t, $Z(\cdot)$ is a.s. continuous at t.

(d) Compute

$$P(Z(\cdot) \text{ is continuous on } [0,t]).$$

Exercise 1.33. Let $Z_n(t)$ be as in the previous exercise with $\lambda = n$ and Y_j with characteristic function $\phi_n(u)$, where

$$\lim_{n\to\infty} n[1 - \phi_n(u)] = |u|^\alpha,$$

with $0 < \alpha \le 2$. Show that the finite-dimensional distributions of Z_n converge to those of the stable process from Exercise 1.31.

The following measurability fact will often be useful. For its statement, we assume that Ω has been modified so that $B(\cdot, \omega)$ is continuous for all ω.

Proposition 1.34. *$B(t, \omega)$ is jointly measurable in (t, ω).*

Proof. Let

$$X_n(t) = B\left(\frac{[nt]}{n}\right),$$

where $[\cdot]$ is the greatest integer function. This is jointly measurable for each n since for any Borel set A,

$$\{(t,\omega) : X_n(t,\omega) \in A\} = \bigcup_{k=0}^{\infty} \left[\frac{k}{n}, \frac{k+1}{n}\right) \times \left\{\omega : B\left(\frac{k}{n}, \omega\right) \in A\right\},$$

and each set on the right is a measurable rectangle. By path continuity,

$$X_n(t,\omega) \to B(t,\omega)$$

uniformly on compact t sets. Therefore B is jointly measurable. □

1.6. Path properties

This section is devoted to regularity properties — first some that Brownian motion fails to have, then some it does have.

Feller Processes

3.1. The basic setup

The first part of this chapter is an analogue of the material in the first part of Chapter 2 for Markov processes with a more general state space S. A reasonable question to raise is, why bother with the part of Chapter 2 that deals with the construction of the process at all? Why not go directly to the general case? There are at least two answers. First, everything is much more concrete when S is countable, so it provides a good setting to use for a first exposure to the construction problem. Second, the results in Chapter 2 are not really special cases of those in this chapter. If one applies the results in this chapter to countable S, one gets only a special class of Markov chains. Recall that the Markov chains, transition functions, and Q-matrices from Section 2.1 are not exactly in one-to-one correspondences with one another. Their analogues in this chapter do have this desirable property. The price that one pays for this is that important examples of Markov chains are excluded from this treatment.

In this chapter, S is either a compact or locally compact separable metric space, and $C(S)$ is the space of continuous real-valued functions on S in the compact case, and the space of continuous real-valued functions vanishing at infinity in the locally compact case. In both cases, we use the uniform norm

$$||f|| = \sup_{x\in S} |f(x)|,$$

which makes $C(S)$ into a Banach space. The main reason for using continuous functions that vanish at infinity rather than bounded continuous functions in case S is locally compact is that uniformity is required in many arguments. Bounded continuous functions are not usually uniformly continuous, for example, while continuous functions vanishing at infinity are. Another reason is that in the proof of Theorem 3.26, we will use the fact that $C(S)$ is separable, which is not in general true for the space of all bounded continuous functions on S. See Section A.3 for definitions of these terms.

3.1.1. The process. We begin by describing the ingredients needed for the definition of the main object of interest in this chapter. Let

$$\Omega = D[0,\infty) = \text{the set of right continuous functions}$$
$$\omega : [0,\infty) \to S \text{ with left limits,}$$

$X(t,\omega) = \omega(t)$ and $(\theta_s\omega)(t) = \omega(t+s)$. The σ-algebra $\mathcal{F}$ on Ω is the smallest such that the mapping $\omega \to \omega(t)$ is measurable for each $t \geq 0$. This is analogous to the σ-algebra we used in Section 1.7 in the discussion of the Markov property for Brownian motion, and in the Markov chain context of Chapter 2.

Definition 3.1. A Feller process on S consists of

(a) a collection of probability measures $\{P^x, x \in S\}$ on $(\Omega, \mathcal{F})$

and

(b) a right continuous filtration $\{\mathcal{F}_t, t \geq 0\}$ on Ω with respect to which the random variables $X(t)$ are adapted,

satisfying

$$P^x(X(0) = x) = 1, \tag{3.1}$$

(3.2) the mapping $x \to E^x f(X(t))$ is in $C(S)$ for all $f \in C(S)$ and $t \geq 0$,

and

$$E^x(Y \circ \theta_s \mid \mathcal{F}_s) = E^{X(s)} Y \text{ a.s. } P^x \tag{3.3}$$

for all $x \in S$ and all bounded measurable Y on Ω.

Of course (3.3) is the usual Markov property that we have encountered before — in Theorem 1.46 and in (2.1). Statement (3.2) is known as the Feller property. Another way of stating the continuity part of it, which makes it appear quite natural, is that $x_n \to x$ implies that the distribution of $X(t)$ for the process starting at x_n converges weakly to that for the process starting at x. The Feller property (together with the right continuity of paths) implies the strong Markov property — see the proof of Theorem 1.68 and Remark 1.70. Recall that the strong Markov property may fail without the Feller property — see Example 1.87.

The third property mentioned in Remark 1.70 is that the function

$$x \to E^x \prod_{m=1}^{n} f_m(\omega(t_m)) \tag{3.4}$$

is continuous whenever $f_m \in C(S)$ for each m. When $n = 1$, this is a consequence of (3.2).

Exercise 3.2. Show that (3.2) implies that the mapping (3.4) is continuous for any n.

Theorem 3.3. *Every Feller process has the strong Markov property: If $Y_s(\omega)$ is bounded and jointly measurable on $[0, \infty) \times \Omega$, and τ is a stopping time, then for every x,*

$$E^x(Y_\tau \circ \theta_\tau \mid \mathcal{F}_\tau) = E^{X(\tau)} Y_\tau \quad a.s.\ P^x \quad on \quad \{\tau < \infty\}.$$

3.1.2. The semigroup. To motivate the next definition, suppose S is countable and $p_t(x, y)$ is a transition function. The transition function can be encoded in terms of a family of operators

$$T(t)f(x) = \sum_y p_t(x, y) f(y), \tag{3.5}$$

for $f \in C(S)$. Returning to general S, we have:

Definition 3.4. A probability semigroup is a family of continuous linear operators $\{T(t), t \geq 0\}$ on $C(S)$ satisfying the following properties:

(a) $T(0)f = f$ for all $f \in C(S)$.

(b) For every $f \in C(S)$, $\lim_{t \downarrow 0} T(t)f = f$.

(c) $T(s + t)f = T(s)T(t)f$ for every $f \in C(S)$.

(d) $T(t)f \geq 0$ for every nonnegative $f \in C(S)$.

(e) If S is compact: $T(t)1 = 1$ for each $t > 0$; if S is not compact: there exist $f_n \in C(S)$ so that $\sup_n ||f_n|| < \infty$, and $T(t)f_n$ converges to 1 pointwise for each $t \geq 0$.

