# ON ARMA PROBABILITY DENSITY ESTIMATION

by

Jeffrey D. Hart and Henry L. Gray

Technical Report No. 156
Department of Statistics ONR Contract

December 1981

Research sponsored by the Office of Naval Research
Contract N00014-75-C-0439

Reproduction in whole or in part is permitted
for any purpose of the United States Government

The document has been approved for public
release and sale; its distribution is unlimited

DEPARTMENT OF STATISTICS
Southern Methodist University
Dallas, Texas 75275

Table of Contents

LIST OF FIGURES

LIST OF TABLES

# ON ARMA PROBABILITY DENSITY ESTIMATION

Jeffrey D. Hart
Department of Mathematics
University of Arkansas
Fayetteville, Arkansas 72701

Henry L. Gray
Department of Statistics
Southern Methodist University
Dallas, Texas 75275

Key Words: Probability density estimation; Kronmal-Tarter method; Autoregressive estimators; Generalized jackknife

## SUMMARY

A new method of probability density estimation is investigated which exploits the Fourier series representation of a density function. The new method employs density estimators $\hat{f}_{p,q}(\cdot)$, $p = 0,1,2,\ldots$ and $q = 0,1,2,\ldots$, which are such that $\hat{f}_{0,q}(\cdot)$ is a Fourier series (Kronmal-Tarter type) estimator and $\hat{f}_{p,0}(\cdot)$ is an autoregressive estimator. Each of the estimators $\hat{f}_{p,q}(\cdot)$ (referred to as ARMA estimators) is shown to depend upon the $e_n$-transform, thus providing a strong motivation for the use of estimators with both $p > 0$ and $q > 0$. Small and large sample properties of ARMA density estimators are obtained and a data-based method of selecting optimal values of p and q is proposed. The results of a simulation study show that, for the densities considered, a savings in integrated square error is attained by using ARMA, rather than Fourier series, density estimation.

# CHAPTER I

# INTRODUCTION

## 1.1 Introduction

The purpose of this work is to investigate a method of probability density estimation which is based upon what will be called the ARMA method of approximating a function. The ARMA method employs representations of the form

$$f_{p,q}(x) = \frac{\sum_{k=-q}^{q} \beta_k e^{ikx}}{|1-\alpha_1 e^{ix} - \ldots - \alpha_p e^{ipx}|^2} \qquad (\beta_{-k} = \overline{\beta}_k) \qquad (1.1)$$

to approximate the real-valued function $f(\cdot)$ over the interval $[-\pi,\pi]$. The acronym ARMA is used because of the fact that, if $f_{p,q}(\cdot)$ is nonnegative, its numerator may be expressed as

$$k|1-\theta_1 e^{ix} - \ldots - \theta_q e^{iqx}|^2 \quad \text{for all} \quad x \in [-\pi,\pi] .$$

Expressed in this way, $f_{p,q}(\cdot)$ is seen to have a form equivalent to the spectrum of an autoregressive, moving average (ARMA) process.

Because of the wide applicability of the ARMA model in time series, representations such as (1.1) have a very natural motivation in spectral estimation. The motivation, to be developed fully in succeeding chapters, for their use in probability density estimation must obviously be somewhat different. For the present we simply point out that the relationship of $f_{p,q}(\cdot)$ to a numerical analysis tool known as the $e_n$-transform implies that ARMA representations are attractive as an approximation scheme. Their value as an approximation scheme in turn suggests their possible value in the estimation setting.

In Chapter II definitions of the constants $\beta_k (k=0,1,\ldots,q)$ and $\alpha_k (k=1,\ldots,p)$ will be given which, for a given function $f(\cdot)$, uniquely define an approximator $f_{p,q}(\cdot)$ for each pair of values $(p,q)$. The approximator so defined depends only upon the Fourier coefficients $\phi(0), \phi(1),\ldots,\phi(p+q)$, where

$$\phi(v) = \int_{-\pi}^{\pi} e^{-ivx} f(x)\,dx, \quad |v| = 0,1,2,\ldots$$

(note that $\phi(-v) = \overline{\phi(v)}$). Thus if $f(\cdot)$ is the probability density function of a random variable with support $[-\pi,\pi]$, estimators $\hat{f}_{p,q}(\cdot)$ of $f(\cdot)$ can be formed by estimating the Fourier coefficients of $f(\cdot)$.

In light of the many existing techniques of density estimation, one might reasonably question the consideration of the class of estimators just described. In order to be of more than simply academic interest, a new technique should either have the potential for improvement over, or shed some informative light on existing techniques. Hopefully, it will be shown that the method of density estimation being proposed satisfies both of these requirements with respect to

(i) Fourier series density estimators, and

(ii) autoregressive density estimators.

It will be seen shortly that these two classes of estimators are members of the general class of ARMA density estimators. Before embarking on an investigation of ARMA estimators, it will thus be expedient to briefly discuss the origin and properties of Fourier series and autoregressive density estimators.

## 1.2 Fourier Series Density Estimation

Cencov (1962) first suggested the use of Fourier series ideas in the estimation of a probability density function. Let $L_2(r)$ be a Hilbert space whose inner product is defined by

$$(\phi,\psi) = \int_{-\infty}^{\infty} \phi(x)\psi(x)r(x)dx,$$

where r is a weight function. Let $f(\cdot)$ be the density of a random variable X and assume that $f(\cdot) \in L_2(r)$. Now, suppose $E_m$ is an arbitrary m-dimensional subspace of $L_2(r)$ with orthonormal basis $\{\xi_{1m},\ldots,\xi_{mm}\}$. The best mean square error approximation of $f(x)$ in $E_m$ is

$$f_m(x) = \sum_{k=1}^{m} \phi_{km}\xi_{km}(x) ,$$

where

$$\phi_{km} = (\xi_{km},f) = \int_{-\infty}^{\infty} \xi_{km}(x)r(x)dF(x).$$

If a random sample $X_1,\ldots,X_n$ is obtained from $f(\cdot)$, then Cencov suggests estimating $f(x)$ by

$$\hat{f}_m(x) = \sum_{k=1}^{m} \hat{\phi}_{km}\xi_{km}(x),$$

where

$$\hat{\phi}_{km} = \frac{1}{n} \sum_{i=1}^{n} \xi_{km}(X_i)r(X_i) .$$

Cencov points out that $E[||\hat{f}_m(x)-f(x)||^2]$ can be made arbitrarily small by choosing a sufficiently good approximating subspace $E_m$ and then taking a large enough number n of observations.

Kronmal and Tarter (1968) have investigated a special case of the above by considering the weight function $r(x) = \frac{2}{b-a} I_{[a,b]}(x)$ and the orthonormal system

$$\left\{ \frac{1}{\sqrt{2}} , \cos \pi\left(\frac{x-a}{b-a}\right) , \ldots, \cos m\pi\left(\frac{x-a}{b-a}\right) \right\} .$$

Based on this system an estimator of $f(x)$ ($x \in [a,b]$) is

$$\hat{f}_m(x) = \frac{\hat{\phi}_o}{2} + \sum_{k=1}^{m} \hat{\phi}_k \cos k\pi \left(\frac{x-a}{b-a}\right) ,$$

where

$$\hat{\phi}_k = \frac{2}{(b-a)n} \sum_{i=1}^{n} \cos \frac{k\pi(X_i-a)}{(b-a)} I_{[a,b]}(X_i) .$$

It can be shown that

$$\operatorname{cov}(\hat{\phi}_j, \hat{\phi}_k) = \frac{1}{n}\left[\frac{1}{b-a} (\phi_{j-k} + \phi_{j+k}) - \phi_j \phi_k\right] \qquad (j \geq k)$$

where

$$\phi_k = \frac{2}{b-a} \int_a^b f(x) \cos \pi k \left(\frac{x-a}{b-a}\right) dx .$$

This leads to a simple expression for the mean integrated square error (MISE) of $\hat{f}_m(\cdot)$, namely

$$E\left[\int_{-\infty}^{\infty} (\hat{f}_m(x) - f(x))^2 r(x)dx\right] = \frac{\phi_o}{2n}\left(\frac{2}{b-a} - \phi_o\right) + \frac{1}{n} \sum_{k=1}^{m} \left(\frac{\phi_o + \phi_{2k}}{b-a} - \phi_k^2\right) + \sum_{k=m+1}^{\infty} \phi_k^2 .$$

Making use of these results Kronmal and Tarter prove the following theorem.

Theorem 1.1 If the Fourier cosine series of the density $f(\cdot)$ converges uniformly and if $m = o(\sqrt{n})$, then

$$\lim_{m\to\infty} E(\hat{f}_m(x) - f(x))^2 = 0 \quad \text{(uniformly in } x\varepsilon[a,b])$$

and

$$\lim_{m\to\infty} E \int_a^b (\hat{f}_m(x) - f(x))^2 dx = 0 .$$

The importance of this theorem is its establishment of the rate at which the truncation point m may increase with the sample size in order for $\hat{f}_m(\cdot)$ to be a consistent estimator of $f(\cdot)$. In addition to this asymptotic result, Kronmal and Tarter devise a procedure for choosing an m which, for a given sample size, minimizes the MISE.

Approaches for estimating $f(\cdot)$ using different orthogonal systems of functions have also been considered. For example, Schwartz (1967) has investigated the use of Hermite polynomials. In the present study, however, our principal interest will be in the trigonometric systems because of their close association with ARMA approximators and estimators.

## 1.3 Autoregressive Density Estimation

Carmichael (1976) has adapted the idea of autoregressive spectral estimation to the estimation of a probability density. In order to briefly outline Carmichael's method, let $f(\cdot)$ be the pdf of a random variable X with support $[-\pi,\pi]$. Define $R(\cdot)$ by

$$R(v) = \int_{-\pi}^{\pi} e^{-ivx} f(x)dx \ , \quad |v| = 0,1,2,\ldots \ .$$

Let $(\alpha_{1m},\alpha_{2m},\ldots,\alpha_{mm})$ be defined as the solution (assumed unique) of the following system of Yule-Walker equations:

$$\begin{bmatrix} 1 & R(-1) & \ldots & R(-m+1) \\ R(1) & 1 & \ldots & R(-m+2) \\ \vdots & \vdots & & \vdots \\ R(m-1) & R(m-2) & \ldots & 1 \end{bmatrix} \begin{bmatrix} \alpha_{1m} \\ \alpha_{2m} \\ \vdots \\ \alpha_{mm} \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ \vdots \\ R(m) \end{bmatrix} \ .$$

The m-th order approximator of f(x) is then defined by Carmichael as

$$f_m(x) = \frac{k_m}{2\pi} \frac{1}{|1-\alpha_{1m}e^{ix}-\alpha_{2m}e^{2ix}-\ldots-\alpha_{mm}e^{mix}|^2} ,$$

where $k_m$ is chosen so that $R(0) = 1$.

The term approximator is appropriate since it can be shown that

$$\int_{-\pi}^{\pi} e^{-ivx} f_m(x)dx = R(v), \quad |v| = 0,1,\ldots,m .$$

When observations are available from $f(\cdot)$, an estimator $\hat{f}_m(\cdot)$ can be similarly obtained by first estimating $R(\cdot)$.

Carmichael provides two motivations for the approach just outlined. One motivation involves regarding $\{R(v):|v| = 0,1,\ldots\}$ as the correlation sequence of a complex-valued, stationary time series. The spectral density $f(\cdot)$ of this hypothetical time series is approximated by the mth order autoregressive scheme $f_m(\cdot)$. Another motivation follows from showing the equivalence of $f_m(\cdot)$ to an approximator formed by constructing a set of polynomials in $e^{ix}$ which are orthogonal with respect to the inner product

$$(g,h) = \int_{-\pi}^{\pi} g(e^{ix}) \overline{h(e^{ix})} f(x)dx.$$

The weak consistency of $\hat{f}_m(\cdot)$ as an estimator of $f(\cdot)$ has also been established by Carmichael. This result may be stated as follows. Let $X_1,\ldots,X_n$ be a random sample from $f(\cdot)$ and

$$\hat{R}(v) = \frac{1}{n} \sum_{j=1}^{n} e^{-ivX_j} .$$

Then $\hat{f}_m(\cdot)$ is formed by replacing $R(\cdot)$ by $\hat{R}(\cdot)$ in the system of

equations presented previously. If $f(\cdot)$ satisfies certain regularity conditions and

$$\lim_{\substack{m\to\infty \\ n\to\infty}} \frac{m^{3/2}}{\sqrt{n}} = 0 \text{ , then}$$

$|\hat{f}_m(x) - f_\infty(x)| \xrightarrow{p} 0$ uniformly in x, where $f_\infty(x)=f(x)$ a.e.$[-\pi,\pi]$.

Parzen (1979) proposes an additional application of autoregressive representations in the estimation of density-quantile, or fQ, functions, where $f(\cdot)$ and $Q(\cdot)$ are respectively, the probability density and quantile function of a random variable X and

$$fQ(u) = f(Q(u)),\ 0 \leq u \leq 1 \ .$$

Although density-quantile estimation will not be investigated in this work, the ARMA method is easily adapted to this problem. It is hoped that some of the forthcoming observations pertaining to density estimation will find applications in the estimation of fQ and other types of functions, such as hazard functions.

# CHAPTER II

# THE DETERMINISTIC SETTING: $f_{p,q}(\cdot)$ AS AN APPROXIMATOR OF $f(\cdot)$

## 2.1 Definitions and Assumptions

In the current chapter we will consider the problem of approximating a function using a finite number of its Fourier coefficients. To facilitate our discussion the following definitions and assumptions are stated. The notation presented here will be followed consistently throughout the remainder of this work.

(i) $f(\cdot)$ denotes a real-valued function with domain of definition $[-\pi,\pi]$, which we wish to approximate or estimate. Unless otherwise stated, it shall be assumed that $f(\cdot)$ is square integrable on $[-\pi,\pi]$, i.e.,

$$\int_{-\pi}^{\pi} f^2(x)\,dx < \infty.$$

(ii) The sequence $\{\phi(v): |v| = 0,1,2,\ldots,\}$ of Fourier coefficients of $f(\cdot)$ is defined as

$$\phi(v) = \int_{-\pi}^{\pi} e^{-ivx} f(x)\,dx, \quad |v| = 0,1,\ldots .$$

Under the integrability condition in (i) $|\phi(v)|$ is finite for all v. Note that if $f(\cdot)$ is a probability density function, $\phi(\cdot)$ is simply its characteristic function evaluated at the integers.

(iii) Unless stated to the contrary, it will be assumed that $f(\cdot)$ satisfies conditions which ensure that

$$f(x) = \frac{1}{2\pi} \sum_{v=-\infty}^{\infty} \phi(v) e^{ivx}, \qquad \text{a.e.}[-\pi,\pi].$$

One such set of conditions (see Apostol (1973)) is that $f(\cdot)$ be continuous and of bounded variation throughout $[-\pi,\pi]$.

(iv) $f(\cdot)$ will be said to have an ARMA representation iff

$$f(x) = \frac{\sum_{-q}^{q} \beta_v e^{ivx}}{|1-\alpha_1 e^{ix}-\ldots-\alpha_p e^{ipx}|^2}, \qquad \text{a.e. } [-\pi,\pi],$$

where $p$ and $q$ are non-negative integers, $\beta_v(|v| = 0,1,\ldots,q)$ and $\alpha_k(k = 1,\ldots,p)$ are complex constants with $\beta_{-v} = \overline{\beta}_v$, and the roots of $1 - \alpha_1 x-\ldots-\alpha_p x^p = 0$ all lie outside the unit circle.

## 2.2 Discussion and Definition of $f_{p,q}(\cdot)$

Before moving to the stochastic setting, the ARMA method will be motivated by demonstrating its value as a deterministic approximation scheme. In the current section the ARMA approximator $f_{p,q}(\cdot)$ is defined and shown to be related to the $e_n$-transform. In Section 2.3 truncated Fourier series, autoregressive, and ARMA approximators will be compared as to their ability to approximate a function $f(\cdot)$. Comparisons will be made on the basis of how well the approximator fits $f(\cdot)$ visually, and also by means of the measure

$$\text{ISE}(f^*) = \int_{-\pi}^{\pi} (f^*(x) - f(x))^2 dx,$$

where $f^*(\cdot)$ approximates $f(\cdot)$.

