NON-LINEAR AUTOREGRESSIVE PROCESSES

by

David Alan Jones

Thesis submitted for the
Diploma of Membership of Imperial College
and the degree of
Doctor of Philosophy of the University of London

September, 1976

ABSTRACT

Models of the form $X_{n+1} = \lambda(X_n) + Z_{n+1}$ are considered for time-series $\{X_n\}$ where $\{Z_n\}$ is an impulse sequence. Some conditions for stationarity and non-stationarity of such processes are given and it is shown by simulated realisations that these processes extend the range of behaviour available with linear autoregressive-moving average models. Methods for approximating the stationary distribution of a process are considered and expressions are found by which the exact moments, joint moments and densities of the stationary processes can be obtained in a wide range of cases. Moments and densities of conditional distributions useful for prediction are also found. The results of these methods have been verified by computer simulations and these and other numerical results are given. The methods found can be extended directly to deal with multidimensional processes of the above form and this is discussed briefly.

ACKNOWLEDGEMENTS

I am grateful to Professor D.R. Cox for suggesting this topic and for his encouragement and advice. I should also like to thank the Science Research Council for their financial support during this research.

TABLE OF CONTENTS

1 INTRODUCTION

The models for a discrete-time time-series $\{X_i\}$ considered in this work are of the form

$$X_{n+1} = \lambda(X_n) + Z_{n+1} \qquad (n = \ldots -1,0,1,\ldots) \tag{1.1}$$

where $\lambda(x)$ is a function of a real variable and where the series $\{Z_n\}$ is a sequence of independent and identically distributed random variables. This is a generalisation of the first-order linear autoregressive model

$$X_{n+1} - \mu = \phi_1(X_n - \mu) + Z_{n+1} \qquad (n = \ldots -1,0,1,\ldots), \tag{1.2}$$

where $\mu = E(X_n)$, $E(Z_n) = 0$ and $|\phi_1| < 1$, and for which

$$E(X_{n+1} | X_n = x) = \mu + \phi_1(x - \mu) . \tag{1.3}$$

One reason for considering models of the above form is that, although the linear form of (1.3) may be a good approximation for a particular real system, it is unlikely that it should hold exactly everywhere. It is therefore of interest to be able to treat situations in which this response is not necessarily linear and the model (1.1) is one way in which this can be done: it retains the constant variance, additive error structure of (1.2).

In the above case functions $\lambda(x)$ which are "nearly linear" would be of most interest although there is no reason to restrict attention to such functions. By allowing functions further from linearity, some interesting types of behaviour can be obtained for realisations of the process (1.1). Some indication of this behaviour can be derived by consideration of the corresponding deterministic sequence $\{x_n\}$ generated by

$$x_{n+1} = \lambda(x_n) \qquad (n = 0,1,2,\ldots). \tag{1.4}$$

For these deterministic sequences the roots of the equations

$$x = \lambda(x), \quad x = \lambda\{\lambda(x)\}, \quad x = \lambda[\lambda\{\lambda(x)\}], \text{ etc.},$$

play a large role in the behaviour of the sequence. A root x^* of the equation $x = \lambda(x)$ is called stable if the slope of $\lambda(x)$ at x^* is strictly less than one in absolute value, for then the sequence (1.4) starting from a value in a neighbourhood of x^* satisfies $x_n \to x^*$ as $n \to \infty$. A root of $x = \lambda(x)$ at which the slope of λ is greater than one is unstable in that a sequence starting close to, but not exactly at, the root will tend to move away from it. Assuming that the sequence $\{Z_n\}$ appearing in (1.1) can take any real value but with a distribution concentrated in a region about zero, it is apparent that the behaviour of the process $\{X_n\}$ can be described by saying that it remains in a region close to a stable root of the equation $x = \lambda(x)$ until a large enough value of the input sequence occurs when the process shifts to a location about a new stable root. However there is also the possibility of stable periodic cycles for the deterministic process and these would also have their analogues for the process (1.1). The types of behaviour that can be obtained from a deterministic process of the form (1.4) are very complex as has been described by May (1976), however in simple cases a general description of the behaviour of the stochastic process (1.1) can be given as above. There need not be a root of the equation $x = \lambda(x)$ (or of $x = \lambda\{\lambda(x)\}$, etc.) in a region for that region to be given fairly large weight in the marginal distribution of the stochastic process: it is enough that the line $y = \lambda(x)$ should be close to the line $y = x$ for then, once that region is reached, the next few values of the process will also be in that region since $X_{n+1} \simeq \lambda(X_n) \simeq X_n$ with large probability.

Some computer realisations of particular processes of the form (1.1) have been obtained and graphical representations of these are given in Chapter 6. Here the sequences are relatively short (151 values) but some of the above type of behaviour can be observed. More complicated types of behaviour would require longer realisations to be discernable. Figure 1.1 gives a realisation of the process (1.1) for which the distribution of the input sequence is normal with zero mean and standard deviation one half, and for which

$$\lambda(x) = \begin{cases} -3.0 & (x \leq -3.5 \text{ or } x > 3.5), \\ x + 0.5 & (-3.5 < x \leq 3.5). \end{cases}$$

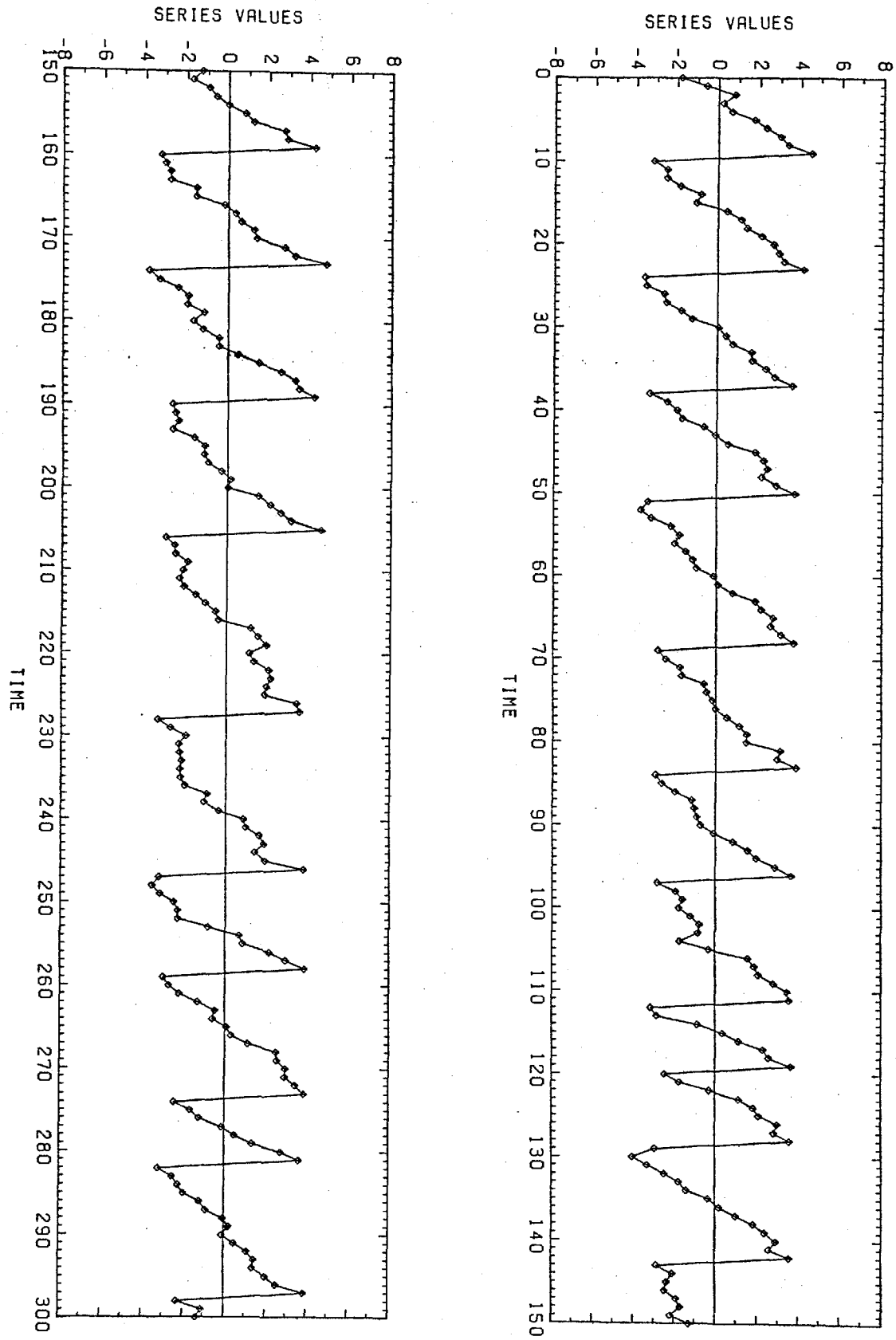

Figure 1.1: Realisation of autoregressive process with

$\lambda(x) = x + 0.5$ $(-3.5 < x \leq 3.5)$, $= -3.0$ (otherwise), $Z_n \sim N(0, \frac{1}{4})$.

It is fairly clear that the model (1.1) can give rise to processes which could not be modelled by the more usual linear autoregressive-moving average processes: this could be done only as far as second-order moment properties are concerned. Therefore, at the very least, these processes give a further wide class of models which could be fitted to approximate the behaviour of some real data series.

A process of the form (1.1) will be called a (non-linear) autoregressive process and the function $\lambda(x)$ will be called the autoregression function of the process. It is not assumed that the distribution of $\{Z_n\}$ is centred at zero. An associated process which may sometimes be more useful is the series $\{Y_n\}$ given by $Y_n = \lambda(X_n)$ for which the following generating equation holds

$$Y_{n+1} = \lambda(Y_n + Z_{n+1}) \qquad (n = \ldots -1,0,1,\ldots), \tag{1.5}$$

and $X_{n+1} = Y_n + Z_{n+1}$. The processes $\{X_n\}$, $\{Y_n\}$ are equivalent in that the distributional properties of one can be found from the other.

An example of a model of the form (1.5) has been given by Feller (1971, p.208). One version of this is that individual pieces of planking are cut in such a way that the angle between two supposedly parallel end faces is a random variable θ, independently and identically distributed for each plank. Suppose that in forming a length of planking the $(n + 1)$'st piece of planking is butted on to the preceding piece in such a way that its contribution, $\pm\theta_{n+1}$, to the angle between the first and last faces of the total length of planking is of opposite sign to that of the total accumulated angle. If the absolute value of this angle is ψ_n after n planks, then $\psi_1 = |\theta_1|$ and

$$\psi_{n+1} = \left|\psi_n - |\theta_{n+1}|\right| \qquad (n = 1,2,3,\ldots) ,$$

which is comparable with (1.5) with $Z_{n+1} = -|\theta_{n+1}|$. Here the equivalent form (1.1) of the process is

$$X_{n+1} = |X_n| - |\theta_{n+1}| \qquad (n = 1,2,3,\ldots)$$

where the sequence $\{X_n\}$ gives the amount of under or over-correction at each stage. Feller gives the stationary distribution of the process.

A further example is given by a single server queue with Poisson arrivals and general service distribution. Let X_n be the number of customers present immediately following the completion of service of the n'th customer and let Z_n be the number of customers arriving during the service of the n'th customer, then with

$$U(x) = \begin{cases} 1 & (x > 0) \\ 0 & (x \leq 0), \end{cases}$$

$$X_{n+1} = X_n - U(X_n) + Z_{n+1} \qquad (n = 0,1,2,\ldots)$$

(Cox and Miller, 1965, pp.88,115). Another example in the same type of situation is given by the waiting time of a customer (Cox and Miller, 1965, p.64; Tweedie, 1975a, §7).

The model (1.1) can be generalised in an obvious way for multi-dimensional processes and special cases of these include non-linear autoregressions with higher orders of dependence and also non-linear "moving-average" models.

2 STATIONARITY OF AUTOREGRESSIVE PROCESSES

2.1 Autoregressive processes as Markov Processes

In this chapter conditions are investigated for the stationarity of the non-linear autoregressive processes generated by models of the form

$$X_{n+1} = \lambda(X_n) + Z_{n+1} \qquad (n = \ldots, -1,0,1,\ldots) \tag{2.1.1}$$

where, here and throughout, the input series $\{Z_n\}$ is a sequence of independent and identically distributed random variables with common distribution function $F_Z(z)$. The function $\lambda(x)$ is assumed fixed, real and measurable, so that $\lambda(X)$ is a well-defined random variable for any random variable X. In model building applications only continuous or sectionally continuous functions would normally be used.

In order to specify more exactly what is required of a model of the type (2.1.1) for it to represent a stationary (doubly infinite) sequence $\{X_n; -\infty < n < \infty\}$, consider a process of the same structure as (2.1.1) but starting at a finite time, say time zero, in some fixed or random state: that is a process $\{U_n; n \geq 0\}$ defined by

$$U_{n+1} = \lambda(U_n) + Z_{n+1} \qquad (n = 0,1,2,\ldots) \tag{2.1.2}$$

with U_0 having some fixed distribution and $\{Z_n\}$ having the same properties as before together with independence of U_0. The sequence $\{U_n\}$ forms a Markov process on the real line having stationary transition probabilities

$$P(y,A) = \Pr[U_{n+1} \in A \mid U_n = y] \qquad (n = 0,1,2,\ldots)$$

which are given by (2.1.2), and an initial distribution given by that of U_0. For these transition probabilities there may exist an invariant probability measure $\pi(\cdot)$ satisfying

$$\pi(A) = \int \pi(dy)P(y,A). \tag{2.1.3}$$

If such an invariant measure exists then the probability structure of the singly infinite process (2.1.2) may be extended to form a well-defined probability measure for the doubly infinite process given by (2.1.1) (Rosenblatt, 1971, p.73). Such a process is then a stationary Markov

process. There may be more than one probability measure satisfying (2.1.3) and each distinct measure determines a different marginal distribution for the process (2.1.1). Thus the transition probability function $P(y,A)$ and the particular invariant measure $\pi(A)$ together determine a process for which (2.1.1) is a representation.

The usual definition of stationarity assumes the existence of each of the joint distribution functions $F_{i_1,i_2,\ldots i_k}$ of $(X_{i_1}, X_{i_2}, \ldots X_{i_k})$ $(k = 1,2,3,\ldots)$, and requires that they satisfy the time invariance property

$$F_{i_1,i_2,\ldots i_k} = F_{i_1+h,\ldots i_k+h} \quad (h = 0,\pm1,\pm2,\ldots;\ k = 1,2,3,\ldots). \tag{2.1.4}$$

In particular this requires the existence of a distribution function F_X such that $F_i = F_X$ for all i, and since

$$F_{n+1}(x) = \int F_Z(x - \lambda(u))dF_n(u)$$

from the form of the process, F_X must satisfy

$$F_X(x) = \int F_Z(x - \lambda(u))dF_X(u) \tag{2.1.5}$$

which is equivalent to (2.1.3) with $F_X(x) = \pi[(-\infty,x)]$. When an invariant measure $\pi(\cdot)$ exists, the corresponding distribution function F_X will be called a stationary or equilibrium distribution function. The existence of an invariant measure implies stationarity in the sense (2.1.4) with probabilities of events given for $k \geq 1$ and any n by

$$\Pr[X_n \in A_0, X_{n+1} \in A_1, \ldots\ldots, X_{n+k} \in A_k]$$

$$= \int_{A_0} \pi(dx_0) \int_{A_1} P(x_0,dx_1) \int_{A_2} P(x_1,dx_2) \ldots\ldots \int_{A_{k-1}} P(x_{k-2},dx_{k-1})$$

$$P(x_{k-1},A_k).$$

If there is no probability measure satisfying (2.1.3) there can be no stationary distribution satisfying (2.1.5) and so there can be no stationary process generated by (2.1.1).

There are two natural (and equivalent) Markov Processes corresponding to each autoregressive process (2.1.1). These are the Markov process $\{X_n\}$

itself, and the process $\{Y_n\}$ where

$$Y_n = \lambda(X_n) \qquad (n = \ldots, -1,0,1,\ldots)$$

and hence

$$Y_{n+1} = \lambda(Y_n + Z_{n+1}) \qquad (n = \ldots, -1,0,1,\ldots).$$

For the process $\{X_n\}$ the transition probabilities are given by

$$P_X(x,A) = \Pr[X_{n+1} \in A \mid X_n = x]$$

$$= \int_{z+\lambda(x)\in A} dF_Z(z) = \int_{z\in A-\lambda(x)} dF_Z(z) \ . \qquad (2.1.6)$$

The transition probabilities for the process $\{Y_n\}$ are

$$P_Y(y,A) = \Pr[Y_{n+1} \in A \mid Y_n = y]$$

$$= \int_{\lambda(y+z)\in A} dF_Z(z) = \int_{z\in\lambda^{-1}(A)-y} dF_Z(z) \ . \qquad (2.1.7)$$

Here for sets A and real numbers x,

$$A - x = \{x' : x+x' \in A\}, \ \lambda^{-1}(A) = \{x' : \lambda(x') \in A\}.$$

It is clear that the state space of the process $\{Y_n\}$ may be reduced to the range of $\lambda(x)$ on the real line if this is finite. Also if a stationary distribution exists for either of the processes $\{X_n\}$ or $\{Y_n\}$ then there is a corresponding stationary distribution for the other. Since the Markov processes on the real line include as a special class the Markov chains with denumerable state spaces, it is clear that such phenomena as recurrence, periodicity and closed sets of states should have their analogues in the wider class. These are discussed in section 2.6 where some conditions for stationarity of the processes are also discussed.

If it can be shown that the sequence $\{U_n\}$, given by the finite starting time process (2.1.2), converges properly in distribution (Feller, 1971, p.248) for some initial distribution of U_0, then the limit distribution is a stationary distribution. When the process concerned is not periodic the convergence in distribution of $\{U_n\}$ is a fairly natural

requirement for stationarity: if the distributions of the sequence converge to an improper distribution this corresponds to a positive probability of "drifting off to infinity". Using this type of criterion the stationarity of linear first-order autoregressive processes will be investigated in section 2.2. It will be shown that, even when the "slope" parameter is strictly less than one in absolute value, not all linear processes are stationary. In section 2.4 the non-stationarity of certain non-linear processes will be demonstrated: this will be done by showing directly that, for the processes concerned, there is a non-zero probability of an event which has a natural interpretation as a "drift to infinity". This method can be extended to investigate more closely the rate of drift of non-stationary processes: this is done in section 2.5.

Finally it may be remarked that (2.1.1) is not in general a "Markov process in the wide sense" (Doob, 1953, pp.90,233) and hence the correlations of (X_n, X_{n+k}) in a stationary process will not necessarily decrease geometrically with k. This is exemplified in section 2.3.

2.2 Linear autoregressive processes

2.2.1 In this section consideration is given to processes generated by

$$X_{n+1} = bX_n + Z_{n+1} \qquad (n = \ldots, -1,0,1,\ldots) \tag{2.2.1}$$

where it is not assumed that the independent random variables Z_n are centred at zero. To investigate the possible stationarity of such a process the corresponding finite starting time process (2.1.2) is set up, i.e.,

$$U_{n+1} = bU_n + Z_{n+1} \qquad (n = 0,1,2,\ldots). \tag{2.2.2}$$

By repeated substitution this expression is equivalent to

$$U_n = Z_n + bZ_{n-1} + \ldots + b^{n-1}Z_1 + b^nU_0 \quad (n = 1,2,3,\ldots), \tag{2.2.3}$$

and the marginal distributions of these quantities are the same, by a simple relabelling, as those of

$$U_n^* = Z_1 + bZ_2 + \ldots + b^{n-1}Z_n + b^nU_0 \quad (n = 1,2,3,\ldots).$$

It is required to know under what circumstances the sequence of random variables $\{U_n^*;\ n \geq 1\}$, and hence $\{U_n;\ n \geq 1\}$, converges in distribution. The sequence $\{S_n\}$ is defined by

$$S_n = S_n(b) = Z_1 + bZ_2 + \ldots + b^{n-1}Z_n \quad (n = 1,2,3,\ldots). \qquad (2.2.4)$$

Clearly if $\{S_n\}$ does not converge in distribution then neither does $\{U_n^*\}$, since b^nU_0 and S_n are independent. However if $\{S_n\}$ does converge in distribution and $|b| < 1$ then $\{U_n^*\}$ also converges in distribution since $b^nU_0 \to 0$ in distribution for any random variable U_0. The quantities $S_n(b)$ are the partial sums of the random power series $\sum_{j=1}^{\infty} b^{j-1}Z_j$, and in particular this is a random power series with independent and identically distributed coefficients. Therefore (Kawata, 1972, p.625; Lukacs, 1975, p.127) the sequence $S_n(b)$ converges almost surely, and hence in distribution, only for

$$\text{(i)} \quad |b| < 1 \quad \text{if} \quad \int_1^{\infty} \log x \, dF_{|Z|}(x) < \infty$$

$$\text{(ii)} \quad b = 0 \quad \text{if} \quad \int_1^{\infty} \log x \, dF_{|Z|}(x) = \infty.$$

Here $F_{|Z|}$ denotes the distribution of $|Z|$ when Z has the distribution F_Z and it is assumed that Z_n is not trivially zero with probability 1. Thus there are input distributions for which $\{U_n\}$ does not converge in distribution for $0 < |b| < 1$. For example let F_Z have a density f_Z,

$$f_Z(z) = \begin{cases} 0 & (|z| < e), \\ \dfrac{1}{2|z|(\log|z|)^2} & (|z| \geq e), \end{cases}$$

then

$$\int_1^{\infty} \log x \, dF_{|Z|}(x) = \int_e^{\infty} \frac{dx}{x \log x} = \infty.$$

A further example is given by Lukacs (1975, p.129). An additional result is that, with probability 1, $\sum_{j=1}^{\infty} b^{j-1}Z_j$ cannot be continued outside its circle of convergence to form an analytic function of b (Lukacs, 1975, p.130; Kawata, 1972, p.629). Hence there is no natural way of defining an autoregressive process with parameter b ($|b| \geq 1$) by extension from the case $|b| < 1$.

2.2.2 The stationary distribution of a linear process is usually most easily found in terms of its characteristic function $\phi_X(s)$. If $\phi_Z(s)$ is the characteristic function of the input distribution,

$$\begin{aligned}\phi_X(s) &= \phi_Z(s)\phi_X(bs), \\ &= \prod_{n=0}^{\infty} \phi_Z(b^n s) ,\end{aligned} \tag{2.2.5}$$

where (Lukacs, 1970, p.58) the convergence, uniform in every finite interval of s, of the infinite product is both necessary and sufficient for the convergence in distribution of S_n of (2.2.4). Except in special cases it would usually be impossible to invert the stationary characteristic function (2.2.5) to obtain an explicit form for the stationary distribution. It may however be noted that if $\phi_Z(s)$ corresponds to a stable distribution then $\phi_X(s)$ is also a stable distribution with the same exponent and whose other parameters may be obtained from infinite summations corresponding to (2.2.5). The general form for the characteristic function of a stable distribution (Lukacs, 1970, p.136) is given by

$$\log \phi_Z(t) = iat - c|t|^{\alpha}\{1 + i\beta \frac{t}{|t|} w(|t|,\alpha)\} \quad (c \geq 0; |\beta| \leq 1; 0 < \alpha \leq 2),$$

where $$w(|t|,\alpha) = \begin{cases} \tan\frac{\pi\alpha}{2} & (\alpha \neq 1), \\ \frac{2}{\pi}\log|t| & (\alpha = 1). \end{cases}$$

The stationary characteristic function is then given, by (2.2.5), as

$$\log \phi_X(t) = ia_X t - c_X|t|^{\alpha}\{1 + i\beta_X \frac{t}{|t|} w(|t|,\alpha)\}$$

where

$$a_X = \begin{cases} \frac{a}{1-b} & (\alpha \neq 1), \\ \frac{a}{1-b} - \frac{2\beta c}{\pi}\frac{b\log|b|}{1-b^2} & (\alpha = 1), \end{cases} \tag{2.2.6}$$

$$c_X = \frac{c}{1-|b|^{\alpha}}, \quad \beta_X = \beta\frac{1-|b|^{\alpha}}{1-\mathrm{sgn}(b)|b|^{\alpha}}. \tag{2.2.7}$$

For the normal distribution ($\alpha = 2$; $w(|t|,\alpha) \equiv 0$) with mean a and variance σ^2, these give the well known result that the stationary distribution is normal with mean $a/(1 - b)$ and variance $\sigma^2/(1 - b^2)$. When the input distribution is a Cauchy distribution with density

$$f_Z(z) = \frac{c}{\pi[(z - a)^2 + c^2]} ,$$

the stationary distribution is of the same form with a replaced by $a/(1 - b)$ and c by $c/(1 - |b|)$. This corresponds to the case $\alpha = 1$, $\beta = 0$. There is one further combination of parameters for which the corresponding stationary distribution can be written in terms of elementary functions. This is the case $\alpha = 1/2$, $\beta = -1$, $\log \phi_Z(t) = iat - c|t|^{1/2}\{1 - i \frac{t}{|t|}\}$. The random variable Z having this characteristic function has the distribution of $2c^2(W + a)$ where W has the density

$$f_W(w) = \begin{cases} \frac{1}{2\sqrt{\pi}} w^{-3/2} \exp\{-\frac{1}{4w}\} & (w > 0), \\ 0 & (w \leq 0), \end{cases}$$

(Lukacs, 1970, p.143). From formulae (2.2.5), (2.2.7) it can be seen that when $0 \leq b < 1$ the stationary distribution is of the same form with the centering parameter a replaced by $a/(1 - b)$ and the scaling parameter c^2 replaced by $c^2/(1 - b^{1/2})^2$. However if b is negative, the parameter β_X of the stationary distribution is

$$\beta_X = -\frac{1 - |b|^{1/2}}{1 + |b|^{1/2}}$$

so that the distribution is not of the above form.

Apart from the stable families of distributions, certain properties of the limiting distribution of $\{U_n\}$ generated by (2.2.2) have been found for one further case, namely that when the input sequence is two-valued with each value having equal probability. The properties of such processes are the same, apart from a shift and change of scale, as those of the process with input sequence taking values +1, -1, each with a probability of one half. That is

$$U_{n+1} = bU_n + Z_{n+1} \qquad (n = 0,1,2,\ldots),$$

$$\Pr(Z_i = 1) = \Pr(Z_i = -1) = 1/2.$$

Since the effect of the term $b^n U_0$ in (2.2.3) vanishes, as $n \to \infty$, and b may be replaced by $|b|$, the limiting distribution of $\{U_n\}$ is a particular case of a symmetric Bernoulli convolution (Lukacs, 1970, p.64; Wintner, 1947, §63) and is one for which the following properties of the limiting distribution F_X are known:

(a) F_X is either purely singular or purely absolutely continuous when $0 < |b| < 1$.

(b) F_X is purely singular when $0 < |b| < \frac{1}{2}$.

(c) when $|b| = \frac{1}{2}$, F_X is the uniform distribution on $(-2,2)$.

(d) when $1 > |b| > \frac{1}{2}$ F_X may be either purely singular or purely absolutely continuous, examples being

(i) $|b| = 2^{-1/k}$ for some integer $k \geq 1$, when F_X is absolutely continuous,

(ii) $|b| = \frac{1}{2}(\sqrt{5} - 1) > \frac{1}{2}$, when F_X is purely singular.

It is therefore clear that, in the general non-linear case, when the input distribution has a discrete component the stationary distribution, if any, may have a singular component. However, if the input distribution is absolutely continuous then, since the stationary distribution must be the convolution of the input distribution with another, it follows that the stationary distribution is itself absolutely continuous (Lukacs, 1970, p.38).

Suppose that the stationary distributions $F_X^{(i)}$ of the linear processes $\{X_n^{(i)}\}$ $(i = 1,\ldots,k)$ with input distributions $F_Z^{(i)}$ are known, where the processes are generated by

$$X_{n+1}^{(i)} = bX_n^{(i)} + Z_{n+1}^{(i)} \qquad (n = \ldots -1,0,1,\ldots \; ; \; i = 1,\ldots k).$$

Then it is clear from (2.2.5) that the stationary distribution of the linear process (with the same parameter b) with an input distribution which is a convolution of distributions amongst the $F_Z^{(i)}$ is just the corresponding convolution of the $F_X^{(i)}$'s. For example, let the input distribution have density

$$\frac{1}{2\sqrt{2\pi}\sigma}\left[\exp\{-\frac{1}{2\sigma^2}(x-1)^2\}+\exp\{-\frac{1}{2\sigma^2}(x+1)^2\}\right]$$

(which corresponds to a convolution of a discrete and a normal distribution) and let the process have parameter $b = 1/2$, then the stationary distribution is that of $X = S + T$ where S is $N(0, \frac{4}{3}\sigma^2)$ and T is uniform on $(-2,2)$. Thus the stationary density is

$$\frac{1}{4\sqrt{2\pi}\sigma^*}\int_{-2}^{2}\exp\{-\frac{1}{2\sigma^{*2}}(x-y)^2\}dy = \frac{1}{4}\{\Phi(\frac{2-x}{\sigma^*}) - \Phi(\frac{-2-x}{\sigma^*})\}$$

where $\sigma^* = \frac{2}{\sqrt{3}}\sigma$ and $\Phi(\cdot)$ is the standard normal distribution function.

2.2.3 Even though the stationary distribution itself cannot, in general, be calculated, its moments are available directly from (2.2.5). Thus, if $\kappa_{m,X}$ and $\kappa_{m,Z}$ are the cumulants of order m of the stationary and input distributions, they are related by

$$\kappa_{m,X} = \sum_{n=0}^{\infty}(b^{mn}\kappa_{m,Z}) = \frac{\kappa_{m,Z}}{1-b^m} \qquad (m \geq 1).$$

Standardising the cumulants of the two series by their respective variances leads to the conclusion that the m'th order ($m \geq 3$) standardised cumulant of $\{X_n\}$, when generated by (2.2.1), is a factor of that of the sequence $\{Z_n\}$. This factor is

$$l_m(b) = \frac{(1-b^2)^{m/2}}{1-b^m}$$

which approaches zero as b nears 1. Also $l_m(b) \to 0$ as $m \to \infty$ ($b \neq 0$), while $(l_m(b))^{1/m} \to (1-b^2)^{1/2}$ as $m \to \infty$. These formulae indicate that the moments of the stationary distribution, when they exist, approach those of a normal distribution as $b \to 1-0$ or $b \to -1+0$. An alternative route to the same result is to consider a standardised version of the sequence (2.2.2) or (2.2.3): for example, when $b > 0$ let

$$W_n = (1 + b + \ldots + b^n)^{-1}U_n.$$

Then, for $b \neq 1$, $W_n = \frac{1-b}{1-b^{n+1}}U_n$ and when $b = 1$, $W_n = \frac{1}{n+1}U_n$

$= \frac{1}{n+1}\{\sum_{i=1}^{n} Z_i + U_0\}$. Under the usual conditions the central limit

theorem shows that when b = 1, W_n converges to a normal distribution, and it can also be shown that W_n converges in distribution for b > 0. If U_0 has the distribution F_Z then $W_n = W_n(b)$ has the same distribution as $W_n(b^{-1})$. The convergence of W_n for $b \geq 1$ indicates that the non-convergence of the distributions of U_n is due to their increasing dispersion since

$$U_n = \begin{cases} \dfrac{b^{n+1} - 1}{b - 1} W_n & (b > 1) \\ (n + 1)\, W_n & (b = 1). \end{cases}$$

This means that the distribution functions converge to a constant function - an improper distribution.

Mallows (1967) considers the closeness to normality of the distributions of linear processes by a different approach, and also considers joint distributions.

2.3 Non-linear processes equivalent to Markov Chains

Besides the linear processes there is a further class of autoregressive processes for which the properties are easily found. If the autoregression function appearing in (2.1.1) takes on only a finite number of different values, then the process is equivalent to a Markov Chain with a finite number of states. Let

$$L_n = \lambda(X_n) \qquad (n = \ldots -1,0,1,\ldots)$$

and suppose that $\lambda(x)$ takes the values $\lambda_1, \lambda_2, \ldots, \lambda_N$ on the sets $A_1, A_2, \ldots, A_N$, where these form a partition of the real line. Then the probabilistic development of the sequence $\{L_n\}$ is given by the (N × N) matrix P whose elements P_{ij} are given by

$$P_{ij} = \Pr[L_{n+1} = \lambda_j \mid L_n = \lambda_i]$$

$$= \Pr[\lambda_i + Z_{n+1} \in A_j] = \Pr[\lambda(\lambda_i + Z_{n+1}) = \lambda_j] .$$

The properties of $\{X_n\}$ then follow from the relation

$$X_{n+1} = L_n + Z_{n+1} \qquad (n = \ldots -1,0,1,\ldots).$$

It is well known (Cox and Miller, 1965, p.80) that a stationary distribution for $\{L_n\}$ is given by $\Pr[L_n = \lambda_i] = \pi_i$ where $\pi = (\pi_1,\ldots,\pi_N)'$ is any solution of

$$\pi'P = \pi',$$

then also $\pi_i = \Pr[X_n \in A_i]$. A stationary distribution of $\{X_n\}$ is then given by the corresponding mixture

$$F_X(x) = \sum_{i=1}^{N} \pi_i F_Z(x - \lambda_i),$$

and, if the input distribution has a density f_Z, the stationary density is

$$f_X(x) = \sum_{i=1}^{N} \pi_i f_Z(x - \lambda_i).$$

The matrix P^k is the k-step transition matrix with elements $P_{ij}^{(k)}$ such that

$$P_{ij}^{(k)} = \Pr[L_{n+k} = \lambda_j \mid L_n = \lambda_i] \qquad (i,j = 1,\ldots,N).$$

Hence a stationary joint distribution of (L_n, L_{n+k}) is given by

$$\Pr[L_n = \lambda_i,\ L_{n+k} = \lambda_j] = \pi_i P_{ij}^{(k)}.$$

The joint distribution of (X_n, X_{n+k}) can also be found: assuming the density functions exist, the joint density $f_k(\cdot,\cdot)$ of (X_n, X_{n+k}) is given by

$$f_k(x,y) = \begin{cases} f_Z(y - \lambda_i) f_X(x) & (x \in A_i;\ k = 1), \\ \sum_{j=1}^{N} f_Z(y - \lambda_j) P_{ij}^{(k-1)} f_X(x) & (x \in A_i;\ k \geq 2). \end{cases}$$

Define the conditional moments

$$\xi_i = E[X_n \mid X_n \in A_i] = E[X_n \mid \lambda(X_n) = \lambda_i]$$

$$= \int_{x \in A_i} x f_X(x)\,dx \qquad (i = 1, \ldots N), \tag{2.3.1}$$

then the moments of the stationary process $\{X_n\}$ are given by

$$E(X_n) = \mu_Z + \Sigma\, \pi_i \lambda_i = \Sigma\, \pi_i \xi_i\,, \tag{2.3.2}$$

$$E(X_n X_{n+1}) = \mu_Z \Sigma\, \pi_i \xi_i + \Sigma\, \lambda_i \pi_i \xi_i\,, \tag{2.3.3}$$

$$E(X_n X_{n+k}) = \mu_Z \Sigma\, \pi_i \xi_i + \Sigma\Sigma\, \lambda_j P_{ij}^{(k-1)} \pi_i \xi_i \qquad (k \geq 2). \tag{2.3.4}$$

For example, let

$$\lambda(x) = \begin{cases} \lambda_1 & (x < x_0), \\ \lambda_2 & (x \geq x_0). \end{cases}$$

Then the transition matrix between the levels λ_1, λ_2 is

$$P = \begin{bmatrix} F_Z(x_0 - \lambda_1) & 1 - F_Z(x_0 - \lambda_1) \\ F_Z(x_0 - \lambda_2) & 1 - F_Z(x_0 - \lambda_2) \end{bmatrix}$$

and the stationary distribution of the process $\{L_n\}$ is given by

$$\pi_1 = \frac{F_Z(x_0 - \lambda_2)}{1 - F_Z(x_0 - \lambda_1) + F_Z(x_0 - \lambda_2)}\,, \qquad \pi_2 = 1 - \pi_1\,.$$

The stationary distribution of the $\{X_n\}$ process is then

$$F_X(x) = \frac{F_Z(x-\lambda_1) F_Z(x_0-\lambda_2) + F_Z(x-\lambda_2)\{1 - F_Z(x_0-\lambda_1)\}}{1 - F_Z(x_0-\lambda_1) + F_Z(x_0-\lambda_2)}$$

and the moments and joint moments are given by the above formulae. When the input distribution is normal the conditional moments ξ_i may be found explicitly by a simple substitution. As a more specific example let

$F_Z(x) = \Phi(x/\sigma)$ where $\Phi(\cdot)$ is the standard normal distribution function and let $x_0 = 0$, $\lambda_1 = -\lambda_2$, $\lambda_2 > 0$, so that the autoregression function is odd (about zero). Then $\pi_1 = \pi_2 = 1/2$ and

$$F_X(x) = \frac{1}{2}\{\Phi(\frac{x + \lambda_2}{\sigma}) + \Phi(\frac{x - \lambda_2}{\sigma})\},$$

$$\xi_2 = E[X_n \mid X_n > 0]$$

$$= \frac{\sigma}{\sqrt{2\pi}} \exp\{-\frac{1}{2}\frac{\lambda_2^2}{\sigma^2}\} + \lambda_2 [\Phi(\frac{\lambda_2}{\sigma}) - \frac{1}{2}]$$

$$\xi_1 = -\xi_2 .$$

It follows that $E(X_n) = E(X_n^3) = 0$, $\mathrm{var}(X_n) = \sigma^2 + \lambda_2^2$ and that $E(X_n^4) = 3\sigma^4 + 6\lambda_2^2\sigma^2 + \lambda_2^4$. Hence $E(X_n^4) - 3\{\mathrm{var}(X_n)\}^2 = -2\lambda_2^4$. For the joint moments, $E(X_n X_{n+1}) = \lambda_2\xi_2$, and, using the symmetry of the transition matrix P, which is

$$P = \begin{bmatrix} \Phi(\lambda_2/\sigma) & 1 - \Phi(\lambda_2/\sigma) \\ 1 - \Phi(\lambda_2/\sigma) & \Phi(\lambda_2/\sigma) \end{bmatrix},$$

and of its powers P^k, $E(X_n X_{n+k}) = \lambda_2\xi_2\{P_{22}^{(k-1)} - P_{21}^{(k-1)}\}$. By definition, and using the symmetry,

$$P_{22}^{(k)} - P_{21}^{(k)} = (P_{21}^{(k-1)}P_{12} + P_{22}^{(k-1)}P_{22}) - (P_{21}^{(k-1)}P_{11} + P_{22}^{(k-1)}P_{21})$$

$$= (P_{22}^{(k-1)} - P_{21}^{(k-1)})(P_{22} - P_{21}) = (P_{22}^{(k-1)} - P_{21}^{(k-1)})$$

$$\cdot(2\Phi(\lambda_2/\sigma) - 1) .$$

It then follows that

$$E(X_n X_{n+k}) = \lambda_2\xi_2(2\Phi(\lambda_2/\sigma) - 1)^{k-1} \quad (k = 1,2,3,\ldots). \tag{2.3.5}$$

Thus the correlations $\rho_0, \rho_1, \rho_2, \ldots$ (in the usual notation) decrease geometrically from ρ_1 onwards.

In the general case it can be seen from (2.3.2)-(2.3.4) by making a decomposition of P in terms of its eigenvalues (using a Jordan canonical

form, see Cox and Miller (1965), p.121), that the covariances can be expressed as a finite sum of terms involving powers of the eigenvalues of P. If P is first reduced to the transition matrix between the states given positive probability by the stationary distribution π and if the chain on this reduced state space is irreducible and not periodic then the covariances, $\mathrm{cov}(X_n, X_{n+k})$, can be written as the sum of a number of terms decreasing geometrically in absolute value to zero (with k) at different rates. If the chain is periodic then the covariances do not converge to zero for large k, but approach values depending on the residue of k on division by the period of the chain.

2.4 Nonstationarity of non-linear processes

2.4.1 In this section the possible nonstationarity of non-linear auto-regressive processes is investigated. For the process under consideration

$$X_{n+1} = \lambda(X_n) + Z_{n+1} \qquad (n = \ldots. -1,0,1,\ldots),$$

the corresponding finite starting time process

$$U_{n+1} = \lambda(U_n) + Z_{n+1} \qquad (n = 0,1,2,\ldots) \tag{2.4.1}$$

is set up and it is shown that, under conditions on $\lambda(x)$ and $F_Z(z)$, events of the type $\{U_0 > 0,\ U_1 > 1,\ U_2 > 2,\ U_3 > 3,\ \ldots\ldots\}$ have a non-zero probability. It is natural to interpret this event as the process "drifting to infinity". In section 2.5 the methods developed here are extended to study more closely the "rate of drift to infinity" of non-stationary processes.

2.4.2 Define $\lambda_{(+)}(x) = \inf_{y>x} \lambda(y)$ and, with $h > 0$, set

$$s_n = F_Z\{(n+1)h - \lambda_{(+)}(nh)\} \qquad (n = \ldots., -1,0,1,\ldots),$$

where the usual conventions apply if $\lambda(y)$ is not bounded below for $y > x$. It will be shown that the process $\{U_n\}$ drifts off to infinity in the above natural sense under the condition that

$$\sum_{n=0}^{\infty} s_n \quad \text{converges.} \tag{2.4.2}$$

Suppose this condition holds. Then there is an integer k_0, which may be negative, such that

$$F_Z((k+1)h - \lambda_{(+)}(kh)) = s_k < 1 \qquad \text{for all } k \geq k_0 . \tag{2.4.3}$$

It will be shown that

$$\Pr\{U_n > (k_0+n)h;\ n = 0,1,2,\ldots\} > 0 \text{ if } \Pr\{U_0 > k_0 h\} > 0 . \tag{2.4.4}$$

Let $U^n = (U_0, U_1, \ldots U_n)$ $(n \geq 0)$ and let $P_n(\cdot)$ be the probability measure for U^n. Define the events $C_n = \{U_j > (k_0+j)h;\ j = 0,1,\ldots n\}$ and $C = \bigcap_{n=0}^{\infty} C_n = \{U_j > (k_0+j)h;\ j = 0,1,2,\ldots\}$. Then, since $C_{n+1} \subseteq C_n$,

$$\Pr\{U_j > (k_0+j)h;\ j = 0,1,2,..\} = \lim_{n\to\infty} \Pr(C_n). \tag{2.4.5}$$

Denote $\Pr(C_n)$ by p_n. Then, for $n \geq 1$,

$$\begin{aligned} p_n = \Pr(C_n) &= \Pr\{U_n > (k_0+n)h \mid C_{n-1}\} \\ &= \int_{C_{n-1}} \Pr\{U_n > (k_0+n)h \mid U^{n-1} = u^{n-1}\} P_{n-1}(du^{n-1}). \end{aligned} \tag{2.4.6}$$

Because of the Markov structure of the process

$$\begin{aligned} \Pr\{U_n > (k_0+n)h \mid U^{n-1} = u^{n-1}\} &= \Pr\{U_n > (k_0+n)h \mid U_{n-1} = u_{n-1}\} \\ &= \Pr\{Z_n > (k_0+n)h - \lambda(u_{n-1})\} \\ &\geq \Pr[Z_n > (k_0+n)h - \lambda_{(+)}\{(k_0+n-1)h\}] \\ &\qquad \text{for } u_{n-1} > (k_0 + n-1)h \\ &= 1 - s_{k_0+n-1} . \end{aligned}$$

Substituting this result into (2.4.6) gives

$$p_n \geq (1 - s_{k_0+n-1}) \int_{C_{n-1}} P_{n-1}(du^{n-1}) = (1 - s_{k_0+n-1}) p_{n-1} \quad (n \geq 1),$$

and therefore

$$p_n \geq p_0 \prod_{j=0}^{n-1} (1 - s_{k_0+j}) \qquad (n \geq 1) \ . \tag{2.4.7}$$

By (2.4.3), $0 < 1 - s_{k_0+j} \leq 1$ $(j \geq 0)$ so that the sequence of products in (2.4.7) is non-zero and decreasing. Since the sequence $\{p_n\}$ converges to the limit (2.4.5),

$$\Pr\{U_j > (k_0+j)h, \ j = 0,1,2,\ldots\} \geq p_0 \prod_{j=0}^{\infty} (1 - s_{k_0+j}) \ . \tag{2.4.8}$$

This infinite product either converges to a positive number or else it diverges to zero and it converges or diverges according to whether $\sum_{j=0}^{\infty} s_{k_0+j}$ converges or diverges (Titchmarsh, 1939, p.14). Therefore, because of the assumption (2.4.2), the right hand side of inequality (2.4.8) is strictly positive if $p_0 = \Pr(U_0 > k_0 h) > 0$.

It is clear that the same arguments apply to processes becoming increasingly large and negative. Defining $\lambda^{(-)}(x) = \sup_{y \leq x} \lambda(y)$ and setting

$$t_n = \Pr(Z > -(n+1)h - \lambda^{(-)}(-nh))$$

$$= 1 - F_Z\{-(n+1)h - \lambda^{(-)}(-nh)\} \qquad (n = \ldots -1,0,1,\ldots) \ ,$$

it may be concluded that, if $\sum_{n=0}^{\infty} t_n$ converges, there is an integer ℓ_0 such that the event $\{U_n \leq -(\ell_0+n)h; \ n = 0,1,2,..\}$ has positive probability whenever $\Pr\{U_0 \leq -\ell_0 h\} > 0$. The number ℓ_0, which may be negative is such that

$$t_{\ell_0+j} < 1 \quad \text{for } j \geq 0 \ . \tag{2.4.9}$$

Processes whose realisations have increasing absolute value with alternating signs can also be dealt with in a similar manner. Let $\lambda_{(-)}(x) = \inf_{y \leq x} \lambda(y)$ and $\lambda^{(+)}(x) = \sup_{y > x} \lambda(y)$ and define

$$q_n = F_Z\{(n+1)h - \lambda_{(-)}(-nh)\} \qquad (n = \ldots. -1,0,1,\ldots) \ ,$$

$$r_n = 1 - F_Z\{-(n+1)h - \lambda^{(+)}(nh)\} \qquad (n = \ldots. -1,0,1,\ldots) \ .$$

Then if $\sum_{n=0}^{\infty} q_n$ and $\sum_{n=0}^{\infty} r_n$ both converge, the event $\{U_{2j} > (k_0 + 2j)h, U_{2j+1} \leq -(k_0 + 2j+1)h;\ j = 0,1,\ldots\}$ has positive probability if $\Pr(U_0 > k_0 h) > 0$. Here k_0 is any number such that

$$q_{k_0+2j} < 1, \quad r_{k_0+2j+1} < 1 \quad \text{for } j \geq 0 . \tag{2.4.10}$$

The condition that $\sum_{n=0}^{\infty} q_n$, $\sum_{n=0}^{\infty} r_n$ should both converge seems somewhat strong since this imposes the same type of restrictions on the function $\lambda(x)$ at both extremes of the real line. It seems reasonable that a similar alternating drift to infinity should hold if $\lambda(x) \simeq -x$ at one extreme with suitable conditions at the other.

The above arguments do not depend essentially on the additive-error structure of the process but they also hold for a wider class of Markovian processes. Thus the conclusion (2.4.4) holds with

$$s_n = \sup_{x>nh} \Pr\{U_{n+1} \leq (n+1)h \,\big|\, U_n = x\},$$

and similar definitions may be made to deal with the other types of drift.

2.4.3 It has been shown that there is a positive probability of the process (2.4.1) drifting to infinity in one or more of the ways described when the corresponding series $\sum_{n=0}^{\infty} s_n$, $\sum_{n=0}^{\infty} t_n$ or $\sum_{n=0}^{\infty} (q_n + r_n)$ converge. For ease of consideration the four types of term may be written in a common form. Let $G(x)$ be a nonnegative decreasing function, not identically zero, such that $G(x) \to 0$ as $x \to \infty$. Then

(i) $s_n = G\{\mu(nh) - (n+1)h\}$ where $G(x) = F_Z(-x)$, $\mu(x) = \lambda_{(+)}(x)$,

(ii) $t_n = G\{\mu(nh) - (n+1)h\}$ where $G(x) = 1 - F_Z(x)$, $\mu(x) = -\lambda^{(-)}(-x)$,

(iii) $q_n = G\{\mu(nh) - (n+1)h\}$ where $G(x) = F_Z(-x)$, $\mu(x) = \lambda_{(-)}(-x)$,

(iv) $r_n = G\{\mu(nh) - (n+1)h\}$ where $G(x) = 1 - F_Z(x)$, $\mu(x) = -\lambda^{(+)}(x)$.

In each case $G(x)$ is one of the tails of the input distribution function. Because of this common form only one case need be treated and this will be the first.

It is clear that, for the series Σs_n to converge, the condition

$$s_n = G(\lambda_{(+)}(nh) - (n+1)h) \to 0 \quad \text{as } n \to \infty$$

must hold. For this $\lambda(x) - x$ must be bounded below as $x \to \infty$. Exact conditions for the convergence of the series depend on both the lower tail of F_Z and the behaviour of $\lambda(x)$ as $x \to \infty$. Conditions on $\lambda(x)$ to ensure the convergence of Σs_n under certain assumptions about F_Z will now be given. Under these conditions there is a drift to infinity as defined by (2.4.4).

Suppose that F_Z has a finite lower tail so that $F_Z(z) = 0$ $(z < -L)$, $F_Z(z) > 0$ $(z > -L)$. Then $G(x) = 0$ for $x > L$, and hence $s_n = 0$ whenever

$$\lambda_{(+)}(nh) - (n+1)h > L. \tag{2.4.11}$$

Thus a sufficient condition for the convergence of the series is that this should hold for all large enough n. Convergence implies that

$$\lim_{n\to\infty} \inf\{\lambda_{(+)}(nh) - (n+1)h\} \begin{cases} > L & (F_Z(-L) > 0), \\ \geq L & (F_Z(-L) = 0). \end{cases}$$

A sufficient condition for the convergence of Σs_n for some $h > 0$, and hence for the result (2.4.4) to hold for some h, is that for some $\varepsilon > 0$ and some x_0,

$$\lambda(x) \geq L + \varepsilon + x \qquad (x \geq x_0).$$

Now suppose that F_Z has a power-like lower tail so that

$$G(x) \leq cx^{-\delta} \tag{2.4.12}$$

for some $c, \delta > 0$ and for large enough x. Then $s_n \leq c\{\lambda_{(+)}(nh) - (n+1)h\}^{-\delta}$. Let $d_n = \lambda_{(+)}(nh) - (n+1)h$, then Σs_n converges if $\Sigma d_n^{-\delta}$ does. By selecting suitable series Σw_n which are known to converge and taking $d_n^{-\delta} \leq w_n$, or equivalently $d_n \leq w_n^{-1/\delta}$, bounds on the behaviour of $\lambda(x)$ to ensure convergence of Σs_n can be found. Thus taking

$$w_n = an^{-(1+\varepsilon)} \qquad (\varepsilon > 0) \tag{2.4.13}$$

leads to the bound

$$\lambda(x) \geq x + ax^{\frac{1+\varepsilon}{\delta}} + h \tag{2.4.14}$$

which when satisfied for large enough x for some $h,a,\varepsilon > 0$ ensures the positive probability of a drift to infinity for some initial distribution. Better bounds with the same property can be obtained from other series Σw_n. Thus the convergent series

$$\Sigma \frac{a}{n(\log n)^{1+\varepsilon}} \qquad (a,\varepsilon > 0) \tag{2.4.15}$$

leads to the bound

$$\lambda(x) \geq x + ax^{1/\delta}(\log x)^{\frac{1+\varepsilon}{\delta}} + h.$$

If now $G(x) \leq c \exp\{-\delta x^q\}$ for large enough x for some $c,\delta,q > 0$, then proceeding in the same way and bounding $\Sigma\exp\{-\delta d_n^{\,q}\}$ by suitable convergent series Σw_n gives bounds of the form

$$d_n \geq \{-\frac{1}{\delta}\log w_n\}^{1/q}.$$

The series given by (2.4.13) leads to the bound

$$\lambda(x) \geq x + b(\log x - k)^{1/q} + h \tag{2.4.16}$$

for large enough x for some k and some $b > \delta^{-1/q}$.

Generally the longer the lower tail of the input distribution is, the more rapidly the function $\lambda(x)$ has to increase as $x \to \infty$ to ensure the property (2.4.4) by the convergence of Σs_n. Thus if it is only assumed that $G(x) \leq c(\log x)^{-1}$ for large enough x, the above methods produce the bound $d_n \geq \exp(c\, w_n^{-1})$. With w_n of the form (2.4.13) this gives

$$\lambda(x) \geq x + \exp\{ax^{1+\varepsilon}\} + h$$

for some $a > 0$, while a weaker bound derived from (2.4.15) is

$$\lambda(x) \geq x + \exp\{ax(\log x)^{1+\varepsilon}\} + h$$

for large enough x for some $a,\varepsilon,h > 0$.

