ESTIMATION TECHNIQUES FOR
ARMA TIME SERIES MODELS AND
RANDOM GEOMETRIC SERIES

by

Peter Jeffry Smith

Royal Holloway and Bedford New College

Thesis submitted to the
University of London
for the Degree of
Doctor of Philosophy

May 1988
ABSTRACT

The thesis falls naturally into two sections. Firstly is a study of estimation techniques for ARMA models. Secondly is the work on random geometric series stemming from industrial collaboration with G.E.C. Telecommunications Laboratories. The unifying theory between these topics is discussed in the introduction.

In the first section of the thesis the problem of estimation for ARMA models is considered. This is of long standing interest and the problem of maximum likelihood estimation is substantially solved. However the relationship between approximate and exact maximum likelihood estimation is less well known. The approximate procedures of A.M. Walker and E.J. Godolphin are considered in detail. Some results are produced which compare the two methods to each other and to various exact procedures. These comparisons are used to evaluate the success of some well known approximations. Finally a new approach to exact maximum likelihood estimation is developed which is simple to formulate and implement. This is used to study some numerical problems in estimation which occur near the boundary of the unit circle.

The random geometric series considered in the second section have applications in both statistics and telecommunications. Specific examples of these series have been used to study infinite Bernoulli convolutions, intersymbol interference, error detection and many other subjects. In most applications it is the distribution of the series that is of interest. Initially the problem of calculating the distribution is considered in detail for a specific example concerning
error detection. Several approaches are developed and compared to existing methods. It is shown that the most effective procedure is dependent on numerical properties of the series. Finally the new methods are extended to give two techniques, which can be used in all situations. These procedures are based on the semi-contraction mapping principles and the use of the Poisson summation formula to invert Fourier transforms.
<table>
<thead>
<tr>
<th>Chapter 1</th>
<th>Introduction</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.1 Motivation and review</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>1.2 Theory and notation for ARMA models</td>
<td>21</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 2</th>
<th>Invariance results relating the estimation procedures of Walker and Godolphin</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.1 Introduction</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>2.2 The AR(p) model</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>2.3 The ARMA(1,1) model</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>2.4 The ARMA(p,q) model</td>
<td>37</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 3</th>
<th>The effects of approximations in maximum likelihood estimation</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.1 Introduction</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>3.2 Approximations to the likelihood equations</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>3.3 The AR(1) model</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>3.4 The MA(1) model</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>3.5 The ARMA(1,1) model</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>3.6 Comparisons of exact and approximate estimators</td>
<td>58</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
<td>Page No.</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>9.3</td>
<td>Implementation of the semicontraction mapping</td>
<td>159</td>
</tr>
<tr>
<td>9.1</td>
<td>Fourier Transforms</td>
<td>165</td>
</tr>
<tr>
<td>9.2</td>
<td>Computing the Fourier transforms</td>
<td>166</td>
</tr>
<tr>
<td>9.3</td>
<td>Inversion of the transforms</td>
<td>168</td>
</tr>
<tr>
<td>9.3</td>
<td>A Comparison of the Fourier transform</td>
<td>182</td>
</tr>
<tr>
<td>10.1</td>
<td>The general random geometric series</td>
<td>206</td>
</tr>
<tr>
<td>10.2</td>
<td>Introduction</td>
<td>206</td>
</tr>
<tr>
<td>10.2</td>
<td>The semicontraction mapping principle</td>
<td>207</td>
</tr>
<tr>
<td>10.3</td>
<td>Fourier transforms</td>
<td>216</td>
</tr>
<tr>
<td>10.4</td>
<td>Examples</td>
<td>219</td>
</tr>
<tr>
<td>I</td>
<td>APPENDIX I</td>
<td>235</td>
</tr>
<tr>
<td>II</td>
<td>APPENDIX II</td>
<td>240</td>
</tr>
<tr>
<td>III</td>
<td>APPENDIX III</td>
<td>242</td>
</tr>
<tr>
<td></td>
<td>REFERENCES</td>
<td>246</td>
</tr>
</tbody>
</table>
ILLUSTRATIONS

Figure 5.2.1 The likelihood contour for Sample 1  
Figure 5.2.2 The $\alpha$ and $\beta$-equations for Sample 1  
Figure 5.2.3 The likelihood contour for Sample 15  
Figure 5.2.4 The $\alpha$ and $\beta$-equations for Sample 15  
Figure 5.2.5 The approximate likelihood contours for Sample 15  
Figure 5.2.6 The approximate $\alpha$ and $\beta$-equations for Sample 15  
Figure 5.2.7 A detail of Figure 5.2.3  
Figure 5.2.8 A detail of Figure 5.2.4  
Figure 6.2.1 The even mark parity error detector  
Figure 6.2.2 Spurious error indication  
Figure 6.2.3 Missing error indications  
Figure 6.2.4 Choosing the decay parameter for the leaky integrator  
Figure 6.3.1 Two state model of the voltage process  
Figure 8.3.1 The "tree" approach for the semi-contraction mapping  
Figure 8.3.2 Distribution functions for the voltage process  
Figure 8.3.3 A Pearson curve approximation  
Figure 8.3.4 Non-convergence of the semi-contraction mapping  
Figure 9.3.1 Fourier transforms for the voltage process  
Figure 9.3.2 Fourier transforms for the voltage process
Figure 10.4.6 Probability density functions for Example 10.4.2

Figure 10.4.7 Probability density functions for Example 10.4.2

Figure 10.4.8 Distribution functions for Example 10.4.2

Figure 10.4.9 Distribution functions for Example 10.4.2

Figure 10.4.10 Normal approximation to the distribution function

Figure 10.4.11 Distribution functions for Example 10.4.2

Figure 10.4.12 Distribution functions for Example 10.4.2

Figure 10.4.13 Distribution functions for Example 10.4.2
### TABLES

<table>
<thead>
<tr>
<th>Table Number</th>
<th>Description</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3.1</td>
<td>A comparison of the estimates of Walker and Godolphin</td>
<td>37</td>
</tr>
<tr>
<td>3.5.1</td>
<td>Parameter estimates for the ARMA(1,1) model</td>
<td>60</td>
</tr>
<tr>
<td>3.5.2</td>
<td>Sample variances from Table 3.5.1</td>
<td>61</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Speeds of convergence for exact estimators</td>
<td>77</td>
</tr>
<tr>
<td>5.1.1</td>
<td>Simulations of the ARMA(1,1) model</td>
<td>82</td>
</tr>
<tr>
<td>5.1.2</td>
<td>Simulations of the ARMA(1,2) model</td>
<td>83</td>
</tr>
<tr>
<td>5.1.3</td>
<td>Simulations of the ARMA(2,1) model</td>
<td>84</td>
</tr>
<tr>
<td>5.1.4</td>
<td>Simulations of the ARMA(2,2) model</td>
<td>85</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Simulations of the ARMA(1,1) model</td>
<td>97</td>
</tr>
<tr>
<td>7.4.1</td>
<td>A comparison of approximate and exact density values</td>
<td>148</td>
</tr>
<tr>
<td>7.4.2</td>
<td>Gram-Charlier expansions</td>
<td>148</td>
</tr>
<tr>
<td>8.3.5</td>
<td>Semi-contraction mapping results</td>
<td>164</td>
</tr>
<tr>
<td>9.2.1</td>
<td>Nearest neighbours for the Poisson summation formula</td>
<td>177</td>
</tr>
<tr>
<td>9.3.1</td>
<td>Density values based on Fourier transform</td>
<td>195</td>
</tr>
<tr>
<td>9.3.2</td>
<td>An empirical inspection of an exact solution</td>
<td>203</td>
</tr>
<tr>
<td>10.4.1</td>
<td>State description of Example 10.4.2</td>
<td>222</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

I am greatly indebted to my supervisors E.J. Godolphin and F.M. Clayton for their help, advice and encouragement.

I would also like to thank the staff of Royal Holloway College and the General Electric Company Systems Studies Group for many helpful discussions.

Finally I am indebted to my family and friends and most of all to my wife for their encouragement and support.
CHAPTER 1

INTRODUCTION

1.1 Motivation and review

The study of autoregressive moving-average (ARMA) time series models has received a great deal of attention in the statistical literature since the 1950's. Problems of identification, estimation and prediction have all been studied in depth and applications for these models have arisen in many different fields.

In this thesis interest is centred on the estimation of parameters for ARMA models and their application to the design and analysis of block codes for digital transmissions.

The estimation problem considered is that of estimating the parameters of a Gaussian ARMA model. A Gaussian ARMA(p,q) model for a time series \( (X_t) \) is defined by

\[
X_t + \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} = \epsilon_t + \beta_1 \epsilon_{t-1} + \ldots + \beta_q \epsilon_{t-q}
\]  

(1.1.1)

where \( \epsilon_t \sim N(0, \sigma^2) \) and \( \mathbb{E}(\epsilon_t \epsilon_{t'}) = 0 \) for \( t \neq t' \). It is assumed that the models considered are stationary and invertible and hence that the polynomials

\[
\alpha(Z) = 1 + \alpha_1 Z + \ldots + \alpha_p Z^p, \quad \beta(Z) = 1 + \beta_1 Z + \ldots + \beta_q Z^q
\]

have zeroes which lie strictly outside the unit circle. It is also assumed that \( \alpha(Z) \) and \( \beta(Z) \) have no common factors.
The parameter set to be estimated is \( \{\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q, \sigma^2\} \) although in what follows the emphasis is placed on estimating the coefficients \( \{\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q\} \).

A brief history of the estimation problem is now given to introduce the ideas involved and to motivate the work of later Sections.

The problem of estimating the parameters of an ARMA process is of long standing interest in time series theory. Various methods of tackling this problem have been proposed over the years but the main concern has been the derivation of maximum likelihood (ML) estimation procedures.

The pioneering work of Whittle [1951, 1953, 1954] established that the ML estimator is consistent, efficient and asymptotically normal. Whittle also suggested a practical approach to deriving the ML estimator. However such an approach was not developed for several years since the procedure was complicated and required considerable computational effort.

The computational and numerical problems associated with such a ML procedure meant that early researchers looked to methods other than ML estimation. Up to 1970 most of the estimation procedures developed were large sample techniques based on the sample serial correlations. The methods of Durbin [1959] and Walker [1961, 1962] are two such examples. Durbin's method remains the only estimation procedure which is non-iterative. The procedures of Walker were not thought to be asymptotically ML (Osborn [1976]) but the work of Godolphin [1980] and some results in Chapter 2 provide some evidence that this is in fact the case.
The work of Box and Jenkins [1970] revolutionized the approach to time series theory and estimation in particular. They showed how to construct the exact likelihood function for AR and MA models. Also several approximations were suggested for the likelihood of ARMA models. Estimation was then performed by numerical maximization of the likelihood functions. In particular the non-linear maximization routine due to Marquardt [1963] was proposed. The increased power and speed of computers made these techniques feasible and led the way to the development of exact ML methods. The estimators of Box and Jenkins received a great deal of interest in the literature and their techniques were used and developed by many authors (McLeod [1977]).

The work of Galbraith and Galbraith [1974] and Newbold [1974] was the next major development. Both authors showed how the exact likelihood of an ARMA process could be calculated. Hence for the first time a general method was available for exact ML estimation of ARMA processes via maximization of the likelihood function. Clearly there can be problems with such numerical maximization routines especially in multiparameter situations. However the development of improved algorithms has made their use increasingly efficient.

Interest in the estimation problem was also apparent in the engineering literature with Akaike [1973] and Caines and Rissanen [1974] both producing estimation procedures.

Following these developments the emphasis for several years was on computationally efficient methods of calculating and maximizing the exact likelihood for an ARMA model. To this end Ali [1977], Dent [1977], Ansley [1979] and Ljung and Box [1979] all made contributions.
The corresponding estimation problem for multivariate time series was also studied. Exact multivariate methods were developed by Phadke and Kedem [1978], Nicholls and Hall [1979], Ansley and Kohn [1983] and others.

A new approach to estimation was to use the state variable form which allows estimation to be performed via Kalman filtering. This technique has been increasingly accepted and several versions have been produced. Akaike [1978] was one of the pioneers of this approach for which several algorithms were developed (Gardner, Harvey and Phillips [1980], Pearlman [1980] and Mélard [1984]).

Despite the proliferation of exact ML theory since 1970 interest in approximate theory has continued. This has partly been due to the complex nature of the exact methods and partly because the properties of exact and approximate estimators are not well known. This is especially the case for small samples and when the parameters lie close to the invertibility boundary.

The large sample procedures of Godolphin [1977, 1984] and Pham Dinh [1979, 1984] were based on the sample serial correlations and were asymptotically ML. Later work by Porat [1987] showed that the sample serial covariances are not usually efficient estimators of the model covariances and suggested implications for their use in estimation.

The approximate techniques of Anderson [1975, 1977] were not based on the sample serial correlations. Instead the common assumption of zero pre-period values was made. Anderson considered estimation in the time and frequency domains for a large class of models. He was also one of the first to use likelihood derivatives to
implement estimation by Newton-Raphson and Scoring methods. In fact Akaike [1973] showed that the method of Hannan [1959] was equivalent to one step of a Newton-Raphson procedure to maximize an approximate likelihood function.

These approaches based on likelihood derivatives have been extended to the exact ML case by several authors including Ansley and Kohn [1985], Wnsek and Reinsel [1986], Porat and Friedlander [1986] and Pham Dinh [1987].

Recent work has also tackled the problems of missing data and contributions have been made by Ansley and Kohn [1983, 1985], Kohn and Ansley [1986] and Wnsek and Reinsel [1986].

Increasingly efficient estimation algorithms have been produced and the recent approach of Shea [1987], implemented by the NAG routine G13DCF, makes some claims to being one of the fastest. This algorithm is another which is based on Kalman filtering.

This then is a brief history of the estimation problem. As can be seen the problem is substantially solved with algorithms available for exact ML estimation for multivariate as well as univariate time series and also for time series with missing data. However there are still areas of interest in the theory which have not been fully investigated.

Despite the overall shift in emphasis from approximate to exact ML methods relatively little work has been done of a comparative nature. The studies which have been produced are predominantly Monte Carlo studies, usually of low order models such as the MA(1) process. The results have not been unanimous but several contributions of interest have been made by Nelson [1972], McClave [1974], Kang [1975],
Osborn (1976), Cooper and Thompson (1977), Phadke and Kedem (1978) and Godolphin and De Gooijer (1982). These authors have concentrated on looking for possible bias and making comparisons of mean-squared errors of parameter estimates. The consensus of opinion seems to be that for large samples there is very little to choose between many different methods. For small samples several studies report larger m.s.e. for approximate techniques although again this is not unanimous. The behaviour of the estimates near the boundary of the unit circle has been of special interest. For small samples differences in m.s.e. seem to be increased here. Also some indications of bias towards zero have been found in approximate methods in this region.

The conclusions made by Ansley and Newbold (1980) were based on empirical forecasting comparisons for finite samples. Their results agree with the above comments to a large degree. They conclude that it is possible for either approximate or exact procedures to be superior in different situations. However they recommend the use of exact techniques on the basis of greater reliability.

On the basis of the above remarks it is of interest to see if any simple analytical comparisons can be made between approximate and exact estimation procedures. To this end the AR(1), MA(1) and ARMA(1,1) models are studied in Chapter 3. The effects of two common approximations are considered there, firstly the exclusion of the determinant term in the likelihood and secondly the use of Shaman's approximate inverse. These two simplifications form the basis of Godolphin's estimation procedures (1977, 1984). Hence it is
Godolphin's "direct" approach which is used as the basis of the comparisons in Chapter 3.

As mentioned above the approach of Walker [1961, 1962] was not originally thought to be ML. In 1980 Godolphin showed that Walker's estimator for MA models is asymptotically ML. This was shown by proving that Walker's and Godolphin's estimation equations are asymptotically equivalent. Since both authors later developed estimation procedures for ARMA models which were extensions of their MA techniques it is of interest to see if the same equivalence holds. This problem is studied in Chapter 2. As well as indicating that Walker's approach may well be asymptotically ML Chapter 2 also provides some new results for the ARMA(1,1) model and serves as a description of the direct approach for the work of Chapter 3.

Chapter 4 concentrates on the current interest in estimation procedures, namely the calculation of the exact likelihood and its derivatives. The methods of Scoring and Newton-Raphson are discussed as well as other possible iterative schemes. The method derived is similar to that of Wincek and Reinsel [1986] relying on various recursive equations. The motivation is similar to that of Pham Dinh [1987] who sought a simple approach to the necessary calculations.

In Chapter 5 the results of Chapters 2, 3 and 4 are illustrated numerically. The behaviour of estimates for approximate and exact procedures are studied. Some likelihood surfaces are considered to explain these properties and special attention is given to the boundary of the unit circle. Comparisons are made to several other Monte Carlo studies and some problems with maximization are considered.
The link between estimation of the correlations and the parameters of an ARMA model is a procedure known as the Cramer-Wold factorization (Godolphin [1976]). This procedure therefore links the estimation methods of Godolphin and Walker as discussed in Chapter 2.

A further application of the Cramer-Wold factorization in digital transmissions has been studied by Justesen [1982]. This application stems from the equivalence of the spectral density function for an ARMA process with the power spectra of a block code.

This equivalence implies that the second order statistical properties of a given block code can be studied within the general framework of ARMA time series theory. Justesen [1982] has shown that results for ARMA models have implications for the design and analysis of block codes. One example is the use of linear minimum mean squared error predictors in designing codes. Hence the estimation and prediction of ARMA models has many interesting applications to digital transmission theory.

The use of ARMA models for the design of block codes is a difficult problem. Justesen [1982] has produced codes which have maximum rate for a given spectra. Some results were also produced on maxentropic codes which is a substantially unsolved problem (Slepian [1972]). However greater practical interest lies in the design of codes for which the spectra is not known in advance. This enables the code to be designed to meet the requirements of the transmission system without the restrictions of a fixed spectra. Hence the direct application of ARMA models to the design problem, which assumes a given spectra, has not been considered.
The analysis of the performance of a block code is of more immediate interest. Once a code is proposed its spectra can be calculated (Cariolaro and Tronca [1974]) and the equivalent ARMA model is available via the Cramer-Wold factorization (Justesen [1972]).

One of the major features of the spectra of a block code is its behaviour near the origin. It is usual to have zero spectral value at the origin and to require small values in a region around the origin. Such behaviour limits the amount of interference in the digital transmissions based on such a code. This behaviour can be observed directly from a knowledge of the spectra. Hence no further information is available from a direct application of ARMA models and a different approach is required.

To gain more detailed information about interference in block coded transmissions it is useful to consider the distribution functions of various random variables related to interference. The variables occurring in this situation can be described as random geometric series. Hence it is the calculation of the distribution functions of random geometric series which is considered in Chapters 6 to 10.

In Chapter 6 applications of random geometric series are introduced in more detail. It is shown that the interference problem can be studied via a specific example related to error detection. The distribution function for this example is calculated by various approximate methods in Chapter 7.

In Chapters 8 and 9 various exact methods are derived to calculate the distribution function. Finally in Chapter 10 the methods discussed previously are generalized to the study of an arbitrary
random geometric series. Various examples of the general case are also given.

1.2 Theory and notation for ARMA models

For a stationary invertible Gaussian ARMA process \((X_t)\), \(t=1, \ldots, n\), the following definitions apply.

The sample serial covariance,

\[ C_k = \frac{1}{n-k} \sum_{t=1}^{n-k} X_t X_{t+k} \]

The sample serial correlation,

\[ r_k = \frac{C_k}{C_0} \]

The covariance, \( \gamma_k = \text{cov}(X_t, X_{t+k}) \)

The correlation, \( \rho_k = \frac{\gamma_k}{\gamma_0} \)

The asymptotic covariance matrix \( V = (v_{ij}) \) of \( n^{-1} (r_1 - \rho_1), \ldots, n^{-1} (r_q - \rho_q), \ldots, n^{-1} (r_{q+1} - \rho_{q+1}), \ldots, n^{-1} (r_m - \rho_m) \) where \( m \leq n-1 \) is due to Bartlett (1946) who showed that

\[ v_{ij} = \delta_{j-i} + \delta_{i+j} + 2(\rho_1 \rho_j \rho_0 \rho_j - \rho_j^2 \rho_j - \rho_j) \] \hspace{1cm} (1.2.1)

where

\[ \delta_k = \begin{cases} \rho_k & \text{if } k < 0 \\ \rho_{-k} & \text{if } k > 0 \end{cases} \]

where \( s_k = \rho s^{k+1} \). Where terms appear for both ARMA models and the corresponding MA model the MA term is specified by the superscript \( M \). For instance the correlations of model (1.1.1) are denoted by \( \rho_k \). The correlations of
the model where \( \alpha_1 = \ldots = \alpha_p = 0 \) are denoted by \( \Phi_k \).

The full likelihood function for a stationary invertible Gaussian ARMA model is given by

\[
L = (2\pi \sigma^2)^{-n/2} (\text{det} \Gamma)^{-1/2} \exp \left( -\frac{1}{2\sigma^2} X^T \Gamma^{-1} X \right)
\]  

Many authors consider estimation procedures which maximize the log likelihood

\[
\log L = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \text{log det} \Gamma - \frac{1}{2\sigma^2} (X^T \Gamma^{-1} X)
\]  

The likelihood equations are gained from differentiating (1.2.4).

\[
\frac{\sigma^2}{n} = \frac{X^T \Gamma^{-1} X}{n}
\]  

\[
\frac{\delta^2}{\delta \theta_1} (\text{det} \Gamma) + \text{det} \Gamma X \frac{\delta}{\delta \theta_1} (\Gamma^{-1}) X = 0
\]  

\[
\theta \in \{ \alpha, \ldots, \alpha, \beta, \ldots, \beta \}
\]

Substituting for \( \sigma^2 \) in (1.2.6) gives

\[
\frac{\delta}{\delta \theta_1} ((X^T \Gamma^{-1} X) (\text{det} \Gamma)^{1/n}) = 0
\]  

It is usually this modified likelihood equation that is discussed in the following Chapters. A solution to (1.2.7) would constitute an exact estimation procedure. Common simplifications to (1.2.7) are given by the following two approximations

\[
(\text{det} \Gamma)^{1/n} \approx 1
\]  

This is described as the determinant approximation and has been used extensively in the literature (Godolphin [1984]).
This is Shaman's approximation. In his papers of 1973, 1975 and 1976, Shaman proved that $\Gamma^{-1}$ has rank $2\max(p, q)$ and that $\Gamma$ can be taken to be the inverse of $\Gamma$ if "end-effects" are ignored (Godolphin [1982]). The matrix $\Gamma$ is the covariance matrix for the ARMA $(q, p)$ model.

The use of both approximations is the basis of Godolphin's direct approach which uses the approximate likelihood equation

$$\frac{\delta}{\delta \theta_1} (X'X) = 0$$  (1.2.9)

If Shaman's approximation is used but not the determinant approximation then the estimation equation is

$$\frac{\delta}{\delta \theta_1} (X'X) (\text{detI})^{1/2} = 0$$

Godolphin (1980) showed that for ARMA models the two estimation procedures are asymptotically equivalent. Hence it was shown that Walker's procedure for moving average models is asymptotically maximum likelihood. In the 1975/1976 paper Godolphin also demonstrates that the equivalence of the two methods in not a simple application of the maximum likelihood convergence result given for example in Hansen (1971).

Both Walker and Godolphin extended their methods for many people to ARMA$(p,q)$ models. It is of interest to see if a similar equivalence holds for the new methods. Before attempting to investigate this equivalence it is necessary to describe in more detail the two methods considered.
CHAPTER 2

INVARINACE RESULTS RELATING THE

ESTIMATION PROCEDURES OF WALKER AND GODOHIN

2.1 Introduction

The approximate estimation techniques of Walker [1961, 1962] and Godolphin [1977, 1984] have been described briefly in Chapter 1. The approach of Walker was based on the likelihood function of the asymptotic distribution. Differentiating the resulting likelihood gave the likelihood equations. Godolphin's approach was to approximate the derivative of the likelihood based on the exact distribution. Walker's estimation procedure was not shown to be asymptotically maximum likelihood. However, Godolphin [1980] showed that for MA(q) models the two estimation procedures are asymptotically equivalent. Hence it was shown that Walker's procedure for moving average models is asymptotically maximum likelihood. In the 1980 invariance paper Godolphin also demonstrates that the equivalence of the two methods is not a simple application of the maximum likelihood invariance result given for example in Zehna [1966].

Since both Walker and Godolphin extended their methods for MA(q) models to ARMA(p,q) models it is of interest to see if a similar equivalence holds for the new methods. Before attempting to investigate this equivalence it is necessary to describe in some detail the two methods considered.
Walker's method

Walker's method is described fully in his 1961 and 1962 papers. However, his notation is not followed here. A more convenient version of Walker's estimation equations is given in T.W. Anderson [1971]. Some of the notation used is repeated below. Walker's procedure estimates the parameters \((\alpha_1, \ldots, \alpha_p, \rho_1, \ldots, \rho_q)\) and estimates of \(\beta_1, \ldots, \beta_q\) are obtained from these by the Cramer-Wold factorization (Godolphin [1976]). The estimates gained are based on the sample serial correlations \((r_1, \ldots, r_m)\) defined in Section 1.2. The number of sample serial correlations used is \(m\) where \(m \leq n-1\) and \(n\) is the number of observations.

Walker considers the statistics \(S_1, \ldots, S_m\) defined by:

\[
\begin{align*}
S_1 &= \langle S_1, \ldots, S_q \rangle' \\
S_2 &= \langle S_{q+1}, \ldots, S_{q+p} \rangle' \\
S_3 &= \langle S_{q+p+1}, \ldots, S_m \rangle'
\end{align*}
\]

where

\[
\begin{align*}
S_1 &= r_i & i &= 1, \ldots, q \\
S_i &= \sum_{s=0}^{p} \alpha_s r_{i-s} & i &= q+1, \ldots, q+p & (2.1.1) \\
S_i &= \prod_{s,t=0}^{p} \alpha_s \alpha_t r_{i-s-t} & i &= q+p+1, \ldots, m
\end{align*}
\]

The statistics defined by (2.1.1) can be written in matrix form.
The distribution of $n^{\frac{1}{2}}(r_1 - \rho_1), \ldots, n^{\frac{1}{2}}(r_q - \rho_q), n^{\frac{1}{2}}r_{q+1}, \ldots, n^{\frac{1}{2}}r_m$ is asymptotically normal with mean zero and covariance matrix $V = (V_{ij})$ given by (1.2.1). Since $T$ is of rank $m$ the distribution of $n^{\frac{1}{2}}(S_1 - \rho_1), \ldots, n^{\frac{1}{2}}(S_q - \rho_q), n^{\frac{1}{2}}S_{q+1}, \ldots, n^{\frac{1}{2}}S_m$ has a limiting distribution which is normal with mean zero and covariance matrix $\bar{V} = TVT'$.

If $\bar{I}$ is partitioned as below

\[
\bar{I} = \begin{bmatrix}
I_{11} & I_{12} & I_{13} \\
I_{21} & I_{22} & I_{23} \\
I_{31} & I_{32} & I_{33}
\end{bmatrix}
\]

then Anderson [1971] shows the estimation equations are given by

\[
S_2 = I_{23} I_{33}^{-1} S_3 \tag{2.1.3}
\]

\[
\rho = S_1 - I_{13} I_{33}^{-1} S_3 \tag{2.1.4}
\]

where $\rho = (\rho_1, \ldots, \rho_q)$.

For a pure MA model the $\sigma$-equations (2.1.3) are irrelevant and the $\beta$-equations (2.1.4) can be written

\[
\bar{m} = r_1 - V_{12} V_{22}^{-1} r_2 \tag{2.1.5}
\]
where \( r_1 = (r_1^{\ldots}, r_q^{\ldots}) \), \( r_2 = (r_{q+1}^{\ldots}, r_m^{\ldots}) \).

\( V_{12} \) and \( V_{22} \) are partitions of the covariance matrix \( V \) defined by

\[
V = \begin{bmatrix}
V_{11} & V_{12} \\
V_{21} & V_{22}
\end{bmatrix}
\]

where \( V \) is the covariance matrix of \( n^x(r_1 - \bar{r}_1), \ldots, n^x(r_q - \bar{r}_q), n^x(r_{q+1}), \ldots, n^x(r_m) \). Hence \( V \) is defined by (1.2.1).

An interesting identity which relates the MA covariance matrix to the ARMA covariance matrix is given by

\[
V_{22} = \frac{Y_0^2}{m} \sum_{m} \Phi_{33}
\]

This identity follows from the definition of \( \Gamma \) in (2.1.2). Since \( \Phi = TWT' \) the elements of \( \Phi_{33} \) can be identified as

\[
(\Phi_{33})_{ij} = \begin{cases} 
\frac{p}{q+p+s+t} & s,t,u,v = 0 \\
\alpha_s \alpha_t \alpha_u \alpha_v & q+p+i-s-t, q+p+j-u-v
\end{cases}
\]

Similarly the elements of \( \Phi_{13} \) and \( \Phi_{23} \) are given by

\[
(\Phi_{13})_{ij} = (\Phi_{23})_{ij} = \begin{cases} 
\frac{p}{q+p+s+t} & u,v = 0 \\
\alpha_t \alpha_u \alpha_v & q+i-s, q+p+j-u-v
\end{cases}
\]

The elements \( \Phi_{ij} \) are given in terms of the \( \emptyset_k \) terms by (1.2.1). Simplification can be achieved by using two well known relations

\[
s,t = 0 \quad s,t \emptyset_k = 0 \quad \text{for } k > 2q+1
\]
Using (2.1.7) and (2.1.8) the elements $x_{ij}, y_{ij}$ and $z_{ij}$ can be written

$$x_{ij} = \sum_{s,t,u,v=0}^{p} \alpha_s \alpha_t \alpha_u \alpha_v \delta_{i-j+u+v-s-t}$$

$$y_{ij} = \sum_{u,v=0}^{p} \alpha_u \alpha_v (\delta_{i+q+p+j-u-v} + \delta_{q+p+j-i-u-v} - 2 \delta_{i+q+p+j-u-v})$$

$$z_{ij} = \sum_{t,u,v=0}^{p} \alpha_t \alpha_u \alpha_v \delta_{p+j-i+s-u-v}$$

The relation (2.1.6) follows from expanding (2.1.9) to give

$$x_{ij} = \sum_{u,v=0}^{p} \alpha_u \alpha_v \delta_{x-u-v} \rho_{x+i-j-s-t}$$

Invoking the result

$$\sum_{s,t,k-s-t=0}^{m} \alpha_s \alpha_t y_{k-s-t} = y_k$$

gives

$$\gamma_{0}^{2}x_{ij} = \gamma_{0}^{2} \sum_{x} \rho_{x} x_{x+i-j} = \gamma_{0}^{2} \delta_{i-j}$$

Thus

$$\gamma_{0}^{2} \Theta_{ij}^{0} = \gamma_{0}^{2} \delta_{i-j} = \gamma_{0}^{2} \Theta_{22}^{0}$$

and the result follows.

Godolphin's method

The notation of Godolphin [1984] is repeated below. The $\alpha$-equations are given by

$$\alpha = -\Psi^{-1}y$$
where
\[ \Psi = \begin{bmatrix} y_0 & \cdots & y_{p-1} \\ \vdots & \ddots & \vdots \\ y_{p-1} & \cdots & y_0 \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix}, \quad \alpha = (\alpha_1, \ldots, \alpha_p)^t \] (2.1.14)

The elements \( y_0, y_1, \ldots, y_p \) can be written
\[ y_i = \Pi_i^p + \sum_{k=1}^{p} \Pi_k^p \Pi_{k+1}^p r_k \] (2.1.15)

where \( \Pi_k^p \) is the autocovariance function for the AR(q) process

\[ X_t + \beta_1 X_{t-1} + \ldots + \beta_q X_{t-q} = \epsilon_t \]

The \( \beta \)-equations are given by
\[ \rho = \epsilon_1^p + D_1^p D_2^p \] (2.1.16)

where \( \epsilon_1^p = (r_1^p, \ldots, r_q^p) \) and \( \epsilon_2^p = (r_{q+1}^p, \ldots, r_m^p) \). The modified sample serial correlations \( r_k^p \) are defined by
\[ r_k^p = \frac{c_k}{c_0} \] (2.1.17)

The matrices \( D_1^p \) and \( D_2^p \) are defined by
\[ C^p_k = \sum_{k=1}^{p} C_{k-1}^p (C_k + C_{k+1}) \] (2.1.18)

Hence \( D_1^p = 0 \). Using this result, Walker's equation collapses to \( D_2^p = 0 \), which is equivalent to the full-Walker equations.
\[
(D_1^*)_{ij} = d_{ij}, \quad i,j = 1,...,p
\]
\[
(D_2^*)_{ij} = d_{ij}, \quad i=1,...,p, \quad j = q+1,...,m \quad (2.1.18)
\]

\[
\delta \Pi_1^* \frac{d_{ij}}{\delta \beta_1} = 0
\]

For a pure MA process the estimation equations are given by
\[
\mathbf{H} = \mathbf{r}_1 + D_1^{-1} D_2^* \mathbf{r}_2 \quad (2.1.19)
\]
where \( \mathbf{r}_1 \) and \( \mathbf{r}_2 \) are defined in (2.1.5).

The invariance result for MA models

The estimation equations due to Walker (2.1.5) and Godolphin (2.1.19) are of a similar form. In his 1980 invariance paper Godolphin showed that \( \Delta_1 \mathbf{V}_2^{-1} + D_1^{-1} D_2^* \) is asymptotically zero. Hence the two estimation equations are asymptotically equivalent and Walker's approach is shown to be asymptotically maximum likelihood.

2.2 The AR(p) model

Using Walker's \( \alpha \)-equation (2.1.3) it is clear that \( H_{23} = 0 \) for an AR model since

\[
Z_{1,j} = \sum_{s=0}^{p} \alpha_s \sum_{u=0}^{j} \alpha_u \sum_{v=0}^{j} \alpha_v \beta_{p+j-1+s-u-v}
\]
from (2.1.11). The bracketed term is zero for \( p+j-1+s > 1 \) using (2.1.7). Hence \( Z_{1,j} = 0 \). Using this result Walker's \( \alpha \)-equation collapses to \( S_2 = 0 \), which is equivalent to the Yule-Walker equations,
\[ \sum_{j=0}^{p} \alpha_j r_{1-j} = 0, \ i=1, \ldots, p \]

Godolphin's equations (2.1.13) also collapse to the Yule-Walker equations since for an AR(p) model \( \Pi_0^* = 1 \) and \( \Pi_k^* = 0, \ k > 1 \). This gives \( \psi_1 = r_1 \) from (2.1.15) and substituting in (2.1.13) gives the Yule-Walker equations.

Hence the two estimation procedures are exactly equivalent for AR(p) models.

2.3 The ARMA(1,1) model

Walker's \( \varphi \)-equations

Walker's \( \varphi \)-equation is given by (2.1.3). For the ARMA(1,1) model the form of \( S_2 \) and \( S_3 \) is simple.

\[ S_2 = r_2 + \alpha r_1 \]

\[ S_3 = \begin{bmatrix} \alpha^2 & 2\alpha & 1 \\ \vdots & \ddots & \vdots \\ \alpha^2 & 2\alpha & 1 \end{bmatrix} \begin{bmatrix} r_1 \\ \vdots \\ r_m \end{bmatrix} \]

(2.3.1)

Hence the formulation of the equation relies on the calculation of \( \underline{L}_{23}^{-1} \) and \( \underline{L}_{33}^{-1} \). \( \underline{L}_{23} \) has only two non zero terms from (2.1.7) and (2.1.11).

Hence \( Z_{11} = 0 \) for \( j > 3 \).

Hence \( Z_{1,1} \), \( Z_{1,2} \) and the first two rows of \( \underline{L}_{33}^{-1} \) are all that is required. Using (2.1.11), (1.2.2) and calculating \( \rho_k \) for an ARMA(1,1) model it can be shown that
\[ Z_{1,1} = \beta(1-\alpha^2)^2 (2+2\beta^2-\alpha \beta) (1+\beta^2-2\alpha \beta)^{-2} \]

\[ Z_{1,2} = \beta^2(1-\alpha^2)^2(1+\beta^2-2\alpha \beta)^{-2} \] (2.3.2)

\[ \Phi_{33}^{-1} \] is a little more difficult. However using (2.1.9) the first 2 rows of \( \Phi_{33}^{-1} \) can be calculated. Let the first 2 rows be denoted \( x' \) and \( y' \) then \( x', y' \) satisfy

\[ \begin{bmatrix} x' \\ y' \end{bmatrix} \Phi_{33}^{-1} \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \beta \\ \beta \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} \] (2.3.3)

This formulation yields two sets of difference equations. Let

\[ x = (x_1), \ y = (y_1), \ i = 1,..., m-2. \]

The difference equation for \( x_1 \)

\[ \varphi_2 x_1 + \varphi_1 x_{i+1} + \varphi_0 x_{i+2} + \varphi_1 x_{i+3} + \varphi_2 x_{i+4} = 0 \quad i = 1,..., m-6 \]

The initial conditions are

\[ \varphi_0 x_1 + \varphi_1 x_2 + \varphi_2 x_3 = 1 \]

\[ \varphi_1 x_1 + \varphi_0 x_2 + \varphi_1 x_3 + \varphi_2 x_4 = 0 \]

The end conditions are

\[ \varphi_2 x_m-4 + \varphi_1 x_m-3 + \varphi_0 x_m-2 = 0 \]

\[ \varphi_2 x_m-5 + \varphi_1 x_m-4 + \varphi_0 x_m-3 + \varphi_1 x_m-2 = 0 \]

The solution to the equation is complicated algebraically but simple to perform using standard techniques giving

\[ x_1 = i(1+\beta^2 - 2\alpha \beta)^2 (-\beta)^{i-1} (1-\alpha^2)^{-2} \] (2.3.4)
The difference equation for \( y_1 \)

Similarly a difference equation for \( y \) can be set up using (2.3.3). This has solution

\[
y_1 = ((1+2\beta^2)1-1)(1+\beta^2-2\alpha\beta)^2(-\beta)^{1-2}(1-\alpha^2)^{-2}
\]

(2.3.5)

Returning to Walker’s \( \alpha \)-equations they can now be written

\[
\alpha_1 + \alpha r_1 = \Phi_{23} \Phi_{33}^{-1} S_3
\]

where \( \Phi_{23} \Phi_{33}^{-1} \) can be gained from (2.3.2), (2.3.4) and (2.3.5) and \( S_3 \) is given in (2.3.1). Substituting in for these equations gives

\[
\frac{m}{m-1} (-\beta)^{1-2} (1 + 1(\alpha\beta-1))r_1 = 0
\]

(2.3.6)

Let (2.3.6) be denoted by \( W_\alpha = 0 \). Solving for \( \alpha \) gives the final version of Walker’s \( \alpha \)-equation.

\[
\alpha = -\frac{m-2}{m-1} (i-1)(-\beta)^{1-2} r_1
\]

(2.3.7)

Walker’s \( \beta \)-equations

Walker’s \( \beta \)-equation is given by (2.1.4). As for the \( \alpha \)-equations, \( S_1 \) and \( S_3 \) are easy to formulate and the work mainly involves calculating \( \Phi_{13} \Phi_{33}^{-1} \). However \( \Phi_{13} \) has only one non-zero term from (2.1.10) and (2.1.7). Hence

\[
y_{1j} = 0 \quad \text{for} \quad j > 2
\]

(2.3.8)

Calculating \( p_\alpha \) for an ARMA(1,1) model enables (2.1.10) to be expanded giving
\[ y_{1,1} = \beta^2 (1-\alpha^2)^2 \left(1 + \beta^2 - 2\alpha\beta\right)^{-2} \]  

Using (2.3.1), (2.3.4) and (2.3.9) Walker's \( \beta \)-equation can be written

\[ \rho_1 = (1-\alpha\beta) \sum_{i=1}^{m-2} (2-1(1-\alpha\beta)) (-\beta)^{i-1} r_i \]  

(2.3.10)

\( \rho_1 \) can be evaluated as \((\beta-\alpha)(1-\alpha\beta)(1+\beta^2-2\alpha\beta)^{-1}\). Hence (2.3.10) can be rewritten

\[ \beta - \alpha - (1+\beta^2-2\alpha\beta) \sum_{i=1}^{m-2} (2-1(1-\alpha\beta)) (-\beta)^{i-1} r_i = 0 \]  

(2.3.11)

Let (2.3.11) be denoted by \( W_\beta = 0 \).