This is the analogue of Definition 2.2. Part (c) is the analogue of the Chapman-Kolmogorov equations (2.3), and is called the semigroup property. One of its consequences is that $T(t)$ and $T(s)$ commute. It follows from parts (d) and (e) that $||T(t)f|| \leq ||f||$ for all $f \in C(S)$, so that each $T(t)$ is a contraction operator. Property (b) is known as strong continuity. Together with (c) and the contraction property, it implies that the function $t \to T(t)f$ from $[0, \infty)$ to $C(S)$ is continuous.

Here is an important example — the Brownian semigroup. Part (b) of the exercise illustrates the reasons for taking functions in $C(S)$ to vanish at infinity.

Exercise 3.5. Take $S = R^1$ and $X(t)$ to be Brownian motion.

(a) Show that $T(t)$ defined by

$$T(t)f(x) = E^x f(X(t))$$

is a probability semigroup.

(b) Explain why this would not be the case if $C(S)$ were the space of bounded continuous functions.

In this chapter, it will be necessary to integrate continuous $C(S)$-valued functions of t. The calculus of such functions is analogous to that of real-valued functions. The basic facts are given in Section A.3.

Associated with a semigroup is its Laplace transform

$$U(\alpha)f = \int_0^\infty e^{-\alpha t} T(t) f \, dt, \quad \alpha > 0, \tag{3.6}$$

which is called the resolvent of the semigroup. The integral in (3.6) is well-defined since the function $t \to e^{-\alpha t} T(t) f$ is continuous, and

$$||e^{-\alpha t} T(t) f|| \le e^{-\alpha t} ||f||$$

by the contraction property. Note that $U(\alpha)$ is a linear operator on $C(S)$, and satisfies

$$||U(\alpha)f|| \le \frac{1}{\alpha} ||f||,$$

and

$$\lim_{\alpha \to \infty} \alpha U(\alpha) f = f.$$

The semigroup property translates into the following useful relation, which is known as the resolvent equation:

$$U(\alpha) - U(\beta) = (\beta - \alpha) U(\alpha) U(\beta). \tag{3.7}$$

To check this, take $\alpha \ne \beta$ and write

$$\begin{aligned} U(\alpha)U(\beta)f &= \int_0^\infty e^{-\alpha t} T(t) U(\beta) f \, dt \\ &= \int_0^\infty e^{-\alpha t} \int_0^\infty e^{-\beta s} T(t) T(s) f \, ds \, dt \\ &= \int_0^\infty e^{-\alpha t} \int_0^\infty e^{-\beta s} T(s+t) f \, ds \, dt \\ &= \int_0^\infty T(r) f \int_0^r e^{-\alpha t - \beta(r-t)} \, dt \, dr \\ &= \int_0^\infty T(r) f \frac{e^{-\alpha r} - e^{-\beta r}}{\beta - \alpha} \, dr. \end{aligned} \tag{3.8}$$

One consequence of (3.7) is that $U(\alpha)$ and $U(\beta)$ commute.

Exercise 3.6. In the second step of (3.8), the order of the action of $T(t)$ and the integration with respect to s was reversed. Justify this step.

Exercise 3.7. Show that the resolvent for Brownian motion is given by

$$U(\alpha)f(x) = \frac{1}{\sqrt{2\alpha}} \int_{-\infty}^{\infty} f(y) e^{-\sqrt{2\alpha}|x-y|} dy.$$

The next two exercises are designed to show that when S is countable, the setting in this chapter is more restrictive than that of Chapter 2.

Exercise 3.8. Suppose $p_t(x, y)$ is a transition function on a finite or countable set S, and let $T(t)$ be defined as in (3.5).

(a) Show that if S is finite, then $T(t)$ is a probability semigroup in the sense of Definition 3.4.

(b) Show that if S is infinite, then $T(t)$ is a probability semigroup in the sense of Definition 3.4 if and only if

$$\lim_{x\to\infty} p_t(x, y) = 0, \text{ for all } y \in S, \ t > 0. \tag{3.9}$$

Exercise 3.9. Take $S = \{0, 1, 2, \dots\}$ and the Q-matrix for a pure death process, given by $q(0, 1) = 1,\ q(0, 0) = -1,\ q(i, i-1) = \delta_i,\ q(i, i) = -\delta_i$ for $i \geq 1$, and $q(i, j) = 0$ otherwise, where $\delta_i > 0$.

(a) Find a necessary and sufficient condition on the δ_i's for (3.9) to hold.

(b) How does the condition in (a) compare with the necessary and sufficient condition for the minimal solution of (2.13) to be stochastic?

3.1.3. Lévy processes. In Chapter 1, we discussed Brownian motion in detail, and briefly described symmetric stable processes and compound Poisson processes. An increasing stable process appeared in Exercise 1.85. All are special cases of Lévy processes, which we describe now, in order to provide a large class of examples of Feller processes. Let

$$\psi(u) = i\beta u - \frac{\sigma^2 u^2}{2} + \int_{-\infty}^{\infty} \left(e^{iux} - 1 - \frac{iux}{1+x^2} \right) \nu(dx), \tag{3.10}$$

where β is real, $\sigma \geq 0$, and ν is a measure on R^1 satisfying

$$\int_{-\infty}^{\infty} \frac{x^2}{1+x^2} \nu(dx) < \infty.$$

Then $e^{\psi(u)}$ is a characteristic function. To see this, note first that if $\nu = 0$, then this is the characteristic function of the normal distribution with mean β and variance σ^2. Recall that Z is compound Poisson if it can be written in the form

$$Z = \sum_{j=1}^{N} Y_j,$$

where $Y_1, Y_2, \dots$ are i.i.d. random variables and N is Poisson distributed with parameter λ, and independent of the Y_j's. In this case, Z has characteristic function $e^{\psi(u)}$, where

$$\psi(u) = \lambda \int_{-\infty}^{\infty} (e^{iux} - 1)\mu(dx)$$

and μ is the distribution of the Y_j's. This is of the form (3.10) with

$$\nu = \lambda\mu, \ \sigma = 0, \ \text{and } \beta = \lambda \int_{-\infty}^{\infty} \frac{x}{1+x^2}\mu(dx).$$

Therefore, if ν is finite, $e^{\psi(u)}$ is the characteristic function of the sum of two independent random variables, one normal and the other compound Poisson. The general case is obtained by taking limits — see Proposition A.25.

The Lévy process corresponding to the ψ in (3.10) is the one with semigroup $T(t)$ on $S = R^1$ defined by

$$T(t)f(x) = Ef(x + \xi_t),$$

where ξ_t is a random variable with characteristic function $e^{t\psi(u)}$. If $\beta = \nu = 0$ and $\sigma^2 = 1$, this is Brownian motion. If $\beta = \sigma^2 = 0$ and

$$\nu(dx) = \frac{1}{|x|^{\alpha+1}} dx, \quad 0 < \alpha < 2,$$

it is a symmetric stable process of index α.

Exercise 3.10. Show that the $T(t)$ defined above is a probability semigroup.

