Time Series Analysis for Directional Data

Peter S. Craig

Thesis submitted for the degree of Ph.D.
Trinity College Dublin
Dept. of Statistics

September, 1988

I declare that this thesis has not been submitted as an exercise for a degree at any university other than the University of Dublin.

I declare that this thesis is entirely my own work.

I agree that the Library of Trinity College Dublin may lend or copy this thesis upon request.

(Peter Craig)

Summary of thesis

This thesis is an account of some aspects of time series analysis for directional data (or, more strictly, circular data), which is an almost totally unexplored area of statistics. The thesis is in four chapters.

The first concerns a family of models for directional time series which is naturally derived from the ARMA family of time series models. The identification problem for the family is discussed and an analogue of the auto-correlation function is defined. The remainder of the chapter is devoted to estimation of that analogue with major attention being given to showing that the estimators used are both consistent and asymptotically normal.

The second chapter examines the estimation problem in detail for the simplest model from the family introduced in the first chapter. A form of moment estimation is described and its asymptotic properties derived. The majority of the chapter is devoted to maximum likelihood estimation. Maximum likelihood estimation is shown to to be consistent and asymptotically normal. The asymptotic properties are quantified and shown to be superior to those for moment estimation, and the chapter closes with a discussion of the computational problems involved in performing maximum likelihood estimation for the model.

The third chapter deals with a number of aspects of Markov models for directional time series. Most of the chapter is given to a discussion of the various bivariate circular distributions to be found in the literature, while stationarity, higher order models, and estimation properties are also considered.

The final chapter is a trial data analysis for a sequence of wind directions. Two useful diagnostic techniques are introduced. The analysis proceeds from the models of the first chapter to the Markov models of the third chapter and the chapter conclude with an attempt to model some of the seasonal behaviour apparent in the data.

Acknowledgements

First and foremost, I wish to acknowledge the help and support of my supervisor, Adrian Raftery, who has put up with so much and who has complained so little. Without his spur and his guidance, I would never have finished this work.

I wish also to acknowledge the kindness and help of all in the Dept. of Statistics at the University of Washington, but most especially Doug Martin, who helped make my visit possible and who helped provide financial support.

My thanks to the Mathematics Dept. of TCD for the use of their typesetting equipment.

Finally, my thanks to all my family and friends for their support.

This work was partially supported, during the summer of 1986, by the Office of Naval Research under contract no. N00014-84-C-0169.

Contents

List of Figures

Chapter 0

Introduction

The title of this thesis is "Time Series Analysis for Angular Data". It might more accurately be described as an initial investigation of this area. While time series analysis has become part of mainstream statistics, relatively little work has been done on non-linear models or models for unusual kinds of data. This parallels the development of non-dependent statistics where only relatively recently has attention turned to statistics for directional or other unusual data. However, directional statistics has been the focus of considerable attention in the last decade, and it seems likely that this is only the beginning in terms of the exploration of new areas of statistical modelling.

Background

The origins of this interest in directional statistics lie in the rich diversity of situations in the physical, biological and earth sciences where such data arise naturally — magnetism in physics, molecular orientation and growth in chemistry, the directions of rock fracture and hence of veins of mineralisation in geology, migrational and homing behaviour of animals in biology (see [1] for many biological examples). Other areas also give rise less frequently to directional phenomena; the time of day at which something occurs may be thought of as a circular variable. It has also been suggested to me that the error involved in truncating a number to a computer representation might usefully be considered to be a circular variable.

Many situations where directions arise naturally do not require the application of special directional statistical methods. For example the direction of magnetic North at a point on the earth's surface is constrained to lie within a few degrees of true North as defined by the Pole star. In this case there is no reason why the small part of the circle which contains observations should not be identified with part of the real line and conventional statistical tools applied. On the other hand, if the whole circle is rich with data then this identification would not be reasonable. To perform the identification it would be necessary to cut the circle at some point in order to lay it flat on the line. The problem is then that points which are close to each other on different sides of the cut would be identified with points a long way apart on the real line. My reason for discussing these two cases is that the work which has been done for directional data can be divided into two categories, depending on which case it deals with. From the point of view of new techniques interesting problems are those where the data to be modelled is truly circular, in the sense that it fills the whole circle, as there is no real need for time series models for circular data which is localised on the circle since the data can be transformed to the real line and normal techniques applied.

Examples of time series of this kind are far less obvious than examples of general directional data. However, this must to some extent be due to the lack of techniques for their analysis, leading researchers to transform data into some other space prior to analysis. For example, most people are sufficiently familiar with the sine and cosine functions that they can transform their data to the real line before attempting analysis. The flaw with that approach is the consequent loss of information. In fact when dealing with time series that approach may fail simply because of the paucity of time series models for data confined to a finite interval. There are some fairly obvious examples where time series of angles arise — the variation of the earth's magnetic field at a fixed location on its surface and the time of day at which peak traffic congestion occurs in a city. The example with which I have chosen to work is the variation of wind direction. This has certain advantages, over other possibilities, in that there is a large quantity of such data readily available and that the data are of high quality, not being flawed by requiring subjective interpretation in the process of measurement. It can be argued that wind directions are a somewhat artificial example, since the wind speed is usually observed together with the direction, thus making possible the construction of velocity vectors which could be analysed as a two-dimensional time series. The latter approach has severe problems of its own. Firstly analysis of multivariate time series is by no means easy, requiring techniques which are more difficult to use than those for univariate time series analysis — the cross spectrum, for example, is much more difficult to interpret than the ordinary spectrum. Secondly, it has been the experience of others that this velocity data does not fit very well with the assumptions of the available two-dimensional linear models.

Time series analysis for directional is an almost totally unexplored area. A certain amount of work has been done in the area of stochastic processes of this type (see [22, 31, 33]). The most notable result is the fact that any random walk on the circle has the uniform distribution as its marginal distribution. This has important implications for time series analysis, rendering the simplest family of models useful only in very limited circumstances.

Dependence measures

Having established the background and need for circular time series analysis, I shall turn to the difficulties involved. The essence of modelling time series is to establish the nature of the dependence between observations at different times. For non-directional data the most commonly used measure of dependence is correlation which measures the degree of linear dependence between variables. In the late 1970's and early 1980's there was a brief flurry of papers ([10, 17, 18, 23, 28, 29]) proposing "correlation" measures for circular data. Most of these were based on the idea of developing circular analogues of the ordinary correlation. It would be unfair to dismiss these proposals, but it is well to point out that there are a number of flaws in the general approach adopted.

All of these authors approach the problem as being one of finding an all-purpose "correlation" measure. No doubt this is due to the almost exclusive use of correlation for measuring dependence between real variables. Strictly speaking, correlation is only a good measure when the nature of the dependence is linear and the random behaviour is Gaussian, i.e. the variables have a bivariate normal distribution. Fortunately, many phenomena fall broadly into that category of dependence. The classical reason given for this is that most phenomena are the result of multiple influences and the central limit theorem indicates that this gives rise to Gaussian behaviour. Where two variables do not have a bivariate normal distribution, other measures of dependence are appropriate. In fact the measure of dependence should be derived for the family of joint distributions under consideration. The latter problem has been discussed in [19]. Where does this leave those who would advocate a standard measure of dependence for circular random variables ? Only a few families of bivariate circular distributions have been proposed. There has not yet been sufficient shared experience to suggest which, if any, of the families might be a standard. The circle does not have a central limit theorem so there is no help available from that quarter. One possible argument is that most directional phenomena arise as projections onto the circle of two-dimensional variables. If those variables are themselves subject to multiple influences, they might have bivariate normal distributions, and this would give rise to a generalisation of the offset normal distribution as a plausible standard family of bivariate circular distributions. The essential point is, however, that there is not yet sufficient evidence to allow a decision in favour of any one family of distributions.

The more specific criticisms which I would make of the proposed "correlation" measures lie in two areas. Firstly they are mostly based on the concept of linearity. Linearity is a very restrictive relationship between circular variables. It is only possible to take linear combinations of circular variables if the coefficients are integers. Secondly, nearly all involve the use of trigonometric functions. As noted earlier, the sine or cosine of an angle contains considerably less information than the angle itself. The conclusion I draw from these facts and the preceding paragraph is that it is more important to start with models from which can be derived the required measures of dependence for identification and other purposes. I have, in fact made extensive use of the measure proposed in [29] as an identification tool for the models discussed in chapter 1. I make no claim that it is any way optimal, but it has the virtue of being mathematically tractable for that family of models.

Outline of thesis

The material discussed divides naturally into two sections each of which subdivides naturally into two chapters. The first two chapters are a discussion of a family of models derived from the ARMA models for conventional time series. The first chapter is a general examination of the family with particular emphasis on the problem of identification within the family. The second chapter is a detailed examination of the estimation problem for the simplest model from the family. The last two chapters deal with Markov models for circular time series. The third chapter contains a general analysis of stationarity, the available families of bivariate distributions, a method for developing higher order Markov models and the asymptotic behaviour of maximum likelihood estimation for Markov models. The final chapter is a description of a trial analysis of a sequence of wind directions. It also contains a description of a general technique for diagnostic analysis of fitted models. Finally there is a short appendix of referenced material drawn from various areas of probability theory.

Chapter 1 exploits the idea of wrapping (well known in circular statistics) to introduce a family of models derived from the ARMA family. The measure of dependence proposed in [29] is used to define a circular version of the auto-correlation function. The asymptotic behaviour of this function is derived in detail, and it is shown that it can be used to identify wrapped moving-average models. The possibility of defining an analogue to the partial auto-correlation function is discussed briefly, but it is clear that a different kind of approach will be required for that problem.

Chapter 2 is a detailed analysis of the simplest model from the family introduced in chapter 1 — the wrapped first order auto-regressive model. Moment estimation is discussed briefly, mainly for the purpose of comparison with maximum likelihood estimation. A detailed understanding of the behaviour of the non-wrapped process conditional upon the wrapped process is developed as a basis for the proofs of strong convergence and asymptotic normality which follow for the maximum likelihood estimates. Numerical values for the covariance structure of the maximum likelihood estimates are computed and compare favourably with those obtained for moment estimation. Finally there is a short account of the computational problems actually encountered when computing and maximising the likelihood.

Chapter 3 is a discussion of Markov models. I commence by considering the existence problem for stationary processes having a given transition probability function and show that under certain mild conditions there is a unique stationary process. I then consider the suitability, for use in Markov models, of a number of families of bivariate distributions which have been proposed in the literature. In many contexts data exhibits more than first order dependence and I discuss some properties of a family of higher order Markov models called linear conditional probability models. These models have the attractive feature that they can be defined in terms of first order models. Finally I show that under certain useful conditions on the transition function maximum likelihood estimation for Markov models is consistent and asymptotically normal.

Chapter 4 is a discussion of a trial analysis of a sequence of wind directions. Two useful methods for diagnostic analysis of models are also introduced. Daily periodic behaviour is observed in the sequence and related to physical processes. For the remainder of the analysis, I work with daily averages. A sequence of Markov models is fitted, followed by a sequence of higher order Markov models using the technique described in chapter 3. There is clear seasonal variation in the sequence and a seasonal model is fitted with limited success in accounting for the seasonality.

Conclusions

Clearly the subject of time series analysis for directional data is only at a beginning. I have only considered the simplest case of circular data which is, in certain ways, a special case. There are new problems of several kinds for the general case. Most importantly, the technique of wrapping is only available for circular data. This means that there is no easy way to generate non-Markovian models for spherical time-series. In fact even Markov models are difficult. The loss of wrapping robs us of large families of bivariate distributions which are available for the circle. Even more fundamental difficulties exist than the lack of models. The problem of how best to display a spherical time series is not an easy one. Totally new techniques will be needed for general directional data.

There is no evidence yet indicating that the wrapped models of chapters 1 and 2 will be of any practical use. It does certainly seem reasonable that some natural circular phenomena might arise via "wrapping". Chapter 1 indicates that, if such phenomena exists, the theory of some aspects of the appropriate wrapped model is not excessively difficult. Chapter 2 is in many ways more interesting, despite the fact that at first it would appear to be much more limited. I believe it to be reasonably obvious that the estimation properties proved in chapter 2 for the wrapped AR(1) could easily be extended to the wrapped AR(n). However, this is not the real interest. The theory developed for the conditional behaviour of the unwrapped process given the wrapped process has potentially wide-ranging implications. It seems likely that it could be extended to apply to a large class of processes which are functions of Markov processes. Obviously not all functions are appropriate, but it should be a large class of functions. Equally, not all Markov processes would be appropriate. A minimum requirement would be stationarity. Almost certainly any Markov process having a strong contractive property like that of the AR(1) could replace it as the process being wrapped. It seems likely also that regularity conditions could be established so as to extend the asymptotic normality proof to a large class of these general "wrapped" processes. The form of the derivatives of the log-likelihood function would retain many of the features of those for the AR(1), in particular the fact that they involve conditional moments of stationary sums of functions.

Some parts of chapter 3 extend naturally to a large group of spaces in which time series might take values. The stationarity properties of Markov processes depend only on the fact that the circle is compact. The proof of consistency and asymptotic normality for maximum likelihood estimation is also based largely on the fact that the circle is compact. The definition of linear conditional probability models is totally general.

Chapter 4 raises far more problems than it solves. Seasonality is very difficult to deal with satisfactorily. Linear conditional probability models almost certainly do not have a sufficiently wide range to cope with the forms of dependence are be observed. However, the real failure of the analysis is probably that it does not model the first order dependence adequately. There is a real need for new families of bivariate distributions. There are some encouraging aspects to the chapter. Spectral analysis is available as a tool for circular time series, as it is for the ordinary case. This enables the detection of periodic effects as was found to be the case for the wind data. The diagnostic approach of the conditional cumulative distribution sequence is available for time series analysis in any space. Admittedly its distributional properties are not known and appear intractable, but it can still be used as a qualitative tool. Further its properties for any given model can be obtained by simulation as was done for the final model fitted in chapter 4.

The real key to time series analysis for circular data must be to find physical processes which generate natural families of models for time-series. If this can be done there will surely be close similarities between the circular and more general situations. Difficult problems are seasonality and long-term variation. These are difficult in all areas of time series analysis, but especially for directional data where the concept of linearity is extremely limited. Solutions to these which are found in other contexts will, however, help to indicate solutions for the directional case. Ultimately, all comes down to the need for more data analysis in order to develop the insight required to make progress.

Notation

$P[A]$	denotes the probability of the event A
$E[X]$	denotes the expectation of the random variable X
$D[X]$	denotes the variance of the random variable X
$C[X,Y]$	denotes the covariance of the random variables X and Y
$\rho[X,Y]$	denotes the correlation between the random variables X and Y
$III[X,Y]$	denotes $E\Big[(X - E[X])(Y - E[Y])(Z - E[Z])\Big]$
f	with a subscript denotes the (possibly joint) probability density function of the random variables in the subscript.
F	with a subscript denotes the (possibly joint) cumulative distribution function of the random variables in the subscript.
$\mathcal{F}_m^n(X)$	where X is a stochastic process denotes the sigma-algebra generated by the random variables $X_m, X_{m+1}, \ldots, X_n$
$\mathbf{x}$	a bold-face letter denotes a vector (possibly infinite dimensional)
$\mathbf{x}_m^n$	a bold-face letter with subscript and superscript generally denotes the vector of values $x_m, x_{m+1}, \ldots, x_n$
$\mathbf{R}$	denotes the space of real numbers (with an integer superscript denotes the product space of that many copies of the real line)
$\mathbf{Z}$	denotes the space of integers
$\mathbf{N}$	denotes the space of positive integers
$O(x)$ as $x \to a$	denotes any quantity g(x) for which there exists C such that $\overline{\lim}_{x \to a} g(x)/x \le C$
∂_x	denotes the partial (or absolute) derivative with respect to x

Chapter 1

Wrapped Linear Processes

In this chapter I introduce a family of models for circle-valued time-series, derived from the familiar ARMA models for real valued time-series. The problem of identification for these models is considered using an approach similar to the use of the auto-correlation function for ARMA models, and statistics are defined for performing the identification. Formulae are derived for the asymptotic distribution of these statistics and its implications are illustrated for models derived from white noise and moving average models. Finally consideration is given to the problems of finding an analogue for the partial auto-correlation function.

1.1 Wrapped Models

One of the earliest methods used to obtain distributions for circle-valued random variables was by "wrapping" a real-valued random variable. The name originates in the idea of wrapping the real line around the circle (anti-clockwise) — like thread on a spool. Thus a given value of the circle-valued random variable arises from an infinite number of values of the real-valued random variable. Let us assume throughout what follows that the circle is chosen to have unit circumference and that values of a circle-valued variable are given as distances along the circumference anti-clockwise from some origin 0. Then, if X is the circle-valued random variable and Y is the real-valued random variable,

$$f_X(x) = \sum_{j \in \mathbf{Z}} f_Y(x + j)$$

if the wrapping is done so that 0 on the circle coincides with 0 in $\mathbf{R}$. An alternative way of understanding this is that X is a partial observation of Y where we see only the fractional part of Y and lose the integer part. i.e.

$$X = Y - [Y]$$

This suggests an obvious way to obtain time series models for the circle. If Y_t is a real-valued stochastic process, we can obtain a circle-valued stochastic process by wrapping each co-ordinate so that

$$X_t = Y_t - [Y_t]$$

and hence

$$f_{(X_{t_1},\ldots,X_{t_n})}(x_{t_1},\ldots,x_{t_n}) = \sum_{\mathbf{j}_1^n \in \mathbf{Z}^n} f_{(Y_{t_1},\ldots,Y_{t_n})}(x_{t_1}+j_1,\ldots,x_{t_n}+j_n)$$

thus defining a model for a circular-valued time series.

1.2 Wrapped Gaussian Linear Models

In this section we will consider the sub-class of wrapped models obtained from Gaussian linear models. Denote by Y_t a zero mean Gaussian linear process. i.e.

$$Y_t = \sum_{j=0}^{\infty} \alpha_j \epsilon_{t-j} \qquad \text{where} \quad \alpha_j \in \mathbf{R}, \text{ and the } \alpha_j \text{ are summable}$$

and the ϵ_t form a sequence of independent Gaussian random variables with mean 0 and variance σ^2. Further, suppose that Y_t is a stationary sequence, i.e.

$$\sum_j |\alpha_j|^2 < \infty$$

and that Y_t is invertible, in other words we can write

$$\epsilon_t = \sum_{j=0}^{\infty} \beta_j X_{t-j}$$

for some real summable sequence β_j. Now let X_t be the corresponding "wrapped" process, i.e.

$$X_t = Y_t - [Y_t]$$

One of the problems that occurs with ARMA models is to identify which model from the family is appropriate. The usual tools are the auto-correlation function and the partial auto-correlation function.

1.2.1 Circular auto-correlation function

In the attempt to find a circular analogue of the auto-correlation function we encounter the problem of defining correlation between circle-valued random variables. While many alternatives have been proposed, as discussed in the introduction, no concensus has yet emerged as to which, if any, is the best. For the purposes of this section I have chosen the one which has the simplest form and which, more importantly, is easy to calculate for wrapped linear models. This is the measure proposed in [29] and is given by

$$\tilde{\rho}(X_1, X_2) = \frac{E[\sin 2\pi(X_1 - \tilde{\mu}(X_1)) \sin 2\pi(X_2 - \tilde{\mu}(X_2))]}{\left\{E[\sin^2 2\pi(X_1 - \tilde{\mu}(X_1))]E[\sin^2 2\pi(X_2 - \tilde{\mu}(X_2))]\right\}^{\frac{1}{2}}} \tag{1.1}$$

where $\tilde{\mu}(X)$ is the circular mean given by $\tilde{\mu}(X) =$ the direction of $E[e^{2\pi iX}]$. When X_1 and X_2 have zero circular mean this becomes

$$\tilde{\rho}(X_1, X_2) = \rho(\sin 2\pi X_1, \sin 2\pi X_2)$$

This suggests the definition of the circular covariance of zero-mean circle-valued random variables by

$$\tilde{C}[X_1, X_2] = E[\sin(2\pi X_1) \sin(2\pi X_2)]$$

The question is now whether the measure in (1.1) provides a useful means for identification of zero-mean wrapped ARMA processes. Define the circular auto-correlation function $\rho_{C,j}$ by

$$\rho_{C,j} = \tilde{\rho}(X_t, X_{t-j}) = \frac{\gamma_{C,j}}{\gamma_{C,0}}$$

where $\tilde{\gamma}_j$ is the circular auto-covariance function given by

$$\begin{aligned} \gamma_{C,j} &= \tilde{C}[X_t, X_{t-j}] = E[\sin(2\pi X_t) \sin(2\pi X_{t-j})] \\ &= E\Big[\sin(2\pi Y_t) \sin(2\pi Y_{t-j})\Big] \\ &= -\tfrac{1}{4}E\left[(e^{2\pi iY_t} - e^{-2\pi iY_t})(e^{2\pi iY_{t-j}} - e^{-2\pi iY_{t-j}})\right] \end{aligned} \tag{1.2}$$

But since the ϵ_t are i.i.d. $N(0,1)$

$$\begin{aligned} E\left[e^{2\pi iY_t} e^{2\pi iY_{t-j}}\right] &= E\left[e^{2\pi i(Y_t + Y_{t-j})}\right] \\ &= E\left[\exp\Big(2\pi i[\textstyle\sum_{k=0}^{\infty} \alpha_k \epsilon_{t-k} + \sum_{k=0}^{\infty}(\alpha_k + \alpha_{k+j})\epsilon_{t-j-k}]\Big)\right] \end{aligned}$$

$$
\begin{aligned}
&= \prod_{k=0}^{j-1} E[e^{2\pi i \alpha_k \epsilon_{t-k}}] \prod_{k=0}^{\infty} E[e^{2\pi i(\alpha_k+\alpha_{k+j})\epsilon_{t-j-k}}] \\
&= \prod_{k=0}^{j-1} e^{-\frac{1}{2}\sigma^2\alpha_k^2(2\pi)^2} \prod_{k=0}^{\infty} e^{-\frac{1}{2}\sigma^2(\alpha_k+\alpha_{j+k})^2(2\pi)^2} \\
&= \exp\Big(-2\pi^2\sigma^2\{2\textstyle\sum_{k=0}^{\infty}\alpha_k^2 + 2\sum_{k=0}^{\infty}\alpha_k\alpha_{j+k}\}\Big)
\end{aligned}
$$

Similarly

$$
E\left[e^{2\pi i Y_t}e^{-2\pi i Y_{t-j}}\right] = \exp\Big(-2\pi^2\sigma^2\{2\textstyle\sum_{k=0}^{\infty}\alpha_k^2 - 2\sum_{k=0}^{\infty}\alpha_k\alpha_{k+j}\}\Big)
$$

Thus

$$
\begin{aligned}
\gamma_{C,j} &= -\tfrac{1}{2}\exp\left(-4\pi^2\sigma^2\textstyle\sum_{k=0}^{\infty}\alpha_k^2\right) \\
&\qquad \times \left\{\exp\left(-4\pi^2\sigma^2\textstyle\sum_{k=0}^{\infty}\alpha_k\alpha_{k+j}\right) - \exp\left(4\pi^2\sigma^2\textstyle\sum_{k=0}^{\infty}\alpha_k\alpha_{k+j}\right)\right\}
\end{aligned}
$$

But

$$
\begin{aligned}
\gamma_j &= E[Y_tY_{t-j}] = E\Big[\sum_{k=0}^{\infty}\alpha_k\epsilon_{t-k}\sum_{l=0}^{\infty}\alpha_l\epsilon_{t-j-l}\Big] \\
&= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\alpha_k\alpha_l E[\epsilon_{t-k}\epsilon_{t-j-l}] = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\alpha_k\alpha_l\sigma^2\delta_{t-k,t-j-l} \\
&= \sum_{l=0}^{\infty}\alpha_l\alpha_{j+l}\sigma^2
\end{aligned}
$$

Therefore

$$
\gamma_{C,j} = e^{-4\pi^2\gamma_0}\sinh 4\pi^2\gamma_j
$$

and

$$
\rho_{C,j} = \frac{\sinh 4\pi^2\gamma_j}{\sinh 4\pi^2\gamma_0}
$$

This compares to $\rho_j = \frac{\gamma_j}{\gamma_0}$. Thus there is a clear similarity between ρ_j and $\rho_{C,j}$ and as one would expect this is especially pronounced when the auto-covariances are small since then the level of information loss through "wrapping" is low. In particular, when Y_t is an MA(k) process (i.e. $\alpha_j = 0$ for $j > k$), then $\gamma_j = 0$ for $j > k$. This means that $\rho_j = \rho_{C,j} = 0$ when $j > k$. This would appear to provide the same basis for identification in either case given the existence of satisfactory estimators for the $\rho_{C,j}$.

1.2.2 Estimation of the circular auto-covariance function

I have chosen to estimate the circular auto-covariance function of the sample $x_1, \ldots, x_n$ by $g_{C,j}$ where

$$g_{C,j} = \frac{1}{n-j} \sum_{k=0}^{n-j} \sin(2\pi x_k) \sin(2\pi x_{k+j})$$

which from (1.2) is obviously an unbiased estimate of $\gamma_{C,j}$. In order to derive various asymptotic properties of the $\{g_{C,j}\}$ we shall need some knowledge of their covariance structure. Now

$$C[g_{C,j}, g_{C,k}]$$

$$= \tfrac{1}{(n-j)(n-k)} \sum_{t=j+1}^{n} \sum_{u=k+1}^{n} E\left[\sin 2\pi X_t \sin 2\pi X_{t-j} \sin 2\pi X_u \sin 2\pi X_{u-k}\right] - \gamma_{C,j}\gamma_{C,k}$$

But

$$E\left[\ \sin 2\pi X_t \sin 2\pi X_{t-j} \sin 2\pi X_u \sin 2\pi X_{u-k}\right]$$

$$= \sum_{\eta_1^4 \in \{-1,1\}^4} \frac{\eta_1\eta_2\eta_3\eta_4}{(2i)^4} E\Big[\exp(2\pi i\{\eta_1 X_t + \eta_2 X_{t-j} + \eta_3 X_u + \eta_4 X_{u-k}\})\Big] \tag{1.3}$$

Suppose that $u < t - j$, then

$$E\Big[\exp(2\pi i\{\eta_1 X_t + \eta_2 X_{t-j} + \eta_3 X_u + \eta_4 X_{u-k}\})\Big]$$

$$\begin{aligned}
= \quad & E\Bigg[\exp\Bigg(2\pi i\Big\{\sum_{m=0}^{j-1} \eta_1^2 \alpha_m \epsilon_{t-m} + \sum_{m=0}^{t-j-u-1} (\eta_1\alpha_{m+j} + \eta_2\alpha_m)^2 \epsilon_{t-j-m} \\
& + \sum_{m=0}^{k-1} (\eta_1\alpha_{m+t-u} + \eta_2\alpha_{m+t-j-u} + \eta_3\alpha_m)^2 \epsilon_{u-m} \\
& + \sum_{m=0}^{\infty} (\eta_1\alpha_{m+t+k-u} + \eta_2\alpha_{m+t+k-j-u} + \eta_3\alpha_{m+k} + \eta_4\alpha_m)^2 \epsilon_{u-k-m}\Big\}\Bigg)\Bigg]
\end{aligned}$$

$$\begin{aligned}
= \quad & \exp\Bigg(-2\pi^2\sigma^2\Big\{\eta_1^2 \sum_{m=0}^{\infty} \alpha_m^2 + \eta_2^2 \sum_{m=0}^{\infty} \alpha_m^2 + \eta_3^2 \sum_{m=0}^{\infty} \alpha_m^2 + \eta_4^2 \sum_{m=0}^{\infty} \alpha_m^2 \\
& +2 \sum_{m=0}^{\infty} \eta_1\eta_2\alpha_m\alpha_{m+j} + 2 \sum_{m=0}^{\infty} \eta_1\eta_3\alpha_m\alpha_{m+t-u} \\
& +2 \sum_{m=0}^{\infty} \eta_2\eta_3\alpha_m\alpha_{m+t-j-u} + 2 \sum_{m=0}^{\infty} \eta_3\eta_4\alpha_m\alpha_{m+k} \\
& + 2 \sum_{m=0}^{\infty} \eta_2\eta_4\alpha_m\alpha_{m+t-j-u+k} + 2 \sum_{m=0}^{\infty} \eta_1\eta_4\alpha_m\alpha_{m+t-u+k}\Big\}\Bigg)
\end{aligned}$$

$$= \exp\Big(-4\pi^2\{2\gamma_0 + \eta_1\eta_2\gamma_j + \eta_1\eta_3\gamma_{t-u} + \eta_2\eta_3\gamma_{t-j-u}$$

$$+\eta_3\eta_4\gamma_k + \eta_2\eta_4\gamma_{t-j-u+k} + \eta_1\eta_4\gamma_{t-u+k}\}\Big)$$

If $t > u > t-j$ we obtain

$$E\Big[\exp(2\pi i\{\eta_1 X_t + \eta_2 X_{t-j} + \eta_3 X_u + \eta_4 X_{u-k}\})\Big]$$

$$\begin{aligned} = \exp\Big(-4\pi^2\{2\gamma_0 + \eta_1\eta_3\gamma_{t-u} & + \eta_1\eta_2\gamma_j + \eta_3\eta_2\gamma_{u-t+j} \\ & + \eta_2\eta_4\gamma_{t-j-u+k} + \eta_3\eta_4\gamma_k + \eta_1\eta_4\gamma_{t-u+k}\}\Big) \end{aligned}$$

which is the same since $\gamma_j = \gamma_{-j}$.

Consider those terms in (1.3) for which $\eta_1\eta_2 = \eta_3\eta_4 = 1$. We have

$$\begin{aligned} e^{-4\pi^2(\gamma_j+\gamma_k)}\Big\{ & \exp(-4\pi^2\{\gamma_{t-u} + \gamma_{t-u+k} + \gamma_{t-j-u} + \gamma_{t-j-u+k}\}) \\ + & \exp(-4\pi^2\{\gamma_{t-u} + \gamma_{t-u+k} + \gamma_{t-j-u} + \gamma_{t-j-u+k}\}) \\ + & \exp(-4\pi^2\{-\gamma_{t-u} - \gamma_{t-u+k} - \gamma_{t-j-u} - \gamma_{t-j-u+k}\}) \\ + & \exp(-4\pi^2\{-\gamma_{t-u} - \gamma_{t-u+k} - \gamma_{t-j-u} - \gamma_{t-j-u+k}\})\Big\} \end{aligned}$$

$$= 4e^{-4\pi^2(\gamma_j+\gamma_k)}\cosh\Big(4\pi^2(\gamma_{t-u} + \gamma_{t-u+k} + \gamma_{t-j-u} + \gamma_{t-j-u+k})\Big)$$

By applying the same method to the other terms in (1.3), the right-hand side becomes

$$\begin{aligned} \tfrac{1}{4}e^{-8\pi^2\gamma_0}\Big\{ & e^{-4\pi^2(\gamma_j+\gamma_k)}\cosh(4\pi^2(\gamma_{t-u} + \gamma_{t-u+k} + \gamma_{t-j-u} + \gamma_{t-j-u+k})) \\ - & e^{-4\pi^2(\gamma_j-\gamma_k)}\cosh(4\pi^2(-\gamma_{t-u} + \gamma_{t-u+k} - \gamma_{t-j-u} + \gamma_{t-j-u+k})) \\ - & e^{-4\pi^2(-\gamma_j+\gamma_k)}\cosh(4\pi^2(-\gamma_{t-u} - \gamma_{t-u+k} + \gamma_{t-j-u} + \gamma_{t-j-u+k})) \\ + & e^{-4\pi^2(-\gamma_j-\gamma_k)}\cosh(4\pi^2(\gamma_{t-u} - \gamma_{t-u+k} - \gamma_{t-j-u} + \gamma_{t-j-u+k})) \quad \Big\} \end{aligned}$$

Also

$$\begin{aligned} \gamma_{C,j}\gamma_{C,k} &= e^{-4\pi^2\gamma_0}e^{-4\pi^2\gamma_0}\sinh 4\pi^2\gamma_j \sinh 4\pi^2\gamma_k \\ &= \tfrac{1}{4}e^{-8\pi^2\gamma_0}\left\{e^{-4\pi^2(\gamma_j+\gamma_k)} - e^{-4\pi^2(\gamma_j-\gamma_k)} - e^{-4\pi^2(-\gamma_j+\gamma_k)} + e^{-4\pi^2(-\gamma_j-\gamma_k)}\right\} \end{aligned}$$

and $\cosh x - 1 = \frac{1}{2}\{e^x + e^{-x} - 2\} = \frac{1}{2}(e^x - e^{-x})^2 = 2\sinh^2\frac{1}{2}x$. Therefore

$$E\Big[\sin 2\pi X_t \sin 2\pi X_{t-j} \sin 2\pi X_u \sin 2\pi X_{u-k}\Big]$$

$$-E\Big[\sin 2\pi X_t \sin 2\pi X_{t-j}\Big] E\Big[\sin 2\pi X_u \sin 2\pi X_{u-k}\Big]$$

$$\begin{aligned} = \tfrac{1}{2} e^{-8\pi^2\gamma_0} \Big\{ & \quad e^{-4\pi^2(\gamma_j+\gamma_k)} \sinh^2(2\pi^2(\gamma_{t-u} + \gamma_{t-u+k} + \gamma_{t-j-u} + \gamma_{t-j-u+k})) \\ & - \; e^{4\pi^2(\gamma_k-\gamma_j)} \sinh^2(2\pi^2(-\gamma_{t-u} + \gamma_{t-u+k} - \gamma_{t-j-u} + \gamma_{t-j-u+k})) \\ & - \; e^{4\pi^2(\gamma_j-\gamma_k)} \sinh^2(2\pi^2(-\gamma_{t-u} - \gamma_{t-u+k} + \gamma_{t-j-u} + \gamma_{t-j-u+k})) \\ & + \; e^{4\pi^2(\gamma_j+\gamma_k)} \sinh^2(2\pi^2(\gamma_{t-u} - \gamma_{t-u+k} - \gamma_{t-j-u} + \gamma_{t-j-u+k})) \quad \Big\} \end{aligned}$$

Suppose that $j \leq k$. Let $h(t, u)$ be some function of two variables which in fact only depends on $t - u$ so that we can actually write $h(t - u)$. Then

$$\begin{aligned} \sum_{u=k+1}^{n} \sum_{t=j+1}^{n} h(t,u) &= \sum_{u=k+1}^{n} \sum_{t=j+1}^{n} h(t-u) \\ &= \sum_{u=k+1}^{n} \sum_{t-u=j+1-u}^{n-u} h(t-u) \\ &= \sum_{u=k+1}^{n} \sum_{d=j+1-u}^{n-u} h(d) \end{aligned} \tag{1.4}$$

From figure 1.1 we can see that when $d = 0, -1, \ldots, -j-k$, that value of d occurs $n-k$ times; when $d = 1, 2, \ldots, n-k-1$, that value occurs $n-k-d$ times; and when $d = j-k-1, j-k-2, \ldots, j-n+1$, that value occurs $n-k-(j-k-d) = n-j+d$ times, i.e. any given value of d occurs $\min(n-k, n-k-d, n-j+d)$ times. So (1.4) becomes

$$\sum_{d=j-n+1}^{n-k-1} \min(n-k, n-k-d, n-j+d) h(d)$$

So we have now shown that

$$E[g_{C,j} g_{C,k}] - E[g_{C,j}]E[g_{C,k}]$$

$$= \tfrac{1}{(n-j)(n-k)} \sum_{t=j+1}^{n} \sum_{u=k+1}^{n} E[\sin 2\pi X_t \sin 2\pi X_{t-j} \sin 2\pi X_u \sin 2\pi X_{u-k}] - \gamma_{C,j}\gamma_{C,k}$$

	$n-k-1$	1	0	. . .	0	0
	$n-k-2$	1	1	. . .	0	0
	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
	2	1	1	. . .	0	0
	1	1	1	. . .	1	0
	0	1	1	. . .	1	1
d	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
	$j-k$	1	1	. . .	1	1
	$j-k-1$	0	1	. . .	1	1
	$j-k-2$	0	0	. . .	1	1
	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
	$j+2-n$	0	0	. . .	1	1
	$j+1-n$	0	0	. . .	0	1
		$k+1$	$k+2$	. . .	$n-1$	n
				u		

Figure 1.1: Diagram indicating which values of d occur for which values of u

$$= \frac{1}{(n-j)(n-k)} \sum_{d=j-n+1}^{n-k-1} \min(n-k, n-k-d, n-j-d) \tfrac{1}{2} e^{-8\pi^2\gamma_0}$$

$$\begin{aligned} \times \quad \Big\{ & e^{-4\pi^2(\gamma_j+\gamma_k)} \sinh^2(2\pi^2(\gamma_d + \gamma_{d+k} + \gamma_{d-j} + \gamma_{d-j+k})) \\ & - e^{4\pi^2(\gamma_k-\gamma_j)} \sinh^2(2\pi^2(-\gamma_d + \gamma_{d+k} - \gamma_{d-j} + \gamma_{d-j+k})) \\ & - e^{4\pi^2(\gamma_j-\gamma_k)} \sinh^2(2\pi^2(-\gamma_d - \gamma_{d+k} + \gamma_{d-j} + \gamma_{d-j+k})) \\ & + e^{4\pi^2(\gamma_j+\gamma_k)} \sinh^2(2\pi^2(\gamma_d - \gamma_{d+k} - \gamma_{d-j} + \gamma_{d-j+k})) \quad \Big\} \end{aligned}$$

which is the first part of the following theorem.

Theorem 1.1 *a)*

$$C[g_{C,j}, g_{C,k}] = \frac{e^{-8\pi^2\gamma_0}}{2(n-j)(n-k)} \sum_{d=j-n+1}^{n-k-1} \min(n-k, n-k-d, n-j+d)$$

$$\begin{aligned} \Big\{ & e^{-4\pi^2(\gamma_j+\gamma_k)} \sinh^2(2\pi^2(\gamma_d + \gamma_{d+k} + \gamma_{d-j} + \gamma_{d-j+k})) \\ & - e^{4\pi^2(\gamma_k-\gamma_j)} \sinh^2(2\pi^2(-\gamma_d + \gamma_{d+k} - \gamma_{d-j} + \gamma_{d-j+k})) \\ & - e^{4\pi^2(\gamma_j-\gamma_k)} \sinh^2(2\pi^2(-\gamma_d - \gamma_{d+k} + \gamma_{d-j} + \gamma_{d-j+k})) \\ & + e^{4\pi^2(\gamma_j+\gamma_k)} \sinh^2(2\pi^2(\gamma_d - \gamma_{d+k} - \gamma_{d-j} + \gamma_{d-j+k})) \Big\} \end{aligned}$$

b) Suppose that $\alpha_j \leq O(j^{-3/2})$ *as* $j \to \infty$*. Then* $C[g_{C,j}, g_{C,k}] \leq O(n^{-1})$ *as* $n \to \infty$*.*

Proof: We have already shown part a). Thus, we have

$$\begin{aligned} C[g_{C,j}, g_{C,k}] &= \frac{e^{-8\pi^2\gamma_0}}{2(n-j)(n-k)} \sum_{d=j-n+1}^{n-k-1} \min(n-k, n-k-d, n-j-d) T_d \\ &\leq \frac{e^{-8\pi^2\gamma_0}}{2(n-j)} \sum_{d=j-n+1}^{n-k-1} T_d \end{aligned}$$

where the definition of T_d is obvious from the statement of part a) of the theorem.

Since $\alpha_j \leq O(j^{-3/2})$, there exists α^* such that $\alpha_j \leq \alpha^*/j^{3/2}$. Therefore we have

$$k\gamma_k = k \sum_j \sigma^2 \alpha_j \alpha_{j+k} \leq \sum_j \frac{k\sigma^2\alpha^{*2}}{j^{3/2}(j+k)^{3/2}} \leq \sum_j \frac{\sigma^2\alpha^{*2}}{j^{3/2}j^{1/2}} \leq \sum_j \frac{\sigma^2\alpha^{*2}}{j^2} = C < \infty$$

and so $\gamma_k = O(k^{-1})$ as $k \to \infty$ which implies that $\gamma_d \pm \gamma_{d+k} \pm \gamma_{d-j} \pm \gamma_{d-j+k} \leq O(d^{-1})$ as $d \to \infty$. Also $\sinh(x) = O(x)$ as $x \to 0$. Combining these gives

$$T_d \leq O(d^{-2}) \qquad \text{as} \quad d \to \infty$$

which implies $\left|\sum_{d=-\infty}^{\infty} T_d\right| < \infty$ and therefore that

$$C[g_{C,j}, g_{C,k}] \leq O(n^{-1})$$

as required.

$\mathcal{Q.E.D.}$

We now know the covariance structure of the estimators of the circular auto-covariance function. The remainder of this section is devoted to showing that the $g_{C,j}$ are, asymptotically, normally distributed, which then completes our knowledge of their asymptotic distribution. The key to this is the use of a central limit theorem for certain strongly mixing stationary processes. The next three proofs show that, under mild conditions on the α_j, the Y_t process is strongly mixing with coefficients which die out sufficiently rapidly. I believe that this has probably been shown before but, in the absence of a suitable reference, I have chosen to derive this explicitly.

Definition 1.1 *A stationary stochastic process $\{Z_t\}$ is said to be strongly mixing if*

$$\psi_Z(\tau) = \sup_{A \in \mathcal{F}_{-\infty}^{0}(Z), B \in \mathcal{F}_{\tau}^{\infty}(Z)} \Big| P[A \cap B] - P[A]P[B] \Big| \to 0 \quad \text{as } \tau \to \infty$$

The $\psi_Z(\tau)$ are called the (strong) mixing coefficients for $\{Z_t\}$.

The following lemma will be used, together with theorem A.1, to find upper bounds for the mixing coefficients of $\{Y_t\}$.

Lemma 1.2 *If the stationary linear process $\{Y_t\}$ is invertible, then the spectral density $f(\lambda)$ is everywhere non-zero and there exists C such that*

$$f(\lambda) \geq C > 0 \quad \text{for all } \lambda$$

Proof: Define $\alpha_j = 0$ for all $j < 0$. Then

$$f(\lambda) \;=\; \sum_{j=-\infty}^{\infty} \gamma_j e^{ij\lambda}$$

$$
\begin{aligned}
&= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \alpha_k \alpha_{k+j} e^{ij\lambda} \\
&= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \alpha_k \alpha_{k+j} e^{-ik\lambda} e^{i(j+k)\lambda} \\
&= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \alpha_k \alpha_j e^{-ik\lambda} e^{ij\lambda} \\
&= |\sum_{j=-\infty}^{\infty} \alpha_j e^{ij\lambda}|^2
\end{aligned}
$$

Suppose that $\sum_{j=-\infty}^{\infty} \alpha_j e^{ijx} = 0$; i.e. $\sum_{j=0}^{\infty} \alpha_j e^{ijx} = 0$ and $\sum_{j=0}^{\infty} \alpha_j e^{-ijx} = 0$. Consider the ϵ_t sequence given by $\epsilon_t = a\cos(tx - b)$. Then

$$
\begin{aligned}
Y_t &= \sum_{j=0}^{\infty} \alpha_j \epsilon_{t-j} = \tfrac{1}{2}a(\sum_{j=0}^{\infty} \alpha_j e^{-ib} e^{i(t-j)x} + \sum_{j=0}^{\infty} \alpha_j e^{ib} e^{-i(t-j)x}) \\
&= \tfrac{1}{2}a(e^{i(tx-b)} \sum_{j=0}^{\infty} \alpha_j e^{-ijx} + e^{i(b-tx)} \sum_{j=0}^{\infty} \alpha_j e^{ijx}) = 0
\end{aligned}
$$

Thus any pair of realisations of the ϵ-process which differ only by $a\cos(tx - b)$ give rise to the same realisation of the Y-process; i.e. the Y-process cannot be invertible, contrary to the hypothesis.

Therefore we have $f(\lambda) > 0$ for any λ. But $f(\lambda)$ is continuous since the γ_j are summable. Therefore $f(\lambda)$ is a continous function on the closed interval $[-\pi, \pi]$ and must attain its infimum somewhere on that interval and so $\inf_\lambda f(\lambda) > 0$ as required.

Q.E.D.

Lemma 1.3 *Let $\alpha_j \leq O(j^{-3})$ as $j \to \infty$. Then $\psi_Y(\tau) \leq O(\tau^{-2})$ as $\tau \to \infty$.*

Proof: From the hypothesis, for some α^*

$$
|k^3\gamma_k| = k^3|\sum_{j=0}^{\infty} \sigma^2 \alpha_j \alpha_{j+k}| \leq k^3 \sigma^2 \alpha^* \sum_{j=0}^{\infty} \frac{1}{j^3(j+k)^3} \leq \alpha^* \sigma^2 \sum_{j=0}^{\infty} \frac{1}{j^3} =^{\text{def}} \gamma^* < \infty
$$

The γ_k form, therefore, an absolutely summable sequence. It is however well known (see [5]) that any complex power series is analytic in any disc where it is absolutely convergent. So setting

$$
\phi_\tau(z) = \sum_{j=0}^{\infty} \gamma_{\tau-j} z^j
$$

defines a function analytic in the unit disc.

Furthermore, since $\{Y_t\}$ is invertible by the original definition, lemma 1.2 shows that there exists $C > 0$ such that $|f_Y(\lambda)| \geq C$ for all λ. Thus, from theorem A.1

$$\begin{aligned}
\psi_Y(\tau) &\leq \inf_\phi \sup_\lambda |f_Y(\lambda) - e^{i\lambda\tau}\phi(e^{-i\lambda})|/f_Y(\lambda) \\
&\leq \sup_\lambda C^{-1}|f_Y(\lambda) - e^{i\lambda\tau}\phi_\tau(e^{-i\lambda})| \\
&= \sup_\lambda C^{-1}|\sum_{j=-\infty}^{\infty}\gamma_j e^{i\lambda j} - \sum_{j=-\infty}^{\tau}\gamma_j e^{i\lambda j}| \\
&= \sup_\lambda C^{-1}|\sum_{j=\tau+1}^{\infty}\gamma_j e^{ij\lambda}| \\
&\leq \sup_\lambda \frac{\gamma^*}{C}\sum_{j=\tau+1}^{\infty} j^{-3} \\
&\leq \frac{\gamma^*}{2C\tau^2}
\end{aligned}$$

as required

$\mathcal{Q.E.D.}$

We have now sufficient resources to prove the following theorem on the asymptotic normality of the $\{g_{C,j}\}$. We shall make use of a central limit theorem for stationary processes which are functions in a local sense of strongly mixing sequences.

Theorem 1.4 *Let $\alpha_j = O(j^{-3})$ as $j \to \infty$. Then the $\{g_{C,j}\}$ are asymptotically jointly normally distributed in the sense that, if $(t_1, \ldots, t_m) \in \mathbf{R}^m$ and $j_1, \ldots, j_m$ are any natural numbers, then*

$$\sqrt{n}\sum_{k=1}^{m} t_k \{g_{C,j_k} - \gamma_{C,j_k}\} \to^d N(0, \tilde{\sigma}^2)$$

where $\tilde{\sigma}^2 = \sum_{l=-\infty}^{\infty} E[G_0 G_l]$ exists and is non-negative and

$$G_l = \sum_{k=1}^{m} t_k\{\sin 2\pi X_l \sin 2\pi X_{l+j_k} - \gamma_{C,j_k}\}$$

Proof: First we let $\{Y_t\}$ and G_0 take the roles of $\{Z_t\}$ and W respectively in theorem A.4, and we show that conditions (1), (2) and (3) of the theorem are satisfied. G_0 is clearly a measurable function of $\{Y_t\}$. $E[G_0] = 0$ by definition, and G_l is obviously obtained by time-shifting G_0.

(1) $|G_l| \leq (\max_{1\leq k\leq m}|\gamma_{C,j_m}| + 1)\sum_{k=1}^{m}|t_k|$ and is clearly uniformly bounded.

(2) G_0 is a function only of $Y_0, Y_1, \ldots, Y_J$ where $J = \max_{1\leq k\leq m} j_m$. Therefore, whenever $T > J$, $E[G_0|\mathcal{F}_{-T}^{T}(Y)] = G_0$ and the summation must converge.

(3) $\sum_{k=1}^{\infty} \psi_Y(k) \leq \sum_{k=1}^{\infty} \gamma^*/(Ck^2) < \infty$ by lemma 1.3.

Therefore $\tilde{\sigma}^2$ is finite and non-negative. We must now consider two cases separately.

a) $\tilde{\sigma}^2 = 0$

$$\tfrac{1}{n}E[(\sum_{l=1}^{n} G_l)^2] = \tfrac{1}{n}\left\{nE[G_0^2] + \sum_{l=1}^{n-2} 2(n-l)E[G_0G_l]\right\}$$

by stationarity of $\{G_l\}$. But we already know that $\sum_{l=1}^{\infty} E[G_0G_l]$ is convergent and so $\lim_{n\to n} \frac{1}{n}\sum_{l=1}^{\infty} lE[G_0G_l] = 0$ which implies that

$$\lim_{n\to\infty} \tfrac{1}{n}E[(\sum_{l=1}^{n} G_l)^2] = \sum_{l=-\infty}^{\infty} E[G_0G_l] = 0$$

i.e. $\frac{1}{\sqrt{n}}\sum_{l=1}^{n} G_l \to 0$ in L_2-norm, which in turn implies that

$$\frac{1}{\sqrt{n}}\sum_{l=1}^{n} G_l \to^d N(0,0) = N(0,\tilde{\sigma}^2)$$

b) $\tilde{\sigma}^2 > 0$ By theorem A.4, we already know that

$$\frac{1}{\sqrt{n}}\sum_{l=1}^{n} G_l \to^d N(0,\tilde{\sigma}^2)$$

The final part of the proof is to show that $\frac{1}{\sqrt{n}}\sum_{l=1}^{n} G_l$ is asymptotically equivalent to $\sqrt{n}\sum_{m=1}^{k} t_m\{g_{C,j_m} - \gamma_{C,j_m}\}$. However

$$\left|\sqrt{n}(g_{C,j_m} - \gamma_{C,j_m}) - \frac{1}{\sqrt{n}}\sum_{l=1}^{n}(\sin 2\pi X_l \sin 2\pi X_{l+j_m} - \gamma_{C,j_m})\right|$$

$$= \left|\frac{\sqrt{n}}{n-j_m}\sum_{l=j_m+1}^{n} \sin 2\pi X_l \sin 2\pi X_{l-j_m} - \frac{1}{\sqrt{n}}\sum_{l=1}^{n} \sin 2\pi X_l \sin 2\pi X_{l+j_m}\right|$$

$$\leq \left|\left(\frac{\sqrt{n}}{n-j_m} - \frac{1}{\sqrt{n}}\right)\sum_{l=1}^{n-j_m} \sin 2\pi X_l \sin 2\pi X_{l-j_m}\right|$$

$$+\frac{1}{\sqrt{n}}\left|\sum_{l=n-j_m+1}^{n} \sin 2\pi X_l \sin 2\pi X_{l+j_m}\right|$$

$$\leq (n-j_m)\left|\frac{\sqrt{n}}{n-j_m} - \frac{1}{\sqrt{n}}\right| + \frac{j_m}{\sqrt{n}}$$

$$= \frac{2j_m}{\sqrt{n}} \to 0$$

and the result follows.

Q.E.D.

1.2.3 Estimation of the circular auto-correlation function

In this section we shall explore the asymptotic behaviour of the $r_{C,j}$. Three main results will be obtained. First it will be shown that the convergence of $r_{C,j}$ to $\rho_{C,j}$ is better than almost sure; $r_{C,j}$ converges rapidly in mean to $\rho_{C,j}$. Secondly the covariance structure of the $r_{C,j}$ is explored in terms of the covariance structure of the $g_{C,j}$. Finally the $r_{C,j}$ are shown to be, asymptotically, normally distributed. The principal tool for the first two of these is the theorem below, taken from [11], on the behaviour of functions of random variables in terms of Taylor series expansions.

I have chosen to estimate the circular auto-correlation function $\rho_{C,j}$ by

$$r_{C,j} = \frac{g_{C,j}}{g_{C,0}}$$

This is clearly consistent since the $\{g_{C,k}\}$ are consistent for the $\{\gamma_{C,j}\}$ and $\gamma_{C,0} > 0$. Throughout this section we shall asume that $\gamma_j \leq O(j^{-2})$ and hence that there exists a constant γ^* such that for all j

$$|\gamma_j| \leq \frac{\gamma^*}{1+j^2}$$

The approach used for deriving properties of the $r_{C,j}$ is based on Taylor expansion. The following lemma will be used to validate the assumptions of a theorem from [11] quoted in the appendix.

Lemma 1.5 *As* $n \to \infty$

$$E[(g_{C,j} - \gamma_{C,j})^4] = O(n^{-2})$$

Proof: For the duration of the proof, let $s(x)$ denote the function $\sin(2\pi x)$. Then by an argument similar to that used in theorem 1.1, we have

$$E[s(X_t)s(X_{t-j})s(X_u)s(X_{u-j})s(X_v)s(X_{v-j})s(X_w)s(X_{w-j})]$$

$$= \frac{1}{256} \sum_{\eta_1^8 \in \{-1,1\}^8} \prod_{l=1}^{8} \eta_l \tag{1.5}$$

$$\times \exp\Big(-4\pi^2\{4\gamma_0 + (\eta_1\eta_2 + \eta_3\eta_4 + \eta_5\eta_6 + \eta_7\eta_8)\gamma_j$$

$$+A + B + C + D + E + F\}\Big)$$

where

$$\begin{aligned}
A &= (\eta_1\eta_3 + \eta_2\eta_4)\gamma_{d_1} + \eta_1\eta_4\gamma_{d_1+j} + \eta_2\eta_3\gamma_{d_1-j} \\
B &= (\eta_1\eta_5 + \eta_2\eta_6)\gamma_{d_1+d_2} + \eta_1\eta_6\gamma_{d_1+d_2+j} + \eta_2\eta_5\gamma_{d_1+d_2-j} \\
C &= (\eta_1\eta_7 + \eta_2\eta_8)\gamma_{d_1+d_2+d_3} + \eta_1\eta_8\gamma_{d_1+d_2+d_3+j} + \eta_2\eta_7\gamma_{d_1+d_2+d_3-j} \\
D &= (\eta_3\eta_5 + \eta_4\eta_6)\gamma_{d_2} + \eta_3\eta_6\gamma_{d_2+j} + \eta_4\eta_5\gamma_{d_2-j} \\
E &= (\eta_3\eta_7 + \eta_4\eta_8)\gamma_{d_2+d_3} + \eta_3\eta_8\gamma_{d_2+d_3+j} + \eta_4\eta_7\gamma_{d_2+d_3-j} \\
F &= (\eta_5\eta_7 + \eta_6\eta_8)\gamma_{d_3} + \eta_5\eta_8\gamma_{d_3+j} + \eta_6\eta_7\gamma_{d_3-j}
\end{aligned}$$

where $d_1 = u - t$, $d_2 = v - u$ and $d_3 = w - v$.

Now by the same approach we find

$$\begin{aligned}
&E[s(X_t)s(X_{t-j})s(X_u)s(X_{u-j})s(X_v)s(X_{v-j)}] \\
&= -\frac{1}{64}\sum_{\eta_1^6\in\{-1,1\}^6}\prod_{l=1}^{6}\eta_l \exp(-4\pi^2\{3\gamma_0 + (\eta_1\eta_2 + \eta_3\eta_4 + \eta_5\eta_6)\gamma_j + A + B + D\})
\end{aligned}$$

We also have

$$\begin{aligned}
\gamma_{C,j} &= e^{-4\pi^2\gamma_0}\sinh 4\pi^2\gamma_j \\
&= \tfrac{1}{2}e^{-4\pi^2\gamma_0}(e^{4\pi^2\gamma_j} - e^{-4\pi^2\gamma_j}) \\
&= -\frac{1}{4}\sum_{\eta_7,\eta_8\in\{-1,1\}^2}\eta_7\eta_8\exp(-4\pi^2(\gamma_0 + \eta_7\eta_8\gamma_j))
\end{aligned}$$

and so

$$\begin{aligned}
&E\Big[s(X_t)s(X_{t-j})s(X_u)s(X_{u-j})s(X_v)s(X_{v-u)}\Big] \times \gamma_{C,j} \qquad (1.6) \\
&= \frac{1}{256}\sum_{\eta_1^8\in\{-1,1\}^8}\prod_{l=1}^{8}\eta_l \\
&\qquad \times \exp\Big(-4\pi^2\{4\gamma_0 + (\eta_1\eta_2 + \eta_3\eta_4 + \eta_5\eta_6 + \eta_7\eta_8)\gamma_j + A + B + D\}\Big)
\end{aligned}$$

If in the last equation we use u, v and w instead of t, u and v we get the same result with D, E and F instead of A, B and D. If we use t, u and w we get A, E and C. If we use t, v and w we B, F and C.