Godolphin's \( \alpha \)-equations

Godolphin's \( \alpha \)-equation can be formulated from a knowledge of the \( \Pi_k^\alpha \) terms. For an ARMA(1,1) process

\[ \Pi_k^\alpha = (-\beta)^k (1-\beta^2)^{-1} \]

Using this result (2.1.13) can be written

\[ \alpha = \beta - (1+\beta^2) \sum_{k=1}^{m} (-\beta)^{k-1} r_k \]  

(2.3.12)

In fact terms involving \((-\beta)^{m+1}\) have been neglected in formulating (2.3.12). Rewriting (2.3.12) gives

\[ \alpha(1+2 \sum_{k=1}^{m} (-\beta) r_k) - \beta + (1+\beta^2) \sum_{k=1}^{m} (-\beta)^{k-1} r_k = 0 \]  

(2.3.13)

Let (2.3.13) be denoted by \( G_\alpha = 0 \).
Godolphin's $\beta$-equations

Godolphin's $\beta$-equation is given by (2.1.19). For the ARMA(1,1) model this takes the form

$$P_1 = r_1 + \left(d_{1,1},1 \right)^{-1} \sum_{k=2}^{\infty} d_{1,k} r_k \tag{2.3.14}$$

$$r_k = C_k^* = \frac{(1+\alpha^2)r_k + \alpha r_{k-1} + r_{k+1}}{C_0 (1+\alpha^2) + 2\alpha r_1}$$

$$d_{1,k} = \frac{\delta P_k}{\delta \beta} = (-\beta)^{k-1} \left(2\beta^2 + k(1-\beta^2)(1+\beta^2)^{-1} \right)$$

Expanding (2.3.14) gives

$$(1+\alpha^2) \beta - \alpha (1+\beta^2) + \sum_{k=1}^{\infty} m_k r_k = 0 \tag{2.3.15}$$

where

$$m_k = (-\beta)^{k-2} \left((k-1)(1-\beta^4) + 4\beta^2 - k\beta(1-\beta^2) - 2\beta^3 \right)$$

and terms in $(-\beta)^m$ have been neglected. Let (2.3.15) be denoted by

$$G_{\beta} = 0.$$}

The invariance result

As outlined in the previous work the estimation equations due to Walker can be written $(W, \ W')^\gamma = 0$, from (2.3.6) and (2.3.11). Those due to Godolphin are $(G_{\alpha}, \ G_{\beta})^\gamma = 0$, from (2.3.13) and (2.3.15). It is simple to show that the relationship between the two sets of equations is given by
where \( K = \begin{bmatrix} \beta(1+\beta^2 - 2\alpha\beta) & (1-\alpha\beta)^{-1} \\ (1-\beta^2)(\alpha-\beta)-\beta(1-\alpha\beta)(1+\beta^2-2\alpha\beta) & 1 \end{bmatrix} \)

\[
det K = (1-\beta^2)(\alpha-\beta)(1-\alpha\beta)^{-1}
\]

Hence there is an asymptotic invariance between Walker’s estimates and Godolphin’s estimates for the ARMA(1,1) model. The relationship is not exact since (2.3.16) is formulated using \( G_\alpha, G_\beta \), which are versions of Godolphin’s equations which neglect terms in \((-\beta)^{m+1}\). The equivalence relies on the non-singularity of \( K \). \( K \) is singular for \( \beta = \pm 1 \) and \( \alpha = \beta \). However this is not a problem since we assume \( |\beta| < 1 \) and \( \alpha \neq \beta \).

The following table illustrates the invariance result. 10 samples of 100 observations were simulated from an ARMA(1,1) model with \( \alpha = -\frac{1}{4}, \beta = \frac{1}{4} \) and \( m = 50 \).
Table 2.3.1

<table>
<thead>
<tr>
<th>Sample No</th>
<th>Godolphin’s Method</th>
<th>Walker’s Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>α</td>
<td>β</td>
</tr>
<tr>
<td>1</td>
<td>-0.71070</td>
<td>0.55373</td>
</tr>
<tr>
<td>2</td>
<td>-0.72544</td>
<td>0.61637</td>
</tr>
<tr>
<td>3</td>
<td>-0.71385</td>
<td>0.61762</td>
</tr>
<tr>
<td>4</td>
<td>-0.62175</td>
<td>0.64070</td>
</tr>
<tr>
<td>5</td>
<td>-0.76959</td>
<td>0.46463</td>
</tr>
<tr>
<td>6</td>
<td>-0.74172</td>
<td>0.53013</td>
</tr>
<tr>
<td>7</td>
<td>-0.52462</td>
<td>0.54106</td>
</tr>
<tr>
<td>8</td>
<td>-0.71457</td>
<td>0.46133</td>
</tr>
<tr>
<td>9</td>
<td>-0.72247</td>
<td>0.33922</td>
</tr>
<tr>
<td>10</td>
<td>-0.76599</td>
<td>0.53166</td>
</tr>
</tbody>
</table>

The largest discrepancy for the simulations is 0.00005 which illustrates the asymptotic equivalence demonstrated by (2.3.16).

2.4 The ARMA(p,q) model

It has been shown that the methods of Godolphin and Walker are asymptotically equivalent for AR, MA and ARMA(1,1) models. The general ARMA(p,q) model is harder to deal with and such invariance results are only partially developed in the following Section.

The invariance result for the MA model is easily stated and suggests that simplifying the estimation equations for the ARMA model...
may be useful. In particular relating the ARMA estimation equations to the MA equations has been effective.

The first identity of use here is that relating the ARMA covariance matrix to the corresponding MA covariance matrix given by (2.1.12). The second is the asymptotic result

$$V_{12} V_{22}^{-1} = D_1^* D_2^*$$

given in Section 2.1.

Thirdly Godolphin's $\alpha$-equation can be rewritten in the form

$$\begin{bmatrix} \alpha_1 & \cdots & \alpha_{p-1} \\ \vdots & \ddots & \vdots \\ \alpha_{p-1} & \cdots & \alpha_1 \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_{p-1} \end{bmatrix} = 0$$

(2.4.2)

The vector of $\gamma_k$ terms is a linear combination of the sample serial correlations with coefficients involving the $\Pi^*$ terms only. Hence (2.4.2) can be written

$$[B \mid A^t] \Pi^* r = 0$$

(2.4.3)

where

$$B = \begin{bmatrix} \alpha_1 & \cdots & \alpha_{p-1} \\ \vdots & \ddots & \vdots \\ \alpha_{p-1} & \cdots & \alpha_1 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & \cdots & \alpha_{p-1} \\ \vdots & \ddots & \vdots \\ \alpha_{p-1} & \cdots & 1 \end{bmatrix}$$

$\Pi^*$ is a matrix of dimension $(2p \times m)$ defined by

$$(\Pi^*)_{i,1} = \Pi_{p-i}^*$$

$$(\Pi^*)_{i,j} = \Pi_{j+i-p-1}^* + \Pi_{j+p-i-1}^*, \quad j > 2$$

and $r = (r_1, \ldots, r_m)$.
The elements of $\Pi^i$ satisfy the relation
\[
\sum_{i=0}^{q} \beta_i \Pi^i_{j-i} = 0
\]
for $j \geq q+1$. This is easily seen from the definition of $\Pi^i_k$. A direct application of this result is that
\[
(\Pi^i V_{22})_{ij} = 0
\]
for $i = 1, ..., 2p$, $2q+p < j < m+1-2q$. In fact
\[
\Pi^i V_{22} = [ U | O | E ]
\]
where $U$ is of dimension $(2p \times 2q+p)$ and $E$ which has dimension $(2p \times 2q)$ is asymptotically negligible. Hence Godolphin's $\alpha$-equation is asymptotically equivalent to
\[
[ B | A' ] [ U | O ] V_{22}^{-1} r = 0
\]
Using the three results (2.1.6), (2.4.1) and (2.4.6) the estimation equations can be written as below. Walker's equations
\[
S_2 - \begin{bmatrix} Y_0 \\ m \end{bmatrix}^2 F_{23} V_{22}^{-1} S_2 = 0
\]
\[
\rho - S_1 + \begin{bmatrix} Y_0 \\ m \end{bmatrix}^2 F_{13} V_{22}^{-1} S_3 = 0
\]
Godolphin's equations
\[
[ B | A' ] [ U | O ] V_{22}^{-1} r = 0
\]
\[
\frac{\rho}{\rho} - \frac{r}{r} + V_{12} V_{22}^{-1} r = 0
\]
Denote Walker's estimation equations by

\[
\begin{bmatrix}
V_\alpha & W_\alpha \\
W_\beta & W_\beta
\end{bmatrix} \begin{bmatrix} r \\ - \end{bmatrix} = 0
\]

where $V_\alpha$ is a matrix of dimension $(p \times q+p+1)$, $W_\alpha$ is $(p \times m-p-q)$, $W_\beta$ is $(q \times q+p+1)$, and $W_\beta$ is $(q \times m-p-q)$. Similarly denote Godolphin's equations by

\[
\begin{bmatrix}
G_\alpha & V_\alpha \\
G_\beta & V_\beta
\end{bmatrix} \begin{bmatrix} r \\ - \end{bmatrix} = 0
\]

To demonstrate the equivalence of the two methods it would suffice to formulate a non-singular matrix $K$ satisfying

\[
\begin{bmatrix}
G_\alpha & G_\alpha \\
G_\beta & G_\beta
\end{bmatrix} = K \begin{bmatrix}
W_\alpha & W_\alpha \\
W_\beta & W_\beta
\end{bmatrix}
\]

where the identity may be an asymptotic result. Let $K$ be partitioned as below

\[
K = \begin{bmatrix}
K_1 & K_2 \\
K_3 & K_4
\end{bmatrix}
\]

then the submatrices must satisfy

\begin{align*}
G_\alpha &= K_1 W_\alpha + K_2 W_\beta \\
G_\beta &= K_3 W_\alpha + K_4 W_\beta
\end{align*}

(2.4.9)

From (2.3.7) $W_\alpha$ and $W_\beta$ can be written as

\begin{align*}
W_\alpha &= \Phi_{33} V^{-1} \Omega_{22} H \\
W_\beta &= \Phi_{13} V^{-1} \Omega_{22} H
\end{align*}

(2.4.10)
where

\[ H = \begin{bmatrix} 1 & \cdots & \alpha_1 & \alpha_2 & \cdots & 1 \\ \vdots & \ddots & \vdots & \vdots & \alpha_1 & \alpha_2 & \cdots & 1 \\ \end{bmatrix} \]

and \( H \) has dimension \((m-q-p)^2\).

Hence the required result can be written

\[
G_1 H^{-1} V_{22} = K_1 \bar{I}_{23} + K_2 \bar{I}_{13}
\]

\[
G_2 H^{-1} V_{22} = K_3 \bar{I}_{23} + K_4 \bar{I}_{13}
\]

\( \bar{I}_{13} \) has dimension \( q \times m-p-q \) with a non-zero submatrix of dimension \( q \times 2q \). This submatrix can be split into a \( q \times q \) matrix \( B \) and a diagonal lower triangular matrix \( C \) of dimension \( q \times q \). Thus

\[
\bar{I}_{13} = \begin{bmatrix} B & C & 0 \end{bmatrix}
\]

where \( 0 \) is a zero matrix. Similarly \( \bar{I}_{23} \) consists of a non-zero submatrix of dimension \( p \times 2q-p \). This submatrix can be split into a \( p \times 2(q-p) \) matrix \( D \) and a diagonal lower triangular matrix \( E \) of dimension \( p \times p \). Thus

\[
\bar{I}_{23} = \begin{bmatrix} D & E & 0 \end{bmatrix}
\]

Hence the right hand side of (2.4.11) has a non-zero submatrix of dimension \( p \times 2q \). This property is shared by the left hand side and so the relation is feasible. Similarly both sides of (2.4.12) have dimension \( q \times 2q \).

The form of \( C \) means that \( K_2 \) and \( K_4 \) are completely or partially specified by a set of \( q \) recurrence relations. Similarly the form of \( E \) specifies \( p \) recurrence relations which \( K_1 \) and \( K_3 \) must satisfy. For...
It has been shown that the methods of Walker and Godolphin are asymptotically equivalent for AR(p), MA(q) and ARMA(1,1) models. The results of Section 2.4 indicate that such an invariance may hold for the general ARMA(p,q) model. However a complete proof for this case is not available.

As discussed in Section 2.1 these invariance results are not direct applications of the maximum likelihood invariance property. This stems from the different approximations made by the two methods.
Walker's method is based on the asymptotic distribution whereas Godolphin's uses an approximation of the exact distribution.

The invariance is an interesting result in several ways. As discussed by Godolphin (1980) the procedure of Walker was not originally shown to be efficient. The invariance results studied in this Chapter show that Walker's method is not only efficient but asymptotically maximum likelihood for the AR(p), MA(q) and ARMA(1,1) models. It may also be conjectured that these properties apply to the general ARMA(p,q) model.

These properties have two implications for a possible joint estimation approach. The direct \( \beta \)-equations can be used to convert Walker's \( \beta \)-equations to give \( \hat{\beta} \) directly. This avoids the need to use the Cramer-Wold factorization at each iteration. Also it is conceivable that Walker's procedure could converge to a value of \( \rho \) which gives complex values of \( \beta \). Hence if the direct approach has convergence problems then estimation of \( \rho \) by Walker's approach may still give a solution.

The results of Section 2.4 are also of interest in their own right. There are numerous simplifications which can be made for both methods. In particular several results show that expressions involving the ARMA model parameters are simple transformations of similar expressions involving the corresponding MA model parameters. These transformations from ARMA to MA terms can in fact reduce the estimation equations to include only MA expressions linked by matrices involving the \( \alpha \) terms. These results may have applications to other methods. If such results could alter the estimation process for ARMA
models by transforming to the MA model then considerable
simplifications could be achieved.

This detailed consideration of the two approximate procedures is
the background for later work. In Chapter 3 the effect of the
approximations used by the direct approach are considered. Also in
Chapter 5 comparisons are made between the direct approach and various
exact procedures.

In this chapter the emphasis is solely on estimates of $\theta$.

3.1. Estimation of $\theta$ not considered. The approximations
discussed will be those used to formulate the direct
version of the determinantal approximation (1.9.8) and Schur's
determinant inverse (1.3.9). Specific work along these lines has been
done by Godolphin and de Uzupj (1963) who considered AR models and
the determinantal approximation in detail.

Other related work has largely been based on Monte Carlo studies.
Numerous such studies have been discussed briefly in Chapter 1. The
consensus of opinion can be summarized by three points.

(i) Approximate procedures have larger A.E.E. than exact
methods (cf. Wang (1945), Bowden and Austin (1961), Olson (1972)
and others).

(ii) In most cases procedures are less reliable near the boundary
of the null surface (cf. Wang (1945), Ljung and Box (1970), Cooper and

(iii) Some approximate procedures show signs of bias towards
mean (cf. Olson (1972), Wang (1945), Analey and Bensol (1978) and
others).
CHAPTER 3

THE EFFECTS OF APPROXIMATIONS IN

MAXIMUM LIKELIHOOD ESTIMATION

3.1 Introduction

In this chapter the emphasis is solely on estimates of \( \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q \). Estimation of \( \sigma^2 \) is not considered. The approximations discussed will be those used to formulate the direct approach, namely the determinant approximation (1.2.8) and Shaman's approximate inverse (1.2.9). Specific work along these lines has been done by Godolphin and De Gooijer (1982) who considered MA models and the determinant approximation in detail.

Other related work has largely been based on Monte Carlo studies. Several such studies have been discussed briefly in Chapter 1. The consensus of opinion can be summarized by three points.

(i) Approximate procedures have larger m.s.e. than exact procedures (cf. Kang [1975], Phadke and Kedem [1978], Nelson [1972] and others).

(ii) Approximate procedures are less reliable near the boundary of the unit circle (cf. Kang [1975], Ljung and Box [1979], Cooper and Thompson [1977], Nicholls and Hall [1979], McClave [1974] and others).

In this chapter three simple examples are considered, the AR(1), MA(1) and ARMA(1,1) models. The two approximations considered are studied analytically to see if any evidence can be found concerning the three points above. Also some empirical results are presented where the analysis becomes difficult. Firstly it is necessary to construct the relevant likelihood functions under the various approximations.

3.2 Approximations to the likelihood equations

There are four sets of likelihood equations of interest. Some have been given in Chapters 1 and 2 but are included here for the sake of completeness.

(i) The exact likelihood equations (1.2.7)

\[
\frac{\delta}{\delta \theta} \left[ (X' \Gamma^{-1} X) (\det \Gamma)^{1/n} \right] = 0
\]

\[\theta \in (\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)\]

(ii) The equations produced by Godolphin [1984], using the determinant approximation (1.2.8) and Shaman's inverse approximation (1.2.9). These approximations give

\[
\frac{\delta}{\delta \theta} (X' \Gamma^{-1} X) = 0, \quad \theta \in (\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)
\]

(iii) The equations produced by using the determinant approximation (1.2.8) only

\[
\frac{\delta}{\delta \theta} (X' \Gamma^{-1} X) = 0, \quad \theta \in (\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)
\]
The equations produced by using Shaman's inverse approximation (1.2.9) only

$$ \frac{\delta}{\delta \theta} \left[ (X'X)(\text{det}I)^{-1/n} \right] = 0 \quad (3.2.5) $$

$$ \theta \in \{\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q\} $$

For ease of notation the four likelihood equations will be referred to, respectively, as the EXACT, DIRECT, DIRINV and DIRDET equations.

As discussed in Chapter 1, there is a large literature on the exact approach. Further work on this technique is given in Chapters 4 and 5. The DIRECT approach has been studied by Godolphin [1977, 1984] and is closely connected with the procedures of Walker [1961, 1962] as shown in Godolphin [1980] and Chapter 2. This method has been considered in some detail in Chapters 1 and 2. Relatively little has been done on the approaches using only one approximation, DIRDET and DIRINV. Some results on the DIRDET approach for moving average models were given in Godolphin and De Gooijer [1982].

In terms of complexity the DIRECT approach is simplest to formulate and implement. It may not be faster in computer time due to differences in speed of convergence. The other three methods are similar in complexity. All three methods require solution by numerical maximization methods.

The purpose of this chapter is to investigate the differences in parameter estimates produced by the four methods. To do this some simple examples are given in detail.
3.3 The AR(1) model

The model is given by $X_t + \alpha X_{t-1} = \epsilon_t$ under all the standard assumptions concerning $\alpha$ and $\epsilon_t$. The autocovariance matrix for the AR(1) model is well known and is given by

$$\Gamma = \alpha^2 (1-\alpha^2)^{-1} \begin{bmatrix} 1 & -\alpha & \cdots & (-\alpha)^{n-1} \\ -\alpha & 1 & \cdots & (-\alpha)^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ (-\alpha)^{n-1} & (-\alpha)^{n-2} & \cdots & 1 \end{bmatrix}$$

$$\Gamma^{-1} = \alpha^{-2} \begin{bmatrix} 1 & \alpha & \alpha^2 & \cdots & \alpha^{n-1} \\ \alpha & 1+\alpha^2 & \cdots & \alpha^{n-2} \\ \alpha^2 & \alpha & \cdots & \alpha^{n-3} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha^{n-1} & \alpha^{n-2} & \cdots & 1 \end{bmatrix}$$

Shaman's approximation to $\Gamma^{-1}$ is given by

$$\Gamma = \alpha^{-2} \begin{bmatrix} 1 & \alpha & \alpha^2 & \cdots & \alpha^{n-1} \\ \alpha & 1+\alpha^2 & \cdots & \alpha^{n-2} \\ \alpha^2 & \alpha & \cdots & \alpha^{n-3} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha^{n-1} & \alpha^{n-2} & \cdots & 1 \end{bmatrix}$$

The DIRECT method

For an AR(1) model, $\Pi_0 = 1, \Pi_k = 0, k \geq 1$. Hence (3.2.3) gives

$$\delta C_0 = 0$$

Thus the solution is

$$\alpha = -\frac{C_1}{C_0}, \text{ represented by } \alpha_D = -r_1.$$  (3.3.1)
The DIRDET method

Equation (3.2.5) gives

\[ \delta \frac{\delta}{\delta \alpha} \left( (1+\alpha^2)C_0 + 2aC_1 \right) \left( \text{det}\Sigma \right)^{-1/n} = 0, \]

where

\[ \text{det}\Sigma = (1 - \alpha^{2n+2})(1-\alpha^-1) \]

\[ \alpha = -r_1 \left[ \frac{n - 2\alpha^2 K_n(\alpha)}{n - (1+\alpha^2)K_n(\alpha)} \right] = -r_1 \left( L_n(\alpha) \right) \]

where

\[ K_n(\alpha) = \frac{1 - (n+1)\alpha^{2n} + n\alpha^{2n+2}}{(1-\alpha)(1-\alpha^{2n+2})} \]

Hence the DIRDET solution is given by

\[ \alpha_{DD} = -r_1 \left( L_n(\alpha_{DD}) \right) \quad (3.3.2) \]

The DIRINV method

Equation (3.2.4) gives

\[ \delta \frac{\delta}{\delta \alpha} (X^{-1}X) = 0 \]

Now \( r^{-1} = I - E \) where \( E_{1,1} = E_{n,n} = \alpha^2 \) and \( E_{1,j} = 0 \) otherwise. Hence the DIRINV method gives

\[ \frac{\delta}{\delta \alpha} (XEX) - \frac{\delta}{\delta \alpha} (XEX) = 0 \]

\[ 2\alpha C_0 + 2C_1 \left[ \frac{2aX_1^2 + 2aX_n^2}{n} \right] = 0 \]

Hence

\[ \alpha = -r_1 \left( M_n \right) \]

where

\[ M_n = 1 + \frac{X_1^2 + X_n^2}{X_2^2 + \ldots + X_n^2} \]
Hence the DIRINV solution is
\[ \alpha_{DI} = -r_1(\mathbb{M}_n) \]  
\eqref{3.3.3}

The EXACT solution

Equation \( (3.2.1) \) can be simplified using the preceding results. Hence

\[ \frac{\delta}{\delta \alpha} \langle X' (\mathbb{E} - \mathbb{B}) X (\text{det}\Gamma)^{1/n} \rangle = 0 \]

\[ \frac{\delta}{\delta \alpha} \langle (1 + \alpha^2)C_0 + 2\alpha C_1 - \frac{\alpha^2}{n} (x_1^2 + x_n^2) (1 - \alpha^2)^{1/n} \rangle = 0 \]

The solution is given by

\[ \alpha_E = -r_1(\mathbb{M}_n(\alpha_E)) \]  
\eqref{3.3.4}

where

\[ \mathbb{M}_n(\alpha) = \frac{(1 - \alpha^2)n^2 C_0 - 2\alpha \alpha C_0}{(1 - \alpha^2)n^2 C_0 - nC_0(1 + \alpha^2) + (x_1^2 + x_n^2)(\alpha^2 - n + \alpha^2)} \]

Comparing the solutions

The four solutions are all closely related and a direct comparison of \( \alpha_D, \alpha_{DD}, \alpha_{DL} \) and \( \alpha_E \) is possible for the AR(1) model.

The empirical results suggesting the possibility of bias towards zero for approximate techniques can be proven for this particular case.

The four solutions are

\( (i) \quad \alpha_D = -r_1 \)

\( (ii) \quad \alpha_{DD} = -r_1(L_n(\alpha_{DD})) \)

\( (iii) \quad \alpha_{DL} = -r_1(\mathbb{M}_n) \)

\( (iv) \quad \alpha_E = -r_1(\mathbb{M}_n(\alpha_E)) \)

Hence to compare the solutions it is necessary to study the functions
L_n(\alpha), N_n(\alpha) and the constant M_n. It is easiest to consider the solutions relative to the DIRECT method since this will indicate the effect of removing the assumptions.

(i) Using both approximations gives \( \sigma_D = -r_1 \)

(ii) Removing the determinant assumption gives

\[
\sigma_{DD} = -r_1(L_n(\alpha_D))
\]

It is easy to show that

\[
\min_{\alpha} \{L_n(\alpha)\} = L_n(0) = 1+(n-1)^{-1}
\]

Similarly \( \sup_{\alpha} \{L_n(\alpha)\} = \lim_{\alpha \to 1} \{L_n(\alpha)\} \)

L'Hôpital's rule enables the limit to be calculated giving

\[
\sup_{\alpha} \{L_n(\alpha)\} = 1 + 3(n-1)^{-1}
\]

Hence \( |\sigma_{DD}| < |\sigma_D| \) and the difference between the estimates satisfies

\[
|\sigma_D - \sigma_{DD}| < \frac{3}{n-1} |\sigma_D|
\]

(iii) Removing the inverse approximation gives \( \sigma_{DI} = -r_1(M_n) \). The value of M depends on the observations \( X_1, \ldots, X_n \) but in general it is clear that \( M_n > 1 \) and for reasonably large n, \( M_n^{-1} \) is small. Hence \( |\sigma_{DI}| > |\sigma_D| \) and the difference between the estimates satisfies

\[
|\sigma_D - \sigma_{DI}| = \frac{x_1^2 + x_n^2}{X_1^2 + \ldots + X_n^2} |\sigma_D|
\]

(iv) Removing both approximations gives

\[
\sigma_F = -r_1(N_n(\alpha_F))
\]

The behaviour of \( N_n(\alpha) \) is more complicated but it can be shown that
\[
\max_{\alpha} \left( n \right) = n_0(n) = \frac{nC_0}{nC_0 - C_0 - (X_1^2 + X_n^2)}
\]

Hence \( n_0(\alpha) > 1 \) and also \( n_0(\alpha) - 1 = O(1/n) \). Hence

\[|\alpha_P| > |\alpha_D| \]
and the difference between the estimates is \( O(1/n) \). The order of this difference is not surprising since

\[
\frac{1}{n} \left( \frac{1}{\alpha^2 + 2} \right)^{1/n} \to 1 \text{ as } n \to \infty. \quad \text{This is valid for}
\]

\[|\alpha| < 1 \]
and also as \( \alpha \to \pm 1 \).

Hence the approximations used for the AR(1) model tend to bias \( \alpha \) towards the centre of the unit circle as suggested by previous empirical conclusions for other models.

It is interesting to note that the direct solution \( \alpha_D = -r_1 \) can lie outside the unit circle as shown by Chanda (1962). Also valid solutions produced by \( \alpha_D = -r_1 \) which are near the unit circle boundary can have corresponding exact solutions which lie outside the unit circle.

3.4 The MA(1) model

The effect of approximations on the MA(1) model has been studied in depth by E.J. Godolphin and J. De Gooijer (1982). Their results were based on a comparison of the DIRECT, DIRDET and EXACT methods. The empirical evidence presented strongly suggested that the three methods produced virtually identical solutions. The simulations were based on 50 observations with \( \beta = 2/3, \gamma = -0.9 \) and \( \gamma = -0.95 \). For these simulations no evidence was found to back up any of the three points made in Section 3.1.
An analytic comparison is not included since the details are similar to the ARMA(1,1) model. However the results of Godolphin and De Gooijer (1982) provide examples of situations where the exact estimate is both larger and smaller in modulus than the direct estimate. Hence it is unlikely that a constant bias towards zero can be found for the MA(1) model as it was for the AR(1).

### 3.5 The ARMA(1,1) model

Even for this simple model the difficulty of analysing the approximations is greatly increased from the AR(1) case. Hence the results presented here are less informative but give an indication of the order of the approximations involved.

**The DIRECT method**

The direct equations are

\[ \alpha_D = - \frac{y_1}{y_0} \]  
\[ \beta_D = \alpha_D^* + D_1^* D_2^* r_2^* \]

which follow directly from (2.1.13) and (2.1.16). The \( \beta \)-equation can be rewritten

\[ \beta_D = \frac{1 + \beta_D^2}{\beta_D} \]

\[ \frac{1}{d_{1,1}} \]  
\[ \prod_{k=1}^{m} d_{1,k} \]

**The DIRDET method**

The determinant of \( \Sigma \) for the ARMA (1,1) model is less straightforward than for the AR(1) case.
\[ \det \xi = \frac{(1-\alpha^2)(1-\beta^2)}{(1-\alpha\beta)^2 - (\alpha-\beta)^2 \beta^{2n}} \]

\((\det \xi)^{1/n} + 1\) as \(n \to \infty\) as long as either \(\alpha\) or \(\beta\) is less than 1 in modulus. However, if both \(\alpha\) and \(\beta\) tend to 1 in modulus then this may not be the case.

Equation (3.2.5) gives

\[ \frac{\delta}{\delta \theta} \left( \frac{X'\Sigma X - \frac{1}{n} \det \xi}{\det \xi} \right) = 0, \quad 0 \in (\alpha, \beta) \quad (3.5.3) \]

Using \(X'\Sigma X = C_0(1+\alpha^2)\theta + 2\alpha C_0\theta\) the \(\alpha\)-equation becomes

\[ \alpha_\theta = -\frac{\gamma_1}{\gamma_0} Q_n(\alpha_\theta, \theta) \quad (3.5.4) \]

where

\[ Q_n(\alpha, \beta) = \frac{2\alpha(1-\alpha q_n(\alpha, \beta))}{2\alpha - q_n(\alpha, \beta)(1+\alpha^2)} \]

and

\[ q_n(\alpha, \beta) = \frac{1}{n \det \xi} \frac{\delta}{\delta \alpha} (\det \xi) \]

Algebraically it is simplest to consider

\[ \alpha(1 - S_n(\alpha, \beta)) = -\frac{\gamma_1}{\gamma_0} \quad (3.5.5) \]

where

\[ 1 - S_n(\alpha, \beta) = [Q_n(\alpha, \beta)]^{-1} \]

The form of \(S_n(\alpha, \beta)\) is complicated but it can be shown that

\[ 0 \leq S_n(\alpha, \beta) \leq 6n(2n^2+1)^{-1} \]

and that this inequality holds whenever \(\alpha, \beta\) lie in the unit circle and also as \(\alpha \to \pm 1\) or \(\beta \to \pm 1\).

The \(\beta\)-equation is more difficult to study. Using (3.5.3) the \(\beta\)-equation becomes
\[ \beta_{DD} = \frac{(1+\beta)^2}{d_{1,1}} \left\{ \begin{array}{c} m \sum_{k=1}^{d} \sum_{r \in x} q_n(\alpha, \beta) \frac{X^TXX}{2} \end{array} \right\} \]  

(3.5.6)

where

\[ R_n(\alpha, \beta) = - \frac{(1+\beta^2) q_n(\alpha, \beta)}{2d_{1,1}} \]

and

\[ q_n(\alpha, \beta) = \frac{1}{n \det \Sigma} \frac{\delta}{\delta \beta} \]

Since the first term provides the DIRECT solution it is of interest to investigate the size of \( R_n(\alpha, \beta) \). The form of \( X'XX \) is a sum of squares and hence \( \frac{X'XX}{n} \) is positive and bounded. Also it can be shown that

\[ 0 < R_n(\alpha, \beta) < 4(2n-1)^{-1} \]

and this inequality is valid whenever \( \alpha \) and \( \beta \) have modules less than 1 or when \( \alpha \to \pm 1 \) or \( \beta \to \pm 1 \).

Comparing the solutions

Equations (3.5.1), (3.5.2), (3.5.5) and (3.5.6) show that the difference between the DIRECT and DIRDET equations is \( O(1/n) \). However since the methods are iterative this is not sufficient to prove that the final solutions have such a relationship. For fixed \( \alpha \) it is clear that for one iteration

\[ \beta_{DD} > \beta_D \text{ and } \beta_{DD} - \beta_D = O(1/n) \]

Also for fixed \( \beta \)
However these differences are valid for only one iteration, hence no conclusions can be drawn about bias in general.

To investigate the relationship between the solutions it would be necessary to study the response of the $\alpha$ and $\beta$ equations to small changes in $\alpha$ and $\beta$. If changes of the order $O(1/n)$ to $\alpha$ and $\beta$ cause changes of the same order in the $\alpha$ and $\beta$ equations then some conclusions could be drawn about the differences $|\alpha_{DD} - \alpha_D|$ and $|\beta_{D} - \beta_{DD}|$. This is beyond the scope of the thesis and some numerical results are presented instead at the end of this section.

For the ARMA$(1,1)$ case no analytic results can be found for suggesting that $|\alpha_{DD}|$ and $|\beta_{DD}|$ are larger than $|\alpha_D|$ and $|\beta_D|$ using this simple approach.

The EXACT and DIRINV methods

It has been shown above that including the determinant in the DIRECT estimation method produces equations which differ by $O(1/n)$ from the DIRECT equations. Also the bounds on $S_n(\alpha, \beta)$ and $R_n(\alpha, \beta)$ were obtained as either $\alpha \to \pm 1$ or $\beta \to \pm 1$. Hence for many cases the differences may be considerably smaller.

To consider either the DIRINV or the EXACT method it is necessary to see how closely $X'XX$ approximates $X'X^{-1}X$. Shaman's approximation is given by
\[
\Gamma = \begin{bmatrix}
\pi_0 & \cdots & \pi_{n-1} \\
\vdots & \ddots & \vdots \\
\pi_{n-1} & \cdots & \pi_0
\end{bmatrix}
\]

where \( \pi_0 = \frac{1 - 2\alpha \beta + \alpha^2}{1 - \beta^2} \), \( \pi_1 = \frac{(1 - \alpha \beta)(\alpha - \beta)}{1 - \beta^2} \), and \( \pi_k = (-\beta)^{k-1} \pi_1 \).

Hence \( I_{r,s} = (-\beta)^{s-r-1} \frac{(1 - \alpha \beta)(\alpha - \beta)}{1 - \beta^2} \) for \( s > r \).

Galbraith and Galbraith (1974) give an explicit form for \( \Gamma^{-1} \) including the formula

\[
(\Gamma^{-1})_{r,s} = (-\beta)^{s-r-1} \frac{(1 - \alpha \beta)(\alpha - \beta)}{1 - \beta^2} \left\{ \frac{(1 + t \beta^{2r-1})(1 + t \beta^2(n-s) + 1)}{1 - t^2 \beta^{2n}} \right\}
\]

where \( t = \frac{\alpha - \beta}{1 - \alpha \beta} \). Hence the accuracy of the approximation can be studied by looking at

\[
D = \frac{(1 + t \beta^{2r-1})(1 + t \beta^2(n-s) + 1)}{1 - t^2 \beta^{2n}} - 1
\]

For \( 0 \leq r \leq s \leq n \) it is clear that \( D = 0 \). However for small values of \( r \) and large values of \( s \) the value of \( D \) can become fairly large. This would suggest that the use of an approximate inverse will have more impact on the likelihood equations than the exclusion of the determinant term. This is difficult to study analytically but the numerical results in the following section tend to support this suggestion.
3.6 Comparisons of exact and approximate estimators

Overall it can be seen that analytic evidence of the effects of approximations is very hard to produce. Even for the MA(1) and ARMA(1,1) models it was necessary to resort to simulations to produce any concrete information.

The AR(1) model was the exception in that it could be shown that the DIRECT approach was biased towards zero compared to the exact ML estimator. A bound on the difference between the two estimators was also produced giving

\[ |\alpha_D - \alpha_E| \leq \frac{C_0 x_1^2 + x_n^2}{n C_0 - C_0 (x_1^2 + x_n^2)} |r_1| \]

These particular points would probably imply that the approximate estimator \( \alpha_D \) is at least as reliable as the exact estimator \( \alpha_E \) near the boundary of the unit circle. The problems of picking up values near and on the boundary have been described by Ansley and Newbold (1980) for MA(1) models. In the light of these problems a slight bias towards zero may well be a not undesirable property.

For the MA(1) example no empirical evidence was found in the simulations of Godolphin and De Gooijer (1982) to support any of the three points proposed in Section 3.1. However several other simulations have concentrated on the MA(1) model and have found relevant evidence. The situation is typically unclear with conflicting empirical evidence concerning all three points. It is fairly certain from the remarks in Section 3.4 that there is no constant bias term in the estimation equations for the MA(1) model of
the form described for the AR(1) model. But this is perhaps the only conclusion that can be drawn.

The ARMA(1,1) model is also difficult to investigate fully. In Section 3.5 it was shown that the approximate estimation equations differ by $O(1/n)$ from the exact equations. However the iterative nature of the equations makes it difficult to compare the eventual solutions. Comparisons of the first iterations suggested that $|\alpha_D| > |\alpha_E|$ and that $\beta_D > \beta_E$. However these results do not necessarily apply to the solutions. The results do show that even in the region close to the boundary of the unit circle the approximations made are still $O(1/n)$. This gives some indication that the approximate procedures should not be too different from the exact methods even for these difficult cases.