2.4.4 In section 2.4.2, k_0 and ℓ_0 are any numbers satisfying the appropriate condition (2.4.3), (2.4.9) or (2.4.10). It is possible that the condition will be satisfied for all values of these numbers. For instance, when $\lambda(x)$ is bounded below on every finite range (so that $\lambda_{(+)}(x) > -\infty$) and F_Z has an infinite upper tail (i.e. $F_Z(z) < 1$ for all finite z), $s_k < 1$ for all positive and negative k. It follows that in this case the process (2.4.1) drifts, with positive probability, to plus infinity for every distribution of the initial value U_0 whenever Σs_k converges. Similarly for the other types of drift there are wide classes of processes for which the drift property holds for any distribution of U_0. The same conclusions follow if, for a fixed k_0 satisfying (2.4.3), some integer p and all distributions of U_0,

$$\Pr[U_p > k_0 h] > 0 \ . \qquad (2.4.17)$$

Thus for certain processes it can be shown that for no distribution of U_0 does the sequence $\{U_n\}$, given by (2.4.1), converge in distribution and hence that there can be no stationary distribution. However in many cases it may only be possible to show this for a restricted class of initial distribution - typically those giving positive weight to suitable extreme values of U_0. Indeed it may be the case that the process $\{U_n\}$ converges in distribution for certain initial distributions while drifting to infinity for others. For example consider a process with

$$\lambda(x) = \begin{cases} 2x & (|x| > 2) \\ 0 & (|x| \leq 2) \end{cases}$$

and with $\{Z_i\}$ taking only values in $[-1,1]$. Then, if the initial distribution is such that U_0 takes only values in $[-2,2]$, $U_n = Z_n$ $(n = 1,2,3,\ldots)$, so that $\{U_n\}$ converges trivially in distribution. However if the initial distribution gives positive probability to the complement of $[-2,2]$, then with exactly that probability the sequence $\{U_n\}$ diverges to $\pm\infty$.

It has been shown in Section 2.2 that there cannot be a stationary process associated with the linear autoregression function $\lambda(x) = bx$, $|b| \geq 1$, for any input distribution which is not concentrated entirely at

zero. With some restrictions on the input distributions the same conclusion holds for processes with polynomial-like autoregression functions. By reversing the arguments of section 2.4.3 and seeking those functions G for which

$$\sum_{n=0}^{\infty} G(\lambda_{(+)}(nh) - (n+1)h) < \infty$$

for a particular autoregression function, it may be shown that a process with an autoregression function satisfying

$$\lambda(x) \geq \gamma x^q \qquad (x > x_0;\ \gamma > 0;\ q > 1) \tag{2.4.18}$$

has a drift to infinity as defined by (2.4.4) if the lower tail of the input distribution satisfies, for some $x_1, c, \delta > 0$, the bound

$$G(x) = F_Z(-x) \leq c\, x^{-\left(\frac{1+\delta}{q}\right)} \qquad (x > x_1)\ , \tag{2.4.19}$$

or a slightly weaker bound

$$G(x) \leq c\, x^{-1/q} (\log x)^{-\frac{1+\delta}{q}} \qquad (x > x_1)\ . \tag{2.4.20}$$

Therefore, if this bound is satisfied and if either (2.4.17) holds or if

$$s_k = G(\lambda_{(+)}(nh) - (n+1)h) < 1 \quad \text{for all } k,$$

there is no stationary process with an autoregression function satisfying (2.4.18). Since the other types of drift to infinity may be dealt with in a similar way it may be concluded that, for an input distribution having infinite tails with exponential-like behaviour, i.e. if for some x_0

$$0 < F_Z(-x),\ \ 1 - F_Z(x) < ce^{-ax} \qquad (x > x_0;\ c,a > 0) \tag{2.4.21}$$

then the autoregressive process (2.1.1) is nonstationary if any of the following hold for some x_1, γ, $\gamma_1 > 0$ and some $q, q_1 > 1$:

(i) $\lambda(x) \geq \gamma x^{q} \quad (x > x_1)$,

(ii) $\lambda(x) \leq -\gamma|x|^{q} \quad (x < -x_1)$

(iii) $\lambda(x) \geq \gamma|x|^{q} \quad (x < -x_1)$ and $\lambda(x) \leq -\gamma_1|x|^{q_1} \quad (x > x_1)$.

Clearly the same conclusion holds with a slightly weaker bound than (2.4.21) and it also holds with q, q_1 restricted to $q, q_1 \geq q_0 > 1$ and the bound on the tails of the distribution replaced by a power-like bound similar to (2.4.19), (2.4.20).

2.5 Rates of Drift to Infinity

The bounds obtained in section 2.4.3 are conditions for the process to have a drift to infinity with the property that

$$\Pr\{U_0 > k_0 h,\ U_1 > (k_0 + 1)h,\ U_2 > (k_0 + 2)h, \ldots\} > 0$$

for some initial distribution. A drift to infinity might equally well be said to hold if

$$\Pr\{U_j > x_j;\ j = 0,1,2,\ldots\} > 0 \tag{2.5.1}$$

where $\{x_j\}$ is any sequence of real numbers such that $x_j \to \infty$ as $j \to \infty$. It is clear that the whole of section (2.4.2) holds when the sequence $\{jh\}$ is replaced by any sequence $\{x_n\}$, whether or not $x_j \to \infty$ as $j \to \infty$. Thus, for example, a sufficient condition for (2.5.1) to hold for some initial distribution is that $\sum_{n=0}^{\infty} s_n$ should converge, where now

$$s_n = F_Z(x_{n+1} - \lambda_{(+)}(x_n)) \quad (n = \ldots -1,0,1,\ldots) .$$

It would appear that a weaker form of drift, say

$$\Pr\{U_j > c \log (j+1);\ j = 0,1,2,\ldots\} > 0,$$

should hold under weaker conditions on $\lambda(x)$ than those obtained in section 2.4.3. However the above criterion that Σs_n should converge does not in general produce weaker bounds on the autoregression function.

The basis of the method is that $\lambda_{(+)}(x_n) - x_{n+1}$ must tend to plus infinity sufficiently rapidly for Σs_n to converge, and this, together with the condition that $x_n \to \infty$ as $n \to \infty$, heavily restricts the combinations of functions $\lambda(x)$ and sequences $\{x_n\}$ for which a "drift to infinity" can be demonstrated by this method. In all, the condition for the drift seems somewhat inconsistent, for, if it is true that

$$\Pr\{U_j > x_j;\ \ j = 0,1,2,\ldots\} > 0 \qquad (2.5.2)$$

and $x_j \geq x_j^*$ $(j = 0,1,2,\ldots)$, then it is certainly true that

$$\Pr\{U_j > x_j^*;\ \ j = 0,1,2,\ldots\} > 0\ , \qquad (2.5.3)$$

and yet the method would fail to show this. For example let $x_j^* = x_0$ in a case where (2.5.2) holds for an increasing sequence $\{x_j\}$, then the appropriate series, Σs_n^* say, is

$$\Sigma s_n^* = \Sigma F_Z\{x_0 - \lambda_{(+)}(x_0)\},$$

which diverges whenever $F_Z\{x_0 - \lambda_{(+)}(x_0)\} \neq 0$. As a further example consider the process with autoregression function $\lambda(x) = \beta x$ $(\beta > 1)$ and for which the input distribution has an exponential lower tail, $F_Z(x) = c \exp\{\delta x\}$ $(x < -x_0;\ \delta > 0)$. Then the convergence of Σs_n shows that (2.5.1) holds for $\{x_n\}$ given by

$$\text{(i)}\quad x_n = k\alpha^n \qquad (1 < \alpha < \beta),$$

$$\text{or}\quad \text{(ii)}\quad x_n = k\beta^n + an \qquad (a > 0),$$

but does not show that it holds for

$$\text{(iii)}\quad x_n = k\alpha^n \qquad (0 \leq \alpha \leq 1),$$

or for (iv) $x_n = k\beta^n$.

Here k is a constant > 0.

Some idea of the possible sequences $\{x_n\}$ for which (2.5.1) holds for a particular process can however be obtained. For convenience it will be assumed that $\lambda_{(+)}(x)$ is strictly increasing and has an inverse

function $\lambda_{(+)}^{-1}(\cdot)$ and it will also be assumed that $G(x) = F_Z(-x)$, the lower tail of the input distribution, has an inverse $G^{-1}(\cdot)$. Let $\sum_{n=0}^{\infty} w_n$ be any convergent series and let $\{v_n\}$ be any sequence of real numbers. Then define

$$u_n = \lambda_{(+)}^{-1}[v_{n+1} + G^{-1}(w_n)] \qquad (n = 0,1,\ldots).$$

Assuming that $u_n \leq v_n$ $(n \geq 0)$, it can be shown that

$$\sum_{n=0}^{\infty} s_n = \sum_{n=0}^{\infty} G\{\lambda_{(+)}(x_n) - x_{n+1}\} < \infty$$

for any sequence $\{x_n\}$ satisfying $u_n \leq x_n \leq v_n$, and hence that (2.5.1) holds for such sequences. For, since G is decreasing,

$$s_n = G\{\lambda_{(+)}(x_n) - x_{n+1}\} < G\{\lambda_{(+)}(u_n) - v_{n+1}\}$$

$$= G\{G^{-1}(w_n)\} = w_n .$$

Hence Σs_n converges. It remains to determine for which sequences $\{v_n\}$

$$\lambda_{(+)}^{-1}[v_{n+1} + G^{-1}(w_n)] \leq v_n$$

for some convergent series Σw_n. It then follows by the above discussion that (2.5.3) holds for any sequence $\{x_j^*\}$ satisfying $x_j^* \leq v_j$ $(j \geq 0)$.

For example consider the linear process with exponential-tailed input distribution described above and let $\Sigma w_n = \Sigma a(n+1)^{-1+\varepsilon}$ $(\varepsilon > 0)$. Then, since

$$G^{-1}(w_n) = b + \frac{1+\varepsilon}{\delta}\log(n+1),$$

where b is a constant depending on a,δ, and c, the condition for $\{v_n\}$ to satisfy is that

$$\frac{1}{\beta}\left[v_{n+1} + b + \frac{1+\varepsilon}{\delta}\log(n+1)\right] \leq v_n \qquad (n = 0,1,2,\ldots)$$

for some constant b (possibly negative or zero) and some $\varepsilon > 0$. Considering sequences $\{v_n\}$ of the form $v_n = k\alpha^n$, this condition is satisfied for

$1 < \alpha < \beta$ but not for $\alpha \leq 1$ or $\alpha \geq \beta$. The same conclusion also holds if the lower tail of the input distribution satisfies a power-like bound (2.4.12) but not if the lower tail $G(x)$ decreases like $(\log x)^{-1}$. Let $F_Z(x) = c(\log |x|)^{-1}$ $(x < x_0)$, then $G^{-1}(u) = e^{c/u}$ and the above condition for the series $\{v_n\}$ is, with w_n as before,

$$\frac{1}{\beta}\,[v_{n+1} + \exp\{b(n+1)^{1+\varepsilon}\}] \leq v_n \qquad (n = 0,1,2,\ldots)$$

for some $b,\varepsilon > 0$. It does not seem possible to find an increasing sequence satisfying this. The criterion that Σs_n should converge does not demonstrate the nonstationarity of the linear process with this type of input distribution as does the theory of random power series in section 2.2.

2.6 Decomposition of State Space and Stationarity

2.6.1 Given the transition probability function $P(\cdot,\cdot)$ of a Markov process $\{X_n\}$ on a state space Ω with a Borel field $\mathcal{A}$, the (n+1)-step transition probabilities,

$$P^{(n+1)}(x,A) = \Pr[X_{n+1} \in A \mid X_0 = x],$$

are given, for all $x \in \Omega$ and $A \in \mathcal{A}$ by

$$P^{(n+1)}(x,A) = \int_\Omega P(x,dy)P^{(n)}(y,A) \qquad (n = 1,2,3,\ldots),$$

and they satisfy the Chapman-Kolmogorov equation

$$P^{(n+m)}(x,A) = \int_\Omega P^{(n)}(x,dy)P^{(m)}(y,A) \qquad (n = 1,2,3,\ldots;\ m = 1,2,3,\ldots).$$

Results concerning the existence of a stationary measure for the process and the possible decomposition of the state space are available under the following condition (Doob, 1953, pp.190-218; Rosenblatt, 1971, p.117). There should exist a finite measure $\phi(\cdot)$ on sets $A \in \mathcal{A}$, with $\phi(\Omega) > 0$, an integer $m \geq 1$ and an $\varepsilon > 0$ such that, whenever $\phi(A) \leq \varepsilon$,

$$P^{(m)}(x,A) \leq 1 - \varepsilon \quad \text{for all } x \in \Omega. \tag{2.6.1}$$

This is Doob's Condition D. In effect the condition insists that from anywhere in the state space the process moves with high probability in a fixed finite number of steps to a region near the "centre" of the state space (Tweedie, 1975a).

Under this condition the following hold.

(i) There exists an invariant probability measure π for the transition function $P(\cdot,\cdot)$, i.e.,

$$\int_{\Omega} \pi(dy)P(y,A) = \pi(A) \qquad (2.6.2)$$

The measure ϕ is absolutely continuous with respect to π (i.e. $\phi(A) = 0$ whenever $\pi(A) = 0$).

(ii) There are at most a finite number of minimal invariant sets E_j $(j = 1,...p)$. An invariant set is a set E with the property that

$$P^{(n)}(x,E) = 1 \text{ for all } n \geq 1 \text{ and all } x \in E.$$

A minimal invariant set is an invariant set which contains no non-empty invariant set with smaller measure under $\phi(\cdot)$. If E is any non-null minimal invariant set it differs from some E_j (above) by at most a set of ϕ-measure zero. For any $y \in \Omega$

$$\lim_{n\to\infty} P^{(n)}(y, \bigcup_j E_j) = 1 ,$$

so that a process commencing at any state y will remain outside the invariant sets for only a finite number of steps in its transitions (with probability 1). The set $\Omega - \bigcup_j E_j$ is transient in the sense that

$$\lim_{n\to\infty} P^{(n)}(y, \Omega - \bigcup_j E_j) = 0 \qquad (y \in \Omega).$$

(iii) Each minimal invariant set E_j may be decomposed into a finite number, d_j, of disjoint cyclically moving sets C_i^j $(i = 1,... d_j)$. These are such that, interpreting $C^j_{d_j+k}$ as C^j_k ,

$$P(y, C^j_{i+1}) = 1 \qquad (y \in C^j_i;\ i = 1,....d_j).$$

(iv) To each cyclically moving subset C_i^j there corresponds a probability measure $\pi_i^j(\cdot)$ such that

$$\lim_{n\to\infty} P^{(nd_j+m)}(\xi,E) = \pi_k^j(E) \quad (\xi \in C_i^j;\ E \in \mathcal{A})$$

where $k = m+i \pmod{d_j}$. The measures π_i^j have the properties that

$$\pi_i^j(C_i^j) = 1, \quad \pi_i^j(E) > 0 \quad \text{if } \phi(E \cap C_i^j) > 0.$$

(v) To each minimal invariant set E_j there corresponds a probability measure $\pi^j(\cdot)$ such that

$$\pi^j(E_j) = 1, \quad \pi^j(E) = \frac{1}{d_j} \sum_{i=1}^{d_j} \pi_i^j(E) \qquad (j = 1,\ldots p),$$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} P^{(m)}(\xi,E) = \pi^j(E) \quad (\xi \in E_j;\ E \in \mathcal{A}).$$

(vi) For every $\xi \in \Omega$, the measure $q(\xi,\cdot)$ defined by

$$q(\xi,E) = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} P^{(m)}(\xi,E)$$

is an invariant probability measure for the transition probabilities $P(\cdot,\cdot)$. A measure $\pi(\cdot)$ is an invariant measure, i.e., a solution of (2.6.2) if, and only if, it is a linear combination of the measures $\pi^j(\cdot)$.

Consideration of condition (2.6.1) for the transition probabilities associated with the autoregressive processes shows that it does not hold for many of the processes of interest. Suppose that $\lambda(x)$ is bounded on every finite interval but is such that $\lambda(x) \to -\infty$ as $x \to -\infty$ (for example, $\lambda(x) = \frac{1}{2}x$), then for any finite measure ϕ on the real line and any $\varepsilon > 0$ there is a set $A = (-\infty,a)$ with $\phi(A) \leq \varepsilon$. Then, for the transition probabilities in the form (2.1.6),

$$P_x(x,A) = \int_{z\in(-\infty,a)-\lambda(x)} dF_Z(z)$$

$$= F_Z(a - \lambda(x)) \to 1 \quad \text{as } x \to -\infty.$$

Hence (2.6.1) does not hold for m = 1: it does not seem possible to make general statements about its validity for m $\geq$ 2. In the particular case for which $\lambda(x)$ is linear, $\lambda(x) = bx$ ($0 < |b| < 1$), and for which the input distribution is normal, N(0,1) say, the transition probabilities are known exactly. These are

$$P_X^{(m)}(x,A) = [2\pi(1 - b^{2m})]^{-1/2} \int_A \exp\{ - \frac{(z-b^m x)^2}{2(1-b^{2m})} \} dz$$

and, for any fixed semi-infinite set A,

$$P_X^{(m)}(x,A) \to 1 \quad \text{either as } x \to +\infty, \text{ or as } x \to -\infty.$$

Thus the condition is invalid for what is, apparently, the simplest case. Clearly, for general autoregressive processes, a theory which applies (at least) to linear processes is needed.

Nonetheless condition (2.6.1) can be shown to hold for the transition probabilities (2.1.6) of autoregressive processes which have uniformly bounded autoregression functions and input distributions which are strictly increasing everywhere (so that any value of Z_n is possible): for then m can be taken as 1, and $\phi(\cdot)$ can be defined by any strictly positive probability density. For other input distributions it seems possible to say little in general.

2.6.2 A condition used by Harris (1956) (see also Rosenblatt, 1971, p.117) is that, for some non-trivial σ-finite measure ϕ, for all points $x \in \Omega$ and all sets $A \in \mathcal{A}$ with $\phi(A) > 0$,

$$\Pr[X_n \in A \text{ infinitely often} \mid X_0 = x] = 1. \qquad (2.6.3)$$

Under this condition there is an invariant measure ψ ($\psi(\Omega) \leq \infty$) such that ϕ is absolutely continuous with respect to ψ. An example would be the symmetric random walk on the integers. However this condition appears to be of little use since showing that it holds for a particular process would be difficult. The condition is known as ϕ-recurrence (Tweedie, 1975a). Assumptions that lead to the conclusion that there does exist an invariant probability measure for the process are that Ω is a compact Hausdorff space and that the transition probabilities are weakly continuous (see below) (Rosenblatt, 1971, p.98). However the

real line is not compact so that this cannot be applied generally. It is of possible use if $\lambda(x)$ is of finite range or if the state space can otherwise be reduced to a compact set. A related theory which holds for locally compact state spaces, but with a further assumption, gives as a result the existence of a not necessarily finite invariant measure (Rosenblatt, 1971, p.102). The existence of a non-finite (finite) invariant measure corresponds to the recurrence (respectively, positive recurrence) of discrete state space Markov Chains.

Tweedie (1975a) has obtained sufficient conditions for the stationarity of Markov processes. The first assumption is that the chain $\{X_n\}$ is ϕ-irreducible for some non-trivial measure ϕ on $\mathcal{A}$: that is

$$\phi(A) > 0 \text{ implies } \sum_{n=1}^{\infty} 2^{-n} P^{(n)}(x,A) > 0 \text{ for every } x \in \Omega. \qquad (2.6.4)$$

This just says that for every set A ($\phi(A) > 0$) there is a positive probability of eventually entering that set from every point x in the state space. This condition implies (Tweedie, 1974) the existence of a measure M such that

(i) $\{X_n\}$ is M-irreducible,

(ii) if $B \in \mathcal{A}$ is such that $M(B) = 0$, then $M(\overline{B}) = 0$, where

$$\overline{B} = \{x : \sum_{n} 2^{-n} P^{(n)}(x,B) > 0\}.$$

$\overline{B}$ is the set of points $x \in \Omega$ from which B can be reached. M can be constructed so that $M(\Omega) < \infty$ and so that it gives positive measure to any set which can be reached with positive probability from points of any set A with $\phi(A) > 0$. The measure ϕ may be such that it gives positive measure to subsets of a particular (possibly bounded) set and zero measure to sets which are disjoint to this set. Even with this allowance it seems difficult in general to find a suitable measure to satisfy (2.6.4): consider the linear processes of section 2.2.2 for which singular measures are the natural measures. However the above condition can be shown to hold in some particular cases: it clearly holds for autoregressive processes on the real line, with transition probabilities in either of the forms (2.1.6), (2.1.7), whenever the input distribution assigns positive probability to all sets of positive Lebesgue measure (with ϕ based on Lebesgue measure).

Under the assumption of ϕ-irreducibility the following dichotomy holds (Tweedie, 1974, 1975a). Either

(i) for every $x \in \Omega$ and $A \in \mathcal{A}$ with $M(A) > 0$

$$\sum_{n=1}^{\infty} P^{(n)}(x,A) = \infty,$$

in which case the process is called recurrent, or

(ii) there is a countable partition $\{A_j\}$ of Ω,

$$A_j \cap A_k = \emptyset \ (j \neq k), \quad \bigcup A_j = \Omega,$$

such that, for every $x \in \Omega$ and every $A \in \mathcal{A}$ with $A \subset A_j$ for some j,

$$\sum_{n=1}^{\infty} P^{(n)}(x,A) < \infty,$$

in which case the process $\{X_n\}$ is called transient. A similar result is obtained by Foguel (1966).

If a process is ϕ-irreducible and recurrent then there is a unique solution (up to constant multiples) of the invariance equation

$$\mu(A) = \int \mu(dy) P(y,A).$$

If the solution is a finite measure then the process $\{X_n\}$ is called ergodic or positive recurrent. Recurrence and ϕ-recurrence (2.6.3) are closely related (Tweedie, 1975a, Lemma 2.2). If $\{X_n\}$ is ergodic then the invariant measure, π say, satisfies the limit property

$$\frac{1}{n} \sum_{m=1}^{n} P^{(m)}(y,A) \to \pi(A) \quad \text{as } n \to \infty$$

for all $y \in \Omega$ excluding a set of M-measure zero. However if $\{X_n\}$ is recurrent but not ergodic, then there is a partition $\{K_j\}$ of Ω such that for every $A \in \mathcal{A}$ with $A \subset K_j$ for some j,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} P^{(m)}(x,A) = 0 \quad \text{for all } x \in \Omega.$$

Further, if the process is ergodic with invariant measure π, the expected recurrence time $R(x,A)$ is finite for π-almost all $x \in A$: here

$$R(x,A) = \sum_{n=1}^{\infty} n \, {}_AP^{(n)}(x,A)$$

with

$${}_AP^{(n)}(x,A) = \Pr[X_n \in A, X_i \in A;\ i = 1,\ldots n-1 \mid X_0 = x].$$

These definitions and results correspond to those of discrete state space Markov chains (Cox and Miller, 1965, §3.4). Tweedie (1975b) considers the effect of changes in the transition probabilities on the recurrence or ergodicity of a process.

The transition law $\{P(x,\cdot)\}$ is called strongly continuous if, for every $A \in \mathcal{B}$, $P(x,A)$ is a continuous function of x. The law $\{P(x,\cdot)\}$ is strongly continuous if and only if $\int P(x,dy)g(y)$ is a continuous bounded function of x whenever g is a bounded measurable function on Ω. The transition law $\{P(x,\cdot)\}$ is called weakly continuous if $\int P(x,dy)g(y)$ is a continuous bounded function whenever g is a continuous bounded function (Tweedie, 1975a).

In his paper, Tweedie (1975a) proves the results quoted below for strong continuity but notes that weak continuity suffices for normed state spaces which are complete in their norm. Examples of such spaces are finite dimensional Euclidean spaces, including the real line. The transition probabilities P_X given by (2.1.6) are weakly continuous whenever $\lambda(x)$ is a continuous function of x (for any input distribution), while the transition probabilities P_Y given by (2.1.7) are strongly continuous for any autoregression function $\lambda(x)$ whenever the input distribution is absolutely continuous.

When the state space is a finite dimensional Banach space with norm $||\cdot||$, and the transition probabilities are ϕ-irreducible and strongly or weakly continuous, a sufficient condition for the ergodicity of the process is the existence of constants $k, \varepsilon > 0$ such that

$$E\{||X_{n+1}|| - ||X_n|| \mid X_n = x\} \leq -\varepsilon \quad (||x|| > k) \tag{2.6.5}$$

and

$$E\{||X_{n+1}|| \,\big|\, X_n = x\} \text{ is bounded above for } ||x|| \leq k.$$

A sufficient condition for the recurrence of the process is the existence of a constant $k > 0$ such that

$$E\{||X_{n+1}|| - ||X_n|| \,\big|\, X_n = x\} \leq 0 \quad (||x|| > k). \tag{2.6.6}$$

More general conditions than the above can be given in terms of expectations of functions other than $||\cdot||$ (Tweedie, 1975a).

For autoregressive processes with transition probabilities given by (2.1.6) the conditions for ergodicity hold if the process is ϕ-irreducible, $\lambda(x)$ is continuous and there are constants $k,\varepsilon > 0$ such that

$$E\{|\lambda(x) + Z| - |x|\} \leq -\varepsilon \quad (|x| > k), \tag{2.6.7}$$

where as usual Z denotes a random variable from the input distribution. Clearly this inequality holds if $|\lambda(x)| \leq |x| - E|Z| - \varepsilon$ for $|x| > k$.

When the transition probabilities are in the form (2.1.7) the conditions for ergodicity hold if the process is ϕ-irreducible with absolutely continuous input distribution and if there exist constants $k,\varepsilon > 0$ such that

$$E\{|\lambda(x + Z)| - |x|\} \leq -\varepsilon \quad (|x| > k). \tag{2.6.8}$$

Sufficient conditions for cases in which the expectations do not exist can be obtained from the generalisations mentioned above: these can also lead to weaker conditions than the above.

Thus the existence of a stationary distribution is assured for some classes of autoregressive processes. As already noted a process is ϕ-irreducible if the input distribution has a strictly increasing absolutely continuous component. Although it is possible to demonstrate ϕ-irreducibility in other particular cases, it does seem to be a difficult condition to work with in general. Loosely speaking ϕ-irreducibility ensures that the sets of the state space which take a continuing part in the process can be identified.

A sufficient condition for the ergodicity of a process with a normal input distribution is just (2.6.8) itself since then ϕ-irreducibility holds. The condition ensures that if the process takes a large value then the next value is likely to be nearer the "centre" of the space, though in a much weaker sense than condition (2.6.1).

For multi-dimensional autoregressive processes (2.6.5) or (2.6.6) can be applied for a suitable norm. Kesten (1976) has given an analysis of recurrence and transience for multidimensional processes based on cone-shaped sets. Tweedie (1976) gives a full account of the classification of general state space Markov Chains including the above material.

3 SOME APPROACHES TO CALCULATING PROPERTIES OF STATIONARY DISTRIBUTIONS

3.1 Introduction

In the last chapter it was shown that, under certain conditions on $\lambda(\cdot)$ and the distribution F_Z of $\{Z_n\}$, the non-linear autoregressive process given by

$$X_{n+1} = \lambda(X_n) + Z_{n+1} \qquad (n = \ldots, -1,0,1,\ldots) \tag{3.1.1}$$

has a stationary distribution; that is there exists a distribution F_X satisfying

$$F_X(x) = \int_{-\infty}^{\infty} F_Z(x - \lambda(y))dF_X(y). \tag{3.1.2}$$

In some cases this solution is unique. Without assuming any particular set of conditions, it will now be assumed that a solution to this equation exists, and various ways of finding this distribution and its properties will be considered. Quantities associated with the stationary and input distributions will be distinguished by appropriate subscripts: thus μ_X, μ_Z stand for the respective means.

The methods described for solving equation (3.1.2) are of two types. Methods of the first type, discussed in section 3.2, are based on the distributions of the finite-starting-time sequence $\{U_n\}$ described earlier. The second type of method attempts to exploit the invariance of the properties of X_n $(n = \ldots, -1,0,1,\ldots)$ when the process has a stationary distribution to achieve a solution: that is it seeks F_X directly as a solution of (3.1.2) rather than as a limiting distribution as in the first method. This is described in section 3.3. However, to employ this second method effectively, the process (3.1.1) is imbedded in a family of processes which are considered together. Section 3.4 introduces one such family and properties of the stationary distributions are obtained by manipulation of the random variables of the processes. In Chapter 4 a different family of processes is used and the stationary distribution functions of the processes are considered. This is equivalent to imbedding equation (3.1.2) in a family of integral equations which are solved together.

3.2 Repeated Convolution Methods

3.2.1 As noted in section 2.1, if a sequence of random variables $\{U_n\}$, generated by the same mechanism as (3.1.1) but starting at a finite time, converges in distribution, it does so to a stationary distribution for the process (3.1.1). For all fixed distributions $F_0(u)$ for U_0, the distributions of the random variables $\{U_n\}$ generated by

$$U_{n+1} = \lambda(U_n) + Z_{n+1} \qquad (n = 0,1,2,\ldots) ,$$

are given recursively by

$$F_{n+1}(x) = \int F_Z(x - \lambda(u))dF_n(u) \qquad (n = 0,1,2,\ldots).$$

This suggests, as one method for finding a stationary distribution F_X for the process, the evaluation of the sequence of distribution functions $(F_0,F_1,F_2,\ldots)$ by numerical integration for some suitably chosen initial distribution F_0. Whether or not the sequence of distributions converges, the distributions F_i calculated may be of direct interest: for, if F_0 corresponds to U_0 taking a single value u_0, say, then F_i $(i = 1,2,3,\ldots)$ are the distributions required for making predictions of $(X_{n+1},X_{n+2},\ldots)$ given an observed value $X_n = u_0$ of the process (3.1.1). An alternative to this choice for F_0 is to take a distribution with large dispersion giving positive weight to all parts of the real line. Should there be more than one stationary distribution for the process this would ensure that all regions having a positive probability under some stationary distribution could be found. The limit of the sequence of distributions, if one existed, would be a mixture of the stationary distributions. Other tactics, such as averaging over "time", could be adopted to deal with periodic processes.

Once a stationary marginal distribution F_X of the process $\{X_n\}$ has been found, the stationary joint distributions of any finite collection $(X_n,X_{n+\ell_1}, \ldots X_{n+\ell_k})$ can be found by further numerical integrations. These integrations would be over the joint distributions of consecutive sequences $(X_n,X_{n+1}, \ldots X_{n+\ell})$; these are given in terms of densities, with the obvious notation, by

$$f_{X_n,X_{n+1},\ldots X_{n+\ell}}(x_0,x_1,\ldots.x_\ell) = f_Z(x_\ell - \lambda(x_{\ell-1}))\ldots\ldots f_Z(x_1 - \lambda(x_0))f_X(x_0).$$

By performing further numerical integrations the moments and joint moments of the process $\{X_n\}$ under the stationary distribution can be calculated. Any aspect of the stationary process $\{X_n\}$ can thus be investigated by numerical methods when enough integrations are performed.

The disadvantage of this method is that, being based entirely on numerical quadrature, it does not provide analytic expressions for any of the quantities required and thus no information as to the results for a particular autoregression function is available unless the above procedure is carried out for this function. Some problems arise from the numerical integrations themselves but these are purely procedural and, in principle, there should be no difficulty in overcoming them.

3.2.2 The above method is based on the assumption that, if F_0 is a distribution close to the stationary distribution, then F_1 given by

$$F_1(x) = \int F_Z(x - \lambda(u))dF_0(u) \qquad (3.2.1)$$

will be closer. For suitable choices of F_0 it might be possible to evaluate this integral analytically, or possibly the corresponding integral when densities exist. The distribution F_0 would be chosen to be itself close to the stationary distribution and (3.2.1) would produce a distribution F_1 which, while not exactly the required distribution, would at least give some indication as to its properties. These properties would often be summarised by the moments and these would be approximations to the moments of the stationary distribution. Even when the integral (3.2.1) for the distribution cannot be evaluated it may be possible to perform the corresponding integrals for the moments exactly. For example, the approximation to the mean is

$$\int \lambda(u)dF_0(u) + \mu_Z ,$$

or $E[\lambda(U_0) + Z]$, where U_0 has distribution F_0 and Z has distribution F_Z.

As a particular case let $\lambda(x) = \exp\{-\frac{1}{2}x^2\}$ and let the input distribution be $N(0,1)$. Consideration of $\lambda(x)$ shows that the mean μ_X of the stationary distribution must satisfy $0 \leq \mu_X \leq 1$ and that a reasonable guess would be $\mu_X \simeq 0.6$. The variance of the stationary distribution must be at least as great as that of the input distribution and the variance of a random variable $(\lambda(X_n))$ concentrated on $[0,1]$ cannot be

greater than 1/4: hence

$$1 \leq \sigma_X^2 = \text{var } X_{n+1} = \text{var } \lambda(X_n) + \text{var } Z_{n+1} \leq 1.25.$$

With the distribution F_0 chosen to be a normal distribution with mean μ and variance σ^2, the integrals for the moments of the distribution F_1 can be evaluated explicitly and give for the mean, variance and third central moment respectively

$$\mu' = (1 + \sigma^2)^{-1/2} \exp\{-\frac{1}{2}\frac{\mu^2}{1+\sigma^2}\} ,$$

$$\sigma'^2 = (2 + \sigma^2)^{-1/2} \exp\{-\frac{1}{2}\frac{2\mu^2}{2+\sigma^2}\} - \mu'^2 + 1,$$

$$\mu_3' = (3 + \sigma^2)^{-1/2} \exp\{-\frac{1}{2}\frac{3\mu^2}{3+\sigma^2}\} - 3\mu'(\sigma'^2 - 1) - \mu'^3.$$

One way of proceeding is to choose μ, σ^2 so that $\mu' = \mu$ and $\sigma'^2 = \sigma^2$ and this solution can be found by an iterative substitution method using the first two equations. This gives the approximations

$$\mu' = 0.62758, \quad \sigma'^2 = 1.1060, \quad \mu_3' = -0.0193$$

compared with the true moments

$$\mu_X = 0.62756, \quad \sigma_X^2 = 1.0995, \quad \mu_{3,X} = -0.0151.$$

These moments have been calculated by a method to be described later and have been confirmed by computer simulations. The above procedure finds a normal distribution, the first two moments of which are left unchanged by the transformation $X \to \lambda(X) + Z$. Since the stationary distribution is itself close to normal in this case, this may explain the apparent success of the method here. More generally F_0 might be chosen as a mixture of distributions of tractable form. The above method of equating moments of the distributions before and after the transformation takes account of only these moments and no other properties of the distributions.

A similar method may be used to obtain approximate results for the joint moments of (X_n, X_{n+1}) under the stationary distribution of the powers (3.1.1). This joint distribution is approximated by that of $(U_0, \lambda(U_0) + Z)$ where U_0 has the approximate stationary distribution F_0. In the above example, the value found for the approximation to $\mathrm{cov}(X_n, X_{n+1})$ when F_0 is $N(\mu, \sigma^2)$ is

$$c_1' = -\sigma^2(1+\sigma^2)^{-3/2}\,\mu \exp\{-\frac{1}{2}\frac{\mu^2}{1+\sigma^2}\} = -\frac{\sigma^2}{1+\sigma^2}\,\mu\mu'$$

and, with μ, σ^2 chosen as above, this gives $\mathrm{cov}(X_n, X_{n+1}) \simeq -0.2068$ compared with the true value -0.2056.

Although the method is based principally on the one step ahead transformation of distribution, the choice of the initial distribution F_0 is based on equating moments of the distributions at successive times.

3.3 Invariance Methods

An attempt will now be made to use the equivalence of the distributions of the random variables $X, \lambda(X) + Z$ to find the exact or approximate properties of the stationary distribution. Here X and Z are random variables from the stationary and input distributions respectively.

One method of obtaining approximations for the moments of the stationary distribution is to use the well-known formulae approximating the moments of a function of a random variable. These are obtained by Taylor series expansions of powers of $\lambda(X)$ about the mean μ_X of X. Truncating the expansions at the usual point gives the approximations

$$E\{\lambda(X)\} \simeq \lambda(\mu_X) + \frac{1}{2}\sigma_X^2\lambda^{(2)}(\mu_X) \qquad (3.3.1)$$

and

$$\mathrm{var}\{\lambda(X)\} \simeq \{\lambda^{(1)}(\mu_X)\}^2\sigma_X^2. \qquad (3.3.2)$$

where $\lambda^{(r)}(x)$ is the r'th derivative of $\lambda(x)$ which function is assumed differentiable as often as necessary. Thus the following equations hold approximately

$$\mu_X = \lambda(\mu_X) + \frac{1}{2}\sigma_X^2\lambda^{(2)}(\mu_X) + \mu_Z,$$

$$\sigma_X^2 = \{\lambda^{(1)}(\mu_X)\}^2\sigma_X^2 + \sigma_Z^2,$$

and from these an approximation to the mean of the stationary distribution is as a solution of

$$\mu_X = \lambda(\mu_X) + \mu_Z + \frac{1}{2}\frac{\lambda^{(2)}(\mu_X)}{1 - \{\lambda^{(1)}(\mu_X)\}^2}\sigma_Z^2, \qquad (3.3.3)$$

with an approximation for the variance given by

$$\sigma_X^2 = \frac{\sigma_Z^2}{1 - \{\lambda^{(1)}(\mu_X)\}^2} \; . \qquad (3.3.4)$$

For the example, used also in section 3.2.2, for which $\lambda(x) = \exp\{-\frac{1}{2}x^2\}$ with input distribution N(0,1), the approximation for the mean is as a solution of the equation

$$\mu = \left[1 + \frac{1}{2} \cdot \frac{\mu^2 - 1}{1 - \mu^2 e^{-\mu^2}}\right] e^{-\frac{1}{2}\mu^2} \; .$$

This solution is rapidly found by repeatedly substituting approximations into the right-hand side. The approximations for the mean and variance are found to be 0.4715 and 1.217 respectively compared with the exact results 0.6276 and 1.099. The root of the equation $\mu = \lambda(\mu)$ is 0.7531 so that the modification (3.3.3) of this equation has resulted in a change in the right direction but one that is rather too large.

The approximations derived by this method require the solution of equation (3.3.3) for μ_X: this is a disadvantage for it is not clear when the equation would have exactly one solution. Also as has been seen, in this case at least, the approximations compare unfavourably with those obtained earlier. It may be noted that the only information used about the input distribution is its mean and variance.

In view of the poor results at this order of approximation one might attempt to extend the result by equating moments of higher order. If the series are truncated to retain terms of order not greater than 3 (relative to the scale of X) then, for a normal input distribution the equations reduce exactly to (3.3.3) and (3.3.4). However retaining terms of 4'th order gives the equations

$$\mu_X = \lambda + \frac{1}{2}\sigma_X^2\lambda^{(2)} + \frac{1}{6}\mu_{3X}\lambda^{(3)} + \frac{1}{24}\mu_{4X}\lambda^{(4)} + \mu_Z \qquad (3.3.5)$$

$$\sigma_X^2 = \sigma_X^2\lambda^{(1)2} + \mu_{3X}\lambda^{(1)}\lambda^{(2)} + \frac{1}{4}(\mu_{4X}-\sigma_X^4)\lambda^{(2)2} + \frac{1}{3}\mu_{4X}\lambda^{(1)}\lambda^{(3)} + \sigma_Z^2$$

$$\mu_{3X} = \mu_{3X}\lambda^{(1)3} + \frac{3}{2}(\mu_{4X}-\sigma_X^4)\lambda^{(1)2}\lambda^{(2)} + \mu_{3Z}$$

$$\mu_{4X} = \mu_{4X}\lambda^{(1)4} + 6[\sigma_X^2\lambda^{(1)2} + \mu_{3X}\lambda^{(1)}\lambda^{(2)} + \frac{1}{4}(\mu_{4X}-\sigma_X^4)\lambda^{(2)2} + \frac{1}{3}\mu_{4X}\lambda^{(1)}\lambda^{(3)}]\sigma_Z^2 + \mu_{4Z}$$

$$= \mu_{4X}\lambda^{(1)4} + 6[\sigma_X^2 - \sigma_Z^2]\sigma_Z^2 + \mu_{4Z}$$

where λ and its derivatives are evaluated at μ_X and μ_{3X}, μ_{4X} are the third and fourth central moments of the stationary distribution. For a fixed value of μ_X the last three equations lead to a quadratic equation for σ_X^2 and from this μ_{3X}, μ_{4X} can be found. Using these quantities to substitute in (3.3.5) produces a new approximation for the mean μ_X and thus an iterative scheme of solution is obtained.

Unfortunately this does not seem to be an improvement in the method. For the above example the method converges to a solution for the first four moments as

0.62600, 0.7948, -0.2220, 1.9022

compared with their true values

0.62756, 1.09945, -0.0151, 3.6154.

The second root of the quadratic for σ_X^2 is near -121 when μ_X is at the above solution. The approximation for the mean seems good but equally good approximations are found by inserting any reasonable values for $\mu_X, \sigma_X^2, \mu_{3X}, \mu_{4X}$ into the right-hand side of (3.3.5). Taking $\mu_X = 0$, $\sigma_X^2 = 1$, $\mu_{3X} = 0$, $\mu_{4X} = 3$ in the example gives 0.625 for the mean.

The Taylor series expansions however are based on the assumption that $(X - \mu_X)$ and its moments are small although in this situation it is not clear what this might mean. The usual applications of formulae (3.3.1) and (3.3.2) are to quantities derived from samples of size n, say,

and the sizes of the terms in the expansion are compared with n^{-1} (as $n \to \infty$) (Kendall and Stuart, 1969, p.232). In particular r'th order moments are usually comparable with $n^{-r/2}$. For the present use there is no such parameter available and therefore the truncation of the expansions as above is not really justified. Possibly a truncation based on different reasoning would be better. It may be noted that the approximations obtained from (3.3.3) and (3.3.4) are exact when the autoregression function is linear, and this might suggest that the method should work for functions which are nearly linear or for which the variance of the stationary distribution is small so that the function is effectively linear in the most important regions. In the next section a method is introduced for incorporating a parameter in terms of which expansions may be made.

3.4 A family of processes with changing input variance

3.4.1 Consider the autoregressive process

$$X_{n+1} = \lambda(X_n) + Z_{n+1} \qquad (n = \ldots -1,0,1,\ldots) \qquad (3.4.1)$$

as before, where it is now assumed that the input distribution has zero mean and moments of all orders and that $\lambda(x)$ is continuously differentiable. If the variance σ_Z^2 of the input distribution is zero then the process $\{X_n\}$ is generated by

$$X_{n+1} = \lambda(X_n) \qquad (n = \ldots -1,0,1,\ldots) \; . \qquad (3.4.2)$$

The assumptions needed are that the equation $x = \lambda(x)$ has exactly one root, ξ say, which is "stable", and that there are no periodic stationary sequences satisfying (3.4.2) so that the only possible stationary sequence is that which takes the fixed value ξ ($\xi = \lambda(\xi)$).

Now suppose that the variance of Z_n is very small. Then the above assumptions make it reasonable to assume that the variance of the stationary sequence is correspondingly small. If there were more than one possible stationary process satisfying (3.4.2) this would not necessarily hold. It is also reasonable to assume that the stationary distributions change continuously as the scale of the input distribution changes. This suggests considering the family of processes $\{X_n(\alpha)\}$ generated by

$$X_{n+1}(\alpha) = \lambda\{X_n(\alpha)\} + \alpha Z_{n+1} \qquad (n = \ldots, -1,0,1,\ldots) \qquad (3.4.3)$$

for α in some interval containing [0,1]. For the present it is assumed that the same values of the input sequence are used to generate the sequences $\{X_n(\alpha)\}$ for different α. By considering a version of (3.4.3) starting in the remote past with the same value for each sequence indexed by α, it is reasonable to expand $X_n(\alpha)$, for each n, as a power-series in α. Thus

$$X_n(\alpha) = X_n^{(0)} + \alpha X_n^{(1)} + \tfrac{1}{2}\alpha^2 X_n^{(2)} + \ldots.$$

This may be substituted into (3.4.3) giving the expansion

$$X_{n+1}^{(0)} + \alpha X_{n+1}^{(1)} + \tfrac{1}{2}\alpha^2 X_{n+1}^{(2)} + \ldots.. = \lambda(X_n^{(0)}) + (\alpha X_n^{(1)} + \tfrac{1}{2}\alpha^2 X_n^{(2)} + \ldots)\lambda^{(1)}(X_n^{(0)})$$
$$+ \tfrac{1}{2}(\alpha X_n^{(1)} + \tfrac{1}{2}\alpha^2 X_n^{(2)} + \ldots)^2 \lambda^{(2)}(X_n^{(0)})$$
$$+ \ldots \qquad + \alpha Z_{n+1}.$$

The equating of powers of α in this expression leads to formulae generating the sequences $\{X_n^{(0)}\}$, $\{X_n^{(1)}\}$, etc. The first of these is $X_{n+1}^{(0)} = \lambda(X_n^{(0)})$. Intuitively at least, if all the sequences $\{X_n(\alpha)\}$ are to be stationary then each of $\{X_n^{(0)}\}$, $\{X_n^{(1)}\}$, $\{X_n^{(2)}\}$ etc., must also be stationary. With this assumption the sequence $\{X_n^{(0)}\}$ must take the fixed value ξ and then the other expressions become, with λ and its derivatives evaluated at ξ,

$$X_{n+1}^{(1)} = \lambda^{(1)} X_n^{(1)} + Z_{n+1} \qquad (3.4.4)$$

$$X_{n+1}^{(2)} = \lambda^{(1)} X_n^{(2)} + \lambda^{(2)} X_n^{(1)2} \qquad (3.4.5)$$

$$X_{n+1}^{(3)} = \lambda^{(1)} X_n^{(3)} + 3\lambda^{(2)} X_n^{(1)} X_n^{(2)} + \lambda^{(3)} X_n^{(1)3} \qquad (3.4.6)$$

$$X_{n+1}^{(4)} = \lambda^{(1)} X_n^{(4)} + \lambda^{(2)}\{3X_n^{(2)2} + 4X_n^{(1)} X_n^{(3)}\} + 6\lambda^{(3)} X_n^{(1)2} X_n^{(2)} + \lambda^{(4)} X_n^{(1)4} \qquad (3.4.7)$$

The sequences $\{X_n^{(r)};\ n = \ldots -1,0,1\ldots\}$ $(r = 0,1,2,\ldots)$ will be referred to as the component processes of $\{X_n(\alpha);\ n = \ldots -1,0,1\ldots\}$. Since Z_{n+1} is independent of $X_k(\alpha)$ $(k \leq n)$ for all α, it must also be

independent of $X_k^{(r)}$ ($k \leq n$; $r = 0,1,2,\ldots$). However it is clear that for k fixed, $X_k^{(r)}$ ($r = 0,1,2,\ldots$) are dependent amongst themselves because, for example, $X_k^{(1)}$ enters into the generation of $X_{k+1}^{(1)}$, $X_{k+1}^{(2)}$, $X_{k+1}^{(3)}$, etc. Because of the stationarity of the sequences the moments of the component processes at different times may be equated, and, by suitably multiplying the above equations together and taking expected values in the resulting expressions, these moments may be found when $|\lambda^{(1)}(\xi)| < 1$.

3.4.2 Define $q_r = (1 - \lambda^{(1)r})^{-1}$ ($r = 1,2,3,\ldots$) and $p_1 = \lambda^{(1)}q_1$, $p_2 = \lambda^{(2)}q_2$, $p_{11} = \lambda^{(1)2}q_2$, and more generally for k subscripts

$$p_{r_1,\ldots r_k} = \lambda^{(r_1)}\lambda^{(r_2)}\ldots\ldots\lambda^{(r_k)}q_{r_1+r_2+\ldots r_k}.$$

Then the moments of the random variables $X_n^{(r)}$ may be found in the following order

$$E(X_n^{(1)}) = 0,$$

$$E(X_n^{(1)2}) = q_2\sigma_Z^2,$$

$$E(X_n^{(1)3}) = q_3\mu_{3Z},$$

$$E(X_n^{(1)4}) = q_4(\mu_{4,Z} + 6p_{11}\sigma_Z^4) = q_4 s_4, \text{ say},$$

$$E(X_n^{(2)}) = q_1p_2\sigma_Z^2$$

$$E(X_n^{(1)}X_n^{(2)}) = q_2p_{12}\mu_{3Z}$$

$$E(X_n^{(3)}) = q_1(3p_{12}p_2 + p_3)\mu_{3Z}$$

$$E(X_n^{(1)2}X_n^{(2)}) = q_3(p_{112}s_4 + q_1p_2\sigma_Z^4) = q_3s_4', \text{ say},$$

$$E(X_n^{(2)2}) = q_2(2p_{12}s_4' + p_{22}s_4),\ E(X_n^{(1)}X_n^{(3)}) = q_2(3p_{12}s_4' + p_{13}s_4),$$

$$E(X_n^{(4)}) = q_1\{6(3p_2p_{12} + p_3)s_4' + (4p_2p_{13} + 3p_2p_{22} + p_4)s_4\}.$$

Here, for example, $E(X_n^{(2)2})$, $E(X_n^{(1)}X_n^{(3)})$, $E(X_n^{(1)2}X_n^{(2)})$ and $E(X_n^{(1)4})$ are all needed to find $E(X_n^{(4)})$.

In order to find approximations for the moments of the process $\{X_n\}$ given by (3.4.1), which is $\{X_n(1)\}$ in (3.4.3), there are two choices.

The first is to express the particular moment of $X_n(\alpha)$ as a power-series in α and to truncate this at some particular power of α and set $\alpha = 1$. For example

$$\begin{aligned}\mathrm{var}(X_n(\alpha)) &= \mathrm{var}(\xi + \alpha X_n^{(1)} + \tfrac{1}{2}\alpha^2 X_n^{(2)} + \tfrac{1}{6}\alpha^3 X_n^{(3)} + \ldots.) \\ &= \alpha^2 \mathrm{var}(X_n^{(1)}) + \alpha^3 \mathrm{cov}(X_n^{(1)}, X_n^{(2)}) + \tfrac{1}{4}\alpha^4 \mathrm{var}(X_n^{(2)}) \\ &\quad + \tfrac{1}{3}\alpha^4 \mathrm{cov}(X_n^{(1)}, X_n^{(3)}) + \tfrac{1}{6}\alpha^5 \mathrm{cov}(X_n^{(2)}, X_n^{(3)}) + \ldots..,\end{aligned}$$

which, truncated to exclude powers greater than 4, gives the approximation

$$\sigma_X^2 \simeq \mathrm{var}(X_n^{(1)}) + \mathrm{cov}(X_n^{(1)}, X_n^{(2)}) + \tfrac{1}{4}\mathrm{var}(X_n^{(2)}) + \tfrac{1}{3}\mathrm{cov}(X_n^{(1)}, X_n^{(3)}). \qquad (3.4.8)$$

The alternative is to approximate the random variable X_n directly by truncating the expansion of $X_n(\alpha)$ at a convenient power. Truncating this after the second power of α gives an approximating random variable as

$$\hat{X}_n = \xi + X_n^{(1)} + \tfrac{1}{2} X_n^{(2)}$$

and this gives as an approximation to the variance

$$\sigma_X^2 \simeq \mathrm{var}(\hat{X}_n) = \mathrm{var}(X_n^{(1)}) + \mathrm{cov}(X_n^{(1)}, X_n^{(2)}) + \tfrac{1}{4}\mathrm{var}(X_n^{(2)}).$$

By comparison with (3.4.8) it is readily seen that this is not an expansion correct to any particular power of α; this happens in general. However this method of truncating the expansion of $X_n(\alpha)$ has the advantage of always producing approximations for the moments which are always the moments of some proper random variable. Thus the approximations for the variance obtained in this way are always non-negative, which is not true of those found by the first method. If the input distribution is symmetric $\mathrm{cov}(X_n^{(1)}, X_n^{(2)}) = 0$, and so in these cases the first few approximations for the variance must increase.