By the same approach we also obtain

$$E\Big[s(X_t)s(X_{t-j})s(X_u)s(X_{u-j})\Big] \times \gamma_{C,j}^2 \qquad (1.7)$$

$$= \frac{1}{256} \sum_{\eta_1^8 \in \{-1,1\}^8} \prod_{l=1}^{8} \eta_l \exp(-4\pi^2 \{4\gamma_0 + (\eta_1\eta_2 + \eta_3\eta_4 + \eta_5\eta_6 + \eta_7\eta_8)\gamma_j + A\})$$

If we use different indices than t and u we get a different term instead of A. Using t and v gives rise to B. Using t and w gives rise to C. Using u and v gives rise to D. Using u and w gives rise to E. Using v and w gives rise to F.

Finally, we have

$$\gamma_{C,j}^4 = \frac{1}{256} \sum_{\eta_1^8 \in \{-1,1\}^8} \prod_{l=1}^{8} \eta_l \exp(-4\pi^2 \{4\gamma_0 + (\eta_1\eta_2 + \eta_3\eta_4 + \eta_5\eta_6 + \eta_7\eta_8)\gamma_j\}) \quad (1.8)$$

Combining (1.5), (1.6), (1.7) and (1.8), we have $(n-j)^4 E[(g_{C,j} - \gamma_{C,j})^4] =$

$$\sum_{t=j+1}^{n} \sum_{u=j+1}^{n} \sum_{v=j+1}^{n} \sum_{w=j+1}^{n} E\Big[(s(X_t)s(X_t - j) - \gamma_{C,j})(s(X_u)s(X_u - j) - \gamma_{C,j})$$

$$\times (s(X_v)s(X_v - j) - \gamma_{C,j})(s(X_w)s(X_w - j) - \gamma_{C,j})\Big]$$

$$= \frac{e^{-16\pi^2\gamma_0}}{256(n-j)^4} \sum_{t=j+1}^{n} \sum_{u=j+1}^{n} \sum_{v=j+1}^{n} \sum_{w=j+1}^{n} \sum_{\eta_1^8 \in \{-1,1\}^8} \prod_{l=1}^{8} \eta_l$$

$$\times \exp\Big(-4\pi^2(\eta_1\eta_2 + \eta_3\eta_4 + \eta_5\eta_6 + \eta_7\eta_8)\gamma_j\Big) \quad (1.9)$$

$$\times \Big\{e^{-4\pi^2(A+B+C+D+E+F)} - e^{-4\pi^2(A+B+D)} - e^{-4\pi^2(D+E+F)} - e^{-4\pi^2(A+C+E)}$$

$$-e^{-4\pi^2(B+C+F)} + e^{-4\pi^2 A} + e^{-4\pi^2 B} + e^{-4\pi^2 C} + e^{-4\pi^2 D} + e^{-4\pi^2 E} + e^{-4\pi^2 F} - 3\Big\}$$

Denote by $g(d_1, d_2, d_3, \eta_1^8)$ the expression in braces in equation 1.9. Then

$$E[(g_{C,j} - \gamma_{C,j})^4]$$

$$\leq \frac{e^{-16\pi^2\gamma_0}}{256(n-j)^4} \sum_{t=j+1}^{n} \sum_{u=j+1}^{n} \sum_{v=j+1}^{n} \sum_{w=j+1}^{n} \sum_{\eta_1^8 \in \{-1,1\}^8} |g(d_1, d_2, d_3, \eta_1^8)|$$

$$\leq \frac{e^{-16\pi^2\gamma_0}}{256(n-j)^4} \sum_{\eta_1^8 \in \{-1,1\}^8} \sum_{t=j+1}^{n} \sum_{d_1=-n}^{n} \sum_{d_2=-n}^{n} \sum_{d_3=-n}^{n} |g(d_1, d_2, d_3, \eta_1^8)| \quad (1.10)$$

Now, by Taylor's theorem, we have

$$e^{z_1} = 1 + e^{\zeta_1} \stackrel{\text{def}}{=} 1 + \xi_1 z_1$$

for some $\zeta_1 \in [0, z_1]$; i.e. some $\xi_1 \in [0, e^{z_1}]$. Therefore putting $z_1 = -4\pi^2 A$, $z_2 = -4\pi^2 B$, ... etc.

$$\begin{aligned} g(d_1, d_2, d_3, \eta_1^8) &= (1+\xi_1 z_1)(1+\xi_2 z_2)(1+\xi_3 z_3)(1+\xi_4 z_4)(1+\xi_5 z_5)(1+\xi_6 z_6) \\ &- (1+\xi_1 z_1)(1+\xi_2 z_2)(1+\xi_4 z_4) \\ &- (1+\xi_4 z_4)(1+\xi_5 z_5)(1+\xi_6 z_6) \\ &- (1+\xi_1 z_1)(1+\xi_3 z_3)(1+\xi_5 z_5) \\ &- (1+\xi_2 z_2)(1+\xi_3 z_3)(1+\xi_6 z_6) \\ &+ (1+\xi_1 z_1)+(1+\xi_2 z_2)+(1+\xi_3 z_3) \\ &+ (1+\xi_4 z_4)+(1+\xi_5 z_5)+(1+\xi_6 z_6) \\ &- 3 \end{aligned} \tag{1.11}$$

Expanding this into a polynomial in $\xi_1 z_1, \ldots, \xi_6 z_6$ yields a complicated expression of 41 terms which has the important feature that every term involves at least two distinct $\xi_j z_j$.

But, by definition of A,

$$\begin{aligned} |4\pi^2 A| &\leq 4\pi^2 \left(2|\gamma_{d_1}| + |\gamma_{d_1+j}| + |\gamma_{d_1-j}|\right) \\ &\leq 4\pi^2 \left(\frac{2\gamma^*}{1+d_1^2} + \frac{\gamma^*}{1+(d_1+j)^2} + \frac{\gamma^*}{1+(d_1-j)^2}\right) \\ &\leq 16\pi^2\gamma^* \stackrel{\text{def}}{=} M > 1 \end{aligned} \tag{1.12}$$

The same holds for $B, C, \ldots, F$. Thus $|z_j| \leq M$ and $|\xi_j| \leq e^M$. Hence the modulus of each term in the polynomial expansion of g is less than

$$M^4 e^{4M}\, e^M |z_{j_1}| e^M |z_{j_2}| \tag{1.13}$$

for some $j_1, j_2 \in \{1, 2, \ldots, 6\}$ with $j_1 \neq j_2$. Consider the case when $j_1 = 5$ and $j_2 = 6$. Then by (1.12)

$$\sum_{d_1=-n}^{n} \sum_{d_2=-n}^{n} \sum_{d_3=-n}^{n} |z_5||z_6| \leq 16\pi^4 \gamma^{*2} \sum_{d_1=-n}^{n} \sum_{d_2=-n}^{n} \sum_{d_3=-n}^{n} h(d_2+d_3)h(d_3) \tag{1.14}$$

where

$$h(l) = \frac{2}{1+l^2} + \frac{1}{1+(l+j)^2} + \frac{1}{1+(l-j)^2}$$

But

$$\sum_{d_2=-n}^{n}\sum_{d_3=-n}^{n} h(d_2+d_3)h(d_3) \leq \sum_{d_2=-\infty}^{\infty}\sum_{d_3=-\infty}^{\infty} h(d_2+d_3)h(d_3) = \sum_{d_2=-\infty}^{\infty} h(d_2) \sum_{d_3=-\infty}^{\infty} h(d_3)$$

and

$$\sum_{l=-\infty}^{\infty} h(l) = 4\sum_{l=-\infty}^{\infty} \frac{1}{1+l^2} \stackrel{\text{def}}{=} C_2 < \infty$$

Thus, from (1.14),

$$\sum_{d_1=-n}^{n}\sum_{d_2=-n}^{n}\sum_{d_3=-n}^{n} |z_5||z_6| \leq (2n+1)16\pi^4\gamma^{*2}C_2^2$$

and similarly for other values of j_1 and j_2 giving

$$\sum_{d_1=-n}^{n}\sum_{d_2=-n}^{n}\sum_{d_3=-n}^{n} |z_{j_1}||z_{j_2}| \leq (2n+1)16\pi^4\gamma^{*2}C_2^2 \tag{1.15}$$

Combining the count of the number of terms in (1.11) with (1.10), (1.13) and (1.15) we find

$$\begin{aligned} &E[(g_{C,j}-\gamma_{C,j})^4] \\ &\leq \frac{e^{-16\pi^2(\gamma_0-\gamma_j)}}{256(n-j)^3} \sum_{\eta_1^8\in\{-1,1\}^8} 41.M^4e^{6M}.256(2n+1)\pi^4\gamma^{*2}C_2^2 \\ &= \frac{10516(2n+1)\pi^4\gamma^{*2}C_2^2M^4e^{6M}e^{-16\pi^2(\gamma_0-\gamma_j)}}{(n-j)^3} \\ &= O(n^{-2}) \end{aligned}$$

as required.

Q.E.D.

This is now used to prove the following theorem which is really a lemma for the corollary which follows. The corollary is the first real result of the section, on the nature of the convergence of $r_{C,j}$ to $\rho_{C,j}$.

Theorem 1.6 *As* $n \to \infty$

$$r_{C,j} = \frac{g_{C,j}}{\gamma_{C,0}} - \frac{\rho_{C,j}\,(g_{C,0}-\gamma_{C,0})}{\gamma_{C,0}} + O_{L^1}(n^{-1})$$

Proof: We shall show that $(r_{C,j} - g_{C,j}/\gamma_{C,0} + (g_{C,0} - \gamma_{C,0})\gamma_{C,j}/\gamma_{C,0}^2)^2$ satisfies the conditions of theorem A.7 with $\mathbf{Z} = (g_{C,0}, g_{C,j})$, $\alpha = 1$, $s = 4$, $N_0 = 1$, $\mu = (\gamma_{C,0}, \gamma_{C,j})$ and $a_n = O(n^{-\frac{1}{2}})$.

1. $E\Big[((g_{C,0}-\gamma_{C,0})^2+(g_{C,j}-\gamma_{C,j})^2)^2\Big] \leq 2E[(g_{C,0}-\gamma_{C,0})^4]+2E[(g_{C,j}-\gamma_{C,j})^4] = O(n^{-2})$ $= O(\sqrt{n}^{-4})$, by lemma 1.5.

2. $$|r_{C,j} - \frac{g_{C,j}}{\gamma_{C,0}} + \frac{\gamma_{C,j}}{\gamma_{C,0}^2}(g_{C,0} - \gamma_{C,0})|$$
   $$\leq (j+1) + \gamma_{C,0}^{-1} + \gamma_{C,0}^{-2}|\gamma_{C,j}|(1+\gamma_{C,0})$$
   which is a uniform bound as required. This is true since $|g_{C,j}| \leq 1$ and
   $$|r_{C,j}| = \frac{n}{n-j}\frac{\sum s_k s_{k+j}}{\sum s_k^2} \leq \frac{j+1}{1}$$

3. Let $S = \{(x,y) : |x-\gamma_{C,0}|^2 + |y-\gamma_{C,j}|^2 \leq \frac{1}{4}\gamma_{C,0}^2\}$. Then, since x is bounded away from zero in S
   $$f_n(x,y) = (\frac{y}{x} - \frac{y}{\gamma_{C,0}} + \frac{\gamma_{C,j}}{\gamma_{C,0}^2}(x-\gamma_{C,0}))^2$$
   has continuous derivatives of any order in S.

4. $(\gamma_{C,0}, \gamma_{C,j})$ is in the interior of S by definition.

5. This holds since f_n is independent of n, the derivatives are continuous in S and S is compact.

Now we apply theorem A.7.

$$f(x,y) = (g(x,y))^2$$

But

$$g(x,y)|_\mu = 0$$

$$g_{,x}(x,y)|_\mu = \frac{-y}{x^2} - \frac{\gamma_{C,j}}{\gamma_{C,0}^2}\Big|_\mu = 0$$

and

$$g_{,y}(x,y)|_\mu = \frac{1}{x} - \frac{1}{\gamma_{C,0}}\Big|_\mu = 0$$

Therefore all derivatives of f_n up to third order are zero at μ. So, by theorem A.7,

$$E\left[\left(r_{C,j} - \frac{g_{C,j}}{\gamma_{C,0}} + (g_{C,0} - \gamma_{C,0})\frac{g_{C,j}}{\gamma_{C,0}^2}\right)^2\right] = O(a_n^4) = O(n^{-2})$$

But for any random variable Z, $E[|Z|] \leq \sqrt{E[Z^2]}$ and so the result follows.

$\mathcal{Q.E.D.}$

Corollary 1.7 $\left|E[r_{C,j}] - \rho_{C,j}\right| \leq O(n^{-1})$.

Proof: From theorem 1.6

$$r_{C,j} = \frac{g_{C,j}}{\gamma_{C,0}} - \frac{\rho_{C,j}\,(g_{C,0} - \gamma_{C,0})}{\gamma_{C,0}} + O_{L^1}(n^{-1})$$

Thus

$$\begin{aligned} E[r_{C,j}] &= \rho_{C,j} - \frac{\rho_{C,j}}{\gamma_{C,0}}.0 + O(n^{-1}) \\ &= \rho_{C,j} + O(n^{-1}) \end{aligned}$$

which is the desired result.

$\mathcal{Q.E.D.}$

The next theorem concerns the covariance structure of the $r_{C,j}$. It can be used together with theorem 1.1 to explicitly calculate the covariance structure of the $r_{C,j}$.

Theorem 1.8

$$\begin{aligned} E[(r_{C,j} - \rho_{C,j})(r_{C,k} - \rho_{C,k})] &= \frac{1}{\gamma_{C,0}^2} C[g_{C,j}, g_{C,k}] + \frac{\gamma_{C,j}\gamma_{C,k}}{\gamma_{C,0}^4} D[g_{C,0}] \\ &\quad - \frac{\gamma_{C,j}}{\gamma_{C,0}^3} C[g_{C,k}, g_{C,0}] - \frac{\gamma_{C,k}}{\gamma_{C,0}^3} C[g_{C,j}, g_{C,0}] + O(n^{-3/2}) \end{aligned}$$

Proof: We shall see that $(r_{C,j} - \rho_{C,j})(r_{C,k} - \rho_{C,k})$ satisfies the hypotheses of theorem A.7 with $\mathbf{Z} = (g_{C,0}, g_{C,j}, g_{C,k})$, $\alpha = 4/3$, $s = 3$, $N_0 = 1$, $\mu = (\gamma_{C,0}, \gamma_{C,j}, \gamma_{C,k})$ and $a_n = O(n^{-\frac{1}{2}})$.

1. $$E\Big[((g_{C,0} - \gamma_{C,0})^2 + (g_{C,j} - \gamma_{C,j})^2 + (g_{C,k} - \gamma_{C,k})^2)^2\Big]$$

   $$\leq 2E[(g_{C,0} - \gamma_{C,0})^4] + 2E[(g_{C,j} - \gamma_{C,j})^4] + 2E[(g_{C,k} - \gamma_{C,k})^4]$$

   $$= O(n^{-2}) = O(\sqrt{n}^{-4}) \qquad \text{by lemma 1.5}$$

2. $$|r_{C,j} - \rho_{C,j}||r_{C,k} - \rho_{C,k}| \le ((j+1) + |\rho_{C,j}|)((k+1) + |\rho_{C,k}|)$$

and so $E\Big[|(r_{C,j} - \rho_{C,j})(r_{C,k} - \rho_{C,k})|^4\Big]$ is uniformly bounded and hence $O(1)$.

3. Let $S = \{(x, y, z) : |x - \gamma_{C,0}|^2 + |y - \gamma_{C,j}|^2 + |z - \gamma_{C,k}|^2 \le \frac{1}{4}\gamma_{C,0}^2\}$. Then since x is bounded away from zero in S, all derivatives of

$$f_n(x, y, z) = \left(\frac{y}{x} - \frac{\gamma_{C,j}}{\gamma_{C,0}}\right)\left(\frac{z}{x} - \frac{\gamma_{C,k}}{\gamma_{C,0}}\right)$$

are continuous is S.

4. $(\gamma_{C,0}, \gamma_{C,j}, \gamma_{C,k})$ is in the interior of S by definition.

5. Since f_n is independent of n, this follows from (3) and (4).

We now apply theorem A.7. The following identities are easily verified.

$$f(\mu) = f(\gamma_{C,0}, \gamma_{C,j}, \gamma_{C,k}) = 0$$

$$f_{,y}(x, y, z) = (1/x)(z/x - \gamma_{C,k}/\gamma_{C,0}) = 0 \qquad \text{at } \mu$$

$$f_{,z}(x, y, z) = (1/x)(y/x - \gamma_{C,j}/\gamma_{C,0}) = 0 \qquad \text{at } \mu$$

$$f_{,x}(x, y, z) = -\frac{y}{x^2}\left(\frac{z}{x} - \frac{\gamma_{C,k}}{\gamma_{C,0}}\right) - \frac{z}{x^2}\left(\frac{y}{x} - \frac{\gamma_{C,j}}{\gamma_{C,0}}\right) = 0 \qquad \text{at } \mu$$

$$f_{,yy}(x, y, z) \equiv 0$$

$$f_{,zz}(x, y, z) \equiv 0$$

$$f_{,yz}(x, y, z) = x^{-2} = \gamma_{C,0}^{-2} \qquad \text{at } \mu$$

$$f_{,xy}(x, y, z) = -\frac{2z}{x^3} + \frac{\gamma_{C,k}}{x^2\gamma_{C,0}} = -\frac{\gamma_{C,k}}{\gamma_{C,0}^3} \qquad \text{at } \mu$$

$$f_{,xz}(x, y, z) = -\frac{2y}{x^3} + \frac{\gamma_{C,j}}{x^2\gamma_{C,0}} = -\frac{\gamma_{C,j}}{\gamma_{C,0}^3} \qquad \text{at } \mu$$

$$f_{,xx}(x, y, z) = \frac{3yz}{x^4} - \frac{2y\gamma_{C,k}}{x^3\gamma_{C,0}} + \frac{3yz}{x^4} - \frac{2z\gamma_{C,j}}{x^3\gamma_{C,0}} = \frac{2\gamma_{C,j}\gamma_{C,k}}{\gamma_{C,0}^4} \qquad \text{at } \mu$$

Therefore

$$E[(r_{C,j} - \rho_{C,j})(r_{C,k} - \rho_{C,k})]$$

$$= \frac{1}{2}\Bigg\{ \quad \frac{2}{\gamma_{C,0}^2}E[(g_{C,j}-\gamma_{C,j})(g_{C,k}-\gamma_{C,k})] - \frac{2\gamma_{C,j}}{\gamma_{C,0}^3}E[(g_{C,k}-\gamma_{C,k})(g_{C,0}-\gamma_{C,0})]$$
$$-\frac{2\gamma_{C,k}}{\gamma_{C,0}^3}E[(g_{C,j}-\gamma_{C,j})(g_{C,0}-\gamma_{C,0})] + \frac{2\gamma_{C,j}\gamma_{C,k}}{\gamma_{C,0}^4}E[(g_{C,0}-\gamma_{C,0})^2] \quad \Bigg\}$$
$$+O(a_n^3)$$

as required since $O(a_n^3) = O(n^{-3/2})$.

$\mathcal{Q.E.D.}$

The final theorem of the section shows asymptotic normality of the $r_{C,j}$. The proof is based on the fact that the asymptotic behaviour of the $r_{C,j}$ is closely related to that of the $g_{C,j}$ and is a simple application of theorem 1.4.

Theorem 1.9 *Let $\alpha_j = O(j^{-2})$ as $j \to \infty$. Then the $r_{C,j}$ are jointly asymptotically normally distributed in the sense that if $t_1, \ldots, t_m$ are real numbers and $j_1, \ldots, j_m$ are any natural numbers, then*

$$\sqrt{n}\sum_{k=1}^{m} t_k\{r_{C,j_k} - \rho_{C,j_k}\} \to^d N(0, \hat{\sigma}^2) \tag{1.16}$$

where

$$\hat{\sigma}^2 = \frac{1}{\gamma_{C,0}^2}\sum_{l=-\infty}^{\infty} E[G_0 G_l]$$

and

$$G_l = \sum_{k=1}^{m} t_k\{\sin 2\pi X_l \sin 2\pi X_{l+j_k} - \gamma_{C,j_k} + \rho_{C,j_k}(\sin^2 2\pi X_l - \gamma_{C,0})\}$$

Proof:

$$\begin{aligned} r_{C,j} - \rho_{C,j} &= \frac{g_{C,j}}{g_{C,0}} - \frac{\gamma_{C,j}}{\gamma_{C,0}} \\ &= \frac{g_{C,j}-\gamma_{C,j}}{g_{C,0}} + \gamma_{C,j}\frac{\gamma_{C,0}-g_{C,0}}{g_{C,0}\gamma_{C,0}} \end{aligned}$$

Thus the left-hand side of equation 1.16 is

$$\frac{\sqrt{n}}{g_{C,0}}\left\{\sum_{k=1}^{m} t_k(g_{C,j_k}-\gamma_{C,j_k}) + (\gamma_{C,0}-g_{C,0})\sum_{k=1}^{m}\frac{\gamma_{C,j_k}t_k}{\gamma_{C,0}}\right\}$$

But

$$\sqrt{n}\left\{\sum_{k=1}^{m} t_k(g_{C,j_k}-\gamma_{C,j_k}) + (\gamma_{C,0}-g_{C,0})\sum_{k=1}^{m}\frac{\gamma_{C,j_k}t_k}{\gamma_{C,0}}\right\} \to^d N(0, \tilde{\sigma}^2)$$

for some $\tilde{\sigma}^2$ by theorem 1.4. Also $g_{C,0} \to \gamma_{C,0}$ almost surely and so the left-hand side of equation 1.16 converges in distribution to $N(0, \tilde{\sigma}^2/\gamma_{C,0}^2)$. The theorem is then proved if $\tilde{\sigma}^2$ is the appropriate value. But, by theorem 1.4, $\tilde{\sigma}^2 = \sum_{l=-\infty}^{\infty} E[G_0 G_l]$ where

$$G_l = \sum_{k=1}^{m} t_k(\sin 2\pi X_l \sin 2\pi X_{l+j_k} - \gamma_{C,j_k}) + (\sin^2 2\pi X_l - \gamma_{C,0}) \sum_{k=1}^{m} t_k \frac{\gamma_{C,j_k}}{\gamma_{C,0}}$$

as required.

Q.E.D.

1.2.4 Uses of the circular auto-correlation function

The reason for constructing and estimating the circular auto-correlation was to facilitate the identification of wrapped moving-average models. I now illustrate that this is in fact possible, although it works well only for the case of an independent sequence.

Sequence of Independent Variables

The estimated circular auto-correlations have certain desirable features for a sequence of independent Gaussian random variables. In this case $\alpha_j = 0$ when $j > 0$ and so $\gamma_j = 0$ when $j \neq 0$. Then by theorem 1.1

$$\begin{aligned} C[g_{C,j}, g_{C,k}] = \frac{e^{-8\pi^2\gamma_0}}{2(n-j)(n-k)} & \sum_{d=j-n+1}^{n-k-1} \min(n-k, n-k-d, n-j-d) \\ \times \; \Big\{ & e^{-4\pi^2(\gamma_j+\gamma_k)} \sinh^2(2\pi^2(\gamma_d + \gamma_{d+k} + \gamma_{d-j} + \gamma_{d-j+k})) \\ & - e^{4\pi^2(\gamma_k-\gamma_j)} \sinh^2(2\pi^2(-\gamma_d + \gamma_{d+k} - \gamma_{d-j} + \gamma_{d-j+k})) \\ & - e^{4\pi^2(\gamma_j-\gamma_k)} \sinh^2(2\pi^2(-\gamma_d - \gamma_{d+k} + \gamma_{d-j} + \gamma_{d-j+k})) \\ & + e^{4\pi^2(\gamma_j+\gamma_k)} \sinh^2(2\pi^2(\gamma_d - \gamma_{d+k} - \gamma_{d-j} + \gamma_{d-j+k})) \Big\} \end{aligned}$$

But each term in the above sum is trivially zero, unless $d = 0$ or $d = j$ or $d = -k$ or $d = j - k$. We shall now consider a number of different cases comprising all the possible different values of j and k. We shall suppose (without loss of generality) that $j \leq k$. Since the only part of the sum which depends on d is that between curly braces, we shall ignore the rest.

1. $j \neq k$, $j > 0$ and $k > 0$. In this case $\gamma_j = \gamma_k = 0$ and so each of the interesting values of d gives rise to a term of the form

$$\sinh^2 2\pi^2\gamma_0 + \sinh^2 2\pi^2\gamma_0 - \sinh^2 2\pi^2\gamma_0 - \sinh^2 2\pi^2\gamma_0$$

which implies $C[g_{C,j}, g_{C,k}] = 0$.

2. $j = k > 0$. This time there are only three terms since $d = 0$ coincides with $d = j - k$. When $d = j$ or $d = -k$ we get

$$\sinh^2 2\pi^2\gamma_0 + \sinh^2 2\pi^2\gamma_0 - \sinh^2 2\pi^2\gamma_0 - \sinh^2 2\pi^2\gamma_0$$

which is 0. When $d = 0$ we get

$$\sinh^2 2\pi^2 2\gamma_0 + \sinh^2 2\pi^2 2\gamma_0 - 0 - 0$$

and so when $j > 0$

$$D[g_{C,j}] = \frac{1}{n-j} e^{-8\pi^2\gamma_0} \sinh^2 4\pi^2\gamma_0 = \frac{1}{n-k}\gamma_{C,0}^2$$

3. $j = 0$ and $k > 0$. This time there are only two terms since $d = 0$ coincides with $d = j$ and $d = -k$ coincides with $d = j - k$. When $d = 0$ we get

$$e^{-4\pi^2\gamma_0} \sinh^2 2\pi^2 2\gamma_0 + e^{4\pi^2\gamma_0} \sinh^2 2\pi^2(0)$$

$$-e^{-4\pi^2\gamma_0} \sinh^2 2\pi^2(-2\gamma_0) - e^{4\pi^2\gamma_0} \sinh^2 2\pi^2(0)$$

which is 0. When $d = -k$ we get

$$e^{-4\pi^2\gamma_0} \sinh^2 2\pi^2 2\gamma_0 + e^{4\pi^2\gamma_0} \sinh^2 2\pi^2(0)$$

$$-e^{-4\pi^2\gamma_0} \sinh^2 2\pi^2 2\gamma_0 - e^{4\pi^2\gamma_0} \sinh^2 2\pi^2(0)$$

which is also 0. Thus $C[g_{C,0}, g_{C,j}] = 0$ when $j > 0$.

4. $j = k = 0$. This time there is only one term which is

$$e^{-8\pi^2\gamma_0} \sinh^2 2\pi^2(4\gamma_0) + e^{8\pi^2\gamma_0} \sinh^2 2\pi^2(0) - \sinh^2 2\pi^2(0) - \sinh^2 2\pi^2(0)$$

and so

$$D[g_{C,0}] = \frac{1}{2n} e^{-16\pi^2\gamma_0} \sinh^2 8\pi^2\gamma_0$$

From the above, and from corollary 1.7 and theorem 1.8 we see that, when j and k are both positive and $j \neq k$

$$D[r_{C,j}] = \frac{1}{\gamma_{C,0}^2} D[g_{C,j}] + O(n^{-\frac{3}{2}}) = \frac{1}{n-k} + O(n^{-\frac{3}{2}})$$

and that

$$C[r_{C,j}, r_{C,k}] = O(n^{-\frac{3}{2}})$$

which together imply that

$$\rho(r_{C,j}, r_{C,k}) = O(n^{-\frac{1}{2}})$$

So, for large samples, we have a simple test for dependence based on the circular auto-correlation function estimates, since we know they are asymptotically normal by theorem 1.9.

Wrapped MA(l) Process

In this section we shall suppose that the $\{Y_t\}$ process is an MA(l) process, i.e. that α_j and therefore γ_j are zero when $|j| > l$. Then by theorem 1.1, when $j > l$

$$\begin{aligned} D[g_{C,j}] = \frac{e^{-8\pi^2\gamma_{C,0}}}{2(n-j)^2} \sum_{d=j-n+1}^{n-j-1} (n-j-|d|) \quad \{ & \sinh^2 2\pi^2(2\gamma_d + \gamma_{d+j} + \gamma_{d-j}) \\ - & \sinh^2 2\pi^2(\gamma_{d+j} - \gamma_{d-j}) \\ - & \sinh^2 2\pi^2(-\gamma_{d+j} + \gamma_{d-j}) \\ + & \sinh^2 2\pi^2(2\gamma_d - \gamma_{d+j} - \gamma_{d-j})\} \end{aligned} \qquad (1.17)$$

Suppose now that $j > 2l$. Then at most one of γ_d, γ_{d+j} and γ_{d-j} can be non-zero for any d. If γ_{d+j} is non-zero the terms in braces in equation 1.17 are

$$\sinh^2 2\pi^2\gamma_{d+j} - \sinh^2 2\pi^2\gamma_{d+j} - \sinh^2 2\pi^2\gamma_{d+j} + \sinh^2 2\pi^2\gamma_{d+j} = 0$$

and similarly if γ_{d-j} is non-zero. Thus equation 1.17 becomes

$$\frac{e^{-8\pi^2\gamma_{C,0}}}{2(n-j)^2} \sum_{d=-l}^{l} (n-j-|d|)\{\sinh^2 2\pi^2 2\gamma_d - 0 - 0 + \sinh^2 2\pi^2 2\gamma_d\}$$

$$\to \frac{e^{-8\pi^2\gamma_{C,0}}}{(n-j)} \sum_{d=-l}^{l} \sinh^2 4\pi^2\gamma_d$$

and so by theorem 1.8

$$D[g_{C,j}] = \frac{1}{n-j}\frac{\sum_{d=-l}^{l} \sinh^2 4\pi^2\gamma_d}{\sinh^2 4\pi^2\gamma_0} + O(n^{-3/2})$$

Further, when $j, k > 2l$ and $k - j > 2l$, for any value of d, only one of γ_d, γ_{d+k}, γ_{d-j} and γ_{d-j+k} can be non-zero and so applying theorem 1.1 $C[g_{C,j}, g_{C,k}] = 0$. Hence by applying theorem 1.8

$$C[r_{C,j}, r_{C,k}] = O(n^{-3/2})$$

This provides a test for greater than l-dependence. If the process is a wrapped MA(l), the variance of $r_{C,2l+1}$ can be approximated by the estimated variance determined from $r_{C,2l+1}, \ldots, r_{C,2l+N}$ for some N. Since, by theorem 1.9, $r_{C,2l+1}$ is asymptotically normally distributed, we can test to see if $r_{C,2l+1}$ lies far in the tails of its estimated distribution.

This is clearly less powerful than the corresponding result for the autocorrelation function of an MA(l). However it does provide some help with the identification problem.

1.2.5 Identification of wrapped auto-regressive models

Identifying wrapped MA models is only half of the requirement. For ordinary ARMA models, the partial auto-correlation function is the usual tool for the identification of auto-regressive processes. The obvious approach in the wrapped case is to try to find some analogue for it. The ordinary partial auto-correlation function is the sequence $p_1, p_2, \ldots$ where p_j is the leading coefficient of the AR(j) model fitted to the data, i.e. $p_j = p_{jj}$ where

$$Y_t = p_{j1}Y_{t-1} + p_{j2}Y_{t-2} + \cdots + p_{jj}Y_{t-j} + \epsilon_t$$

is the maximum likelihood AR(j) model for the observed sequence. Clearly, if the data is actually observed from an AR(k) process, we expect p_j to be nearly zero when $j > k$. The *pacf* can be estimated in two ways. One is simply to perform maximum likelihood estimation for each value of j. The other, which is equivalent, is to calculate the *pacf* from the *acf* since there is a known bijective relationship between them.

In the case of the wrapped model, the first approach still works provided maximum likelihood estimation works for wrapped auto-regressive models. This is shown to be the case in chapter 2. However, this process is computationally extremely expensive. Alternatively, one might attempt to estimate the auto-correlation function from the circular auto-correlation function and thence estimate the partial auto-correlation function. Two difficulties are encountered.

To estimate the auto-correlation function from the circular auto-correlation function, it seems obvious to exploit the bijective relationship

$$\gamma_{C,j} = e^{-4\pi^2\gamma_0} \sinh 4\pi^2\gamma_j$$

between $\gamma_0, \gamma_1, \ldots$ and $\gamma_{C,1}, \gamma_{C,2}, \ldots$ Inverting we obtain

$$\gamma_j = \frac{1}{4\pi^2} \sinh^{-1}\left(\frac{\gamma_{C,j}}{[1 - 2\gamma_{C,0}]^{\frac{1}{2}}}\right)$$

which suggests estimating the *acf* by

$$g_j = \frac{1}{4\pi^2} \sinh^{-1}\left(\frac{g_{C,j}}{[1 - 2g_{C,0}]^{\frac{1}{2}}}\right) \tag{1.18}$$

Unfortunately, it is not always the case that $g_{C,0} < \frac{1}{2}$. It is clear from its definition that $g_{C,0}$ can be as large as 1. Of course this can be overcome by using some different function to calculate g_0 from $g_{C,0}$ perhaps derived by some pseudo-Bayesian approach.

However even supposing that this can be done, there is another problem. The estimated auto-correlation function in the non-wrapped case is always a positive definite sequence. This is neccesary to calculate the estimated partial auto-correlation function. On the other hand there is no reason why, in general, a sequence obtained through (1.18) should be positive definite. This would seem to rule out this whole approach, leaving only the method of successive maximum likelihood estimates.

Chapter 2

Estimation for the Wrapped AR(1)

In this chapter I examine the estimation problem for the simplest member of the wrapped linear family of models — the wrapped AR(1). The chapter opens with a brief discussion of a crude form of moment estimation. The remainder is devoted to maximum likelihood estimation. Two sections are devoted to consistency and asymptotic normality, both of which draw heavily on the examination of the behaviour of the the unwrapped AR(1) conditional upon the wrapped process. The chapter closes with a description of some of the computational properties of the maximum likelihood estimates.

2.1 Estimation by Moments

For the purpose of comparison with maximum likelihood estimation — in particular to demonstrate that the latter is justifiable despite its computational complexity – I commence this chapter by briefly considering a form of moment estimation based on the ideas in chapter 1.

Moment estimation for the AR(1) is easily formulated from the relationships

$$\phi = \frac{\gamma_1}{\gamma_0} \qquad \text{and} \qquad \sigma^2 = \gamma_0(1 - \phi^2)$$

using the notation of chapter 1. Hence provided suitable estimators g_0 and g_1 (of γ_0 and γ_1 respectively) exist, moment estimation can be performed for the wrapped AR(1). As described in section 1.2.5 it is obvious to exploit the relationships

$$\gamma_0 = -\frac{1}{8\pi^2} \ln[1 - 2\gamma_{C,0}]$$

$$\gamma_1 = \frac{1}{4\pi^2} \sinh^{-1} \frac{\gamma_{C,1}}{\sqrt{1 - 2\gamma_{C,0}}}$$

to define g_0 and g_1 in terms of $g_{C,0}$ and $g_{C,1}$. The problem, as previously noted, is that $g_{C,0}$ is not necessarily less than $\frac{1}{2}$. However, for sufficiently large values of n, this will occur with very small probability and in what follows I shall ignore this possibility. Since the estimates are undefined when $g_{C,0} \geq \frac{1}{2}$, allowing for this case can only worsen the properties of the estimators, and since the intention is ultimately to show that moment estimation is a poor procedure, ignoring this case is of no consequence.

Before proceeding any further, note that $g_{C,j} - \gamma_{C,j}$ is asymptotically of order $n^{-\frac{1}{2}}$ by theorem 1.4 and hence all products of two or more of the $g_{C,j} - \gamma_{C,j}$ are of order n^{-1} or less. Thus, using Taylor series, from $g_0 = -(8\pi^2)^{-1}\ln[1 - 2g_{C,0}]$ we have

$$g_0 = \gamma_0 + \frac{1}{4\pi^2}(1 - 2\gamma_{C,0})^{-1}(g_{C,0} - \gamma_{C,0}) + O(n^{-1})$$

and from $g_1 = (4\pi^2)^{-1}\sinh^{-1} g_{C,1}/\sqrt{1 - 2g_{C,0}}$ we have

$$g_1 = \gamma_1 + \frac{1}{4\pi^2}(\gamma_{C,1}^2 + 1 - 2\gamma_{C,0})^{-\frac{1}{2}}(g_{C,1} - \gamma_{C,1})$$
$$+\frac{\gamma_{C,1}}{4\pi^2}(1 - 2\gamma_{C,0})^{-1}(\gamma_{C,1}^2 + 1 - 2\gamma_{C,0})^{-\frac{1}{2}}(g_{C,0} - \gamma_{C,0}) + O(n^{-1})$$

These last two equations can be rewritten, by putting $A = \frac{1}{4\pi^2}(\gamma_{C,1}^2 + 1 - 2\gamma_{C,0})^{-\frac{1}{2}}$, $B = \frac{\gamma_{C,1}}{4\pi^2}(1 - 2\gamma_{C,0})^{-1}(\gamma_{C,1}^2 + 1 - 2\gamma_{C,0})^{-\frac{1}{2}}$ and $C = \frac{1}{4\pi^2}(1 - 2\gamma_{C,0})^{-1}$ as

$$g_0 = \gamma_0 + C(g_{C,0} - \gamma_{C,0}) + O(n^{-1})$$

and

$$g_1 = \gamma_1 + A(g_{C,1} - \gamma_{C,1}) + B(g_{C,0} - \gamma_{C,0}) + O(n^{-1})$$

Hence, by the binomial theorem

$$\hat{\phi} = \frac{g_1}{g_0} = \phi + \frac{A}{\gamma_0}(g_{C,1} - \gamma_{C,1}) + \frac{1}{\gamma_0}(B - \phi C)(g_{C,0} - \gamma_{C,0}) + O(n^{-1})$$

and

$$\begin{aligned}\hat{\sigma} &= \sqrt{g_0(1 - \hat{\phi}^2)} \\ &= \sigma - \sigma^{-1}\phi A(g_{C,1} - \gamma_{C,1}) + \tfrac{1}{2}\sigma^{-1}\{(1 + \phi^2)C - 2\phi B\}(g_{C,0} - \gamma_{C,0}) + O(n^{-1})\end{aligned}$$

Therefore, by theorem 1.4, $\hat{\phi}$ and $\hat{\sigma}$ are consistent and have, asymptotically, a bivariate normal distribution. Further, by ignoring the $O(n^{-1})$ terms, the covariance matrix of

$\hat{\phi}$ and $\hat{\sigma}$ can be computed in terms of the $(g_{C,1} - \gamma_{C,1})$ and $(g_{C,0} - \gamma_{C,0})$. However the latter covariance matrix is order n^{-1} by theorem 1.4 and so the covariance matrix of $\sqrt{n}\phi$ and $\sqrt{n}\hat{\sigma}$ converges to some positive definite matrix. From theorem 1.1, it is easy to see that

$$\begin{aligned} nC[g_{C,j}, g_{C,k}] \to \tfrac{1}{2}e^{-8\pi^2\gamma_0} \sum_{d\in\mathbf{Z}} \quad & \Big\{ e^{-4\pi^2(\gamma_j+\gamma_k)} \sinh^2(2\pi^2(\gamma_d + \gamma_{d+k} + \gamma_{d-j} + \gamma_{d-j+k})) \\ & - e^{4\pi^2(\gamma_k-\gamma_j)} \sinh^2(2\pi^2(-\gamma_d + \gamma_{d+k} - \gamma_{d-j} + \gamma_{d-j+k})) \\ & - e^{4\pi^2(\gamma_j-\gamma_k)} \sinh^2(2\pi^2(-\gamma_d - \gamma_{d+k} + \gamma_{d-j} + \gamma_{d-j+k})) \\ & + e^{4\pi^2(\gamma_j+\gamma_k)} \sinh^2(2\pi^2(\gamma_d - \gamma_{d+k} - \gamma_{d-j} + \gamma_{d-j+k})) \Big\} \end{aligned}$$

and the limit form of the covariance matrix of $\sqrt{n}\phi$ and $\sqrt{n}\hat{\sigma}$ is thence quite easy to compute numerically. Figure 2.1 shows grey-scale plots of the asymptotic log of standard deviations of $\sqrt{n}\phi$ and $\sqrt{n}\hat{\sigma}$ and the asymptotic correlation between $\sqrt{n}\phi$ and $\sqrt{n}\hat{\sigma}$, for a range of values of the true parameters σ and ϕ.

2.2 Maximum Likelihood Estimation

The remainder of the chapter is devoted to maximum likelihood estimation. I begin with a discussion of the likelihood function and its derivatives. There is then a long and very mathematical section which explores the behaviour of the AR(1) conditional upon a realisation of the wrapped process and leads to strong results concerning the decay of dependence with time. These results are used in the two following sections to show consistency and asymptotic normality of the maximum likelihood estimates. The latter requires more analysis of the conditional behaviour of the AR(1) dependent on the wrapped process and is quite involved. The chapter closes with a discussion of some computational aspects of maximum likelihood estimation.

The mathematical presentation which follows is extremely intricate. There are a very large number of lemmata and theorems, the main purposes of which I have attempted to explain in short paragraphs at the beginning of each subsection. To further aid the reader in comprehending this material, I have included 3 figures (2.2, 2.3, 2.4), which display the dependencies between the various lemmata and theorems in each of the major sections of mathematical material.

Figure 2.1: Asymptotic log of standard deviation and correlation between the parameter estimates from moment estimation

Throughout this section let Y_t denote an AR(1) process, i.e.

$$Y_t = \phi Y_{t-1} + \epsilon_t$$

where ϵ_t is a sequence of independent normal random variables with mean 0 and variance σ^2 and ϵ_t is independent of $Y_{t-1}, Y_{t-2}, \ldots$. Let X_t denote the corresponding wrapped process and K_t denote the integer difference between Y_t and X_t, i.e. $Y_t = X_t + K_t$ where $K_t = [Y_t]$ is the nearest integer to Y_t. Note that this implies that $X_t \in [-\frac{1}{2}, \frac{1}{2})$.

Some other notation needs to be defined. θ denotes the parameter pair (σ, ϕ) and the space of allowable values of θ will be denoted by $\mathcal{P}$, i.e. $\mathcal{P} = \{\theta : \sigma > 0 \text{ and } |\phi| \leq 1\}$. The symbol ω will be encountered frequently. ω denotes a realisation of the wrapped AR(1) process, and will usually be encountered in the form "for all ω", meaning "for all realisations of the wrapped AR(1)". Finally the pseudo-norm $\|x\|_{\mathbf{Z}}$ is used to denote the distance from x to the nearest integer.

Two very important properties of the wrapped AR(1) which derive from the AR(1) will be used frequently, often without reference. The first, which has been extensively used in chapter 1, is stationarity. The second is time-reversibility, i.e. if $\ldots, Y_t, Y_{t+1}, \ldots$ is an AR(1) with parameters ϕ and σ^2, then so is the time-reversed sequence $\ldots, Y_t, Y_{t-1}, \ldots$

2.2.1 The Likelihood Function

The likelihood function for an AR(1) model is given by

$$f_{\mathbf{Y}_1^n}(\mathbf{y}_1^n) = \frac{\sqrt{1-\phi^2}}{\sigma^n (2\pi)^{n/2}} \exp\Big(-\tfrac{1}{2}\sigma^{-2}\mathbf{y}_1^{nT}\mathbf{M}_n\mathbf{y}_1^n\Big) \tag{2.1}$$

where $\mathbf{M}_n$ is the $n \times n$ matrix

$$\mathbf{M}_n = \begin{pmatrix} 1 & -\phi & & & & \\ -\phi & 1+\phi^2 & -\phi & & & \\ & -\phi & 1+\phi^2 & -\phi & & \\ & & \ddots & \ddots & \ddots & \\ & & & -\phi & 1+\phi^2 & -\phi \\ & & & & -\phi & 1 \end{pmatrix}$$

and all the unspecified entries farther off the diagonal are 0. The following lemma puts an important bound on $\mathbf{M}_n$.

Lemma 2.1 *For any value of n and all $\mathbf{x} \in \mathbf{R}^n$*

$$(1-|\phi|)^2\|\mathbf{x}\|^2 \leq \mathbf{x}^T\mathbf{M}_n\mathbf{x} \leq (1+|\phi|)^2\|\mathbf{x}\|^2$$

where $\|\mathbf{x}\|$ denotes the Euclidean norm on $\mathbf{R}^n$.

Proof: Since $\mathbf{M}_n$ is positive definite real symmetric, all of its eigenvalues are positive reals and, for $\mathbf{x} \in \mathbf{R}^n$,

$$\mathbf{x}^T\mathbf{M}_n\mathbf{x} \geq \lambda_{\min}\|\mathbf{x}\|^2$$

where $\lambda_{\min}$ is the smallest eigenvalue of $\mathbf{M}_n$. By theorem A.8, all of the eigenvalues of $\mathbf{M}_n$ lie in the union of the n complex disks defined by

$$|\mathbf{M}_{nii} - re^{i\theta}| \leq \sum_{j\neq i}|\mathbf{M}_{nij}| \quad i = 1,\ldots,n$$

Therefore $|1-\lambda_{\min}| \leq |\phi|$ or $|1+\phi^2-\lambda_{\min}| \leq 2|\phi|$, i.e.

$$\lambda_{\min} \geq \min(1-|\phi|,(1-|\phi|)^2) = (1-|\phi|^2)$$

The upper bound follows similarly.

$\mathcal{Q.E.D.}$

From (2.1), the likelihood function for a wrapped AR(1) is given by

$$\begin{aligned} f_{\mathbf{X}_1^n}(\mathbf{x}_1^n) &= \sum_{\mathbf{k}_1^n\in\mathbf{Z}^n} f_{\mathbf{Y}_1^n}(\mathbf{x}_1^n+\mathbf{k}_1^n) \\ &= \frac{\sqrt{1-\phi^2}}{\sigma^n(2\pi)^{n/2}}\sum_{\mathbf{k}_1^n\in\mathbf{Z}^n}\exp\Big(-\tfrac{1}{2}\sigma^{-2}(\mathbf{x}_1^n+\mathbf{k}_1^n)^T\mathbf{M}_n(\mathbf{x}_1^n+\mathbf{k}_1^n)\Big) \end{aligned} \tag{2.2}$$

One obvious difficulty is that the X-likelihood function cannot be calculated since it requires the summation of infinitely many terms. Further even if some finite approximation were adequate, say by summing each k_j from $-N$ to N for some N, this would still appear to require the calculation of N^n terms which would be impractical for all except the smallest values of n.

Fortunately, there exists a factorisation of the likelihood function which makes it possible to calculate for large sample sizes and which is crucial to the proofs given later of properties of the maximum likelihood estimator. It arises as follows:

$$f_{X_t|\mathbf{X}_1^{t-1}}(x_t|\mathbf{x}_1^{t-1})$$

$$= f_{\mathbf{X}_1^t}(\mathbf{x}_1^t)/f_{\mathbf{X}_1^{t-1}}(\mathbf{x}_1^{t-1})$$

$$= \sum_{\mathbf{k}_1^t \in \mathbf{Z}^t} f_{\mathbf{Y}_1^t}(\mathbf{x}_1^t + \mathbf{k}_1^t)/f_{\mathbf{X}_1^{t-1}}(\mathbf{x}_1^{t-1})$$

$$= \sum_{\mathbf{k}_1^t \in \mathbf{Z}^t} f_{Y_t|Y_{t-1}}(x_t + k_t | x_{t-1} + k_{t-1}) f_{\mathbf{Y}_1^{t-1}}(\mathbf{x}_1^{t-1} + \mathbf{k}_1^{t-1})/f_{\mathbf{X}_1^{t-1}}(\mathbf{x}_1^{t-1})$$

$$= \sum_{k_t, k_{t-1} \in \mathbf{Z}} f_{Y_t|Y_{t-1}}(x_t + k_t | x_{t-1} + k_{t-1}) f_{Y_{t-1}, \mathbf{X}_1^{t-2}}(x_{t-1} + k_{t-1}, \mathbf{x}_1^{t-2})/f_{\mathbf{X}_1^{t-1}}(\mathbf{x}_1^{t-1})$$

$$= \sum_{j,k \in \mathbf{Z}} f_{Y_t|Y_{t-1}}(x_t + j | x_{t-1} + k) P[Y_{t-1} = x_{t-1} + k | \mathbf{X}_1^{t-1} = \mathbf{x}_1^{t-1}] \tag{2.3}$$

Also

$$P[Y_t = x_t + j | \mathbf{X}_1^t = \mathbf{x}_1^t]$$

$$= f_{Y_t, \mathbf{X}_1^{t-1}}(x_t + j, \mathbf{x}_1^{t-1})/f_{\mathbf{X}_1^t}(\mathbf{x}_1^t)$$

$$= \sum_{k \in \mathbf{Z}} \frac{f_{Y_t|Y_{t-1}}(x_t + j | x_{t-1} + k) f_{Y_{t-1}, \mathbf{X}_1^{t-2}}(x_{t-1} + k, \mathbf{x}_1^{t-2})}{f_{X_t|\mathbf{X}_1^{t-1}}(x_t | \mathbf{x}_1^{t-1}) f_{\mathbf{X}_1^{t-1}}(\mathbf{x}_1^{t-1})}$$

$$= \sum_{k \in \mathbf{Z}} f_{Y_t|Y_{t-1}}(x_t + j | x_{t-1} + k) P[Y_{t-1} = x_{t-1} + k | \mathbf{X}_1^{t-1} = \mathbf{x}_1^{t-1}]/f_{X_t|\mathbf{X}_1^{t-1}}(x_t | \mathbf{x}_1^{t-1})$$

From this it is clear that, provided we keep track of $P[Y_t = x_t + j | \mathbf{X}_1^t = \mathbf{x}_1^t]$ at every stage, it is possible to calculate each $f_{X_t|\mathbf{X}_1^{t-1}}(x_t | \mathbf{x}_1^{t-1})$ with a fixed amount of effort independent of t. In other words the likelihood function takes order n calculations. These conditional probabilities are of the utmost importance in the rest of the chapter and we shall denote them by $\mathbf{a}_t$, where the vector $\mathbf{a}_t$ has components

$$a_{t,k} = P[Y_t = X_t + k | \mathbf{X}_1^t]$$

i.e., for $t \geq 1$,

$$a_{t,j} = \sum_{k \in \mathbf{Z}} f_{Y_t|Y_{t-1}}(X_t + j | X_{t-1} + k) a_{t-1,k}/f_{X_t|\mathbf{X}_1^{t-1}}(X_t | \mathbf{X}_1^{t-1}) \tag{2.4}$$

Note that the $a_{t,k}$ depend, in general, on the sequence $X_1, \ldots, X_t$. This dependence will not usually be made explicit. The appropriate sequence should be obvious from the context. The following lemma puts an important uniform bound on $f_{X_t|\mathbf{X}_1^{t-1}}(x_t | \mathbf{x}_1^{t-1})$.

Lemma 2.2 *There exists a positive continuous function $K(.)$ such that $K(\sigma) \to 1$ as $\sigma \to \infty$ and such that for all real y*

$$K^{-1}(\sigma) \leq \sum_{k\in\mathbf{Z}} (\sigma\sqrt{2\pi})^{-1} \exp\left(-\tfrac{1}{2}\sigma^{-2}(y+k)^2\right) \leq K(\sigma) \tag{2.5}$$

Further, for all t and ω and for all $\theta \in \mathcal{P}$,

$$K^{-1}(\sigma) \leq f_{X_t|\mathbf{X}_1^{t-1}}(X_t|\mathbf{X}_1^{t-1}) \leq K(\sigma)$$

Proof: Denote by $h(y,\sigma)$ the function occurring in the centre of (2.5). Clearly h exists and is positive for all y and all $\sigma > 0$. Also, for each σ, h is a periodic function of y, with period 1. So we need only consider whether the inequalities are satisfied for all y in $[-\frac{1}{2}, \frac{1}{2}]$. But h is the absolutely convergent sum of functions, uniformly continuous at each σ and y, and is therefore continuous. Any continuous function on a closed interval attains its infimum and supremum on that interval. Therefore, for each σ,

$$\inf_{y\in[-\frac{1}{2},\frac{1}{2}]} h(y,\sigma) \qquad \text{and} \qquad \sup_{y\in[-\frac{1}{2},\frac{1}{2}]} h(y,\sigma)$$

both exist and are positive continuous functions of σ. Thus the function K is easily chosen to satisfy the inequalities in (2.5). As σ tends to infinity, the sum in (2.5) tends uniformly in y to the integral of the normal probability density function having that variance. For, if $k \neq 1$,

$$\begin{aligned} &(\sigma\sqrt{2\pi})^{-1}\left|\int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\exp(-\tfrac{1}{2}\sigma^{-2}(y+k)^2) - \exp(-\tfrac{1}{2}\sigma^{-2}(x+k)^2)\right] dx\right| \\ &\leq (\sigma\sqrt{2\pi})^{-1}\left[\exp(-\tfrac{1}{2}\sigma^{-2}(|k|-\tfrac{1}{2})^2) - \exp(-\tfrac{1}{2}\sigma^{-2}(|k|+\tfrac{1}{2})^2)\right] \end{aligned}$$

and so the difference between the integral and the sum is less than

$$(\sigma\sqrt{2\pi})^{-1}\left|\int_{-\frac{1}{2}}^{\frac{1}{2}} \exp(-\tfrac{1}{2}\sigma^{-2}x^2)\,dx - \exp(-\tfrac{1}{2}\sigma^{-2}y^2)\right| + 2(\sigma\sqrt{2\pi})^{-1}\exp(-\tfrac{1}{8}\sigma^2)$$

which clearly tends to 0, uniformly in y, as $\sigma \to \infty$. But the integral is always 1, which is the desired limit.

The final part of the proposition follows from the fact that the conditional density is a mixture of the functions in (2.5) evaluated at different values of y. Since the bound is uniform in y, it must hold for the mixture.

Q.E.D.