Finally what conclusions concerning the ARMA(1,1) model can be drawn from simulations? Table (3.5.1) shows the results of 20 simulations of 50 observations from an ARMA(1,1) model with parameters $\alpha = \frac{1}{2}$ and $\beta = -\frac{1}{2}$. The estimators $\alpha$ and $\beta$ with subscripts D, DD and E were defined previously. The observations were simulated using the NAG routine G05EWF. The DIRECT and DIRDET methods were implemented using a Newton-Raphson method and the EXACT procedure used was the NAG routine G13BEF. The EXACT procedure FMLAMS available at L.S.E. (University of London) gave identical results to G13BEF.

The DIRINV approach was not examined since in terms of complexity it approaches the EXACT method. Hence the simulations are based on comparisons of the EXACT procedure with the simpler approximate methods DIRECT and DIRDET.
### Table 3.5.1

<table>
<thead>
<tr>
<th>Sample</th>
<th>DIRDET</th>
<th>DIRECT</th>
<th>EXACT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha_{DD}$</td>
<td>$\beta_{DD}$</td>
<td>$\alpha_D$</td>
</tr>
<tr>
<td>1</td>
<td>0.52084</td>
<td>-0.59307</td>
<td>0.53241</td>
</tr>
<tr>
<td>2</td>
<td>0.54306</td>
<td>-0.31956</td>
<td>0.55372</td>
</tr>
<tr>
<td>3</td>
<td>0.33003</td>
<td>-0.48493</td>
<td>0.33799</td>
</tr>
<tr>
<td>4</td>
<td>0.79666</td>
<td>-0.08089</td>
<td>0.81113</td>
</tr>
<tr>
<td>5</td>
<td>0.76948</td>
<td>-0.32695</td>
<td>0.78415</td>
</tr>
<tr>
<td>6</td>
<td>0.74693</td>
<td>-0.09508</td>
<td>0.76039</td>
</tr>
<tr>
<td>7</td>
<td>0.61545</td>
<td>-0.13154</td>
<td>0.62734</td>
</tr>
<tr>
<td>8</td>
<td>0.18811</td>
<td>-0.55116</td>
<td>0.19137</td>
</tr>
<tr>
<td>9</td>
<td>0.44163</td>
<td>-0.49320</td>
<td>0.44998</td>
</tr>
<tr>
<td>10</td>
<td>0.59068</td>
<td>-0.40762</td>
<td>0.60110</td>
</tr>
<tr>
<td>11</td>
<td>0.50000</td>
<td>-0.56501</td>
<td>0.50979</td>
</tr>
<tr>
<td>12</td>
<td>0.40121</td>
<td>-0.56034</td>
<td>0.41148</td>
</tr>
<tr>
<td>13</td>
<td>0.56019</td>
<td>-0.60630</td>
<td>0.57235</td>
</tr>
<tr>
<td>14</td>
<td>0.36188</td>
<td>-0.48718</td>
<td>0.36641</td>
</tr>
<tr>
<td>15</td>
<td>0.39540</td>
<td>-0.50498</td>
<td>0.40381</td>
</tr>
<tr>
<td>16</td>
<td>0.49559</td>
<td>-0.69737</td>
<td>0.50719</td>
</tr>
<tr>
<td>17</td>
<td>0.63380</td>
<td>-0.48790</td>
<td>0.64642</td>
</tr>
<tr>
<td>18</td>
<td>0.59366</td>
<td>-0.63886</td>
<td>0.59979</td>
</tr>
<tr>
<td>19</td>
<td>0.71925</td>
<td>-0.32198</td>
<td>0.73206</td>
</tr>
<tr>
<td>20</td>
<td>0.47713</td>
<td>-0.60801</td>
<td>0.48837</td>
</tr>
</tbody>
</table>
The theoretical means and variances of the estimators are compared with the sample values below.

Table 3.5.2

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theoretical</td>
<td>α</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>β</td>
<td>-0.5</td>
</tr>
<tr>
<td></td>
<td>α_D</td>
<td>0.54436</td>
</tr>
<tr>
<td></td>
<td>β_D</td>
<td>-0.45743</td>
</tr>
<tr>
<td></td>
<td>α_DD</td>
<td>0.53405</td>
</tr>
<tr>
<td></td>
<td>β_DD</td>
<td>-0.44810</td>
</tr>
<tr>
<td></td>
<td>α_E</td>
<td>0.53880</td>
</tr>
<tr>
<td></td>
<td>β_E</td>
<td>-0.54424</td>
</tr>
</tbody>
</table>

Several other simulations of the ARMA(1,1) model were performed with sample sizes varying from 30 to 100 and values of (α,β) in the range (0.1,−0.1) to (0.9,−0.9). In the following discussion it is the overall results which are considered and not just those given in Table 3.5.1.
In all simulations the m.s.e. or sample variances matched the theoretical values reasonably well. In general the difference in m.s.e. values for the 3 methods was very small. The only occasions when there were significant differences occurred when the exact procedure picked up values at ±1 and hence had a larger sample variance. This is exemplified by samples 15 and 18 in Table 3.5.1. These results boost the sample variance of $\beta_E$ to 0.05003 as compared with 0.03409 for $\beta_D$. In the remaining cases the sample variances were usually smallest for the EXACT approach and largest for the DIRECT approach. However differences were generally of the order of $10^{-4}$. Hence there is some very slight indication that the m.s.e. for approximate procedures is larger although this appears to be of small importance.

The question of reliability of parameter estimates near the boundary of the unit circle is difficult to observe. The behaviour of samples 15 and 18 in Table 3.5.1 is also representative of the other simulations. When the EXACT procedure produces estimates near to ±1 the approximate methods give values considerably smaller in modulus. This behaviour only occurs regularly when the EXACT estimates are almost on the boundary of the unit circle. Hence it does not appear that this is due to a general bias towards zero. Perhaps it is only for parameter values virtually on this boundary that such a discrepancy occurs. For values of $|\beta_E|$ up to 0.98 close relationship has been found with $\beta_D$.

Certainly no constant bias towards zero exists for the ARMA(1,1) model as it does for the AR(1) model. However there are the problems outlined above where approximate estimates are considerably smaller.
than EXACT estimates near ± 1. This could be said to constitute bias and unreliability but the situation is not that straightforward. In all such cases the approximate values were closer to the simulated values. Also in each case the exact likelihood equation is extremely difficult to solve. This is shown in more detail in Chapter 5.

The simulations seem to show that the determinant approximation has less effect than the use of Shaman's approximate inverse. In Table 3.5.1 the \((\alpha_D, \beta_D)\) values are generally very close to the \((\alpha_{DD}, \beta_{DD})\) values. Hence including the determinant has had a small effect. The differences between \((\alpha_E, \beta_E)\) and \((\alpha_{DD}, \beta_{DD})\) are usually larger indicating that including the correct inverse has a greater effect.

In terms of speed the fastest procedure was DIRECT and the slowest EXACT. However differences in computer time are not particularly relevant and the approximate procedures were programmed in a straightforward way without attempting to save on computation.

The work of Section 3.5 showed that the EXACT equations differed from the DIRECT equations by \(O(n^{-1})\). However using the estimation equations iteratively results in larger discrepancies than this. In Table 3.5.1 sample 6 has \(\beta_E - \beta_D = -0.37053\). Clearly for a sample size of 50 this would not occur if the difference between solutions were \(O(n^{-1})\). Also there are the large differences near the boundary of the unit circle. However in general the agreement between the different estimates was good for sample sizes ranging from 30 upwards.
4.1 Introduction

The development of exact ML estimation techniques was briefly summarized in Chapter 1. As outlined the more recent work has concentrated on the calculation of the exact likelihood function and its derivatives. These have been used to develop efficient maximization procedures. To this end several Kalman filtering algorithms have been derived (Shea [1987], Ansley and Kohn [1985], Mélard [1984] and others).

There has also been interest in simpler procedures which are based on the fundamental ideas of ARMA models and which do not rely on the state variable approach. Such techniques have recently been proposed by Pham Dinh [1987], Porat and Friedlander [1986] and Wincek and Reinsel [1986]. Another such method is proposed in Section 4.2. This alternative procedure is similar in that it relies on several recursive equations, some of which have appeared in the previous papers. However it is a more basic approach and its calculations do not depend on either the periodogram of Pham Dinh or the innovations transform of Wincek and Reinsel.

The following Section continues the approach of T.W. Anderson [1975,1977] in that the likelihood and its derivatives are calculated with a view to using Newton-Raphson or Scoring methods for maximization. It is not the intention to amend the procedure to
produce a computationally efficient algorithm. Instead a simple approach is used which gives considerable simplifications when the method of Scoring is implemented.

In Section 4.3 various methods of implementing the approach of Section 4.2 are compared and some numerical results are presented. Some comparisons with other exact estimation procedures are made.

### 4.2 Computation of the likelihood function and its derivatives

The likelihood function is given by

\[ L = \frac{1}{\sqrt{2\pi \sigma^2}} \cdot \frac{1}{n/2} (\text{det} \Gamma)^{-1/n} \exp(-\frac{1}{2\sigma^2} X \cdot \Gamma^{-1} X) \]

In Section 1.2 it was shown that maximizing this function is equivalent to solving the likelihood equation (1.2.7) involving the expression

\[ L^* = (X \cdot \Gamma^{-1} X) (\text{det} \Gamma)^{1/n} \]

Hence in this section we concentrate on \( L^* \) and the resulting likelihood equations.

The likelihood function

The computation of \( L^* \) relies on the ability to formulate \( X \cdot \Gamma^{-1} X \) and \( \text{det} \Gamma \). Computationally efficient techniques for these calculations have been available for several years (Galbraith and Galbraith [1974], Dent [1977], Box and Ljung [1979]).

Hence the formulation of \( L^* \) is not considered here and the derivatives are considered in greater detail. However in considering the derivatives it is necessary to include the calculation of \( \Gamma^{-1} \).
Hence only the determinant calculation is not given explicitly.

The likelihood equations

The likelihood equation (1.2.7) gives

\[
(detr)^X (\frac{\delta}{\delta \theta} \Gamma^{-1}X + \frac{X \Gamma^{-1}X}{n} \frac{\delta}{\delta \theta} \det \Gamma) = \frac{\delta L^*}{\delta \theta} = 0
\]

where \( \theta \in \{\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q\} \). The derivatives of \( \Gamma^{-1} \) and \( \det \Gamma \) cannot be conveniently calculated and so the following substitutions are used.

\[
\frac{\delta \Gamma^{-1}}{\delta \theta} = -\Gamma^{-1} \frac{\delta \Gamma}{\delta \theta} \Gamma^{-1}
\]

\[
\frac{\delta \det \Gamma}{\delta \theta} = \det \Gamma \text{ tr}(\Gamma^{-1} \frac{\delta \Gamma}{\delta \theta})
\]

These have been used before to simplify the likelihood equations (T.W. Anderson [1977]). This gives the following equation

\[
\frac{X \Gamma^{-1}X}{n} \text{ tr}(\Gamma^{-1} \frac{\delta \Gamma}{\delta \theta}) - Y \frac{\delta \Gamma}{\delta \theta} Y = 0
\]

where \( Y = \Gamma^{-1}X \). Hence the likelihood equation can be computed by calculating \( \Gamma^{-1} \) and \( \frac{\delta \Gamma}{\delta \theta} \).

For (4.2.1) to be solved by an iterative technique such as Newton-Raphson it is necessary to compute its derivatives.

If \( \theta \in \{\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q\} \) the derivative of (4.2.1) w.r.t \( \theta \) is given by

\[
2Y \frac{\delta \Gamma}{\delta \theta} \Gamma^{-1} \frac{\delta \Gamma}{\delta \theta} Y - Y \frac{\delta^2 \Gamma}{\delta \theta^2} Y + \frac{X \Gamma^{-1}X}{n} \text{ tr}(\Gamma^{-1} \frac{\delta^2 \Gamma}{\delta \theta^2})
\]

\[
-1/n Y \frac{\delta \Gamma}{\delta \theta} \text{ tr}(\Gamma^{-1} \frac{\delta \Gamma}{\delta \theta}) - \frac{X \Gamma^{-1}X}{n} \text{ tr}(\Gamma^{-1} \frac{\delta \Gamma}{\delta \theta} \Gamma^{-1} \frac{\delta \Gamma}{\delta \theta})
\]

(4.2.2)
Hence the derivative of (4.2.1) can be computed by calculating $\Delta^{-1}$, $\frac{\partial}{\partial \delta} \Delta$, $\frac{\partial^2}{\partial \delta^2} \Delta$. In the rest of the section the calculation of these three basic expressions is considered.

Calculating $\Gamma^{-1}$

The method and notation of Galbraith and Galbraith (1974) is repeated except where stated. Their main result stated that

$$
\Gamma^{-1} = A_n (I_n - H(D + H' H))^{-1} A_n^{-1}
$$

(4.2.3)

where $A_n$ is lower triangular with $(i,j)$th element $a_{i,j}$ satisfying

$$
a_n = \begin{cases} 
0 & u < 0 \\
\alpha_u - \beta_j \alpha_{u-1} - \cdots - \beta_q \alpha_{u-q} & u > 0
\end{cases}
$$

(4.2.4)

It is assumed throughout that $\alpha_0 = 1$ and that $\alpha_u = 0$ for $u > p$. $H$ is calculated by the following recurrence relation on $K = (IH')^{-1}$. If $K = (k_{i,j})$ then

$$
k_{i,j} = \begin{cases} 
\delta_{i,j} & 1 \leq i,j \leq p+q \\
\alpha_{i-j-q} - \beta_i \kappa_{i-1,j} - \cdots - \beta_q \kappa_{i-q,j} & 1 \leq j \leq p, \quad p+q+1 \leq i \leq p+q+n \\
-\beta_i \kappa_{i-1,j} - \cdots - \beta_q \kappa_{i-q,j} & p+1 \leq j \leq p+q, \quad p+q+1 \leq i \leq p+q+n
\end{cases}
$$

(4.2.5)

$D$ is calculated from the identity

$$
D^{-1} = \begin{bmatrix} 
\Gamma & B_{pxq} \\
B_{pxq} & I_q 
\end{bmatrix}
$$

$B_{pxq}$ has $(i,j)$th element $b_{i-j-p+q}$ where the $b_k$ terms are calculated from the recurrence relation.
Throughout it is assumed that $\beta_0 = 1$ and $\beta_u = 0$ for $u > q$. The matrix $\Gamma_p$ is due to the notation of Box and Ljung [1979], Galbraith and Galbraith used $\mathbf{M}_p^{-1}$. To calculate $\Gamma_p$ the method of Box and Ljung is used. Consider the basic model

$$X_t + \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \ldots + \beta_q \varepsilon_{t-q}$$

Multiplying throughout by $X_{t-1}$ and taking expectations gives

$$X_{t-1} + \alpha_1 X_{t-2} + \ldots + \alpha_p X_{t-p} = \varepsilon_{t-1} + \beta_1 \varepsilon_{t-2} + \ldots + \beta_q \varepsilon_{t-q}$$

where $\varepsilon_t = E(X_t \varepsilon_t)$. The $y_k$ terms are calculated from the recurrence relation

$$y_{t-k} = \begin{cases} 0 & k > t \\ 1 & k = t \\ \beta_1 \gamma_{t-1} - \ldots - \alpha_p \gamma_{t-p} & t = k+1 \\ 0 & l > 0 \end{cases}$$

The paper by Box and Ljung confuses the derivation by using both $y^0 = 0$ and $y^1 = 1$. However this does not affect the result.

Solving (4.2.7) for $i = 0, 1, \ldots, p$ gives

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_p \end{bmatrix} = A^{-1} V$$

where $V = (v_1, \ldots, v_{p+1})^T$, $v_j = \sum_{i=0}^{p} \beta_i y_{i+j}$

and $A$ is given by

$$A = \begin{bmatrix} 1 \\ \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix} + \begin{bmatrix} 0 & \alpha_1 & \alpha_2 & \ldots & \alpha_p \\ 0 & 0 & \alpha_1 & \alpha_2 & \ldots & \alpha_p \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & 0 & \alpha_1 \\ 0 & 0 & \cdots & \cdots & 0 & 1 \end{bmatrix}$$
Using equations (4.2.4) to (4.2.10) $\Gamma^{-1}$ can be calculated from (4.2.3). A direct implementation of this approach will involve 5 sets of recurrence relations to compute 5 matrices, 2 matrix inversions of order $(p+1)^2$ and $(p+q)^2$ and some matrix arithmetic.

Calculating $\frac{\delta \Gamma}{\delta \theta}$

The autocovariance generating function $\Gamma(z)$ has a basic role in calculating these derivatives. The definition of $\Gamma(z)$ is

$$\Gamma(z) = \frac{\beta(z)\beta(z^{-1})}{\alpha(z)\alpha(z^{-1})}$$

where

$$\alpha(z) = 1 + \alpha_1 z + \ldots + \alpha_p z^p$$
$$\beta(z) = 1 + \beta_1 z + \ldots + \beta_q z^q$$

An equivalent definition is

$$\Gamma(z) = \gamma_0 + \sum_{i=1}^p \gamma_i (z^{-1} + z^{-i})$$

Equating the two formulations gives

$$\left( g_0 + \sum_{k=1}^p g_k (z^{k} - z^{-k}) \right) \left( \gamma_0 + \sum_{i=1}^p \gamma_i (z^{-1} + z^{-i}) \right) = b_0 + \sum_{j=1}^p b_j (z^{j} + z^{-j})$$

Equating coefficients in (4.2.11) gives

$$g_k \gamma_n + \sum_{i=1}^p g_k (\gamma_{n-k} + \gamma_{n+k}) = b_n \quad n = 0, 1, \ldots$$
This can be rewritten to give a recurrence relation

\[ Y_{n+p} = \frac{-1}{\xi_p^p} \left( \xi_0 Y_n + \xi_2 Y_{n-p} - b_n^{\frac{k}{l}} g_k(Y_{n-k} + Y_{n+k}) \right) \quad (4.2.12) \]

Note that \( \xi_p = \alpha_p \neq 0 \).

The relation (4.2.12) can be used to compute \( Y_{p+k} \), \( k \geq 1 \) in terms of \( Y_0, \ldots, Y_p \) which are available from (4.2.9). However the relation given below is simpler and provides an easier method of calculating \( Y_{q+k} \), \( k \geq 1 \)

\[ Y_n = - \sum_{i=1}^{p} \alpha_i Y_{n-i} \quad n \geq q+1 \quad (4.2.13) \]

Hence for \( p > q \) it is convenient to use (4.2.9) and (4.2.13) to calculate \( Y_1 \), \( i \geq 0 \). For \( q > p \) it is necessary to use (4.2.9) and (4.2.12) to calculate \( Y_0, Y_1, \ldots, Y_q \) and then use (4.2.13) to calculate \( Y_{q+k} \), \( k \geq 1 \).

To calculate \( \delta \Gamma/\delta \theta \) the relations involved in calculating \( \Gamma \) are differentiated. Firstly consider the calculation of \( \delta \Gamma/\delta \alpha_j \). In the following work the operator \( D \) represents \( \delta /\delta \alpha_j \).

Differentiating (4.2.7) gives

\[ Y_{i-j} + \alpha_1 Y_{i-1} + \ldots + \alpha_p Y_{i-p} = D Y_{i-1} + \beta_1 D Y_{i-1} + \ldots + \beta_q D Y_{i-q} \quad (4.2.14) \]

Differentiating (4.2.8) gives

\[ D Y_{t-k} = \begin{cases} 0 & k \geq t \\ -Y_{t-j} - \alpha_1 Y_{t-1} - \ldots - \alpha_p Y_{t-p} & t = k+1, \ 1 \geq 0 \end{cases} \quad (4.2.15) \]

Solving (4.2.14) for \( i = 0, 1, \ldots, p \) gives

\[ \begin{bmatrix} D Y_0 \\ \vdots \\ D Y_p \end{bmatrix} = A^{-1} \begin{bmatrix} D Y_1 - Y_{-j} \\ \vdots \\ D Y_{p+1} - Y_{p-j} \end{bmatrix} \quad (4.2.16) \]
where \(v_1', \ldots, v_{p+1}'\) and \(A\) are defined in (4.2.10) and \(Dv_{i}', i=1, \ldots, p+1\) are calculated from (4.2.16). The other derivatives are available recursively from either (4.2.12) or (4.2.13). Differentiating (4.2.12) gives for \(n = 1, 2, \ldots\):

\[
Dv_{n+p} = -\frac{1}{g_p} (2\alpha_j Y_n + g_q DV_n + g_p DV_{n-p})
\]

\[
+ \sum_{k=1}^{p-1} (\alpha_{k+j} + \alpha_{j-k}) (\gamma_{n-k} + \gamma_{n+k})
\]

\[
+ \prod_{k=1}^{p} \gamma_k (DY_{n-k} + \gamma_{n+k})
\]

Differentiating (4.2.13) gives:

\[
DY_n = v_j - \sum_{i=1}^{p} \alpha_i DY_{n-i}, \quad n \geq q+1
\]

Using equations (4.2.16), (4.2.17) and (4.2.18) \(\delta f / \delta \alpha_j\) can be calculated. As described above (4.2.17) is only necessary for the calculation of \(DY_{p+1}' \ldots, DY_q\). Thereafter (4.2.18) is used. A direct implementation of this approach will involve 3 sets of recursive relations and a matrix inversion of order \((p+1)^2\).

Secondly consider \(\delta f / \delta \beta_j\), where the operator \(D\) now represents \(\delta f / \delta \beta_j\). Differentiating (4.2.7) gives for \(i \geq 0\):

\[
DY_{1} + \alpha_1 DY_{i-1} \ldots + \alpha_p DY_{i-p} = \gamma_{j-i} + \gamma_{j-p} + \gamma_{j-q} + \gamma_{j-q-1}
\]

Differentiating (4.2.8) gives:

\[
DY_{i-k} = \begin{cases} 
0 & \text{if } 0 \leq k \leq t \\
\delta_{1,j} - \alpha_1 DY_{i-1} \ldots - \alpha_p v_{i-p} & \text{if } t = k+1 \\
0 & \text{if } t > 0
\end{cases}
\]

Solving (4.2.19) for \(i = 0, 1, \ldots, p\) gives
The other derivatives are available recursively from either (4.2.12) or (4.2.13). Differentiating (4.2.12) gives

\[
D_Y^{n+p} = - \frac{1}{\varepsilon_p} \left( \varepsilon_0 D_Y^n + \varepsilon_p D_Y^{n-p} - \beta_{n-j} \beta_j \right) - p^i \sum_{k=1}^{p-1} \varepsilon_k \left( D_Y^{n-k} D_Y^{n+k} \right) 
\]  

(4.2.22)

Differentiating (4.2.13) gives

\[
D_Y^n = - \frac{p-1}{\varepsilon_p} \alpha_i D_Y^{n-i} \quad \text{for } n > q+1
\]  

(4.2.23)

As for \( \delta T/\delta \alpha_j \), the calculation of \( \delta T/\delta \beta_j \) relies on 3 recursive relations and a matrix inversion.

Calculating \( \delta^2 \Gamma \)/\( \delta \theta \delta \theta \)

These second derivatives are obtained in exactly the same way as the first derivatives. Namely by differentiating (4.2.7), (4.2.12) and (4.2.13) twice. The results are summarized below.

(a) \( \theta = \alpha_k, \theta = \alpha_j, D_2 = \delta^2/\delta \theta \delta \theta, D_k = \delta/\delta \alpha_k, D_j = \delta/\delta \alpha_j \)

\[
\left[ \begin{array}{c}
D_2 \gamma_0 \\
D_2 \gamma_p \\
\end{array} \right] = A^{-1} \left[ \begin{array}{c}
D_2 \gamma_1 - D_k \gamma_{j-k} - D_j \gamma_{k-} \\
D_2 \gamma_{p+1} - D_k \gamma_{p-j} - D_j \gamma_{p-k} \\
\end{array} \right]
\]

\[
D_2 \gamma_{t-k} = \left\{ \begin{array}{ll}
0 & \text{if } 0 < t \\
-D_j \gamma_{t-k} - D_k \gamma_{t-j} - \alpha_1 D_2 \gamma_{t-1} - \ldots - \alpha_1 D_2 \gamma_{t-1} - p D_2 \gamma_{t-1} & \text{if } t = k+1
\end{array} \right.
\]
\[ D_{2}Y_{n+p} = \frac{1}{z_{p}} \begin{vmatrix} (2\alpha (j-k)Y_{n} + 2\alpha \Delta_{j,k}Y_{n} + 2\alpha \Delta_{j,k}Y_{n} \\
+ \Delta_{0}Y_{n} + \Delta_{p}D_{2}Y_{n-p} + E_{n} \end{vmatrix} \]

\[ E_{n} = \sum_{i=1}^{p} \left( \delta(j+1-k) + \delta(j-1-k) \right) (Y_{n-i} + Y_{n+i}) \]

\[ + \sum_{i=1}^{p} \left( \alpha_{i} + \alpha_{j-1} \right) (D_{k}Y_{n-i} + D_{k}Y_{n+i}) \]

\[ + \sum_{i=1}^{p} \left( \alpha_{k-i} + \alpha_{k+1} \right) (D_{j}Y_{n-i} + D_{j}Y_{n+i}) \]

\[ + \sum_{i=1}^{p} \Delta_{i} \frac{D_{2}Y_{n-i} + D_{2}Y_{n+i}}{n > 1} \]

\[ D_{2}Y_{n} = -D_{k}Y_{n-j} - D_{j}Y_{n-k} - \sum_{i=1}^{p} \alpha_{i} D_{2}^{2}Y_{n-i} \quad n > q+1 \]

(b) \[ \emptyset = \alpha_{k}, \quad \theta = \beta_{j}, \quad D_{2} = \delta/\delta_{k} \delta_{j}, \quad D_{k} = \delta/\delta_{k}, \quad D_{j} = \delta/\delta_{j} \]

\[ \begin{bmatrix} D^{2}Y_{0} \\
\vdots \\
D^{2}Y_{p} \end{bmatrix} = A^{-1} \begin{bmatrix} D^{2}Y_{1} - D_{j}Y_{-k} \\
\vdots \\
D^{2}Y_{p+1} - D_{j}Y_{p-k} \end{bmatrix} \]

\[ D_{2}Y_{t-k} = \begin{cases} 0 & t = 0 \\
-D_{j}Y_{1-k} - \alpha_{l} D_{2}^{2}Y_{1-1} - \ldots - \alpha_{p} D_{2}^{2}Y_{1-p} & t = k+1, l > 0 \end{cases} \]

\[ D_{2}Y_{n+p} = \frac{1}{z_{p}} \begin{vmatrix} (2\alpha \Delta_{j,k}Y_{n} + \delta^{2} \left( \alpha_{k-1} + \alpha_{k+1} \right) (D_{j}Y_{n-i} + D_{j}Y_{n+i}) \\
+ \delta_{0}D^{2}Y_{n} + \delta_{p}D^{2}Y_{n-p} \\
+ \sum_{i=1}^{p} \Delta_{i} \frac{D^{2}Y_{n-i} + D^{2}Y_{n+i}}{n > 1} \end{vmatrix} \]

\[ D_{2}Y_{n} = -D_{j}Y_{n-k} - \sum_{i=1}^{p} \alpha_{i} D_{2}^{2}Y_{n-i} \quad n > q+1 \]
(c) $\theta = \beta_k$, $\theta = \beta_j$, $D = \delta/\delta \beta_k \delta \beta_j$, $D_j = \delta/\delta \beta_j$, $D_k = \delta/\delta \beta_k$

$$
\begin{bmatrix}
D^2 \gamma_0 \\
D^2 \gamma_p
\end{bmatrix} = A^{-1}
\begin{bmatrix}
D^2 \gamma_1 \\
D^2 \gamma_{p+1}
\end{bmatrix}
$$

$D_{2t} \gamma_{t-k} = \begin{cases} 
0 & t = 0 \\
-\alpha_1 D_{2} \gamma_{1-t-1} - \ldots - \alpha_p D_{2} \gamma_{1-p} & t = k+1, 1 > 0
\end{cases}$

$D_2 \gamma_{n+p} = \frac{1}{\varepsilon_p} \langle \varepsilon_0 D_2 \gamma_{n} - \delta(n+j-k) - \delta(j-n-k) + \varepsilon_p D_2 \gamma_{n-p} \\
+ \sum_{i=1}^{p-1} \varepsilon_1 (D_2 \gamma_{n-i} + D_2 \gamma_{n+p}) \rangle, \ n > 1$

$D_2 \gamma_{n} = - \sum_{i=1}^{p} \alpha_1 D_2 \gamma_{n-1}, \ n > q+1$

4.3 Solutions of the likelihood equations

As described in Section 4.2 the likelihood equation (4.2.1) and its derivative can be formulated via a set of recursive relations and matrix inversions of order $(p+1)^2$ and $(p+q)^2$. Hence a Newton-Raphson scheme can certainly be used.

However the computations necessary to produce $\delta^2 / \delta \theta \delta \theta$ are of a complicated if basically simple nature. To avoid such details 2 alternatives have been investigated and are described below.

(a) Approximating the likelihood derivatives

Using the assumptions discussed in Chapter 3 $L^t$ can be approximated by $M^t = X' L X$. The derivatives of $M^t$ depend only on the calculation of $\delta L / \delta \theta$ and $\delta^2 L / \delta \theta \delta \theta$ and this is a much simpler calculation than the corresponding exact calculations. The derivatives of $M^t$ can be
used in the following scheme.

\[
\delta L^* = \frac{(\det \Gamma)^{1/n}}{\delta \theta} \left(-Y_1 \delta \Gamma + \frac{X \Gamma^{-1} X}{n} \operatorname{tr}(\Gamma^{-1} \delta \Gamma)\right)
\]

Hence using the approximation \((\det \Gamma)^{1/n} = 1\) the likelihood equation (4.2.1) can be approximated by \(X' \frac{\delta \Sigma}{\delta \theta} X\) and the derivative (4.2.2) can be approximated by \(X' \frac{\delta^2 L}{\delta \theta \delta \theta} X\). This approach gives considerable simplification since an iterative scheme now uses \(X' \frac{\delta^2 L}{\delta \theta \delta \theta} X\) instead of (4.2.2) to solve (4.2.1). However its use is essentially conjectural since \(M^* \neq L^*\) does not necessarily imply that \(\frac{\delta^2 M^*}{\delta \theta \delta \theta} = \frac{\delta^2 L^*}{\delta \theta \delta \theta}\). In fact the simulations discussed below for the ARMA(1,1) model show that the method is unpredictable and less efficient than the methods of Scoring and Newton Raphson.

(b) Scoring

The work of T.W. Anderson [1977] clearly shows the advantages of using the method of Scoring for his approximate likelihood equations. The same simplifications are produced for the exact likelihood equation. Using the result

\[
P'MP = \operatorname{tr}(MPP')
\]

the likelihood equation's derivative (4.2.2) can be rewritten

\[
\operatorname{tr}(\Gamma^{-1} \frac{\delta \Sigma}{\delta \theta} \Gamma^{-1} \Gamma^{-1} XX') - \operatorname{tr}(\Gamma^{-1} \frac{\delta^2 \Gamma}{\delta \theta \delta \theta} \Gamma^{-1} XX') + \operatorname{tr}(\Gamma^{-1} \frac{\delta \Gamma}{\delta \theta} \Gamma^{-1} \Gamma^{-1} XX') \\
- \frac{1}{n} \operatorname{tr}(\Gamma^{-1} \frac{\delta \Gamma}{\delta \theta} \Gamma^{-1} XX') \operatorname{tr}(\Gamma^{-1} \frac{\delta^2 \Gamma}{\delta \theta \delta \theta} \Gamma^{-1} XX') \operatorname{tr}(\Gamma^{-1} \frac{\delta \Gamma}{\delta \theta} \Gamma^{-1} \Gamma^{-1} \Gamma^{-1} XX') \\
+ \frac{\operatorname{tr}(\Gamma^{-1} XX')}{n} \operatorname{tr}(\Gamma^{-1} \frac{\delta^2 \Gamma}{\delta \theta \delta \theta} \Gamma^{-1} XX')
\]
Taking expectations gives

\[ \text{tr}(\Gamma^{-1} \frac{\delta \Gamma}{\delta \theta} \Gamma^{-1} \frac{\delta \Gamma}{\delta \theta}) -\frac{1}{n} \text{tr}(\Gamma^{-1} \frac{\delta \Gamma}{\delta \theta}) \text{tr}(\Gamma^{-1} \frac{\delta \Gamma}{\delta \theta}) \]  

(4.3.1)

since \( E(X'X) = \Gamma \). This is a considerable saving on the Newton-Raphson method since the formulation of (4.3.1) requires no further terms than those calculated already for the likelihood equation (4.2.1). Hence the iterative process can be implemented using only the calculations required for the likelihood.

Comparing methods of implementation

Consider the following four methods for solving the likelihood equation (4.2.1).

(a) FD - A Newton-Raphson approach using finite differences to calculate the derivatives.

(b) AD - A Newton-Raphson approach using the approximate derivatives of Section 4.3(a).

(c) NR - The Full Newton-Raphson approach derived in Section 4.2.

(d) SC - The method of Scoring derived in Section 4.3(b).

It can very quickly be shown that the AD method is unreliable. 20 simulations of an ARMA(1,1) model produced the results in Table 4.3.1. The same simulations are used here as in Table 3.5.1. All four methods when convergent agreed with the exact solutions given in Table 3.5.1 except for small variations due to differences in convergence. However for samples 15 and 18 convergence was not reached for any of the methods.
Table 4.3.1

<table>
<thead>
<tr>
<th>Sample</th>
<th>Number of Iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FD</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>4</td>
</tr>
<tr>
<td>14</td>
<td>5</td>
</tr>
<tr>
<td>15</td>
<td>-</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td>18</td>
<td>-</td>
</tr>
<tr>
<td>19</td>
<td>4</td>
</tr>
<tr>
<td>20</td>
<td>3</td>
</tr>
</tbody>
</table>

As would be expected the differences between the methods were in speed of convergence rather than in the actual parameter estimates. The AD method is clearly inferior with sample 6 requiring 124 iterations compared to the 4 iterations required by the FD method. In general twice as many iterations was not unusual.

In terms of computer time the method of Scoring is more economical with the finite difference and Newton-Raphson methods fairly similar. Convergence properties for all three methods are
similar and there seems no evidence to prefer the full Newton-Raphson method to the method of Scoring. However simulations have only been considered in detail for the ARMA(1,1) model.

The lack of convergence for samples 15 and 18 serves to emphasise the problems of parameter values on the boundary of the unit circle. In Chapter 5 these samples are considered in detail and the lack of convergence is investigated. It is suggested that this is not a drawback of these iterative methods but an indication of the difficulty in finding a true solution.
5.1 Introduction

In Chapter 5 the estimation techniques considered in Chapters 2, 3 and 4 are discussed further. In Chapter 2 results were produced which indicate that the two large sample procedures due to Godolphin (1984) and Walker (1962) may be asymptotically equivalent. Both procedures rely on the data transformation from the observations \(X_1, \ldots, X_n\) to the sample serial correlations \(r_k, k=1, \ldots, m\) where \(m < n-1\). The DIRECT approach of Godolphin depends on two common assumptions, namely neglecting "end effects" and employing Shaman's approximate inverse. In Section 5.2 the effect of these approximations is considered by comparing the likelihood contours of the DIRECT method with the corresponding EXACT likelihood contours.

In Chapter 3 it was shown that an analytic comparison of the DIRECT and EXACT methods is difficult even for the simplest of models. The estimation equations were shown to differ by \(O(1/n)\) but the relationship does not always apply to the parameter estimates except for the AR(1) model. The empirical studies mentioned in Section 1.1 suggested larger M.S.E. for approximate estimates and also possible bias. This bias has been identified for the AR(1) model but the evidence for more general models is not conclusive. However there is certainly a marked difference in estimates when the EXACT solution lies on the boundary of the unit circle. This characteristic of
approximate procedures has been mentioned in the literature on several occasions and has usually been suggested as evidence that approximate procedures are unsatisfactory or unreliable. Samples 15 and 18 in Table 3.5.1 are further examples of this behaviour for the DIRECT approach in particular. Closer study of these cases in Section 5.2 suggests that in these cases the approximate solutions may be preferable.

In Chapter 4 a new method for ML estimation was developed. The procedure is iterative and the methods of Newton Raphson and Scoring were outlined in Section 4.3. As mentioned in Section 3.5 any convergent parameter estimates produced by this approach exactly coincided with the results of G13BEF in Table 3.5.1. However for samples 15 and 18 the procedure failed to converge, irrespective of which iterative scheme was used. These results suggest further evidence that values on the boundary of the unit circle can cause unusual problems for the estimation routines.

It is interesting that a relaxation of the convergence criterion for the exact procedure of Chapter 4 resulted in convergence for samples 15 and 18. The original criterion stated that

\[
|\alpha_n - \alpha_{n-1}| < \epsilon \\
|\beta_n - \beta_{n-1}| < \epsilon \\
|\frac{\delta L}{\delta \alpha}| < \epsilon, \quad |\frac{\delta L}{\delta \beta}| < \epsilon
\]

Hence as well as insisting that the \(n\) iteration was \(O(\epsilon)\) the \(\alpha\) and \(\beta\) equations were also required to be small. If this second criterion is removed then convergence is achieved as below.
Sample 15:  \( \alpha = 0.32195 \)
\( \beta = -0.97388 \)
No. of iterations = 191

Sample 18:  \( \alpha = 0.61890 \)
\( \beta = -0.99450 \)
No. of iterations = 30

For these samples the likelihoods are very flat in a region close to \( \beta = -1 \). Hence each iteration moves only a small distance and convergence can be achieved without \( \frac{\delta L}{\delta \alpha} = 0 \) being satisfied. The \( \alpha \) and \( \beta \) equations were evaluated at the above points and in both cases only one equation was approximately zero. This behaviour is further illustrated in Section 5.2.

An arithmetic difficulty with the procedure of Chapter 4 is the appearance of the term \( \delta_p^{-1} = \alpha_p^{-1} \) in the recursive equations. If \( \alpha \rightarrow 0 \) then it may be necessary to escape from the routine to prevent exploding parameter estimates. However this difficulty has not arisen in any of the simulations performed, even for values of \( \alpha \) in \((0.01, 0.1)\).

As discussed in Chapter 3 there is a good overall match between the DIRECT and EXACT procedures for most case, both in terms of parameter estimates and M.S.E.. These results are valid for more general models as well as the simple ARMA(1,1) case. This is illustrated in Tables 5.1.1 to 5.1.4 for the ARMA(1,1), ARMA(1,2), ARMA(2,1) and ARMA(2,2) models. In all cases the EXACT results were produced by G13BBF and replicated by the procedure of Chapter 4.
Table 5.1.1

20 simulations of the ARMA(1,1) model with $\alpha = -\frac{1}{4}$ and $\beta = \frac{1}{4}$ are tabulated below. Sample size = 50.

