show that the probability of extinction is

$$p_{ext} = \frac{1 - \sqrt{1 - 4p(1-p)}}{2p}.$$

8. Use Fig. 10.4b to show graphically that $x_n \to x^*$ when $\mu > 1$.
9. A branching process has initially one individual. Use the law of total probability in the form

$$\Pr(\text{extinction}) = \sum_k \Pr(\text{extinction} \mid k \text{ descendants}) \Pr(k \text{ descendants})$$

to deduce that the extinction probability x is a solution of $x = P(x)$.

10. Let $\{X_n, n = 0, 1, 2, \ldots\}$ be a branching process with $X_0 = 1$ and with the number of offspring per individual $0, 1, 2$ with probabilities p, q, r, respectively, where $p + q + r = 1$ and $p, q, r > 0$. Show that if $q + 2r > 1$, the probability of extinction is

$$x^* = \frac{1 - q - \sqrt{(1-q)^2 - 4pr}}{2r}.$$

11. Assume, very roughly speaking, that a human population is a branching process. What is the probability of extinction if the proportion of families having 0, 1 or 2 children are 0.2, 0.4 and 0.4 respectively?

11

Stochastic processes and an introduction to stochastic differential equations

11.1 DETERMINISTIC AND STOCHASTIC DIFFERENTIAL EQUATIONS

A **differential equation** usually expresses a relation between a function and its derivatives. For example, if $t \geqslant t_0$ represents time, and the rate of growth of a quantity $y(t)$ is proportional to the amount $y(t)$ already present, then we have

$$\frac{dy}{dt} = ky, \tag{11.1}$$

where k is a constant of proportionality. Equation (11.1) is called a **first-order differential equation** because the highest order derivative appearing is the first derivative. It is also called **linear** because both y and its derivative occur raised to power 1.

Equation (11.1) may be viewed as a prescription or **mathematical model** for finding y at all times subsequent to (or before) a given time t_0 at which the value y_0 of y is known. This is expressed in the solution of (11.1),

$$y(t) = y_0 e^{k(t-t_0)}, \tag{11.2}$$

which has the same form as the Malthusian population growth law of Section 9.1. It is also a formula for finding an asset value with compound interest when the initial value is y_0.

In the natural sciences (biology, chemistry, physics, etc.), differential equations have provided a concise method of summarizing physical principles. An important example of a **nonlinear first-order differential equation** is **Verhulst's logistic equation**:

$$\frac{dy}{dt} = ry\left(1 - \frac{y}{y^*}\right), \tag{11.3}$$

with $r \geqslant 0$. This equation is frequently used to model the growth of populations of organisms. The quantity y^* is called the **carrying capacity** whereas r is called the **intrinsic growth rate**. It will be seen in Exercise 1 that the solution of (11.3) which passes through the value y_0 at time $t = t_0$ is

$$y(t) = \frac{y^*}{\left[1 + \left(\frac{y^*}{y_0} - 1\right)e^{-rt}\right]}. \tag{11.4}$$

Figure 11.1 shows how populations evolve for different starting values. As $t \to \infty$ the population approaches the value y^* asymptotically, which explains the term carrying capacity. Since its inception by the Belgian mathematician Verhulst (1838), the logistic equation has been used for many different populations, including those of cancer cells (Thompson and Brown, 1987) as well as human populations over countries (Pearl and Reed, 1920), continents and the world (Tuckwell and Koziol, 1992, 1993).

The differential equations (11.1) and (11.3) we have thus far considered are called **deterministic** because a given initial value **determines** the solution completely for **all** subsequent times. The behaviour of the solution is totally predictable and there are no chance elements. Put another way, the trajectory $y(t)$ is fixed (it is a particular function) and there are no haphazard or **random fluctuations**.

Deterministic differential equations proved to be extremely powerful in

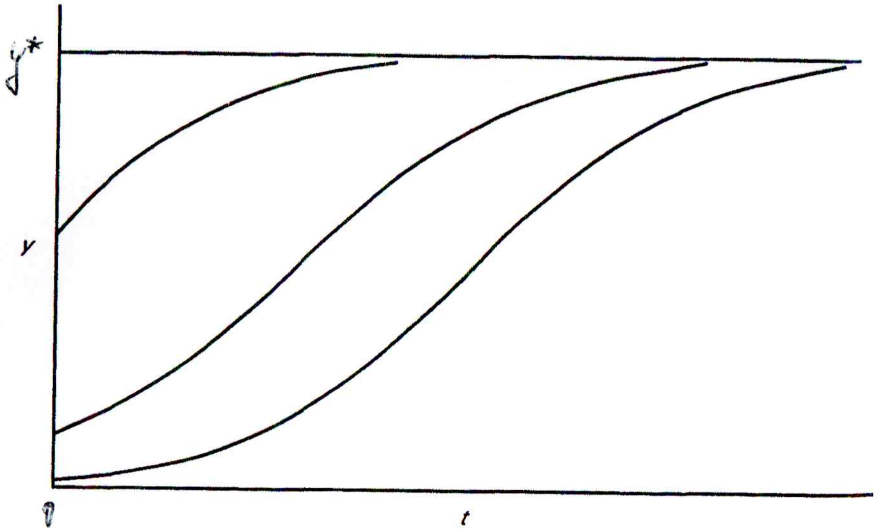


Figure 11.1 Showing solutions of a logistic differential equation for various initial population sizes.

