8

Random processes

Summary. This brief introduction to random processes includes elementary previews of stationary processes, renewal processes, queueing processes, and the Wiener process (Brownian motion). It ends with a discussion of the Kolmogorov consistency conditions.

8.1 Introduction

Recall that a 'random process' X is a family $\{X_t : t \in T\}$ of random variables which map the sample space Ω into some set S. There are many possible choices for the index set T and the state space S, and the characteristics of the process depend strongly upon these choices. For example, in Chapter 6 we studied discrete-time ($T = \{0, 1, 2, \dots\}$) and continuous-time ($T = [0, \infty)$) Markov chains which take values in some countable set S. Other possible choices for T include $\mathbb{R}^n$ and $\mathbb{Z}^n$, whilst S might be an uncountable set such as $\mathbb{R}$. The mathematical analysis of a random process varies greatly depending on whether S and T are countable or uncountable, just as discrete random variables are distinguishable from continuous variables. The main differences are indicated by those cases in which

(a) $T = \{0, 1, 2, \dots\}$ or $T = [0, \infty)$,

(b) $S = \mathbb{Z}$ or $S = \mathbb{R}$.

There are two levels at which we can observe the evolution of a random process X.

(a) Each X_t is a function which maps Ω into S. For any fixed $\omega \in \Omega$, there is a corresponding collection $\{X_t(\omega) : t \in T\}$ of members of S; this is called the *realization* or *sample path* of X at ω. We can study properties of sample paths.

(b) The X_t are not independent in general. If $S \subseteq \mathbb{R}$ and $\mathbf{t} = (t_1, t_2, \dots, t_n)$ is a vector of members of T, then the vector $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ has joint distribution function $F_{\mathbf{t}} : \mathbb{R}^n \to [0, 1]$ given by $F_{\mathbf{t}}(\mathbf{x}) = \mathbb{P}(X_{t_1} \le x_1, \dots, X_{t_n} \le x_n)$. The collection $\{F_{\mathbf{t}}\}$, as $\mathbf{t}$ ranges over all vectors of members of T of any finite length, is called the collection of *finite-dimensional distributions* (abbreviated to *fdds*) of X, and it contains all the information which is available about X from the distributions of its component variables X_t. We can study the distributional properties of X by using its fdds.

These two approaches do not generally yield the same information about the process in question, since knowledge of the fdds does not yield complete information about the properties of the sample paths. We shall see an example of this in the final section of this chapter.

We are not concerned here with the general theory of random processes, but prefer to study certain specific collections of processes which are characterized by one or more special properties. This is not a new approach for us. In Chapter 6 we devoted our attention to processes which satisfy the Markov property, whilst large parts of Chapter 7 were devoted to sequences $\{S_n\}$ which were either martingales or the partial sums of independent sequences. In this short chapter we introduce certain other types of process and their characteristic properties. These can be divided broadly under four headings, covering 'stationary processes', 'renewal processes', 'queues', and 'diffusions'; their detailed analysis is left for Chapters 9, 10, 11, and 13 respectively.

We shall only be concerned with the cases when T is one of the sets $\mathbb{Z}$, $\{0, 1, 2, \dots\}$, $\mathbb{R}$, or $[0, \infty)$. If T is an uncountable subset of $\mathbb{R}$, representing continuous time say, then we shall usually write $X(t)$ rather than X_t for ease of notation. Evaluation of $X(t)$ at some $\omega \in \Omega$ yields a point in S, which we shall denote by $X(t; \omega)$.

8.2 Stationary processes

Many important processes have the property that their finite-dimensional distributions are invariant under time shifts (or space shifts if T is a subset of some Euclidean space $\mathbb{R}^n$, say).

(1) Definition. The process $X = \{X(t) : t \geq 0\}$, taking values in $\mathbb{R}$, is called **strongly stationary** if the families

$$\{X(t_1), X(t_2), \dots, X(t_n)\} \quad \text{and} \quad \{X(t_1 + h), X(t_2 + h), \dots, X(t_n + h)\}$$

have the same joint distribution for all $t_1, t_2, \dots, t_n$ and $h > 0$.

Note that, if X is strongly stationary, then $X(t)$ has the same distribution for all t.

We saw in Section 3.6 that the covariance of two random variables X and Y contains some information, albeit incomplete, about their joint distribution. With this in mind we formulate another stationarity property which, for processes with $\text{var}(X(t)) < \infty$, is weaker than strong stationarity.

(2) Definition. The process $X = \{X(t) : t \geq 0\}$ is called **weakly** (or **second-order** or **covariance) stationary** if, for all t_1, t_2, and $h > 0$,

$$\mathbb{E}(X(t_1)) = \mathbb{E}(X(t_2)) \quad \text{and} \quad \text{cov}\big(X(t_1), X(t_2)\big) = \text{cov}\big(X(t_1 + h), X(t_2 + h)\big).$$

Thus, X is weakly stationary if and only if it has constant means, and its *autocovariance function*

(3) $$c(t, t + h) = \text{cov}\big(X(t), X(t + h)\big)$$

satisfies

$$c(t, t + h) = c(0, h) \quad \text{for all } t, h \geq 0.$$

We emphasize that the autocovariance function $c(s, t)$ of a weakly stationary process is a function of $t - s$ only.

Definitions similar to (1) and (2) hold for processes with $T = \mathbb{R}$ and for discrete-time processes $X = \{X_n : n \geq 0\}$; the autocovariance function of a weakly stationary discrete-time process X is just a sequence $\{c(0, m) : m \geq 0\}$ of real numbers.