Exercise 3.11. Show that Lévy processes on R^1 have stationary independent increments.

In fact, Lévy processes are the most general Feller processes on R^1 with stationary independent increments. For more on this important topic, see [**1**].

3.1.4. The generator. So far, these definitions should be quite natural. The next one is less so, but it turns out to be the right analogue of the definition of a Q-matrix. To make the connection, given a Q-matrix on a countable S, let

$$\mathcal{L}f(x) = \sum_y q(x,y)f(y) = \sum_y q(x,y)[f(y) - f(x)] \tag{3.11}$$

for an appropriate set of functions f on S. In general, some restriction on f will clearly be needed to make the series converge.

We will use $\mathcal{D}$ and $\mathcal{R}$ to denote the domain and range of an operator respectively.

Definition 3.12. A probability generator is a (usually unbounded) linear operator $\mathcal{L}$ on $C(S)$ satisfying the following properties:

(a) $\mathcal{D}(\mathcal{L})$ is dense in $C(S)$.

(b) If $f \in \mathcal{D}(\mathcal{L}), \lambda \geq 0$, and $f - \lambda\mathcal{L}f = g$, then

$$\inf_{x\in S} f(x) \geq \inf_{x\in S} g(x).$$

(c) $\mathcal{R}(I - \lambda\mathcal{L}) = C(S)$ for all sufficiently small $\lambda > 0$.

(d) If S is compact: $1 \in \mathcal{D}(\mathcal{L})$ and $\mathcal{L}1 = 0$; if S is not compact: for small positive λ there exist $f_n \in \mathcal{D}(\mathcal{L})$ (that may depend on λ) so that $g_n = f_n - \lambda\mathcal{L}f_n$ satisfies $\sup_n ||g_n|| < \infty$, and both f_n and g_n converge to 1 pointwise.

Note that property (b) has the following consequence:

$$f \in \mathcal{D}(\mathcal{L}), \lambda \geq 0, f - \lambda\mathcal{L}f = g \text{ imply } ||f|| \leq ||g||. \tag{3.12}$$

To see this, write

$$\inf_{x\in S} g(x) \leq \inf_{x\in S} f(x) \leq \sup_{x\in S} f(x) \leq \sup_{x\in S} g(x),$$

with the last inequality coming from (b) with f and g replaced by $-f$ and $-g$ respectively. So, for sufficiently small positive λ, $(I - \lambda\mathcal{L})^{-1}$ is an everywhere defined contraction that maps nonnegative functions to nonnegative functions.

Since Definition 3.12 is a bit abstract, it may be helpful to consider the following example, which turns out to be the generator of the process on the line that moves to the right at unit speed — see Exercise 3.18. Note that the hardest property to check is (c). This is usually the case.

Exercise 3.13. Suppose that $S = R^1$,

$$\mathcal{D}(\mathcal{L}) = \{f \in C(S) : f' \in C(S)\},$$

and $\mathcal{L}f = f'$. Show that $\mathcal{L}$ is a probability generator.

Exercise 3.14. Suppose that Q is a Q-matrix on the finite or countable set S, as defined in Definition 2.3, and let $\mathcal{L}f$ be defined as in (3.11) for those $f \in C(S)$ for which the series converges for each x, and the resulting $\mathcal{L}f$ is in $C(S)$.

(a) Show that if S is finite, then $\mathcal{L}$ is a probability generator.

(b) Take $S = \{0, 1, 2, \dots\}$ and the Q-matrix for a pure death chain given by $q(0, 1) = 1,\ q(0, 0) = -1,\ q(i, i - 1) = \delta_i,\ q(i, i) = -\delta_i$ for $i \geq 1$, and $q(i, j) = 0$ otherwise, where $\delta_i > 0$. Show that properties (a), (b), and (d) of Definition 3.12 are satisfied, and find a necessary and sufficient condition on the δ_i's so that $\mathcal{R}(I - \lambda\mathcal{L})$ contains all functions with finite support.

Chapter 4

Interacting Particle Systems

4.1. Some motivation

The topic of this chapter is a more recently developed field than Brownian motion, Markov chains, and Feller processes, which are the subjects of the first three chapters of this book. The field began about 1970 — see [**42**] and [**16**] — and is still the subject of vigorous development. It had a number of distinct motivations:

(a) *Mathematical.* Most interacting particle systems are Feller processes, but of a nature that is very different from that of the Markov processes that motivated their study prior to 1970. Those earlier processes typically had a state space that is countable, or R^n, or perhaps a more general manifold. Interacting particle systems usually live on the totally disconnected state space $\{0,1\}^S$, where S is a countable set. The earlier processes tended to have strong irreducibility properties, but this is no longer the case in the present context. A consequence of this is that relatively little of the earlier theory has much to say about the new systems. While the entire infinite system is Markovian, the evolution of an individual particle is not. Thus the new field represents a departure from the Markovian world, without leaving it entirely.

(b) *Applied.* Roughly in parallel with the mathematical development, models of this new type appeared in the scientific literature. Certain spin systems provided natural time evolutions for statistical mechanical systems such as the Ising model, and served as useful tools in Monte Carlo studies. Research in areas as different as traffic behavior and polymers suggested models of the exclusion type. Contact and biased voter models were proposed to improve understanding of the mechanisms of spread of infection and tumor growth. Many of the applied papers were nonrigorous, but through both heuristic arguments and simulations, they suggested profitable directions for the more theoretical research.

This chapter is an introduction to this large and active field. It starts from the foundations provided by Chapter 3, but goes on to investigate issues that are of particular relevance to these models. Much more about the subject can be found in [**31**] and [**32**], and the references therein. These books should be consulted for information about original sources for this material — I have generally not repeated here the credits that are given there.

Our notation will change slightly to conform with the notation commonly used in this field. In this chapter, S will be a countable set, and the state space of the process will be $\{0,1\}^S$. It is given the product topology, and then becomes compact. Points in the state space will be denoted by η, ξ, ζ rather than x, y, z, and the particle systems will be denoted by η_t, ξ_t, ζ_t rather than $X(t), Y(t)$. This latter notation will be used for the evolution of individual particles.

Much of our interest will be in studying the stationary distributions for various classes of interacting particle systems. The situation here is quite different than it was for Markov chains in Chapter 2. In that context, one of the main issues was to determine whether or not a stationary distribution exists. Once one did exist, it was unique under a mild irreducibility assumption — see Corollary 2.67. Now stationary distributions always exist by Theorem 3.38. The issue becomes the determination of the structure of the class of all stationary distributions. This is the first step in analyzing the long-time behavior of the process.

Section 4.2 should be read before Sections 4.3 and 4.4, but these latter two are independent of each other. Sections 4.2 and 4.3 are more elementary than the later sections, so the reader who wants a more accessible introduction to the subject may choose to cover just these two sections. Section 4.5 is independent of the preceding sections, except for the construction of exclusion processes in Proposition 4.66 and Theorem 4.68, whose proofs are modeled after that of Theorem 4.3. If the reader wants to cover only this section, the construction can be taken for granted.