some branches of classical physics and chemistry, but at the beginning of the twentieth century the study of atomic and subatomic systems indicated that deterministic theories were inadequate. Thus **quantum mechanics**, which is fundamentally probabilistic, was formulated to describe changes in very small systems (see for example, Schiff, 1955). Furthermore, in complex systems, containing millions or billions of interacting particles, the application of deterministic methods would have been so laborious that scientists also devised probabilistic methods for them. Such considerations for large collections of atoms or molecules lead to the discipline of **statistical mechanics** (see for example, Reichl, 1980).

In the latter part of the twentieth century quantitative methods have become increasingly widely used in the study of intrinsically complex systems such as arise in biology and economics. The use of deterministic methods is limited so there has been a large and rapid development in the application of probabilistic methods. One such very useful concept has been that of **stochastic differential equations**.

In the case of deterministic differential equations which are useful for quantitatively describing the evolution of natural systems, the solution is uniquely determined, usually by imposing a starting value and possibly other constraints. In the case of stochastic differential equations there are several possible trajectories or paths over which the system of interest may evolve. It is not known which of these trajectories will be followed, but one can often find the probabilities associated with the various paths. The situation

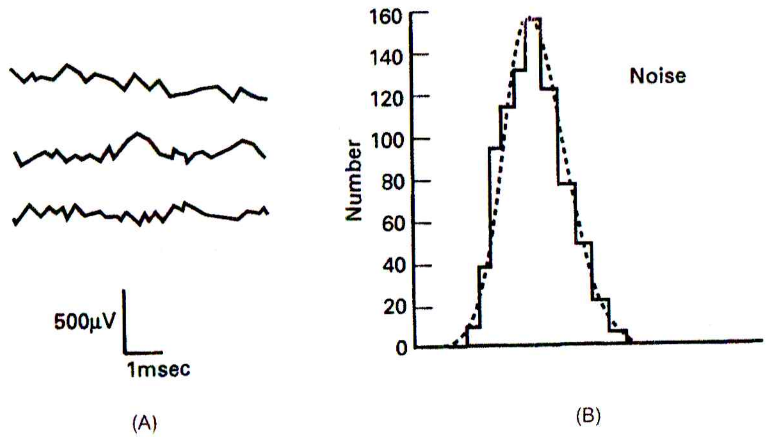


Figure 11.2a The three records on the left (A) show the fluctuations in the resting electrical potential difference across a nerve cell membrane. These fluctuations can be modelled with a stochastic differential equation involving a Wiener process – see section 12.7. On the right (B) is shown a histogram of amplitudes of the fluctuations, fitted with a normal density (from Jack, Redman and Wong, 1981).

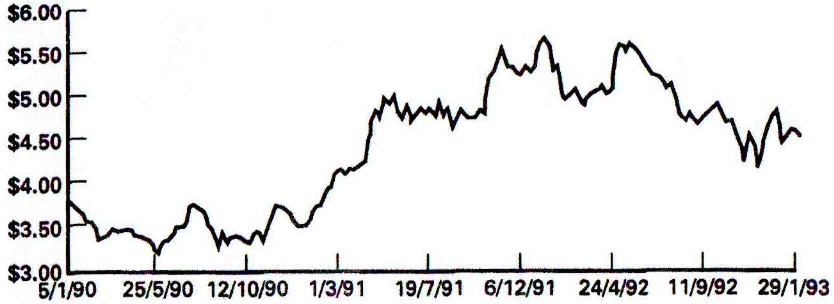


Figure 11.2b Here are shown the fluctuations in the price of a share (Coles–Myer Limited) from week to week over a period of a few years. Such fluctuations can also be modelled using a stochastic differential equation – see section 12.7.

is similar to that in the simple random walk which we studied in Chapter 7, except that in most cases the time variable is continuous rather than discrete. We could say that the quantity we are looking at wanders all over the place in a random and thus unpredictable fashion.

Physical examples of quantities which might be modelled with stochastic differential equations are illustrated in Figs 11.2a and 11.2b. In the first of these we show a record of fluctuations in the electrical potential difference across the membrane of a nerve cell in a cat's spinal cord (a spinal motorneurone which receives messages from the brain and sends messages to a muscle fibre which may result in a movement). In the second example, the weekly variations in the price of an industrial share are shown from May 1990 to January 1993.

11.2 THE WIENER PROCESS (BROWNIAN MOTION)

The most useful stochastic differential equations have proven to be those which involve either Wiener processes or Poisson processes. When Wiener processes are involved, the solutions are usually continuous whereas when Poisson processes are involved the solutions exhibit jumps. Most of our discussion focuses on continuous processes so that our immediate concern is to define Wiener processes and discuss their properties.