Weak stationarity interests us more than strong stationarity for two reasons. First, the condition of strong stationarity is often too restrictive for certain applications; secondly, many substantial and useful properties of stationary processes are derivable from weak stationarity alone. Thus, the assertion that X is *stationary* should be interpreted to mean that X is *weakly stationary*. Of course, there exist processes which are stationary but not strongly stationary (see Example (5)), and conversely processes without finite second moments may be strongly stationary but not weakly stationary.

(4) Example. Markov chains. Let $X = \{X(t) : t \geq 0\}$ be an irreducible Markov chain taking values in some countable subset S of $\mathbb{R}$ and with a unique stationary distribution $\boldsymbol{\pi}$. Then (see Theorem (6.9.21))

$$\mathbb{P}\big(X(t) = j \,\big|\, X(0) = i\big) \to \pi_j \quad \text{as} \quad t \to \infty$$

for all $i, j \in S$. The fdds of X depend on the initial distribution $\boldsymbol{\mu}^{(0)}$ of $X(0)$, and it is not generally true that X is stationary. Suppose, however, that $\boldsymbol{\mu}^{(0)} = \boldsymbol{\pi}$. Then the distribution $\boldsymbol{\mu}^{(t)}$ of $X(t)$ satisfies $\boldsymbol{\mu}^{(t)} = \boldsymbol{\pi}\mathbf{P}_t = \boldsymbol{\pi}$ from equation (6.9.19), where $\{\mathbf{P}_t\}$ is the transition semigroup of the chain. Thus $X(t)$ has distribution $\boldsymbol{\pi}$ for all t. Furthermore, if $0 < s < s+t$ and $h > 0$, the pairs $(X(s), X(s+t))$ and $(X(s+h), X(s+t+h))$ have the same joint distribution since:

(a) $X(s)$ and $X(s+h)$ are identically distributed,

(b) the distribution of $X(s+h)$ (respectively $X(s+t+h)$) depends only on the distribution of $X(s)$ (respectively $X(s+t)$) and on the transition matrix $\mathbf{P}_h$.

A similar argument holds for collections of the $X(u)$ which contain more than two elements, and we have shown that X is strongly stationary. ●

(5) Example. Let A and B be uncorrelated (but not necessarily independent) random variables, each of which has mean 0 and variance 1. Fix a number $\lambda \in [0, \pi]$ and define

(6) $$X_n = A\cos(\lambda n) + B\sin(\lambda n).$$

Then $\mathbb{E}X_n = 0$ for all n and $X = \{X_n\}$ has autocovariance function

$$\begin{aligned} c(m, m+n) &= \mathbb{E}(X_m X_{m+n}) \\ &= \mathbb{E}\Big(\big[A\cos(\lambda m) + B\sin(\lambda m)\big]\big[A\cos\{\lambda(m+n)\} + B\sin\{\lambda(m+n)\}\big]\Big) \\ &= \mathbb{E}\big(A^2\cos(\lambda m)\cos\{\lambda(m+n)\} + B^2\sin(\lambda m)\sin\{\lambda(m+n)\}\big) \\ &= \cos(\lambda n) \end{aligned}$$

since $\mathbb{E}(AB) = 0$. Thus $c(m, m+n)$ depends on n alone and so X is stationary. In general X is not strongly stationary unless extra conditions are imposed on the joint distribution of A and B; to see this for the case $\lambda = \frac{1}{2}\pi$, simply calculate that

$$\{X_0, X_1, X_2, X_3, \dots\} = \{A, B, -A, -B, \dots\}$$

which is strongly stationary if and only if the pairs (A, B), $(B, -A)$, and $(-A, -B)$ have the same joint distributions. It can be shown that X is strongly stationary for any λ if A and B are $N(0, 1)$ variables. The reason for this lies in Example (4.5.9), where we saw that normal variables are independent whenever they are uncorrelated. ●

Two major results in the theory of stationary processes are the 'spectral theorem' and the 'ergodic theorem'; we close this section with a short discussion of these. First, recall from the theory of Fourier analysis that any function $f : \mathbb{R} \to \mathbb{R}$ which

(a) is periodic with period 2π (that is, $f(x + 2\pi) = f(x)$ for all x),

(b) is continuous, and

(c) has bounded variation,

has a unique Fourier expansion

$$f(x) = \tfrac{1}{2}a_0 + \sum_{n=1}^{\infty}\big[a_n \cos(nx) + b_n \sin(nx)\big]$$

which expresses f as the sum of varying proportions of regular oscillations. In some sense to be specified, a stationary process X is similar to a periodic function since its autocovariances are invariant under time shifts. The spectral theorem asserts that, subject to certain conditions, stationary processes can be decomposed in terms of regular underlying oscillations whose magnitudes are random variables; the set of frequencies of oscillations which contribute to this combination is called the 'spectrum' of the process. For example, the process X in (5) is specified precisely in these terms by (6). In spectral theory it is convenient to allow the processes in question to take values in the complex plane. In this case (6) can be rewritten as

$$(7) \qquad X_n = \operatorname{Re}(Y_n) \quad \text{where} \quad Y_n = Ce^{i\lambda n};$$

here C is a complex-valued random variable and $i = \sqrt{-1}$. The sequence $Y = \{Y_n\}$ is stationary also whenever $\mathbb{E}(C) = 0$ and $\mathbb{E}(C\overline{C}) < \infty$, where $\overline{C}$ is the complex conjugate of C (but see Definition (9.1.1)).

The ergodic theorem deals with the partial sums of a stationary sequence $X = \{X_n : n \geq 0\}$. Consider first the following two extreme examples of stationarity.