4.2. Spin systems

The property that distinguishes spin systems from other Feller processes on $\{0,1\}^S$ is that individual transitions involve one site only. In this section, $c(x, \eta)$ will denote a nonnegative, uniformly bounded (unless otherwise stated) function on $S \times \{0,1\}^S$ that is continuous in η for each $x \in S$. For $\eta \in \{0,1\}^S$ and $x \in S$, let $\eta_x \in \{0,1\}^S$ be the configuration obtained from η by "flipping" the coordinate $\eta(x)$ to $1-\eta(x)$, leaving the other coordinates unchanged. The interpretation of $c(x,\eta)$ is that it gives the rate at which η changes to η_x. If S were finite, the corresponding process η_t would be a finite state, continuous time Markov chain with Q-matrix given by

$$q(\eta,\eta_x) = c(x,\eta),$$

and therefore generator

$$\mathcal{L}f(\eta) = \sum_x c(x,\eta)[f(\eta_x) - f(\eta)] \tag{4.1}$$

— see (3.11). (Note a slight ambiguity in the notation: Subscripts x, y, z are used to indicate coordinate flips. Subscripts r, s, t will indicate time parameters. The meaning of the subscript should be clear from the context.)

For general countable S, some restriction on f is needed to make the series in (4.1) converge. To deal with this issue, let

$$D = \left\{ f \in C(\{0,1\}^S) : |||f||| := \sum_x \sup_\eta |f(\eta_x) - f(\eta)| < \infty \right\}.$$

Note that $f \in D$ is a type of Lipschitz condition on f. To see this, recall that a metric that metrizes the product topology on $\{0,1\}^S$ is given by

$$\rho(\eta,\zeta) = \sum_x \alpha(x)|\eta(x) - \zeta(x)|,$$

where α is strictly positive and summable. A continuous function f is in D if and only if there is an α (depending on f) so that f is Lipschitz continuous with respect to the corresponding metric. The relation between f and α is

$$\alpha(x) = \sup_\eta |f(\eta_x) - f(\eta)|$$

if this quantity is strictly positive. Otherwise, one can modify it by increasing it slightly. It is not natural to fix the α, since that would have the effect of favoring some x's over others. Using

$$\sum_x \sup_\eta |f(\eta_x) - f(\eta)| < \infty \tag{4.2}$$

in the definition of D has the advantage of treating all sites equally.

Exercise 4.1. Show that (4.2) alone does not imply the continuity of f.

Now we can define $\mathcal{L}$ on D by (4.1). Note that since the convergence of the series is uniform, the resulting $\mathcal{L}f$ is continuous. If $c(x,\cdot) \equiv 0$ for all but finitely many $x \in S$, $\mathcal{L}f$ is defined by (4.1) for any continuous f.

The voter model and contact process that are treated in the next two sections are important special classes of spin systems. Exclusion processes are not spin systems, since individual transitions involve two sites, rather than one.

4.2.1. Construction of spin systems. Our first objective is to find natural conditions on $c(x,\eta)$ that guarantee that the closure of $\mathcal{L}$ is a probability generator. Conditions (a), (b), and (d) of Definition 3.12 are easy to check and do not require further assumptions — see the next paragraph. The real challenge is (c).

For condition (a), use the Stone-Weierstrass theorem: D is an algebra of continuous functions on a compact set that separates points ($\eta \neq \zeta$ implies that there is an $x \in S$ such that $\eta(x) \neq \zeta(x)$; the function $f(\xi) = \xi(x)$ separates η and ζ) and contains the constant functions, so $\overline{D} = C(\{0,1\}^S)$. For (b), suppose $f \in D, \lambda \geq 0$, and $f - \lambda \mathcal{L} f = g$. Since $\{0,1\}^S$ is compact and f is continuous, there is an η at which f attains its minimum. But then $\mathcal{L}f(\eta) \geq 0$ since all the summands in (4.1) are nonnegative, so that

$$\min_\zeta f(\zeta) = f(\eta) \geq g(\eta) \geq \min_\zeta g(\zeta).$$

Part (d) follows from $1 \in D$ and $\mathcal{L}1 = 0$.

To check (c) of Definition 3.12, we need to derive an *a priori* bound on solutions of the equation $f - \lambda \mathcal{L} f = g$, and in order to state it, we need some notation:

$$\epsilon = \inf_{u,\eta}[c(u,\eta) + c(u,\eta_u)] \quad \text{and} \quad \gamma(x,u) = \sup_\eta |c(x,\eta_u) - c(x,\eta)|.$$

Note that $\gamma(x,u)$ is a measure of the extent to which the flip rate at site x depends on the configuration at site u. One might guess that if something is to go wrong, it will be because there is too much of this dependence.

Let $l_1(S)$ be the Banach space of functions $\alpha : S \to R^1$ that satisfy

$$||\alpha|| := \sum_x |\alpha(x)| < \infty.$$

The matrix γ defines an operator Γ on $l_1(S)$ by

$$\Gamma\alpha(u) = \sum_{x:x\neq u} \alpha(x)\gamma(x,u).$$

It is well-defined and bounded provided that

$$M := \sup_x \sum_{u:x\neq u} \gamma(x,u) < \infty,$$

and then $||\Gamma|| = M$.

For $f \in C(\{0,1\}^S)$ and $x \in S$, let

$$\Delta_f(x) = \sup_\eta |f(\eta_x) - f(\eta)|.$$

Then $|||f||| = ||\Delta_f||_{l_1(S)}$. Here is the *a priori* bound we need.

Proposition 4.2. *Suppose that either*

(a) $f \in D$,

or

(b) f *is continuous and*

$$c(x,\cdot) \equiv 0 \text{ for all but finitely many } x \in S. \tag{4.3}$$

If $f - \lambda\mathcal{L}f = g \in D$, $\lambda > 0$, *and* $\lambda M < 1 + \lambda\epsilon$, *then*

$$\Delta_f \le \big[(1+\lambda\epsilon)I - \lambda\Gamma\big]^{-1}\Delta_g, \tag{4.4}$$

where the inequality is interpreted coordinatewise, and the inverse is defined by the infinite series

$$\big[(1+\lambda\epsilon)I - \lambda\Gamma\big]^{-1}\alpha = \frac{1}{1+\lambda\epsilon}\sum_{k=0}^{\infty}\left(\frac{\lambda}{1+\lambda\epsilon}\right)^k \Gamma^k\alpha. \tag{4.5}$$

Proof. Note that the series in (4.5) converges for $\alpha \in l_1(S)$ by the assumption $\lambda M < 1 + \lambda\epsilon$. Writing $f - \lambda\mathcal{L}f = g$ at η and η_u and subtracting, noting that $(\eta_u)_u = \eta$, gives

$$\begin{aligned}[f(\eta_u) - f(\eta)]\big[1 + \lambda c(u,\eta) + \lambda c(u,\eta_u)] &= [g(\eta_u) - g(\eta)] \\ + \lambda \sum_{x:x\neq u}\Big\{c(x,\eta_u)[f((\eta_u)_x) - f(\eta_u)] &- c(x,\eta)[f(\eta_x) - f(\eta)]\Big\}.\end{aligned} \tag{4.6}$$