In Section 7.8 we considered a simple random walk and let the step size get smaller as the rate of their occurrence increased. We took this to the limit of zero step sizes and an infinite rate of occurrence, but did so in such a way that the variance at any time neither vanished nor became unbounded. In fact, the variance of the limiting random process at time t was made to equal t.

The symbol we employ for the limiting process, which we call the **Wiener process**, is $W = \{W(t), t \geq 0\}$. However, this process can be defined in a more general way, which makes no reference to limiting operations. In this section we will give this a more general definition, and discuss some of the elementary yet important properties of W.

Before we give this definition we will define a large class of processes to which both the Wiener process and Poisson process belong. This consists of those processes whose behaviour during any time interval is independent of their behaviour during any non-overlapping time interval. We will restrict our attention to processes whose index set (see section 7.1) is continuous.

Definition **Let $X = \{X(t)\}$ be a random process with a continuous parameter set $[0, T]$, where $0 < T < \infty$. Let $n \geq 2$ be an integer and suppose $0 \leq t_0 < t_1 < t_2 < \cdots < t_n \leq T$. Then X is said to be a random process with independent increments if the n random variables**

$$X(t_1) - X(t_0), X(t_2) - X(t_1), \ldots, X(t_n) - X(t_{n-1}),$$

are independent.

Thus, increments in X which occur in disjoint time intervals are independent. This implies that the evolution of the process after any time $s > 0$ is independent of the history up to and including s. Thus any process with independent increments is a **Markov process** as will be shown formally in the exercises. The converse is not true.

We have already encountered one example of an independent-increment process in section 9.2 – the **Poisson process.** Before defining a Wiener process, we mention that if the distributions of the increments of a process in various time intervals depend only on the lengths of those intervals and not their locations (i.e., their starting values), then the increments are said to be **stationary.** In section 9.2 we saw that for a Poisson process $N = \{N(t), t \geq 0\}$, the random increment $N(t_2) - N(t_1)$ is Poisson distributed with a parameter proportional to the length of the interval $(t_1, t_2]$. Thus a Poisson process has stationary independent increments.

Definition **A standard Wiener process $W = \{W(t), t \geq 0\}$, on $[0, T]$, is a process with stationary independent increments such that for any $0 \leq t_1 < t_2 \leq T$, the increment $W(t_2) - W(t_1)$ is a Gaussian random variable with mean zero and variance equal to $t_2 - t_1$; i.e.,**

$$E[W(t_2) - W(t_1)] = 0,$$
$$\text{Var}[W(t_2) - W(t_1)] = t_2 - t_1.$$

Furthermore,

$$W(0) = 0,$$

with probability 1.

The probability density $p(x;t_1,t_2)$ of the increment of W in the interval (t_1,t_2) is defined through

$$\Pr\{W(t_2)-W(t_1)\in(x,x+\Delta x]\}=p(x;t_1,t_2)\Delta x+o(\Delta x).$$

From the definition of W we see that this is given by

$$p(x;t_1,t_2)=\frac{1}{\sqrt{(2\pi(t_2-t_1))}}\exp\left[-\frac{x^2}{2(t_2-t_1)}\right]. \qquad (11.5)$$

In the case $t_1=0$, it is seen that the random variable $W(t_2)$ has mean 0 and variance t_2. Thus, **for any $t>0$, $W(t)$ is a Gaussian random variable with mean 0 and variance t, so that its probability density $p(x;t)$ is given by the simple expression**

$$p(x;t)=\frac{1}{\sqrt{(2\pi t)}}\exp\left[\frac{-x^2}{2t}\right].$$

The word 'standard' in the definition refers to the fact that the mean is zero, the variance at t is t and the initial value is zero.

Sample paths

It can be proved for the random process defined above, that the **sample paths** or **trajectories** are continuous with probability one. Sample paths are also called **realizations** and correspond to a 'value' of the process when an experiment is performed. That is, supposing it is possible to observe a standard Wiener process over the time interval $[0,T]$, we would see, with probability one, a continuous function starting at the origin, wandering around haphazardly and reaching some random end-value $W(T)$ – as in Fig. 11.3.

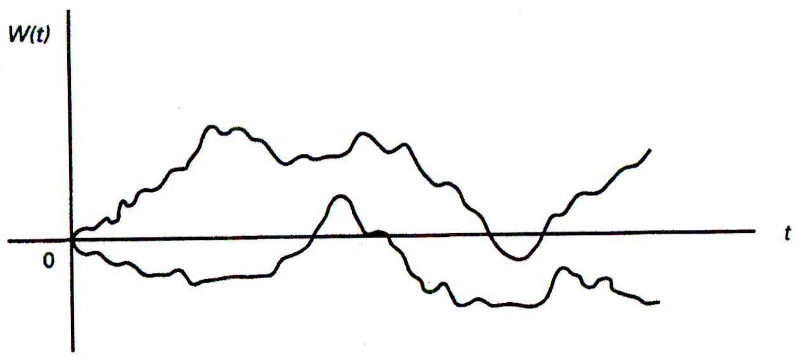


Figure 11.3 A depiction of a few sample paths for a Wiener process.