2.2.2 Derivatives of the log-likelihood function

For the purposes of maximum likelihood estimation the derivatives of the likelihood function are critically important. The complicated likelihood function gives rise to complicated derivatives. The purpose of this section is largely to introduce notation simplifying their algebraic form. From (2.2)

$$\partial_\sigma f_{\mathbf{X}_1^n}(\mathbf{x}_1^n) = \frac{-n}{\sigma} f_{\mathbf{X}_1^n}(\mathbf{x}_1^n) + \frac{1}{\sigma^3} \sum_{\mathbf{k}_1^n \in \mathbf{Z}^n} (\mathbf{x}_1^n + \mathbf{k}_1^n)^T \mathbf{M}_n (\mathbf{x}_1^n + \mathbf{k}_1^n) f(\mathbf{x}_1^n + \mathbf{k}_1^n)$$

and so

$$\partial_\sigma \ln f_{\mathbf{X}_1^n}(\mathbf{x}_1^n) = \frac{-n}{\sigma} + \frac{1}{\sigma^3} \frac{\sum_{\mathbf{k}_1^n \in \mathbf{Z}^n} (\mathbf{x}_1^n + \mathbf{k}_1^n)^T \mathbf{M}_n (\mathbf{x}_1^n + \mathbf{k}_1^n) f(\mathbf{x}_1^n + \mathbf{k}_1^n)}{\sum_{\mathbf{k}_1^n \in \mathbf{Z}^n} f(\mathbf{x}_1^n + \mathbf{k}_1^n)}$$

But

$$\frac{f(\mathbf{x}_1^n + \mathbf{k}_1^n)}{\sum_{\mathbf{k}_1^n \in \mathbf{Z}^n} f(\mathbf{x}_1^n + \mathbf{k}_1^n)} = P[\mathbf{Y}_1^n = \mathbf{x}_1^n + \mathbf{k}_1^n | \mathbf{X}_1^n = \mathbf{x}_1^n]$$

and so

$$\partial_\sigma \ln f_{\mathbf{X}_1^n}(\mathbf{X}_1^n) = \frac{-n}{\sigma} + \frac{1}{\sigma^3} E[(\mathbf{Y}_1^n)^T \mathbf{M}_n \mathbf{Y}_1^n | \mathbf{X}_1^n] \tag{2.6}$$

By a similar argument

$$\partial_\phi \ln f_{\mathbf{X}_1^n}(\mathbf{X}_1^n) = \frac{-\phi}{1-\phi^2} - \frac{1}{2\sigma^2} E[(\mathbf{Y}_1^n)^T \partial_\phi \mathbf{M}_n \mathbf{Y}_1^n | \mathbf{X}_1^n] \tag{2.7}$$

Now, writing

$$S_n(\mathbf{Y}_1^n) = (\mathbf{Y}_1^n)^T \mathbf{M}_n \mathbf{Y}_1^n = (1-\phi^2) Y_1^2 + \sum_{j=2}^{n} (Y_j - \phi Y_{j-1})^2$$

$$T_n(\mathbf{Y}_1^n) = (\mathbf{Y}_1^n)^T \partial_\phi \mathbf{M}_n \mathbf{Y}_1^n = -2\phi Y_1^2 + -2 \sum_{j=2}^{n} Y_{j-1}(Y_j - \phi Y_{j-1})$$

$$U_n(\mathbf{Y}_1^n) = (\mathbf{Y}_1^n)^T \partial_\phi^2 \mathbf{M}_n \mathbf{Y}_1^n = 2 \sum_{j=2}^{n-1} Y_j^2$$

it is obvious that

$$\partial_\phi S_n = T_n \qquad \partial_\phi T_n = U_n \qquad \partial_\phi U_n = 0$$

and

$$\partial_\sigma S_n = \partial_\sigma T_n = \partial_\sigma U_n = 0$$

Further, if $H(\mathbf{Y}_1^n)$ is any function of $\mathbf{Y}_1^n$, σ, and ϕ,

$$\partial_\sigma E[H(\mathbf{Y}_1^n)|\mathbf{X}_1^n] = \frac{\partial}{\partial\sigma}\frac{\sum_{\mathbf{k}_1^n\in\mathbf{Z}^n} H(\mathbf{x}_1^n+\mathbf{k}_1^n)f(\mathbf{x}_1^n+\mathbf{k}_1^n)}{\sum_{\mathbf{k}_1^n\in\mathbf{Z}^n} f(\mathbf{x}_1^n+\mathbf{k}_1^n)}$$

$$= \frac{\sum_{\mathbf{k}_1^n\in\mathbf{Z}^n} \frac{\partial}{\partial\sigma}H(\mathbf{x}_1^n+\mathbf{k}_1^n)f(\mathbf{x}_1^n+\mathbf{k}_1^n)}{\sum_{\mathbf{k}_1^n\in\mathbf{Z}^n} f(\mathbf{x}_1^n+\mathbf{k}_1^n)}$$

$$+\frac{\sum_{\mathbf{k}_1^n\in\mathbf{Z}^n} H(\mathbf{x}_1^n+\mathbf{k}_1^n)\Big(\frac{-n}{\sigma}+\frac{1}{\sigma^2}S_n(\mathbf{x}_1^n+\mathbf{k}_1^n)\Big)f(\mathbf{x}_1^n+\mathbf{k}_1^n)}{\sum_{\mathbf{k}_1^n\in\mathbf{Z}^n} f(\mathbf{x}_1^n+\mathbf{k}_1^n)}$$

$$-\frac{\sum_{\mathbf{k}_1^n\in\mathbf{Z}^n} H(\mathbf{x}_1^n+\mathbf{k}_1^n)f(\mathbf{x}_1^n+\mathbf{k}_1^n)}{\sum_{\mathbf{k}_1^n\in\mathbf{Z}^n} f(\mathbf{x}_1^n+\mathbf{k}_1^n)}\cdot\frac{\sum_{\mathbf{k}_1^n\in\mathbf{Z}^n} (\frac{-n}{\sigma}+\frac{1}{\sigma^2}S_n(\mathbf{x}_1^n+\mathbf{k}_1^n))f(\mathbf{x}_1^n+\mathbf{k}_1^n)}{\sum_{\mathbf{k}_1^n\in\mathbf{Z}^n} f(\mathbf{x}_1^n+\mathbf{k}_1^n)}$$

$$= E[\frac{\partial}{\partial\sigma}H(\mathbf{Y}_1^n)|\mathbf{X}_1^n] - \frac{n}{\sigma}E[H(\mathbf{Y}_1^n)|\mathbf{X}_1^n] + \frac{1}{\sigma^3}E[H(\mathbf{Y}_1^n)S_n(\mathbf{Y}_1^n)|\mathbf{X}_1^n]$$

$$-E[H(\mathbf{Y}_1^n)|\mathbf{X}_1^n]\left(-\frac{n}{\sigma}+\frac{1}{\sigma^3}E[S_n(\mathbf{Y}_1^n)|\mathbf{X}_1^n]\right)$$

$$= E[\frac{\partial}{\partial\sigma}H(\mathbf{Y}_1^n)|\mathbf{X}_1^n] + \frac{1}{\sigma^3}C[H(\mathbf{Y}_1^n), S_n(\mathbf{Y}_1^n)|\mathbf{X}_1^n] \quad (2.8)$$

and, by a similar argument,

$$\frac{\partial}{\partial\phi}E[H(\mathbf{Y}_1^n)|\mathbf{X}_1^n] = E[\frac{\partial}{\partial\phi}H(\mathbf{Y}_1^n)|\mathbf{X}_1^n] - \frac{1}{2\sigma^2}C[H(\mathbf{Y}_1^n), T_n(\mathbf{Y}_1^n)] \quad (2.9)$$

Thus, by repeated application of (2.8) and (2.9) to (2.6) and (2.7)

$$\partial_\sigma \ln f_{\mathbf{X}_1^n}(\mathbf{x}_1^n) = \frac{-n}{\sigma} + \frac{1}{\sigma^3}E[S_n(\mathbf{Y}_1^n)|\mathbf{X}_1^n] \quad (2.10)$$

$$\partial_\phi \ln f_{\mathbf{X}_1^n}(\mathbf{x}_1^n) = \frac{-\phi}{1-\phi^2} - \frac{1}{2\sigma^2}E[T_n(\mathbf{Y}_1^n)|\mathbf{X}_1^n] \quad (2.11)$$

$$\partial_\sigma^2 \ln f_{\mathbf{X}_1^n}(\mathbf{x}_1^n) = \frac{n}{\sigma^2} - \frac{3}{\sigma^4}E[S_n(\mathbf{Y}_1^n)|\mathbf{X}_1^n] + \frac{1}{\sigma^6}D[S_n(\mathbf{Y}_1^n)|\mathbf{X}_1^n] \quad (2.12)$$

$$\partial_\phi\partial_\sigma \ln f_{\mathbf{X}_1^n}(\mathbf{x}_1^n) = \frac{1}{\sigma^3}E[T_n(\mathbf{Y}_1^n)|\mathbf{X}_1^n] - \frac{1}{2\sigma^5}C[T_n(\mathbf{Y}_1^n), S_n(\mathbf{Y}_1^n)|\mathbf{X}_1^n] \quad (2.13)$$

$$\partial_\phi^2 \ln f_{\mathbf{X}_1^n}(\mathbf{x}_1^n) = -\frac{1+\phi^2}{(1-\phi^2)^2} - \frac{1}{2\sigma^2}E[U_n(\mathbf{Y}_1^n)|\mathbf{X}_1^n] + \frac{1}{4\sigma^4}D[T_n(\mathbf{Y}_1^n)|\mathbf{X}_1^n] \quad (2.14)$$

$$\partial_\sigma^3 \ln f_{\mathbf{X}_1^n}(\mathbf{x}_1^n) = \frac{-2n}{\sigma^3} + \frac{12}{\sigma^5} E[S_n(\mathbf{Y}_1^n)|\mathbf{X}_1^n] - \frac{9}{\sigma^7} D[S_n(\mathbf{Y}_1^n)|\mathbf{X}_1^n] + \frac{1}{\sigma^9} III[S_n(\mathbf{Y}_1^n), S_n(\mathbf{Y}_1^n), S_n(\mathbf{Y}_1^n)|\mathbf{X}_1^n] \tag{2.15}$$

$$\partial_\phi \partial_\sigma^2 \ln f_{\mathbf{X}_1^n}(\mathbf{x}_1^n) = \frac{-3}{\sigma^4} E[T_n(\mathbf{Y}_1^n)|\mathbf{X}_1^n] + \frac{7}{2\sigma^6} C[T_n(\mathbf{Y}_1^n), S_n(\mathbf{Y}_1^n)|\mathbf{X}_1^n] - \frac{1}{2\sigma^8} III[S_n(\mathbf{Y}_1^n), S_n(\mathbf{Y}_1^n), T_n(\mathbf{Y}_1^n)|\mathbf{X}_1^n] \tag{2.16}$$

$$\partial_\phi^2 \partial_\sigma \ln f_{\mathbf{X}_1^n}(\mathbf{x}_1^n) = \frac{1}{\sigma^3} E[U_n(\mathbf{Y}_1^n)|\mathbf{X}_1^n] - \frac{1}{\sigma^5} D[T_n(\mathbf{Y}_1^n)|\mathbf{X}_1^n] - \frac{1}{2\sigma^5} C[U_n(\mathbf{Y}_1^n), S_n(\mathbf{Y}_1^n)|\mathbf{X}_1^n] + \frac{1}{4\sigma^7} III[T_n(\mathbf{Y}_1^n), T_n(\mathbf{Y}_1^n), S_n(\mathbf{Y}_1^n)|\mathbf{X}_1^n] \tag{2.17}$$

$$\partial_\phi^3 \ln f_{\mathbf{X}_1^n}(\mathbf{x}_1^n) = -\frac{2\phi(\phi^2+3)}{(1-\phi^2)^3} + + \frac{3}{4\sigma^4} C[T_n(\mathbf{Y}_1^n), U_n(\mathbf{Y}_1^n)|\mathbf{X}_1^n] - \frac{1}{8\sigma^6} III[T_n(\mathbf{Y}_1^n), T_n(\mathbf{Y}_1^n), T_n(\mathbf{Y}_1^n)|\mathbf{X}_1^n] \tag{2.18}$$

2.2.3 The AR(1) conditional upon the wrapped AR(1)

The behaviour of $\mathbf{Y}_1^n$ conditional upon $\mathbf{X}_1^n$ is the key to analysis of the wrapped AR(1). We shall see that the Markovian behaviour remains, although it is non-homogeneous, and that a number of uniform inequalities can be found for the decay of information over time. The key idea is to show the existence of an appropriate metric on infinite dimensional vectors which is strictly reduced by the Markov transition operators. The inequalities obtained will be used to derive bounds for strong mixing coefficients and hence some bounds on expectations and covariances of polynomials.

Theorem 2.3 *Conditional upon $\mathbf{X}_1^n$, $K_1, \ldots, K_n$ form a Markov chain. Further the transition matrix from K_j to K_{j+1} conditional upon $\mathbf{X}_1^n$ is a function only of $\mathbf{X}_j^n$ and not of $\mathbf{X}_1^{j-1}$.*

Proof: Let $P_n[.]$ denote probabilities conditional upon $\mathbf{X}_1^n$. Then

$$P_n[K_{j+1} = k_{j+1} | \mathbf{K}_1^j = \mathbf{k}_1^j]$$

$$= P_n[\mathbf{K}_1^{j+1} = \mathbf{k}_1^{j+1}] / P_n[\mathbf{K}_1^j = \mathbf{k}_1^j]$$

Figure 2.2: Tree illustrating the dependencies between the various lemmata and theorems in section 2.2.3

$$= \frac{f_{\mathbf{X}_{j+2}^n,\mathbf{Y}_1^{j+1}}(\mathbf{x}_{j+2}^n, \mathbf{x}_1^{j+1} + \mathbf{k}_1^{j+1}) / f_{\mathbf{X}_1^n}(\mathbf{x}_1^n)}{f_{\mathbf{X}_{j+1}^n,\mathbf{Y}_1^{j}}(\mathbf{x}_{j+1}^n, \mathbf{x}_1^{j} + \mathbf{k}_1^{j}) / f_{\mathbf{X}_1^n}(\mathbf{x}_1^n)}$$

$$= \frac{f_{\mathbf{X}_{j+2}^n|Y_{j+1}}(\mathbf{x}_{j+2}^n | x_{j+1} + k_{j+1}) f_{Y_{j+1}|Y_j}(x_{j+1} + k_{j+1} | x_j + k_j) f_{\mathbf{Y}_1^j}(\mathbf{x}_1^j + \mathbf{k}_1^j)}{f_{\mathbf{X}_{j+1}^n|Y_j}(\mathbf{x}_{j+1}^n | x_j + k_j) f_{\mathbf{Y}_1^j}(\mathbf{x}_1^j + \mathbf{k}_1^j)}$$

$$= \frac{f_{\mathbf{X}_{j+2}^n|Y_{j+1}}(\mathbf{x}_{j+2}^n | x_{j+1} + k_{j+1}) f_{Y_{j+1}|Y_j}(x_{j+1} + k_{j+1} | x_j + k_j)}{f_{\mathbf{X}_{j+1}^n|Y_j}(\mathbf{x}_{j+1}^n | x_j + k_j)}$$

which does not involve $\mathbf{k}_1^{j-1}$ or $\mathbf{x}_1^{j-1}$. Thus the results follow.

$\mathcal{Q.E.D.}$

Notation

Before proceeding further, some notation must be introduced. The reasons for these notations should become clear as they are used. Frequent use will be made of the transition matrices of theorem 2.3. Therefore, for $j \le n$, define ${}_nT^{(j)}$ to be the matrix, depending on ω and θ, given by

$$\left({}_nT^{(j)}\right)_{k_1k_2} = P_{\mathbf{X}_m^n}[K_j = k_1 | K_{j-1} = k_2] \tag{2.19}$$

where $m \le j-1$. Note that there is no reference to m attached to T because of the second result of theorem 2.3. In many contexts there is no risk of confusion as to the value of n, and so it will be dropped from T for typographical convenience. For any $B \subset \mathcal{P}$, define $\mathcal{T}(B) = \{{}_nT^{(j)} : n \ge j; \theta \in B \text{ and some } \omega\}$.

The vector $P[K_j = .|\mathbf{X}_m^n]$ also will occur frequently and will be denoted by $\mathbf{p}_j^{(m,n)}$ where

$$\left(\mathbf{p}_j^{(m,n)}\right)_k = P[K_j = k|\mathbf{X}_m^n] \tag{2.20}$$

A special case of this is the vector $\mathbf{a}_j$ which has already been introduced in section 2.2.1.

$$\mathbf{a}_j = \mathbf{p}_j^{(1,j)} = P[K_j = .|\mathbf{X}_1^j]$$

Denote by $f^{(t)}$ the matrix generated from the conditional density of Y_t by

$$f_{jk}^{(t)} = f_{Y_t|Y_{t-1}}(X_t + j | X_{t-1} + k) \tag{2.21}$$

Related to this define the matrix $F(\epsilon)$, for all $\theta \in \mathcal{P}$ and all ϵ by

$$F_{jk}(\epsilon) = \exp\left[-\tfrac{1+\epsilon}{2\sigma^2}(j - \phi k)^2\right] \tag{2.22}$$

Finally, denote by $\|.\|_\gamma$ the norm defined by

$$\|\mathbf{a}\|_\gamma = \sum_{j\in\mathbf{Z}} e^{\gamma j^2}|a_j| \tag{2.23}$$

and, by D_γ, the metric defined by

$$D_\gamma(\mathbf{a},\mathbf{b}) = \|\mathbf{a}-\mathbf{b}\|_\gamma$$

Uniform domination for the $\mathbf{p}_j^{(m,n)}$

In this section I present some lemmas leading to uniform upper and lower bounds for the elements of the $\mathbf{p}_j^{(m,n)}$. The sequence is to show the existence of appropriate bounds for the elements of the $f^{(t)}$, then for the $\mathbf{a}_t$ vectors and hence for the $\mathbf{p}_j^{(m,n)}$.

Lemma 2.4 *Let $B \subset \mathcal{P}$ be compact, and $\epsilon > 0$ be sufficiently small. There exists C such that*

$$C^{-1}F_{jk}(\epsilon) \le f_{jk}^{(t)} \le CF_{jk}(-\epsilon) \tag{2.24}$$

for all j, k, t, ω and $\theta \in B$.

Proof: From (2.21)

$$f_{jk}^{(t)} = (\sigma\sqrt{2\pi})^{-1}\exp\left[-\tfrac{1}{2}\sigma^{-2}(j+x-\phi(k+y))^2\right]$$

where $x, y \in [-\frac{1}{2}, \frac{1}{2}]$. However, for any $\epsilon > 0$,

$$\begin{aligned} C_1 &= \sup\,(j+x-\phi(y+k))^2 - (1+\epsilon)(j-\phi k)^2 \\ &= \sup -\epsilon(j-\phi k)^2 + 2(j-\phi k)(x-\phi y) + (x-\phi y)^2 \end{aligned}$$

exists, where the supremum is taken over $j, k \in \mathbf{Z}$, $x, y \in [-\frac{1}{2}, \frac{1}{2}]$ and $\theta \in B$. Hence

$$f_{jk}^{(t)} \ge (\sigma\sqrt{2\pi})^{-1}\exp\left[-\tfrac{1}{2}\sigma^{-2}C_1\right]\exp\left[\tfrac{1+\epsilon}{2\sigma^2}(j-\phi k)^2\right]$$

But since B is compact, $\inf_{\theta\in B}(\sigma\sqrt{2\pi})^{-1}\exp\left[-\frac{1}{2}\sigma^{-2}C_1\right]$ is positive. This proves the left inequality in (2.24). The proof of the right inequality is similar.

$\mathcal{Q.E.D.}$

Lemma 2.5 *Let $B \subset \mathcal{P}$ be compact. Then, if $\Gamma > 0$ is such that $\Gamma < 1-\phi^2$ for all $\theta \in B$, for all sufficiently small $\epsilon > 0$ there exists C such that for all k, $\theta \in B$ and $0 < \gamma < (1-\phi^2-\Gamma)/2\sigma^2$*

$$\|F_{.k}(-\epsilon)\|_\gamma \le Ce^{(1-\epsilon)\gamma k^2}$$

Proof: From (2.22)

$$\begin{aligned}
\|F_{\cdot k}(-\epsilon)\|_\gamma &= \sum_{j\in\mathbf{Z}} e^{\gamma j^2} \exp\left[-\tfrac{1-\epsilon}{2\sigma^2}(j-\phi k)^2\right] \\
&= \exp\left[\tfrac{\gamma(1-\epsilon)}{1-\epsilon-2\gamma\sigma^2}\phi^2 k^2\right] \sum_{j\in\mathbf{Z}} \exp\left[-\tfrac{(1-\epsilon-2\gamma\sigma^2)}{2\sigma^2}(j-\tfrac{1-\epsilon}{1-\epsilon-2\gamma\sigma^2}\phi k)^2\right] \\
&\leq \frac{\sigma\sqrt{2\pi}K(\sigma/\sqrt{1-\epsilon-2\gamma\sigma^2})}{\sqrt{1-\epsilon-2\gamma\sigma^2}} \exp\left[\tfrac{\gamma(1-\epsilon)}{1-\epsilon-2\gamma\sigma^2}\phi^2 k^2\right]
\end{aligned}$$

where K is the continuous function of lemma 2.2. Provided that $\epsilon < \Gamma$, $1 \geq 1-\epsilon-2\gamma\sigma^2 \geq \phi^2$ if $0 < \gamma < (1-\phi^2-\Gamma)/2\sigma^2$. Therefore, since B is compact

$$\frac{\sigma\sqrt{2\pi}K(\sigma/\sqrt{1-\epsilon-2\gamma\sigma^2})}{\sqrt{1-\epsilon-2\gamma\sigma^2}}$$

is bounded, by C say. Thus

$$\|F_{\cdot k}(-\epsilon)\|_\gamma \leq C e^{\gamma(1-\epsilon)k^2}$$

as required.

Q.E.D.

Lemma 2.6 *Let $B \subset \mathcal{P}$ be compact. Then, if $\Gamma > 0$ and $\epsilon > 0$ there exists C such that for all k, $\theta \in B$ and $\gamma \in (0,\Gamma)$*

$$\|F_{\cdot k}(\epsilon)\|_{-\gamma} \geq C e^{-\gamma\phi^2 k^2}$$

Proof: From (2.22)

$$\begin{aligned}
&\|F_{\cdot k}(\epsilon)\|_{-\gamma} \\
&= \sum_{j\in\mathbf{Z}} e^{-\gamma j^2} \exp\left[-\tfrac{1+\epsilon}{2\sigma^2}(j-\phi k)^2\right] \\
&= \exp\left[-\tfrac{\gamma(1+\epsilon)}{1+\epsilon+2\gamma\sigma^2}\phi^2 k^2\right] \sum_{j\in\mathbf{Z}} \exp\left[-\tfrac{(1+\epsilon+2\gamma\sigma^2)}{2\sigma^2}(j-\tfrac{1+\epsilon}{1+\epsilon+2\gamma\sigma^2}\phi k)^2\right] \\
&\geq \frac{\sigma\sqrt{2\pi}K^{-1}(\sigma/\sqrt{1+\epsilon+2\gamma\sigma^2})}{\sqrt{1+\epsilon+2\gamma\sigma^2}} \exp\left[-\tfrac{\gamma(1+\epsilon)}{1+\epsilon+2\gamma\sigma^2}\phi^2 k^2\right]
\end{aligned}$$

where K is the continuous function in lemma 2.2. But γ is bounded and B is compact, so $\sigma\sqrt{2\pi}K^{-1}(\sigma/\sqrt{1+\epsilon+2\gamma\sigma^2})/\sqrt{1+\epsilon+2\gamma\sigma^2}$ has a lower bound, C say. Thus

$$\|F_{\cdot k}(\epsilon)\|_{-\gamma} \geq C e^{-\gamma\phi^2 k^2}$$

as required.

Q.E.D.

Lemma 2.7 *Let $B \subset \mathcal{P}$ be compact. If $d > 0$ is such that $1 - \phi^2 > d > 0$ for all $\theta \in B$, there exists $C > 0$ such that for all $\theta \in B$ and for all t and ω*

$$\sum_{j \in \mathbf{Z}} \exp\left[\tfrac{(1-\phi^2-d)}{2\sigma^2} j^2\right] a_{t,j} \leq C$$

Proof: By lemma 2.2 and (2.4) there exists a continuous function $K(\sigma)$ such that

$$\begin{aligned}\sum_{j \in \mathbf{Z}} \exp\left[\tfrac{(1-\phi^2-d)}{2\sigma^2} j^2\right] a_{t+1,j} &\leq K(\sigma) \sum_{j,k \in \mathbf{Z}} \exp\left[\tfrac{(1-\phi^2-d)}{2\sigma^2} j^2\right] f_{jk}^{(t+1)} a_{t,k} \\ &= K(\sigma) \sum_{k \in \mathbf{Z}} \|f_{.k}^{(t+1)}\|_{(1-\phi^2-d)/2\sigma^2} a_{t,k}\end{aligned}$$

and, by lemmas 2.5 and 2.4, there exists C_1 and $\epsilon > 0$ so that this last is dominated by

$$C_1 K(\sigma) \sum_{k \in \mathbf{Z}} \exp\left[\tfrac{(1-\epsilon)(1-\phi^2-d)}{2\sigma^2} k^2\right] a_{t,k}$$

for all t, ω and $\theta \in B$. Since K is continuous and B compact, there exists C_2 so that $K(\sigma) \leq C_2$ for all $\theta \in B$. However, from the definition of d, the infimum of $(1 - \phi^2 - d)/2\sigma^2$ over $\theta \in B$ exists and is positive. Hence there exists J so that for all $|k| \geq J$ and $\theta \in B$

$$\exp\left[\tfrac{(1-\epsilon)(1-\phi^2-d)}{2\sigma^2} k^2\right] \leq \tfrac{1}{2} C_1^{-1} C_2^{-1} \exp\left[\tfrac{(1-\phi^2-d)}{2\sigma^2} k^2\right]$$

Therefore

$$\begin{aligned}&\sum_{j \in \mathbf{Z}} \exp\left[\tfrac{(1-\phi^2-d)}{2\sigma^2} j^2\right] a_{t+1,j} \\ &\leq \tfrac{1}{2} \sum_{|j| \geq J} \exp\left[\tfrac{(1-\phi^2-d)}{2\sigma^2} j^2\right] a_{t,j} + C_1 C_2 \sum_{|j| < J} \exp\left[\tfrac{(1-\phi^2-d)}{2\sigma^2} j^2\right] a_{t,j} \\ &\leq \tfrac{1}{2} \sum_{j \in \mathbf{Z}} \exp\left[\tfrac{(1-\phi^2-d)}{2\sigma^2} j^2\right] a_{t,j} + C_1 C_2 \exp\left[\tfrac{(1-\phi^2-d)}{2\sigma^2} J^2\right]\end{aligned}$$

The result then follows from the contractive nature of this inequality, provided there is an upper bound in the case of $\mathbf{a}_1$. But, again by lemma 2.2, there exists continuous K such that

$$\begin{aligned}&\sum_{j \in \mathbf{Z}} \exp\left[\tfrac{(1-\phi^2-d)}{2\sigma^2} j^2\right] a_{1,j} \\ &= \frac{\sum_{j \in \mathbf{Z}} \exp\left[\tfrac{(1-\phi^2-d)}{2\sigma^2} j^2\right] \frac{\sqrt{1-\phi^2}}{\sigma\sqrt{2\pi}} \exp\left[-\tfrac{(1-\phi^2)}{2\sigma^2}(j + X_1)^2\right]}{\sum_{j \in \mathbf{Z}} \frac{\sqrt{1-\phi^2}}{\sigma\sqrt{2\pi}} \exp\left[-\tfrac{(1-\phi^2)}{2\sigma^2}(j + X_1)^2\right]}\end{aligned}$$

$$\leq \frac{\sqrt{1-\phi^2}}{\sigma\sqrt{2\pi}} K(\sigma/\sqrt{1-\phi^2}) \exp\left[\tfrac{(1-\phi^2)(1-\phi^2-d)}{d} X_1^2\right] \sum_{j\in\mathbf{Z}} \exp\left[-\tfrac{d}{\sigma^2}(j+\tfrac{1-\phi^2}{d}X_1)^2\right]$$

$$\leq d^{-\frac{1}{2}}\sqrt{1-\phi^2} K(\sigma/\sqrt{1-\phi^2}) K(\sigma/\sqrt{d}) \exp\left[\tfrac{(1-\phi^2)(1-\phi^2-d)}{4d}\right]$$

which is a continuous function of θ and hence has an upper bound for $\theta \in B$ since B is compact.

$\mathcal{Q.E.D.}$

Lemma 2.8 *Let $B \subset \mathcal{P}$ be compact. If $d > 0$ is such that $1-\phi^2 > d > 0$ for all $\theta \in B$, there exists $C > 0$ such that*

$$a_{t,j} \leq C \exp\left[-\tfrac{(1-\phi^2-d)}{2\sigma^2} j^2\right]$$

for all $j \in \mathbf{Z}$, t, ω and $\theta \in B$.

Proof: By lemma 2.7, there exists C such that

$$\sum_{j\in\mathbf{Z}} a_{t,j} \exp\left[\tfrac{(1-\phi^2-d)}{2\sigma^2} j^2\right] \leq C$$

and so

$$a_{t,j} \exp\left[\tfrac{(1-\phi^2-d)}{2\sigma^2} j^2\right] \leq C$$

for any j, which is the desired result.

$\mathcal{Q.E.D.}$

Lemma 2.9 *Let $B \subset \mathcal{P}$ be compact. Then there exists a sequence $c_j > 0$, $j \in \mathbf{Z}$ such that $a_{t,j} \geq c_j$ for all j, t, ω and all $\theta \in B$.*

Proof: By lemma 2.7 there exists some $d > 0$ and some C such that

$$\sum_{j\in\mathbf{Z}} \exp\left[\tfrac{(1-\phi^2-d)}{2\sigma^2} j^2\right] a_{t,j} \leq C \tag{2.25}$$

for all t, w and $\theta \in B$. For $j \in \mathbf{N}$ set

$$b_j = \inf_{\theta\in B} \exp\left[\tfrac{(1-\phi^2-d)}{2\sigma^2} j^2\right]$$

The sequence b_j, $j \in \mathbf{N}$ increases monotonically to infinity. Let J be such that $b_J > C$. Then

$$\sum_{|j|<J} a_{t,j} \geq \frac{b_J - C}{b_J}$$

for all t, w and $\theta \in B$, because otherwise (2.25) is contradicted. But, by lemma 2.2,

$$\begin{aligned} a_{t+1,j} &\geq K^{-1}(\sigma) \sum_{k \in \mathbf{Z}} f_{jk}^{(t+1)} a_{t,k} \\ &\geq K^{-1}(\sigma) \sum_{|k|<J} f_{jk}^{(t+1)} a_{t,k} \\ &\geq \frac{K^{-1}(\sigma)}{\sigma\sqrt{2\pi}} \exp\left[-\tfrac{1}{2\sigma^2}(|j| + |\phi|J + 1)^2\right] \sum_{|k|<J} a_{t,k} \\ &\geq \frac{K^{-1}(\sigma)}{\sigma\sqrt{2\pi}} \exp\left[-\tfrac{1}{2\sigma^2}(|j| + |\phi|J + 1)^2\right] (b_J - C)/b_J \end{aligned}$$

But C, J and b_J are independent of j, t, ω and θ. Hence this positive lower bound is continuous in $\theta \in B$ and has a positive infimum for $\theta \in B$ as required.

$\mathcal{Q.E.D.}$

Lemma 2.10 *Let $B \subset \mathcal{P}$ be compact. If $d > 0$, there exists $C > 0$ such that*

$$a_{t,j} \geq C \exp\left[-\tfrac{1-\phi^2+d}{2\sigma^2} j^2\right]$$

for all $j \in \mathbf{Z}$, t, ω and $\theta \in B$.

Proof: Let $\alpha(\mathbf{a}_t)$ be the largest number such that, for all j,

$$a_{t,j} \geq \alpha \exp\left[-\tfrac{(1-\phi^2+d)}{2\sigma^2} j^2\right]$$

Then, by lemma 2.2 and (2.4)

$$\begin{aligned} a_{t+1,j} &\geq K^{-1}(\sigma) \sum_{k \in \mathbf{Z}} f_{jk}^{(t+1)} a_{t,k} \\ &\geq \alpha(\mathbf{a}_t) K^{-1}(\sigma) \sum_{k \in \mathbf{Z}} f_{jk}^{(t+1)} \exp\left[-\tfrac{(1-\phi^2+d)}{2\sigma^2} k^2\right] \end{aligned}$$

and, by lemma 2.4, for any sufficiently small $\epsilon_1 > 0$ there exists C_1 such that this last is bounded below by

$$\alpha(\mathbf{a}_t) K^{-1}(\sigma) C_1 \sum_{k \in \mathbf{Z}} \exp\left[-\tfrac{(1+\epsilon_1)}{2\sigma^2}(j - \phi k)^2\right] \exp\left[-\tfrac{1-\phi^2+d}{2\sigma^2} k^2\right]$$

$$= \alpha(\mathbf{a}_t) K^{-1}(\sigma) C_1 \exp\left[-\tfrac{(1+\epsilon_1)(1-\phi^2+d)}{2\sigma^2(1+d+\epsilon_1\phi^2)} j^2\right] \sum_{k \in \mathbf{Z}} \exp\left[-\tfrac{(1+d+\epsilon\phi^2)}{2\sigma^2}(k - \tfrac{\phi(1+\epsilon_1}{1+d+\epsilon_1\phi^2} j)^2\right]$$

$$\geq \alpha(\mathbf{a}_t) K^{-1}(\sigma) C_1 K^{-1}(\sigma/\sqrt{1 + \epsilon_1\phi^2 + d}) \exp\left[-\tfrac{(1+\epsilon)(1-\phi^2+d)}{2\sigma^2(1+d+\epsilon\phi^2)} j^2\right]$$

for all j, t, ω and $\theta \in B$. Take $\epsilon_1 < d$. Then, from the definition of d and compactness of B, the infimum, C_2, over $\theta \in B$ of $(1-\phi^2+d)/2\sigma^2$ is positive. Since $C_2 > 0$, there exists J so that for all $|j| \geq J$ and $\theta \in B$

$$\exp\left[-\tfrac{(1+\epsilon)(1-\phi^2+d)}{2\sigma^2(1+d+\epsilon\phi^2)}j^2\right] \geq K(\sigma)C_1^{-1}K(\sigma/\sqrt{1+d+\epsilon_1\phi^2})\exp\left[-\tfrac{(1-\phi^2+d)}{2\sigma^2}j^2\right]$$

Also, by lemma 2.9, there exists C_3 such that, for all $|j| < J$, t, ω and $\theta \in B$,

$$a_{t+1,j} \geq C_3 \exp\left[-\tfrac{1-\phi^2+d}{2\sigma^2}j^2\right]$$

These two equations together imply that, for all t, ω and $\theta \in B$,

$$\alpha(\mathbf{a}_{t+1}) \geq \min(\alpha(\mathbf{a}_t), C_3))$$

The result then follows if $\alpha(\mathbf{a}_1)$ has a lower bound. But, by lemma 2.2, for some continuous function K

$$\begin{aligned}
\alpha(\mathbf{a}_1) &= \inf_{k\in\mathbf{Z}} \frac{\exp\left[\tfrac{1+d-\phi^2}{2\sigma^2}k^2\right]\tfrac{\sqrt{1-\phi^2}}{\sigma\sqrt{2\pi}}\exp\left[-\tfrac{1-\phi^2}{2\sigma^2}(X_1+k)^2\right]}{\tfrac{\sqrt{1-\phi^2}}{\sigma\sqrt{2\pi}}\sum_{j\in\mathbf{Z}}\exp\left[-\tfrac{1-\phi^2}{2\sigma^2}(X_1+k)^2\right]} \\
&\geq \inf_{x_1,k} \frac{\sqrt{1-\phi^2}}{\sigma\sqrt{2\pi}K(\sigma/\sqrt{1-\phi^2})}\exp\left[\tfrac{1}{2\sigma^2}(dk^2 - 2(1-\phi^2)kX_1 - (1-\phi^2)X_1^2)\right] \\
&\geq \frac{\sqrt{1-\phi^2}}{\sigma\sqrt{2\pi}K(\sigma/\sqrt{1-\phi^2})}\exp\left[-\tfrac{(1-\phi^2)}{8\sigma^2}(1+(1-\phi^2)d^{-1})\right]
\end{aligned}$$

and hence, since B is compact, there exists a lower bound which is independent of ω and $\theta \in B$ as required.

Q.E.D.

The bounds derived so far are only for the $\mathbf{a}_t$. However $\mathbf{p}_j^{(m,n)}$ can be written in terms of time-shifted and time-reversed $\mathbf{a}_t$. For

$$\begin{aligned}
\left(\mathbf{p}_j^{(m,n)}\right)_k &= P[K_j = k|\mathbf{X}_m^n] \\
&= \frac{f_{\mathbf{X}_m^{j-1},Y_j,\mathbf{X}_{j+1}^n}(\mathbf{X}_m^{j-1}, X_j+k, \mathbf{X}_{j+1}^n)}{f_{\mathbf{X}_m^n}(\mathbf{X}_m^n)}
\end{aligned}$$

and

$$f_{\mathbf{X}_m^{j-1},Y_j,\mathbf{X}_{j+1}^n}(\mathbf{X}_{j+1}^n, X_j+k, \mathbf{X}_{j+1}^n)$$

$$= f_{\mathbf{X}_{j+1}^n|Y_j}(\mathbf{X}_{j+1}^n|X_j+k)f_{Y_j,\mathbf{X}_m^{j-1}}(X_j+k|\mathbf{X}_m^{j-1})$$

$$= \frac{f_{Y_j,\mathbf{X}_{j+1}^n}(X_j+k,\mathbf{X}_{j+1}^n) f_{Y_j,\mathbf{X}_m^{j-1}}(X_j+k,\mathbf{X}_m^{j-1})}{f_{Y_j}(X_j+k)}$$

$$= \frac{P[K_j=k|\mathbf{X}_j^n]P[K_j=k|\mathbf{X}_m^j]}{P[K_j=k|X_j]} . \frac{f_{\mathbf{X}_j^n}(\mathbf{X}_j^n) f_{\mathbf{X}_m^j}(\mathbf{X}_m^j)}{f_{X_j}(X_j)}$$

and so

$$(\mathbf{p}_j^{(m,n)})_k \propto \frac{P[K_j=k|\mathbf{X}_j^n]P[K_j=k|\mathbf{X}_m^j]}{P[K_j=k|X_j]} \tag{2.26}$$

The following theorem exploits this identity to provide upper and lower bounds on the $\mathbf{p}_j^{(m,n)}$.

Theorem 2.11 *Let $B \subset \mathcal{P}$ be compact. Then if $d > 0$ such that $(1-\phi^2) > 3d > 0$ for all $\theta \in B$, there exist $C_1 > 0$ and C_2 such that*

$$C_1 \exp\left[-\tfrac{(1-\phi^2+3d)}{2\sigma^2}k^2\right] \leq (\mathbf{p}_j^{(m,n)})_k \leq C_2 \exp\left[-\tfrac{(1-\phi^2-3d)}{2\sigma^2}k^2\right]$$

for all m, n, j, k, ω and $\theta \in B$.

Proof: By stationarity and time-reversibility of Y_t, the vectors $P[K_j = .|\mathbf{X}_j^n]$, $P[K_j = .|\mathbf{X}_m^j]$ and $P[K_j = .|X_j]$ have the same distribution (and hence the same bounds) as $\mathbf{a}_{n-j+1}$, $\mathbf{a}_{j-m+1}$ and $\mathbf{a}_1$ respectively. Therefore by lemmas 2.8 and 2.10 there exist $C_1 > 0$ and C_2 such that

$$\begin{aligned} \frac{C_1^2}{C_2} \exp\left[-\tfrac{(1-\phi^2+3d)}{2\sigma^2}k^2\right] &\leq \frac{P[K_j=k|\mathbf{X}_j^n]P[K_j=k|\mathbf{X}_m^j]}{P[K_j=k|X_j]} \\ &\leq \frac{C_2^2}{C_1} \exp\left[-\tfrac{(1-\phi^2-3d)}{2\sigma^2}k^2\right] \end{aligned}$$

for all j, m, n, k, ω and $\theta \in B$. But there exists $C_3 > 0$ such that

$$\begin{aligned} C_3 &\leq \frac{C_1^2 \sigma\sqrt{2\pi}}{C_2 K(\sigma/\sqrt{1-\phi^2+3d})\sqrt{1-\phi^2+3d}} \\ &\leq \frac{C_1^2}{C_2} \sum_{k\in\mathbf{Z}} \exp\left[-\tfrac{(1-\phi^2+3d)}{2\sigma^2}k^2\right] \end{aligned}$$

for all $\theta \in B$, and there exists $C_4 < \infty$ such that

$$C_4 \geq \frac{C_2^2}{C_1} \sum_{k\in\mathbf{Z}} \exp\left[-\tfrac{(1-\phi^2-3d)}{2\sigma^2}k^2\right]$$

for all $\theta \in B$. Hence

$$\frac{C_1^2}{C_2C_4} \exp\left[-\tfrac{(1-\phi^2+3d)}{2\sigma^2}k^2\right] \leq (\mathbf{p}_j^{(m,n)})_k \leq \frac{C_2^2}{C_1C_3} \exp\left[-\tfrac{(1-\phi^2-3d)}{2\sigma^2}k^2\right]$$

for all j, m, n, k, ω and $\theta \in B$, as required.

Q.E.D.

Time-decay of dependence

In this section I show that the influence of X_j on the conditional distribution of K_k given some collection of X_t's including X_j diminishes geometrically as $j \to -\infty$ (or ∞). The key is to show that for an appropriate choice of metric, the transition operators ${}_nT^{(t)}$ strictly reduce the distance between probability distributions on $\mathbf{Z}$.

Lemma 2.12 *Let $B \subset \mathcal{P}$ be compact. Then, for all sufficiently small γ, there exists $C > 0$ such that for all m, n, j, ω and all $\theta \in B$*

$$\|\mathbf{p}_j^{(m,n)}\|_\gamma \leq C$$

Proof: By theorem 2.11, for any sufficiently small d there exists C_1 such that

$$(\mathbf{p}_j^{(m,n)})_k \leq C_1 \exp\left[-\tfrac{(1-\phi^2-d)}{2\sigma^2}k^2\right]$$

for all k, j, m, n, ω and $\theta \in B$. Therefore

$$\begin{aligned}
&\|\mathbf{p}_j^{(m,n)}\|_\gamma \\
&\leq C_1 \sum_{k\in\mathbf{Z}} e^{\gamma k^2} \exp\left[-\tfrac{(1-\phi^2-d)}{2\sigma^2}k^2\right] \\
&= C_1 \sum_{k\in\mathbf{Z}} \exp\left[-\tfrac{(1-\phi^2-d-2\gamma\sigma^2)}{2\sigma^2}k^2\right] \\
&\leq C_1 \frac{\sigma\sqrt{2\pi}K(\sigma/\sqrt{1-\phi^2-d-2\gamma\sigma^2})}{\sqrt{1-\phi^2-d-2\gamma\sigma^2}}
\end{aligned}$$

where K is the continuous function of lemma 2.2. Provided, $d + 2\sigma^2\gamma < (1-\phi^2)$ for all $\theta \in B$, this has a uniform upper bound for $\theta \in B$ since B is compact.

$\mathcal{Q.E.D.}$

Lemma 2.13 *Let $B \subset \mathcal{P}$ be compact. Then, for all sufficiently small d and ϵ, there exists C such that, for all j, k and $T \in \mathcal{T}(B)$,*

$$C^{-1}e^{-(1-\epsilon)dk^2}e^{-dj^2}F_{jk}(\epsilon) \leq T_{jk} \leq Ce^{\phi^2 dk^2}e^{dj^2}F_{jk}(-\epsilon)$$

Proof: From the definition of $\mathcal{T}(B)$, T is ${}_nT^{(t)}$ for some n, t, ω and $\theta \in B$. Thus, from the proof of theorem 2.3, we have

$${}_nT_{jk}^{(t)} = \frac{f_{\mathbf{X}_{t+1}^n|Y_t}(\mathbf{x}_{t+1}^n|x_t+j)f_{Y_t|Y_{t-1}}(x_t+j|x_{t-1}+k)}{f_{\mathbf{X}_t^n|Y_{t-1}}(\mathbf{x}_t^n|x_{t-1}+k)}$$

i.e.

$$ {}_nT^{(t)}_{jk} = \frac{g_j f^{(t)}_{jk}}{\sum_{j\in\mathbf{Z}} g_j f^{(t)}_{jk}} \tag{2.27} $$

where

$$ \begin{aligned} g_j &= f_{\mathbf{X}^n_{t+1}|Y_t}(\mathbf{x}^n_{t+1}|x_t + j) \\ &\propto \frac{P[K_t = j|\mathbf{X}^n_t]}{P[K_t = j|X_t]} \end{aligned} $$

Then, by theorem 2.11, since (2.27) is not affected by multiplying $\mathbf{g}$ by an arbitrary scalar, given any sufficiently small δ, there exist $C_1 > 0$ and $C_2 < \infty$ such that

$$ C_1 \exp\left[-\tfrac{\delta}{2\sigma^2} j^2\right] \le g_j \le C_2 \exp\left[\tfrac{\delta}{2\sigma^2} j^2\right] $$

for all j, and $T \in \mathcal{T}(B)$. Take $d = \sup_{\theta\in B} \delta/2\sigma^2$. By lemma 2.4, for all sufficiently small $\epsilon > 0$, there exists $C_3 > 0$ and C_4 such that, for all j, k and $T \in \mathcal{T}(B)$

$$ C_3 F_{jk}(\epsilon) \le f^{(t)}_{jk} \le C_4 F_{jk}(-\epsilon) $$

Lemma 2.5 implies the existence of C_5 such that

$$ \|F_{\cdot k}(-\epsilon)\|_d \le C_5 e^{(1-\epsilon)dk^2} $$

and lemma 2.6 implies the existence of C_6 such that

$$ \|F_{\cdot k}(\epsilon)\|_{-d} \ge C_6 e^{-\phi^2 dk^2} $$

Hence

$$ \frac{C_1 C_3 e^{-dj^2} F_{jk}(\epsilon)}{C_2 C_4 C_5 e^{(1-\epsilon)dk^2}} \le T_{jk} \le \frac{C_2 C_4 e^{dj^2} F_{jk}(-\epsilon)}{C_1 C_3 C_6 e^{-\phi^2 dk^2}} $$

from which the result follows.

$\mathcal{Q.E.D.}$

Lemma 2.14 *Let $B \subset \mathcal{P}$ be compact. Then there exist $L \in \mathbf{N}$, $\Gamma > 0$ and $\delta < 1$ such that for all $\gamma \in (0, \Gamma)$, $|k| \ge L$, and $T \in \mathcal{T}(B)$,*

$$ \|T_{\cdot k}\|_\gamma \le e^{\delta\gamma k^2} $$

Proof: For every $T \in \mathcal{T}(B)$ and integer k, T_{jk} as a function of j defines a probability distribution on $\mathbf{Z}$ and hence on $\mathbf{R}$. Denote this distribution function by F_k. Thus $\|T_{.k}\|_\gamma$ is simply the expectation of $e^{\gamma x^2}$ with respect to F_k, i.e.

$$\|T_{.k}\|_\gamma = \int_{-\infty}^{\infty} e^{\gamma x^2} dF_k(x) = \int_0^{\infty} e^{\gamma x^2} d\tilde{F}_k(x)$$

where $\tilde{F}_k(x) = F_k(x) - F_k(-x)$. Now note that, if h is a positive increasing function on $\mathbf{R}^+$ and F and G are distribution functions on $\mathbf{R}^+$ such that $F(x) \geq G(x)$ for all $x \in \mathbf{R}^+$,

$$\begin{aligned}
\int_0^\infty h(x)dF(x) &= \int_0^1 h(F^{-1}(z))dz \\
&\leq \int_0^1 h(G^{-1}(z))dz \\
&= \int_0^\infty h(x)dG(x)
\end{aligned}$$

The proof of this lemma is then largely one of finding an appropriate sequence of distribution functions G_k such that $\tilde{F}_k(x) \geq G_k(x)$ for all $x \geq 0$ and all possible $\tilde{F}_k(.)$.

Let W_k be a random variable having distribution function $F_k(.)$. Then, applying lemma 2.13, for sufficiently small d and ϵ_1, there exists C_1 such that, for all $T \in \mathcal{T}(B)$

$$P[|W_k| \geq J] = \sum_{|j| \geq J} T_{jk}$$

$$\leq \sum_{|j|\geq J} C_1 e^{\phi^2 dk^2} e^{dj^2} F_{jk}(-\epsilon_1)$$

$$= C_1 e^{\phi^2 dk^2} \exp\left[\tfrac{d(1-\epsilon_1)}{(1-\epsilon_1-2d\sigma^2)}\phi^2 k^2\right] \sum_{|j|\geq J} \exp\left[-\tfrac{1-\epsilon_1-2d\sigma^2}{2\sigma^2}(j - \tfrac{1-\epsilon_1}{1-\epsilon_1-2d\sigma^2}\phi k)^2\right]$$

$$\leq 2C_1 e^{\phi^2 dk^2} \exp\left[\tfrac{d(1-\epsilon_1)}{1-\epsilon_1-2d\sigma^2}\phi^2 k^2\right] \sum_{j\geq J} \exp\left[-\tfrac{1-\epsilon_1-2d\sigma^2}{2\sigma^2}(j - \tfrac{1-\epsilon_1}{1-\epsilon_1-2d\sigma^2}|\phi||k|)^2\right]$$

Now let d and ϵ_1 be so small that, for all $\theta \in B$

$$\sqrt{|\phi|} > \frac{1-\epsilon_1}{1-\epsilon_1-2d\sigma^2}|\phi|$$

and

$$\phi^2 d + \frac{d(1-\epsilon_1)}{1-\epsilon_1-2d\sigma^2} < \left(\sqrt{|\phi|} - \frac{1-\epsilon_1}{1-\epsilon_1-2d\sigma^2}|\phi|\right)^2 \frac{1-\epsilon_1-2d\sigma^2}{2d\sigma^2} \tag{2.28}$$

Since B is compact and the functions on each side of (2.28) are continuous, there exists a minimum difference, C_2 say, between the two sides. Then, putting $\epsilon_2 = \sup_{\theta \in B} \epsilon_1 + 2d\sigma^2$ and $\tilde{\sigma} = \sup_{\theta \in B} \sigma$

$$P[|W_k| \geq J] \leq 2C_1 e^{-C_2 k^2} \sum_{j \geq J} \exp\left[-\tfrac{1-\epsilon_2}{2\tilde{\sigma}^2}(j - \sqrt{|\phi|}|k|)^2\right]$$

for all k, $J \geq \sqrt{|\phi|}|k|$ and $T \in \mathcal{T}(B)$. And so for any real $w \geq \sqrt{|\phi|}|k| + 1$

$$P[|W_k| \geq w] \leq 2C_1 e^{-C_2 k^2} \int_{w-1}^{\infty} \exp\left[-\tfrac{1-\epsilon_2}{2\tilde{\sigma}^2}(w - \sqrt{|\phi|}k)^2\right] dw$$

Let L_1 be such that $C_1 e^{-C_2 k^2} \leq \sqrt{1-\epsilon_2}/(\tilde{\sigma}\sqrt{2\pi})$ for $k \geq L_1$ and all $T \in \mathcal{T}(B)$. Put $\psi_k = |k|\sqrt{|\phi|} + 1$ and define

$$G_k(z) = \begin{cases} 0 & : 0 < z < \psi_k \\ 2\frac{\sqrt{1-\epsilon_2}}{\tilde{\sigma}\sqrt{2\pi}} \int_0^{z-\psi_k} \exp\left[-\frac{(1-\epsilon_2)}{2\tilde{\sigma}^2}x^2\right] dx & : \psi_k \leq z \leq \infty \end{cases}$$

Then, for $z > \psi_k$

$$\begin{aligned} G_k(z) &= 1 - 2\frac{\sqrt{1-\epsilon_2}}{\tilde{\sigma}\sqrt{2\pi}} \int_{z-\psi_k}^{\infty} \exp\left[-\tfrac{(1-\epsilon_2)}{2\tilde{\sigma}^2}x^2\right] dx \\ &= 1 - \frac{\sqrt{1-\epsilon_2}}{\tilde{\sigma}\sqrt{2\pi}} \int_{z-1}^{\infty} \exp\left[-\tfrac{(1-\epsilon_2)}{2\tilde{\sigma}^2}(x - \sqrt{\phi}k)^2\right] dx \\ &\leq P[|W_k| \leq z] \end{aligned}$$

for all $k \geq L_1$ and $T \in \mathcal{T}(B)$. But

$$\int_0^{\infty} \exp\left[\tfrac{\gamma(1-\epsilon_2)}{2\tilde{\sigma}^2}x^2\right] dG_k(x)$$

$$= 2\frac{\sqrt{1-\epsilon_2}}{\tilde{\sigma}\sqrt{2\pi}} \int_{\psi_k}^{\infty} \exp\left[\tfrac{\gamma(1-\epsilon_2)}{2\tilde{\sigma}^2}x^2\right] \exp\left[-\tfrac{1-\epsilon_2}{2\tilde{\sigma}^2}(x-\psi_k)^2\right] dx$$

$$= 2\frac{\sqrt{1-\epsilon_2}}{\tilde{\sigma}\sqrt{2\pi}} \exp\left[\tfrac{\gamma(1-\epsilon_2)}{2(1-\gamma)\tilde{\sigma}^2}\psi_k^2\right] \int_{\psi_k}^{\infty} \exp\left[-\tfrac{(1-\gamma)(1-\epsilon_2)}{2\tilde{\sigma}^2}(x - \tfrac{\psi_k}{1-\gamma})^2\right] dx$$

$$\leq \exp\left[\tfrac{\gamma(1-\epsilon_2)}{2(1-\gamma)\tilde{\sigma}^2}\psi_k^2\right] \left\{\frac{1}{\sqrt{1-\gamma}} + 2\frac{\sqrt{1-\epsilon_2}}{\tilde{\sigma}\sqrt{2\pi}}\psi_k(\frac{1}{1-\gamma} - 1)\right\}$$

Choose $\Gamma \leq \frac{1}{2}$ and $L_2 \geq L_1$ so that, for all $\gamma \in (0, \Gamma)$ and $|k| \geq L_2$, $\sqrt{|\phi|}k^2 \geq (\sqrt{|\phi|}k + 1)^2/(1-\gamma)$. Then there exists $C_3 > 0$ such that, for all $\gamma \in (0, \Gamma)$, $(1-\gamma)^{-\frac{1}{2}} \leq 1 + C_3\gamma$, and so

$$\int_0^{\infty} \exp\left[\tfrac{\gamma(1-\epsilon_2)}{2\tilde{\sigma}^2}x^2\right] dG_k(x)$$

$$\leq \exp\left[\tfrac{\gamma(1-\epsilon_2)}{2\tilde{\sigma}^2}\sqrt{|\phi|}k^2\right]\{1 + C_3\gamma + C_4(\sqrt{|\phi|}k + 1)\gamma\} \tag{2.29}$$

where $C_4 = 4\sqrt{1-\epsilon_2}/(\tilde{\sigma}\sqrt{2\pi})$. Now choose $L_3 \geq L_2$ so that , for all $|k| \geq L_3$,

$$C_3 + C_4(\sqrt{|\phi|}k + 1) \leq \frac{(1-\epsilon_2)}{2\tilde{\sigma}^2}k^2\tfrac{1}{2}(1 - \sqrt{|\phi|})$$

Then the right-side of (2.29) is bounded above by

$$\exp\left[\tfrac{\gamma(1-\epsilon_2)}{2\tilde{\sigma}^2}\tfrac{1}{2}(1 + \sqrt{|\phi|})k^2\right]$$

for $\gamma \in (0, \Gamma)$ and $k \geq L_3$ and so

$$\begin{aligned} \|T_{.k}\|_{\gamma(1-\epsilon_2)/2\sigma^2} &\leq \int_0^\infty \exp\left[\tfrac{\gamma(1-\epsilon_3)}{2\tilde{\sigma}^2}x^2\right] dG_k(x) \\ &\leq \exp\left[\tfrac{1}{2}(1 + \sqrt{|\phi|})\tfrac{\gamma(1-\epsilon_2)}{2\tilde{\sigma}^2}k^2\right] \end{aligned}$$

for all $\gamma \in (0, \Gamma)$, $|k| \geq L_3$ and $T \in \mathcal{T}(B)$, as required.

$\mathcal{Q.E.D.}$

Lemma 2.15 *Let $B \subset \mathcal{P}$ be compact. Then, for all sufficiently small γ, there exists C such that for all n, all $T_1, \ldots, T_n \in \mathcal{T}(B)$ and all positive $\mathbf{a}$*

$$\|T_1 \cdots T_n \mathbf{a}\|_\gamma \leq C\|\mathbf{a}\|_\gamma$$

Proof: By lemma 2.14, there exists L and $\delta < 1$ such that, for all $|k| > L$ and all $T \in \mathcal{T}(B)$, $\|T_{.k}\|_\gamma \leq e^{\delta\gamma k^2}$. But, by lemma 2.13, for any sufficiently small d and ϵ, there exists C_1 such that, for all $T \in \mathcal{T}(B)$

$$T_{jk} \leq C_1 e^{\phi^2 dk^2} e^{dj^2} F_{jk}(-\epsilon)$$

and so, by lemma 2.5, there exists C_2 and C_3 such that

$$\|T_{.k}\|_\gamma \leq C_1 e^{\phi^2 dk^2}\|F_{.k}(-\epsilon)\|_{\gamma+d} \leq C_1 C_2 e^{\phi^2 dk^2} e^{(1-\epsilon)(\gamma+d)k^2} \leq C_3$$

for all $T \in \mathcal{T}(B)$ and $|k| \leq L$. Thus, for all positive $\mathbf{a}$ and all $T \in \mathcal{T}(B)$

$$\|T\mathbf{a}\|_\gamma \leq C_3 \sum_{|j|\leq L} e^{\gamma j^2}|a_j| + e^{-(1-\delta)\gamma L^2} \sum_{|j|>L} e^{\gamma j^2}|a_j|$$

From this it is easily shown that $\|T\mathbf{a}\|_\gamma \leq \|\mathbf{a}\|_\gamma$ if

$$\frac{\sum_{|j|\leq L} e^{\gamma j^2}|a_j|}{\|\mathbf{a}\|_\gamma} \leq \frac{1 - e^{-(1-\delta)\gamma L^2}}{C_3 - e^{-(1-\delta)\gamma L^2}} \tag{2.30}$$

Restrict for the moment to those $\mathbf{a}$ for which $\|\mathbf{a}\|_0 = 1$. Clearly $\sum_{|j|\le L} e^{\gamma j^2}|a_j| \le e^{\gamma L^2}$. Therefore, by (2.30), there exists C_4 so that, for all $\mathbf{a}$ satisfying $\|\mathbf{a}\|_\gamma \ge C_4$ and all $T \in \mathcal{T}(B)$, $\|T\mathbf{a}\|_\gamma \le \|\mathbf{a}\|_\gamma$. But T being stochastic implies $\|T\mathbf{a}\|_0 = \|\mathbf{a}\|_0$ and so if $\|\mathbf{a}\|_\gamma \ge C_4$

$$\|T_1 \cdots T_n \mathbf{a}\|_\gamma \le \|\mathbf{a}\|_\gamma$$

for all $T_1, \ldots, T_n \in \mathcal{T}(B)$. However $\|\mathbf{a}\|_0 = 1$ implies $\|\mathbf{a}\|_\gamma \ge 1$ and so, for all $\mathbf{a}$,

$$\|T_1 \cdots T_n \mathbf{a}\|_\gamma \le C_4 \|\mathbf{a}\|_\gamma \tag{2.31}$$

Now remove the restriction that $\|\mathbf{a}\|_0 = 1$. (2.31) implies that

$$\|T_1 \cdots T_n \mathbf{b}\|_\gamma \le \|\mathbf{b}\|_\gamma$$

where $\mathbf{b} = \mathbf{a}/\|\mathbf{a}\|_0$. The result follows from the linearity of $\|.\|_\gamma$ and the operators $T_1, \ldots, T_n$ if we multiply through by $\|\mathbf{a}\|_0$.