(8) Example. Independent sequences. Let $X = \{X_n : n \geq 0\}$ be a sequence of independent identically distributed variables with zero means and unit variances. Certainly X is stationary, and its autocovariance function is given by

$$c(m, m+n) = \mathbb{E}(X_m X_{m+n}) = \begin{cases} 1 & \text{if } n = 0, \\ 0 & \text{if } n \neq 0. \end{cases}$$

The strong law of large numbers asserts that $n^{-1}\sum_{j=1}^{n} X_j \xrightarrow{\text{a.s.}} 0$. ●

(9) Example. Identical sequences. Let Y be a random variable with zero mean and unit variance, and let $X = \{X_n : n \geq 0\}$ be the stationary sequence given by $X_n = Y$ for all n. Then X has autocovariance function $c(m, m+n) = \mathbb{E}(X_m X_{m+n}) = 1$ for all n. It is clear that $n^{-1}\sum_{j=1}^{n} X_j \xrightarrow{\text{a.s.}} Y$ since each term in the sum is Y itself. ●

These two examples are, in some sense, extreme examples of stationarity since the first deals with independent variables and the second deals with identical variables. In both examples, however, the averages $n^{-1}\sum_{j=1}^{n} X_j$ converge as $n \to \infty$. In the first case the limit is constant, whilst in the second the limit is a random variable with a non-trivial distribution. This indicates a shared property of 'nice' stationary processes, and we shall see that any stationary sequence $X = \{X_n : n \geq 0\}$ with finite means satisfies

$$\frac{1}{n}\sum_{j=1}^{n} X_j \xrightarrow{\text{a.s.}} Y$$

for some random variable Y. This result is called the ergodic theorem for stationary sequences. A similar result holds for continuous-time stationary processes.

The theory of stationary processes is important and useful in statistics. Many sequences $\{x_n : 0 \leq n \leq N\}$ of observations, indexed by the time at which they were taken, are suitably modelled by random processes, and statistical problems such as the estimation of unknown parameters and the prediction of the future values of the sequence are often studied in this context. Such sequences are called 'time series' and they include many examples which are well known to us already, such as the successive values of the Financial Times Share Index, or the frequencies of sunspots in successive years. Statisticians and politicians often seek to find some underlying structure in such sequences, and to this end they may study 'moving average' processes Y, which are smoothed versions of a stationary sequence X,

$$Y_n = \sum_{i=0}^{r} \alpha_i X_{n-i},$$

where $\alpha_0, \alpha_1, \dots, \alpha_r$ are constants. Alternatively, they may try to fit a model to their observations, and may typically consider 'autoregressive schemes' Y, being sequences which satisfy

$$Y_n = \sum_{i=1}^{r} \alpha_i Y_{n-i} + Z_n$$

where $\{Z_n\}$ is a sequence of uncorrelated variables with zero means and constant finite variance.

An introduction to the theory of stationary processes is given in Chapter 9.

Exercises for Section 8.2

1. Flip–flop. Let $\{X_n\}$ be a Markov chain on the state space $S = \{0, 1\}$ with transition matrix

$$\mathbf{P} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix},$$

where $\alpha + \beta > 0$. Find:

(a) the correlation $\rho(X_m, X_{m+n})$, and its limit as $m \to \infty$ with n remaining fixed,

(b) $\lim_{n\to\infty} n^{-1}\sum_{r=1}^{n} \mathbb{P}(X_r = 1)$.

Under what condition is the process strongly stationary?

2. Random telegraph. Let $\{N(t) : t \geq 0\}$ be a Poisson process of intensity λ, and let T_0 be an independent random variable such that $\mathbb{P}(T_0 = \pm 1) = \frac{1}{2}$. Define $T(t) = T_0(-1)^{N(t)}$. Show that $\{T(t) : t \geq 0\}$ is stationary and find: (a) $\rho(T(s), T(s+t))$, (b) the mean and variance of $X(t) = \int_0^t T(s)\,ds$.

3. **Korolyuk–Khinchin theorem.** An integer-valued counting process $\{N(t) : t \geq 0\}$ with $N(0) = 0$ is called *crudely stationary* if $p_k(s, t) = \mathbb{P}(N(s+t) - N(s) = k)$ depends only on the length t and not on the location s. It is called *simple* if, almost surely, it has jump discontinuities of size 1 only. Show that, for a simple crudely stationary process N, $\lim_{t \downarrow 0} t^{-1}\mathbb{P}(N(t) > 0) = \mathbb{E}(N(1))$.

8.3 Renewal processes

We are often interested in the successive occurrences of events such as the emission of radioactive particles, the failures of light bulbs, or the incidences of earthquakes.

(1) Example. Light bulb failures. This is the archetype of renewal processes. A room is lit by a single light bulb. When this bulb fails it is replaced immediately by an apparently identical copy. Let X_i be the (random) lifetime of the ith bulb, and suppose that the first bulb is installed at time $t = 0$. Then $T_n = X_1 + X_2 + \cdots + X_n$ is the time until the nth failure (where, by convention, we set $T_0 = 0$), and

$$N(t) = \max\{n : T_n \leq t\}$$

is the number of bulbs which have failed by time t. It is natural to assume that the X_i are independent and identically distributed random variables. ●

(2) Example. Markov chains. Let $\{Y_n : n \geq 0\}$ be a Markov chain, and choose some state i. We are interested in the time epochs at which the chain is in the state i. The times $0 < T_1 < T_2 < \cdots$ of successive visits to i are given by

$$T_1 = \min\{n \geq 1 : Y_n = i\},$$
$$T_{m+1} = \min\{n > T_m : Y_n = i\} \quad \text{for} \quad m \geq 1;$$

they may be defective unless the chain is irreducible and persistent. Let $\{X_m : m \geq 1\}$ be given by

$$X_m = T_m - T_{m-1} \quad \text{for} \quad m \geq 1,$$

where we set $T_0 = 0$ by convention. It is clear that the X_m are independent, and that $X_2, X_3, \ldots$ are identically distributed since each is the elapsed time between two successive visits to i. On the other hand, X_1 does *not* have this shared distribution in general, unless the chain began in the state $Y_0 = i$. The number of visits to i which have occurred by time t is given by $N(t) = \max\{n : T_n \leq t\}$. ●

Both examples above contain a continuous-time random process $N = \{N(t) : t \geq 0\}$, where $N(t)$ represents the number of occurrences of some event in the time interval $[0, t)$. Such a process N is called a 'renewal' or 'counting' process for obvious reasons; the Poisson process of Section 6.8 provides another example of a renewal process.