Note, however, that there are possibly discontinuous paths but these have zero probability associated with them. Usually, attention is restricted to those paths which are in fact continuous and in fact continuity of sample paths is often included in the definition. This is a convenient way to discard the problem of the discontinuous paths.

Although the probability of finding a continuous trajectory for W is one, *the probability is zero* that at any time $t\in[0,T]$ the path is differentiable. This is considered to be a pathological property and is one reason why a study of the Wiener process has been so interesting to mathematicians. This, and the fact that sample paths have *unbounded variation*, are proved and elaborated on in, for example, Hida (1980). An elementary consideration is given in Exercise 3.

Mean value and covariance function

An important property of a random process X is its mean at time t, $E(X(t))$, which is often called its **mean value function,** being a function of t alone. We have the mean and variance of $W(t)$ immediately from the above definition.

To further understand the behaviour of a random process, it is useful to know how its value at any time is connected with its value at any other time. Although knowing the joint probability distribution of these values would be nice, we may be content with a rougher indication. To this end we make the following definition.

Definition The covariance function of a random process is the covariance (cf. Chapter 1) of the values of the process at two arbitrary times.

Note that sometimes the covariance function is called an *autocovariance function* to distinguish it from a covariance between two different processes.

It is also useful to define a class of processes whose covariance function depends only on the difference between the times at which it is evaluated and not on their location.

Definition If the covariance function $\mathrm{Cov}(X(s),X(t))$ depends only on $|t-s|$, the random process X is said to be covariance stationary. Other terms for this are wide-sense stationary or weakly stationary.

If X is a weakly stationary process, we may put

$$\mathrm{Cov}(X(s),X(s+\tau))=R(\tau).$$

We can see for such a process that (see Exercises):

(a) the mean value function is a *constant*; and,

(b) the covariance function is an *even function*:

$$R(\tau) = R(-\tau).$$

In the case of a standard Wiener process we will see that the following is true.

The covariance function of a standard Wiener process is

$$\mathrm{Cov}(W(s), W(t)) = \min(s, t),$$

where min(.,.) is defined as the smaller of the two arguments.

Proof We utilize the fact that the increments of a Wiener process over disjoint (nonoverlapping) time intervals are independent random variables and hence have covariance equal to zero. With $s < t$ we have

$$\mathrm{Cov}[W(s), W(t) - W(s)] = 0.$$

The quantity we seek can be written

$$\mathrm{Cov}[W(s), W(t) + W(s) - W(s)].$$

But in general, if A, B, and C are three random variables (see Exercises),

$$\mathrm{Cov}[A, B + C] = \mathrm{Cov}[A, B] + \mathrm{Cov}[A, C].$$

Thus,

$$\begin{aligned}\mathrm{Cov}[W(s), W(t)] &= \mathrm{Cov}[W(s), W(t) - W(s)] + \mathrm{Cov}[W(s), W(s)]\\ &= \mathrm{Cov}[W(s), W(s)]\\ &= \mathrm{Var}[W(s)] = s.\end{aligned}$$

Had t been less than s we would have obtained t instead of s. Hence the covariance is the smaller of s and t, which proves the result.

Note that the Wiener process is not therefore covariance-stationary as the covariance of $W(s)$ and $W(t)$ depends directly on the magnitude of the smaller of s or t. For further information on the topics we have dealt with in this section, the reader may consult Papoulis (1965), Parzen (1962) and Yaglom (1973).

11.3 WHITE NOISE

Although the Wiener process is of central importance in the theory of stochastic differential equations, there is a useful related concept, called white noise, which we introduce in this section.

The paths traced out by a Wiener process are with probability one not differentiable. However, it is often convenient to talk about the derivative of

W as if it did exist. We use the symbol $w(t)$ for the 'derivative' of $W(t)$, and we call the random process $w = \{w(t), t \geqslant 0\}$, (Gaussian) **white noise**. However, it must be remembered that, strictly speaking, this process does not have a well-defined meaning – it is nevertheless heuristically useful.

The word *noise*, of course, refers to unwanted signals. If you are in a crowded cafeteria or football stadium or surrounded by dense city traffic, close your eyes and listen, you will hear a noise that seems an amorphous assortment of meaningless sounds; you generally won't be able to pick out particular signals unless they originate close-by. This kind of background noise is an acoustic approximation to white noise. Sound engineers have devices called white noise generators which are used to test the acoustic properties of rooms – the basic idea is to subject the chamber to all frequencies at once.