Q.E.D.

Theorem 2.16 *Let $B \subset \mathcal{P}$ be compact. For all sufficiently small $\gamma > 0$ there exists $\epsilon > 0$ such that for all $T \in \mathcal{T}(B)$ and all probability vectors* $\mathbf{a}$, $\mathbf{b}$.

$$D_\gamma(T\mathbf{a}, T\mathbf{b}) \le (1-\epsilon) D_\gamma(\mathbf{a}, \mathbf{b})$$

Proof: Clearly the theorem is equivalent to showing that, for all sufficiently small $\gamma > 0$, there exists $\epsilon > 0$ such that

$$\sup_{\mathbf{c}} \frac{\|T\mathbf{c}\|_\gamma}{\|\mathbf{c}\|_\gamma} \le 1 - \epsilon$$

where the supremum is taken over those $\mathbf{c}$ for which $\sum c_j = 0$. The proof proceeds in two stages. First we shall see that

$$\sup_{\mathbf{c}} \frac{\|T\mathbf{c}\|_\gamma}{\|\mathbf{c}\|_\gamma} \le \sup_{j,k \in \mathbf{Z}} \frac{\|T_{.j} - T_{.k}\|_\gamma}{e^{\gamma j^2} + e^{\gamma k^2}} \tag{2.32}$$

To show this we note that, if $\mathbf{c}$ can be written as $\mathbf{c} = \mathbf{c}_1 + \mathbf{c}_2$ where $\|\mathbf{c}_1 + \mathbf{c}_2\|_\gamma = \|\mathbf{c}_1\|_\gamma + \|\mathbf{c}_2\|_\gamma$, then

$$\frac{\|T\mathbf{c}\|_\gamma}{\|\mathbf{c}\|_\gamma} \le \frac{\|T\mathbf{c}_1\|_\gamma + \|T\mathbf{c}_2\|_\gamma}{\|\mathbf{c}_1\|_\gamma + \|\mathbf{c}_2\|_\gamma} \le \max\left(\frac{\|T\mathbf{c}_1\|_\gamma}{\|\mathbf{c}_1\|_\gamma}, \frac{\|T\mathbf{c}_2\|_\gamma}{\|\mathbf{c}_2\|_\gamma}\right)$$

Set $c^+ = \max_j c_j$ and $c^- = \min_j c_j$ and $\alpha = \min(c^+, |c^-|)$. Let k_1 be such that $c_{k_1} \geq \alpha$ and k_2 be such that $c_{k_2} \leq -\alpha$. Then, putting $\mathbf{c}_1$ to be the vector

$$(\mathbf{c})_j = \begin{cases} \alpha & : j = k_1 \\ -\alpha & : j = k_2 \\ 0 & : \text{otherwise} \end{cases}$$

and $\mathbf{c}_2 = \mathbf{c} - \mathbf{c}_1$, is a decomposition of $\mathbf{c}$ of the above kind. However

$$\frac{\|T\mathbf{c}_1\|_\gamma}{\|\mathbf{c}_1\|_\gamma} = \frac{\|T_{.k_1} - T_{.k_2}\|_\gamma}{e^{\gamma k_1^2} + e^{\gamma k_2^2}}$$

Repeating this decomposition on $\mathbf{c}_2$ inductively, proves (2.32). In fact the inequality in (2.32) is easily seen to be an equality.

Lemma 2.14 shows that there exists $\delta < 1$, $\Gamma_1 > 0$ and $L \in \mathbf{N}$ such that if $|k| > L$ and $\gamma <= \Gamma_1$

$$\frac{\|T_{.k}\|_\gamma}{e^{\gamma k^2}} \leq e^{-(1-\delta)\gamma k^2}$$

Therefore, if $|j|, |k| > L$,

$$\frac{\|T_{.j} - T_{.k}\|_\gamma}{e^{\gamma j^2} + e^{\gamma k^2}} \leq e^{-(1-\delta)\gamma L^2}$$

Lemma 2.15 shows that there exists C_1 and γ_2 such that

$$\|T_{.k}\|_{\Gamma_2} \leq C_1 e^{\Gamma_2 k^2} \leq C_1 e^{\Gamma_2 L^2} \stackrel{\text{def}}{=} C_2$$

when $|k| \leq L$. From the definition of the norm

$$\frac{d}{d\gamma}\|T_{.k}\|_\gamma = \sum j^2 e^{\gamma j^2} T_j k$$

Clearly we can choose C_3 so that $j^2 \leq C_3 e^{\frac{1}{2}\Gamma j^2}$ for all j. Then, provided $|k| \leq L$ and $\gamma \leq \frac{1}{2}\Gamma_2$,

$$\|T_{.k}\|_\gamma \leq 1 + C_3 C_2 \gamma \tag{2.33}$$

Therefore, if $|k| \leq L$, $|j| > L$ and $\gamma \leq \Gamma_3 \stackrel{\text{def}}{=} \min(\Gamma_1, \frac{1}{2}\Gamma_2)$

$$\begin{aligned} \frac{\|T_{.j} - T_{.k}\|_\gamma}{e^{\gamma j^2} + e^{\gamma k^2}\|} &\leq \frac{e^{\delta\gamma j^2} + (1 + C_4\gamma)}{e^{\gamma j^2} + 1} \\ &\leq \min\left(\frac{e^{\delta\gamma j^2}}{\delta e^{\gamma j^2} + (1-\delta)}, \frac{1 + C_4\gamma}{(1-\delta)e^{\gamma j^2} + \delta}\right) \end{aligned}$$

where $C_4 = C_2 C_3$.

But, by Taylor expansion, it is easily seen that $e^{\delta x} < \delta e^x + (1-\delta)$ for all $x > 0$. Further $g_1(x) \stackrel{\text{def}}{=} e^{\delta x} / \Big(\delta e^x + (1-\delta)\Big) \to 0$ as $x \to \infty$. Hence, since g_1 is continuous, given $x > 0$ there exists $\epsilon_1(x) > 0$ such that $g_1(y) \le 1 - \epsilon_1(x)$ when $y \ge x$. Hence, given M, there exists $\epsilon_2(\gamma)$ such that $g_1(\gamma j^2) \le 1 - \epsilon_2$ whenever $|j| \ge M$. On the other hand

$$\frac{1 + C_4\gamma}{(1-\delta)e^{\gamma j^2} + \delta} \le \frac{1 + C_4\gamma}{1 + (1-\delta)\gamma j^2}$$

Hence, if $(1-\delta)M^2 > C_4$, there exists $\epsilon_3(\gamma)$ so that, whenever $|j| > M$,

$$\frac{1 + C_4\gamma}{(1-\delta)e^{\gamma j^2} + \delta} \le 1 - \epsilon_3$$

Finally, we must consider the case when $|k| \le L$ and $|j| \le M$. By lemma 2.13, there exist C_5, d and $\epsilon_4 > 0$ such that, for all $T \in \mathcal{T}(B)$

$$\begin{aligned} T_{0k} &\ge C_5 F_{0k}(\epsilon_4) e^{-(1-\epsilon_4)dk^2} \\ &= C_5 \exp\left[-\tfrac{1+\epsilon_4}{2\sigma^2}\phi^2 k^2\right] e^{-(1-\epsilon_4)dk^2} \end{aligned}$$

Hence, since B is compact, for each k, there exists $b_k > 0$ such that $T_{0k} \ge b_k$ for all $T \in \mathcal{T}(B)$. Denote by b' the minimum of the b_k for $|k| \le M$. By the same argument as led to (2.33), there exist C_6 and Γ_4 so that whenever $|k| \le M$ and $\gamma \le \Gamma_4$

$$\|T_{.k}\|_\gamma \le 1 + C_6\gamma$$

for all $T \in \mathcal{T}(B)$.

Therefore, if $|j| \le L$, $|k| \le M$ and $\gamma \le \Gamma_4$

$$\begin{aligned} \|T_{.j} - T_{.k}\|_\gamma &= \sum_l e^{\gamma l^2} |T_{lj} - T_{lk}| \\ &\le \sum_l e^{\gamma l^2}(T_{lj} + T_{lk}) - b_j - b_k \\ &\le 2 + 2\gamma C_6 - 2b' \end{aligned}$$

Hence, provided $\gamma \le \min(b'/2C_6, \Gamma_4) \stackrel{\text{def}}{=} \Gamma_5$ and $|j| \le L$, $|k| \le M$

$$\frac{\|T_{.j} - T_{.k}\|_\gamma}{e^{\gamma j^2} + e^{\gamma k^2}} \le 1 - b'$$

The desired result follows for any $\gamma \le \min(\Gamma_1, \Gamma_3, \Gamma_5)$ by taking

$$\epsilon = \min(1 - e^{-(1-\delta)\gamma L^2}, \epsilon_2(\gamma), \epsilon_3(\gamma), b')$$

Q.E.D.

Mixing for the K_t

Theorem 2.16 is the crucial component for the proof of the theorem below, which shows that independent of the realisation of the X-process the decay of dependence on the past is extremely rapid.

Lemma 2.17 *Let Z_t, $t = 1, \ldots, n$ be an integer-valued Markov chain having initial distribution $\mathbf{p}_1$ for Z_1 and transition matrices $T^{(t)}$, $t = 2, \ldots, n$ from Z_{t-1} to Z_t such that*

$$D_\gamma(T^{(t)}\mathbf{a}, T^{(t)}\mathbf{b}) \leq (1-\epsilon)D_\gamma(\mathbf{a}, \mathbf{b})$$

for all $\mathbf{a}$, $\mathbf{b}$ and t, and such that the time-reversed process Z_t, $t = n, \ldots, 1$ is also a Markov chain. Let $\mathbf{p}_j$ denote the vector $P[Z_j = .]$ obtained by applying the transition matrices to $\mathbf{p}_1$. Then there exists C depending only on $sup_{1\leq j\leq n}\|\mathbf{p}_j\|_\gamma$ such that for all $1 \leq j_1, j_2 \leq n$, $\mathcal{A} \in \mathcal{F}_{j_2}^t(Z)$ and $\mathcal{B} \in \mathcal{F}_1^{j_1}(Z)$

$$|P_Z[\mathcal{A} \cap \mathcal{B}] - P_Z[\mathcal{A}]P_Z[\mathcal{B}]| \leq C(1-\epsilon)^{j_2 - j_1}$$

Proof: Corresponding to $\mathcal{A}$ and $\mathcal{B}$ are sets $A \subset \mathbf{Z}^{n-j_2+1}$ and $B \subset \mathbf{Z}^{j_1}$ where $\mathcal{A} = \{\mathbf{Z}_{j_2}^n \in A\}$ and $\mathcal{B} = \{\mathbf{Z}_1^{j_1} \in B\}$. A and B can be decomposed as

$$A = \bigcup_{k_{j_2} \in \mathbf{Z}} A_{k_{j_2}} \times \{k_{j_2}\}$$

and

$$B = \bigcup_{k_{j_1} \in \mathbf{Z}} \{k_{j_1}\} \times B_{k_{j_1}}$$

Then

$$\begin{aligned} P[\mathbf{Z}_{j_2}^n \in A] &= \sum_{k_{j_2} \in \mathbf{Z}} P[\mathbf{Z}_{j_2+1}^n \in A_{k_{j_2}} \cap Z_{j_2} = k_{j_2}] \\ &= \sum_{k_{j_2} \in \mathbf{Z}} P[\mathbf{Z}_{j_2+1}^n \in A_{k_{j_2}} | Z_{j_2} = k_{j_2}] P[Z_{j_2} = k_{j_2}] \end{aligned}$$

and, by time-reversibility,

$$P[\mathbf{Z}_1^{j_1} \in B] = \sum_{k_{j_1} \in \mathbf{Z}} P[\mathbf{Z}_1^{j_1 - 1} \in B_{k_{j_1}} | Z_{j_1} = k_{j_1}] P[Z_{j_1} = k_{j_1}]$$

and

$$P[\mathbf{Z}_{j_2}^n \in A \cap \mathbf{Z}_1^{j_1} \in B]$$

$$= \sum_{k_{j_1}, k_{j_2} \in \mathbf{Z}} P[\mathbf{Z}_{j_2+1}^t \in A_{k_{j_2}} | Z_{j_2} = k_{j_2}] P[\mathbf{Z}_1^{j_1-1} \in B_{k_{j_1}} | Z_{j_1} = k_{j_1}] \times P[Z_{k_2} = k_{j_2} \cap Z_{j_1} = k_{j_1}]$$

Therefore since all probabilities are less than 1

$$\begin{aligned}
&\Big| P[\mathbf{Z}_{j_2}^n \in A \cap \mathbf{Z}_1^{j_1} \in B] - P[\mathbf{Z}_{j_2}^n \in A] P[\mathbf{Z}_1^{j_1} \in B] \Big| \\
\leq& \sum_{k_{j_2}, k_{j_1} \in \mathbf{Z}} \Big| P[Z_{j_2} = k_{j_2} \cap Z_{j_1} = k_{j_1}] - P[Z_{j_2} = k_{j_2}] P[Z_{j_1} = k_{j_1}] \Big| \\
=& \sum_{k_{j_2}, k_{j_1} \in \mathbf{Z}} \Big| P[Z_{j_2} = k_{j_2} | Z_{j_1} = k_{j_1}] - \sum_{k \in \mathbf{Z}} P[Z_{j_2} = k_{j_2} | Z_{j_1} = k] P[Z_{j_1} = k] \Big| \\
&\times P[Z_{j_1} = k_{j_1}]
\end{aligned}$$

But

$$P[Z_{j_2} = k_{j_2} | Z_{j_1} = k_{j_1}] = (T^{(j_2)} T^{(j_2-1)} \cdots T^{(j_1+1)})_{k_{j_2} k_{j_1}}$$

Define the operator Q by $Q = T^{(j_2)} T^{(j_2-1)} \cdots T^{(j_1+1)}$. Let $\mathbf{d}_j$ be the vector $(\mathbf{d}_j)_k = \delta_{jk}$. Then

$$\begin{aligned}
&\sum_{k_{j_2} \in \mathbf{Z}} | P[Z_{j_2} = k_{j_2} | Z_{j_1} = k_{j_1}] - \sum_{k \in \mathbf{Z}} P[Z_{j_2} = k_{j_2} | Z_{j_1} = k] P[Z_{j_1} = k] | \\
=& \sum_{k_{j_2} \in \mathbf{Z}} | (Q\mathbf{d}_{k_{j_1}})_{k_{j_2}} - (Q\mathbf{p}_{j_1})_{k_{j_2}} | \\
\leq& \sum_{k_{j_2} \in \mathbf{Z}} e^{\gamma k_{j_2}^2} | (Q\mathbf{d}_{k_{j_1}})_{k_{j_2}} - (Q\mathbf{p}_{j_1})_{k_{j_2}} | \\
=& D_\gamma (Q\mathbf{d}_{k_{j_1}}, Q\mathbf{p}_{j_1}) \\
\leq& (1-\epsilon) D_\gamma (T^{(j_2-1)} \cdots T^{(j_1+1)} \mathbf{d}_{k_{j_1}}, T^{(j_2-1)} \cdots T^{(j_1+1)} \mathbf{p}_{j_1}) \\
\leq& (1-\epsilon)^{j_2-j_1} D_\gamma (\mathbf{d}_{k_{j_1}}, \mathbf{p}_{j_1}) \\
\leq& (1-\epsilon)^{j_2-j_1} \{ \|\mathbf{d}_{k_{j_1}}\|_\gamma + \|\mathbf{p_{j_1}}\|_\gamma \} \\
=& (1-\epsilon)^{j_2-j_1} \{ e^{\gamma k_{j_1}^2} + \|\mathbf{p}_{j_1}\|_\gamma \}
\end{aligned}$$

and so

$$\begin{aligned}
&| P[\mathbf{Z}_{j_2}^n \in A \cap \mathbf{Z}_1^{j_1} \in B] - P[\mathbf{Z}_{j_2}^n \in A] P[\mathbf{Z}_1^{j_1} \in B] | \\
\leq& (1-\epsilon)^{j_2-j_1} \sum_{k_{j_1} \in \mathbf{Z}} \{ e^{\gamma k_{j_1}^2} + \|\mathbf{p}_{j_1}\|_\gamma \} (\mathbf{p}_{j_1})_{k_{j_1}} \\
=& 2(1-\epsilon)^{j_2-j_1} \|\mathbf{p}_{j_1}\|_\gamma
\end{aligned}$$

which is the desired result.

Q.E.D.

Theorem 2.18 *Let $B \subset \mathcal{P}$ be compact. There exist C and $\rho < 1$ such that for any $1 \le j_1 \le j_2 \le n$, $\mathcal{A} \in \mathcal{F}_{j_2}^t(Z)$, $\mathcal{B} \in \mathcal{F}_1^{j_1}(Z)$, ω and $\theta \in B$*

$$|P[\mathcal{A} \cap \mathcal{B}|\mathbf{X}_1^n] - P[\mathcal{A}|\mathbf{X}_1^n]P[\mathcal{B}|\mathbf{X}_1^n]| \le C\rho^{j_2-j_1}$$

i.e. conditional upon $\mathbf{X}_1^n$, K_t is a strong mixing process with mixing coefficients tending geometrically to zero, uniformly in ω, n and $\theta \in B$.

Proof: This is an immediate consequence of lemma 2.17 together with lemma 2.12, theorems 2.16 and 2.3 and the time-reversibility of the AR(1).

$\mathcal{Q.E.D.}$

Polynomials

In much of the remainder of the chapter we shall be concerned with the behaviour of polynomials in the Y's conditional upon the X-process. The following lemmas will be used extensively.

Lemma 2.19 *Let $B \subset \mathcal{P}$ be compact. Let $g(\mathbf{y}_1^n)$ and $h(\mathbf{y}_1^n)$ be polynomials in $y_1, \ldots, y_n$. Then for all sufficiently small $\gamma > 0$ there exists C such that for all j, $N \ge n + j$, k, ω and $\theta \in B$*

$$E[|g(\mathbf{Y}_{j+1}^{j+n})h(\mathbf{Y}_1^n)||\mathbf{X}_2^N, Y_1 = X_1 + k] \le Ce^{\gamma k^2}$$

Proof: Clearly it suffices to prove the lemma for any g of the form $y_n^{m_n} \cdots y_1^{m_1}$ and any h of the form $y_n^{l_n} \cdots y_1^{l_1}$. But

$$E[|Y_{j+n}|^{m_n} \cdots |Y_{j+1}|^{m_1}|Y_n|^{l_n} \cdots |Y_1^{l_1}||\mathbf{X}_2^N, Y_1 = X_1 + k_1]$$

$$= \sum_{\mathbf{k}_2^{j+n} \in \mathbf{Z}^{n+j-1}} |x_{j+n} + k_{j+n}|^{m_n} \cdots |x_{j+1} + k_{j+1}|^{m_1}|x_n + k_n|^{l_n} \cdots |x_1 + k_1|^{l_1}$$

$$\times P[\mathbf{K}_1^{j+n} = \mathbf{k}_1^{j+n}|\mathbf{X}_1^N, K_1 = k_1]$$

$$\le \sum_{\mathbf{k}_2^{j+n}} (1 + |k_{j+n}|)^{m_n} \cdots (1 + |k_{j+1}|)^{m_1}(1 + |k_n|)^{l_n} \cdots (1 + |k_1|)^{l_1} T_{k_{j+n}k_{j+n-1}}^{(j+n)} \cdots T_{k_2k_1}^{(2)}$$

$$\le \sum_{\mathbf{k}_2^{j+n}} (1 + |k_{j+n}|)^{M} \cdots (1 + |k_{j+1}|)^{M}(1 + |k_n|)^{M} \cdots (1 + |k_1|)^{M} T_{k_{j+n}k_{j+n-1}}^{(j+n)} \cdots T_{k_2k_1}^{(2)}$$

where $M = \max(m_1, \ldots, m_n, l_1, \ldots, l_n)$. Now define Q to be the diagonal matrix with elements $Q_{jj} = (1+|j|)^M$ and let $\mathbf{d}_j$ denote the vector $(\mathbf{d}_j)_k = \delta_{jk}$. Then

$$E[|Y_{j+n}|^{m_n} \cdots |Y_{j+1}|^{m_1} |Y_n|^{l_n} \cdots |Y_1^{l_1}||\mathbf{X}_2^N, Y_1 = X_1 + k]$$

$$\leq \|QSQ\mathbf{d}_k\|_0$$

where S is an operator consisting of the product of $T^{(2)}, \ldots, T^{(j+n)}$ and $2n-2$ copies of the Q operator. Now let Γ be as in lemma 2.14. Then, by lemma 2.15, for any $\gamma \in (0, \Gamma)$, there exists c_γ such that $\|T_1 \cdots T_K \mathbf{a}\|_\gamma \leq c_\gamma \|\mathbf{a}\|_\gamma$ for any positive $\mathbf{a}$, any K, and any $T_1, \ldots, T_K \in \mathcal{T}(B)$. Let γ be so given. Clearly there exists C such that $(1+|k|)^M \leq C \exp[\gamma k^2/2n]$ for all k, and hence $\|Q\mathbf{a}\|_\alpha \leq C\|\mathbf{a}\|_{\alpha+\gamma/2n}$ for any $\alpha \geq 0$ and positive $\mathbf{a}$. Therefore,

$$\|QSQ\mathbf{d}_k\|_0$$

$$\leq C\|SQ\mathbf{d}_k\|_{\gamma/2n}$$

$$\leq Cc_{\gamma/2n}Cc_{2\gamma/2n}C \cdots Cc_{\gamma-\gamma/2n}\|Q\mathbf{d}_k\|_\gamma$$

$$= C^{2n-1}\left(\prod_{j=1}^{2n-1} c_{j\gamma/2n}\right)\|Q\mathbf{d_k}\|_{\gamma-\gamma/2n}$$

$$\leq C^{2n} e^{\gamma k^2} \prod_{j=1}^{2n-1} c_{j\gamma/2n}$$

as required.

Q.E.D.

Lemma 2.20 *Let $B \subset \mathcal{P}$ be compact. Let $g(\mathbf{y}_1^n)$ be a polynomial in $y_1, \ldots, y_n$. Then there exists C such that for all $m \leq 1$, $M \geq n$, ω and $\theta \in B$*

$$E[|g(\mathbf{Y}_1^n)||\mathbf{X}_m^M] \leq C$$

Proof: By lemma 2.19, for all sufficiently small γ, there exists C_1 such that

$$\begin{aligned} E[|g(\mathbf{Y}_1^n)||\mathbf{X}_m^M] &= \sum_{k_1 \in \mathbf{Z}} E[|g(\mathbf{Y}_1^n)||\mathbf{X}_2^M, Y_1 = X_1 + k_1] P[K_1 = k_1|\mathbf{X}_m^M] \\ &\leq C_1 e^{\gamma k_1^2} P[K_1 = k_1|\mathbf{X}_m^M] \\ &= C_1 \|\mathbf{p}_1^{(m,M)}\|_\gamma \end{aligned}$$

for all $m \leq 1$, $M \geq n$, ω and $\theta \in B$. The result follows by applying lemma 2.12.

Q.E.D.

Lemma 2.21 *Let $B \subset \mathcal{P}$ be compact. Let $g(\mathbf{y}_1^n)$ and $h(\mathbf{y}_1^n)$ be polynomials in $y_1, \ldots, y_n$. Then there exists C and $\rho < 1$ such that for all $m_1, m_2 < 1$, $j \geq 1$, $M \geq n + j$, ω and $\theta \in B$*

$$|E[g(\mathbf{Y}_{j+1}^{j+n})h(\mathbf{Y}_1^n)|\mathbf{X}_{m_1}^M] - E[g(\mathbf{Y}_{j+1}^{j+n})h(\mathbf{Y}_1^n)|\mathbf{X}_{m_2}^M]| \leq C\rho^{\min(|m_1|,|m_2|)}$$

Proof: As before

$$E[g(\mathbf{Y}_{j+1}^{j+n})h(\mathbf{Y}_1^n)|\mathbf{X}_{m_1}^M]$$

$$= \sum_{k_1 \in \mathbf{Z}} E[g(\mathbf{Y}_{j+1}^{j+n})h(\mathbf{Y}_1^n)|\mathbf{X}_2^M, Y_1 = X_1 + k_1]P[K_1 = k_1|\mathbf{X}_{m_1}^M]$$

and so, by lemma 2.19, there exists C_1 such that for all $m_1, m_2 < 1$, $j \geq 1$, $M \geq n+j$, ω and $\theta \in B$

$$|E[g(\mathbf{Y}_{j+1}^{j+n})h(\mathbf{Y}_1^n)|\mathbf{X}_{m_1}^M] - E[g(\mathbf{Y}_{j+1}^{j+n})h(\mathbf{Y}_1^n)|\mathbf{X}_{m_2}^M]|$$

$$\leq \sum_{k_1 \in \mathbf{Z}} E[|g(\mathbf{Y}_{j+1}^{j+n})h(\mathbf{Y}_1^n)||\mathbf{X}_2^M, Y_1 = X_1 + k_1]$$

$$\left|P[K_1 = k_1|\mathbf{X}_{m_1}^M] - P[K_1 = k_1|\mathbf{X}_{m_2}^M]\right|$$

$$\leq C_1 D_\gamma(\mathbf{p}_1^{m_1,M}, \mathbf{p}_1^{m_2,M})$$

Suppose without loss of generality that $|m_1| \leq |m_2|$. Then

$$\begin{aligned} D_\gamma(\mathbf{p}_1^{m_1,M}, \mathbf{p}_1^{m_2,M}) &= D_\gamma(T^{(1)}T^{(0)} \cdots T^{(m_1+1)}\mathbf{p}_{m_1}^{m_1,M}, T^{(1)} \cdots T^{(m_1+1)}\mathbf{p}_{m_1}^{m_2,M}) \\ &\leq (1-\epsilon)^{|m_1+1|} D_\gamma(\mathbf{p}_{m_1}^{m_1,M}, \mathbf{p}_{m_1}^{m_2,M}) \\ &\leq (1-\epsilon)^{|m_1+1|}\{\|\mathbf{p}_{m_1}^{m_1,M}\|_\gamma + \|\mathbf{p}_{m_1}^{m_2,M}\|_\gamma\} \end{aligned}$$

by lemma 2.16. The result follows by applying lemma 2.12.

$\mathcal{Q.E.D.}$

Lemma 2.22 *Let $B \subset \mathcal{P}$ be compact and let $g(\mathbf{y}_1^m)$ and $h(\mathbf{y}_1^m)$ be polynomials in $y_1, \ldots, y_m$. There exists C and $\rho < 1$ such that*

$$\left|C[g(\mathbf{Y}_{j+1}^{j+m}), h(\mathbf{Y}_{k+1}^{k+m})|\mathbf{X}_1^n] - C[g(\mathbf{Y}_{j+1}^{j+m}), h(\mathbf{Y}_{k+1}^{k+m})|\mathbf{X}_1^\infty]\right| \leq C\rho^{n-\max(j,k)}$$

for all ω and $0 \leq j, k \leq n - m$.

Proof: By lemma 2.21 and time reversibility, there exist C_1 and $\rho_1 < 1$ such that, for all ω and n

$$\left|E[g(\mathbf{Y}_{j+1}^{j+m})h(\mathbf{Y}_{k+1}^{k+m})|\mathbf{X}_1^n] - E[g(\mathbf{Y}_{j+1}^{j+m})h(\mathbf{Y}_{k+1}^{k+m})|\mathbf{X}_1^\infty]\right| \leq C_1\rho_1^{n-m-\max(j,k)}$$

$$\left|E[g(\mathbf{Y}_{j+1}^{j+m})|\mathbf{X}_1^n] - E[g(\mathbf{Y}_{j+1}^{j+m})|\mathbf{X}_1^\infty]\right| \leq C_1\rho_1^{n-m-j}$$

and

$$\left|E[h(\mathbf{Y}_{k+1}^{k+m})|\mathbf{X}_1^n] - E[h(\mathbf{Y}_{k+1}^{k+m})|\mathbf{X}_1^\infty]\right| \leq C_1\rho_1^{n-m-k}$$

Also, by lemma 2.20, there exists C_2 such that, for all $N \geq j+m$, $M \leq j+1$ and ω

$$\left|E[g(\mathbf{Y}_{j+1}^{j+m})|\mathbf{X}_M^N]\right| \leq C_2$$

and

$$\left|E[h(\mathbf{Y}_{k+1}^{k+m})|\mathbf{X}_M^N]\right| \leq C_2$$

Hence

$$\left|C[g(\mathbf{Y}_{j+1}^{j+m}), h(\mathbf{Y}_{k+1}^{k+m})|\mathbf{X}_1^n] - C[g(\mathbf{Y}_{j+1}^{j+m}), h(\mathbf{Y}_{k+1}^{k+m})|\mathbf{X}_1^\infty]\right|$$

$$\leq C_1\rho_1^{n-m-\max(j,k)} + C_3C_1\rho_1^{n-m-j} + C_3C_1\rho_1^{n-m-k}$$

$$\leq (1+2C_3)C_1\rho_1^{n-m-\max(j,k)}$$

from which C and ρ can be obtained.

Q.E.D.

Lemma 2.23 *Let $g(\mathbf{y}_1^m)$ and $h(\mathbf{y}_1^m)$ be polynomials in $y_1, \ldots, y_m$. Then there exists C and $\rho < 1$ such that for all n, $0 \leq j_1, j_2 \leq n-m$ and ω*

$$|C[g(\mathbf{Y}_{j_1+1}^{j_1+m}), h(\mathbf{Y}_{j_2+1}^{j_2+m})|\mathbf{X}_1^n]| \leq C\rho^{|j_2-j_1|}$$

Proof: Without loss of generality, suppose $j_2 \geq j_1$. By theorem 2.18 and lemma A.2, there exists ρ_1 such that, for $j_2 \geq j_1 + m$,

$$|C[g(\mathbf{Y}_{j_1+1}^{j_1+m}), h(\mathbf{Y}_{j_2+1}^{j_2+m})|\mathbf{X}_1^n]| \leq \rho_1^{j_2+1-j_1-m} E[|g(\mathbf{Y}_{j_1+1}^{j_1+m})|^2|\mathbf{X}_1^n] E[|h(\mathbf{Y}_{j_2+1}^{j_2+m})|^2|\mathbf{X}_1^n]$$

But the square of a polynomial is a polynomial and hence, by lemma 2.20, there exists C_1 so that, for $j_2 \geq j_1 + m$,

$$|C[g(\mathbf{Y}_{j_1+1}^{j_1+m}), h(\mathbf{Y}_{j_2+1}^{j_2+m})|\mathbf{X}_1^n]| \leq C_1 \rho_1^{j_2+1-j_1-m}$$

But lemma 2.20 also implies the existence of c_j such that for any j_1 and j_2

$$|C[g(\mathbf{Y}_{j_1+1}^{j_1+m}), h(\mathbf{Y}_{j_2+1}^{j_2+m})|\mathbf{X}_1^n]| \leq c_{j_2-j_1}$$

Hence by defining C and ρ appropriately in terms of C_1 and $c_1, \ldots, c_{m-1}$, the result follows.

Q.E.D.

2.2.4 Consistency

The maximum likelihood estimator will now be shown to be consistent — in fact the stronger result of strong convergence will be shown. The proof is in the style of [32] rather than that of [6]. The work in [32] was only for the case of independent and identically distributed sequence of random variables, so considerable extra effort is required. However, the body of the proof of the main statement of the theorem uses the same ideas as those in [32].

Extension of the Likelihood Function

Many of the complications which arise are at the boundary of the parameter space $\mathcal{P}$. The boundaries at $\sigma = 0$ and $\sigma = \infty$ are dealt with directly. The boundary at $|\phi| = 1$ is more difficult and is dealt with by extending the likelihood function to $|\phi| = 1$. The extension is necessary because, for $|\phi| \geq 1$, the AR(1) is not stationary, and so only conditional, as opposed to marginal, distributions exist for the process; hence there is no likelihood function. However, it is both convenient and possible to define the likelihood function for the wrapped AR(1) when $|\phi| = 1$, not just when $|\phi| < 1$. This is because, for $|\phi| = 1$, the wrapped AR(1) is a Markov process which has a stationary distribution. Consider the case when $\phi = 1$. Then

$$\begin{aligned} f_{X_t|Y_{t-1}}(x_t|y_{t-1}) &= \sum_{j \in \mathbf{Z}} f_{Y_t|Y_{t-1}}(x_t + j|y_{t-1}) \\ &= \sum_{j \in \mathbf{Z}} f_{\epsilon_t}(x_t + j - y_{t-1}) \end{aligned}$$

Figure 2.3: Tree illustrating the dependencies between the various lemmata and theorems in section 2.2.4

$$
\begin{aligned}
&= \sum_{j\in\mathbf{Z}} f_{\epsilon_t}(x_t + j - y_{t-1} - k) \\
&= f_{X_t|Y_{t-1}}(x_t|y_{t-1} + k)
\end{aligned}
$$

for any integer k. Hence defining $f_{X_t|X_{t-1}}(x_t|x_{t-1}) = f_{X_t|Y_{t-1}}(x_t|x_{t-1})$ is valid, since the dependence of X_t on Y_{t-1} involves only the information contained in X_{t-1}. Further, since Y_t is a Markov process, $f_{X_t|\mathbf{Y}_1^{t-1}}(x_t|\mathbf{x}_1^{t-1} + \mathbf{k}_1^{t-1}) = f_{X_t|Y_{t-1}}(x_t|x_{t-1} + k_{t-1}) = f_{X_t|X_{t-1}}(x_t|x_{t-1})$ and so

$$
f_{X_t|\mathbf{X}_1^{t-1}}(x_t|\mathbf{x}_1^{t-1}) = f_{X_t|X_{t-1}}(x_t|x_{t-1}) \tag{2.34}
$$

Consequently X_t is a Markov process when $\phi = 1$. In addition the X_t process has a stationary marginal distribution (the uniform distribution), for

$$
\begin{aligned}
\int_{-\frac{1}{2}}^{\frac{1}{2}} f_{X_t|X_{t-1}}(x|y).1\, dy &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \sum_{j\in\mathbf{Z}} f_{\epsilon_t}(x + j - y)\, dy \\
&= \int_{-\infty}^{\infty} f_{\epsilon_t}(y)\, dy \\
&= 1
\end{aligned}
$$

and so the likelihood function is well-defined for $\phi = 1$. The preceding argument requires only trivial modificiation to show that, for $\phi = -1$, X_t is again a Markov process having the uniform distribution as equilibrium distribution.

Now, when $\phi = 1$,

$$
\begin{aligned}
f_{X_t|\mathbf{X}_1^{t-1}}(x_t|\mathbf{x}_1^{t-1}) &= f_{X_t|X_{t-1}}(x_t|x_{t-1}) \\
&= \sum_{j\in\mathbf{Z}} (\sigma\sqrt{2\pi})^{-1} \exp\Big(-\tfrac{1}{2}\sigma^{-2}(x_t + j - (x_{t-1} + k))^2\Big)
\end{aligned}
$$

for any integer k, and so $f_{\mathbf{X}_1^t}(\mathbf{x}_1^t)$ can be written for any value of $k_1 \in \mathbf{Z}$ as

$$
\begin{aligned}
&(\sigma\sqrt{2\pi})^{1-n} \sum_{\mathbf{k}_2^n \in \mathbf{Z}^{n-1}} \exp\Big(-\tfrac{1}{2}\sigma^{-2} \sum_{j=2}^{n} (x_j + k_j - (x_{j-1} + k_{j-1}))^2\Big) \\
&= (\sigma\sqrt{2\pi})^{1-n} \sum_{\mathbf{k}_2^n \in \mathbf{Z}^{n-1}} \exp\Big(-\tfrac{1}{2}\sigma^{-2}(\mathbf{x}_1^n + \mathbf{k}_1^n)^T \mathbf{B}_n^+ (\mathbf{x}_1^n + \mathbf{k}_1^n)\Big) \qquad (2.35)
\end{aligned}
$$

where $\mathbf{B}_n^+$ is the $n \times n$ matrix

$$
\mathbf{B}_n^+ = \begin{pmatrix}
1 & -1 & & & & \\
-1 & 2 & -1 & & & \\
 & -1 & 2 & -1 & & \\
 & & \ddots & \ddots & \ddots & \\
 & & & -1 & 2 & -1 \\
 & & & & -1 & 1
\end{pmatrix}
$$

and all farther off diagonal entries are 0. When $\phi = -1$ the same equation holds with $\mathbf{B}_n^+$ replaced by $\mathbf{B}_n^-$ where

$$\mathbf{B}_n^- = \begin{pmatrix} 1 & 1 & & & & \\ 1 & 2 & 1 & & & \\ & 1 & 2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & 2 & 1 \\ & & & & 1 & 1 \end{pmatrix}$$

Note that $\mathbf{B}_n^+ = \lim_{\phi\to 1}\mathbf{M}_n$ and $\mathbf{B}_n^- = \lim_{\phi\to -1}\mathbf{M}_n$. In fact (2.35) is a continuous extension of the likelihood function since lemma 2.2 shows that

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

uniformly in x as $|\phi| \to 1$. That this extension is, in fact, uniformly continuous in n and ω will be shown as part of the proof of a later lemma. The following lemma will be used as part of that proof.

Lemma 2.24 *Let $\theta \in \mathcal{P}$. Then if $\phi > 0$*

$$\mathbf{x}^T\mathbf{M}_n\mathbf{x} \geq \phi\mathbf{x}^T\mathbf{B}_n^+\mathbf{x} + (1-\phi)x_1^2$$

Otherwise

$$\mathbf{x}^T\mathbf{M}_n\mathbf{x} \geq -\phi\mathbf{x}^T\mathbf{B}_n^-\mathbf{x} + (1+\phi)x_1^2$$

Proof: Consider the case where $\phi > 0$. Then by definition of $\mathbf{M}_n$ and $\mathbf{B}_n^+$

$$\begin{aligned} (\mathbf{x}_1^n)^T\mathbf{M}_n(\mathbf{x}_1^n) &= \phi(\mathbf{x}_1^n)^T\mathbf{B}_n^+(\mathbf{x}_1^n) + (1-\phi)(x_1^2 + x_n^2) + (1+\phi^2-2\phi)\sum_{j=2}^{n-1} x_j^2 \\ &\geq \phi(\mathbf{x}_1^n)^T\mathbf{B}_n^+(\mathbf{x}_1^n) + (1-\phi)x_1^2 \end{aligned}$$

as required. The case when $\phi < 0$ is similar.

$\mathcal{Q.E.D.}$

A Stationary Approximation

In the case of a sequence of independent random variables, the log-likelihood is the sum of a sequence of independent and identically distributed random variables. For a stationary Markov process this becomes the sum of a stationary m-dependent sequence. In the general case such as the wrapped AR(1) things are more complicated. However, for the wrapped AR(1), there exists a useful approximation to the log-likelihood function by a stationary sequence which satisfies the strong law of large numbers. The following lemmas concern this approximation.

Lemma 2.25 *Let $\theta \in \mathcal{P}'$. There exists C and $\rho < 1$ such that, for all t, $j \geq 0$ and ω*

$$\left| f_{X_t|\mathbf{X}_{t-j}^{t-1}}(X_t|\mathbf{X}_{t-j}^{t-1}) - f_{X_t|\mathbf{X}_{t-j-1}^{t-1}}(X_t|\mathbf{X}_{t-j-1}^{t-1}) \right| < C\rho^j$$

Proof: If θ is such that $|\phi| = 1$ the result is a trivial consequence of the fact that X_t is then a Markov process. Therefore, assume $\theta \in \mathcal{P}$. From (2.3), (2.19), (2.20) and theorem 2.3

$$\begin{aligned}
&\left| f_{X_t|\mathbf{X}_{t-j}^{t-1}}(X_t|\mathbf{X}_{t-j}^{t-1}) - f_{X_t|\mathbf{X}_{t-j-1}^{t-1}}(X_t|\mathbf{X}_{t-j-1}^{t-1}) \right| \\
&= \left| \sum_{j,k \in \mathbf{Z}} f_{jk}^{(t)} \left\{ (\mathbf{p}_{t-1}^{(t-j,t-1)})_k - (\mathbf{p}_{t-1}^{(t-j-1,t-1)})_k \right\} \right| \\
&\leq K(\sigma) \|\mathbf{p}_{t-1}^{(t-j,t-1)} - \mathbf{p}_{t-1}^{(t-j-1,t-1)}\|_0 \\
&\leq K(\sigma) \|\mathbf{p}_{t-1}^{(t-j,t-1)} - \mathbf{p}_{t-1}^{(t-j-1,t-1)}\|_\gamma \\
&= K(\sigma) D_\gamma (T^{(t-1)} T^{(t-2)} \cdots T^{(t-j+1)} \mathbf{p}_{t-j}^{(t-j,t-1)}, T^{(t-1)} T^{(t-2)} \cdots T^{(t-j+1)} \mathbf{p}_{t-j}^{(t-j-1,t-1)}) \\
&\leq K(\sigma)(1-\epsilon)^{j-1} \|\mathbf{p}_{t-j}^{(t-j,t-1)} - \mathbf{p}_{t-j}^{(t-j-1,t-1)}\|_\gamma \\
&\leq 2C_1 K(\sigma)(1-\epsilon)^{j-1}
\end{aligned}$$

by invoking lemmas 2.2, 2.16 and 2.12. But C_1 and ϵ are independent of t, j and ω as required.

Q.E.D.

Theorem 2.26 *The definition*

$$f_{X_t|\mathbf{X}_{-\infty}^{t-1}}(x_t|\mathbf{x}_{-\infty}^{t-1}) = \lim_{n \to \infty} f_{X_t|\mathbf{X}_{t-n}^{t-1}}(x_t|\mathbf{x}_{t-n}^{t-1})$$

is well defined for all $\theta \in \mathcal{P}'$. Further, there exists C such that

$$\left| \ln f_{\mathbf{X}_1^n}(\mathbf{X}_1^n) - \sum_{j=1}^{n} \ln f_{X_j|\mathbf{X}_{-\infty}^{j-1}}(X_j|\mathbf{X}_{-\infty}^{j-1}) \right| \leq C$$

for all n and ω.

Proof: This is a obvious consequence of lemmas 2.2 and 2.25.

$\mathcal{Q.E.D.}$

Lemma 2.27 *Let $\theta \in \mathcal{P}'$. There exists $\epsilon > 0$ such that*

$$\frac{1}{n}\ln f_{\mathbf{X}_1^n}(\mathbf{X}_1^n) \stackrel{a.s.}{\to} \epsilon$$

Proof: By theorem 2.26

$$\frac{1}{n}|\ln f_{\mathbf{X}_1^n}(\mathbf{X}_1^n) - \sum_{j=1}^{n} \ln f_{X_j|\mathbf{X}_{-\infty}^{j-1}}(X_j|\mathbf{X}_{-\infty}^{j-1})| \stackrel{\text{a.s.}}{\to} 0$$

But the sequence $\ln f_{X_t|\mathbf{X}_{-\infty}^{t-1}}(X_t|\mathbf{X}_{-\infty}^{t-1})$ is a stationary sequence which, by lemma 1.3 and lemmas 2.2 and 2.25 satisfies the hypotheses of theorem A.3. Hence

$$\frac{1}{n}\sum_{j=1}^{n} \ln f_{X_j|\mathbf{X}_{-\infty}^{j-1}}(X_j|\mathbf{X}_{-\infty}^{j-1}) \stackrel{\text{a.s.}}{\to} E[\ln f_{X_t|\mathbf{X}_{-\infty}^{t-1}}(X_t|\mathbf{X}_{-\infty}^{t-1})]$$

But

$$E[\ln f_{X_t|\mathbf{X}_{-\infty}^{t-1}}(X_t|\mathbf{X}_{-\infty}^{t-1})] = E[E[\ln f_{X_t|\mathbf{X}_{-\infty}^{t-1}}(X_t|\mathbf{X}_{-\infty}^{t-1})|\mathbf{X}_{-\infty}^{t-1}]] \tag{2.36}$$

and, by Jensen's inequality,

$$\begin{aligned} E[-\ln f_{X_t|\mathbf{X}_{-\infty}^{t-1}}(X_t|\mathbf{X}_{-\infty}^{t-1})|\mathbf{X}_{-\infty}^{t-1}] &\leq \ln \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{f_{X_t|\mathbf{X}_{-\infty}^{t-1}}(x_t|\mathbf{X}_{-\infty}^{t-1})} dF_{X_t|\mathbf{X}_{-\infty}^{t-1}}(x_t|\mathbf{X}_{-\infty}^{t-1}) \\ &= 0 \end{aligned}$$

with equality only if $f_{X_t|\mathbf{X}_{-\infty}^{t-1}}(x_t|\mathbf{X}_{-\infty}^{t-1}) = 1$ for almost all $x_t \in [-\frac{1}{2}, \frac{1}{2}]$. Hence (2.36) is negative unless $f_{X_t|\mathbf{X}_{-\infty}^{t-1}}(X_t|\mathbf{X}_{-\infty}^{t-1}) = 1$ almost surely, which would contradict X_t having a wrapped normal distribution.

$\mathcal{Q.E.D.}$

The boundary of the parameter space

Most of the machinery required for the proof has been established in the preceding sections. The following three lemmas deal with problems relating to the behaviour of the likelihood function at the edges of the parameter space $\mathcal{P}$.

Lemma 2.28 $P_{\theta_0}\left[\overline{\lim}_{n\to\infty}\hat{\sigma}_n = \infty\right] = 0$ *for all $\theta_0 \in \mathcal{P}$.*

Proof: By lemma 2.27 there exists $\epsilon > 0$ such that

$$P_{\theta_0}\Big[\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \ln f_{X_j|\mathbf{X}_1^{j-1}}(X_j|\mathbf{X}_1^{j-1};\theta_0) = \epsilon\Big] = 1 \tag{2.37}$$

and by lemma 16, there exists a function $K(\sigma)$ such that for all $\theta \in \mathcal{P}'$, t and ω

$$K^{-1}(\sigma) < f_{X_t|\mathbf{x}_1^{t-1}}(x_t|\mathbf{x}_1^{t-1}) < K(\sigma)$$

and such that $\lim_{\sigma\to\infty} K(\sigma) = 1$. Thus we can choose σ_1 so that for all $\theta \in \mathcal{P}'$ with $\sigma > \sigma_1$ and for all t and ω

$$|\ln f_{X_t|\mathbf{X}_1^{t-1}}(X_t|\mathbf{X}_1^{t-1})| < \epsilon/2 \tag{2.38}$$

Therefore, combining (2.37) and (2.38)

$$P_{\theta_0}\Big[\lim_{n\to\infty} \sup_{\theta\in\mathcal{H}} \frac{1}{n}\sum_{j=1}^{n} \ln f_{X_j|\mathbf{x}_1^{j-1}}(x_j|\mathbf{x}_1^{j-1}) - \ln f_{X_j|\mathbf{x}_1^{j-1}}(X_j|\mathbf{X}_1^{j-1};\theta_0) < -\epsilon/2\Big] = 1$$

where $\mathcal{H} = \{\theta \in \mathcal{P}' : \sigma > \sigma_1\}$, and so

$$P_{\theta_0}\Big[\lim_{n\to\infty} \sup_{\theta\in\mathcal{H}} f_{\mathbf{X}_1^n}(\mathbf{x}_1^n) \;/\; f_{\mathbf{X}_1^n}(\mathbf{X}_1^n;\theta_0) = 0\Big] = 1$$

from which the result follows.

$\mathcal{Q.E.D.}$

Lemma 2.29 $P_{\theta_0}[\underline{\lim}_{n\to\infty} \hat{\sigma}_n = 0] = 0$ *for all* $\theta_0 \in \mathcal{P}$.

Proof: $\hat{\theta}_n$ maximises $f_{\mathbf{X}_1^n}(\mathbf{X}_1^n)$. Thus $\partial_\sigma f_{\mathbf{X}_1^n}(\mathbf{X}_1^n)\Big|_{\hat{\theta}_n} = 0$. But, for $\theta \in \mathcal{P}$

$$\begin{aligned}\frac{\partial}{\partial\sigma} f_{\mathbf{X}_1^n}(\mathbf{x}_1^n) &= \frac{\sqrt{1-\phi^2}}{\sigma^n(2\pi)^{n/2}} \sum_{\mathbf{k}_1^n\in\mathbf{Z}^n} \Big\{\frac{-n}{\sigma} + \frac{1}{\sigma^3}(\mathbf{x}_1^n+\mathbf{k}_1^n)^T\mathbf{M}_n(\mathbf{x}_1^n+\mathbf{k}_1^n)\Big\} \\ &\qquad \times \exp\Big(-\tfrac{1}{2}\sigma^{-2}(\mathbf{x}_1^n+\mathbf{k}_1^n)^T\mathbf{M}_n(\mathbf{x}_1^n+\mathbf{k}_1^n)\Big)\end{aligned}$$

and, by lemma 2.1, $\mathbf{x}^T\mathbf{M}_n\mathbf{x} \geq (1-|\phi|)^2\|\mathbf{x}\|^2$. Therefore, if $\sigma^2 < (1-|\phi|)^2\|\mathbf{x}\|^2/n$,

$$\begin{aligned}\frac{-n}{\sigma} + \frac{1}{\sigma^3}(\mathbf{x}_1^n+\mathbf{k}_1^n)^T\mathbf{M}_n(\mathbf{x}_1^n+\mathbf{k}_1^n) &\geq \frac{-n}{\sigma} + \frac{(1-|\phi|)^2}{\sigma^3}\|\mathbf{x}_1^n+\mathbf{k}_1^n\|^2 \\ &\geq \frac{-n}{\sigma} + \frac{(1-|\phi|)^2}{\sigma^3}\|\mathbf{x}_1^n\|^2 \\ &> 0\end{aligned}$$

and hence $\dfrac{\partial}{\partial\sigma} f_{\mathbf{X}_1^n}(\mathbf{x}_1^n) > 0$. Thus

$$\hat{\sigma}_n^2 \geq (1-|\hat{\phi}_n|)^2\frac{\|\mathbf{X}_1^n\|^2}{n}$$

since this holds trivially when $|\hat{\phi}_n| = 1$. But, since the AR(1) process is ergodic, $n^{-1}\|\mathbf{X}_1^n\|^2 \overset{\text{a.s.}}{\to} C_1$ in θ_0-measure where $C_1 = E_{\theta_0}[X_t^2] > 0$. Thus, with probability 1, $\underline{\lim}_{n\to\infty} \hat{\sigma}_n^2 = 0$ implies $\overline{\lim}_{n\to\infty} |\hat{\phi}_n| = 1$. Let ω be some realisation of the process for which $\underline{\lim}_{n\to\infty} \hat{\sigma}_n^2 = 0$ and $\overline{\lim}_{n\to\infty} \hat{\phi}_n = 1$. Then there exist $n_1, n_2 \ldots \in \mathbf{N}$ such that $\lim_{j\to\infty} \hat{\sigma}_{n_j} = 0$ and $\lim_{j\to\infty} \hat{\phi}_{n_j} = 1$. By lemma 9, for any positive ϕ

$$(\mathbf{x}_1^n + \mathbf{k}_1^n)^T \mathbf{M}_n (\mathbf{x}_1^n + \mathbf{k}_1^n) \geq \phi(\mathbf{x}_1^n + \mathbf{k}_1^n)^T \mathbf{B}_n^+ (\mathbf{x}_1^n + \mathbf{k}_1^n) + (1-\phi)(x_1 + k_1)^2$$

Therefore, for sufficiently large j,

$$\begin{aligned} f_{\mathbf{X}_1^{n_j}}(\mathbf{X}_1^{n-j}; \hat{\theta}_{n_j}) \leq \quad & \frac{\sqrt{1-\hat{\phi}_{n_j}^2}}{\hat{\sigma}_{n_j}\sqrt{2\pi}} \sum_{k_1 \in \mathbf{Z}} \exp(-\tfrac{1}{2}\hat{\sigma}_{n_j}^{-2}(1-\hat{\phi}_{n_j})(X_1 + k_1)^2) \\ & \times \prod_{l=2}^{n_j} \Big\{ (\hat{\sigma}_{n_j}\sqrt{2\pi})^{-1} \sum_{k_l \in \mathbf{Z}} \exp(-\tfrac{1}{2}\hat{\sigma}_{n_j}^{-2}\hat{\phi}_{n_j}(X_l + k_l - X_{l-1})^2) \Big\} \end{aligned} \tag{2.39}$$

where $\hat{\phi}_{n_j} < 1$, and the first term is replaced by 1 if $\hat{\phi}_{n_j} = 1$. Now, since $\hat{\phi}_{n_j}/\hat{\sigma}_{n_j}^2 \to \infty$ as $j \to \infty$, there exist $J \in \mathbf{N}$ and positive C_2 such that for any $j > J$ and $x_{l-1}, x_l \in [-\frac{1}{2}, \frac{1}{2}]$

$$(\hat{\sigma}_{n_j}\sqrt{2\pi})^{-1} \sum_{k_l \in \mathbf{Z}} \exp(-\tfrac{1}{2}\hat{\sigma}_{n_j}^{-2}\hat{\phi}_{n_j}(x_l + k_l - x_{l-1})^2)$$

$$\leq \frac{C_2}{\hat{\sigma}_{n_j}\sqrt{2\pi}} \exp(-\tfrac{1}{2}\hat{\sigma}_{n_j}^2 \hat{\phi}_{n_j} \|x_l - x_{l-1}\|_{\mathbf{Z}}^2)$$

and hence, for any $j > J$, the product term in (2.39) is less than

$$\frac{C_2^{n_j-1}}{\hat{\sigma}_{n_j}^{n_j-1}(2\pi)^{\frac{1}{2}(n_j-1)}} \exp\Big(-\tfrac{1}{2}\hat{\sigma}_{n_j}^{-2}\hat{\phi}_{n_j} \sum_{l=2}^{n_j} \|X_l - X_{l-1}\|_{\mathbf{Z}}^2\Big)$$

which limits to 0, as $j \to \infty$, if

$$\underline{\lim}_{n\to\infty} n^{-1} \sum_{l=2}^{n} \|X_{l+1} - X_l\|_{\mathbf{Z}}^2 > 0$$

which, since the AR(1) is ergodic, is an event of probability 1. But lemma 2.27 shows that, for almost all ω, for sufficiently large n, $f_{\mathbf{X}_1^n}(\mathbf{X}_1^n; \hat{\theta}_n) \geq 1$. So, if we can show that the first term in (2.39) almost surely cannot tend to infinity, the result follows. But, writing $\psi = \hat{\sigma}_{n_j}^2(1-\hat{\phi}_{n_j})$, that term is

$$\frac{\sqrt{1-\hat{\phi}_{n_j}^2}}{\sqrt{1-\hat{\phi}_{n_j}}} (\psi\sqrt{2\pi})^{-1} \sum_{k \in \mathbf{Z}} \exp(-\tfrac{1}{2}\psi^{-2}(X_1 + k)^2)$$

The first part is $\sqrt{1+\hat{\phi}_{n_j}}$ which is bounded above by 2. Provided $X_1 \neq 0$, the rest is a continuous function of ψ, which tends to 0 as $\psi \to 0$ and tends to 1 as $\psi \to \infty$, by lemma 2.2. Therefore, with probability 1, the expression is bounded above and cannot tend to infinity as $j \to \infty$.

A similar argument applies for $\underline{\lim}_{n\to\infty} \hat{\phi}_n = -1$ and so this completes the proof.

Q.E.D.