3.4.3 Approximations for joint moments of the stationary process $\{X_n\}$ may also be found, for example, by approximating the joint distribution of $(X_n(\alpha), X_{n+k}(\alpha))$ by that of $(X_n^{(0)} + \alpha X_n^{(1)} + \frac{1}{2}\alpha^2 X_n^{(2)}, X_{n+k}^{(0)} + \alpha X_{n+k}^{(1)} + \frac{1}{2}\alpha^2 X_{n+k}^{(2)})$. In the same way as before the following expressions are obtained for $k \geq 1$

$$\text{cov}(X_n^{(1)}, X_{n+k}^{(1)}) = \lambda^{(1)k} q_2 \sigma_Z^2,$$

$$\text{cov}(X_n^{(2)}, X_{n+k}^{(1)}) = \lambda^{(1)k} q_2 p_{12} \mu_{3Z},$$

$$\text{cov}(X_n^{(1)}, X_{n+k}^{(1)2}) = \lambda^{(1)2k} q_3 \mu_{3Z},$$

$$\text{cov}(X_n^{(1)}, X_{n+k}^{(2)}) = \lambda^{(1)k} q_2 p_{12} \mu_{3Z} + \lambda^{(1)k-1}(1 - \lambda^{(1)k})\lambda^{(2)} q_1 q_3 \mu_{3,Z},$$

$$\text{cov}(X_n^{(2)}, X_{n+k}^{(1)2}) = \lambda^{(1)2k}(q_3 s_4' - q_1 q_2 p_2 \sigma_Z^4),$$

$$\text{cov}(X_n^{(2)}, X_{n+k}^{(2)}) = \lambda^{(1)k-1}\{\lambda^{(1)} q_2 (2p_{12} s_4' + p_{22} s_4) + (1 - \lambda^{(1)k})\lambda^{(2)} q_1 q_3 s_4'$$
$$- (1 + \lambda^{(1)} - \lambda^{(1)k}) q_1^2 p_2^2 \sigma_Z^4\}.$$

These may be used to approximate $\text{cov}(X_n(\alpha), X_{n+k}(\alpha))$ by $\text{cov}(\alpha X_n^{(1)} + \frac{1}{2}\alpha^2 X_n^{(2)}, \alpha X_{n+k}^{(1)} + \frac{1}{2}\alpha^2 X_{n+k}^{(2)})$ (since $X_n^{(0)} = \xi$ is fixed). In particular, for $k = 1$, this gives

$$\text{cov}(X_n(\alpha), X_{n+1}(\alpha)) \simeq \alpha^2 \lambda^{(1)} q_2 \sigma_Z^2 + \alpha^3(\lambda^{(1)} q_2 p_{12} + \tfrac{1}{2}\lambda^{(2)} q_3)\mu_{3,Z}$$
$$+ \tfrac{1}{4}\alpha^4\{\lambda^{(1)} q_2 (2p_{12} s_4' + p_{22} s_4) + \lambda^{(2)} q_3 s_4' - q_1^2 p_2^2 \sigma_Z^4\}$$

and, to the same order in α,

$$\text{var}(X_n(\alpha)) \simeq \alpha^2 q_2 \sigma_Z^2 + \alpha^3 q_2 p_{12} \mu_{3Z} + \tfrac{1}{4}\alpha^4\{q_2(2p_{12} s_4' + p_{22} s_4) - q_1^2 p_2^2 \sigma_Z^4\}.$$

Hence

$$\text{corr}(X_n(\alpha), X_{n+1}(\alpha)) \simeq \lambda^{(1)} + \frac{\frac{1}{2}\alpha\lambda^{(2)} q_3 \mu_{3,Z} + \frac{1}{4}\alpha^2\{\lambda^{(2)} q_3 s_4' - q_1 p_2^2 \sigma_Z^4\}}{q_2 \sigma_Z^2 + \alpha q_2 p_{12} \mu_{3Z} + \frac{1}{4}\alpha^2\{q_2(2p_{12} s_4' + p_{22} s_4) - q_1^2 p_2^2 \sigma_Z^4\}}$$

$$= \lambda^{(1)} + \frac{1}{2}\alpha \frac{\lambda^{(2)} q_3 \mu_{3Z}}{q_2 \sigma_Z^2} + \frac{1}{4}\alpha^2 \frac{\lambda^{(2)} q_3 s_4' - q_1 p_2^2 \sigma_Z^4 - 2p_{12}(\mu_{3Z}/\sigma_Z^2)}{q_2 \sigma_Z^2} + \cdots$$

3.4.4 In the special case that $\lambda^{(1)}(\xi) = 0$, the random variables $X_n^{(r)}$ may be expressed in terms of a finite number of the values of the input sequence $\{Z_n\}$. Thus, from (3.4.4)-(3.4.7),

$$X_{n+1}^{(1)} = Z_{n+1}$$

$$X_{n+1}^{(2)} = \lambda^{(2)} X_n^{(1)2} = \lambda^{(2)} Z_n^2$$

$$X_{n+1}^{(3)} = 3\lambda^{(2)2} Z_n Z_{n-1}^2 + \lambda^{(3)} Z_n^3$$

$$X_{n+1}^{(4)} = \lambda^{(2)3}(3Z_{n-1}^4 + 12 Z_n Z_{n-1} Z_{n-2}^2) + \lambda^{(2)}\lambda^{(3)}(4 Z_n Z_{n-1}^3 + 6 Z_n^2 Z_{n-1}^2) + \lambda^{(4)} Z_n^4 .$$

From expressions such as these it would be possible to find the exact distribution of any of the random variables $\xi + X_n^{(1)}$, $\xi + X_n^{(1)} + \frac{1}{2}X_n^{(2)}$, $\xi + X_n^{(1)} + \frac{1}{2}X_n^{(2)} + \frac{1}{6}X_n^{(3)}$, ... which approximate the random variable $X_n = X_n(1) = \Sigma \frac{1}{r!} X_n^{(r)}$. It would also be possible to find approximations for the joint distributions of the sequence $\{X_n\}$ by the same method.

When $\lambda^{(1)}(\xi) \neq 0$ each $X_n^{(r)}$ $(r = 1,2,3,\ldots)$ may be expressed as a summation involving an infinite number of $\{Z_n, Z_{n-1}, Z_{n-2}, \ldots\}$ and, if a similar method to that above is to be applied, it is not clear whether some truncation of these series would be suitable.

3.4.5 An alternative method of finding the stationary marginal distribution would be to attempt a series expansion of the stationary distributions of the $\{X_n(\alpha)\}$ processes. Once again a truncated version of this expansion is not the same as the distribution of a truncated expansion of $X_n(\alpha)$; indeed it is generally not a distribution function at all. However the truncated version may be easy to find. Let $F_X(x;\alpha)$, $\phi_X(s;\alpha)$ denote the stationary marginal distribution function and the corresponding characteristic function of the process $\{X_n(\alpha)\}$ given by (3.4.3). Then

$$\phi_X(s;\alpha) = \int e^{isx} dF_X(x;\alpha) = \phi_Z(\alpha s) \int e^{is\lambda(x)} dF_X(x;\alpha) . \qquad (3.4.9)$$

Expanding F_X and ϕ_X as power series in α,

$$F_X(x;\alpha) = F_0(x) + \alpha F_1(x) + \alpha^2 F_2(x) + \ldots \qquad (3.4.10)$$

$$\phi_X(s;\alpha) = \phi_0(s) + \alpha\phi_1(s) + \alpha^2\phi_2(s) + \ldots$$

and assuming that F_Z has moments of all orders to that $\phi_Z(s)$ is continuously differentiable, these equations may be solved formally by equating powers of α. However this leads to expansions in terms of the Dirac delta function and its derivatives. The function $F_0(x)$ is a unit step function at ξ and, when $\mu_Z = 0$, $F_1(x) = 0$, and

$$\phi_2(s) = \{i\lambda^{(2)}q_1 s - s^2\} \frac{q_2\sigma_Z^2}{2} e^{i\xi s}$$

$$dF_2(x) = \frac{q_2\sigma_Z^2}{2}\{-\lambda^{(2)}q_1\delta^{(1)}(x-\xi) + \delta^{(2)}(x-\xi)\}dx . \qquad (3.4.11)$$

The expansion (3.4.10) of the distribution function does not give a useful approximation as it stands; however it may be employed to produce other more useful approximations. Inserting the power series expression into (3.4.9) produces for the characteristic function

$$\phi_X(s;\alpha) = \phi_Z(\alpha s)\{e^{is\xi} + \alpha^2 \frac{q_2\sigma_Z^2}{2}(i\lambda^{(2)}p_1 s - \lambda^{(1)2}s^2)e^{is\xi} + \ldots\}$$

This is not a power series expansion but, on truncating and setting $\alpha = 1$, it gives a function to which corresponds the approximation

$$f_X(x) \simeq f_Z(x - \xi) - \frac{1}{2}p_1p_2\sigma_Z^2 f_Z^{(1)}(x-\xi) - \frac{1}{2}p_{11}\sigma_Z^2 f_Z^{(2)}(x-\xi)$$

for the stationary marginal density function in terms of the input density function f_Z.

In other methods, to be presented in chapter 4, it will be seen that expansions of distribution functions $F(x;\alpha)$ in powers of α involve derivatives of $F(x;0)$ with respect to x. In this case $F(x;0)$ is a step function so that, in one sense, the δ-functions appearing in (3.4.11) are caused by this choice of $F(x;0)$. An alternative is to consider a rescaling of $X_n(\alpha)$ given by

$$W_n(\alpha) = \frac{X_n(\alpha) - \xi}{\alpha} .$$

Then corresponding to (3.4.3) is the generating relation

$$W_{n+1}(\alpha) = \frac{1}{\alpha}\{\lambda(\xi + \alpha W_n(\alpha)) - \xi\} + Z_{n+1} \quad (n = \ldots -1,0,1,\ldots).$$

As $\alpha \to 0$, this becomes

$$W_{n+1}(0) = \lambda^{(1)}(\xi)W_n(0) + Z_{n+1} \qquad (n = \ldots -1,0,1,\ldots),$$

so that in this case $F_W(x,0)$ is a non-degenerate distribution function being the stationary distribution of a linear autoregressive process.

3.4.6 In the above three assumptions have been made, first that the input distribution has zero mean, secondly that the equation $x = \lambda(x)$ has a single root ξ with the conditions discussed in section 3.4.1 holding, and thirdly that $|\lambda^{(1)}(\xi)| < 1$. The first two assumptions may be weakened considerably by forming, instead of (3.4.3), the family of processes $\{U_n(\alpha)\}$ given by

$$U_{n+1}(\alpha) = \lambda(U_n(\alpha)) - a + \alpha(Z_{n+1} + a) \qquad (n = \ldots -1,0,1,\ldots).$$

Here a is any fixed real number and $\{Z_n\}$ may have any mean. Once again $\{U_n(1)\}$ corresponds to the process $\{X_n\}$ given by (3.4.1). The same methods may then be applied and similar approximations obtained provided that the equation $x = \lambda(x) - a$ has a single root ξ_a (and also that the discussion earlier applies) and that $|\lambda^{(1)}(\xi_a)| < 1$. Obviously these conditions may hold for many different values of a. The question of the choice of a will not be discussed except to suggest that it be chosen so that the first order approximations to the mean and variance $(\xi_a + (\mu_Z + a)/\{1 - \lambda^{(1)}(\xi_a)\}$, $\sigma_Z^2/\{1 - \lambda^{(1)2}(\xi_a)\}$, respectively) should be close to some intuitively reasonable values.

The reason for the assumption that there is only one stationary sequence $\{X_n\}$ satisfying

$$X_{n+1} = \lambda(X_n) \qquad (n = \ldots,-1,0,1,\ldots), \tag{3.4.12}$$

is so that there is only a small effect on the stationary distribution caused by introducing a small-scale additive error. That is, the stationary distribution of the process

$$X_{n+1} = \lambda(X_n) + Z_{n+1} \qquad (n = \ldots, -1,0,1,\ldots) \tag{3.4.13}$$

is almost the same as that of (3.4.12) when the scale of the Z-sequence is small. When there is more than one stationary sequence $\{X_n\}$ satisfying (3.4.12) any mixture of the corresponding distribution is also a stationary distribution. The problem is to know which of these mixtures is the limiting stationary distribution of (3.4.13) as the scale of $\{Z_n\}$ decreases to zero. This limiting distribution must be known in order to find the moments of the components $X_n^{(0)}$, $X_n^{(1)}$, $X_n^{(2)}$,... from the equations

$$X_{n+1}^{(0)} = \lambda(X_n^{(0)})$$

$$X_{n+1}^{(1)} = \lambda^{(1)}(X_n^{(0)})X_n^{(1)} + Z_{n+1}$$

$$X_{n+2}^{(2)} = \lambda^{(1)}(X_n^{(0)})X_n^{(2)} + \lambda^{(2)}(X_n^{(0)})X_n^{(1)2}, \text{ etc.}$$

However there are cases in which it is clear which mixture of stationary distributions is appropriate. Suppose that $\lambda(x)$ is odd about zero, $\lambda(-x) = -\lambda(x)$, and that the equation $x = \lambda(x)$ has three roots, one at zero and the others at ξ and $-\xi$, and suppose that $|\lambda^{(1)}(\xi)| < 1$ and $|\lambda^{(1)}(0)| > 1$. Suppose further that the distribution function of the input sequence is symmetric about zero. Then clearly the stationary distribution is itself symmetric about zero, and, as the variance of the input distribution decreases, the stationary distribution becomes more concentrated about the roots ξ and $-\xi$ but not about zero since this root is "unstable". Hence it may be concluded that the appropriate distribution for $\{X_n^{(0)}\}$ is given by

$$\text{pr}\{X_n^{(0)} = \xi\} = \text{pr}\{X_n^{(0)} = -\xi\} = \frac{1}{2}$$

with $X_{n+1}^{(0)}$ taking the same value as $X_n^{(0)}$ (with probability 1). Thus $\{X_n^{(0)}\}$ is completely deterministic. One way of proceeding is to first find the required moments conditional on $X_n^{(0)}$. These are essentially those found earlier when $\text{Pr}\{X_n^{(0)} = \xi\} = 1$. Because of the above assumptions $\lambda^{(1)}(X_n^{(0)}) = \lambda^{(1)}(\xi)$ whether $X_n^{(0)} = \xi$ or $-\xi$. Therefore, with the same notation as before,

$$E(X_n^{(1)} \mid X_n^{(0)}) = 0,$$

$$E(X_n^{(1)2} \mid X_n^{(0)}) = q_2\sigma_Z^2,$$

$$E(X_n^{(1)3} \mid X_n^{(0)}) = 0,$$

$$E(X_n^{(1)4} \mid X_n^{(0)}) = q_4(\mu_{4Z} + 6p_{11}\sigma_Z^4),$$

$$E(X_n^{(2)} \mid X_n^{(0)}) = q_1q_2\lambda^{(2)}(X_n^{(0)})\sigma_Z^2,$$

$$E(X_n^{(1)}X_n^{(2)} \mid X_n^{(0)}) = 0, \quad E(X_n^{(3)} \mid X_n^{(0)}) = 0,$$

$$E(X_n^{(2)2} \mid X_n^{(0)}) = q_2q_3q_4\{\mu_{4Z}(1+\lambda^{(1)3}) + 2\lambda^{(1)}(1+3\lambda^{(1)}-2\lambda^{(1)2}+3\lambda^{(1)3})q_1\sigma_Z^4\}\{\lambda^{(2)}(X_n^{(0)})\}^2.$$

Since $\lambda^{(2)}(\xi) = -\lambda^{(2)}(-\xi)$, the following formulae for the unconditional moments hold.

$$E(X_n^{(1)}) = E(X_n^{(2)}) = E(X_n^{(3)}) = 0,$$

$$E(X_n^{(1)2}) = q_2\sigma_Z^2, \quad E(X_n^{(1)3}) = 0,$$

$$E(X_n^{(1)4}) = q_4(\mu_{4Z} + 6p_{11}\sigma_Z^4), \quad E(X_n^{(1)}X_n^{(2)}) = 0,$$

$$E(X_n^{(2)2}) = q_2q_3p_{22}\{\mu_{4Z}(1+\lambda^{(1)3}) + 2\lambda^{(1)}(1+3\lambda^{(1)}-2\lambda^{(1)2}+3\lambda^{(1)3})q_1\sigma_Z^4\},$$

The following moments are also required

$$E(X_n^{(0)}) = E(X_n^{(0)}X_n^{(1)}) = 0, \quad E(X_n^{(0)2}) = \xi^2,$$

$$E(X_n^{(0)}X_n^{(2)}) = q_1p_2\xi\sigma_Z^2.$$

From these it may be seen that the mean of any truncation of $X_n = X_n^{(0)} + X_n^{(1)} + \frac{1}{2}X_n^{(2)} + \ldots$ is always zero and that the variance of the truncation $\hat{X}_n = X_n^{(0)} + X_n^{(1)} + \frac{1}{2}X_n^{(2)}$ is

$$\mathrm{var}(\hat{X}_n) = \xi^2 + q_2\{1 + q_1\xi\lambda^{(2)}(\xi)\}\sigma_Z^2$$

$$+ \tfrac{1}{4}q_2q_3q_4\{\mu_{4Z}(1+\lambda^{(1)3}) + 2\lambda^{(1)}(1+3\lambda^{(1)}-2\lambda^{(1)2}+3\lambda^{(1)3})q_1\sigma_Z^4\}\{\lambda^{(2)}(\xi)\}^2 .$$

With the extensions described in this subsection, approximations for the moments of the process (3.4.1) can be found in some circumstances, a major requirement being that the autoregression function is continuously differentiable. It is also required that there should be a suitable root of an equation of the form $x = \lambda(x) - a$. In the next two chapters a method is found which relies on much less restrictive conditions.

4 FAMILIES OF PROCESSES WITH VARYING AUTOREGRESSION FUNCTION

4.1 Introduction

In the last chapter various methods of approximating the stationary distribution of an autoregressive process were discussed. The method of section 3.4 was to consider the required non-linear process,

$$X_{n+1} = \lambda(X_n) + Z_{n+1} \qquad (n = \ldots -1,0,1,\ldots), \qquad (4.1.1)$$

as one of a family of processes $\{X_n(\alpha)\}$ generated by

$$X_{n+1}(\alpha) = \lambda(X_n(\alpha)) + \alpha Z_{n+1} \qquad (n = \ldots -1,0,1,\ldots). \qquad (4.1.2)$$

In this chapter the process (4.1.1) is imbedded in a different family of processes. This new method does not have the disadvantages of the previous method which needed restrictions on the possible stationary distributions of the process $X_{n+1} = \lambda(X_n)$ $(n = \ldots -1,0,1,\ldots)$ and on the derivative of $\lambda(x)$. In section 4.2, power-series expansions of the values of the processes are used to approximate the moments of the processes by the same method as used in section 3.4. In sections 4.4 and 4.5 a solution for the distribution of the stationary process is found by using a power series expansion for the stationary distribution functions of the family of processes. A similar method is used in section 4.6 to give expressions for conditional distributions of the processes. The expressions derived in this chapter are used later to give practical procedures for calculating the moments and distributions of the processes.

The family of processes (4.1.2) allows the scale of the input process to vary continuously. Earlier this was shown to be equivalent to considering the family of processes $\{W_n(\alpha)\}$ generated by

$$W_{n+1}(\alpha) = \frac{1}{\alpha}[\lambda\{\xi + \alpha W_n(\alpha)\} - \xi] + Z_{n+1} \qquad (n = \ldots -1,0,.,\ldots),$$

where $\xi = \lambda(\xi)$. In this family the autoregression function itself is allowed to vary. This suggests defining a family of processes $\{X_n(\beta)\}$ generated by

$$X_{n+1}(\beta) = \mu\{X_n(\beta);\beta\} + Z_{n+1} \qquad (n = \ldots -1,0,1,\ldots), \qquad (4.1.3)$$

where $\mu(x;\beta)$ is a suitable family of functions of x, indexed by β. These should have the properties that, for fixed x, $\mu(x;\beta)$ varies continuously with β and that $\mu(x;1) = \lambda(x)$, the required autoregression function. Further $\mu(x;0)$, which will be called the initial autoregression function, should be such that the corresponding process, $\{X_n(0)\}$, has a stationary distribution which is known: this will be called the initial (stationary) distribution and the process $\{X_n(0)\}$ will be called the initial process. Other requirements are that the process $\{X_n(\beta)\}$ should be stationary for every β in some interval covering $[0,1]$ and that the properties of the processes should change continuously with β.

It may be noted that the choice

$$\mu(x;\beta) = \xi + \beta^{-1}[\lambda\{\xi + \beta(x-\xi)\} - \xi] \qquad (4.1.4)$$

corresponds to the process $\{\xi + W_n(\alpha)\}$ with $\{W_n(\alpha)\}$ as above. In this case $\xi = \lambda(\xi)$. An obvious extension of this is to choose $\mu(x;\beta) = \lambda(\xi)+\beta^{-1}[\lambda\{\xi + \beta(x-\xi)\} - \lambda(\xi)]$, where now ξ is not necessarily a root of $x = \lambda(x)$. This does not correspond to a family of processes with a varying scale input distribution unless $\xi = \lambda(\xi)$. After some consideration it can be seen that there would be a discontinuity, at $\beta = 0$, in the behaviour of the processes $\{X_n(\beta)\}$ if the equation, $x = \lambda(x)$, has more than one root. This is because the equations $x = \mu(x;\beta)$ will all, except for $\beta = 0$, also have more than one root. The above modification does not overcome this problem.

A good choice for $\mu(x;\beta)$, and the one that will be most used from now on, is given by

$$\mu(x;\beta) = a + bx + \beta\{\lambda(x) - a - bx\} \qquad (4.1.5)$$

where a and b are two fixed real numbers, arbitrary except that the initial process (with autoregression function $a + bx$) should be stationary: for most input distributions this just means $|b| < 1$, see section 2.2.1. Thus the family of processes $\{X_n(\beta)\}$ under consideration is generated by

$$\begin{aligned} X_{n+1}(\beta) &= a + bX_n(\beta) + \beta[\lambda\{X_n(\beta)\} - a - bX_n(\beta)] + Z_{n+1} \\ &= a + bX_n(\beta) + \beta\Lambda\{X_n(\beta)\} + Z_{n+1} \qquad (n = \ldots -1,0,1,\ldots) . \end{aligned} \qquad (4.1.6)$$

In the next section the method of section 3.4 will be applied to this family of processes. For this it is assumed that exactly the same sequence $\{Z_n\}$ is used in the generation of each process $\{X_n(\beta)\}$. In the method to be described later, involving expansions of the stationary distribution functions, this assumption need not be made.

4.2 Power-series expansion of the process itself

4.2.1 By formally taking derivatives throughout (4.1.6), and setting $\beta = 0$, expressions are found for the components $\{X_n^{(r)}\}$ of the expansion of the processes $\{X_n(\beta)\}$, where

$$X_n(\beta) = X_n^{(0)} + \beta X_n^{(1)} + \frac{1}{2}\beta^2 X_n^{(2)} + \ldots . \tag{4.2.1}$$

Let

$$\Lambda^{(n)}(x) = \frac{d^n}{dx^n}\Lambda(x) , \quad \Lambda(x) = \lambda(x) - a - bx ,$$

where it is assumed that the derivatives exist. Then the formulae obtained are

$$X_{n+1}^{(0)} = a + bX_n^{(0)} + Z_{n+1} \tag{4.2.2}$$

$$X_{n+1}^{(1)} = bX_n^{(1)} + \Lambda(X_n^{(0)}) \tag{4.2.3}$$

$$X_{n+1}^{(2)} = bX_n^{(2)} + 2\Lambda^{(1)}(X_n^{(0)})X_n^{(1)} , \text{ etc.}$$

Again Z_{n+1} is independent of all random variables with smaller subscript. Here $\{X_n^{(0)}\}$ is a linear autoregressive process (equation 4.2.2) and, as such, has known properties. By the same method as used in section 3.4, the moments of the other component processes are found to be

$$E(X_n^{(1)}) = \frac{1}{1-b} E[\Lambda(X_n^{(0)})] ,$$

$$E(X_n^{(2)}) = \frac{2}{1-b} E[\Lambda^{(1)}(X_n^{(0)})X_n^{(1)}]$$

$$= \frac{2}{1-b} E[\Lambda^{(1)}(X_n^{(0)})\{bX_{n-1}^{(1)} + \Lambda(X_{n-1}^{(0)})\}]$$

$$= \frac{2}{1-b} \sum_{r=0}^{\infty} b^r E[\Lambda^{(1)}(X_{n+1}^{(0)})\Lambda(X_{n-r}^{(0)})] ,$$

where reliance is made on the asymptotic independence of $X_n^{(0)}$, $X_{n-k}^{(1)}$ for large k. Since the joint distributions of the process $\{X_n^{(0)}\}$ are known, the above expectation may (theoretically at least) be calculated. Similarly

$$\text{cov}[X_n^{(1)}, \Lambda(X_n^{(0)})] = \sum_{r=0}^{\infty} b^r \text{cov}[\Lambda(X_{n+1}^{(0)}), \Lambda(X_{n-r}^{(0)})] ,$$

$$\text{var}(X_n^{(1)}) = b^2 \text{var}(X_n^{(1)}) + 2b \text{cov}[X_n^{(1)}, \Lambda(X_n^{(0)})] + \text{var}[\Lambda(X_n^{(0)})]$$

$$= \frac{1}{1-b^2}\{2b \sum_{r=0}^{\infty} b^r \text{cov}[\Lambda(X_{n+1}^{(0)}), \Lambda(X_{n-r}^{(0)})] + \text{var}[\Lambda(X_n^{(0)})]\} \quad (4.2.4)$$

$$= \frac{1}{1-b^2} \sum_{r=-\infty}^{\infty} b^{|r|} \text{cov}[\Lambda(X_n^{(0)}), \Lambda(X_{n-r}^{(0)})] ,$$

$$\text{cov}(X_n^{(0)}, X_n^{(1)}) = \frac{b}{1-b^2} \text{cov}[X_n^{(0)}, \Lambda(X_n^{(0)})] . \quad (4.2.5)$$

From these approximations for the stationary moments of the process $\{X_n\} \equiv \{X_n(1)\}$ are available, these being based on truncations of the expansion (4.2.1) of $X_n(\beta)$. Thus

$$E\{X_n^{(0)} + \beta X_n^{(1)} + \frac{1}{2}\beta^2 X_n^{(2)}\} = E(X_n^{(0)}) + \frac{\beta}{1-b} E[\Lambda(X_n^{(0)})]$$

$$+ \frac{\beta^2}{1-b^2} \sum_{r=0}^{\infty} b^r E[\Lambda^{(1)}(X_{n+1}^{(0)})\Lambda(X_{n-r}^{(0)})] \quad (4.2.6)$$

and

$$\text{var}\{X_n^{(0)} + \beta X_n^{(1)}\} = \text{var}(X_n^{(0)}) + \frac{2\beta b}{1-b^2} \text{cov}[X_n^{(0)}, \Lambda(X_n^{(0)})]$$

$$+ \frac{\beta^2}{1-b^2} \sum_{r=-\infty}^{\infty} b^{|r|} \text{cov}[\Lambda(X_n^{(0)}), \Lambda(X_{n-r}^{(0)})] .$$

In the same way approximations for the joint moments can be found. From (4.2.3), for $k \geq 1$,

$$X_{n+k}^{(1)} = \sum_{r=1}^{k} b^{r-1} \Lambda(X_{n+k-r}^{(0)}) + b^k X_n^{(1)} , \qquad (4.2.7)$$

and hence, using (4.2.5), for $k \geq 1$,

$$\begin{aligned} \operatorname{cov}(X_{n+k}^{(1)}, X_n^{(0)}) &= \sum_{r=1}^{k} b^{r-1} \operatorname{cov}[X_n^{(0)}, \Lambda(X_{n+k-r}^{(0)})] + b^k \frac{b}{1-b^2} \operatorname{cov}[X_n^{(0)}, \Lambda(X_n^{(0)})] \\ &= \sum_{r=0}^{k-1} b^{k-r-1} \operatorname{cov}[\Lambda(X_{n+r}^{(0)}), X_n^{(0)}] + \frac{b^{k+1}}{1-b^2} \operatorname{cov}[X_n^{(0)}, \Lambda(X_n^{(0)})], \end{aligned}$$

$$\operatorname{cov}(X_n^{(1)}, X_{n+k}^{(0)}) = b \operatorname{cov}(X_n^{(1)}, X_{n+k-1}^{(0)}) = b^k \operatorname{cov}(X_n^{(1)}, X_n^{(0)})$$

$$= \frac{b^{k+1}}{1-b^2} \operatorname{cov}[X_n^{(0)}, \Lambda(X_n^{(0)})] .$$

For $k \geq 1$, (4.2.7) gives

$$\operatorname{cov}(X_{n+k}^{(1)}, X_n^{(1)}) = \sum_{r=0}^{k-1} \sum_{\ell=0}^{\infty} b^{\ell+k-r-1} \operatorname{cov}[\Lambda(X_{n+r}^{(0)}), \Lambda(X_{n-\ell-1}^{(0)})] + b^k \operatorname{var}(X_n^{(1)})$$

$$= \frac{1}{1-b^2} \sum_{q=0}^{\infty} (b^{|q-k+1|} - b^{k+q+1}) \operatorname{cov}[\Lambda(X_{n+1}^{(0)}), \Lambda(X_{n-q}^{(0)})] + b^k \operatorname{var}(X_n^{(1)})$$

$$= \frac{1}{1-b^2} \{\Sigma (b^{|q-k+1|} + b^{k+q+1}) \operatorname{cov}[\Lambda(X_{n+1}^{(0)}), \Lambda(X_{n-q}^{(0)})] + b^k \operatorname{var}(X_n^{(0)})\}$$

on using (4.2.4). From these the approximation $\operatorname{cov}(X_n^{(0)}+\beta X_n^{(1)}, X_{n+k}^{(0)}+\beta X_{n+k}^{(1)})$ for $\operatorname{cov}(X_n(\beta), X_{n+k}(\beta))$ can be calculated. Using also the expression for the variance, the approximation for the correlation is

$$\operatorname{corr}(X_n(\beta), X_{n+k}(\beta)) \simeq b^k + \Big\{\beta \sum_{r=0}^{k-1} b^{k-r-1} \operatorname{cov}[X_n^{(0)}, \Lambda(X_{n+r}^{(0)})]$$

$$+ \frac{\beta^2}{1-b^2} \sum_{q=0}^{\infty} (b^{|q-k+1|} - b^{q+k+1}) \operatorname{cov}[\Lambda(X_{n+1}^{(0)}), \Lambda(X_{n-q}^{(0)})]\Big\} / \operatorname{var}[X_n^{(0)} + \beta X_n^{(1)}] .$$

For k = 1, this is, to order 2 in β,

$$\operatorname{corr}(X_n(\beta), X_{n+1}(\beta)) \simeq b + \beta \frac{\operatorname{cov}[X_n^{(0)}, \Lambda(X_n^{(0)})]}{\operatorname{var}[X_n^{(0)}]}$$

$$+ \beta^2 \left\{ \sum_{q=0}^{\infty} b^q \frac{\operatorname{cov}[\Lambda(X_{n+1}^{(0)}), \Lambda(X_{n-q}^{(0)})]}{\operatorname{var}[X_n^{(0)}]} - \frac{2b}{1-b^2} \left(\frac{\operatorname{cov}[X_n^{(0)}, \Lambda(X_n^{(0)})]}{\operatorname{var}[X_n^{(0)}]} \right)^2 \right\}.$$

The moments of higher order truncations of (4.2.1) may be found in the same way, but the resulting expressions become increasingly complicated.

4.2.2 In the special case that b = 0 the equations for the components of $X_n(\beta)$ are

$$\begin{aligned} X_{n+1}^{(0)} &= a + Z_{n+1} , \\ X_{n+1}^{(1)} &= \Lambda(X_n^{(0)}) , \\ X_{n+1}^{(2)} &= 2\Lambda^{(1)}(X_n^{(0)}) X_n^{(1)} , \\ X_{n+1}^{(3)} &= 3\Lambda^{(1)}(X_n^{(0)}) X_n^{(2)} + 3\Lambda^{(2)}(X_n^{(0)}) X_n^{(1)2} . \end{aligned} \tag{4.2.8}$$

Hence the values of the sequence $\{X_n^{(0)}\}$ are independent with distribution given by (4.2.8), and, substituting amongst the equations, gives

$$\begin{aligned} X_{n+1}^{(2)} &= 2\Lambda^{(1)}(X_n^{(0)}) \Lambda(X_{n-1}^{(0)}) \\ X_{n+1}^{(3)} &= 6\Lambda^{(1)}(X_n^{(0)}) \Lambda^{(1)}(X_{n-1}^{(0)}) \Lambda(X_{n-2}^{(0)}) + 3\Lambda^{(2)}(X_n^{(0)}) \{\Lambda(X_{n-1}^{(0)})\}^2 \end{aligned}$$

The expansion of $\{X_n(\beta)\}$ is then given by

$$\begin{aligned} X_{n+1}(\beta) = a + Z_{n+1} &+ \beta\Lambda(a+Z_n) + \beta^2 \Lambda^{(1)}(a+Z_n)\Lambda(a+Z_{n-1}) \\ &+ \beta^3 [\Lambda^{(1)}(a+Z_n)\Lambda^{(1)}(a+Z_{n-1})\Lambda(a+Z_{n-2}) + \tfrac{1}{2}\Lambda^{(2)}(a+Z_n)\{\Lambda(a+Z_{n-1})\}^2] \\ &+ \ldots . \end{aligned} \tag{4.2.9}$$

Approximations for the stationary moments of the process $\{X_n(1)\}$ can be found directly from this. Thus

$$E(X_n^{(0)} + \beta X_n^{(1)} + \tfrac{1}{2}\beta^2 X_n^{(2)} + \tfrac{1}{6}\beta^3 X_n^{(3)}) = E\{a + Z\} +$$

$$+ E\{\Lambda(a+Z)\}(\beta + \beta^2 E\{\Lambda^{(1)}(a+Z)\} + \beta^3[E\{\Lambda^{(1)}(a+Z)\}]^2)$$

$$+ \tfrac{1}{2}\beta^3 E\{\Lambda(a+Z)^2\} E\{\Lambda^{(2)}(a+Z)\}$$

and, letting $\hat{X}_n = X_n^{(0)} + \beta X_n^{(1)} + \frac{1}{2}\beta^2 X_n^{(2)}$,

$$\begin{aligned}\operatorname{var}(\hat{X}_n) = {} & \operatorname{var}(Z) + \beta^2 \operatorname{var}\{\Lambda(a+Z)\} \\ & + 2\beta^3 \operatorname{cov}\{\Lambda(a+Z), \Lambda^{(1)}(a+Z)\}E\{\Lambda(a+Z)\} \\ & + \beta^4[\operatorname{var}\{\Lambda(a+Z)\}E\{\Lambda^{(1)}(a+Z)^2\} + E\{\Lambda(a+Z)^2\}\operatorname{var}\{\Lambda^{(1)}(a+Z)\}] .\end{aligned} \tag{4.2.10}$$

Also, for joint moments,

$$\begin{aligned}\operatorname{cov}(\hat{X}_{n+1}, \hat{X}_n) = {} & \beta \operatorname{cov}\{Z, \Lambda(a+Z)\} + \beta^2 E\{\Lambda(a+Z)\}\operatorname{cov}\{Z, \Lambda^{(1)}(a+Z)\} \\ & + \beta^3 E\{\Lambda^{(1)}(a+Z)\}\operatorname{var}\{\Lambda(a+Z)\} \\ & + \beta^4 E\{\Lambda^{(1)}(a+Z)\}E\{\Lambda(a+Z)\}\operatorname{cov}\{\Lambda(a+Z), \Lambda^{(1)}(a+Z)\},\end{aligned} \tag{4.2.11}$$

$$\operatorname{cov}(\hat{X}_{n+2}, \hat{X}_n) = \beta^2 E\{\Lambda^{(1)}(a+Z)\}\operatorname{cov}\{Z, \Lambda(a+Z)\}, \tag{4.2.12}$$

$$\operatorname{cov}(\hat{X}_{n+k}, \hat{X}_n) = 0 \qquad (k > 2).$$

Taking instead a truncation of (4.2.9) to order 3 in β would introduce a term in β^4 into the approximation for the variance and a term in β^3 into the approximation for each covariance of lag up to 3. In this sense (4.2.10) is correct to order β^3 and (4.2.11-3) are correct to order β^2 as approximations to the variances and covariances of $\{X_n(\beta)\}$.

Clearly the approximations for the moments are of a much simpler form when b = 0 than otherwise and the simplicity carries over when other approaches are used, as will be seen later.

4.3 Power-series expansions for the stationary distribution

4.3.1 In this section the stationary distributions, $F(x;\beta)$, of the processes $\{X_n(\beta)\}$ are considered. It will be assumed that the family of stationary distributions is expressible as a power-series in β, convergent at $\beta = 1$. It is not clear exactly when this will hold but it seems a reasonable assumption to make in some cases. There are certainly some families of autoregression functions $\{\mu(x;\beta)\}$ and input distributions F_Z for which the assumption holds. Let the input sequence be normally distributed with zero mean and variance σ_Z^2 and let $\{\mu(x;\beta)\}$ be given by

$$\mu(x;\beta) = a_0 + b_0 x + \beta\{(a_1-a_0) + (b_1-b_0)x\}$$

with $|b_0|, |b_1| < 1$. Then the stationary distributions $F(x;\beta)$ are normal with mean and variance given by, respectively,

$$\frac{a_0 + \beta(a_1-a_0)}{1 - \{b_0+\beta(b_1-b_0)\}} \quad , \quad \frac{\sigma_Z^2}{1 - \{b_0+\beta(b_1-b_0)\}^2} \quad .$$

In this case not only is $\{F(x;\beta)\}$ continuous in β but also it is possible to make a power-series expansion in β, convergent whenever $|b_0 + \beta(b_1-b_0)| < 1$.

4.3.2 The stationary distributions $F(x;\beta)$ of the processes given by (4.1.3) must satisfy

$$F(x;\beta) = \int F_Z\{x - \mu(y;\beta)\}F(dy;\beta), \qquad (4.3.1)$$

and one method of finding the required distribution $F(x;1)$ is to expand $F(x;\beta)$ as a power-series in β and to equate powers of β in the above expression. In fact it is more convenient to work with the corresponding characteristic functions: this also has the advantage of avoiding the ambiguity of the definition of distribution functions at isolated points. Let $\phi(s;\beta)$ be the characteristic function corresponding to $F(x;\beta)$ and let $\phi_X(s)$ correspond to the stationary distribution $F_X(x)$ required. Then $\phi_X(s) = \phi(s;1)$.

In this section expressions for the first few terms in the expansions will be derived to show how the method works. In a later section a more complete expansion will be derived and some attention will be given to its validity. For the present various assumptions will be made for convenience: it will be assumed that all distributions involved possess densities which are continuously differentiable.

Under certain circumstances, for instance if $\phi(s;\beta)$ is absolutely integrable (with respect to s), the characteristic functions $\phi(s;\beta)$ must satisfy the equation, equivalent to (4.3.1),

$$\phi(s;\beta) = \frac{\phi_Z(s)}{2\pi} \iint \exp\{is\mu(x;\beta) - ixt\}\phi(t;\beta)dt\, dx.$$

For the particular choice of $\mu(x;\beta)$,

$$\mu(x;\beta) = a + bx + \beta(\lambda(x) - a - bx) = a + bx + \beta\Lambda(x)$$

this equation is

$$\phi(s;\beta) = \frac{\phi_Z(s)}{2\pi} \iint \exp\{ix(bs-t) + isa + is\beta\Lambda(x)\}\phi(t;\beta)dt\, dx.$$

Let $\phi(s;\beta) = \sum_{r=0}^{\infty} \theta_r(s)\beta^r$, then

$$\Sigma\theta_r(s)\beta^r = \frac{e^{isa}\phi_Z(s)}{2\pi} \iint \exp\{ix(bs-t)\} \sum_{j=0}^{\infty} \frac{\{i\beta s\Lambda(x)\}^j}{j!} \sum_{r=0}^{\infty} \theta_r(t)\beta^r dt\, dx.$$

Equating powers of β gives, for $r \geq 0$,

$$\theta_r(s) = \frac{e^{isa}\phi_Z(s)}{2\pi} \iint \exp\{ix(bs-t)\} \sum_{j=0}^{r} \frac{\{is\Lambda(x)\}^j}{j!} \theta_{r-j}(t)dt\, dx.$$

The term for $j = 0$ in this expression may be readily found, giving

$$\theta_0(s) = e^{isa}\phi_Z(s)\theta_0(bs) \tag{4.3.2}$$

and, for $r \geq 1$,

$$\theta_r(s) = e^{isa}\phi_Z(s)\theta_r(bs)$$

$$+ \frac{e^{isa}\phi_Z(s)}{2\pi} \iint \exp\{ix(bs-t)\} \sum_{j=1}^{r} \frac{\{is\Lambda(x)\}^j}{j!} \theta_{r-j}(t)\,dt\,dx \qquad (4.3.3)$$

The solution to (4.3.2) is

$$\theta_0(s) = \prod_{r=0}^{\infty} e^{isb^r a}\phi_Z(b^r s) = \exp\{isa/(1-b)\} \prod_{r=0}^{\infty} \phi_Z(b^r s),$$

where the infinite product converges since $\theta_0(s)$ is the characteristic function of the stationary distribution of the initial process

$$X_{n+1}(0) = a + bX_n(0) + Z_{n+1} \quad (n = \ldots -1,0,1,\ldots),$$

a linear process, see section 2.2.2. Let $f_0(x)$ be the density corresponding to $\theta_0(s)$.

Since $\phi(s;\beta)$ is, for every β, continuous at $s = 0$ with $\phi(0;\beta) = 1$ and $\theta_0(0) = 1$, it is assumed that each $\theta_r(s)$ is continuous at $s = 0$ with $\theta_r(0) = 0 (r \geq 1)$. Since $|b| < 1$ this implies, for $r \geq 1$ and all s, $\theta_r(b^n s) \to 0$ as $n \to \infty$. Using this limit, iteration of equation (4.3.3) gives, for $r \geq 1$,

$$\theta_r(s) = \sum_{\ell=0}^{\infty} \frac{\rho_\ell(s)}{2\pi} \iint \exp\{ix(b^{\ell+1}s-t)\} \sum_{j=1}^{r} \frac{\{ib^\ell s\Lambda(x)\}^j}{j!} \theta_{r-j}(t)\,dt\,dx, \qquad (4.3.4)$$

where

$$\rho_\ell(s) = \prod_{m=0}^{\ell} e^{isb^m a}\phi_Z(b^m s) = \exp\{\frac{isa(1-b^{\ell+1})}{1-b}\} \prod_{m=0}^{\ell} \phi_Z(b^m s)$$

$$= \frac{\theta_0(s)}{\theta_0(b^{\ell+1}s)} . \qquad (4.3.5)$$

The function $\rho_\ell(s)$ is the characteristic function of a random variable Y_ℓ given by

$$Y_n = a + bY_{n-1} + Z_n \qquad (n = 0,1,\ldots,\ell) ,$$

with $Y_{-1} = 0$. Let $p_\ell(y)$ be the density corresponding to $\rho_\ell(s)$: then $p_\ell(y - b^{\ell+1}x)$, which appears later, is the density of Y_ℓ when $Y_{-1} = x$, and $p_\ell(y - b^{\ell+1}x)f_0(x)$ is the stationary joint density, at lag $\ell + 1$, of the initial (linear) process.

The expression (4.3.4) for θ_r does not involve θ_r on the right-hand side. Thus, since θ_0 is known, this formula gives $\theta_1, \theta_2, \theta_3, \ldots$ successively in terms of the earlier functions. For $r = 1$ (4.3.4) gives

$$\theta_1(s) = \sum_{\ell=0}^{\infty} \frac{\rho_\ell(s)}{2\pi} \iint \exp\{ix(b^{\ell+1}s-t)\} ib^\ell s \Lambda(x) \theta_0(t) dt\, dx$$

$$= \sum_{\ell=0}^{\infty} \rho_\ell(s) \int \exp\{ixb^{\ell+1}s\} ib^\ell s \Lambda(x) f_0(x) dx. \qquad (4.3.6)$$

If $f_1(x)$ is the inverse Fourier transform of $\theta_1(s)$,

$$f_1(y) = -\sum_{\ell=0}^{\infty} \int b^\ell \frac{\partial}{\partial y} p_\ell(y - b^{\ell+1}x) \Lambda(x) f_0(x) dx$$

$$= \int_{-\infty}^{\infty} R_1(y,x) \Lambda(x) f_0(x) dx \qquad (4.3.7)$$

where

$$R_1(y,x) = -\sum_{\ell=0}^{\infty} b^\ell \frac{\partial}{\partial y} p_\ell(y - b^{\ell+1}x) \qquad (4.3.8)$$

On taking the inverse Fourier transform of $\theta_r(s)$, (4.3.4) leads to

$$f_r(y) = \int \sum_{j=1}^{r} R_j(y,x) \{\Lambda(x)\}^j f_{r-j}(x) dx \qquad (r \geq 1)$$

where

$$R_j(y,x) = \frac{(-1)^j}{j!} \sum_{\ell=0}^{\infty} b^{\ell j} \frac{\partial^j}{\partial y^j} p_\ell(y - b^{\ell+1}x) \qquad (j \geq 1).$$

Hence in particular

$$f_2(y) = \int R_1(y,x) \Lambda(x) f_1(x) dx + \int R_2(y,x) \{\Lambda(x)\}^2 f_0(x) dx \qquad (4.3.9)$$

and so, to the second order in β, the density of $F(y,\beta)$ is

$$f(y;\beta) \simeq f_0(y) + \beta f_1(y) + \beta^2 f_2(y)$$

with f_1 and f_2 given by (4.3.7), (4.3.9) respectively. This approximation has been found to sometimes give negative values.

Approximations for the moments of the stationary distributions may be obtained either by differentiating the approximation for the characteristic function or by integrating the above approximation for the stationary density. These lead to the following approximations for the mean, $\mu(\beta)$, and variance, $\sigma^2(\beta)$, of the process $\{X_n(\beta)\}$.

$$\mu(\beta) \simeq \mu_0 + \frac{\beta}{1-b}\int \Lambda(x) f_0(x)\,dx + \frac{\beta^2}{1-b}\iint \Lambda(x)\Lambda(y) R_1(y,x) f_0(x)\,dx\,dy \qquad (4.3.10)$$

$$\sigma^2(\beta) \simeq \sigma_0^2 + \beta\,\frac{2b}{1-b^2}\int (x-\mu_0)\Lambda(x) f_0(x)\,dx + \beta^2\left[\frac{2b}{1-b^2}\iint (y-\mu_0)\Lambda(x)\Lambda(y) R_1(y,x) f_0(x)\,dx\,dy + \frac{1}{1-b^2}\int\{\Lambda(x)\}^2 f_0(x)\,dx - \left\{\frac{1}{1-b}\int \Lambda(x) f_0(x)\,dx\right\}^2\right] \qquad (4.3.11)$$

where μ_0, σ_0^2 are the mean and variance of the density $f_0(x)$. If $\Lambda(y)$ is differentiable it is possible to simplify the expressions slightly by an integration by parts. The above approximation for the variance need not be positive.

4.3.3 It was seen in section 4.2.2 that the choice $b = 0$ leads to a great reduction in the complexity of the expressions derived by the method used there. This is so also for the present method. Choosing $b = 0$ means that the initial approximation for the autoregression function $\lambda(x)$ is now a constant function ($= a$) rather than a linear function ($= a + bx$). Hence, given that some linear function will usually be a better approximation to $\lambda(x)$ (in some sense) than simply a constant function, it is clear that restricting b to be zero will lead to a worse approximation for the stationary distribution in the sense that more terms

will be needed to attain the same accuracy. However the relative simplicity with which these terms may be calculated when b = 0 seems to outweigh this disadvantage when more than just one or two terms are needed by a choice of $b \neq 0$. Indeed, for many input distributions, the case b = 0 may be the only one for which the stationary distribution of the initial (linear) process can be found explicitly, see section 2.2.2.

When b = 0 the above formulae reduce to

$$\theta_0(s) = e^{isa}\phi_Z(s), \quad f_0(y) = f_Z(y - a)$$

and, for $r \geq 1$,

$$\theta_r(s) = \frac{e^{isa}\phi_Z(s)}{2\pi} \iint \exp\{ixt\} \sum_{j=1}^{r} \frac{\{is\Lambda(x)\}^j}{j!} \theta_{r-j}(t)\,dt\,dx$$

$$= \theta_0(s) \int \sum_{j=1}^{r} \frac{\{is\Lambda(x)\}^j}{j!} f_{r-j}(x)\,dx$$

$$f_r(y) = \sum_{j=1}^{r} \left[\frac{(-1)^j}{j!} \frac{d^j}{dy^j} f_0(y) \int_{-\infty}^{\infty} \{\Lambda(x)\}^j f_{r-j}(x)\,dx \right]$$

where now $\Lambda(x) = \lambda(x) - a$. These lead to the following expressions for the first few terms of the expansion of the density $f(x;\beta)$, where these have been written in terms of the initial density f_0 and integration by parts has been employed.

$$f_1(y) = -f_0^{(1)}(y) \int \{\lambda(x) - a\} f_0(x)\,dx,$$

$$f_2(y) = -f_0^{(1)}(y) \int \lambda^{(1)}(x) f_0(x)\,dx \int \{\lambda(x) - a\} f_0(x)\,dx$$

$$+ \frac{1}{2} f_0^{(2)}(y) \int \{\lambda(x) - a\}^2 f_0(x)\,dx,$$

$$f_3(y) = -f_0^{(1)}(y) \left\{ \int \lambda^{(1)}(x) f_0(x)\,dx \right\}^2 \int \{\lambda(x) - a\} f_0(x)\,dx$$

$$- \frac{1}{2} f_0^{(1)}(y) \int \lambda^{(2)}(x) f_0(x)\,dx \cdot \int \{\lambda(x) - a\}^2 f_0(x)\,dx$$

$$+ f_0^{(2)}(y) \int \lambda^{(1)}(x)\{\lambda(x) - a\} f_0(x)dx. \int \{\lambda(x) - a\} f_0(x)dx$$

$$- \frac{1}{6} f_0^{(3)}(y) \int \{\lambda(x) - a\}^3 f_0(x)dx.$$

Expansions for the mean and variance, to order 3, are

$$\mu(\beta) = E\{Z + a\} + \beta E\{\lambda(Z + a) - a\}$$

$$+ \beta^2 E\{\lambda(Z + a) - a\} E\{\lambda^{(1)}(Z + a)\}$$

$$+ \beta^3 E\{\lambda(Z + a) - a\}[E\{\lambda^{(1)}(Z + a)\}]^2 + \frac{1}{2}\beta^3 E[\{\lambda(Z + a) - a\}^2] E\{\lambda^{(2)}(Z+a)\}$$

$$\sigma^2(\beta) = \text{var}(Z) + \beta^2 \text{var}\{\lambda(Z + a)\}$$

$$+ 2\beta^3 E\{\lambda(Z + a) - a\} \text{cov}\{\lambda(Z + a) - a, \lambda^{(1)}(Z + a)\}.$$

Note that the approximation (4.2.10) for the variance is the same as this except that a term of order β^4 has been added. However the extension of the above method to terms of order 4 would yield not only this term but another term also of order 4. (this is essentially $\text{cov}(X_n^{(1)}, X_n^{(3)})$ where $X_n(\beta) = \Sigma \frac{\beta^r}{r!} X_n^{(r)}$ as in section 4.2.2). The earlier approximation has the advantage of always being positive.

4.3.4 Equation (4.3.4) gives an alternative integral equation for the characteristic functions $\phi(s;\beta)$: thus

$$\phi(s;\beta) = \theta_0(s) + \sum_{\ell=0}^{\infty} \frac{\rho_\ell(s)}{2\pi} \iint \exp\{ix(b^{\ell+1}s-t)\}[\exp\{ib^\ell \beta s \Lambda(x)\} - 1]\phi(t;\beta)dtdx.$$

The equivalent equation for the stationary density function is

$$f(y;\beta) = f_0(y) + \sum_{\ell=0}^{\infty} \int [p_\ell\{y-b^{\ell+1}x-b^\ell\beta\Lambda(x)\} - p_\ell\{y-b^{\ell+1}x\}] f(x;\beta)dx.$$

When b = 0 this reduces to the equation

$$f(y;\beta) = f_0(y) + \int_{-\infty}^{\infty} [f_0(y-\beta\Lambda(x)) - f_0(y)] f(x;\beta)dx,$$

and this is equivalent to

$$f(y;\beta) = \int f_0(y-\beta\Lambda(x)) f(x;\beta)dx,$$

which is just the equilibrium equation for the density of the process $\{X_n(\beta)\}$ since $f_0(y) = f_Z(y-a)$. However the interpretation of the equation is not clear when $b \neq 0$.

4.3.5 In this section the approximations for the moments produced by section 4.3.2 are considered for a simple case in which they can be evaluated simply and for which the true values are known. Let $\lambda(x) = a^* + b^*x$ and let the input distribution have zero mean and variance σ_Z^2. Then the true mean and variance of the stationary distribution are

$$\mu^* = \frac{a^*}{1-b^*}, \quad \sigma^{*2} = \frac{\sigma_Z^2}{1-b^{*2}},$$

respectively. Suppose that the method of section 4.3.2 is applied with a and b not necessarily taking their correct values a^*, b^*. Then the approximations to the mean μ^* are, from (4.3.10), to zero, first and second powers of β,

$$\mu_0 = \frac{a}{1-b}, \quad \frac{1}{1-b}\{a^* + \frac{b^*-b}{1-b}a\}$$

$$\frac{1}{1-b}\{a^* + (\frac{b^*-b}{1-b})a^* + (\frac{b^*-b}{1-b})^2 a\}$$

respectively. For comparison of the approximations their ratios to the true value are taken and the terms grouped appropriately: this gives, respectively,

$$\frac{a}{a^*}\{1 - \frac{b^*-b}{1-b}\}, \quad 1 + (\frac{a}{a^*} - 1)(\frac{b^*-b}{1-b}) - \frac{a}{a^*}(\frac{b^*-b}{1-b})^2$$

$$1 + (\frac{a}{a^*} - 1)(\frac{b^*-b}{1-b})^2 - \frac{a}{a^*}(\frac{b^*-b}{1-b})^3 .$$

These expansions are appropriate when the initial autoregression function is specified in terms of its intercept and slope. However it may be more natural to specify the slope of the line together with the mean, μ_0, of the initial process. In this case appropriate expressions for the comparison of the three approximations are

$$\frac{\mu_0}{\mu^*}\ ,\quad 1-\frac{b^*-b}{1-b}\left(1-\frac{\mu_0}{\mu^*}\right),\quad 1-\left(\frac{b^*-b}{1-b}\right)^2\left(1-\frac{\mu_0}{\mu^*}\right).$$

From these it can be seen that, provided $b < \frac{1}{2}(b^* + 1)$, the method gives an improving sequence of approximations for the mean and further that if either the slope, b, or the mean, μ_0, are chosen correctly then the approximations found using these initial values are exact.

A similar comparison may be made for the approximations to the variance of the stationary distribution. The initial, first and second order approximations are

$$\frac{\sigma_Z^2}{1-b^2}\ ,\quad \frac{\sigma_Z^2}{1-b^2}\left\{1+\frac{2b(b^*-b)}{1-b^2}\right\}$$

$$\frac{\sigma_Z^2}{1-b^2}\left\{1+\frac{2b(b^*-b)}{1-b^2}+\frac{(1+3b^2)(b^*-b)^2}{(1-b^2)^2}\right\}$$

respectively. The ratios of these to the true value σ^{*2} are

$$1-\frac{b^{*2}-b^2}{1-b^2}\ ,\quad 1-\frac{(b^*-b)^2}{1-b^2}\left\{1+\frac{2b(b^*+b)}{1-b^2}\right\}$$

$$1-\frac{(b^*-b)^3}{1-b^2}\left\{1+3b+b^2(b^*+b)\right\}$$

respectively. Once again the method produces an improving sequence of approximations when b is sufficiently close to b*. However the errors of the initial and first approximations are of the same order when either b* or b is zero.