Since the values of $f(\eta_u) - f(\eta)$ as η varies for fixed u form a symmetric set, and this difference is a continuous function of η, for each u there exists an η so that

$$f(\eta_u) - f(\eta) = \sup_\zeta |f(\zeta_u) - f(\zeta)| = \Delta_f(u).$$

Therefore,

$$f(\zeta_u) - f(\zeta) \le f(\eta_u) - f(\eta)$$

for each ζ. Applying this to $\zeta = \eta_x$ and rearranging gives

$$f((\eta_u)_x) - f(\eta_u) = f((\eta_x)_u) - f(\eta_u) \le f(\eta_x) - f(\eta),$$

and then using this inequality in (4.6) leads to

$$
\begin{aligned}
\Delta_f(u)(1+\lambda\epsilon) &\le \Delta_f(u)\big[1+\lambda c(u,\eta)+\lambda c(u,\eta_u)\big] \\
&\le \Delta_g(u)+\lambda\sum_{x:x\ne u}\big[c(x,\eta_u)-c(x,\eta)\big]\big[f(\eta_x)-f(\eta)\big] \\
&\le \Delta_g(u)+\lambda\sum_{x:x\ne u}\gamma(x,u)\Delta_f(x).
\end{aligned}
\tag{4.7}
$$

If (4.3) holds, then there are only finitely many nonzero terms in the sum on the right, so under either assumption of the proposition, $f\in D$. Therefore, (4.7) can be written as

$$(1+\lambda\epsilon)\Delta_f \le \Delta_g+\lambda\Gamma\Delta_f.$$

Iterating this inequality and using the positivity of Γ leads to

$$\Delta_f \le \frac{1}{1+\lambda\epsilon}\sum_{k=0}^{n}\left(\frac{\lambda}{1+\lambda\epsilon}\right)^k\Gamma^k\Delta_g+\left(\frac{\lambda}{1+\lambda\epsilon}\right)^{n+1}\Gamma^{n+1}\Delta_f.$$

Now let $n\to\infty$ to get (4.4). □

Here is the main existence theorem for spin systems.

Theorem 4.3. *Suppose $M<\infty$. Then $\overline{\mathcal{L}}$ generates a probability semigroup $T(t)$. Furthermore,*

$$\Delta_{T(t)f}\le e^{-\epsilon t}e^{t\Gamma}\Delta_f. \tag{4.8}$$

In particular, if $f\in D$, then so is $T(t)f$ and

$$|||T(t)f|||\le e^{(M-\epsilon)t}|||f|||. \tag{4.9}$$

Proof. As observed earlier, properties (a), (b), and (d) in Definition 3.12 are easy to check for $\mathcal{L}$, and are inherited by $\overline{\mathcal{L}}$ in view of Proposition 3.30(a). For property (c), take S_n finite, increasing to S, and let

$$\mathcal{L}_nf(\eta)=\sum_{x\in S_n}c(x,\eta)[f(\eta_x)-f(\eta)],\quad f\in C(\{0,1\}^S). \tag{4.10}$$

This is the generator for the spin system in which the coordinates

$$(\eta_t(x),x\notin S_n)$$

are frozen. Since $\mathcal{L}_n$ is a bounded generator, it satisfies

$$\mathcal{R}(I-\lambda\mathcal{L}_n)=C(\{0,1\}^S)$$

for all $\lambda>0$ by Propositions 3.22 and 3.30(c). So, given $g\in D$, we can define $f_n\in C(\{0,1\}^S)$ by $f_n-\lambda\mathcal{L}_nf_n=g$. Since $\mathcal{L}_n$ satisfies (4.3), if λ is sufficiently small that $\lambda M<1+\lambda\epsilon$, then $f_n\in D$ by Proposition 4.2. Therefore, we can define

$$g_n=f_n-\lambda\mathcal{L}f_n\in\mathcal{R}(I-\lambda\mathcal{L}).$$

Letting $K = \sup_{x,\eta} c(x, \eta)$, we then have by Proposition 4.2

$$
\begin{aligned}
||g_n - g|| = \lambda||(\mathcal{L} - \mathcal{L}_n)f_n|| &\le \lambda K \sum_{x \notin S_n} \Delta_{f_n}(x) \\
&\le \lambda K \sum_{x \notin S_n} \big[(1 + \lambda\epsilon)I - \lambda\Gamma\big]^{-1} \Delta_g(x).
\end{aligned} \tag{4.11}
$$

Since $\Delta_g \in l_1(S)$, the right side of (4.11) tends to zero as $n \to \infty$, so $g_n \to g$. It follows that $g \in \overline{R(I - \lambda\mathcal{L})}$, so we conclude that $D \subset \overline{R(I - \lambda\mathcal{L})}$. Since D is dense in $C(\{0,1\}^S)$, we see that $R(I - \lambda\mathcal{L})$ is also dense. Therefore

$$R(I - \lambda\overline{\mathcal{L}}) = C(\{0,1\}^S)$$

by Proposition 3.30(d). This completes the verification that $\overline{\mathcal{L}}$ has properties (a)–(d) in Definition 3.12, and so it is a probability generator.

Turning to the second statement, write (4.4) as

$$\Delta_{(I-\lambda\mathcal{L})^{-1}g} \le \big[(1 + \lambda\epsilon)I - \lambda\Gamma\big]^{-1} \Delta_g,$$

and then iterate to get

$$\Delta_{(I-\frac{t}{n}\mathcal{L})^{-n}g} \le \left[\left(1 + \frac{t}{n}\epsilon\right)I - \frac{t}{n}\Gamma\right]^{-n} \Delta_g.$$

Passing to the limit, using Theorem 3.16(c), gives (4.8). □

4.2.2. Ergodicity of spin systems. One of the most important issues concerning spin systems is ergodicity. Here is the relevant definition. Recall that by Theorem 3.38, a Feller process on $\{0,1\}^S$ always has at least one stationary distribution. As in Chapter 3, $\mathcal{I}$ will denote the set of stationary distributions, and $\mathcal{I}_e$ its extreme points.

Definition 4.4. The spin system η_t with semigroup $T(t)$ is said to be ergodic, if its stationary distribution μ is unique, and

$$\lim_{t\to\infty} T(t)f(\eta) = \int f\, d\mu$$

for every η and $f \in C(\{0,1\}^S)$. In other words, it is ergodic if $\nu T(t) \Rightarrow \mu$ for any probability measure ν.

If there are no interactions among the coordinates, ergodicity is easy to check. With interactions, the issue is much more subtle. Here is the easy case.