Given the Fourier coefficients $\phi(0)$, $\phi(1),\ldots,\phi(m)$ (note $\phi(-v) = \overline{\phi(v)}$) of a function $f(\cdot)$ with a series representation as in the previous section, the most obvious choice for an approximator of $f(x)$ is

$$f_m(x) = \frac{1}{2\pi} \sum_{v=-m}^{m} \phi(v) e^{ivx} \quad .$$

The error associated with this approximation is

$$|f(x)-f_m(x)| = \frac{1}{2\pi} \left| \sum_{|v|>m} \phi(v) e^{ivx} \right| ,$$

which can be made arbitrarily small by choosing m large enough. The convergence of $f_m(\cdot)$ to $f(\cdot)$ is uniform if $f(\cdot)$ is continuous and of bounded variation (see Apostol (1973)). In addition to the pointwise error of $f_m(\cdot)$, we have, by Parseval's theorem,

$$ISE(f_m) = \frac{1}{\pi} \sum_{v=m+1}^{\infty} |\phi(v)|^2 .$$

In certain applications or for certain functions, a suitable choice for m may be prohibitively large. In other words, $f_{m_0}(\cdot)$ based upon a reasonable number $m_0$ of Fourier coefficients may not provide an adequate approximation to $f(\cdot)$. Suppose, however, that $\phi(m_0+1)$, $\phi(m_0+2),\ldots$ are in some sense related to the previous Fourier coefficients. It may then be possible to exploit this relationship and construct an approximator based on $\phi(0),\phi(1),\ldots,$ $\phi(m_0)$ which has better error properties than does $f_{m_0}(\cdot)$.

A model for the relationship between the Fourier coefficients of $f(\cdot)$ which is often at least approximately satisfied is

$$\{\phi(v)\} \in L(p,\Delta) \text{ for } v > q \quad . \tag{2.1}$$

where $\{f_m\} \in L(n,\Delta)$ for $m > m_0$ if there exists a smallest integer $n > 0$ and a set of $c_i$'s such that

$$f_m + c_1 f_{m-1} + \ldots + c_n f_{m-n} = 0, \quad m > m_0.$$

In the following theorem we establish the equivalence of functions whose Fourier coefficients satisfy (2.1) and functions having ARMA representations.

Theorem 2.1 Suppose the roots of

$$1 - \alpha_1 x - \ldots - \alpha_p x^p = 0$$

all lie outside the unit circle. Then $f(\cdot)$ has an ARMA representation of the form

$$f(x) = \frac{\sum_{v=-q}^{q} \beta_v e^{ivx}}{|1-\alpha_1 e^{ix} - \ldots - \alpha_p e^{ipx}|^2} \qquad \text{a.e. } [-\pi,\pi]$$

iff $\phi(v) - \alpha_1\phi(v-1) - \ldots - \alpha_p\phi(v-p) = 0, \quad v > q$ .

Proof: Suppose first that $\phi(v)$ satisfies the prescribed difference equation. Now consider the function $f^*_{p,q}(\cdot)$ satisfying

$$f^*_{p,q}(x) = \frac{\sum_{v=-q}^{q} \beta_v e^{ivx}}{|1 - \alpha_1 e^{ix} - \ldots - \alpha_p e^{ipx}|^2} \qquad \text{for } x \in [-\pi,\pi]$$

where the $\beta_v$ are chosen so that

$$\int_{-\pi}^{\pi} e^{-ijx} f^*_{p,q}(x)dx = \phi^*_{p,q}(j) = \phi(j) \quad \text{for } |j| = 0,1,\ldots,q.$$

The system of $2q+1$ equations which must be solved to find the $\beta_v$ is readily seen to be linear, and it is tacitly assumed that the system has a solution.

Now consider, for $v > q$,

$$\phi^*_{p,q}(v) - \alpha_1\phi^*_{p,q}(v-1) - \ldots - \alpha_p\phi^*_{p,q}(v-p)$$

$$= \int_{-\pi}^{\pi} (e^{-ivx} - \alpha_1 e^{-i(v-1)x} - \ldots - \alpha_p e^{-i(v-p)x}) f^*_{p,q}(x)\,dx$$

$$= \int_{-\pi}^{\pi} e^{-ivx}(1 - \alpha_1 e^{ix} - \ldots - \alpha_p e^{ipx}) f^*_{p,q}(x)\,dx$$

$$= \int_{-\pi}^{\pi} \frac{e^{-ivx} e^{iqx} \sum_{v=-q}^{q} \beta_v e^{-ix(q-v)}\,dx}{(1 - \bar{\alpha}_1 e^{-ix} - \ldots - \bar{\alpha}_p e^{-ipx})}$$

$$= \frac{-1}{i} \oint \frac{z^{v-q-1} \sum_{v=-q}^{q} \beta_v z^{q-v}\,dz}{(1 - \bar{\alpha}_1 z - \ldots - \bar{\alpha}_p z^p)} \,.$$

Since $v > q$ and the roots of $1 - \bar{\alpha}_1 z - \ldots - \bar{\alpha}_p z^p = 0$ are outside the unit circle, it follows that the above integrand is analytic on and inside the unit circle. Thus, by the Cauchy-Goursat theorem the integral is zero. It follows that $\phi^*_{p,q}(v)$ satisfies the same difference equation as $\phi(v)$ for $v > q$. Since $\phi(v) = \phi^*_{p,q}(v)$ for $v = 0,1,2,\ldots,q$ we must then have $\phi(v) = \phi^*_{p,q}(v)$ for $|v| = 0,1,2,\ldots$ . By the uniqueness of the Fourier coefficients of square integrable functions (and it is easily shown that a function having an ARMA representation is square integrable), it follows that

$$f(x) = f^*_{p,q}(x) \qquad \text{a.e. } [-\pi,\pi] \,.$$

One part of the theorem is thus proven. By mimicking a portion of the above argument it is easily shown that

$$\phi(v)-\alpha_1\phi(v-1) - \ldots - \alpha_p\phi(v-p) = 0 \ , \quad v > q \ ,$$

whenever $f(\cdot)$ has the stated ARMA representation.

Implicit in Theorem 2.1 is a method for forming an approximator of $f(\cdot)$ in the situation where

$$\phi(v) - \alpha_1\phi(v-1)-\ldots-\alpha_p\phi(v-p) \simeq 0, \ v > q.$$

Given $\phi(0)$, $\phi(1)$, $\phi(-1),\ldots,\phi(-p-q)$, $\phi(p+q)$ an approximator $f^*_{p,q}(\cdot)$ can be constructed by first solving the system of equations

$$\begin{aligned}
&\alpha_1\phi(q) + \alpha_2\phi(q-1)+\ldots+\alpha_p\phi(q-p+1) = \phi(q+1)\\
&\alpha_1\phi(q+1)+\alpha_2\phi(q) +\ldots+\alpha_p\phi(q-p+2) = \phi(q+2)\\
&\qquad\qquad\vdots\\
&\alpha_1\phi(q+p-1)+\alpha_2\phi(q+p-2)+\ldots+\alpha_p\phi(q) = \phi(q+p)
\end{aligned} \tag{2.2}$$

for $\alpha_1,\ldots,\alpha_p$. The coefficients $\beta_0,\beta_{-1},\beta_1,\ldots,\beta_{-q},\beta_q$ can then be found by forcing $\phi^*_{p,q}(v) = \phi(v)$, $|v| = 0,1,\ldots,q$, where

$$\phi^*_{p,q}(v) = \int_{-\pi}^{\pi} e^{-ivx} f^*_{p,q}(x)dx \quad \text{and}$$

$$f^*_{p,q}(x) = \frac{\sum_{v=-q}^{q} \beta_v e^{ivx}}{|1 - \alpha_1 e^{ix} - \ldots - \alpha_p e^{ipx}|^2} \ .$$

Under the assumption that the roots of $1 -\alpha_1 x-\ldots-\alpha_p x^p = 0$ lie outside the unit circle, the approximator $f^*_{p,q}(\cdot)$ satisfies

$$\phi^*_{p,q}(v) = \phi(v), \ |v| = 0,1,\ldots,p+q \ . \tag{2.3}$$

This property follows from the fact that by (2.2) and Theorem 2.1, $\phi(v)$ and $\phi^*_{p,q}(v)$ both satisfy the difference equation

$$y(v) - \alpha_1 y(v-1) - \ldots - \alpha_p y(v-p) = 0$$

for $v = q+1,\ldots,q+p$ subject to the initial conditions $y(v) = \phi(v)$, $|v| = 0,1,\ldots,q$.

Property (2.3) justifies the use of the term approximator for $f^*_{p,q}(\cdot)$ even when $\phi(\cdot)$ is not well modeled as the solution to a difference equation. The following error properties of $f^*_{p,q}(\cdot)$ are a simple consequence of (2.3).

$$\begin{aligned} |f(x) - f^*_{p,q}(x)| &= \frac{1}{2\pi}\Big| \sum_{|v|>p+q} (\phi(v) - \phi^*_{p,q}(v))e^{ivx}\Big| \\ \mathrm{ISE}(f^*_{p,q}) &= \frac{1}{\pi} \sum_{v=p+q+1}^{\infty} |\phi(v) - \phi^*_{p,q}(v)|^2 \ . \end{aligned} \tag{2.4}$$

Although the method discussed above for constructing $f^*_{p,q}(\cdot)$ is informative, it can be quite cumbersome analytically. The approximator $f_{p,q}(\cdot)$ to be defined below will be shown to be identical to $f^*_{p,q}(\cdot)$ under the assumption that the roots of $1 - \alpha_1 x - \ldots - \alpha_p x^p = 0$ lie outside the unit circle. However, $f_{p,q}(\cdot)$ has the advantage of being much simpler to construct than $f^*_{p,q}(\cdot)$. In addition, the dependence of $f_{p,q}(\cdot)$ upon a numerical analysis tool known as the $e_n$-transform provides important insight into why the ARMA method is of value as an approximation scheme.

Before defining $f_{p,q}(\cdot)$ we give the following definition of

the $e_n$-transform.

Definition 2.1 Given the sequence $\{a_k, a_{k+1}, \ldots,\}$ of complex numbers and the partial sums $A_j = \sum_{v=k}^{j} a_v$, we define (for $m \geq n+k-1$)

$$e_n(A_m) = \frac{\begin{vmatrix} A_{m-n} & A_{m-n+1} & \cdots & A_m \\ a_{m-n+1} & a_{m-n+2} & \cdots & a_{m+1} \\ \vdots & \vdots & & \vdots \\ a_m & a_{m+1} & \cdots & a_{m+n} \end{vmatrix}}{\begin{vmatrix} 1 & 1 & \cdots & 1 \\ a_{m-n+1} & a_{m-n+2} & \cdots & a_{m+1} \\ \vdots & \vdots & & \vdots \\ a_m & a_{m+1} & \cdots & a_{m+n} \end{vmatrix}}$$

whenever this quantity is defined. If both numerator and denominator are zero, then define $e_n(A_m) = e_{n-1}(A_m)$. If only the denominator is zero, then $e_n(A_m) = \infty$.

The important result associated with the $e_n$-transform is that in a wide class of problems $e_n(A_m)$ is a better approximation to $A_\infty$ than is $A_{n+m}$. With Definition 2.1 we are now in a position to define the approximator $f_{p,q}(\cdot)$.

Definition 2.2 Let $\{\phi(k), \phi(k+1), \ldots\}$ be a sequence of Fourier coefficients of $f(\cdot)$. Then the approximator $f_{p,q}(\cdot)$ of $f(\cdot)$ is

defined as

$$f_{p,q}(x) = \frac{1}{2\pi}[\phi(0)+2\text{Real}\{e_p(F_q(x))-F_0(x)\}], \quad x \in [-\pi,\pi],$$

where

$$k = \begin{cases} q - p + 1, & q + 1 - p \leq 0 \quad \text{and} \\ 1, & q + 1 - p > 0 \end{cases}$$

$$F_j(x) = \begin{cases} \sum_{v=k}^{j} \phi(v)e^{ivx}, & j \geq k \\ 0, & j < k \end{cases}.$$

Since

$$f(x) = \frac{1}{2\pi}\sum_{v=-\infty}^{\infty} \phi(v)e^{ivx} = \frac{1}{2\pi}[\phi(0)+2\text{Real}(\sum_{v=1}^{\infty} \phi(v)e^{ivx})],$$

$e_p(F_q(x)) - F_0(x)$ is seen to approximate $\sum_{v=1}^{\infty} \phi(v)e^{ivx}$. The extent to which $e_p(F_q(x)) - F_0(x)$ is a better approximator of this quantity than $\sum_{v=1}^{p+q}\phi(v)e^{ivx}$ depends upon the particular sequence $\{\phi(v)\}$. Conditions under which $e_n(A_m)$ converges more rapidly (as $m \to \infty$) to $\sum_{v=k}^{\infty} a_v$ than does $\sum_{v=k}^{n+m} a_v$ have been established by different authors, including Shanks (1955), McWilliams (1969), and Gray, Houston, and Morgan (1978). Except for Theorem 2.2, however, the discussion of these conditions will be postponed until Chapter IV. For the present, we simply note that they provide an important motivation for using the $e_n$-transform in situations where $\{a_m\}$ is not the solution of a difference equation. The strongest result concerning the $e_n$-transform is the following.

Theorem 2.2 Suppose the complex sequence $\{a_m\}$ is an element of $L(n,\Delta)$ for $m > m_0$ and that the roots of the associated characteristic equation are outside the unit circle. Then

$$e_n(A_m) = \sum_{v=k}^{\infty} a_v \quad \text{for all } m \geq m_0 .$$

Proof: See Gray, Houston, and Morgan (1978) for the case of $\{a_m\}$ real. The extension of the proof to include $\{a_m\}$ complex is trivial.

By applying the results of Theorems 2.1 and 2.2, the equivalence of $f^*_{p,q}(\cdot)$ and $f_{p,q}(\cdot)$ is easily shown. Morton (1981) has also proven this result in the context of power spectral density estimation.

Theorem 2.3 Let $f_{p,q}(\cdot)$, $f^*_{p,q}(\cdot)$, and $\alpha_1,\alpha_2,\ldots,\alpha_p$ be as defined previously, and suppose that the roots of $1-\alpha_1 x-\ldots-\alpha_p x^p = 0$ are outside the unit circle. Then we have

$$f^*_{p,q}(\cdot) \equiv f_{p,q}(\cdot) .$$

Proof: Let k be as in Definition 2.2 and consider

$$\sum_{v=k}^{\infty} \phi^*_{p,q}(v)e^{ivx}, \text{ where } \phi^*_{p,q}(v)$$

is the vth Fourier coefficient of $f^*_{p,q}(\cdot)$. Since $\{\phi^*_{p,q}(v)\}$ satisfies a pth order difference equation for $v > q$, then so does $\{\phi^*_{p,q}(v)e^{ivx}\}$. Therefore, by Theorem 2.2, we have

$$e_p(F^*_q(x)) = \sum_{v=k}^{\infty} \phi^*_{p,q}(v)e^{ivx} .$$

This implies that

$$\frac{1}{2\pi}[\phi^*_{p,q}(0) + 2\text{Real}\{ e_p(F^*_q(x)) - F_0{}^*(x)\}]$$

$$= \frac{1}{2\pi}[\phi^*_{p,q}(0) + 2\text{Real}( \sum_{v=1}^{\infty} \phi^*_{p,q}(v)e^{ivx})] = f^*_{p,q}(x) .$$

However, since $\phi^*_{p,q}(v) = \phi(v)$ for $|v| = 0,1,\ldots,p+q$ it follows that

$$\frac{1}{2\pi}[\phi^*_{p,q}(0) + 2\text{Real}\{ e_p(F^*_q(x)) - F^*_0(x)\}]$$

$$= \frac{1}{2\pi}[\phi(0) + 2\text{Real}\{ e_p(F_q(x))-F_0(x) \}] = f_{p,q}(x).$$

Thus, $f_{p,q}(x) = f^*_{p,q}(x)$.

Since $f^*_{p,q}(\cdot)$ and $f_{p,q}(\cdot)$ are equivalent under the condition (which shall henceforth be referred to as condition S) that the roots of $1 - \alpha_1 x - \ldots - \alpha_p x^p = 0$ lie outside the unit circle, it follows that $f_{p,q}(\cdot)$ satisfies the error properties of (2.4) under condition S. However, the following two important facts are noted at this time.

(a) If condition S is not satisfied, then $f^*_{p,q}(\cdot)$ and $f_{p,q}(\cdot)$ are not in general equivalent.

(b) If condition S is not satisfied, then neither $f^*_{p,q}(\cdot)$ nor $f_{p,q}(\cdot)$ possess the property that their first $p + q + 1$ Fourier coefficients are equal to $\phi(0)$, $\phi(1),\ldots, \phi(p+q)$.

Because of fact (b), it is not clear in what sense $f_{p,q}(\cdot)$ is

approximating $f(\cdot)$ when condition S is not satisfied. When using $f_{p,q}(\cdot)$ for approximation purposes it is thus important to always verify whether or not this condition is met.

In concluding this section two special cases of $f_{p,q}(\cdot)$ are noted. When $p = 0$

$$f_{0,q}(x) = \frac{1}{2\pi}[\phi(0) + 2\text{Real}(\sum_{v=1}^{q} \phi(v)e^{ivx})] ,$$

a Fourier series approximator, and when $q = 0$

$$f_{p,0}(x) = \frac{k_p}{2\pi} \frac{1}{|1 - \alpha_1 e^{ix} - \ldots - \alpha_p e^{ipx}|^2} ,$$

an autoregressive approximator. The first of these two relationships follows trivially from the definition of $e_n(A_m)$. The second follows from the fact, proven by Pagano (1973), that condition S is always satisfied whenever $q = 0$ (assuming $\{\phi(v)\}$ is positive definite), and thus, by Theorem 2.3, $f_{p,0}(\cdot) \equiv f^*_{p,0}(\cdot)$. Autoregressive approximators have an advantage over ARMA approximators in that they always satisfy condition S, which of course implies that $\phi_{p,0}(v) = \phi(v)$ for $|v| = 0,1,\ldots,p$. However, as will be illustrated in the next and succeeding sections, there is much to be gained in considering $f_{p,q}(\cdot)$ for $q > 0$.