4.3.6 Even for low order terms many of the expressions of section 4.3.2 are not particularly simple even when the conditional density $p_\ell(y)$ can be written explicitly. Suppose that the input distribution is normal with zero mean and variance σ_Z^2, then $\phi_Z, \theta_0, f_0, \rho_\ell$ and p_ℓ can be found. In particular, setting $\mu_0 = a/(1-b)$ and $\sigma_0^2 = \sigma_Z^2/(1-b^2)$,

$$p_\ell(y) = \{2\pi(1-b^{2\ell+2})\sigma_0^2\}^{-1/2}\exp\left[-\frac{\{y-\mu_0(1-b^{\ell+1})\}^2}{2\sigma_0^2(1-b^{2\ell+2})}\right]\quad(\ell \geq 0)\ .$$

When the autoregression function is $\lambda(x) = \exp\{-\frac{1}{2}x^2\}$ the first correction term $f_1(y)$ (4.3.7) to the density $f_0(y)$ can eventually be shown to be given by $f_1(y) = \frac{\partial}{\partial y} F_1(y)$ where

$$F_1(y) = \frac{-1}{\sqrt{2\pi}\sigma_0} \exp\{-\frac{(y-\mu_0)^2}{2\sigma_0^2}\}\left(\sum_{\ell=0}^{\infty} \frac{b^\ell \sigma_0}{\{1+\sigma_0^2(1-b^{2\ell+2})\}^{1/2}} \exp\left[-\frac{\{\mu_0+b^{\ell+1}(y-\mu_0)\}^2}{2\{1+\sigma_0^2(1-b^{2\ell+2})\}}\right] - \frac{\mu_0}{1-b} - \frac{b^2(y-\mu_0)}{1-b^2}\right) .$$

The approximation to order one for the mean is, from either (4.2.6) or (4.3.10), with $\beta = 1$,

$$\mu_0 + \frac{\beta}{1-b}\left[(1+\sigma_0^2)^{-1/2}\exp\{-\frac{\mu_0^2}{2(1+\sigma_0^2)}\} - a - b\mu_0\right]$$

$$= \frac{1}{1-b}\left[(1+\sigma_0^2)^{-1/2}\exp\{-\frac{\mu_0^2}{2(1+\sigma_0^2)}\} - b\mu_0\right] .$$

The next term in the expansion is, from (4.2.6),

$$\frac{1}{1-b^2}\sum_{r=0}^{\infty} b^r E\left[\{\exp(-\tfrac{1}{2}X_{n-r}^{(0)2}) - a - bX_{n-r}^{(0)}\}\{-X_{n+1}^{(0)}\exp(-\tfrac{1}{2}X_{n+1}^{(0)2}) - b\}\right]$$

where the joint density $f_{r+1}(x,y)$ of $(X_{n-r}^{(0)}, X_{n+1}^{(0)})$ is the product of $f_0(x)$ and $p_r(y-b^{r+1}x)$.

4.4 The marginal characteristic function

4.4.1 In section 4.3 a solution was obtained by formal methods for $\phi(s;\beta)$, the characteristic functions of the stationary distributions of the processes $\{X_n(\beta)\}$ given by (4.1.6). Here it is shown that this expression satisfies the necessary equations for the characteristic functions of the stationary processes.

In finding the solution for the characteristic functions $\phi(s;\beta)$ it seems natural to work with the Fourier transforms of the functions $\Lambda^j(x)$ $(j = 1,2,3,\ldots)$. However these need not exist since it is not even assumed that $\Lambda(x) \to 0$ as $x \to \pm\infty$. There are several ways of avoidng this difficulty. One possibility is to approximate the

autoregression function by another function or sequence of functions such that the Fourier transforms involved for these new processes all exist and have other suitable properties. Other possibilities are the use of Generalised Fourier Transforms (Titchmarsh, 1937, p4) or the theory of distributions (generalised functions). However, here some restrictive conditions are placed on $\Lambda(x)$ to ensure good properties for the Fourier transforms. No conditions are then needed on the input distribution. It would also be possible to demonstrate the validity of the solution (in a slightly different form) by placing conditions on the input distribution with much less restrictive conditions on the autoregression function.

In section 4.6 a different approach leads to a different form for the solution $\phi(s;\beta)$ which avoids forming the Fourier transforms mentioned and which, at least when $b = 0$, is equivalent to the present solution. This method indicates that the solution should hold under a wide range of conditions.

4.4.2 The expressions found in section 4.3.2 can be rearranged to give, as the possible solution for the characteristic function $\phi(s;\beta)$, the formulae

$$\phi(s;\beta) = \theta_0(s) + \sum_{N=1}^{\infty} \beta^N \theta_N(s) \tag{4.4.1}$$

$$\theta_N(s) = \sum_{j=1}^{N} \sum_{\ell=0}^{\infty} (ib^{\ell}s)^j T^{(N,j)}(b^{\ell+1}s)\rho_{\ell}(s) \qquad (N = 1,2,3,\ldots) \tag{4.4.2}$$

$$T^{(N,j)}(s) = \frac{1}{2\pi j!} \int L_j(u)\theta_{N-j}(s-u)\,du \qquad (1 \leq j \leq N), \tag{4.4.3}$$

with

$$\theta_0(s) = \exp\{\frac{isa}{1-b}\} \prod_{r=0}^{\infty} \phi_Z(b^r s) \tag{4.4.4}$$

$$\rho_{\ell}(s) = \exp\{\frac{isa(1-b^{\ell+1})}{1-b}\} \prod_{r=0}^{\ell} \phi_Z(b^r s) \qquad (\ell \geq 0). \tag{4.4.5}$$

The functions $L_j(u)$ are the Fourier transforms of the functions $\Lambda^j(x)$,

$$L_j(u) = \int \Lambda^j(x) e^{iux} dx \qquad (j \geq 1). \tag{4.4.6}$$

As has been stated, these Fourier transforms do not in general exist for all real values of the argument: expressions for $T^{(N,j)}(s)$ equivalent to the above are

$$T^{(N,j)}(s) = \frac{1}{2\pi j!} \iint \exp\{ix(s-t)\} \Lambda^j(x) \theta_{N-j}(t) dt\, dx$$

$$= \frac{1}{j!} \int e^{isx} \Lambda^j(x) dF_{N-j}(x)$$

where $\theta_q(s) = \int e^{isx} dF_q(x)$. Thus it is not essential to work with the Fourier transforms since these alternative expressions are available. However they would be more complicated to use since $\theta_q(s)$ would have to be inverted to give F_q: it is not clear when this can be done.

Equations (4.4.2), (4.4.3) give expressions for $\theta_1, \theta_2, \theta_3, \ldots$ in turn. Since θ_0 is known, (4.4.3) gives $T^{(1,1)}$ and with this (4.4.2) gives θ_1: then using (4.4.3) again gives $T^{(2,1)}$, $T^{(2,2)}$ and these are used by (4.4.2) to give θ_2. Proceeding in this way the sequence $\theta_0(s), \theta_1(s), \theta_2(s), \ldots\ldots$ is given.

4.4.3 One way of ensuring that the integral appearing in formula (4.4.3) exists is to place conditions on $\Lambda(x)$ so that $L_j(u)$ decreases rapidly at infinity and then no condition need be put on the distribution of the input random variables. The behaviour of the functions $L_j(u)$ at infinity is related to the smoothness of $\Lambda(x)$ throughout its range (Titchmarsh, 1937, p.174). Assume that $\Lambda(x)$ may be extended into the complex plane to a function, $\Lambda(z)$, analytic in a strip, $-c_1 < \mathrm{Im} z < c_2$ $(c_1, c_2 > 0)$, containing the real axis and that

$$\int_{-\infty}^{\infty} |\Lambda(x+iy)| dx < \infty \qquad (-c_1 < y < c_2).$$

Then by changing the contour of integration it may be shown that for any c, $0 \leq c < \min(c_1, c_2)$ there are constants A,B such that

$$|L_j(u)| \leq BA^j e^{-c|u|} \qquad (-\infty < u < \infty). \tag{4.4.7}$$

A fixed $c > 0$ is chosen. Also under these assumptions the inversion formula holds

$$\Lambda^{j}(x) = \frac{1}{2\pi}\int L_j(u)e^{-iux}du \qquad (-\infty < x < \infty). \tag{4.4.8}$$

Using (4.4.7) and the inequality

$$\int_{-\infty}^{\infty}|u+y|^k e^{-h|u|}du = \int |u|^k e^{-h|u-y|}du$$

$$\leq \int|u|^k e^{-h\{|u|-|y|\}}du = e^{h|y|}\frac{2k!}{h^{k+1}} \qquad (h > 0), \tag{4.4.9}$$

it may be shown that the following bound for $\theta_N(s)$ holds for $N \geq 1$

$$|\theta_N(s)| \leq KA^N\{K\sum_{j=1}^{N-1}\frac{|s|^j}{j!}d^{N-j}(1+K)^{N-j-1} + \frac{|s|^N}{N!}\}e^{c|bs|} \tag{4.4.10}$$

where $d = \{c(1 - |b|)\}^{-1}$, $K = Bd\{\pi(1 - |b|)\}^{-1}$ and where the summation appearing is zero for $N = 1$. If the bound holds it follows that the series (4.4.1) is absolutely convergent for β lying in a certain interval about zero.

Firstly, since $\theta_0(s)$ is a characteristic function,

$$|T^{(j,j)}(s)| \leq \frac{1}{2\pi j!}\int|L_j(u)||\theta_0(s-u)|du$$

$$\leq \frac{1}{2\pi j!}\int_{-\infty}^{\infty} BA^j e^{-c|u|}du = \frac{BA^j}{j!2\pi}\cdot\frac{2}{c}.$$

However, in order to obtain a bound with a manageable form, the following inequality (which holds since $c^{-1} = d(1 - |b|) \leq d$) is used

$$|T^{(j,j)}(s)| \leq \frac{BdA^j}{\pi j!}e^{c|s|} \qquad (j \geq 1). \tag{4.4.11}$$

In particular $|T^{(1,1)}(s)| \leq BdA\pi^{-1}e^{c|s|}$ and hence, from (4.4.2),

$$|\theta_1(s)| \leq \sum_{\ell=0}^{\infty}|b^{\ell}s||T^{(1,1)}(b^{\ell+1}s)| \leq \sum_{\ell=0}^{\infty}|s||b|^{\ell}\frac{BdA}{\pi}e^{c|b^{\ell+1}s|}$$

$$\leq \sum_{\ell=0}^{\infty}|s||b|^{\ell}\frac{BdA}{\pi}e^{c|bs|} = \frac{BdA}{\pi(1 - |b|)}|s|e^{c|bs|}$$

$$= AK|s|e^{c|bs|}.$$

Thus (4.4.10) holds for N = 1. Assume that (4.4.10) holds for N = 1,...N*, then, for $j \geq 1$ and $1 \leq N \leq N^*$,

$$|T^{(N+j,j)}(s)| \leq \frac{1}{2\pi j!}\int |L_j(u)||\theta_N(s-u)|du$$

$$\leq \frac{1}{2\pi j!}\int BA^j e^{-c|u|} KA^N\{K\sum_{p=1}^{N-1}\frac{|s-u|^p}{p!}d^{N-p}(1+K)^{N-p-1} + \frac{|s-u|^N}{N!}\}e^{-c|bs-bu|}du.$$

Using the inequality $-c|u| - c|bs-bu| \leq -c(1-|b|)|u| - c|bs|$, the following is obtained by applying (4.4.9) with $h = c(1-|b|) = d^{-1}$:

$$|T^{(N+j,j)}(s)| \leq \frac{BA^{N+j}K}{2\pi j!}\{K\sum_{p=1}^{N-1} d^{N-p}(1+K)^{N-p-1}2d^{p+1} + 2d^{N+1}\}e^{-c|bs|+c(1-|b|)|s|}$$

$$= \frac{2BA^{N+j}Kd^{N+1}}{2\pi j!}\{K.\frac{(1+K)^{N-1}-1}{(1+K)-1} + 1\}\, e^{c|s|}$$

$$= \frac{BdK}{j!\pi}.A^{N+j}d^N(1+K)^{N-1}e^{c|s|} \quad (j \geq 1;\ 1 \leq N \leq N^*). \qquad (4.4.12)$$

The expression (4.4.2) defining $\theta_{N^*+1}(s)$ involves the terms $T^{(N^*+1,j)}(s)$ $(j = 1,...N^*+1)$. For $j = N^*+1$ the term is bounded by (4.4.11) while the bounds for the terms for $j = 1,...N^*$ are provided by (4.4.12) since here $1 \leq N = N^* + 1 - j \leq N^*$. Thus, using also the inequalities

$$|b|^{\ell j} \leq |b|^{\ell} \quad (j \geq 1), \quad e^{c|b^{\ell+1}s|} \leq e^{c|bs|} \quad (\ell \geq 0),$$

$$|\theta_{N^*+1}(s)| \leq \sum_{j=1}^{N^*+1}\sum_{\ell=0}^{\infty}|b^\ell s|^j\,|T^{(N^*+1,j)}(b^{\ell+1}s)|$$

$$\leq \sum_{j=1}^{N^*}\frac{|s|^j}{1-|b|}\frac{BdK}{j!\pi}A^{N^*+1}d^{N^*-j+1}(1+K)^{N^*-j}e^{c|bs|}$$

$$+ \frac{|s|^{N^*+1}}{1-|b|}\cdot\frac{BdA^{N^*+1}}{\pi(N^*+1)!}\, e^{c|bs|}$$

$$= KA^{N^*+1}\{K\sum_{j=1}^{N^*}\frac{|s|^j}{j!}d^{N^*+1-j}(1+K)^{N^*-j} + \frac{|s|^{N^*+1}}{(N^*+1)!}\}e^{c|bs|}. \qquad (4.4.13)$$

This is (4.4.10) for $N = N^* + 1$, hence the bound holds for all $N \geq 1$. Since (4.4.10) holds, another bound is

$$|\theta_N(s)| \leq \left[\frac{K^2\{Ad(1+K)\}^N}{1+K} \exp\left\{\frac{|s|}{d(1+K)}\right\} + \frac{KA^N|s|^N}{N!}\right] e^{c|bs|}$$

and this implies that $\Sigma\beta^N|\theta_N(s)|$ converges for β satisfying $|\beta| < \{Ad(1+K)\}^{-1}$. It would be unreasonable to expect this interval to be a good indication of the region of convergence of the series (4.4.1) since it takes no account at all of the input distribution. Further many of the inequalities on which it is based may be improved considerably. The interval is

$$|\beta| < \frac{c^2(1-|b|)^3}{A\{c(1-|b|)^2 + B\pi^{-1}\}} ,$$

however, since A and B depend upon c, this does not imply that the series (4.4.1) converges for all β when $\Lambda(z)$ is an entire function.

4.4.4 If $F(x;\beta)$ is the marginal stationary distribution function of the process $\{X_n(\beta)\}$ generated by (4.1.6), then $\phi(s;\beta)$, the corresponding characteristic function, must satisfy the equations

$$\phi(s;\beta) = \int e^{isx} dF(x;\beta)$$

$$= \phi_Z(s) \int \exp[is\{a + bx + \beta\Lambda(x)\}] dF(x;\beta)$$

where $\phi_Z(s)$ is the characteristic function of the input distribution. Under the assumptions already made about $\Lambda(x)$, the exponential may be expanded as a series and the integral and summation reversed, giving

$$\phi(s;\beta) = \phi_Z(s) \sum_{j=0}^{\infty} \int \exp\{is(a+bx)\} \frac{(is\beta)^j}{j!} \Lambda^j(x) dF(x;\beta)$$

$$= \phi_Z(s) \{e^{isa}\phi(bs;\beta) + \sum_{j=1}^{\infty} \frac{e^{isa}}{2\pi} \frac{(is\beta)^j}{j!} \int L_j(u)\phi(bs-u;\beta)du\} \quad (4.4.14)$$

where this last follows on substituting for $\Lambda^j(x)$ from (4.4.8) and using the absolute integrability of the resulting double integral to justify reversing the integrations to give (4.4.14).

The argument giving expression (4.4.13) shows that the infinite series defining each $\theta_N(s)$ is absolutely convergent and hence that changing the order of integrations and summations involving these terms is justified. Substitution of the series (4.4.1) for $\phi(s;\beta)$ into the right hand side of (4.4.14) and use of the relations (4.4.2-5) leads, after a little manipulation, to the conclusion that $\phi(s;\beta)$ as defined does satisfy the above equations.

4.4.5 The special forms of the above expressions for $\phi(s;\beta)$ in the important special case when $b = 0$ are now given. Firstly

$$\theta_0(s) = \rho_\ell(s) = e^{isa}\phi_Z(s) \qquad (\ell \geq 0).$$

Since $T^{(N,j)}(s)$ appears only as $T^{(N,j)}(0)$ this will be replaced by $T^{(N,j)}$. Then (4.4.2) becomes

$$\theta_N(s) = \theta_0(s) \sum_{j=1}^{N} (is)^j T^{(N,j)} \qquad (N \geq 1), \tag{4.4.15}$$

and $T^{(N,j)}$ is given by

$$T^{(N,j)} = \frac{1}{2\pi j!}\int L_j(u)\theta_{N-j}(-u)\,du \qquad (1 \leq j \leq N).$$

Substituting for θ_N in this expression gives

$$T^{(N,j)} = \begin{cases} P_j^{(0)} & (j = N), \\ \sum_{n=1}^{N-j} T^{(N-j,n)} P_j^{(n)} & (1 \leq j < N), \end{cases} \tag{4.4.16}$$

where

$$P_j^{(n)} = \frac{1}{2\pi j!}\int L_j(u)(-iu)^n\theta_0(-u)\,du \qquad (j \geq 1;\ n \geq 0).$$

This expression for $P_j^{(n)}$ may be written in various other equivalent ways when the required quantities exist; thus alternatives are

$$P_j^{(n)} = \frac{1}{j!}\int \frac{\partial^n \Lambda^j(x)}{\partial x^n}\, dF_0(x)$$

$$= \frac{1}{j!}\int \Lambda^j(x)\, \frac{\partial^n f_0(x)}{\partial x^n}\, dx$$

$$= \frac{1}{j!}\, \frac{d^n}{dy^n}\left[\int \Lambda^j(x+y)\, dF_0(x)\right]_{y=0} \qquad (4.4.17)$$

The last alternative turns up in a natural way in the method described in section 4.6.2. The derivatives exist for all n if either $\Lambda(x)$ or $F_0(x)$ is continuously differentiable and if the integral exists. Here F_0 is the stationary distribution of the initial process and in this case $F_0(x) = F_Z(x-a)$.

The expression for $\phi(s;\beta)$ can be written in the form

$$\phi(s;\beta) = \theta_0(s)h(s;\beta)$$

$$h(s;\beta) = 1 + \sum_{N=1}^{\infty}\sum_{j=1}^{N} \beta^N (is)^j T(N,j). \qquad (4.4.18)$$

4.5 Joint characteristic functions

4.5.1 The marginal distribution of the process (4.1.1) is given by the characteristic function $\phi(s;1)$ as found in section 4.4 and, once this is known, the joint stationary distribution of any finite collection of values $(X_n, X_{n+\ell_1}, \ldots, X_{n+\ell_k})$ can be found immediately. For the pair (X_n, X_{n+k}) this would involve integrating the joint stationary distribution of $(X_n, X_{n+1}, \ldots, X_{n+k})$ which is known from the structure of the process and the marginal distribution. Thus it would be necessary first to invert the characteristic function and then to perform a number of integrations: it is doubtful whether even the first step of inverting $\phi(s;1)$ could be performed analytically except in very special cases. An alternative which is discussed in this section is to continue to consider the required process $\{X_n\}$ as one of the family of processes $\{X_n(\beta)\}$ and to obtain the joint distribution of $(X_n(\beta), X_{n+k}(\beta))$ as an expansion in powers of β. This is a natural extension of the method for finding the stationary marginal distribution and, in the next chapter, leads to simple procedures for calculating the moments of the stationary joint distributions.

4.5.2 The solution

$$\phi(s;\beta) = \theta_0(s) + \sum_{N=1}^{\infty} \beta^N \theta_N(s)$$

(where $\theta_N(s)$ are given by (4.4.2-6)) for the stationary characteristic functions of the processes $\{X_n(\beta)\}$ may be used as a basis for determining the stationary joint distributions. Let $\phi_k(s,t;\beta)$ and $F_k(x,y;\beta)$ denote the joint stationary characteristic and distribution functions of $(X_n(\beta), X_{n+k}(\beta))$. Then, for $k \geq 1$,

$$\phi_k(s,t;\beta) = \iint \exp\{isx + ity\} F_k(dx,dy;\beta)$$

$$= \phi_Z(t) \iint \exp[isx + it\{a+by+\beta\Lambda(y)\}] F_{k-1}(dx,dy;\beta).$$

This gives, on making a power-series expansion in β and using (4.4.6), for $k \geq 1$,

$$\phi_k(s,t;\beta) = \phi_Z(t)\{e^{ita}\phi_{k-1}(s,bt;\beta) + \frac{e^{ita}}{2\pi} \sum_{j=1}^{\infty} \frac{(it\beta)^j}{j!} \int L_j(u)\phi_{k-1}(s,bt-u;\beta)\, du\}.$$

This defines $\phi_k(s,t;\beta)$ recursively for increasing k commencing with

$$\phi_0(s,t;\beta) = \phi(s+t;\beta).$$

The coefficients of an expansion in powers of β of $\phi_k(s,t;\beta)$ may be identified by equating powers of β in the above equation. Thus

$$\phi_k(s,t;\beta) = \theta_{0,k}(s,t) + \sum_{N=1}^{\infty} \beta^N \theta_{N,k}(s,t), \tag{4.5.1}$$

$$\theta_{N,0}(s,t) = \theta_N(s+t) \qquad (N \geq 0) \tag{4.5.2}$$

$$\theta_{0,k}(s,t) = e^{ita}\phi_Z(t)\theta_{0,k-1}(s,bt) \qquad (k \geq 1). \tag{4.5.3}$$

With (4.5.2) and (4.4.4) this gives, for $k \geq 1$,

$$\theta_{0,k}(s,t) = \exp\{\frac{i(s+t)a}{1-b}\} \prod_{m=0}^{k-1} \phi_Z(b^m t) \prod_{n=0}^{\infty} \phi_Z(b^n(s+b^k t)),$$

which is the joint characteristic function of $(X_n(0), X_{n+k}(0))$ from the initial linear process. The other coefficients $\theta_{N,k}$ $(N \geq 1;\ k \geq 1)$ are given by

$$\theta_{N,k}(s,t) = e^{ita}\phi_Z(t)\{\theta_{N,k-1}(s,bt) + \sum_{j=1}^{N} (it)^j R_k^{(N,j)}(s,bt)\} \qquad (4.5.4)$$

where, for $k \geq 1$,

$$R_k^{(N,j)}(s,t) = \frac{1}{2\pi j!}\int L_j(u)\theta_{N-j,k-1}(s,t-u)du \qquad (1 \leq j \leq N). \qquad (4.5.5)$$

Note that $R_1^{(N,j)}(s,t) = T^{(N,j)}(s+t)$ where $T^{(N,j)}(\cdot)$ is given by (4.4.3).

Since $\theta_N(s)$ given by (4.4.2) satisfies, for $N \geq 1$,

$$\theta_N(s) = e^{isa}\phi_Z(bs)\{\theta_N(bs) + \sum_{j=1}^{N} (is)^j T^{(N,j)}(bs)\},$$

the above formulae imply that, for all $k \geq 0$,

$$\theta_{N,k}(0,t) = \theta_N(t), \quad \phi_k(0,t;\beta) = \phi(t;\beta).$$

Obviously this should hold since the marginal distributions of $X_n(\beta)$ and $X_{n+k}(\beta)$ are the same by the stationarity assumption.

The above expressions give the expansion in powers of β of the joint characteristic function of the pair $(X_n(\beta), X_{n+k}(\beta))$ from the stationary process $\{X_n(\beta)\}$. Joint characteristic functions of larger (finite) collections of values from the process can also be found in the obvious way.

4.5.3 The special form of the above expressions in the case $b = 0$ will now be given. Formulae (4.5.2), (4.5.3) give

$$\theta_{0,k}(s,t) = \begin{cases} \exp\{i(s+t)a\}\phi_Z(s+t) = \theta_0(s+t) & (k = 0), \\ \exp\{i(s+t)a\}\phi_Z(s)\phi_Z(t) = \theta_0(s)\theta_0(t) & (k \geq 1). \end{cases} \qquad (4.5.6)$$

Since $R_k^{(N,j)}$ appears only as $R_k^{(N,j)}(s,0)$ this is written as $R_k^{(N,j)}(s)$. Then (4.5.4) becomes

$$\theta_{N,k}(s,t) = \theta_0(t)\{\theta_{N,k-1}(s,0) + \sum_{j=1}^{N} (it)^j R_k^{(N,j)}(s)\} \qquad (N,k \geq 1).$$

This implies that $\theta_{N,k}(s,0) = \theta_{N,k-1}(s,0) = \theta_{N,0}(s,0) = \theta_N(s)$ using (4.5.2). Hence

$$\theta_{N,k}(s,t) = \theta_0(t)\{\theta_N(s) + \sum_{j=1}^{N} (it)^j R_k^{(N,j)}(s)\} \qquad (N,k \geq 1). \tag{4.5.7}$$

Let the functions $P_j(y,s)$ be defined by

$$P_j(y,s) = \frac{1}{j!} \int e^{isx} \Lambda^j(x+y) dF_0(x) \qquad (j \geq 1). \tag{4.5.8}$$

This is equivalent to the formula

$$P_j(y,s) = \frac{1}{2\pi j!} \int e^{-iuy} L_j(u) \theta_0(s-u) du \qquad (j \geq 1)$$

when Λ^j is written in terms of its transform. Then

$$P_j^{(n)}(0,s) = \frac{\partial^n}{\partial y^n} P_j(y,s)\Big|_{y=0} \tag{4.5.9}$$

$$= \frac{1}{2\pi j!} \int \{-iu\}^n L_j(u) \theta_0(s-u) du$$

and the quantities $P_j^{(n)}$ used earlier (equations 4.4.16-7) are given by

$$P_j^{(n)} = P_j^{(n)}(0,0) \qquad (j \geq 1;\ n \geq 0).$$

Using these quantities it is possible to substitute for $\theta_{N,k}$ in formula (4.5.5) for $R_k^{(N,j)}$ to obtain the expressions

$$R_k^{(N,j)}(s) = \begin{cases} \begin{cases} P_j^{(0)}(0,s) & (j = N) \\ \sum_{n=1}^{N-j} T^{(N-j,n)} \sum_{q=0}^{n} \binom{n}{q} (is)^q P_j^{(n-q)}(0,s) & (1 \leq j < N) \end{cases} & (k = 1), \\ \begin{cases} \theta_0(s) P_j^{(0)} & (j = N) \\ \theta_{N-j}(s) P_j^{(0)} + \sum_{n=1}^{N-j} R_{k-1}^{(N-j,n)}(s) P_j^{(n)} & (1 \leq j < N) \end{cases} & (k \geq 2). \end{cases} \tag{4.5.10}$$

With (4.5.6-7), (4.5.1) becomes, for $k \geq 1$,

$$\phi_k(s,t;\beta) = \theta_0(s)\theta_0(t) + \sum_{N=1}^{\infty} \beta^N \theta_0(t)\{\theta_N(s) + \sum_{j=1}^{N} (it)^j R_k^{(N,j)}(s)\}$$

$$= \theta_0(t)\{\phi(s;\beta) + \sum_{N=1}^{\infty} \sum_{j=1}^{N} \beta^N (it)^j R_k^{(N,j)}(s)\}. \qquad (4.5.11)$$

An alternative and useful form can be obtained by defining $T_k^{(N,j)}(s)$ by

$$R_k^{(N,j)}(s) = \theta_0(s) T_k^{(N,j)}(s).$$

Then

$$\phi_k(s,t;\beta) = \theta_0(s)\theta_0(t)\{1 + \sum_{N=1}^{\infty} \sum_{j=1}^{N} \beta^N (is)^j T^{(N,j)} + \sum_{N=1}^{\infty} \sum_{j=1}^{N} \beta^N (it)^j T_k^{(N,j)}(s)\} \qquad (4.5.12)$$

where

$$T_k^{(N,j)}(s) = \begin{cases} P_j^{(0)}(0,s)/\theta_0(s) & (j = N) \\ & (k = 1), \\ \sum_{n=1}^{N-j} T^{(N-j,n)} \sum_{q=0}^{n} \binom{n}{q} (is)^q \{P_j^{(n-q)}(0,s)/\theta_0(s)\} & (1 \leq j < N) \end{cases} \qquad (4.5.13)$$

$$\begin{cases} P_j^{(0)} & (j = N) \quad (k \geq 2). \\ P_j^{(0)} \sum_{n=1}^{N-j} (is)^n T^{(N-j,n)} + \sum_{n=1}^{N-j} T_{k-1}^{(N-j,n)}(s) P_j^{(n)} & (1 \leq j < N) \end{cases}$$

4.6 Stationary distribution as the limit of conditional distributions

4.6.1 In section 4.4 the stationary characteristic functions $\phi(s;\beta)$ were found by using the identity of the distribution of $X_n(\beta)$ and $X_{n+1}(\beta)$ when the process is stationary. Now the distribution is found as that obtained by conditioning on an event increasingly far in the past. This method avoids the assumptions of both section 4.2 and section 4.4 and leads also to expressions for conditional characteristic functions. However, it has not been extended to give expressions for joint characteristic functions similar to those found in section 4.5.

For the process $\{X_n(\beta)\}$ generated by (4.1.6) let

$$Y_n(\beta) = bX_n(\beta) + \beta\Lambda\{X_n(\beta)\}$$

and $$Z_n^* = a + Z_n \qquad (n = \ldots -1,0,1,\ldots).$$

Then $$X_{n+1}(\beta) = Y_n(\beta) + Z_{n+1}^*$$

and $$Y_{n+1}(\beta) = b\{Y_n(\beta) + Z_{n+1}^*\} + \beta\Lambda\{Y_n(\beta) + Z_{n+1}^*\} \quad (n = \ldots -1,0,1,\ldots). \tag{4.6.1}$$

Here the random variables $Y_n(\beta)$, Z_{n+1}^* are independent for each n. Let $\phi_Y(s;\beta)$ denote the stationary characteristic function of $Y_n(\beta)$ and let ϕ_{Z^*}, F_{Z^*} be the characteristic function and distribution function of Z_n^*. Then $\phi(s;\beta)$, the characteristic function of $X_n(\beta)$, is given by

$$\phi(s;\beta) = \phi_{Z^*}(s)\phi_Y(s;\beta).$$

The function

$$\phi_{Y,0}(s;y;\beta) = L_{0,0}(s,y) = e^{isy} \tag{4.6.2}$$

is, trivially, the conditional characteristic function for each n and β of $Y_n(\beta)$ given that $Y_n(\beta) = y$. Using (4.6.1) the conditional characteristic function of $Y_{n+1}(\beta)$ given that $Y_n(\beta) = y$ is

$$\phi_{Y,1}(s;y;\beta) = \int L_{0,0}(s,by+bz+\beta\Lambda(y+z))dF_{Z^*}(z).$$

Because of the definition (4.6.2) the integral may be expanded as a power series, giving

$$\phi_{Y,1}(s;y;\beta) = \int \sum_{j=0}^{\infty} \beta^j \frac{\Lambda^j(y+z)}{j!} L_{0,0}^{(j)}(s,by+bz)dF_{Z^*}(z),$$

where

$L_{0,0}^{(j)}(s,y) = \frac{\partial^j}{\partial y^j} L_{0,0}(s,y)$ $(j \geq 0)$. Then, writing

$$L_{j,1}(s;y) = \int \frac{\Lambda^j(y+z)}{j!} L_{0,0}^{(j)}(s,by+bz)dF_{Z^*}(z), \tag{4.6.3}$$

the conditional characteristic function is

$$\phi_{Y,1}(s;y;\beta) = \sum_{j=0}^{\infty} \beta^j L_{j,1}(s;y).$$

Because of the convolution-like form of (4.6.3), $L_{j,1}(s,y)$ is continuously differentiable (with respect to y) if one or both of $\Lambda(\cdot)$ or $F_{Z^*}(\cdot)$ are. It will be assumed that there is a Taylor series expansion for $L_{j,1}(s,y)$ about any point in the y-range. Then, extending the notation as required, the conditional characteristic function of $Y_{n+2}(\beta)$ given that $Y_n(\beta) = y$ is

$$\phi_{Y,2}(s;y;\beta) = \int \sum_{j=0}^{\infty} \beta^j L_{j,1}(s,by+bz+\beta\Lambda(y+z))dF_{Z^*}(z)$$

$$= \int \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \beta^{j+k} \frac{\Lambda^k(y+z)}{k!} L_{j,1}^{(k)}(s,by+bz)dF_{Z^*}(z)$$

$$= \sum_{j=0}^{\infty} \beta^j L_{j,2}(s,y),$$

where $L_{j,2}(s,y)$ is given by an integral over the distribution F_{Z^*}.

The above procedure may be repeated, obtaining the conditional characteristic function of $Y_{n+m+1}(\beta)$ given that $Y_n(\beta) = y$ as

$$\phi_{Y,m+1}(s;y;\beta) = \sum_{j=0}^{\infty} \beta^j L_{j,m+1}(s,y)$$

where $L_{j,m+1}(s,y)$ is given in terms of $L_{j,m}(s,y)$ the corresponding functions from the previous step. This relation is given by

$$L_{j,m+1}(s,y) = \sum_{p=0}^{j} \int L^{(p)}_{j-p,m}(s,by+bz) \frac{\Lambda^p(y+z)}{p!} dF_{Z^*}(z) \quad (j,m \geq 0), \quad (4.6.4)$$

where $$L_{j,0}(s,y) = \begin{cases} e^{isy} & (j = 0), \\ 0 & (j \geq 1), \end{cases} \quad (4.6.5)$$

and where $L^{(p)}_{j,m}(s,y) = \frac{\partial^p}{\partial y^p} L_{j,m}(s,y) \qquad (j,m,p \geq 0).$

Then, by the assumption of stationarity, when the processes are not periodic

$$\phi_Y(s;\beta) = \lim_{m\to\infty} \sum_{j=0}^{\infty} \beta^j L_{j,m}(s,y) \quad (4.6.6)$$

since this is the limit as $m \to \infty$ of the conditional characteristic function of $Y_n(\beta)$ given that $Y_{n-m}(\beta) = y$. It follows that the right hand side of (4.6.6) should be independent of y unless the process has more than one stationary distribution. Even if the process is periodic with more than one possible stationary distribution, for any fixed y the stationary distribution is defined by the characteristic function

$$\phi_Y(s;\beta) = \lim_{M\to\infty} \frac{1}{M} \sum_{m=1}^{M} \sum_{j=0}^{\infty} \beta^j L_{j,m}(s,y) .$$

It is not possible to find $L_{j,\infty}(s) = \lim_{M\to\infty} L_{j,m}(s,y)$ directly as a solution to (4.6.4) since, taking into account that $L_{j,\infty}$ is independent of y, the equation becomes

$$L_{j,\infty}(s) = \int L_{j,\infty}(s) dF_{Z^*}(z) \equiv L_{j,\infty}(s)$$

which is trivially satisfied by any function $L_{j,\infty}(s)$. Thus the solution must be found explicitly using (4.6.5). Equation (4.6.4) is easily solved for $L_{0,n}(s,y)$ $(n \geq 0)$ since these are given by

$$L_{0,m+1}(s,y) = \int L_{0,m}(s,by+bz) dF_{Z^*}(z) \qquad (m \geq 0)$$

and, since $L_{0,0}(s,y) = e^{isy}$, this leads to the formula

$$L_{0,n}(s,y) = \prod_{r=1}^{n} \phi_{Z^*}(b^r s) e^{isb^n y} \qquad (n \geq 1). \quad (4.6.7)$$

The limit of this as $n \to \infty$ is independent of y.

By repeated substitution in (4.6.4) the following set of formulae may be derived. Define for $j,m,p \geq 0$,

$$\ell^{(p)}_{j,m,0}(s,y) = \int L^{(p)}_{j,m}(s,by+bz)\,\frac{\Lambda^p(y+z)}{p!}\,dF_{Z^*}(z)$$

$$\ell^{(p)}_{j,m,q}(s,y) = \int \ell^{(p)}_{j,m,q-1}(s,by+bz)\,dF_{Z^*}(z) \qquad (q = 1,2,3,\ldots).$$

Then

$$L_{j,m+1}(s,y) = \sum_{n=0}^{m}\sum_{p=1}^{j} \ell^{(p)}_{j-p,m-n,n}(s,y) \qquad (j \geq 1;\ m \geq 0). \quad (4.6.8)$$

This gives the coefficient of β^j in terms of those of lower order. However it does not seem possible to find the limit as $m \to \infty$ of $L_{j,m}(s,y)$ explicitly.

Define

$$\rho^*_n(s) = \begin{cases} 1 & (n = 0), \\ \prod_{r=1}^{n} \phi_{Z^*}(b^r s) & (n \geq 1). \end{cases}$$

These are related to the quantities $\rho_n(s)$, used earlier in (4.3.5) and (4.4.5), by

$$\rho_n(s) = \phi_{Z^*}(s)\rho^*_n(s).$$

Then, using (4.6.7) and (4.6.8),

$$L_{1,m+1}(s,y) = \sum_{n=0}^{m} \ell^{(1)}_{0,m-n,n}(s,y) = \sum_{n=0}^{m} \ell^{(1)}_{0,n,m-n}(s,y) \qquad (4.6.9)$$

where, for $q \geq 0$,

$$\ell^{(1)}_{0,m,q}(s,y) = \rho^*_m(s)\,ib^m s \int \exp\{isb^m(b^{q+1}y + b^{q+1}z_0 + b^q z_1 + \ldots + bz_q)\}$$
$$\cdot \Lambda(b^q y + b^q z_0 + b^{q-1}z_1 + \ldots + z_q) \prod_{i=0}^{q} dF_{Z^*}(z_i)\,.$$

Substituting this in (4.6.9)

$$L_{1,m+1}(s,y) = \sum_{n=0}^{m} \rho_n^*(s) ib^n s \int \exp\{isb^n(b^{m-n+1}y + b\sum_{j=0}^{m-n} b^j z_j)\}$$
$$\Lambda(b^{m-n}y + \sum_{j=0}^{m-n} b^j z_j) \prod_{i=0}^{m-n} dF_{Z^*}(z_i) .$$

The distribution of $\sum_{j=0}^{q} b^j Z_j^*$ tends, as $q \to \infty$, to the stationary distribution F_0 of the initial process $\{X_n(0)\}$ given by

$$X_{n+1}(0) = a + bX_n(0) + Z_{n+1} = bX_n(0) + Z_{n+1}^* ,$$

and hence

$$\lim_{m\to\infty} L_{1,m+1}(s,y) = \sum_{n=0}^{\infty} \rho_n^*(s) ib^n s \int \exp\{ib^{n+1}sx\}\Lambda(x)dF_0(x).$$

4.6.2 Once again the expressions simplify greatly when $b = 0$. Formula (4.6.4) becomes

$$L_{j,m+1}(s,y) = \sum_{q=0}^{j} L_{j-q,m}^{(q)}(s,0)P_q(y) \qquad (j,m \geq 0). \tag{4.6.10}$$

where $$P_q(y) = \int \frac{\Lambda^q(y+z)}{q!} dF_{Z^*}(z)$$

and in particular $P_0(y) = 1$. From either (4.6.7) or (4.6.10)

$$L_{0,n}(s,y) = \begin{cases} e^{isy} & (n = 0), \\ 1 & (n \geq 1). \end{cases} \tag{4.6.11}$$

Hence (4.6.10) becomes, for $j \geq 1$,

$$L_{j,m+1}(s,y) = \begin{cases} (is)^j P_j(y) & (m = 0), \\ \sum_{q=0}^{j-1} L_{j-q,m}^{(q)}(s,0)P_q(y) & (m \geq 1). \end{cases} \tag{4.6.12}$$

From this,

$$L_{1,1}(s,y) = is\,P_1(y)$$

$$L_{1,2}(s,y) = L_{11}(s,0)P_0(y) = is\,P_1(0)$$

and $L_{1,m+1}(s,y) = L_{1,m}(s,0)P_0(y) = L_{1,m}(s,0) \quad (m = 2,3,\ldots.)$.

Thus, for $m \geq 2$, $L_{1,m}(s,y)$ is independent of y and $L_{1,m}(s,y) = L_{1,1}(s,0)$ $(m \geq 2)$. It may be shown that, for all j,

$$L_{j,j+r}(s,y) = L_{j,j}(s,0) \qquad (r \geq 1)$$

and thus that $L_{j,j+r}(s,y)$ is independent of y for $r \geq 1$. Suppose that this is true for $j = 1,\ldots j^*-1$. Then

$$L^{(q)}_{j^*-q,j^*+r}(s,0) = 0 \qquad (q = 1,\ldots j^*-1;\ r \geq 0),$$

and hence (4.6.12) gives, for $r \geq 0$,

$$L_{j^*,j^*+r+1}(s,y) = L_{j^*,j^*+r}(s,0)P_0(y) = L_{j^*,j^*+r}(s,0).$$

Hence $L_{j^*,j^*+r+1}(s,y) = L_{j^*,j^*}(s,0) \quad (r \geq 0)$. (4.6.13)

However note that $L_{j^*,j^*}(s,y)$ is not independent of y. Because of these results (4.6.12) now becomes, for $j \leq 1$,

$$L_{j,m+1}(s,y) = \begin{cases} (is)^j P_j(y) & (m = 0), \\ \sum_{q=0}^{j-m} L^{(q)}_{j-q,m}(s,0)P_q(y) & (j > m \geq 1), \\ L_{j,m}(s,0) = L_{j,j}(s,0) & (m \geq j). \end{cases} \qquad (4.6.14)$$

Define $P_j^{(n)} = \dfrac{d^n}{dy^n}\{P_j(y)\}_{y=0}$ $(n,j \geq 0)$, then as can be seen by substitution in the above,

$$L_{j,m+1}(s,y) = L_{j,m}(s,0) + \Sigma (is)^{r_1} P_{r_1}^{(r_2)} P_{r_2}^{(r_3)} \ldots\ldots P_{r_m}^{(r_{m+1})} P_{r_{m+1}}(y)$$

where the summation is over all sets of (m+1) integers $r_1,\ldots r_{m+1} \geq 1$, such that $r_1 + r_2 + \ldots + r_{m+1} = j$. Note that $L_{j,0}(s,y)$ is given by (4.6.5) and hence that $L_{j,0}(s,y) = 0$ $(j \geq 1)$. Therefore

$$L_{j,m}(s,0) = \sum_{n=1}^{m} \Sigma (is)^{r_1} P_{r_1}^{(r_2)} P_{r_2}^{(r_3)} \ldots\ldots P_{r_{n-1}}^{(r_n)} P_{r_n}^{(0)} \qquad (1 \leq m \leq j)$$

where now the combined summation is over all sets of $n \leq m$ integers $(r_1,\ldots r_n \geq 1)$ satisfying $r_1 + \ldots + r_n = j$.

Let $T^{(N,j)}$, $N \geq j \geq 1$, be given by (4.4.16), i.e.

$$T^{(N,j)} = \begin{cases} P_j^{(0)} & (j = N), \\ \sum_{n=1}^{N-j} T^{(N-j,n)} P_j^{(n)} & (1 \leq j < N), \end{cases}$$

then

$$T^{(N,j)} = \Sigma P_{r_1}^{(0)} P_{r_2}^{(r_1)} \ldots\ldots P_{r_k}^{(r_{k-1})}$$

where the summation is over all sets of $k \geq 1$ integers $(r_1,\ldots,r_k \geq 1)$ satisfying $r_1 + r_2 + \ldots + r_k = N$ and $r_k = j$. With these restrictions k can be no larger than N-j+1. By comparison, it can be seen that,

$$L_{N,m}(s,0) = \sum_{j=N+1-m}^{N} (is)^j T^{(N,j)} \qquad (1 \leq m \leq N)$$

and, in particular,

$$L_{N,N}(s,0) = \sum_{j=1}^{N} (is)^j T^{(N,j)} \qquad (N \geq 1).$$

Because of the results (4.6.13) and (4.6.11) the limit in (4.6.6) becomes, if conditions of uniform convergence hold,

$$\phi_Y(s;\beta) = 1 + \sum_{N=1}^{\infty} \beta^N L_{NN}(s,0)$$

$$= 1 + \sum_{N=1}^{\infty} \sum_{j=1}^{N} \beta^N(is) j_T(N,j) \tag{4.6.15}$$

$$= 1 + \sum_{j=1}^{\infty} \sum_{N=j}^{\infty} \beta^N(is) j_T(N,j) \tag{4.6.16}$$

Then (4.6.5) gives

$$\phi(s;\beta) = e^{isa} \phi_Z(s) \{1 + \sum_{N=1}^{\infty} \sum_{j=1}^{N} \beta^N(is) j_T(N,j)\} \tag{4.6.17}$$

and this is exactly the same result as (4.4.18), obtained by a different method.

The m-step conditional characteristic function of $Y_{n+m}(\beta)$ given that $Y_n(\beta) = y$ is, for $m \geq 1$,

$$\phi_{Y,m}(s;y;\beta) = 1 + \sum_{N=1}^{\infty} \beta^N L_{N,m}(s,y)$$

and using the above results this becomes, for $m \geq 1$,

$$\phi_{Y,m}(s;y;\beta) = 1 + \sum_{N=1}^{m-1} \beta^N L_{N,N}(s,0) + \sum_{N=m}^{\infty} \beta^N L_{N,m}(s,y)$$

$$= 1 + \sum_{N=1}^{m-1} \sum_{j=1}^{N} \beta^N(is) j_T(N,j) + \sum_{N=m}^{\infty} \sum_{j=N+2-m}^{N} \beta^N(is) j_T(N,j)$$

$$+ \sum_{N=m}^{\infty} \sum_{\substack{r_1+\ldots+r_m=N \\ r_i \geq 1}} \beta^N(is)^{r_1} P_{r_1}^{(r_2)} P_{r_2}^{(r_3)} \ldots P_{r_{m-1}}^{(r_m)} P_{r_m}(y) \tag{4.6.18}$$

Here the first two summations are empty for m = 1. A particular case, the conditional characteristic function of $Y_{n+m}(\beta)$ given that $Y_n(\beta) = 0$ (equivalently, $\Lambda\{X_n(\beta)\} = 0$), is found by identifying the last part of the above expression as

$$\sum_{N=m}^{\infty} \sum_{r_m=N+1-m}^{N} \beta^N(is)^{r_m} T^{(N,r_m)} .$$

This gives

$$\phi_{Y,m}(s;0;\beta) = 1 + \sum_{N=1}^{m-1}\sum_{j=1}^{N} \beta^N(is)^j T(N,j) + \sum_{N=m}^{\infty}\sum_{j=N+1-m}^{N} \beta^N(is)^j T(N,j) \tag{4.6.19}$$

$$= 1 + \sum_{j=1}^{\infty}\sum_{N=j}^{j+m-1} \beta^N(is)^j T(N,j).$$

This may be compared with (4.6.16).

The same result for the stationary characteristic function has been found in two ways, firstly by considering the process (4.1.6) when it is in equilibrium and secondly by considering the process at two points in time which become increasingly separated. This second method requires that the functions

$$P_j(y) = \int \frac{\{\Lambda(y+z)\}^j}{j!} dF_{Z^*}(z) \qquad (j \geq 1) \tag{4.6.20}$$

should be continuously differentiable, a condition which holds when either or both of $\Lambda(x)$, $F_Z(z)$ are continuously differentiable. The function $\Lambda(x)$ need not be bounded at $\pm\infty$, although this condition was necessary for the existence of $L_j(u)$ in section 4.4. The convergence of the series obtained has not been demonstrated but clearly the validity of the interchange of limits

$$\lim_{n\to\infty} \sum_{j=1}^{\infty}\sum_{N=j}^{j+m-1} \beta^N(is)^j T(N,j) = \sum_{j=1}^{\infty}\sum_{N=j}^{\infty} \beta^N(is)^j T(N,j)$$

is necessary. Since the convergence of the series and its validity as a solution to the problem have been demonstrated to hold in some circumstances it is not unreasonable to assume that, whenever the series and others derived from it converge, they converge to the correct answer. Results obtained by making this assumption have been checked by simulations in some cases, see Chapter 6.

4.7 Choice of constants

The constants a and b appearing throughout have been arbitrary (except for the condition $|b| < 1$). The choice $b = 0$ considerably simplifies the formulae involved and greatly reduces the amount of calculation needed to find the solution $\phi(s;\beta)$ to any given order in β: in many cases this would be a reasonable choice to make. There then remains the constant a. Although this too is arbitrary, it affects the speed of convergence, and indeed the convergence itself, of the series. Hence a may need to be chosen with some care.

It is reasonably clear that the sequence $\{\phi_{Y,m}(s;y;\beta);\ m = 1,2,3,\ldots\}$ will converge fastest to its limit $\phi_Y(s;\beta)$ if the value y is a typical (rather than extreme) value of $Y_n(\beta)$ under the stationary distribution. For otherwise, starting at an extreme value, the succeeding values $Y_{n+1}(\beta)$, $Y_{n+2}(\beta)$,... would take some time to drift back to the more probable regions of the process $\{Y_n(\beta)\}$. Comparing (4.6.15) with (4.6.18) and (4.6.19), it can be seen that, for any m, $\phi_Y(s;\beta)$ agrees with $\phi_{Y,m}(s;y;\beta)$ in terms up to β^{m-1} and with $\phi_{Y,m}(s;0;\beta)$ in terms up to order m. Thus it might be expected that the series (4.6.15) would converge well if $Y_n(\beta) = 0$ is a very likely, or at least typical, value. Usually the process of interest has $\beta = 1$ and so a good choice of a is such that

$$Y_n(1) = \Lambda(X_n(1)) = \lambda\{X_n(1)\} - a = 0 \qquad (4.7.1)$$

is a typical event under the stationary distribution of $\{X_n(1)\} \equiv \{X_n\}$. This requirement gives some form to the intuitive idea that a is best chosen so that the graphs of $y = \lambda(x)$ and $y = a$ are close.

However, even with the above condition, a is still a disposable constant and quantities derived from truncations of $\phi(s;\beta)$ in (4.6.17) may all be found for different choices of a. Since the true values of these quantities should all be independent of a, this provides a check, not only on whether the series have converged, but also on the validity of the algorithms used to calculate these quantities.

Equation (4.6.20) may be written as

$$P_j(y) = \frac{1}{j!}\int \{\lambda(z + a + y) - a\}^j dF_Z(z)$$

and, since the final expressions depend only on the derivatives of this at zero, it seems that, if F_Z is constant over certain regions of the line, the values of $\lambda(x)$ may be changed on corresponding regions without affecting the solution. It has been assumed that a Taylor series expansion of $P_j(y)$ about zero converges over a region large enough to cover all values $\{\lambda(z + a) - a\}$ where z is a point of increase of F_Z. (An even stricter condition may be needed to ensure convergence of the series for $\phi(s;\beta)$.) There still may be some regions where the values of $\lambda(x)$ may be changed without affecting the validity of the method and it is intuitively reasonable that this should be the case. For, suppose that the stationary process never visits a certain set, then the autoregression function may be changed on this set without affecting the invariance of the original distribution. If the autoregression function and input distribution are such that there are two or more different stationary distributions, it is possible that the above property would allow these distributions to be calculated by using different choices of a where, as discussed above, the choice of a would identify the particular stationary process (equation (4.7.1)).

It is clear that, in some sense, the closeness of the line $y = \lambda(x)$ and the initial autoregression function $y = a + bx$ affects the convergence of the series derived. It has been seen that the choice $b = 0$ greatly reduces the work of finding each coefficient and thus that, even at the expense of evaluating more coefficients, taking $b = 0$ is best in many cases. However, when the "overall slope" of $\lambda(x)$ (in relation to the unknown stationary distribution) is near ± 1 the approximation of $y = \lambda(x)$ by $y = a$ is poor, and an extremely large number of terms in the expansion of powers of β may be required with the choice $b = 0$ and this may not always be best. In particular if $\lambda(x)$ is very close to some line $a + bx$, then choosing these particular values for a and b may result in only one or two terms in the expansion being required.

The following formula for $\theta_1(s)$ in (4.4.1) is obtained by both methods presented, i.e. from (4.4.2-6) and from section 4.6,

$$\theta_1(s) = \sum_{\ell=0}^{\infty} ib^{\ell} s \rho_{\ell}(s) \int \Lambda(x) e^{ib^{\ell+1}sx} dF_0(x) \quad . \tag{4.7.2}$$

This illustrates two difficulties involved in a choice $b \neq 0$. Firstly the stationary distribution F_0 of the initial (linear) process must be known: only in special cases can this be written in explicit form

(section 2.2.2). Secondly there is the evaluation of the infinite summation appearing in (4.7.2): this cannot generally be done explicitly unless $\Lambda(x)$ is a linear function. The expressions for $\theta_N(s)$ each contain N infinite summations. It is usually not of interest to find $\phi(s;\beta)$ itself but only quantities derived from it; for example, moments of the processes are given by the derivatives (with respect to s) at zero. The same difficulties arise when dealing with the corresponding expansions.

