CHAPTER 3: STOCHASTIC INTEGRALS, ITO'S LEMMA

Consider the present discounted value of a stream of returns over an infinite horizon,

$$\begin{aligned} v(x_0) &\equiv \int_0^\infty e^{-\rho t}\pi(x(t))dt \qquad (1)\\ \text{with}\quad \dot{x}(t) &= g\left(x(t)\right), \quad t \geq 0, \\ x(0) &= x_0, \end{aligned}$$

where $\rho > 0$ is a discount rate, $\pi(x)$ is a return function, $x(t)$ is a state variable that evolves according to the law of motion $g(x)$, and x_0 is the initial state. For example, $v(x_0)$ might be the value of a firm, where ρ is the interest rate, $x(t)$ describes the size of the market for the firm's product, $\pi(x)$ is the profit flow as a function of market size, and $g(x)$ describes the evolution of market size. Since the horizon is infinite, ρ is constant, and π and g are stationary (time-invariant), the function $v(x_0)$ representing the discounted value of the profit flow depends only on the initial state.

A standard and fairly simple argument, to be developed below, establishes that v satisfies the continuous-time Bellman equation

$$\rho v(x) = \pi(x) + v'(x)g(x). \qquad (2)$$

If v is the value of an asset, the interpretation of (2) is straightforward: the return on the asset, the term on the left, is the sum of the dividend and the capital gain, the terms on the right. A key step in deriving (2) from (1) involves using a first order

Taylor series approximation to the change in the value over a short interval of time Δt:

$$\begin{aligned} \Delta v &\equiv v(x(t+\Delta t)) - v(x(t)) \\ &\approx v'(x(t))\left[x(t+\Delta t) - x(t)\right] \\ &\approx v'(x(t))g(x(t))\Delta t. \end{aligned}$$

In many applications where value functions like v arise, it is natural to think of the exogenous state variable $x(t)$ as stochastic. Thus, instead of assuming that $x(t)$ is governed by a deterministic differential equation, as in (1), it is more accurately modeled as including a component that is a Brownian motion, geometric Brownian motion, or some other diffusion. Then to develop the analog of (2), the analog of a Taylor series approximation is needed for the case where the state variable is a diffusion. Ito's lemma is the key ingredient for this task.

Ito's lemma and related results are developed in this chapter. In section 1 the stochastic analog of (2) is derived using a heuristic approach. A brief discussion of stochastic integrals is provided in section 2, and in section 3 Ito's lemma is formally stated and used to obtain the stochastic analog of (2) more rigorously. In section 4 Ito's lemma is used to derive some results about geometric Brownian motion. Occupation measure and local time are defined in section 5 and used in section 6 to develop Tanaka's formula, an extension of Ito's lemma that applies to functions with kinks. Sections 7 and 8 develop the Kolmogorov backward and forward equations.

1. The Hamilton-Jacobi-Bellman equation

Fix a filtered probability space $(\Omega, \mathbb{F}, P)$, and let $X(t,\omega)$ be a Brownian motion with initial value $X(0) = x_0$ and parameters μ, σ^2. Then X can be written as

$$X(t) = X(0) + \mu t + \sigma W(t), \quad \text{all } t, \text{ all } \omega, \tag{3}$$

where W is a Wiener process, and where for notational simplicity the ω is suppressed. A shorthand notation for (3) is the differential form

$$dX(t) = \mu dt + \sigma dW(t), \quad \text{all } t, \text{ all } \omega. \tag{4}$$

Note that given the initial condition $X(0)$, (4) is simply an alternative way of writing (3).

More generally, suppose X is a diffusion with initial value $X(0) = x_0$ and infinitesimal parameters $\mu(t,x), \sigma(t,x)$. Then the differential form, the analog of (4) is

$$dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dW(t), \quad \text{all } t, \text{ all } \omega. \tag{5}$$

There is also an analog of (3), which will be developed later.

Let $F(t,x)$ be a function that is differentiable at least once in t and twice in x. The 'total differential' of $F(t, X(t,\omega))$, call it dF, can be approximated with a Taylor series expansion. Use $F_t \equiv \partial F/\partial t$, $F_x \equiv \partial F/\partial x$, etc. to denote the derivatives of F and substitute from (5) to find that

$$\begin{aligned} dF &= F_t\,dt + F_x\,dX + \frac{1}{2}F_{xx}\,(dX)^2 + ... \\ &= F_t\,dt + F_x\,[\mu\,dt + \sigma\,dW] \\ &\quad + \frac{1}{2}F_{xx}\left[\mu^2\,(dt)^2 + 2\mu\sigma\,dt\,dW + \sigma^2\,(dW)^2\right] + ..., \end{aligned}$$

where the dots indicate higher order terms, and where μ and σ are evaluated at $(t, X(t))$. Then rearrange terms and drop those of order higher than dt or $(dW)^2$ to obtain

$$dF = F_t dt + \mu F_x dt + \sigma F_x\,dW + \frac{1}{2}\sigma^2 F_{xx}\,(dW)^2\,. \tag{6}$$

Notice that dF is the sum of four components, two in dt, one in dW, and one in $(dW)^2$.

Since $\mathrm{E}[dW] = 0$ and $\mathrm{E}\left[(dW)^2\right] = dt$, taking expectations in (6) gives

$$\begin{aligned} \mathrm{E}\left[dF\right] &= \left[F_t + \mu F_x + \frac{1}{2}\sigma^2 F_{xx}\right] dt, \\ \mathrm{Var}\left[dF\right] &= \mathrm{E}\left[\, dF - \mathrm{E}\left[dF\right]\,\right]^2 \\ &= \sigma^2 F_x^2 dt. \end{aligned}$$

A case of particular interest is one where the drift and variance $\mu(x)$ and $\sigma(x)$ are stationary and F is the discounted value of a stationary function. That is, $F(t, x) = e^{-rt} f(x)$, where $r \geq 0$ is the discount rate. For this case

$$\mathrm{E}\left[d\ e^{-rt} f\right] = \left[-rf + \mu f' + \frac{1}{2}\sigma^2 f''\right] e^{-rt} dt, \tag{7}$$

where $\mu, \sigma^2, f,\ f'$ and f'' are evaluated at $X(t)$. For $r = 0$ this equation is simply

$$\mathrm{E}\left[df\right] = \left(\mu f' + \frac{1}{2}\sigma^2 f''\right) dt. \tag{8}$$