## 2.3 Examples Comparing Fourier Series, Autoregressive, and ARMA Approximators

By way of illustration we will now compare the Fourier series, autoregressive, and ARMA ($p > 0$ and $q > 0$) methods of approximating a function. Since these methods are of interest to us in the context of density estimation, the examples to follow

involve, for the most part, functions which are commonly used as models for probability densities. Although they are certainly not exhaustive, the examples given serve to illustrate the value of the ARMA method as an approximation scheme. Numerous additional examples already exist which show dramatically how the $e_n$-transform accelerates the rate of convergence of slowly convergent sequences, and in some cases induces convergence of divergent sequences (see Gray, Houston, and Morgan (1978)). Since the sequences $\{F_m(x)\}$ associated with the functions of this section are not what would usually be considered slowly convergent, the examples which follow are not as dramatic as those just mentioned, but nonetheless interesting.

In our first example, we investigate how well the Fourier series and autoregressive methods fare in approximating a density for which there exists an error-free ARMA approximator. Consider the function

$$f^{(1)}(x) = \frac{1}{2}f_1(x) + \frac{1}{2}f_2(x) \quad (x \in [-\pi,\pi])$$

a mixture of the densities

$$f_1(x) = \frac{1}{2\pi}\left\{\frac{(.4742)|1-.50e^{ix}|^2|1-(.40e^{-i(\pi/8)})e^{ix}|^2}{|1-(.80e^{i(\pi/4)})e^{ix}|^2}\right\}$$

and

$$f_2(x) = \frac{1}{2\pi}\left\{\frac{(.2775)}{|1 + (.85i)e^{ix}|^2}\right\} .$$

A result which will be proven in Chapter III is that the mixture of densities having ARMA representations itself has an ARMA representation. With this result it is easily verified that

$f^{(1)}(\cdot)$ has an ARMA (2,3) representation. By the earlier results of this chapter, it then follows that the approximator $f^{(1)}_{2,3}(\cdot)$ is identical to $f^{(1)}(\cdot)$, or, in other words, $f^{(1)}(\cdot)$ is completely determined by its first five Fourier coefficients. Of interest, though, is a determination of how well the Fourier series and autoregressive approximation schemes perform in this situation.

In Figures 2.1 and 2.2, respectively, the Fourier series and autoregressive approximators based on $\phi^{(1)}(1),\ldots,\phi^{(1)}(5)$ have been plotted with $f^{(1)}(\cdot)$. Figure 2.3 shows a plot of $f^{(1)}_{0,10}(\cdot)$ and $f^{(1)}(\cdot)$, and in Table 2.1 a comparison of ISE is given for the two methods being considered. The ISE for each approximator has been approximated numerically by Simpson's rule using 201 function evaluations on $[-\pi,\pi]$. (The ISE in all the examples to follow has been calculated in the same way.) The autoregressive method is seen to perform considerably better in this instance than does the Fourier series method. In a visual sense $f^{(1)}_{6,0}(\cdot)$, $f^{(1)}_{7,0}(\cdot),\ldots,f^{(1)}_{15,0}(\cdot)$ are virtually indistinguisable from $f^{(1)}(\cdot)$ (and hence a plot of $f^{(1)}_{10,0}(\cdot)$ has been omitted). The Fourier series approximators, however, have difficulty in resolving the peaks of $f^{(1)}(\cdot)$ without introducing spurious variation. This shortcoming is even more important in the stochastic setting where it is desirable to limit the cause of spurious variation in a fitted curve to sampling variability. In Chapter VI a data set is discussed which verifies the practical importance of densities such as $f^{(1)}(\cdot)$ which have rather sharp peaks.

In our last three examples we compare the three different approximation schemes on functions which do not have ARMA repre-

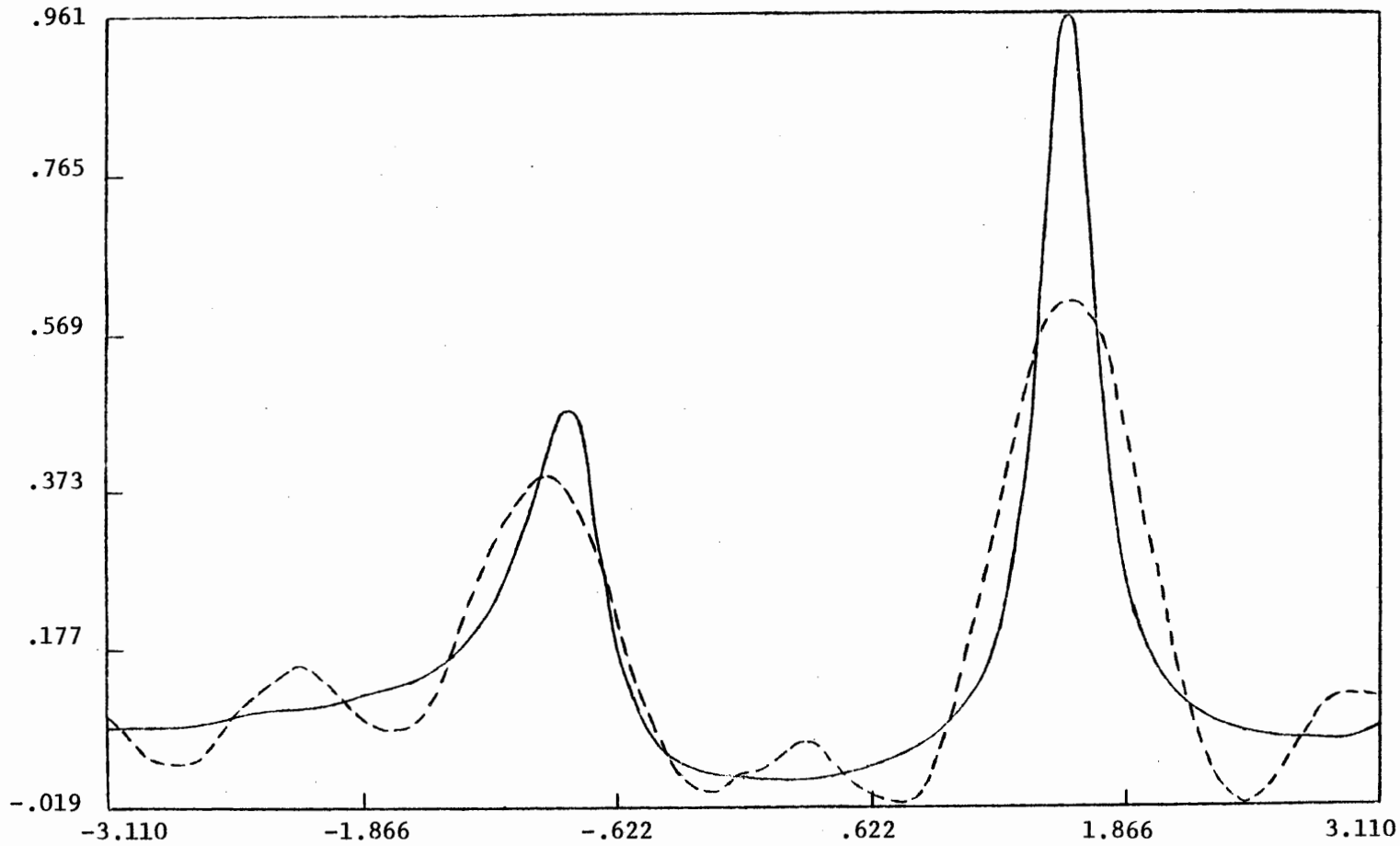


FIGURE 2.1

The Function $f^{(1)}(\cdot)$ and Fourier Series Approximator $f^{(1)}_{0,5}(\cdot)$

——— $f^{(1)}(\cdot)$ ------ $f^{(1)}_{0,5}(\cdot)$

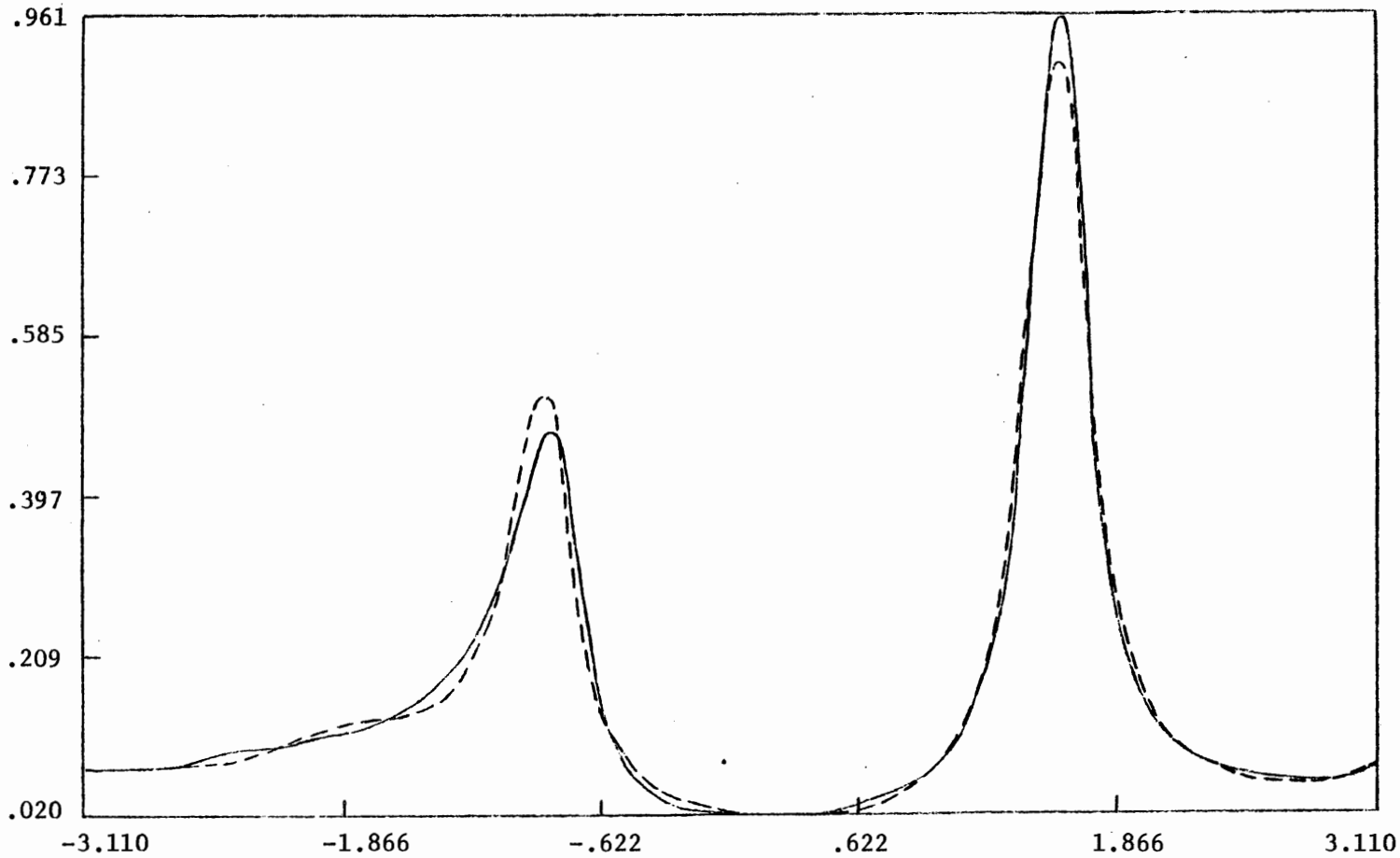


FIGURE 2.2

The Function $f^{(1)}(\cdot)$ and Autoregressive Approximator $f^{(1)}_{5,0}(\cdot)$

———— $f^{(1)}(\cdot)$ ----- $f^{(1)}_{5,0}(\cdot)$

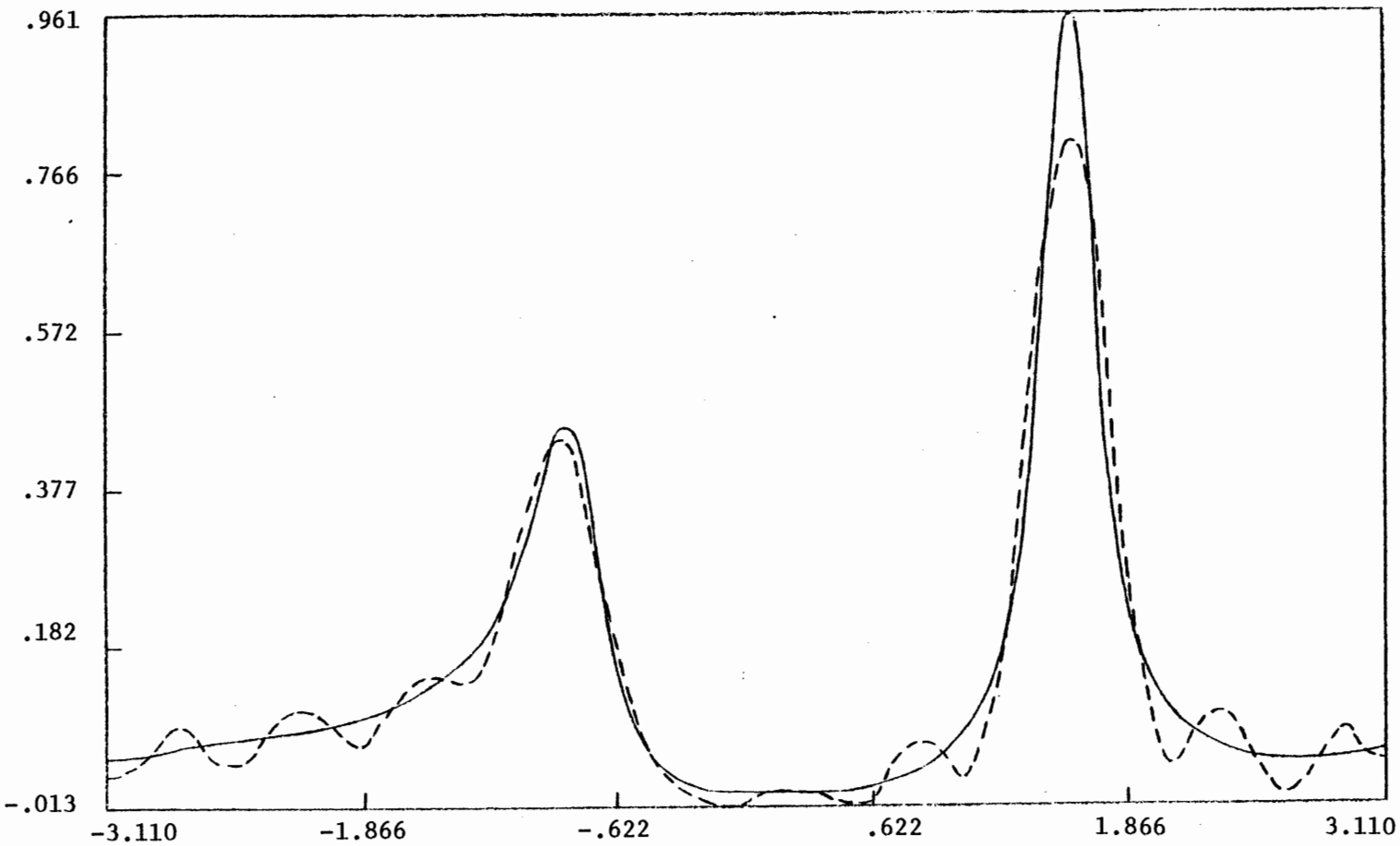


FIGURE 2.3

The Function $f^{(1)}(\cdot)$ and Fourier Series Approximator $f^{(1)}_{0,10}(\cdot)$

———— $f^{(1)}(\cdot)$ ----- $f^{(1)}_{0,10}(\cdot)$

TABLE 2.1

ISE COMPARISON FOR FOURIER SERIES AND AUTOREGRESSIVE APPROXIMATORS OF THE FUNCTION $f^{(1)}(\cdot)$

| k | ISE $(f^{(1)}_{0,k})$ | ISE $(f^{(1)}_{k,0})$ |
|---|---|---|
| 5 | .04216 | .00238 |
| 6 | .03332 | .00055 |
| 7 | .02574 | .00025 |
| 8 | .01495 | .00008 |
| 9 | .01307 | .00003 |
| 10 | .00865 | .00001 |
| 11 | .00577 | .00000 |
| 12 | .00487 | .00000 |
| 13 | .00299 | .00000 |
| 14 | .00230 | .00000 |
| 15 | .00174 | .00000 |

sentations. In this way the versatility of the ARMA method is investigated by examining its performance in situations which are other than ideal for it. The functions considered are

$$f^{(2)}(x) = \frac{1224}{\pi}[\frac{x+\pi}{2\pi}]^{15} \quad [1 - (\frac{x+\pi}{2\pi})]^2 I_{[-\pi,\pi]}(x),$$

$$f^{(3)}(x) = 2e^{-4|x|} I_{[-\pi,\pi]}(x),$$

and

$$f^{(4)}(x) = 6[\frac{x+\pi}{2}]^5 \exp\{-2[(x+\pi)/2]^6\} .$$

The function $f^{(2)}(\cdot)$ is simply a Beta (16,3) density which has been shifted and rescaled so that its support is the interval $[-\pi,\pi]$. The second function, $f^{(3)}(\cdot)$, is a truncated double exponential (or Laplace) density, and $f^{(4)}(\cdot)$ is a Weibull density (with scale parameter 2 and shape parameter 6) which has been truncated at $\pi$ and then shifted and rescaled to have support $[-\pi,\pi]$. Since $f^{(3)}(\cdot)$ and $f^{(4)}(\cdot)$ exclude, respectively, only .00035% and less than $10^{-97}$% of the area of the original densities, the comparisons to follow may be regarded as comparisons of the ARMA, Fourier series, and autoregressive density estimation methods in the absence of stochastic errors.