5 MOMENTS AND OTHER DISTRIBUTIONAL PROPERTIES

5.1 Introduction

The problem of finding the stationary distribution of non-linear autoregressive processes has been considered and this has led to expressions for the characteristic functions of the marginal and joint distributions. These have been given in Chapter 4. Although, in principle, the characteristic functions completely determine the corresponding distributions, the expressions found are not directly of use as they are in the form of infinite series. However, from the expressions for the characteristic functions, expansions can be derived for quantities of interest such as the moments and joint moments of the stationary process and for the stationary distributions themselves. These are considered here and lead to simple computational procedures for calculating the stationary moments and distributions. The most useful procedures considered are based on the family of processes with varying autoregression function starting from the process of independent and identically distributed values $\{X_n(0)\}$ given by

$$X_{n+1}(0) = a + Z_{n+1} \qquad (n = \ldots -1,0,1,\ldots).$$

5.2 Expansions for moments and joint moments

5.2.1 In Chapter 4 the coefficients $\theta_N(s)$ in the expansion

$$\phi(s;\beta) = \sum_{N=0}^{\infty} \beta^N \theta_N(s)$$

of the stationary characteristic functions of the processes $\{X_n(\beta)\}$ (generated by (4.1.6)) have been given. Then, provided the moments exist,

$$E\{X_n^k(\beta)\} = (-i)^k \frac{\partial^k}{\partial s^k} \phi(s;\beta)\Big|_{s=0} = \sum_{N=0}^{\infty} \beta^N (-i)^k \frac{\partial^k}{\partial s^k} \theta_N(s)\Big|_{s=0}$$

$$= k! \sum_{N=0}^{\infty} \beta^N \theta_N^{(k)}(0) \qquad (5.2.1)$$

where

$$\theta_N^{(k)}(s) = \frac{(-i)^k}{k!}\frac{\partial^k}{\partial s^k}\theta_N(s) \qquad (N,k \geq 0).$$

In particular $k!\theta_0^{(k)}(0) = E\{X_n^k(0)\}$, the moments of the initial process. The extra factor of $(-i)^k/k!$ is used here and later in order to simplify the expressions obtained.

For the family of processes $\{X_n(\beta)\}$ corresponding to the family of autoregression functions $\mu(x;\beta) = a + bx + \beta\{\lambda(x) - a - bx\}$, the coefficients $\theta_N(s)$ are given by (4.4.2-6). Using these and defining for $k \geq 0$,

$$T^{(N,j)(k)}(s) = \frac{(-i)^k}{k!}\frac{\partial^k}{\partial s^k}T^{(N,j)}(s) \qquad (1 \leq j \leq N),$$

$$\rho_\ell^{(k)}(s) = \frac{(-i)^k}{k!}\frac{\partial^k}{\partial s^k}\rho_\ell(s) \qquad (\ell \geq 0),$$

the expression for $\theta_N^{(k)}(s)$ is found to be, for $k \geq 0$, $N \geq 1$,

$$\theta_N^{(k)}(s) = \sum_{j=1}^{N}\sum_{\ell=0}^{\infty}\sum_{q=0}^{\min(k,j)}\sum_{p=0}^{k-q} b^{(\ell+1)(k-q-p)+q\ell}(ib^\ell s)^{j-q}\binom{j}{q} \cdot T^{(N,j)(k-q-p)}(b^{\ell+1}s)\rho_\ell^{(p)}(s)$$

and hence

$$\theta_N^{(k)}(0) = \sum_{j=1}^{\min(N,k)}\sum_{\ell=0}^{\infty}\sum_{p=0}^{k-j} b^{(\ell+1)(k-j-p)+j\ell}T^{(N,j)(k-j-p)}(0)\rho_\ell^{(p)}(0)$$

$$= \sum_{j=1}^{\min(N,k)}\sum_{\ell=0}^{\infty}\sum_{p=0}^{k-j} b^{(\ell+1)p+j\ell}T^{(N,j)(p)}(0)\rho_\ell^{(k-j-p)}(0). \qquad (5.2.2)$$

The terms $\rho_\ell^{(p)}(0)$ are essentially moments of the distributions described after (4.3.5). The cumulants $\kappa_j^{(\ell)}$ of these distributions are related to those of the random variables $Z_n^* = a + Z_n$, κ_j^* say, by

$$\kappa_j^{(\ell)} = \frac{1 - b^{(\ell+1)j}}{1 - b^j}\kappa_j^*.$$

Using the expressions relating non-central moments to cumulants (Kendall and Stuart, 1969, p.68) the infinite summation over the index ℓ appearing in (5.2.2) may, in principle, be evaluated analytically so that this need be no problem. However, defining $L_j^{(p)}(u) = \frac{(-i)^p}{p!}\frac{\partial^p}{\partial u^p} L_j(u)$,

$$T^{(N,j)(p)}(s) = \frac{1}{2\pi j!}\int L_j(u)\theta_{N-j}^{(p)}(s-u)du$$

$$= \frac{1}{2\pi j!}\int L_j^{(p)}(u)\theta_{N-j}(s-u)du$$

and thus it can be seen that the quantities $T^{(N,j)(p)}(0)$ cannot be evaluated without finding $\theta_{N-j}(u)$ for all u (these are given by (4.4.2)). Because of the much greater simplicity obtained by taking b = 0 this has not been pursued. It is however possible that using either a numerical integration or an expansion about zero of part of the integrand would provide an effective method for calculating the coefficients $\theta_N^{(k)}(0)$ of powers of β in (5.2.1).

In section 4.2 the first terms in expansions of the means and variances of the family of processes were given. The infinite expansions for the moments produced by the different methods would be the same, however formulae (4.2.6) onwards choose a natural way of truncating the expansions which is not equivalent to a truncation to a fixed power of β. For the methods employed here there are two other possible ways of truncating the series. Firstly the expressions for the moments about zero may be truncated to a fixed power of β and then these approximations used in forming central moments. Secondly the expressions for the central moments may themselves be truncated to a fixed power of β. These two procedures lead to different approximations of which the second is better but the first more convenient for computations.

5.2.2 For the characteristic function of the marginal distribution of the stationary process

$$X_{n+1}(\beta) = a + \beta[\lambda\{X_n(\beta)\} - a] + Z_{n+1} \quad (n = \ldots -1,0,1,\ldots), \qquad (5.2.3)$$

which is (4.1.6) with b = 0, the following formulae were derived in section 4.4.5

$$\phi(s;\beta) = e^{isa}\phi_Z(s)h(s;\beta) = \theta_O(s)h(s;\beta) \tag{5.2.4}$$

where $h(s;\beta) = \phi_Y(s;\beta)$ is the characteristic function of $Y_n(\beta) = \beta[\lambda\{X_n(\beta)\} - a]$ and is given by

$$h(s;\beta) = 1 + \sum_{N=1}^{\infty}\sum_{j=1}^{N} \beta^N (is)^j T^{(N,j)},$$

$$= 1 + \sum_{j=1}^{\infty}\sum_{N=j}^{\infty} \beta^N (is)^j T^{(N,j)}. \tag{5.2.5}$$

The quantities $T^{(N,j)}$ are given recursively by

$$T^{(N,j)} = \begin{cases} P_j^{(O)} & (j = N), \\ \sum_{n=1}^{N-j} T^{(N-j,n)} P_j^{(n)} & (1 \leq j < N), \end{cases} \tag{5.2.6}$$

with $P_j^{(n)}$ $(j \geq 1;\ n \geq 0)$ defined by (4.4.17). From these, expressions for the moments of the two processes $\{X_n\}$, $\{\lambda(X_n)\}$ can be found from the following with $\beta = 1$. Let

$$h^{(r)}(s;\beta) = \frac{(-i)^r}{r!}\frac{\partial^r}{\partial s^r} h(s;\beta) \qquad (r \geq O),$$

then $$E[Y_n^r(\beta)] = E(\beta^r[\lambda\{X_n(\beta)\} - a]^r)$$

$$= r!h^{(r)}(O;\beta). \tag{5.2.7}$$

From (5.2.5),

$$h^{(r)}(O;\beta) = \begin{cases} 1 & (r = O), \\ \sum_{N=r}^{\infty} \beta^N T^{(N,r)} & (r \geq 1). \end{cases} \tag{5.2.8}$$

Defining $\theta_O^{(r)}(s) = \frac{(-i)^r}{r!}\frac{\partial^r}{\partial s^r}\theta_O(s)$ $(r \geq O)$ it follows from (5.2.4) that

$$E\{X_n^r(\beta)\} = (-i)^r \frac{\partial^r}{\partial s^r} \phi(s;\beta)\Big|_{s=0} = r!\phi^{(r)}(0;\beta),$$

$$= r! \sum_{\ell=0}^{r} \theta_0^{(r-\ell)}(0) h^{(\ell)}(0;\beta) \qquad (r \geq 0) . \qquad (5.2.9)$$

The means of the processes $\{Y_n(\beta)\}$, $\{X_n(\beta)\}$ are $\mu_Y(\beta) = h^{(1)}(0;\beta)$ and $\mu_X(\beta) = \theta_0^{(1)}(0) + h^{(1)}(0;\beta)$ respectively, and then the central moments are given by

$$E[\{Y_n(\beta) - \mu_Y(\beta)\}^r] = r! \sum_{j=0}^{r} h^{(r-j)}(0;\beta) \frac{\{-\mu_Y(\beta)\}^j}{j!} \qquad (r \geq 2) , \qquad (5.2.10)$$

$$E[\{X_n(\beta) - \mu_X(\beta)\}^r] = r! \sum_{j=0}^{r} \phi^{(r-j)}(0;\beta) \frac{\{-\mu_X(\beta)\}^j}{j!} \qquad (r \geq 2) . \qquad (5.2.11)$$

5.2.3 It is clear that expressions for the joint moments of the $\{X_n(\beta)\}$ processes can be found from the formulae (4.5.1-5) for the joint characteristic functions when the family of autoregression functions is the general one starting with a line of slope b and intercept a. However, since these expressions suffer from the same disadvantages as do the formulae for the marginal moments, they will not be given. Once again the choice b = 0 provides a workable set of formulae. In this case the joint characteristic functions are given by (4.5.12-3), namely,

$$\phi_k(s,t;\beta) = \theta_0(s)\theta_0(t)h_k(s,t;\beta) \qquad (k \geq 1), \qquad (5.2.12)$$

where $h_k(s,t;\beta) = h(s;\beta) + \sum_{N=1}^{\infty} \sum_{j=1}^{N} \beta^N (it)^j T_k^{(N,j)}(s)$.

Define, for $\ell,m \geq 0$,

$$\phi_k^{(\ell,m)}(s,t;\beta) = \frac{(-i)^{\ell+m}}{\ell!m!} \frac{\partial^{\ell+m}}{\partial s^\ell \partial t^m} \phi_k(s,t;\beta) ,$$

and

$$h_k^{(\ell,m)}(s,t;\beta) = \frac{(-i)^{\ell+m}}{\ell!m!} \frac{\partial^{\ell+m}}{\partial s^\ell \partial t^m} h_k(s,t;\beta).$$

Then $$E\{X_n^{\ell}(\beta)X_{n+k}^{m}(\beta)\} = \ell!m!\phi_k^{(\ell,m)}(0,0;\beta)$$

where $$\phi_k^{(\ell,m)}(0,0;\beta) = \sum_{p=0}^{\ell}\sum_{q=0}^{m}\theta_0^{(\ell-p)}(0)\theta_0^{(m-q)}(0)\,h_k^{(p,q)}(0,0;\beta)\ . \qquad (5.2.13)$$

The quantities $h_k^{(\ell,m)}(0,0;\beta)$ $(\ell,m \geq 0)$ are given by

$$h_k^{(\ell,0)}(0,0;\beta) = h^{(\ell)}(0;\beta) \qquad (k \geq 1;\ \ell \geq 0)$$

and for $k,m \geq 1$, $\ell \geq 0$,

$$h_k^{(\ell,m)}(0,0;\beta) = \sum_{N=m}^{\infty}\beta^N T_k^{(N,m)(\ell)}(0)\ , \qquad (5.2.14)$$

where $$T_k^{(N,j)(\ell)}(s) = \frac{(-i)^{\ell}}{\ell!}\frac{\partial^{\ell}}{\partial s^{\ell}}\,T_k^{(N,j)}(s) \qquad (1 \leq j \leq N;\ \ell \geq 0;\ k \geq 1).$$

From (4.5.13) the quantities $T_k^{(N,j)\ (\ell)} = T_k^{(N,j)(\ell)}(0)$ are given by, for $\ell \geq 1$,

$$T_k^{(N,j)(\ell)} = \begin{cases} \begin{cases} Q_j^{(0,\ell)} & (N = j) \\ \sum_{n=1}^{N-j}\sum_{q=0}^{\min(n,\ell)}\binom{n}{q}T^{(N-j,n)}Q_j^{(n-q,\ell-q)} & (N > j) \end{cases} & (k = 1) \\ \begin{cases} 0 & (N = j) \\ P_j^{(0)}T^{(N-j,\ell)} + \sum_{n=1}^{N-j}T_{k-1}^{(N-j,n)(\ell)}P_j^{(n)} & (N > j) \end{cases} & (k \geq 2), \end{cases} \qquad (5.2.15)$$

where the term $T^{(N-j,\ell)}$ is zero if $\ell > N-j$, and where

$$Q_j^{(n,\ell)} = \frac{(-i)^{\ell}}{\ell!}\frac{\partial^{\ell}}{\partial s^{\ell}}\left\{\frac{P_j^{(n)}(0,s)}{\theta_0(s)}\right\}_{s=0}$$

$$= \frac{(-i)^{\ell}}{j!\ell!}\frac{\partial^{n+\ell}}{\partial y^n\partial s^{\ell}}\left\{\frac{\int e^{isx}\Lambda^j(x+y)\,dF_0(x)}{\int e^{isx}dF_0(x)}\right\}_{y,s=0} \qquad (j \geq 1;\ n,\ell \geq 0).$$

with $Q_j^{(n,0)} = P_j^{(n)}$ $(j \geq 1;\ n \geq 0)$. The terms $T_k^{(N,j)(\ell)}$ for $\ell = 0$ are given by

$$T_k^{(N,j)(\ell)} = T^{(N,j)} \qquad (\ell = 0;\ 1 \leq j \leq N;\ k \geq 1) \tag{5.2.16}$$

and hence

$$h_k^{(0,m)}(0,0;\beta) = \sum_{N=m}^{\infty} \beta^N T^{(N,m)} = h^{(m)}(0;\beta) = h_k^{(m,0)}(0,0;\beta) \quad (m,k \geq 1).$$

Assuming that $Q_j^{(n,\ell)}$ can be found, with (5.2.6), (5.2.15) gives a simple recursive procedure for calculating the numbers $T_k^{(N,j)(\ell)}$ $(1 \leq j \leq N;\ k \geq 1;\ \ell \geq 0)$ and hence the moments and joint moments of the processes, by (5.2.13).

Because of the zero value appearing in (5.2.15) it can be shown that

$$T_k^{(N,j)(\ell)} = 0 \qquad (N-j < \ell;\ 0 \leq N-j \leq k-2)$$

and therefore the number of terms appearing in the summation in (5.2.15) can be reduced.

5.2.4 There are no practical difficulties in turning the formulae of sections 5.2.2 and 5.2.3 into computer algorithms. The infinite summations appearing in (5.2.8) and (5.2.14) are power series in β which enter linearly into equations (5.2.7), (5.2.9), (5.2.13) giving the moments and joint moments (about zero) of the processes. Therefore it is convenient to obtain these moments correct to a power, N* say, of β by truncating the series for $h^{(r)}(0;\beta)$ and $h_k^{(\ell,m)}(0,0;\beta)$ to that same power. However, approximations for central moments formed from these approximations for moments about zero are not truncations of the corresponding expansions to any fixed power of β since the approximation for the mean enters non-linearly, see equations (5.2.10), (5.2.11).

Suppose that it is required to form the moments of the process up to order N* in β and that marginal moments up to degree J* and joint moments, $E[X_n^{\ell}(\beta)X_{n+k}^{m}(\beta)]$, for $\ell \leq L^*$ and $m \leq M^*$ are required. For convenience assume that $J^* \geq L^*, M^*$ and that $L^*, M^* \neq 0$: simple modifications cover any exceptions. Then the following procedure is available and is the

basis of the numerical results described in Chapter 6.

(i) Choose a number a.

(ii) Find $\theta_0^{(j)}(0)$ for $0 \leq j \leq J^*$.

(iii) Form $P_j^{(n)}, Q_j^{(n,\ell)}$ for $0 \leq n \leq N^*-j$, $1 \leq j \leq N^*$, $0 \leq \ell \leq L^*$, noting that $Q_j^{(n,0)} = P_j^{(n)}$.

(iv) Form $T^{(N,j)}$ for $1 \leq j \leq N \leq N^*$, in the order

$$T^{(1,1)},$$

$$T^{(2,1)}, T^{(2,2)},$$

$$T^{(3,1)}, T^{(3,2)}, T^{(3,3)},$$

$$T^{(4,1)}, T^{(4,2)}, T^{(4,3)}, T^{(4,4)}, \text{ etc.}$$

In fact $T^{(N,j)}$ need only be calculated for $1 \leq j \leq N \leq N^* - 1$, and for $1 \leq j \leq J^*$ when $N = N^*$ if $J^* < N^*$.

(v) From these form the approximations $h^{(j,0)}$ for $h^{(j)}(0;\beta)$ for $0 \leq j \leq J^*$, where

$$h^{(j,0)} = \begin{cases} 1 & (j = 0), \\ \sum_{N=j}^{N^*} \beta^N T^{(N,j)} & (1 \leq j \leq N^*), \\ 0 & (j > N^*). \end{cases} \qquad (5.2.17)$$

(vi) The approximations for $E[Y_n^r(\beta)]$, $E[X_n^r(\beta)]$ may then be formed by analogy with (5.2.7) and (5.2.9).

The joint moments may be found recursively, the following steps being repeated for $k = 1,2,\ldots$.

(vii) Form $T_k^{(N,j)(\ell)}$ for $1 \leq j \leq N \leq N^*$ and $1 \leq \ell \leq L^*$ using (5.2.15).

(viii) Form the approximations $h^{(\ell,m)}$ for $h_k^{(\ell,m)}(0,0,\beta)$ for $0 \leq \ell \leq L^*$, $0 \leq m \leq M^*$ by

$$h^{(0,m)} = h^{(m,0)} \qquad (m \geq 0),$$

$$h^{(\ell,m)} = \begin{cases} \sum_{N=m}^{N^*} \beta^N T_k^{(N,m)(\ell)} & (\ell \geq 1;\ 1 \leq m \leq N^*) \\ 0 & (\ell \geq 1;\ m > N^*). \end{cases} \qquad (5.2.18)$$

(ix) The approximations for the joint moments may then be formed from (5.2.13).

These steps may be repeated for increasing k as far as is required. It is possible to arrange computer calculations of step (vii) so that the quantities $T_k^{(N,j)}(\ell)$ overwrite those of the previous value of k, and thus no limitation of space is imposed when calculating moments at large lags. Approximations for central moments can be found by formulae (5.2.10-1) and the equivalent formulae for joint moments.

It can be seen from (5.2.17) that the approximation for $E\{Y_n^j(\beta)\}$ is zero whenever $j > N^*$, the first term in the full expansion being of order β^j. This term in the expansion of $E\{Y_n^j(\beta)\} = \beta^j E[\{\lambda(X_n(\beta)) - a\}^j]$ is

$$j!\beta^j T^{(j,j)} = j!\beta^j P_j^{(0)} = \beta^j \int \Lambda^j(x)dF_0(x) = \beta^j E[\{\lambda(a+Z) - a\}^j]. \qquad (5.2.19)$$

From (5.2.18) it can be seen that, similarly, the terms $h^{(\ell',m')}$ contributing to $E[X_n^{\ell}(\beta)X_{n+k}^{m}(\beta)]$ are zero for $m' > N^*$, though this is not true for the index ℓ'. It would always be sensible to choose $N^* \geq J^*, M^*$, and also presumably $N^* \geq L^*$; usually N^* would have to be much larger than the degree of the moments to be calculated.

In an obvious manner the equivalent approximations for the moments to orders lower than N^* in β can be found from (5.2.17), (5.2.18) without recalculating the T-arrays. This provides a check on the convergence of the power-series and a guide to the accuracy of the approximations for the moments. By inserting different values of β in (5.2.17-8) the moments of any of the processes

$$X_{n+1}(\beta) = a + \beta\{\lambda(X_n(\beta)) - a\} + Z_{n+1} \quad (n = \ldots -1,0,1,\ldots)$$

may be obtained rather than just those of the process $\{X_n(1)\}$.

All the above assumes that the power series (5.2.8) and (5.2.14) converge for the particular values of β and a chosen. It has been found in practice that this is not always so, but that the properties of the process with autoregression function $\lambda(x)$ can often be found by making some other choice of the constant a. This constant needs to be chosen with some care; see sections 4.7 and 6.2. A possible procedure is to choose a on the basis of results obtained using several trial values of a to calculate just the marginal moments for small J^* and N^*.

5.2.5 An alternative but equivalent set of expressions for the joint moments may be derived from (4.5.11), i.e.,

$$\phi_k(s,t;\beta) = \theta_0(t) g_k(s,t;\beta)$$

where $g_k(s,t;\beta) = \phi(s;\beta) + \sum_{N=1}^{\infty} \sum_{j=1}^{N} \beta^N (it)^j R_k^{(N,j)}(s)$.

Define $g_k^{(\ell,m)}(s,t;\beta)$ and $R_k^{(N,j)(\ell)}(s)$ analogously to $h_k^{(\ell,m)}(s,t;\beta)$ and $T_k^{(N,j)(\ell)}(s)$; then

$$\phi_k^{(\ell,m)}(0,0;\beta) = \ell! m! \sum_{m'=0}^{m} \theta_0^{(m-m')}(0) g_k^{(\ell,m')}(0,0;\beta)$$

where $g_k^{(\ell,0)}(0,0;\beta) = \phi^{(\ell)}(0;\beta)$ $(\ell \geq 0)$ and, for $m \geq 1$,

$$g_k^{(\ell,m)}(0,0;\beta) = \sum_{N=m}^{\infty} \beta^N R_k^{(N,m)(\ell)}.$$

The quantities $R_k^{(N,m)(\ell)} = R_k^{(N,m)(\ell)}(0)$ are given for $\ell \geq 1$ by

$$R_k^{(N,j)(\ell)} = \begin{cases} \begin{cases} P_j^{(0,\ell)} & (j = N) \\ \sum_{n=1}^{N-j} \sum_{q=0}^{\min(n,\ell)} \binom{n}{q} T^{(N-j,n)} P_j^{(n-q,\ell-q)} & (j < N) \end{cases} & (k = 1), \\ \begin{cases} P_j^{(0)} \theta_0^{(\ell)}(0) & (j = N) \\ P_j^{(0)} \sum_{n=1}^{\min(\ell,N-j)} \theta_0^{(\ell-n)}(0) T^{(N-j,n)} + \sum_{n=1}^{N-j} R_{k-1}^{(N-j,n)(\ell)} P_j^{(n)} & (j < N) \end{cases} & (k \geq 2) \end{cases}$$

and, for $\ell = 0$, $R_k^{(N,j)(\ell)} = T^{(N,j)}$ $(k \geq 1;\ 1 \leq j \leq N)$.

$$\text{Here } P_j^{(n,\ell)} = \frac{(-i)^\ell}{\ell!}\frac{\partial^\ell}{\partial s^\ell} P_j^{(n)}(0,s)\Big|_{s=0}$$

$$= \frac{(-i)^\ell}{\ell!}\frac{\partial^{n+\ell}}{\partial y^n \partial s^\ell}\Big\{P_j(y,s)\Big\}_{y,s=0}$$

$$= \frac{(-i)^\ell}{j!\ell!}\frac{\partial^{n+\ell}}{\partial y^n \partial s^\ell}\Big\{\int e^{isx}\Lambda^j(x+y)\,dF_0(x)\Big\}_{y,s=0} \qquad (j \geq 1;\ n,\ell \geq 0). \tag{5.2.20}$$

Clearly the sets of expressions (5.2.12-16) and the above must be equivalent. The quantities $P_j^{(n,\ell)}, Q_j^{(n,\ell)}$ are related by

$$P_j^{(n,\ell)} = \frac{(-i)^\ell}{\ell!}\frac{\partial^\ell}{\partial s^\ell}\left\{\theta_0(s)\,\frac{P_j^{(n)}(0,s)}{\theta_0(s)}\right\}_{s=0}$$

$$= \sum_{k=0}^{\ell} \theta_0^{(\ell-k)}(0)\,\frac{(-i)^k}{k!}\frac{\partial^k}{\partial s^k}\left\{\frac{P_j^{(n)}(0,s)}{\theta_0(s)}\right\}_{s=0}$$

$$= \sum_{k=0}^{\ell} \theta_0^{(\ell-k)}(0)\,Q_j^{(n,k)} \qquad (j \geq 1;\ n,\ell \geq 0). \tag{5.2.21}$$

The algorithm derived from (5.2.12-16) has the advantage that it leads naturally to an approximation for the joint densities of the process $\{X_n(\beta)\}$, see the next section, and it also has a simpler form. However it requires the quantities $Q_j^{(n,\ell)}$ which have a relatively more complicated expression compared with that for $P_j^{(n,\ell)}$. The expression (5.2.20) can be reduced to

$$P_j^{(n,\ell)} = \frac{1}{j!\ell!}\frac{\partial^n}{\partial y^n}\Big\{\int x^\ell \Lambda^j(x+y)\,dF_0(x)\Big\}_{y=0} \qquad (j \geq 1;\ n,\ell \geq 0).$$

Nonetheless, as will be seen later, when the input distribution is normal the numbers $Q_j^{(n,\ell)}$ have a very simple expression in terms of $P_j^{(n)}$ so that in this case they may be eliminated from the right-hand-side of (5.2.15) giving an expression for $T_k^{(N,j)}(\ell)$ in terms of just $P_j^{(n)}$.

5.3 Expansions for marginal and joint distributions

5.3.1 Methods of obtaining approximations for the stationary distribution functions of the processes $\{X_n(\beta)\}$ will now be considered. This will only be done for the family of processes generated by (5.2.3). Similar expansions could be made, theoretically at least, for the more general family of processes for which the initial process is a linear auto-regression: the first few terms of such expressions have been given in section 4.3.2.

The expansion for the characteristic function of the stationary distribution is

$$\phi(s;\beta) = \theta_0(s)\Big[1 + \sum_{N=1}^{\infty}\sum_{j=1}^{N} \beta^N (is)^j T^{(N,j)}\Big] \qquad (5.3.1)$$

and the problem is to invert this to obtain the corresponding distribution. Suppose that the integrals $\int \phi_Z(s)|s|^p ds$ are finite for all $p \geq 0$. Then the input distribution has a density function $f_Z(z)$ which is continuously differentiable everywhere and the initial density $f_0(z) = f_Z(z-a)$ has the same property: let $f_0^{(r)}(z)$ denote the r'th derivative. Since the input distribution has a density so too does the stationary distribution. Let this density be $f_X(x;\beta)$. Then formally inverting (5.3.1) gives

$$f_X(x;\beta) = f_0(x) + \sum_{N=1}^{\infty}\sum_{j=1}^{N} \beta^N (-1)^j f_0^{(j)}(x) T^{(N,j)}, \qquad (5.3.2)$$

$$= f_0(x) + \sum_{j=1}^{\infty} \{(-1)^j f_0^{(j)}(x) \sum_{N=j}^{\infty} \beta^N T^{(N,j)}\}.$$

Thus the approximation for the density to order N* in β is

$$f_X(x;\beta) \simeq f_0(x) + \sum_{j=1}^{N^*} (-1)^j f_0^{(j)}(x) \sum_{N=j}^{N^*} \beta^N T^{(N,j)},$$

$$= f_0(x) + \sum_{j=1}^{N^*} (-1)^j f_0^{(j)}(x) h^{(j,0)}, \qquad (5.3.3)$$

where $h^{(j,0)}$ is given by (5.2.17). Therefore the algorithm given earlier can easily be extended to calculate approximations, to any required power

of β, for the marginal density of the stationary process. This approximation is not in general positive everywhere.

There does not seem to be a natural interpretation of the above formula when f_Z is not continuously differentiable everywhere. However it may be interpreted as the result of formally inverting the characteristic function

$$\phi_Y(s;\beta) = 1 + \sum_{N=1}^{\infty} \sum_{j=1}^{N} \beta^N (is)^j T^{(N,j)} \tag{5.3.4}$$

to give the density $f_Y(y;\beta)$ a representation as an infinite series composed of derivatives of the delta function.

$$f_Y(y;\beta) = \delta(y) + \sum_{N=1}^{\infty} \sum_{j=1}^{N} \beta^N (-1)^j \delta^{(j)}(y) T^{(N,j)}. \tag{5.3.5}$$

The expression (5.3.2) is then the formal convolution of $f_Z(x-a)$ with this density. This suggests that a better approximation for the stationary distribution may be found by constructing alternative approximations to the distribution having characteristic function (5.3.4) and then forming the convolution of these with the initial distribution. It seems intuitively better to use ordinary functions rather than generalised functions, as in (5.3.5), for the distribution of $Y_n(\beta)$. One possibility is to express (5.3.4) as the product of the characteristic function of a normal distribution with a further series in $(is)^j$. This would lead to an expression for $f_Y(y;\beta)$ as a series of derivatives of normal distributions and it would be natural to truncate this series according to powers of β. A second possibility would be to approximate the distribution by a discrete distribution. This might be done by fitting such a distribution to the approximations for the moments of $Y_n(\beta)$ produced in section 5.2.4. To form a final approximation for the distribution of $Y_n(\beta)$ it would seem best to first form a good approximation for the stationary distribution of $X_n(\beta)$ and then to transform this using $Y_n(\beta) = \beta[\lambda\{X_n(\beta)\} - a]$.

5.3.2 Once a stationary distribution of $\{X_n(\beta)\}$ is found the joint distribution of $(X_n(\beta), X_{n+1}(\beta))$ follows immediately. If the input distribution has a density and an approximation $\hat{f}_X(x;\beta)$ for the stationary density is available then an approximation for the joint density $f_1(x,y;\beta)$ is

$$\hat{f}_1(x,y;\beta) = f_Z[y - \{a+\beta(\lambda(x) - a)\}]\hat{f}_X(x;\beta) \ . \qquad (5.3.6)$$

This is possibly the best way of approximating this distrubiton but it is not easily generalised to pairwise distributions at larger lags.

The method that produced the approximation (5.3.3) for the marginal density can be extended to obtain approximations for the joint densities at any required lag. From (5.2.12)

$$\phi_k(s,t;\beta) = \theta_0(s)\theta_0(t) \sum_{\ell=0}^{\infty} \sum_{m=0}^{\infty} (is)^{\ell}(it)^{m} h_k^{(\ell,m)}(0,0;\beta) \qquad (k \geq 1),$$

and thus, corresponding to (5.3.1), an expression for the joint density $f_k(x,y;\beta)$ of $(X_n(\beta), X_{n+k}(\beta))$ is

$$f_k(x,y;\beta) = \sum_{\ell=0}^{\infty} \sum_{m=0}^{\infty} (-1)^{\ell+m} h_k^{(\ell,m)}(0,0;\beta) f_0^{(\ell)}(x) f_0^{(m)}(y) \ .$$

Truncating this to a fixed power, N*, in β gives

$$\hat{f}_k(x,y;\beta) = \sum_{\ell=0}^{\infty} \sum_{m=0}^{N^*} (-1)^{\ell+m} h^{(\ell,m)} f_0^{(\ell)}(x) f_0^{(m)}(y) \qquad (5.3.7)$$

where $h^{(\ell,m)}$ is given by (5.2.18). There is still an infinite summation for the ℓ-index since $h^{(\ell,m)}$ is non zero for all ℓ (when $1 \leq m \leq N^*$). It is therefore necessary to choose some method of truncating this summation and one way of doing this is to use the approximation

$$\hat{f}_k(x,y;\beta) = \sum_{\ell=0}^{L^*} \sum_{m=0}^{N^*} (-1)^{\ell+m} h^{(\ell,m)} f_0^{(\ell)}(x) f_0^{(m)}(y) \ . \qquad (5.3.8)$$

This uses the quantities $h^{(\ell,m)}$ $(0 \leq \ell \leq L^*;\ 0 \leq m \leq N^*)$ which can be found from the algorithm given in section 5.2.

Given that these approximations for the joint and marginal densities of the stationary process are available, it is then easy to calculate the conditional density of $X_{n+k}(\beta)$ given $X_n(\beta)$, and also the conditional density of $X_n(\beta)$ given $X_{n+k}(\beta)$ (by using the approximations) with the formulae

$$f_k(x,y;\beta)/\{f_X(y;\beta)\}, \quad f_k(x,y;\beta)/\{f_X(y;\beta)\} \qquad (5.3.9)$$

respectively. The forward conditional density is required for predicting $X_{n+k}(\beta)$ from observed values of $X_n(\beta)$, however the backward conditional

densities describing the distribution of past values given the current value do not seem to be of such immediate statistical interest.

An alternative expression for the forward conditional density can be found from the formulae obtained in section 4.6 for the conditional characteristic functions. The conditional characteristic function of $X_{n+m+1}(\beta)$ given that $Y_n(\beta) = y$ is, for $m \geq 0$,

$$e^{isa}\phi_Z(s)\phi_{Y,m}(s;y;\beta) = \theta_0(s)\phi_{Y,m}(s;y;\beta)$$

where $\phi_{Y,m}(s;\beta)$ is given by (4.6.18) and is the conditional characteristic function of $Y_{n+m}(\beta)$ given $Y_n(\beta) = y$. Here $Y_n(\beta) = \beta\Lambda\{X_n(\beta)\}$ and $X_{n+1}(\beta) = Y_n(\beta) + Z_{n+1}$, hence the conditional characteristic function of $X_{n+m+1}(\beta)$, given that $X_n(\beta) = x$, is

$$\theta_0(s)\phi_{Y,m}(s;\beta\Lambda(x);\beta) = \theta_0(s)\phi_{Y,m}(s;\beta(\lambda(x) - a);\beta).$$

It is possible to formally invert these expressions to obtain expansions for the conditional densities. For $m = 0$ this involves inverting the characteristic function $\theta_0(s)\exp\{is\beta(\lambda(x) - a)\}$ which gives the conditional density

$$f_1(y|x) = f_Z(y - a - \beta\{\lambda(x) - a\}). \tag{5.3.10}$$

For the cases $m \geq 1$, $\phi_{Y,m}(s;y;\beta)$ is given by (4.6.18) in terms of powers of (is), and therefore the conditional densities can be obtained in terms of derivatives of f_0 or f_Z. Expanding the term $P_{r_m}\{\beta\Lambda(x)\}$ in (4.6.18) in powers of β and rearranging gives, for $m \geq 1$,

$$f_{m+1}(y|x) = f_0(y) + \sum_{j=1}^{\infty}\sum_{N=j}^{j+m-1} \beta^N T(N,j)(-1)^j f_0^{(j)}(y)$$

$$+ \sum_{N=m+1}^{\infty} \sum_{\substack{r_1+\ldots+r_{m+1}=N \\ r_i \geq 1}} \beta^N (-1)^{r_1} f_0^{(r_1)}(y) P_{r_1}^{(r_2)} \ldots P_{r_{m-1}}^{(r_m)} P_{r_m}^{(r_{m+1})} \frac{\{\Lambda(x)\}^{r_{m+1}}}{r_{m+1}!} \tag{5.3.11}$$

Alternatively it may be better to leave this in its original form

$$f_{m+1}(y|x) = f_0(y) + \sum_{j=1}^{\infty} \sum_{N=j}^{j+m-2} \beta^N T(N,j) (-1)^j f_0^{(j)}(y)$$

$$+ \sum_{N=m}^{\infty} \sum_{\substack{r_1+\ldots+r_m=N \\ r_i \geq 1}} \beta^N (-1)^{r_1} f_0^{(r_1)}(y) P_{r_1}^{(r_2)} \ldots P_{r_{m-1}}^{(r_m)} P_{r_m} \{\beta\Lambda(x)\}. \tag{5.3.12}$$

From these conditional densities alternative expressions for the joint densities can be found using (5.3.2) for the marginal density. Using (5.3.10) produces a better approximation for the joint density at lag 1 than (5.3.8) possibly because there is no further error due to truncation once the marginal density is found. However, for lags greater than 1, using (5.3.11), (5.3.12) would necessitate a further truncation: in this sense (5.3.12) might well give better results than (5.3.11). An advantage of these forms is that truncating to a fixed power of β leads immediately to finite summations, unlike (5.3.7).

5.3.3 Consider expressions (5.3.2) onwards in the special case when the initial density is normal with zero mean and unit variance. Then the expansion of the density is of the form

$$f_X(x;\beta) = e^{-\frac{1}{2}x^2} \sum_{j=0}^{\infty} H_j(x) A_j(\beta) \tag{5.3.13}$$

where $H_j(x)$, $j \geq 0$, are Hermite polynomials. Such series are considered by Kendall and Stuart (1969; pp.155-63). It is known that, when $f_X(x;\beta)$ is a normal distribution with zero mean and variance greater than two, the above series does not converge. Such a case arises when applying the method of section 5.3.1 to a linear autoregression with slope greater than $1/\sqrt{2}$. Nonetheless applying (5.3.3) for increasing N* does converge.

Let the autoregression function be $\lambda(x) = \lambda x$, λ constant with $|\lambda| < 1$, and let the input distribution be N(0,1). Then, with the choice a = 0, the stationary density of the process $\{X_n(\beta)\}$, generated by $X_{n+1}(\beta) = \beta\lambda X_n(\beta) + Z_{n+1}$, is

$$f_X(x;\beta) = \left(\frac{1-\beta^2\lambda^2}{2\pi}\right)^{1/2} \exp\{-\tfrac{1}{2}(1-\beta^2\lambda^2)x^2\}, \tag{5.3.14}$$

and $\phi(s;\beta)$ is the characteristic function corresponding to this. The terms $A_j(\beta)$ in (5.3.13) may be identified from

$$\phi_Y(s;\beta) = \frac{\phi(s;\beta)}{\theta_O(s)} = \exp\left\{\frac{-\beta^2\lambda^2 s^2}{2(1-\beta^2\lambda^2)}\right\} = \sum_{j=0}^{\infty} (is)^j A_j(\beta),$$

giving

$$A_{2j}(\beta) = \frac{1}{2!}\left\{\frac{\beta^2\lambda^2}{2(1-\beta^2\lambda^2)}\right\}^j, \quad A_{2j+1}(\beta) = 0 \qquad (j = 0,1,2,\ldots.),$$

and thus (5.3.12) is

$$f_X(x;\beta) = e^{-\frac{1}{2}x^2} \sum_{j=0}^{\infty} H_{2j}(x)\, \frac{1}{j!}\left\{\frac{\beta^2\lambda^2}{2(1-\beta^2\lambda^2)}\right\}^j. \qquad (5.3.15)$$

It is possible to show that (5.3.15) converges to (5.3.14) if $|\beta^2\lambda^2/(1-\beta^2\lambda^2)| < 1$, or equivalently if $\beta^2\lambda^2 < \frac{1}{2}$, the same result as that mentioned above.

However, in the form that is used (i.e. 5.3.3) the series (5.3.15) is truncated to a fixed power of β. Hence the sequence of values actually found for the approximation to the density is the sequence of truncations of (5.3.14) expressed as a power series in β. This sequence converges for $|\beta\lambda| < 1$.

5.4 Basic Quantities

5.4.1 To use either of the algorithms of section 5.2 one of the sets of numbers $Q_j^{(n,\ell)}, P_j^{(n,\ell)}$ must be found. The quantities $Q_j^{(n,\ell)}$ required by the first algorithm are given by

$$Q_j^{(n,\ell)} = \frac{(-i)^\ell}{j!\ell!}\frac{\partial^{n+\ell}}{\partial y^n \partial s^\ell}\left\{\frac{\int e^{isx}\Lambda^j(x+y)\,dF_O(x)}{\int e^{isx}dF_O(x)}\right\}_{y,s=0} \qquad (j \geq 1;\ n,\ell \geq 0), \qquad (5.4.1)$$

while those required by the second form are given by a relatively simple expression

$$P_j^{(n,\ell)} = \frac{1}{j!\ell!}\frac{\partial^n}{\partial y^n}\left\{\int x^\ell \Lambda^j(x+y)\,dF_O(x)\right\}_{y=0} \qquad (j \geq 1;\ n,\ell \geq 0). \qquad (5.4.2)$$

In both cases $\Lambda(x) = \lambda(x) - a$ and $F_0(x) = F_Z(x-a)$. The other numbers required, $P_j^{(n)} = Q_j^{(n,0)} = P_j^{(n,0)}$, are included in both the above sets. The two sets are related by the equality (5.2.21), and from this or directly from (5.4.1)

$$Q_j^{(n,0)} = P_j^{(n)} = \frac{1}{j!}\frac{\partial^n}{\partial y^n}\{\int \Lambda^j(x+y)dF_0(x)\}_{y=0}$$

$$Q_j^{(n,1)} = \frac{1}{j!}\frac{\partial^n}{\partial y^n}\{\int (x-\mu_0)\Lambda^j(x+y)dF_0(x)\}_{y=0},$$

$$Q_j^{(n,2)} = \frac{1}{j!2}\frac{\partial^n}{\partial y^n}\{\int[(x-\mu_0)^2 - \sigma_0^2]\Lambda^j(x+y)dF_0(x)\}_{y=0},$$

$$Q_j^{(n,3)} = \frac{1}{j!3!}\frac{\partial^n}{\partial y^n}\{\int [(x-\mu_0)^3 - 3\sigma_0^2(x-\mu_0) - \mu_{3,0}]\Lambda^j(x+y)dF_0(x)\}_{y=0},$$

where μ_0, σ_0^2, $\mu_{3,0}$ are the mean, variance and third central moment of the initial distribution F_0.

Certain useful relations can be found by writing $Q_j^{(n,\ell)}$ of (5.4.1) as explicit functionals of $\Lambda(x)$,

$$Q_j^{(n,\ell)}[\Lambda(x)] = Q_j^{(n,\ell)}.$$

For any constant c and functions $\Lambda_1(x)$, $\Lambda_2(x)$,

$$Q_j^{(n,\ell)}[c\Lambda_1(x)] = c^j Q_j^{(n,\ell)}[\Lambda_1(x)], \tag{5.4.3}$$

$$Q_j^{(n,\ell)}[\Lambda_1(x) + c] = \sum_{p=0}^{j} \frac{c^p}{p!} Q_{j-p}^{(n,\ell)}[\Lambda_1(x)], \tag{5.4.4}$$

$$Q_j^{(n,\ell)}[\Lambda_1(x) + \Lambda_2(x)] = \sum_{p=0}^{j} Q_1^{(n,\ell)}\left[\frac{\Lambda_1^p(x)}{p!} \cdot \frac{\Lambda_2^{j-p}(x)}{(j-p)!}\right]. \tag{5.4.5}$$

Further, if $\Lambda_1(x)\Lambda_2(x) = 0$ everywhere, so that at least one function is zero at each point,

$$Q_j^{(n,\ell)}[\Lambda_1(x) + \Lambda_2(x)] = Q_j^{(n,\ell)}[\Lambda_1(x)] + Q_j^{(n,\ell)}[\Lambda_2(x)]. \tag{5.4.6}$$

In particular a useful relation, since $\Lambda(x) = \lambda(x) - a$, is

$$Q_j^{(n,\ell)}[\lambda(x)-a] = \sum_{p=0}^{j} \frac{(-a)^p}{p!} Q_{j-p}^{(n,\ell)}[\lambda(x)], \tag{5.4.7}$$

however it should be noted that the functional $Q_j^{(n,\ell)}[\ ]$ itself depends upon a. Exactly similar expressions hold for the functionals $P_j^{(n,\ell)}[\Lambda(x)] = P_j^{(n,\ell)}$ of (5.4.2).

Some relations for simple functions are

$$Q_j^{(n,\ell)}[c] = \begin{cases} \dfrac{c^j}{j!} & (n = 0,\ \ell = 0), \\ 0 & (\text{otherwise}), \end{cases}$$

$$Q_j^{(n,\ell)}[x] = \begin{cases} Q_{j-n}^{(0,\ell)}[x] & (j \geq n), \\ 0 & (n > j), \end{cases}$$

$$P_j^{(n,\ell)}[x] = P_j^{(0,\ell)}[x] = \frac{1}{(j-n)!\ell!} \int x^{\ell+j-n} dF_0(x)$$

$$= \binom{\ell+j-n}{\ell} P_{\ell+j-n}^{(0,0)}[x] \qquad (\ell + j - n \geq 0).$$

In the above formulae the definitions (5.4.1), (5.4.2) have been extended to include values of the index $j = 0$ in the natural way

$$Q_0^{(n,\ell)} = \begin{cases} 1 & (n = \ell = 0), \\ 0 & (\text{otherwise}). \end{cases} \qquad P_0^{(n,\ell)} = \begin{cases} \theta_0^{(\ell)}(0) & (n = 0), \\ 0 & (\text{otherwise}). \end{cases}$$

5.4.2 Some useful results about the quantities $Q_j^{(n,\ell)}$, $T^{(N,j)}$ for special classes of autoregressive processes can be established. Suppose that the autoregression function $\lambda(x)$ is odd about a point a, that is

$$\lambda(x+a) - a = -\{\lambda(-x+a) - a\} \qquad (-\infty < x < \infty),$$

and suppose further that the input distribution is symmetric about zero. When these conditions hold this number a is a natural choice for the number a required by the algorithm of section 5.2. In this case it is easily seen from (5.4.1) and (5.4.2) that

$$Q_j^{(n,\ell)} = (-1)^{n+\ell+j} Q_j^{(n,\ell)} \qquad (j \geq 1;\ n,\ell \geq 0), \tag{5.4.8}$$

$$\text{and} \quad P_j^{(n,\ell)} = (-1)^{n+\ell+j} P_j^{(n,\ell)} \qquad (j \geq 1;\ n,\ell \geq 0),$$

These imply that

$$Q_j^{(n,\ell)} = 0 \qquad \text{when } n + \ell + j \text{ is odd}$$

and in particular that

$$P_j^{(n)} = 0 \qquad \text{when } n + j \text{ is odd.}$$

Then from (5.2.6) and (5.2.15) it can be shown that

(i) $T^{(N,j)} = 0$ unless N and j are both even (5.4.9)

(ii) $T_k^{(N,j)(\ell)} = 0$ unless (a) k is odd, N is even and j,ℓ are both odd or both even

or (b) k is even and N,j,ℓ are all even or all odd.

It is clear that, when the autoregression function is odd about some point, making use of these results can greatly reduce the computational effort required by the methods of section 5.2. Similar identities may be demonstrated for the quantities $R_k^{(N,j)(\ell)}$ involved in the expressions of section 5.2.5.

From (5.4.9) and expression (5.2.17) it can be seen that the approximations for odd degree moments of the marginal distribution are all zero and that the approximations to order $(2N^* + 1)$ in β for the even degree central moments are the same as those to order $2N^*$. It may also be concluded that the quantities $h^{(\ell,m)}$ of (5.2.18) for any k satisfy

$$h^{(\ell,m)} = 0 \quad \text{unless} \quad \ell + m \text{ is even.}$$

Hence the approximations for $E[X_n^{\ell}(\beta)X_{n+k}^{m}(\beta)]$ are all zero unless ℓ and m are both even or both odd. Again the sequence of approximations for the other joint moments to increasing orders of β are equal in pairs.

It is possible to make use of special relations among the numbers $Q_j^{(n,\ell)}$ in another special case, this time when the autoregression function is even about some point; that is

$$\lambda(x+a) - a = \lambda(-x+a) - a \qquad (-\infty < x < \infty)$$

for some value a. Here this value is not in general a natural choice for the arbitrary value a required by the algorithm, however it may be possible to use it. If the input distribution is again symmetric about zero then, corresponding to (5.4.8), the following holds

$$Q_j^{(n,\ell)} = (-1)^{n+\ell} Q_j^{(n,\ell)} \qquad (j \geq 1;\ n,\ell \geq 0)$$

so that

$$Q_j^{(n,\ell)} = 0 \text{ for } n + \ell \text{ odd}$$

and in particular $P_j^{(n)} = 0$ for n odd. Then it can be shown that

$$T^{(N,j)} = 0 \qquad \text{unless } N - j \text{ is even}$$

$$T_k^{(N,j)(\ell)} = 0 \qquad \text{unless } N - j + \ell \text{ is even } (k \geq 1).$$

It may then be concluded that the sequences of approximations to the moments about zero are equal in pairs, but that the sequences for the central moments are not (if (5.2.10-11) are used).

5.5 Normal input distribution

5.5.1 A class of autoregressive processes which might be of particular interest is that for which the input distribution is normal. Reasons for this interest would be mainly those which lead to the importance of the normal distribution throughout statistics; thus each value of the input sequence might be considered the result of a large number of small and independent effects. Considering the simplicity of the distributional properties of linear autoregression processes with normal input distributions it might be thought that the properties of non-linear processes might be relatively more simple to find than those of processes whose input distributions are non-normal. Some simplification does occur but the main effect of this is to simplify slightly the expressions leading to the joint moments. It is possible that for other particular input distributions similar simplifications might be found.

Suppose that the input distribution is normal. Then the initial distribution is also normal with mean μ_0, say, and variance equal to that of the input distribution, σ^2 say. With some rearrangement, for any autoregression function $\lambda(x)$ $(\Lambda(x) = \lambda(x) - a)$,

$$\frac{\int e^{isx}\Lambda^j(x+y)\,dF_0(x)}{\int e^{isx}\,dF_0(x)} = \frac{1}{\sqrt{2\pi}\sigma}\int \exp\{-i\mu_0 s + \frac{\sigma^2 s^2}{2} + is(z-y) - \frac{1}{2\sigma^2}(z-y-\mu_0)^2\}\Lambda^j(z)\,dz,$$

$$= \frac{1}{\sqrt{2\pi}\sigma}\int \exp\{-\frac{1}{2\sigma^2}(z-y-\mu_0-is\sigma^2)^2\}\Lambda^j(z)\,dz,$$

and hence, for $n,\ell \geq 0$ and $j \geq 1$,

$$Q_j^{(n,\ell)} = \frac{(-i)^\ell}{j!\ell!}\frac{\partial^{n+\ell}}{\partial y^n \partial s^\ell}\left\{\frac{\int e^{isx}\Lambda^j(x+y)\,dF_0(x)}{\int e^{isx}\,dF_0(x)}\right\}_{y,s=0}$$

$$= \frac{\sigma^{2\ell}}{\ell!}P_j^{(n+\ell)}, \qquad (5.5.1)$$

where, as before,

$$P_j^{(n)} = \frac{1}{j!}\frac{\partial^n}{\partial y^n}\{\int \Lambda^j(x+y)\,dF_0(x)\}_{y=0}.$$

In this case these quantities are

$$P_j^{(n)} = \frac{1}{\sqrt{2\pi q}} \int \frac{\Lambda^j(z)}{j!} \frac{1}{\sigma^n} H_n\left[\frac{z-\mu_O}{\sigma}\right] \exp\{-\frac{1}{2\sigma^2}(z-\mu_O)^2\} dz. \qquad (5.5.2)$$

The result (5.5.1) may be used to simplify the algorithm given earlier for this special case. For then, instead of calculating $Q_j^{(n,\ell)}$ for $0 \leq n \leq N^* - j$, $1 \leq j \leq N^*$ and $0 \leq \ell \leq L^*$, it is sufficient to calculate $P_j^{(n)}$ for $0 \leq n \leq N^* + L^* - j$, $1 \leq j \leq N^*$ and to replace $Q_j^{(n,\ell)}$ when it occurs by the expression (5.5.1) giving in particular as part of (5.2.15)

$$T_1^{(N,j)(\ell)} = \begin{cases} \dfrac{\sigma^{2\ell}}{\ell!} P_j^{(\ell)} & (j = N), \\ \displaystyle\sum_{n=1}^{N-j} \sum_{q=0}^{\min(n,\ell)} \binom{n}{q} T^{(N-j,n)} \frac{\sigma^{2(\ell-q)}}{(\ell-q)!} P_j^{(n+\ell-2q)} & (j < N). \end{cases}$$

This appears to be the only structural simplification that occurs. However the other quantities required by the algorithm, namely $\theta^{(j)} = \theta_O^{(j)}(O)$ $(j = 0,1,\ldots)$, can be found by simple formulae in this special case. If the input distribution has mean μ_Z and variance σ^2, then the initial distribution has mean $\mu_O = a + \mu_Z$ and

$$\theta_O(s) = \exp\{i\mu_O s - \frac{\sigma^2 s^2}{2}\} = \exp\{-\frac{1}{2}\sigma^2(s - \frac{i\mu_O}{\sigma^2})^2 + \frac{\mu_O^2}{2\sigma^2}\}.$$

Then

$$\theta^{(n)} = \theta_O^{(n)}(O) = \frac{(-i)^n}{n!} \frac{d^n}{ds^n} \theta_O(s)\Big|_{s=O} = \frac{(i\sigma)^n}{n!} H_n(-\frac{i\mu_O}{\sigma}) \qquad (n = 0,1,2,\ldots).$$

The Hermite Polynomials satisfy the recurrence formula

$$H_n(x) = xH_{n-1}(x) - (n-1)H_{n-2}(x) \qquad (n \geq 2) \qquad (5.5.3)$$

with $H_O(x) = 1$, $H_1(x) = x$. Using this a recurrence formula for $\theta^{(n)}$ is

$$\theta^{(n)} = \frac{1}{n}\{\mu_O\theta^{(n-1)} + \sigma^2\theta^{(n-2)}\} \qquad (n \geq 2)$$

with the initial values given by $\theta^{(O)} = 1$, $\theta^{(1)} = \mu_O$. It is also easy to calculate the initial density and its derivatives required by the procedures of section 5.3 for finding the stationary marginal and joint densities.