One application of (7) is in deriving the analog of the Bellman equation (2). Consider an infinite stream of returns as in (1), but where X is a diffusion with infinitesimal parameters $\mu(x)$ and $\sigma(x)$. Let π and ρ be as before, with π bounded and continuous and $\rho > 0$. Define $v(x_0)$ to be the expected discounted value of the stream of returns given the initial state $X(0) = x_0$:

$$v(x_0) \equiv \mathrm{E}\left[\int_0^\infty e^{-\rho t} \pi\left(X(t, \omega)\right) dt \mid X(0) = x_0\right], \quad \text{all } x_0. \tag{9}$$

The integral on the right, an integral over time for a fixed sample path, is an ordinary Riemann integral. To see this note that since $X(\cdot, \omega)$ is the sample path of a diffusion, it is a continuous function. Since by assumption π is continuous, the integral over the horizon $[0, T]$ exists for any finite T. Moreover, since π is bounded and $\rho > 0$, the limit exists as $T \to \infty$, so the integral over the infinite horizon is also well defined. The integral is then a bounded random variable defined on Ω, so the expected value is well defined.

For any small interval of time Δt, (9) has the Bellman-type property

$$v(x_0) \approx \pi\left(x_0\right)\Delta t + \frac{1}{1+\rho\Delta t}\mathrm{E}\left[v\left(X(0+\Delta t)\ \mid X(0)=x_0\right]\right. .$$

Multiply this equation by $(1+\rho\Delta t)$ and subtract $v(x_0)$ from each side to get

$$\rho v(x_0)\Delta t \approx \pi(x_0)\left(1+\rho\Delta t\right)\Delta t + \mathrm{E}\left[\Delta v \mid X(0)=x_0\right],$$

where

$$\Delta v \equiv v(X(\Delta t, \omega)) - v(x_0) \tag{10}$$

is the change in the value function over the interval Δt. Then divide by Δt and let $\Delta t \to 0$ to find that

$$\begin{aligned} \rho v(x_0) &= \lim_{\Delta t \to 0}\left\{\pi(x_0)\left(1+\rho\Delta t\right) + \frac{1}{\Delta t}\mathrm{E}\left[\Delta v \mid X(0)=x_0\right]\right\} \\ &= \pi(x_0) + \frac{1}{dt}\mathrm{E}\left[dv \mid X(0)=x_0\right]. \end{aligned}$$

The function v does not depend directly on time, so using (8) to evaluate $\mathrm{E}[dv]$ and dropping the subscript on the initial condition produces the **Hamilton-Jacobi-Bellman equation**

$$\rho v(x) = \pi(x) + \mu(x)v'(x) + \frac{1}{2}\sigma^2(x)v''(x), \quad \text{all } x. \tag{11}$$

This equation is the stochastic counterpart of the Bellman equation (2), and with $\mu(x)=g(x)$ and $\sigma^2(x)=0$ the two are identical.

In the next section the differentials dX and dF in (5) and (6) are described in more detail. As with dX in (4), each is a shorthand notation for a more precise expression.

2. Stochastic integrals

In this section the definition of the stochastic integral is first examined briefly. Doing so is useful because it shows why stochastic integrals have a certain fundamental property. Propositions 3.1 and 3.2 then characterize a class of integrable functions and state several properties of stochastic integrals.

All integrals are defined by first considering functions where the integral of interest is an easily calculated sum, and then extending the definition to a broader class of functions. For example, the Riemann integral is first defined for step functions. It is then extended to a broader class by approximating any other function of interest with a sequence of step functions, where the sequence is chosen so that the approximations become arbitrarily good. The integral of each function in the sequence is easily computed, and the integral of the function of interest is defined as the limit of the integrals of the approximating sequence. Similarly, integrals on measure spaces are defined first for simple functions, and the definition is then extended by approximating other measurable functions with sequences of simple functions.

The Ito integral is defined in a similar way. As with other integrals, the main issues are to show that the integral defined in this way exists and is unique, i.e., that there exists at least one approximating sequence, and that if there are many such sequences all have a common limit. Stated a little differently, the key step in defining an integral is to determine for which functions the approximation process leads to a uniquely defined value, i.e., to determine what class of functions is integrable.

Fix a filtered probability space $(\Omega, \mathbb{F}, P)$, let W be a Wiener process on this space, and let Y be a stochastic process adapted to it. That is, $Y : [0, \infty) \times \Omega \to \mathbf{R}$ is jointly measurable in (t, ω). The goal is to define a **stochastic integral** of Y with respect to W:

$$I_Y(t, \omega) = \int_0^t Y(s, \omega)\, dW(s, \omega), \quad \text{all } t > 0, \text{ all } \omega. \tag{12}$$

Notice that $I_Y(t, \omega)$ is an integral of Y along a particular sample path ω up to date t. Since it is a function of (t, ω), assuming the required joint measurability condition holds, I_Y is itself a stochastic process adapted to $(\Omega, \mathbb{F}, P)$.

Since Y is jointly measurable, if it is also suitably bounded there is no problem integrating it with respect to t and then taking the expected value. To this end assume that

$$\mathrm{E}\left[\int_0^t Y^2(s,\omega)\,ds\right] < \infty, \quad \text{all } t > 0. \tag{13}$$

Let H^2 be the set of all stochastic processes Y satisfying (13). Any Brownian motion satisfies (13).

The stochastic integral in (12) is first defined in terms of sums for simple (step-type) functions, and then extended to a broader class through approximating sequences. A stochastic process Y is called **simple** if there exists a countable sequence $\{t_k\}$ with $0 = t_0 < t_1 < ... < t_k \to \infty$ such that

$$Y(s,\omega) = Y(t_{k-1},\omega), \quad \text{all } s \in [t_{k-1}, t_k), \quad \text{all } \omega.$$

Note that the t_k's do *not* vary with ω. Let $S^2 \subset H^2$ be the set of all simple stochastic processes that satisfy (13).