Exercise 4.5. *Independent spin systems.* Let

$$c(x, \eta) = \begin{cases} \beta(x) & \text{if } \eta(x) = 0; \\ \delta(x) & \text{if } \eta(x) = 1, \end{cases}$$

Chapter 5

Stochastic Integration

5.1. Some motivation

Stochastic calculus has become an important tool for both theory and applications during the past half century. Many models in economics, engineering and other areas can be written informally as

$$dZ(t) = f(t, Z(t))dt + g(t, Z(t))\, dM(t), \tag{5.1}$$

where the first term on the right represents the infinitesimal drift of the process $Z(t)$, which is the process of interest, and the second term represents its variability. Since the first term accounts for the drift, it is reasonable to assume that the second term has no drift. So, it is natural to take $M(t)$ to be a martingale.

In order to make sense of (5.1), it is necessary to define integrals of the form

$$\int_0^t Y(s, \omega)\, dM(s), \tag{5.2}$$

where $M(t)$ is a martingale, and $Y(s)$ is a stochastic process with reasonable properties. If $M(t)$ were of bounded variation on finite t intervals, this could be defined as a Stieltjes integral. We saw in Theorem 1.36 that Brownian paths are nowhere differential, and hence not of bounded variation. And, of course, Brownian motion is a very important martingale, and the most natural choice as integrator in (5.2). This problem that arises from using Brownian motion as integrator is quite a bit more general, as the following result shows.

Proposition 5.1. *If $M(t)$ is a martingale with continuous paths that are of bounded variation on finite t intervals, then $M(t) = M(0)$ for each t a.s.*

Proof. Without loss of generality, take $M(0) = 0$. Let $V(t)$ be the total variation of $M(\cdot)$ on the interval $[0, t]$. Consider the stopping times

$$\tau_n = \inf\{t > 0 : V(t) \geq n\}.$$

Then $M_n(t) = M(t \wedge \tau_n)$ is a martingale by Theorem 1.93, and has continuous paths of total variation on $[0, \infty)$ bounded by n — since the paths are continuous. Since $\tau_n \to \infty$,

$$M(t) = \lim_{n\to\infty} M_n(t).$$

Therefore, it is enough to prove the proposition under the assumption that $V = \lim_{t\to\infty} V(t) \leq K$ for some constant K. By Exercise 1.89, if

$$\pi = \{0 = s_0 < s_1 < \cdots < s_m = t\}$$

is a partition of $[0, t]$, then

$$\begin{aligned} EM^2(t) &= E\sum_{j=1}^{m}(M^2(s_j) - M^2(s_{j-1})) = E\sum_{j=1}^{m}(M(s_j) - M(s_{j-1}))^2 \\ &\leq KE\max_j |M(s_j) - M(s_{j-1})|. \end{aligned}$$

As a consequence of path continuity, the right side tends to 0 as $|\pi|$ tends to zero. Therefore $M(t) = 0$ a.s. for each t, and using path continuity again, $P(M \equiv 0) = 1$. □

Remark 5.2. The continuity assumption that appears in Proposition 5.1 is important, since if $N(t)$ is a Poisson process with parameter 1, $N(t) - t$ is a martingale with paths that are of bounded variation on finite intervals.

This chapter is devoted to the construction of integrals such as the one in (5.2) and the development of (part of) their theory. For years, such integrals have been named for K. Itô, who defined and studied them in the 1940's. Recent information indicates that W. Doeblin made closely related discoveries at about the same time — see [**10**] for the interesting story.

As was the case in Chapter 4, this chapter is only an introduction to a very large and active field. There are a number of excellent books on the subject that can be consulted for a more comprehensive treatment. Among them are [**13**], [**19**], [**27**], [**40**], and [**43**].

Stochastic calculus can be developed with various degrees of generality. Throughout this chapter, we will assume for simplicity that $M(t)$ is a continuous martingale with respect to the right continuous filtration $\{\mathcal{F}_t\}$, and that $EM^2(t) < \infty$ for all t. Assumptions in the main results will often be stronger than necessary, in order to simplify the exposition. As in Chapter 1, $B(t)$ will always denote standard Brownian motion — it starts at the origin and has variance t at time t.

5.2. The Itô integral

Before defining the integral for general integrators, it is necessary to do some work that is not needed if $M(t)$ is Brownian motion. We do that next.

5.2.1. The variance process. Recall from Theorems 1.95 and 1.40 that standard Brownian motion $B(t)$ satisfies the following two properties:

(a) $B^2(t) - t$ is a martingale,

and

(b) $S^2(\pi_n, B) \to t$ in L_2 (or a.s. under further assumptions) as $|\pi_n| \to 0$, where π_n is a sequence of partitions of $[0, t]$ and

$$S^2(\pi, B) = \sum_{j=1}^{m} \big[B(s_j) - B(s_{j-1})\big]^2, \ \pi = \{0 = s_0 < s_1 < \cdots < s_m = t\}.$$

Our first objective is to extend these statements to martingales $M(t)$.

Theorem 5.3. *There exists a unique increasing, continuous process $A(t)$ such that $A(0) = 0$ and $M^2(t) - A(t)$ is a martingale.*

Remark 5.4. Of course, if $M(t)$ is Brownian motion, then the increasing process is just $A(t) = t$ by (a) above. Observation (b) above suggests how one might construct $A(t)$ more generally.

Proof of Theorem 5.3. For the uniqueness statement, suppose that both $A_1(t)$ and $A_2(t)$ have these properties. Then $A_1(t) - A_2(t)$ satisfies the assumptions of Proposition 5.1, so $A_1(t) - A_2(t) = A_1(0) - A_2(0) = 0$.

In the existence proof below, it will be convenient to be able to assume that $M(t)$ is uniformly bounded. So, we first reduce the problem to the bounded case. Note that we may assume that $M(0) = 0$, since $M(t) - M(0)$ and $[M(t) - M(0)]^2 - M^2(t)$ are both continuous martingales, so the same process $A(t)$ works for both $M(t)$ and $M(t) - M(0)$.

Suppose now that the theorem has been proved for uniformly bounded martingales. Given $M(t)$, let

$$\tau_n = n \wedge \inf\{t \geq 0 : |M(t)| \geq n\}.$$

These are stopping times that increase to ∞. Furthermore, by Theorem 1.93, $M_n(t) = M(t \wedge \tau_n)$ is a martingale for each n. It satisfies $|M_n(t)| \leq n$, so by the result for uniformly bounded martingales, there exist unique increasing, continuous processes $A_n(t)$ with $A_n(0) = 0$ so that $M_n^2(t) - A_n(t)$ is a martingale for each n. If $m < n$, then by Theorem 1.93,

$$M_n^2(t \wedge \tau_m) - A_n(t \wedge \tau_m)$$

is a martingale. Since $M_n(t\wedge\tau_m) = M_m(t)$, the uniqueness statement implies that $A_n(t\wedge\tau_m) = A_m(t)$, i.e., that $A_n(t) = A_m(t)$ for $t \le \tau_m$. So, we may define $A(t)$ by $A(t) = A_n(t)$ for $t < \tau_n$. Note that

$$E\big[A_n(t) - A_n(\tau_n), t > \tau_n\big] = E\big[M_n^2(t) - M_n^2(\tau_n), t > \tau_n\big] = 0,$$

so $A_n(t)$ is constant for $t \ge \tau_n$. Therefore, $A_n(t) \uparrow A(t)$. By Jensen's inequality, Proposition A.34,

$$M_n^2(t) = M^2(t\wedge\tau_n) \le E(M^2(t) \mid \mathcal{F}_{t\wedge\tau_n}),$$

so $\{M_n^2(t), n \ge 1\}$ is uniformly integrable for each t by Proposition A.39. Therefore, we can pass to the limit in the martingale property

$$E\big[M_n^2(t) - A_n(t) \mid \mathcal{F}_s\big] = M_n^2(s) - A_n(s), \quad s < t,$$

to conclude that $M^2(t) - A(t)$ is a martingale.