Pictured in Figures 2.4 - 2.12 are plots of various approximators along with the functions $f^{(i)}(\cdot)$, $i = 2,3,4$. Comparisons of ISE are given in Tables 2.2 - 2.4. Both visually and in terms of ISE, the ARMA approximators display a decided advantage over the other two approximation schemes. A hallmark of the ARMA method which surfaces in these three examples is the ability of ARMA approximators to correctly fit both the

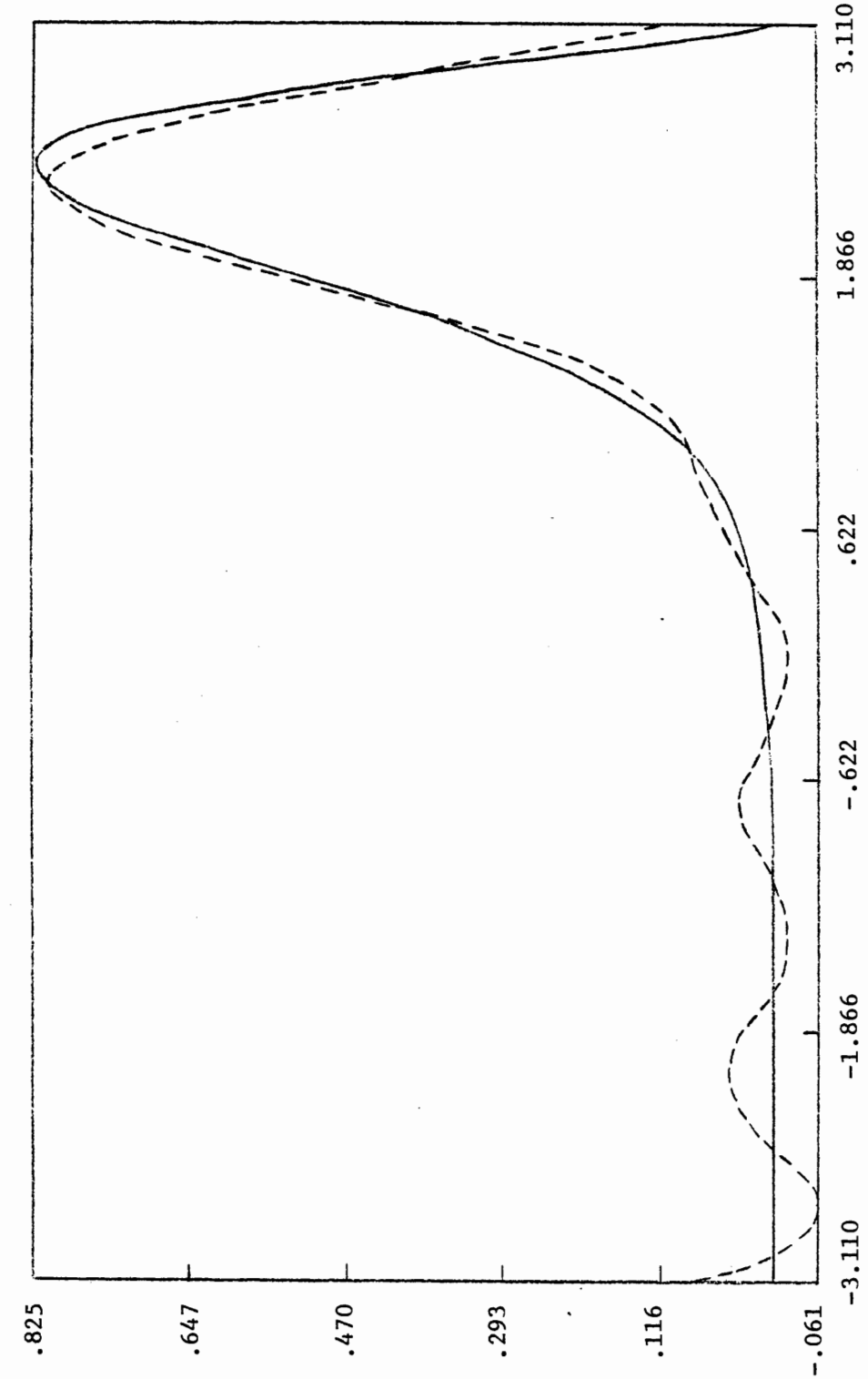


FIGURE 2.4

The Function $f^{(2)}(\cdot)$ and Fourier Series Approximator $f^{(2)}_{0,4}(\cdot)$

—— $f^{(2)}(\cdot)$ ------- $f^{(2)}_{0,4}(\cdot)$

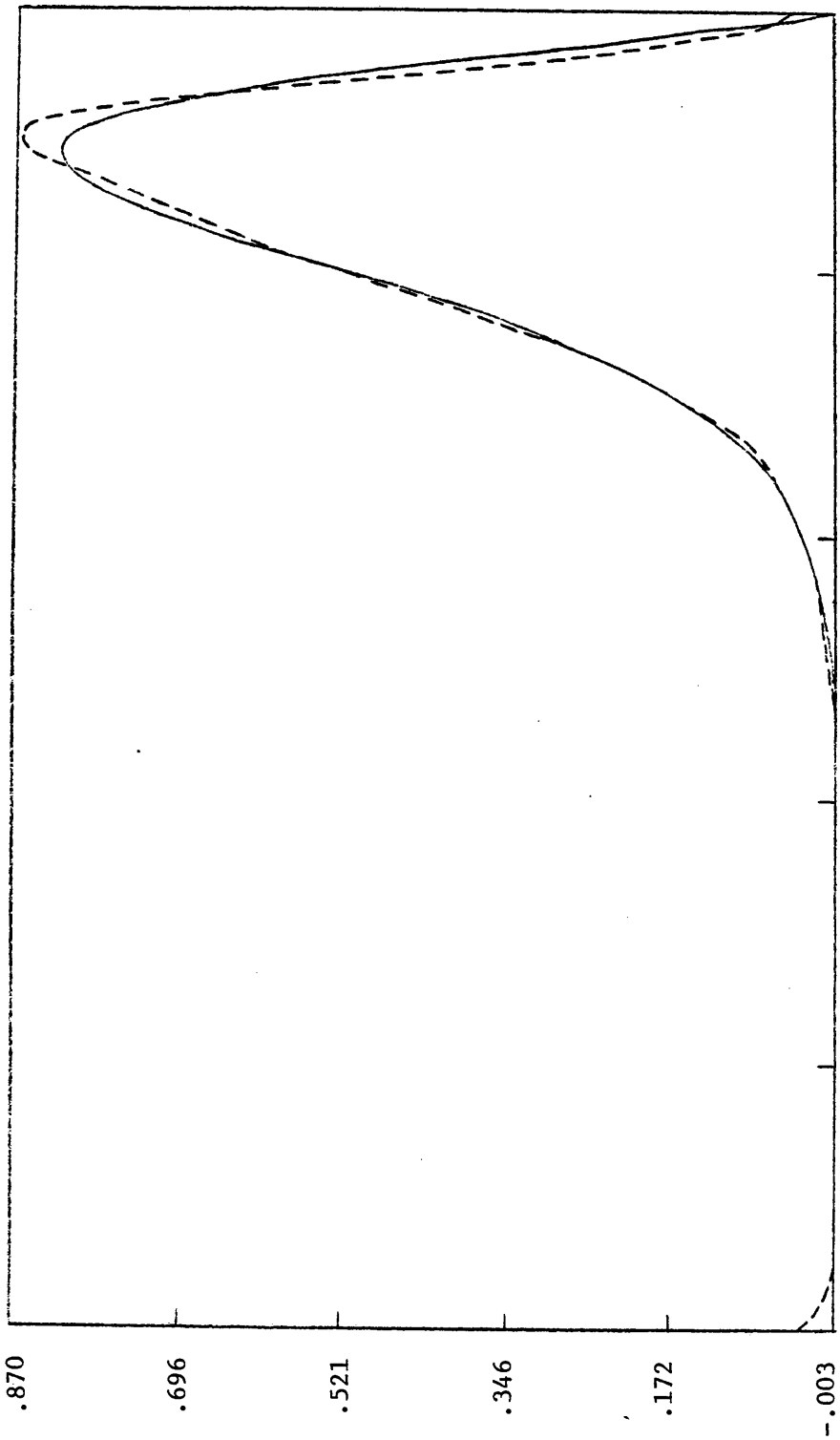


FIGURE 2.5

The Function $f^{(2)}(\cdot)$ and ARMA Approximator $f^{(2)}_{2,2}(\cdot)$

—— $f^{(2)}(\cdot)$ ----- $f^{(2)}_{2,2}(\cdot)$

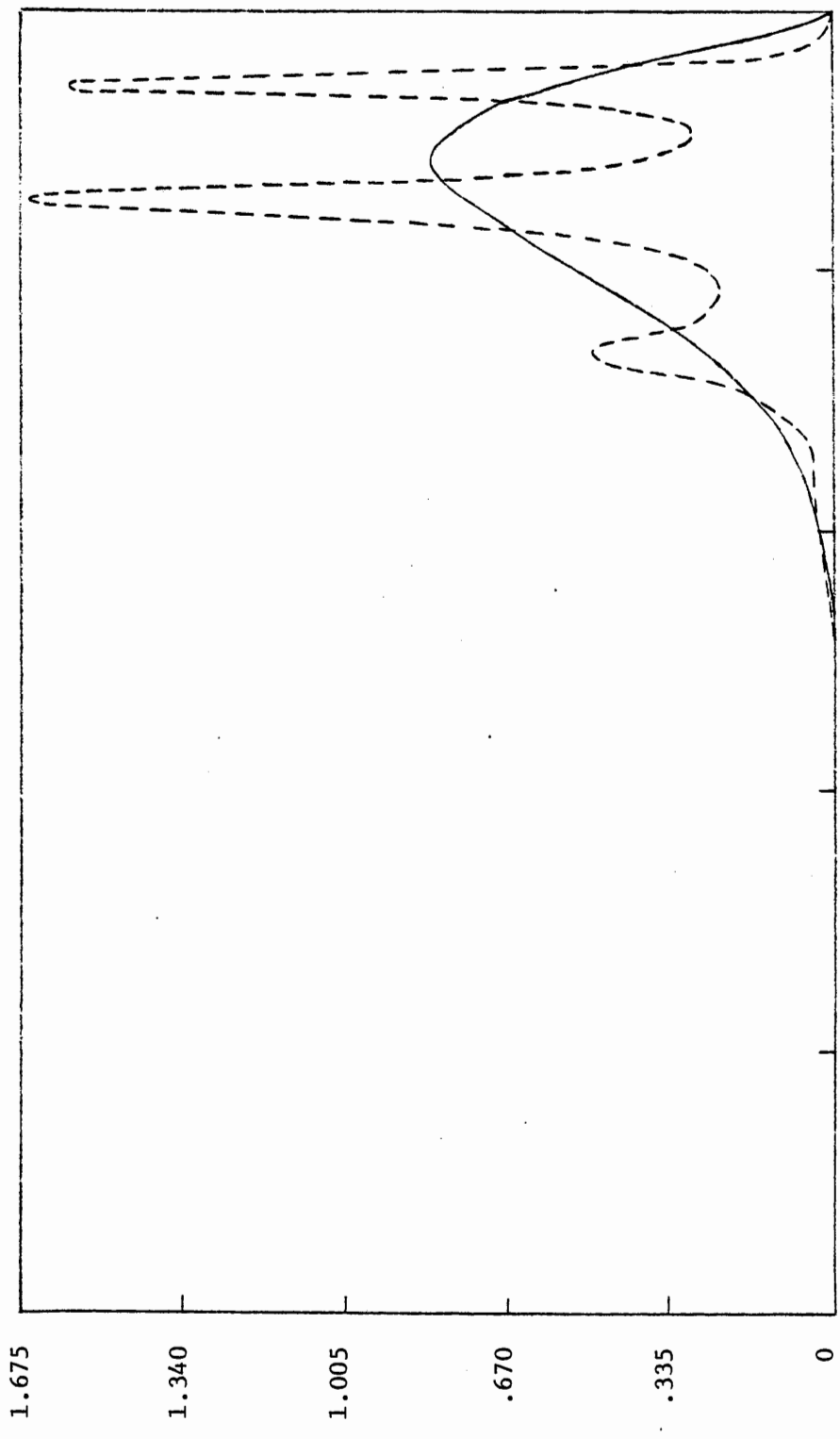


FIGURE 2.6

The Function $f^{(2)}(\cdot)$ and Autoregressive Approximator $f^{(2)}_{4,0}(\cdot)$

——— $f^{(2)}(\cdot)$ ----- $f^{(2)}_{4,0}(\cdot)$

TABLE 2.2

ISE COMPARISON FOR APPROXIMATORS OF THE FUNCTION $f^{(2)}(\cdot)$

| k | $ISE(f^{(2)}_{0,k})$ | $ISE(f^{(2)}_{1,k-1})$ | $ISE(f^{(2)}_{2,k-2})$ | $ISE(f^{(2)}_{3,k-3})$ | $ISE(f^{(2)}_{4,k-4})$ | $ISE(f^{(2)}_{k,0})$ |
|---|---|---|---|---|---|---|
| 3 | .02198 | .00733 | .02311 | .32789 | | .32789 |
| 4 | .00833 | .00167 | .00250 | .01416 | .26344 | .26344 |
| 5 | .00356 | .00050 | .00044 | .00120 | .00953 | .22365 |
| 6 | .00169 | .00019 | .00011 | .00017 | .00068 | .20246 |
| 7 | .00088 | .00008 | .00004 | .00004 | .00009 | .18367 |
| 8 | .00049 | .00004 | .00002 | .00001 | .00002 | .17349 |
| 9 | .00029 | .00002 | .00001 | .00000 | .00001 | .19042 |
| 10 | .00018 | .00001 | .00000 | .00000 | .00000 | .32217 |