<table>
<thead>
<tr>
<th>Sample No.</th>
<th>EXACT estimates</th>
<th>DIRECT estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\alpha}_E$</td>
<td>$\hat{\beta}_E$</td>
</tr>
<tr>
<td>1</td>
<td>-0.35196</td>
<td>0.59729</td>
</tr>
<tr>
<td>2</td>
<td>-0.25541</td>
<td>0.72053</td>
</tr>
<tr>
<td>3</td>
<td>-0.50993</td>
<td>0.51142</td>
</tr>
<tr>
<td>4</td>
<td>-0.64178</td>
<td>0.59990</td>
</tr>
<tr>
<td>5</td>
<td>-0.52733</td>
<td>0.63114</td>
</tr>
<tr>
<td>6</td>
<td>-0.43059</td>
<td>0.45711</td>
</tr>
<tr>
<td>7</td>
<td>-0.35217</td>
<td>0.65906</td>
</tr>
<tr>
<td>8</td>
<td>-0.49140</td>
<td>0.46333</td>
</tr>
<tr>
<td>9</td>
<td>-0.55220</td>
<td>0.65493</td>
</tr>
<tr>
<td>10</td>
<td>-0.85516</td>
<td>0.00412</td>
</tr>
<tr>
<td>11</td>
<td>-0.06968</td>
<td>0.75698</td>
</tr>
<tr>
<td>12</td>
<td>-0.58918</td>
<td>0.65336</td>
</tr>
<tr>
<td>13</td>
<td>-0.58359</td>
<td>0.59412</td>
</tr>
<tr>
<td>14</td>
<td>-0.60538</td>
<td>0.81862</td>
</tr>
<tr>
<td>15</td>
<td>-0.58396</td>
<td>0.53243</td>
</tr>
<tr>
<td>16</td>
<td>-0.16517</td>
<td>0.52605</td>
</tr>
<tr>
<td>17</td>
<td>-0.63145</td>
<td>0.40546</td>
</tr>
<tr>
<td>18</td>
<td>-0.61216</td>
<td>0.43592</td>
</tr>
<tr>
<td>19</td>
<td>-0.69182</td>
<td>0.30636</td>
</tr>
<tr>
<td>20</td>
<td>-0.26029</td>
<td>0.44493</td>
</tr>
</tbody>
</table>

$\bar{\alpha}_E = -0.48196$  $\bar{\alpha}_D = -0.45855$

$\bar{\beta}_E = 0.53880$  $\bar{\beta}_D = 0.51398$
The sample variances are:

\[ \alpha_E = 0.044141 \]
\[ \beta_E = 0.032036 \]
\[ \alpha_D = 0.045572 \]
\[ \beta_D = 0.035326 \]

The theoretical variances are 0.058594.

Table 5.1.2

8 simulations of the ARMA(1,2) model with \( \alpha = 0.7, \beta_1 = -0.5, \beta_2 = 0.5 \) are tabulated below. Sample size = 50

<table>
<thead>
<tr>
<th>Sample no.</th>
<th>EXACT estimates</th>
<th>DIRECT estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{\alpha} )</td>
<td>( \hat{\beta}_1 )</td>
</tr>
<tr>
<td>1</td>
<td>0.6983</td>
<td>-0.5093</td>
</tr>
<tr>
<td>2</td>
<td>0.7210</td>
<td>-0.4891</td>
</tr>
<tr>
<td>3</td>
<td>0.7103</td>
<td>-0.5216</td>
</tr>
<tr>
<td>4</td>
<td>0.8294</td>
<td>-0.3186</td>
</tr>
<tr>
<td>5</td>
<td>0.7092</td>
<td>-0.6531</td>
</tr>
<tr>
<td>6</td>
<td>0.8002</td>
<td>-0.5407</td>
</tr>
<tr>
<td>7</td>
<td>0.8135</td>
<td>-0.4883</td>
</tr>
<tr>
<td>8</td>
<td>0.7629</td>
<td>-0.5092</td>
</tr>
<tr>
<td><strong>AVERAGE VALUES</strong></td>
<td><strong>0.7556</strong></td>
<td><strong>0.5037</strong></td>
</tr>
</tbody>
</table>
TABLE 5.1.3

4 simulations of the ARMA(2,1) model with $\alpha_1 = -0.5$, $\alpha_2 = 0.5$, $\beta = 0.7$. Sample size = 50.

<table>
<thead>
<tr>
<th>Sample No.</th>
<th>EXACT estimates</th>
<th>DIRECT estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\alpha}_1$</td>
<td>$\hat{\alpha}_2$</td>
</tr>
<tr>
<td>1</td>
<td>-0.5219</td>
<td>0.5186</td>
</tr>
<tr>
<td>2</td>
<td>-0.4683</td>
<td>0.5401</td>
</tr>
<tr>
<td>3</td>
<td>-0.4120</td>
<td>0.5340</td>
</tr>
<tr>
<td>4</td>
<td>-0.6428</td>
<td>0.5479</td>
</tr>
<tr>
<td>AVERAGE VALUES</td>
<td>-0.5113</td>
<td>0.5652</td>
</tr>
</tbody>
</table>
8 simulations of the ARMA(2,2) model with $\alpha_1 = -0.5$, $\alpha_2 = 0.5$, $\beta_1 = -0.1$, $\beta_2 = 0.3$. Sample size = 50.

<table>
<thead>
<tr>
<th>Sample No</th>
<th>EXACT estimates</th>
<th>DIRECT estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\alpha}_1$</td>
<td>$\hat{\alpha}_2$</td>
</tr>
<tr>
<td>1</td>
<td>-0.5293</td>
<td>0.6318</td>
</tr>
<tr>
<td>2</td>
<td>-0.6287</td>
<td>0.6602</td>
</tr>
<tr>
<td>3</td>
<td>-0.7740</td>
<td>0.5198</td>
</tr>
<tr>
<td>4</td>
<td>-0.4503</td>
<td>0.4320</td>
</tr>
<tr>
<td>5</td>
<td>-0.5604</td>
<td>0.5286</td>
</tr>
<tr>
<td>6</td>
<td>-0.6003</td>
<td>0.4293</td>
</tr>
<tr>
<td>7</td>
<td>-0.6419</td>
<td>0.5064</td>
</tr>
<tr>
<td>8</td>
<td>-0.4602</td>
<td>0.5981</td>
</tr>
<tr>
<td>AVERAGE</td>
<td>-0.5806</td>
<td>0.5383</td>
</tr>
</tbody>
</table>

5.2 Simulation results

The results given in Tables 5.1.1, 5.1.2, 5.1.3, 5.1.4 are unsurprising, showing an agreement between the large sample DIRECT procedure and the EXACT procedure to be expected with a sample size of 50. It is the comparison of methods for small or unusual samples that is of interest. The question posed by Samples 15 and 18 is a difficult one. Should estimates for $|\beta| = 1$ be accepted when some
exact procedures fail to converge and approximate procedures give totally different answers? It is difficult to recognize a true ML solution in these cases since convergence may be due to numerical difficulties. Hence it is suggested that it may be dangerous to accept these particular estimates. The question is considered in detail below.

Consider Sample 1 of Table 3.5.1. This is a representative sample in the sense that both approximate and exact estimates are similar. For this sample the likelihood function is described by a contour diagram in Figure 5.2.1. The likelihood function plotted is

$$\frac{1}{n} (x^T \Gamma^{-1} x) (\det \Gamma)^{1/n}$$

and approximate likelihood values are given on the contours. The contour diagrams for Samples 2-14, 16-17, 19-20 are all of the same form.

The form of Figure 5.2.1 is that of a valley centred on the exact solution at \((\hat{a}, \hat{b}) = (0.53496, -0.61398)\). For likelihoods of this form maximization is straightforward by any approach.

From Section 4.2 the \(a\) and \(b\)-equations can also be calculated over the same range. For Sample 1 the lines \(\frac{\partial L}{\partial a} = 0\) and \(\frac{\partial L}{\partial b} = 0\) are given in Figure 5.2.2. A clear intersection of the two lines at the exact solution is apparent corresponding to the form of the likelihood in Figure 5.2.1. Again this is representative of the plots for Samples 2-14, 16-17, 19-20.

For these Samples the corresponding likelihood contours and plots of the \(a\) and \(b\)-equations can be produced for the DIRECT approach. Here the likelihood function is
Figure 5.2.1
Figure 5.2.2

\[ \frac{\delta L}{\delta \alpha} = 0 \]

\[ \frac{\delta L}{\delta \beta} = 0 \]
\[
\frac{1}{n} \mathbf{X}' \mathbf{X}
\]

and its derivatives give the \( \alpha \) and \( \beta \)-equations. In all cases the results were of the same form as that illustrated by Figure 5.2.1 and 5.2.2. Hence the agreement in parameter estimates.

For Samples 15 and 18 the situation is substantially different. For Sample 15 the EXACT likelihood contours are displayed in Figure 5.2.3. On such a scale it is difficult to pick out any obvious minima, but the behaviour is clearly a decreasing gradient towards \( \beta = -1 \), where the EXACT solution is \((\hat{\alpha}, \hat{\beta}) = (0.29735, -1.05542)\). The simple "valley" contours of Sample 1 are no longer present. The behaviour of the likelihood near to \( \beta = -1 \) is problematical and is described on a magnified scale in Figure 5.2.7.

The \( \alpha \) and \( \beta \)-equations for Sample 15 are plotted in Figure 5.2.4. Again the clear intersection of Sample 1 is missing here and there is apparently no real minimum. However the behaviour of the equations near \( \beta = -1 \) cannot be seen with this scale and a closer inspection is given in Figure 5.2.8.

For Sample 15 the DIRECT approach resulted in completely different parameter estimates, \((0.40381, -0.51434)\) as opposed to \((0.29735, -1.05542)\). The corresponding likelihood contours and \( \alpha \) and \( \beta \)-equations are given in Figures 5.2.5 and 5.2.6. Again for the DIRECT approach the results are of the same form as in Sample 1. It is interesting that in both samples 15 and 18 the large discrepancies in these plots resulted in DIRECT estimates close to the simulated values. This behaviour is repeated in Table 5.2.1 and seems to
Figure 5.2.3
Figure 5.2.4

\[ \frac{\delta L}{\delta \alpha} = 0 \]

\[ \frac{\delta L}{\delta \beta} = 0 \]
Figure 5.2.5
Figure 5.2.6

The significant differences between Figures 5.2.5 and 5.2.6 may result from the flatter nature of the HINT-contours. The HINT contours are widely spaced and small alterations in absolute value can change the form of the plot to a large degree.

For the equations of special interest such negative sign implies that for an increase in the a-equations are large (illustrative data in Figures 5.2.3 and 5.2.4). For example let the range in a be 0.00 - 0.50, and b be 0.00 - 0.50. Similarly the a and b-equations of the situation obtaining in the presence of local minima. In the situation as a complex still with increasing tendency producing more minima. In an instance of a large region where the minima form a trough and the a and b-equations are unable to zero, converting to a true minimum is only situation is considered.

In the case of this behaviour whether or whether it produces further simulations or necessary. By Table 5.2.15, 17, although are considered as the results of 5 simulations of 5 minima are obtained. The final model is the procedure of choosing 5 sets of the parameter variation convergent in adding toll be expected.
constitute the only major difference between the DIRECT and EXACT methods.

The significant differences between Figures 5.2.3, 5.2.4 and Figures 5.2.5, 5.2.6 may result from the flatness of Figure 5.2.3. The EXACT contours are widely spaced and small alterations in absolute value can change the form of the plot to a large degree.

For the EXACT approach values of $\beta$ in the region close to -1 are of special interest. On a magnified scale the likelihood contours and $\alpha$ and $\beta$-equations are more informative than in Figures 5.2.3 and 5.2.4. For Sample 18 the region $\alpha \in (0.55, 0.6), \beta \in (-0.99, -0.9996)$ is considered in Figures 5.2.7 and 5.2.8. The likelihood contours show the flat bumpy behaviour which causes difficulties and there are 3 local minima virtually on $\beta = -1$. Similarly the $\alpha$ and $\beta$-equations show a double intersection implying the presence of local minima. In fact the situation is more complex still with increasing accuracy producing more minima. In fact there is a large region where the likelihood forms a trough and the $\alpha$ and $\beta$-equations are close to zero. Converging to a true minimum in this situation is extremely difficult.

To see if this behaviour is atypical or whether it presents a real problem further simulations are necessary. In Table 5.2.1 10 ARMA(1,1) models are considered and the results of 40 simulations of each model are tabulated. For each model the EXACT approach is implemented by both G13BBF and the procedure of Chapter 4. Also the DIRECT method is used for each case. For each set of simulations the number of parameter estimates convergent in modulus to 1 is reported.
Figure 5.2.7

-1.0

β

1.19 - 1.24
1.19 - 1.20
1.18 - 1.19
1.19 - 1.18
1.17

0

0.55

-0.99

0.6

α

1.24 - 1.26
1.20 - 1.23
1.19 - 1.20

Figure 5.2.8

\[
\frac{\delta L}{\delta \alpha} = 0
\]

\[
\frac{\delta L}{\delta \beta} = 0
\]
Table 5.2.1

<table>
<thead>
<tr>
<th>Simulated values</th>
<th>Sample size</th>
<th>No. of estimates of modulus 1</th>
<th>G13BEF</th>
<th>DIRECT</th>
<th>Chapter 4 procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>β</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.1</td>
<td>-0.5</td>
<td>30</td>
<td>5</td>
<td>0(2)</td>
<td>0(3)</td>
</tr>
<tr>
<td>-0.1</td>
<td>0.1</td>
<td>30</td>
<td>9</td>
<td>1(4)</td>
<td>1(8)</td>
</tr>
<tr>
<td>-0.1</td>
<td>-0.9</td>
<td>30</td>
<td>14</td>
<td>0(2)</td>
<td>2(12)</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>50</td>
<td>0</td>
<td>0(0)</td>
<td>0(0)</td>
</tr>
<tr>
<td>0.5</td>
<td>-0.5</td>
<td>50</td>
<td>0</td>
<td>0(1)</td>
<td>0(0)</td>
</tr>
<tr>
<td>0.5</td>
<td>-0.5</td>
<td>30</td>
<td>3</td>
<td>0(1)</td>
<td>1(2)</td>
</tr>
<tr>
<td>0.9</td>
<td>-0.5</td>
<td>30</td>
<td>0</td>
<td>0(0)</td>
<td>0(0)</td>
</tr>
<tr>
<td>0.9</td>
<td>-0.5</td>
<td>50</td>
<td>0</td>
<td>1(0)</td>
<td>0(0)</td>
</tr>
<tr>
<td>-0.9</td>
<td>0.9</td>
<td>30</td>
<td>21</td>
<td>3(10)</td>
<td>1(20)</td>
</tr>
<tr>
<td>-0.1</td>
<td>-0.9</td>
<td>30</td>
<td>14</td>
<td>1(4)</td>
<td>4(10)</td>
</tr>
</tbody>
</table>

The figures in brackets represent the number of times a simulation resulted in non-convergence.

In total, 400 simulations resulted in 64 different estimates having modulus approximately 1 (1 ± 0.001) using the NAG routine G13BEF. Of these 64 estimates, the EXACT approach of Chapter 4 agreed with 9 and failed to converge on 55 occasions. The DIRECT approach converged to estimates similar to the simulated values on 34 occasions and failed to converge 24 times. These results strongly suggest that
the problem is reasonably common and that the cases studied in detail (Samples 15 and 18) are representative of the general situation.

An obvious conclusion is that estimates lying on the boundary of the unit circle are unreliable and should be treated with caution. Further it can be seen that when the DIRECT procedure converges to an acceptable solution then this is a preferable estimate than an EXACT estimate of modulus 1. This situation occurs fairly regularly and warrants further study. Perhaps the transformation from the observations \( (X_1, \ldots, X_n) \) to the sample serial correlations \( (r_k, k=1, \ldots, m) \) renders the direct method more resistant to this particular problem. Certainly it seems that approximate procedures have an important role to play in parameter estimation despite the concentration in recent years on the exact ML procedure.
6.1 Introduction

In Chapter 1 the application of random geometric series to block codes and digital transmissions was introduced. The work of Cariolaro and Tronca [1974] and Justesen [1972] indicated the use of spectral density functions of ARMA processes in the study of block codes. Specifically the spectral density functions contain information relevant to the amount of interference present in digital transmissions. However practical applications of the ARMA spectral density functions are difficult to produce. Hence an alternative approach has been used.

The work of Huzii and Sugiyama [1970] and Hill and Blanco [1973] is a good introduction to the problem of intersymbol interference and random geometric series. It is shown that probabilistic information concerning intersymbol interference in block coded digital transmissions is based on the random variable $X$ given by

$$X = \sum_{n=1}^{\infty} a^{n-1} X_n$$

(6.1.1)

where $0 < a < 1$. $(X_n)$ is a regular stationary Markov pulse train with states $\alpha_1, \ldots, \alpha_m$ and transition probability matrix $P = (p_{ij})$.

Hence in Chapters 6 to 10 the aim is to produce a set of techniques to evaluate the probability density function and the distribution function of the variable $X$. 


Variables of the form given by (6.1.1) occur in several different disciplines as outlined in Section 6.4. The exact form of $X$ is different in various applications but the methods available for calculating the distribution are the same.

The same general $X$ variable has been considered by Huzii and Sugiyama [1970]. However their work concentrated on some specific examples where considerable simplifications occur. Most of the work related to this problem has concentrated on a slight variation of (6.1.1), namely on the variable $Z$ given by

$$Z = (1-a) \sum_{n=0}^{\infty} a^n X_n$$

(6.1.2)

where $0 < a < 1$. This is a simpler situation where the sequence $\{X_n\}$ is assumed to consist of independent and identically distributed random variables taking the values $\pm 1$ with equal probability. This "equiprobable" case has had a lot of attention in the mathematical literature where studies have looked at finding values of $a$ for which $Z$ has a non-singular distribution. Interest has also been apparent in the engineering literature. The tutorial type paper by Hill and Blanco [1973] outlines the practical applications of the $Z$ variable.

Recently the analysis of an error detector by the General Electric Company has also stimulated interest in a particular example of an $X$ variable of the form (6.1.1). Limited techniques are available to find the distributions of certain $X$ variables. However these techniques are only feasible in certain situations. The error detector study at G.E.C. is based on a particularly intractable $X$ variable. Hence the aim of the approaches developed in later Chapters is to be applicable to all situations of interest.
The remainder of Chapter 6 concerns the background to the general problem. Current contributions to the knowledge of the distributions of \( X \) and \( Z \) are discussed and some applications are introduced. It is shown how the performance of an error detector in digital communications can be maximized by knowledge of the distribution of \( X \).

The work stemming from the error detector considered by G.E.C. is a special case of (6.1.1) with a specific type of Markov train and transition matrix. The study of this "G.E.C." case is instructive for the general class of Markov trains and transition matrices. Hence this example is considered in detail in Chapters 7, 8 and 9. Throughout these Chapters all the work will refer to the G.E.C. case. Chapter 10 will then generalize the applicable methods to the general case given by (6.1.1).
6.2 The even mark parity error detector

An introduction to this type of error detection is given by Sharland and Stevenson [1983]. The original detection scheme was invented by Jessop [1978]. A simplified scheme is given in Figure 6.2.1.

A description of the operation of this error detector is given below.

(i) Coded symbols - the digital transmission is coded into a sequence of binary digits. The binary 0 and 1 are referred to as spaces and marks respectively. The coding has the property that sequences of digits containing an even number of marks are more probable than sequences containing an odd number. The coding also introduces a preferred parity associated with the marks as explained below.

(ii) Parity state decoder - the decoder is a device which has two states. In the up state it outputs a voltage of 1. In the down state it outputs a voltage of -1. It changes state on receipt of a mark and is unchanged on receipt of a space. The predominance of even sequences of marks will make the decoder output more likely to be in the same state as it started in. Hence there is a preferred parity and the output voltage is more likely to be in one state than the other.

(iii) Errors - the decoder has a preferred parity which will remain the same if there are no errors. Consider the situation when errors occur and marks are replaced by spaces or vice-versa. Either of these errors cause the decoder to be in the opposite state to the correct state. Then, the predominance of even sequences of marks
Figure 6.2.1

Coded Symbols

- Parity State Decoder
  - Down (-1)  Up (+1)

Leaky Integrator

Decision Device

Error Indications
will make the decoder output more likely to be in the opposite state to the one in which it started. Hence single errors, or in general an odd number of errors, cause the preferred parity to be changed.

(iv) Leaky integrator - the purpose of this device is to smooth the output from the parity state decoder. The device used has the property that for arbitrary input \( h(t) \), it produces an output \( y(t) \) satisfying

\[
y'(t) + \alpha y(t) = \frac{\alpha}{2} h(t) \tag{6.2.1}
\]

where \( \alpha \) is a parameter of the integrator. When fed with the ±1 output from the decoder Wray [1982] has shown that the integrator produces a response \( V \) that satisfies

\[
V_{n+1} = a V_n + (1-a) X_n \tag{6.2.2}
\]

\( X_n = \pm 1 \) output from decoder, \( a = e^{-\alpha} \).

The formula (6.2.2) is a discrete time, \( V_n \) rather than \( V(t) \), since symbols are assumed to enter the decoder at a rate of 1 per second.

(v) Decision device - the integrator smooths the ±1 output from the decoder into a continuous DC voltage in the range \([-1, +1]\). The preferred parity means that this voltage settles towards an equilibrium non-zero value. An odd number of errors will cause this equilibrium to be disturbed. After the error the voltage again settles towards an equilibrium value but of opposite sign to the previous one. Hence the integrator output is expected to cross zero after an odd number of errors. This zero crossing is used by the decision device to detect errors.
The design problem

There are two problems that such a decision device will suffer.

(a) Actual errors occurring at low rates. Here there will be
spurious error indications caused by zero crossings resulting
from random fluctuations in the smoothed output.
Diagrammatically this is demonstrated in Figure 6.2.2.

(b) Actual errors occurring at high rates. Here the smoothed output
may change direction without crossing zero as a result of two
adjacent errors. This is illustrated by Figure 6.2.3.

Thus the design problem is to choose the parameter \( \alpha \) from (6.2.1)
to give the best compromise between the two shortcomings. This can
best be seen by considering a generalized graph of the log of the
indicated error rate against the log of the actual error rate. This
is shown in Figure 6.2.4.

The upper asymptote is caused by the error detector having an
upper bound on the number of errors it can detect (see problem (b)).
The lower asymptote is caused by the occurrence of spurious errors at
a certain rate swamping the actual errors (see problem (a)).

The choice of \( \alpha \) is demonstrated in Figure 6.2.4. For an actual
log error rate of \( e_0 \) the value \( \alpha = \alpha_0 \) is the best compromise
compared to \( \alpha_1 \) and \( \alpha_2 \).

Wray (1982), has shown, using a result in Blake and Lindsay
(1973), that the probability density of the voltage process (6.2.2)
can yield information about detected error rates and spurious error
Figure 6.2.2

Integrator Output (V)

Equilibrium Voltage

Time

Spurious Error Indication

Figure 6.2.3

Integrator Output (V)

Error Events

Time

Error Indications

Missing Error Indications
Figure 6.2.4

The voltage process (6.3.3) is of the same form as the basic random variable in (6.1.1). Thus the design problem is precisely that outlined in the introduction, namely to find the probability density function of a random variable satisfying an equation of type (6.1.1). Then the density function can be used to choose a value for \( \alpha \) which is optimal in some sense.

\[
\log (\text{indicated error rate})
\]

\[
\log (\text{actual error rate})
\]
rates (see Section 6.4). Also in Section 6.3 it will be shown that the voltage process (6.2.2) is of the same form as the basic random variable in (6.1.1). Thus the design problem is precisely that outlined in the introduction, namely to find the probability density function of a random variable satisfying an equation of type (6.1.1). Then the density function can be used to choose a value for $\alpha$ which is optimal in some sense.

$$P = \begin{bmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \cdots & p_{nn} \end{bmatrix}$$

If there is a preferred parity in terms of $\alpha$ we work and $\varepsilon$ is represented as a bias in the transients.

$$P = \begin{bmatrix} p_{11} & q_{12} & \cdots & q_{1n} \\ q_{21} & p_{22} & \cdots & q_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ q_{n1} & q_{n2} & \cdots & p_{nn} \end{bmatrix}$$

In the work that follows the lower-left $m \times m$ block need not be computed.

$$P = \begin{bmatrix} p_{11} & \cdots & p_{1m} \\ \vdots & \ddots & \vdots \\ p_{m1} & \cdots & p_{mm} \end{bmatrix}$$

where $p_{ij} = \text{prob} \text{ that } i \text{ cannot switch to } j$.

Theorem 6.3.3

$$p_{ij}(t) = p_{ij}(t)^{\infty} = \lim_{n \to \infty} p_{ij}(t^n)$$

$$p_{ij}(t) = p_{ij}(t)^{\infty} = \lim_{n \to \infty} p_{ij}(t^n)$$

$$a_{ij}(t) = \frac{p_{ij}(t)}{p_{ii}(t)}$$

$$b_{ij}(t) = \frac{a_{ij}(t)}{a_{jj}(t)}$$

$$c_{ij}(t) = \frac{b_{ij}(t)}{b_{jj}(t)}$$

$$d_{ij}(t) = \frac{c_{ij}(t)}{c_{jj}(t)}$$

$$e_{ij}(t) = \frac{d_{ij}(t)}{d_{jj}(t)}$$

$$f_{ij}(t) = \frac{e_{ij}(t)}{e_{jj}(t)}$$

$$g_{ij}(t) = \frac{f_{ij}(t)}{f_{jj}(t)}$$

$$h_{ij}(t) = \frac{g_{ij}(t)}{g_{jj}(t)}$$

$$i_{ij}(t) = \frac{h_{ij}(t)}{h_{jj}(t)}$$

$$j_{ij}(t) = \frac{i_{ij}(t)}{i_{jj}(t)}$$

$$k_{ij}(t) = \frac{j_{ij}(t)}{j_{jj}(t)}$$

$$l_{ij}(t) = \frac{k_{ij}(t)}{k_{jj}(t)}$$

$$m_{ij}(t) = \frac{l_{ij}(t)}{l_{jj}(t)}$$

$$n_{ij}(t) = \frac{m_{ij}(t)}{m_{jj}(t)}$$

$$o_{ij}(t) = \frac{n_{ij}(t)}{n_{jj}(t)}$$

$$p_{ij}(t) = \frac{o_{ij}(t)}{o_{jj}(t)}$$

$$q_{ij}(t) = \frac{p_{ij}(t)}{p_{jj}(t)}$$

$$r_{ij}(t) = \frac{q_{ij}(t)}{q_{jj}(t)}$$

$$s_{ij}(t) = \frac{r_{ij}(t)}{r_{jj}(t)}$$

$$t_{ij}(t) = \frac{s_{ij}(t)}{s_{jj}(t)}$$

$$u_{ij}(t) = \frac{t_{ij}(t)}{t_{jj}(t)}$$

$$v_{ij}(t) = \frac{u_{ij}(t)}{u_{jj}(t)}$$

$$w_{ij}(t) = \frac{v_{ij}(t)}{v_{jj}(t)}$$

$$x_{ij}(t) = \frac{w_{ij}(t)}{w_{jj}(t)}$$

$$y_{ij}(t) = \frac{x_{ij}(t)}{x_{jj}(t)}$$

$$z_{ij}(t) = \frac{y_{ij}(t)}{y_{jj}(t)}$$

$$| \begin{array}{cc} a & b \\ c & d \end{array} |$$

| \begin{array}{cc} e & f \\ g & h \end{array} |
6.3 Representations of the voltage process

The state of the parity state decoder in Figure 6.2.1 can be represented by a transition matrix or a transition diagram as shown in Figure 6.3.1. Let the probability that a coded symbol is a mark be \( q \) and the probability of a space be \( p \).

If there is no preferred parity induced by the coding then the transition matrix is

\[
P = \begin{bmatrix}
p & q \\
q & p
\end{bmatrix}
\]

If there is a preferred parity in favour of the marks then this can be represented as a bias in the transition matrix

\[
P = \begin{bmatrix}
p - \delta & q + \delta \\
q - \delta & p + \delta
\end{bmatrix}
\]

In the work that follows the transition matrix will be represented by

\[
P = \begin{bmatrix}
p_{-} & p_{+} \\
p_{+-} & p_{++}
\end{bmatrix}
\]

where \( p_{ij} = \text{prob} (\text{decoder is now in state } j | \text{ previous state was } i) \).

Proposition 6.3.1.

\[
af_n(v) = p_{-} f_{n-1} \left[ \frac{v + 1 - a}{a} \right] + p_{+} g_{n-1} \left[ \frac{v - 1 + a}{a} \right]
\]

\[
ag_n(v) = p_{+} f_{n-1} \left[ \frac{v + 1 - a}{a} \right] + p_{++} g_{n-1} \left[ \frac{v - 1 + a}{a} \right]
\]

\[
p_{-} = 1 - q - \delta, \quad p_{+} = q - \delta, \quad p_{+-} = q + \delta, \quad p_{++} = 1 - q + \delta
\]

\[
f_n(v) = \text{prob} (V_n = v \text{ and } X_n = -1)
\]

\[
g_n(v) = \text{prob} (V_n = v \text{ and } X_n = +1)
\]
Figure 6.3.1

Proof

Consider the (x,y) pair for the two points in Figure 6.

Suppose that \( V_x \neq 0 \).

Then if \( V_{x-1} = V_x \),

it follows from (2.3) that

\[
V = y^{m-1} (x^{m-1}) dy^{m-1} dx^{m-1}
\]

and

\[
V + yV = y^{m-1} (x^{m-1}) dy^{m-1} dx^{m-1}
\]

Hence it follows that

\[
V = y^{m-1} (x^{m-1}) dy^{m-1} dx^{m-1}
\]
Proof

Consider the \((n-1)\)th and \(n\)th instant that symbols enter the integrator in Figure 1.1.1. The voltage responses of the integrator at these two points are \(V_{n-1}\) and \(V_n\).

Suppose that \(V_n \in (V, V + \delta V)\) for small \(\delta V\).

Then if \(V_{n-1} \in (V^*, V^* + \Delta V)\) for small \(\Delta V\), it follows from (6.2.2) that

\[
V = aV^* + (1-a)X_{n-1}
\]

and

\[
V + \delta V = a(V^* + \Delta V) + (1-a)X_{n-1}.
\]

Thus

\[
V^* = \frac{V - (1-a)X_{n-1}}{a}
\]

and

\[
\delta V = a\Delta V.
\]

Hence it follows that

\[
\text{prob} \left( V_n \in (V, V + \delta V) \text{ and } X_n = -1 \right)
\]

\[
= p_- \text{prob} \left( V_{n-1} \in \left( \frac{v+1-a}{a}, \frac{v+1-a}{a} + \Delta V \right) \text{ and } X_{n-1} = -1 \right)
\]

\[
+ p_+ \text{prob} \left( V_{n-1} \in \left( \frac{v-1+a}{a}, \frac{v-1+a}{a} + \Delta V \right) \text{ and } X_{n-1} = +1 \right)
\]

From the notation of the proposition this can be rewritten in the form

\[
f_n(v) \delta V = p_- f_{n-1} \left( \frac{v+1-a}{a} \right) \Delta V + p_+ g_{n-1} \left( \frac{v-1+a}{a} \right) \Delta V
\]

or more simply as

\[
a f_n(v) = p_- f_{n-1} \left( \frac{v+1-a}{a} \right) + p_+ g_{n-1} \left( \frac{v-1+a}{a} \right)
\]

Similarly by considering \(\text{prob} \left( V_n \in (V, V + \delta V) \text{ and } X_n = +1 \right)\)

the other equation follows

\[
a g_n(v) = p_+ f_{n-1} \left( \frac{v+1-a}{a} \right) + p_+ g_{n-1} \left( \frac{v-1+a}{a} \right)
\]
The equations (6.3.2) form the first representation of the G.E.C. design problem. They are of the same form as the equations derived by Hill and Blanco [1973], for the equiprobable case and by Huzii and Sugiyama [1970], for the general case. If the voltage process settles towards an equilibrium distribution such that

\( f_n(v) \to f(v), \quad g_n(v) \to g(v) \) as \( n \to \infty \), then the equilibrium distribution will satisfy

\[
af(v) = p_+ f\left(\frac{v+1-a}{a}\right) + p_+ g\left(\frac{v-1+a}{a}\right)
\]

(6.3.3)

Integrating throughout gives

\[
F(v) = p_- F\left(\frac{v+1-a}{a}\right) + p_+ G\left(\frac{v-1+a}{a}\right)
\]

(6.3.4)

where

\[
F(v) = \int_{-\infty}^{\infty} f(x)dx, \quad G(v) = \int_{-\infty}^{\infty} g(x)dx
\]

In Chapter 8 it is shown that such an equilibrium situation does occur. This is done by showing that (6.3.4) represents a semicontraction mapping. Hence the solutions of (6.3.3) and \( \text{or} \) (6.3.4) give the information required for the G.E.C. design problem via the identity

\[
p(v) = \text{equilibrium probability density function of the voltage process} = f(v) + g(v).
\]
The second representation of the voltage process is of the same form as (6.1.1). Equation (6.2.2) gives

\[ V_n = aV_{n-1} + (1-a) X_{n-1} \]

Operating this recursion \( n \) times gives

\[ V_n = a^n V_0 + (1-a)^{n-1} a^i X_{n-1-i} \]  \hspace{2cm} (6.3.5)

To see the similarity of (6.3.5) to (6.1.1) consider reversing the sequence \( X_0, X_1, \ldots, X_{n-1} \) so it is written \( X_{n-1}, X_{n-2}, \ldots, X_0 \).

Using this notion \( V_n \) becomes

\[ V_n = a^n V_0 + (1-a)^{n-1} a^i X_{i} \]

Letting \( n \to \infty \) gives

\[ V = (1-a) \sum_{i=0}^{\infty} a^i X_{i} \]  \hspace{2cm} (6.3.6)

where the reversed sequence \( (X_i) \) now has probability transition matrix

\[ P' = \begin{bmatrix} P_{-} & P_{+} \\ P_{-} & P_{++} \end{bmatrix} \]

Hence this change of notation has produced (6.3.6) which is only a slight modification of (6.1.1).

Finally the Markov nature of the voltage process is discussed. The voltage response of the integrator at the \((n+1)\)th instant is governed by equation (6.2.2). Hence \( V_{n+1} \) depends on the voltage and the state of the decoder at the \( n \)th instant. Thus the process is a Markov process in discrete time with state space \((-1, +1) \times (-1, +1)\). This is the same state space as reported by Wray (1982), but he attempted to consider the process in continuous time.
A summary of the error detector problem.

The voltage output of the integrator is a random variable satisfying (6.3.6). The equilibrium probability density and distribution functions satisfy respectively

\[ af(v) = p_{--} f \left( \frac{v+1-a}{a} \right) + p_{+} g \left( \frac{v-a}{a} \right) \]

\[ ag(v) = p_{-+} f \left( \frac{v-1+a}{a} \right) + p_{-} g \left( \frac{v-a}{a} \right) \]

\[ F(v) = p_{--} F \left( \frac{v+1-a}{a} \right) + p_{+} G \left( \frac{v-a}{a} \right) \]

\[ G(v) = p_{-+} F \left( \frac{v-1+a}{a} \right) + p_{-} G \left( \frac{v-a}{a} \right) \]

where \( 0 < a < 1 \), \( p_{--} = 1-q-\delta \), \( p_{-+} = q + \delta \), \( p_{+} = q-\delta \), \( p_{++} = 1-q+\delta \).

Thus the design problem motivates the solution of the above equations given the parameter set \( (a,q,\delta) \).

For codes in use or contemplated the parameters take values in the following regions

\[ q \approx \% \]

\[ 0.01q < \delta < 0.1q \quad (6.3.7) \]

\[ 1-10^{-3} < a < 1-10^{-5} \]

This then gives a typical parameter set for the error detector problem. The value of \( a \) is so close to 1 that it causes considerable problems in later work. In the notation that follows \( a = 1 - \varepsilon \) is often used. Thus \( \varepsilon = 1 - a \) is usually in the range

\[ 10^{-5} < \varepsilon < 10^{-3} \].
The notation \((p_-, p_+)\) is used for the stationary probability vector of the transition matrix \(P\), where

\[
(p_-, p_+) = \left( \frac{q - \delta}{2q}, \frac{q + \delta}{2q} \right)
\]  

(6.3.8)
6.4 Review of current literature.

Wray [1982] is the only author to consider the G.E.C. problem specifically. He used a version of (6.2.2) in continuous time
\[ V(t+T) = aV(t) + (1-a)X(t). \]
Here \( X(t) \) is the state of the decoder in time \((t, t+T)\). If there is a change of state in \((t, t+r)\) then this equation is not valid. Hence the continuous time process has the property that the voltage at time \( t+T \) depends upon \( V(t) \), \( X(t) \) and also on whether a change of state occurs in \((t, t+r)\). Hence the simpler Markov nature of the discrete time process is complicated. Thus the derivation of the forward equations treating \( V(t) \) as a Markov process will produce at most an approximation to the real problem.

The expressions produced by Wray for the equilibrium probability density functions are
\[
\begin{align*}
  f(v) &= K(1-v)^r(1+v)^s - \alpha T, \\
  g(v) &= K(1-v)^r-1(1+v)^s - \alpha T, \\
\end{align*}
\]
where \( e = a \), \( r = \frac{q-\delta}{\alpha T} \), \( s = \frac{q+\delta}{\alpha T} \),
\[
K = 2 \frac{-2q}{\alpha T} \left[ \begin{array}{c}
  2q \\
  - \frac{r}{\langle \alpha T \rangle} \\
  - \frac{q-\delta}{\langle \alpha T \rangle} \\
  - \frac{q+\delta}{\langle \alpha T \rangle}
\end{array} \right]
\]
Together the two equations give
\[ p(v) = 2K (1-v)^r(1+v)^s \]
This form for the voltage process is that of a Beta distribution. The corresponding distribution function is an incomplete Beta function. It is interesting that two methods outlined in Sections 7.1 and 7.2 also result in Beta distributions although with different parameters. Another similarity with some of the work in Section 7.2 is the identity \((1+v)f(v) = (1-v)g(v)\), which is considered later.

In Section 3 of Wray’s report it is shown how the probability density function of the voltage process can give information about error rates and hence help to solve the design problem.