Assume now that $|M(t)| \le K$ for all t. To prove the theorem in this case, we are motivated by the above two observations about Brownian motion. In that case, the increasing process we are looking for is the quadratic variation up to time t. Given a partition

$$\pi = \{0 = s_0 < s_1 < s_2 < \cdots\}$$

of $[0,\infty)$ with $s_k \to \infty$, define a process $A_\pi(t)$ by

$$A_\pi(t) = \sum_{j=1}^{k} \big[M(s_j) - M(s_{j-1})\big]^2 + \big[M(t) - M(s_k)\big]^2$$

if $s_k \le t < s_{k+1}$. If $s < t$, choose $k \le l$ so that $s_k \le s < s_{k+1}$ and $s_l \le t < s_{l+1}$. Then

$$\begin{aligned} A_\pi(t) - A_\pi(s) = &\sum_{j=k+1}^{l} \big[M(s_j) - M(s_{j-1})\big]^2 \\ &+ \big[M(t) - M(s_l)\big]^2 - \big[M(s) - M(s_k)\big]^2. \end{aligned}$$

It follows that

$$E\big[A_\pi(t) - A_\pi(s) \mid \mathcal{F}_s\big] = E\big[M^2(t) - M^2(s) \mid \mathcal{F}_s\big]. \tag{5.3}$$

To see this, take for example the case $k < l$. It follows from Exercise 1.89 that

$$\begin{aligned} E\big[A_\pi(t) - A_\pi(s) \mid \mathcal{F}_s\big] = &\sum_{j=k+2}^{l} E\big[M^2(s_j) - M^2(s_{j-1}) \mid \mathcal{F}_s\big] \\ &+ E\big[[M(s_{k+1}) - M(s_k)]^2 \mid \mathcal{F}_s\big] \\ &+ E\big[M^2(t) - M^2(s_l) \mid \mathcal{F}_s\big] - \big[M(s) - M(s_k)\big]^2, \end{aligned}$$

and the sum above telescopes. As a consequence of (5.3),

$$M^2(t) - A_\pi(t) \text{ is a (continuous) martingale.} \tag{5.4}$$

However, A_π is not increasing, though it is when restricted to the points s_j.

In order to get an increasing process, we need to take a limit on π as the partition gets finer, and for that, the main issue is to bound

$$E\big[A_\pi(t) - A_{\pi'}(t)\big]^2,$$

where π and π' are two partitions that satisfy $t \in \pi \subset \pi'$. If

$$\pi' = \{0 = s_0 < s_1 < s_2 < \cdots\},$$

let $\xi_i = M(s_i) - M(s_{i-1})$ be the increments corresponding to π'. Then the increments corresponding to π are of the form

$$\xi_1 + \cdots + \xi_{k_1}, \xi_{k_1+1} + \cdots + \xi_{k_2}, \ldots,$$

where $0 < k_1 < k_2 < \cdots$. We will use the following easy consequence of the martingale property, recalling that $M(t)$ is uniformly bounded: If $1 \le i_1 \le i_2 \le \cdots \le i_{l-1} < i_l$, then

$$E\xi_{i_1} \cdots \xi_{i_l} = E(\xi_{i_1} \cdots \xi_{i_{l-1}} E(\xi_{i_l} \mid \mathcal{F}_{s_{i_l-1}})) = 0. \tag{5.5}$$

If m is defined by $s_{k_m} = t$, then

$$A_\pi(t) - A_{\pi'}(t) = \sum_{i=1}^{m} (\xi_{k_{i-1}+1} + \cdots + \xi_{k_i})^2 - \sum_{j=1}^{k_m} \xi_j^2 = 2\sum_{i=1}^{m} T_i,$$

where

$$T_l = \sum_{k_{l-1} < i < j \le k_l} \xi_i \xi_j.$$

By (5.5), $ET_iT_j = 0$ for $i \ne j$. Therefore,

$$E\big[A_\pi(t) - A_{\pi'}(t)\big]^2 = 4\sum_{l=1}^{m} ET_l^2. \tag{5.6}$$

Using (5.5) again,

$$ET_l^2 = \sum_{k_{l-1} < i, i' < j \le k_l} E\xi_i\xi_{i'}\xi_j^2 = \sum_{j=k_{l-1}+1}^{k_l} E\big[M(s_{j-1}) - M(s_{k_{l-1}})\big]^2 \xi_j^2.$$

Letting $\delta_{M,t}(\epsilon)$ be the modulus of continuity of the path $M(\cdot)$ on the interval $[0,t]$, this gives

$$\sum_{l=1}^{m} ET_l^2 \le E\delta_{M,t}^2(|\pi|) \sum_{j=1}^{k_m} \xi_j^2 \le \Big(E\delta_{M,t}^4(|\pi|)\Big)^{\frac{1}{2}} \bigg(E\Big(\sum_{j=1}^{k_m} \xi_j^2\Big)^2\bigg)^{\frac{1}{2}}. \tag{5.7}$$

Next, we need to bound the second term on the right of (5.7). To do so, use (5.5) twice more to write

$$
\begin{aligned}
E\Big(\sum_{j=1}^{k_m}\xi_j^2\Big)^2 &= \sum_{j=1}^{k_m}E\xi_j^4 + 2\sum_{i=1}^{k_m}E\xi_i^2\sum_{j=i+1}^{k_m}\xi_j^2 \\
&= \sum_{j=1}^{k_m}E\xi_j^4 + 2\sum_{i=1}^{k_m}E\xi_i^2\Big(\sum_{j=i+1}^{k_m}\xi_j\Big)^2 \\
&= \sum_{j=1}^{k_m}E\xi_j^2[M(s_j)-M(s_{j-1})]^2 \qquad (5.8)\\
&\quad + 2\sum_{i=1}^{k_m}E\xi_i^2[M(s_{k_m})-M(s_i)]^2 \\
&\le 12K^2E\sum_{j=1}^{k_m}\xi_j^2 = 12K^2E\Big(\sum_{j=1}^{k_m}\xi_j\Big)^2 \le 48K^4.
\end{aligned}
$$

Combining (5.6), (5.7), and (5.8) gives

$$
E\big[A_\pi(t)-A_{\pi'}(t)\big]^2 \le 28K^2\Big(E\delta^4_{M,t}(|\pi|)\Big)^{\frac{1}{2}}. \tag{5.9}
$$

Since $M(\cdot)$ is uniformly bounded and continuous, the right side above tends to 0 as $|\pi|\to 0$.