(3) Definition. **A renewal process** $N = \{N(t) : t \geq 0\}$ **is a process for which**

$$N(t) = \max\{n : T_n \leq t\}$$

where

$$T_0 = 0, \quad T_n = X_1 + X_2 + \cdots + X_n \quad \text{for} \quad n \geq 1,$$

and the X_m are independent identically distributed non-negative random variables.

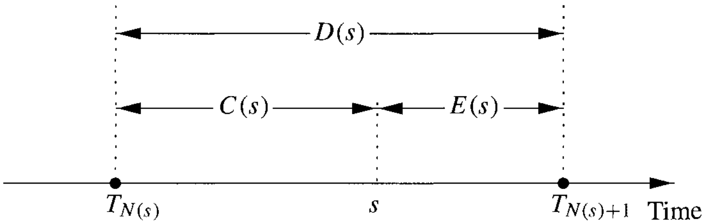

Figure 8.1. Excess, current, and total lifetimes at time s.

This definition describes N in terms of an underlying sequence $\{X_n\}$. In the absence of knowledge about this sequence we can construct it from N; just define

(4)
$$T_n = \inf\{t : N(t) = n\}, \quad X_n = T_n - T_{n-1}.$$

Note that the finite-dimensional distributions of a renewal process N are specified by the distribution of the X_m. For example, if the X_m are exponentially distributed then N is a Poisson process. We shall try to use the notation of (3) consistently in Chapter 10, in the sense that $\{N(t)\}$, $\{T_n\}$, and $\{X_n\}$ will always denote variables satisfying (4).

It is sometimes appropriate to allow X_1 to have a different distribution from the shared distribution of $X_2, X_3, \dots$; in this case N is called a *delayed* (or *modified*) renewal process. The process N in (2) is a delayed renewal process whatever the initial Y_0; if $Y_0 = i$ then N is an ordinary renewal process.

Those readers who paid attention to Claim (6.9.13) will be able to prove the following little result, which relates renewal processes to Markov chains.

(5) Theorem. *Poisson processes are the only renewal processes which are Markov chains.*

If you like, think of renewal processes as a generalization of Poisson processes in which we have dropped the condition that interarrival times be exponentially distributed.

There are two principal areas of interest concerning renewal processes. First, suppose that we interrupt a renewal process N at some specified time s. By this time, $N(s)$ occurrences have already taken place and we are awaiting the $(N(s) + 1)$th. That is, s belongs to the random interval

$$I_s = [T_{N(s)}, T_{N(s)+1}).$$

Here are definitions of three random variables of interest.

(6) The *excess* (or *residual*) *lifetime* of I_s: $E(s) = T_{N(s)+1} - s$.

(7) The *current lifetime* (or *age*) of I_s: $C(s) = s - T_{N(s)}$.

(8) The *total lifetime* of I_s: $D(s) = E(s) + C(s)$.

We shall be interested in the distributions of these random variables; they are illustrated in Figure 8.1.

It will come as no surprise to the reader to learn that the other principal topic concerns the asymptotic behaviour of a renewal process $N(t)$ as $t \to \infty$. Here we turn our attention to the *renewal function* $m(t)$ given by

(9)
$$m(t) = \mathbb{E}(N(t)).$$

For a Poisson process N with intensity λ, Theorem (6.8.2) shows that $m(t) = \lambda t$. In general, m is *not* a linear function of t; however, it is not too difficult to show that m is asymptotically linear, in that

$$\frac{1}{t}m(t) \to \frac{1}{\mu} \quad \text{as} \quad t \to \infty, \qquad \text{where} \quad \mu = \mathbb{E}(X_1).$$

The 'renewal theorem' is a refinement of this result and asserts that

$$m(t+h) - m(t) \to \frac{h}{\mu} \quad \text{as} \quad t \to \infty$$

subject to a certain condition on X_1.

An introduction to the theory of renewal processes is given in Chapter 10.

Exercises for Section 8.3

1. Let $(f_n : n \geq 1)$ be a probability distribution on the positive integers, and define a sequence $(u_n : n \geq 0)$ by $u_0 = 1$ and $u_n = \sum_{r=1}^{n} f_r u_{n-r}$, $n \geq 1$. Explain why such a sequence is called a *renewal sequence*, and show that u is a renewal sequence if and only if there exists a Markov chain U and a state s such that $u_n = \mathbb{P}(U_n = s \mid U_0 = s)$.

2. Let $\{X_i : i \geq 1\}$ be the inter-event times of a discrete renewal process on the integers. Show that the excess lifetime B_n constitutes a Markov chain. Write down the transition probabilities of the sequence $\{B_n\}$ when reversed in equilibrium. Compare these with the transition probabilities of the chain U of your solution to Exercise (1).