(i) Spurious errors - the model is based on a two state transition matrix which assumes that the marks and spaces which pass into the decoder are correct. Hence the model produces an output voltage which crosses zero due to random fluctuations only. In section 2.4 of Wray’s report it is shown that the mean rate of zero crossing for the voltage process is
\[ af(0) + ag(0) = ap(0). \]
Hence knowing \( p(0) \) gives the mean rate of spurious indication of errors.

(ii) Real errors - as described in (i) the model description does not allow for real errors. However approximate information about real errors can be gained as described by Wray. The approach is to set \( \delta = 0 \) and \( q = \) probability of error. This gives a voltage output
without even-mark parity bias which changes state with the same probability as that of the occurrence of an error. Hence each error will cause a zero crossing if random fluctuations causing spurious errors can be ignored. Hence this is an approximate method of gaining the mean rate of error indication for an error rate \( q \). Again this rate is given by \( \alpha p(0) \) with the suitable definitions for \( \delta \) and \( q \). However since random fluctuations are ignored the results are only appropriate when the real error rate is considerably larger than the spurious error rate.

The two state model considered gives, through knowledge of \( p(0) \), information primarily about the asymptotes in Figure 6.2.4. From (i), above, the lower asymptote can be found caused by spurious errors. From (ii), above, the upper asymptote can be found caused by real errors swamping the spurious errors. The central part of the graph can be approximated by (ii) but with low accuracy. Hence the design problem of choosing a suitable value for \( \alpha \) can be approached by the two state model as demonstrated in Section 6.2. This gives good results but the lack of accuracy in producing Figure 6.2.4 means that better solutions are available. In Section 10.4 it will be shown that a four state model can represent the process more fully and produce a more accurate version of Figure 6.2.4.

The main authors to consider the general problem of the random variable in (6.1.1) are Huzii and Sugiyama [1970]. If \( F_i(x) \) is defined by

\[
F_i(x) = \text{prob} \left( X_i < x \mid X_i = x_i \right) \quad i = 1, \ldots, m
\]
then they produced the following set of equations

\[
F_i(x) = \sum_{j=1}^{m} P_{ij} F_j \left[ \frac{x - \alpha_j}{a} \right] \quad i = 1, \ldots, m \quad (6.4.2)
\]

Comparison of (6.3.4) with (6.4.2) shows that the G.E.C. equations are a special case of the general equations. This can be seen by setting

\[
m = 2, \quad F_i(x) = F(x), \quad F_j(x) = G(x)
\]

\[
\alpha_1 = -(1-a), \quad \alpha_2 = 1-a
\]

Since \(F_i(x)\) conditions on \(x_0\) and \(F(x)\), \(G(x)\) involve the current state, setting \(F_i(x) = F(x), \quad F_j(x) = G(x)\) implies a reversal of the \(\{X_i\}\) sequence. This explains why the matrix

\[
\begin{bmatrix}
P_{11} & P_{12} \\
P_{21} & P_{22}
\end{bmatrix}
\]

becomes

\[
\begin{bmatrix}
P_{--} & P_{+-} \\
P_{-+} & P_{++}
\end{bmatrix}
\]

and not

\[
\begin{bmatrix}
P_{--} & P_{-+} \\
P_{+-} & P_{++}
\end{bmatrix}
\]

Huzii and Sugiyama concentrated on two special cases of (6.4.2) where the \(m\) equations collapse to a single equation of the form

\[
F_1(x) = \sum_{j=1}^{m} P_{j} F_1 \left[ \frac{x - \alpha_j}{a} \right] \quad (6.4.3)
\]

Their method of solution of (6.4.3), specifically for the case \(m = 2\), is to consider semicontraction mappings in metric spaces of distribution functions. Some of the details are given below since this method will be extended in Chapter 8.
Definition 6.4.4.

A mapping \( U \) from a metric space \((R, \rho)\) into itself is called a **semicontraction** mapping if and only if the following two conditions hold

1. \( \rho(Us,Ut) \leq \rho(s,t) \quad \forall s, t \in R \)
2. \( l \in \mathbb{Z}^+ \) s.t. \( \rho(U^l s, U^l t) \leq r \rho(s,t) \)

for some \( 0 < r < 1 \) and \( \forall s, t \in R \).

The semicontraction mapping principle is used to solve (6.4.3) and consists of the following two propositions.

**Proposition 6.4.5.**

If \((R, \rho)\) is a complete metric space then a semicontraction mapping \( U \) has a unique fixed point in \( R \).

**Proposition 6.4.6.**

Let \((R, \rho)\) be a complete metric space and \( U \) a semicontraction mapping operating on \( R \). If \( t_0 \) is an arbitrary point in \( R \) then let the sequence \( \{t_n\} \) be defined by

\[
U t_n = t_{n+1} \quad n=0,1,2,\ldots.
\]

The sequence is convergent to the unique fixed point \( t \) in \( R \) in the sense that \( \lim_{n \to \infty} \rho(t_n, t) = 0 \).

In applying the semicontraction mapping principle to equations (6.4.3) Huzii and Sugiyama use the following definitions.
\( R \) is the set of all continuous and non-decreasing distribution functions such that each function increases only in the interval \([0,1]\).

\[
\rho(F, F') = \sup_{x \in \mathbb{R}} \left\{ |F(x) - F'(x)| \right\}.
\]

Firstly they prove that \((R, \rho)\) as defined is a complete metric space. Then they show that (6.4.3) can be represented as

\[
\langle UF_1 \rangle(x) = \sum_{j=1}^{m} P_j\left[ \frac{x - \alpha_j}{a} \right] 
\]

where \(U\) is a semicontraction mapping operating on \(R\). Hence using the semicontraction mapping principle it can be seen that repeated application of (6.4.7), starting from any element of \(R\), will result in a sequence of distribution functions converging to \(F_1(x)\). Thus they have provided a convergent technique to solve (6.4.2) in the cases where (6.4.2) collapses to (6.4.3). In fact this innovative approach can be extended to nearly all cases of interest as shown in Chapter 10.

The only drawback to this important technique is that of speed of convergence. As \(a \to 1\) the speed of convergence slows down and for values of \(a\) very close to 1 the procedure becomes impractical to use. For these situations other techniques are developed.
Problems of singularity are mentioned by Huzii and Sugiyama for the special case (6.4.3) with \( m = 2 \). For this example they show that for \( 0 < a < 1/2 \) the resulting distribution is singular. Garsia [1962], was the first to prove this. This singularity for \( 0 < a < 1/2 \) is also true for the G.E.C. equations (6.3.4) as demonstrated later in this section.

Most of the other contributions of interest have been brought together in the tutorial paper by Hill and Blanco [1973]. This has superseded the original report by Rice [1957], which first expressed interest in the problem from an engineering standpoint. The paper is based on the equiprobable case, that is the variable \( Z \) from (6.1.2).

Some of the points of interest to the general problem and the G.E.C. problem are mentioned below.

(i) The only situation where the distribution of \( Z \) is known explicitly is when \( a = 1/2 \). Here \( Z \) has a uniform distribution (Chung [1968]). This points to the difficulty of attempting analytic solutions to the distribution of \( Z \).

(ii) It has been shown that the distribution of \( Z \) is either purely singular or absolutely continuous (Jessen and Wintner [1935]). They report that finding values of \( a \) for which \( Z \) has an absolutely continuous distribution is a very difficult and substantially unsolved problem. It is known that the distribution is singular for \( a < 1/2 \), Garcia [1962]), and when \( a \) is the reciprocal of a Pisot-Vijayaraghavan number (Erdos [1939]). A P-V number is an algebraic integer all of whose conjugates have absolute value less than 1. It is also known
that the smallest P-V number is the solution of $x^3 - x - 1 = 0$, (Siegel [1944]). The reciprocal of this P-V number is approximately 0.755. Hence the largest value of $a$ for which the distribution is known to be singular is approximately 0.755. Garsia [1962], has conjectured that the distribution may be singular only when $a$ is algebraic and satisfies a polynomial equation with coefficients ±1 or 0.

How much of this work can be extended to consider the singularity of $F(x)$ in (6.3.4) is not known. Some work on more general variables than $Z$ has been done by Garsia [1963] and Marsaglia [1971]. As far as the G.E.C. problem is concerned the only result that clearly applies is that $F(x)$ and $G(x)$ are singular for $a < 1/2$. This follows since in Huzii and Sugiyama the proof depends solely on the fact that $F(x)$ depends on $F_1 \left( \frac{x}{a} \right)$ and $F_1 \left( \frac{x-1+a}{a} \right)$.

By changing the variable to $w = 2x - 1$ Hill and Blanco show that $F_1(w)$ depends on $F_1 \left( \frac{w+1-a}{a} \right)$ and $F_1 \left( \frac{w-1+a}{a} \right)$. This is the case for the G.E.C. equations and so the singularity proof applies.

It would appear that for values of $a > 1/2$ the distributions of $Z$ (6.1.2) and $V$ (6.3.6) are absolutely continuous almost everywhere. For the voltage process this has the implication that even if a certain value of $a$ leads to a singular distribution a slight perturbation to $a$ can remove this singularly. This means that the
problems of singularity (where the probability density function is zero almost everywhere) are of mathematical interest rather than engineering importance. In the work that follows where computational results are produced for the distributions it is assumed that the value of \( a \) corresponds to a non-singular distribution.

(iii) The moments of the variable \( Z \) are produced following Rice [1957]. However no mention of techniques to compute the distribution using the moments is made. Some such methods are discussed in Chapter 7.

(iv) For the special case where \( a \) is an integer root of \( 1/2 \), \( a = 2^{-1/K} \), some analytic results are produced. For the \( Z \) variable a functional form is produced for the tails of the distribution.

\[
F(z) = A(1+z)^K, \quad z \in [1-2a, 1]
\]

where \( a = 2^{-1/K} \) and \( A \) is a constant. Note that the absolutely continuous form of \( F(z) \) in \([-1, 1-2a] \) ensures that the distribution is absolutely continuous everywhere in \([-1, 1]\). This follows from Jessen and Wintner [1935], as stated in (ii) above. Equation (6.4.8) can thus be used to evaluate \( F(z) \) exactly in \([-1, 1-2a]\). Hill and Blanco show how these values and expressions can be used recursively to generate values and expressions outside \([-1, 1-2a]\). This recursion is generated by the equation.

\[
F(z) = \frac{1}{2} F\left(\frac{z+1-a}{a}\right) + \frac{1}{2} F\left(\frac{z-1+a}{a}\right)
\]
which is the version of (6.4.2) corresponding to the Z variable. As with the semicontraction mapping principle this method has considerable implementation difficulties for values of a close to 1, as shown in Section 8.1.

(v) An extended form of (6.4.8) is produced for the case a ≠ 2^(-1/K), following some work by Rice [1957]. It is not known whether this solution is general enough to be the correct form and hence its usefulness is limited.
Applications of the general problem.

Several applications of the variable $X$, as defined in (6.1.1), have been given in the introductory sections so far. For the sake of completeness these uses as well as some new ones are brought together in the summary below.

(i) The error detector analysed at G.E.C. can be made more efficient by a knowledge of the probability density function of an $X$ variable. Wray [1982].

(ii) The intersymbol interference between symbols transmitted in digital transmissions is a random variable of the $X$ type. Hence finding the distribution of this variable is important to the knowledge of the interference. Huzii and Sugiyama [1970], Hill and Blanco [1973].

(iii) A special case of the $X$ variable is produced by infinite Bernoulli convolutions. Hence the study of these convolutions leads to the distribution of the variable. Garsia [1962], Erdos [1939], Jessen and Wintner [1935].

(iv) Some models in learning theory involving random walks have produced $X$ variables. Karlin [1953], Kemeny and Snell [1957].

(v) The survival of information in filters is a topic in information theory which has produced an $X$ type variable, although with a finite sum. Golomb [1972].

(vi) The original report to express interest in the $X$ variable from an engineering point of view was by Rice [1957]. This was
concerned with the response of a resonant circuit to a train of random pulses.
CHAPTER 7.

APPROXIMATE SOLUTIONS

In this Chapter various approaches are considered to provide approximate solutions to the equations (6.3.3). The interest in approximate methods stems from the difficulty in obtaining analytic solutions. In Chapters 8 and 9 accurate techniques are developed to find \( f(v), g(v) \). However for many parameters of interest these methods are slow to converge and require considerable computer time. Hence a good approximation would be worthwhile in terms of trading accuracy for simplicity and reduced computer time. Finally methods using initial estimates of \( f(v), g(v) \) will be speeded up by using an approximate solution as the first iteration. Such a procedure is developed in Chapter 9, and is an extension of the semi-contraction mapping approach of Huzii and Sugiyama [1970].

7.1 Moment based techniques

The moments of the probability density function produced from the equations (6.4.2) were formulated by Rice [1957]. In a similar fashion the moments of the probability density function arising from the error detector can be gained. Taking Fourier transforms of equations (6.3.3) gives

\[
\phi(t) = p_- e^{-i\xi t} \phi(at) + p_+ e^{i\xi t} \psi(at)
\]

(7.1.1)

\[
\psi(t) = p_- e^{-i\xi t} \phi(at) + p_+ e^{i\xi t} \psi(at)
\]
where \( \Phi(t) = \int_{-\infty}^{\infty} e^{itv} f(v) dv \)

\( \gamma(t) = \int_{-\infty}^{\infty} e^{itv} g(v) dv \)

Differentiating (7.1.1) \( n \) times and letting \( t = 0 \) gives

\[
\mu_n (1-a^n_{p-}) + \lambda_n (a^n_{p+}) = n \cdot \frac{r}{r-1} \cdot \varepsilon^r a^{n-r} (p_{-} - \lambda_{n-r} + (-1)^r p_{-} \mu_{n-r})
\]

\[
\mu_n (a^n_{p-}) + \lambda_n (1-a^n_{p+}) = n \cdot \frac{r}{r-1} \cdot \varepsilon^r a^{n-r} (p_{+} + \lambda_{n-r} + (-1)^r p_{+} \mu_{n-r})
\]

\[
\ldots \quad (7.1.2)
\]

where \( \mu_n = \int_{-\infty}^{\infty} v^n f(v) dv \), \( \lambda_n = \int_{-\infty}^{\infty} v^n g(v) dv \).

Hence all the moments \( \mu_n, \lambda_n \) are available from the recursive formula (7.1.2). The overall moments \( w_n \) are obtained from

\[
w_n = \int_{-\infty}^{\infty} v^n f(v) dv = \mu_n + \lambda_n.
\]

To start the recursion \( \mu_0, \lambda_0 \) are gained by setting \( n = 0 \).

There is a computational difficulty in that the left hand side of (7.1.2) is

\[
\left[\begin{array}{cc}
1-a^n_{p-} & -a^n_{p+} \\
-a^n_{p-} & 1-a^n_{p+}
\end{array}\right] \left[\begin{array}{c}
\mu_n \\
\lambda_n
\end{array}\right]
\]

The determinant of the matrix of coefficients is \( O(\epsilon) \) which could cause problems in the solution of (7.1.2) when \( \epsilon \) takes very small values. However working in double precision and programming to avoid this problem will avoid any difficulties and give accurate answers.

Using (7.1.2) the following moments are gained
\begin{equation}
\mu_0 = \frac{q-\delta}{2q}, \quad \lambda_0 = \frac{q+\delta}{2q}, \quad \nu_0 = 1. \tag{7.1.3}
\end{equation}

\begin{equation}
\mu_1 = \frac{q-\delta}{2q}\left\{\frac{2q\delta-\epsilon(1-2q)}{2q+\epsilon(1-2q)}\right\}, \quad \lambda_1 = \frac{q+\delta}{2q}\left\{\frac{2q\delta+\epsilon(1-2q)}{2q+\epsilon(1-2q)}\right\}, \quad \nu_1 = \frac{\delta}{q}. \tag{7.1.4}
\end{equation}

\begin{equation}
\lim_{\epsilon \to 0} \langle \mu_1 \rangle = \left( \frac{q-\delta}{2q} \right) \frac{\delta}{q}, \quad \lim_{\epsilon \to 0} \langle \lambda_2 \rangle = \left( \frac{q+\delta}{2q} \right) \frac{\delta}{q}.
\end{equation}

\begin{equation}
\nu_2 = \frac{1}{2q}\left\{\frac{6\delta^2 + 4(q-q^2 - 2\delta^2)\epsilon - 2q(1-2q)\epsilon^2}{4q + 2(1-3q)\epsilon - (1-2q)\epsilon^2}\right\}
\end{equation}

The overall variance is

\begin{equation}
\nu_2 - \nu_1^2 = \frac{\left(\frac{q^2 - \delta^2}{(2-\epsilon)q^2}\right) \epsilon}{\frac{2}{2q+\epsilon(1-2q)}} - 1. \tag{7.1.5}
\end{equation}

\begin{equation}
\lim_{\epsilon \to 0} \nu_2 = \frac{\delta^2}{q}, \quad \lim_{\epsilon \to 0} \nu_2 - \nu_1^2 = 0.
\end{equation}

Since values of \( \epsilon \) corresponding to the error detector can be as low as \( 10^{-5} \), the behaviour of the moments as \( \epsilon \to 0 \) is of interest. A typical set of parameters would be \( q = 0.5, \quad 0.005 < \delta < 0.05, \quad 10^{-5} < \epsilon < 10^{-3} \).

These give moments of the order

\begin{equation}
0.01 < \nu_1 < 0.1,
\end{equation}

\begin{equation}
5 \times 10^{-6} < \text{variance} < 5 \times 10^{-4},
\end{equation}

\begin{equation}
0.002 < \text{s.d.} < 0.02.
\end{equation}

For these values the probability density function is likely to be sharply peaked at the mean and to drop away very quickly to near zero. As \( \epsilon \to 0 \) it might be expected that the density function would behave like a Dirac delta function, since the variance would also tend to zero. This kind of behaviour will have implications for the methods used to evaluate the density. Some effects are described below:
(i) Some procedures have difficulty in the tails of a distribution. For the error detector one value of special interest is $f(0)$. A typical set of parameters will give zero as being between 5 and 50 standard deviations away from the mean. Hence zero can be a long way out in the tails of the distribution.

(ii) Some procedures work best for distributions with only moderate skewness. For the error detector the distribution exhibits a large degree of skewness, as shown in the following section on Pearson curves.

(iii) The non-zero mean, $W = \delta / q$, causes the Fourier transform of the probability density function to have an oscillatory nature. Pearson curves.

Since all the moments are available a Pearson curve can be produced which will have the same first four moments as the real density. Pearson curves work best for distributions of moderate skewness, Kendall and Stuart [1963]. However here the skewness tends to be large as shown below.

One measure of skewness is given by

$$Sk = \frac{\beta_2^{1/2} (\beta_2 + 3)}{2(5\beta_2 - 6\beta_1 - 9)}$$

where $\beta_r$ is the $r$th moment about the mean.

For the parameters $q = 0.5$, $\delta = 0.05$, $\epsilon = 10^{-5}$, $Sk = -15$
indicating considerable skewness. For the typical values \((6.3.7)\) the skewness is always of this order.

The Pearson curve is defined by the equation

\[
\frac{df(x)}{dx} = \frac{(x-a) f(x)}{b_0 + b_1 x + b_2 x^2}
\]

(7.1.6)

This yields the recursive relation for the moments of the Pearson curve.

\[
n b_0 \mu_{n-1} + \left( (n+1) b_1 - a \right) \mu_n + \left( (n+2) b_2 + 1 \right) \mu_{n+1} = 0.
\]

Using this equation, \(a\), \(b_0\), \(b_1\), \(b_2\) can be identified in terms of the first four moments of the density function to be approximated.

Hence by using a Pearson curve an approximation is available which has the first four moments correct. One drawback is that the range of the curve will not in general be \((-1, +1)\). Replacing \(x-a\) by \(u\) in (7.1.6) gives

\[
\frac{df(u)}{du} = \frac{u}{(b_0 + b_1 u + b_2 u^2) + u (b_1 + 2 a b_2) + b_2 u^2}.
\]

(7.1.6)

If the roots of \(B_0 + B_1 u + B_2 u^2\) are \(Z_1\) and \(Z_2\) then the range of the curve is \((Z_1, Z_2)\). The type of roots found determines the type of Pearson curve. If \(Z_1 < 0, Z_2 > 0\) then the curve is that of a Beta-distribution, sometimes referred to as a Pearson's Type 1 distribution. In the case of equations (6.3.3) it seems that this is always the case for the parameters of interest. For various values of \(q, \delta, \epsilon\) in the respective ranges \((0.45, 0.55), (0.005, 0.05),\)
the range of the Pearson curve remained between 
\((-0.4, 0.4)\) and 
\((-0.6, 0.6)\). This truncation of the range from 
\((-1, +1)\) is not unexpected in an approximation since the moments 
\((7.1.4)\) and 
\((7.1.5)\) imply that \(f(x)\) is close to zero away from the mean.

It is interesting to note that the approximation due to Wray [1982] was also a Beta-distribution, but with different parameters.

To summarize, Pearson curves can provide an approximation in the form of a Beta-distribution. They have four correct moments but usually a different range to the true density. Some examples and an evaluation of their accuracy are given in Section 7.4.

If the distribution function is required rather than the density then the Pearson Type 1 curve has an incomplete Beta-function as its corresponding distribution function. This cannot be gained in a straightforward way from the Pearson curve but several expansions are available to calculate it. Some of these are found in Abramowitz and Stegun [1972].

Edgeworth and Gram-Charlier expansions.

Both of these methods seek to express a probability density function as a series in the derivatives of the normal density function. Formally the two expansions are identical. Practically the Edgeworth expansion is more desirable since successive terms are smaller than the preceding terms.

Both types of expansion are possible since they rely only on a knowledge of the moments of the original distribution. These are
available from (7.1.2). Both expansions give a density function whose first \( n \) moments are the same as the real distribution for any chosen \( n \). However there are computational difficulties.

The Gram-Charlier expansion is easily computed as a straightforward infinite series. The Edgeworth expansion however involves the exponential of an infinite series in the differential operator \( \frac{d}{dx} \). This can be computed for a few finite terms but the general form is less easy to compute. The best way to implement the Edgeworth expansion is to use an algebraic manipulation package as described by Sibley in the discussion of the paper by Cox and Barndorff Nielsen (1979). This is especially the case when a large number of terms are required to give a good approximation.

Kendall and Stuart (1963) report a theorem of Cramér (1926) which gives the following conditions for the convergence of these expansions. If \( f(x) \) is of bounded variation in every finite interval and if

\[
\int_{-\infty}^{\infty} |f(x)| e^{1/4x^2} \, dx
\]

exists, then the expansions for \( f(x) \) converge everywhere to the sum \( 1/2(f(x-0) + f(x+0)) \). For the distributions of interest it is assumed that the parameters yield an absolutely continuous distribution. Also the probability density functions vary over a finite interval and are themselves bounded. Hence the conditions are satisfied and the expansions are valid. In Section (7.4) it is shown that these expansions are useful but suffer from the drawback that convergence can be very slow and a large number of terms may be needed to ensure a good approximation.
This means that the expansions are more efficient as approximations of single values rather than of the functions over their whole range.

Forward equations. As discussed in Section 4.3 this is an approximation to the true nature of the voltage process. A similar approach based on the equilibrium equations of a further differential equation is considered. Here the approximation is made by the use of a one-term Taylor expansion. This procedure relies on the absolute continuity of the distribution function. Since in what follows it is assumed there are no singularity problems.

Rewriting (6.34) gives

\[ F(v) = p_{-1} \left[ \frac{v + \theta}{1 - \epsilon} \right] + p_0 \left[ \frac{v}{1 - \epsilon} \right] \]

\[ G(v) = p_{-1} \left[ \frac{v - \theta}{1 - \epsilon} \right] + p_0 \left[ \frac{v}{1 - \epsilon} \right] \]

Since \( \epsilon \) is an equal for the voltage process, a one-term Taylor expansion is assumed valid for \( \left[ \frac{v + \theta}{1 - \epsilon} \right] \) and \( \left[ \frac{v - \theta}{1 - \epsilon} \right] \). Hence

\[ F(v) = p_{-1} \left( S(v) + \epsilon (v^2) R^2 (v) \right) + p_0 \left( \frac{v}{1 - \epsilon} \right) \]

\[ G(v) = p_{-1} \left( S(v) - \epsilon (v^2) R^2 (v) \right) + p_0 \left( \frac{v}{1 - \epsilon} \right) \]

Adding these two equations gives

\[ F(v) + G(v) = \frac{v}{1 - \epsilon} \]

Adding these two equations gives

\[ F(v) - G(v) = \epsilon \left( v^2 \right) R^2 (v) \]

Substituting these two equations gives

\[ 2p_{-1} \left( S(v) \right) + p_0 \left( \frac{v}{1 - \epsilon} \right) = \epsilon \left( v^2 \right) R^2 (v) \]
7.2 A differential equation approach.

Wray produced a differential equation by considering (6.2.2) as a continuous time Markov process and formulating the corresponding forward equations. As discussed in Section 6.3 this is an approximation to the true nature of the voltage process. In a similar approach based on the equilibrium equations (6.3.4) another differential equation is considered. Here the approximation is made by the use of a one-term Taylor expansion. This procedure relies on the absolute continuity of the distribution function. Hence in what follows it is assumed there are no singularity problems.

Rewriting (6.3.4) gives

\[
F(v) = p_{--} F\left(\frac{v+\epsilon}{1-\epsilon}\right) + p_{+-} G\left(\frac{v-\epsilon}{1-\epsilon}\right)
\]

\[
G(v) = p_{-+} F\left(\frac{v+\epsilon}{1-\epsilon}\right) + p_{++} G\left(\frac{v-\epsilon}{1-\epsilon}\right)
\]

Since \( \epsilon \) is so small for the voltage process, a one-term Taylor expansion is assumed valid for \( F\left(\frac{v+\epsilon}{1-\epsilon}\right) \) and \( G\left(\frac{v-\epsilon}{1-\epsilon}\right) \). Hence

\[
F(v) = p_{--} \left( F(v) + \epsilon (1+v) F'(v) \right) + p_{+-} \left( G(v) + \epsilon (v-1) G'(v) \right)
\]

\[
G(v) = p_{-+} \left( F(v) + \epsilon (1+v) F'(v) \right) + p_{++} \left( G(v) + \epsilon (v-1) G'(v) \right)
\]

Adding these last equations gives

\[
(1-v) G'(v) = (1+v) F'(v)
\]

(7.2.1)

Substituting (7.2.2) into (7.2.1) gives

\[
(1-p_{-})F(v) = p_{--}(1+v)F'(v) + p_{+-}G(v) - p_{--}(1+v) F'(v)
\]

(7.2.2)
The equations now become easier to handle using the parameters $q$ and $\delta$.

\[(q+\delta)F(v) = (1-q-\delta)e(1+v)F'(v) + (q-\delta)G(v) - (q-\delta)e(1+v)F'(v)\]

... (7.2.3)

Using (7.2.3) $G'(v)$ can be calculated as

\[G'(v) = \frac{1}{q-\delta} \left\{ (q+\delta)F'(v) - \epsilon(1-2q) \langle F'(v) + (1+v)F''(v) \rangle \right\} \]

... (7.2.4)

Equating (7.2.4) and (7.2.2) gives

\[(1-v)\left\{ (q+\delta)F'(v) - \epsilon(1-2q) \langle F'(v) + (1+v)F''(v) \rangle \right\} = (1+v)F'(v).\]

Rewriting using $f(v) = F'(v)$ gives for $q \neq 1/2$

\[\frac{f'(v)}{f(v)} = \frac{1}{\epsilon(1-2q)} \left( \frac{q+\delta}{1+v} - \frac{q-\delta}{1-v} - \epsilon(1-2q) \right).\]

Integrating gives the solution

\[\log f(v) = \log(1+v)^{r-1} + \log(1-v)^{s} + \text{const} \]

where \( r = \frac{q+\delta}{\epsilon(1-2q)} \), \( s = \frac{q-\delta}{\epsilon(1-2q)} \).

Hence $f(v) = C(1+v)^{r-1}(1-v)^{s}$. From (7.2.2) this yields

\[g(v) = C(1+v)^{r}(1-v)^{s-1} \]

Thus $p(v) = f(v)+g(v) = 2C(1+v)^{r-1}(1-v)^{s-1}$, $q \neq 1/2$

Hence the Taylor series approximation has produced a Beta distribution, as did Wray and the Pearson curves. The equation is

\[p(v) = K(1+v)^{r-1}(1-v)^{s-1}, \ q \neq 1/2 \] (7.2.5)

In fact for this to be meaningful it is necessary to restrict the value of $q$ even further. If one of $r-1$, $s-1$ is less than zero the corresponding distribution is J-shaped. If both of $r-1$, $s-1$ are less
than zero the function (7.2.5) is U shaped. Hence, it is necessary to restrict the values of r,s so that r-1 >> s-1 >> 0, and this will give a unimodal distribution. The condition s-1 > 0 implies

\[ q - \delta > \epsilon \]

\[ 1 - 2q \]

For the specific case of the typical parameter set (6.3.7) where \( q - \delta >> \epsilon \) it is sufficient that \( 1 - 2q > 0 \), that is \( q < 1/2 \). The constant \( K \) can be easily found, as in Wray, giving

\[
p(V) = \frac{\Gamma(r+s)2^{1-r-s}(1+v)^{r-1}(1-v)^{s-1}}{\Gamma(r)\Gamma(s)}, \quad q < 1/2
\]

As with the Pearson curves the corresponding distribution function is an incomplete Beta-function. Various expansions and formulae for calculating this are in Chapter 26 of Abramowitz and Stegun [1972].

Overall this approach does not seem particularly successful. The solution (7.2.6) is only valid for a restricted range of values of \( q \) and is of the same form as two approximations already considered. Further comparisons are made in Section 7.4.

One approach which could be of interest is based on (7.2.2). Both the Taylor expansion method and Wray's approach produce this equation. If this equation actually holds or is a good approximation in (6.3.3) then the equations collapse to a single equation which may be easier to handle.
In the region $v \in [-1, 1-2a]$ (6.3.3) becomes

$$a f(v) = p f \left( \frac{v+1-a}{a} \right)$$

(7.2.7)

$$a g(v) = p f \left( \frac{v+1-a}{a} \right)$$

Substituting using (7.2.2) gives

$$a f(v) = p_+ f \left( \frac{v+1-a}{a} \right)$$

$$a \frac{1+v}{1-v} f(v) = p_- f \left( \frac{v+1-a}{a} \right)$$

Hence in $[-1, 1-2a]$ (7.2.2) provides a good approximation if

$$p_+ (1-v) \approx p_- (1+v).$$

But as $v$ approaches $-1$, $p_+ (1-v) \gg p_- (1+v)$, so the approximation is useless in $[-1, 1-2a]$. A similar approach shows that in $[2a-1, 1]$ the approximation is also of no use. Hence Wray's solution and the Taylor expansion solution, which rely on (7.2.2), are likely to be poor approximations in the tails.

In fact the simplification which holds in $[-1, 1-2a]$ is

$$f(v) = p_- g(v)$$

which follows from (7.2.7). Substituting this in (6.3.3) gives a contradiction unless $p_+ p_- = p_- p_+$. This is satisfied for $q = 1/2$. Hence for $q = 1/2$ this may yield a useful approximation.

However even if the pair of equations (6.3.3) were approximated by a single equation the problem is not simplified greatly. For instance Huzii and Sugiyama [1970] considered various special cases where there was only one equation. Even in this situation the methods developed were not straightforward. The general techniques
discussed in Chapters 8 and 9 yield just as convenient an approach without approximation.

In Section 7.4 a comparison is made of the approximate solutions produced in this Chapter. One criterion could be how closely the moments compare to the actual moments of (7.1.2). Hence for the Pearson curves, the Taylor series approximation and Wray's solution it is necessary to find the moments of a Beta distribution. Consider equation (7.2.6).

\[
p(v) = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} \frac{2^{1-r-s}(1+v)^{r-1}(1-v)^{s-1}}{v^\epsilon \mathbb{[-1, +1]}}
\]

Here \(r, s\) are arbitrary constants satisfying \(r, s > 1\), since the general distribution is being considered. The \(k^{th}\) moment is

\[
E(v^k) = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} \int_{-1}^{+1} v^k (1+v)^{r-1}(1-v)^{s-1} dv
\]

Using a standard result this gives

\[
E(v^k) = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} \sum_{i=0}^{k} \binom{k}{i} 2^i (-1)^{k-i} \frac{\Gamma(r+i)\Gamma(s)}{\Gamma(r+s+i)}
\]  

Hence all the moments of these three approximations are available from (7.2.8).
7.3 The Markov diffusion approximation.

In essence Wray's approach was to approximate a discrete time Markov process by one in continuous time. A standard way of doing this is to approximate using a Markov diffusion process. Two approaches to this are given in Cox and Miller [1965] both producing differential equations. In one the probability density function is used and in the other the moment generating function. Useful approximations can sometimes be gained for values of the time variable large compared to the intervals between transitions. Clearly this is the case for the voltage process where the equilibrium state is of interest and possible transitions occur in unit time.

Using the notation of Cox and Miller [1965] the two differential equations are given below

\[ \frac{1}{2} \frac{\delta^2}{\delta x^2} \left( \alpha(x,t)p(x,t) \right) - \frac{\delta}{\delta t} \left( \beta(x,t)p(x,t) \right) = \frac{\delta p(x,t)}{\delta t} \]  

\[ \delta \left( \phi(\Theta,t) \right) = -\theta \sum_{r=0}^{\infty} \beta_r(t)(-1)^r \delta^r \phi(\Theta,t) + \theta^2 \sum_{r=0}^{\infty} \alpha_r(t)(-1)^r \delta^r \phi(\Theta,t) \]

For the voltage process \( p(x,t) \) represents the probability density function of the voltage \( x \), at time \( t \). \( \phi(\Theta,t) \) is the moment generating function at time \( t \). In (7.3.1) \( \beta(x,t) \) and \( \alpha(x,t) \) are the equivalents of the infinitesimal mean and variance of a continuous time process, namely the mean and variance of the increment in unit time.

In (7.3.2) \( \alpha(x,t) \), \( \beta(x,t) \) are also assumed to take the form
\[\alpha(x,t) = \sum_{r=0}^{\infty} \alpha(r)x^r, \quad \beta(x,t) = \sum_{r=0}^{\infty} \beta(r)x^r.\]

Since it is the equilibrium situation that is of interest it is assumed that \(p(x,t) \to p(x), \emptyset(\theta,t) \to \emptyset(\theta)\) as \(t \to \infty\). With this assumption the equilibrium versions of (7.3.1) and (7.3.2) can be set up.

If \(\alpha(x,t) \to \alpha(x), \beta(x,t) \to \beta(x)\) as \(t \to \infty\) then \(\alpha(x)\) and \(\beta(x)\) are given below.

\[\beta(x) = \lim_{n \to \infty} \mathbb{E} [V_{n+1} - V_n \mid V_n = x]\]
\[= \lim_{n \to \infty} \mathbb{E} [\varepsilon (X_n - x)] \quad \text{from (6.2.2)}\]
\[= \varepsilon (p_+ - p_- - x) \quad \text{from (6.3.8)}\]
\[= \varepsilon \left(\frac{\delta}{q} - x\right) \quad (7.3.3)\]

\[\alpha(x) = \lim_{n \to \infty} \text{Var} [V_{n+1} - V_n \mid V_n = x] \]
\[= \lim_{n \to \infty} \mathbb{E} [\varepsilon^2 (X_n - x)^2] - \beta^2(x)\]
\[= \varepsilon^2 \left(1 - \left(\frac{\delta}{q}\right)^2\right) \quad (7.3.4)\]

Hence (7.3.1) becomes
\[\frac{1}{2} \frac{d^2}{dx^2} \left(\varepsilon^2 (1 - \left(\frac{\delta}{q}\right)^2)p(x)\right) - \frac{d}{dx} \left(\varepsilon \left(\frac{\delta}{q}\right) x p(x)\right) = 0.\]

Rewriting gives
\[\frac{d^2 p(x)}{dx^2} - (a-bx) \frac{dp(x)}{dx} + b p(x) = 0. \quad (7.3.5)\]
where $a = \frac{\delta q}{\varepsilon (q^2 - \delta^2)}$, $b = \frac{q^2}{\varepsilon (q^2 - \delta^2)}$.

For the voltage process $q > \delta$, so that $a, b > 0$. Given that $p(x)$ is a solution of (7.3.5) it follows that $q(x)$ is a solution of

$$\frac{d^2 q(x)}{dx^2} + \left[ \frac{b - 1}{2} (a-bx)^2 \right] q(x) = 0.$$  

where $q(x) = p(x) \exp \left(-\frac{1}{2}(ax - bx^2)\right)$

The substitution $y = b^{1/2} x - ab^{-1/2}$ gives

$$\frac{d^2 q(y)}{dy^2} - \frac{y^2}{4} - 1 \frac{q(y)}{2} = 0. \quad (7.3.6.)$$

The solution of (7.3.6) is a parabolic cylinder function as discussed in Abramowitz and Stegun (1972).

In their notation

$$\frac{d^2 y}{dx^2} - \frac{1}{4} x^2 + a y = 0.$$  

has the solutions

$$y_1 = e^{-1/4x^2} M \left( \frac{1}{2} a + \frac{1}{4}, \frac{1}{4}, \frac{1}{4} x^2 \right)$$

$$y_2 = xe^{-1/4x^2} M \left( \frac{1}{2} a + \frac{3}{4}, \frac{3}{4}, \frac{1}{4} x^2 \right)$$

where $M (r,s,z) = \sum_{i=0}^{\infty} \frac{(r)i z^i}{(s)i i!}$, $\imath = 1, (r)i = r(r+1) \ldots (r+i-1)$ for $i > 1$.

For the voltage process
\[ q_1 = e^{-1/4y^2} N(0, \frac{1}{2}, \frac{1}{2} y^2) \]

\[ q_2 = ye^{-1/4y^2} N(\frac{1}{2}, \frac{3}{2}, \frac{1}{2} y^2) \text{ since } a = -1/2. \]

Hence

\[ q_1 = e^{-1/4y^2} \]

\[ q_2 = e^{-1/4y^2} \sum_{i=0}^{\infty} \frac{y^{2i+1}}{(2i+1)2^i i!} \]

Thus \( Aq_1 + Bq_2 \) is the general solution to (7.3.6). One of the boundary conditions is given by \( p(-1) = 0 \) which corresponds to

\[ A - B \sum_{i=0}^{\infty} \left( \frac{b^i}{(1 + a/b)} \right) \frac{2i+1}{(2i+1) 2^i i!} = 0. \]

Another boundary condition is \( p(1) = 0 \) which implies

\[ A + B \sum_{i=0}^{\infty} \left( \frac{b^i}{(1 - a/b)} \right) \frac{2i+1}{(2i+1) 2^i i!} = 0. \]

Only the trivial solution \( q_1 = q_2 = 0 \) satisfies these conditions exactly. However it might be that the family of functions \( Aq_1 + Bq_2 \) could still provide a reasonable approximation.