Lemma 2.30 *Let $\theta_0 \in \mathcal{P}$ and let $\theta \in \mathcal{P}'$ be such that $|\phi| = 1$. Then, given $\epsilon > 0$, there exists an open neighbourhood $B(\theta)$ of θ in $\mathcal{P}'$ such that*

$$P_{\theta_0}\left[\lim_{n\to\infty} \sup_{\theta_1 \in B(\theta)} \frac{f_{\mathbf{X}_1^n}(\mathbf{X}_1^n;\theta_1)}{f_{\mathbf{X}_1^n}(\mathbf{X}_1^n;\theta_0)} = 0\right] = 1$$

Proof: By the same appeal to Jensen's inequality as in the proof of lemma 2.27

$$\ln f_{X_t|\mathbf{X}_{-\infty}^{t-1}}(X_t|\mathbf{X}_{-\infty}^{t-1}) - \ln f_{X_t|\mathbf{X}_{-\infty}^{t-1}}(X_t|\mathbf{X}_{-\infty}^{t-1};\theta_0) \tag{2.40}$$

has negative θ_0-expectation, $-\epsilon$ say, unless it is almost surely 1. But X_t has different marginal distributions for θ and θ_0 (uniform and wrapped normal), so (2.40) is not almost surely 1. Further (2.40) is stationary, and so, by lemma 1.3 and lemmas 2.2 and 2.25, satisfies the hypotheses of theorem A.3. Thus, applying lemma 2.26,

$$P_{\theta_0}\left[\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \ln f_{X_j|\mathbf{X}_1^{j-1}}(x_j|\mathbf{x}_1^{j-1}) - \ln f_{X_j|\mathbf{X}_1^{j-1}}(X_j|\mathbf{X}_1^{j-1};\theta_0) = -\epsilon\right] = 1 \tag{2.41}$$

We must now extend this to a neighbourhood of θ in $\mathcal{P}$. Let θ_1 be such that $\phi_1 = 1$. Then

$$f_{\mathbf{X}_1^n}(\mathbf{x}_1^n;\theta_1) = \prod_{j=2}^{n}(\sigma_1\sqrt{2\pi})^{-1}\sum_{k_j\in\mathbf{Z}} \exp(-\tfrac{1}{2}\sigma_1^{-2}(x_j + k_j - x_{j-1})^2) \tag{2.42}$$

Since the wrapped normal density is a continuous function of σ and x, there exists $\delta > 0$ so that for all n and ω, $\ln f_{\mathbf{X}_1^n}(\mathbf{X}_1^n;\theta_1) \le \ln f_{\mathbf{X}_1^n}(\mathbf{x}_1^n) + \frac{1}{3}n\epsilon$, whenever $|\sigma_1^2 - \sigma^2| < \delta$.

Now let $\theta_2 \in \mathcal{P}$ be such that $0 < \phi_2 < 1$. By lemma 9

$$(\mathbf{x}_1^n + \mathbf{k}_1^n)^T\mathbf{M}_n(\mathbf{x}_1^n + \mathbf{k}_1^n) \ge \phi_2\,(\mathbf{x}_1^n + \mathbf{k}_1^n)^T\mathbf{B}_n^+(\mathbf{x}_1^n + \mathbf{k}_1^n) + (1-\phi_2)(x_1 + k_1)^2$$

and therefore

$$\begin{aligned} f_{\mathbf{X}_1^n}(\mathbf{x}_1^n;\theta_2) \;\le\; & \frac{\sqrt{1-\phi_2^2}}{\sigma_2\sqrt{2\pi}}\sum_{k_1\in\mathbf{Z}} \exp(-\tfrac{1}{2}(1-\phi_2)\sigma_2^{-2}(x_1+k_1)^2) \\ & \times \prod_{j=2}^{n}(\sigma_2\sqrt{2\pi})^{-1}\sum_{k_j\in\mathbf{Z}} \exp\left(-\tfrac{1}{2}\phi_2\sigma_2^{-2}(x_j + k_j - x_{j-1})^2\right) \end{aligned} \tag{2.43}$$

Here the first term is simply $\sqrt{1+\phi_2}$ times a wrapped normal density function and hence, by lemma 2.2, there exists C such that if $\sigma_2^2/(1-\phi_2) \geq C$ the log of the term is bounded above by $\ln 2 + \epsilon/3$. The product in (2.43) is of the form (2.42) multiplied by ϕ_2^{1-n}. Hence, provided $-\ln \phi_2 < \epsilon/3$, $|\phi_2^{-1}\sigma_2^2 - \sigma^2| < \delta$ and $(1-\phi_2) \leq C^{-1}\sigma_2^2$

$$\ln f_{\mathbf{X}_1^n}(\mathbf{x}_1^n;\theta_2) \leq \ln f_{\mathbf{X}_1^n}(\mathbf{x}_1^n) + \ln 2 + \tfrac{1}{3}\epsilon + \tfrac{2}{3}(n-1)\epsilon \tag{2.44}$$

It is clear, from (2.43), (2.44) and (2.41), that taking B to be the set of $\theta_2 \in \mathcal{P}$ satisfying these last three conditions produces a set of the required type.

A similar argument suffices for the case $\phi = -1$.

Q.E.D.

Proof of consistency

The following two lemmas are the core of the proof of consistency. The first shows that for each parameter value other than the true value the likelihood function tends almost surely to a smaller value than at the true parameter. The second lemma shows that the rate of change of the likelihood function as the parameter changes can in a certain sense be bounded uniformly, and uses this to show that for each parameter value other than the true one a neighbourhood can be found where the maximum of the likelihood function tends almost surely to a smaller value than at the true parameter.

Lemma 2.31 *Let θ_0, $\theta \in \mathcal{P}$, $\theta \neq \theta_0$. There exists $\epsilon > 0$ such that*

$$P_{\theta_0}[\overline{\lim_{n\to\infty}}\, n^{-1}\left\{\ln f_{\mathbf{X}_1^n}(\mathbf{x}_1^n) - \ln f_{\mathbf{X}_1^n}(\mathbf{x}_1^n;\theta_0)\right\} < -\epsilon] = 1$$

Proof: By lemma 2.26

$$\lim_{n\to\infty} n^{-1}|\sum_{j=1}^{n} \ln f_{X_j|\mathbf{X}_1^{j-1}}(X_j|\mathbf{X}_1^{j-1}) - \sum_{j=1}^{n} \ln f_{X_j|\mathbf{X}_{-\infty}^{j-1}}(X_j|\mathbf{X}_{-\infty}^{j-1})| = 0$$

for all ω and $\theta \in \mathcal{P}$. But, by the same appeal to Jensen's inequality as in the proof of lemma 2.27

$$E_{\theta_0}[\ln f_{X_t|\mathbf{X}_{-\infty}^{t-1}}(X_t|\mathbf{X}_{-\infty}^{t-1}) - \ln f_{X_t|\mathbf{X}_{-\infty}^{t-1}}(X_t|\mathbf{X}_{-\infty}^{t-1};\theta_0)] \leq 0 \tag{2.45}$$

with equality only if

$$\ln f_{X_t|\mathbf{X}_{-\infty}^{t-1}}(X_t|\mathbf{X}_{-\infty}^{t-1};\theta) - \ln f_{X_t|\mathbf{X}_{-\infty}^{t-1}}(X_t|\mathbf{X}_{-\infty}^{t-1};\theta_0) \tag{2.46}$$

is zero θ_0-almost surely. However, that would imply that the X_t process is indistinguishable for θ and θ_0, which is contradicted by the existence of consistent estimators for σ and ϕ as was shown in chapter 1. Hence (2.45) is a strict inequality. But (2.46) is stationary and by lemma 1.3 and lemmas 2.2 and 2.25 satisfies the hypotheses of theorem A.3. Thus

$$n^{-1}\left|\sum_{j=1}^{n} \ln f_{X_j|\mathbf{X}_{-\infty}^{j-1}}(X_t|\mathbf{X}_{-\infty}^{j-1};\theta) - \ln f_{X_j|\mathbf{X}_{-\infty}^{j-1}}(X_t|\mathbf{X}_{-\infty}^{j-1};\theta_0)\right|$$

converges θ_0-almost surely to its negative θ_0-expectation, which implies the result.

Q.E.D.

Lemma 2.32 *Let $\theta_0 \in \mathcal{P}$. Let $\epsilon > 0$ be given. There exists a neighbourhood O of θ_0 in $\mathcal{P}$ such that, for all ω, n and $\theta_1 \in O$*

$$n^{-1}|\ln f_{\mathbf{X}_1^n}(\mathbf{X}_1^n;\theta_1) - \ln f_{\mathbf{X}_1^n}(\mathbf{X}_1^n;\theta_0)| \le \epsilon$$

Proof: Let O_1 be a bounded open set in $\mathcal{P}$ containing θ_0 such that $\bar{O}_1 \subset \mathcal{P}$. Such a choice is possible since $\mathcal{P}$ is itself open. From the choice of O_1, $\bar{O}_1$ is compact. Hence by lemma 2.20 there exists C_1 such that

$$E[\epsilon_j^2|\mathbf{X}_1^n] \le C_1 \text{ and } E[(1-\phi^2)Y_1^2|\mathbf{X}_1^n] \le C_1$$

for all n, $1 \le j \le n$, ω and $\theta \in \bar{O}_1$. Hence, from (2.10)

$$\left|\partial_\sigma \ln f_{\mathbf{X}_1^n}(\mathbf{x}_1^n)\right| \le \frac{n}{\sigma} + \frac{nC_1}{\sigma^3}$$

for all $\theta \in \bar{O}_1$, ω, and n. Since $\bar{O}_1$ is compact, there exists C_2 such that $\sigma \ge C_2$ for all $\theta \in \bar{O}_1$ and hence there exists C_3 such that

$$\left|\partial_\sigma \ln f_{\mathbf{X}_1^n}(\mathbf{x}_1^n)\right| \le nC_3$$

for all $\theta \in \bar{O}_1$, n and ω. Similarly there exists C_4 such that

$$\left|\partial_\phi \ln f_{\mathbf{X}_1^n}(\mathbf{x}_1^n)\right| \le nC_4$$

for all $\theta \in \bar{O}_1$, n and ω.

Thus

$$n^{-1}|\ln f_{\mathbf{X}_1^n}(\mathbf{X}_1^n;\theta_1) - \ln f_{\mathbf{X}_1^n}(\mathbf{X}_1^n;\theta_0)| \le C_3|\sigma_1 - \sigma_0| + C_4|\phi_1 - \phi_0|$$

for all $\theta_1 \in \bar{O}_1$, n and ω. Let $O_2 = \{\theta : |\sigma - \sigma_0| < \epsilon/2C_3 \text{ and } |\phi - \phi_0| < \epsilon/2C_4\}$. Put $O = O_1 \cap O_2$. Clearly O is the desired neighbourhood.

$\mathcal{Q.E.D.}$

All the machinery is now in place. The proof of consistency follows.

Theorem 2.33 *The maximum likelihood estimator for the wrapped $AR(1)$ converges, θ_0-almost surely, to θ_0, for all $\theta_0 \in \mathcal{P}$.*

Proof: For all ω, either $\overline{\lim}_{n\to\infty} \hat{\sigma}_n^2(\omega) = \infty$ or the sequence $\hat{\theta}(\omega)$ has a limit point in $\overline{\mathcal{P}'}$. So, by lemma 1, the sequence $\hat{\theta}_n(\omega)$ has, θ_0 almost surely, a limit point in $\overline{\mathcal{P}'}$. Lemma 2 shows that, with probability 1, this limit point is not at $\sigma = 0$. Hence, with probability 1, $\{\hat{\theta}_n(\omega)\}$ has a limit point in $\mathcal{P}'$. To complete the proof we must show that this limit point is unique for almost all ω, and that it is the true parameter value θ_0.

By lemma 2.32, any θ in $\mathcal{P}$ has a neighbourhood $B(\theta)$ in $\mathcal{P}$ (and therefore in $\mathcal{P}'$) such that

$$P_{\theta_0}\left[\lim_{n\to\infty} \sup_{\theta_1 \in B(\theta)} \frac{f_{\mathbf{X}_1^n}(\mathbf{X}_1^n;\theta_1)}{f_{\mathbf{X}_1^n}(\mathbf{X}_1^n;\theta_0)} = 0\right] = 1 \tag{2.47}$$

Furthermore, by lemma 4, any θ with $|\phi| = 1$ has a neighbourhood $B(\theta)$ in $\mathcal{P}'$ with the same property. For each $m \in \mathbf{N}$, define the set $\mathcal{P}_m$ and $\mathcal{P}_m^I$ as follows:

$$\mathcal{P}_m = \{(\sigma, \phi) \mid \sigma \in [\tfrac{1}{m}, m]; \phi \in [-1, 1]; |\phi - \phi_0| + |\sigma - \sigma_0| \geq \tfrac{1}{m}\}$$

and

$$\mathcal{P}_m^I = \{(\sigma, \phi) \mid \sigma \in (\tfrac{1}{m}, m); \phi \in [-1, 1]; |\phi - \phi_0| + |\sigma - \sigma_0| > \tfrac{1}{m}\}$$

Then $\mathcal{P}_m$ is closed and bounded, and is therefore compact. Also we note that

$$\mathcal{P}' \setminus \{\theta_0\} = \bigcup_{m=1}^{\infty} \mathcal{P}_m^I$$

The collection of all the $B(\theta)$ for θ in $\mathcal{P}_m$ forms an open cover of $\mathcal{P}_m$. Since $\mathcal{P}_m$ is compact there exists a finite subcover $B(\theta_1), \ldots, B(\theta_M)$. Then, from (2.47)

$$P_{\theta_0}\left[\lim_{n\to\infty} \sup_{\theta_1 \in \mathcal{P}_m} \frac{f_{\mathbf{X}_1^n}(\mathbf{X}_1^n;\theta_1)}{f_{\mathbf{X}_1^n}(\mathbf{X}_1^n;\theta_0)} = 0\right] = 1$$

Therefore $P_{\theta_0}[\hat{\theta}_n$ has a limit point in $\mathcal{P}^I_m] = 0$, and so

$$P_{\theta_0}[\hat{\theta}_n \text{ has a limit point in } \mathcal{P}' \setminus \{\theta_0\}] \leq \sum_{m \in \mathbf{N}} P_{\theta_0}[\hat{\theta}_n \text{ has a limit point in } \mathcal{P}^I_m] = 0$$

Thus, θ_0-a.s., θ_0 is the only limit point of $\hat{\theta}_n$, and we have the desired result:

$$P_{\theta_0}\left[\lim_{n\to\infty} \hat{\theta}_n = \theta_0\right] = 1$$

$\mathcal{Q.E.D.}$

2.2.5 Asymptotic Normality

The maximum likelihood estimator for the wrapped AR(1) will now be shown to have, asymptotically, a normal distribution. The proof is based on the ideas in [7]. Unfortunately, there is an error in the section of that paper which deals with the particular case of mixing processes (for further details see the discussion in the appendix). The proof which follows is therefore considerably more complicated than might be expected by readers familiar with [7].

The log-likelihood function is denoted by $L_n(\theta)$, i.e.

$$L_n(\theta) = f_{\mathbf{X}_1^n}(\mathbf{X}_1^n; \theta)$$

$L_n'(\theta)$ denotes the vector of first derivatives, i.e.

$$L_n'(\theta) = \begin{pmatrix} \partial_\sigma L_n(\theta) \\ \partial_\phi L_n(\theta) \end{pmatrix}$$

and $L_n''(\theta)$ denotes the matrix of second derivatives, i.e.

$$L_n''(\theta) = \begin{pmatrix} \partial_\sigma^2 L_n(\theta) & \partial_\sigma \partial_\phi L_n(\theta) \\ \partial_\phi \partial_\sigma L_n(\theta) & \partial_\phi^2 L_n(\theta) \end{pmatrix}$$

As is usual $L_n'(\theta)$ is expanded as a Taylor series around θ_0 (the true parameter).

$$L_n'(\theta) = L_n'(\theta_0) + L_n''(\theta_0, \theta)(\theta - \theta_0)$$

where $L_n''(\theta_0, \theta)$ denotes the matrix of second derivatives with elements being evaluated, at possibly different points, on the line segment joining θ_0 and θ. If $\hat{\theta}_n$ is the maximum likelihood estimate, $L_n'(\hat{\theta}_n) = 0$ and so

$$(\hat{\theta}_n - \theta_0) = -L_n''(\theta_0, \hat{\theta}_n) L_n'(\theta_0)$$

Figure 2.4: Tree illustrating the dependencies between the various lemmata and theorems in section 2.2.4

The proof shows $\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{\text{d}}{\to} N(0, \mathcal{L}^{-1})$ for some positive definite covariance matrix $\mathcal{L}^{-1}$ in three stages:

1. $n^{-1}L''_n(\theta_0, \theta_0) \overset{\text{a.s.}}{\to} -\mathcal{L}$.

2. $n^{-1}\{L''(\theta_0, \hat{\theta}_n) - L''_n(\theta_0, \theta_0)\} \overset{\text{a.s.}}{\to} 0$.

3. $n^{-\frac{1}{2}}L'_n(\theta_0) \overset{\text{d}}{\to} N(0, \mathcal{L})$.

In most of what follows, we will actually work with a family of approximations, $\tilde{L}'(\theta)$ and $\tilde{L}''(\theta)$, to the derivatives of the likelihood function which are defined by

$$(\tilde{L}'_n(\theta))_\sigma = \frac{-n}{\sigma} + \frac{1}{\sigma^3}E[\tilde{S}_n(\mathbf{Y}_1^n)|\mathbf{X}_1^n]$$

$$(\tilde{L}'_n(\theta))_\phi = -\frac{1}{2\sigma^2}E[\tilde{T}_n(\mathbf{Y}_1^n)|\mathbf{X}_1^n]$$

$$(\tilde{L}''_n(\theta))_{\sigma\sigma} = \frac{n}{\sigma^2} - \frac{3}{\sigma^4}E[\tilde{S}_n(\mathbf{Y}_1^n)|\mathbf{X}_1^n] + \frac{1}{\sigma^6}D[\tilde{S}_n(\mathbf{Y}_1^n)|\mathbf{X}_1^n]$$

$$(\tilde{L}''_n(\theta))_{\sigma\phi} = \frac{1}{\sigma^3}E[\tilde{T}_n(\mathbf{Y}_1^n)|\mathbf{X}_1^n] - \frac{1}{2\sigma^5}C[\tilde{T}_n(\mathbf{Y}_1^n), \tilde{S}_n(\mathbf{Y}_1^n)|\mathbf{X}_1^n]$$

$$(\tilde{L}''_n(\theta))_{\phi\phi} = -\frac{1}{2\sigma^2}E[U_n(\mathbf{Y}_1^n)|\mathbf{X}_1^n] + \frac{1}{4\sigma^4}D[\tilde{T}_n(\mathbf{Y}_1^n)|\mathbf{X}_1^n]$$

where

$$\tilde{S}_n = \sum_{j=2}^{n} \epsilon_j^2 = S_n - (1-\phi^2)Y_1^2$$

and

$$\tilde{T}_n = -2\sum_{j=2}^{n} \epsilon_j Y_{j-1} = T_n + 2\phi Y_1^2$$

The following lemma will be used to validate this approximation.

Lemma 2.34 *Let $B \subset \mathcal{P}$ be compact. There exists C such that for all ω n and $\theta \in B$*

$$|L'_n(\theta) - \tilde{L}'_n(\theta)| \leq C \tag{2.48}$$

$$|L''_n(\theta) - \tilde{L}''_n(\theta)| \leq C \tag{2.49}$$

Proof: The σ-element of $L'(\theta) - \tilde{L}'(\theta)$ is

$$\sigma^{-3}E[S_n|\mathbf{X}_1^n] - \sigma^{-3}E[\tilde{S}_n|\mathbf{X}_1^n] = \sigma^{-3}(1-\phi^2)E[Y_1^2|\mathbf{X}_1^n]$$

which, by lemma 2.20 has a uniform bound for all n, ω and $\theta \in B$. The ϕ-element is similar. This proves (2.48).

Consider the σ-ϕ element of $L_n''(\theta) - \tilde{L}_n''(\theta)$ which is

$$\sigma^{-3}E[T_n|\mathbf{X}_1^n] - \tfrac{1}{2}\sigma^{-5}C[T_n, S_n|\mathbf{X}_1^n] - \sigma^{-3}E[\tilde{T}_n|\mathbf{X}_1^n] + \tfrac{1}{2}\sigma^{-5}C[\tilde{T}_n, \tilde{S}_n|\mathbf{X}_1^n]$$

From the argument in the previous paragraph we need only consider the covariance terms. But

$$\begin{aligned} C[T_n, S_n|\mathbf{X}_1^n] - C[\tilde{T}_n, \tilde{S}_n|\mathbf{X}_1^n] &= C[\tilde{T}_n, (1-\phi^2)Y_1^2|\mathbf{X}_1^n] \\ &\quad + C[\tilde{S}_n, -2\phi Y_1^2|\mathbf{X}_1^n] \\ &\quad + C[(1-\phi^2)Y_1^2, -2\phi Y_1^2|\mathbf{X}_1^n] \end{aligned} \tag{2.50}$$

The last term is uniformly bounded by lemma 2.20. But, if h is a polynomial in two variables

$$\begin{aligned} \Big|C[Y_1^2, \sum_{j=1}^{n-1} h(\mathbf{Y}_j^{j+1})|\mathbf{X}_1^n]\Big| &\leq \sum_{j=1}^{n-1}\Big|C[Y_1^2, h(\mathbf{Y}_j^{j+1})]|\mathbf{X}_1^n]\Big| \\ &\leq C\sum_{j=1}^{n-1}\rho^j \leq C\frac{\rho}{1-\rho} \end{aligned}$$

for all n, ω, $\theta \in B$ and some $\rho < 1$ by lemma 2.23. Thus the first two terms on the right-side of (2.50) are uniformly bounded and therefore the σ-ϕ term 0f $L_n''(\theta) - \tilde{L}_n''(\theta)$ is uniformly bounded. The σ-σ and ϕ-ϕ components are similar. Hence the result follows.

Q.E.D.

Conditional behaviour of polynomial sums

The expressions for the derivatives of the log-likelihood function involve cross-moments of sums of polynomials in the Y-process conditional upon the X-process. This section contains a number of lemmas concerning the asymptotic behavious of such quantities. These lemmas are really the basic units of the proof of asymptotic normality.

Lemma 2.35 *Let $B \subset \mathcal{P}$ be compact. Let $g(.,.)$ and $h(.,.)$ be polynomials in 2 variables. Then there exists C such that*

$$\left|C\Big[\sum_{j=0}^{n-2} g(\mathbf{Y}_{j+1}^{j+2}), \sum_{j=0}^{n-2} h(\mathbf{Y}_{j+1}^{j+2})|\mathbf{X}_1^n\Big]\right| \leq Cn$$

for all n, ω and $\theta \in B$.

Proof: By lemma 2.23, there exists C_1 and ρ_1 such that for all n, ω, $0 \leq j_1, j_2 \leq n-2$ and $\theta \in B$

$$\left|C\Big[g(\mathbf{Y}_{j_1+1}^{j_1+2}), h(\mathbf{Y}_{j_2+1}^{j_2+2})|\mathbf{X}_1^n\Big]\right| \leq C\rho^{|j_2-j_1|}$$

Thus

$$\begin{aligned}\left|C\Big[\sum_{j=0}^{n-2} g(\mathbf{Y}_{j+1}^{j+2}), \sum_{j=0}^{n-2} h(\mathbf{Y}_{j+1}^{j+2})|\mathbf{X}_1^n\Big]\right| &\leq \sum_{j_1=0}^{n-2}\sum_{j_2=0}^{n-2} C\rho^{|j_2-j_1|} \\ &\leq nC\sum_{j\in\mathbf{Z}} \rho^{|j|}\end{aligned}$$

as required.

$\mathcal{Q.E.D.}$

Lemma 2.36 *Let $g(.,.)$ and $h(.,.)$ be polynomials in two variables. There exists a stationary sequence $W_1, \ldots$ and C such that*

$$\left|C[\sum_{j=0}^{n-2} g(\mathbf{Y}_{j+1}^{j+2}), \sum_{k=0}^{n-2} h(\mathbf{Y}_{k+1}^{k+2})|\mathbf{X}_1^n] - \sum_{j=1}^{n} W_j\right| < C$$

for all ω and n.

Proof: Put

$$W_j = \sum_{k=-\infty}^{\infty} C[g(\mathbf{Y}_{j+1}^{j+2}), h(\mathbf{Y}_{k+1}^{k+2})|\mathbf{X}_{-\infty}^{\infty}]$$

Then, using time-reversibility and applying lemma 2.22 twice and lemma 2.23,

$$\sum_{j=1}^{n-1}\sum_{k=1}^{n-1}\left|C[g(\mathbf{Y}_j^{j+1}), h(\mathbf{Y}_k^{k+1})|\mathbf{X}_1^n] - C[g(\mathbf{Y}_j^{j+1}), h(\mathbf{Y}_k^{k+1})|\mathbf{X}_{-\infty}^{\infty}]\right|$$

$$\leq \sum_{j=0}^{n-2}\sum_{k=0}^{n-2} \min(C_1\rho_1^{\min(j,k)} + C_1\rho_1^{n-2-\max(j,k)}, C_2\rho_2^{|j-k|})$$

$$\leq \sum_{j=0}^{n-2}\sum_{k=j}^{n-2}(C_1(\rho_1^j + \rho_1^{n-2-k})C_2\rho_2^{k-j})^{\frac{1}{2}} + \sum_{j=0}^{n-2}\sum_{k=0}^{j-1}(C_1(\rho_1^k + \rho_1^{n-2-j})C_2\rho_2^{j-k})^{\frac{1}{2}}$$

$$\leq C_3 \sum_{j=0}^{n-2} \sum_{k=j}^{n-2} (\rho_3^k + \rho_3^{n-2-j}) + C_3 \sum_{j=0}^{n-2} \sum_{k=0}^{j-1} (\rho_3^j + \rho_3^{n-2-k})$$

$$\leq C_3 \Big\{ \sum_{j=0}^{\infty} \frac{\rho_3^j}{1-\rho_3} + \sum_{j=0}^{\infty} (j+1)\rho_3^j + \sum_{j=0}^{\infty} j\rho_3^j + \sum_{j=0}^{\infty} \frac{\rho_3^j}{1-\rho_3} \Big\}$$

$$\stackrel{\text{def}}{=} C_4 \tag{2.51}$$

Also, by lemma 2.23,

$$\sum_{j=1}^{n-1} \Big\{ \sum_{k=-\infty}^{0} + \sum_{k=n}^{\infty} \Big\} |C[g(\mathbf{Y}_j^{j+1}), h(\mathbf{Y}_k^{k+1}) | \mathbf{X}_{-\infty}^{\infty}]|$$

$$\leq \sum_{j=1}^{n-1} \Big\{ \sum_{k=-\infty}^{0} C_2 \rho_2^{j-k} + \sum_{k=n}^{\infty} C_2 \rho^{k-j} \Big\}$$

$$\leq \frac{C_2\rho_2}{(1-\rho_2)^2} + \frac{C_2\rho_2}{(1-\rho_2)^2} \tag{2.52}$$

and the result follows from (2.51) and (2.52).

$\mathcal{Q.E.D.}$

Lemma 2.37 *Let $g(.,.)$ and $h(.,.)$ be polynomials in two variables. Then for any $\theta \in \mathcal{P}$*

$$D\Big[n^{-1} C[\sum_{j=1}^{n-1} g(\mathbf{Y}_j^{j+1}), \sum_{k=1}^{n-1} h(\mathbf{Y}_k^{k+1}) | \mathbf{X}_1^n]\Big] \to 0 \tag{2.53}$$

Proof: Let g_j denote $g(\mathbf{Y}_j^{j+1})$, h_k denote $h(\mathbf{Y}_k^{k+1})$. Then the expression on the left side of (2.53) is $n^{-2} \sum_{j_1,j_2,j_3,j_4=1}^{n-1} R_{j_1j_2j_3j_4}$ where

$$R_{j_1j_2j_3j_4} = C\Big[C[g_{j_1}, h_{j_2} | \mathbf{X}_1^n], C[g_{j_3}, h_{j_4} | \mathbf{X}_1^n]\Big]$$

By lemma 2.23, there exists C_1 and $\rho_1 < 1$ such that, for all ω, n and $1 \leq j_1, j_2 \leq n-1$

$$|C[g_{j_1}, h_{j_2} | \mathbf{X}_1^n]| \leq C_1 \rho_1^{|j_2 - j_1|} \tag{2.54}$$

Suppose for the moment that $j_1 \leq j_3$. Let J_1 denote $\max(j_1, j_2)$ and J_2 denote $\min(j_3, j_4)$. By lemma 2.22 there exists C_2 and $\rho_2 < 1$ so that, for all n, $1 \leq j_1, j_2, j_3, j_4 \leq n$, $J_1 + \Delta \leq n$ and $J_2 - \Delta \geq 1$

$$|C[g_{j_1}, h_{j_2} | \mathbf{X}_1^n] - C[g_{j_1}, h_{j_2} | \mathbf{X}_1^{J_1+\Delta}]| \leq C_2 \rho_2^{\Delta} \tag{2.55}$$

and, by time-reversibility,

$$|C[g_{j_3}, h_{j_4}|\mathbf{X}_1^n] - C[g_{j_3}, h_{j_4}|\mathbf{X}_{J_2-\Delta}^n]| \leq C_2\rho_2^\Delta \tag{2.56}$$

Further, by lemma 1.3, lemma A.2 and (2.54), there exists C_3 and ρ_3 such that

$$|C[C[g_{j_1}, h_{j_2}|\mathbf{X}_1^{J_1+\Delta}], C[g_{j_3}, h_{j_4}|\mathbf{X}_{J_2-\Delta}^n]]| \leq C_3\rho_3^{J_2-J_1-2\Delta} \tag{2.57}$$

provided $J_2 - J_1 - 2\Delta \geq 0$.

But, for any random variables A_1, A_2, B_1 and B_2

$$\begin{aligned}
&\Big|C[A_1, B_1] - C[A_2, B_2]\Big| \\
&\quad\leq \Big|E[(A_1 - A_2)B_1]\Big| + \Big|E[A_2(B_1 - B_2)]\Big| \\
&\qquad\quad + \Big|E[A_1 - A_2]E[B_1]\Big| + \Big|E[A_2]E[B_1 - B_2]\Big| \\
&\quad\leq 4\sup_\omega \max(|A_1|, |A_2|, |B_1|, |B_2|) \max(|A_1 - A_2|, |B_1 - B_2|)
\end{aligned}$$

Hence from (2.54), (2.55), (2.56) and (2.57)

$$\begin{aligned}
&|C[C[g_{j_1}, h_{j_2}|\mathbf{X}_1^n], C[g_{j_3}, h_{j_4}|\mathbf{X}_1^n]]| \\
&\leq |C[C[g_{j_1}, h_{j_2}|\mathbf{X}_1^{J_1+\Delta}], C[g_{j_3}, h_{j_4}|\mathbf{X}_{J_2-\Delta}^n]]| + 4C_1C_2\rho_2^\Delta \\
&\leq 4C_1C_2\rho_2^\Delta + C_3\rho_3^{J_2-J_1-2\Delta}
\end{aligned}$$

provided $J_2 - J_1 - 2\Delta \geq 0$. Thus, if $J_2 \geq J_1$, taking $\Delta = \lfloor (J_2 - J_1)/3 \rfloor$ implies the existence of C_4 and $\rho_4 < 1$ such that

$$|R_{j_1j_2j_3j_4}| \leq C_4\rho_4^{J_2-J_1}$$

when $J_2 \geq J_1$. However (2.54) implies that

$$C[C[g_{j_1}, h_{j_2}|\mathbf{X}_1^n], C[g_{j_3}, h_{j_4}|\mathbf{X}_1^n]] \leq C_1^2\rho_1^{|j_2-j_1|+|j_4-j_3|} \leq C_1^2 \tag{2.58}$$

and so there exist C_5 and $\rho_5 < 1$ so that, even when $J_2 < J_1$,

$$C[C[g_{j_1}, h_{j_2}|\mathbf{X}_1^n], C[g_{j_3}, h_{j_4}|\mathbf{X}_1^n]] \leq C_5\rho_5^{J_2-J_1} \tag{2.59}$$

Let $J = \max(j_1, j_2, j_3, j_4) - \min(j_1, j_2, j_3, j_4)$, i.e. $J = |j_3 - j_4| + J_2 - J_1 + |j_2 - j_1|$. Hence $\max(|j_3 - j_4|, J_2 - J_1, |j_2 - j_1|) \geq J/3$ and so by (2.54) and (2.59), there exists

C_6 and $\rho_6 < 1$ such that $|R_{j_1j_2j_3j_4}| \le C_6\rho_6^J$. So far, this holds only when $j_1 \le j_3$. If $j_3 \ge j_1$, it is clear that the same argument holds with the roles of j_1, j_2 and j_3, j_4 reversed. However $J \ge (|j_1 - j_2| + |j_1 - j_3| + |j_1 - j_4|)/3$ and so

$$\begin{aligned} n^{-2}\sum_{j_1,j_2,j_3,j_4=1}^{n-1}|R_{j_1j_2j_3j_4}| &\le n^{-2}C_6\sum_{j_1=0}^{n-2}\sum_{j_2\in\mathbf{Z}}\sum_{j_3\in\mathbf{Z}}\sum_{j_4\in\mathbf{Z}}\rho_6^{|j_1-j_2|/3}\rho_6^{|j_1-j_3|/3}\rho_6^{|j_1-j_4|/3} \\ &\le n^{-1}C_8\Big(\sum_{j=-\infty}^{\infty}\rho_6^{|j|/3}\Big)^3 \\ &\to 0 \end{aligned}$$

as required.

Q.E.D.

Lemma 2.38 *Let $B \subset \mathcal{P}$ be compact. Let $g(.,.)$, $h(.,.)$ and $e(.,.)$ be polynomials in 2 variables. There exists C so that*

$$\Big|III[\sum_{j=0}^{n-2} g(\mathbf{Y}_{j+1}^{j+2}), \sum_{j=0}^{n-2} h(\mathbf{Y}_{j+1}^{j+2}), \sum_{j=0}^{n-2} e(\mathbf{Y}_{j+1}^{j+2})|\mathbf{X}_1^n]\Big| \le Cn \tag{2.60}$$

for all n, ω and $\theta \in B$.

Proof: For $0 \le j_1, j_2, j_3 \le n-2$ define

$$A_{j_1j_2j_3} = III[g(\mathbf{Y}_{j_1+1}^{j_1+2}), h(\mathbf{Y}_{j_2+1}^{j_2+2}), e(\mathbf{Y}_{j_3+1}^{j_3+2})|\mathbf{X}_1^n]$$

Without loss of generality, by re-arranging g, h and e, assume $j_1 \le j_2 \le j_3$ and $j_2 - j_1 \ge j_3 - j_2$. From the definition of the III operator

$$\begin{aligned} A_{j_1j_2j_3} &= C[g(\mathbf{Y}_{j_1+1}^{j_1+2}), h(\mathbf{Y}_{j_2+1}^{j_2+2})e(\mathbf{Y}_{j_3+1}^{j_3+2})|\mathbf{X}_1^n] \\ &\quad -C[g(\mathbf{Y}_{j_1+1}^{j_1+2}), h(\mathbf{Y}_{j_2+1}^{j_2+2})|\mathbf{X}_1^n]E[e(\mathbf{Y}_{j_3+1}^{j_3+2})|\mathbf{X}_1^n] \\ &\quad -C[g(\mathbf{Y}_{j_1+1}^{j_1+2}), e(\mathbf{Y}_{j_3+1}^{j_3+2})|\mathbf{X}_1^n]E[h(\mathbf{Y}_{j_2+1}^{j_2+2})|\mathbf{X}_1^n] \end{aligned} \tag{2.61}$$

By lemma 2.23, there exists C_1 and $\rho_1 < 1$ depending only on g, h and e such that

$$|C[g(\mathbf{Y}_{j_1+1}^{j_1+2}), h(\mathbf{Y}_{j_2+1}^{j_2+2}), e(\mathbf{Y}_{j_3+1}^{j_3+2})|\mathbf{X}_1^n]| \le C_1\rho_1^{j_2-j_1} \tag{2.62}$$

and by lemmas 2.23 and 2.20, there exist C_2 and $\rho_2 < 1$ depending only on g, h and e such that

$$|C[g(\mathbf{Y}_{j_1+1}^{j_1+2}), h(\mathbf{Y}_{j_2+1}^{j_2+2})|\mathbf{X}_1^n]E[e(\mathbf{Y}_{j_3+1}^{j_3+2})|\mathbf{X}_1^n]| \le C_2\rho_2^{j_2-j_1} \tag{2.63}$$

and

$$|C[g(\mathbf{Y}_{j_1+1}^{j_1+2}), e(\mathbf{Y}_{j_3+1}^{j_3+2})|\mathbf{X}_1^n]E[h(\mathbf{Y}_{j_2+1}^{j_2+2})|\mathbf{X}_1^n]| \leq C_2\rho_2^{j_3-j_1} \tag{2.64}$$

Hence $j_3 - j_1 \geq j_2 - j_1$ implies the existence of C_3 and $\rho_3 < 1$ such that $|A_{j_1j_2j_3}| \leq C_3\rho_3^{j_2-j_1}$. But, by assumption, $j_2 - j_1 \geq \frac{1}{2}(j_3 - j_1) = \frac{1}{2}\{\max(j_1, j_2, j_3) - \min(j_1, j_2, j_3)\}$, i.e.

$$|A_{j_1j_2j_3}| \leq C_3\rho_3^{\frac{1}{2}\{\max(j_1,j_2,j_3)-\min(j_1,j_2,j_3)\}}$$

which holds for any $0 \leq j_1, j_2, j_3 \leq n-2$, redefining C_3 and ρ_3 to allow for any permutation of the polynomials g, h and e in (2.61), (2.62), (2.63), and (2.64). Hence

$$\Big|III[\sum_{j=0}^{n-2} g(\mathbf{Y}_{j+1}^{j+2}), \sum_{j=0}^{n-2} h(\mathbf{Y}_{j+1}^{j+2}), \sum_{j=0}^{n-2} e(\mathbf{Y}_{j+1}^{j+2}), |\mathbf{X}_1^n]\Big|$$

$$\leq C_5 3 \sum_{j_1=0}^{n-2}\sum_{j_2=0}^{n-2}\sum_{j_3=0}^{n-2} \rho_3^{\frac{1}{2}\{\max(j_1,j_2,j_3)-\min(j_1,j_2,j_3)\}} \tag{2.65}$$

But $\max(j_1, j_2, j_3) - \min(j_1, j_2, j_3) \geq \frac{1}{2}|j_1 - j_2| + \frac{1}{2}|j_1 - j_3|$ implies (2.65) is bounded by

$$C_3 \sum_{j_1=0}^{n-2}\sum_{j_2=0}^{n-2}\sum_{j_3=0}^{n-2} \rho_3^{\frac{1}{4}|j_1-j_2|}\rho_3^{\frac{1}{4}|j_1-j_3|} \leq nC_3\Big(\sum_{j\in\mathbf{Z}} \rho_3^{\frac{1}{4}|j|}\Big)^2$$

as required.

$\mathcal{Q.E.D.}$

Convergence of the second derivative

Note that $E[\tilde{S}_n] = (n-1)\sigma^2$, $E[T_n] = 0$ and $E[U_n] = 2(n-2)\sigma^2/(1-\phi^2)$. Further $D[\tilde{S}_n] = 2(n-1)\sigma^4$ and $D[\tilde{T}_n = (n-1)\sigma^4/(1-\phi^2)$. In the remaining lemmas of this section, two quantities Z_n and H_n play a crucial part. Z_n is defined by $Z_n = \sum_{j=2}^n e^{2\pi i Y_j} = \sum_{j=2}^n e^{2\pi i X_j}$. H_n is defined by $H_n = \sum_{j=2}^n e^{2\pi i(Y_j+Y_{j-1})} = \sum_{j=2}^n e^{2\pi i(X_j+X_{j-1})}$.

Lemma 2.39 *The following limits hold for any $\theta \in \mathcal{P}$.*

1. $n^{-1}C[\tilde{S}_n, Z_n] \to -4\pi^2\sigma^4(1-\phi^2)^{-1}\exp\Big[-\frac{2\pi^2\sigma^2}{(1-\phi^2)}\Big]$
2. $n^{-1}C[\tilde{T}_n, Z_n] \to 8\pi^2\phi\sigma^4(1-\phi^2)^{-1}\exp\Big[-\frac{2\pi^2\sigma^2}{(1-\phi^2)}\Big]$
3. $n^{-1}C[\tilde{S}_n, H_n] \to -8\pi^2\sigma^4(1-\phi)^{-1}\exp\Big[-\frac{4\pi^2(1+\phi)\sigma^2}{(1-\phi^2)}\Big]$
4. $n^{-1}C[\tilde{T}_n, H_n] \to 8\pi^2\sigma^4(1-\phi)^{-2}\exp\Big[-\frac{4\pi^2(1+\phi)\sigma^2}{(1-\phi^2)}\Big]$

Proof: Commence by noting that, if $W \sim N(0, \sigma^2)$,

$$\begin{aligned} E[e^{itW}W] &= -i\partial_t E[e^{itW}] \\ &= -i\partial_t e^{-\frac{1}{2}t^2\sigma^2} \\ &= it\sigma^2 e^{-\frac{1}{2}t^2\sigma^2} \\ &= it\sigma^2 E[e^{itw}] \end{aligned}$$

and

$$\begin{aligned} E[e^{itW}W^2] &= -\partial_t^2 E[e^{itW}] \\ &= -\partial_t^2 e^{-\frac{1}{2}t^2\sigma^2} \\ &= (\sigma^2 - t^2\sigma^4)e^{-\frac{1}{2}t^2\sigma^2} \\ &= (\sigma^2 - t^2\sigma^4)E[e^{itW}] \end{aligned}$$

Note also that

$$E[e^{2\pi i Y_j}] = e^{-2\pi^2\sigma^2/(1-\phi^2)}$$

and

$$E[e^{2\pi i(Y_j + Y_{j-1})}] = e^{-4\pi^2(1+\phi)\sigma^2/(1-\phi^2)}$$

Define $e_{jkl} = E[e^{2\pi i(Y_j + bY_{j-1})}\epsilon_{j-k}\epsilon_{j-k-l}]$. The computation of e_{jkl} can be broken into a number of cases of interest. Note that since $Y_t = \sum_{l=0}^{\infty} \phi^l \epsilon_{t-l}$

$$e_{jkl} = E\Big[\exp\left[2\pi i(\epsilon_j + \textstyle\sum_{m=0}^{\infty}(b+\phi)\phi^m \epsilon_{j-m-1})\right] \epsilon_{j-k}\epsilon_{j-k-l}\Big]$$

1. $k < 0$ and $l = 0$

$$e_{jkl} = \sigma^2 E[e^{2\pi i(Y_j + bY_{j-1})}]$$

2. $k < 0$ and $l > 0$

$$e_{jkl} = 0$$

3. $k = 0$ and $l = 0$

$$e_{jkl} = (\sigma^2 - 4\pi^2\sigma^4)E[e^{2\pi i(Y_j + bY_{j-1})}]$$

4. $k = 0$ and $l > 0$

$$e_{jkl} = (2\pi i \sigma^2)(2\pi i(b+\phi)\phi^{l-1}\sigma^2)E[e^{2\pi i(Y_j+bY_{j-1})}]$$
$$= -4\pi^2\sigma^4(b+\phi)\phi^{l-1}E[e^{2\pi i(Y_j+bY_{j-1})}]$$

5. $k > 0$ and $l = 0$

$$e_{jkl} = (\sigma^2 - 4\pi^2(b+\phi)^2\phi^{2k-2}\sigma^4)E[e^{2\pi i(Y_j+bY_{j-1})}]$$

6. $k > 0$ and $l > 0$

$$e_{jkl} = (2\pi i(b+\phi)\phi^{k-1}\sigma^2)(2\pi i(b+\phi)\phi^{k+l-1}\sigma^2)E[e^{2\pi i(Y_j+bY_{j-1})}]$$
$$= -4\pi^2\sigma^4(b+\phi)^2\phi^{l+2k-2}E[e^{2\pi i(Y_j+bY_{j-1})}]$$

Hence, we have

1. $$E[e^{2\pi iY_j}\tilde{S}_n] = (n-j)\sigma^2 e^{2\pi iY} + \sum_{k=0}^{j-2}(\sigma^2 - 4\pi^2\phi^{2k}\sigma^4)e^{2\pi iY}$$

   and

   $$C[e^{2\pi iY_j}, \tilde{S}_n] = -\sum_{k=0}^{j-2} 4\pi^2\phi^{2k}\sigma^4 e^{2\pi iY}$$

   Therefore

   $$\begin{aligned} n^{-1}C[Z_n, \tilde{S}_n] &= -n^{-1}\sum_{j=2}^{n}\sum_{k=0}^{j-2} 4\pi^2\phi^{2k}\sigma^4 e^{2\pi iY} \\ &\to -\sum_{k=0}^{\infty} 4\pi^2\phi^{2k}\sigma^4 e^{2\pi iY} \\ &= \frac{-4\pi^2\sigma^4}{(1-\phi^2)}e^{2\pi iY} \end{aligned}$$

2. $$E[e^{2\pi iY_j}\tilde{T}_n] = (n-j).0 + \sum_{k=0}^{j-2}\sum_{l=1}^{\infty} 8\pi^2\phi^{l-1}\phi^{2k+l}\sigma^4 e^{2\pi iY}$$

   Therefore

   $$\begin{aligned} n^{-1}C[Z_n, \tilde{S}_n] &= n^{-1}\sum_{j=1}^{n}\sum_{k=0}^{j-2} 8\pi^2\frac{\phi^{2k+1}}{(1-\phi^2)}\sigma^4 e^{2\pi iY} \\ &\to \sum_{k=0}^{\infty} 8\pi^2\frac{\phi^{2k+1}}{(1-\phi^2)}\sigma^4 e^{2\pi iY} \\ &= \frac{8\pi^2\phi\sigma^4}{(1-\phi^2)^2}e^{2\pi iY} \end{aligned}$$

3. For $2 \leq j \leq n$

$$
\begin{aligned}
E[e^{2\pi i(Y_j+Y_{j-1})}\tilde{S}_n] &= (n-j)\sigma^2 e^{2\pi i(Y_1+Y_0)} \\
&\quad +(\sigma^2 - 4\pi^2\sigma^4)e^{2\pi i(Y_1+Y_0)} \\
&\quad +\sum_{k=1}^{j-2}(\sigma^2 - 4\pi^2(1+\phi)^2\phi^{2k-2}\sigma^4)e^{2\pi i(Y_1+Y_0)}
\end{aligned}
$$

Therefore

$$
\begin{aligned}
(n-1)^{-1}C[H_n, \tilde{S}_n] &= -4\pi^2\sigma^4 e^{2\pi i(Y_1+Y_0)} \\
&\quad -\frac{4\pi^2}{n-1}\sigma^4 \sum_{j=2}^{n}\sum_{k=1}^{j-1}(1+\phi)^2\phi^{2k-2}e^{2\pi i(Y_1+Y_0)} \\
&\to -4\pi^2\sigma^4\Big\{1+\sum_{k=1}^{\infty}(1+\phi)^2\phi^{2k-2}\Big\}e^{2\pi i(Y_1+Y_0)} \\
&= \frac{-8\pi^2\sigma^4}{1-\phi}e^{2\pi i(Y_1+Y_0)}
\end{aligned}
$$

4. For $2 \leq j \leq n$

$$
\begin{aligned}
E[e^{2\pi i(Y_j+Y_{j-1})}\tilde{T}_n] &= (n-j).0 \\
&\quad +\sum_{l=1}^{\infty}\phi^{l-1}8\pi^2\sigma^4(1+\phi)\phi^{l-1}e^{2\pi i(Y_1+Y_0)} \\
&\quad +\sum_{k=1}^{j-2}\sum_{l=1}^{\infty}\phi^{l-1}.8\pi^2\sigma^4(1+\phi)^2\phi^{l+2k-2}e^{2\pi i(Y_1+Y_0)} \\
&= \frac{8\pi^2\sigma^4}{1-\phi}e^{2\pi i(Y_1+Y_0)} \\
&\quad +\frac{8\pi^2\sigma^4}{1-\phi}\sum_{k=1}^{j-2}\phi^{2k-1}(1+\phi)e^{2\pi i(Y_1+Y_0)}
\end{aligned}
$$

Therefore

$$
\begin{aligned}
(n-1)^{-1}C[H_n, \tilde{T}_n] &= \frac{8\pi^2\sigma^4}{1-\phi}e^{2\pi i(Y_1+Y_0)} \\
&\quad +\frac{8\pi^2\sigma^4}{(n-1)(1-\phi)}\sum_{j=2}^{n}\sum_{k=1}^{j-1}\phi^{2k-1}(1+\phi)e^{2\pi i(Y_1+Y_0)} \\
&\to 8\pi^2\sigma^4\Big\{\frac{1}{1-\phi}+\frac{1+\phi}{1-\phi}\sum_{k=1}^{\infty}\phi^{2k-1}\Big\}e^{2\pi i(Y_1+Y_0)} \\
&= \frac{8\pi^2\sigma^4}{(1-\phi)^2}e^{2\pi i(Y_1+Y_0)}
\end{aligned}
$$

as required.

Q.E.D.

Lemma 2.40 *For each $\theta \in \mathcal{P}$, there exists a negative definite matrix $\mathcal{L}$ such that*

$$n^{-1}\tilde{L}''_n(\theta) \stackrel{L_2}{\to} \mathcal{L}$$

Proof: By lemmas 2.21 and 2.36, $n^{-1}\tilde{L}''_n(\theta)$ converges to the average of a stationary sequence. There are then two stages to the proof of this lemma. One is to show that the expectation of the sequence is negative definite, and the second is to show that $D[n^{-1}\tilde{L}''_n(\theta)] \to 0$ as $n \to \infty$.

To prove the second part, note that $n^{-2}D[E[\tilde{S}_n|\mathbf{X}_1^n]] \leq n^{-2}D[\tilde{S}_n] = 2\sigma^4/n \to 0$. By lemma 2.37

$$D[n^{-1}D[\tilde{S}_n|\mathbf{X}_1^n]] \to 0$$

Hence the variance of the $\sigma - \sigma$ component of $n^{-1}\tilde{L}''_n(\theta) \to 0$. The other components of the matrix are similar.

Now

$$E[\tilde{L}''_n] = \begin{pmatrix} -\sigma^{-6}D[E[\tilde{S}_n|\mathbf{X}_1^n]] & \frac{1}{2}\sigma^{-5}C[E[\tilde{S}_n|\mathbf{X}_1^n], E[\tilde{T}_n|\mathbf{X}_1^n]] \\ \frac{1}{2}\sigma^{-5}C[E[\tilde{S}_n|\mathbf{X}_1^n], E[\tilde{T}_n|\mathbf{X}_1^n]] & -\frac{1}{4}\sigma^{-4}D[E[\tilde{T}_n|\mathbf{X}_1^n]] \end{pmatrix}$$

But $D[E[\tilde{S}_n|\mathbf{X}_1^n]] \leq 2n\sigma^4$ and similarly for other terms. Hence $n^{-1}E[\tilde{L}''_n]$ is bounded and therefore from the stationary approximation adduced earlier it converges to some matrix $\mathcal{L}$. It remains to show that $\mathcal{L}$ is negative definite.

To complete the proof, we first show that $\mathcal{L}$ is non-zero. To do this, note that $D[E[\tilde{S}_n|\mathbf{X}_1^n]] \geq |C[E[\tilde{S}_n|\mathbf{X}_1^n], Z_n]|^2.D[Z_n]$. However $C[E[\tilde{S}_n|\mathbf{X}_1^n], Z_n] = C[\tilde{S}_n, Z_n]$ since Z_n is $\mathbf{X}_1^n$-measurable. Further $D[Z_n] = E[Z_n\bar{Z}_n] - |E[Z_n]|^2$ and

$$\begin{aligned} E[Z_n\bar{Z}_n] &= \sum_{j=1}^{n}\sum_{k=1}^{n} E[e^{2\pi i(Y_j - Y_k)}] \\ &= \sum_{j=1}^{n}\sum_{k=1}^{n} \exp\left[-2\pi^2\{1 + 1 - 2\phi^{|j-k|}\}\sigma^2/(1-\phi^2)\right] \\ &= e^{-4\pi^2\sigma^2/(1-\phi^2)} \sum_{j=1}^{n}\sum_{k=1}^{n} \exp\left[4\pi^2\phi^{|j-k|}\sigma^2/(1-\phi^2)\right] \end{aligned}$$

Hence

$$\frac{1}{n}D[Z_n] \quad \to \quad e^{-4\pi^2\sigma^2/(1-\phi^2)} \sum_{k\in\mathbf{Z}} \left\{\exp\left[4\pi^2\phi^{|k|}\right] - 1\right\}$$

$$
\begin{aligned}
&= e^{-4\pi^2\sigma^2/(1-\phi^2)} \sum_{k\in\mathbf{Z}} \sum_{j=1}^{\infty} \left(\frac{4\pi^2\sigma^2}{1-\phi^2}\right)^j \frac{\phi^{j|k|}}{j!} \\
&= e^{-4\pi^2\sigma^2/(1-\phi^2)} \sum_{j=1}^{\infty} \left(\frac{4\pi^2\sigma^2}{1-\phi^2}\right)^j \frac{1}{j!}\frac{1+\phi^j}{1-\phi^j} \\
&> 0
\end{aligned}
$$

and, by applying lemma 2.39

$$
\lim_{n\to\infty} \frac{1}{n} D[E[\tilde{S}_n|\mathbf{X}_1^n]] > 0
$$

Similarly $\lim_{n\to\infty} \frac{1}{n} D[E[\tilde{T}_n|\mathbf{X}_1^n]] > 0$.

The final part of the proof is to show that the matrix $\mathcal{L}$ is negative definite. This follows if the determinant is positive. In order to show this, we must first show that the limit of $\frac{1}{n}D[H_n]$ is positive. However

$$
\begin{aligned}
E[H_n\bar{H}_n] &= \sum_{j=1^n} \sum_{k=1}^{n} E[\exp[2\pi i(Y_j + Y_{j-1} - Y_k - Y_{k-1})]] \\
&= n + 2(n-1)e^{-2\pi^2(1+1-2\phi^2)\sigma^2/(1-\phi^2)} \\
&\quad + \sum_{j=2}^{n-1} 2(n-j) \exp[-2\pi^2(4+4\phi-2\phi^{j-1}-4\phi^j-2\phi^{j+1})\sigma^2/(1-\phi^2)]
\end{aligned}
$$

Hence

$$
\begin{aligned}
\frac{1}{n}D[H_n] &\to e^{-8\pi^2\sigma^2/(1-\phi)} \sum_{j\in\mathbf{Z}} \left\{\exp\left[4\pi^2(\phi^{|j-1|}+2\phi^{|j|}+\phi^{|j+1|})\right]-1\right\} \\
&= e^{-8\pi^2\sigma^2/(1-\phi)} \left\{ e^{8\pi^2(1+\phi)} - 1 + 2\sum_{j=0}^{\infty} \{\exp[4\pi^2(1+\phi)^2\phi^j]-1\} \right\} \\
&= e^{-8\pi^2\sigma^2/(1-\phi)} \left\{ e^{8\pi^2(1+\phi)} - 1 + 2\sum_{k=1}^{\infty} \frac{(4\pi^2(1+\phi)^2)^k}{k!(1-\phi^k)} \right\} \\
&> 0
\end{aligned}
$$

as required.

Put $\hat{S}_n = E[\tilde{S}_n|\mathbf{X}_1^n] - E[\tilde{S}_n]$ and $\hat{T}_n = E[\tilde{T}_n|\mathbf{X}_1^n] - E[\tilde{T}_n]$. But, for any α,

$$
E[(\hat{S}_n - \alpha\hat{T}_n)^2] \geq |C[\hat{S}_n, Z_n]|^2/D[Z_n]
$$

and

$$
E[(\hat{S}_n - \alpha\hat{T}_n)^2] \geq |C[\hat{S}_n, H_n]|^2/D[H_n]
$$

Therefore, since H_n and Z_n are $\mathbf{X}_1^n$-meaurable

$$\begin{aligned} \frac{1}{n}E[(\hat{S}_n - \alpha\hat{T}_n)^2] &\geq \frac{n^{-2}\min(|C[\tilde{S}_n - \alpha T_n, Z_n]|^2, |C[\tilde{S}_n - \alpha T_n, H_n]|^2)}{n^{-1}\max(D[Z_n], D[H_n])} \\ &\rightarrow C_1 > 0 \end{aligned}$$

by lemma 2.39. Hence there exists $C_2 < 1$ such that

$$\lim_{n\to\infty} \rho(E[\tilde{S}_n|\mathbf{X}_1^n], E[\tilde{T}_n|\mathbf{X}_1^n]) = C_2$$

Hence the determinant of $\mathcal{L}$ is simply

$$\lim_{n\to\infty} n^{-2}(1 - C_2^2)D[E[\tilde{S}_n|\mathbf{X}_1^n]]D[E[\tilde{T}_n|\mathbf{X}_1^n]]$$

which is the required result.