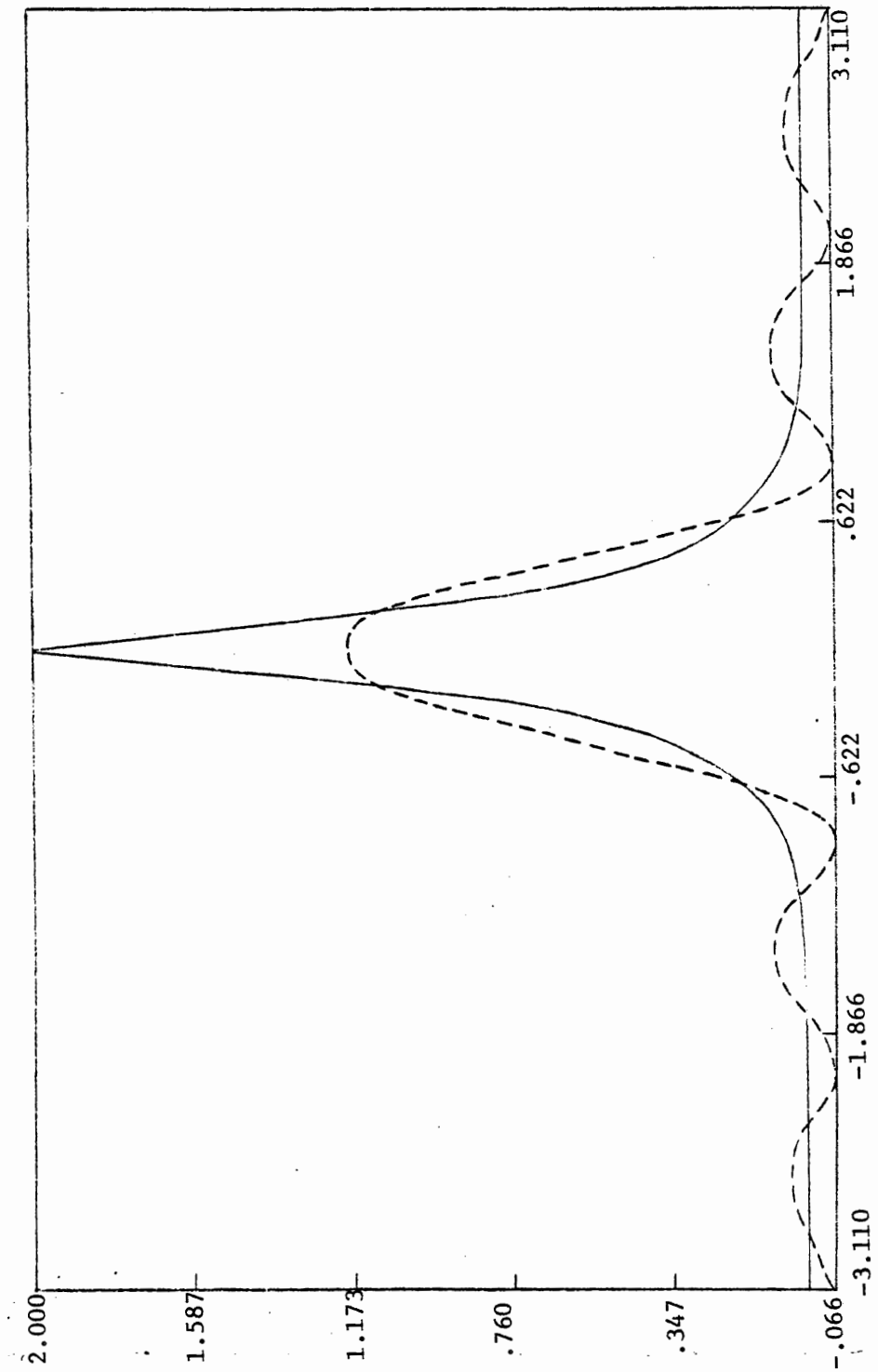


FIGURE 2.7

The Function $f^{(3)}(\cdot)$ and Fourier Series Approximator $f^{(3)}_{0,5}(\cdot)$

——— $f^{(3)}(\cdot)$ ----- $f^{(3)}_{0,5}(\cdot)$

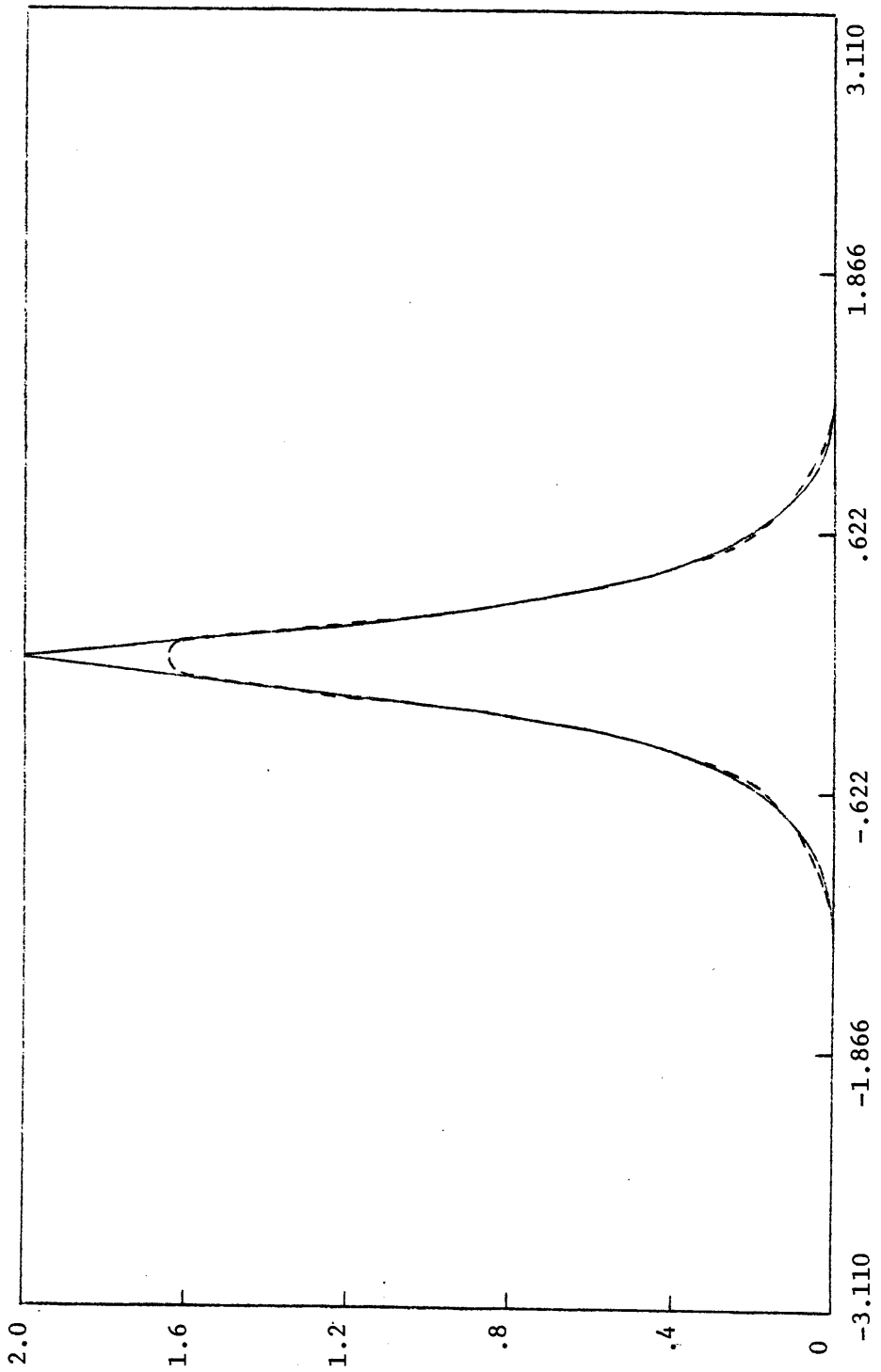


FIGURE 2.8

The Function $f^{(3)}(\cdot)$ and ARMA Approximator $f_{1,4}^{(3)}(\cdot)$

——— $f^{(3)}(\cdot)$ ------ $f_{1,4}^{(3)}(\cdot)$

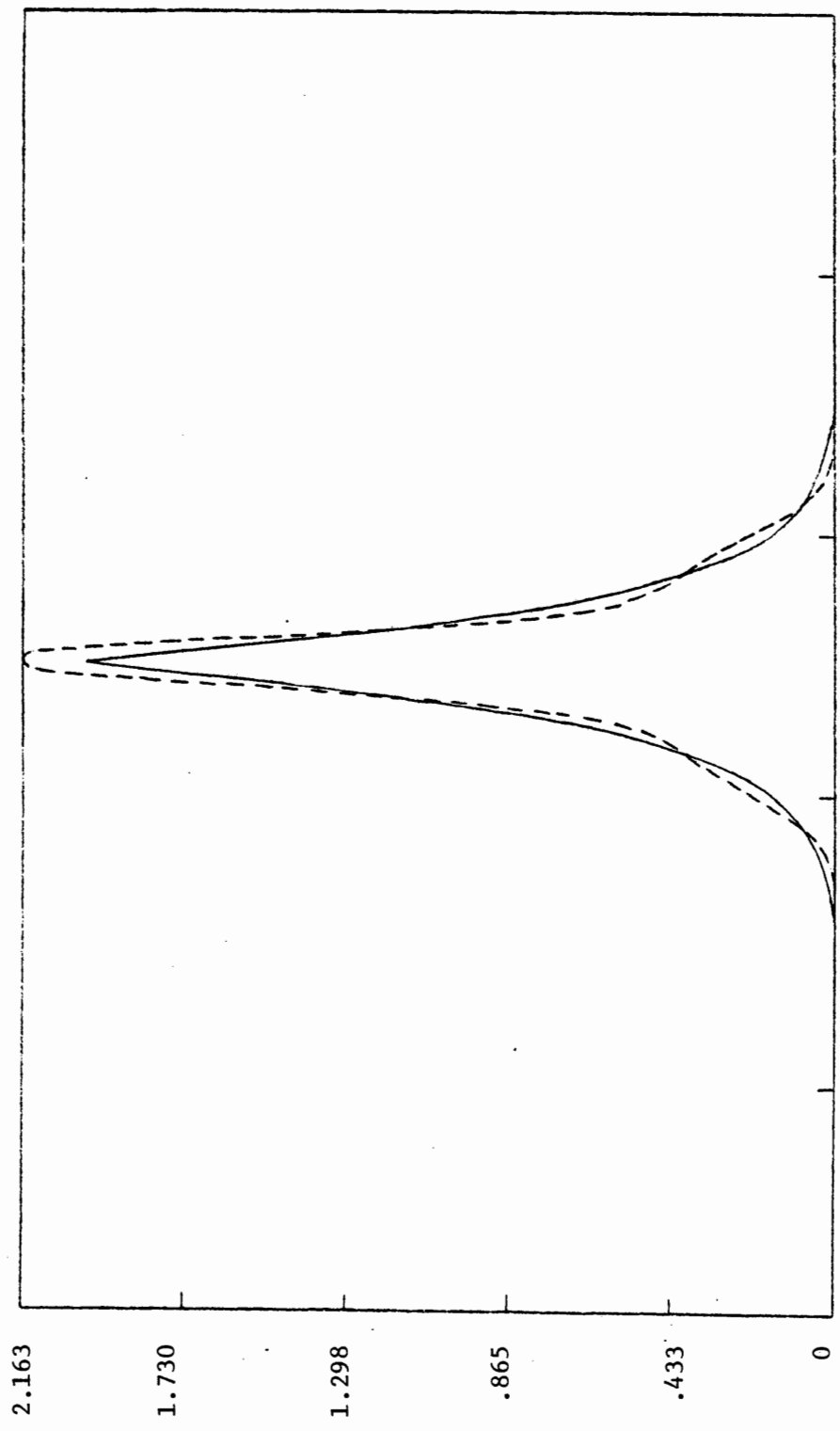


FIGURE 2.9

The Function $f^{(3)}(\cdot)$ and Autoregressive Approximator $f_{5,0}^{(3)}(\cdot)$

——— $f^{(3)}(\cdot)$ ----- $f_{5,0}^{(3)}(\cdot)$

TABLE 2.3

ISE COMPARISON FOR APPROXIMATORS OF THE FUNCTION $f^{(3)}(\cdot)$

| k | $ISE(f^{(3)}_{0,k})$ | $ISE(f^{(3)}_{1,k-1})$ | $ISE(f^{(3)}_{k,0})$ |
|---|---|---|---|
| 1 | .55890 | .84191 | .84191 |
| 2 | .35518 | .02652 | .31226 |
| 3 | .22480 | .00403 | .10221 |
| 4 | .14523 | .00602 | .08036 |
| 5 | .09675 | .00613 | .01784 |
| 6 | .06662 | .00517 | .02481 |
| 7 | .04733 | .00410 | .00407 |
| 8 | .03460 | .00319 | .00969 |
| 9 | .02595 | .00249 | .00151 |
| 10 | .01989 | .00196 | .00472 |
| 11 | .01555 | .00157 | .00095 |
| 12 | .01237 | .00127 | .00274 |
| 13 | .00999 | .00105 | .00077 |
| 14 | .00818 | .00088 | .00180 |
| 15 | .00678 | .00075 | .00067 |

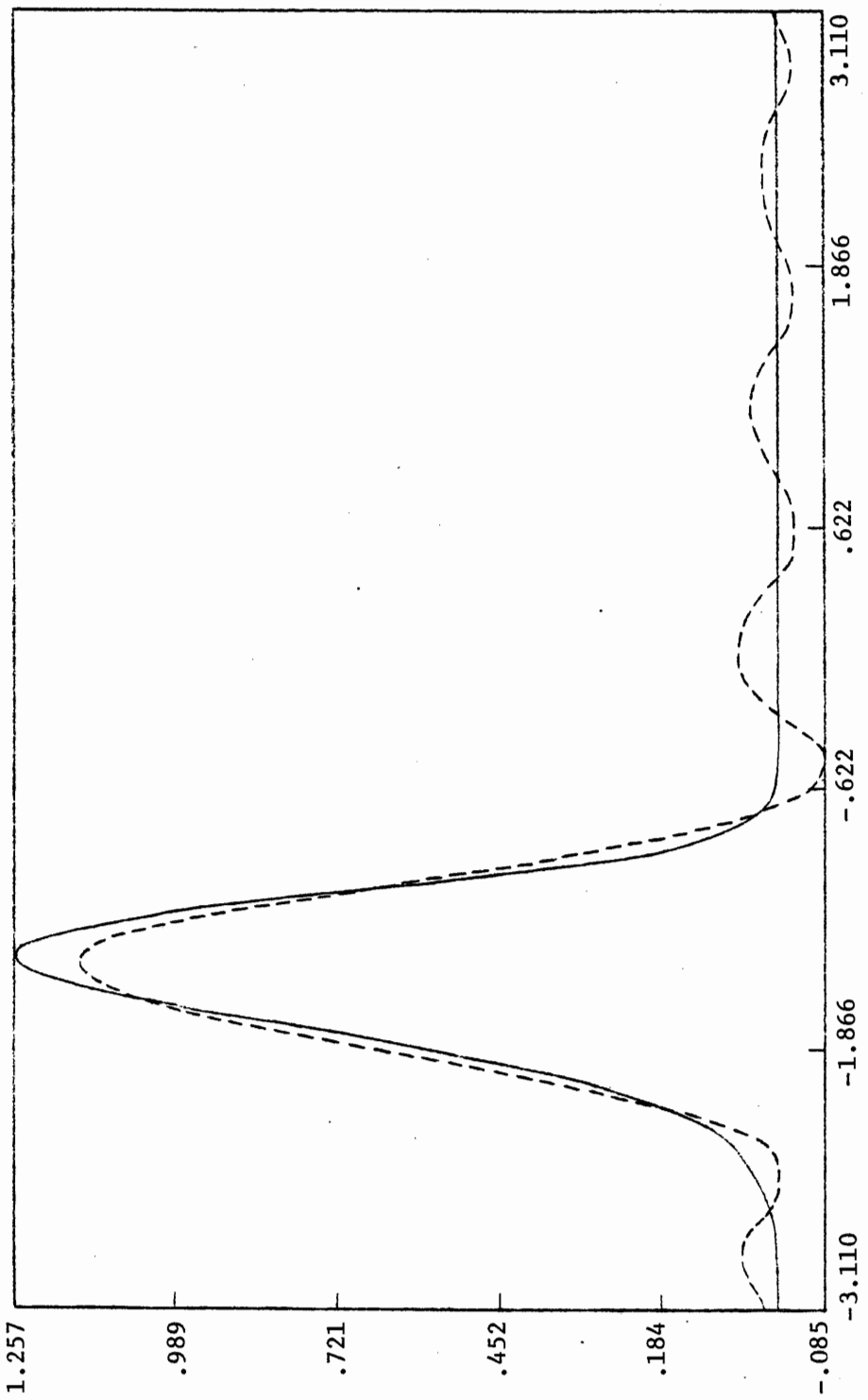


FIGURE 2.10

The Function $f^{(4)}(\cdot)$ and Fourier Series Approximator $f^{(4)}_{0,5}(\cdot)$

