3

Probability models for time series

3.1 STOCHASTIC PROCESSES

This chapter describes various probability models for time series, which are collectively called stochastic processes. Most physical processes in the real world involve a random element in their structure, and a **stochastic process** can be described as 'a statistical phenomenon that evolves in time according to probabilistic laws'. Well-known examples are the length of a queue, the size of a bacterial colony, and the air temperature on successive days at a particular site. The word 'stochastic', which is of Greek origin, is used to mean 'pertaining to chance', and many writers use 'random process' as a synonym for stochastic process.

Mathematically, a stochastic process may be defined as a collection of random variables which are ordered in time and defined at a set of time points which may be continuous or discrete. We will denote the random variable at time t by $X(t)$ if time is continuous (usually $-\infty < t < \infty$), and by X_t if time is discrete (usually $t=0, \pm 1, \pm 2, \ldots$).

The theory of stochastic processes has been extensively developed and is discussed in many books including Papoulis (1984), written primarily for engineers, Parzen (1962), Cox and Miller (1968, especially Chapter 7), Yaglom (1962) and Grimmett and Stirzaker (1992). In this chapter we concentrate on those aspects particularly relevant to time-series analysis.

Most statistical problems are concerned with estimating the properties of a population from a sample. In time-series analysis there is a rather different situation in that, although it may be possible to vary the **length** of the observed time series – the sample – it is usually impossible to make more than one observation at any given time. Thus we only have a single outcome of the process and a single observation on the random variable at time t. Nevertheless we may regard the observed time series as just one example of the infinite set of time series which might have been observed. This infinite set of time series is sometimes called the **ensemble**. Every member of the ensemble is a possible **realization** of the stochastic process. The observed time series can be thought of as one particular realization, and will be denoted by $x(t)$ for

$(0 \leqslant t \leqslant T)$ if observations are continuous, and by x_t for $t=1, \ldots, N$ if observations are discrete.

Because there is only a notional population, time-series analysis is essentially concerned with evaluating the properties of the probability model which generated the observed time series.

One way of describing a stochastic process is to specify the joint probability distribution of $X(t_1), \ldots, X(t_n)$ for any set of times $t_1, \ldots, t_n$ and any value of n. But this is rather complicated and is not usually attempted in practice. A simpler, more useful way of describing a stochastic process is to give the **moments** of the process, particularly the first and second moments, which are called the mean, variance and autocovariance functions. These will now be defined for continuous time, with similar definitions applying in discrete time.

Mean The mean function $\mu(t)$ is defined by

$$\mu(t) = E[X(t)]$$

Variance The variance function $\sigma^2(t)$ is defined by

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

Autocovariance The variance function alone is not enough to specify the second moments of a sequence of random variables. In addition, we must define the autocovariance function $\gamma(t_1, t_2)$, which is the covariance of $X(t_1)$ with $X(t_2)$, namely

$$\gamma(t_1, t_2) = E\{[X(t_1) - \mu(t_1)][X(t_2) - \mu(t_2)]\}$$

(Readers who are unfamiliar with the term 'covariance' should read Appendix C. When applied to a sequence of random variables, it is called an autocovariance.) Note that the variance function is a special case of the autocovariance function when $t_1 = t_2$.

Higher moments of a stochastic process may be defined in an obvious way, but are rarely used in practice, since a knowledge of the two functions $\mu(t)$ and $\gamma(t_1, t_2)$ is usually adequate.

3.2 STATIONARY PROCESSES

An important class of stochastic processes are those which are stationary. A heuristic idea of stationarity was introduced in Section 2.2.

A time series is said to be **strictly stationary** if the joint distribution of $X(t_1), \ldots, X(t_n)$ is the same as the joint distribution of $X(t_1 + \tau), \ldots, X(t_n + \tau)$ for all $t_1, \ldots, t_n, \tau$. In other words, shifting the time origin by an amount τ has no effect on the joint distributions, which must therefore depend only on the intervals between $t_1, t_2, \ldots, t_n$. The above definition holds for any value of n.