3. Let $(u_n : n \geq 1)$ satisfy $u_0 = 1$ and $u_n = \sum_{r=1}^{n} f_r u_{n-r}$ for $n \geq 1$, where $(f_r : r \geq 1)$ is a non-negative sequence. Show that:

(a) $v_n = \rho^n u_n$ is a renewal sequence if $\rho > 0$ and $\sum_{n=1}^{\infty} \rho^n f_n = 1$,

(b) as $n \to \infty$, $\rho^n u_n$ converges to some constant c.

4. Events occur at the times of a discrete-time renewal process N (see Example (5.2.15)). Let u_n be the probability of an event at time n, with generating function $U(s)$, and let $F(s)$ be the probability generating function of a typical inter-event time. Show that, if $|s| < 1$:

$$\sum_{r=0}^{\infty} \mathbb{E}(N(r))s^r = \frac{F(s)U(s)}{1-s} \quad \text{and} \quad \sum_{t=0}^{\infty} \mathbb{E}\left[\binom{N(t)+k}{k}\right] s^t = \frac{U(s)^k}{1-s} \quad \text{for } k \geq 0.$$

5. Prove Theorem (8.3.5): Poisson processes are the only renewal processes that are Markov chains.

8.4 Queues

The theory of queues is attractive and popular for two main reasons. First, queueing models are easily described and draw strongly from our intuitions about activities such as shopping or dialling a telephone operator. Secondly, even the solutions to the simplest models use much of the apparatus which we have developed in this book. Queues are, in general, non-Markovian, non-stationary, and quite difficult to study. Subject to certain conditions, however, their analysis uses ideas related to imbedded Markov chains, convergence of sequences of random variables, martingales, stationary processes, and renewal processes. We present a broad account of their theory in Chapter 11.

Customers arrive at a service point or counter at which a number of servers are stationed. [For clarity of exposition we have adopted the convention, chosen by the flip of a coin, that customers are male and servers are female.] An arriving customer may have to wait until one of these servers becomes available. Then he moves to the head of the queue and is served; he leaves the system on the completion of his service. We must specify a number of details about this queueing system before we are able to model it adequately. For example,

(a) in what manner do customers enter the system?
(b) in what order are they served?
(c) how long are their service times?

For the moment we shall suppose that the answers to these questions are as follows.

(a) The number $N(t)$ of customers who have entered by time t is a renewal process. That is, if T_n is the time of arrival of the nth customer (with the convention that $T_0 = 0$) then the *interarrival times* $X_n = T_n - T_{n-1}$ are independent and identically distributed.
(b) Arriving customers join the end of a single line of people who receive attention on a 'first come, first served' basis. There are a certain number of servers. When a server becomes free, she turns her attention to the customer at the head of the waiting line. We shall usually suppose that the queue has a single server only.
(c) Service times are independent identically distributed random variables. That is, if S_n is the service time of the nth customer to arrive, then $\{S_n\}$ is a sequence of independent identically distributed non-negative random variables which do not depend on the arriving stream N of customers.

It requires only a little imagination to think of various other systems. Here are some examples.

(1) *Queues with baulking.* If the line of waiting customers is long then an arriving customer may, with a certain probability, decide not to join it.

(2) *Continental queueing.* In the absence of queue discipline, unoccupied servers pick a customer at random from the waiting mêlée.

(3) *Airline check-in.* The waiting customers divide into several lines, one for each server. The servers themselves enter and leave the system at random, causing the attendant customers to change lines as necessary.

(4) *Last come, first served.* Arriving documents are placed on the top of an in-tray. An available server takes the next document from the top of the pile.

(5) *Group service.* Waiting customers are served in batches. This is appropriate for lift queues and bus queues.

(6) *Student discipline.* Arriving customers jump the queue, joining it near a friend

We shall consider mostly the single-server queues described by (a), (b), and (c) above. Such queues are specified by the distribution of a typical interarrival time and the distribution of a typical service time; the method of analysis depends partly upon how much information we have about these quantities.

The state of the queue at time t is described by the number $Q(t)$ of waiting customers ($Q(t)$ *includes* customers who are in the process of being served at this time). It would be unfortunate if $Q(t) \to \infty$ as $t \to \infty$, and we devote special attention to finding out when this occurs. We call a queue *stable* if the distribution of $Q(t)$ settles down as $t \to \infty$ in some well-behaved way; otherwise we call it *unstable*. We choose not to define stability more precisely at this stage, wishing only to distinguish between such extremes as

(a) queues which either grow beyond all bounds or enjoy large wild fluctuations in length,

(b) queues whose lengths converge in distribution, as $t \to \infty$, to some 'equilibrium distribution'.

Let S and X be a typical service time and a typical interarrival time, respectively; the ratio

$$\rho = \frac{\mathbb{E}(S)}{\mathbb{E}(X)}$$

is called the *traffic intensity*.

(7) Theorem. *Let $Q = \{Q(t) : t \geq 0\}$ be a queue with a single server and traffic intensity ρ.*

(a) *If $\rho < 1$ then Q is stable.*

(b) *If $\rho > 1$ then Q is unstable.*

(c) *If $\rho = 1$ and at least one of S and X has strictly positive variance then Q is unstable.*

The conclusions of this theorem are intuitively very attractive. Why?

A more satisfactory account of this theorem is given in Section 11.5.

Exercises for Section 8.4

1. The two tellers in a bank each take an exponentially distributed time to deal with any customer; their parameters are λ and μ respectively. You arrive to find exactly two customers present, each occupying a teller.

(a) You take a fancy to a randomly chosen teller, and queue for that teller to be free; no later switching is permitted. Assuming any necessary independence, what is the probability p that you are the last of the three customers to leave the bank?

(b) If you choose to be served by the quicker teller, find p.

(c) Suppose you go to the teller who becomes free first. Find p.