——— $f^{(4)}(\cdot)$ ------ $f^{(4)}_{0,5}(\cdot)$

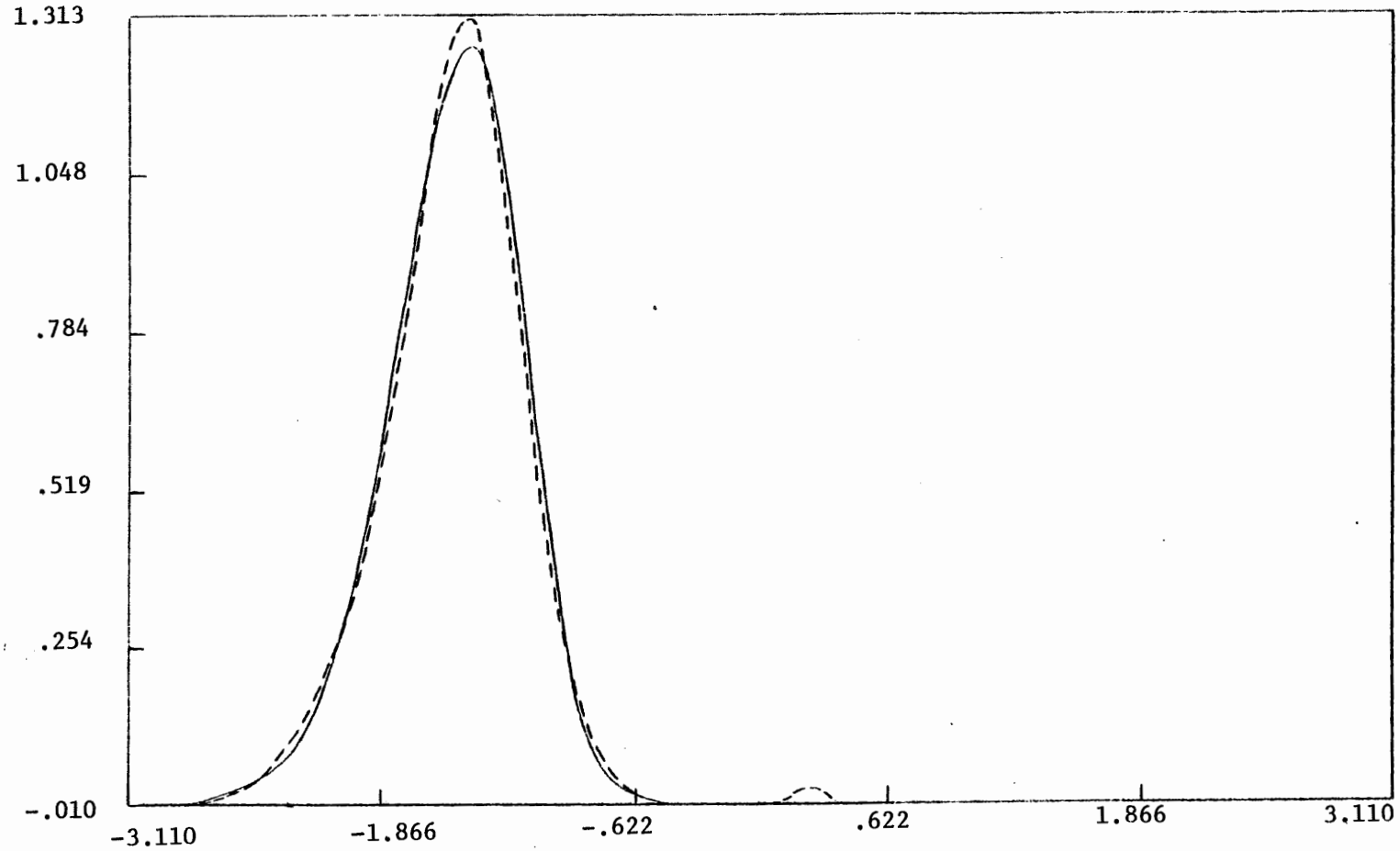


FIGURE 2.11

The Function $f^{(4)}(\cdot)$ and ARMA Approximator $f^{(4)}_{1,4}(\cdot)$

——— $f^{(4)}(\cdot)$ ----- $f^{(4)}_{1,4}(\cdot)$

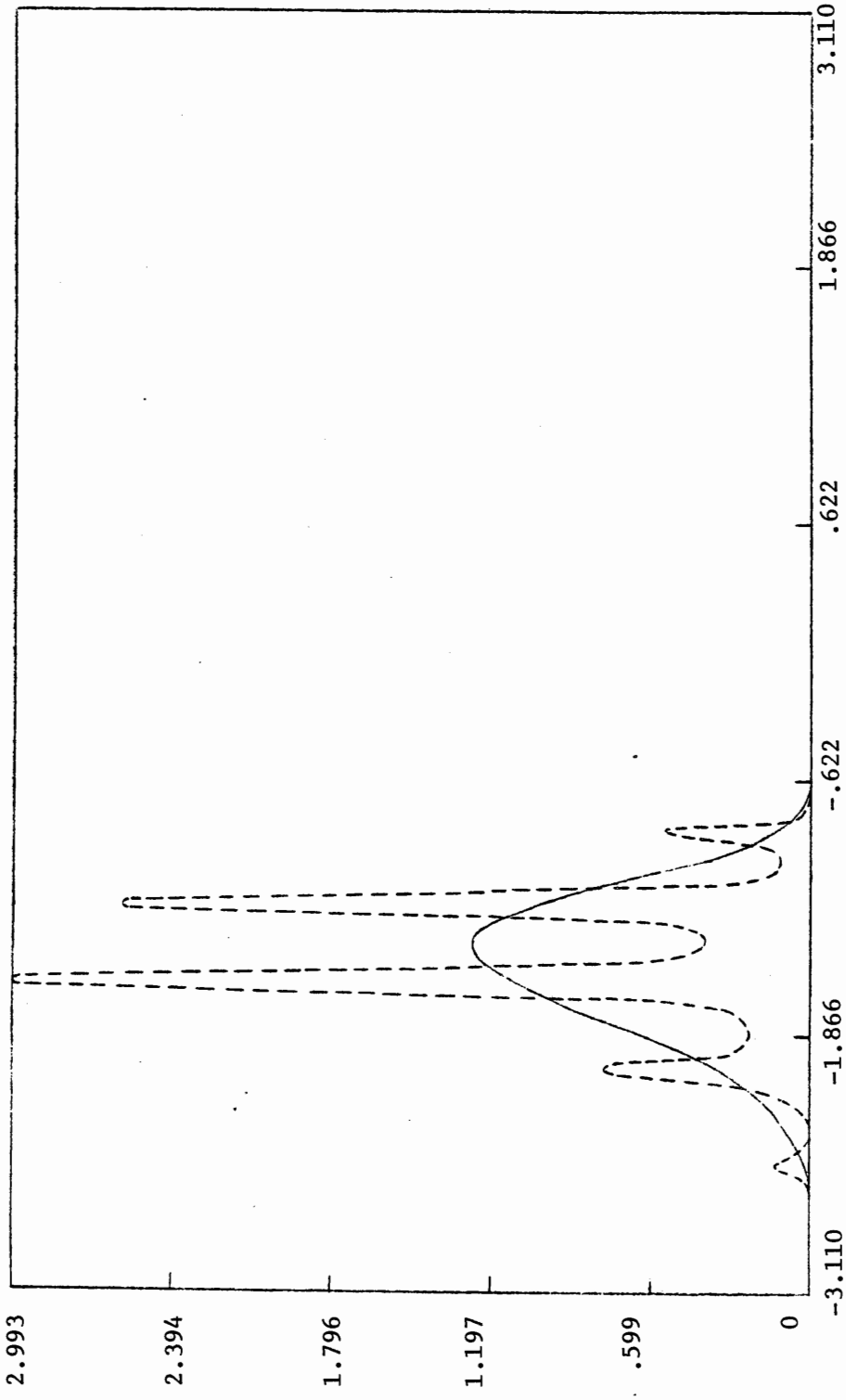


FIGURE 2.12

The Function $f^{(4)}(\cdot)$ and Autoregressive Approximator $f_{5,0}^{(4)}(\cdot)$

—— $f^{(4)}(\cdot)$ ----- $f_{5,0}^{(4)}(\cdot)$

TABLE 2.4

ISE COMPARISONS FOR APPROXIMATORS OF THE FUNCTION $f^{(4)}(\cdot)$

| k | $ISE(f^{(4)}_{0,k})$ | $ISE(f^{(4)}_{1,k-1})$ | $ISE(f^{(4)}_{2,k-2})$ | $ISE(f^{(4)}_{k,0})$ |
|---|---|---|---|---|
| 1 | .44243 | 1.84975 | | 1.84975 |
| 2 | .23112 | .20863 | 1.22354 | 1.22354 |
| 3 | .10407 | .03968 | .11457 | .92122 |
| 4 | .04104 | .00804 | .01615 | .87385 |
| 5 | .01463 | .00155 | .00228 | .68997 |
| 6 | .00488 | .00028 | .00030 | .66224 |
| 7 | .00155 | .00005 | .00003 | .28571 |
| 8 | .00047 | .00001 | .00000 | 1.27134 |
| 9 | .00013 | .00000 | .00000 | .54683 |
| 10 | .00004 | .00000 | .00000 | .90297 |

tails and the peak of a function. In Figures 2.4, 2.7, and 2.10 the Fourier series approximators are seen to correctly (or nearly correctly) fit the peak of each function only at the expense of incorrectly fitting the tails. By contrast, the ARMA approximators of Figures 2.5, 2.8 and 2.11 (based in each case on the same number of Fourier coefficients as the corresponding Fourier series approximator) smooth out variation in the tails while still correctly fitting the peaks.

The autoregressive method performs quite well on the function $f^{(3)}(\cdot)$ but does very poorly on $f^{(2)}(\cdot)$ and $f^{(4)}(\cdot)$. This phenomenon can be explained quite simply by examining the Fourier series representation of the approximator $f_{k,0}(\cdot)$. By property (2.4) we have

$$|f(x)-f_{k,0}(x)| = \frac{1}{2\pi} \left| \sum_{|v|>k} (\phi(v)-\phi_{k,0}(v))e^{ivx} \right|$$

and

$$ISE(f_{k,0}) = \frac{1}{\pi} \sum_{v=k+1}^{\infty} |\phi(v) - \phi_{k,0}(v)|^2 .$$

The approximator $f_{k,0}(\cdot)$ obviously, then, performs poorly if it does a poor job of extrapolating the Fourier coefficients $\phi(k+1)$, $\phi(k+2),\ldots$ . This is clearly what has occurred in the examples involving $f^{(2)}(\cdot)$ and $f^{(4)}(\cdot)$. Our examples seem to indicate that, in general, fixing the autoregressive order and allowing the moving average order to increase is the best scheme for reducing the error inherent in Fourier series approximators.

Having examined the advantages of using the ARMA (as opposed to Fourier series or autoregressive) approximation method the remainder of this work is devoted to an investigation of the ARMA method in the stochastic setting of probability density estimation.

# CHAPTER III

# SMALL SAMPLE PROPERTIES OF ARMA DENSITY ESTIMATORS

## 3.1 Introduction

We now formally begin our study of probability density estimation via ARMA representations. In the current chapter we introduce the estimation problem and define an ARMA estimator $\hat{f}_{p,q}(\cdot)$. Alternative ways of expressing $\hat{f}_{p,q}(\cdot)$ are derived which serve to motivate ARMA estimators and show explicitly their relationship to Fourier series estimators. The main result of this chapter, however, will be establishing the relationship between $\hat{f}_{p,q}(\cdot)$ and the generalized jackknife statistic. It will be shown that ARMA estimators employ an adaptive, higher order generalized jackknife scheme.

Chapter III is concluded with a result concerning the mixture of densities having ARMA representations. The mixture of autoregressive densities is seen, in general, to be an ARMA density. This result conveys the necessity of ARMA representations to a theory based on the representation of densities by autoregressive schemes.

## 3.2 Definition of the Estimation Problem and $\hat{f}_{p,q}(\cdot)$

Suppose Y is a random variable with continuous probability density function $g(\cdot)$ and that a random sample $Y_1,\ldots,Y_n$ is obtained from $g(\cdot)$. In the remainder of this work we shall be concerned with

the problem of estimating the function $g(\cdot)$.

All theoretical results will be based upon the assumption that Y has the finite support [a,b]. To be consistent with previous notation, we shall in this situation consider estimating the density $f(\cdot)$ of the random variable

$$X = \frac{\pi}{(b-a)}[2Y - (b+a)]$$

which has support $[-\pi,\pi]$. As before it is also assumed that $f(\cdot)$ has the Fourier series representation

$$f(x) = \frac{1}{2\pi} \sum_{v=-\infty}^{\infty} \phi(v) e^{ivx} , \qquad \text{a.e. } [-\pi,\pi] .$$

Tapia and Thompson (1978) note that the finite support assumption is only a small liability in practice since, in the absence of any prior information about $g(\cdot)$, it would be unreasonable to estimate the density outside the range of the data. If the support of Y is indeed infinite, or unknown, then a and b may be replaced, for a given data set, by $y_{(0)}$ and $y_{(n+1)}$, where

(i) $y_{(0)}$ and $y_{(n+1)}$ are "natural" minimum and maximum values for the random variable Y, or

(ii) $y_{(0)} = y_{(1)}$ and $y_{(n+1)} = y_{(n)}$ ($y_{(i)}$ denotes the ith order statistic of the random sample $y_1,\ldots,y_n$).

The density $g(\cdot)$ is then estimated over the interval $[y_{(0)},y_{(n+1)}]$ by first estimating an associated $f(\cdot)$ over the interval $[-\pi,\pi]$ using the transformed sample

$$x_i = \frac{\pi}{(y_{(n+1)}-y_{(0)})} [2y_i - (y_{(n+1)} + y_{(0)})] \quad (i = 1,\ldots,n).$$

Under the finite support assumption $f(\cdot)$ is characterized by the Fourier coefficients

$$\phi(v) = \int_{-\pi}^{\pi} e^{-ivx} dF(x), \quad v = 1,2,\ldots,$$

where $F(\cdot)$ is the cumulative distribution function (cdf) of X. Given a random sample $X_1,\ldots,X_n$ from $f(\cdot)$ we shall estimate $\phi(v)$ by forming an appropriate functional of the empirical cdf $F_n(\cdot)$, i.e.

$$\hat{\phi}(v) = \int_{-\pi}^{\pi} e^{-ivx} dF_n(x)$$

$$= \frac{1}{n} \sum_{j=1}^{n} e^{-ivX_j}, \quad v = 1,2,\ldots .$$

The empirical characteristic function $\hat{\phi}(v)$ is obviously unbiased for $\phi(v)$ and also possesses the following easily established properties (see Tarter and Kronmal (1970)):

$$\begin{aligned} \operatorname{var}(\hat{\phi}(v)) &= \frac{1}{n}(1 - |\phi(v)|^2) \\ \operatorname{cov}(\hat{\phi}(v_1), \hat{\phi}(v_2)) &= E(\hat{\phi}(v_1)\overline{\hat{\phi}(v_2)}) - \phi(v_1)\overline{\phi(v_2)} \\ &= \frac{1}{n}[\phi(v_1-v_2)-\phi(v_1)\phi(-v_2)], \quad v_1 \neq v_2 . \end{aligned} \tag{3.1}$$

For the situation where the support of the original random variable Y is infinite or unknown we have

$$\hat{\phi}(v) = \frac{1}{n} \sum_{j=1}^{n} e^{-ivX_j}$$

$$= \frac{1}{n} \sum_{j=1}^{n} \exp\{ \frac{-iv\pi}{(Y_{(n+1)}-Y_{(0)})} [2Y_j-(Y_{(n+1)}+Y_{(0)})]\}.$$

In this case $\hat{\phi}(v)$ is unbiased for the parameter

$$\phi^*(v) = E(e^{-ivX})$$

$$= \int_{-\pi}^{\pi} e^{-ivx} f(x)\, dx ,$$

where $f(\cdot)$ is the density of the random variable

$$X = \frac{\pi}{(Y_{(n+1)} - Y_{(0)})} [2Y_1 - (Y_{(n+1)} + Y_{(0)})] .$$

The density being estimated on $[-\pi,\pi]$ by the methods to be discussed below is thus

$$f(x) = \frac{1}{2\pi} \sum_{v=-\infty}^{\infty} \phi^*(v) e^{ivx}$$

(where it is assumed that this Fourier series converges). We note that if $Y_{(0)}$ and $Y_{(n+1)}$ are nonstochastic the properties in (3.1) hold if $\phi(v)$ is replaced by $\phi^*(v)$.

We are now ready to define the ARMA estimator $\hat{f}_{p,q}(\cdot)$ of $f(\cdot)$, where it is understood that $f(\cdot)$ arises in one of the two ways described above.

Definition 3.1 Let $\{\hat{\phi}(k), \hat{\phi}(k+1), \ldots\}$ be a sequence of estimated Fourier coefficients. Then the ARMA estimator $\hat{f}_{p,q}(\cdot)$ of $f(\cdot)$ is defined by

$$\hat{f}_{p,q}(x) = \frac{1}{2\pi}[1 + 2\text{Real}\{e_p(\hat{F}_q(x)) - \hat{F}_0(x)\}], \quad x \in [-\pi,\pi]$$

where

$$k = \begin{cases} q + 1 - p, & q + 1 - p \leq 0 \\ 1, & q + 1 - p > 0 \end{cases}$$

and

$$\hat{F}_j(x) = \begin{cases} \sum_{v=k}^{j} \hat{\phi}(v)e^{ivx} , & j \geq k \\ 0 , & j < k \end{cases} .$$

It is seen that $\hat{f}_{p,q}(\cdot)$ is simply the stochastic analog of the approximator $f_{p,q}(\cdot)$. Just as in the deterministic setting we have the two special cases

$$\hat{f}_{0,q}(x) = \frac{1}{2\pi}[1 + 2\text{Real}(\sum_{v=1}^{q} \hat{\phi}(v)e^{ivx})]$$

and

$$\hat{f}_{p,0}(x) = \frac{\hat{k}_p}{2\pi} \cdot \frac{1}{|1-\hat{\alpha}_1 e^{ix} - \ldots - \hat{\alpha}_p e^{ipx}|^2} ,$$

a Fourier series and autoregressive estimator respectively.

## 3.3 The Generalized Jackknife Property of $\hat{f}_{p,q}(\cdot)$

Schucany, Gray and Owen (1971) introduced a generalized notion of the jackknife statistic which greatly enhances the effectiveness of the jackknife as a bias reduction tool. Their work exploits the specific form of the bias expansion of an estimator and gives the proper notion for reapplication of the jackknife.