The mean value and covariance functions of white noise can be obtained from those of a Wiener process – as will be seen in the exercises. These turn out to be

$$\begin{aligned}E[w(t)] &= 0,\\ \mathrm{Cov}[w(s), w(t)] &= \delta(t - s).\end{aligned} \tag{11.6}$$

Thus the covariance is zero whenever $s \neq t$ and is very very large when $s = t$.

Covariance functions are often decomposed to see if there are regularities present, especially in the form of periodicities or harmonics of various frequencies. Such a decomposition is done using the following definition. Note that we restrict our attention to real-valued processes.

Definition The spectral density $S(k)$ of a covariance-stationary random process whose covariance function is $R(t), t \geqslant 0$, is given by the integral

$$S(k) = \int_{-\infty}^{\infty} \cos(kt) R(t)\, dt. \tag{11.7}$$

The reader may recognize this as the *Fourier transform* of $R(t)$, recalling that the latter is here an *even* function of t. Another name for $S(k)$ is the **power spectrum** – it indicates the contributions from various frequencies to the total activity of the process.

A knowledge of the spectral density can be used to obtain the covariance function using the following **inversion formula** which is proved in courses of analysis (see for example Wylie, 1960),

$$R(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(k) \cos(kt)\, dk. \tag{11.8}$$

Let us see how various harmonics in $R(t)$ manifest themselves in $S(k)$. Suppose $S(k)$ were very much concentrated around the single frequency k_0

so we might put $S(k) = \delta(k - k_0)$. Then

$$R(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \delta(k - k_0)\cos(kt)\,dk$$
$$= \frac{1}{2\pi}\cos(k_0 t),$$

where we have used the substitution property of the delta function (formula (3.13)). Thus we see that a very large peak in the spectral density $S(k)$ comes about at k_0 if there is a single dominant frequency $k_0/2\pi$ in the covariance function $R(t)$.

Let us consider white noise $w(t)$ from this point of view. We have from Equation (11.6), $R(\tau) = \delta(\tau)$. Substituting this in the definition of the spectral density gives

$$S(k) = \int_{-\infty}^{\infty} \cos(kt)\delta(t)\,dt$$
$$= 1,$$

where we have used the substitution property and the fact that $\cos(0) = 1$.

This tells us that the spectral density of white noise is a constant, independent of the frequency. That is, all frequencies contribute equally, from $-\infty$ to ∞, whereby we can see the analogy with 'white light'. Hence the description of the derivative of a Wiener process as (Gaussian) white noise. It is realized that it is not physically possible to have frequencies over such a huge range. In engineering practice white noise generators have cut-off frequencies at finite values – they are called **band-limited** white noises. Sometimes white noise is called **delta-correlated** noise.

11.4 THE SIMPLEST STOCHASTIC DIFFERENTIAL EQUATIONS – THE WIENER PROCESS WITH DRIFT

In this section we will take a first look at stochastic differential equations involving Wiener processes. A more detailed account will be given in the next chapter.

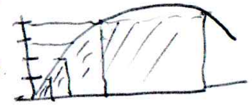
The increment in a standard Wiener process in a small time interval $(t, t + \Delta t]$ is

$$\Delta W(t) = W(t + \Delta t) - W(t),$$

and we know from above that ΔW is normally distributed with mean zero and variance Δt. We use a similar notation as in differential calculus and

use the symbol $dW(t)$ or dW to indicate the limiting increment or **stochastic differential** as $\Delta t \to 0$. The simplest stochastic differential equation involving a Wiener process is thus:

$$dX = dW \tag{11.9}$$



which states that the increments in X are those of W. The solution of (11.9) is

$$X(t) = X(0) + W(t),$$

which states that the value of the process X at time t, namely the random variable $X(t)$, is equal to the sum of two random variables: the initial value $X(0)$ and the value of a standard Wiener process at time t. Equation (11.9) is interpreted more rigorously as the corresponding integral

$$\int_0^t dX(t') = \int_0^t dW(t'),$$

whose meaning will be explained in section 12.5. This gives

$$X(t) - X(0) = W(t) - W(0) = W(t),$$

which is the same as (11.10) because from the definition, $W(0) = 0$, identically.

Notice that when writing stochastic differential equations involving a Wiener process, we usually avoid writing time derivatives because, as we have seen, these do not, strictly speaking, exist. However, we can, if we are careful in our interpretation, just as well write (11.9) as

$$\frac{dX}{dt} = w(t),$$

where w is white noise.

We may perform simple algebraic operations on a standard Wiener process. For example, we can form a new process whose value at time t is obtained by multiplying $W(t)$ by a constant σ, usually assumed to be positive; adding a linear function of time μt, where μ can be negative, zero, or positive; and giving a particular initial value $X(0) = x_0$:

$$X(t) = x_0 + \mu t + \sigma W(t) \tag{11.10}$$

This defines a *Wiener process with drift μt and variance parameter σ*. The drift function μt here is linear, though any other deterministic function of time can be added instead of μt. For the random process defined by (11.10) we write the *stochastic differential equation*

$$dX = \mu\,dt + \sigma\,dW, \tag{11.11}$$

and say that (11.10) is a solution of (11.11) with a particular initial value.