By varying \( A \) and \( B \), \( Aq_1 + Bq_2 \) could be forced into having the right location and satisfying \( p(-1) = 0 \), \( p(1) = 0 \). Such a procedure starts to involve as much work as the exact techniques and hence its use as an approximation is questionable.
The differential equation based on the probability density function (7.3.1) has not provided a convenient approximation. Hence (7.3.2) is considered. The equilibrium version of (7.3.2) is

\[
0 = -\theta \left\{ \frac{\epsilon^2}{q} \varnothing(\theta) + \epsilon \varnothing'(\theta) \right\} + \theta^2 \epsilon^2 (1 - \left( \frac{\epsilon}{q} \right)^2) \varnothing(\theta)
\]

\[
\frac{\varnothing'(\theta)}{\varnothing(\theta)} = -\delta + \epsilon \left( 1 - \left( \frac{\epsilon}{q} \right)^2 \right) \theta
\]

\[
\varnothing(\theta) = A \exp \left[ -\left( \frac{\epsilon}{q} \right)^q + \frac{\epsilon}{4} (1 - \left( \frac{\epsilon}{q} \right)^2, \theta^2) \right] \text{ for } \theta \neq 0
\]

However this approach gives an expression for the moment generating function, \( \varnothing(\theta) \) such that \( \varnothing(\theta) \to \pm \infty \) as \( \theta \to \pm \infty \). This is useless since any moment generating function has to satisfy

\( \varnothing(\theta) \to 0 \) as \( \theta \to \pm \infty \).

Since neither of the two equations representing the Markov diffusion approximation have produced useful results this approach is not developed any further.
7.4 A comparison of approximate solutions.

The approximations considered have fallen into two classes.

(i) Beta distributions - three distributions have resulted from Wray's report, Pearson curves and the differential equation of Section 7.2.

(ii) Gram-Charlier and Edgeworth expansions.

A simple comparison of the approximations resulting in Beta distributions can be made using the moments.

The Pearson curve is the only distribution to have the first four moments correct. The other two Beta distributions only have the mean correct. Calculating the first four moments over a typical spread of parameter values shows that Wray's approximation is nearly always better than that of Section 7.2. Despite this, large differences are still found between the true moments and the moments of Wray's function.

Using (7.2.8) the second moment of a Beta distribution is

\[ \frac{(r+s)(r+s+1) - 4rs}{(r+s)(r+s+1)} \]

and the mean is \[ \frac{r-s}{r+s} \]

Hence the variance is \[ \frac{4rs}{(r+s)(r+s+1)} \]. For Wray's approximation \[ r = \frac{q - \delta}{\alpha T} \]

and \[ s = \frac{q + \delta}{\alpha T} \]. Assuming unit time as before, \( T = 1 \), and the variance becomes \[ \frac{\alpha (q^2 - \delta^2)}{q^2 (2q + \alpha)} \]. The true variance is given by \[ q^2 (2q + \alpha) \]
(7.1.5). To compare the two consider the quotient of the exact variance $V_e$ and the approximate variance $V_w$.

$$
\frac{V_e}{V_w} = \frac{\varepsilon (2(1-q) - \varepsilon (1-2q)(2q + \alpha))}{(2-\varepsilon)(2q+\varepsilon(1-2q))}\alpha
$$

For the typical values of the parameters (6.3.7) terms in $\varepsilon^2$ can be ignored relative to terms in $\varepsilon$. Also $\alpha$ is related to $\varepsilon$ by $\varepsilon = 1 - e^{-\alpha}$ and so $\varepsilon \approx \alpha$. Hence

$$
\frac{V_e}{V_w} \approx \frac{2(1-q)2q}{2(2q)} = 1-q = 1/2.
$$

So even the variance is approximately twice the correct size in Wray's approximation. It can be shown that higher moments have even worse agreement in relative size.

Finally there is an exact analytic check that can be done on a certain set of parameters. Choosing $q = 1/2$ and setting $\delta = 0$ causes the G.E.C. equations (6.3.4) to become the equations (6.4.9) which relate to the equiprobable case. For this case Hill and Blanco (1973) have shown that the solution (6.4.8) holds in a certain region. Hence numerical comparisons can be made for these particular parameters. Some typical results are given in Table 7.4.1.

From both comparisons it seems that the Pearson curve is the best of the Beta distributions. However the analytic check makes it clear that if reasonable accuracy is required then the Pearson curve alone is not sufficient. Hence the curves can be used as a first iteration in the semicontraction mapping approach of Chapter 8.

The expansions in (ii) are not suited for use as approximations to the whole probability density function. This is because both
### TABLE 7.4.1

$q = \frac{1}{2}, \delta = 0, \epsilon = 0.12945$

<table>
<thead>
<tr>
<th>$x$</th>
<th>Pearson curve solution</th>
<th>Wray's solution</th>
<th>Analytic solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.75</td>
<td>$1.7535 \times 10^{-2}$</td>
<td>0.246143</td>
<td>$1.74097 \times 10^{-2}$</td>
</tr>
<tr>
<td>-0.8</td>
<td>$2.09 \times 10^{-3}$</td>
<td>0.148068</td>
<td>$7.16420 \times 10^{-3}$</td>
</tr>
<tr>
<td>-0.85</td>
<td>$1.8263 \times 10^{-4}$</td>
<td>0.075127</td>
<td>$2.26680 \times 10^{-3}$</td>
</tr>
<tr>
<td>-0.9</td>
<td>$1.0563 \times 10^{-5}$</td>
<td>0.027987</td>
<td>$4.47763 \times 10^{-4}$</td>
</tr>
<tr>
<td>-0.95</td>
<td>$3.39747 \times 10^{-7}$</td>
<td>$4.916572 \times 10^{-3}$</td>
<td>$2.79852 \times 10^{-5}$</td>
</tr>
</tbody>
</table>

### TABLE 7.4.2

$q = \frac{1}{2}, \delta = 0$

<table>
<thead>
<tr>
<th>$x$</th>
<th>$\epsilon$</th>
<th>Gram-Charlier expansion</th>
<th>Analytic solution</th>
<th>No. of moments required</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.4</td>
<td>0.29289</td>
<td>0.63896</td>
<td>0.618198</td>
<td>50</td>
</tr>
<tr>
<td>-0.6</td>
<td>0.2063</td>
<td>0.04219</td>
<td>0.037965</td>
<td>50</td>
</tr>
<tr>
<td>-0.75</td>
<td>0.12945</td>
<td>$-1.8326 \times 10^{-2}$</td>
<td>$1.74907 \times 10^{-2}$</td>
<td>50</td>
</tr>
<tr>
<td>-0.96</td>
<td>0.0358</td>
<td>$7.3428 \times 10^{-2}$</td>
<td>$4.9765 \times 10^{-25}$</td>
<td>80</td>
</tr>
</tbody>
</table>
Gram-Charlier and Edgeworth expansions require a large number of terms before their finite truncations are good approximations. However, they are more accurate than any of the Beta distributions and so can be used for individual values. Some examples corresponding to those in Table 7.4.1 are given in Table 7.4.2. The examples are solely from Gram-Charlier expansions due to the ease of implementation of this approach. Some shorter Edgeworth expansions have been computed. These give similar results but do not appear to converge at a much faster rate.

As can be seen from Table 7.4.2 the expansions are accurate but can be very slow to converge. This problem is clearly worse for the particular values of ε of interest for the voltage process.

Overall it would seem that none of the approximations are satisfactory in all cases. The Pearson curves provide simple functions over the whole range which can be refined by using the semicontraction mapping property. Gram-Charlier expansions can be used for individual values but have convergence problems for all parameter sets, especially as ε → 0. These techniques may be useful for some X variables of the general form (6.1.1), but in the cases considered it would appear that the exact solutions of Chapters 8 and 9 are just as convenient.
CHAPTER 8
ITERATIVE SOLUTIONS

8.1 A recursive procedure

In this Chapter extensions are considered to two approaches originally proposed by Hill and Blanco [1973] and Huzii and Sugiyama [1970]. Firstly a recursive procedure is assessed which seeks to use an analytic result valid in the tails of the distribution. This result can be used to evaluate the distribution elsewhere by using (6.3.3) or (6.3.4) as recursive equations. Secondly the semicontraction mapping principle is used to produce a convergent sequence of distribution functions. As mentioned in Chapter 7 this convergent sequence can use the approximations developed there as first iterations.

In Section 6.4 an exact recursive procedure due to Hill and Blanco [1973] was described. This applied to the equiprobable case with \( a = 2^{-1/K} \). For \( a \neq 2^{-1/K} \) such a method was not proven but seemed possible. For the voltage process the situation is similar although less clear. Consider the equations (6.3.4) in the region

\[
\frac{v-1+a}{a} \leq -1
\]

\[
F(v) = p_\text{-}F\left(\frac{v+1-a}{a}\right) \quad v \in [-1, -1+2\varepsilon] \quad (8.1.1)
\]

\[
G(v) = p_+F\left(\frac{v+1-a}{a}\right)
\]
\( F(v) = A(1+v)^K \) is clearly a solution in this region with the condition \( a^K = p_{-+} \). \( G(v) \) follows using \( G(v) = \frac{F(v)}{P_{-}} \). But whether this is the correct solution is not known. The symmetry of the equiprobable case which enabled this solution to be proven for \( a^K = 1/2 \) is missing here. Nevertheless numerical results shown in Section 9.3. show that if a solution of this form is accepted then a good approximation results.

Given that such a solution to (8.1.1) is accepted how viable are the recursive equations (6.3.3) and (6.3.4)? For the error detector problem such a recursion is prohibitive. This is shown by the following example. In Section 6.4 it was shown that \( f(o), g(o) \) are of special interest in the design problem. Hence the recursive process would need to take values of \( f(v), g(v) \) in \([-1, -1+2e]\) and iterate out to \( f(o) \) and \( g(o) \).

The two equations (6.3.3) can be rewritten to provide a recursion involving only one function. Letting \( w = \frac{v+1-a}{a} \) and rewriting only in terms of \( f(w) \) gives

\[
f(w) = \frac{a}{P_{+-+}} a f(a^2w-a^2+a^2+1-a) - \frac{p_{-}f(aw-1+a+2(1-a))}{a} \\
- \frac{p_{++}}{P_{+-+}} a f(aw-2(1-a)+1-a) - \frac{p_{-}f(w)}{a} \quad (8.1.2)
\]

Equation (8.1.2) can be used recursively if the four arguments of \( f \) are ordered. The four arguments are

\[ w, \ a^2w+(1-a)^2, \ aw-(1-a), \ \frac{aw-(1-a) + 2(1-a)}{a} \]
In the region $[-1, 0]$ the arguments are, in order
\[ a w - (1-a) \leq w < a^2 w + (1-a)^2 < a w - (1-a) + 2(1-a). \]

Hence if (8.1.2) is used recursively in $[-1, 0]$ the smallest distance moved at each stage is
\[ (a w - (1-a) + 2(1-a) - (a^2 w + (1-a)^2) \leq (1-a) (a w + 2 - 2a + a^2). \]

Rewriting the difference as $A w + B$ it is straightforward to show that the recursion moves a minimum distance of $B \sum_{i=0}^{n-1} (1+A)^i$ in $n$ steps.

This is assuming the recursion starts at zero. Hence to achieve values for $f(a)$, $g(a)$ it is necessary to perform $N$ recursions where
\[ N = \frac{B \sum_{i=0}^{n-1} (1+A)^i - 1 - 2\varepsilon}{1-A} \]

Now $A = a (1-a) = \varepsilon (1-\varepsilon)$ and
\[ B = \frac{1-a (2-2a+a^2)}{\varepsilon (1+\varepsilon)} = \varepsilon (1+\varepsilon) \]

Hence $N$ satisfies
\[ (1+\varepsilon (1-\varepsilon))^N \geq \frac{2 (1-\varepsilon)^2}{1+\varepsilon}. \quad (8.1.3) \]

For typical values of $\varepsilon$ in the range $10^{-5} < \varepsilon < 10^{-3}$, this corresponds to $691 < N < 69,312$. Hence an upper bound on the number of evaluations of $f(w)$ in $[-1, -1+2\varepsilon]$ is $3^N$ where $N$ satisfies (8.1.3).

This is because each recursion relates the value of $f(w)$ at one point to three other points.
Similarly a lower bound can be obtained. The largest distance moved at each stage is

\[ aw - (1-a) + 2(1-a) = \frac{2(1-a)}{a}. \]

Hence the recursion moves a maximum distance of \( \frac{2(1-a)}{a} \) m in m steps. To calculate \( f(o) \), \( g(o) \), \( M \) steps are required where

\[ \frac{2(1-a)}{a} M \geq 1-2\varepsilon \]

\[ (2\varepsilon)M \geq (1-2\varepsilon)(1-\varepsilon) \quad (8.1.4) \]

For typical values of \( \varepsilon \) in the range \( 10^{-5} < \varepsilon < 10^{-3} \) this corresponds to \( 498 < M < 49,998 \). Hence the lower bound on the number of evaluations of \( f(w) \) in \([-1, -1+2\varepsilon]\) is \( 3^M \) where \( M \) satisfies (8.1.4).

Clearly this procedure is impractical for these values of \( \varepsilon \). However for other applications where \( a \) is not so close to 1 it may be viable.

Since the numerical implementation of the recursion is prohibitive any attempt to develop formulae recursively, following Hill and Blanco [1973], would be pointless.
8.2 The semicontraction mapping principle.

The semicontraction mapping approach of Huzii and Sugiyama [1970], applied to certain examples of the equations (6.4.2) which could be expressed as a single equation. In fact the G.E.C. equations (6.3.4) also form a semicontraction mapping although they do not collapse to one equation. Some notation and a proof of this result follows.

Notation

Let \( R_1 = \{ F | F \) is a non-decreasing continuous function on \([-1, 1]\) such that \( F(x) = 0 \) for \( x \leq -1 \) and \( F(x) \) is constant and less than 1 for \( x > 1 \}\)

Let \( \rho_1 \) be the supremum metric on \( R_1 \). Hence

\[
\rho_1(F, G) = \sup_{-\infty < x < \infty} \{ | F(x) - G(x) | \} \quad \text{for} \quad F, G \in R_1
\]

\((R_1, \rho_1)\) forms a complete metric space as shown by Huzii and Sugiyama [1970] for a very similar metric space.

Define the set \( R \subseteq R_1 \times R_1 \) by

\[
R = \{ (F, G) | F+G \) is a distribution function, \( F, G \) \in R_1 \}
\]

Also define a distance measure \( \rho \) on \( R \) by

\[
\rho [(F, G), (J, K)] = \rho_1(F, J) + \rho_1(G, K).
\]

Proposition 8.2.1.

\((R, \rho)\) is a complete metric space.

Proof

It is obvious that \( R \) is a non-empty set and that \( \rho : R \times R \rightarrow \mathbb{R} \)

where \( \mathbb{R} \) is the set of real numbers. Hence to show \((R, \rho)\) is a metric
space it suffices to prove the three properties of a metric.

(i) \( \rho [(F,G), (J,K)] = \rho_1 (F,J) + \rho_1 (G,K) \geq 0 \) for any \( (F,G), (J,K) \in \mathbb{R} \) since \( \rho_1 \) is a metric on \( \mathbb{R}_1 \)

\[ \rho [(F,G), (J,K)] = 0 \iff \rho_1 (F,J) + \rho_1 (G,K) = 0 \]

\[ \iff F=J, G=K \text{ since } \rho_1 \text{ is a metric on } \mathbb{R}_1 . \]

Hence \( \rho [(F,G), (J,K)] = 0 \iff (F,G) = (J,K) \)

(ii) \( \rho [(F,G), (J,K)] = \rho_1 (F,J) + \rho_1 (G,K) \)

\[ = \rho_1 (J,F) + \rho_1 (K,G) \text{ for any } (F,G), (J,K) \in \mathbb{R} \text{ since } \rho_1 \text{ is a metric on } \mathbb{R}_1 . \]

\[ = \rho [(J,K), (F,G)] . \]

(iii) \( \rho [(F,G), (J,K)] + \rho [(J,K), (L,M)] \)

\[ = \rho_1 (F,J) + \rho_1 (G,K) + \rho_1 (J,L) + \rho_1 (K,M) \]

\[ \geq \rho_1 (F,L) + \rho_1 (G,M) \text{ for any } (F,G), (J,K), (L,M) \in \mathbb{R} \text{ since } \rho_1 \text{ is a metric on } \mathbb{R}_1 . \]

\[ = \rho [(F,G), (L,M)] . \]

Also, convergence in \( \mathbb{R} \) with respect to \( \rho \) is equivalent to convergence in \( \mathbb{R}_1 \) with respect to \( \rho_1 \). Hence convergence is uniform and \((\mathbb{R}, \rho)\) is a complete metric space. This is a simple generalization of the proof in Huzii and Sugiyama [1970].

Proposition 8.2.2.

The G.E.C. equations (6.3.4) represent a mapping \( U : \mathbb{R} \rightarrow \mathbb{R} \).
Proof.

For any \((F,G) \in \mathbb{R}\) define the mapping \(U\) on \(\mathbb{R}\) by

\[
U \left( F(x), G(x) \right) = p_+ F\left( \frac{x+1-a}{a} \right) + p_- G\left( \frac{x-1+a}{a} \right) \quad (8.2.3)
\]

\[
= \left[ F^*(x), G^*(x) \right]
\]

Since \((F,G) \in \mathbb{R}\) it follows that \(F^*(x)\) and \(G^*(x)\) are continuous and non-decreasing functions. If \(x < -1\), \(F^*(x) = G^*(x) = 0\) since

\[
\frac{x+1-a}{a} < -1, \quad \frac{x-1+a}{a} < -1.
\]

If \(x > 1\), \(F^*(x)\) and \(G^*(x)\) are constant since

\[
\frac{x+1-a}{a} > 1, \quad \frac{x-1+a}{a} > 1.
\]

\[
F^*(1) + G^*(1) = F\left( \frac{1+1-a}{a} \right) + G\left( \frac{1-1+a}{a} \right)
\]

\[
= F(1) + G(1)
\]

\[
= 1 \text{ since } F+G \text{ is a distribution function on } [-1, 1].
\]

Thus \((F^*,G^*) \in \mathbb{R}\), concluding the proof.

Theorem 8.2.4.

The mapping \(U\) defined by (8.2.3) is a contraction mapping for \(a > 1/2\).

Proof.

In the notation of Propositions 8.2.1. and 8.2.2., \(U\) is a mapping such that \(U : \mathbb{R} \to \mathbb{R}\) and \((\mathbb{R}, \rho)\) is a complete metric space. Hence it suffices to prove that the conditions of Definition 6.4.4. hold.
Let \((F,G),(J,K)\) be any elements of \(R\).

\[
\rho \left[ \bigcup (F,G), \bigcup (J,K) \right] = \rho \left[ (F^*, G^*), (J^*, K^*) \right]
\]

\[
= \rho_1 (F^*, J^*) + \rho_1 (G^*, K^*)
\]

\[
\rho_1 (F^*, J^*) = \sup_{-\infty < x < \infty} \left\{ \left| p_-(F-J) \left( \frac{v+1-a}{a} \right) + p_+(G-K) \left( \frac{x-1+a}{a} \right) \right| \right\}
\]

\[
\leq \sup_{-\infty < x < \infty} \left\{ \left| F(x)-J(x) \right| \right\} + \sup_{-\infty < x < \infty} \left\{ \left| G(x)-K(x) \right| \right\}
\]

\[
= \rho_1 (F,J) + \rho_1 (G,K)
\]

Similarly

\[
\rho_1 (G^*, K^*) \leq \rho_1 (F,J) + \rho_1 (G,K)
\]

Hence

\[
\rho \left[ \bigcup (F,G), \bigcup (J,K) \right] \leq \rho_1 (F,J) + \rho_1 (G,K) = \rho \left[ (F,G), (J,K) \right]
\]

Thus the first condition holds.

(ii) It is required to prove that for some \(m \in \mathbb{Z}^+\), \(\exists \right\), \(0 < r < 1\) such that

\[
\rho \left[ \bigcup^m (F,G), \bigcup^m (J,K) \right] \leq r \rho \left[ (F,G), (J,K) \right]
\]

for all \((F,G), (J,K) \in R\). The proof is included in Appendix I.

Hence the G.E.C. equations \((6.3.4)\) represent a semicontraction mapping in a certain metric space of distribution functions. This property is very useful since it guarantees the convergence of a sequence of distribution functions, \(\text{see (Proposition 6.4.6.)}\), to the unique solution, irrespective of the starting distribution used. The restriction of \(a > 1/2\) for this proof is unimportant since for \(a < 1/2\) the distribution is purely singular, and for the G.E.C. problem \(a\) is always very close to 1.

In formulating the basic equations \((6.3.3)\) and \((6.3.4)\) it was assumed that an equilibrium solution to \((6.3.2)\) existed. The semicontraction mapping property satisfied by \((6.3.4)\) ensures that
this equilibrium solution does exist for absolutely continuous distributions.

The semicontraction mapping property cannot be applied to the probability density functions (6.3.3) despite the similarity between (6.3.3) and (6.3.4). If the same sort of approach is attempted as in Theorem 8.2.4. then the equivalent of (i) is

\[ p \left[ U(f, g), U(j, k) \right] \leq \frac{1}{a} p \left[ (f, g), (j, k) \right] \]

Hence the first condition is not necessarily satisfied. Similarly it is difficult to satisfy condition (ii) except for values of \( a \) very close to 1. This is one of the drawbacks of this approach. If values of \( f(v) \) and \( g(v) \) are required then they have to be gained by approximate derivates of \( F(v) \) and \( G(v) \).
8.3 Implementation of the semicontraction mapping.

The semicontraction mapping can be implemented in two ways. Globally it can be used to give the whole distribution function. Also it can be used to give one specific value. To give the whole distribution function convergence must be fast and large values of \( \epsilon \) are required. Note that Huzii and Sugiyama (1970) include some graphs but only for values of \( \epsilon > 0.1 \). Individual values do not require such fast convergence but even so moderately large \( \epsilon \) values are needed. These comments apply irrespective of whether the approximate Beta functions of Chapter 7 are used as initial estimates. These initial estimates do help but even with them small values of \( \epsilon \) cause the convergence to be prohibitively slow.

The dependence of the speed of convergence on the value of \( \epsilon \) can easily be seen from (6.3.4). The distribution function at \( v \) is related to its values at

\[
\frac{v+1-a}{a} = \frac{v+\epsilon}{1-\epsilon} \quad \text{and} \quad \frac{v-1+a}{a} = \frac{v-\epsilon}{1-\epsilon}.
\]

Hence each iteration of (6.3.4) gives the new values at

\[
\frac{v \pm \epsilon}{1 - \epsilon}
\]

in terms of the old values at \( v \). For small values of \( \epsilon \) this means that each step only causes a small change to the function. Hence many iterations are required before the distribution function is substantially altered at \( v \).

The global approach can be implemented by starting with a piecewise linear distribution function. This can be based on a set of values given by the approximate methods of Chapter 7. Equation (6.3.4) can then be used stepwise to provide successive
approxi mates. For accurate results this must be done in a fine grid over the whole range [-1, 1]. The number of calculations required for this can become very large. This semicontraction mapping approach suffers from the same convergence problems as the recursive technique of Section 8.1.

For individual values the best technique is to set up a tree of values for \( v \) stemming from the point of interest. This ensures that the maximum number of iterations is used to reach that particular point. Schematically this is shown below.

![Tree Diagram](image)

Figure 8.3.1.

The nature of the tree implies the need for \( 2^n \) values of the initial distribution function if \( n \) iterations are used. Clearly only a few iterations are possible before this becomes impractical on a computer. The problems faced by this technique are purely computational. The method guarantees a unique solution independent of the initial distribution functions used. Hence if an infinite number of iterations were possible then the method would be as
accurate as required in all cases. The problems of convergence are exemplified by the following examples.

Figures 8.3.2 and 8.3.3 show that good results are possible for large \( \epsilon \) values. However even a small change in \( \epsilon \) can cause convergence to be impossible as shown in Figure 8.3.4. Similarly the tree approach can produce accurate values as shown in Table 8.3.5., but for smaller values of \( \epsilon \) convergence is not nearly reached. The analytic results given in Table 8.3.5 are from the solution (6.4.8). In Chapter 9 techniques are produced which can deal with these problem cases.

The applicability of the recursion of Section 8.1. is similar. In Section 9.3. it is shown that a solution of the form (6.4.8) seems to hold with reasonable accuracy in the tails of the distribution. Hence for large values of \( \epsilon \) the recursive procedure will be viable as discussed in Section 8.1.
Figure 8.3.2

Convergence reached after 22 iterations from a uniform initial estimate.

$q = 0.48$

$\delta = 0.05$

$\epsilon = 0.1$

---

Figure 8.3.3

Convergence reached after 10 iterations from a Pearson curve initial estimate.

$q = 0.48$

$\delta = 0.05$

$\epsilon = 0.1$
Figure 8.3.4

10 iterations have minimal
effect on uniform initial estimate.

\[ q = 0.48 \]
\[ \delta = 0.05 \]
\[ \epsilon = 0.00001 \]

<table>
<thead>
<tr>
<th>( p )</th>
<th>( q(p) )</th>
<th>( F(v) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2001</td>
<td>0.4800</td>
<td>0.43761</td>
</tr>
<tr>
<td>0.3094</td>
<td>0.6090</td>
<td>0.49204</td>
</tr>
<tr>
<td>0.4082</td>
<td>0.7831</td>
<td>0.54312</td>
</tr>
<tr>
<td>0.4800</td>
<td>0.8000</td>
<td>0.55555</td>
</tr>
<tr>
<td>0.6090</td>
<td>0.9000</td>
<td>0.60000</td>
</tr>
<tr>
<td>0.7831</td>
<td>1.0000</td>
<td>0.70000</td>
</tr>
</tbody>
</table>

Cases marked * indicate that convergence was not achieved by the
semi-contraction mapping in 10 iterations. These steps involve more
calls to the initial approximation which is closer for the example.
$F_T(v)$ denotes the solution from the tree approach to the semi-contraction mapping. $F_E(v)$ denotes the exact solution given by (6.4.8).

<table>
<thead>
<tr>
<th>$\epsilon$</th>
<th>v</th>
<th>$F_T(v)$</th>
<th>$F_E(v)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2063</td>
<td>-0.6</td>
<td>0.037924</td>
<td>0.037965</td>
</tr>
<tr>
<td>0.1294</td>
<td>-0.8</td>
<td>2.7931x10^-4</td>
<td>2.8657x10^-4</td>
</tr>
<tr>
<td>0.0611</td>
<td>-0.9</td>
<td>#</td>
<td>4.3381x10^-11</td>
</tr>
<tr>
<td>0.0358</td>
<td>-0.96</td>
<td>#</td>
<td>4.9765x10^-25</td>
</tr>
</tbody>
</table>

Cases marked * indicate that convergence was not reached by the semicontraction mapping in 12 iterations. Twelve steps involve 4096 calls to the initial approximation which is linear for this example.
In Chapter 7 several approximate solutions were developed for the G.E.C. problem. As discussed in Section 7.4 none of these approaches are sufficiently accurate to be used on their own. Hence the semicontraction mapping principle described in Chapter 8 can be used to refine these solutions iteratively. This approach has several advantages discussed in Section 8.2. Unfortunately for typical values of $\varepsilon$ given in (6.3.7) the rate of convergence is too slow to be of use. Hence the methods discussed so far are useful to the problem, but only for certain values of $\varepsilon$ outside the range of interest for the error detector.

In this Chapter methods of finding $f(v)$, $g(v)$ numerically are produced which are useful for even the smallest values of $\varepsilon$. These methods are based on the Fourier transforms of the functions $f(v)$ and $g(v)$. As shown in (7.1.2) all the moments of these functions are available. Hence the Fourier transform is also available. In fact there are two ways of producing the transforms as discussed in Section 9.1. Also the problem of retrieving the functions from the transforms can be performed in several different ways, some more suitable than others, depending on the parameters $q, \delta, \varepsilon$. 

Using the notation of Chapter 7

\[ \Phi(t) = \int_{-\infty}^{\infty} e^{i\nu t} f(\nu) d\nu, \Psi(t) = \int_{-\infty}^{\infty} e^{i\nu t} g(\nu) d\nu \]

\[ \Phi(t) = \sum_{n=0}^{\infty} \frac{(it)^n}{n!} \mu_n \]

Similarly

\[ \Psi(t) = \sum_{n=0}^{\infty} \frac{(it)^n}{n!} \lambda_n \]  \hspace{1cm} (9.1.1)

Since all the moments are available from (7.1.2) these infinite series can be truncated to \( s \) terms where \( s \) is large enough to ensure the required convergence. Since it is assumed that the parameters yield an absolutely continuous distribution such representations are valid and can be inverted to give approximations to \( f(\nu) \) and \( g(\nu) \).

Equations (9.1.1) give the Fourier transform by a truncation of an infinite series at each point. In Section 9.3 it is shown that this method has two potential drawbacks:

(i) Inaccuracy for small values of the Fourier transforms.

(ii) Expensive use of computer time when large values of \( s \) are required in (9.1.1).

An alternative method is to calculate \( \Phi(t), \Psi(t) \) at certain points using (9.1.1) and then use equations (7.1.1) iteratively to find \( \Phi, \Psi \) elsewhere. This is demonstrated below.
Rewriting (7.1.1) gives

\[
\begin{bmatrix}
\varnothing(t) \\
y(t)
\end{bmatrix} =
\begin{bmatrix}
p_{-e^{-i\epsilon t}} & p_{+e^{-i\epsilon t}} \\
p_{-e^{-i\epsilon t}} & p_{+e^{-i\epsilon t}}
\end{bmatrix}
\begin{bmatrix}
\varnothing((1-\epsilon)t) \\
y((1-\epsilon)t)
\end{bmatrix}
\]  

(9.1.2)

Equation (9.1.2) iterates outwards in the sense that if \(\varnothing((1-\epsilon)t), y((1-\epsilon)t)\) are known then (9.1.2) gives \(\varnothing(t), y(t)\). Repeated application of (9.1.2) gives the transforms at \(t\) in terms of their values at \((1-\epsilon)^n t\). Inverting the equation gives

\[
\begin{bmatrix}
\varnothing((1-\epsilon)t) \\
y((1-\epsilon)t)
\end{bmatrix} =
\begin{bmatrix}
p_{+e^{i\epsilon t}} & -p_{e^{i\epsilon t}} \\
-p_{-e^{i\epsilon t}} & p_{e^{i\epsilon t}}
\end{bmatrix}
\begin{bmatrix}
\varnothing(t) \\
y(t)
\end{bmatrix}
\]  

(9.1.3)

In a similar way (9.1.3) can be used to iterate backwards to give the transforms at \((1-\epsilon)^n t\) in terms of their values at \(t\).

The simplest way to implement such a procedure is to use (9.1.1) to give values of \(\varnothing(t), y(t)\) close to the origin. Then iterate out to give the transforms at all other points of interest. Clearly this may involve a considerable number of iterations. Some of the examples of interest shown in Section 9.3 require values of the transforms at \(t > 700\). If (9.1.2) is used to iterate out from \(t = 1\) to \(t = 700\) then \(N\) iterations are required where \(700(1-\epsilon)^N < 1\). For \(10^{-5} < \epsilon < 10^{-3}\) this corresponds to \(6,547 < N < 655,104\). With so many iterations the effect of rounding error in any computations has to be considered. For values of \(\epsilon\) in this range the above iterations were carried out on an ICL 2988 computer in double and quadruple precision. Even when \(\epsilon = 10^{-5}\) the results in both precisions agreed.
to ten or more decimal places. This seems good evidence that rounding errors are at worst $0(10^{-10})$. In fact this can be improved on by imposing stricter convergence criteria. Hence although this procedure uses considerable computer time, accuracy does not seem to be a problem. In fact results tend to be superior to the truncation method.

9.2 Inversion of the transforms.

The standard way of inverting the Fourier transforms is by the inversion formulae

\[
\begin{align*}
    f(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixt} \Phi(t) dt \\
    g(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixt} \Psi(t) dt
\end{align*}
\]  

(9.2.1)

Clearly for this example $\Phi$ and $\Psi$ are known numerically from the calculations in Section 9.1. Computationally $\Phi$ and $\Psi$ are known very accurately as demonstrated in Section 9.1. Gaining $f(x)$ and $g(x)$ to the same degree of accuracy is more difficult. Implementing the representation (9.2.1) involves numerical integration of $e^{-ixt}\Phi(t)$ and $e^{-ixt}\Psi(t)$ over the range $(-\infty < t < \infty)$. In the work that follows all the numerical integration has been performed by NAG routine D01GAF(1987).

In Section 9.3 it is shown that such numerical integration methods are very accurate for larger values of $\epsilon$ but impractical for the small $\epsilon$ values of interest. Also when the Fourier transform
exhibits a highly oscillatory nature numerical integration may be inaccurate.

To cope with the transforms from the error detector problem a different approach to inversion is required. A suitable technique can be developed from the Poisson Summation Formula (P.S.F.). Various versions of the P.S.F. are useful in different situations but the basic form is the formal relation

$$\sum_{n=-\infty}^{\infty} f(n) = \sum_{m=-\infty}^{\infty} F(2\pi m)$$

where $F$ is the Fourier transform of $f$ defined by

$$F(t) = \int_{-\infty}^{\infty} e^{itx} f(x) \, dx$$

In the development of this technique $f$ will refer to a general function and not to the error detector density function.

The formula has found applications in many areas of numerical work due to its linking of discrete values of functions and their transforms. In the engineering literature the P.S.F. is known as the basis of Shannon's sampling theorem concerning the transmission of information. (Papoulis [1962], Shannon [1949]). Several techniques have been developed to find values of a function or its transform based on the P.S.F. These methods have been summarized by Arsac [1966] in his book on distribution theory. This is an excellent introduction to the use of such methods in the study of probability density functions.

An alternative procedure has been suggested by Medhurst [1960] for the specific example of calculating radio frequency spectra. It is this approach which is developed in this Section. Medhurts
technique was essentially formal in nature and a general procedure was not given. His ideas are generalized in Theorem 9.2.8 where the approach is given a more rigorous treatment.

The usefulness of these applications is obviously dependent on the relation (9.2.2) being valid. Many sets of sufficient conditions have been given varying from the general conditions of Linfoot [1928], involving $C_1$ integrability, to the stricter conditions of Lighthill [1958]. A convenient set of conditions given by Titchmarsh [1937] states that (9.2.2) is valid if $f$ is bounded and continuous everywhere and $\int_{-\infty}^{\infty} f(x) \, dx$ exists.

In the following work it is assumed that (9.2.2) is a true relation and any further conditions are imposed to ensure the relevant procedures are valid.

The motivation behind these procedures lies in the possible difficulties in inverting or calculating Fourier transforms by numerical integration routines. When the integration is over the real axis or a large finite interval then quadrature can have storage problems. If the integrand is oscillatory then there can be numerical accuracy problems due to the addition and subtraction of similar areas. These problems are bypassed by the P.S.F. since it involves discrete rather than continuous values.

Arsac [1966] gives several versions of the P.S.F. from which the following can be obtained.

$$\sum_{n=-\infty}^{\infty} e^{-\ln^2} f(n\beta) = \frac{1}{\beta} \sum_{n=-\infty}^{\infty} \left[ \frac{2n\pi}{\beta} \right] \left( \beta - \gamma \right)$$  (9.2.3)
It is this representation which is used in the following work. The problem to be solved is the evaluation of the function \( f \) at the point \( \alpha \) in terms of its Fourier transform \( F \). Here \( \alpha \) is any real number and is not related to the error detector parameters. It should be noted that the subsequent procedure is easily amended to give the transform in terms of the function. The procedure is described in more detail in Smith, Clayton and Bennett [1988]. Before the technique is given some notation and an introductory Lemma are required.

**Notation**

\[
L (\beta, y) = \sum_{n=-\infty}^{\infty} e^{-i n \beta y} f(n \beta) = \frac{1}{\beta} \sum_{n=-\infty}^{\infty} \left( \frac{2n\pi}{\beta} - y \right)
\]

\[
\mathcal{M}_{2i+1} = \frac{1}{2} \left( L ((2i+1)\alpha, 0) - L ((4i+2)\alpha, 0) \right)
\]

\[
= (-1)^i \text{Im} \left\{ \text{sgn}(\alpha) \left( L \left( (2i+1)\alpha, \frac{\pi}{2\alpha(2i+1)} \right) \right) \right\}
\]

Using this notation the following Lemma is necessary.

**Lemma 9.2.4.**

\[
\text{If } \sum_{n=-\infty}^{\infty} d(|n|) f(n\alpha) \text{ is absolutely convergent, where } d(r) \text{ represents the number of divisors of } r, \text{ and } f \text{ satisfies the P.S.F. then } \sum_{i=0}^{\infty} \mathcal{M}_{2i+1} \text{ is absolutely convergent.}
\]
Proof.

The proof for $\alpha > 0$ is given, the proof for $\alpha < 0$ follows immediately.

Consider $M_{4j+1}$ for any non-negative integer $j$.

\[
M_{4j+1} = \frac{1}{2} \left[ \sum_{n=-\infty}^{\infty} f((4j+1)n\alpha) - \sum_{n=-\infty}^{\infty} f((8j+2)n\alpha) \right]
- \text{Im} \left\{ \sum_{n=-\infty}^{\infty} e^{-in\pi/2} f((4j+1)n\alpha) \right\}
\]

\[
= \frac{1}{2} \left[ \sum_{n=-\infty}^{\infty} f((4j+1)(2n+1)\alpha) + \sum_{n=-\infty}^{\infty} (-1)^n f((4j+1)(2n+1)\alpha) \right]
\]

\[
= \sum_{n=-\infty}^{\infty} f((4j+1)(4n+1)\alpha) \quad (9.2.5)
\]

Similarly $M_{4j+3}$ is given by

\[
M_{4j+3} = \sum_{n=-\infty}^{\infty} f((4j+3)(4n+3)\alpha) \quad (9.2.6)
\]

(9.2.5) and (9.2.6) are valid representations since the conditions of the Lemma ensures the absolute convergence of $L(\beta, y)$ for the relevant values of $\beta$ and $y$.

If \( \sum_{j=0}^{\infty} M_{4j+1} \) and \( \sum_{j=0}^{\infty} M_{4j+3} \) are absolutely convergent then the proof is complete. Consider \( \sum_{j=0}^{\infty} M_{4j+1} \) only, the proof for \( \sum_{j=0}^{\infty} M_{4j+3} \) following automatically.