2. Customers arrive at a desk according to a Poisson process of intensity λ. There is one clerk, and the service times are independent and exponentially distributed with parameter μ. At time 0 there is exactly one customer, currently in service. Show that the probability that the next customer arrives before time t and finds the clerk busy is

$$\frac{\lambda}{\lambda + \mu}(1 - e^{-(\lambda+\mu)t}).$$

3. Vehicles pass a crossing at the instants of a Poisson process of intensity λ; you need a gap of length at least a in order to cross. Let T be the first time at which you could succeed in crossing to the other side. Show that $\mathbb{E}(T) = (e^{a\lambda} - 1)/\lambda$, and find $\mathbb{E}(e^{\theta T})$.

Suppose there are two lanes to cross, carrying independent Poissonian traffic with respective rates λ and μ. Find the expected time to cross in the two cases when: (a) there is an island or refuge between the two lanes, (b) you must cross both in one go. Which is the greater?

4. Customers arrive at the instants of a Poisson process of intensity λ, and the single server has exponential service times with parameter μ. An arriving customer who sees n customers present (including anyone in service) will join the queue with probability $(n+1)/(n+2)$, otherwise leaving for ever. Under what condition is there a stationary distribution? Find the mean of the time spent in the queue (not including service time) by a customer who joins it when the queue is in equilibrium. What is the probability that an arrival joins the queue when in equilibrium?

5. Customers enter a shop at the instants of a Poisson process of rate 2. At the door, two representatives separately demonstrate a new corkscrew. This typically occupies the time of a customer and the representative for a period which is exponentially distributed with parameter 1, independently of arrivals and other demonstrators. If both representatives are busy, customers pass directly into the shop. No customer passes a free representative without being stopped, and all customers leave by another door. If both representatives are free at time 0, show the probability that both are busy at time t is $\frac{2}{5} - \frac{2}{3}e^{-2t} + \frac{4}{15}e^{-5t}$.

8.5 The Wiener process

Most of the random processes considered so far are 'discrete' in the sense that they take values in the integers or in some other countable set. Perhaps the simplest example is simple random walk $\{S_n\}$, a process which jumps one unit to the left or to the right at each step. This random walk $\{S_n\}$ has two interesting and basic properties:

(a) *time-homogeneity*, in that, for all non-negative m and n, S_m and $S_{m+n} - S_n$ have the same distribution (we assume $S_0 = 0$); and

(b) *independent increments*, in that the increments $S_{n_i} - S_{m_i}$ $(i \geq 1)$ are independent whenever the intervals $(m_i, n_i]$ are disjoint.

What is the 'continuous' analogue of this random walk? It is reasonable to require that such a 'continuous' random process has the two properties above, and it turns out that, subject to some extra assumptions about means and variances, there is essentially only one such process which is called the *Wiener process*. This is a process $W = \{W(t) : t \geq 0\}$, indexed by continuous time and taking values in the real line $\mathbb{R}$, which is time-homogeneous with independent increments, and with the vital extra property that $W(t)$ has the normal distribution with mean 0 and variance $\sigma^2 t$ for some constant σ^2. This process is sometimes called *Brownian motion*, and is a cornerstone of the modern theory of random processes. Think about it as a model for a particle which diffuses randomly along a line. There is no difficulty in constructing Wiener processes in higher dimensions, leading to models for such processes as the Dow–Jones index or the diffusion of a gas molecule in a container. Note that $W(0) = 0$; the definition of a Wiener process may be easily extended to allow more general starting points.

What are the finite-dimensional distributions of the Wiener process W? These are easily calculated as follows.

(1) Lemma. *The vector of random variables $W(t_1), W(t_1), \dots, W(t_n)$ has the multivariate normal distribution with zero means and covariance matrix (v_{ij}) where $v_{ij} = \sigma^2 \min\{t_i, t_j\}$.*

Proof. By assumption, $W(t_i)$ has the normal distribution with zero mean and variance $\sigma^2 t_i$. It therefore suffices to prove that $\text{cov}(W(s), W(t)) = \sigma^2 \min\{s, t\}$. Now, if $s < t$, then

$$\mathbb{E}\big(W(s)W(t)\big) = \mathbb{E}\big(W(s)^2 + W(s)\big[W(t) - W(s)\big]\big) = \mathbb{E}(W(s)^2) + 0,$$

since W has independent increments and $\mathbb{E}(W(s)) = 0$. Hence

$$\text{cov}\big(W(s), W(t)\big) = \text{var}(W(s)) = \sigma^2 s. \qquad \blacksquare$$

A Wiener process W is called *standard* if $W(0) = 0$ and $\sigma^2 = 1$. A more extended treatment of Wiener processes appears in Chapter 13.

Exercises for Section 8.5

1. For a Wiener process W with $W(0) = 0$, show that

$$\mathbb{P}\big(W(s) > 0,\ W(t) > 0\big) = \frac{1}{4} + \frac{1}{2\pi}\sin^{-1}\sqrt{\frac{s}{t}} \quad \text{for } s < t.$$

Calculate $\mathbb{P}(W(s) > 0,\ W(t) > 0,\ W(u) > 0)$ when $s < t < u$.

2. Let W be a Wiener process. Show that, for $s < t < u$, the conditional distribution of $W(t)$ given $W(s)$ and $W(u)$ is normal

$$N\left(\frac{(u-t)W(s) + (t-s)W(u)}{u-s}, \frac{(u-t)(t-s)}{u-s}\right).$$

Deduce that the conditional correlation between $W(t)$ and $W(u)$, given $W(s)$ and $W(v)$, where $s < t < u < v$, is

$$\sqrt{\frac{(v-u)(t-s)}{(v-t)(u-s)}}.$$

3. For what values of a and b is $aW_1 + bW_2$ a standard Wiener process, where W_1 and W_2 are independent standard Wiener processes?