In particular, if $n = 1$, strict stationarity implies that the distribution of $X(t)$ is the same for all t, so that, provided the first two moments are finite, we have

$$\mu(t)=\mu$$

$$\sigma^2(t)=\sigma^2$$

are both constants which do not depend on the value of t.

Furthermore, if $n=2$ the joint distribution of $X(t_1)$ and $X(t_2)$ depends only on (t_2-t_1), which is called the **lag**. Thus the autocovariance function $\gamma(t_1, t_2)$ also depends only on (t_2-t_1) and may be written as $\gamma(\tau)$, where

$$\begin{aligned}\gamma(\tau)&=E\{[X(t)-\mu][X(t+\tau)-\mu]\}\\&=\text{Cov}[X(t), X(t+\tau)]\end{aligned}$$

is called the autocovariance coefficient at lag τ. In future, 'autocovariance function' will be abbreviated to acv.f.

The size of an autocovariance coefficient depends on the units in which $X(t)$ is measured. Thus, for interpretative purposes, it is useful to standardize the acv.f. to produce a function called the **autocorrelation** function, which is given by

$$\rho(\tau)=\gamma(\tau)/\gamma(0)$$

and which measures the correlation between $X(t)$ and $X(t+\tau)$. Its empirical counterpart was introduced in Section 2.7. In future, 'autocorrelation function' will be abbreviated to ac.f. Note that the argument τ of $\gamma(\tau)$ and $\rho(\tau)$ is discrete if the time series is discrete and continuous if the time series is continuous.

At first sight it may seem surprising to suggest that there are processes for which the distribution of $X(t)$ should be the same for all t. However, readers with some knowledge of stochastic processes will know that there are many processes $\{X(t)\}$ which have what is called an **equilibrium** distribution as $t\to\infty$, in which the probability distribution of $X(t)$ tends to a limit which does **not** depend on the initial conditions. Thus once such a process has been running for some time, the distribution of $X(t)$ will change very little. Indeed if the initial conditions are specified to be identical to the equilibrium distribution, the process is stationary in time and the equilibrium distribution is then the stationary distribution of the process. Of course the **conditional** distribution of $X(t_2)$ given that $X(t_1)$ has taken a particular value, say $x(t_1)$, may be quite different from the stationary distribution, but this is perfectly consistent with the process being stationary.

3.2.1 Second-order stationarity

In practice it is often useful to define stationarity in a less restricted way than that described above. A process is called second-order stationary (or weakly stationary) if its mean is constant and its acv.f. depends only on the lag, so that

$$E[X(t)] = \mu$$

and

$$\mathrm{Cov}[X(t), X(t+\tau)] = \gamma(\tau)$$

No assumptions are made about higher moments than those of second order. By letting $\tau = 0$, we note that the above assumption about the acv.f. implies that the variance, as well as the mean, is constant. Also note that both the variance and the mean must be finite.

This weaker definition of stationarity will generally be used from now on, as many of the properties of stationary processes depend only on the structure of the process as specified by its first and second moments. One important class of processes where this is particularly true is the class of **normal** processes where the joint distribution of $X(t_1), \ldots, X(t_n)$ is multivariate normal for all $t_1, \ldots, t_n$. The multivariate normal distribution is completely characterized by its first and second moments, and hence by $\mu(t)$ and $\gamma(t_1, t_2)$, and so it follows that second-order stationarity implies strict stationarity for normal processes. However, μ and $\gamma(\tau)$ may not adequately describe processes which are very 'non-normal'.

3.3 THE AUTOCORRELATION FUNCTION

We have already noted in Section 2.7 that the sample autocorrelation coefficients of an observed time series are an important set of statistics for describing the time series. Similarly the (theoretical) autocorrelation function (ac.f.) of a stationary stochastic process is an important tool for assessing its properties. This section investigates the general properties of the ac.f.

Suppose a stationary stochastic process $X(t)$ has mean μ, variance σ^2, acv.f. $\gamma(\tau)$, and ac.f. $\rho(\tau)$. Then

$$\rho(\tau) = \gamma(\tau)/\gamma(0) = \gamma(\tau)/\sigma^2$$

Note that $\rho(0) = 1$.