Let $S_N$ be defined by

\[
S_N = \sum_{j=0}^{N} M_{4j+1} = \sum_{j=0}^{N} \sum_{n=-\infty}^{\infty} f((4j+1)(4n+1)\alpha) \quad (9.2.7)
\]
Since each $\mathbb{M}_{4j+1}$ is absolutely convergent (9.2.7) can be rearranged as follows.

$g(n\alpha)$ appears at most once in $\mathbb{M}_{4j+1}$ if $4j+1$ is a divisor of $n$.

Hence

$$|S_N| \leq \sum_{j=0}^{N} |\mathbb{M}_{4j+1}| \leq \sum_{n=-\infty}^{\infty} d_{4N+1}(|4n+1|) |f((4n+1)\alpha)|$$

where $d_{N}(x)$ is the number of divisors of $x$ that are less than or equal to $N$. Hence

$$|S_N| \leq \sum_{n=-\infty}^{\infty} d(|4n+1|) |f((4n+1)\alpha)| \forall N$$

$$\leq \sum_{n=-\infty}^{\infty} d(|n|) |f(n\alpha)| \forall N$$

Hence the conditions of the Lemma ensure that $\lim_{N \to \infty} S_N$ is an absolutely convergent series as required.

The conditions required for Lemma 9.2.4 are the absolute convergence of $\sum_{n=-\infty}^{\infty} d(|n|) f(n\alpha)$ as well as the validity of the P.S.F. Since $d(|n|)$ grows very slowly compared to $n \alpha$ the first of these conditions is not much more restrictive than the absolute convergence of $\sum_{n=-\infty}^{\infty} f(n\alpha)$. Hardy and Wright [1960] show that

$$\frac{d(n)}{n^\delta} \to 0 \text{ as } n \to \infty$$

for any positive $\delta$. Hence a more useful version of the condition could be to insist that $f(n\alpha) = O(n^{-1-\varepsilon})$ for some $\varepsilon > 0$. Using the
Theorem 9.2.8

If \( f(x) \) satisfies the P.S.F. and the conditions of Lemma 9.2.4 then

\[
f(\alpha) = \sum_{i=0}^{\infty} \delta_{2i+1} M_{2i+1}
\]

where \( \delta_s \) is defined by

\[
\delta_1 = 1
\]

For \( s>1, \delta_s = \begin{cases} (-1)^r & \text{if } |s| \text{ is a product of } r \text{ distinct primes} \\ 0 & \text{otherwise} \end{cases}
\]

Proof

The proof for \( \alpha > 0 \) is given, the corresponding result for \( \alpha < 0 \) follows immediately.

Let \( T_N \) be defined by

\[
T_N = \sum_{i=0}^{N} \delta_{4i+1} M_{4i+1} + \sum_{i=0}^{N} \delta_{4i+3} M_{4i+3} = \sum_{i=0}^{N} \delta_{4i+1} \sum_{n=-\infty}^{\infty} f((4i+1)(4n+1)\alpha) + \sum_{i=0}^{N} \delta_{4i+3} \sum_{i=0}^{\infty} f((4i+3)(4n+3)\alpha)
\]

Since each \( M_{2i+1} \) is absolutely convergent \( T_N \) can be rewritten,

\[
T_N = \sum_{m=-\infty}^{\infty} C_{4n+1,N} f((4n+1)\alpha)
\]

where coefficients \( C_{p,N} \) are defined for \( p \equiv 1 \pmod{4} \) by

\[
C_{p,N} = \sum_{\delta_{2i+1} \mid p, 0 < 2i+1 \leq 4N+3}
\]

(9.2.9)
The summation here is over all odd positive divisors of \( p \) that are less than or equal to \( 4N+3 \).

From the definition of \( \delta_s \) (9.2.9) gives

\[
C_{1,N} = \delta_1 = 1
\]

\[
C_{p,N} = \sum_{s \leq 4N+3} (-1)^s + \delta_1 \text{ for } |p| > 1.
\]

The summation is over all values of \( s \) that are odd, positive, divisors of \( p \), less than or equal to \( 4N+3 \) and the product of \( r \) distinct primes. Hence

\[
C_{p,N} = \sum_{r=1}^{t} \left( \begin{array}{c} \frac{t}{r} \\ r \end{array} \right) (-1)^r + 1 \text{ for } |p| \leq 4N+3
\]

where \( t \) is the number of prime divisors of \( p \). So for \( |p| < 4N+3 \)

\[
C_{p,N} = \sum_{r=0}^{t} \left( \begin{array}{c} \frac{t}{r} \\ r \end{array} \right) (-1)^r = 0
\]

For \( |p| > 4N+3 \) it is clear that

\[
|C_{p,N}| \leq d(|p|). \text{ Hence } T_N \text{ can be written}
\]

\[
T_N = f(\alpha) + r_N
\]

where \( r_N \) satisfies

\[
|r_N| \leq \sum_{|n| > N} d(|4n+1|) \left| f((4n+1)\alpha) \right|
\]

From Lemma 9.2.4 it is clear that \( \lim_{n \to \infty} r_n = 0 \) and that \( \sum_{i=0}^{\infty} \delta_{2i+1} M_{2i+1} \) is absolutely convergent. Hence the result follows

\[
f(\alpha) = \lim_{N \to \infty} T_N = \sum_{i=0}^{\infty} \delta_{2i+1} M_{2i+1} \tag{9.2.10}
\]
Hence Theorem 9.2.8 gives a formulation for \( f(\alpha) \) in terms of an infinite sum of discrete values of the Fourier transform. This avoids some of the problems of oscillatory integrands and storage requirements involved in numerical integration. However further improvements can be made by considering different methods of implementing this approach.

**Implementation of Theorem 9.2.8.**

The great advantage of this approach is that the Fourier transform only needs to be calculated at the discrete set of points

\[
\frac{2\pi}{\alpha} - y, \quad -\infty < n < \infty.
\]

For probability density functions the Fourier transform tends to zero as the variable tends to infinity. Hence each infinite series \( 0^n \) or \( 1^n \) involving the transforms can be truncated to a finite form. Equation (9.2.10) however involves a double summation. If the expressions

\[
\frac{1}{\delta} \delta_{4i+1} \begin{bmatrix} 0^{4i+1} - 0^{8i+2} & -\text{sgn}(\alpha) \cdot 1^{4i+1} \end{bmatrix}
\]

and

\[
\frac{1}{\delta} \delta_{4i+3} \begin{bmatrix} 0^{4i+3} - 0^{8i+6} & +\text{sgn}(\alpha) \cdot 1^{4i+3} \end{bmatrix}
\]

are considered as single operations then (9.2.10) represents an infinite series of such operations. As shown above each operation can be truncated to a finite number of calculations. Theorem 9.2.8 shows that the number of operations can also be truncated to give a convergent answer. It is useful however to have a guide to the number of operations required.

In what follows the operations are numbered as follows:
operation 1 \[-\frac{1}{2} \delta_1 \left( 0 - 0^2 - \text{sgn}(\alpha)I^2 \right)\]

operation 2 \[-\frac{1}{2} \delta_3 \left( 0^3 - 0^6 + \text{sgn}(\alpha)I^3 \right)\]

operation 3 \[-\frac{1}{2} \delta_5 \left( 0^5 - 0^{10} - \text{sgn}(\alpha)I^5 \right)\]

operation 4 \[-\frac{1}{2} \delta_7 \left( 0^7 - 0^{14} + \text{sgn}(\alpha)I^7 \right)\]

operation 5 \[-\frac{1}{2} \delta_{11} \left( 0^{11} - 0^{22} + \text{sgn}(\alpha)I^{11} \right)\]

etc.

Note that operation 5 is not \[-\frac{1}{2} \delta_9 \left( 0^9 - 0^{18} - \text{sgn}(\alpha)I^9 \right)\]

since \(\delta_9 = 0\) and this operation is not required. The successive gain of each operation is shown below by listing the nearest term left to the required \(f(\alpha)\) term. For example operation 1 results in

\[f(\alpha) + f(-3\alpha) + f(5\alpha) + f(-7\alpha) + \ldots\]

The nearest term to \(f(\alpha)\) is \(f(-3\alpha)\).

Table 9.2.1.

<table>
<thead>
<tr>
<th>No. of operations</th>
<th>Nearest term</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-3\alpha</td>
</tr>
<tr>
<td>2</td>
<td>5\alpha</td>
</tr>
<tr>
<td>3</td>
<td>-7\alpha</td>
</tr>
<tr>
<td>4</td>
<td>-11\alpha</td>
</tr>
<tr>
<td>5</td>
<td>13\alpha</td>
</tr>
<tr>
<td>6</td>
<td>-15\alpha</td>
</tr>
<tr>
<td>7</td>
<td>17\alpha</td>
</tr>
<tr>
<td>8</td>
<td>-19\alpha</td>
</tr>
</tbody>
</table>
For the voltage process \( f(v) \) varies on \([-1, +1]\). Hence when the nearest term \( f(\pm\alpha) \) satisfies
\[
| \alpha | > 1
\]
the remainder terms are all zero. This has the consequence that, for example, \(| \alpha | > 1/3\) requires only one operation. For any \(| \alpha | > 0.05263\) at most eight iterations are required. Also this is a conservatively high estimate since \( f(x) \) can be very close to zero even inside the range \([-1, +1]\). This is especially true of the probability density function of the voltage process which was shown to behave similarly to a Dirac delta function in Section 7.1.

The conclusion is that except for very small values of \( \alpha \) a small number of operations will suffice. This is important since the truncations of \( 0^n \) and \( 1^n \) can contain a great many terms and it would slow down the procedure if such series had to be calculated many times. If \( \alpha \) is very small then a large number of operations may be required to achieve convergence. However this is not usually a problem since reducing the size of \( \alpha \) has the effect of reducing the number of terms required in \( 0^n \) and \( 1^n \). For very small \( \alpha \) the transforms at \( \frac{2\pi k}{\alpha} \) will soon tend to zero.

So far the problem of finding the function at zero has not been considered. This is of special interest to the voltage process as discussed in Section 6.4. Theorem 9.2.8 gives a method for any \( \alpha \neq 0 \) and the same sort of approach can be used for \( \alpha = 0 \) as shown below.

Calculating \( f(0) \) from the Poisson summation formula.

1. A very simple method would be to set \( \alpha = 1 \) and then
\[ f(a) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} f(k) = \sum_{l=-\infty}^{\infty} F(2\pi l). \]

This is similar to numerical integration and may suffer from the same problems if \( F(x) \) is highly oscillatory or is very slow to decay. Otherwise this method is usually acceptable.

(ii) A more complicated method is based on shifting the functions before taking Fourier transforms. In fact this approach can be useful to the Poisson summation technique for all values of \( \alpha \), not only zero.

Consider the error detector problem. Instead of the normal functions \( f(v), g(v) \) consider the shifted functions \( f^*(v) \) and \( g^*(v) \) given by
\[ f^*(v) = f(v + \mu_1), \quad g^*(v) = g(v + \lambda_1) \quad (9.2,12) \]

Let their Fourier transforms be \( \hat{f}^*(x) \) and \( \hat{g}^*(x) \). Then
\[ \hat{f}^*(x) = e^{-i\mu_1 x} \hat{f}(x), \quad \hat{g}^*(x) = e^{-i\lambda_1 x} \hat{g}(x) \quad (9.2,13) \]

where \( \hat{f}(x) \) and \( \hat{g}(x) \) are the Fourier transforms of \( f(v) \) and \( g(v) \).

Considering these shifted functions has two advantages.

(i) The oscillations in the Fourier transforms of \( f(v) \) and \( g(v) \) caused by \( \mu_1 \neq 0 \) and \( \lambda_1 \neq 0 \) are removed. The resulting transform is smoother and may be more amenable to accurate calculation.

(ii) If \( f(0), g(0) \) are required then this corresponds to \( f^*(-\mu_1) \) and \( g^*(-\lambda_1) \). These values can be gained by application of Theorem 9.2.8 since the points of interest are now non-zero.

Examples of this approach are given in Section 9.3.
The Vandemonde matrix approach

A third method of inversion is based on a different approach to dealing with the Poisson summation formula. Theorem 9.2.8 gives one method of recovering \(f(\alpha)\) from the formula (9.2.3). There is another approach based on inverting a Vandemonde matrix. This is suggested by the property (9.2.11).

Suppose that \(f(\alpha)\) is required when \(\alpha \neq 0\) that \(f(x)\) is non-zero in the range \([a, b]\). Choose \(l\) such that \(l \alpha = \max (a, b)\). If \(f(x)\) satisfies these properties then (9.2.3) becomes

\[
\int_{-\infty}^{\infty} e^{-i\alpha \gamma} f(n\alpha) = 1 \sum_{|\alpha| = 1}^\infty F \left( \frac{2n\pi}{\alpha} - \gamma \right)
\]

Let \(H(\alpha, \gamma) = \int_{-\infty}^{\infty} e^{-i\alpha \gamma} f(n\alpha) = 1 \sum_{|\alpha| = 1}^\infty F \left( \frac{2n\pi}{\alpha} - \gamma \right)\).

Choose \(2l+1\) different values for \(\gamma\) denoted by \(\gamma_1, \gamma_2, \ldots, \gamma_{2l+1}\), such that \(e^{-i\alpha \gamma} \neq e^{-i\alpha \gamma'}\) for \(\gamma \neq \gamma'\). Then the Poisson summation formula can be written in matrix form

\[
\begin{bmatrix}
-1 1 & 1 1 \\
1 & \ddots & \ddots & \ddots \\
\vdots & \ddots & \ddots & \ddots \\
1 & \ddots & \ddots & \ddots \\
1 & \ldots & 1 1 & 1
\end{bmatrix}
\begin{bmatrix}
f(-\alpha) \\
f(\alpha) \\
f(0) \\
f(-\alpha)
\end{bmatrix}
= \begin{bmatrix}
H(\alpha, \gamma_1) \\
\vdots \\
\vdots \\
H(\alpha, \gamma_{2l+1})
\end{bmatrix} \tag{9.2.14}
\]

For ease of notation denote (9.2.14) by

\[
Yf = H \tag{9.2.15}
\]

\(Y\) is similar to a Vandemonde matrix and is non-singular since
Let $e^{i\alpha_k} = e^{i\alpha}$, then

$$Y = \begin{bmatrix}
  a_{-1} & \cdots & a_1 \\
  \vdots & \ddots & \vdots \\
  a_{-1} & \cdots & a_1 \\
  21+1 & 21+1 & \ldots & 21+1
\end{bmatrix}$$

$$\det Y = \prod_{k=1}^{21+1} a_k^{-1} (\det Z)$$

where $Z$ is a Vandemonde matrix given by

$$Z = \begin{bmatrix}
  a_1 & \cdots & a_1 \\
  \vdots & \ddots & \vdots \\
  a_{21+1} & \cdots & a_{21+1}
\end{bmatrix}$$

Using a standard result for the determinant of a Vandemonde matrix, it follows that

$$\det Y = \prod_{k=1}^{21+1} a_k^{-1} \prod_{r<s} (a_r - a_s).$$

Hence the non-singularity of $Y$.

The solution of (9.2.15) can be produced analytically or numerically. A general form for the inverse of Vandemonde matrices can be derived as shown in Appendix III. Alternatively it may be easier to use a computer package which handles systems of linear equations.

Clearly the drawback here will be the size of the matrix $Y$. For large $\alpha$ values it will be small and vice versa. In many ways the method is similar to that of the Poisson summation formula. Both techniques derive $f(\alpha)$ in terms of linear combinations of infinite series in the Fourier transforms of $f(x)$. Also both techniques require more complicated and lengthy linear combinations when $\alpha$ is
small. In general the technique would seem to be as useful as that of Theorem 9.2.8 for moderate $\alpha$ but less efficient for small $\alpha$. Some examples are given in Section 9.3. It does have the advantage of giving values of $f(n\alpha)$ for $n = -1, \ldots, 0, \ldots 1$ rather than just $f(\alpha)$.

9.3 A comparison of the Fourier transform techniques

Several methods have been produced in Chapter 9 to gain the probability density function of the voltage process via its Fourier transform. Some indications of the relative importance of these techniques have already been given. In this section some examples are given and some numerical comparisons made. Also the applicability of the methods for various values of the parameters is investigated.

Firstly some Fourier transforms are illustrated to show the behaviour of the transforms for different values of the parameters. All the transforms shown have been calculated by the iterative method of Section 9.1. In all cases except where stated only the real part of $\phi(t)$ is given.

Figures 9.3.1 and 9.3.2 suggest that increasing $\delta$ increases the oscillations in the transform. This is to be expected since increasing $\delta$ is equivalent to increasing the non-zero first moments. Figures 9.3.3 and 9.3.4 suggest that increasing $q$ has the effect of spreading the transform over a wider range, "stretching" the transform. Several other plots have confirmed this. Similarly decreasing the value of $\epsilon$ seems to have the same effect, Figures
Figure 9.3.1

\[
\text{Re}\{\phi(t)\}
\]

\[
q = \frac{1}{6}
\]

\[
\sigma = 0.05
\]

\[
\epsilon = 0.053
\]
Figure 9.3.2

\[ \text{Re}\{\phi(t)\} \]

- \( q = \frac{1}{6} \)
- \( \theta = 0.2 \)
- \( \epsilon = 0.053 \)
$q = 0.3$

$\delta = 0$

$\epsilon = 0.053$
Figure 9.3.4

The Fourier transforms based on the typical values (9.3.5) are shown in Figures 9.3.5 and 9.3.6. They illustrate the large range of the transforms for these values. The problems with numerical integration described in Section 9.3.2 are certainly relevant here. If the integral is required then (9.3.12) must be integrated over the whole range. When attempting to do this for Figure 9.3.5 it is found that significant modifications occur in the tails out to $t = 150$. This causes problems with accuracy due to the addition and subtraction of small areas resulting in a net error of the same order of magnitude.

For these reasons the Poisson summation formula is a more suitable approach for the error detection process.

$q = 0.7$

$\delta = 0$

$c = 0.053$

This is shown by Figure 9.3.12. The advantage of shifting can be seen by considering Figure 9.3.13. Here the transform decays to zero at $t = 0.6$. Here, soon after $t = 0.125$ the transform will be negligible. The unshifted transform however will have significant values near $0.125$ due to the oscillations. Also there is an advantage in the calculation of $C_n^p$ for the Poisson summation formula. If shifted the form will go to $C_n^p$ as a series of positive values. This means the unshifted transform will also go to $C_n^p$ as a series of positive values which are easy to sum.
9.3.5, 9.3.6, and 9.3.7. Again this is expected since $\mu_n$ and $\lambda_n$ decay more slowly for large $q$ and small $\varepsilon$.

Some Fourier transforms based on the typical values (6.3.7) are shown in Figures 9.3.8 and 9.3.9. They illustrate the large range of the transforms for these values. The problems with numerical integration described in Section 9.2 are certainly relevant here. If $f(0)$ is required then $\mathcal{0}(t)$ must be integrated over the whole range. When attempting to do this for Figure 9.3.8 it is found that significant oscillations occur in the tails out to $t = \pm 700$. This causes problems with accuracy due to the addition and subtraction of small areas resulting in a solution of the same order of magnitude. For these reasons the Poisson Summation Formula is a more suitable approach for the error detector problem.

In Section 9.2 it was suggested that shifting the probability density functions by (9.2.12) gave the transforms smoother properties. This is shown by Figures 9.3.11 and 9.3.12. The advantage of shifting can be seen by considering Figure 9.3.12. Here the transform decays to zero smoothly. Hence soon after $\pm 120$ the transform will be negligible. The unshifted version however will have significant values after $\pm 140$ due to the oscillations. Also there is an advantage in the calculation of $\mathcal{O}^n$ for the Poisson summation formula. The shifted transform will give $\mathcal{O}^n$ as a series of positive values decaying to zero. The unshifted transform will give $\mathcal{O}^n$ as a series of positive and negative values which may be less accurate.

So far all the transforms have been calculated from the iterative method of (9.1.2). Figures 9.3.1 to 9.3.9 were also calculated using
Figure 9.3.5

\[ q = \frac{1}{6} \]
\[ \sigma = 0 \]
\[ \epsilon = 0.15 \]

\[ \text{Re}\{\phi(t)\} \]
Figure 9.3.6

\[ q = \frac{1}{4} \]
\[ \delta = 0 \]
\[ \epsilon = 0.053 \]

\[ \text{Re}\{\phi(t)\} \]
Figure 9.3.7

\[ q = \frac{1}{2} \]
\[ \delta = 0 \]
\[ \epsilon = 0.02 \]
Figure 9.3.8

\[ q = 0.48 \]
\[ \delta = 0.05 \]
\[ \epsilon = 0.001 \]
Figure 9.3.9

\[ q = 0.48 \]

\[ \delta = 0.05 \]

\[ \epsilon = 0.0001 \]
Figure 9.3.11

\[ q = 0.48 \]
\[ \delta = 0.05 \]
\[ \epsilon = 0.001 \]
The truncation method is used. For moderate values of $s$ (Figures 9.3.11 or 9.3.9) both methods attain to a high degree of accuracy. For Figures 9.3.4 and 9.3.7 the truncation method is less accurate. In the table, more important oscillations contribute to the iterative integral. Hence for small $s$, the iterative method is preferable and has been used in all subsequent calculations.

Figures 9.3.4 to 9.3.7 give an indication of the behaviour of the transform for moderate values of the parameter $q$. In particular, Figures 9.3.6 and 9.3.7 illustrate the difficulty of inverting by numerical integration. Therefore the various methods of solution the analytic solution of 4.51 in a nearly identical. For this to be true, it is necessary change the parameters as below

$q = 0.48$

$s = 0.05$

$e = 0.001$

The results are recorded in the following table:

<table>
<thead>
<tr>
<th>$t$</th>
<th>$Re{\exp(\text{i}u_1t)\phi(t)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

The values in the table are computed by the exact solution.
the truncation method of (9.1.1). For moderate values of $\varepsilon$ (Figures 9.3.1 to 9.3.7) both methods coincide to a high degree of accuracy. For Figures 9.3.8 and 9.3.9 the truncation method is less accurate in the tails where important oscillations contribute to the inversion integral. Hence for small $\varepsilon$ the iterative method is preferable and has been used in all subsequent calculations.

Figures 9.3.1 to 9.3.19 give an indication of the behaviour of the transforms for various values of the parameters $q$, $\delta$, $\varepsilon$. In particular Figures 9.3.8 and 9.3.9 illustrate the difficulty of inversion by numerical integration. To compare the various methods of solution the analytic solution (6.4.8) is a useful baseline. For this to be valid it is necessary to set the parameters as below

$$q = 1/2, \quad \delta = 0, \quad \varepsilon = 1 - 2^{-1/K}$$

The techniques compared are:

- **E** - exact solution given by (6.4.8).
- **N** - numerical integration of transform.
- **P** - Poisson summation technique given by Theorem 9.2.8.
- **V** - Vandemonde matrix approach given by 9.2.15.

The results are recorded in the following table.

**Table 9.3.1.**

<table>
<thead>
<tr>
<th>K</th>
<th>$X$</th>
<th>$f(x)$</th>
<th>$N$</th>
<th>$P$</th>
<th>$V$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>$-1 + 2\varepsilon$</td>
<td>0.3029575314</td>
<td>0.3029577124</td>
<td>0.3029576297</td>
<td>0.3029576308</td>
</tr>
<tr>
<td>7</td>
<td>$-1 + 2\varepsilon$</td>
<td>$4.603782852 \times 10^{-4}$</td>
<td>$4.603801297 \times 10^{-4}$</td>
<td>$4.603779218 \times 10^{-4}$</td>
<td>$4.603780123 \times 10^{-4}$</td>
</tr>
<tr>
<td>13</td>
<td>$-1 + 2\varepsilon$</td>
<td>$1.570607242 \times 10^{-10}$</td>
<td>$1.571245320 \times 10^{-10}$</td>
<td>$1.570512492 \times 10^{-10}$</td>
<td>$1.570513009 \times 10^{-10}$</td>
</tr>
</tbody>
</table>
In all the preceding examples the transforms have been $\varnothing(t)$ the transform of $f(v)$ of the voltage process. For completeness the case $q = 1/2$, $\delta = 0.1$, $\epsilon = 0.2$ is given in detail.

Hence the real and imaginary parts of $\varnothing(t)$ and $\varphi(t)$ are given as well as the resulting densities.

The value of $\epsilon$ was chosen to be high so that the densities $f(v)$ and $g(v)$ could be plotted fairly quickly. The densities were gained at 20 points in $[-1, +1]$ by numerical integration of the transforms. The graphs are given in Figures 9.3.14 to 9.3.19.

Figures 9.3.11 and 9.3.12 also show the behaviour of the function $f(v)$ suggested in Sections 7.1. For small values of $\epsilon$ it was shown that $p(v) = f(v) + g(v)$ would behave like a Dirac delta function. $f(v)$ and $g(v)$ also share this property. Integrating Figure 9.3.11 would give $f(0)$. Integrating Figure 9.3.12 would give $f(\mu_1)$. It is clear from the graphs that $f(\mu_1) >> f(0)$. Hence even though $\mu_1 \approx 0.1$ the value of $f(v)$ at zero is already much smaller than that at $\mu_1$. This property is shown graphically for a more general example in Section 10.4.

Table 9.3.1 shows the order of accuracy for all three methods $(N, P, V)$. Clearly for these parameter values the exact solution is being gained with errors of $O(10^{-10})$ for all three approaches. However the smallest value of $\epsilon$ used was $1 - 2^{-1/13} = 0.0519$. This is still of moderate size and hence inversion by numerical integration is accurate. Also $x$ has to be in the tails of the distribution for the exact solution to be valid. This means that the
Figure 9.3.14

$\text{Re}(\phi(t))$
Figure 9.3.15

The diagram represents the function $\text{Im}(\phi(t))$ over the range of $t$ from $-35$ to $35$. The y-axis is labeled $\text{Im}(\phi(t))$ and the x-axis is labeled $t$. The curve peaks at a maximum value of approximately 0.5 and minimum values close to -0.5.
Figure 9.3.16
Figure 9.3.17

Image of a graph showing the imaginary part of a function $\text{Im}(\varphi(t))$.
Figure 9.3.19

Warendooh matrix solution requires only a $2 \times 2$ matrix inversion. Hence all three methods give good results.

Consider Figure 9.3.9 where the parameters are

\[ a = 0.4, \quad b = 0.01, \quad c = 0.0001 \]

This is a typical parameter set and poses difficulties diagnosed previously. Here numerical integration is difficult on a practical level due to massive problems. Also for very small $v$ of $e_{100}$ is calculated as a negative solution. Clearly the oscillations in the tails are causing errors to grow in the solution. Similarly the Warendooh matrix approach becomes less useful since in calculating $I(s)$ for $0.4 < 0.01$ a matrix inversion of at least $300 \times 300$ is required. The Poisson summation rule seems to perform reasonably well even for these difficult cases where numerical integration is possible there is always close agreement with the E.G.F. But for small $v$ values only the E.G.F. is a consistent method.

The exact solution (6.4.8) values in Table 9.1 has only been proven for the case $a = 0.1$. However Rice [12] suggested that the general form of solution is the correct for general $a$. If this is the case then (6.4.8) for $x = 0.01$ at least must be expected to constant where $e$ is in a rate manner. To study this possibility values of $\phi(x, x' = 1)$ were looked at for varying $x$. The values were used for $a = 0.1$.
Vandemonde matrix solution requires only a 3 x 3 matrix inversion. Hence all three methods give good results.

Consider Figure 9.3.9 where the parameters are

\[ q = 0.48, \quad \delta = 0.05, \quad \epsilon = 0.0001. \]

This is a typical parameter set and poses difficulties discussed previously. Here numerical integration is difficult on a practical level due to storage problems. Also for many values of \( x \) \( f(x) \) is calculated as a negative solution. Clearly the oscillations in the tails are causing errors to swamp the solution. Similarly the Vandemonde matrix approach becomes less useful since in calculating \( f(x) \) for \( 0 < x < 0.01 \) a matrix inversion of at least 200 x 200 is required. The Poisson Summation Formula seems to perform reasonably well even for these difficult cases. Where numerical integration is possible there is always close agreement with the P.S.F. But for small \( \epsilon \) values only the P.S.F. gives consistent results.

The exact solution (6.4.8) provided in Table 9.3.1 has only been proven for the case \( a = 2^{-1/K} \). However Rice [1957] suggested that the general form of solution may be correct for general \( a \). If this is the case then for fixed \( a \) the quotient \( f(x) (1+x)^{-q} \) should be constant where \( q \) is a real number. To study this possibility values of \( f(x) (1+x)^{-q} \) are tabulated below for varying \( x \). The values of \( f(x) \) are gained by numerical integration.

**Table 9.3.2**

The example tabulated below has parameters

\[ a = 0.8, \quad p_{--} = 1/2, \quad p_{+} = 1/2, \quad p_{++} = 0.4, \quad p_{++} = 0.6. \]
The tabulated values are similar but not as close as expected were the solution an exact one. Experience with numerical integration suggests that $f(x)$ is correct to within 8 decimal places. Hence similar accuracy is expected of $f(x)(1+x)^{-q}$. The agreement is not of this order and suggests empirical evidence that the solution is only an approximate one.

The results of Chapters 7, 8 and 9 imply the following use of the techniques discussed (Figure 9.3.20).

The problem which remains unsolved is the calculation of the distribution function for small values of $\varepsilon$. At present only numerical integration of the density is feasible. But this is impractical since each value of the density is computationally intensive.

Insofar as the error detector problem is based on a knowledge of $p(0) = f(0) + g(0)$ then the P.S.P. provides a complete solution. However as discussed in Section 6.4 a more precise approach uses a 4 state model and this is considered in Section 10.4.
Figure 9.3.20

In Chapter 6 to 9, the problem studied was based on the error detector example. In Chapter 6 it was shown that designing the error detector relies on a knowledge of the voltage process produced by the task, which is shown in Figure 8.2.1. This problem considered 1, 2, and 3 errors of these. In Chapter 6 the technique used for error detector problems provided a Poisson summation formula (6.3.1). A schematic of the application of this technique is given in Figure 9.3.1. The derivation of the Poisson summation formula (6.3.1) is also given.

It is with reference to the chapter on general proof in constrained and the details in Chapter 7...
CHAPTER 10

THE GENERAL RANDOM GEOMETRIC SERIES

10.1 INTRODUCTION

In Chapters 6 to 9 the problem studied was based on the error detector example. In Chapter 6 it was shown that designing the error detector relied on a knowledge of the voltage process produced by the leaky integrator in Figure 6.2.1. This in turn required the evaluation of the density of a random variable satisfying (6.3.3) which can be expressed as a random geometric series (6.3.6).

Various approaches to calculating this density have been considered in Chapters 7 to 9. Figure 9.3.20 summarizes the uses of these methods. In this Chapter the techniques developed for the error detector problem are generalized where possible to the general case.

The general random geometric series is given in (6.1.1) and some of its applications are summarized in Section 6.4. In particular its relevance to intersymbol interference is of interest since the application of ARMA time series models to coding and interference problems provided the initial motivation.

In Section 10.2 the semi-contraction mapping approach of Chapter 8 is extended to the general case. A general proof is complicated and the details are given in Appendix III. Section 10.3 uses the methods of Chapter 9 and outlines the use of Fourier transforms and the P.S.F. Finally in Section 10.3 some examples of the general case are given and various methods of solution illustrated.
As in the case of the error detector problem assumptions have to be made concerning the distributions of interest. Throughout Chapter 10 it is assumed that the states of the Markov pulse train \((\alpha_1, \ldots, \alpha_m)\) and the constant \(a\) are such that the distribution functions (6.4.2) are absolutely continuous. Following the discussion of singularity problems for the "equiprobable" case (6.1.2) it seems likely that finding values for \((a, \alpha_1, \ldots, \alpha_m)\) which give absolutely continuous distributions will be a very difficult problem. However it is likely that as before there is a range of values for \(a\) where the distributions are absolutely continuous almost everywhere.

Before considering the semicontraction mapping approach there is a simple approximation which can be used in certain cases. In the error detector problem the parameter \(a = 1 - \epsilon\) is very close to 1. Hence the sum \(\sum_{n=1}^{\infty} a^{n-1} X_n\) is similar to an infinite sum of Markov states. In general as \(a \to 1\) the random geometric series tends to the sum of Markov states. In this situation a Normal approximation is valid as shown in O'Brien [1974]. In the 1974 paper O'Brien produces a central limit theorem for certain classes of chain dependent processes. These results apply to this case but their accuracy has not been investigated. However by using a Normal approximation to the distribution function it is found that convergence by the semicontraction mapping approach is speeded up considerably. This indicates the successful application of Normality at least for an initial estimate.

10.2 The semicontraction mapping principle

Consider the variable \(X\) given by (6.1.1). \(X\) takes values in the
range $[\frac{a_1}{1-a}, \frac{a_m}{1-a}] = [l_1, l_2] = L$

**Notation**

Let $R$ be a set of vectors of functions given by

$$R = \{F(x) = (F_i(x), \ldots, F_m(x)) \mid F_i(x) \text{ is a continuous, non-decreasing distribution function increasing only on } (l_i, l_i^+), i=1, \ldots, m\}$$

$p$ is a distance function on $R$ given by

$$p[F(x), G(x)] = \max_{i=1, \ldots, m} \left( \sup_{x \in L} |F_i(x) - G_i(x)| \right)$$

**Proposition 10.2.1**

$(R, p)$ is a complete metric space.

**Proof:**

Clearly $R$ is a non-empty set and $p: R \times R \to \mathbb{R}$, where $\mathbb{R}$ is the set of real numbers. Hence it suffices to verify the three properties of a metric to prove that $(R, p)$ is a metric space.

1. $p[F(x), G(x)] \geq 0$ for any $F, G \in R$ since each

   $\sup_{x \in L} |F_i(x) - G_i(x)| \geq 0$

2. $p[F(x), G(x)] = p[G(x), F(x)]$ for any $F, G \in R$ since

   $|F_i(x) - G_i(x)| = |G_i(x) - F_i(x)|$

3. $p[F(x), H(x)] \leq p[F(x), G(x)] + p[G(x), H(x)]$

   for any $F, G, H \in R$ since

   $|F_i(x) - H_i(x)| \leq |F_i(x) - G_i(x)| + |G_i(x) - H_i(x)|$

Also, convergence in $R$ depends upon convergence of each element of the vector. The convergence of each element is uniform since the metric
is so similar to the supremum metric. Hence convergence in $\mathbb{R}$ is uniform and so the metric space $(\mathbb{R}, \rho)$ is complete.

Proposition 10.2.2

The equations (6.4.2) represent a mapping $U : \mathbb{R} \to \mathbb{R}$.

Proof:

For any $F(x) = (F_1(x), \ldots, F_m(x)) \in \mathbb{R}$ define the mapping $U$ on $\mathbb{R}$ by

$$U(F(x)) = (j_{i=1}^m \sum \limits_{j=1}^m P_{ij} F_j \left( \frac{x-a}{a} \right), \ldots, j_{i=1}^m \sum \limits_{j=1}^m P_{ij} F_j \left( \frac{x-a}{a} \right)) \quad (10.2.1)$$

Since $F(x) \in \mathbb{R}$, each $F_i(x)$ is continuous and non-decreasing. Hence each $j_{i=1}^m \sum \limits_{j=1}^m P_{ij} F_j \left( \frac{x-a}{a} \right)$ is continuous and non-decreasing. Each $F_i(x)$ satisfies $F_i(x) = 0$, $x \leq 1$. Thus

$$j_{i=1}^m \sum \limits_{j=1}^m P_{ij} F_j \left( \frac{x-a}{a} \right) = 0, \quad \frac{x-a}{a} \leq 1 \quad \text{for each } j = 1, \ldots, m$$

But $\frac{x-a}{a} \leq 1 \Rightarrow x \leq a_1 + \alpha_j$.

Also $a_1 + \alpha_j \geq a_1 + \alpha_1 = 1$

Hence $j_{i=1}^m \sum \limits_{j=1}^m P_{ij} F_j \left( \frac{x-a}{a} \right) = 0$ for $x \leq 1_i$, $\alpha_j = 1, \ldots, m$.

Similarly each $F_i(x)$ satisfies $F_i(x) = 1$ for $x > 1_2$. Thus

$$j_{i=1}^m \sum \limits_{j=1}^m P_{ij} F_j \left( \frac{x-a}{a} \right) = 1 \text{ for } \frac{x-a}{a} > 1_2, \quad j=1, \ldots, m$$

But $\frac{x-a}{a} > 1_2 \Rightarrow x \geq a_2 + \alpha_j$.

Also $a_2 + \alpha_j \geq a_2 + \alpha_m = 1_2$.
210

Hence \( \prod_{j=1}^{m} P_{ij} F_j \left( \frac{x-a^j}{a} \right) = 1 \) for \( x > 1 \), \( i=1, \ldots, m \).

This completes the proof.

**Condition 10.2.4**

Consider a Markov chain with states \( \alpha_1, \ldots, \alpha_m \) and transition matrix \( P = (p_{ij}) \). Let an allowable sequence be any sequence of states \( \{ \alpha_i \} \) where \( h_1 \in \{ 1, \ldots, m \} \) and \( P_{h_i} h_{i+1} \neq 0 \) for \( i = 0, 1, \ldots \).

A Markov chain satisfies Condition 10.2.4 for a given value of \( a \) if for every state \( \alpha_{h_1} \), two allowable sequences \( \{ \alpha_i \} \) and \( \{ \alpha'_{k_i} \} \) with \( h_0 = k_0 = i \) satisfying

\[
\prod_{i=0}^{+\infty} \frac{1}{a} \alpha_{h_i} \neq \prod_{i=0}^{+\infty} \frac{1}{a} \alpha_{k_i}
\]

**Theorem 10.2.5**

The mapping \( U \) defined by (10.2.3) is a semicontraction mapping for values of \( a \) and Markov chains satisfying Condition 10.2.4.

Proof:

In the notation of Propositions 10.2.1 and 10.2.2, \( U \) is a mapping such that \( U : R \rightarrow R \) and \( (R, \rho) \) is a complete metric space. Hence it suffices to prove that the conditions of Definition 6.4.4 hold.