$\mathcal{Q.E.D.}$

Lemma 2.41 *For all* $\theta_0 \in \mathcal{P}$

$$\frac{1}{n}|L_n''(\theta_0, \hat{\theta}_n) - L_n''(\theta_0, \theta_0)| \stackrel{a.s.}{\rightarrow} 0$$

in θ_0*-measure.*

Proof: From lemmas 2.38, 2.35 and 2.20, it is easy to see that if B is a compact set containing θ_0, there exists C such that all the third derivatives of $L(\theta)$ are bounded by nC for all n, ω and $\theta \in B$. Hence

$$|L_n''(\theta_0, \hat{\theta}_n) - L_n''(\theta_0, \theta_0)| \leq Cn|\hat{\theta}_n - \theta_0|$$

for $\hat{\theta}_n \in B$. But theorem 2.33 shows that $\hat{\theta}_n \stackrel{\text{a.s.}}{\rightarrow} \theta_0$, i.e. $|\hat{\theta}_n - \theta_0| \stackrel{\text{a.s.}}{\rightarrow} 0$ and the result follows.

$\mathcal{Q.E.D.}$

Theorem 2.42 *For all* $\theta_0 \in \mathcal{P}$*, there exists a positive definite matrix* $\mathcal{L}$ *such that*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \stackrel{d}{\rightarrow} N(0, \mathcal{L}^{-1})$$

in θ_0*-measure.*

Proof: By Taylor expansion

$$\hat{\theta}_n - \theta_0 = -(L''_n(\theta_0, \hat{\theta}_n))^{-1} L'_n(\theta_0)$$

and so

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = (-n^{-1} L''_n(\theta_0, \hat{\theta}_n))^{-1} \frac{1}{\sqrt{n}} L'_n(\theta_0)$$

By lemmas 2.34 and 2.40, there exists a positive definite matrix $\mathcal{L}$ such that

$$\frac{1}{n} L''_n(\theta_0, \theta_0) \stackrel{\text{a.s.}}{\to} -\mathcal{L}$$

and by lemma 2.41

$$\frac{1}{n} |L''_n(\theta_0, \hat{\theta}_n) - L''_n(\theta_0, \theta_0)| \stackrel{\text{a.s.}}{\to} 0$$

So

$$(\frac{1}{n} L''_n(\theta_0, \hat{\theta}_n))^{-1} \stackrel{\text{a.s.}}{\to} -\mathcal{L}$$

It remains to show that

$$\frac{1}{\sqrt{n}} L'_n(\theta_0) \stackrel{\text{d}}{\to} N(0, \mathcal{L})$$

But

$$\begin{aligned} E[\partial_\sigma L_n(\theta)] &= E[\frac{\partial_\sigma f_{\mathbf{X}_1^n}(\mathbf{X}_1^n)}{f_{\mathbf{X}_1^n}(\mathbf{X}_1^n)}] \\ &= \partial_\sigma \int f_{\mathbf{X}_1^n}(\mathbf{x}_1^n) d\mathbf{x}_1^n \\ &= 0 \end{aligned}$$

and similarly for ϕ. Hence

$$\begin{aligned} \frac{1}{n} C[\partial_\sigma L_n, \partial_\phi L_n] &= \frac{1}{n} E[(\partial_\sigma L_n)(\partial_\phi L_n)] \\ &= \frac{1}{n} E[\frac{\partial_\sigma^2 f_{\mathbf{X}_1^n}(\mathbf{X}_1^n)}{f_{\mathbf{X}_1^n}(\mathbf{X}_1^n)}] - \frac{1}{n} E[\partial_\sigma^2 L_n] \\ &= -\frac{1}{n} E[\partial_\sigma^2 L_n] \end{aligned}$$

which converges to the $\sigma - \sigma$ component of $\mathcal{L}$. The result follows provided $L'_n(\theta)$ is asymptotically normal.

But $n^{-\frac{1}{2}} \partial_\sigma L_n(\theta)$ is asymptotically equivalent to $\sigma^{-3} n^{-\frac{1}{2}} E[\tilde{S}_n - n\sigma^2 | \mathbf{X}_{-\infty}^{\infty}]$ by lemma 2.21. Similarly $n^{-\frac{1}{2}} \partial_\phi L_n(\theta)$ is asymptotically equivalent to $\sigma^{-3} n^{-\frac{1}{2}} E[\tilde{T}_n - n\sigma^2 | \mathbf{X}_{-\infty}^{\infty}]$.

Therefore any linear combination of $n^{-\frac{1}{2}}\partial_\sigma L_n(\theta)$ and $n^{-\frac{1}{2}}\partial_\phi L_n(\theta)$ is asymptotically equivalent to

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n} E[t_1(\epsilon_j^2-\sigma^2)+t_2\epsilon_j Y_{j-1}|\mathbf{X}_{-\infty}^{\infty}] \tag{2.66}$$

But, by lemmas 2.20, 2.21 and theorem 2.18 together with lemma 1.3 the terms in (2.66) form a stationary sequence which satisfies the assumptions of theorem A.4 and hence (2.66) converges in distribution to a normal random variable.

$\mathcal{Q.E.D.}$

2.2.6 Quantification of Estimation Properties

The maximum likelihood estimates having been shown to be asymptotically normal, it is important to establish the asymptotic covariance structure of the estimates. From the point of view of using the estimates, empirical standard errors can, of course, be obtained by the usual method of regarding the second derivative matrix of the log-likelihood function at the maximum as a satisfactory approximation to the inverse of the true covariance matrix. However, for the purpose of evaluating the quality of maximum likelihood estimation in comparison to, for example, the crude moment based estimation procedure discussed in section 2.1, it is desirable to compute the covariance matrix of the estimates for a range of values of σ and ϕ.

It is, I believe, clear that this matrix cannot be constructed easily in closed form. Indeed, if it were possible, much of the earlier effort expended would be unnecessary. There are many ways to estimate the covariance structure by simulation, all of which appear to require a large quantity of computer time. The procedure I have chosen is to use the fact that the second derivative matrix divided by the series length converges almost surely to the inverse of the covariance matrix. The series length has to be chosen arbitrarily. I have chosen a series length of 200. The simulation was performed 20 times at each point on a grid of (σ, ϕ) values. σ ranged from 0.015 to 0.405 in steps of 0.015. ϕ ranged from 0 to 0.85 in steps of 0.05. Clearly this will give rise to estimates of the asymptotic second derivative which have error which will in turn give rise to estimates of the correlation matrix with error. The results of the simulation are presented in figure 2.5 in the form of asymptotic log standard deviations for $\sqrt{n}\hat{\phi}$ and $\sqrt{n}\hat{\sigma}$ and also

Figure 2.5: Asymptotic log of standard deviation and correlation between the parameter estimates for maximum likelihood estimation

the asymptotic correlation between these two variables. The perspective plots shown are noisy because of the estimation procedure. The results compare favourably with those shown for moment estimation in figure 2.1. Figure 2.6 is a comparative table giving asymptotic log standard deviations of the parameter estimates and their correlation for the two estimation procedures at a number of values of the parameters. Note that the asymptotic behaviour of the maximum likelihood estimate is always better and that for extreme values of the parameters it is enormously better. Admittedly some allowance needs to be made for error in these figures because of the way in which they were computed for maximum likelihood estimation.

2.2.7 Computational Difficulties

The difficulties which arise with maximum likelihood estimation for this model fall into two classes — calculating the likelihood function and maximising it.

Calculating the likelihood

As seen before, the likelihood cannot be exactly calculated since it involves infinite summations in the formulae

$$f_{X_t|\mathbf{X}_1^{t-1}}(x_t|\mathbf{x}_1^{t-1}) = \sum_{j,k\in\mathbf{Z}} (\sigma\sqrt{2\pi})^{-1}\exp\Big(-\tfrac{1}{2}\sigma^{-2}(x_t+j-\phi(x_{t-1}+k))^2\Big)a_{t-1,k}$$

$$a_{t,j} = \frac{\sum_{k\in\mathbf{Z}}(\sigma\sqrt{2\pi})^{-1})\exp\Big(-\tfrac{1}{2}\sigma^{-2}(x_t+j-\phi(x_{t-1}+k))^2\Big)a_{t-1,k}}{f_{X_t|\mathbf{X}_1^{t-1}}(x_t|\mathbf{x}_1^{t-1})}$$

$$f_{X_1}(x_1) = \frac{\sqrt{1-\phi^2}}{\sigma\sqrt{2\pi}}\sum_{j\in\mathbf{Z}}\exp\left(-\frac{(1-\phi^2)}{2\sigma^2}(x_1+j)^2\right)$$

$$a_{1,j} = \frac{\sqrt{1-\phi^2}}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(1-\phi^2)}{2\sigma^2}(x_1+j)^2\right)/f_{X_1}(x_1)$$

which hold for $t > 1$.

However, the likelihood can obviously be approximated by performing each sum from $-N$ to N for some N. This does not allow for the propagation of errors in the $a_{t,j}$ which will result. This last is rendered less serious by three facts:

1. Lemma 2.8 guarantees that for the true $a_{t,j}$ (as opposed to the approximations) we can make the finite sum arbitrarily close to 1 by an appropriate choice of N, and this N can be chosen independent of the sample sequence or its length.

Figure 2.6: Table of values of log standard deviations of the parameter estimates and their correlation for maximum likelihood and moment estimation

2. Since the rate of convergence in the above to 1 is extremely fast, a small increase in N will permit a large increase in the sample length and still yield the same accuracy of approximation.

3. The rate of convergence is so rapid that, for any sensible sample size (say less than 10000) on a reasonably large computer such as a VAX 11/780, the approximate likelihood can be calculated for a value of N sufficiently large so as to render all dropped terms smaller than can be represented on the computer.

Maximising the likelihood

Simulation shows that for some parameter values (those with large σ^2 or $|\phi|$ close to 1) the likelihood surface is so flat away from the maximum that without an extremely good choice of initial point, conventional gradient based function maximisers simply stop and do not find the true maximum of the likelihood. This is perhaps due to errors in numerical approximations to the gradient of the likelihood function.

An alternative to the straightforward maximisation of the likelihood function is the EM-algorithm of [8] which uses the statistical structure of the model to find the maximum of the surface. It is suitable for those models where the likelihood function results from an underlying likelihood function by the loss of information and where the underlying likelihood is easy to maximise.

This is exactly the situation for the wrapped AR(1). The likelihood function arises from that of a true AR(1) by the loss of the information contained in the integer part of each observation. Further, the likelihood function for an AR(1) is easy to maximise, requiring only the computation of the sufficient statistics g_1 and g_0 from the sequence, and then setting $\hat{\phi} = r_1 = g_1/g_0$ and $\hat{\sigma}^2 = g_0 - \hat{\phi} g_1$.

The EM-algorithm consists of iteration of two simple steps. Beginning with some initial guess for the parameters

1. Calculate the expectations of the sufficient statistics conditional upon the data using the current parameter guess

2. Use these expected sufficient statistics to obtain new parameter estimates and repeat the previous step.

These two steps are then repeated until convergence of the parameter estimates occurs.

In the case of the wrapped AR(1) this becomes

- A. Obtain initial parameter estimates $\hat{\phi}_1$ and $\hat{\sigma}_1^2$, perhaps by moment estimation.

- B. On the j^{th} time through, calculate

$$\psi_{1,j} = E_{\hat{\theta}_j}[\frac{1}{n}\sum_{k=1}^{n} Y_k^2 | \mathbf{X}_1^n = \mathbf{x}_1^n]$$

  and

$$\psi_{2,j} = E_{\hat{\theta}_j}[\frac{1}{n}\sum_{k=2}^{n} Y_k Y_{k-1} | \mathbf{X}_1^n = \mathbf{x}_1^n]$$

- C. Set $\hat{\phi}_{j+1} = \psi_{1,j}/\psi_{2,j}$ and $\hat{\sigma}_{j+1}^2 = \psi_{2,j} - \hat{\phi}_{j+1}\psi_{1,j}$.

- D. Test if the sequence $\hat{\theta}_1, \ldots, \hat{\theta}_{j+1}$ has converged. If not go to B.

The obvious issue is whether the sequence necessarily converges to the parameter value maximising the likelihood function. In [8] a proof is given showing that it does for certain simple models. For the wrapped AR(1) that proof does not suffice. However simulations suggest that in fact the convergence is to the maximum.

A flaw of the EM-algorithm which has been noted is that it converges extremely slowly in many cases (see [21]). This is the case for the wrapped AR(1). Depending on the initial guess, perhaps as many as 10^6 iterations might be required. This would appear to render the method useless for this problem. Fortunately the convergence, althhough slow, is extremely regular. This enables the use of a convergence acceleration algorithm, to speed the process. Numerical analysts have long been familiar with the ϵ-algorithm for speeding up the convergence of real-valued sequences. A number of algorithms derived from this for accelerating the convergence of vector-valued sequences may be found in [4]. One of these, called the vector ϵ-algorithm, has proven to be extremely successful in simulations.

The vector ϵ-algorithm is defined as follows. Given a sequence $\mathbf{x}_1, \mathbf{x}_2, \ldots$ we wish to accelerate the convergence of the sequence to its limit. the principle behind the algorithm is that the original sequence should be converging approximately according to some power law. The algorithm produces a collection of improved estimates

$\epsilon_0^{(j)}, \epsilon_2^{(j)}, \epsilon_4^{(j)}, \ldots$ at each point of the original sequence. $\epsilon_0^{(j)}$ is simply the original sequence. $\epsilon_2^{(j)}$ is computed from $\mathbf{x}_j, \mathbf{x}_{j+1}, \mathbf{x}_{j+2}$. $\epsilon_4^{(j)}$ is computed from $\mathbf{x}_j, \mathbf{x}_{j+1}, \mathbf{x}_{j+2}, \mathbf{x}_{j+3}, \mathbf{x}_{j+4}$. The definition is as follows

$$\epsilon_{-1}^{(j)} = 0$$

$$\epsilon_0^{(j)} = \mathbf{x}_j$$

$$\epsilon_{n+1}^{(j)} = \epsilon_{n-1}^{(j+1)} + \left(\epsilon_n^{(j+1)} - \epsilon_n^{(j)}\right)^{-1}$$

for $n >= 1$. Since the quantities involved are vectors a definition needs to be supplied for the inverse of a vector. The suggested form is

$$\mathbf{x}^{-}1 = \frac{\mathbf{x}}{\mathbf{x}.\mathbf{x}}$$

where . denotes the scalar product.

My use of this procedure has been somewhat *ad hoc*. I found that in practise the best results were obtained with ϵ_2 or ϵ_4. The procedure I followed was

1. Start with some initial guess for the parameter values
2. Iterate with the EM-algorithm computing $\epsilon_2^{(j)}$ for the sequence until the ϵ_2 values stabilised sufficiently.
3. Use the stable point of ϵ_2 as a new starting point for the EM-algorithm.
4. Repeat the last two steps until the sequence of stable ϵ_2 values converged.

The benefit of this procedure was obvious in the case when the initial parameter guess was bad. The EM-algorithm would take at least some thousands of iterations. For a large series or large values of σ and ϕ this would take more computer time than was actually available. The accelerated procedure would converge after only one or two hundred evaluations of the likelihood function — a very considerable improvement. If the initial guess was good, — that is, on the upward slope to the maximum — there was no appreciable difference between the ordinary EM-algorithm and the accelerated procedure. The accelerated procedure failed altogether in certain circumstances. It is possible for the accelerated estimate to so overshoot the true point that the new starting point for the EM-algorithm is actually further away from the true maximum than the original guess. This occurred rarely and it should be possible to detect this sort of wild swing and avoid it. Using ϵ_6 or higher order methods was not successful. I believe the reason for this to lie in the assumption of power series convergence for the unaccelerated sequence, which does not neccesarily hold for the EM-iterations. The higher order ϵ-accelerations are more sensitive to irregularities in the convergence of the original series. My overall conclusion is that the ϵ-algorithm is worth the extra programming required but that it should be used with a certain amount of care, as intervention by the user may be required.

Chapter 3

Markov Models

This chapter examines some aspects of Markov processes on the circle. The existence of stationary processes for given transition functions is analysed, and it is seen that, as in the case of finite Markov chains, a unique stationary measure exists under mild conditions on the transition function. A number of bivariate circular distributions have been proposed in the literature. These naturally give rise to transition functions for Markov processes. Some details of these distributions are considered and a comparison is made of the types of conditional behaviour which would arise from them. First order Markov behaviour is often too restrictive for modelling dependence in series and a method for deriving higher order models is proposed which does not require the use of families of multivariate distributions for three or more circular random variables. Finally, brief consideration is given to the question of estimation for Markov models, in which it is shown that maximum likelihood estimation is consistent and asymptotically normal under certain useful conditions.

3.1 Requirements for Stationary Models

In this section I examine the constraints imposed upon Markov models by stationarity, in particular the constraints imposed upon the transition function.

Let X_t be a homogeneous Markov process on the circle defined by

$$f_{X_t|\mathbf{X}_1^{t-1}}(x_t|\mathbf{x}_1^{t-1}) = g(x_t|x_{t-1}) \tag{3.1}$$

where g is some conditional density and where $f_{X_1}(x_1)$ is some initial distribution.

This will only define a stationary Markov process for certain (if any) choices of the initial distribution. Fortunately, Markov processes on any compact space, such as the circle, have a great deal in common with Markov chains on finite sets. The following material, which is drawn from [9], shows that, for most choices of the transition function, there exists a unique initial distribution which gives rise to a stationary process.

Definition 3.1 *A transition function P on a space X is said to satisfy the Doeblin hypothesis if there exists a finite measure ϕ on X with $\phi(X) > 0$, an integer $n \geq 1$ and an $\epsilon > 0$ such that*

$$P^{(n)}(x, A) \leq 1 - \epsilon \quad \text{if} \quad \phi(A) \leq \epsilon$$

Definition 3.2 *Let P be a transition function on a space X which satisfies the Doeblin hypothesis. A set E is said to be an invariant set if*

$$P(x, E) = 1 \quad \textit{for all } x \in E$$

E is said to be a minimal invariant set if $E' \subset E$ and E' invariant implies $\phi(E') = \phi(E)$.

Theorem 3.1 *If the only minimal invariant subset of X with respect to P is X, then there exists a unique stationary probability distribution.*

Proof: This is a corollary of theorem A.5

Q.E.D.

Now I shall apply this to the case of the circle. Suppose that the transition density $f(x|y)$ has a uniform upper bound C. It is easily seen that the Doeblin hypothesis is satisfied. For, taking ϕ to be Lebesgue measure, $n = 1$ and $\epsilon = \frac{1}{2C} \leq \frac{1}{2}$ since $C > 1$ by hypothesis..

$$P(x, A) = \int_A f(y|x)dy \leq \int_A C\, dy \leq C\phi(A)$$

So $P(x, A) \leq \frac{C}{2C} = \frac{1}{2} \leq 1 - \epsilon$ when $\phi(A) \leq \epsilon$.

Further, if $f(x|y) > 0$ for all x and y, then

$$\int_E f(y|x)\, dy = 1 \quad \Rightarrow \quad \int_{[0,1]\setminus E} f(x|y)\, dy = 0$$

which implies $[0,1] \setminus E$ has null Lebesgue measure. So we have $\phi(E) = \phi([0,1])$. That is $[0,1]$ is a minimal invariant set. So, by theorem 3.1, there exists a unique stationary probability distribution.

Sometimes, instead of defining the transition density, one will specify a function which will be the joint density of X_t and X_{t-1} for any choice of t. Clearly this determines the transition density in (3.1), since this is merely the conditional density of X_t with respect to X_{t-1}. It also clearly defines the initial distribution. For the definition to be consistent we must require that X_t and X_{t-1} have the same marginal distribution. Thus any bivariate circular distribution defines a unique stationary process if, and only if, its marginal distributions are identical.

3.2 Bivariate circular distributions

In this section, I shall examine a number of bivariate distributions which have been proposed by others, and consider their suitability for use as transition distributions for Markov processes. A number of figures are included in this section, which require some explanation. In each case a bivariate distribution of two circular random variables is under consideration. Five of the small sub-figures show a graph of $f_{Y|X}(y|x)$ for a given value of x. The sixth shows the marginal density of Y. The value of x is indicated by a short vertical line from the horizontal axis. For the sake of clarity the range of y for each plot is $[0,2]$ thus giving two copies, side by side, of the density.

3.2.1 Bivariate von Mises distribution

Proposed by Mardia in [24], this distribution is an attempt to define a bivariate analogue of the von Mises distribution on the circle. The density is given by

$$f_{\Theta,\Phi}(\theta,\phi) =$$

$$C \exp\left(\kappa_1 \cos 2\pi(\theta-\mu) + \kappa_2 \cos 2\pi(\phi-\nu) + \rho\sqrt{\kappa_1\kappa_2}\cos 2\pi(\theta \pm \phi - \psi)\right) \qquad (3.2)$$

where $\rho \leq 1$, C is a normalising constant, and κ_1 and κ_2 are positive. This density has the mathematically attractive property of being the maximum entropy density under the constraints $E[e^{i\theta}] = e^{i\mu}$, $E[e^{i\phi}] = e^{i\nu}$ and $E[e^{i(\theta\pm\phi)}] = \rho e^{i\psi}$. It is clear that (3.2) also defines a bivariate density when $\rho > 1$. However, in this case $f(\theta, \phi)$ will not have the property of maximum entropy.

We shall now examine what constraints are imposed on the parameters of the density by the requirement that θ and ϕ should have the same marginal distribution. The marginal density of Φ is known to be (see [24])

$$f_\Phi(\phi) = Ce^{\kappa_2 \cos 2\pi(\phi - \nu)} I_0(Q_\Phi^{\frac{1}{2}})$$

where

$$\begin{aligned} Q_\Phi = \kappa_1^2 &+ \rho^2 \cos^2 2\pi\phi + \rho^2 \sin^2 2\pi\phi \\ &+ 2\kappa_1\rho(\cos 2\pi\psi \cos 2\pi\mu + \sin 2\pi\psi \sin 2\pi\mu) \cos 2\pi\phi \\ &+ 2\kappa_1\rho(-\sin 2\pi\psi \cos 2\pi\mu + \cos 2\pi\psi \sin 2\pi\mu) \sin 2\pi\phi \end{aligned}$$

and I_0 is the incomplete Bessel function of order 0. A similar expression holds for the marginal density of Θ. We see therefore that for the marginal densities to be the same it is necessary and sufficient that $\kappa_1 = \kappa_2$ and $\mu = \nu$. Mardia shows in [24] that the marginals cannot be von Mises unless the variables are independent. This is clear, for if the marginal distribution of Φ is von Mises, then Q_Φ must be constant as a function of ϕ, and so ρ must be zero. So in this respect, at least, the analogy between the bivariate normal and the bivariate von Mises distribution is incomplete.

Now assume that $\kappa_1 = \kappa_2$ and $\mu = \nu$. We shall examine the conditional and marginal distributions in more detail. First, the conditional densities are unimodal, since for any given ϕ

$$f_{\Theta|\Phi}(\theta|\phi) \propto \exp\left(\kappa \cos 2\pi(\theta - \mu) + \kappa \cos 2\pi(\phi - \mu) + \rho\kappa \cos 2\pi(\theta - \phi - \psi)\right)$$

and so

$$\begin{aligned} \frac{\partial}{\partial\theta} f_{\Theta|\Phi}(\theta|\phi) \quad \propto \quad & \{-2\pi\kappa \sin 2\pi(\theta - \mu) - 2\pi\rho\kappa \sin 2\pi(\theta - \phi - \psi)\} \\ & \exp\left(\kappa \cos 2\pi(\theta - \mu) + \kappa \cos 2\pi(\theta - \phi - \psi)\right) \end{aligned}$$

Equating the derivative to zero gives

$$\sin 2\pi(\theta - \mu) + \rho \sin 2\pi(\theta - \phi - \psi) = 0$$

which, by expansion, gives

$$\sin 2\pi(\theta-\mu)\{1+\rho\cos 2\pi(\phi-\mu+\psi)\}-\cos 2\pi(\theta-\mu)\rho\sin 2\pi(\phi-\mu+\psi)=0$$

Therefore, either $1+\rho\cos 2\pi(\theta-\mu+\psi)=0$ or

$$\tan 2\pi(\theta-\mu)=\frac{\rho\sin 2\pi(\phi-\mu+\psi)}{1+\rho\cos 2\pi(\phi-\mu+\psi)} \tag{3.3}$$

1. $1+\rho\cos 2\pi(\theta-\mu+\psi)=0$ implies that $\cos 2\pi(\theta-\mu)=0$. This is only possible for two values of θ, one of which must correspond to a maximum of the density, and the other to a minimum, since it is continuous.

2. Equation (3.3) is also only possible for two values of θ and again one must correspond to a maximum of the density and the other to a minimum.

Figure 3.1 shows graphs of the conditional density for certain values of ϕ and the marginal density, when $\kappa = 0.4$, $\rho = 0.8$, $\psi = 0$ and $\mu = 0$. Certain features of these graphs appear to be typical of the behaviour of this density. The mode of the conditional density appears to be "pulled" towards the value of ϕ being conditioned upon. However, as ϕ passes through 0.5, the mode must move from one side of the circle to the other. This happens by having the mode move back towards 0 as ϕ gets close to 0.5. It is also noteworthy that while the marginal density of Θ is not von Mises, it is at least unimodal and does not seem an unreasonable circular density, though it is possible that this may not be true for other values of the parameters.

3.2.2 Johnson and Wehrly distributions

This family of distributions was proposed in [33]. The joint density of Θ and Φ is given by

$$f_{\Theta,\Phi}(\theta,\phi)=g(F_1(\theta)-F_2(\phi))f_1(\theta)f_2(\phi)$$

where g, f_1 and f_2 are univariate circular densities, and F_1 and F_2 are the distribution functions corresponding to f_1 and f_2.

The marginal densities of Θ and Φ are f_1 and f_2 respectively. For

$$f_\Theta(\theta) \;=\; \int_0^1 f_{\Theta,\Phi}(\theta,\phi)\,d\phi$$

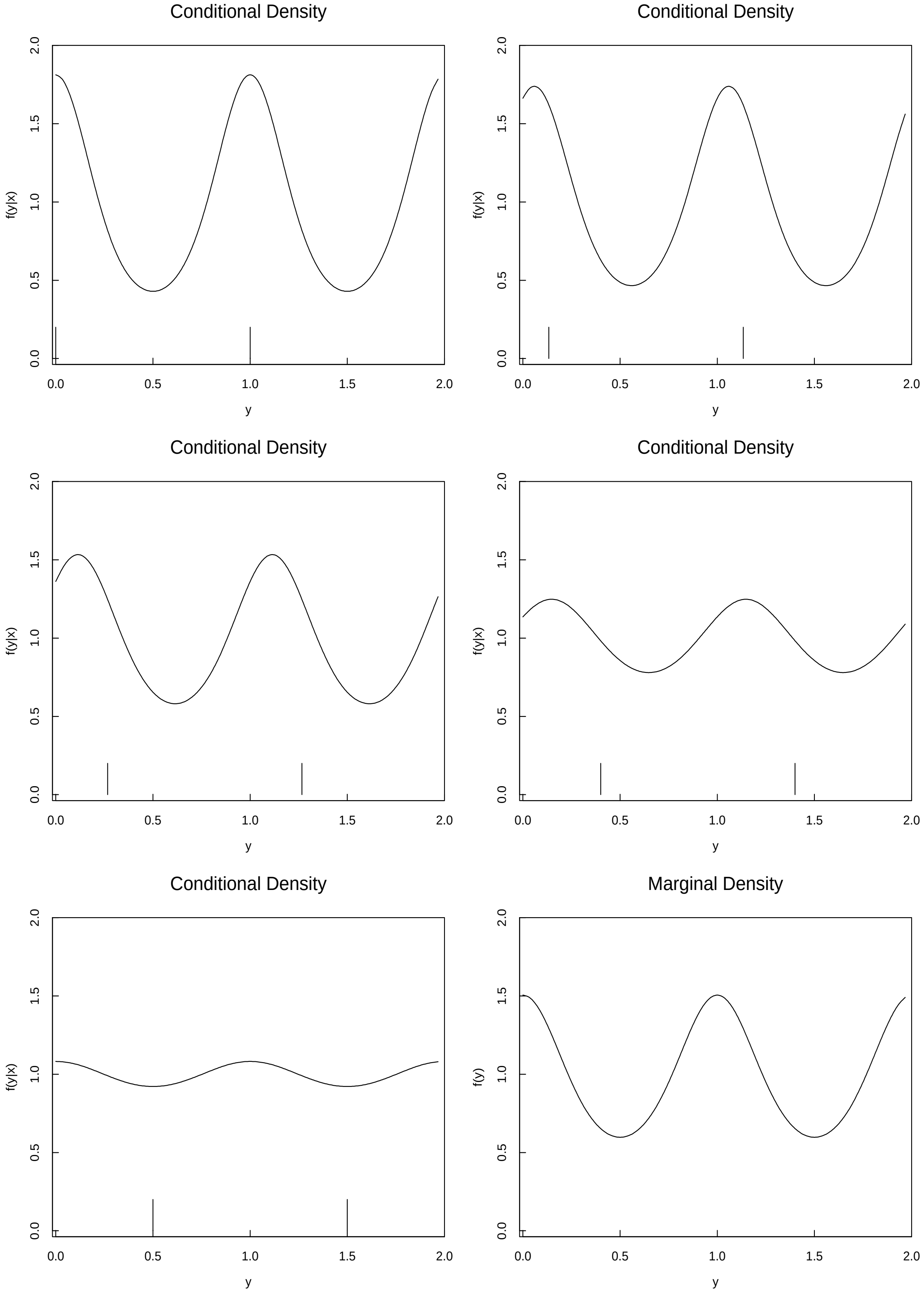

Figure 3.1: Graphs of the marginal and conditional densities for the bivariate von Mises distribution with $\kappa = 0.4$, $p = 0.8$, $\psi = 0$ and $\mu = 0$.
Each figure shows a density function. The range of values of y ranges from 0 to 2, thus giving two copies of the density side by side. The first five plots are of the density $f_{Y|X}(y|x)$ for different values of x, the value of x being indicated by the short vertical lines. The sixth plot shows the marginal density $f_Y(y)$.

$$
\begin{aligned}
&= \int_0^1 g(F_1(\theta) - F_2(\phi)) f_1(\theta) f_2(\phi)\, d\phi \\
&= f_1(\theta) \int_0^1 g(F_1(\theta) - z) dz \\
&= f_1(\theta)
\end{aligned}
$$

and, similarly $f_\Phi(\phi) = f_2(\phi)$.

If we now impose the constraint that the marginals of Θ and Φ be the same, we obtain

$$f_{\Theta,\Phi}(\theta, \phi) = g(F(\theta) - F(\phi)) f(\theta) f(\phi)$$

One might think that this would be a very rich family of bivariate distributions, but this is not the case. Suppose we have a Markov process Θ_t with this distribution defining the transition function. Then

$$f_{\Theta_t|\mathbf{\Theta}_1^{t-1}}(\theta_t|\theta_1^{t-1}) = g(F(\theta_t) - F(\theta_{t-1})) f(\theta_t)$$

If we suppose that $f(\theta)$ is positive for almost all θ, then the inverse function of F exists, and we can define a new process Z_t by $Z_t = F(\Theta_t)$ without any loss of information. But

$$f_{Z_t|\mathbf{Z}_1^{t-1}}(z_t|\mathbf{z}_1^{t-1}) = g(z_t - z_{t-1})$$

or, in other words Z_t is a random walk. So by using this family of distributions, the only Markov processes we can obtain are those which are univariate transformations of random walks. Consider the special case when g and f are von Mises densities. i.e.

$$f(x) = \frac{1}{I_0(\kappa_f)} e^{-\kappa_f \cos 2\pi x}$$

Figure 3.2 shows graphs of the conditional densities and the marginal density when $\kappa_g = 0.6$ and $\kappa_f = 0.5$. We see that, as in the case of the bivariate von Mises, Θ is "pulled" toward Φ which is what we would expect when we almost have random walk behaviour. However the "pull" does not weaken as Φ tends towards 0.5. The bimodality of the conditional density may be seen as a consequence of this fact, since as Φ passes through 0.5 the mode of the distribution must change from one side of the circle to the other. Therefore either there must be a discontinuity or the density must be bimodal. In the next section we shall see an example of a distribution which has the discontinuity instead.

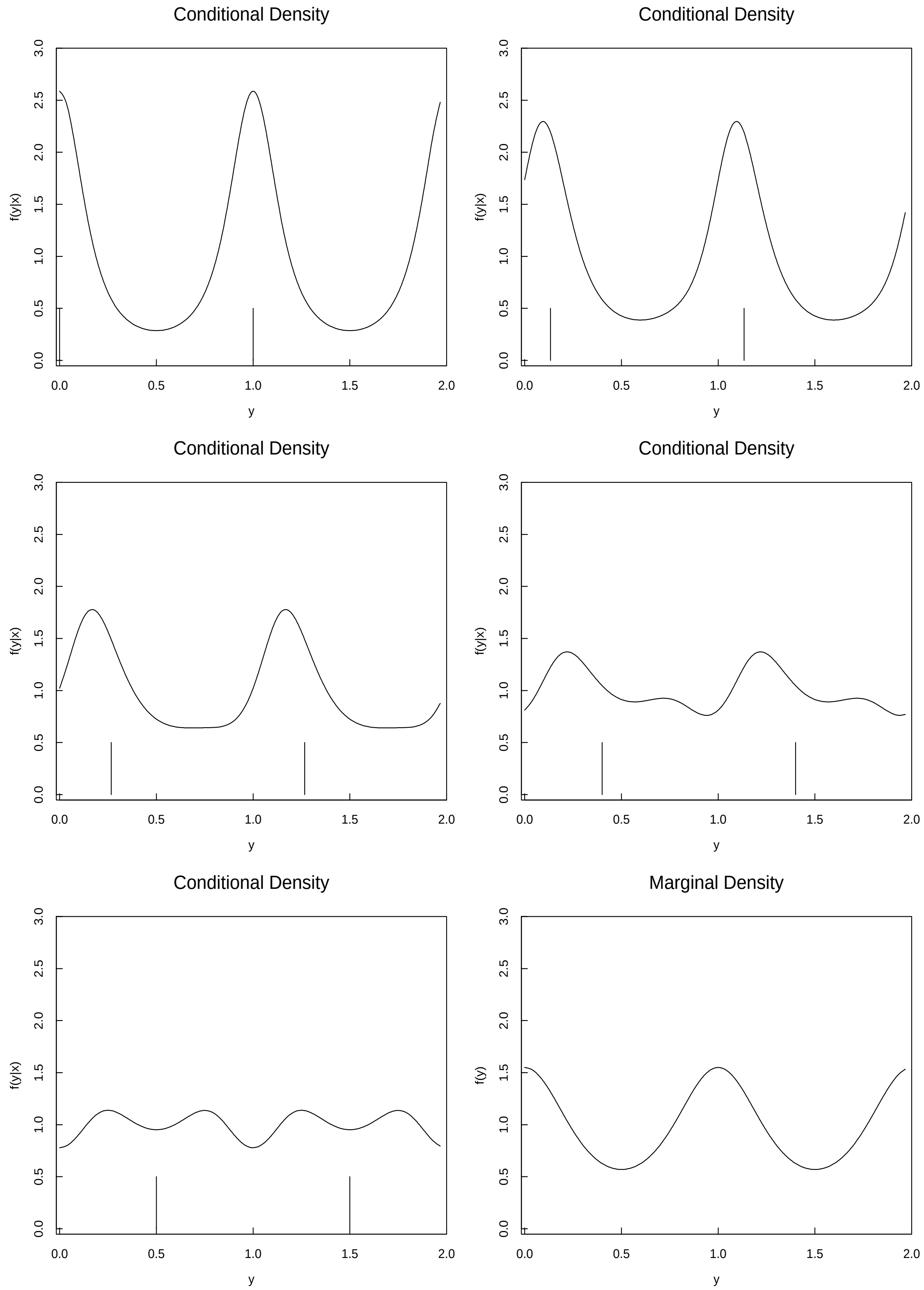

Figure 3.2: Graphs of the conditional and marginal densities for the von Mises based Wehrly and Johnson distribution with $\kappa_f = 0.5$ and $\kappa_g = 0.6$. For an explanation of te information displayed see figure 3.1

3.2.3 Saw's Distributions

In [30] Saw proposed two families of bivariate distributions on the n-dimensional hypersphere and noted that the two approaches he used could be combined. For the circle this reduces to using a bivariate distribution on $[-1, 1]$ to determine the distribution of $(\cos\Theta, \cos\Phi)$, and independently using some bivariate distribution on $\{-1, 1\}$ to generate the sign of $\sin\Theta$ and the sign of $\sin\Phi$. i.e.

$$f_{\Theta,\Phi}(\theta, \phi) = P[\mathrm{sign}(\Theta) = \mathrm{sign}(\theta) \cap \mathrm{sign}(\Phi) = \mathrm{sign}(\phi)] f_{|\Theta|,|\Phi|}(|\theta|, |\phi|)$$

providing we assume that Θ and Φ are in $[-\frac{1}{2}, \frac{1}{2}]$.

He suggested further a way for generating distributions for $f_{|\Theta|,|\Phi|}$ having given marginal distributions, based on collections of functions orthogonal to the marginal distributions of $(\cos\Theta, \cos\Phi)$. In the particular case where we want uniform marginals for Θ and Φ, the appropriate set of functions are the Gegenbauer polynomials, which for the circle are proportional to $\cos n\theta$. The net result of this is the density

$$f_{|\Theta|,|\Phi|}(|\theta|, |\phi|) = 1 + \sum_{j=1}^{\infty} \alpha_j \cos j\theta \cos j\phi$$

where the α_j are chosen so that this is a density. Clearly this has uniform marginals for $|\Theta|$ and $|\Phi|$. In fact, it can be generalised further to

$$f_{|\Theta|,|\Phi|}(|\theta|, |\phi|) = 1 + \sum_{j,k=1}^{\infty} \alpha_{j,k} \cos j\theta \cos k\phi$$

and still have the desired marginality.

A choice must also be made for the distribution of $\mathrm{sign}(\Theta)$ and $\mathrm{sign}(\Phi)$. The most general possible case is

	sign(Θ)	
sign(Φ)	-1	1
-1	p	q
1	r	$1-p-q-r$

However if we impose the constraint of Θ and Φ having the same marginal distributions, then this must also apply to the marginal distributions of $\mathrm{sign}(\Theta)$ and $\mathrm{sign}(\Phi)$ and so we get

	sign(Θ)	
sign(Φ)	-1	1
-1	p	q
1	q	$1-p-2q$

However, this will generally give rise to marginal and conditional densities for Θ and Φ having discontinuities unless we also require $P[\text{sign}(\Theta) = -1] = P[\text{sign}(\Theta) = 1] = \frac{1}{2}$, in which case we get

	sign(Θ)	
sign(Φ)	-1	1
-1	p	q
1	q	p

where $p + q = \frac{1}{2}$. However, we shall see that the conditional densities have discontinuities unless $p = q = \frac{1}{4}$.

Figure 3.3 illustrates the conditional densities and the marginal behaviour in the case where the $\alpha_{j,k}$ are zero except for $\alpha_{11} = 0.7$. The marginals are continuous with $p = 0.7$ and $q = 0.3$. We see that the conditional densities have discontinuities at 0 and 0.5. In the (somewhat dubious) sense that they have a mode, we can see that Θ is "pulled" towards Φ, as Φ tends toward 0.5. The problem of what happens to the mode as Φ passes through 0.5 is solved in this case by the discontinuity at 0.5.

The question arises as to whether the Saw distribution is truly a circular one. Since $\cos\Theta$ and $\text{sign}(\sin(\Theta))$ are independent it can only have a limited range of marginal densities. Also because the dependence between $\cos(\Phi)$ and $\cos(\Theta)$ is unconnected with the dependence between $\text{sign}(\Phi)$ and $\text{sign}(\Theta)$, it can only model a very restricted form of dependence between circular random variables. On the other hand, the same is true of Wehrly and Johnson's family of distributions. The Saw family can obviously model a very great range of dependence between $\cos\Theta$ and $\cos\Phi$. This of course is due simply to the large number (potentially infinite) number of $\alpha_{j,k}$ parameters. It is merely a

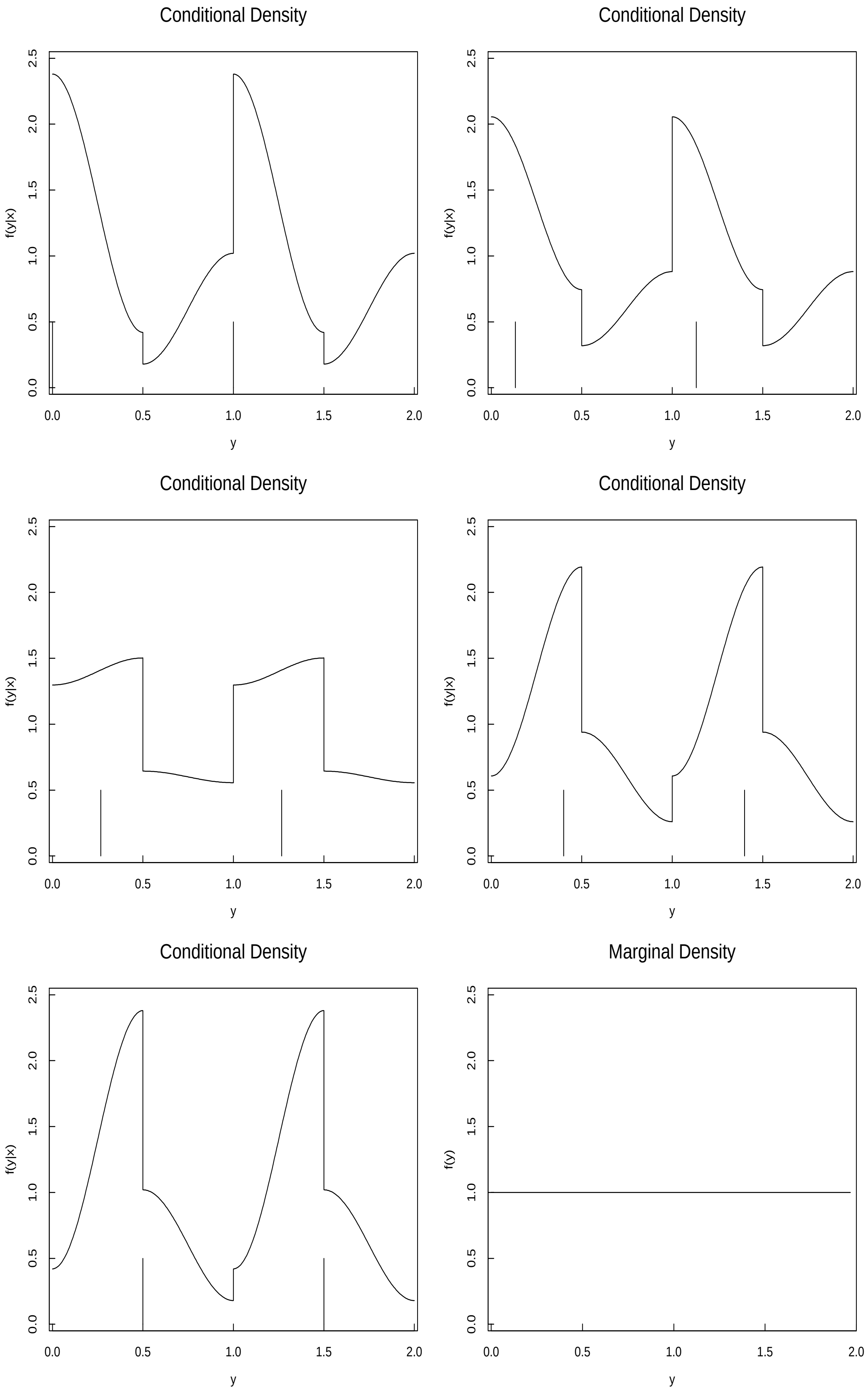

Figure 3.3: Conditional and marginal densities for Saw's distribution, with the only non-zero α_{ij} being $\alpha_{11} = 0.7$ and with $p = 0.7$. For an explanation of the display see figure 3.1.

special case of writing the density as a Fourier series, which is discussed in the section 3.3 and if one is going to use this style of distribution one may as well have the full generality. In higher dimensions, however, the Saw distribution appears to have much nicer properties. The discontinuities disappear. The independence property remains, but the Saw distribution could be useful in testing for exactly such a phenomenon in data.

3.2.4 The Wrapped Bivariate Normal

This distribution is the obvious equivalent of the wrapped univariate normal. The joint density is then

$$f_{\Theta,\Phi}(\theta,\phi) = \sum_{j,k\in\mathbf{Z}} f_{X,Y}(\theta+j,\phi+k)$$

where $f_{X,Y}(x,y)$ is bivariate normal density,

$$X, Y \sim N(0, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix})$$

The marginals are wrapped normal with variances σ_1^2 and σ_2^2 respectively. Under the constraint of identical marginals, obviously we must have $\sigma_1 = \sigma_2$. The conditional densities appear to be unimodal from examination of a few cases. Figure 3.4 shows the conditional densities and marginal distribution in the case when $\sigma = 0.3$ and $\phi = 0.5$. There seems to be a great deal of similarity between this distribution and the bivariate von Mises. Perhaps this is not surprising in view of the similarity between the univariate versions of these distributions.

3.2.5 Comparison and Conclusions

When considering which, if any, of the preceding families of bivariate distributions may be most useful, a number of considerations arise. One must consider marginal behaviour, conditional behaviour, computational convenience and mathematical convenience.

The marginal behaviour of the wrapped normal and bivariate von Mises distributions is similar, and must be unimodal. The wrapped normal has a certain advantage,

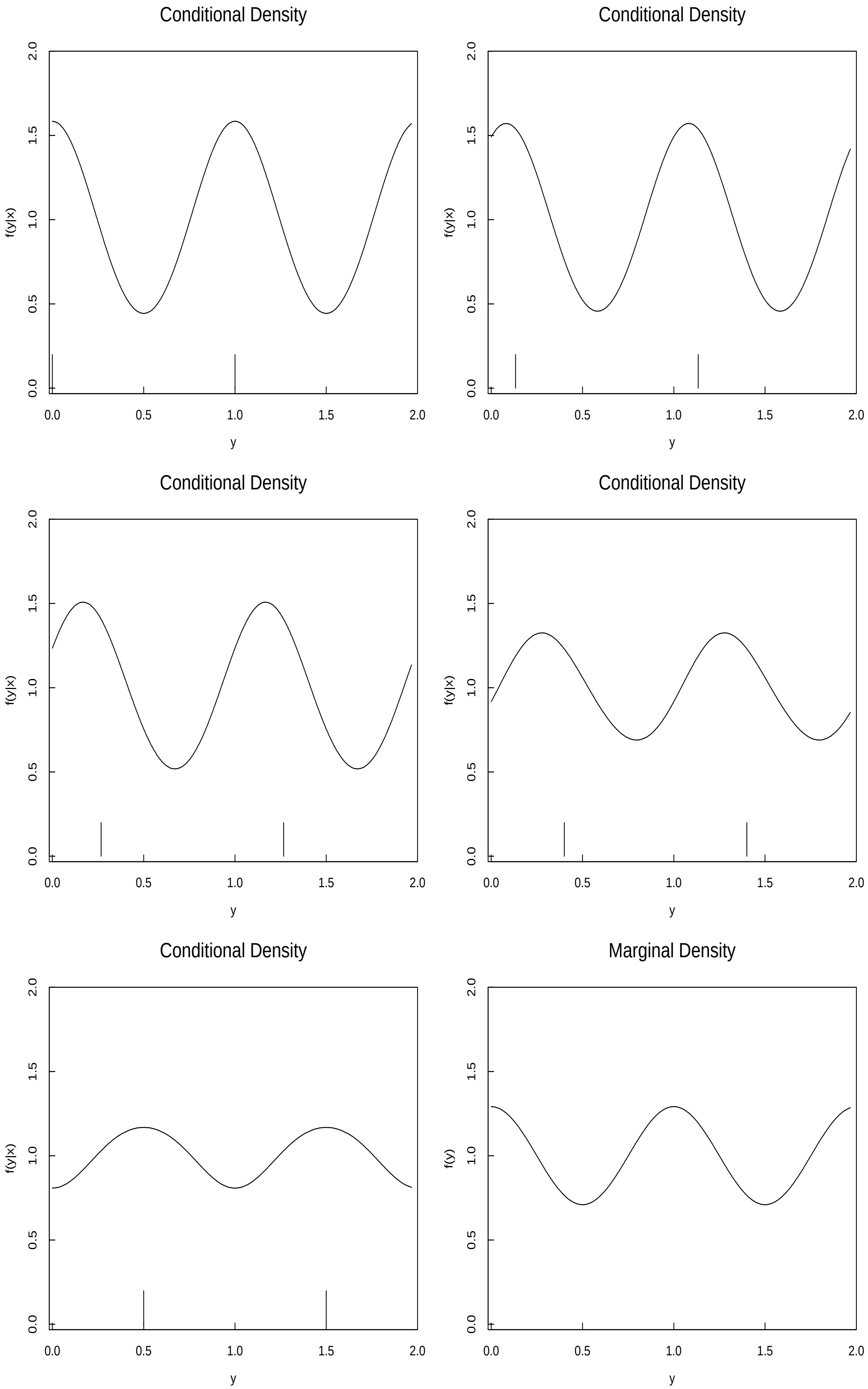

Figure 3.4: Conditional and marginal densities for the wrapped bivariate normal distribution with $\sigma^2 = 0.25$ and $\phi = 0.6$. For an explanation of the display see figure 3.1

in that its marginal behaviour is the familiar wrapped normal. The Wehrly and Johnson family, being simple transformations of random walks, are capable of any marginal behaviour, and the same is true for the Saw family due to the large number of parameters.

The conditional behaviour of these families is the most interesting feature from the point of view of time series analysis. One might consider that the simplest form of dependence other than a pure random walk is that where the Y is pulled towards X away from zero, but does not actually have its mode as far round the circle as X. All of these families, except the Saw distribution, display this property to some degree. In fact no bivariate circular distribution can have this property totally and be completely continuous in its behaviour. The reason is quite simple. Consider the density $f_{Y|X}(y|x)$. The attraction towards X may be expressed as saying that the mode of f as a function of y for fixed x should be found somewhere around the circle in between 0 and x. Then there is clearly a difficulty as y passes through 0.5, as the mode must "jump" from one side of the circle to the other. Each of the families gets around this in a different way.

1. The bivariate von Mises lets the mode be pulled towards x more strongly the further x is from 0, except that as x gets close to 0.5, the pull weakens again until, when $x = 0.5$ the mode has returned to 0.

2. The Wehrly and Johnson family have unimodal conditional densities for x close to 0 and the mode is pulled towards x away from 0. However, as x approaches 0.5 a second mode appears on the other side of the circle and increases in size as x gets closer to 0.5 until, when $x = 0.5$, the two modes are of equal size, enabling a smooth transition.

3. The wrapped normal has unimodal conditional densities. Again there is pull of the mode towards x away from 0. This time however, the solution is that as x approaches 0.5, the pull exerted on the mode increases to the point where the mode is itself 0.5 when $x = 0.5$.

4. The conditional behaviour of the Saw distribution can only be said to be ugly. It seems very unlikely to be useful.

5. A form of behaviour which is not exhibited by any of these families is for the conditional density of Y given X to be continuous for each X but to have a change of behaviour at some value of X. An example where this is the case appears in chapter 4.

Computationally, the wrapped normal and Wehrly and Johnson distributions are easy to use. The summations required for the wrapped normal converge extremely quickly. On the other hand, the von Mises has complicated normalisation, and a difficult marginal distribution. The Saw distribution, with its large number of parameters could require the computation of many transcendental functions.

Mathematically, the wrapped normal and Wehrly and Johnson families are very attractive. The former retains its un-wrapped stability properties and infinite divisibility and has a convenient characteristic function. The latter is easy to work with because it gives rise to stochastic process which are transformations of random walks. The von Mises and Saw distributions do not appear to have any particularly convenient properties.

In conclusion, the wrapped normal distribution appears the most interesting with the Wehrly and Johnson family a good second. However it must be emphasised that a great deal depends on the type of dependence to be modelled.

3.3 Densities as Fourier Series

We can write any bivariate circular density as a Fourier series in its variables. We have

$$f_{\Theta,\Phi}(\theta,\phi) = \sum_{j,k=-\infty}^{\infty} c_{jk} e^{2\pi ij\theta} e^{2\pi ik\phi} \tag{3.4}$$

where various (unknown) constraints are placed on the $c_{j,k}$ in order for this to be a density. In particular, we do know that $c_{00} = 1$ in order for the function to integrate to 1. The marginal distributions are

$$f_{\Theta}(\theta) = \sum_{j=-\infty}^{\infty} c_{j0} e^{2\pi ij\theta}$$

and

$$f_{\Phi}(\phi) = \sum_{k=-\infty}^{\infty} c_{0k} e^{2\pi ik\phi}$$

Thus to satisfy the constraint that Θ and Φ have the same marginal distribution, it is necessary and sufficient that $c_{0j} = c_{j0}$ for all j.

To use this in practice we should have to make a finite approximation to the infinite series. There are severe problems. The number of parameters increases rapidly as we increase the degree of the finite approximation to the sum. Because it takes a large number of terms to adequately Fourier approximate a straight line, simple phenomena of random walk type would require a very large number of the c_{jk} to be non-zero. From the point of view of numerical maximisation of likelihoods, one might be concerned that there would be a "hole" in the space of the c_{jk} where the functions generated were not densities. Fortunately this particular problem does not exist. The set of values of the c_{jk} for which equation 3.4 defines a density is convex since a linear combination with positive coefficients of densities is a density when appropriately normalised.

Log Densities as Fourier Series

A simple modification of the above approach is to write the log of the density as a Fourier series. i.e.

$$\log f_{\Theta,\Phi}(\theta, \phi) = \sum_{j,k=-\infty}^{\infty} c_{jk} e^{2\pi i j\theta} e^{2\pi i k\phi}$$

Here we have no issue of positivity. Integration to 1 is more difficult. Any arbitrary choice of those c_{jk} for which either j or k is non-negative determines a unique value of c_{00} for which we obtain a density. Further, it is not clear what constraints are imposed on the c_{jk} by requiring that the marginals to be the same. This might seem to make this approach useless. However for a large sample, we have the advantage that calculating the log likelihood for a large number of different values does not get more difficult as the sample size grows. The hard part of maximising the likelihood is therefore the constraint on c_{00} which does not depend on the sample size. It may be that, for large samples, this approach is more suitable.

3.4 Markov processes of higher order

First order Markov processes are often inadequate for modelling purposes. Higher orders of dependence are required. In this section a family of higher order Markov models is discussed. These models are known as linear conditional probability models. They were originally proposed in [27] for finite Markov chains and further discussed in a more general context in [25]. The advantages of the approach are that we do not have to find explicit forms for general multivariate circular distributions, but rather can generate a family of higher order conditional distributions using only bivariate distributions. It does however impose severe restrictions upon the type of higher order dependence which results.

Define a model for l^{th} order Markov processes as follows.

$$f_{X_t|\mathbf{X}_1^{t-1}}(x_t|\mathbf{x}_1^{t-1}) = f_{X_t|\mathbf{X}_{t-l}^{t-1}}(x_t|\mathbf{x}_{t-l}^{t-1}) = \sum_{j=1}^{l} \lambda_j g_j(x_t|x_{t-j}) \tag{3.5}$$

where the g_j are some bivariate conditional densities, and choose some initial distribution for $\mathbf{X}_1^l$. Clearly this defines a Markov process whenever (3.5) defines a valid conditional density. Integrating (3.5) over x_t we get $\sum_{j=1}^{l} \lambda_j = 1$. Also since the density must be non-negative this imposes some constraint upon the λ_j. Provided we take $\lambda_j \geq 0$ for all j there is obviously no difficulty. For the remainder of this section consider the case when all the g_j are the same, i.e. $g_j = g$ for all j.