It is easy to integrate functions in S^2 with respect to W over any finite time interval. Choose $Y \in S^2$, with steps at $\{t_k\}_{k=0}^{\infty}$. Fix $t > 0$ and $\omega \in \Omega$, and choose $n \geq 0$ such that $t_n < t \leq t_{n+1}$. Define the integral

$$\begin{aligned} \int_0^t Y(s,\omega)dW(s,\omega) \quad &\equiv \quad \sum_{k=0}^{n-1} Y(t_k,\omega)\left[W(t_{k+1},\omega) - W(t_k,\omega)\right] \\ &\quad\quad +Y(t_n,\omega)\left[W(t,\omega) - W(t_n,\omega)\right]. \end{aligned} \tag{14}$$

The terms on the right side of (14) have a very important feature: the value of the state Y at date t_k is multiplied by the increment to W that comes *after* that date, the increment between t_k and $t_{k+1.}$

As noted above, the extension to a broader class of functions involves the use of approximating sequences, and the argument has two parts. First, it must be shown that for any function in the broader class there exists at least one sequence in S^2 that approximates it. Second, it must be shown that if there are many approximating sequences, all converge to a common value. The integral is then defined to be that common value. Both of these arguments are rather complicated, and they are not especially useful in the applications here. Instead, several main results are simply stated.

The first part of Proposition 3.1 identifies a class of integrable functions, those in H^2. As noted above the integral $I_Y(t,\omega)$ is a stochastic process. The second part of the proposition concerns expected values of this process.

PROPOSITION 3.1: (a) If $Y \in H^2$, then Y is integrable, i.e., there exists a stochastic process $I_Y(t,\omega)$ satisfying (12).

(b) If Y is integrable, then the expected value of $I_Y(t,\omega)$ is zero at all dates:

$$\mathrm{E}\left[\int_0^t Y\left(s,\omega\right) dW(s,\omega)\right] = 0, \qquad \text{all } t \geq 0.$$

Part (b) of Proposition 3.1 asserts a remarkable property: *the expected value of a stochastic integral is identically zero.* Although this claim looks astonishing at first sight, the idea behind it is really very simple. Consider again the definition in (14) of the stochastic integral of a simple function. Each term in (14) involves $Y(t_k)$ multiplied by the increment to $W(t)$ over the *subsequent* interval, from t_k to t_{k+1}. The expected value of this increment at date t_k is zero, and the expected value of the integral is just the sum of these terms. Hence the zero expectation property holds for any simple function. But the stochastic integral of any other function is simply the limit of a sequence of integrals of simple functions, so the zero expectation property holds for all integrable functions.

The next proposition establishes two additional properties of stochastic integrals.

PROPOSITION 3.2: If $X, Y \in H^2$, and $a, b \in \mathbf{R}$, then

(a) the stochastic integral of the weighted sum $(aX + bY)$ is the weighted sum of the stochastic integrals of X and Y:

$$\int_0^t (aX + bY)\, dW = a \int_0^t X\, dW + b \int_0^t Y\, dW, \quad \text{all } \omega, \quad \text{all } t \geq 0;$$

(b) the expected value of the product of the stochastic integrals of X and Y is the expected value of the (Riemann) integral of XY:

$$\mathrm{E}\left[\int_0^t X\, dW \int_0^t Y\, dW\right] = \mathrm{E}\left[\int_0^t X\, Y\, ds\right], \qquad \text{all } t \geq 0.$$

The first part of Proposition 3.2 says that, like other types of integration, stochastic integration is a linear operator. The second part says that the expected value of the product of two stochastic integrals is equal to the expected value of the integral of their product. To see why this is so, consider approximating each of the integrals on the left with a finite sum. Since each of the increments dW over various subintervals has mean zero and all are mutually independent, all of the terms in the product have expected value zero except those that involve common time increments. For these, $\mathrm{E}\left[(dW)^2\right] = ds$, so the integral becomes an ordinary Riemann integral. Also note that since the integrals $\int_0^t X\, dW$ and $\int_0^t Y\, dW$ are random variables with means of zero, the expression in (b) is simply their covariance.

3. Ito's Lemma

Let W be a Wiener process on the filtered space $(\Omega, \mathbb{F}, P)$, let $\mu\,(t, x)$ and $\sigma(t, x) > 0$ be continuous functions, and let $X(0, \omega) = x_0(\omega)$ be a measurable function. The stochastic process X satisfying

$$X(t, \omega) = X(0, \omega) + \int_0^t \mu(s, X(s, \omega)) ds + \int_0^t \sigma(s, X(s, \omega))\, dW(s, \omega), \quad \text{all } t, \omega, \quad (15)$$

is a diffusion. Notice that the first integral in (15) is a Riemann integral, while the second is a stochastic integral of the type defined in the previous section. This equation is the integral form of the differential in (5), and a Brownian motion is simply the special case where μ and σ are constant.

The next result, Ito's lemma, is the basis for calculating values of a function $F(t,x)$ that has a diffusion as its second argument.

PROPOSITION 3.3 (ITO'S LEMMA): Let $F\colon \mathbf{R}_+ \times \mathbf{R} \to \mathbf{R}$ be once continuously differentiable in its first argument and twice continuously differentiable in its second, and let X be the diffusion in (15). Then

$$\begin{aligned} F\left(t, X(t,\omega)\right) &= F\left(0, X(0,\omega)\right) + \int_0^t F_t\left(s, X\right)\, ds \\ &\quad + \int_0^t F_x\left(s, X\right) \mu(s,X)\, ds \\ &\quad + \int_0^t F_x\left(s, X\right) \sigma(s,X)\, dW(s,\omega) \\ &\quad + \frac{1}{2}\int_0^t F_{xx}\left(s, X\right) \sigma^2(s,X)\, ds, \quad \text{all } t, \omega, \end{aligned} \tag{16}$$

where the arguments of $X(s,\omega)$ have been suppressed.

The right side of (16) contains four integrals. The third is a stochastic integral and the others are Riemann integrals. These four terms are the counterparts of the four terms in (6).

If μ and σ are stationary and $F(t,x) = e^{-rt}f(x)$, where $r \geq 0$ is a constant discount rate, then (16) takes the form

$$\begin{aligned} e^{-rt} f\left(X(t)\right) &= f\left(X(0)\right) - r\int_0^t e^{-rs} f\left(X\right)\, ds \\ &\quad + \int_0^t e^{-rs} f'(X)\mu(X)\, ds + \int_0^t e^{-rs} f'(X)\sigma(X)\, dW \\ &\quad + \frac{1}{2}\int_0^t e^{-rs} f''(X)\sigma^2(X)\, ds. \end{aligned} \tag{17}$$

The expected value of the differential form of this equation is in (7).