5.5.2 For specific autoregression functions it may be possible to obtain explicit expressions for $P_j^{(n)}$, from (5.5.2) for example. However it is often more appropriate to look for simple recurrence relations for calculating these numbers. The relations (5.4.3-6) are useful for computing the functionals $P_j^{(n)}$ for particular autoregression functions from those corresponding to simpler forms. Using the functional notation of section 5.4.1, the quantities $P_j^{(n)}$ corresponding to an autoregression function $\lambda^*(x) + c$, with the choice of the arbitrary constant a, can be found from

$$P_j^{(n)}[\lambda^*(x) + c - a] = \sum_{p=0}^{j} \frac{(c-a)^p}{p!} P_{j-p}^{(n)}[\lambda^*(x)] . \qquad (5.5.4)$$

In the following examples the input distribution is normal with zero mean and variance σ^2, so, for the choice of constant a, the initial distribution is $N(a,\sigma^2)$. For the autoregression function $\lambda(x)$, the quantities required are $P_j^{(n)}[\lambda(x) - a]$ $(j \geq 1;\ n \geq 0)$ which are given by (5.5.4) and

$$P_j^{(n)}[\lambda(x)] = \frac{1}{\sqrt{2\pi}\sigma} \frac{d^n}{dy^n} \{ \int \frac{\lambda^j(z)}{j!} \exp\{-\frac{1}{2\sigma^2}(z - a - y)^2\}dz \}_{y=0} \qquad (5.5.5)$$

It is sometimes also convenient to consider the functionals $P_j[\lambda(x)](y)$, where here the notations of earlier sections are combined,

$$P_j[\lambda(x)](y) = \frac{1}{\sqrt{2\pi}\sigma} \int \frac{\lambda^j(z)}{j!} \exp\{-\frac{1}{2\sigma^2}(z - a - y)^2\}dz.$$

The initial density is

$$f_0(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{1}{2\sigma^2}(x - a)^2\}$$

and the initial distribution function, $F_0(x)$, is given by $F_0(x) = \Phi(\frac{x-a}{\sigma})$ where $\Phi(\cdot)$ is the standard normal distribution function. In the following it is often necessary to calculate (numerically) derivatives of F_0 at specified points. This can be done in a simple manner by using recurrence relations among the derivatives. Let

$$F_0^{(n)}(x) = \frac{d^n}{dx^n} F_0(x) \qquad (n \geq 0),$$

then, using the recurrence relation (5.5.3) for the Hermite polynomials, the following generating formula holds

$$F_0^{(n)}(x) = (-\frac{1}{\sigma^2})\{(x-a)F^{(n-1)}(x) + (n-2)F^{(n-2)}(x)\} \quad (n \geq 3) \qquad (5.5.6)$$

with $F_0^{(1)}(x) = f_0(x)$, $F_0^{(2)}(x) = -\frac{(x-a)}{\sigma^2}f_0(x)$. It is sometimes more convenient to work with $G_0^{(n)}(x)$ defined by

$$G_0^{(n)}(x) = (-1)^n \frac{d^n}{dx^n} F_0(x) = (-1)^n F_0^{(n)}(x)$$

or, equivalently, $G_0^{(n)}(-x) = \frac{d^n}{dx^n} F_0(-x)$. The recurrence relation for the $G_0^{(n)}(x)$ is

$$G_0^{(n)}(x) = \frac{1}{\sigma^2}\{(x-a)G_0^{(n-1)}(x) - (n-2)G_0^{(n-2)}(x)\} \quad (n \geq 3) \qquad (5.5.7)$$

with $G_0^{(1)}(x) = -f_0(x)$, $G_0^{(2)}(x) = -\frac{x-a}{\sigma^2} f_0(x)$.

Example 5.1 Suppose $\lambda(x)$ is of the form

$$\lambda(x) = \begin{cases} \mu & (x_1 < x < x_2), \\ 0 & (\text{otherwise}), \end{cases}$$

where $[x_1, x_2]$ may be a finite or semi-infinite interval. In this case, for $j \geq 1$,

$$P_j[\lambda(x)](y) = P_j(y) = \frac{1}{\sqrt{2\pi}\sigma} \int_{x_1}^{x_2} \frac{\mu^j}{j!} \exp\{-\frac{1}{2\sigma^2}(z-a-y)^2\}dz,$$

$$= \frac{\mu^j}{j!}\{F_0(x_2-y) - F_0(x_1-y)\}.$$

Then $P_j^{(n)}[\lambda(x)] = P_j^{(n)} = \frac{d^n}{dy^n}\{P_j(y)\}_{y=0}$, and using the above definitions

$$P_j^{(n)} = \frac{\mu^j}{j!}\{G_0^{(n)}(x_2) - G_0^{(n)}(x_1)\} \qquad (j \geq 1;\ n \geq 0). \qquad (5.5.8)$$

The recurrence relation (5.5.7) provides a simple method of calculating these quantities. Note that these depend on a, since for example $G_0^{(0)}(x) = F_0(x) = \Phi(\frac{x-a}{\sigma})$. To form $P_j^{(n)}[\lambda(x)-a]$ similar procedures to the above may be applied directly to the function $\lambda(x) - a$, otherwise (5.5.4) may be employed, noting that

$$P_0^{(n)}[\cdot] = \begin{cases} 1 & (n = 0), \\ 0 & (\text{otherwise}) \end{cases} \tag{5.5.9}$$

This is not (5.5.8) with j = 0. The following are found.

$$P_j^{(0)}[\lambda(x)+c-a] = \frac{(\mu+c-a)^j-(c-a)^j}{j!}\{G_0(x_2) - G_0(x_1)\} + \frac{(c-a)^j}{j!} \qquad (j \geq 1)$$

and, for $n \geq 1$,

$$P_j^{(n)}[\lambda(x)+c-a] = \frac{(\mu+c-a)^j-(c-a)^j}{j!}\{G_0^{(n)}(x_2) - G_0^{(n)}(x_1)\} \qquad (j \geq 1) .$$

These may be checked directly from (5.5.5) for the function $\lambda(x) + c - a$ which is

$$\lambda(x) + c - a = \begin{cases} \mu+c-a & (x_1 < x < x_2), \\ c - a & (\text{otherwise}) . \end{cases}$$

Example 5.2 From the above formulae the quantities $P_j^{(n)}$ corresponding to any step-function may be found. Let $-\infty = x_0 < x_1 < x_2 \ldots < x_{K-1} < x_K = \infty$ be a partition of the real line and let the function $\lambda(x)$ be of the form

$$\lambda(x) = \mu_k \text{ for } x \in (x_{k-1}, x_k),$$

where μ_k $(k = 1,\ldots K)$ are K specified constants. Then using the obvious extension of (5.4.6) and the result (5.5.8),

$$P_j^{(n)}[\lambda(x)] = \sum_{k=1}^{K} \frac{\mu_k^j}{j!}\{G_0^{(n)}(x_k) - G_0^{(n)}(x_{k-1})\} \qquad (n \geq 0;\ j \geq 1).$$

Alternatively, using the special values of x_0 and x_K,

$$P_j^{(0)}[\lambda(x)] = \sum_{k=1}^{K-1} \frac{\mu_k^j - \mu_{k+1}^j}{j!} G_0(x_k) + \frac{\mu_K^j}{j!} \qquad (j \geq 1),$$

$$P_j^{(n)}[\lambda(x)] = \sum_{k=1}^{K-1} \frac{\mu_k^j - \mu_{k+1}^j}{j!} G_0(x_k) \,. \qquad (n \geq 1;\ j \geq 1).$$

The quantities $P_j^{(n)}[\lambda(x) - a]$ actually required by the algorithm can be seen to be equivalent to the above with μ_k replaced by $\mu_k - a$, $k = 1,\ldots K$.

Example 5.3 Let $\lambda(x)$ be the function

$$\lambda(x) = \begin{cases} \lambda x + \mu & (x_1 < x < x_2), \\ 0 & (\text{Otherwise}), \end{cases}$$

where λ, μ are constants. Then, for $j \geq 1$,

$$P_j[\lambda(x)](y) = \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{j!} \int_{x_1}^{x_2} (\lambda z+\mu)^j \exp\{-\frac{1}{2\sigma^2}(z-a-y)^2\}dz = R_j(y) \,. \qquad (5.5.10)$$

This has been denoted $R_j(y)$ rather than $P_j(y)$ since a recurrence relation for generating $P_j^{(n)}[\lambda(x)]$ $(j \geq 1;\ n \geq 0)$ can be formed by defining

$$R_0(y) = \frac{1}{\sqrt{2\pi}\sigma} \int_{x_1}^{x_2} \exp\{-\frac{1}{2\sigma^2}(z-a-y)^2\}dz$$

$$= G_0(x_2-y) - G_0(x_1-y)$$

whereas, to accord with (5.5.9), $P_0[\lambda(x)](y) = 1$.

On integrating by parts, (5.5.10) gives

$$R_1(y) = (\mu + \lambda a + \lambda y)R_0(y) + \lambda\sigma^2\{G_0^{(1)}(x_2-y) - G_0^{(1)}(x_1-y)\}$$

$$R_j(y) = \frac{1}{j}\{(\mu + \lambda a + \lambda y)R_{j-1}(y) + \lambda^2\sigma^2 R_{j-2}(y)\}$$

$$+ \frac{\lambda\sigma^2}{j!}\{(\lambda x_2+\mu)^{j-1}G_0^{(1)}(x_2-y) - (\lambda x_1+\mu)^{j-1}G_0^{(1)}(x_1-y)\} \qquad (j \geq 2).$$

Let $R_j^{(n)} = \dfrac{d^n}{dy^n}\{R_j(y)\}_{y=0}$ $\qquad (n,j \geq 0)$,

$= P_j^{(n)}[\lambda(x)]$ $\qquad (j \geq 1;\ n \geq 0)$.

Then, from the above, for $n \geq 0$,

$$R_0^{(n)} = G_0^{(n)}(x_2) - G_0^{(n)}(x_1)$$

$$R_1^{(n)} = (\mu+\lambda a)R_0^{(n)} + \lambda n R_0^{(n-1)} + \lambda\sigma^2\{G_0^{(n+1)}(x_2) - G_0^{(n+1)}(x_1)\}$$

$$R_j^{(n)} = \frac{1}{j}\{(\mu+\lambda a)R_{j-1}^{(n)} + \lambda n R_{j-1}^{(n-1)} + \lambda^2\sigma^2 R_{j-2}^{(n)}\}$$

$$+ \frac{\lambda\sigma^2}{j!}\{(\lambda x_2+\mu)^{j-1}G_0^{(n+1)}(x_2) - (\lambda x_1+\mu)^{j-1}G_0^{(n+1)}(x_1)\} \quad (j \geq 2).$$

Here the terms containing a factor with superscript (n-1) are zero for $n = 0$. These provide recurrence relations for calculating the numbers $P_j^{(n)}[\lambda(x)]$ and then (5.5.4) may be used to calculate $P_j^{(n)}[\lambda(x)-a]$. This can also be used to calculate the appropriate numbers $P_j^{(n)}$ for any function of the form

$$\lambda(x) = \begin{cases} \lambda x + \mu + c & (x_1 < x < x_2), \\ c & (\text{otherwise}). \end{cases}$$

Example 5.4 If it is required to compute the numbers $P_j^{(n)}$ for a function which is sectionally linear this may be done by splitting $\lambda(x) - a$ into a number of functions, each of which is linear on an interval and zero elsewhere. Suppose $\lambda(x)$ is the function

$$\lambda(x) = \lambda_k x + \mu_k \qquad \text{for } x \in (x_{k-1}, x_k),$$

where $-\infty = x_0 < x_1 < x_2 \ldots < x_{K-1} < x_K = \infty$. Write $\lambda(x) - a = \sum_{k=1}^{K} \lambda_k^*(x)$,

where, for $k = 1,\ldots K$,

$$\lambda_k^*(x) = \begin{cases} \lambda_k x + \mu_k - a & (x_{k-1} < x < x_k), \\ 0 & (\text{otherwise}). \end{cases}$$

Then

$$P_j^{(n)}[\lambda(x) - a] = P_j^{(n)}[\sum_{k=1}^{K} \lambda_k^*(x)]$$

$$= \sum_{k=1}^{K} P_j^{(n)}[\lambda_k^*(x)]$$

and the numbers $P_j^{(n)}[\lambda_k^*(x)]$ can be found from the recurrence relations in example 5.3 with (x_1, x_2) replaced by (x_{k-1}, x_k) and with (λ, μ) replaced by $(\lambda_k, \mu_k - a)$.

Example 5.5 Let the autoregression function be a linear function, i.e.

$$\lambda(x) = \lambda x + \mu \qquad (-\infty < x < \infty;\ \lambda \neq 0).$$

Then, by using suitable generating functions, it may be shown that

$$P_j^{(n)}[\lambda x + \mu - a] = \frac{\lambda^j (-i\sigma)^{j-n}}{(j-n)!} H_{j-n}\left(\frac{i(\lambda a + \mu - a)}{\lambda\sigma}\right) \qquad (j \geq n),$$

$$0 \qquad (n > j).$$

Making use of the recurrence relation for Hermite Polynomials it may be shown that

$$P_j^{(n)} = \lambda^n S_{j-n} \qquad (n \leq j)$$

where

$$S_n = \{(\lambda a + \mu - a)\, S_{n-1} + \lambda^2\sigma^2 S_{n-2}\}/n \qquad (n \geq 2)$$

and $S_0 = 1$, $S_1 = \lambda a + \mu - a$.

Example 5.6 Let $\lambda(x)$ be given by

$$\lambda(x) = \lambda \exp\{-\tfrac{1}{2} w^2 (x-d)^2\} \qquad (-\infty < x < \infty)$$

where $w^2 > 0$. Then, for $j \geq 1$,

$$P_j[\lambda(x)](y) = \frac{\lambda^j}{j!}\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}\exp\{-\frac{1}{2}jw^2(z-d)^2 - \frac{1}{2\sigma^2}(z-a-y)^2\}dz,$$

$$= \frac{\lambda^j}{j!}\frac{1}{(jw^2\sigma^2+1)^{1/2}}\exp\left\{-\frac{jw^2(y+a-d)^2}{2(jw^2\sigma^2+1)}\right\},$$

and hence, for $j \geq 1$ and $n \geq 0$,

$$P_j^{(n)}[\lambda(x)] = \frac{\lambda^j}{j!}\frac{(jw^2)^{n/2}}{(jw^2\sigma^2+1)^{n+1/2}}H_n\left[\left(\frac{jw^2}{jw^2\sigma^2+1}\right)^{1/2}(a-d)\right]\exp\left\{\frac{-jw^2(a-d)^2}{2(jw^2\sigma^2+1)}\right\}.$$

Using (5.5.3) simple recurrence relations for these quantities can be found and then $P_j^{(n)}[\lambda(x)-a]$ can be found using (5.5.4).

Example 5.7 In order to calculate the quantities $P_j^{(n)}$ for more complicated functions (5.4.3) can be used. This is, for any two functions $\lambda_1(x)$, $\lambda_2(x)$ and $j \geq 1$,

$$P_j^{(n)}[\lambda_1(x) + \lambda_2(x)] = \sum_{p=0}^{j}\frac{1}{p!(j-p)!}P_1^{(n)}[\lambda_1^p(x)\lambda_2^{j-p}(x)].$$

Thus, for example, for the autoregression function

$$\lambda(x) = \lambda_1 x^{k_1}\exp\{-\frac{1}{2}w_1^2(x-d_1)^2\} + \lambda_2 x^{k_2}\exp\{-\frac{1}{2}w_2^2(x-d_2)^2\},$$

it is enough to be able to calculate

$$P_1^{(n)}[x^k\exp\{-\frac{1}{2}w^2(x-d)^2\}] \qquad (n \geq 0;\ k \geq 0) \qquad (5.5.11)$$

for these are essentially the terms of the summation with $k = pk_1 + (j-p)k_2$, $w^2 = pw_1^2 + (j-p)w_2^2$ and $d = \{pw_1^2d_1 + (j-p)w_2^2d_2\}/w^2$ $(p = 0,\ldots j)$. After some manipulations with generating functions it is possible to show that $P_1^{(n)}$ $(n \geq 0)$ of (5.5.11) are given by

$$P_1^{(n)} = \exp\left\{-\frac{w^2(a-d)^2}{2\tau^2}\right\}\sum_{q=0}^{\min(n,k)}\frac{k!}{(k-q)!}\binom{n}{q}\frac{(-i\sigma)^{k-q}w^{n-q}}{\tau^{n+k+1}}$$

$$\cdot H_{k-q}\left\{i\frac{a+d\sigma^2w^2}{\sigma\tau}\right\}H_{n-q}\left\{\frac{w(d-a)}{\tau}\right\}$$

where $\tau^2 = 1 + \sigma^2w^2$ $(\tau, w > 0)$.

6 NUMERICAL RESULTS

6.1 Introduction

In Chapter 5 numerical procedures were given for calculating sequences of approximations for the moments, joint moments and densities of stationary non-linear processes. While these procedures have been justified in restrictive circumstances (section 4.4) it is natural to try to apply them to a much wider range of processes: in either case it remains to determine the numerical behaviour of the algorithms. In the next section the role played by the arbitrary constant a is examined. The results of these numerical algorithms are compared with those obtained from simulations in section 6.3 and also with the exact results which are known for some processes. Throughout, only processes with a normally distributed input sequence are considered, and the algorithm is that based on the choice $b = 0$ rather than the more general case which appears intractable.

6.2 Role of the arbitrary constant

In the algorithm of section 5.2 the arbitrary constant a appears. Different values of a correspond to different families of approximating processes $\{X_n(\beta)\}$

$$X_{n+1}(\beta) = a + \beta[\lambda\{X_n(\beta)\} - a] + Z_{n+1} \qquad (n = \ldots -1,0,1,\ldots).$$

For each family the expression generating the process $\{X_n(1)\}$ is the same. However, while the exact moments of these processes are the same (but see section 4.7), the sequences of approximations to the moments are not. Thus the choice of a may well affect the speed of convergence of the procedure.

In Table 6.1 the sequences of approximations are given, according to the highest power of β included, for the mean of three different processes with several different choices of the constant a. For each of the processes considered the input distribution is normal with zero mean and unit variance: since the mean is zero it is natural to interpret the constant a as the approximation of zero order in β. The autoregression function of the process in Table 6.1(a) is $\lambda(x) = \exp\{-\frac{1}{2}x^2\}$. For this process the approximations converge rapidly for a good choice of a, the sequence having converged to eight decimal places at the eleventh term

ORDER	APPROXIMATIONS TO MEAN					
0	0.0	0.3	0.6	0.62756	0.9	1.2
1	.7071	.6914	.6462	.6408	.5775	.4933
2	.7071	.6508	.6373	.6381	.6613	.7025
3	.6050	.6161	.62622	.62634	.62137	.6194
9	.62832	.627595	.62756067	.62756075	.6275672	.62717
10	.62832	.627546	.62756060	.62756058	.6275610	.62764
11	.62731	.627557	.62756062	.62756062	.6275593	.62761
12	.62731	.627563	.62756062	.62756062	.6275613	.62751
14	.627644	.62756015			.62756065	.62756034
16	.627533	.62756069			.62756061	.62756434
18	.627570	.62756061			.62756062	.62756023
20	.62755750	.62756062				.62756040
22	.62756166					.62756067
24	.62756027					.62756063

Table 6.1(a): $\lambda(x) = \exp\{-\frac{1}{2}x^2\}$.

ORDER	APPROXIMATIONS TO MEAN					
0	0.0	0.6	0.9	1.02412	1.2	1.5
1	1.4142	1.2925	1.1550	1.0880	0.9867	0.8058
2	1.4142	1.0240	1.0225	1.0524	1.1130	1.2253
3	.5977	.9010	1.0065	1.0143	1.0036	.9950
12	-0.1544	1.0159	1.024187	1.024133	1.024143	1.0227
13	2.5950	1.0198	1.024174	1.024114	1.024097	1.0254
14	2.5950	1.0315	1.024067	1.024120	1.024129	1.0236
15	-1.0719	1.0227	1.024136	1.024121	1.024116	1.0240
16	-1.0719	1.0205	1.024127	1.024120	1.024121	1.0245
18	3.8221	1.0247	1.024123		1.024120	1.024188
20	-2.7116	1.0249	1.024119			1.024041
22	6.0123	1.0232	1.024120			1.024116
24	-5.6364	1.0247	1.024120			1.024137

Table 6.1(b): $\lambda(x) = 2\exp\{-\frac{1}{2}x^2\}$.

ORDER	APPROXIMATION TO MEAN						
0	1.5	1.8	1.9	2.0	2.5	3.0	3.3
1	1.76	2.10	2.21	2.32	2.85	3.27	3.47
2	2.04	2.44	2.56	2.68	3.18	3.47	3.57
3	2.50	2.78	2.87	2.95	3.26	3.43	3.51
4	2.86	3.07	3.12	3.18	3.34	3.43	3.48
5	3.29	3.33	3.32	3.31	3.30	3.38	3.44
19	3.326	3.3117	3.3051	3.3012	3.3076	3.3178	3.3043
20	3.356	3.3136	3.3063	3.3029	3.3045	3.3195	3.3093
21	3.370	3.3121	3.3071	3.3052	3.3030	3.3203	3.3139
22	3.365	3.3109	3.3076	3.3063	3.3025	3.3177	3.3201
23	3.354	3.3081	3.3068	3.3065	3.3025	3.3128	3.3242
24	3.331	3.3063	3.3072	3.3074	3.3031	3.3090	3.3223
25	3.309	3.3057	3.3069	3.3071	3.3038	3.3077	3.3164

Table 6.1(c): $\lambda(x) = 4 - 4\exp\{-\frac{1}{4}x^2\}$.

Table 6.1 Approximations to the means of the processes

$$X_{n+1} = \lambda(X_n) + Z_{n+1},\ Z_n \sim N(0,1),$$

to different orders of β and for various choices of the constant a, a = zero order approximation.

for a = 0.6. Although here the difference is small it appears that the sequence for a chosen to be very close to the mean converges marginally less quickly than that for a = 0.6. It seems generally true that the choice of a to be the mean of the process would not necessarily be best, even if this were known in advance. For table 6.1(b) the autoregression function is $\lambda(x) = 2\exp\{-\frac{1}{2}x^2\}$. Here making the choice a = 0 leads to a sequence of approximations for the mean which does not converge; the approximations apparently oscillate with increasing magnitude, possibly about some central value. In this case the best choice of a appears to be near the value eventually found for the mean, i.e. 1.02412, although the best speed of convergence of the sequence is considerably slower than that attainable for the first process. The autoregression function for the process of Table 6.1(c) is $\lambda(x) = 4\{1 - \exp(-\frac{1}{4}x^2)\}$. Here the

convergence of the approximations is much slower than for either of the other two processes and the best choice of the arbitrary constant a appears to be near a = 1.9, whereas the mean of the process, to which the sequence converges, is about 3.3. Usually the sequences converge while oscillating about the final value and this oscillation can be fairly slow.

In section 4.7 the question of the best choice of the constant a was discussed and it was concluded that, for a random variable X having the stationary distribution, a should be a "typical" value of $\lambda(X)$. This should be interpreted as being only a vague guide for choosing a. For the process of Table 6.1(a) $\lambda(x) = \exp\{-\frac{1}{2}x^2\}$, the value a = 1.2 is well outside the distribution of $\lambda(X)$ and yet the corresponding sequence of approximations does converge. However for the process of part (b) of the table, with $\lambda(x) = 2\exp\{-\frac{1}{2}x^2\}$, the value a = 0 is just on the boundary of the distribution in question and yet this leads to a sequence of approximations which diverges. Below in Table 6.2 are given the mean μ_X, variance σ_X^2, coefficient of skewness $\gamma = E\{(X - \mu_X)^3/\sigma_X^3\}$, and the correlation ρ at lag 1 of the process $\{X_n\}$ and also the skewness γ_λ of the process $\{\lambda(X_n)\}$.

$\lambda(x)$	μ_X	σ_X^2	γ	ρ	γ_λ
$\exp\{-\frac{1}{2}x^2\}$	.628	1.10	-0.013	-.187	-.481
$2\exp\{-\frac{1}{2}x^2\}$	1.024	1.49	-0.007	-.430	-.038
$4-4\exp\{-\frac{1}{4}x^2\}$	3.306	2.14	-0.73	.64	-1.9

Table 6.2: see text

These were all calculated by the process under consideration from sequences of approximations which had converged to at least the number of figures given. The skewness of the third process is large and negative and so is that of the process $\{\lambda(X_n)\}$: thus the relevant distribution (of $\lambda(X)$) has a heavy lower tail and this may explain why the best choice of a is considerably lower than the mean.

It appears to be generally true that the convergence of the sequences of approximations for the moments is slow when the variance of the process is appreciably larger than that of the input process or when the lag 1 correlation of the process is large (compared with 0.6, say).

Considering the power-series in β for the means of the first two of the above processes with $\lambda(x) = \exp\{-\frac{1}{2}x^2\}$, $\lambda(x) = 2\exp\{-\frac{1}{2}x^2\}$ respectively and with the choice $a = 0$ in both cases, it is clear that that for the second process is the same as the power series for the first but with β replaced by 2β. That is, the sequences in the tables are the partial sums of the same series at $\beta = 1$ and $\beta = 2$. Hence it follows that the corresponding power series converges at $\beta = 1$ but diverges at $\beta = 2$. It has been found numerically that the series converges at $\beta = 1.5$ (although only very slowly) but a more accurate estimate of the radius of convergence is not available.

In Table 6.3 the sequences of approximations for further moments of the particular process

$$X_{n+1} = 2\exp(-\tfrac{1}{2}X_n^2) + Z_{n+1} \quad (n = \ldots -1,0,1,\ldots)$$

with $Z_n \sim N(0,1)$ are given. In the column denoted '4' are the approximations for $E\{(X_n - \mu_X)^4\}$ and in the column denoted '2,1' for lag 1 moments are the approximations for $E[(X_n - \mu_X)^2(X_{n+1} - \mu_X)]$ and so forth, where μ_X is the mean of the process. Here the approximations to a particular order for the central moments have been calculated from the approximations to the same order for the moments about zero as produced by the algorithm of section 5.2. The arbitrary constant was chosen as $a = 1.0$. All the sequences for the moments converge at roughly the same rate with those for moments of higher degree converging slightly more slowly. This seems to be generally true. Each sequence had converged by the twentieth term.

ORDER	APPROXIMATIONS FOR CENTRAL MOMENTS			
	DEGREE = 1	2	3	4
10	1.024094	1.490149	-0.013445	6.304928
15	1.024122	1.490184	-0.013088	6.304585
20	1.024120	1.490188	-0.013080	6.304587
	CENTRAL MOMENTS AT LAG 1			
	1,1	1,2	2,1	2,2
10	-0.640986	0.111988	-0.503209	2.475243
15	-0.640914	0.112328	-0.503232	2.475301
20	-0.640911	0.112325	-0.503235	2.475300
	CENTRAL MOMENTS AT LAG 2			
	1,1	1,2	2,1	2,2
10	0.289313	-0.030621	0.198550	2.207313
15	0.289447	-0.030466	0.198698	2.207149
20	0.289457	-0.030463	0.198695	2.207147

Table 6.3: Approximations for moments of process $X_{n+1} = 2 \exp\{-\frac{1}{2} X_n^2\} + Z_{n+1}$, $Z_n \sim N(0,1)$ with a = 1.0.

6.3 Comparison with simulation results

6.3.1 The algorithm of section 5.2 provides values for the moments and joint moments of non-linear autoregressive processes. Although the correctness of these values has not been proved analytically, they have been verified by computer simulation for several different processes.

Taking into account the availability of unbiased estimators, it was decided to work with the joint cumulants of the process, and in particular, with the cumulants κ_{ijk} of (X_n, X_{n+1}, X_{n+2}) under the stationary distribution. Only twelve different combinations of the triple subscript (i,j,k) are chosen, namely (0,0,1), (0,0,2), (0,1,1), (1,0,1), (0,0,3), (0,1,2), (0,2,1), (1,0,2), (2,0,1), (0,0,4), (0,2,2), (2,0,2). Clearly, because of the assumed stationarity, many of the cumulants are equal, for example $\kappa_{100} = \kappa_{010} = \kappa_{001}$, $\kappa_{012} = \kappa_{120}$, etc., and the above avoids such combinations. The values of the cumulants are found from the values for the central moments produced by the algorithm and these are compared with the estimates of the cumulants obtained from simulations.

The estimates for the cumulants are based on the corresponding k-statistics (Kendall and Stuart, 1969, pp.280, 308) which are exactly unbiased if a number of independent samples of the random vectors concerned are available. In the present situation, that of estimating from simulated series, it is convenient to use triples $(X_{jh}, X_{jh+1}, X_{jh+2})$ $(j = 1,2,3,\ldots)$ which because of the stationarity of the series are effectively independent for large enough h. In order to increase the precision of the estimates of the cumulants it is possible to use the known structure of the process to relate the cumulants κ_{ijk} of (X_n, X_{n+1}, X_{n+2}) to those, κ^*_{ijk} say, of the triple $(X_n, X_{n+1}, \lambda(X_{n+1}))$ and to estimate these cumulants instead. When the input distribution is normal with zero mean and variance σ^2, using the independence of Z_{n+2} and (X_n, X_{n+1}),

$$\kappa_{ijk} = \kappa^*_{ijk} \text{ unless } i = j = 0 \text{ and } k = 2,$$

$$\kappa_{002} = \kappa^*_{002} + \sigma^2 .$$

Letting ave(·) denote the average of a number of terms of the same form as those inside the parentheses but spaced far apart in the series, the estimate k_{00i} of the mean $(= \kappa_{001})$ of the process

$$k_{001} = \text{ave}(X_n) = \text{ave}\{\lambda(X_{n-1}) + Z_n\} = \text{ave}\{\lambda(X_{n-1})\} + \text{ave}(Z_n)$$

is replaced by the estimate

$$k^*_{001} = \text{ave}\{\lambda(X_{n-1})\}.$$

If the average is composed of N terms, then (approximately, since the terms are only approximately independent)

$$\text{var}(k_{001}) = \frac{1}{N}[\text{var}\{\lambda(X_n)\} + \sigma^2] ,$$

$$\text{var}(k^*_{001}) = \frac{1}{N} \text{var}\{\lambda(X_n)\},$$

so, depending on the relative sizes of $\text{var}\{\lambda(X_n)\}$ and σ^2, the difference in precision of the estimates can be considerable and the above modification of the estimator is always worthwhile. The same sort of reduction in variance holds for the other k-statistics as can be seen from the expressions for the variances and other cumulants of the k-statistics given by Kendall and Stuart (1969, pp.290,311). These expressions for the variances of the k-statistics could be used to derive estimates of their precision: however values of higher order cumulants are required and these are not available if the theoretical values produced by the algorithm are in doubt. The procedure actually adopted is to form the k-statistics separately for several independent simulations and to take the sample mean, and its standard error based on the variation between the samples, as the final estimate of the cumulant and an estimate of its precision.

In the simulations performed, for given integers N, h, p the specified autoregressive process is used to generate values X_n, $\lambda(X_n)$ for $n = 1,\ldots,(N+p)h$ starting from a fixed number X_0. Of these the first ph values of X_n are discarded. Then from the N triples $(X_n, X_{n+1}, \lambda(X_{n+1}))$ $(n = rh+1;\ r = p,\ldots N+p-1)$ the twelve k-statistics are formed. The precision of these estimates can be improved by calculating each k-statistic for each of the h sets of N triples, namely $\{(X_n, X_{n+1}, \lambda(X_{n+1}));$ $n = rh+j;\ r = p,\ldots N+p-1\}$ $(j = 1,\ldots h)$. Then unbiased estimates of the cumulants are formed by taking the average of the h values of each k-statistic. In general the terms in each average are not independent so that the variance of the resulting estimate is not necessarily reduced

by a factor h^{-1}, though it is decreased. Whether it is better to adopt this procedure, or to calculate each k-statistic just once and to increase the size of N in order to obtain a better estimate, depends both on the amount of dependence amongst the terms of the series and on the cost of generating the further terms required for the longer sequence. The cost of computing h sets of k-statistics from N triples is about the same as that for computing one set from Nh triples. For the simulations here the average of h k-statistics was taken as the estimate for the cumulants, though this may have been the wrong choice. Once estimates of the cumulants have been found the whole procedure is repeated with an independent input series. Because of the discarding of the initial section of the series the choice of X_0 is of little importance (except that it should not be too extreme) and a convenient value to take is that of the last term of the previous series, namely $X_{(N+p)h}$. In all the procedure is applied to M independent sections of the process producing M independent estimates of the twelve cumulants. For each of the cumulants a final estimate is provided by the mean of the M independent estimates and an estimate of its variance is given by the obvious estimate based on the sample variance of the M estimates. For large N the k-statistics are approximately normally distributed and taking the average of h values results in a distribution closer to normality. Hence each set of M estimates of the cumulants is a set of independent and identically distributed observations from an approximately normal distribution and thus the sample variance of each set is approximately distributed as χ^2 with M-1 degrees of freedom.

The processes simulated required normally distributed random variables for the input series and pseudo-random variables for these were generated by the 'Fast' method of Marsarglia, Maclaren and Bray (1964): the subroutine used for this was programmed in Fortran by Mr. M.C. Pearce and the uniformly distributed pseudo-random variables required by the procedure were obtained from the standard RANF function as implemented on the CDC 6400/Cyber 7314 system at Imperial College (Atkinson and Pearce, 1976). For each different process simulated a different value was chosen (haphazardly) for the initial seed for the RANF function so that essentially different input sequences were generated and used for the different processes. Of course the different sections of the same process were generated using independent sets of input values: in fact these were taken as the continuation of the sequence used for the earlier sections.

The k-statistics themselves were calculated using an updating technique found by expressing the particular k-statistic based on n triples in terms of the n'th triple and k-statistics based on the first (n-1) triples. This reduces the possibility of truncation errors arising from adding and subtracting relatively large sums of powers and products to form the cumulants by the standard formulae (Kendall and Stuart, 1969, p.84). For the relatively moderate sizes of N and (N < 2000), the particular processes simulated and the length of the computer word used (equivalent to about 14 decimal significant figures), the effect of this would probably be of no importance for the accuracy required.

For all four processes simulated the input distribution was normal with zero mean and unit variance. The autoregression functions for the four processes are given below in Table 6.4 together with the values p, N, h and M used for each simulation. The computer time taken by each of these simulations was between 300 and 350 seconds.

AUTOREGRESSION FUNCTION	p	N	h	M
(i) $\lambda(x) = \exp(-\frac{1}{2}x^2)$	2	1600	12	21
(ii) $\lambda(x) = \frac{1}{2}x\{1 + \exp(-\frac{1}{2}x^2)\}$	2	1500	14	21
(iii) $\lambda(x) = x\{\frac{1}{2} - \exp(-\frac{1}{2}x^2)\}$	2	1600	14	21
(iv) $\lambda(x) = \begin{cases} x + \frac{1}{2} & (\text{for } -1.5<x<1) \\ -1.0 & (\text{otherwise}) \end{cases}$	2	1700	14	21

Table 6.4: see text

Tables 6.5(a)-(d) give the results of the simulations. The values for the cumulants found from the algorithm under test are given together with their estimates from the simulations. Also given are the absolute value of the difference between these values and the estimate of the standard error of the simulation result. For two of the processes the autoregression function is odd about zero and, since the input distribution is symmetric, it is therefore known that the cumulants of odd total degree are all zero. Using this fact could have resulted in more accurate estimates of the other cumulants as well as a saving in computing time. However this was not done: the results produced for these known cumulants are a partial check on the simulation and estimation procedures.

Cumulant	value from algorithm	simulation estimate	|difference|	standard error
001	.627561	.627957	.000396	.000401
002	1.099454	1.099512	.000058	.000159
011	-.205555	-.204810	.000745	.000561
101	.033657	.033174	.000483	.000483
003	-.015102	-.015140	.000038	.000075
012	.045580	.045309	.000271	.000263
021	-.292949	-.293319	.000370	.001617
102	-.006883	-.006690	.000193	.000157
201	.044806	.045542	.000736	.000988
004	-.010992	-.011014	.000022	.000062
022	.002316	.003064	.000748	.000598
202	-.011322	-.011434	.000112	.000193

Table 6.5(a): Simulation results for process with autoregression function $\lambda(x) = \exp(-\frac{1}{2}x^2)$, $Z_n \sim N(0,1)$.

Cumulant	value from algorithm	simulation estimate	|difference|	standard error
001	0.0	.00193	.00193	.00260
002	1.65597	1.65659	.00053	.00108
011	1.02333	1.02426	.00092	.00199
101	.64436	.64427	.00010	.00172
003	.0	-.00292	.00292	.00196
012	.0	-.00477	.00477	.00272
021	.0	-.00790	.00790	.00409
102	.0	-.00318	.00318	.00205
201	.0	-.00465	.00465	.00325
004	-.51426	-.51522	.00096	.00219
022	-.82743	-.82759	.00016	.00531
202	-.49864	-.49729	.00134	.00396

Table 6.5(b): Simulation results for process with autoregression function $\lambda(x) = \frac{1}{2}x\{1 + \exp(-\frac{1}{2}x^2)\}$, $Z_n \sim N(0,1)$.

Cumulant	value from algorithm	simulation estimate	|difference|	standard error
001	0.0	-.00008	.00008	.00040
002	1.10855	1.10966	.00111	.00052
011	.19407	.19659	.00252	.00114
101	.04620	.04746	.00126	.00067
003	.0	-.00050	.00050	.00067
012	.0	-.00104	.00104	.00136
021	.0	-.00191	.00191	.00257
102	.0	.00012	.00012	.00092
201	.0	.00082	.00082	.00170
004	.07741	.07925	.00184	.00111
022	.37629	.38385	.00756	.00437
202	.08318	.08592	.00274	.00186

Table 6.5(c): Simulation results for process with autoregression function $\lambda(x) = x\{\frac{1}{2} - \exp(-\frac{1}{2}x^2)\}$, $Z_n \sim N(0,1)$.

Cumulant	value from algorithm	simulation estimate	|difference|	standard error
001	-.18216	-.18207	.00009	.00113
002	1.65098	1.65191	.00093	.00078
011	.24206	.24453	.00247	.00119
101	-.02125	-.02343	.00218	.00137
003	.25630	.25701	.00071	.00130
012	.26524	.26559	.00035	.00058
021	-1.00806	-1.01074	.00268	.00257
102	.05030	.05040	.00011	.00120
201	-.25849	-.26047	.00198	.00272
004	-.48992	-.49234	.00242	.00230
022	-.00245	-.00486	.00241	.00208
202	-.21442	-.21475	.00033	.00177

Table 6.5(d): Simulation results for process with autoregression function $\lambda(x) = x + 0.5$ for $-1.5 < x \leq 1$, $= -1.0$ otherwise, $Z_n \sim N(0,1)$.

(Further checks on the cumulant estimation procedure were made using a process of independent normal pseudo-random values.)

The estimates of the cumulants produced by the simulations agree very well with the values produced by the algorithm of section 5.2. The differences between the two values are within just over two (about 2¼) standard errors of zero, the accuracy of the algorithm being verified to two or three decimal places for the different cumulants. More formal tests could be applied based on the ratios of the differences to the standard errors which would be distributed approximately as Student's t with M-1 = 20 degrees of freedom. The estimates of the different cumulants for the same process are not independent so that the tests for the cumulants cannot easily be combined.

6.3.2 Besides the comparison with the results of simulations, the algorithm for determining the moments of non-linear autoregressive processes may be validated by using it to generate the moments of processes whose properties are known.

The simplest processes that can be used for this purpose are the linear processes. The sequences of approximations for the moments all converged to the correct values for all linear stationary processes (with normal input distributions) that were tried. In fact it can be shown algebraically that the algorithm gives the correct results for the moments of any linear process (with any input distribution possessing moments of all orders) with slope strictly less than one in absolute value.

It is also possible to calculate the exact stationary distribution for certain other non-linear processes: those for which the autoregression function takes on only a finite number of distinct values, see section 2.3. In particular the moments and covariances of the process $\{x_n\}$ with a two-level, odd about zero, autoregression function can be calculated when the input distribution is normal (equation 2.3.5). This has been done for several different processes and the theoretical values agree exactly (to 8 decimal places anyway) with the values determined from the algorithm.

For processes for which the autoregression function $\lambda(x)$ is either equal to x or to zero, the first covariance and the variance obey a simple relationship and this holds also for the approximations produced by the algorithm.

6.4 Results for some specific processes

Figures 6.1-6 give simulations of various processes and the approximations for the stationary densities given by the methods of Chapter 5. Part (a) of each figure consists of a computer realisation $\{X_n,\ n = 0,\ldots 150\}$ of the process concerned together with plots of X_{n+1} against X_n and of X_{n+2} against X_n for the series obtained. In the first of these a graph of the autoregression function is also given. In part (b) of each figure some densities associated with the stationary processes are plotted. The marginal density of the process is shown in the lower left-hand part of the diagram and above it is a contour plot of the joint stationary density of (X_n,X_{n+1}). This was obtained from the approximation for the marginal density using formula (5.3.6). At the top right-hand side of the diagram is a contour plot of the joint density of (X_n,X_{n+2}): this was obtained from formula (5.3.8) based essentially on the approximations for the joint moments. Below this plot the two-step ahead predictive densities are given, these being obtained from the joint density of (X_n,X_{n+2}). In each of the contour diagrams the levels of the contours drawn are at 0.1, 0.05, 0.01, 0.005, 0.001, 0.0005 and 0.0001 units of joint probability density. Not all of these contours appear on every diagram when the density is moderately dispersed but the outside contour always corresponds to 0.0001.

For each of the six processes the input distribution was normal with zero mean and unit variance. Were the processes linear the two-step ahead predictive densities would also be normal with only the mean depending on the value on which the process is conditioned. This is clearly not so for these non-linear processes. For the processes of figures 6.5 and 6.6 the autoregression function is discontinuous and the joint densities should reflect this. The joint density at lag 1 of the process of figure 6.5 has been drawn by a computer program taking into account the discontinuity whereas that in figure 6.6 has not allowed for the discontinuity. The difference can be clearly seen. For both of these processes the joint density at lag 2 should also have a discontinuity but the method adopted for these approximations does not deal too well with these discontinuities. It is evident that, for the process of figure 6.5, the predictive densities conditioned on any value less than zero should be the same. However for the approximations given this is not quite so; the densities conditioned on values nearest zero are further from the

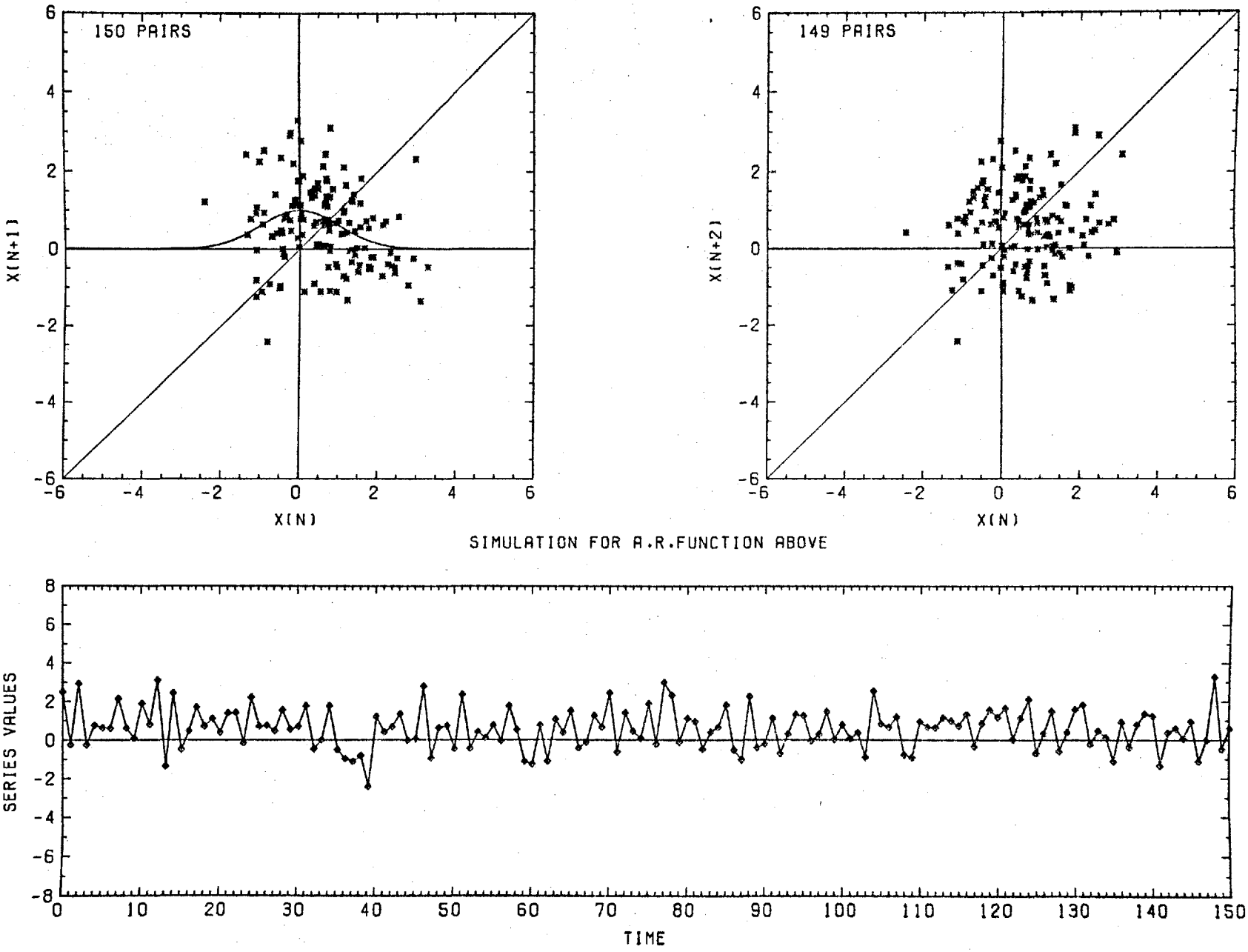

Figure 6.1(a): $\lambda(x) = \exp(-\frac{1}{2}x^2)$, $Z_n \sim N(0,1)$.

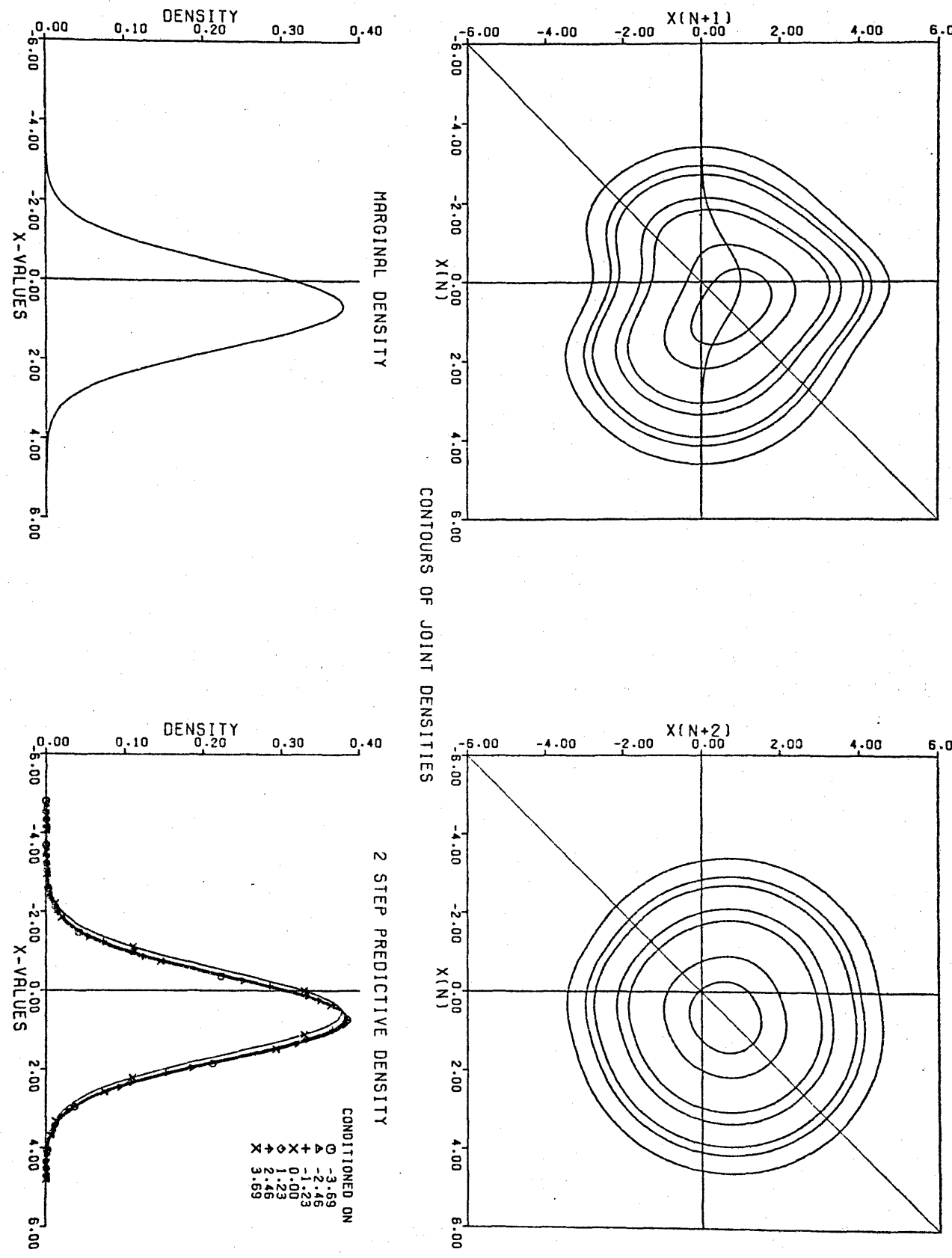

Figure 6.1(b): $\lambda(x) = \exp(-\frac{1}{2}x^2)$, $Z_n \sim N(0,1)$.

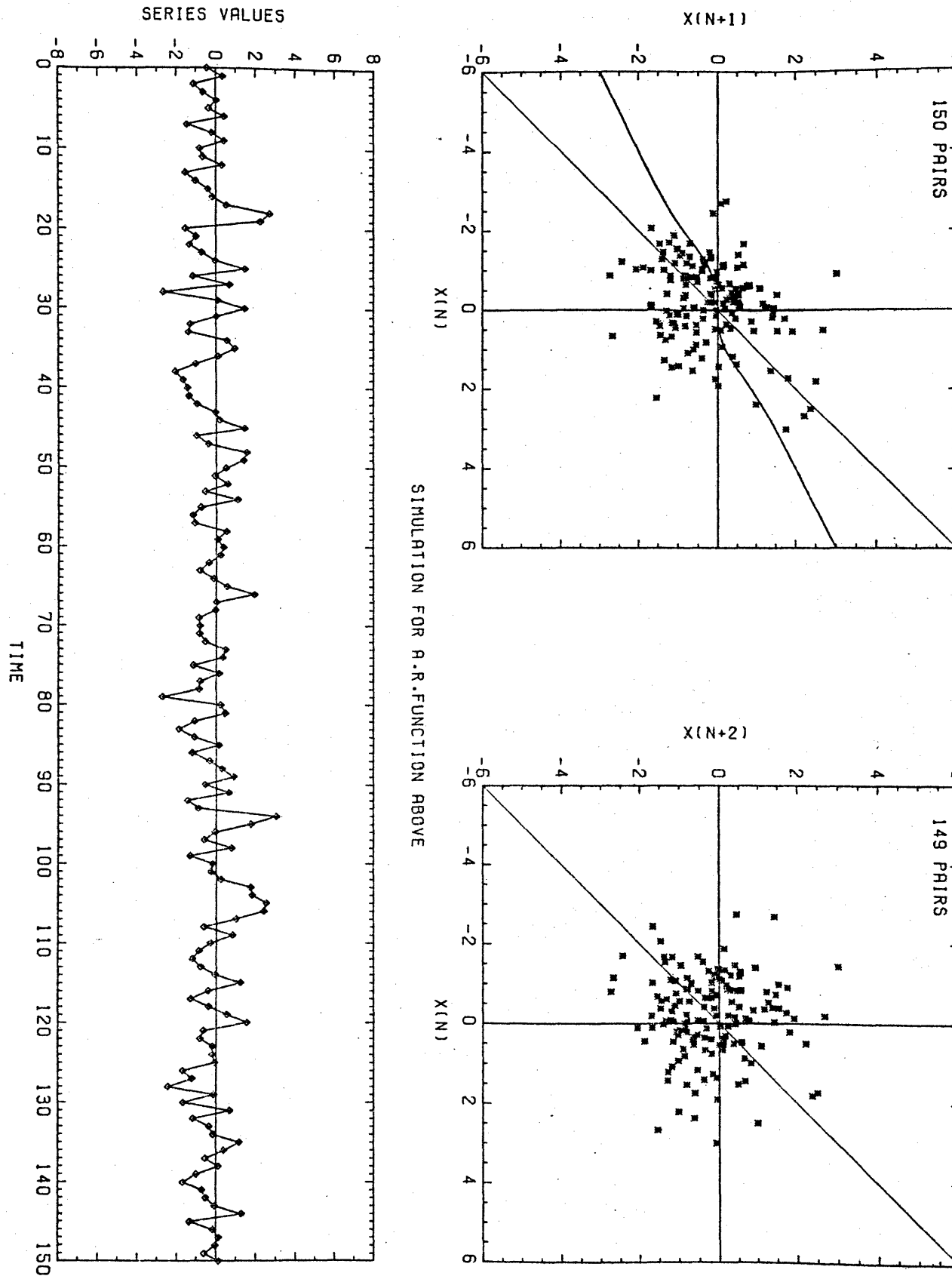

Figure 6.2(a): $\lambda(x) = \frac{1}{2}x\{1 - \exp(-\frac{1}{2}x^2)\}$, $Z_n \sim N(0,1)$.