Stationarity

As earlier, stationarity imposes some constraint on the choice of initial distribution. We shall use the theorems of section 3.1 to show that whenever there exists C such that $0 < g(x|y) < C$ for all x and y there is a unique stationary distribution for $X_1, \ldots, X_l$. Note that $g(x|y) \leq C$ implies that $f_{X_t|\mathbf{X}_{t-l}^{t-1}}(x_t|\mathbf{x}_{t-l}^{t-1}) \leq C$. Consider the vector process $Z_t = (X_t, X_{t+1}, \ldots, X_{t+l-1}) = \mathbf{X}_t^{t+l-1}$ Then Z_t is a first order Markov process since the first $l-1$ components of Z_{t+1} $(X_{t+1}, \ldots, X_{t+l-1})$ are determined by Z_t and the dependence of the l^{th} component X_{t+l} on the past is captured entirely in Z_t.

Z_t satisfies the Doeblin hypothesis with ϕ being Lebesgue measure on the product of l copies of the circle, $n = l$, and $\epsilon = \min(\frac{1}{2}, \frac{1}{2C^l})$. This is true because

$$P[Z_{t+l} \in A | Z_t = z_t]$$

$$= P[\mathbf{X}_{t+l}^{t+2l-1} \in A | \mathbf{X}_t^{t+l-1} = \mathbf{x}_t^{t+l-1}]$$

$$= \int_A f_{X_{t+2l-1}|\mathbf{X}_{t+l-1}^{t+2l-2}}(x_{t+2l-1}|\mathbf{x}_{t+l-1}^{t+2l-2}) \times \cdots \times f_{X_{t+l}|\mathbf{X}_t^{t+l-1}}(x_{t+l}|\mathbf{x}_t^{t+l-1})\, d\mathbf{X}_{t+l}^{t+2l-1}$$

$$\leq \int_A C^l \, d\mathbf{X}_{t+l}^{t+2l-1}$$

$$= C^l \phi(A)$$

$$\leq \frac{C^l}{2C^l} = \frac{1}{2}$$

$$\leq 1 - \epsilon$$

when $\phi(A) \leq \epsilon$.

Now suppose E is such that $P[Z_{t+1} \in E | Z_t = z_t] = 1$ whenever $z_t \in E$. Iterating this we get

$$P[Z_{t+l} \in E | Z_t = z_t] = 1 \qquad \text{whenever } z_t \in E$$

But as before

$$P[Z_{t+l} \in E | Z_t = z_t]$$

$$= \int_E f_{X_{t+2l-1}|\mathbf{X}_{t+l-1}^{t+2l-2}}(x_{t+2l-1}|\mathbf{x}_{t+l-1}^{t+2l-2}) \times \cdots \times f_{X_{t+l}|\mathbf{X}_t^{t+l-1}}(x_{t+l}|\mathbf{x}_t^{t+l-1}) \, d\mathbf{X}_{t+l}^{t+2l-1}$$

Clearly $g(x|y) > 0$ for all x and y implies that

$$f_{X_t|\mathbf{X}_{t-l}^{t-1}}(x_t|\mathbf{x}_{t-l}^{t-1}) > 0 \qquad \text{for all } \mathbf{x}_{t-l}^{t-1}$$

and therefore integration to 1 implies E is equal to the whole space of the product of l copies of the circle. Applying corollary 3.1 shows there exists a unique stationary distribution.

Marginal behaviour

It is sometimes important to know the marginal behaviour of the process, i.e. the distribution of a single X_t without reference to any other X_T, $T \neq t$. Now by definition

$$\begin{aligned} f_{\mathbf{X}_{t-l+1}^t}(\mathbf{x}_{t-l+1}^t) &= \int f_{X_t|\mathbf{X}_{t-l}^{t-1}}(x_t|\mathbf{x}_{t-l}^{t-1}) f_{\mathbf{X}_{t-l}^{t-1}}(\mathbf{x}_{t-l}^{t-1}) \, dx_{t-l} \\ &= \sum_{j=1}^{l} \lambda_j \int g(x_t|x_{t-j}) f_{\mathbf{X}_{t-l}^{t-1}}(\mathbf{x}_{t-l}^{t-1}) \, dx_{t-l} \end{aligned}$$

Now integrate withe respect to $x_{t-1}, x_{t-2}, \ldots, x_{t-l+1}$ to get

$$f_{X_t}(x_t) = \sum_{j-1}^{l} \lambda_j \int g(x_t|x_{t-j}) f_{X_{t-j}}(x_t - j) \, dx_{t-j}$$

But assuming the X_t process is stationary $f_{X_{t-j}}(x_{t-j}) = f_{X_t}(x_{t-j})$ and so we get

$$\begin{aligned} f_{X_t}(x_t) &= \sum_{j=1}^{l} \lambda_j \int g(x_t|x_{t-j}) f_{X_t}(x_{t-j})\, dx_{t-j} \\ &= \sum_{j=1}^{l} \lambda_j \int g(x_t|y) f_{X_t}(y)\, dy \\ &= \int g(x_t|y) f_{X_t}(y)\, dy \end{aligned}$$

Thus the marginal distribution of the higher order process is simply the stationary distribution of the first order Markov process defined by g.

3.5 Estimation for Markov Processes

In the next chapter Markov processes wil be used to model data. It will be assumed that maximum likelihood estimation is consistent and that the estimates obtained have, asymptotically, a normal distribution with covariance given by the inverse of the hessian matrix of the log-likelihood function. i.e. the usual properties will be assumed. A brief justification is given here for those assumptions.

Definition 3.3 *Let $P(.,.)$ be a transition probability function on a space X. Then the coefficient of ergodicity $\alpha(P)$ is defined by*

$$\alpha(P) = 1 - \sup_{x,y \in X, B \subset X} |P(x,B) - P(y,B)|$$

The point behind this definition is that if $\alpha(P) > 0$ then the resulting Markov process is strongly mixing with geometrically decreasing mixing coefficients. (For an explanation of mixing see the appendix). For a derivation of this fact see [15]. The rest of the proof draws on the ideas in [7]

Let $f_{X_t|X_{t-1}}(.|.;\theta)$ be a transition density for a circular Markov process with parameters $\theta \in \Theta$. Then, under the following assumptions, the maximum likelihood estimates are weakly consistent and asymptotically normal.

1. $\theta_1 \neq \theta_2$ implies $f_{X_t|X_{t-1}}(.|.;\theta)$ and $f_{X_t|X_{t-1}}(.|.;\theta)$ are not almost everywhere equal.

2. Given $\theta_0 \in \Theta$, there exists a neighbourhood O of θ_0 such that $\inf_{\theta \in O} \alpha(f) > 0$.

3. There exists a continuous $C_1(\theta) < \infty$ such that

$$\left|\ln f_{X_t|X_{t-1}}(x|y;\theta)\right| < C_1 \text{ for all } x,\ y \text{ and } \theta \in \Theta.$$

4. There exists a continuous $C_2(\theta) < \infty$ such that

$$\left|\partial_{\theta_j} \ln f_{X_t|X_{t-1}}(x|y;\theta)\right| < C_2 \text{ for all } j,\ x,\ y \text{ and } \theta \in \Theta.$$

5. There exists a continuous $C_3(\theta) < \infty$ such that

$$\left|\partial_{\theta_j}\partial_{\theta_k} \ln f_{X_t|X_{t-1}}(x|y;\theta)\right| < C_3 \text{ for all } j,\ k,\ x,\ y \text{ and } \theta \in \Theta.$$

6. There exists a continuous $C_4(\theta) < \infty$ such that

$$\left|\partial_{\theta_j}\partial_{\theta_k}\partial_{\theta_l} \ln f_{X_t|X_{t-1}}(x|y;\theta)\right| < C_4 \text{ for all } j,\ k,\ l,\ x,\ y \text{ and } \theta \in \Theta.$$

Denote the log likelihood by $L_n(\theta)$, the vector of derivatives by $L_n'(\theta)$ and the matrix of second derivatives by $L_n''(\theta, \theta_1)$ where the elements are evaluated on the line segment between θ and θ_1.

Since $L_n'(\theta)$ is the sum of a stationary sequence where each term depends only on two observations, and each term is bounded by assumption 4, it converges in distribution to a bivariate normal by theorem A.4. This is the first assumption of theorem A.6.

It is easily shown (and well known) that $n^{-1}E[-L_n''(\theta,\theta)]$ is the covariance matrix of the vector with coefficients $\partial_{\theta_j} \ln f_{X_t|X_{t-1}}(x|y;\theta)$. But

$$E[|\partial_{\theta_j} \ln f_{X_t|X_{t-1}}(x|y;\theta)|^2] > 0$$

for otherwise assumption 1 is contradicted. Equally

$$\rho\Big(\partial_{\theta_j} \ln f_{X_t|X_{t-1}}(x|y;\theta), \partial_{\theta_k} \ln f_{X_t|X_{t-1}}(x|y;\theta)\Big) < 1$$

or again, assumption 1 is contradicted. Thus the smallest eigenvalue of $E[-L_n''(\theta,\theta)]$ converges to infinity. This is the second assumption of theorem A.6.

However, by assumptions 5 and 2 and the fact that $L_n''(\theta,\theta)$ is the sum of a stationary sequence, each term of which depends only on two observations, the strong law of large numbers for mixing processes (theorem A.3) can be applied to show that $n^{-1}L_n''(\theta,\theta)$ converges almost surely to its expectation. This is the third assumption of theorem A.6.

Finally, since the third derivatives are bounded by assumption 6, the fourth assumption of theorem A.6 holds even without the probabilistic statement.

Thus the maximum likelihood estimates are weakly consistent and asymptotically normal.

Chapter 4

Analysis of Wind Directions

In this chapter I attempt to model a series of wind directions observed at Roche's Point in Co. Cork on the south coast of Ireland. There is clear evidence of dependence on the time of day. For this reason the rest of the analysis proceeds using daily averages. It is seen that the models of Chapter 1 are inappropriate to this data. Modelling the data as a Markov process appears to be reasonably successful, using a conditional distribution which is not of any of the types discussed in Chapter 3. By applying the methods of section 3.4, a superior model is obtained. Diagnostic techniques show that there is probably an annual effect in the data which is modelled by allowing some parameters to vary in a seasonal fashion. It is seen that this is not in itself sufficient to account for all of the seasonal behaviour.

A few general points should be noted before proceeding to the actual analysis. Graphical techniques are of great importance. Since the context is that of circular data, it is important to remember that in many cases the left and right sides (and sometimes the top and bottom) of graphs need to be identified together. For example, in figure 4.7, the small clusters of points in the top-left and bottom-right corners are overflows from the bottom-left and top-right corners.

In much of this chapter the methods used are *ad hoc*. In particular, significance tests are rarely performed. This is due largely to the difficulty of performing tests on periodograms. Also the conditional distribution function method of section 4.1.2 is, while potentially powerful, unexplored. Its estimation properties are unknown and it cannot be used quantitatively.

As far as possible, angles are represented as numbers in $[0, 1)$. The exception to this is the raw hourly data for which degrees are used.

4.1 Useful Tools

In order to perform analysis of a circular time series we shall need a couple of tools. The first is merely a clarification of the application of the usual methods of frequency analysis to this kind of data.

4.1.1 The Circular Periodogram

One of the most important ways of detecting periodic behaviour in time series analysis is by the periodogram. This is an estimate of the spectrum of a zero mean stochastic process, given by

$$\hat{g}(p) = \frac{1}{n} \left| \sum_{j=1}^{n} x_j e^{2\pi i p j/n} \right|^2 \qquad \text{for } p = 0, 1, \ldots, n/2 \tag{4.1}$$

for the sequence $x_1, \ldots, x_n$. We expect $\hat{g}(p)$ to be large for those values of p for which the periodic component at frequency p/n is large.

The definition of $\hat{g}(p)$ holds good in the case when $x_1, \ldots, x_n$ are complex, instead of real, numbers. However, any sequence of angles can be considered as a sequence of points on the unit circle in the complex plane. Thus equation 4.1 can be used to define a version of the periodogram for sequences of angles. Since our process may not have zero mean, we shall remove the average from the series (considered as being in the complex plane) before calculating the periodogram. Thus for a sequence of angles $x_1, \ldots, x_n$ given as numbers in the range $[0, 1)$ (i.e. distances along the circumference of a circle of radius $\frac{1}{2\pi}$) we define

$$\hat{g}(p) = \frac{1}{n} \left| \sum_{j=1}^{n} \left(e^{2\pi i x_j} - \tilde{x} \right) e^{2\pi i p j/n} \right|^2$$

where $\tilde{x}$ is the complex average of the complex numbers $e^{2\pi i x_1}, \ldots, e^{2\pi i x_n}$.

We shall use this tool in two circumstances. In the first case we will be looking for evidence of some periodic (daily or annual) behaviour in a sequence of angles. Secondly, we shall use it to test for dependence in a sequence of angles. If $X_1, \ldots, X_n$ is a sequence of uncorrelated (in particular, independent) circular random variables then, by the standard asymptotic theory, $\hat{g}(p)$ is, for large samples, an independent sequence of approximately identically distributed χ_2^2 random variables. This provides a test for dependence in a sequence based on the circular periodogram. For further detail see [16]. Both these uses are largely *ad hoc*.

In practice we shall often use the cumulative periodogram which is the cumulative sum of the $\hat{g}(p)$ normalised so that $\sum_{p=0}^{n/2} \hat{g}(p) = 1$. The most important characteristic of this sequence is that, for a large uncorrelated sequence, it should be approximately a straight line — the Kolmogorov-Smirnov approach for empirical distribution functions can be used to test whether the deviation is significant.

4.1.2 The Conditional Distribution Function

A most important aspect of the practical use of the ARMA models of [3] is the availability of a simple diagnostic technique for evaluating the appropriateness of a fitted model — residual analysis. The simplest case is the AR(1) model, given by

$$X_t = \phi X_{t-1} + \epsilon_t$$

where the ϵ_t are an independent sequence of identically distributed mean 0 normal random variables having variance σ^2. Having estimated ϕ and σ for a sequence $x_1, \ldots, x_n$, we can estimate the ϵ_t by e_t where

$$e_t = x_t - \phi x_{t-1}$$

We can then test the suitability of the model by testing to see if the $e_2, \ldots, e_n$ are compatible with being from an i.i.d sequence of normals.

For general stochastic models, however, no such technique has been available. We can, in fact, define a useful sequence as follows. For any stochastic process X_t define ϵ_t by

$$\epsilon_t = F_{X_t|\mathbf{X}_1^{t-1}}(X_t|\mathbf{X}_1^{t-1})$$

where $F_{X_t|\mathbf{X}_1^{t-1}}$ is the distribution function of X_t conditional upon $X_1, \ldots, X_{t-1}$ for the true model of the process. But, provided it is absolutely continuous, the distribution of $F_{X_t|\mathbf{X}_1^{t-1}}(X_t|\mathbf{X}_1^{t-1})$ is the uniform distribution, for

$$E[\exp(i\alpha F_{X_t|\mathbf{X}_1^{t-1}}(X_t|\mathbf{X}_1^{t-1}))]$$

$$= \int_{\Omega^t} \exp(i\alpha F_{X_t|\mathbf{X}_1^{t-1}}(x_t|\mathbf{x}_1^{t-1}))dF_{\mathbf{X}_1^t}(\mathbf{x}_1^t)$$

$$= \int_{\Omega^{t-1}} \int_{\Omega} \exp(i\alpha F_{X_t|\mathbf{X}_1^{t-1}}(x_t|\mathbf{x}_1^{t-1}))dF_{X_t|\mathbf{X}_1^{t-1}}(x_t|\mathbf{x}_1^{t-1})dF_{\mathbf{X}_1^{t-1}}(\mathbf{x}_1^{t-1})$$

$$= \int_{\Omega^{t-1}} \int_0^1 \exp(i\alpha z)dzdF_{\mathbf{X}_1^{t-1}}(\mathbf{x}_1^{t-1})$$

$$= \int_0^1 \exp(i\alpha z)dz$$

Also, provided $s < t$,

$$E[\exp(i\alpha F_{X_t|\mathbf{X}_1^{t-1}}(X_t|\mathbf{X}_1^{t-1}) + i\beta F_{X_s|\mathbf{X}_1^{s-1}}(X_s|\mathbf{X}_1^{s-1}))]$$

$$= \int \exp(i\alpha F_{X_t|\mathbf{X}_1^{t-1}}(x_t|\mathbf{x}_1^{t-1})) \exp(i\beta F_{X_s|\mathbf{X}_1^{s-1}}(x_s|\mathbf{x}_1^{s-1}))dF_{\mathbf{X}_1^t}(\mathbf{x}_1^t)$$

$$= \int \exp(i\alpha z) \exp(i\beta F_{X_s|\mathbf{X}_1^{s-1}}(x_s|\mathbf{x}_1^{s-1}))dzdF_{\mathbf{X}_1^{t-1}}(\mathbf{x}_1^{t-1})$$

$$= \int \exp(i\alpha z)dz \int \exp(i\beta F_{X_s|\mathbf{X}_1^{s-1}}(x_s|\mathbf{x}_1^{s-1}))dF_{\mathbf{X}_1^s}\mathbf{x}_1^s$$

$$= E[\exp(i\alpha F_{X_t|\mathbf{X}_1^{t-1}}(X_t|\mathbf{X}_1^{t-1}))]E[\exp(i\beta F_{X_s|\mathbf{X}_1^{s-1}}(X_s|\mathbf{X}_1^{s-1}))]$$

and so $F_{X_t|\mathbf{X}_1^{t-1}}(X_t|\mathbf{X}_1^{t-1})$ is independent of $F_{X_s|\mathbf{X}_1^{s-1}}(X_s|\mathbf{X}_1^{s-1})$. Thus the ϵ_t sequence is an independent sequence of uniformly distributed random variables.

We can use this property as a test for fit of a model. If we have a fitted model, we put

$$e_t = \hat{F}_{X_t|\mathbf{X}_1^{t-1}}(x_t|\mathbf{x}_1^{t-1})$$

where $\hat{F}_{X_t|\mathbf{X}_1^{t-1}}$ is the conditional distribution of X_t given $X_1, \ldots, X_{t-1}$ according to the fitted model. Then we can test the sequence $e_1, \ldots, e_n$ for incompatibility with being a realisation of an i.i.d uniform sequence, and this is a test for incompatibility of the fitted model with the data. We shall call the e_t sequence the conditional distribution function sequence.

An issue which arises in the case of modelling an angular sequence is whether the ϵ_t are uniformly distributed on $[0, 1)$ or on a circle of radius $\frac{1}{2\pi}$. The distinction is that in the latter case 0 and 1 are identified with each other. The answer must be that the circle is appropriate, since the origin on the circle with respect to which the distribution function is defined can only be chosen in an arbitrary manner.

4.2 The Raw Data

The data consists of almost nine years of hourly observations of wind direction at the Irish Meteorological Office's station at Roche's Point in Ireland. A plot of part of the series is shown in figure 4.1. The directions are given in degrees measured clockwise from North. As we shall see, it is worth noticing that this station is located on the coast. The total number of observations is 77241. The direction was not measured exactly but is given to the nearest multiple of $10°$, that is, there are 36 possible observed directions. This means that there is a severe grouping effect. This is not apparent from the marginal distribution (a histogram of the data is shown in figure 4.2). If, however, we examine the conditional behaviour we see that the data is heavily discretised as in figure 4.3, which shows a histogram of the observations at time t given that the observation at time $t-1$ was $180°$. On those occasions when no measurable wind was blowing no observation could be made, giving rise to the difficulty of "missing data".

Also, the data shows clear signs of a 24-hour periodic effect. This can be seen from figure 4.4 which is a plot of the cumulative periodogram of the data. This periodogram was obtained by taking the complex Fourier transform of the data regarded as points on the unit circle in the complex plane. There is a marked jump (large when compared to the smoothness of the rest of the curve) at a frequency which corresponds to a 24-hour period. This is to be expected because the directions were observed at the coast. Any sailor is well aware of the phenomena known as "land-breeze" and "sea-breeze". During the day the wind tends to blow from the sea to the land, because the land is colder than the sea, having cooled more quickly during the night. As the day proceeds the land warms more quickly, eventually becoming warmer than the sea and the wind direction then changes and blows from the land to the sea. One would expect there to be some form of yearly behaviour in the wind directions since this temperature phenomenon is more pronounced during the warmer parts of the year, in particular the summer.

If we take 24-hour averages of the data we eliminate this 24-hour non-stationary behaviour. The data also effectively becomes continuous through this operation as can be seen in figure 4.5, which is a plot of the empirical marginal cumulative distribution

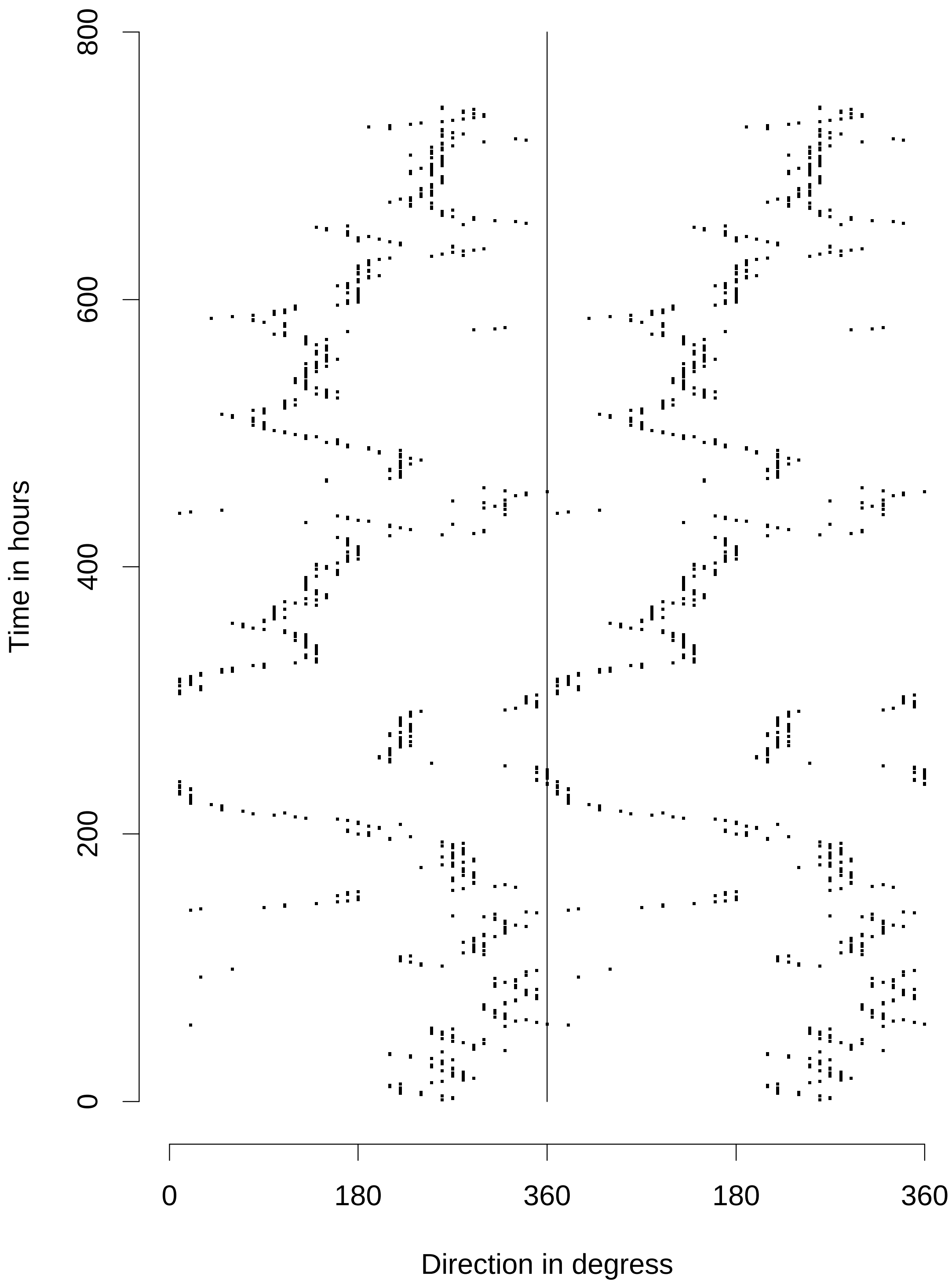

Figure 4.1: Time plot of one month of hourly wind directions

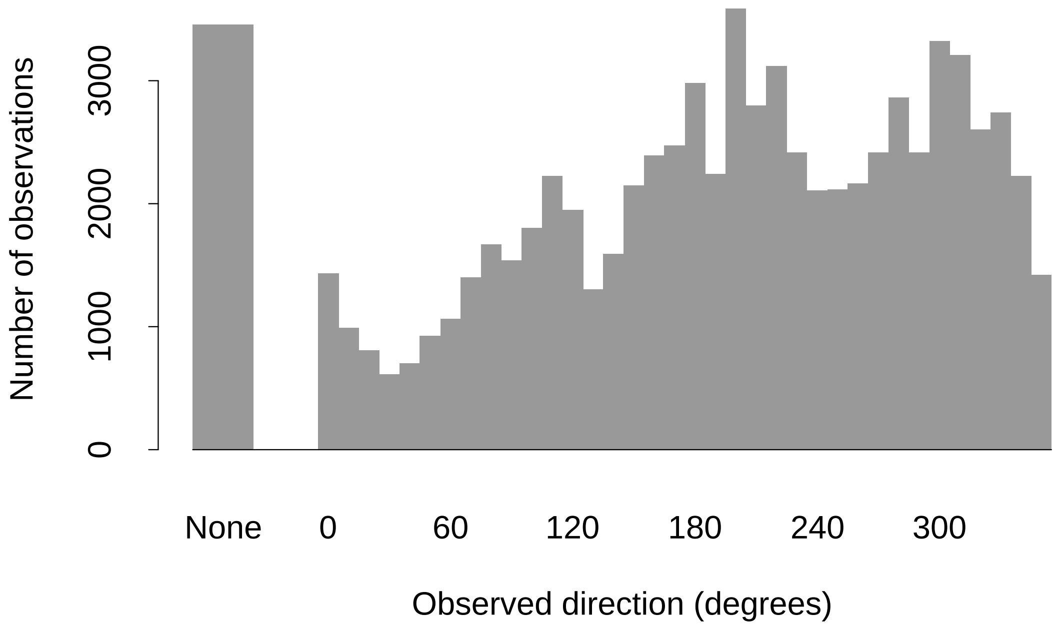

Figure 4.2: A barplot of the frequencies of directions for the raw hourly data

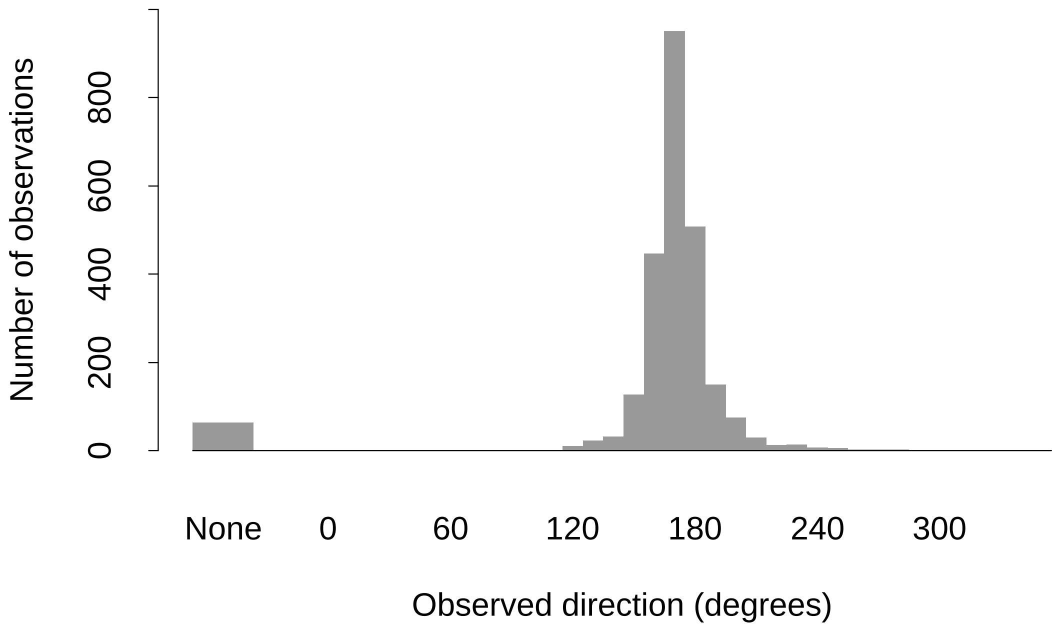

Figure 4.3: A barplot of the frequencies of directions for the raw hourly data given that the previous observation was at 180°

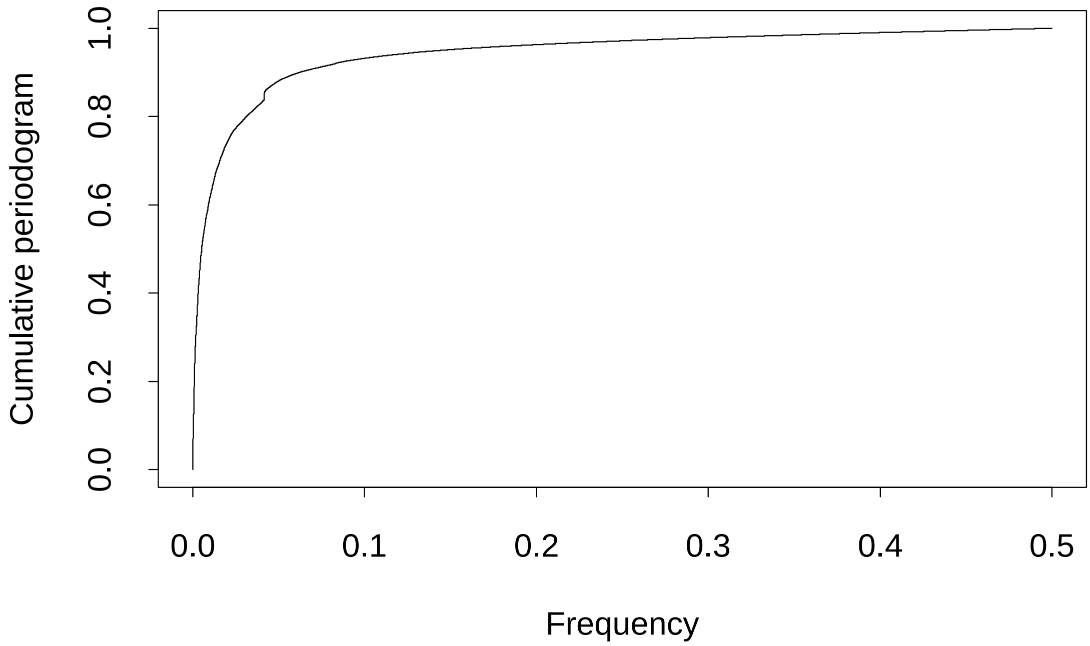

Figure 4.4: Cumulative periodogram for the raw hourly data

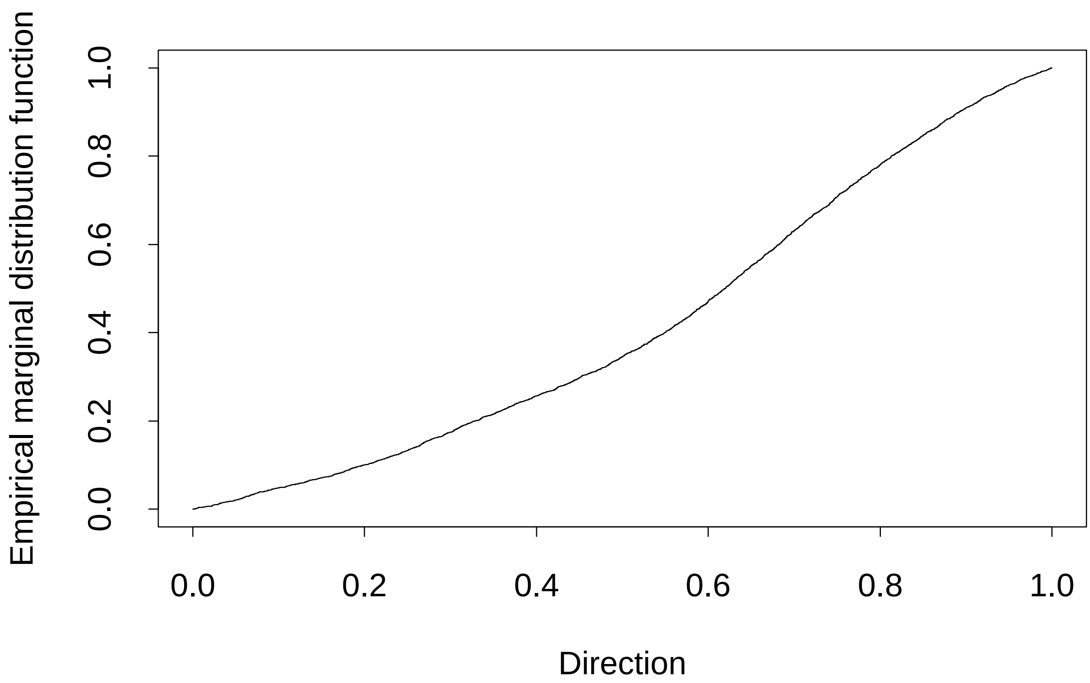

Figure 4.5: The empirical marginal distribution function of the sequence of daily averages

function for the averaged data. The data was averaged using the usual average for circular data (see [22]),

$$Y_t = \text{the direction of } \sum_{j=0}^{23} e^{2\pi i X_{24t-j}}$$

where X_t is the raw data and Y_t is the average, giving a sequence of 3218 observations of the daily average wind direction. There are very few missing values since a 24 hour calm is rare. Those few have been replaced by the average for the previous 24 hours. It is this sequence which we shall attempt to model and which shall be known henceforth as the "data", the term "raw data" being used for the original series.

4.3 Modelling the Daily Average Directions

In figure 4.6 we see the estimated marginal density of the data. This is not wrapped normal in appearance, since it is not symmetric. This would appear to preclude the use of the wrapped linear models discussed in chapters 1 and 2. However for the sake of comparison with later models, the wrapped AR(1) model was fitted to the data, having first transformed the series by rotation to have circular average 0. The result of maximum likelihood estimation was a log-likelihood of 1242 at $\sigma = 0.168$ and $\phi = 0.9975$, when the maximisation stopped due to the difficulty of calculating the likelihood for ϕ near 1. The path through parameter space being followed was tending to ϕbeing 1, while σ was changing very little. The increases in the likelihood for the last few iterations were very small and it is unlikely that finding the true maximum would result in an increase of more than 1.

Further examination of the data was made after transforming the series to have the uniform distribution as its marginal distribution by setting

$$Z_t = F(Y_t)$$

where F is the empirical marginal distribution function of the sequence of averages. There are two reasons for doing this. Firstly, we can confine our attention to the dependence which exists in the series and ignore the difficulties involved in modelling the marginal behaviour. Secondly it makes the interpretation of graphical displays such as lag scatterplots less difficult. This is because, if we now plot Z_{t+1} against Z_t,

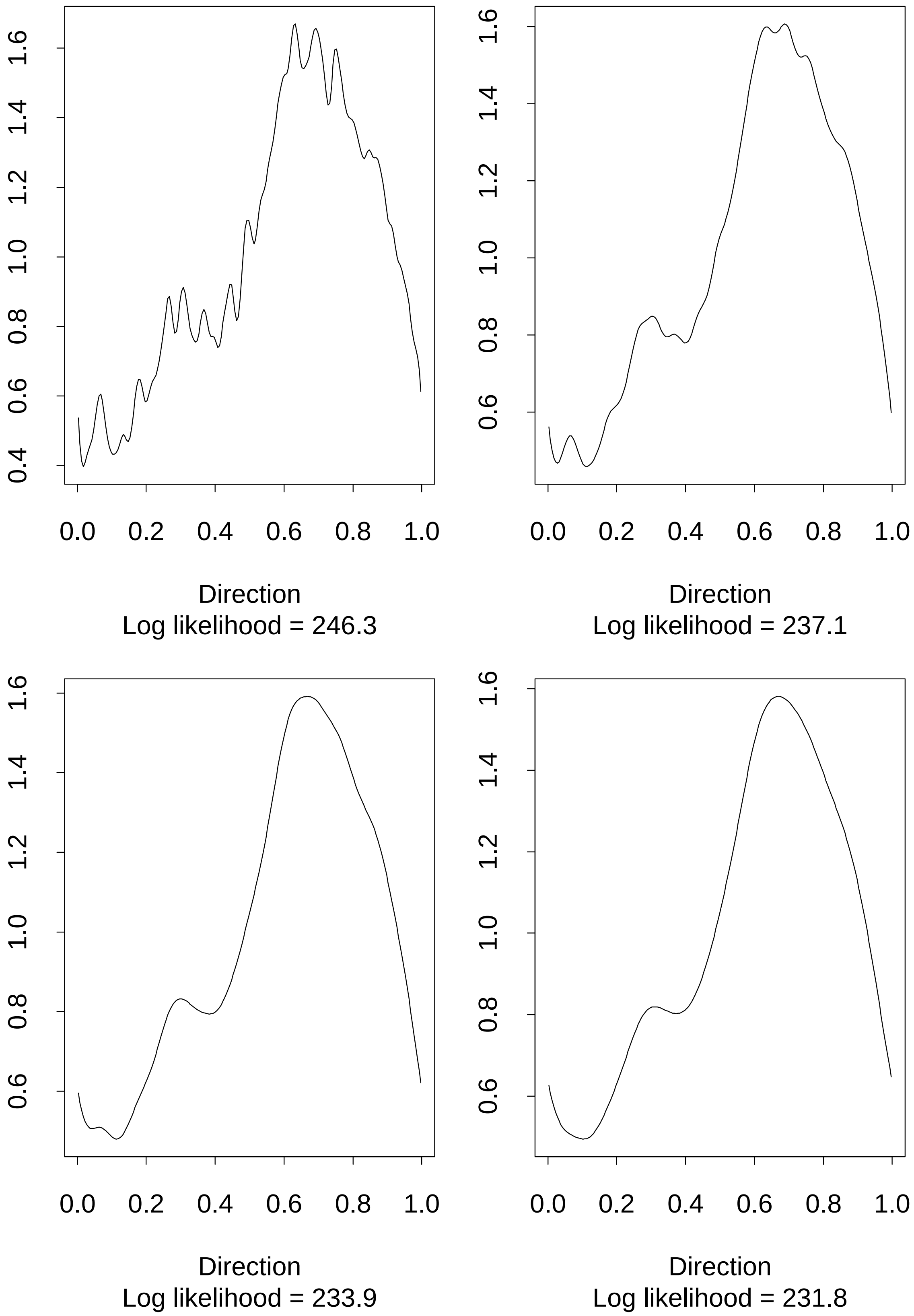

Figure 4.6: Estimates of the marginal density of the daily average sequence. The log-likelihood values are those obtained for the sequence of averages from these densities under independence

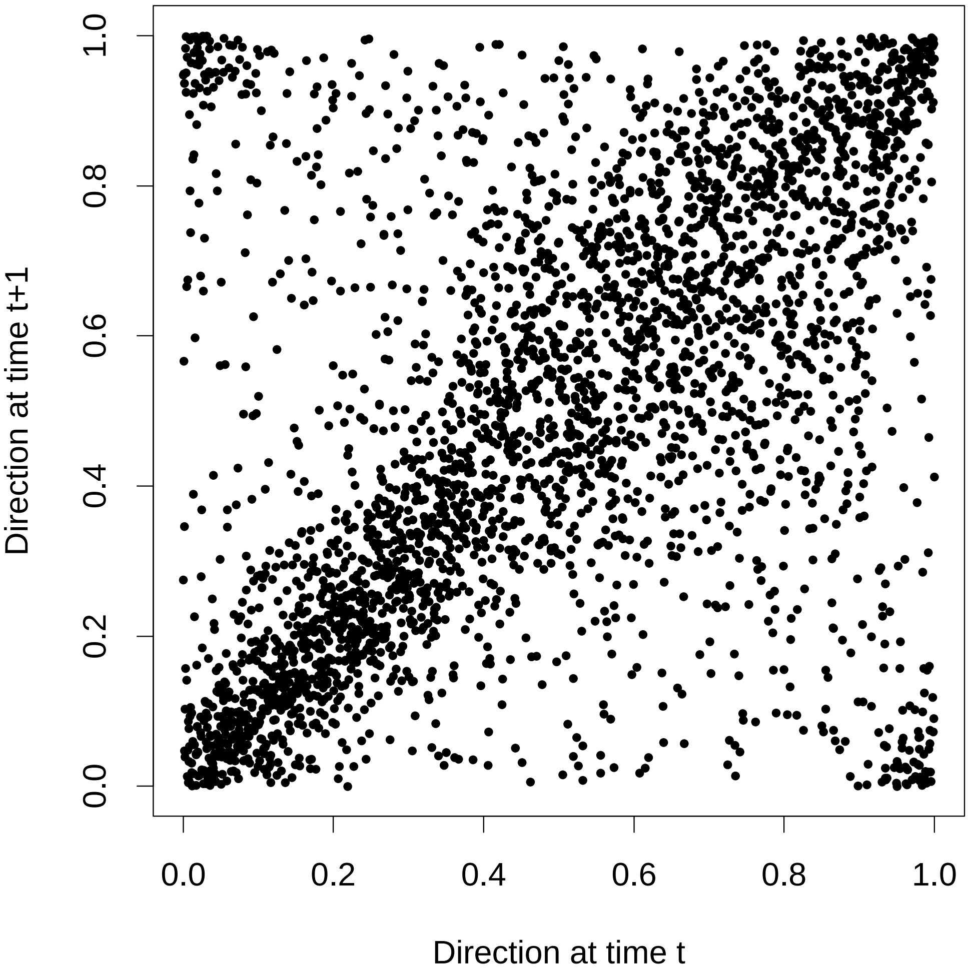

Figure 4.7: Lag 1 scatterplot of the *cdfs* sequence

the density of observations on each axis is constant which is advantageous, for we can now place the same reliability on conclusions drawn about the behaviour of Z_{t+1} for each different value of Z_t. This new series will be known as the *cdfs* for the remainder of the chapter.

We shall now proceed to examine the first order dependence of the *cdfs*. Figure 4.7 shows the lag 1 scatterplot of the *cdfs*. The next three subsections describe and estimate models based on the most prominent features of this plot. All the models from now on will be semi-parametric models for the sequence of daily averages. They will be parametric models for the *cdfs* and the non-parametric component is the transformation between Y_t and Z_t. The problem of comparing these models with parametric models for the daily averages arises naturally in the next section and will be considered there.

4.3.1 Random Walk Model

The most striking aspect of figure 4.7 is the high density of points along the main diagonal (the small clusters of points in the NW and SE corners are artifacts of the fact that the edges of the graph should be identified, i.e. they are spill-over from the main diagonal). This would appear to imply some sort of random walk model based on some unimodal distribution such as those of wrapped normal type. That is, the conditional distribution of Z_{t+1} given Z_t is wrapped normal with mean Z_t and some fixed variance independent of Z_t.

$$Z_{t+1}|Z_t \sim \mathrm{WN}(Z_t, \sigma^2)$$

The result of fitting this model is a loglikelihood value of 933 at $\sigma = 0.187$.

It should be noted that this value for the log-likelihood cannot be directly compared to that obtained for the wrapped AR(1) previously, since the two values arise from different sets of data. However, since the Z_t are the image of the Y_t under the bijective transformation $Z_t = F(Y_t)$, any density on the Z_t induces a density on the Y_t by

$$f_{Y_t|\mathbf{Y}_1^{t-1}}(y_t|\mathbf{y}_1^{t-1}) = F'(y_t) f_{Z_t|\mathbf{Z}_1^{t-1}}(F(y_t)|F(y_{t-1}), \ldots, F(y_1))$$

and so the likelihood of the Y_t series is given by

$$f_{\mathbf{Y}_1^n}(\mathbf{y}_1^n) = f_{\mathbf{Z}_1^n}(\mathbf{z}_1^n) \prod_{j=1}^{n} F'(y_j)$$

The product term is the likelihood of the averages under the assumption of independence, i.e. the difference between the log-likelihoods for the averages and the *cdfs* will be the log-likelihood function of the averages under independence. The remaining difficulty is that F is the *empirical* cumulative distribution function and F' corresponds to the *estimated* marginal density, the estimation of which is governed by subjectivity. Figure 4.6 shows four different estimates of the marginal density with the corresponding values of $\log \prod_{j=1}^{3218} F'(y_j)$ underneath. Using the second one we find that the log likelihood of the current model for the Y_j is $1165 = 933 + 237$ which is less than that obtained for the wrapped AR(1). The subjectivity of the density estimates is, in fact, unimportant, since the models considered later perform better than the wrapped

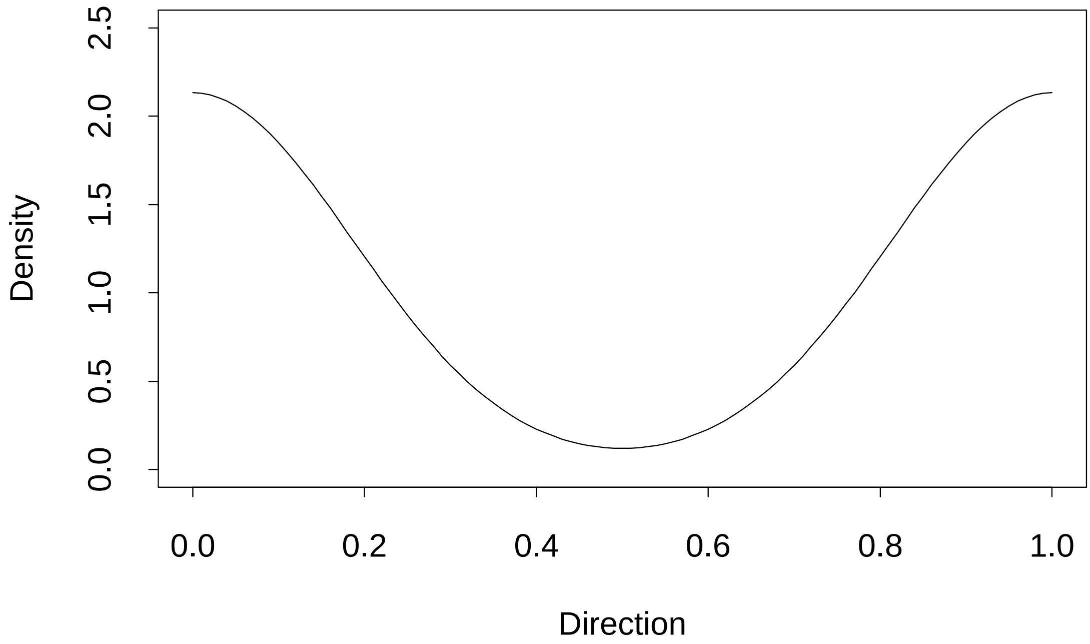

Figure 4.8: Density function of the wrapped normal density with mean 0 and $\sigma = 0.187$

AR(1), no matter which of these density estimates is chosen. It might seem surprising that this likelihood is different from that obtained for the AR(1), since the parameter values obtained for the latter correspond to a random walk having a wrapped normal for the conditional distribution. The difference lies in the fact that the two models are fitted to different series. The AR(1) model is a random walk for the Y_t series but not for the Z_t series. The model of this section is a random walk for the Z_t, but not for the Y_t.

4.3.2 Uniformly contaminated Random Walk Model

A closer examination of figure 4.7 shows that more points are located on or near to the main diagonal than would be expected for the value of σ obtained for the previous model. Further, at larger distances from the diagonal the points appear to be scattered more or less uniformly which is certainly not what should occur for a wrapped normal density. The wrapped normal density for $\sigma = 0.187$ is shown in figure 4.8. It is clearly seen that there is no area of roughly constant density. This suggests the use of a different density for the random walk model. By contaminating the wrapped normal

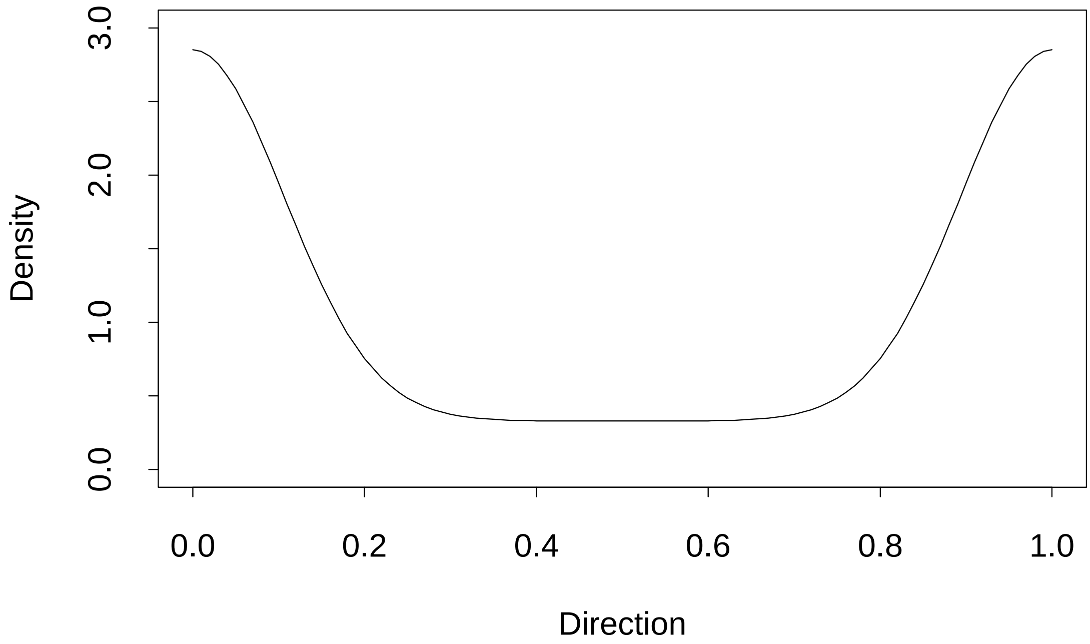

Figure 4.9: Density function of the uniformly contaminated wrapped normal distribution with mean 0, $\sigma = 0.106$ and contamination level $p = 0.33$

with an amount of uniform distribution a density is obtained which corresponds to the requirements described for small values of σ. The model is then given by

$$f_{Z_t|Z_{t-1}}(z_t|z_{t-1}) = p.1 + (1-p).g_\sigma(z_t - z_{t-1})$$

where g_σ is the density of a wrapped normal having mean 0 and variance σ^2. p is the proportion of contamination by the uniform distribution.

Fitting this model by maximum likelihood yields a log-likelihood of 1076 at $\sigma = 0.106$ and $p = 0.330$. Figure 4.9 shows the conditional density for these values of the parameters. The model now incorporates the basic features of figure 4.7. There are, however, still significant divergences.

4.3.3 Pseudo-Regression Model

Figure 4.10 is the same scatterplot as in figure 4.7, but with the addition of boundary lines containing the area where the highest density of points is to be found. This is purely for visual convenience and was not derived by any analytic procedure. It does, however, suggest an improvement to the model. The outline seems to show that

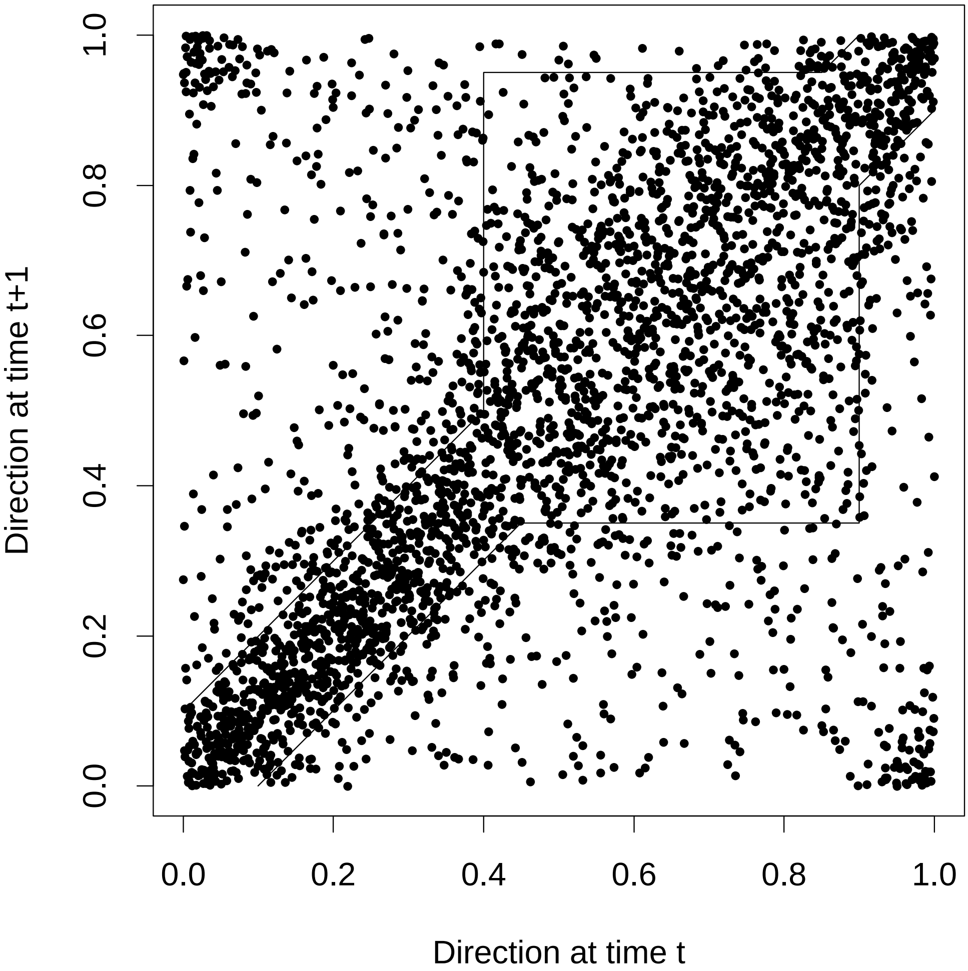

Figure 4.10: Lag 1 scatterplot of the *cdfs* sequence

when $Z_{t-1} < 0.4$ or $Z_{t-1} > 0.9$ the random walk model describes the behaviour of Z_t accurately enough. But when $0.4 < Z_{t-1} < 0.9$ there are two changes. Firstly the mean of Z_t appears to be pulled towards 0.65 or thereabouts. Also the conditional dispersion of Z_t appears to be greater. The uniform behaviour is about the same far from the diagonal in either case. This suggests the following model which incorporates a regression-like approach.

$$f_{Z_t|Z_{t-1}}(z_t|z_{t-1}) = \begin{cases} p.1 + (1-p).g_{\sigma_1}(z_t - z_{t-1}) \\ \qquad\qquad \text{when } 0 \leq z_{t-1} < a \text{ or } b < z_{t-1} < 1 \\ p.1 + (1-p).g_{\sigma_2}(z_t - \{\alpha + k(z_{t-1} - \alpha)\}) \\ \qquad\qquad \text{when } a \leq z_{t-1} \leq b \end{cases}$$

where g_σ is the density function of the wrapped normal having mean 0 and variance σ^2. This model has 7 parameters which is many more than for the previous models. p is the proportion of uniform contamination. σ_1 and σ_2 are variances. α and k are mean and coefficient of the regression and a and b are the end-points of the interval in which the regression behaviour occurs. Estimation of this model is extremely difficult because the likelihood function is not continuous in the parameters a and b, since the likelihood will jump as a point moves from one region to the other. For this reason it is neccesary to introduce an extra parameter δ as follows. Define

$$f_1(x|y) = p + (1-p)g_{\sigma_1}(x - y)$$

and

$$f_2(x|y) = p + (1-p)g_{\sigma_2}(x - \{\alpha + k(y - \alpha)\})$$

Then the model as so far proposed so far is

$$f_{Z_t|Z_{t-1}}(z_t|z_{t-1}) = \begin{cases} f_1(z_t|z_{t-1}) & \text{when } 0 \leq z_{t-1} < a \text{ or } b < z_{t-1} < 1 \\ f_2(z_t|z_{t-1}) & \text{when } a \leq z_{t-1} \leq b \end{cases}$$

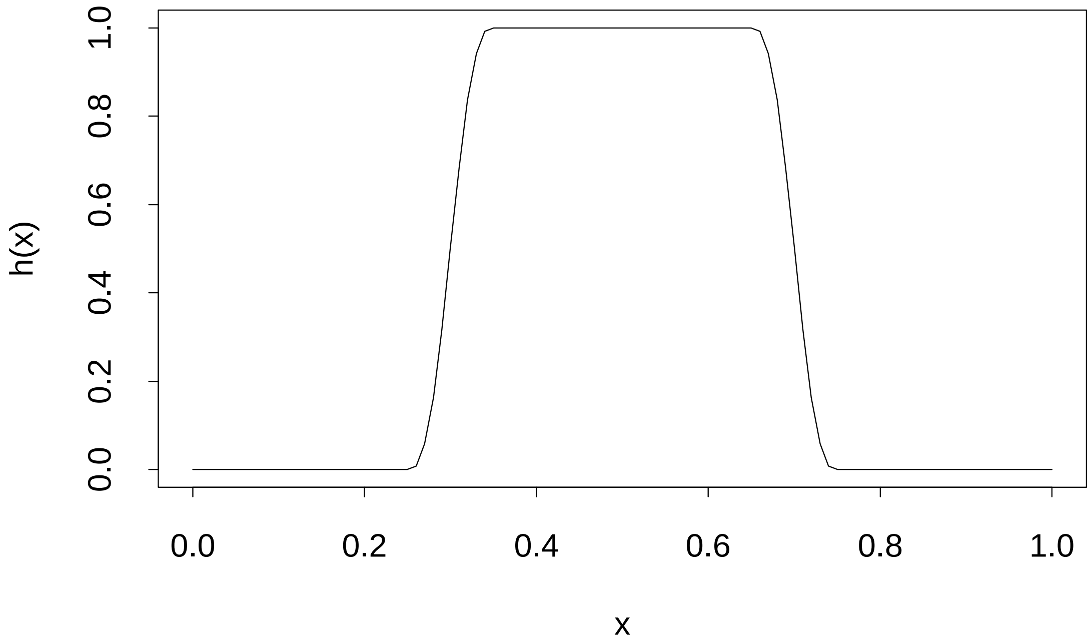

Figure 4.11: Graph of the smoothing function $h(x)$ for $a = 0.25$, $b = 0.75$ and $\delta = 0.1$.