4. Show that a Wiener process W with variance parameter σ^2 has finite quadratic variation, which is to say that

$$\sum_{j=0}^{n-1}\big\{W\big((j+1)t/n\big) - W(jt/n)\big\}^2 \xrightarrow{\text{m.s.}} \sigma^2 t \quad \text{as } n \to \infty.$$

5. Let W be a Wiener process. Which of the following define Wiener processes?
(a) $-W(t)$, (b) $\sqrt{t}W(1)$, (c) $W(2t) - W(t)$.

8.6 Existence of processes

In our discussions of the properties of random variables, only scanty reference has been made to the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$; indeed we have felt some satisfaction and relief from this omission. We have often made assumptions about hypothetical random variables without even checking that such variables exist. For example, we are in the habit of making statements such as 'let $X_1, X_2, \ldots$ be independent variables with common distribution function F', but we have made no effort to show that there exists some probability space on which such variables can be constructed. The foundations of such statements require examination. It is the purpose of this section to indicate that our assumptions are fully justifiable.

First, suppose that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and that $X = \{X_t : t \in T\}$ is some collection of random variables mapping Ω into $\mathbb{R}$. We saw in Section 8.1 that to any vector $\mathbf{t} = (t_1, t_2, \ldots, t_n)$ containing members of T and of finite length there corresponds a joint distribution function $F_{\mathbf{t}}$; the collection of such functions $F_{\mathbf{t}}$, as $\mathbf{t}$ ranges over all possible vectors of any length, is called the set of *finite-dimensional distributions*, or *fdds*, of X. It is clear that these distribution functions satisfy the two *Kolmogorov consistency conditions*:

(1) $F_{(t_1,\ldots,t_n,t_{n+1})}(x_1, \ldots, x_n, x_{n+1}) \to F_{(t_1,\ldots,t_n)}(x_1, \ldots, x_n)$ as $x_{n+1} \to \infty$,

(2) if π is a permutation of $(1, 2, \ldots, n)$ and $\pi\mathbf{y}$ denotes the vector $\pi\mathbf{y} = (y_{\pi(1)}, \ldots, y_{\pi(n)})$ for any n-vector $\mathbf{y}$, then $F_{\pi\mathbf{t}}(\pi\mathbf{x}) = F_{\mathbf{t}}(\mathbf{x})$ for all $\mathbf{x}$, $\mathbf{t}$, π, and n.

Condition (1) is just a higher-dimensional form of (2.1.6a), and condition (2) says that the operation of permuting the X_t has the obvious corresponding effect on their joint distributions. So fdds always satisfy (1) and (2); furthermore (1) and (2) characterize fdds.

(3) Theorem. *Let T be any set, and suppose that to each vector $\mathbf{t} = (t_1, t_2, \ldots, t_n)$ containing members of T and of finite length, there corresponds a joint distribution function $F_{\mathbf{t}}$. If the collection $\{F_{\mathbf{t}}\}$ satisfies the Kolmogorov consistency conditions, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a collection $X = \{X_t : t \in T\}$ of random variables on this space such that $\{F_{\mathbf{t}}\}$ is the set of fdds of X.*

The proof of this result lies in the heart of measure theory, as the following sketch indicates.

Sketch proof. Let $\Omega = \mathbb{R}^T$, the product of T copies of $\mathbb{R}$; the points of Ω are collections $\mathbf{y} = \{y_t : t \in T\}$ of real numbers. Let $\mathcal{F} = \mathcal{B}^T$, the σ-field generated by subsets of the form $\prod_{t \in T} B_t$ for Borel sets B_t all but finitely many of which equal $\mathbb{R}$. It is a fundamental result in measure theory that there exists a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ such that

$$\mathbb{P}\Big(\big\{\mathbf{y} \in \Omega : y_{t_1} \le x_1,\ y_{t_2} \le x_2, \ldots,\ y_{t_n} \le x_n\big\}\Big) = F_{\mathbf{t}}(\mathbf{x})$$

for all $\mathbf{t}$ and $\mathbf{x}$; this follows by an extension of the argument of Section 1.6. Then $(\Omega, \mathcal{F}, \mathbb{P})$ is the required space. Define $X_t : \Omega \to \mathbb{R}$ by $X_t(\mathbf{y}) = y_t$ to obtain the required family $\{X_t\}$. ■

We have seen that the fdds are characterized by the consistency conditions (1) and (2). But how much do they tell us about the sample paths of the corresponding process X? A simple example is enough to indicate some of the dangers here.

(4) Example. Let U be a random variable which is uniformly distributed on $[0, 1]$. Define two processes $X = \{X_t : 0 \le t \le 1\}$ and $Y = \{Y_t : 0 \le t \le 1\}$ by

$$X_t = 0 \quad \text{for all } t, \qquad Y_t = \begin{cases} 1 & \text{if } U = t, \\ 0 & \text{otherwise.} \end{cases}$$

Clearly X and Y have the same fdds, since $\mathbb{P}(U = t) = 0$ for all t. But X and Y are different processes. In particular $\mathbb{P}(X_t = 0 \text{ for all } t) = 1$ and $\mathbb{P}(Y_t = 0 \text{ for all } t) = 0$. ●

One may easily construct less trivial examples of different processes having the same fdds; such processes are called *versions* of one another. This complication should not be overlooked with a casual wave of the hand; it is central to any theory which attempts to study properties of sample paths, such as first-passage times. As the above example illustrates, such properties are not generally specified by the fdds, and their validity may therefore depend on which version of the process is under study.