Following Gray and Schucany (1972) the generalized jackknife may be defined as follows.

Definition 3.2 Let $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_{k+1}$ be $k + 1$ estimators for $\theta$ based on the random sample $X_1, \ldots, X_n$. Further, let $a_{ij}$, $i=1,\ldots,k$ and $j = 1,\ldots,k+1$, be real numbers satisfying

$$\begin{vmatrix} 1 & 1 & \cdots & 1 \\ a_{11} & a_{12} & \cdots & a_{1,k+1} \\ \vdots & \vdots & & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{k,k+1} \end{vmatrix} \neq 0 \quad . \tag{3.2}$$

Then the generalized jackknife $G(\hat{\theta}_1,\hat{\theta}_2,\ldots,\hat{\theta}_{k+1})$ is defined by

$$G(\hat{\theta}_1,\hat{\theta}_2,\ldots,\hat{\theta}_{k+1}) = \frac{\begin{vmatrix} \hat{\theta}_1 & \hat{\theta}_2 & \cdots & \hat{\theta}_{k+1} \\ a_{11} & a_{12} & \cdots & a_{1,k+1} \\ \vdots & \vdots & & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{k,k+1} \end{vmatrix}}{\begin{vmatrix} 1 & 1 & \cdots & 1 \\ a_{11} & a_{12} & \cdots & a_{1,k+1} \\ \vdots & \vdots & & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{k,k+1} \end{vmatrix}} \quad .$$

A simple form for the bias of the generalized jackknife is obtained in the following theorem.

Theorem 3.1 If

$$E(\hat{\theta}_j) - \theta = \sum_{i=1}^{\infty} h_{ij}(n)b_i(\theta), \quad j = 1,2,\ldots,k+1$$

and (5.2) is satisfied with $a_{ij} = h_{ij}(n)$, then

$$E[G(\hat{\theta}_1,\hat{\theta}_2,\ldots,\hat{\theta}_{k+1})] = \theta + B_G(n,\theta),$$

where

$$B_G(n,\theta) = \frac{\begin{vmatrix} B_1 & B_2 & \cdots & B_{k+1} \\ h_{11}(n) & h_{12}(n) & \cdots & h_{1,k+1}(n) \\ \vdots & \vdots & & \vdots \\ h_{k1}(n) & h_{k2}(n) & \cdots & h_{k,k+1}(n) \end{vmatrix}}{\begin{vmatrix} 1 & 1 & \cdots & 1 \\ h_{11}(n) & h_{12}(n) & \cdots & h_{1,k+1}(n) \\ \vdots & \vdots & & \vdots \\ h_{k1}(n) & h_{k2}(n) & \cdots & h_{k,k+1}(n) \end{vmatrix}}$$

and

$$B_j = \sum_{i=k+1}^{\infty} h_{ij}(n) b_i(\theta), \quad j = 1,2,\ldots,k+1.$$

Proof: See Gray and Schucany (1972).

An immediate corollary to Theorem 3.1 is that $G(\hat{\theta}_1,\hat{\theta}_2,\ldots,\hat{\theta}_{k+1})$ is unbiased for $\theta$ if

$$E(\hat{\theta}_j) = \theta + \sum_{i=1}^{k} h_{ij}(n) b_i(\theta), \quad j = 1,2,\ldots,k+1.$$

In order to see the sense in which $\hat{f}_{p,q}(\cdot)$ is related to the generalized jackknife recall that

$$\hat{f}_{p,q}(x) = \frac{1}{2\pi}\left[1 + 2\text{Real}\{e_p(\hat{F}_q(x)) - \hat{F}_0(x)\}\right]$$

where

$$e_p(\hat{F}_q(x)) = \frac{\begin{vmatrix} \hat{F}_{q-p}(x) & \hat{F}_{q-p+1}(x) & \cdots & \hat{F}_q(x) \\ \hat{\phi}(q-p+1)e^{i(q-p+1)x} & \hat{\phi}(q-p+2)e^{i(q-p+2)x} & \cdots & \hat{\phi}(q+1)e^{i(q+1)x} \\ \vdots & \vdots & & \vdots \\ \hat{\phi}(q)e^{iqx} & \hat{\phi}(q+1)e^{i(q+1)x} & \cdots & \hat{\phi}(q+p)e^{i(q+p)x} \end{vmatrix}}{\begin{vmatrix} 1 & 1 & \cdots & 1 \\ \hat{\phi}(q-p+1)e^{i(q-p+1)x} & \hat{\phi}(q-p+2)e^{i(q-p+2)x} & \cdots & \hat{\phi}(q+1)e^{i(q+1)x} \\ \vdots & \vdots & & \vdots \\ \hat{\phi}(q)e^{iqx} & \hat{\phi}(q+1)e^{i(q+1)x} & \cdots & \hat{\phi}(q+p)e^{i(q+p)x} \end{vmatrix}}.$$

Now, if $G(\hat{F}_{q-p}(x), \hat{F}_{q-p+1}(x),\ldots,\hat{F}_q(x))$ is the statistic obtained by replacing $\hat{\phi}(j)$ in the above determinants by fixed, known quantities, then G is a generalized jackknife statistic. More importantly, we note that $e_p(\hat{F}_q(x))$ and each of $\hat{F}_j(x)$, $j = q-p,\ldots,q$, are estimators of

$$\sum_{v=k}^{\infty} \phi(v)e^{ivx}$$

(where it is assumed that the support of Y is finite and known), and that $\hat{F}_j(x)$ has the bias expansion

$$E[\hat{F}_j(x)] - \sum_{v=k}^{\infty} \phi(v)e^{ivx} = -\sum_{v=j+1}^{\infty} \phi(v)e^{ivx} = -\sum_{v=1}^{\infty} \phi(v+j)e^{i(v+j)x}.$$

In the notation of Theorem 3.1, and allowing $h_{mj}(n)$ to depend on unknown parameters, we then have

$$\phi(m+j+q-p-1)e^{i(m+j+q-p-1)x} = h_{mj}(n),$$

$j = 1,\ldots, p+1$, $m = 1,2,\ldots$ and $b_m(\theta) \equiv -1$.

$$\begin{vmatrix} -\sum_{v=q-p+1}^{\infty} \phi(v)e^{ivx} & -\sum_{v=q-p+2}^{\infty} \phi(v)e^{ivx} & \cdots & -\sum_{v=q+1}^{\infty} \phi(v)e^{ivx} \\ \phi(q-p+1)e^{i(q-p+1)x} & \phi(q-p+2)e^{i(q-p+2)x} & \cdots & \phi(q+1)e^{i(q+1)x} \\ \vdots & \vdots & & \vdots \\ \phi(q)e^{iqx} & \phi(q+1)e^{i(q+1)x} & \cdots & \phi(q+p)e^{i(q+p)x} \end{vmatrix}$$

$$\begin{vmatrix} 1 & 1 & \cdots & 1 \\ \phi(q-p+1)e^{i(q-p+1)x} & \phi(q-p+2)e^{i(q-p+2)x} & \cdots & \phi(q+1)e^{i(q+1)x} \\ \vdots & \vdots & & \vdots \\ \phi(q)e^{iqx} & \phi(q+1)e^{i(q+1)x} & \cdots & \phi(q+p)e^{i(q+p)x} \end{vmatrix}$$

Therefore $e_p(\hat{F}_q(x))$ may be regarded as estimating a random variable G* whose bias has the same form as that of a generalized jackknife.

As has been pointed out previously, the ARMA method is especially effective in approximating functions whose Fourier coefficients are well approximated by the solution of a linear, homogeneous, difference equation with constant coefficients. Of interest, then, is the bias of $e_p(\hat{F}_q(x))$ under the assumption that the density $f(\cdot)$ has an ARMA(p,q) representation. Under this assumption we have, by Theorem 2.2,

$$\sum_{v=k}^{\infty} \phi(v)e^{ivx} = e_p(F_q(x)) \ .$$

Thus, we have immediately that

$$\text{Bias}[e_p(\hat{F}_q(x))] = E[e_p(\hat{F}_q(x)) - e_p(F_q(x))] \ .$$

It is still possible, however, to gain some additional insight into $e_p(\hat{F}_q(x))$ by showing precisely how it is related to the jackknife under the ARMA assumption. To illustrate this relationship, we note that, again by Theorem 2.2,

$$\text{Bias}[\hat{F}_j(x)] = -\sum_{v=j+1}^{\infty} \phi(v)e^{ivx} = -e_p\left(\sum_{v=j+1}^{j+p} \phi(v)e^{ivx}\right),$$

$j = q-p,\ldots,q$. An alternative form for expressing the $e_n$-transform is

$$e_n(A_m) = \frac{c_{m-n}A_{m-n} + c_{m-n+1}A_{m-n+1} + \ldots + c_m A_m}{c_{m-n} + c_{m-n+1} + \ldots + c_m},$$

which is obtained by expanding the determinants in the definition of $e_n(A_m)$ by cofactors of the first row. Using this form of $e_n$ we have (to be proven in Section 3.4)

$$e_p\left(\sum_{v=j+1}^{j+p} \phi(v)e^{ivx}\right)$$

$$= \frac{\sum_{v=j+1}^{j+p} \phi(v)e^{ivx} - \alpha_1 e^{ix}\sum_{v=j+1}^{j+p-1} \phi(v)e^{ivx} - \ldots - \alpha_{p-1}e^{i(p-1)x}\phi(j+1)e^{i(j+1)x}}{1 - \alpha_1 e^{ix} - \ldots - \alpha_p e^{ipx}}$$

where the $\alpha_j$ are as in Theorem 2.1. $\hat{F}_j(x)$ obviously, then, has a finite bias expansion, and using the notation of Theorem 3.1 we take

$$b_m(\theta) = \begin{cases} \dfrac{\alpha_{p-m}e^{i(p-m)x}}{1-\alpha_1 e^{ix} - \ldots - \alpha_p e^{ipx}}, & m = 1,\ldots,p-1 \\ \dfrac{-1}{1-\alpha_1 e^{ix} - \ldots - \alpha_p e^{ipx}}, & m = p \end{cases}$$

and $h_{mj}(n) = \sum_{v=q-p+j}^{q-p+j+m-1} \phi(v)e^{ivx}$, $m=1,\ldots,p$ and $j = 1,\ldots,p+1$.

By Theorem 3.1 it follows that

$$E[G^{**}(\hat{F}_{q-p}(x),\ldots,\hat{F}_{q}(x))] = \sum_{v=k}^{\infty} \phi(v)e^{ivx}$$

where

$$G^{**}(\hat{F}_{q-p}(x),\ldots,\hat{F}_{q}(x)) =$$

$$= \frac{\begin{vmatrix} \hat{F}_{q-p}(x) & \hat{F}_{q-p+1}(x) & \cdots & \hat{F}_{q}(x) \\ \phi(q-p+1)e^{i(q-p+1)x} & \phi(q-p+2)e^{i(q-p+2)x} & \cdots & \phi(q+1)e^{i(q+1)x} \\ \sum_{v=q-p+1}^{q-p+2} \phi(v)e^{ivx} & \sum_{v=q-p+2}^{q-p+3} \phi(v)e^{ivx} & \cdots & \sum_{v=q+1}^{q+2} \phi(v)e^{ivx} \\ \vdots & \vdots & & \vdots \\ \sum_{v=q-p+1}^{q} \phi(v)e^{ivx} & \sum_{v=q-p+2}^{q+1} \phi(v)e^{ivx} & \cdots & \sum_{v=q+1}^{q+p} \phi(v)e^{ivx} \end{vmatrix}}{\begin{vmatrix} 1 & 1 & \cdots & 1 \\ \phi(q-p+1)e^{i(q-p+1)x} & \phi(q-p+2)e^{i(q-p+2)x} & \cdots & \phi(q+1)e^{i(q+1)x} \\ \sum_{v=q-p+1}^{q-p+2} \phi(v)e^{ivx} & \sum_{v=q-p+2}^{q-p+3} \phi(v)e^{ivx} & & \sum_{v=q+1}^{q+2} \phi(v)e^{ivx} \\ \vdots & \vdots & & \vdots \\ \sum_{v=q-p+1}^{q} \phi(v)e^{ivx} & \sum_{v=q-p+2}^{q+1} \phi(v)e^{ivx} & & \sum_{v=q+1}^{q+p} \phi(v)e^{ivx} \end{vmatrix}}$$

$$= \frac{\left|A_{1,0}(x)\right|}{\left|A_{2,0}(x)\right|} \quad .$$

Let $A_{i,(j+1)}(x)$ be the matrix obtained by subtracting the (p-j)th row of $A_{i,j}(x)$ from the (p+1-j)th row of $A_{i,j}(x)$, i = 1,2 and j = 0,1,...,p-2. Then

$$\frac{\left|A_{1,(p-1)}(x)\right|}{\left|A_{2,(p-1)}(x)\right|} = G^*(\hat{F}_{q-p}(x),\ldots,\hat{F}_q(x)) \ ;$$

but, by a basic property of determinants, we also have

$$\frac{\left|A_{1,(p-1)}(x)\right|}{\left|A_{2,(p-1)}(x)\right|} = \frac{\left|A_{1,0}(x)\right|}{\left|A_{2,0}(x)\right|}$$

and therefore

$$G^*(\hat{F}_{q-p}(x),\ldots,\hat{F}_q(x)) = G^{**}(\hat{F}_{q-p}(x),\ldots,\hat{F}_q(x)).$$

As noted previously, $e_p(\hat{F}_q(x))$ estimates the random variable G*. Therefore, under the assumption that f(·) has an ARMA (p,q) representation, $e_p(\hat{F}_q(x))$ is seen to estimate a random variable G** which is constructed by the generalized jackknife scheme in such a way that

$$E[G^{**}] = \sum_{v=k}^{\infty} \phi(v)e^{ivx} \ .$$

Although it is not possible to obtain a simple expression for the bias of $e_p(\hat{F}_q(x))$, the following observations are possible. Suppose f(·) has an ARMA (p,q) representation. Then the bias of $e_p(\hat{F}_q(x))$ is a result of the error inherent in the estimation of $\phi(v)$, $|v| = 1,2,\ldots,p+q$. This source of bias may essentially be removed by taking a large enough sample size n. By contrast, the bias of $\hat{F}_j(x)$, a logical competitor of $e_p(\hat{F}_q(x))$, is $-\sum_{v=j+1}^{\infty} \phi(v)e^{ivx}$

regardless of the sample size. These observations provide a motivation for considering ARMA estimators as a possible alternative to Fourier series estimators.

To this point we have considered only the bias of $e_p(\hat{F}_q(x))$. In concluding this section we note that the bias of $\hat{f}_{p,q}(x)$ depends only upon $\text{Bias}[e_p(\hat{F}_q(x))]$. We have

$$\begin{aligned}
\text{Bias}[\hat{f}_{p,q}(x)] &= E[\hat{f}_{p,q}(x)] - f(x) \\
&= \frac{1}{2\pi}[1+2\text{Real}\{E(e_p(\hat{F}_q(x))) - E(\hat{F}_0(x))\}] \\
&\quad - \frac{1}{2\pi}[1+2\text{Real}(\sum_{v=k}^{\infty} \phi(v)e^{ivx} - F_0(x))] \\
&= \frac{1}{\pi} \text{Real}\{E(e_p(\hat{F}_q(x))) - F_0(x)\} \\
&\quad - \frac{1}{\pi} \text{Real}\{\sum_{v=k}^{\infty} \phi(v)e^{ivx} - F_0(x)\} \\
&= \frac{1}{\pi} \text{Real}\{E(e_p(\hat{F}_q(x))) - \sum_{v=k}^{\infty} \phi(v)e^{ivx}\} \\
&= \frac{1}{\pi} \text{Real}\{\text{Bias}[e_p(\hat{F}_q(x))]\} \quad .
\end{aligned}$$

## 3.4 Alternative Ways of Expressing $\hat{f}_{p,q}(\cdot)$

In this section some different ways of expressing $\hat{f}_{p,q}(\cdot)$ are derived which will be useful in later chapters and also show explicitly how ARMA estimators are related to Fourier series estimators. The basic result involves using the alternative form of expressing $e_n(A_m)$ referred to in the previous section. This result is stated in the following theorem.