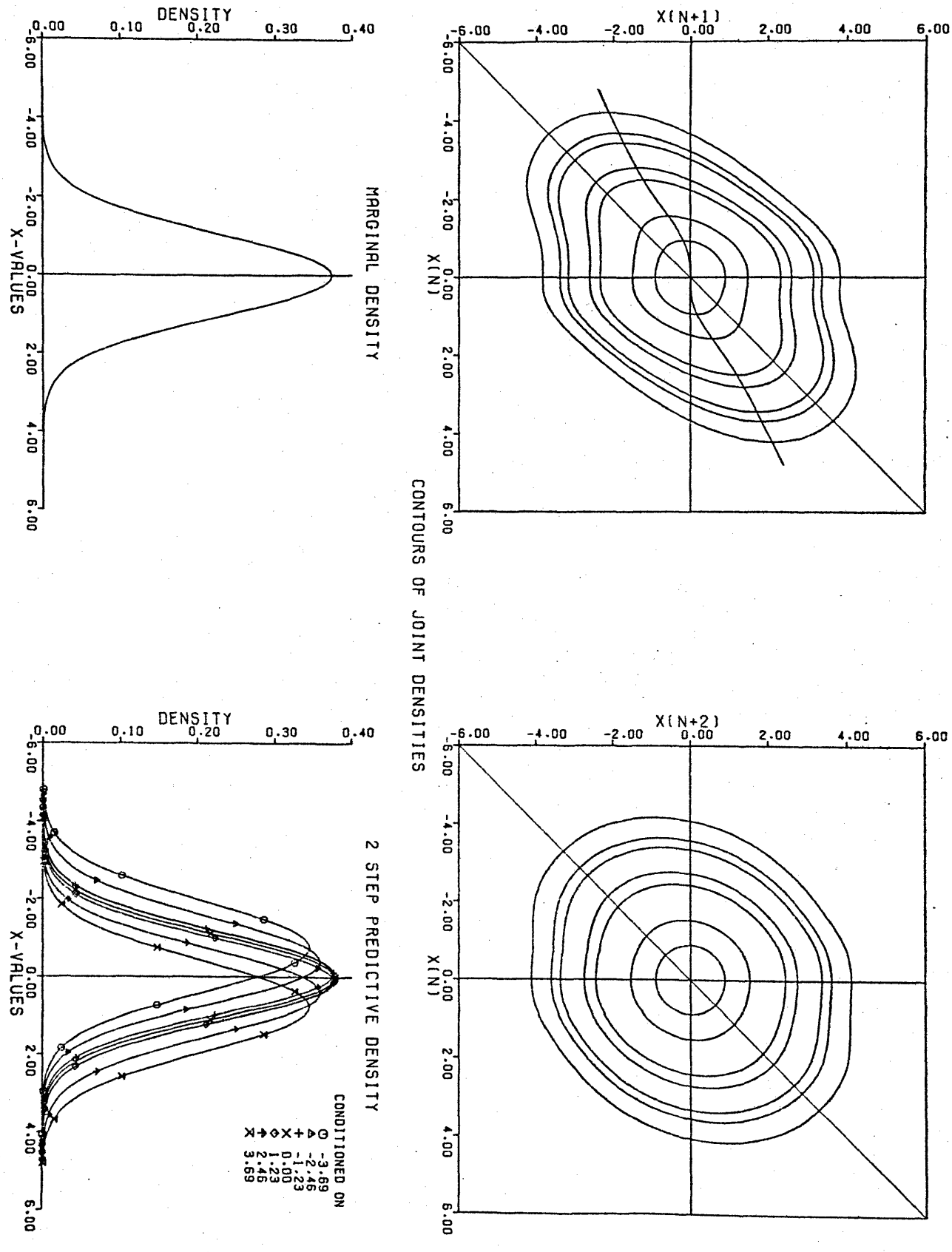

Figure 6.2(b): $\lambda(x) = \frac{1}{2} x \{1 - \exp(-\frac{1}{2} x^2)\}$, $Z_n \sim N(0,1)$.

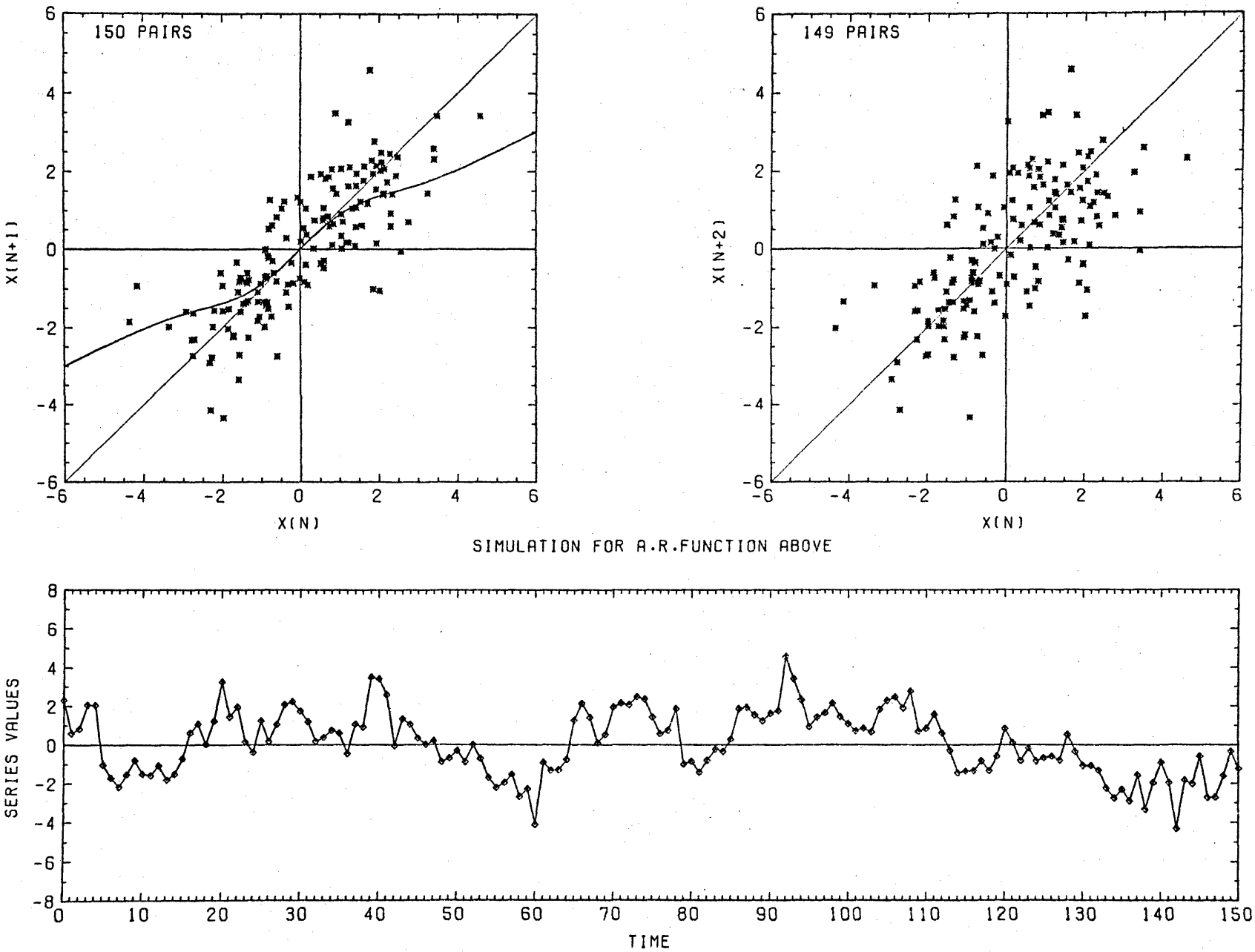

Figure 6.3(a): $\lambda(x) = \frac{1}{2}x\{1 + \exp(-\frac{1}{4}x^2)\}$, $Z_n \sim N(0,1)$.

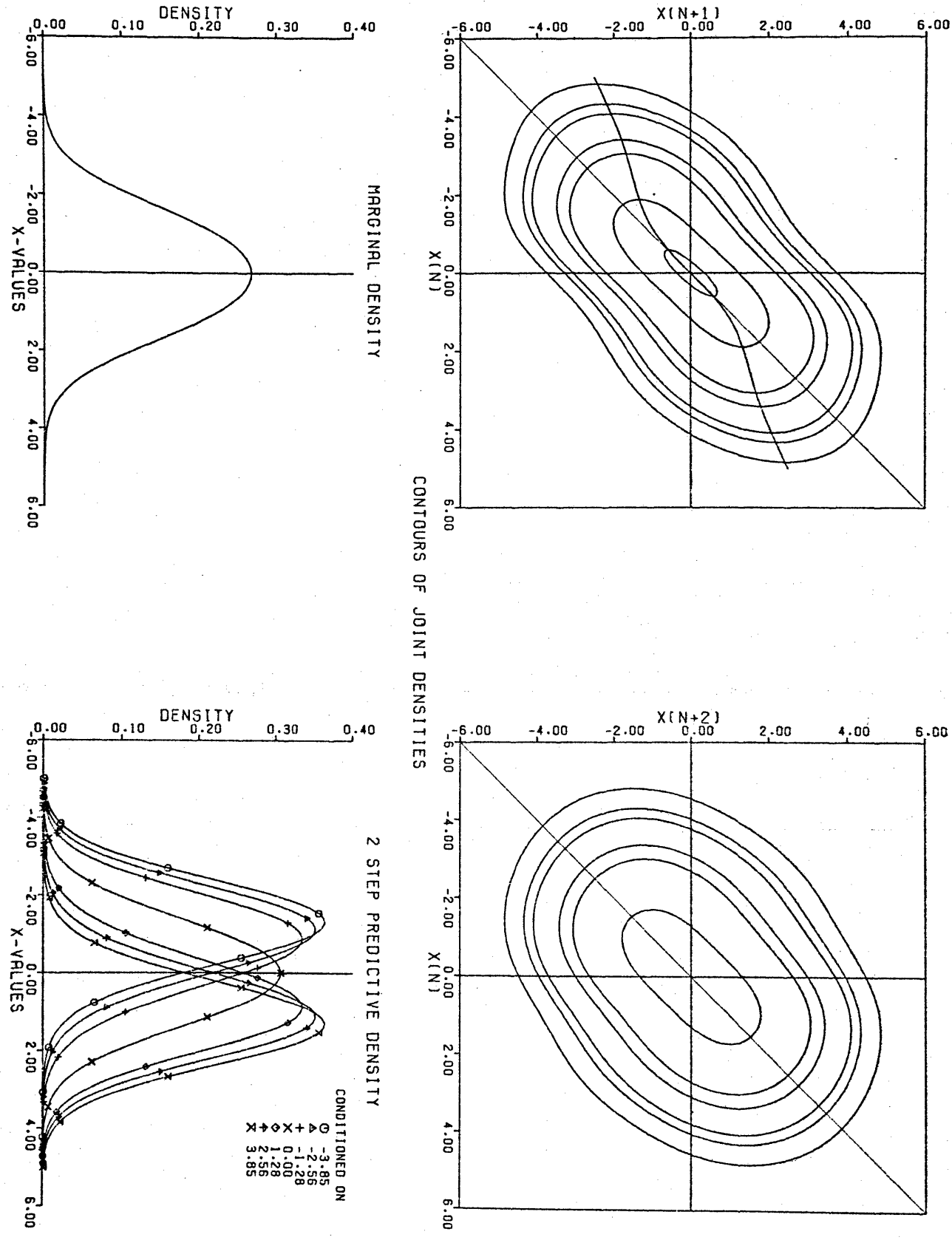

Figure 6.3(b): $\lambda(x) = \frac{1}{2} x\{1 + \exp(-\frac{1}{4} x^2)\}$, $Z_n \sim N(0,1)$.

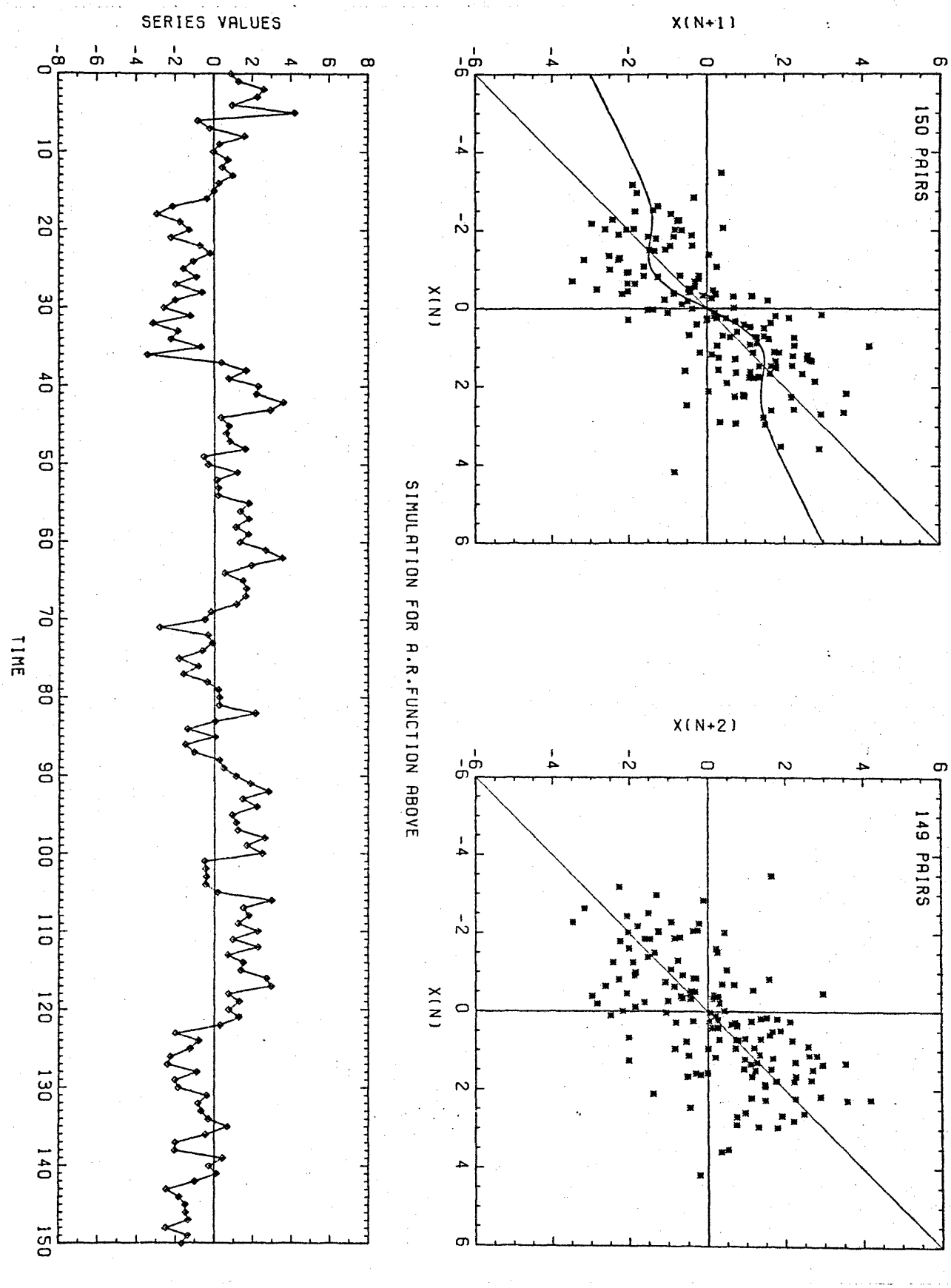

Figure 6.4(a) $\lambda(x) = \frac{1}{2} x\{1 + 3\exp(-\frac{1}{2} x^2)\}$, $Z_n \sim N(0,1)$.

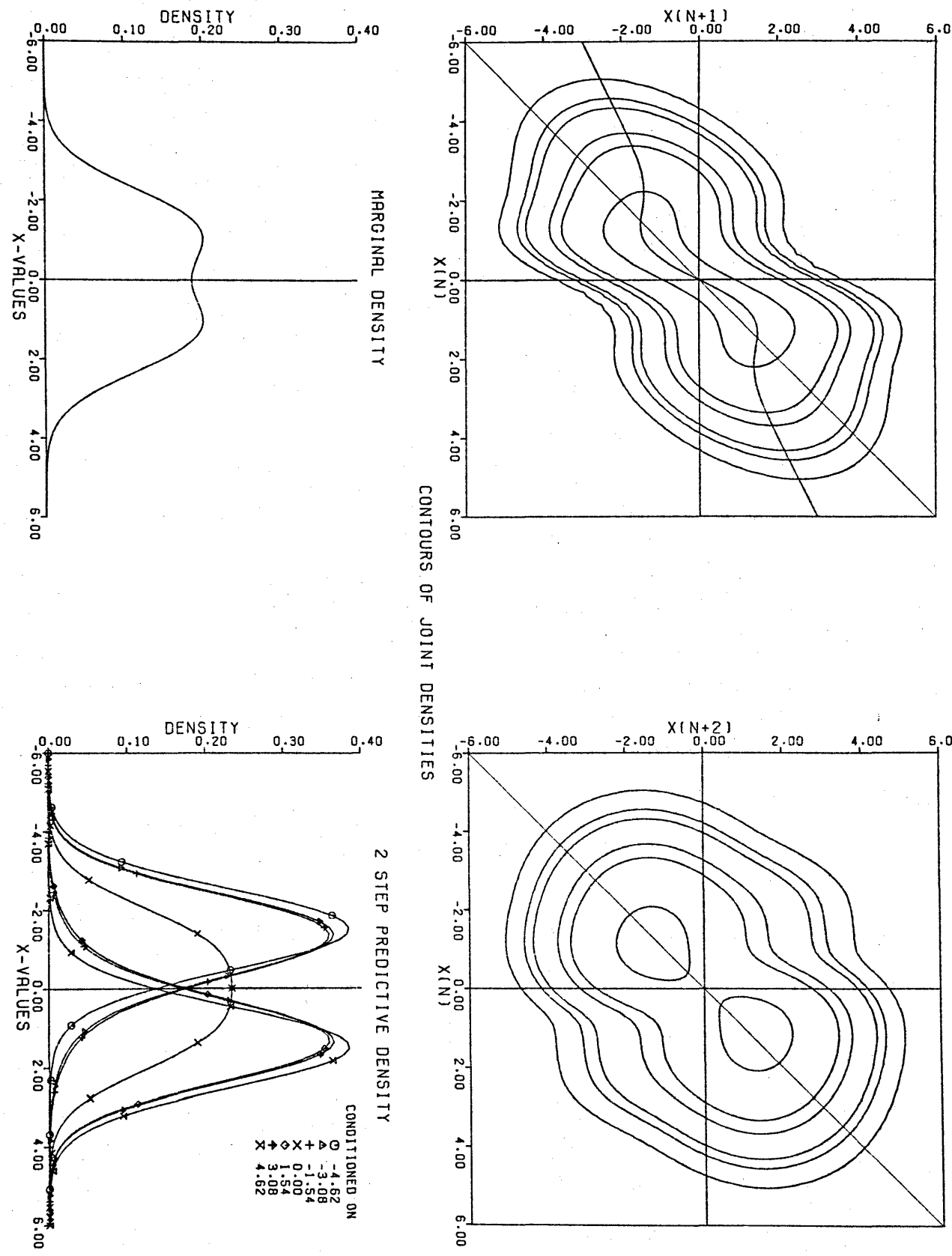

Figure 6.4(b): $\lambda(x) = \frac{1}{2}\{1 + 3\exp(-\frac{1}{2}x^2)\}$, $Z_n \sim N(0,1)$.

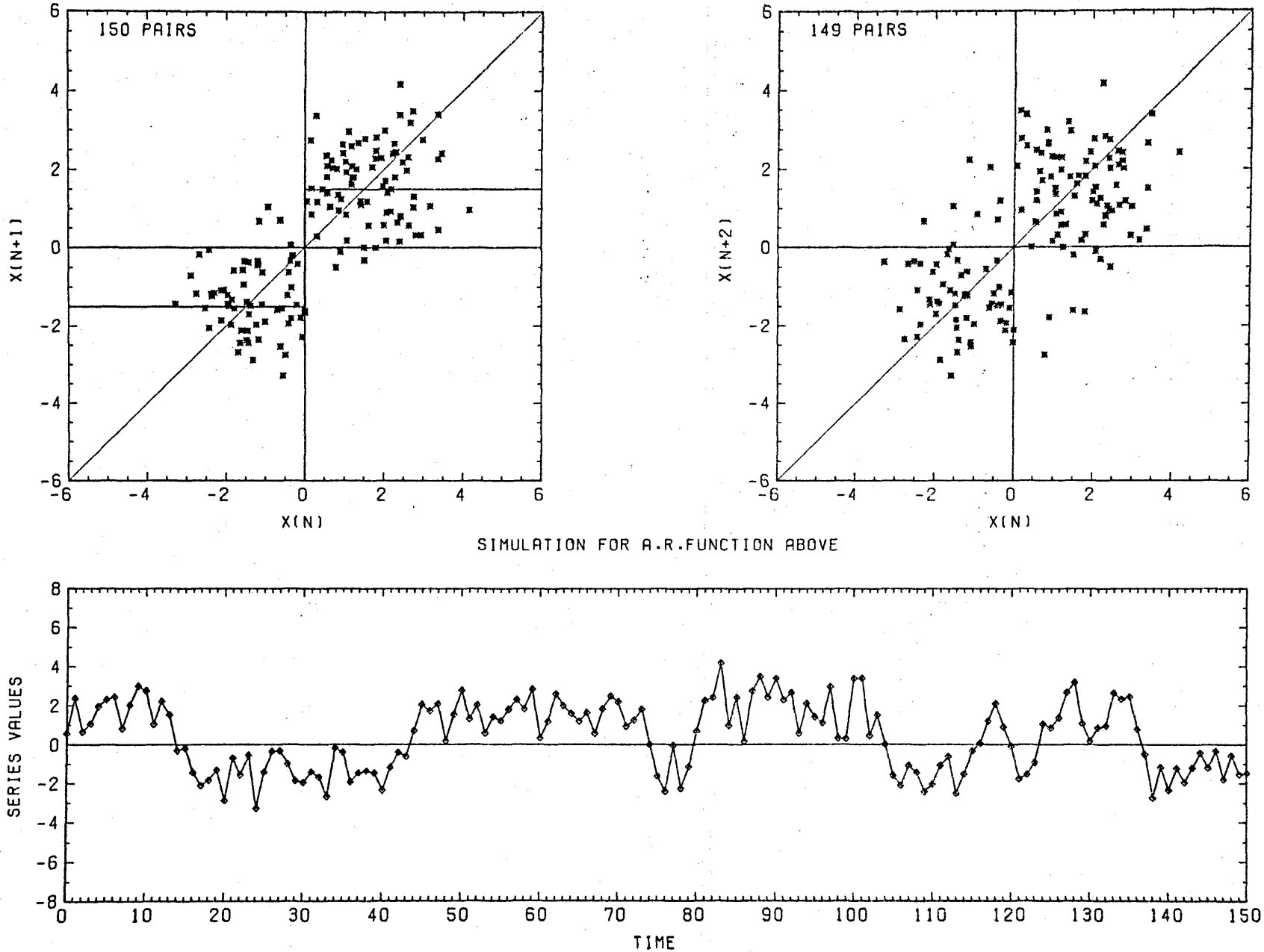

Figure 6.5(a): $\lambda(x) = 1.5\ \mathrm{sgn}(x)$, $Z_n \sim N(0,1)$.

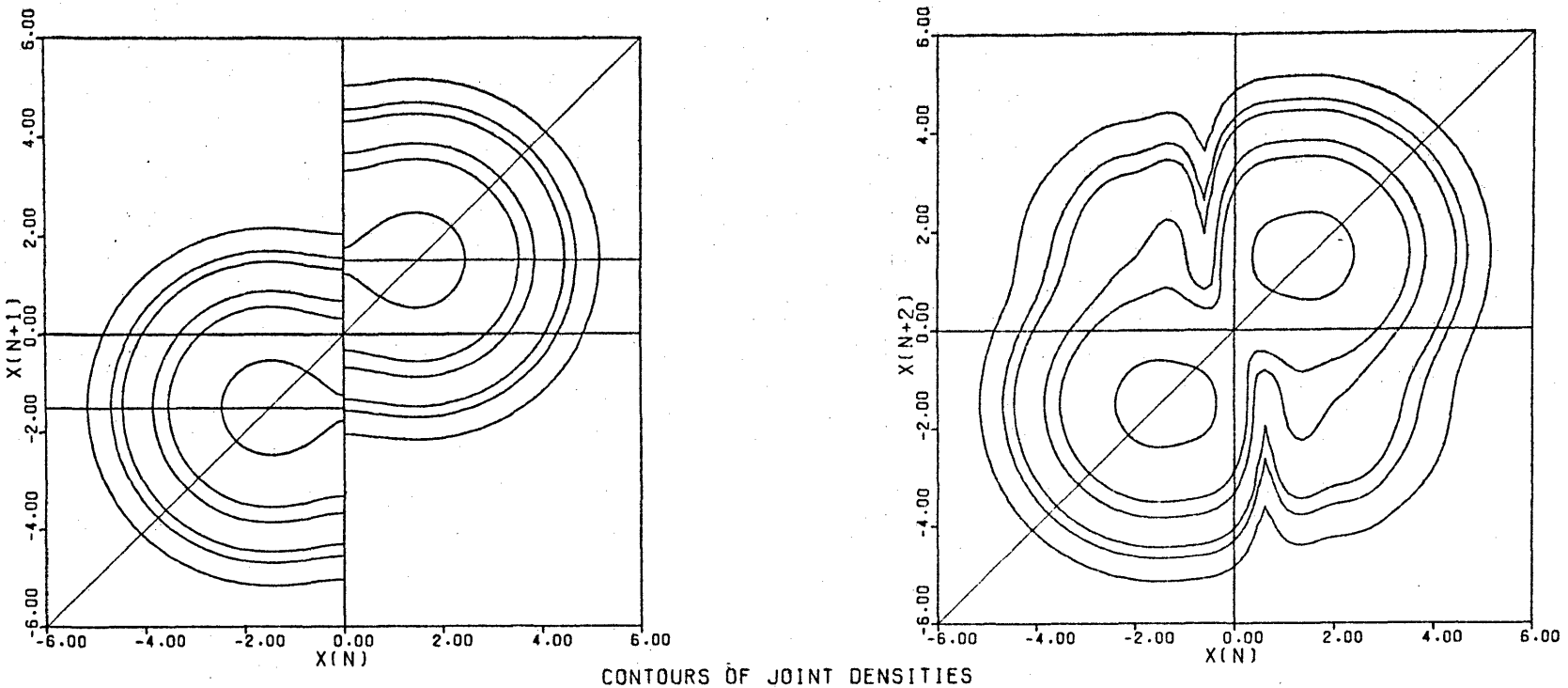

Figure 6.5(b): $\lambda(x) = 1.5\ \mathrm{sgn}(x)$, $Z_n \sim N(0,1)$.

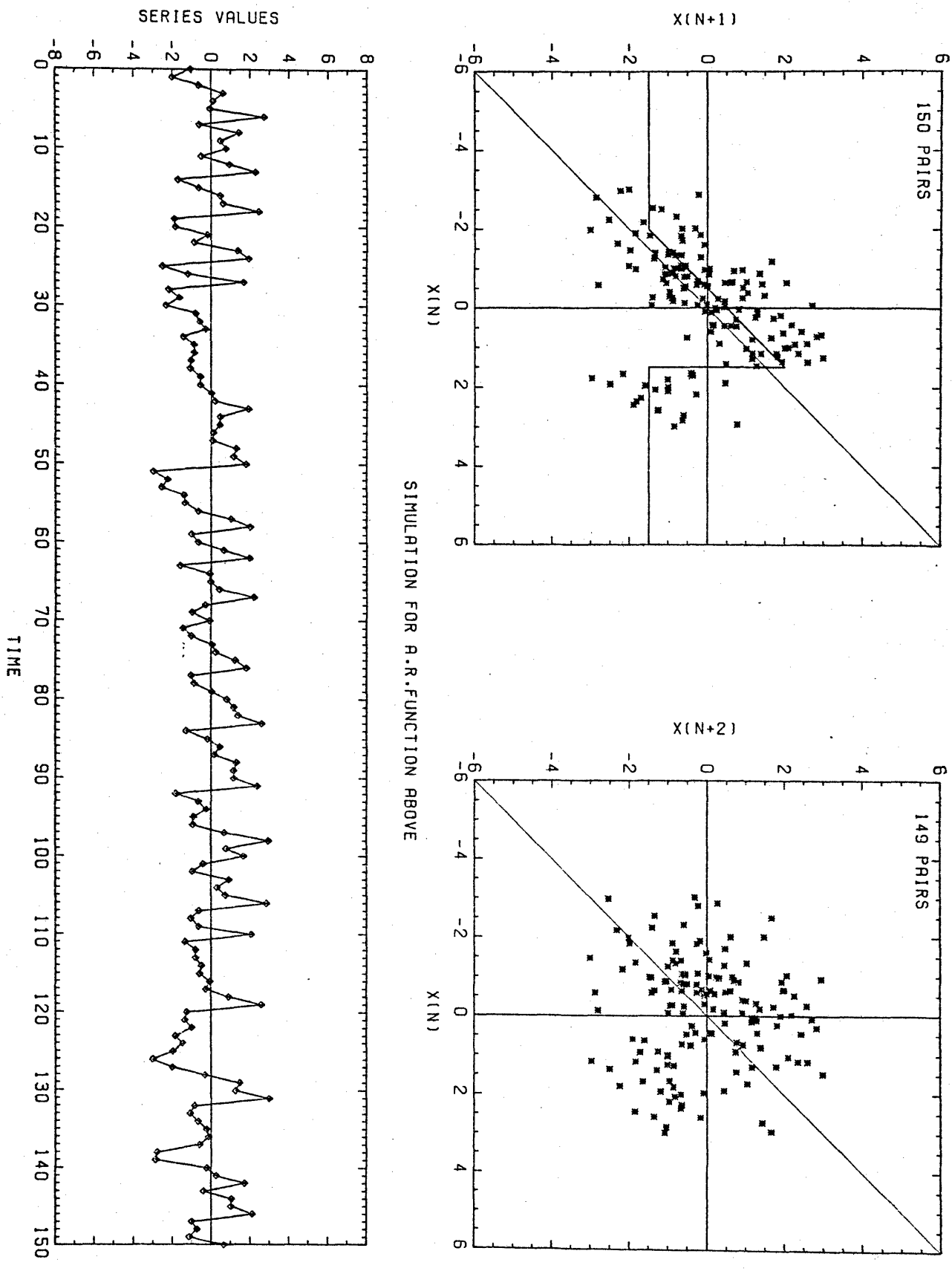

Figure 6.6(a): $\lambda(x) = x + 0.5$ $(-2 < x \leq 1.5)$, $= -1.5$ (otherwise),

$Z_n \sim N(0,1)$.

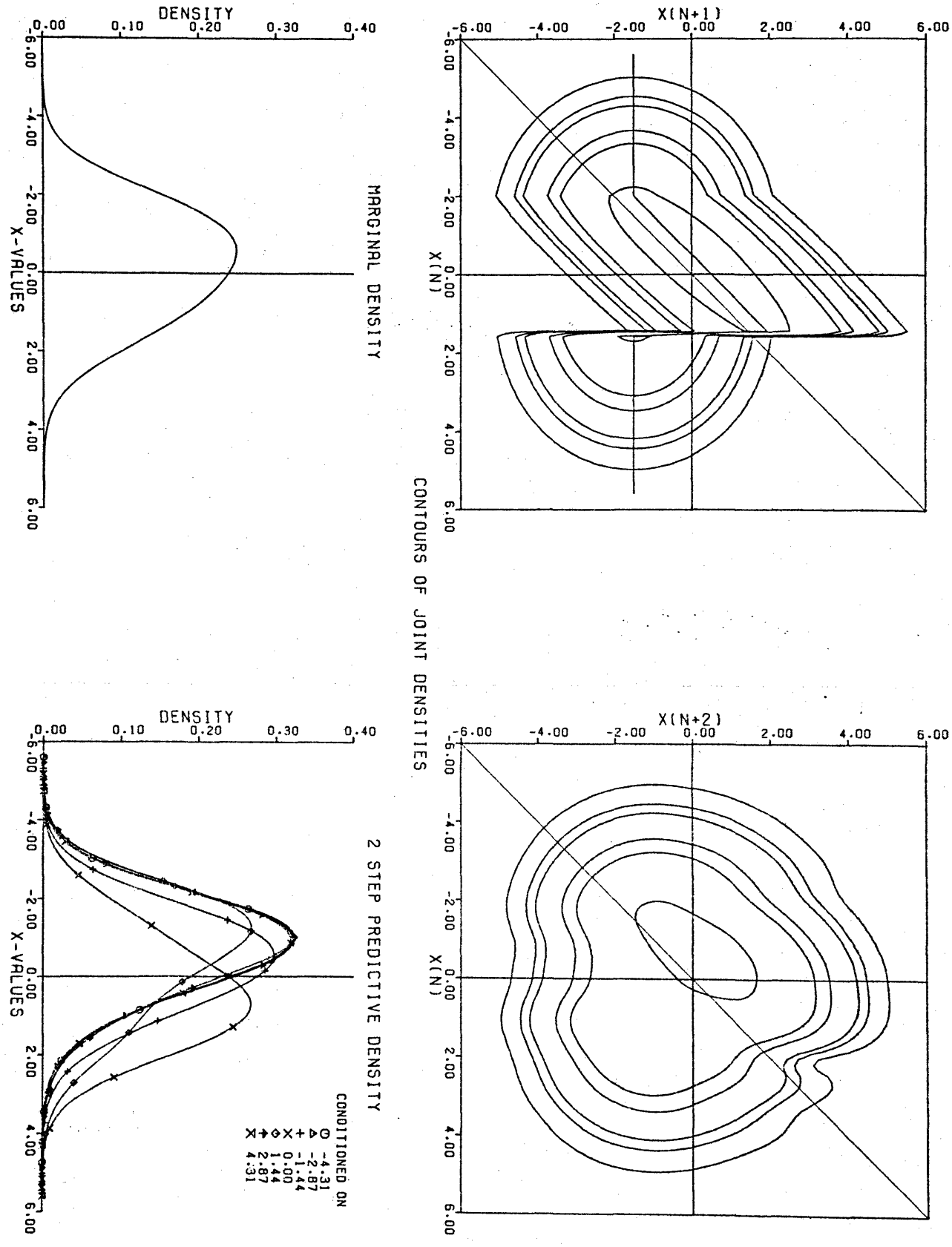

Figure 6.6(b): $\lambda(x) = x + 0.5$ $(-2 < x \leq 1.5)$, $= -1.5$ (otherwise), $Z_n \sim N(0,1)$.

other pairs of densities which are themselves almost coincident. It is interesting that the predictive density conditioned on a value at zero which is produced by this method should coincide with the marginal density. It is clear that the algorithm of Chapter 5 does not distinguish between autoregression functions differing at a single point but, in this case at least, the approximations produced behave as if the autoregression function were defined to take a value halfway between the left- and right-hand limits at the discontinuity. It seems likely that first forming approximations for the conditional densities and then using the approximation for the marginal density to form the joint density would yield better results. However the method used here is easier to program and approximations for joint moments of any order can be calculated at the same time. In using the other method sets of conditional moments would be available.

Figure 6.7 gives an example of the backward in time conditional densities of X_n given X_{n+1} for the autoregressive process with $\lambda(x) = \frac{1}{2} x\{1 + \exp(-\frac{1}{2} x^2)\}$ and with input normally distributed with zero mean and unit variance. The fine line on this diagram is the marginal density of the process. Clearly the backward process cannot be represented as a non-linear autoregression since the variance and shape of the conditional distributions vary. However the backward processes are Markov Processes with one-step transition probabilities given by these backward conditional densities.

All the above approximations for the densities have been obtained by the method of Chapter 5 with J*, L*, M* and N* always chosen to be about 21.

In figures 6.8-10 some examples of the spectra of non-linear autoregressive processes are given. Approximations for the covariances of a particular process can easily be obtained from the algorithm described in section 5.2.4 and from these the corresponding approximation for the spectral density function can be formed in the obvious way. In the examples given here it has not been found necessary to include covariances at lags greater than twelve or thirteen. Also given in each diagram, as a fine line, is the spectrum of a first-order linear autoregressive process having the same variance and lag 1 covariance as the non-linear process of that figure. The process corresponding to figure 6.8 has autoregression function

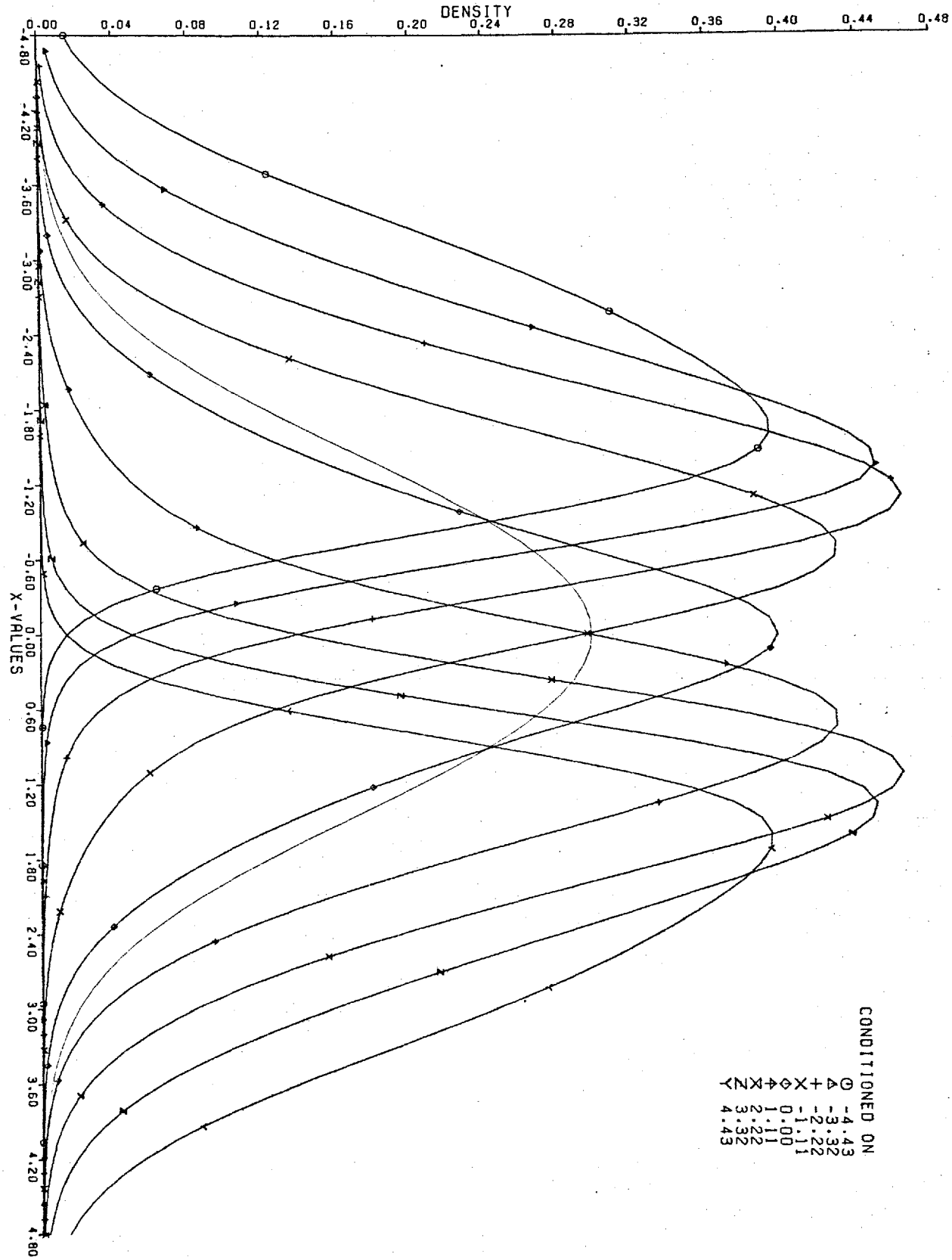

Figure 6.7: Backward conditional densities at lag 1 for autoregressive process with $\lambda(x) = \frac{1}{2}\, x\{1 + \exp(-\frac{1}{2}\, x^2)\}$, $Z_n \sim N(0,1)$.

$$\lambda_1(x) = \begin{cases} x + 0.5 & (-1.5 < x \leq 1.0), \\ -1.0 & (\text{otherwise}), \end{cases}$$

the autoregression function of the process of figure 6.9 is

$$\lambda_2(x) = \begin{cases} x + 0.5 & (-2.0 < x \leq 1.5), \\ -1.5 & (\text{otherwise}), \end{cases}$$

and that corresponding to figure 6.10 is

$$\lambda_3(x) = \begin{cases} x & (-2.0 < x < 2.0), \\ 0 & (\text{otherwise}). \end{cases}$$

Again the input distribution is the standard normal distribution. The first two processes can each be described as being like a random walk with upward drift and with a threshold level such that once this is passed the process returns to a low level. Thus the processes should rise slowly and drop sharply once the threshold is reached. This behaviour is just discernable in Figure 6.6 although, because of the relatively large variation of the input process, it is somewhat obscured. Figure 1.1 gives a realisation of a similar process for which this type of structure is much more clearly apparent. Unfortunately the sequences of approximations for the moments converge only slowly, if at all, for processes for which this behaviour is more distinct and so the spectra cannot be calculated. If the structure described above does underlie the behaviour of the processes it would be expected that the spectra would reflect this by having a peak located away from zero, or at least a region of relatively high spectral density compared with a linear process. The spectra for the processes with autoregression functions λ_1 and λ_2 both have a peak away from zero. The maximum in figure 6.8 is near 0.12π although the maximum value is only 0.0004 larger than that at zero. The maximum of the spectrum in figure 6.9 is near 0.18π. For these two processes the spectra are fairly different from the spectra of the linear process also shown in the diagrams although this is possibly because of the special structure of the processes. The spectra of many other processes have been found to be close to the spectra of the fitted first-order linear autoregressive processes, an example being given in figure 6.10. It is clear that, if the spectrum of a process is close to that of a first-order autoregressive

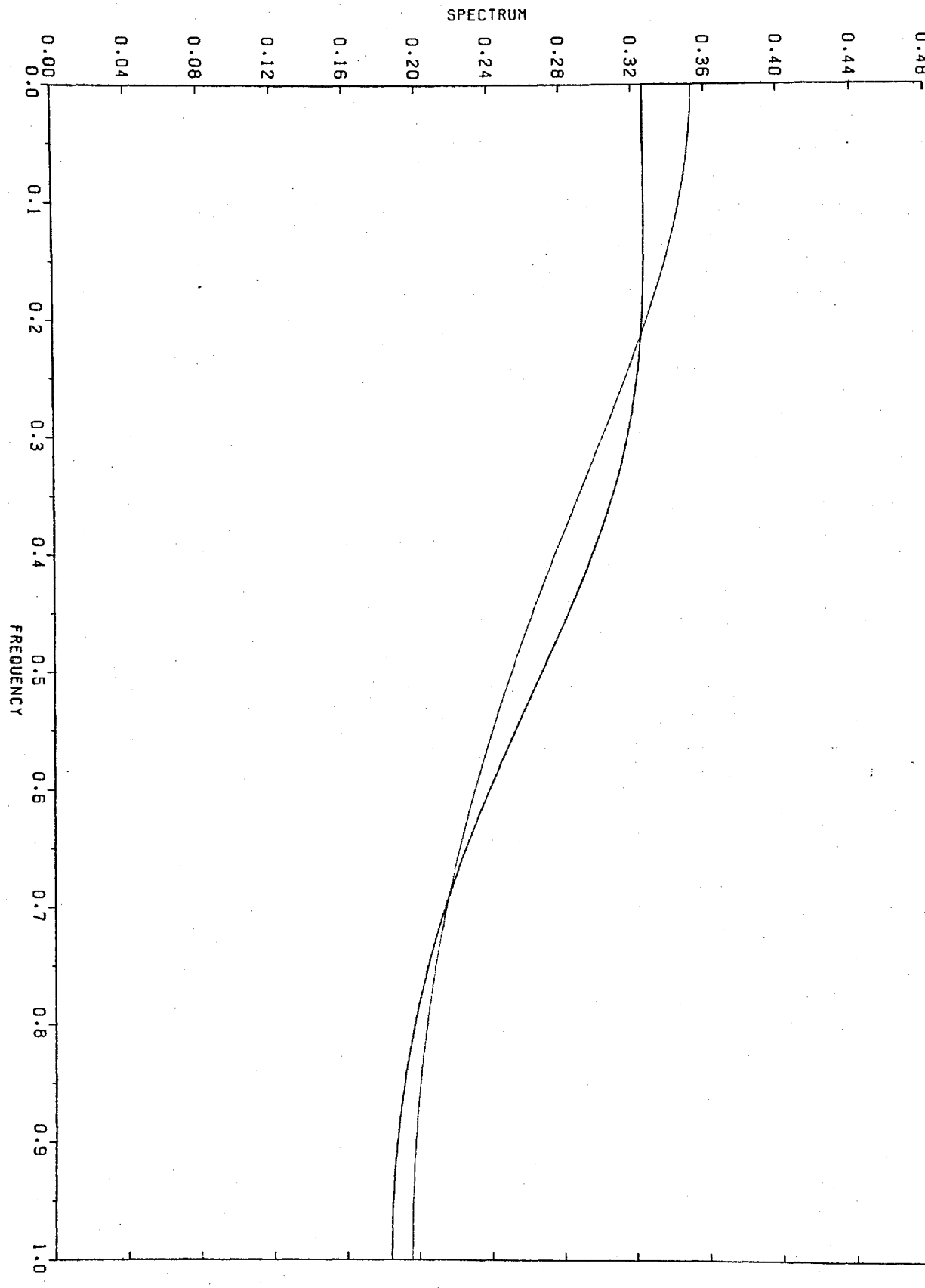

Figure 6.8: Spectrum for process with $\lambda(x) = x + 0.5$ $(-1.5 < x \leq 1.0)$, $= -1.0$ (otherwise), $Z_n \sim N(0,1)$. Frequency in units of π.

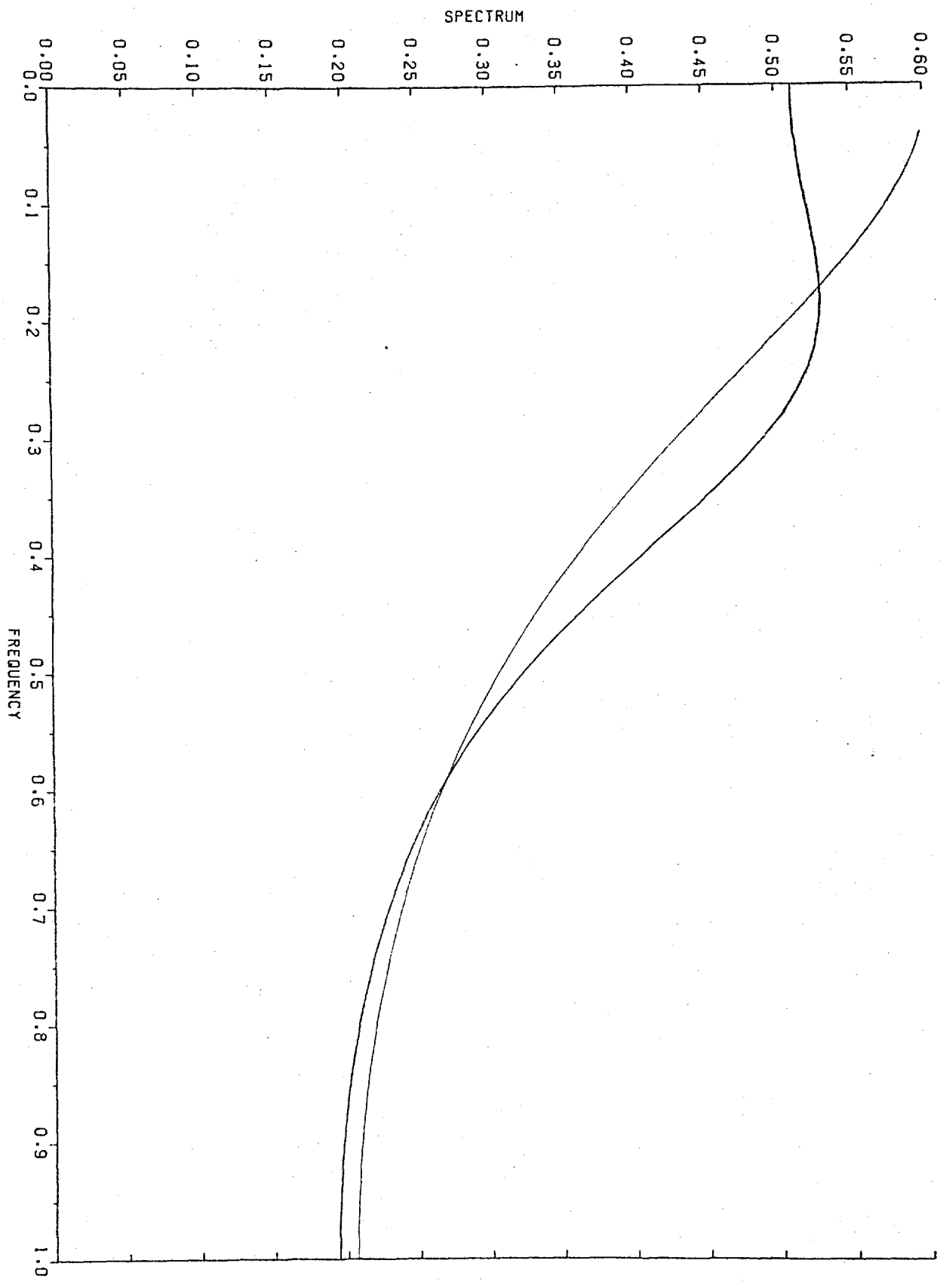

Figure 6.9: Spectrum for process with $\lambda(x) = x + 0.5$ $(-2.0 < x \leq 1.5)$, $= -1.5$ (otherwise), $Z_n \sim N(0,1)$. Frequency in units of π.

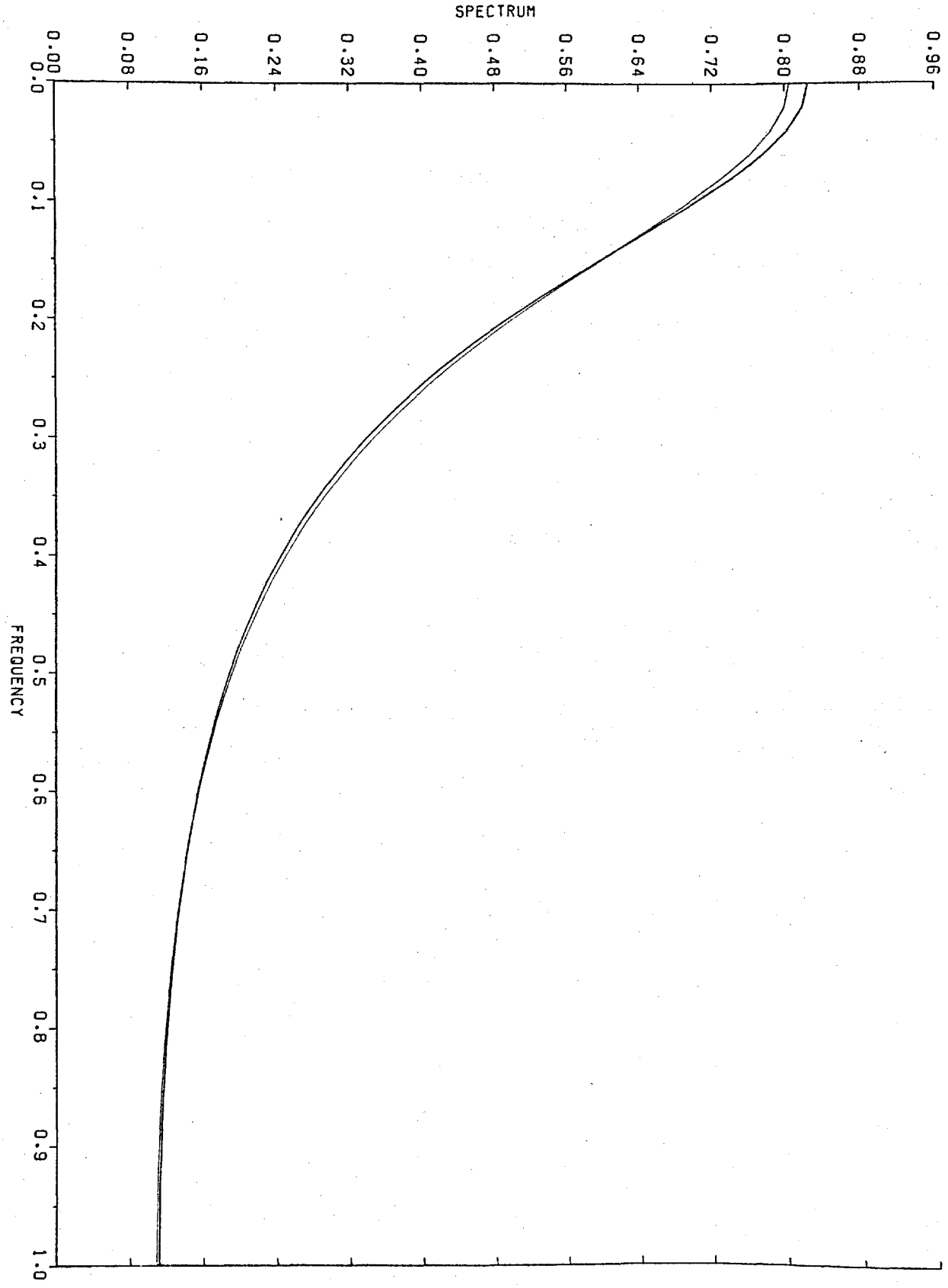

Figure 6.10: Spectrum for process with $\lambda(x) = x$ $(-2.0 < x \leq 2.0)$, $= 0$ (otherwise), $Z_n \sim N(0,1)$. Frequency in units of π.

process, it would be difficult to distinguish between the two processes using only second order moment properties.