These results can be used to derive the Hamilton-Jacobi-Bellman equation (11) more rigorously. Consider the term Δv in (10) Use (17) with $r = 0$, $X(0) = x_0$, and constant functions for μ and σ to find that

$$\begin{aligned}\Delta v &\equiv v(X(\Delta t)) - v(x_0) \\ &= \mu \int_0^{\Delta t} v'(X(s))\, ds + \sigma \int_0^{\Delta t} v'(X(s))\, dW \\ &\quad + \frac{1}{2}\sigma^2 \int_0^{\Delta t} v''(X(s))\, ds.\end{aligned}$$

Then take the expected value, use the zero expectation property, and divide by Δt to obtain

$$\frac{1}{\Delta t}\mathrm{E}\left[\Delta v\right] = \frac{1}{\Delta t}\left[\mu \int_0^{\Delta t} \mathrm{E}\left[v'(X(s))\right]\, ds + \frac{1}{2}\sigma^2 \int_0^{\Delta t} \mathrm{E}\left[v''(X(s))\right]\, ds\right].$$

Finally, take the limit to get

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t}\mathrm{E}\left[\Delta v\right] = \mu v'(x_0) + \frac{1}{2}\sigma^2 v''(x_0),$$

which agrees with the earlier result.

The next section shows how Ito's lemma can be used to characterize the moments of a geometric Brownian motion.

4. Geometric Brownian motion

Recall from Chapter 1 that a **geometric Brownian motion** is a diffusion $X(t)$ with infinitesimal parameters $\mu(x) = \mu x$ and $\sigma(x) = \sigma x$. Thus, for a geometric Brownian motion

$$dX = \mu X dt + \sigma X dW, \tag{18}$$

so the *relative* increments dX/X are i.i.d. with fixed mean and variance. Several facts about this family can be proved using Ito's lemma.

Let X be as in (18), with $X(0) \equiv 1$, and consider the stochastic process $Y = \ln(X)$. Then (16) with $r = 0$ implies that

$$\begin{aligned}
Y(t,\omega) &= \ln(X(t,\omega)) \\
&= \ln(X(0)) + \int_0^t \left(\frac{1}{X}\mu X - \frac{1}{2}\frac{1}{X^2}\sigma^2 X^2 \right) ds + \int_0^t \frac{1}{X}\sigma X \, dW \\
&= 0 + \int_0^t \left(\mu - \frac{1}{2}\sigma^2 \right) ds + \int_0^t \sigma \, dW \\
&= \left(\mu - \frac{1}{2}\sigma^2 \right) t + \sigma W(t), \qquad \text{all } t, \omega.
\end{aligned}$$

Thus, $Y = \ln(X)$ is an ordinary Brownian motion with drift and variance

$$\hat{\mu} = \mu - \frac{1}{2}\sigma^2, \quad \text{and} \quad \hat{\sigma}^2 = \sigma^2.$$

To understand the downward adjustment in the drift, notice that since the logarithm is a concave function, Jensen's inequality implies that $\mathrm{E}[\ln(X(t))] < \ln(\mathrm{E}[X(t)])$. Moreover, since the variance of $X(t)$ increases linearly with t, the difference between the two increases over time. Hence it is the drift that must be adjusted.

Exercise 3.1: a. Let $Y(t)$ be a (μ, σ^2) Brownian motion. Use Ito's lemma to show that for any $\rho \neq 0$, $X(t) = \exp\{\rho Y(t)\}$ is a geometric Brownian motion with parameters $\left(\rho\mu + (\rho\sigma)^2/2, \ (\rho\sigma)^2\right)$.

b. Let $X(t)$ be a geometric Brownian motion with parameters (m, s^2). Show that for any $\lambda \neq 0$, $p(t) = X^\lambda(t)$ is a geometric Brownian motion with parameters $(\lambda m + s^2\lambda(\lambda - 1)/2, \ \lambda^2 s^2)$.

c. Verify that these two formulas agree.

The mean, variance, and higher moments of a geometric Brownian motion can be computed by using the fact that the expected values $\mathrm{E}\left[X^k(t)\right]$, $k = 1, 2, ...$, are deterministic functions of time that satisfy simple ordinary differential equations.

Fix the initial value $X(0) = x_0$ and assume $|\mu|, \sigma^2 < \infty$. Then $\mathrm{E}[X^2(t)]$ is bounded, for all t, so stochastic integrals involving X are well defined. Integrate (18) to get

$$X(t) = x_0 + \mu \int_0^t X(s)ds + \sigma \int_0^t X(s)dW, \quad \text{all } t.$$

Then take the expected value and use the zero expectation property in part (b) of Proposition 3.1 to find that

$$\mathrm{E}\left[X(t)\right] = x_0 + \mu \int_0^t \mathrm{E}\left[X(s)\right] ds, \quad \text{all } t. \tag{19}$$

Define $h(t) \equiv \mathrm{E}[X(t)]$, and write (19) as

$$h(t) = x_0 + \mu \int_0^t h(s)ds, \quad \text{all } t.$$

Then h satisfies the differential equation

$$h'(t) = \mu h(t), \quad \text{all } t,$$

with boundary condition $h(0) = y_0$. Hence

$$\mathrm{E}\left[X(t)\right] = h(t) = x_0 e^{\mu t}, \quad \text{all } t.$$

The argument also works for higher moments, with Ito's lemma providing a key step. For example, consider the second moment. Since $\mathrm{E}[X^4(t)]$ is bounded, for all t, stochastic integrals involving X^2 are well defined. Applying Ito's lemma to the function $f(x) = x^2$, one finds that (17) implies

$$X^2(t) = x_0^2 + \left(2\mu + \sigma^2\right) \int_0^t X^2(s)ds + 2\sigma \int_0^t X^2 dW, \quad \text{all } t, \omega.$$

Taking the expected value and using the zero expectation property then gives

$$\mathrm{E}\left[X^2(t)\right] = x_0^2 + \left(2\mu + \sigma^2\right) \int_0^t \mathrm{E}\left[X^2(s)\right] ds.$$