Property 1

The ac.f. is an **even** function of the lag in that

$$\rho(\tau) = \rho(-\tau)$$

This property simply says that the correlation between $X(t)$ and $X(t+\tau)$ is the same as that between $X(t)$ and $X(t-\tau)$. The result is easily proved using $\gamma(\tau) = \rho(\tau)\sigma^2$ by

$$\gamma(\tau) = \mathrm{Cov}[X(t), X(t+\tau)]$$

$$= \mathrm{Cov}[X(t-\tau), X(t)] \qquad \text{since } X(t) \text{ stationary}$$
$$= \gamma(-\tau)$$

Property 2

$|\rho(\tau)| \leqslant 1$. This is the 'usual' property of a correlation. It is proved by noting that

$$\mathrm{Var}[\lambda_1 X(t) + \lambda_2 X(t+\tau)] \geqslant 0$$

for any constants λ_1, λ_2, since a variance is always non-negative. This variance is equal to

$$\lambda_1^2 \mathrm{Var}[X(t)] + \lambda_2^2 \mathrm{Var}[X(t+\tau)] + 2\lambda_1\lambda_2 \mathrm{Cov}[X(t), X(t+\tau)]$$
$$= (\lambda_1^2 + \lambda_2^2)\sigma^2 + 2\lambda_1\lambda_2\gamma(\tau)$$

When $\lambda_1 = \lambda_2 = 1$, we find

$$\gamma(\tau) \geqslant -\sigma^2$$

so that $\rho(\tau) \geqslant -1$. When $\lambda_1 = 1$, $\lambda_2 = -1$, we find

$$\sigma^2 \geqslant \gamma(\tau)$$

so that $\rho(\tau) \leqslant +1$.

Property 3

Lack of uniqueness. Although a given stochastic process has a unique covariance structure, the converse is not in general true. It is usually possible to find many normal and non-normal processes with the same ac.f. and this creates further difficulty in interpreting sample ac.f.s. Jenkins and Watts (1968, p. 170) give an example of two different stochastic processes which have the same ac.f. Even for stationary normal processes, which are completely determined by the mean, variance and ac.f., the invertibility condition introduced in Section 3.4.3 is required to ensure uniqueness.

3.4 SOME USEFUL STOCHASTIC PROCESSES

This section describes several different types of stochastic process which are sometimes useful in setting up a model for a time series.

3.4.1 A purely random process

A discrete-time process is called a purely random process if it consists of a sequence of random variables $\{Z_t\}$ which are mutually independent and identically distributed. From the definition it follows that the process has constant mean and variance and that

$$\gamma(k) = \text{Cov}(Z_t, Z_{t+k})$$
$$= 0 \qquad \text{for } k = \pm 1, 2, \ldots$$

As the mean and acv.f. do not depend on time, the process is second-order stationary. In fact it is clear that the process is also strictly stationary. The ac.f. is given by

$$\rho(k) = \begin{cases} 1 & k = 0 \\ 0 & k = \pm 1, \pm 2, \ldots \end{cases}$$

A purely random process is sometimes called **white noise**, particularly by engineers. Processes of this type are useful in many situations, particularly as building blocks for more complicated processes such as moving average processes (Section 3.4.3).

The possibility of defining a continuous-time purely random process is discussed in Section 3.4.8.

3.4.2 Random walk

Suppose that $\{Z_t\}$ is a discrete, purely random process with mean μ and variance σ_Z^2. A process $\{X_t\}$ is said to be a random walk if

$$X_t = X_{t-1} + Z_t \tag{3.1}$$

The process is customarily started at zero when $t = 0$, so that

$$X_1 = Z_1$$

and

$$X_t = \sum_{i=1}^{t} Z_i$$

Then we find that $E(X_t) = t\mu$ and that $\text{Var}(X_t) = t\sigma_Z^2$. As the mean and variance change with t, the process is non-stationary.

However, it is interesting to note that the first differences of a random walk, given by

$$\nabla X_t = X_t - X_{t-1} = Z_t$$

form a purely random process, which is therefore stationary.

The best-known examples of time series which behave like random walks are share prices on successive days. A model which often gives a good approximation to such data is

share price on day t = share price on day $(t-1)$ + random error