Let $\psi(x) = 6x^5 - 15x^4 + 10x^3$. Define $h(x)$ by

$$h(x) = \begin{cases} 0 & 0 \leq x \leq a \\ \psi\left(\frac{x-a}{\delta}\right) & a < x \leq a+\delta \\ 1 & a+\delta < x \leq b-\delta \\ \psi\left(\frac{b-x}{\delta}\right) & b-\delta < x \leq b \\ 0 & b < x < 1 \end{cases}$$

$h(x)$ is shown in figure 4.11 for $a = 0.25$, $b = 0.75$ and $\delta = 0.1$. h has the important property (for gradient maximisation routines) of continuous first and second derivatives. We now finally define the new model by

$$f_{Z_t|Z_{t-1}}(z_t|z_{t-1}) = h(z_{t-1})f_2(z_t|z_{t-1}) + (1 - h(z_{t-1}))f_1(z_t|z_{t-1})$$

This is the same as before except when $z_{t-1} \in (a, a+\delta)$ or $z_{t-1} \in (b-\delta, b)$, and deforms continuously between f_1 and f_2 in these small intervals. It is worth noticing that while the likelihood function is now smooth, there may be local maxima as a point moves from $a + \delta$ to a.

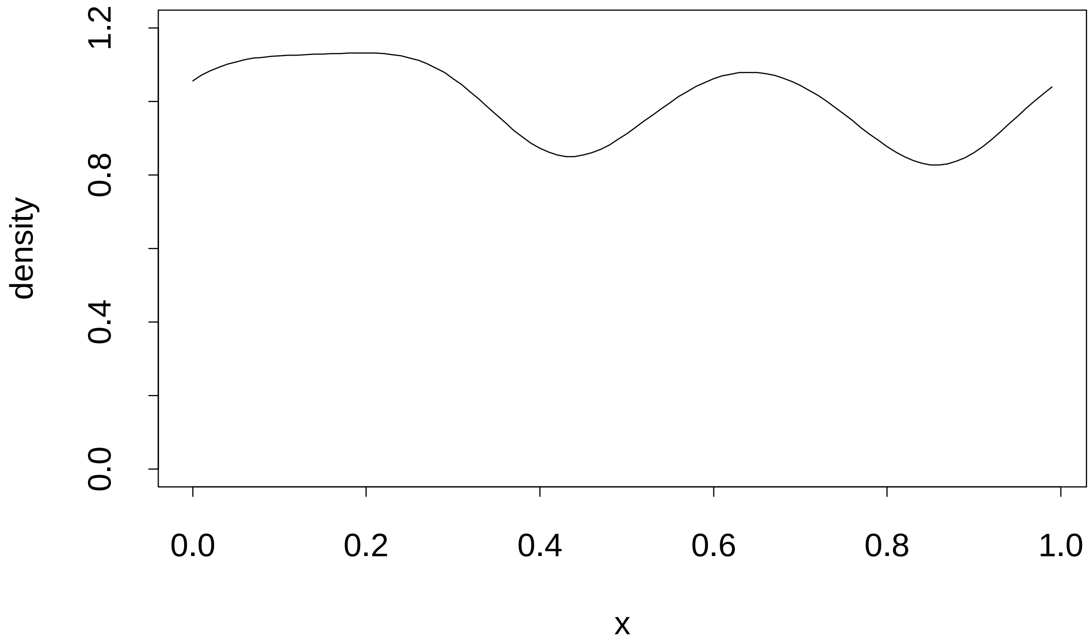

Figure 4.12: The stationary marginal density of the pseudo-regression conditional density at the estimated parameter values

This model was fitted to the data by maximum likelihood. δ was not allowed to vary, but was fixed at 0.01, for δ is not a true parameter of the model, but a contrivance to ease the estimation process. The value of 0.01 seems reasonable. It is not so large as to severely distort the behaviour of the basic model, yet large enough to ensure there being 30 or so points in the range $(a, a + \delta)$. The more points that are in this interval, the smoother is the likelihood function and hence there is less chance of multiple local maxima. The result of the maximisation was a log-likelihood of 1256 at

$$p = 0.253 \quad a = 0.362 \quad b = 0.916 \quad \sigma_1 = 0.076$$

$$\sigma_2 = 0.158 \quad k = 0.645 \quad \alpha = 0.634$$

One problem, not yet mentioned, with this model is that it does not have as its stationary marginal distribution the uniform distribution. Figure 4.12 shows the marginal density for the estimated values of the parameters. However the model does fit significantly better than the previous ones.

An interesting question is the one of physical motivation for this model. The answer may lie in the prevailing meteorological pattern over Ireland. Weather in Ireland is dominated by a regular sequence of pressure systems which sweep in over the west coast of Ireland from the Atlantic. This pattern should be reflected in the sequence of wind directions. As a pressure system passes over the country, there should be some predictable change in the wind direction, because the wind is blowing clockwise or anti-clockwise around the centre of the pressure system. I have not been able to explain the precise form of the pseudo-regression model, but I believe the regression part arises to some extent from the pressure system pattern.

4.4 Diagnostics

We now turn to the issue of diagnostics for the models so far considered. We shall use the tool described at the outset of the paper, the conditional cumulative distribution function. As shown there, if $f_{Z_t|\mathbf{Z}_1^{t-1}}$ is a model for a circular time series, the sequence $F_{Z_t|\mathbf{Z}_1^{t-1}}(z_t|\mathbf{z}_1^{t-1})$ should be a sequence of i.i.d. uniformly distributed circular values if the model is the correct one for the sequence $z_1, \ldots, z_n$. Figure 4.13 shows the estimated marginal density and the cumulative complex periodogram of this sequence for each of the three models already fitted. The marginal behaviour is not as good as might be desired, though this is a purely subjective judgement. The periodograms are more difficult to interpret, comparison being made difficult by their closeness to being straight lines. Figure 4.14 shows for each of these models the difference between the cumulative periodogram and a straight line. Most of the improvement is contained in the addition of uniform contamination. However the pseudo-regression model is an improvement, though by a much smaller amount.

However, there are still clear signs of dependence in the periodogram for the pseudo-regression model. This suggests the possibility either of higher order short-term dependence or, perhaps, of the need to fit a seasonal model.

4.5 Higher Order Markov Models

In this section I shall use the linear conditional probability approach described in section 3.4 for developing Markov models with more than lag 1 dependence from first

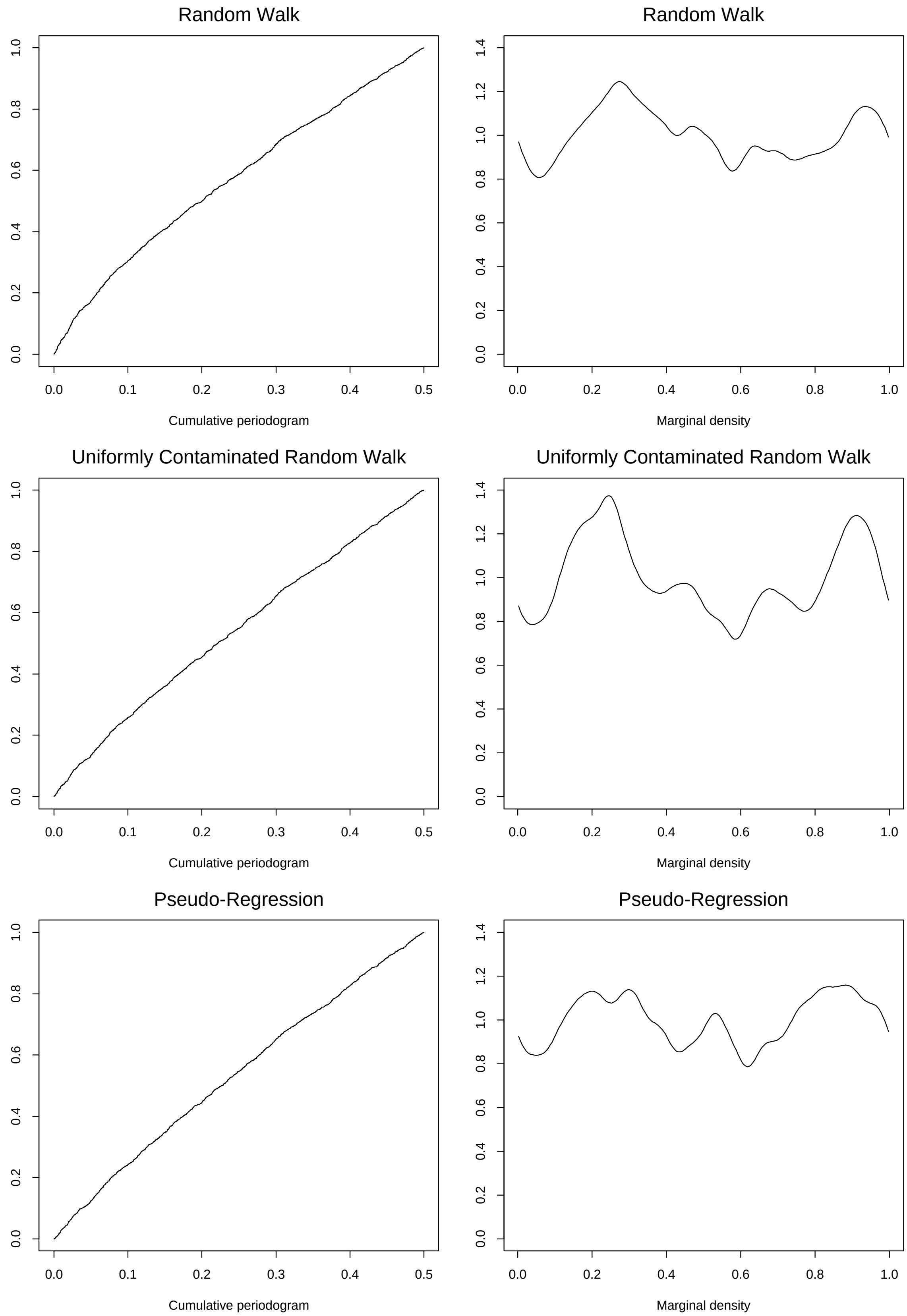

Figure 4.13: Cumulative periodograms and densities of the conditional distribution sequences for several models

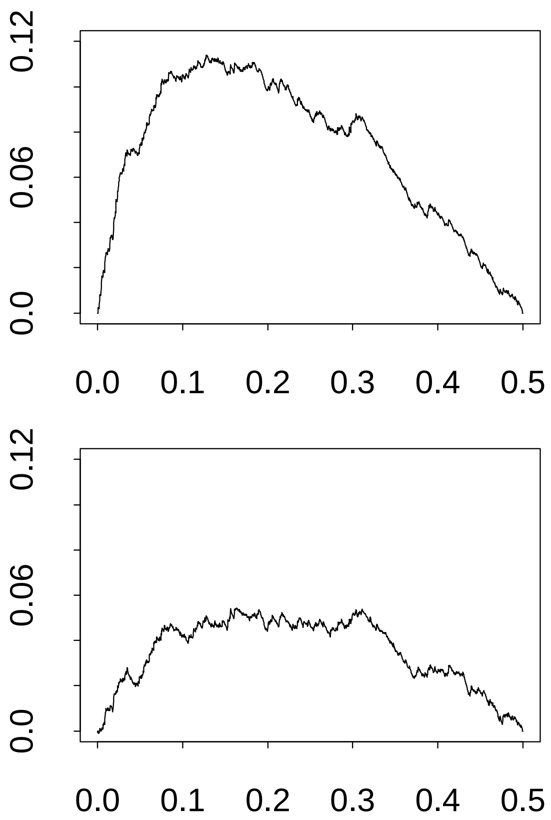

Figure 4.14: Difference between the cumulative periodogram and a straight line for the conditional cumulative distribution sequence from the random walk, contaminated random walk and the pseudo-regression model respectively

order Markov models. That is

$$f_{Z_t|\mathbf{Z}_{t-l}^{t-1}}(z_t|\mathbf{z}_{t-l}^{t-1}) = \sum_{j=1}^{l} \lambda_j g(z_t|z_{t-j}) \tag{4.2}$$

where g is som transition density.

We shall use the conditional densities of the pseudo-regression type of section 4.3.3 in conjunction with this definition as a family of models for the *cdfs*. The parameters l and λ_j, $j = 1, \ldots, l$ are as defined by equation 4.2 The results of fitting this model for a number of values of l were as follows. (λ_l can be obtained by subtraction from 1)

$l = 2$: $p = 0.236$ $a = 0.362$ $b = 0.929$ $\sigma_1 = 0.074$

$\sigma_2 = 0.155$ $k = 0.663$ $\alpha = 0.640$ $\lambda_1 = 0.932$

log-likelihood = 1264.

$l = 3$: $p = 0.213$ $a = 0.362$ $b = 0.930$ $\sigma_1 = 0.075$

$\sigma_2 = 0.153$ $k = 0.668$ $\alpha = 0.641$ $\lambda_1 = 0.894$

$\lambda_2 = 0.033$

log-likelihood = 1274.

$l = 4$: $p = 0.203$ $a = 0.362$ $b = 0.930$ $\sigma_1 = 0.074$

$\sigma_2 = 0.152$ $k = 0.671$ $\alpha = 0.640$ $\lambda_1 = 0.879$

$\lambda_2 = 0.031$ $\lambda_3 = 0.057$

log-likelihood = 1276.

These are a nested family of models and since the increase in the log-likelihood from $l = 3$ to $l = 4$ is not significantly large, fitting ceased at this point.

In case the dependence could be more accurately captured by allowing the dependence on time $t - 2$ to be different from that on time $t - 1$, the following model was fitted.

$$f_{Z_t|Z_{t-1},Z_{t-2}}(z_t|z_{t-1}, z_{t-2}) = \lambda f^{(1)}(z_t|z_{t-1}) + (1-\lambda) f^{(2)}(z_t|z_{t-2}) \quad \lambda > 0$$

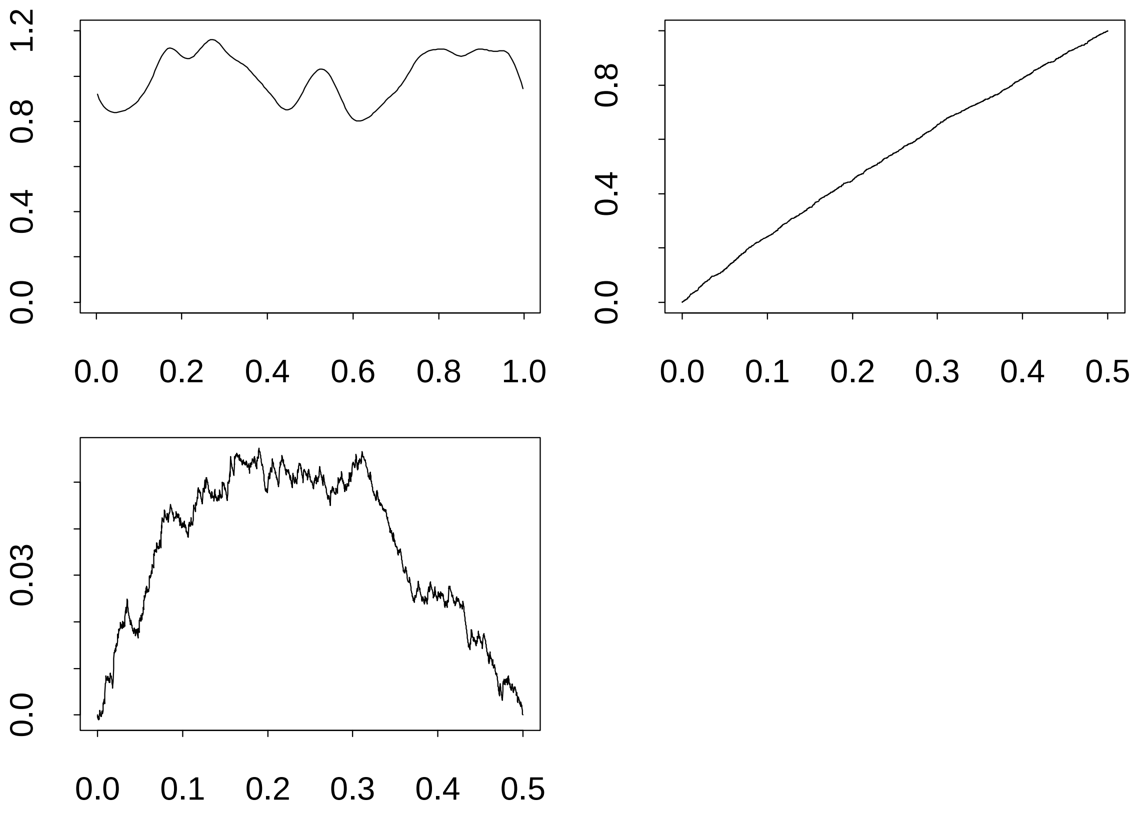

Figure 4.15: Density, cumulative periodogram and difference between the latter and a straight line for the conditional distribution sequence from the order 3 mixture model

where $f^{(1)}$ and $f^{(2)}$ are both conditional densities of the pseudo-regression type, but with different values of the parameters. The result of the fit was as follows.

$p^{(1)} = 0.024 \quad a^{(1)} = 0.372 \quad b^{(1)} = 0.929 \quad \sigma_1^{(1)} = 0.074$

$\sigma_2^{(1)} = 0.162 \quad k^{(1)} = 0.683 \quad \alpha^{(1)} = 0.632$

$p^{(2)} = 0.037 \quad a^{(2)} = 0.291 \quad b^{(2)} = 0.897 \quad \sigma_1^{(2)} = 700\ (\infty)$

$\sigma_2^{(2)} = 0.269 \quad k^{(2)} = 0.438 \quad \alpha^{(2)} = 0.685$

$\lambda = 0.745$

log-likelihood $= 1268$

This is not a significant increase in the log-likelihood over the original pseudo-regression model.

The usual diagnostic method was applied to the above mixture model with $l = 3$. Figure 4.15 shows the estimated marginal density, the cumulative periodogram and the difference between the latter and a straight line for the conditional cumulative distribution sequence. There is no noticeable improvement in the shape of the periodogram over that obtained previously. There are two likely reasons for this. Firstly, the nature of any higher than first order dependence may be too complicated to be captured by this mixture approach. Secondly there may be some seasonal behaviour which will be discussed in the next section.

4.6 A Seasonal Model

What evidence, if any, is there that a seasonal model is needed for this data ? There is, of course, the intuitive feeling that all aspects of the weather vary according to the time of year. More scientifically, there is the "land breeze" and "sea breeze" effect which must be affected by temperature changes and these certainly exhibit annual variation. It would be preferable if we could observe this in a quantitative way.

Classical time series analysis incorporates seasonality as an additive or multiplicative term in the model and, in the case of ARMA models, by a lag 365 coefficient in a linear model. For the circle we do not have available both algebraic operations. There is only one operation, which can be viewed as either addition or multiplication depending on the context. A simple change in the mean direction is an inadequate form of seasonality for this data. Figure 4.16 shows the smoothed periodogram of a sequence of moving variances derived from the sequence of daily averages. While this does not have its peak at an exact frequency corresponding to a period of 1 year or a precise number of months we should not find this surprising since smoothing introduces bias into periodogram estimates which makes precise interpretation of this kind impossible. What is clear is that there is a high concentration of energy at very low frequencies corresponding to periods greater than about two months. Almost all high frequency variation has disappeared due to the taking of moving variances which automatically damps the high frequency changes in the original sequence.

To facilitate understanding any seasonal behaviour, the second-order pseudo-regression model was fitted separately to each month's data. Figure 4.17 shows the smoothed (using the smoother 4(3RSR)2H twice to remove outliers) trajectories of each of the

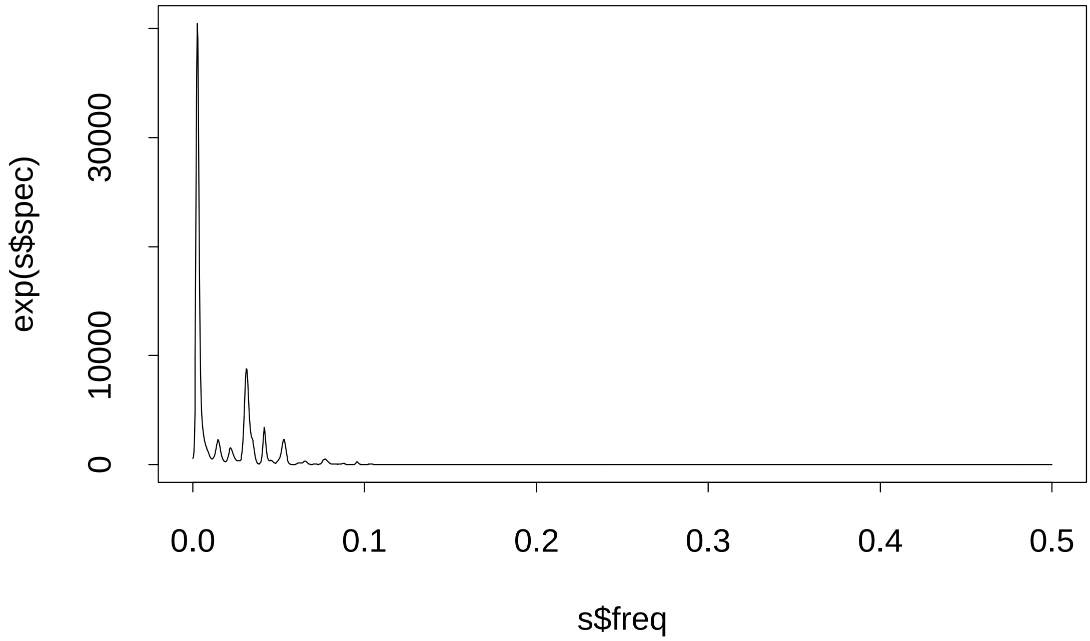

Figure 4.16: Smoothed periodogram for a sequence of moving variances of the daily average data

parameters and of the log-likelihood. The log-likelihood is clearly seasonal as may be the uniform contamination parameter p. That p is the most clearly seasonal parameter is not surprising since its major effect is to control the variance of the process.

I now extend the current model — the third-order pseudo-regression model — to incorporate this fact as follows. The parameter p is made to depend on time in a sinusoidal fashion, according to the formula

$$p = p_M + p_A . \sin(2\pi(t - p_P)/365.25)$$

Here p_M is the mean value of p, p_A the amplitude of its variation and p_P is a phase parameter. Estimating this model by maximum likelihood gives the following parameter values.

$p_M = 0.217 \quad p_A = 0.112 \quad p_P = 77.1 \quad a = 0.372$

$b = 0.948 \quad \sigma_1 = 0.074 \quad \sigma_2 = 0.151 \quad k = 0.696$

$\alpha = 0.651 \quad \lambda_1 = 0.893 \quad \lambda_2 = 0.027$

log-likelihood $= 1281$

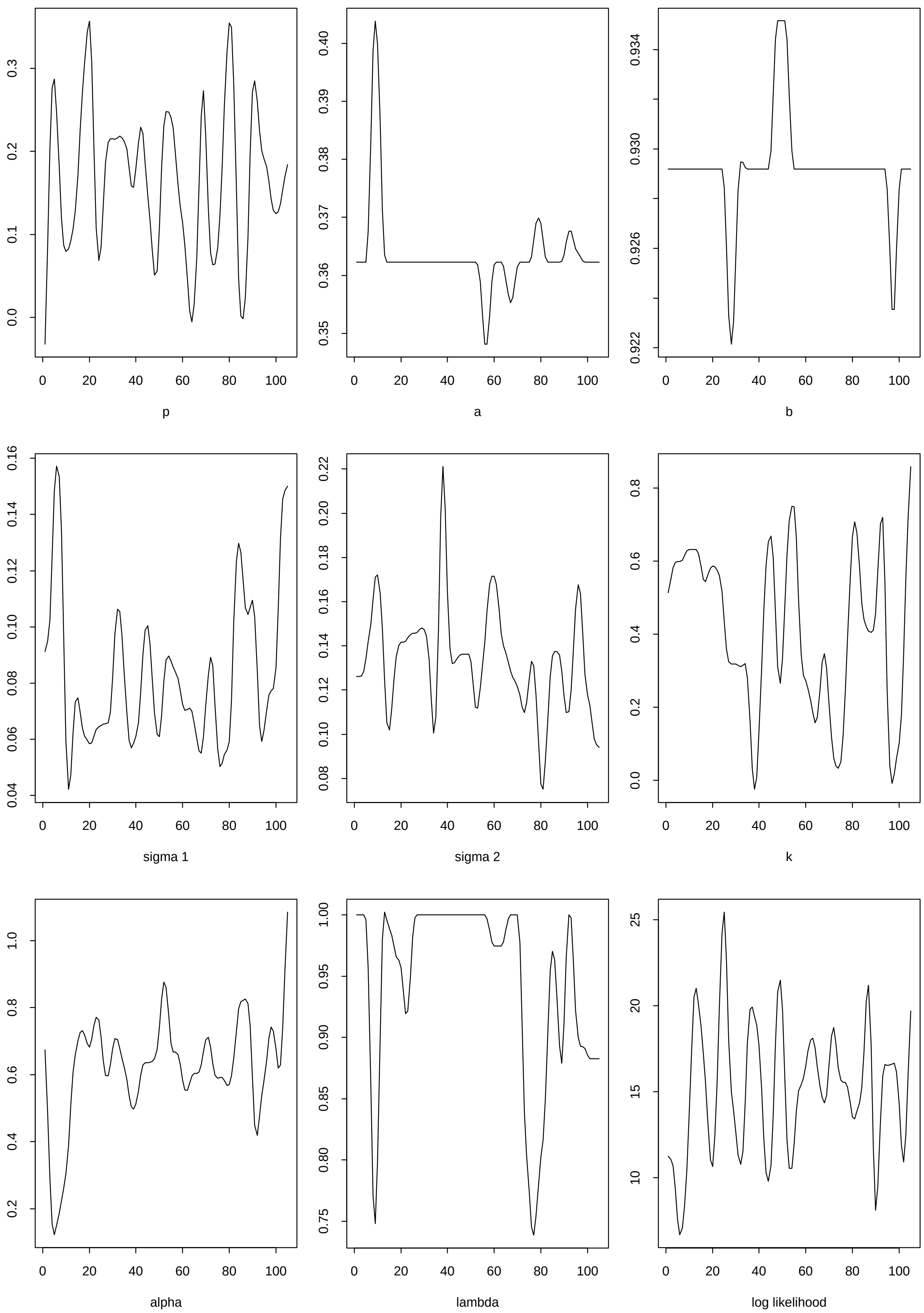

Figure 4.17: Trajectories of the model parameters. The x-axis units are years.

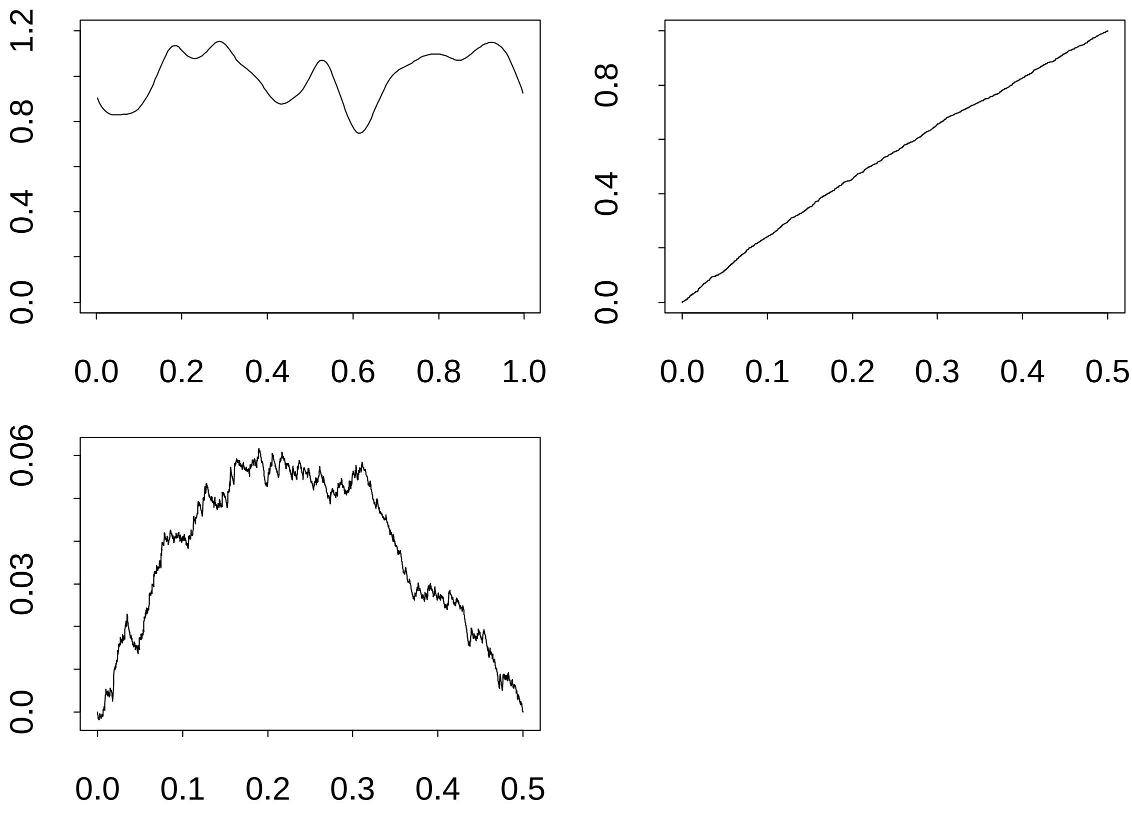

Figure 4.18: Density, cumulative periodogram and difference between the latter and a straight line for the conditional distribution sequence from the seasonal model

The usual diagnostic is used. Figure 4.18 shows the estimated marginal density, the cumulative periodogram and the difference between the latter and a straight line for the conditional distribution function sequence. It is not possible to ascertain visually whether the periodogram is an improvement on those obtained previously, but figure 4.19 shows the difference between the periodogram in figure 4.18 and that in figure 4.15. That for the seasonal model lies below the other at low frequencies and above it for high frequencies. This makes sense since the seasonal model should improve the low frequency fit. The differences at higher frequencies are more difficult to explain. It is possible that in fact the high frequency behaviour of the series does not vary seasonally and so the variation in the model parameters is worsening the high frequency fit. Whether this is true or not, the improvement in the cumulative periodogram is small. To illustrate the kind of change required, I have used simulation, generating 8 sequences from the seasonal model, fitting the model to those sequences and examining the diagnostic sequence. Figure 4.20 shows the marginal distribution of the diagnostic sequence for each simulation, and figure 4.21 shows the difference be-

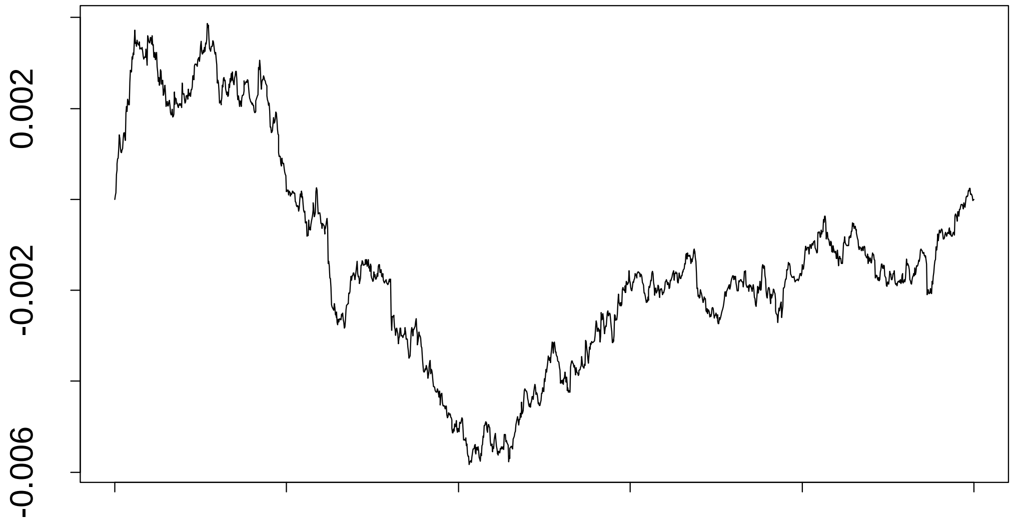

Figure 4.19: Graph of the difference between the cumulative periodograms for the order 3 pseudo-regression model and the seasonal model

tween the cumulative periodogram and a straight line for each simulation. It is obvious that the behaviour exhibited is quite different from that in figure 4.18. The modelling process has gone no further than this. Clearly more attention needs to be given to both higher order dependence and to seasonality. There may also be an issue of long-term variation in behaviour as discussed in [13] for wind speeds in Ireland.

In conclusion, it is true to say that significant aspects of the behaviour of the wind directions have been modelled. It would be ludicrous to suggest that this provides a method of forecasting, but perhaps the model may throw some light on the sequential behaviour of wind directions. Comparison with other sequences of wind observations would be interesting, for it seems likely that some of the features could be explained by the local geography around the meteorological station where the measurement was performed. There is obviously scope for exploration of more appropriate ways of modelling second and higher order dependence than the mixture method. Residual seasonality would appear to be a difficult problem. Unless some way can be found to define more interesting seasonal models progress seems unlikely. Also, there remains the fact that

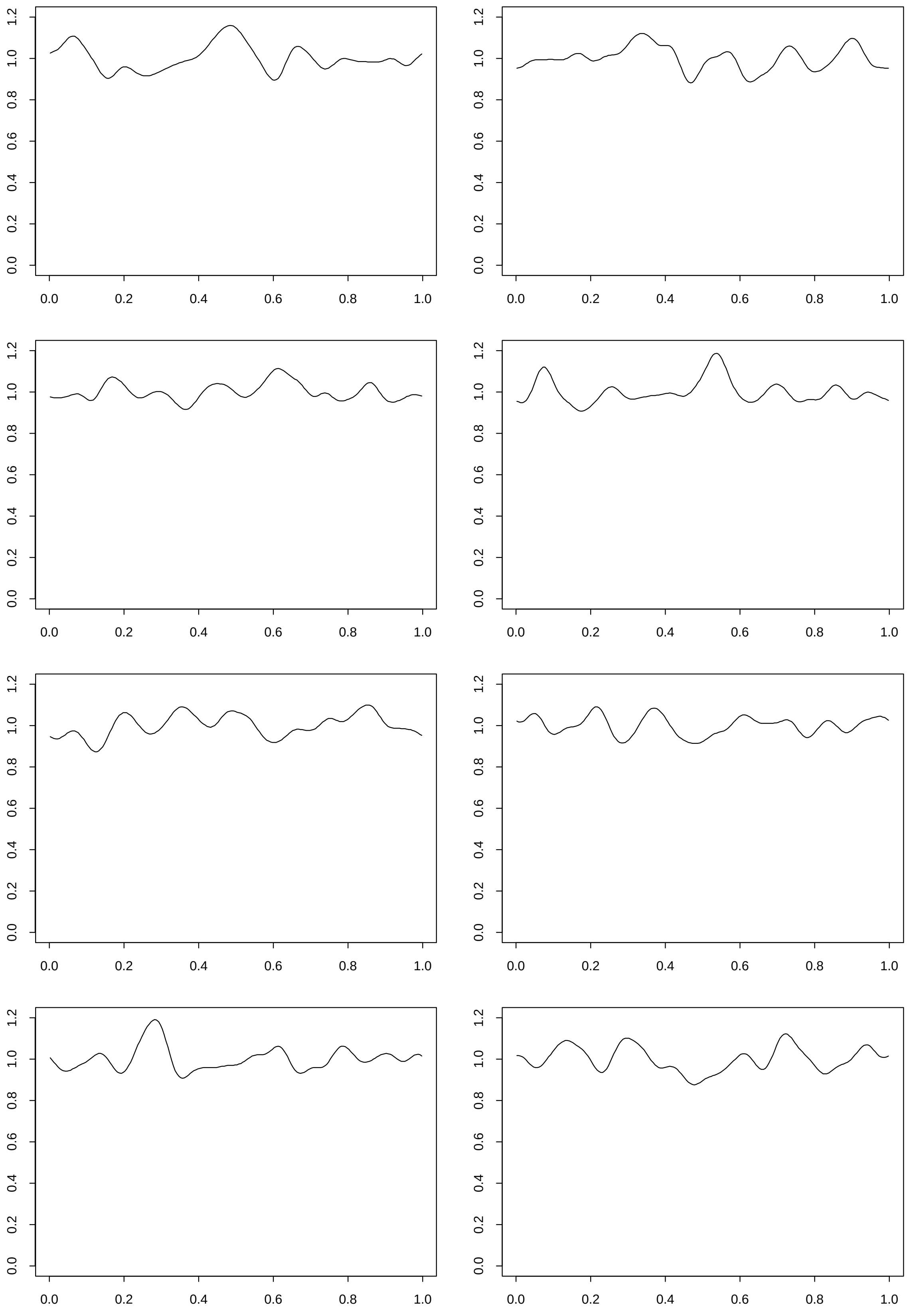

Figure 4.20: Marginal densities for the conditional cumulative distribution sequence obtained from simulations of the seasonal model

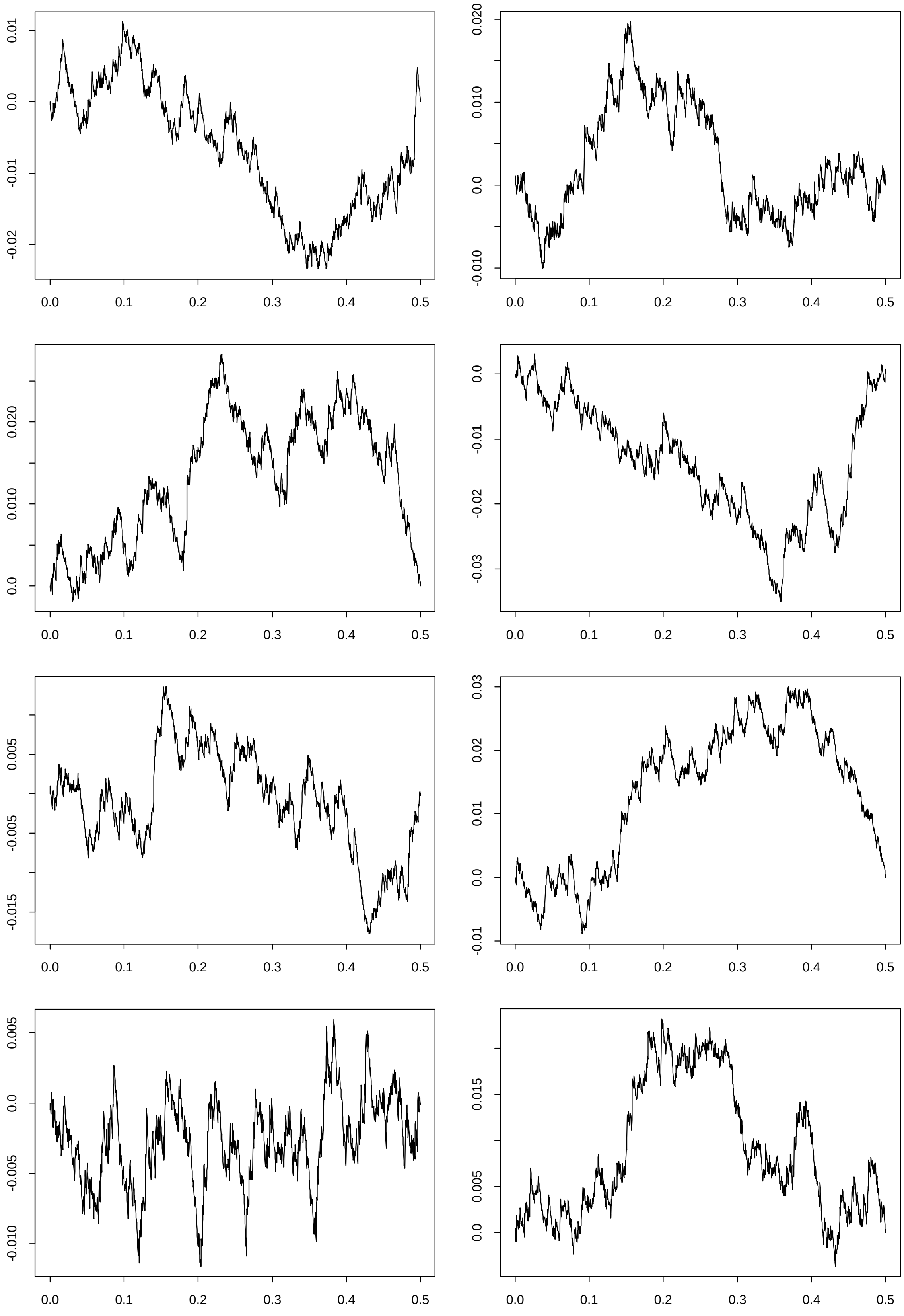

Figure 4.21: Differences between the cumulative periodogram and a straight line for the conditional cumulative distribution sequence obtained from simulations of the seasonal model

best models proposed here do not have the correct marginal distributions. There are clear advantages to modelling a sequence with uniform marginal distribution, subject to availability of a sufficiently large class of conditional distributions having this property. A most interesting avenue of exploration is the properties of the conditional distribution function sequence. While its estimation properties under the null hypothesis of the correct model seem likely to be extremely difficult to calculate, it is potentially so useful (in any multivariate application, not just time series) that they are worthy of some effort.

Appendix A

Borrowed Material

A.1 Mixing Processes

This section brings together a number of results on mixing processes from various sources.

Definition A.1 *A stationary stochastic process* $\{Z_t\}$ *is said to be strongly mixing if*

$$\psi_Z(\tau) = \sup_{A \in \mathcal{F}^0_{-\infty}(Z), B \in \mathcal{F}^\infty_\tau(Z)} \Big| P[A \cap B] - P[A]P[B] \Big| \to 0 \quad as\ \tau \to \infty$$

The $\psi_Z(\tau)$ *are called the (strong) mixing coefficients for* $\{Z_t\}$.

Theorem A.1 *Let* Z_t *be a stationary Gaussian sequence. Then*

$$\psi_Z(\tau) \le \inf_\phi \sup_\lambda \frac{|f_Z(\lambda) - e^{i\lambda\tau}\phi(e^{-i\lambda})|}{f_Z(\lambda)} \le 2\pi\psi_Z(\tau)$$

where the infimum is taken over those functions ϕ *which are analytic in the unit disc.*

Proof: Theorems 1, 2 and 3 of [20].

Theorem A.2 *Let* Z_t *be a strongly mixing stochastic process. Let* X *and* Y *be random variables measurable with respect to* $\mathcal{F}^a_{-\infty}$ *and* $\mathcal{F}^\infty_b$ *resapectively, and that* $E[|X|^p] < \infty$, $E[|Y|^q] < \infty$ *where* $p, q > 1$ *and* $p^{-1} + q^{-1} < 1$. *Then*

$$|C[X,Y]| \le 8E[|X|^p]^{\frac{1}{p}} E[|Y|^q]^{\frac{1}{q}} \psi_Z(b-a)^{1-p^{-1}-q^{-1}}$$

Proof: Corollary A.2 of [12]

Theorem A.3 *Let Z_t be a strongly mixing stationary process with geometrically diminishing mixing coefficients $\psi_Z(\tau)$. Let $f_0(Z_{-\infty}, \ldots, Z_\infty)$ be a function of the Z_t such that*

$$E\left[\left|E[f_0|\mathbf{Z}_{-M}^M] - f_0\right|^2\right]$$

tends geometrically to 0. Then, if f_n is the time-shifted version of f_0

$$\frac{1}{n}\sum_{j=1}^{n} f_j \stackrel{a.s.}{\to} E[f_0]$$

Proof: A trivial corollary of theorem 3.1 of [26]

Theorem A.4 *Let $\{Z_t\}$ be a strongly mixing stationary sequence. Let W be a zero-mean measurable function of $\{Z_t\}$. Let $\{W_t\}$ be the stationary process obtained by time-shifting W. Then, if*

(1) There exists $C > 0$ such that $P[|W| < C] = 1$.

(2) $\sum_{k=1}^{\infty} E[|W - E[W|\mathcal{F}_{-k}^{k}(Z)]|] < \infty$.

(3) $\sum_{k=1}^{\infty} \psi_Z(k) < \infty$,

there exists $\tilde{\sigma}^2 = E[W_0^2] + 2\sum_{k=1}^{\infty} E[W_0 W_k]$ which is finite and non-negative. Further, if $\tilde{\sigma}^2$ is positive,

$$\frac{\sum_{k=1}^{n} W_t}{\tilde{\sigma}\sqrt{n}} \to^d N(0,1)$$

Proof: Theorem 18.6.3 of [14]

A.2 Markov processes

The following material drawn from [9] provides simple criteria for the existence and uniqueness of stationary distributions for Markov processes.

Definition A.2 *A transition function P on a space X is said to satisfy the Doeblin hypothesis if there exists a finite measure ϕ on X with $\phi(X) > 0$, an integer $n \geq 1$ and an $\epsilon > 0$ such that*

$$P^{(n)}(x, A) \leq 1 - \epsilon \quad \text{if} \quad \phi(A) \leq \epsilon$$

Definition A.3 *Let P be a transition function on a space X which satisfies the Doeblin hypothesis. A set E is said to be an invariant set if*

$$P(x, E) = 1 \quad \textit{for all } x \in E$$

E is said to be a minimal invariant set if $E' \subset E$ and E' invariant implies $\phi(E') = \phi(E)$.

Theorem A.5 *Let P be a transition function on a space X satisfying the Doeblin hypothesis. Then*

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

defines for each x a stationary probability distribution. As a function of x, $q(x, E)$ depends only on the minimal invariant set to which x belongs. Further every stationary probability distribution is a linear combination of the $q(x, E)$ with non-negative coefficients .

Proof: Theorem V.5.7 of [9].

A.3 M.L.E for Dependence

The following is a summary of certain parts of [7]. There is a flaw in that paper, where reference is made to a theorem in [15]. Unfortunately the theorem does not apply to the case being considered. There are no problems with the material in the paper until the middle of section 4 — more precisely until equation (4.11). This equation is stated to be a consequence of a theorem found on page 17 of [15]. That theorem is strong law statement for a sequence of functions of a mixing process X_t under certain conditions. One of the conditions is that the functions concerned involve a finite time range from the mixing sequence, i.e. f_t is a function only of $X_{t-n}, \ldots, X_{t+n}$ where n is independent of t. The sequence of functions to which the theorem is being applied in [7] is the second derivative of the log of the conditional density function — $\partial_\theta^2 \ln f_{X_t|\mathbf{X}_1^t}$ — which, in general, for each time point involves the whole of the past. A considerable amount of the work involved in showing consistency and asymptotic normality lies in actually showing that dependence on the distant past is slight for these functions. It is not apparent to me that this is a simple consequence of mixing. Despite this problem much of the earlier part of the paper is extremely useful and can be paraphrased as follows.

Suppose a model has parameter vector θ. Denote the log-likelihood function by $L_n(\theta)$, the derivative vector by $L_n'(\theta)$ and the second derviative matrix by $L_n''(\theta, \theta_1)$ where the elements of the matrix are evaluated at points on the line segment joining θ and θ_1. Denote by $B_n(\theta)$ the matrix

$$E[-L_n''(\theta, \theta)]$$

Write $d_n(\theta, \theta_1)$ for $L_n''(\theta, \theta_1) - L_n''(\theta, \theta)$

Theorem A.6 *Under the following conditions the maximum likelihood estimates are weakly consistent and asymptotically normal*

1. *$L_n'(\theta)$ converges in distribution to a normal distribution with mean 0.*
2. *The smallest eigenvalue a_n of B_n converges to infinity.*
3. *$-B_n^{-1} L_n''(\theta, \theta)$ converges in probability to the identity matrix.*
4. *given $\epsilon > 0$ there exists $\delta > 0$ such that*

   $$P[a_n^{-1} |d_n(\theta, \theta_1)| < \epsilon] \to 1$$

   as n tends to infinity whenever $|\theta_1 - \theta| < \delta$.

A.4 Miscellaneous

A.4.1 Taylor expansions

The following theorem provides a version of Taylor's theorem for functions of random variables.

Theorem A.7 *Let $\mathbf{Z}_n$ be a sequence of k-dimensional random variables with distribution functions $F_n(\mathbf{z})$ and let $f_n(\mathbf{z})$ be a sequence of functions from $\mathbf{R}^k$ to $\mathbf{R}$. Let $\delta > 0$ and $\alpha = \delta^{-1}(1 + \delta)$. Suppose that for some positive s and N_0 in $\mathbf{N}$*

1. *$\int |\mathbf{z} - \mu|^{\alpha s} dF_n(\mathbf{z}) = a_n^{\alpha s}$ where $a_n \to 0$ as $n \to \infty$.*

2. $\int |f_n(\mathbf{z})|^{1+\delta} dF_n(\mathbf{z}) = O(1)$ *as* $n \to \infty$.

3. $f_n^{(i_1,\ldots,i_s)}(\mathbf{z})$ *is continuous in* $\mathbf{z}$ *over a closed and bounded sphere* S*, for all* $n > N_0$.

4. μ *is in the interior of* S.

5. *There exists real* K *such that for* $n > N_0$

$$|f_n^{(i_1,\ldots,i_s)}(\mathbf{z})| \le K \quad \textit{for all } \mathbf{z} \textit{ in } S.$$

$$|f_n^{(i_1,\ldots,i_r)}(\mathbf{z})| \le K \quad \textit{for } r = 1, \ldots, s-1.$$

$$|f_n(\mu)| \le K$$

Then

$$\int f_n(\mathbf{z}) dF_n(\mathbf{z}) = f_n(\mu) + \sum_{j=1}^{s-1} \frac{1}{j!} \int D^j f_n(\mu)(\mathbf{z} - \mu)^j dF_n(\mathbf{z}) + O(a_n^s)$$

The theorem also holds when $\alpha = 1$*, if (2) is replaced by the requirement that* $f_n(\mathbf{z})$ *be uniformly bounded.*

Proof: Theorem 5.4.3 of [11].

A.4.2 Gershgorin's theorem

The following well-known theorem places crude bounds on the eigenvalues of a matrix in terms of its elements.

Theorem A.8 *A an $n \times n$ square matrix. Each e-value lies in one of the complex discs*

$$|A_{ii} - re^{i\theta}| \le \sum_{j \ne i} |A_{ij}|$$

Proof: This is known as Gershgorin's theorem. See books on numerical linear algebra (e.g. [2]).

Bibliography

[1] E. Batschelet, *Circular Statistics in Biology*, Academic Press (1981)

[2] E.F. Beckenbach and R. Bellman, *Inequalities*, Springer-Verlag (1965)

[3] G.E.P. Box and G.M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day (1970)

[4] C. Brezinski, *Acceleration de la Convergence en Analyse Numerique*, Springer (1977)

[5] J.B. Conway, *Functions of One Complex Variable*, Springer (1973)

[6] H. Cramer, *Mathematical Methods of Statistics*, Princeton Univ. (1946)

[7] M.J. Crowder, *Maximum likelihood estimation for dependent observations*, J. R. Statist. Soc. B, **38**(1976), pp. 45–53

[8] A.P. Dempster, N.M. Laird and D.B. Rubin, *Maximum likelihood from incomplete data via the EM algorithm*, J. R. Statist. Soc. B, **39**(1977), pp. 1–38

[9] J.L. Doob, *Stochastic Processes*, Wiley (1953)

[10] N.I. Fisher and A.J. Lee, *A correlation coefficient for circular data*, Biometrika, **70**(1983), pp. 327-32

[11] W.A. Fuller, *Introduction to Statistical Time Series*, Wiley

[12] P. Hall and C.C. Heyde, *Martingale Limit Theory and its Applications*, Academic Press (1980)

[13] J. Haslett and A.Raftery, *Space-time modelling with long-memory dependence; assessing Ireland's wind power resource*, JRSS C (to appear)

[14] I.A. Ibragimov and Yu.V. Linnik, *Independent and Stationary Sequences of Random Variables*, Wolters-Noordhoff (1971)

[15] M. Iosifescu and R. Theodorescu, *Random Processes and Learning*, Springer-Verlag (1969)

[16] G.M. Jenkins and D.G. Watts, *Spectral Analysis and Its Applications*, Holden-Day (1968)

[17] R.A. Johnson and T. Wehrly, *Measures and models for angular correlation and angular-linear correlation*, J. R. Statist. Soc. B, **39**(1977), pp. 222-29

[18] P.E. Jupp and K.V. Mardia, *A general correlation coefficient for directional data and related regression problems*, Biometrika, **67**(1980), pp. 163-73

[19] J.T. Kent, *Information gain and a general measure of correlation*, Biometrika, **70**(1983), pp. 163-73

[20] A.N. Kolmogorov and Yu.A. Rozanov, *On strong mixing conditions for stationary Gaussian processes*, Theor. Prob. Appl., **5**(1960), p.204-208

[21] T.A. Louis, *Finding the observed information matrix when using the EM algorithm*, J. R. Statist. Soc. B, **44**(1982), pp.226–233

[22] K.V. Mardia, *Statistics of Directional Data*, Academic Press (1972)

[23] K.V. Mardia, *Statistics of Directional Data*, J. R. Statist. Soc. B, **37**(1975), pp.349–371.

[24] K.V. Mardia, *Characterizations of directional distributions.* In *Statistical Distributions for Scientific Work, Vol. 3*, edited by G.P. Patil, S. Kotz, J.K. Ord, Boston-Reidel (1975)

[25] R.D. Martin and A.E. Raftery, *Outliers, computation and non-Euclidean models*, JASA **82**(1987), pp.1044-50

[26] D.L. McLeish, *A maximal inequality and dependent strong laws*, Ann. Probab. **3**(1975), pp.829-39

[27] A.E. Raftery, *A model for high-order Markov chains*, J. R. Statist. Soc. B, **47**(1985), pp. 528–539

[28] L.-P. Rivest, *Some Statistical Methods for Bivariate Circular Data*, J. R. Statist. Soc. B, **44**(1982), pp. 81-90

[29] J.S. Rao and Y.R.Sarma, *A measure of association for circular data*, Int. Statist. Inst. 44th Session — Contributed Papers vol. 2, pp. 571-574.

[30] J.G. Saw, *Dependent unit vectors*, Biometrika **70**(1983), pp.665-71

[31] M.A. Stephens, *Random walk on a circle*, Biometrika **50**(1963), pp.385-90

[32] A. Wald *Note on the consistency of the maximum likelihood estimate*, Ann. Math. Statist., **20**(1949), pp.595–601

[33] T.E. Wehrly and R.A. Johnson, *Bivariate models for dependence of angular observations and related Markov processes*, Biometrika **66**(1979), pp. 255–6