Define $h_2(t) = \mathrm{E}[X^2(t)]$. Then

$$h_2(t) = x_0^2 + \left(2\mu + \sigma^2\right) \int_0^t h_2(s)ds,$$

so h_2 satisfies the ordinary differential equation

$$h_2'(t) = \left(2\mu + \sigma^2\right) h_2(t), \quad \text{all } t,$$

with initial condition $h_2(0) = x_0^2$. Hence

$$\mathrm{E}\left[X^2(t)\right] = h_2(t) = x_0^2 e^{\left(2\mu+\sigma^2\right)t}, \quad \text{all } t.$$

The variance of $X(t)$ can be computed by combining the formulas above:

$$\begin{aligned} \mathrm{Var}\left[X(t)\right] &= \mathrm{E}\left[X^2(t)\right] - \mathrm{E}\left[X(t)\right]^2 \\ &= x_0^2 e^{2\mu t}\left(e^{\sigma^2 t} - 1\right), \quad \text{all } t. \end{aligned}$$

Note that if the drift is negative and sufficiently large relative to the variance, so $2\mu + \sigma^2 < 0$, then $\mathrm{E}[X^2(t)]$ decreases over time. In this case $\lim_{t\to\infty}\mathrm{E}[X^2(t)] = 0$ and $\lim_{t\to\infty}\mathrm{Var}[X(t)] = 0$. For large t, the (nonnegative) value $X(t)$ is, with high probability, very close to zero.

Higher moments of $X(t)$ can be computed in the same way.

Exercise 3.2: Calculate the skewness measure $\mathrm{E}[X(t) - \mathrm{E}\left[X(t)\right]]^3$.

5. Occupancy measure, local time

Let $X(s)$ be a Brownian motion on the filtered space $(\Omega, \mathbb{F}, P)$. The **occupancy measure** of the process X is the function m: $\mathfrak{B} \times [0,\infty) \times \Omega \to \mathbf{R}_+$ defined by

$$m(A,t,\omega) = \int_0^t 1_A\left(X\left(s,\omega\right)\right) ds, \quad \text{all } A \in \mathfrak{B},\ t \geq 0,\ \omega \in \Omega, \tag{20}$$

where 1_A is the indicator function for the set A, and where $\mathfrak{B}$ denotes the Borel sets. The value $m(A,t,\omega)$ is the total time the sample path $X(\cdot,\omega)$ has spent in the set A up to date t. Thus,

—for any fixed (t,ω), the mapping $m(\cdot,t,\omega)$: $\mathfrak{B} \to [0,1]$ is a measure (and hence the name) with total mass $m(\mathbf{R},t,\omega) = t$;

—for any fixed (A, t), the function $m(A, t, \cdot)$: $\Omega \rightarrow \mathbf{R}_+$ is a random variable; and

—for any fixed (A, ω), the function $m(A, \cdot, \omega)$: $\mathbf{R}_+ \rightarrow \mathbf{R}$ is continuous and nondecreasing, and is strictly increasing only when $X(t, \omega) \in A$.

The following result says that the function m defined above is absolutely continuous with respect to Lebesgue measure, with a continuous density. That is, it can be written as the integral of a continuous function.

THEOREM 3.4: There exists a function $\ell : \mathbf{R} \times [0, \infty) \times \Omega \rightarrow \mathbf{R}_+$ with the property that $\ell(x, t, \omega)$ is jointly continuous in (x, t) for almost every ω, and

$$m(A, t, \omega) = \int_A \ell(x, t, \omega) dx, \quad \text{all } A \in \mathfrak{B},\ t \geq 0,\ \omega \in \Omega. \tag{21}$$

See Chung and Williams (1990) for a proof.

The process $\ell(x, \cdot, \cdot)$ is called the **local time** of X at level x. For fixed x, $\ell(x, \cdot, \cdot)$ is a continuous stochastic process with the property that it is positive if and only if $X(t, \omega) = x$. Thus, the local time $\ell(x, t, \omega)$ is a measure of the time that the process has spent at state x. Note that

$$\ell(x, t, \omega) = \lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} m\left((x - \varepsilon, x + \varepsilon), t, \omega\right). \tag{22}$$

Theorem 3.4 suggests that ℓ can play the role of a density function. The following theorem shows that this conjecture is correct, leading to a useful fact about integrals along sample paths.

THEOREM 3.5: Let $f : \mathbf{R} \rightarrow \mathbf{R}$ be a bounded, measurable function. Then

$$\int_0^t f(X(s, \omega)) ds = \int_{\mathbf{R}} f(x)\, \ell(x, t, \omega)\, dx, \quad \text{all } t \geq 0; \text{ a.e. } \omega \in \Omega. \tag{23}$$

Theorem 3.5 says that an integral over time can be replaced with an integral over states, weighting outcomes by their local time ℓ. In this sense ℓ plays exactly the role of a density function.

For many economic applications these definitions and results must be extended in two ways, to allow final dates that are stopping times rather than fixed dates, and to incorporate discounting. Both extensions are straightforward. For the former, it suffices to note that since (20) - (23) hold for all t and all ω, they also hold if t is replaced by a stopping time $\tau(\omega)$. To incorporate discounting, let $r > 0$ be an interest rate. At the risk of being tedious the definitions and results so far will be restated.

Define the **discounted occupancy measure** of the process X, call it $\hat{m}\left(\cdot;r\right)$: $\mathfrak{B}\times \mathbf{R}_+ \times \Omega \to \mathbf{R}_+$ by

$$\hat{m}(A,t,\omega;r) \equiv \int_0^t e^{-rs} 1_A\left(X\left(s,\omega\right)\right)ds, \quad \text{all } A \in \mathfrak{B},\ t \geq 0,\ \omega \in \Omega. \tag{24}$$

The value $\hat{m}(A,t,\omega;r)$ is the total *discounted* time, discounted at the rate r, that the sample path $X\left(\cdot,\omega\right)$ has spent in the set A up to date t. Note that $\hat{m}$ has the same three properties as m, except that it has total mass $\hat{m}(\mathbf{R},t,\omega;r) = \left[1 - e^{-rt}\right]/r$. Note that $\hat{m}\left(\cdot;r\right)$ is continuous at $r = 0$: $\lim_{r\to 0}\hat{m}(A,t,\omega;r) = m(A,t,\omega)$.

Like m, the function $\hat{m}$ can be written as the integral of a continuous function.