Now take π_n to be the partition with $s_k = k/2^n$, and t to be a positive integer. Since $A_{\pi_n}(s)-A_{\pi_m}(s)$ is a continuous martingale by (5.4), Doob's inequality, Theorem A.42, gives

$$
E\Big(\max_{0\le s\le t}[A_{\pi_n}(s)-A_{\pi_m}(s)]\Big)^2 \le 4E[A_{\pi_n}(t)-A_{\pi_m}(t)]^2. \tag{5.10}
$$

(Strictly speaking, Theorem A.42 applies to discrete time martingales. However, it does give (5.10) in which the max over all $0\le s\le t$ is replaced by the max over $s_1,\dots,s_k$, where $s_1,s_2,\dots$ is an ordering of the rationals in $[0,t]$. One can then pass to the limit on k and use path continuity to get the inequality as stated.) The right side of (5.10) tends to 0 as $m,n\to\infty$ by (5.9), so there is a subsequence n_k such that

$$
\sum_k E\Big[\max_{0\le s\le k}\big|A_{\pi_{n_{k+1}}}(s)-A_{\pi_{n_k}}(s)\big|\Big] < \infty,
$$

which implies that

$$
A(s) = \lim_{k\to\infty}A_{\pi_{n_k}}(s)
$$

exists a.s. and uniformly on bounded s sets. This subsequence has all the required properties. In particular, it is increasing because $A_\pi(s)$ is an increasing function of s on π. For future reference, we record the following fact, which comes from passing to the limit in (5.9):

$$E\big[A_\pi(t) - A(t)\big]^2 \le 28K^2\big(E\delta^4_{M,t}(|\pi|)\big)^{\frac{1}{2}}. \tag{5.11}$$ □

Remark 5.5. (a) The technique used to reduce the proofs of Proposition 5.1 and Theorem 5.3 to the uniformly bounded case is known as localization. We will often omit the details of such reductions in the future. Localization is also used in the general theory to remove the integrability assumptions on $M(t)$. Thus the general theory is usually framed in the context of local martingales — processes $X(t)$ with the property that there exist stopping times $\tau_n \uparrow \infty$ so that $X(t \wedge \tau_n)$ is a martingale for each n.

(b) Theorem 5.3 is a special case of a much more general result, known as the Doob-Meyer decomposition. It asserts under weak assumptions that any submartingale can be written uniquely as the sum of a martingale and an increasing "predictable" process — see page 74 of [**20**]. (Recall that $M^2(t)$ is a submartingale by Proposition 1.91.) This theorem is elementary in discrete time, but is quite deep in continuous time.

(c) The increasing process produced in Theorem 5.3 is called the quadratic variation of $M(t)$ or the variance process.

Corollary 5.6. *If τ is a stopping time, the variance process associated with the martingale $M(t \wedge \tau)$ is $A(t \wedge \tau)$.*

Proof. The process $A(t \wedge \tau)$ us clearly increasing and continuous. Since $M^2(t) - A(t)$ is a martingale, so is $M^2(t \wedge \tau) - A(t \wedge \tau)$. It follows that $A(t \wedge \tau)$ satisfies all the properties required of the (unique) variance process for $M(t \wedge \tau)$. □

Exercise 5.7. Show that if $M(t)$ is a bounded continuous martingale, then $M^2(t) - A(t)$ is a uniformly integrable martingale.

Exercise 5.8. Suppose that τ is a stopping time for which $A(\tau) = 0$. Show that $M(t) = M(0)$ for all $t \le \tau$.

5.2.2. Construction of the integral. For the time being, we will consistently use $A(t)$ to denote the increasing process constructed in Theorem 5.3 corresponding to $M(t)$. The starting point for the construction of the Itô integral is the following theorem, for which we need a definition.

Definition 5.9. (a) A predictable step function is a stochastic process of the following type:

$$Y(t,\omega) = Y_0(\omega)1_{\{0\}}(t) + \sum_{i=1}^{\infty} Y_i(\omega)1_{(t_i,t_{i+1}]}(t), \tag{5.12}$$

where $0 = t_1 < t_2 < \cdots$ with $t_k \uparrow \infty$, and $Y_i \in \mathcal{F}_{t_i}$ and $|Y_i(\omega)| \le K$ for each i and some constant K.

(b) For such a predictable step function, the Itô integral is

$$\int_0^t Y(s)\, dM(s) = \sum_{i=1}^{\infty} Y_i[M(t \wedge t_{i+1}) - M(t \wedge t_i)]. \tag{5.13}$$

Note that this is a finite sum.

Theorem 5.10. *If $Y(t)$ is a predictable step function, then*

$$M^*(t) = \int_0^t Y(s)\, dM(s)$$

is a continuous square integrable martingale. The variance process corresponding to $M^*(t)$ is

$$A^*(t) = \int_0^t Y^2(s)\, dA(s). \tag{5.14}$$

Finally,

$$E \sup_{0 \le s \le t} \big[M^*(s)\big]^2 \le 4EA^*(t). \tag{5.15}$$

Remark 5.11. (a) The integral in (5.14) is defined pathwise — no theory of stochastic integration is needed. For each ω, this is just a Riemann-Stieltjes integral. This will continue to be the case later when $Y(t)$ is not a step function. It will always have at most countably many discontinuities, so since $A(s)$ is continuous, the integral will be well-defined.

(b) One consequence of (5.14) is

$$E\left[\int_0^t Y(s)\, dM(s)\right]^2 = E\int_0^t Y^2(s)\, dA(s), \tag{5.16}$$

which says that two L_2 (semi) norms are equal. This isometry is the key to the extension of the Itô integral to more general integrands.

Proof of Theorem 5.10. The continuity and square integrability statements, as well as the fact that $M^*(t)$ is adapted, are clear from (5.13). For the martingale property, take $s < t$ and note that for each i,

$$M(t \wedge t_{i+1}) - M(t \wedge t_i) - M(s \wedge t_{i+1}) + M(s \wedge t_i)$$

$$= \begin{cases} M(t) - M(t_i) & \text{if } s < t_i < t < t_{i+1}; \\ M(t_{i+1}) - M(t_i) & \text{if } s < t_i < t_{i+1} \le t; \\ M(t) - M(s) & \text{if } t_i \le s < t < t_{i+1}; \\ M(t_{i+1}) - M(s) & \text{if } t_i \le s < t_{i+1} \le t; \\ 0 & \text{otherwise.} \end{cases} \tag{5.17}$$