The following properties of a Wiener process with drift will be verified in the exercises:

$$E[X(t)] = x_0 + \mu t,$$
$$\mathrm{Cov}[X(s), X(t)] = \sigma^2 \min(s, t).$$
$$\mathrm{Var}[X(t)] = \sigma^2 t.$$

To obtain the probability density function for the Wiener process with drift, as defined by (11.10), we note, as proven in introductory probability theory, that linear operations on a Gaussian random variable produce another Gaussian random variable. Thus $X(t)$ must be a Gaussian random variable with mean and variance as given above. Its probability density function, conditioned on an initial value x_0, is defined through either

$$p(x, t|x_0) = \lim_{\Delta x \to 0} \frac{\Pr\{x < X(t) < x + \Delta x | X(0) = x_0\}}{\Delta x},$$

or

$$p(x, t|x_0)\Delta x = \Pr\{x < X(t) < x + \Delta x\} + o(\Delta x),$$

where $t > 0$, $-\infty < x_0, x < \infty$. This density must be given by

$$p(x, t|x_0) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left[-\frac{(x - x_0 - \mu t)^2}{2\sigma^2 t}\right]. \tag{11.12}$$

(Note that when dealing with continuous random variables as we are here, we can put $<$ rather than $\leqslant$ in inequalities because single points make no contribution.)

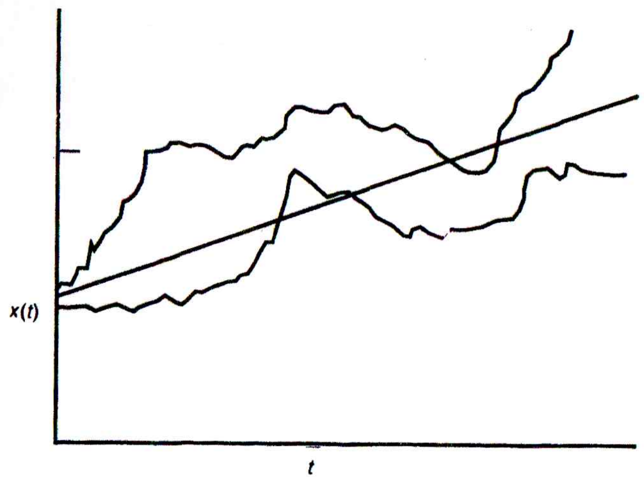


Figure 11.4 A depiction of a few sample paths for a Wiener process with drift $X(t) = x_0 + \mu t + \sigma W(t)$ with $x_0 = 1$, $\mu = \frac{1}{2}$, and $\sigma = 1$.

In anticipation of the material in the next chapter, we mention that the function $p(x, t|x_0)$, given in (11.12), satisfies a simple partial differential equation called a heat equation. This will be familiar to students either from calculus or physics courses and here takes the form,

$$\frac{\partial p}{\partial t} = -\mu \frac{\partial p}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2 p}{\partial x^2} \tag{11.13}$$

as will be verified in the exercises. It can be seen, therefore, that asserting that the probability density of a Markov process satisfies this partial differential equation, is, for all intents and purposes, the same as saying that the process is a Wiener process with drift.

Figure 11.4 illustrates how a Wiener process with drift might behave in the case of a positive drift, with drift parameter $\mu = \frac{1}{2}$ and variance parameter $\sigma = 1$ when $x_0 = 1$.

11.5 TRANSITION PROBABILITIES AND THE CHAPMAN–KOLMOGOROV EQUATION

Before considering a wide class of random processes which can be succinctly described in the language of stochastic differential equations, we will lay the groundwork for an analytical approach to studying their properties.

We saw in Chapter 8 that the fundamental descriptive quantity for Markov chains in discrete time was the set (or matrix) of transition probabilities. For the processes we considered, it was sufficient to specify the one-step transition probabilities, as the probabilities of all other transitions could be obtained from them. In particular, if the initial probability distribution was specified, the probability distribution of the process could be obtained at any time point – see Equation (8.11). Similarly, in Chapter 9, we saw that a set of transition probabilities could be used to quantitatively describe the evolution of Markov chains in continuous time.

The processes we are concerned with here are **Markov processes in continuous time which take on a continuous set of values.** The evolution of such processes is also specified by giving a set of transition probabilities as alluded to in the case of a Wiener process with drift. In general, let $\{X(t), t \geqslant 0\}$ be such a process. Then the **transition probability distribution function** gives the probability distribution of the value of the process at a particular time, conditioned on a known value of the process at some earlier time.

Definition **Let X be a continuous time random process taking on a continuous set of values. The transition probability distribution function $P(y, t|x, s)$, with $s \leqslant t$, is the distribution function of $X(t)$ conditioned on the event $X(s) = x$.**

Thus,

$$P(y, t \mid x, s) = \Pr\{X(t) \leqslant y \mid X(s) = x\},$$

where usually $0 \leqslant s \leqslant t \leqslant \infty$ **and** $-\infty < x, y < \infty$.