Theorem 3.6: There exists a function $\hat{\ell}\left(\cdot;r\right) : \mathbf{R} \times [0,\infty) \times \Omega \to \mathbf{R}_+$ with the property that $\hat{\ell}(x,t,\omega;r)$ is jointly continuous in (x,t) for almost every ω, and

$$\hat{m}(A,t,\omega;r) = \int_A \hat{\ell}(x,t,\omega;r)dx, \quad \text{all } A \in \mathfrak{B},\ t \geq 0,\ \omega \in \Omega. \tag{25}$$

The stochastic process $\hat{\ell}(x,\ \cdot\ ,\ \cdot\ ;r)$ will be called the **discounted local time** of X at level x. As before, it follows from (24) and (25) that

$$\hat{\ell}(x,t,\omega;r) = \lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}\hat{m}\left(\left(x-\varepsilon,x+\varepsilon\right),t,\omega;r\right). \tag{26}$$

Like ℓ, $\hat{\ell}$ plays the role of a density function, giving the following analog of Theorem 3.5.

Theorem 3.7: Let $f : \mathbf{R} \to \mathbf{R}$ be a bounded, measurable function. Then

$$\int_0^t e^{-rs} f(X(s,\omega))ds = \int_{\mathbf{R}} f(x)\,\hat{\ell}(x,t,\omega;r)\,dx, \quad \text{all } t \geq 0;\ \text{a.e. } \omega \in \Omega. \tag{27}$$

Thus, an integral of discounted values over time can be written as an integral over states, weighting outcomes by their discounted local time $\hat{\ell}$. In Chapter 5 explicit formulas for ℓ and $\hat{\ell}$ will be developed for the case where X is a Brownian motion.

6. Tanaka's formula

In this section Tanaka's formula is derived, an extension of Ito's lemma that applies to functions with kinks, i.e., with discontinuous first derivatives.

Let $X(t)$ be a (μ, σ^2) Brownian motion, and recall that for any twice continuously differentiable function f, Ito's lemma says

$$\begin{aligned} f(X(t)) &= f(X(0)) + \mu \int_0^t f'(X)\, ds \\ &\quad + \sigma \int_0^t f'(X)\, dW(s) + \frac{1}{2}\sigma^2 \int_0^t f''(X)\, ds. \end{aligned} \tag{28}$$

Consider the function

$$f(x) = \max\{0, cx\}, \quad c > 0. \tag{29}$$

This function is continuous, but it has a kink at $x = 0$, so

$$f'(x) = \begin{cases} 0, & x < 0, \\ \text{undefined}, & x = 0, \\ c, & x > 0, \end{cases}$$

and

$$f''(x) = \begin{cases} 0, & x < 0, \\ \text{undefined}, & x = 0, \\ 0, & x > 0. \end{cases}$$

Calculating the first and second integrals on the right side in (28) poses no problem: f' is discontinuous at $x = 0$, but a function with a finite number of jumps can be integrated in the usual way. The problem is the last term: f'' has an 'impulse' at

$x = 0$. This suggests that it may be helpful to think of f' as analogous to the CDF for a (signed) measure. Tanaka's formula develops this observation more rigorously.

To begin, note that the function in (29) can be approximated arbitrarily closely with functions having continuous first derivatives. For example, for any $\varepsilon > 0$, define the function

$$f_\varepsilon(x) = \begin{cases} 0, & x < -\varepsilon, \\ c\left(x+\varepsilon\right)^2/4\varepsilon, & -\varepsilon \le x \le +\varepsilon, \\ cx, & x > +\varepsilon. \end{cases}$$

This function is continuously differentiable, even at $x = \pm\varepsilon$:

$$f'_\varepsilon(x) = \begin{cases} 0, & x < -\varepsilon, \\ c\left(x+\varepsilon\right)/2\varepsilon, & -\varepsilon \le x \le +\varepsilon, \\ c, & x > +\varepsilon. \end{cases}$$

Hence $f''_\varepsilon(x)$ is also well defined and continuous, except at the points $x = \pm\varepsilon$:

$$f''_\varepsilon(x) = \begin{cases} 0, & x < -\varepsilon, \\ c/2\varepsilon, & -\varepsilon < x < +\varepsilon, \\ 0, & x > +\varepsilon. \end{cases}$$

For computing the last integral term in (28), it does not matter that f''_ε is discontinuous at these two points. Figure 3.1 displays f_ε and its first two derivatives, for two values of ε.

The idea is to approximate the last term in (28) using f''_ε and then to take the limit as $\varepsilon \to 0$. The approximating functions f''_ε take only two values, zero and $c/2\varepsilon$. Therefore, using the definition of the occupancy measure in (20) one finds that

$$\int_0^t f''_\varepsilon(X(s))ds = \frac{c}{2\varepsilon} m\left(\left(-\varepsilon, +\varepsilon\right), t, \omega\right).$$

Hence

$$\begin{aligned} \lim_{\varepsilon\to 0} \int_0^t f''_\varepsilon(X(s,\omega))ds &= \lim_{\varepsilon\to 0} \frac{c}{2\varepsilon} m\left(\left(-\varepsilon, +\varepsilon\right), t, \omega\right) \\ &= c\,\ell\left(0, t, \omega\right), \end{aligned}$$

where the second line uses the expression for local time in (22). Hence for the function in (29), the analog of (28) is

$$f(X(t)) = f(X(0)) + \mu \int_0^t f'(X)ds + \sigma \int_0^t f'(X)\,dW(s) + \frac{\sigma^2}{2} c\ell\,(0,t)\,, \quad \text{all } \omega.$$

More generally, the following result holds.

Theorem 3.8 (Tanaka's formula): Let X be a (μ, σ^2) Brownian motion. Let f be a continuous function, with a derivative f' that is well defined and continuous except at a finite number of points. Define the signed measure ν on $(\mathbf{R}, \mathfrak{B})$ by $\nu\left((a,b]\right) = f'(b) - f'(a)$, for $-\infty < a < b < +\infty$. Then

$$f(X(t)) = f(X(0)) + \mu \int_0^t f'(X)\,ds + \sigma \int_0^t f'(X)\,dW(s) + \frac{\sigma^2}{2} \int_{\mathbf{R}} \ell\,(x,t)\ \nu(dx).$$

See Harrison (1983) or Chung and Williams (1990) for more general versions of this result. In the example above the signed measure takes the simple form

$$\nu(A) = \begin{cases} c, & \text{if } 0 \in A, \\ 0, & \text{if } 0 \notin A. \end{cases}$$

Notice that if f is twice continuously differentiable, Tanaka's formula returns the expression in Ito's lemma:

$$\begin{aligned} \int_{\mathbf{R}} \ell\,(x,t)\ \nu(dx) &= \int_{\mathbf{R}} \ell\,(x,t)\ f''(x)\,dx \\ &= \int_0^t f''(X(s))\,ds, \end{aligned}$$

where the first line uses the definition of ν and the second uses Theorem 3.4.