For the random process $\{X(t) : t \in T\}$, where $T = [0, \infty)$ say, knowledge of the fdds amounts to being given a probability space of the form $(\mathbb{R}^T, \mathcal{B}^T, \mathbb{P})$, as in the sketch proof of (3) above. Many properties of sample paths do not correspond to events in $\mathcal{B}^T$. For example, the subset of Ω given by $A = \{\omega \in \Omega : X(t) = 0 \text{ for all } t \in T\}$ is an *uncountable* intersection of events $A = \bigcap_{t \in T}\{X(t) = 0\}$, and may not itself be an event. Such difficulties would be avoided if all sample paths of X were continuous, since then A is the intersection of $\{X(t) = 0\}$ over all *rational* $t \in T$; this is a *countable* intersection.

(5) Example. Let W be the Wiener process of Section 8.5, and let T be the time of the first passage of W to the point 1, so that $S = \inf\{t : W(t) = 1\}$. Then

$$\{S > t\} = \bigcap_{0 \le s \le t} \{W(s) \ne 1\}$$

is a set of configurations which does not belong to the Borel σ-field $\mathcal{B}^{[0,\infty)}$. If all sample paths of W were continuous, one might write

$$\{S > t\} = \bigcap_{\substack{0 \le s \le t \\ s \text{ rational}}} \{W(s) \ne 1\},$$

the countable intersection of events. As the construction of Example (4) indicates, there are versions of the Wiener process which have discontinuous sample paths. One of the central results of Chapter 13 is that there exists a version with continuous sample paths, and it is with this version that one normally works. ●

It is too restrictive to require continuity of sample paths in general; after all, processes such as the Poisson process most definitely do not have continuous sample paths. The most which can be required is continuity from either the left or the right. Following a convention, we go for the latter here. Under what conditions may one assume that there exists a version with right-continuous sample paths? An answer is provided by the next theorem; see Breiman (1968, p. 300) for a proof.

(6) Theorem. *Let $\{X(t) : t \ge 0\}$ be a real-valued random process. Let D be a subset of $[0, \infty)$ which is dense in $[0, \infty)$. If:*

(i) *X is continuous in probability from the right, that is, $X(t + h) \xrightarrow{\text{P}} X(t)$ as $h \downarrow 0$, for all t, and*

(ii) *at any accumulation point a of D, X has finite right and left limits with probability 1, that is $\lim_{h \downarrow 0} X(a + h)$ and $\lim_{h \uparrow 0} X(a + h)$ exist, a.s.,*

then there exists a version Y of X such that:

(a) *the sample paths of Y are right-continuous,*

(b) *Y has left limits, in that $\lim_{h \uparrow 0} Y(t + h)$ exists for all t.*

In other words, if (i) and (ii) hold, there exists a probability space and a process Y defined on this space, such that Y has the same fdds as X in addition to properties (a) and (b). A process which is right-continuous with left limits is called càdlàg by some (largely French speakers), and a Skorokhod map or R-process by others.

8.7 Problems

1. Let $\{Z_n\}$ be a sequence of uncorrelated real-valued variables with zero means and unit variances, and define the 'moving average'

$$Y_n = \sum_{i=0}^{r} \alpha_i Z_{n-i},$$

for constants $\alpha_0, \alpha_1, \dots, \alpha_r$. Show that Y is stationary and find its autocovariance function.

2. Let $\{Z_n\}$ be a sequence of uncorrelated real-valued variables with zero means and unit variances. Suppose that $\{Y_n\}$ is an 'autoregressive' stationary sequence in that it satisfies $Y_n = \alpha Y_{n-1} + Z_n$, $-\infty < n < \infty$, for some real α satisfying $|\alpha| < 1$. Show that Y has autocovariance function $c(m) = \alpha^{|m|}/(1 - \alpha^2)$.

3. Let $\{X_n\}$ be a sequence of independent identically distributed Bernoulli variables, each taking values 0 and 1 with probabilities $1 - p$ and p respectively. Find the mass function of the renewal process $N(t)$ with interarrival times $\{X_n\}$.

4. Customers arrive in a shop in the manner of a Poisson process with parameter λ. There are infinitely many servers, and each service time is exponentially distributed with parameter μ. Show that the number $Q(t)$ of waiting customers at time t constitutes a birth–death process. Find its stationary distribution.

5. Let $X(t) = Y\cos(\theta t) + Z\sin(\theta t)$ where Y and Z are independent $N(0, 1)$ random variables, and let $\widetilde{X}(t) = R\cos(\theta t + \Psi)$ where R and Ψ are independent. Find distributions for R and Ψ such that the processes X and $\widetilde{X}$ have the same fdds.

6. **Bartlett's theorem.** Customers arrive at the entrance to a queueing system at the instants of an inhomogeneous Poisson process with rate function $\lambda(t)$. Their subsequent service histories are independent of each other, and a customer arriving at time s is in state A at time $s + t$ with probability $p(s, t)$. Show that the number of customers in state A at time t is Poisson with parameter $\int_{-\infty}^{t} \lambda(u)p(u, t - u)\, du$.

7. In a Prague teashop (U Myšáka), long since bankrupt, customers queue at the entrance for a blank bill. In the shop there are separate counters for coffee, sweetcakes, pretzels, milk, drinks, and ice cream, and queues form at each of these. At each service point the customers' bills are marked appropriately. There is a restricted number N of seats, and departing customers have to queue in order to pay their bills. If interarrival times and service times are exponentially distributed and the process is in equilibrium, find how much longer a greedy customer must wait if he insists on sitting down. Answers on a postcard to the authors, please.