The variables (x, s) which refer to the state of affairs at the earlier time are called *backward variables*, whereas those which refer to the later time are called *forward variables*.

Furthermore, if $P(y, t \mid x, s)$ is differentiable with respect to y, then its derivative, with t, x and s fixed,

$$p(y, t \mid x, s) = \frac{\partial P}{\partial y},$$

is called the **transition probability density function** (tpdf) of the process X.

The Chapman–Kolmogorov integral equation

In section 8.5 the *Chapman–Kolmogorov equations* were established for discrete-time Markov chains. These equations imply that a transition from one state to another in more than one time step must involve an intermediate passage to one of the permissible values of the process. For the type of

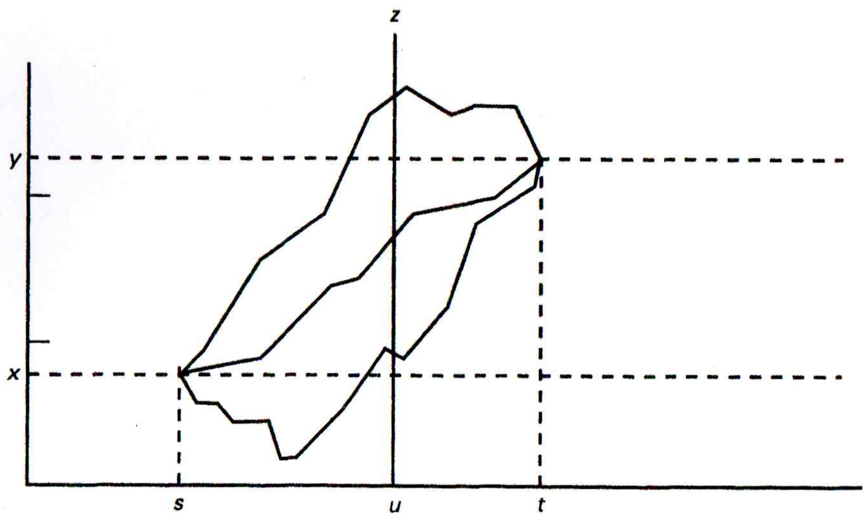


Figure 11.5 Showing how passage from (x, s) to (y, t) must involve passage to one of a continuum of intermediate values, z, at some intermediate time, u. Summing (integrating) over all possible paths gives the Chapman–Kolmogorov equation (11.14).

process of concern to us here, the corresponding equation involves not a sum but an integral over intermediate possible values, reflecting the fact that the state-space is continuous. Refer to Fig. 11.5.

We see that a transition from state x at time s to state y at a later time t, with associated probability density $p(y, t \mid x, s)$, must, at some intermediate time $u \in (s, t)$, involve passage to one of the permissible values z, here assumed to be any real number. The probability of a transition from (x, s) to (z, u) followed by a transition from (z, u) to (y, t) is proportional to the product $p(z, u \mid x, s)p(y, t \mid z, u)$. Integrating over these intermediate values gives the **Chapman–Kolmogorov equation,**

$$p(y, t \mid x, s) = \int_{z=-\infty}^{z=\infty} p(y, t \mid z, u)p(z, u \mid x, s)\,dz. \tag{11.14}$$

It can be seen that this is an integral over all possible paths from (x, s) to (y, t). The Chapman–Kolmogorov equation is useful for deriving differential equations satisfied by the transition density function.

Using similar reasoning, we may find the (absolute) probability distribution of the process at time t from a knowledge of the initial distribution and the transition probability density function or the transition probability distribution function. Let $f(x)$, where $-\infty < x < \infty$, be the density of $X(0)$. Then, to get the probability of being in state y at $t > 0$, we have to integrate over all possible initial values x, weighted with $f(x)\,dx$ and with the probability of a transition from x to y:

$$\Pr\{y < X(t) < y + dy\} = \left\{\int_{x=-\infty}^{x=\infty} f(x)p(y, t \mid x, 0)\,dx\right\}dy.$$

(Note that this is *not* a conditional probability.)

Similarly, the distribution function of $X(t)$ is given by

$$\Pr\{X(t) \leqslant y\} = \int_{z=-\infty}^{z=y}\left(\int_{-\infty}^{\infty} f(x)p(z, t \mid x, 0)\,dx\right)dz$$
$$= \int_{-\infty}^{\infty} f(x)\left(\int_{-\infty}^{y} p(z, t \mid x, 0)\,dz\right)dx.$$

That is,

$$\Pr\{X(t) \leqslant y\} = \int_{-\infty}^{\infty} f(x)P(y, t \mid x, 0)\,dx,$$

where P is the transition probability distribution function.

In the special case, often encountered, where the initial value is not random but a particular specified value x_0, say, so that $f(x) = \delta(x - x_0)$, we have, by the substitution property of the delta function,

$$\Pr\{X(t) \leqslant y\} = \int_{-\infty}^{\infty} \delta(x - x_0) P(y, t | x, 0)\, dx$$

$$= P(y, t | x_0, 0),$$

as expected. Thus the absolute and transition probability distributions coincide.