(1) Let \( F(x), G(x) \) be any elements of \( R \).

\[
\rho[U F, U G] = \max_i \left\{ \sup_{x \in L} \left| \prod_{j=1}^{m} P_{ij} F_j \left( \frac{x-a^j}{a} \right) - \prod_{j=1}^{m} P_{ij} G_j \left( \frac{x-a^j}{a} \right) \right| \right\}
\]

\[
= \max_i \left\{ \sup_{x \in L} \left| \int_j P_{ij} G_j \left( \frac{x-a^j}{a} \right) \right| \right\}
\]

\[
\leq \max_i \left\{ \prod_{j=1}^{m} \sup_{x \in L} \left| G_j \left( \frac{x-a^j}{a} \right) \right| \right\}
\]
Thus the first condition holds

(ii) It is required to prove that for some \( m \in Z^+ \),

\[ \exists r, \quad 0 < r < 1, \text{ such that} \]

\[ \rho(U^m R, U^m G) \leq r \rho(F, G) \quad \text{for all} \ F, G \in R. \]

The proof is included in Appendix III.

Hence for Markov chains satisfying Condition 10.2.4 the semicontraction mapping principle outlined in Propositions 6.4.5 and 6.4.6 can be used. As described in Chapter 8, this iterative technique has the advantage of ensuring a unique solution independent of the initial estimates used.

Clearly the restrictions placed on this technique by Condition 10.2.4 are of interest. It seems that although chains can be produced which do not satisfy this condition, they tend to be quite trivial examples. However any chain possessing an absorbing state will clearly not satisfy condition 10.2.4. In general most chains of at least moderate complexity seem to satisfy the condition and it is usually very easy to verify this.

Example 10.2.6

Consider a Markov pulse train with transition matrix given by
and 5 states given by
\[
(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5) = (-2, -1, 0, 1, 2)
\]
Starting in state \(\alpha_3\) the possible sequences are combinations of the transitions

(i) \(\alpha_3, \alpha_3\)
(ii) \(\alpha_3, \alpha_2, \alpha_5, \alpha_3\)
(iii) \(\alpha_3, \alpha_4, \alpha_1, \alpha_3\)

The contributions made by these transitions to the random variable \(X = \sum_{n=0}^{\infty} a^{n-1} \alpha_n\) are respectively

\[
\begin{align*}
\alpha^1(0 + 0a) \\
\alpha^1(0 - a + 2a^2 + 0) \\
\alpha^1(0 + a - 2a^2 + 0)
\end{align*}
\]

If \(a = \frac{1}{3}\), then all three contributions are zero. Hence for \(a = \frac{1}{3}\) the Markov chain does not satisfy Condition 10.2.4 since starting in state \(\alpha_3\) all possible sequences of states yield \(X = 0\). It is easy to see that only the single value \(a = \frac{1}{3}\) has this property.
Example 10.2.7

Consider the general 2 state Markov pulse train with transition matrix

\[ P = \begin{bmatrix} P_1 & P_2 \\ P_3 & P_4 \end{bmatrix} \]

and two states \( \alpha_1 \) and \( \alpha_2 \). Four possible situations are considered.

(i) \( p_i \neq 0 \), \( i = 1, 2, 3, 4 \).

Starting in state \( \alpha_1 \) two possible sequences are

\((\alpha_1, \alpha_1, \ldots)\) and \((\alpha_1, \alpha_2, \alpha_2, \ldots)\).

These give respectively the variables

\[ X_1 = \frac{\alpha_1}{1-a} \quad \text{and} \quad X_2 = \alpha_1 + \frac{a\alpha_2}{1-a} \]

These are distinct values since \( X_1 = X_2 \) if and only if \( \alpha_1 = \alpha_2 \) (since \( 0 < a < 1 \)).

Similarly starting in state \( \alpha_2 \) gives two possible variables

\[ X_1 = \frac{\alpha_2}{1-a} \quad \text{and} \quad X_2 = \alpha_2 + \frac{a\alpha_1}{1-a} \]

which have distinct values as shown above.

Hence all Markov chains of this type satisfy Condition 10.2.4.

(ii) Clearly if \( p_2 \) or \( p_3 \) were zero then one state would be absorbing and Condition 10.2.4 would not hold.

(iii) \( p_1 = 0, \ p_i \neq 0 \), \( i = 2, 3, 4 \).
Starting in state $\alpha_1$ two possible sequences are 
$\langle \alpha_1, \alpha_2, \alpha_2, \ldots \rangle$ and $\langle \alpha_1, \alpha_2, \alpha_1, \alpha_2, \ldots \rangle$. These give respectively the variables 

$$X_1 = \alpha_1 + \frac{\alpha \alpha_2}{1-a} \quad \text{and} \quad X = \frac{\alpha_1}{1-a^2} + \frac{\alpha \alpha_2}{1-a^2}$$

Again these are discrete values since $X_1 = X_2$ if and only if $\alpha_1 = \alpha_2$, (discounting $a=0$ since $0 < a < 1$).

Similarly starting in state $\alpha_2$ gives two possible variables 

$$X_1 = \frac{\alpha_2}{1-a} \quad \text{and} \quad X_2 = \frac{\alpha_2}{1-a^2} + \frac{\alpha \alpha_1}{1-a^2}$$

which have distinct values for the same reasons.

(iv) $p_i \neq 0, \ i = 1, 2, 3, \ p_4 = 0$

By symmetry this is the same as case (iii).

Hence all $2 \times 2$ chains without an absorbing state satisfy Condition 10.2.4.

There are pointers in the work of Golomb [1972], that Condition 10.2.4 will usually be satisfied. He considered finite sums of the form $f_N = \sum_{a \neq 0} a_1^N \beta^{N-1}$, $a_1 \in \{0,1\}$, where $\beta$ is some complex number $\beta \neq 0$. This is similar to Example 10.2.7 type (i). Hence it has already been shown to satisfy Condition 10.2.4. However Golomb [1972] shows that all $2^{N+1}$ possible values of $f_N$ are different unless $\beta$ is a unit in the ring of algebraic integers that satisfies a polynomial equation with coefficients restricted to +1, -1 and 0. For more general situations with more than two states it seems likely that the
same kind of condition on \(a\) will apply when considering Condition 10.2.4. For example in Example 10.2.6 Condition 10.2.4 is satisfied for all values of \(a\) except when \(a\) satisfies the polynomial equation \(0 - a + 2a^2 = 0\).

Furthermore if the value of \(a\) can be chosen then such difficulties can be avoided by a small alteration to the value of \(a\).

Methods of implementation

Much of the work done for the error detector problem is applicable to the general case. In Section 10.3 the moments of the general distribution functions are derived. Hence as in Chapter 7, Pearson curves, Edgeworth and Gram-Charlier expansions are all available. So the initial estimates of the distribution functions can be produced via these methods based on the moments. Then the semicontraction mapping (10.2.3) can be used iteratively to converge to the unique distribution function required.

As studied in detail in Chapters 7 and 8 there are many possible problems to this approach. The convergence and suitability of Pearson curves and Edgeworth and Gram-Charlier expansions can be in question. Also the rate of convergence of the iterative process can be prohibitive.

Many problems are made worse by increasing the number of states. In the familiar case of two states the basic equations relate one value of the distribution function to two other values. When there are \(m > 2\) states then the semicontraction mapping procedure needs to estimate \(m\) functions simultaneously. The particular problem of values of \(a\) close to 1 will again cause the convergence to be slow, and this
will be made worse by holding \( m \) functions instead of 2. To summarize, problems of convergence in the mapping are worse in the general case, for the same values of \( a \), due to the increased number of functions.

One practical point of importance is the range of each distribution function. Theorem 10.2.5 uses the range
\[
[l_1', l_2'] = \frac{\alpha_1}{1-a}, \frac{\alpha_2}{1-a}
\]
since all the distribution functions are constant outside this region. When \( 1-a \) is small this is a large range. Using initial estimates which vary over \([l_1', l_2']\) can be wasteful since many functions vary over a smaller interval. For example if \( p_{11} = 0 \) then all the distribution functions vary over some region \([l_1', l_2']\) where \( l_1' > l_1 \). If \( p_{mm} = 0 \) then \( l_1' < l_2' \). Convergence will be best if an accurate range can be used for each of the initial estimates. For some chains it is easy to calculate the exact ranges. For complicated chains this may be difficult and approximate values can be used.

10.3 Fourier transforms

The moments of the general distribution functions defined by (6.4.2) are gained by taking Fourier transforms of the derivative of (6.4.2). This gives
\[
\sigma_k(x) = \sum_{j=1}^{\infty} p_k \exp(i\alpha_j x) \sigma_j(ax)
\]
for \( k = 1, \ldots, m \), where
\[
\sigma_k(t) = \int_{-\infty}^{\infty} e^{ix} f_k(x) \, dx.
\]
Differentiating \( n \) times and setting \( x = 0 \) gives
\[
\sigma_k^{(n)}(0) = \sum_{j=1}^{\infty} p_k \prod_{r=0}^{n-1} \left( \frac{\alpha_j}{\alpha_j} \right)^{n-r} \sigma_j^{(n-r)}(0)
\]
Using (10.3.2) recursively all the moments can be calculated from the
values $\phi_k^{(0)}(0) = 1, \ k = 1, \ldots, m$. As described in Chapter 7 the moments can be used to produce approximations to the distribution functions. As in the case of the G.E.C. equations these approximations tend to be unsatisfactory on their own and are more profitably used as initial estimates for the semicontraction mapping approach.

**Calculating the Fourier transforms**

As described in Chapter 9 there are several ways of producing the Fourier transform. Three approaches are given below

(i) \( \phi_k(t) = \sum_{n=0}^{\infty} x^n f_k(x)dx \)

\[ = \sum_{n=0}^{\infty} \frac{(it)^n}{n!} \mu_n^{(k)} \]  \( (10.3.3) \)

where \( \mu_n^{(k)} = \sum_{n=0}^{\infty} x^n f_k(x)dx \)

Hence \( \phi_k(t) \) can be calculated at any value of \( t \) by a suitable truncation of equation \( (10.3.3) \). The moments \( \mu_n^{(k)} \) can be gained recursively from \( (10.3.2) \) as described above.

(ii) Equation \( (10.3.3) \) can be used to give values of the Fourier transform in a region close to the origin. Then \( (10.3.1) \) can be used to iterate out and give values wherever required. Equation \( (10.3.1) \) can be written in matrix form.
Repeated application of (10.3.4) gives the values of $\varphi_i(x)$, $i=1,...,m$, in terms of $\varphi_i(ax)$, $i=1,...,m$, for any $r \in \mathbb{Z}$. The advantage of method (ii) is that the only approximation made is in the truncation of (10.3.3) close to the origin. In this region $t^N$ drops away quickly and so a very accurate value is expected. The iterative process is exact in that (10.3.4) stems from the basic equations. Method (i) however requires a truncation of an infinite sum at every point.

Inverting the Fourier transforms

All the methods of inversion discussed in Section 9.2 are applicable here. The comparisons made in that section are also valid. Briefly the two useful approaches are

(i) Inversion by numerical integration.

(ii) Inversion using the Poisson summation formula.

Applicability

(i) Numerical integration relies on accurate values of the Fourier transform over the whole range. For values of $a$ close to 1 or for transforms exhibiting a highly oscillatory nature this approach can be inaccurate. For smoother transforms or for smaller values of $a$ the method is acceptable.
(ii) The P.S.F. provides a method which is accurate for nearly all values of \( a \) and types of transform. In particular when numerical integration is undesirable the P.S.F. technique will usually still be accurate.

(iii) As discussed in Section 9.2 shifting the Fourier transforms can remove the oscillatory behaviour caused by non-zero means. This can give increased accuracy to method (ii).

10.4 Examples

In this Section the two main applications of random geometric series are illustrated further. Example 10.4.1 takes the analysis of the error detector a step further by considering a 4 state process. In Example 10.4.2 the intersymbol interference problem is considered in detail for a particular coding scheme.

Example 10.4.1

The error detector of interest is given schematically in Figure 6.2.1. In Chapter 6 the application of a 2 state model is described. In particular the mean rate of spurious error indication can be calculated. Also an approximation to the mean rate of error indication for a given error rate is available. By modelling these errors in a 4 state model these approximations can be removed by building the error probabilities into the description of the model itself. The new 4 state model is given in Figure 10.4.1. The different errors indicated by the transitions are given by

Type I error - SPACES replaced by MARKS
Type II error - MARKS replaced by SPACES
If the two types of error are equiprobable then the transition diagram, Figure 10.4.1, can be represented by the transition matrix

\[ P = \begin{pmatrix} (1-e)P & eP \\ eP & (1-e)P \end{pmatrix} \]

where e is the probability of a type I or type II error and P is the transition matrix of the two state model

\[ P = \begin{pmatrix} P_- & P_+ \\ P_+ & P_+ \end{pmatrix} \]

The four states are numbered on Figure 10.4.1.

In this model the error probability e, forms part of the definition of the model. Hence the resulting rate of zero crossings will be the rate caused by the true errors and the spurious errors. This is exactly the information required by the design problem.

Hence the error detector design problem is best described by Figure 10.4.1. From Sections 10.2 and 10.3 it is clear that the distributions involved can be formulated in the general settings considered there. Hence for various parameters the semi-contraction mapping principle or the Poisson Summation Formula can be used to give the required values of the function.

Specifically the 4 state model yields the equations

\[ F_i(v) = p_{1i}F_1\left(\frac{v+1-a}{a}\right) + p_{2i}F_2\left(\frac{v-1+a}{a}\right) \]

\[ + p_{3i}F_3\left(\frac{v+1-a}{a}\right) + p_{4i}F_4\left(\frac{v-1+a}{a}\right) \]

where \( i = 1, 2, 3, 4. \)
Example 10.4.2

This Example is based on a block code which codes binary input into the alphabet \{-3,-2,-1,0,+1,+2,+3\}. The code has 24 states described below.

<table>
<thead>
<tr>
<th>STATE NUMBER</th>
<th>CODED OUTPUT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-3</td>
</tr>
<tr>
<td>2</td>
<td>-3</td>
</tr>
<tr>
<td>3</td>
<td>-3</td>
</tr>
<tr>
<td>4</td>
<td>-2</td>
</tr>
<tr>
<td>5</td>
<td>-2</td>
</tr>
<tr>
<td>6</td>
<td>-2</td>
</tr>
<tr>
<td>7</td>
<td>-1</td>
</tr>
<tr>
<td>8</td>
<td>-1</td>
</tr>
<tr>
<td>9</td>
<td>-1</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>2</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>21</td>
<td>2</td>
</tr>
<tr>
<td>22</td>
<td>3</td>
</tr>
<tr>
<td>23</td>
<td>3</td>
</tr>
<tr>
<td>24</td>
<td>3</td>
</tr>
</tbody>
</table>

The transition matrix describing transitions between the 24 states is given by
Hence 24 simultaneous equations can be constructed as below:

\[ F_i(x) = \frac{1}{24} \sum_{j=1}^{24} p_{ij} F_j \left( \frac{x-a_j}{a} \right) \quad i = 1, \ldots, 24 \]

The \( p_{ij} \) values are gained from the matrix above and the \( a_j \) values are tabulated in Table 10.4.1. There is sufficient symmetry in the transition matrix for the problem to be reduced to the three equations.
The remaining distribution functions are available from $F_1(x), F_2(x), F_3(x)$ as described below:

\[
F_i(x) = \frac{24}{\sum_{j=1} F_{ij} F_j^{\alpha_j}} x - \alpha_j, \quad i = 1, 2, 3 \tag{10.4.3}
\]

Clearly reducing the 24 equations to 3 makes the analysis much simpler.

Taking Fourier Transforms of the derivative of (10.4.3) enables the moments to be calculated as described in (10.3.2). If $\mu_1, \mu_2$ and $\mu_3$ are the means of the distributions $F_1, F_2$ and $F_3$ then the means
are given by

\[ \mu_1 = -\frac{3}{2} \left( \frac{4+a}{4-a} \right) \]

\[ \mu_2 = -\frac{3}{2} \]

\[ \mu_3 = -\frac{3}{2} \left( \frac{4-3a}{4-a} \right) \]

If the densities are denoted by \( f_1, f_2, f_3 \) then their respective transforms \( \varphi_1, \varphi_2, \varphi_3 \) are illustrated in Figures 10.4.2, 10.4.3 and 10.4.4. The parameter \( a \) has value 0.7 for these examples.

For this value of \( a \) the transforms are easily inverted by numerical integration. Figures 10.4.5, 10.4.6 and 10.4.7 show the densities \( f_1, f_2, f_3 \) gained in this manner.

The density \( f_3(x) \) has been integrated by quadrature to give the distribution function \( F_3(x) \) displayed in Figure 10.4.8. Also in Figure 10.4.9 \( F_3(x) \) is shown calculated by the semi-contraction mapping method. This iterative process used an initial estimate of a uniform distribution. The close agreement of the two distribution functions illustrates that the two methods of calculation are interchangeable for values of \( a \) not too close to 1.

The work of O'Brien [1974] was discussed in Section 10.1. He showed that as \( a \to 1 \) the distributions under certain conditions would tend to Normality. For the value \( a = 0.9 \) and the distribution \( F_3 \), Figures 10.4.10 and 10.4.11 compare the normal approximation with the convergent function based on the semi-contraction mapping. The normal approximation is very close to the exact solution and the mapping required only 19 iterations to converge using this approximation as an
initial estimate. The similarity of the two graphs suggests that even for \(a = 0.9\) the approximation is worthwhile and further consideration of this approach would be productive.

For the sake of completeness \(F_1\) and \(F_2\) are also given in Figures 10.4.12 and 10.4.13. In both cases \(a = 0.7\) and the results are from semi-contraction mappings using a uniform initial estimate.

Finally the familiar behaviour of the distributions as \(a \to 1\) is seen by a comparison of Figures 10.4.11 and 10.4.9. In both Figures \(F_3(x)\) is plotted but increasing \(a\) from 0.7 to 0.9 has made the distribution function increase more sharply. This is representative of the overall case where increasing \(a\) decreases the variance of the distributions.
Figure 10.4.2

Re($\phi_1(t)$)

Im($\phi_1(t)$)
Figure 10.4.3

$\text{Re}\{\phi_2(t)\}$

$\text{Im}\{\phi_2(t)\}$
Figure 10.4.4

$\text{Re}(\phi_3(t))$

$\text{Im}(\phi_3(t))$
Figure 10.4.5

Figure 10.4.6
Figure 10.4.7

\[ f_3(x) \]

Figure 10.4.8

\[ F_3(x) \]
Figure 10.4.9

Figure 10.4.10

Normal Approximation to \( F_3(x) \)
Figure 10.4.11

Figure 10.4.12
The general form of $\chi^m(G, \gamma)$ is given by

$$\chi^m(G, \gamma) = \left\{ \chi^1 G(p_{ne}), \ldots, \chi^m G(p_{ne}) \right\}$$

where

$$m = 2^{m+1}\sum_{i=1}^{n} (\chi_i + \varepsilon_i) \geq \sum_{i=1}^{n} (\chi_i + \varepsilon_i) = 1.$$
APPENDIX I.
PROOF OF THEOREM 8.2.4.

Let \( u(x) = \frac{x^{1-a}}{a} \), \( v(x) = \frac{x^{1+a}}{a} \), then

\[
U(F,G) = \left\{ p_+ F(u(x)) + p_- G(v(x)), p_+ F(u(x)) + p_+ G(v(x)) \right\}
\]

The general form of \( U^m(F,G) \) is given by

\[
U^m(F,G) = \left\{ \sum_{i=1}^{\ell} (r_i F(p_i(x)) + s_i G(q_i(x))) \right\}
\]

where

\[
\ell = 2^{m-1}, \quad \sum_{i=1}^{\ell} (r_i + \ell_i) = 1, \quad \sum_{i=1}^{\ell} (s_i + u_i) = 1.
\]

The coefficients \( r_1, s_1, t_1, u_1 \) are of the form

\[
p_1 a_1 - p_2 a_2 + p_3 a_3 + p_4 a_4 = m
\]

The arguments \( p_1(x), q_1(x) \) are combinations of \( m \) functions of the form \( u(x), v(x) \).

Four of these arguments and their coefficients are important to the proof and are given below.

\[
u(x) = p_+ \, m, \quad q(x) = V^m(x) = \frac{x-1+am}{am}
\]

\[
\ell(x) = p_+ \, p_-, \quad p(x) = \frac{x+1-2a+am}{am}
\]

\[
r_1 = p_-, \quad p_1(x) = \frac{x+1-a}{am}
\]

\[
s_1 = p_,- p_+, \quad q_1(x) = U^m(x) = \frac{x-1+2a-3am}{am}
\]
With these definitions let $F_m^*(x)$, $G_m^*(x)$ be defined by

$$U^m(F,G) = \left\{ \sum_{i=1}^{e} f_i F(p_i(x)) + \sum_{i=1}^{e} s_i G(q_i(x)) + \sum_{i=1}^{e} e_i F(p_i(x)) \right\} = \left\{ F_m^*(x), G_m^*(x) \right\}$$

Also define $J_m^*(x)$, $K_m^*(x)$ by

$$U^m(J,K) = \left\{ J_m^*(x), K_m^*(x) \right\}$$

Then

$$\rho[U^m(F,G), U^m(J,K)] = \rho[(F_m^*, G_m^*), (J_m^*, K_m^*)]$$

$$= \sup_{-\infty < x < \infty} |F_m^*(x) - J_m^*(x)| + \sup_{-\infty < x < \infty} |G_m^*(x) - K_m^*(x)| \quad (A1.1)$$

Let the interval $[-1, 1]$ be split into two parts such that

$$[-1, 1] = P_m \cup P'_m$$

where

$$P_m = [-1, 2a-1-2a^m], \quad P'_m = [2a-1-2a^m, 1]$$

This division of the interval is valid for any $a \in (0, 1)$, $m \in \mathbb{Z}^+$. Using the intervals $P_m$ and $P'_m$ (A1.1) can be rewritten

$$\rho[U^m(F,G), U^m(J,K)] = \max \left\{ \sup_{x \in P_m} |F_m^* - J_m^*|, \sup_{x \in P'_m} |F_m^* - J_m^*| + \sup_{x \in P_m} |G_m^* - K_m^*| \right\} \quad (A1.2)$$

For ease of notation let the right hand side of (A1.2) be denoted by

$$\max \{ R_1, R_2 \}$$

Consider the expression $R_1$,

$$R_1 = \sup_{x \in P_m} \left| \sum_{i=1}^{e} f_i (F-J)(p_i(x)) + \sum_{i=1}^{e} s_i (G-K)(q_i(x)) \right| + \sup_{x \in P'_m} \left| \sum_{i=1}^{e} e_i (F-J)(p_i(x)) + \sum_{i=1}^{e} u_i (G-K)(q_i(x)) \right|$$
In fact the last two terms above are zero as shown below:

\[ x \in P_m \Rightarrow x \leq 2a - 1 - 2a^m \]

\[ \Rightarrow p_e(x) \leq \frac{2a - 1 - 2a^m + 1 - 2a + a^m}{a^m} = -1 \]

\[ \Rightarrow (F - J)(p_e(x)) = 0 \text{ for } x \in P_m. \]

Also:

\[ q_e(x) \leq \frac{2a - 1 - 2a^m - 1 + a^m}{a^m} = \frac{2a - 2 - a^m}{a^m} < -1 \]

\[ \Rightarrow (G - K)(q_e(x)) = 0 \text{ for } x \in P_m. \]

These results show that:

\[ R_1 \leq \max \left\{ (1 - \epsilon_e), (1 - \epsilon_e) \right\} \times \left( \sup_{x \in P_m} |F - J| + \sup_{x \in P_m} |G - K| \right) \]

Consider the expression \( R_2 \):

\[ R_2 = \sup_{x \in P_m} \left| \sum_{i=1}^{e} \Gamma_i (F - J)(p_i(x)) + \sum_{i=1}^{e} S_i (G - K)(q_i(x)) \right| \]

\[ + \sup_{x \in P_m} \left| \sum_{i=1}^{e} \epsilon_i (F - J)(p_i(x)) + \sum_{i=1}^{e} U_i (G - K)(q_i(x)) \right| \]
The last term above can be simplified as shown below

\[ x \in P_m' \Rightarrow x > 2a - 1 - 2a^m \]
\[ \Rightarrow p_i(x) > \frac{2a - 1 - 2a^m + 1 - a^m}{a^m} \]
\[ = \frac{2a - 3a^m}{a^m} \]
\[ \Rightarrow p_i(x) > 1 \text{ for } m > \frac{(\log 0.5)/(\log a) + 1}{(1/2)} \]

Also

\[ q_i(x) < \frac{2a - 1 - 2a^m - 1 + 2a - a^m}{a^m} = \frac{4a - 2 - 3a^m}{a^m} \]
\[ \Rightarrow q_i(x) > 1 \text{ for } m > \frac{\log(1 - \frac{1}{2}a)}{\log a} + 1 \]

This is where the criterion \( a > 1/2 \) is necessary to ensure \( q_i(x) > 1 \). This simplification shows that

\[ R_a \leq (1 - \Gamma_i) \sup_{x \in P_m'} |F - J| + (1 - S_i) \sup_{x \in P_m'} |G - K| \]
\[ + \sup_{x \in P_m'} \left| \Gamma_i (F - J)(1) + S_i (G - K)(1) \right| \]
Again the last term can be simplified. Suppose that \( r_1 > s_1 \) (alternatively \( q < 1/2 \)).

\[
\sup_{x \in \mathcal{F}_m} |\Gamma_1(F-J)(1) + S_1(G-K)(1)| = |\Gamma_1(F(l)-J(l)) + S_1(G(l)-K(l))| \\
= |(r-s_1)(F(l)-J(l)) + S_1(F(l)-J(l)) + S_1(G(l)-K(l))| \\
= (r_1-s_1)|F(l)-J(l)| \text{ since } F(l)+G(l) = J(l)+K(l) = 1 \\
\leq (r_1-s_1) \sup_{x \in \mathcal{F}_m} |F-J|
\]

Hence

\[
R_2 \leq (1-r_1) \sup_{x \in \mathcal{F}_m} |F-J| + (1-s_1) \sup_{x \in \mathcal{F}_m} |G-K| + (r_1-s_1) \sup_{x \in \mathcal{F}_m} |F-J|
\]

\[
\Rightarrow R_2 \leq (1-s_1) \left\{ \sup_{x \in \mathcal{F}_m} |F-J| + \sup_{x \in \mathcal{F}_m} |G-K| \right\} \tag{A1.4}
\]

Similarly if \( s_1 > r_1 \) then

\[
R_2 \leq (1-r_1) \left\{ \sup_{x \in \mathcal{F}_m} |F-J| + \sup_{x \in \mathcal{F}_m} |G-K| \right\} \tag{A1.5}
\]

Combining (A1.2), (A1.3), (A1.4) and (A1.5) gives

\[
\rho[U^m(F,G), U^m(J,K)] \leq \max \left\{ 1-r_1, 1-s_1, 1-\epsilon, 1-\mu \right\} \times \left\{ \sup_{-\infty < x < \infty} |F-J| + \sup_{-\infty < x < \infty} |G-K| \right\} = \rho[(F,G), (J,K)]
\]

for values of \( m \) satisfying

\[
m \geq \max \left\{ \frac{\log 0.5}{\log a} + 1, \frac{\log (1-\epsilon)}{\log a} + 1 \right\} \text{ and } 0 < r < 1
\]

Thus the second condition holds and this completes the proof.
APPENDIX II.

INVERSION OF THE VANDERMONDE MATRIX

Consider the matrix $A$ defined by

$$A = \begin{bmatrix} a_1 & a_1^2 & \cdots & a_1^{n-1} \\ a_2 & a_2^2 & \cdots & a_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ a_n & a_n^2 & \cdots & a_n^{n-1} \end{bmatrix}$$

Now

$$1 = \prod_{2 \leq i < j \leq n} (a_i - a_j) \prod_{1 \leq i < j \leq n} (a_i - a_j) \prod_{k=2}^{n} (a_1 - a_k)$$

$$= \prod_{2 \leq i < j \leq n} (a_i - a_j) \sum_{m=0}^{n-1} a_1^m \left( \left( -1 \right)^{n-1-m} \sum_{j_1, j_2, \ldots, j_{n-1-m}} a_{j_1} a_{j_2} \cdots a_{j_{n-1-m}} \right)$$

where the second sum extends over all \( \binom{n-1}{n-1} \) choices of \( n-1 \)-distinct indices \( j_1, \ldots, j_{n-1} \) from the indices 2, 3, \ldots, \( n-1 \).

Also if \( 2 \leq l \leq n \) then

$$0 = \prod_{2 \leq i < j \leq n} (a_i - a_j) \prod_{1 \leq i < j \leq n} (a_i - a_j) \prod_{k=2}^{n} (a_e - a_k)$$

$$= \prod_{2 \leq i < j \leq n} (a_i - a_j) \sum_{m=0}^{n-1} a_e^m \left( \left( -1 \right)^{n-1-m} \sum_{j_1, j_2, \ldots, j_{n-1-m}} a_{j_1} a_{j_2} \cdots a_{j_{n-1-m}} \right)$$

where the second sum extends over all \( \binom{n-1}{n-1} \) choices of \( n-1 \)-distinct indices \( j_1, \ldots, j_{n-1} \) from the indices 2, 3, \ldots, \( n-1 \).
Hence if
\[
X_1 = \begin{bmatrix} X_m \end{bmatrix} = \prod_{a_i \leq j \leq n} (a_i - a_j) \left( \prod_{1 \leq i < j \leq n} (a_i - a_j) \right)^{n-1-m} \sum_{j_1 \cdots j_{n-l-1} = 1} a_{j_1} a_{j_2} \cdots a_{j_{n-l-1}} \right]^{n-1-m}
\]
is a column \(m = 0, 1, \ldots, n-1\) then
\[
AX_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}
\]
If \(X_2\) is obtained from \(X_1\) by applying the permutation \((1,2,3,\ldots,n)\) to the indices of \(a_1, \ldots, a_n\) then
\[
AX_2 = \begin{bmatrix} 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}
\]
Applying the cycle \((1,2,\ldots,n), n-1\) times to produce the columns \(X_2, X_3, \ldots, X_n\) the matrix \(B\) can be formed, \(B = (X_1, \ldots, X_n)\) and \(B\) satisfies \(AB = I\). Hence the inverse of \(A\) is \(B\).

Note that extending this inverse to the matrix
\[
\begin{bmatrix}
  a_1^{-m} & \cdots & a_1^{-1} & 1 & a_1 & \cdots & a_1^m \\
  a_2^{-m} & \cdots & a_2^{-1} & 1 & a_2 & \cdots & a_2^m \\
  \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
  a_{m+1}^{-m} & \cdots & a_{m+1}^{-1} & 1 & a_{m+1} & \cdots & a_{m+1}^m \\
\end{bmatrix}
\]
is simply a matter of suitable multiplication of the columns of \(B\).
APPENDIX III

PROOF OF THEOREM 10.2.5.

Considering only cases where Condition 10.2.4 is satisfied then for each state \( \alpha_i \) there are two sequences \( \{\alpha_{ni}\} \) and \( \{\alpha_{ki}\} \) such that

\[
\sum_{i=1}^{\infty} a_i^{-1} \alpha_{ni} = S_{ii}, \quad \sum_{i=1}^{\infty} a_i^{-1} \alpha_{ki} = S_{2i} \quad \text{where} \quad S_1 < S_2
\]

and \( h_0 = k_0 = i \).

Without loss of generality let \( S_{1i} < S_{2i} \).

The general form of \( U^n(F) \) is given by \( U^n(F) = F_n^*(x) \) where the
dth element of \( F_n^*(x) \) is given by

\[
F_n^*(x) = \sum_{j_1=1}^{m} \cdots \sum_{j_n=1}^{m} P_{j_1} P_{j_2} \cdots P_{j_n} \frac{(x-\alpha_{j_1} - \alpha_{j_2} - \cdots - \alpha_{j_n})}{a^n}
\]

Hence the general form

of \( \rho[U^n_F, U^n_G] \) is given by \( \rho[U^n_F, U^n_G] = \max_i \{ \sup_{x \in A_{ni}} |F_n^*(x) - G_n^*(x)| \} \) \( (A3.1) \)

Let the interval \([l_1, l_2]\) be split into two parts such that

\[
[l_1, l_2] = A_{ii} \cup A_{ii} \quad \text{where}
\]

\[
A_{ii} = [l_1, S_{ii} + \frac{S_{ii} - S_{ii}}{2}], \quad A_{ii} = [S_{ii} + \frac{S_{ii} - S_{ii}}{2}, l_2]
\]

This division of the interval is valid for all possible values of \( S_{11} \) and \( S_{21} \). Using the intervals \( A_{ii} \) and \( A_{ii} \) \( (A3.1) \) can be rewritten

\[
\rho[U^n_F, U^n_G] = \max_i \{ \max_{x \in A_{ii}} \sup_{x \in A_{ii}} |F_n^*(x) - G_n^*(x)| \}
\]

\[
(A3.2)
\]
For ease of notation let the right hand side of (A3.2) be donated by

\[ \max \{ \max (R_{ii}, R_{il}) \} \]

Consider the expression \( R_{ii} \):

\[
R_{ii} = \sup_{x \in A_{ii}} \left| \sum_{j_1=1}^{m} \cdots \sum_{j_n=1}^{m} p_{j_1} \ldots p_{j_n} (F_{j_n} - G_{j_n}) \left( \frac{x - \alpha_{j_1} - \cdots - \alpha_{j_n}}{a^n} \right) \right|
\]

\[
\leq \sum_{j_1=1}^{m} \cdots \sum_{j_n=1}^{m} p_{j_1} \ldots p_{j_n} \sup_{x \in A_{ii}} \left| (F_{j_n} - G_{j_n}) \left( \frac{x - \alpha_{j_1} - \cdots - \alpha_{j_n}}{a^n} \right) \right|
\]

\[
= \sum_{j_1=1}^{m} \cdots \sum_{j_n=1}^{m} p_{j_1} \ldots p_{j_n} \sup_{x \in A_{ii}} \left| (F_{j_n} - G_{j_n}) \left( \frac{x - \alpha_{j_1} - \cdots - \alpha_{j_n}}{a^n} \right) \right|
\]

\[
(j_1, \ldots, j_n) \neq (k_1, \ldots, k_m)
\]

\[
+ \sup_{x \in A_{ii}} \left| (F_{k_m} - G_{k_m}) \left( \frac{x - \alpha_{k_1} - \cdots - \alpha_{k_m}}{a^n} \right) \right|
\]

\[
\leq \left( 1 - \prod_{i=1}^{m} p_{k_i} \right) \max \left\{ \sup_{x \in A_{ii}} \left| (F_{j_n} - G_{j_n})(x) \right| \right\}
\]

\[
+ \sup_{x \in A_{ii}} \left| (F_{k_m} - G_{k_m}) \left( \frac{x - \alpha_{k_1} - \cdots - \alpha_{k_m}}{a^n} \right) \right|
\]

Similarly it can be shown for \( R_{il} \) that

\[
R_{il} \leq \left( 1 - \prod_{i=1}^{m} p_{h_i} \right) \max \left\{ \sup_{x \in A_{il}} \left| (F_{j_n} - G_{j_n})(x) \right| \right\}
\]

\[
+ \sup_{x \in A_{il}} \left| (F_{h_m} - G_{h_m}) \left( \frac{x - \alpha_{h_1} - \cdots - \alpha_{h_m}}{a^n} \right) \right|
\]

The last two inequalities can be simplified by choosing \( n \) such that the remainder terms in both become zero.

Choose \( n_{ii} \) such that

\[
|\alpha_{h_1} + a \alpha_{h_2} + \cdots + a^{n-1} \alpha_{h_n} - S_i| < \frac{S_2 - S_1}{4}
\]

\[ \forall n > n_{ii} \]
Then let

\[ N_i = \begin{cases} 
\max(1, n_{ii}) & \text{if } \ell_2 \leq 0 \\
\max(n_{ii}, \log \left( \frac{s_2 - s_i}{4\ell_2} \right) / \log \alpha) & \text{if } \ell_2 > 0
\end{cases} \]

For \( x \in A_{21} \) it follows that for \( n \geq n_1 \)

\[
\frac{x - \alpha_{k_1} - \ldots - a^{n-1} \alpha_{k_n}}{a^n} \leq \frac{s_2 - s_i}{4a^n} \leq \ell_2
\]

Hence for \( n \geq n_1 \) and \( x \in A_{21} \)

\[
(F_{hn} - G_{hn}) \left( \frac{x - \alpha_{k_1} - \ldots - a^{n-1} \alpha_{k_n}}{a^n} \right) = 0
\]

Similarly choose \( N_{11} \) such that \( \forall n > N_{11} \)

\[
| \alpha_{k_1} + a \alpha_{k_2} + \ldots + a^{n-1} \alpha_{k_n} - s_2 | < \frac{s_2 - s_i}{4}
\]

Then let

\[ N_i = \begin{cases} 
\max(1, N_{ii}) & \text{if } \ell_1 > 0 \\
\max(N_{ii}, \log \left( \frac{s_2 - s_i}{4\ell_1} \right) / \log \alpha) & \text{if } \ell_1 < 0
\end{cases} \]

For \( x \in A_{11} \) it follows that for \( n \geq N_1 \)

\[
\frac{x - \alpha_{k_1} - \ldots - a^{n-1} \alpha_{k_n}}{a^n} \leq \frac{s_1 - s_2}{4a^n} \leq \ell_1
\]

Hence for \( n \geq N_1 \) and \( x \in A_{11} \)

\[
(F_{kn} - G_{kn}) \left( \frac{x - \alpha_{k_1} - \ldots - a^{n-1} \alpha_{k_n}}{a^n} \right) = 0
\]
Having shown the remainder terms are zero (A3.2) can be rewritten

\[
\rho[U^n F, U^n G] \leq \max_i \left\{ \max \left[ (1 - P_{k_i} P_{k_i+1} \cdots P_{n-1-k_n}) \max_j \left( \sup_{x \in A_{ij}} | (F_j - G_j)(x) | \right) \right] \right\}
\]

\[
(1 - P_{k_1} P_{k_2} \cdots P_{n-1-k_n}) \max_j \left( \sup_{x \in A_{k_1}} | (F_j - G_j)(x) | \right) \}
\]

for \( n \geq \max_i \left\{ \max (n_i, N_i) \right\} \)

Hence

\[
\rho[U^n F, U^n G] \leq \rho[F, G]
\]

where

\[
r = \max_i \left\{ \max \left[ (1 - P_{k_1} P_{k_2} \cdots P_{n-1-k_n}) \right] \right\}
\]

for \( n \geq \max_i \left\{ \max (n_i, N_i) \right\} \)

By the definition of an allowable sequence

\[
P_{k_i+1} \neq 0, \quad P_{k_i+1} \neq 0 \quad \text{for} \quad i = 0, 1, 2, \ldots
\]

Hence \( 0 < r < 1 \) and the proof is complete.
REFERENCES