7. The Kolmogorov backward equation

The Kolmogorov backward equation (KBE) is a second-order partial differential equation (PDE) satisfied by the densities at dates $t > 0$ generated by diffusions with different initial values x at date 0. The distribution functions corresponding to these densities also satisfy the KBE. The argument is as follows.

Let $\{X(t), t \geq 0\}$ be a regular diffusion on the open interval (ℓ, r), with infinitesimal parameters $\mu(x), \sigma^2(x)$. Fix any bounded, piecewise continuous function g on (ℓ, r), and define

$$u(t,x) \equiv \mathrm{E}\left[g(X(t)) \mid X(0)=x\right],$$

the expectation of $g(X)$ at date t, conditional on the initial value x at date 0. The first step is to show that u satisfies a certain PDE.

Fix any $t > 0$. By the law of iterated expectations, for any $h > 0$

$$\begin{aligned} u(t+h,x) &= \mathrm{E}\left[g(X(t+h)) \mid X(0)=x\right] \\ &= \mathrm{E}\left\{\mathrm{E}\left[g(X(t+h)) \mid X(h)\right] \mid X(0)=x\right\} \\ &= \mathrm{E}\left[u(t,X(h)) \mid X(0)=x\right], \end{aligned}$$

where the second line inserts an inner expectation conditioned on information at date $h > 0$, and the third rewrites the second in term of u. It follows that for any $h > 0$,

$$\frac{1}{h}\left[u(t+h,x)-u(t,x)\right] = \frac{1}{h}\mathrm{E}\left[u(t,X(h))-u(t,x) \mid X(0)=x\right]. \tag{30}$$

Taking the limit as $h \to 0$ and using Ito's lemma on the right leads to the PDE

$$\frac{\partial u(t,x)}{\partial t} = \mu(x)\frac{\partial u(t,x)}{\partial x} + \frac{1}{2}\sigma^2(x)\frac{\partial^2 u(t,x)}{\partial x^2}. \tag{31}$$

The initial condition for this PDE is $u(0,x) = g(x)$, all x.

For the indicator function $g = 1_{(\ell,y]}$ this construction leads to $u(t,x) = P(t,x,y)$, where

$$P(t,x,y) \equiv \Pr\left[X(t) \leq y \mid X(0)=x\right]$$

is the probability that the process is below y at date t, given the initial condition x at date 0. In this case (31) implies

$$\frac{\partial P(t,x,y)}{\partial t} = \mu(x)\frac{\partial P(t,x,y)}{\partial x} + \frac{1}{2}\sigma^2(x)\frac{\partial^2 P(t,x,y)}{\partial x^2}, \quad t>0,\ x \in (\ell, r). \tag{32}$$

Eq. (32) is called the **Kolmogorov backward equation.** The boundary condition for P is

$$P(0,x,y)=\begin{cases}1, & \text{if } y\leq x,\\ 0, & \text{if } y>x.\end{cases}$$

Since X is a regular diffusion, for each $t>0$ and $x\in(\ell,r)$, P has a density: $\partial P(t,x,y)/\partial y=p(t,x,y)$. Consequently (32) can be differentiated with respect to y to get

$$\frac{\partial p(t,x,y)}{\partial t}=\mu(x)\frac{\partial p(t,x,y)}{\partial x}+\frac{1}{2}\sigma^2(x)\frac{\partial^2 p(t,x,y)}{\partial x^2},\quad t>0,\ x\in(\ell,r)\,. \tag{33}$$

That is, the transition density also satisfies the Kolmogorov backward equation. In this case the boundary condition is different, however: as $t\downarrow 0$, the density function $p(t,x,y)$ collapses to a mass point at $y=x$.

Notice that (32) and (33) do not involve y, which enters only through the limiting conditions. The KBE itself involves only the date t and the initial condition x. It describes, for fixed y, how the density at y, the value $p(t,x,y)$ varies with (t,x).

The PDE in (32) and (33) has many solutions. Indeed, there are many solutions that are c.d.f's and densities. Thus, a boundary condition is needed to identify the one of interest in any particular context.

To illustrate more concretely what (32) and (33) imply, it is useful to look at specific examples: a Brownian motion, a geometric Brownian motion, and an Ornstein-Uhlenbeck process.

Recall that a Normal distribution with mean and variance (m,v) has density

$$\phi(v,m;y)=\frac{1}{\sqrt{2\pi v}}\exp\left\{\frac{-\left(y-m\right)^2}{2v}\right\},\qquad y\in(-\infty+\infty)\,. \tag{34}$$

Example 1: For a Brownian motion the state space is $\mathbf{R}$ and the infinitesimal parameters are $\mu(x)=\mu$, $\sigma(x)=\sigma$. Hence (33) takes the form

$$\frac{\partial p}{\partial t}=\mu\frac{\partial p}{\partial x}+\frac{1}{2}\sigma^2\frac{\partial^2 p}{\partial x^2},\quad t>0,\ x\in\mathbf{R}. \tag{35}$$

A normal distribution with parameters $(m, v) = (x + \mu t, \sigma^2 t)$ has density $\phi(\sigma^2 t, x + \mu t; y)$, where ϕ is as in(34). Thus, the claim is that the function

$$p(t, x, y) = \phi(\sigma^2 t,\ x + \mu t,\ y)$$

satisfies (35). Let $\phi_v = \partial\phi/\partial v$, $\phi_m = \partial\phi/\partial m$, etc. denote the partial derivatives of ϕ. Then it follows from the chain rule for differentiation that p satisfies (35) if

$$\sigma^2 \phi_v + \mu\phi_m = \mu\phi_m + \frac{1}{2}\sigma^2\phi_{mm},$$

or

$$\phi_v = \frac{1}{2}\phi_{mm}, \quad \text{all } v, m, y.$$

Using (34) it is straightforward to show that this condition holds.