When one is seeking the properties of a process X, one may work with the random variables $\{X(t), t \geqslant 0\}$ directly, or one may work with the transition probability functions. The latter approach is called the *analytical approach* and is often more useful than the direct approach because it involves solving differential equations, which is a long and much-studied discipline. The direct approach usually involves stochastic integrals which we shall consider in section 12.5.

REFERENCES

Hida, T. (1980). *Brownian Motion.* Springer-Verlag, New York.

Jack, J.J.B., Redman, S.J. and Wong, K. (1981). The components of synaptic potentials evoked in cat spinal motoneurones by impulses in single group Ia afferents. *J. Physiol.* **321**, 65–96.

Papoulis, A. (1965). *Probability, Random Variables and Stochastic Processes.* McGraw-Hill, New York.

Parzen, E. (1962). *Stochastic Processes.* Holden-Day, San Francisco.

Pearl, R. and Reed, L.J. (1920). On the rate of growth of the population of the United States since 1790 and its mathematical representation. *Proc. Natl. Acad. Sci. USA*, **6**, 275–288.

Reichl, L.E. (1980). *A Modern Course in Statistical Physics.* University of Texas, Austin.

Schiff, L.I. (1955). *Quantum Mechanics.* McGraw-Hill, New York.

Thompson, J.R. and Brown, B.W. (eds) (1987). *Cancer Modeling.* Marcel Dekker, New York.

Tuckwell, H.C. and Koziol, J.A. (1992). World population. *Nature*, **359**, 200.

Tuckwell, H.C. and Koziol, J.A. (1993). World and regional populations. *BioSystems* **31**, 59–63.

Verhulst, P.F. (1838). Notice sur la loi que la population suit dans son accroissement. *Corr. Math. Phys.*, **10**, 113–121.

Wylie, C.R. (1960). *Advanced Engineering Mathematics.* McGraw-Hill, New York.

Yaglom, A.M. (1973). *An Introduction to the Theory of Stationary Random Functions.* Dover, New York.

EXERCISES

1. Prove that the solution of the logistic differential equation (11.3) is in fact given by (11.4). (*Hint*: Separate the variables by putting the equation in the form $f(y)\,dy = g(t)\,dt$, and integrate each side.)

2. Show that a continuous time process with independent increments is a Markov process. (*Hint*: It will suffice to proceed as in Exercise 7.2; examine $\Pr(X(t_3) = z | X(t_2) = y, X(t_1) = x)$, where $t_1 < t_2 < t_3$.)
3. Let $\Delta W = W(t + \Delta t) - W(t)$ be an increment in a standard Wiener process. Examine the limits as $\Delta t \to 0$ of (i) $E[\Delta W]$, (ii) $\mathrm{Var}[\Delta W]$, (iii) $E[\Delta W/\Delta t]$, and (iv) $\mathrm{Var}[\Delta W/\Delta t]$ to provide an indication that W has continuous but not differentiable sample paths.
4. Show that if A, B, and C are three random variables,

$$\mathrm{Cov}[A, (B + C)] = \mathrm{Cov}[A, B] + \mathrm{Cov}[A, C].$$

5. Prove that if a and b are constants, and X and Y are random variables, then

   (a) $\mathrm{Cov}[X + a, Y + b] = \mathrm{Cov}[X, Y]$;
   (b) thus $\mathrm{Cov}[X, Y] = \mathrm{Cov}[X - E[X], Y - E[Y]]$, so means can always be subtracted when calculating covariances;
   (c) $\mathrm{Cov}[a, X, bY] = ab\,\mathrm{Cov}[X, Y]$.
6. Show that for a covariance (weakly) stationary process, the mean value function is constant and the covariance function satisfies $R(\tau) = R(-\tau)$.
7. The derivative $\{X'(t)\}$ of a random process $\{X(t)\}$ can be defined in the usual way by

$$X'(t) = \lim_{\Delta t \to 0} \frac{X(t + \Delta t) - X(t)}{\Delta t}.$$

Use the results

$$E[X'(t)] = \frac{d}{dt} E[X(t)],$$

$$\mathrm{Cov}[X'(s), X'(t)] = \frac{\partial^2}{\partial s \partial t}(\mathrm{Cov}[X(s), X(t)]),$$

to verify the given formulas for the mean and covariance function of white noise.

8. Establish the following results for a Wiener process with drift μt and variance parameter σ:

$$E[X(t)] = x_0 + \mu t,$$

$$\mathrm{Var}[X(t)] = \sigma^2 t,$$

$$\mathrm{Cov}[X(s), X(t)] = \sigma^2 \min(s, t).$$

9. Compute the probability that the Wiener process with drift $X = 3W + t$ satisfies the inequality

$$0 \leqslant X(t) \leqslant 1.$$