The approximations for the covariances used in the calculation of the spectra in figures 6.8-10 are given in Table 6.6. For each of the first two processes the autocovariance function has a damped oscillatory behaviour even though the first correlation is positive. The moments of the first process have been partly checked by simulation (Table 6.3(d)) and simulation and density plots for the second process appear in figure 6.6.

lag	autoregression function		
	λ_1	λ_2	λ_3
0	1.65098	2.21893	1.85998
1	0.24206	0.58111	0.85998
2	-0.02125	0.04384	0.41509
3	-0.0192	-0.0720	0.200
4	-0.00291	-0.0468	0.0967
5	0.52×10^{-3}	-0.0135	0.0467
6	0.28×10^{-3}	0.48×10^{-3}	0.0226
7	0.32×10^{-4}	0.25×10^{-2}	0.0109
8	-0.10×10^{-4}	0.13×10^{-2}	0.53×10^{-2}
9	-0.42×10^{-5}	0.29×10^{-3}	0.25×10^{-2}
10	-0.29×10^{-6}	-0.6×10^{-4}	0.12×10^{-2}
11	0.19×10^{-6}	-0.1×10^{-3}	0.59×10^{-3}
12	0.56×10^{-7}	-0.2×10^{-4}	0.29×10^{-3}
13			0.14×10^{-3}

Table 6.6: Covariances of processes in figures 6.8-10.

7 STATISTICAL TREATMENT OF NON-LINEAR AUTOREGRESSIVE PROCESSES

7.1 Estimation and testing

7.1.1 The theoretical problem of estimation for non-linear autoregressive processes is relatively simple when the process is assumed to be one of a suitably parameterised family of such processes. Maximum likelihood estimators may be used and the usual results about the asymptotic properties of their distributions (as the length of the observed series increases) are available under fairly general conditions.

Suppose that the unknown autoregression function $\lambda(x)$ is one of the family of functions $\lambda(x;\theta)$, with $\lambda(x) = \lambda(x;\theta_0)$, and suppose that the input distribution has a density of the form $f_Z(z;\psi)$. Here θ,ψ may be vector parameters. Suppose that a series of observations $(x_0,\ldots,x_N)$ are obtained from a stationary process whose mechanism is one of the processes (indexed by θ,ψ)

$$X_{n+1} = \lambda(X_n;\theta) + Z_{n+1} \qquad (n = 0,1,\ldots N-1)$$

with $\{Z_1,\ldots Z_N\}$ independent of x_0 and independent amongst themselves with common density $f_Z(z,\psi)$. Two possible ways of analysing this problem are either to treat x_0 as being fixed, effectively taking distributions conditional on this value, or to treat x_0 as having the stationary distribution of the process under consideration. Let $f_X(x;\theta,\psi)$ denote the stationary density. In the first case the likelihood function is

$$\prod_{n=0}^{N-1} f_Z(x_{n+1} - \lambda(x_n;\theta);\psi) \qquad (7.1.1)$$

and in the second there is an additional factor $f_X(x_0;\theta,\psi)$. The contribution of this term is negligible asymptotically and, since in general it would be difficult to evaluate, it may be convenient to discard it for moderately large N. In other situations, for example for a sample consisting of many relatively short sections of independent series, such factors might have considerable importance.

Assuming that maximum likelihood estimators $\hat{\theta},\hat{\psi}$ can be found from the likelihood equations derived from (7.1.1) the question arises as to their distributions. Provided that the problem is regular (Cox and Hinkley, 1974, p.107) these estimators have an asymptotic multivariate normal distribution centred at their true values. The information matrix

associated with the j'th observation is (Cox and Hinkley, 1974, p.300)

$$i_j(\theta,\psi) = \begin{bmatrix} i_{j,\theta\theta} & i_{j,\theta\psi} \\ i_{j,\theta\psi} & i_{j,\psi\psi} \end{bmatrix}$$

$$i_{j,\theta\theta} = -E\{\frac{\partial^2}{\partial z^2} \log f_Z(z;\psi)\}E[\{\frac{\partial}{\partial\theta}\lambda(X_j;\theta)\}^2 | X_0 = x_0]$$

$$i_{j,\theta\psi} = -E\{\frac{\partial^2}{\partial z\partial\psi} \log f_Z(z;\psi)\}E[\frac{\partial}{\partial\theta}\lambda(X_j;\theta) | X_0 = x_0]$$

$$i_{j,\psi\psi} = -E\{\frac{\partial^2}{\partial\psi^2} \log f_Z(z;\psi)\}$$

where it is assumed that θ,ψ are single parameters and where Z denotes a random variable with density $f_Z(z;\psi)$. In forming the total information matrix $i_.(\theta,\psi) = \sum_{j=1}^{N} i_j(\theta,\psi)$, it is convenient to replace the conditional expectations appearing above by expectations treating X_j as if had the stationary distribution. Since the process is stationary this is a good approximation for large N. Hence to terms comparable with N,

$$i_.(\theta,\psi) = N \begin{bmatrix} i_{.\theta\theta} & i_{.\theta\psi} \\ i_{.\theta\psi} & i_{.\psi\psi} \end{bmatrix} \qquad (7.1.2)$$

where

$$i_{.\theta\theta} = -E\{\frac{\partial^2}{\partial z^2} \log f_Z(z;\psi)\}E[\{\frac{\partial}{\partial\theta}\lambda(X;\theta)\}^2],$$

$$i_{.\theta\psi} = -E\{\frac{\partial^2}{\partial z\partial\psi} \log f_Z(z,\psi)\}E[\frac{\partial}{\partial\theta}\lambda(X;\theta)],$$

$$i_{.\psi\psi} = -E\{\frac{\partial^2}{\partial\psi^2} \log f_Z(z;\psi)\},$$

and where X has the stationary density $f_X(x;\theta,\psi)$. The results when θ,ψ are vectors are easily obtained.

For example let the input density be normal with zero mean and variance ψ. Then the information matrix (7.1.2) reduces to

$$N \begin{bmatrix} \psi^{-1} E[\{\frac{\partial}{\partial\theta}\lambda(X;\theta)\}^2] & 0 \\ 0 & \frac{1}{2}\psi^{-2} \end{bmatrix}$$

and hence the asymptotic variance of the estimate of θ is

$$\text{var}(\tilde{\theta}) \simeq \psi/(NE[\{\frac{\partial}{\partial\theta}\lambda(X;\theta)\}^2]) ,$$

and $$\text{var}(\tilde{\psi}) \simeq \frac{2\psi^2}{N} .$$

Also in this case the likelihood equations simplify to

$$\sum_{n=0}^{N-1} \{x_{n+1} - \lambda(x_n;\hat{\theta})\} \frac{\partial}{\partial\theta}\lambda(x_n;\hat{\theta}) = 0 , \qquad (7.1.3)$$

$$\hat{\psi} = \frac{1}{N}\sum_{n=0}^{N-1} \{x_{n+1} - \lambda(x_n;\hat{\theta})\}^2 .$$

In particular, when $\lambda(x;\theta)$ is linear in the parameter θ, $\lambda(x;\theta) = \theta\lambda(x)$ say, then

$$\hat{\theta} = \{\sum_{n=0}^{N-1} x_{n+1}\lambda(x_n)\} / \sum_{n=0}^{N-1} \{\lambda(x_n)\}^2 , \qquad (7.1.4)$$

and

$$\text{var}(\hat{\theta}) \simeq \frac{\psi}{NE[\lambda^2(X)]} .$$

Once the maximum likelihood estimators have been found, their covariance matrix may be approximated by that of the asymptotic distribution given by inverting the matrix (7.1.2). Since this matrix is itself a function of θ and ψ these would usually be replaced by their estimators to produce a quantitative estimate for their variation. To do so involves evaluating expectations of certain functions under the stationary distribution corresponding to $\hat{\theta},\hat{\psi}$. These could possibly be obtained from the algorithm described in Chapter 5. An alternative would be to estimate,

for example, $E[\lambda^2(X)\ ]$ by $\frac{1}{N}\sum_{n=0}^{N-1}\lambda^2(x_n)$, producing as an estimate for the variance of $\hat{\theta}$ in (7.1.4),

$$\text{est}[\text{var}(\hat{\theta})] \;=\; \hat{\psi}\{\sum_{n=0}^{N-1}\lambda^2(x_n)\}^{-1}\ .$$

Such approximations for the estimator variances would then not require the evaluation of the expectations under the stationary distribution.

Thus, to the extent that the process under consideration can be assumed to be of a particular known parametric form, the problem of estimation for non-linear autoregressive processes can be met by maximum likelihood procedures.

7.1.2 Assuming that, for a particular set of data, a non-linear autoregressive process is thought to hold, an appropriate form for the parameterisation of the model must be found. It may first be remarked that the structure of the likelihood function (7.1.1) is exactly that which would be obtained from a regression experiment with observations $x_1,\ldots x_N$ at the values of an independent variable $t_1,\ldots t_N$, with $t_i = x_{i-1}$, for which the following holds

$$x_j \;=\; \lambda(t_j;\theta) + \varepsilon_j \qquad\qquad (j = 1,\ldots,N),$$

with the observation errors ε_j having the density $f_Z(\cdot;\psi)$. This suggests employing least squares estimates for the (vector of) parameters θ. In the scalar case, with $f_Z(\cdot;\psi)$ having zero mean, this leads to the least squares estimate $\tilde{\theta}$ given by

$$\sum_{n=0}^{N-1}\{x_{n+1} - \lambda(x_n;\tilde{\theta})\}\,\frac{\partial}{\partial\theta}\lambda(x_n;\tilde{\theta}) \;=\; 0\ . \qquad (7.1.5)$$

This estimate coincides with the maximum likelihood estimate when the input distribution is normal (compare (7.1.3), (7.1.5)). The comparison with the ordinary regression situation suggests that, if the densities $f_Z(\cdot;\psi)$ are not too far from normal, little is lost in terms of efficiency by using the least-squares estimate rather than the maximum likelihood estimate. The least-squares estimates are always available and do not depend on estimating parameters of the input distribution when $f_Z(\cdot;\psi)$ has zero mean; in fact the only assumptions needed about this distribution

are that it have zero mean and finite variance. Beran (1976) gives an adaptation of the procedure for non-normal distributions and linear processes.

In considering the parameterisation of the autoregression function it would often be natural to ensure that the linear functions (of x) are included amongst the family $\lambda(x;\theta)$ as a special subspace of the parameter space: in this way likelihood ratio tests for the presence of a non-linear part to the autoregression function may be constructed. These tests would be expected to perform well in testing against non-linear functions of the type used in their construction. In the same way as with ordinary regression, the form of the autoregression functions actually fitted would be regarded as being only an approximation to the shape of the "true" autoregression function.

A possible natural choice for the family of functions $\lambda(x;\theta)$ is the family of quadratic (and, more generally, polynomial) functions: with $\theta = (\alpha,\beta,\gamma)$,

$$\lambda(x;\ \alpha,\beta,\gamma) = \alpha + \beta x + \gamma x^2 \ . \qquad (7.1.6)$$

As was shown in section 2.4.4, a process with such an autoregression function, with $\gamma \neq 0$, would be non-stationary for almost all input distributions, exceptions being finite tailed distributions for special ranges of the parameters α, β, γ. However even for non-stationary processes the maximum likelihood estimators may still be defined though the asymptotic properties above no longer hold. Since only a finite number of observations are available, stationarity has little relevance except in providing the asymptotic distributions for the estimates. The function estimated from the family of functions (7.1.6) would anyway be regarded as just an approximation over some central range to the actual autoregression function. Here a possible curvature in the autoregression function is approximated locally by a quadratic function: further polynomial terms could also be included to extend the range behaviour covered.

An alternative approach to testing for non-linearity of the autoregression function would be to attempt to fit it by linear functions over different sections of the range. If the ends of the sections are taken as prespecified constants and if the function is not constrained to be continuous this leads to simple estimates of the slope and intercept parameters of each section and a likelihood ratio test for the equality of

the parameters of the different sections gives a simple test for non-linearity.

Just as in the ordinary regression situation, the calculation of least squares estimates is greatly simplified if the functions $\lambda(x;\theta)$ are chosen to be linear in the parameters in the vector θ, and other considerations concerning the form of the regression functions apply here also. An orthogonalisation of the model might be attempted: for the model (7.1.6) let

$$\bar{x} = \frac{1}{N}\sum_{i=0}^{N-1} x_i, \quad s^2 = \frac{1}{N}\sum_{i=0}^{N-1} (x_i-\bar{x})^2, \quad t = \sum_{i=0}^{N-1} (x_i-\bar{x})^3 / \sum_{i=0}^{N-1} (x_i-\bar{x})^2,$$

then the usual orthogonalisation leads to the model

$$\lambda^*(x;\alpha^*,\beta^*,\gamma^*) = \alpha^* + \beta^*(x-\bar{x}) + \gamma^*[(x-\bar{x})^2 - t(x-\bar{x}) - s^2].$$

While this might be used to calculate the estimates of the parameters in the original model, regarding $\bar{x}$, s^2, t as fixed, and to calculate the estimated autoregression function, the parameters in this model are difficult to interpret since they are connected to the "real" parameters by the sample quantities $\bar{x}$, s^2, t. However the above type of orthogonalisation would help to avoid the occurrence of numerical errors in calculating the least squares estimates. Checking of the model by examination of the residuals, $r_i = x_i - \lambda(x_{i-1},\tilde{\theta})$, can also be extended from the ordinary regression case.

One basis of a non-parametric approach would be to estimate the joint stationary density of (X_n,X_{n+1}) by a suitable kernel estimate formed from the sample $\{(x_i,x_{i+1}); i = 0,\ldots N-1\}$. Here the sample pairs are not independent but the stationarity of the process would enable a consistent estimate to be formed.

7.2 Identification and Prediction

7.2.1 In the above it is assumed that, somehow, it is known that a specific model, namely a first order non-linear autoregressive process should be fitted. Clearly some way of deciding which type of model is to be used is needed and it is natural to try to extend the identification procedure of Box and Jenkins (1970, Ch.6) to fill this requirement. It would be difficult to give a complete description of a procedure which

would purport to distinguish between linear and non-linear autoregressive processes having different orders of dependence on past terms, especially if autoregressive-moving average and non-stationary (integrated) processes of various types are also to be considered as possibilities. Here a simple and fairly obvious generalisation of the Box-Jenkins procedure is given.

The following extended notation for partial correlation coefficients will be used.

$$\rho(X,Y) = \text{corr}(X,Y)$$

$$\rho(X,Y;\, U_1,\ldots U_k) = \text{corr}(X - \alpha_1 U_1 \ldots -\alpha_k U_k,\; Y - \beta_1 U_1 \ldots \beta_k U_k)$$

where $\alpha_1,\ldots \alpha_k$, $\beta_1,\ldots \beta_k$ minimise $\text{var}(X - \alpha_1 U_1 \ldots -\alpha_k U_k)$ and $\text{var}(Y - \beta_1 U_1 \ldots \beta_k U_k)$. Also, for partial regression coefficients

$$\beta(X,Y) = \text{cov}(X,Y)/\text{var}(Y),$$

$$\beta(X,Y;\, U_1,\ldots U_k) = \text{cov}(X, Y-\beta_1 U_1 \ldots -\beta_k U_k)/\text{var}(Y-\beta_1 U_1 \ldots -\beta_k U_k)\ ,$$

and for partial error variances and covariances,

$$\sigma^2(Y) = \sigma(Y,Y) = \text{var}(Y)$$

$$\sigma^2(Y;U_1,\ldots U_k) = \text{var}(Y-\beta_1 U_1 \ldots -\beta_k U_k)$$

$$\sigma(X,Y;U_1,\ldots U_k) = \text{cov}(X-\alpha_1 U_1 \ldots -\alpha_k U_k, Y-\beta_1 U_1 \ldots -\beta_k U_k)$$

where again $\beta_1,\ldots \beta_k$ minimise $\text{var}(Y-\beta_1 U_1 \ldots -\beta_k U_k)$ and $\alpha_1,\ldots \alpha_k$ minimise $\text{var}(X-\alpha_1 U_1 \ldots -\alpha_k U_k)$. This notation is more cumbersome than the usual notation but is more convenient for the purposes here. The following relations hold (Kendall and Stuart, 1973, p.338).

$$\sigma(X,Y;U_1,\ldots U_k) = \sigma(X,Y;U_1,\ldots U_{k-1}) - \frac{\sigma(X,U_k;U_1,\ldots U_{k-1})\,\sigma(Y,U_k;U_1,\ldots U_{k-1})}{\sigma^2(U_k;U_1,\ldots U_{k-1})} \qquad (7.2.1)$$

$$\beta(X,Y;U_1,\ldots U_k) = \sigma(X,Y;U_1,\ldots U_k)/\sigma^2(Y;U_1,\ldots U_k), \tag{7.2.2}$$

$$\rho(X,Y;U_1,\ldots U_k) = \sigma(X,Y;U_1,\ldots U_k)/\{\sigma^2(X;U_1,\ldots U_k)\sigma^2(Y;U_1,\ldots U_k)\}^{1/2} \tag{7.2.3}$$

The above, which are not the usual recurrence formulae for partial correlation coefficients, give a useful form for computation. Formula (7.2.1) is the equivalent to the "sweeping-in" of the variable U_k in a multiple regression situation. An iterpretation of the partial correlation coefficients is given by

$$\sigma^2(X;Y,U_1,\ldots U_k) = \sigma^2(X;U_1,\ldots U_k)\{1 - \rho^2(X,Y;U_1,\ldots U_k)\}. \tag{7.2.4}$$

Hence ρ^2 is the fractional reduction in error variance achieved by including the variable Y in the regression of X on $U_1,\ldots U_k$. The above expressions all hold for any random variables $X,Y,U_1,\ldots U_k$ and not just for multivariate normal distributions.

In a time series context, interest may be in predicting the value X_{n+1} of a stationary series $\{X_i\}$ by a linear function of other variables $U_1^{(n)},\ldots U_k^{(n)}$, where here the index n denotes time and $U_1^{(n)},\ldots U_k^{(n)}$ are random variables for which observed values are available at time n. Then, to minimise the mean square error of prediction, the best linear predictor $\hat{X}_{n+1}$ of X_{n+1} is

$$\hat{X}_{n+1} = \mu_X + \beta_1^{(n)}\{U_1^{(n)} - \mu_1^{(n)}\} + \ldots + \beta_k^{(n)}\{U_k^{(n)} - \mu_k^{(n)}\}$$

where $\mu_X = E(X_n)$, $\mu_i^{(n)} = E(U_i^{(n)})$ $\quad (i = 1,\ldots k)$,

and $\beta_1^{(n)} = \beta(X_{n+1},U_1^{(n)}; U_2^{(n)},\ldots U_k^{(n)})$,

$\beta_2^{(n)} = \beta(X_{n+1},U_2^{(n)}; U_1^{(n)},U_3^{(n)},\ldots U_k^{(n)})$, etc. .

The mean square error of prediction is then $\sigma^2(X_{n+1};U_1^{(n)},\ldots U_k^{(n)})$. The best, not necessarily linear, predictor of X_{n+1} (minimising mean square error) is the conditional expectation $E(X_{n+1}|U_1^{(n)},\ldots U_k^{(n)})$. Though the variables $U_1^{(n)},\ldots U_k^{(n)}$ may be any set of relevant quantities, here they will be restricted to be stationary functions of the past values (up to time n) of the series $\{X_i\}$. In this case the correlation, regression and variance coefficients above are invariant to shifts of time scale.

7.2.2 The Box-Jenkins procedure commences by approximating the covariance structure of the process by the corresponding sample quantities. Usually this is just done for the ordinary covariances, $\mathrm{cov}(X_n, X_{n+k})$, but sample estimates of the quantities $\sigma(U_i^{(n)}, U_j^{(n)}) = \mathrm{cov}(U_i^{(n)}, U_j^{(n)}) = \mathrm{cov}(U_i^{(n+h)}, U_j^{(n+h)})$ can also be formed for any stationary functions $U_j^{(n)}$ of the data series. Here the same notation will be used for the estimated and theoretical values. The above formulae may be applied just as well to the estimated covariances and the quantities $\beta(.,.), \rho(.,.)$ derived by (7.2.1-3) will be estimates of the partial regression and correlation coefficients.

The standard method proceeds by considering the two sequences of coefficients (Box and Jenkins, 1970, p.64),

$$\rho_h = \rho(X_n, X_{n-h}) \qquad (h = 0,1,2,\ldots)$$

$$\phi_{hh} = \rho(X_n, X_{n-h}; X_{n-1}, X_{n-2}, \ldots X_{n-h+1}) \qquad (h = 0,1,2,\ldots) \qquad (7.2.5)$$

The reason for studying the sequence ϕ_{hh} is essentially that, if the process $\{X_n\}$ is generated by a linear autoregressive process giving X_n in terms of $X_{n-1}, \ldots X_{n-p}$ only, plus an additive input, then, for any input distribution, theoretically

$$\phi_{hh} = 0 \qquad (h > p).$$

The problem of deciding on the number of terms to fit in a linear autoregressive process is analogous to the problem of deciding which set of independent variables to include in a multiple regression situation (Box and Jenkins, 1970, p.64). Considering the set of partial correlation coefficients (7.2.5) is really a short-cut to considering more extensive sets of coefficients

e.g. $\rho(X_n, X_{n-h}; X_{n-1}, \ldots X_{n-q}) \qquad (q \geq 1;\ h = q+1, q+2, \ldots)$

or $\rho(X_n, X_{n-h}; X_{n-i_1}, \ldots X_{n-i_q}) \qquad (i_1, \ldots i_q > 0;\ h = 1,2,3,\ldots;\ h \neq i_1, i_2, \ldots i_q).$

These last coefficients give (by their square) the value of including the term X_{n-h} into an existing prediction formula for X_n in terms of $X_{n-i_1}, \ldots X_{n-i_q}$.

Having put the problem into a regression-like situation it is natural to consider introducing polynomial type terms into the prediction equations and this is likely to be appropriate if the process is a non-linear autoregressive process. Clearly terms other than polynomials can be considered. Including polynomial terms in the above procedure is the same as considering prediction formula of the type

$$\hat{X}_n = c + \beta_1 X_{n-1} + \ldots + \beta_p X_{n-p} + \gamma_{11} X_{n-1}^2 + \gamma_{12} X_{n-1} X_{n-2} + \ldots + \gamma_{qq} X_{n-q}^2 + \ldots \tag{7.2.6}$$

and, intuitively at least, such predictors would be expected to perform well if the estimated autoregression function (the right-hand side of (7.2.6)) is close to the true function for values of $(X_{n-1}, X_{n-2}, \ldots)$ which occur most often. Clearly it is perfectly respectable to fit a polynomial predictor even though a process with such an autoregression function would be non-stationary.

It would be possible to decide which subset of a number of lagged linear and polynomial terms to include in the prediction formula by the standard iterative procedures of stepwise regression. However in this time-series context it may be reasonable to suppose that if a term X_{n-q} appears in the prediction equation then so should all of $X_{n-1}, \ldots X_{n-q+1}$, or it may be that an equation containing only terms linear in the series values is to be preferred to one containing non-linear terms. Also a cubic term in a lagged variable would probably not be included unless the same variable appeared as linear and quadratic terms as well. In any case, as far as the prediction mean square error is concerned, the value of including a term into, or excluding a term from, a prediction equation is given by the corresponding partial correlation coefficient or, equivalently (by 7.2.4), by a pair of partial error variances. This would mean considering partial correlation coefficients such as

$$\rho(X_n, X_{n-1}^2;\ X_{n-1}, X_{n-2}, \ldots X_{n-q}),$$

$$\rho(X_n, X_{n-1}^3;\ X_{n-1}, X_{n-2}, \ldots X_{n-q}, X_{n-1}^2),$$

$$\rho(X_n, X_{n-2}^2; X_{n-1}, X_{n-2}, \ldots X_{n-q}),$$

or, starting from a non-linear predictor,

$$\rho(X_n, X_{n-2};\ X_{n-1}, X_{n-1}^2, X_{n-1}^3),$$

$$\rho(X_n, X_{n-3};\ X_{n-1}, X_{n-1}^2, X_{n-1}^3, X_{n-2}) .$$

Two points concerned with the above calculations are, firstly, that when sample values are used for the above procedure it would usually be necessary to adjust the estimates of the prediction variances in the usual way to take account of the number of parameters fitted. Secondly the estimates of the parameters (i.e. the estimated regression coefficients) are not exactly those which would be obtained by the least squares method (Anderson, 1971, §5.4).

8 MULTIDIMENSIONAL PROCESSES

8.1 Introduction

In earlier chapters one dimensional processes have been considered. Here the same methods are applied to multidimensional processes to enable properties of their stationary distributions to be calculated.

Let the process $\{X_n\}$ taking values in a d-dimensional vector space be defined by

$$X_{n+1} = \lambda(X_n) + Z_{n+1} \qquad (n = \ldots -1,0,1,\ldots) \qquad (8.1.1)$$

where $\lambda(\cdot)$ is a d-dimensional vector-valued function of a d-dimensional vector variable and where $\{Z_n\}$ is a sequence of independent and identically distributed d-dimensional vector-valued random variables. The components of each Z_n-vector need not be independent amongst themselves. Much of the discussion of stationarity in Chapter 2 applies also to multidimensional processes so that conditions for the existence of stationary processes with the above representation can be obtained.

Analogues of the methods suggested in Chapter 3 can obviously be developed but here only the generalisation of the successful algorithm obtained by varying the autoregression function is given. There are several ways in which a generalisation of the family of autoregressive processes (4.1.6) may be introduced, for it would be possible to introduce a separate parameter for each dimension. However it seems simplest to work with just a single parameter, and this is done. This may be inappropriate if some dimensions of the autoregression function are exactly linear.

Considering the values of the process as column vectors, a possible family of autoregression functions is

$$\mu(x;\beta) = a + B'x + \beta[\lambda(x) - a - B'x],$$

where a is an arbitrary but fixed d-dimensional column vector and B is a dxd matrix, arbitrary excepting that the initial process (corresponding to $\beta = 0$),

$$X_{n+1}(0) = a + B'X_n(0) + Z_{n+1} \qquad (n = \ldots -1,0,1,\ldots)$$

should be stationary: for well behaved distributions this would usually imply simply the condition $|\det B| < 1$. Properties of the initial stationary distribution can be found using characteristic functions. If $\theta_0(s)$, $\phi_Z(s)$ are the characteristic functions of the initial stationary distribution (of $\{X_n(0)\}$) and of the input distribution respectively, where s is a d-dimensional vector variable, then

$$\theta_0(s) = E\{\exp(is'X_{n+1}(0))\} = E\{\exp(is'a + is'B'X_n(0) + is'Z_{n+1})\}$$

$$= \exp\{ia's\}\,\theta_0(Bs)\phi_Z(s)$$

$$= \exp\{ia'(I-B)^{-1}s\} \prod_{r=0}^{\infty} \phi_Z(B^r s)$$

in exactly the same way as for the one dimensional processes. Even for a normally distributed input it does not seem possible to evaluate this in a simple form for general B (Anderson, 1971, p.181). If the covariance matrix of the input distribution (whether normal or not) is C, that of the initial linear process is $\sum_{r=0}^{\infty} B'^r C\, B^r$.

For the general choice of B it is possible to derive direct analogues of formulae (4.4.1-6) giving the terms in an expansion in powers of β of the stationary characteristic functions $\phi(s;\beta)$ of the processes

$$X_{n+1}(\beta) = a + B'X_n(\beta) + \beta\{\lambda(X_n(\beta)) - a - B'X_n(\beta)\} + Z_{n+1} \quad (n = \ldots -1,0,1,\ldots).$$

(8.1.2)

The derivation is not given here but the formula can be found in exactly the same way as for the one-dimensional processes. Using the notation introduced in the next section this can be done reasonably simply.

8.2 Vector notation

The following notation will be used. For a real or complex one dimensional function f of a vector variable $x = (x_1,\ldots x_d)'$ and a vector r of integers, $r = (r_1,\ldots r_d)'$ $(r_i \geq 0)$, define the symbol for the derivative

$$\frac{\partial^r}{\partial x^r} f(x) = \frac{\partial^{r_1+\ldots+r_d}}{\partial x_1^{r_1} \partial x_2^{r_2} \ldots \partial x_d^{r_d}} f(x_1,\ldots x_d).$$

For a vector-values function $g(x) = (g_1(x), g_2(x),\ldots g_d(x))'$ of a vector variable define, with r as above,

$$\frac{\partial^r}{\partial x^r} g(x) = \{ \frac{\partial^r}{\partial x^r} g_1(x), \ldots\ldots, \frac{\partial^r}{\partial x^r} g_d(x)\}'.$$

Define factorial and combinatorial coefficients for vectors of integers $n = (n_1,\ldots n_d)'$, $r = (r_1,\ldots r_d)'$, by

$$n! = n_1!\, n_2! \ldots n_d!\ ,$$

$$\binom{n}{r} = \frac{n!}{(n-r)!r!} = \binom{n_1}{r_1}\binom{n_2}{r_2}\ldots\binom{n_d}{r_d}\ ,$$

$$= \frac{n_1!n_2!\ldots n_d!}{(n_1-r_1)!\ldots(n_d-r_d)!r_1!\ldots r_d!}\ .$$

Define for vectors $x = (x_1,\ldots x_d)'$, $r = (r_1,\ldots r_d)'$ the powers

$$x^r = x_1^{r_1} x_2^{r_2} \ldots x_d^{r_d} = \prod_{i=1}^{d} x_i^{r_i}\ ,$$

and for matrices $A = \{a_{ij}\}$, $M = \{M_{ij}\}$,

$$A^M = a_{11}^{m_{11}} a_{12}^{m_{12}} \ldots\ldots a_{dd}^{m_{dd}} = \prod_{i,j=1}^{d} a_{ij}^{m_{ij}}\ . \qquad (8.2.1)$$

Since for scalar α and vector x,

$$(\alpha x)^r = (\alpha x_1)^{r_1} \ldots (\alpha x_d)^{r_d} = \alpha^{r_1+\ldots+r_d} x^r,$$

it is natural to define $\alpha^r = \alpha^{r_1+\ldots r_d}$ for $r = (r_1,\ldots r_d)'$ and α scalar. It follows from the above definition that $x^m.x^n = x^{m+n}$ where x,m,n have the same dimension and x^m is scalar.

The following formulae can be shown to hold. Here j and q are d-dimensional vectors of integers, j_1 is a scalar integer, A is a $p\times d$ matrix, r is a p-dimensional vector of integers and x is a d-dimensional (continuous valued) vector.

$$\frac{\partial^q}{\partial x^q} x^j = \begin{cases} \frac{j!}{(j-q)!} x^{j-q} & (j \geq q), \\ 0 & (\text{otherwise}). \end{cases}$$

$$\frac{\partial^q}{\partial x^q} (a'x)^{j_1} = \begin{cases} \frac{j_1!}{(j_1-q'1)!} a^q (a'x)^{j_1-q'1} & (j_1 \geq q'1), \\ 0 & (\text{otherwise}). \end{cases}$$

$$\frac{\partial^q}{\partial x^q} (Ax)^r = \sum_R \frac{q!}{R!} \frac{r!}{(r-R1)!} A^R (Ax)^{r-R1}.$$

This summation is over all $p\times d$ matrices R of non-negative integers satisfying the constraints $R'1 = q$, $R1 \leq r$. Here and elsewhere 1 represents a vector of ones of suitable dimension for multiplication where it occurs. Inequalities between vectors are taken to hold coordinate by coordinate. Similarly if $F(u)$ is a function of a p-dimensional variable

$$\frac{\partial^q}{\partial x^q} F(Ax) = \sum_{R'1=q} \frac{q!}{R!} A^R \frac{\partial^{R1}}{\partial u^{R1}} F(u)\Big|_{u=Ax}.$$

It may be noted that (8.2.1) implies that, for the identity matrix I and the matrix $\emptyset$ composed entirely of zeroes,

$$I^R = \begin{cases} 1 & (\text{if } R \text{ is diagonal}), \\ 0 & (\text{otherwise}), \end{cases}$$

$$\phi^R = \begin{cases} 1 & (R = \phi), \\ 0 & (\text{otherwise}) . \end{cases}$$

For the "ordinary" scalar power it is natural to take

$$\phi^m = \begin{cases} I & (m = 0), \\ \phi & (\text{otherwise}) , \end{cases}$$

and for any matrix A, $A^0 = I$.

For vectors m,n, the symbol

$$\sum_{m=0}^{n} \quad \text{will mean} \quad \sum_{m_1=0}^{n_1} \sum_{m_2=0}^{n_2} \cdots \sum_{m_d=0}^{n_d}$$

and

$$\sum_{m=0}^{\infty} \quad \text{will mean} \quad \sum_{m_1=0}^{\infty} \sum_{m_2=0}^{\infty} \cdots \sum_{m_d=0}^{\infty} .$$

Then it may be checked that, for a function h (possibly vector valued) of a d-dimensional vector,

$$h(x) = \sum_{r=0}^{\infty} \frac{x^r}{r!} \frac{\partial^r}{\partial u^r} h(u) \Big|_{u=0} .$$

Also, for the functions f,g, one of which is scalar valued, the derivative of the product h = fg is given by

$$\frac{\partial^n}{\partial x^n} h(x) = \sum_{r=0}^{n} \binom{n}{r} \frac{\partial^r}{\partial x^r} f(x) \frac{\partial^{n-r}}{\partial x^{n-r}} g(x),$$

where n and r have the same dimension as the variable x.

For two vectors a,t the following expansion holds

$$\exp\{a't\} = \sum_{r=0}^{\infty} \frac{a^r t^r}{r!}$$

and

$$\frac{\partial^r}{\partial t^r} \exp\{a't\} = a^r \exp\{a't\}.$$

Integrals involving a vector variable x will be denoted by

$$\int (\quad) dx = \iint \ldots \int (\quad) dx_1 dx_2 \ldots dx_d$$

and in the following these integrals will always be taken over the entire d-dimensional space. For a probability distribution function $F(x)$ of a d-dimensional variable, the characteristic function is given by

$$\phi(s) = \int \exp\{is'x\}\, dF(x),$$

where s is a d-dimensional vector variable, and the inversion formula

$$f(x) = \frac{1}{(2\pi)^d} \int \exp\{-is'x\}\phi(s)\, ds$$

often holds if $F(x)$ has the density $f(x) = F^{(1)}(x) = \dfrac{\partial^d}{\partial x_1 \ldots \partial x_d} F(x_1, \ldots x_d)$.

8.3 Expansions for distributional quantities

8.3.1 On writing $\Lambda(x)$ for the vector function of vector variable x defined by

$$\Lambda(x) = \lambda(x) - a - B'x,$$

the equations that the characteristic function $\phi(s;\beta)$ of the process $\{X_n(\beta)\}$, given by (8.1.2), must satisfy are

$$\phi(s;\beta) = \int e^{is'x} dF(x;\beta)$$

$$= \phi_Z(s) \int \exp\{is'(a + B'x + \beta\Lambda(x))\} dF(x;\beta). \qquad (8.3.1)$$

Then making an expansion of $\phi(s;\beta)$ similar to (4.4.1), namely

$$\phi(s;\beta) = \theta_0(s) + \sum_{N=1}^{\infty} \beta^N \theta_N(s),$$

the following formulae, defining the coefficient functions $\theta_N(s)$, can be shown to satisfy (8.3.1):

$$\theta_N(s) = \sum_{1\le j'1\le N} \sum_{\ell=0}^{\infty} (iB^{\ell}s)^{j} T^{(N,j)} (B^{\ell+1}s)\rho_{\ell}(s) \qquad (N = 1,2,3,\ldots) \tag{8.3.2}$$

$$T^{(N,j)}(s) = \frac{1}{j!(2\pi)^d} \int L_j(u)\theta_{N-j'1}(s-u)du \qquad (1 \le j'1 \le N),$$

where

$$\theta_0(s) = \exp\{ia'(I-B)^{-1}s\} \prod_{r=0}^{\infty} \phi_Z(B^r s)$$

$$\rho_{\ell}(s) = \prod_{r=0}^{\ell} \exp\{ia'B^r s\}\phi_Z(B^r s) \qquad (\ell = 0,1,2,\ldots)$$

and

$$L_j(u) = \int \Lambda^j(x) e^{iux} dx \qquad (j'1 \ge 1).$$

Here the quantities N, ℓ and r are scalar integers, whereas j is a vector of integers. For example $T^{(N,j)}(s)$ may be written more fully as

$$T^{(N,j)}(s) = T^{(N,j_1,j_2,\ldots j_d)}(s_1,s_2,\ldots s_d) \qquad (1 \le j_1+j_2+\ldots+j_d \le N).$$

As before, important simplifications take place when B is chosen to have special values. There appear to be two distinct cases when this happens. In the first B is identically zero and in the second B is non-zero, but such that the powers B^{ℓ} are zero for $\ell \ge \ell_1 > 1$ say. In either case the infinite summation in (8.3.2) reduces to a finite summation. Only the first case will be treated here although the second might also be useful.

Exactly as in section 4.4.5 the above formulae can be reduced to the following when $B = \emptyset$.

$$\phi(s;\beta) = \theta_0(s)h(s;\beta),$$

$$\text{where} \quad h(s;\beta) = 1 + \sum_{N=1}^{\infty} \sum_{1\le j'1\le N} \beta^N (is)^{j} T^{(N,j)} \tag{8.3.3}$$

$$T^{(N,j)} \begin{cases} P_j^{(0)} & (j'1 = N) \\ \sum_{1\le n'1\le N-j'1} T^{(N-j'1,n)} P_j^{(n)} & (1 \le j'1 < N) \end{cases}$$

and

$$P_j^{(n)} = \frac{1}{j!}\frac{\partial^n}{\partial y^n}\left[\int \{\Lambda(x+y)\}^j \, dF_0(x)\right]_{y=0} . \qquad (8.3.4)$$

Here $\theta_0(s) = \exp\{is'a\}\phi_Z(s)$ and $F_0(x) = F_Z(x-a)$.

As in section 4.5.2, formulae for the joint characteristic function $\phi_k(s,t;\beta)$ of $(X_n(\beta), X_{n+k}(\beta))$ can be obtained. In the general case $(B \neq \emptyset)$,

$$\phi_k(s,t;\beta) = \theta_{0,k}(s,t) + \sum_{N=1}^{\infty} \beta^N \theta_{N,k}(s,t),$$

where

$$\theta_{0,k}(s,t) = \begin{cases} \exp\{it'a\}\phi_Z(t)\theta_{0,k-1}(s,Bt) & (k \geq 1) \\ \theta_0(s+t) & (k = 0) \end{cases}$$

$$\theta_{N,0}(s,t) = \theta_N(s+t) \qquad (N \geq 1)$$

$$\theta_{N,k}(s,t) = e^{it'a}\phi_Z(t)\{\theta_{N,k-1}(s,Bt) + \sum_{1 \leq j'1 \leq N} (it)^j R_k^{(N,j)}(s,Bt)\} \qquad (N,k \geq 1)$$

$$R_k^{(N,j)}(s,t) = \frac{1}{j!(2\pi)^d}\int L_j(u)\theta_{N-j'1,k-1}(s,t-u)\,du \qquad (1 \leq j'1 \leq N;\ k \geq 1).$$

When $B = \emptyset$, the formulae corresponding to (4.5.12-3) are

$$\phi_k(s,t;\beta) = \theta_0(s)\theta_0(t)\left\{h(s;\beta) + \sum_{N=1}^{\infty} \sum_{1 \leq j'1 \leq N} \beta^N (it)^j T_k^{(N,j)}(s)\right\}$$

$$T_k^{(N,j)}(s)=\begin{cases} \begin{cases} Q_j^{(0)}(0,s) & (j'1 = N) \\ & (k = 1), \\ \sum\limits_{1\le n'1\le N-j'1} T^{(N-j'1,n)} \sum\limits_{q=0}^{n} \binom{n}{q} (is)^q Q_j^{(n-q)}(0,s) & (1 \le j'1 < N) \end{cases} \\ \begin{cases} P_j^{(0)} & (j'1 = N) \\ & (k \ge 2). \\ P_j^{(0)} \sum\limits_{1\le n'1\le N-j'1} (is)^n T^{(N-j'1,n)} + \sum\limits_{1\le n'1\le N-j'1} T_{k-1}^{(N-j'1,n)}(s) P_j^{(n)} & \\ & (1 \le j'1 < N) \end{cases} \end{cases}$$

Here

$$Q_j(y,s) = \frac{1}{j!\theta_0(s)} \int e^{is'x} \Lambda^j (x+y)\, dF_0(x)$$

and $$Q_j^{(n)}(0,s) = \frac{\partial^n}{\partial y^n} Q_j(y,s)\Big|_{y=0}.$$

Clearly the methods of section 4.7 could also be employed and it follows that $h(s;\beta)$ given by (8.3.3) is the stationary characteristic function of the multidimensional process given by

$$Y_{n+1}(\beta) = \beta\Lambda\{ Y_n(\beta) + a + Z_{n+1}\} \qquad (n = \ldots -1,0,1,\ldots) \qquad (8.3.5)$$

$$= \beta\Lambda\{ X_n(\beta)\}.$$

8.3.2 From the above formulae, expressions for moments and joint moments of the stationary processes can be derived, just as in Chapter 5. For the choice $B = \emptyset$, the following hold.

$$E\{Y_n^r(\beta)\} = E\{Y_{n,1}^{r_1}(\beta) Y_{n,2}^{r_2}(\beta) \ldots Y_{n,d}^{r_d}(\beta)\},$$

$$= r!h^{(r,0)},$$

where

$$h^{(r,0)} = \begin{cases} 1 & (r'1 = 0), \\ \sum_{N=r'1}^{\infty} \beta^N T(N,r) & (r'1 \geq 1). \end{cases} \tag{8.3.6}$$

$$E\{X_n^r(\beta)\} = r! \sum_{q=0}^{r} \theta_0^{(r-q)}(0) h^{(q,0)},$$

$$\text{where } \theta_0^{(r)}(s) = \frac{(-i)^r}{r!} \frac{\partial^r}{\partial s^r} \theta_0(s). \tag{8.3.7}$$

$$E\{X_n^{\ell}(\beta) X_{n+k}^{m}(\beta)\} = \ell!m! \sum_{p=0}^{\ell} \sum_{q=0}^{m} \theta_0^{(\ell-p)}(0) \theta_0^{(m-q)}(0) h_k^{(p,q)} \tag{8.3.8}$$

where

$$h_k^{(\ell,m)} = \sum_{N=m'1}^{\infty} \beta^N T_k^{(N,m)}(\ell)$$

and where, for $\ell'1 = 0$, $T_k^{(N,m)}(\ell) = T^{(N,m)}$, and for $\ell'1 \geq 1$,

$$T_k^{(N,j)}(\underline{\ell}) \begin{cases} \begin{cases} Q_j^{(0,\ell)} & (j'1 = N) \\ \sum_{1 \leq n'1 \leq N-j'1} \sum_{q=0}^{\min(n,\ell)} \binom{n}{q} T^{(N-j'1,n)} Q_j^{(n-q,\ell-q)} & (1 \leq j'1 < N) \end{cases} & (k = 1), \\ \begin{cases} 0 & (j'1 = N) \\ P_j^{(0)} T^{(N-j'1,\ell)} + \sum_{1 \leq n'1 \leq N-j'1} T_{k-1}^{(N-j'1,n)}(\ell) P_j^{(n)} & (1 \leq j'1 < N) \end{cases} & (k \geq 2), \end{cases}$$

where $T^{(N-j'1,\ell)} = 0 \quad (\ell'1 > N-j'1)$,

$$Q_j^{(n,\ell)} = \frac{(-i)^{\ell}}{j!\,\ell!} \frac{\partial^{n+\ell}}{\partial y^n \partial s^{\ell}} \left\{ \frac{1}{\theta_0(s)} \int e^{is'x} \Lambda^j(x+y)\, dF_0(x) \right\}_{y,s=0}$$

and where $\min(n,\ell) = \{\min(n_1,\ell_1), \min(n_2,\ell_2), \ldots \min(n_d,\ell_d)\}'$. These are all

exactly analogous with the formulae for the one-dimensional case.

Moments and joint moments corresponding to a choice $B \neq \emptyset$ can be found from the expressions for the characteristic functions using the formulae for derivatives given earlier. The expressions obtained are then of a form equivalent to (5.2.2) and, like these, appear impractical for use unless special circumstances hold.

8.4 An Example

In order to demonstrate the interpretation of the above notation consider the 2-dependent process $\{X_n\}$, assumed to be stationary, defined by

$$X_{n+1} = \lambda_1(X_n) + \gamma X_{n-1} + Z_{n+1} \qquad (n = \ldots -1,0,1,\ldots) \; . \qquad (8.4.1)$$

This can be written as a multidimensional process $\{(X_{n,1},X_{n,2})'\}$ generated by

$$X_{n+1,1} = \lambda_1(X_{n,1}) + \gamma X_{n,2} + Z_{n+1}$$

$$X_{n+1,2} = X_{n,1}$$

where $\{Z_n\}$ has the distribution function $G_Z(z)$ and characteristic function $\psi_Z(s)$, say. This process is equivalent to (8.1.1) with $d = 2$, $\lambda(x)$ given by

$$\lambda(x) = \begin{bmatrix} \lambda_1(x_1,x_2) \\ \lambda_2(x_1,x_2) \end{bmatrix} = \begin{bmatrix} \lambda_1(x_1) + \gamma x_2 \\ x_1 \end{bmatrix} \; ,$$

and with $F_Z(z_1,z_2)$ having a degenerate second component,

$$F_Z(z_1,z_2) = \begin{cases} 0 & (z_2 < 0), \\ G_Z(z_1) & (z_2 \geq 0). \end{cases}$$

Then choosing $a = (a_1,a_2)'$,

$$\Lambda_1(x_1,x_2) = \lambda_1(x_1) + \gamma x_2 - a_1, \quad \Lambda_2(x_1,x_2) = x_1 - a_2 ,$$

and

$$F_0(x) = F_0(x_1,x_2) = \begin{cases} 0 & (x_2 < a_2), \\ G_Z(x_1 - a_1) & (x_2 \geq a_2). \end{cases}$$

Also $\phi_Z(s_1,s_2) = \psi_Z(s_1)$ and

$$\theta_0(s_1,s_2) = \exp\{i(s_1a_1 + s_2a_2)\}\psi_Z(s_1) ,$$

so that $\theta_0^{(r_1,r_2)}(0,0)$ defined by (8.3.7) is given by

$$\theta_0^{(r_1,r_2)}(0,0) = \frac{a_2^{r_2}}{r_2!} \cdot \frac{(-i)^{r_1}}{r_1!} \frac{\partial^{r_1}}{\partial s_1^{r_1}} \{e^{is_1a_1}\psi_Z(s_1)\}_{s_1=0} .$$

The quantities $P_j^{(n)}$ of (8.3.4) are given by

$$P_{j_1,j_2}^{(n_1,n_2)} = \frac{1}{j_1!j_2!} \frac{\partial^{n_1+n_2}}{\partial y_1^{n_1}\partial y_2^{n_2}} \left[\int \{\lambda_1(x_1+y_1)+\gamma(x_2+y_2)-a_1\}^{j_1}(x_1+y_1-a_2)^{j_2} dF_0(x_1,x_2)\right]_{y_1=y_2=0}$$

$$= \frac{1}{j_1!j_2!} \frac{\partial^{n_1+n_2}}{\partial y_1^{n_1}\partial y_2^{n_2}} \left[\int \{\lambda_1(x_1+a_1+y_1)+\gamma(a_2+y_2)-a_1\}^{j_1}(x_1+a_1+y_1-a_2)^{j_2} dG_Z(x_1)\right]_{y_1=y_2=0},$$

$$= \frac{\gamma^{n_2}}{j_2!(j_1-n_2)!} \frac{\partial^{n_1}}{\partial y^{n_1}} \left[\int \{\lambda_1(x_1+a_1+y_1)+\gamma a_2-a_1\}^{j_1-n_2}(x_1+a_1+y_1-a_2)^{j_2} dG_Z(x_1)\right]_{y_1=0}$$

$$= \begin{cases} \frac{j_1!}{(j_1-n_2)!}\gamma^{n_2} P_{j_1-n_2,j_2}^{(n_1,0)} & (n_2 \leq j_1) \\ 0 & (\text{otherwise}). \end{cases}$$

The formula for $T^{(N,j)}$ now is

$$T^{(N,j_1,j_2)} = \begin{cases} P_{j_1,j_2}^{(0,0)} & (j_1+j_2 = N), \\ \sum_{1 \le n_1+n_2 \le N-j_1-j_2} T^{(N-j_1-j_2,n_1,n_2)} P_{j_1,j_2}^{(n_1,n_2)} & (1 \le j_1+j_2 < N). \end{cases}$$

and $h^{(r,0)}$ given by (8.3.6) is

$$h^{(r_1,r_2,0,0)} = \begin{cases} 1 & (r_1,r_2 = 0), \\ \sum_{N=r_1+r_2}^{\infty} \beta^N T^{(N,r_1,r_2)} & (r_1+r_2 \ge 1). \end{cases}$$

It may be noted that moments of the process (8.4.1) can be obtained either from the moments of the first component process $\{X_{n,1}(1)\}$ or from the moments of the transformed process $\{Y_{n,2}(1)\}$ of (8.3.5),

$$Y_{n+1,2}(\beta) = \beta\Lambda_2\{X_{n,1}(\beta), X_{n,2}(\beta)\} = \beta\{X_{n,1}(\beta) - a_2\}.$$

The first method leads to

$$E\{X_n^{r_1}\} = E\{X_{n,1}^{r_1}(1)\} = r_1! \sum_{q_1=0}^{r_1} \theta_0^{(r_1-q_1,0)} h^{(q_1,0,0,0)}$$

while the second leads to

$$E\{(X_n - a_2)^{r_2}\} = E\{Y_{n,2}^{r_2}(1)\} = r_2! h^{(0,r_2,0,0)},$$

where the numbers h are formed with $\beta = 1$. It is not obvious that these two expressions are equivalent. From the pairwise moments of the two dimensional process, given by (8.3.8), it is possible to find any moments of the form

$$E\{X_n^{\ell_1} X_{n-1}^{\ell_2} X_{n+k}^{m_1} X_{n+k-1}^{m_2}\} \qquad (\ell_1,\ell_2,m_1,m_2 \ge 0;\ k \ge 1)$$

for the process $\{X_n\}$ given by (8.4.1).

References

Anderson, T.W. (1971). The Statistical Analysis of Time Series. New York, Wiley.

Atkinson, A.C. and Pearce, M.C. (1976). The computer generation of Beta, Gamma and Normal random variables. J.R. Statist. Soc., A, 139.

Beran, R. (1976). Adaptive estimates for autoregressive processes. Ann. Inst. Statist. Math., 28, 77-89.

Box, G.E.P. and Jenkins, G.M. (1970). Time Series Analysis, Forecasting and Control. San Francisco, Holden-Day.

Cox, D.R. and Hinkley, D.V. (1974). Theoretical Statistics. London, Chapman and Hall.

Cox, D.R. and Miller, H.D. (1965). Theory of Stochastic Processes. London, Methuen.

Doob, J.L. (1953). Stochastic Processes. New York, Wiley.

Feller, W. (1971). An Introduction to Probability Theory and Its Applications. Vol.2, 2nd edition. New York, Wiley.

Foguel, S.R. (1966). Limit theorems for Markov processes. Trans. Amer. Math. Soc., 121, 200-209.

Foguel, S.R. (1968). Existence of a σ-finite invariant measure for a Markov process on a locally compact space. Israel J. Math., 6, 1-5.

Harris, T.E. (1956). The existence of stationary measures for certain Markov processes. Proc. 3rd Berkeley Symp. Math. Statist. Prob., 2, 113-124.

Kawata, T. (1972). Fourier Analysis in Probability Theory. New York, Academic Press.

Kendall, M.G. and Stuart, A. (1969). Advanced Theory of Statistics. Vol.1, 3rd edition. London, Griffin.

Kendall, M.G. and Stuart, A. (1973). Advanced Theory of Statistics. Vol.2, 3rd edition. London, Griffin.

Kesten, H. (1976). Recurrence criteria for multidimensional Markov Chains and multidimensional linear birth and death processes. Adv. Appl. Prob., 8, 58-87.

Lukacs, E. (1970). Characteristic Functions. 2nd edition. London, Griffin.

Lukacs, E. (1975). Stochastic Convergence. 2nd edition. New York, Academic Press.

Mallows, C.L. (1967). Linear processes are nearly Gaussian. J. Appl. Prob., 4, 313-329.

Marsaglia, G., MacLaren, M.G. and Bray, T.A. (1964). A fast procedure for generating normal random variables. Commun. Ass. Comp. Mach., 7, 4-10.

May, R.M. (1976). Simple mathematical models with very complicated dynamics. Nature, 261, 459-467.

Rosenblatt, M. (1971). Markov Processes, Structure and Asymptotic Behaviour. Berlin, Springer-Verlag.

Titchmarsh, E.C. (1939). Theory of Functions. Oxford University Press.

Titchmarsh, E.C. (1948). Introduction to the Theory of Fourier Integrals. 2nd edition. Oxford University Press.

Tweedie, R.L. (1974). R-theory for Markov Chains on a general state space (I). Ann. Prob., 2, 840-864.

Tweedie, R.L. (1975a). Sufficient conditions for egodicity and recurrence of Markov Chains on a general state space. Stoch. Proc. Appl., 3, 385-403.

Tweedie, R.L. (1975b). The robustness of positive recurrence and recurrence of Markov Chains under perturbations of the transition probabilities. J. Appl. Prob., 12, 744-752.

Tweedie, R.L. (1976). Criteria for classifying general Markov Chains. (To appear.)

Wintner, A. (1947). The Fourier Transforms of Probability Distributions. Baltimore.