Example 2: For a geometric Brownian motion the state space is $\mathbf{R}_+$ and the infinitesimal parameters are $\mu(x) = \mu x$, $\sigma(x) = \sigma x$. Hence (33) takes the form

$$\frac{\partial p}{\partial t} = \mu x \frac{\partial p}{\partial x} + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 p}{\partial x^2}, \quad t > 0,\ x \in \mathbf{R}_+. \tag{36}$$

It is straightforward to verify that

$$p(t, x, y) = \phi\left[\sigma^2 t,\ \ln x + \left(\mu - \frac{1}{2}\sigma^2\right) t\ , \ln y\right],$$

satisfies (36).

Example 3: For an Ornstein-Uhlenbeck (OU) process the state space is $\mathbf{R}$ and the infinitesimal parameters are $\mu(x) = -\alpha x$, $\sigma(x) = \sigma$. Hence (33) takes the form

$$\frac{\partial p}{\partial t} = -\alpha x \frac{\partial p}{\partial x} + \frac{1}{2}\sigma^2 \frac{\partial^2 p}{\partial x^2}, \qquad t > 0, \quad x \in \mathbf{R}_+. \tag{37}$$

An OU process has increments that are Gaussian but not independent, and it is straightforward to verify that the transition density

$$p(t, x, y) = \phi\left(\frac{\sigma^2}{2\alpha}\left(1 - e^{-2\alpha t}\right),\ x e^{-\alpha t}, y\right),$$

satisfies (37).

8. The Kolmogorov forward equation

The backward equation involves time t and the initial condition x, with the current state y held fixed. A similar PDE, the Kolmogorov forward equation (KFE), involves t and y, with the initial state x fixed. The forward equation is useful for characterizing the limiting distribution, if one exists. It is worth emphasizing that while the backward equation holds for all regular diffusions, the forward equation does not. For example, it may not hold if the state space is bounded and probability accumulates at the boundaries. It does hold for Brownian motions (both ordinary and geometric) and for Ornstein-Uhlenbeck processes.

To derive the forward equation, start by considering any smooth function ϕ satisfying

$$\phi(t+s,y)=\int \phi(t,\xi)p(s,\xi,y)d\xi, \quad \text{all } t,s,y. \tag{38}$$

For example, the density $p\left(t,x,\xi\right)$ satisfies (38), for any initial value x. The stationary density $\psi(\xi)$ also satisfies (38), if one exists.

For diffusions that are well behaved in the sense noted above, it can be shown that for any function ϕ satisfying (38),

$$\frac{\partial \phi(t,y)}{\partial t}=-\frac{\partial}{\partial y}\left[\mu(y)\phi(t,y)\right]+\frac{1}{2}\frac{\partial^2}{\partial y^2}\left[\sigma^2(y)\phi(t,y)\right]. \tag{39}$$

For the density $p\left(t,x,\xi\right)$, (39) takes the form

$$\frac{\partial p(t,x,y)}{\partial t}=-\frac{\partial}{\partial y}\left[\mu(y)p(t,x,y)\right]+\frac{1}{2}\frac{\partial^2}{\partial y^2}\left[\sigma^2(y)p(t,x,y)\right], \tag{40}$$

where x is fixed. This equation is called the Kolmogorov forward equation.

Perhaps the most important use of (39) is to characterize the stationary density $\psi(y)$, if one exists. If it does exist, (39) implies that it satisfies

$$0=-\frac{d}{dy}\left[\mu(y)\psi(y)\right]+\frac{1}{2}\frac{d^2}{dy^2}\left[\sigma^2(y)\psi(y)\right]. \tag{41}$$

Exercise 3.3: Use (41) to show that a Brownian motion does not have a stationary density.

If there is a stationary density, it can be found as follows. Integrate (41) once to get

$$c_1 = \frac{d}{dy}\left[\sigma^2(y)\psi(y)\right] - 2\mu(y)\psi(y),$$

where c_1 is a constant that must be determined. Then use the integrating factor

$$s(y) = \exp\left\{-\int^y \frac{2\mu(\xi)}{\sigma^2(\xi)} d\xi\right\}$$

to write this equation as

$$\frac{d}{dy}\left[s(y)\sigma^2(y)\psi(y)\right] = c_1 s(y),$$

and integrate again to get

$$\psi(x) = \frac{1}{s(x)\sigma^2(x)}\left[c_1 \int^x s(y)dy + c_2\right].$$

The constant c_1, c_2 must be chosen so that $\psi(x) \geq 0$, all x, and

$$\int \psi(x)dx = 1.$$

Example 3: For an Ornstein-Uhlenbeck process $\sigma(x) = \sigma$ and $\mu(x) = -\alpha x$, so the integrating factor is

$$\begin{aligned} s(y) &= \exp\left\{\frac{\alpha}{\sigma^2}\int^y 2\xi d\xi\right\} \\ &= \exp\left\{\gamma y^2\right\}, \end{aligned}$$

where $\gamma \equiv \alpha/\sigma^2$. Hence

$$\psi(x) = \hat{c}_1 \int_0^x e^{\gamma\left(y^2 - x^2\right)} dy + \hat{c}_2 e^{-\gamma x^2},$$

where the variance has been absorbed into the constants. If $\hat{c}_1 \neq 0$, the first term diverges as $|x|$ gets large, and $\psi(x)$ cannot be a density. Hence $\hat{c}_1 = 0$, and the stationary density has the form $\psi(x) = \hat{c}_2 e^{-\gamma x^2}$, where $\hat{c}_2$ is chosen so the $\int \psi = 1$. Thus, $\psi(x) = \phi(1/2\gamma, 0; x)$, and the limiting distribution is a normal with mean $m = 0$ and variance $v = 1/2\gamma$.

9. Notes

Fleming and Rishel (1975), Harrison (1985), Fleming and Soner (1993) and Krylov (1995) all provide good treatments of stochastic integration. Karatzas and Shreve (1991), which is very complete, is useful as a reference. The discussion of geometric Brownian motion in section 4 follows Karlin and Taylor (1975, section 7.4). See Chung and Williams (1990, Chapter 7) for a rigorous development of occupancy measure and local time. The discussion of Tanaka's formula in section 6 follows Harrison (1985, sect. 4.6). There are many good treatments of the Kolmogorov backward and forward equations. The discussion in sections 7 and 8 is based on Karlin and Taylor (1981, Sect. 15.5).