Theorem 3.2

$$e_p(F_q(x)) = \frac{\hat{F}_q(x) - \hat{\alpha}_1 e^{ix}\hat{F}_{q-1}(x) - \ldots - \hat{\alpha}_p e^{ipx}\hat{F}_{q-p}(x)}{1 - \hat{\alpha}_1 e^{ix} - \ldots - \hat{\alpha}_p e^{ipx}}$$

where $(\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_p)$ is the solution of the system of equations

$$\begin{bmatrix} \hat{\phi}(q) & \hat{\phi}(q-1) & \ldots & \hat{\phi}(q-p+1) \\ \hat{\phi}(q+1) & \hat{\phi}(q) & \ldots & \hat{\phi}(q-p+2) \\ \vdots & \vdots & & \vdots \\ \hat{\phi}(q+p-1) & \hat{\phi}(q+p-2) & \ldots & \hat{\phi}(q) \end{bmatrix} \underline{y} = \begin{bmatrix} \hat{\phi}(q+1) \\ \hat{\phi}(q+2) \\ \vdots \\ \hat{\phi}(q+p) \end{bmatrix}. \qquad (3.4)$$

Proof:

$$e_p(\hat{F}_q(x)) = \frac{c_{q-p}(x)\hat{F}_{q-p}(x) + c_{q-p+1}(x)\hat{F}_{q-p+1}(x) + \ldots + c_q(x)\hat{F}_q(x)}{c_{q-p}(x) + c_{q-p+1}(x) + \ldots + c_q(x)}$$

where the $c_{q-j}(x)$ are cofactors of the first row in either the numerator or denominator determinant of $e_p(\hat{F}_q(x))$. By performing approriate row and column operations within these cofactors it is easily verified that

$$e_p(\hat{F}_q(x)) = \frac{a_0\hat{F}_q(x) - a_1 e^{ix}\hat{F}_{q-1}(x) - \ldots - a_p e^{ipx}\hat{F}_{q-p}(x)}{a_0 - a_1 e^{ix} - \ldots - a_p e^{ipx}} \qquad (3.5)$$

where

$$a_j = \begin{vmatrix} \hat{\phi}(q) & \hat{\phi}(q-1) & \ldots & \hat{\phi}(q-j+2) & \hat{\phi}(q+1) & \hat{\phi}(q-j) & \ldots & \hat{\phi}(q-p+1) \\ \hat{\phi}(q+1) & \hat{\phi}(q) & \ldots & \hat{\phi}(q-j+3) & \hat{\phi}(q+2) & \hat{\phi}(q-j+1) & \ldots & \hat{\phi}(q-p+2) \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \hat{\phi}(q+p-1) & \hat{\phi}(q+p-2) & \ldots & \hat{\phi}(q+p-j+1) & \hat{\phi}(q+p) & \hat{\phi}(q+p-j-1) & \ldots & \hat{\phi}(q) \end{vmatrix},$$

$j = 1,2,\ldots,p$ and

$$a_0 = \begin{vmatrix} \hat{\phi}(q) & \hat{\phi}(q-1) & \ldots & \hat{\phi}(q-p+1) \\ \hat{\phi}(q+1) & \hat{\phi}(q) & \ldots & \hat{\phi}(q-p+2) \\ \vdots & \vdots & & \vdots \\ \hat{\phi}(q+p-1) & \hat{\phi}(q+p-2) & \ldots & \hat{\phi}(q) \end{vmatrix} .$$

It follows that $\left( \frac{a_1}{a_0}, \frac{a_2}{a_0}, \ldots, \frac{a_p}{a_0} \right)$ is the Cramer's rule solution to the system in (3.4). By dividing numerator and denominator of (3.5) by $a_0$ the result follows.

By the previous theorem $\hat{f}_{p,q}(x)$ may be expressed as

$$\hat{f}_{p,q}(x) = \frac{1}{2\pi} \left[1 + 2\text{Real}\left( \frac{\hat{F}_q(x) - \hat{\alpha}_1 e^{ix} \hat{F}_{q-1}(x) - \ldots - \hat{\alpha}_p e^{ipx} \hat{F}_{q-p}(x)}{1 - \hat{\alpha}_1 e^{ix} - \ldots - \hat{\alpha}_p e^{ipx}} - \hat{F}_0(x) \right) \right].$$

The results of Chapter II, show that, if $\hat{\alpha}_1, \ldots, \hat{\alpha}_p$ satisfy condition S, $\hat{f}_{p,q}(\cdot)$ satisfies

$$\hat{\phi}_{p,q}(v) = \hat{\phi}(v) , \quad |v| = 0,1,\ldots,p+q$$

where

$$\hat{\phi}_{p,q}(v) = \int_{-\pi}^{\pi} e^{-ivx} \hat{f}_{p,q}(x)\, dx.$$

Using this fact and the result of Theorem 3.2, it is informative to re-express $\hat{f}_{p,q}(x)$ as

$$\hat{f}_{p,q}(x) = \frac{1}{2\pi} \left[1 + 2\text{Real}\left\{ e_p(\hat{F}_q(x)) - \hat{F}_0(x) + \sum_{v=1}^{p+q} \hat{\phi}(v) e^{ivx} - \sum_{v=1}^{p+q} \hat{\phi}(v) e^{ivx} \right\} \right]$$

$$= \frac{1}{2\pi} \sum_{v=-p-q}^{p+q} \hat{\phi}(v) e^{ivx} + \frac{1}{\pi} \text{Real}\left( e_p(\hat{F}_q(x)) - \hat{F}_{p+q}(x) \right)$$

$$= \hat{f}_{0,p+q}(x)$$

$$- \frac{1}{\pi}\text{Real}\left[\frac{\sum_{v=q+1}^{p+q}\hat{\phi}(v)e^{ivx} - \hat{\alpha}_1 e^{ix}\sum_{v=q}^{p+q}\hat{\phi}(v)e^{ivx} - \ldots - \hat{\alpha}_p e^{ipx}\sum_{v=q-p+1}^{p+q}\hat{\phi}(v)e^{ivx}}{1 - \hat{\alpha}_1 e^{ix} - \ldots - \hat{\alpha}_p e^{ipx}}\right]$$

$$= \hat{f}_{0,p+q}(x)$$

$$+ \frac{1}{\pi}\text{Real}\left[\frac{\hat{\alpha}_1 e^{ix}\hat{\phi}(p+q)e^{i(p+q)x} + \hat{\alpha}_2 e^{2ix}\sum_{v=p+q-1}^{p+q}\hat{\phi}(v)e^{ivx} + \ldots + \hat{\alpha}_p e^{ipx}\sum_{v=q+1}^{p+q}\hat{\phi}(v)e^{ivx}}{1 - \hat{\alpha}_1 e^{ix} - \ldots - \hat{\alpha}_p e^{ipx}}\right]$$

$$= \hat{f}_{0,p+q}(x) + \hat{g}_{p,q}(x) \quad . \tag{3.6}$$

Expression (3.6) shows $\hat{f}_{p,q}(\cdot)$ to be the sum of a Fourier series estimator $\hat{f}_{0,p+q}(\cdot)$ and a function $\hat{g}_{p,q}(\cdot)$ which, under condition S, has the Fourier series expansion

$$\frac{1}{2\pi}\sum_{|v|>p+q}\hat{\phi}_{p,q}(v)e^{ivx} \quad ,$$

where the $\hat{\phi}_{p,q}(v)$ are extrapolated from $\hat{\phi}(v)$, $|v| = 0,1,\ldots,p+q$, using the difference equation $y(v) - \hat{\alpha}_1 y(v-1) - \ldots - \hat{\alpha}_p y(v-p) = 0$. We note, though, that (3.6) is valid regardless of whether or not condition S holds, although $\hat{g}_{p,q}(\cdot)$ does not have the same interpretation in this case. The validity of (3.6) will be useful in Chapter V when we consider estimating the MISE of $\hat{f}_{p,q}(\cdot)$.

A simple example which illustrates the consequences of (3.6) will be helpful at this point. Consider the ARMA estimator $\hat{f}_{1,q}(\cdot)$. By (3.6), we have

$$\hat{f}_{1,q}(x) = \hat{f}_{0,q+1}(x) + \frac{1}{\pi}\text{Real}\left[\frac{\hat{\alpha}_1 e^{ix}\hat{\phi}(q+1)e^{i(q+1)x}}{1 - \hat{\alpha}_1 e^{ix}}\right] \quad ,$$

$$\hat{\alpha}_1 = \frac{\hat{\phi}(q+1)}{\hat{\phi}(q)} \quad .$$

If $|\hat{\alpha}_1| < 1$, $\frac{1}{1-\hat{\alpha}_1 e^{ix}} = \sum_{v=0}^{\infty} \hat{\alpha}_1^v e^{ivx}$ and thus

$$\frac{\hat{\alpha}_1 e^{ix} \hat{\phi}(q+1) e^{i(q+1)x}}{1 - \hat{\alpha}_1 e^{ix}} = \sum_{v=0}^{\infty} [\hat{\alpha}_1 \hat{\phi}(q+1)] \hat{\alpha}_1^v e^{i(v+q+2)x}$$

$$= \sum_{v=q+2}^{\infty} [\hat{\alpha}_1 \hat{\phi}(q+1)] \hat{\alpha}_1^{v-q-2} e^{ivx} \quad .$$

Therefore,

$$\hat{f}_{1,q}(x) = \hat{f}_{0,q+1}(x) + \frac{1}{2\pi} \sum_{|v|>q+1} \hat{\phi}_{1,q}(v) e^{ivx}$$

where

$$\hat{\phi}_{1,q}(v) = [\hat{\alpha}_1 \hat{\phi}(q+1)] \, \hat{\alpha}_1^{v-q-2} \quad , \quad v = q+2, \, q+3, \ldots \, . \tag{3.7}$$

Obviously $\hat{\phi}_{1,q}(v) - \hat{\alpha}_1 \hat{\phi}_{1,q}(v-1) = 0$ for $v > q + 2$, but we also have (by (3.7))

$$\hat{\phi}_{1,q}(q+2) = \hat{\alpha}_1 \hat{\phi}(q+1) \implies \hat{\phi}_{1,q}(q+2) - \hat{\alpha}_1 \hat{\phi}(q+1) = 0 \, .$$

This shows explicity how the $\hat{\phi}_{1,q}(v)$, $v = q+2, q+3, \ldots$, are extrapolated from $\hat{\phi}(q)$ and $\hat{\phi}(q+1)$ by using $y(v) - \hat{\alpha}_1 y(v-1) = 0$.

Suppose now that in the above case condition S is not satisfied, i.e. suppose $|\hat{\alpha}_1| > 1$. We then have

$$\frac{\hat{\alpha}_1 \hat{\phi}(q+1) e^{i(q+2)x}}{1 - \hat{\alpha}_1 e^{ix}} = \frac{-\hat{\phi}(q+1) e^{i(q+1)x}}{1 - \hat{\alpha}_1^{-1} e^{-ix}}$$

$$= \sum_{v=0}^{\infty} -\hat{\phi}(q+1) \hat{\alpha}_1^{-v} e^{i(q+1-v)x}$$

$$= \sum_{v=0}^{q+1} -\hat{\phi}(q+1)\hat{\alpha}_1^{v-q-1} e^{ivx}$$

$$+ \sum_{v=-\infty}^{-1} -\hat{\phi}(q+1)\hat{\alpha}_1^{v-q-1} e^{ivx}.$$

This implies that

$$\frac{1}{\pi}\text{Real}\left[\frac{\hat{\alpha}_1\hat{\phi}(q+1)e^{i(q+2)x}}{1-\hat{\alpha}_1 e^{ix}}\right] = \frac{1}{\pi}\text{Real}\left\{\sum_{v=0}^{q+1} -\hat{\phi}(q+1)\hat{\alpha}_1^{v-q-1} e^{ivx}\right.$$

$$\left. + \sum_{v=1}^{\infty} -\overline{\hat{\phi}(q+1)}\,\overline{\hat{\alpha}}_1^{-v-q-1} e^{ivx}\right\}.$$

Using this expression and (3.6) it follows that

$$\hat{f}_{1,q}(x) = \frac{1}{2\pi}[\hat{\phi}_{1,q}(0) + 2\text{Real}(\sum_{v=1}^{\infty}\hat{\phi}_{1,q}(v))]$$

where

$$\hat{\phi}_{1,q}(0) = 1 - 2\text{Real}[\hat{\phi}(q+1)\hat{\alpha}_1^{-q-1}]$$

and (3.8)

$$\hat{\phi}_{1,q}(v) = \begin{cases} \hat{\phi}(v) - \hat{\phi}(q+1)\hat{\alpha}_1^{v-q-1} - \overline{\hat{\phi}(q+1)}\,\overline{\hat{\alpha}}_1^{-v-q-1}, & v = 1,\ldots,q+1 \\ -\overline{\hat{\phi}(q+1)}\,\overline{\hat{\alpha}}_1^{-v-q-1}, & v = q+2, q+3,\ldots . \end{cases}$$

It is easily verified that $\hat{\phi}_{1,q}(v) - \overline{\hat{\alpha}}_1^{-1}\hat{\phi}_{1,q}(v-1) = 0$ for $v > q$. However, by (3.8), $\hat{f}_{1,q}(\cdot)$ does not integrate to 1 and does not satisfy

$$\hat{\phi}_{1,q}(v) = \hat{\phi}(v), \quad |v| = 1,\ldots,q+1.$$

Therefore, $\hat{f}_{1,q}(\cdot)$ is not as easily interpreted in this case as it is when condition S is satisfied. It should be pointed out, though, that the efficacy of $\hat{f}_{p,q}(\cdot)$ as an estimate of $f(\cdot)$ may be assessed

regardless of whether or not condition S is satisfied, as will be shown in Chapter V.

## 3.5 The Mixture of Densities Having ARMA Representations

As pointed out in Chapter I, Carmichael approaches the density estimation problem by using autoregressive schemes to represent the density $f(\cdot)$. Under fairly mild smoothness conditions on $f(\cdot)$ Carmichael shows that

$$\lim_{p\to\infty} f_{p,0}(x) = f(x), \text{ uniformly in } x, \tag{3.9}$$

which implies the existence of a $p_0$ (for $\varepsilon$ arbitrarily small) such that $|f(x) - f_{p_0,0}(x)| < \varepsilon$, a.e. $[-\pi,\pi]$. This result provides a justification for using autoregressive representations in the estimation of probability densities. However, it also leads indirectly to a justification for considering ARMA representations. In order to show why this is so we state and prove the following theorem.

Theorem 3.3 Let $f_{p_j,q_j}(\cdot)$ $(j = 1,2)$ be a probability density function (defined on $[-\pi,\pi]$) having the ARMA $(p_j,q_j)$ representation

$$\frac{\sum_{v=-q_j}^{q_j} \beta_{jv} e^{ivx}}{|1-\alpha_{j1}e^{ix}-\ldots-\alpha_{jp_j}e^{ip_jx}|^2},$$

and let $0 < \gamma < 1$.

Then the mixture density $\gamma f_{p_1,q_1}(\cdot) + (1-\gamma) f_{p_2,q_2}(\cdot)$ has an ARMA$(p_1 + p_2, k)$ representation, where $k \leq \max(q_1+p_2, q_2+p_1)$.

Proof: Let $|1-\alpha_{j1}e^{ix} - \ldots - \alpha_{jp_j}e^{ip_jx}|^2 = a_j(x)$, $j = 1,2$.

Then,

$$\gamma f_{p_1,q_1}(x) + (1-\gamma)f_{p_2,q_2}(x) = \frac{\gamma \sum_{v=-q_1}^{q_1} \beta_{1v}e^{ivx}}{a_1(x)} + \frac{(1-\gamma)\sum_{v=-q_2}^{q_2}\beta_{2v}e^{ivx}}{a_2(x)}$$

$$= \frac{\gamma a_2(x) \sum_{v=-q_1}^{q_1} \beta_{1v}e^{ivx} + (1-\gamma)a_1(x) \sum_{v=-q_2}^{q_2} \beta_{2v}e^{ivx}}{a_1(x)a_2(x)}$$

The denominator is obviously of the form

$$|1-\alpha_1 e^{ix} - \ldots - \alpha_{p_1+p_2}e^{i(p_1+p_2)x}|^2 .$$

In addition, since $a_j(x)$ may be expressed as $\sum_{v=-p_j}^{p_j} b_{jv}e^{ivx}$, the numerator is of the form $\sum_{v=-k}^{k} \beta_v e^{ivx}$ where k does not exceed $\max(q_1+p_2, q_2+p_1)$. The result thus follows.

By induction a similar result follows for the mixture of m ARMA densities ($m \geq 3$).

A special case of Theorem 3.3 which is of interest is the mixture of autoregressive densities. The mixture of $f_{p_1,0}(\cdot)$ and $f_{p_2,0}(\cdot)$ is, by Theorem 3.3, an ARMA $(p_1 + p_2, k)$ density where $k \leq \max(p_1,p_2)$ and, in general, $k > 0$. Carmichael's result, (3.9), and Theorem 3.3 are thus seen to provide a strong motivation for

ARMA representations in situations where $f(\cdot)$ arises as the mixture of densities. Of course (3.9) shows that even the mixture of densities may be well approximated by an autoregressive scheme. However, the ARMA $(p_1+p_2,k)$ representation will necessarily be more parsimonious than a satisfactory autoregressive representation. This is important in the stochastic setting where fitting too many parameters is to be avoided.

# CHAPTER IV

# LARGE SAMPLE PROPERTIES OF ARMA DENSITY ESTIMATORS

## 4.1 Introduction

We continue our study of ARMA density estimators by establishing some of their large sample properties. In Theorem 4.1 conditions are stated under which $\hat{f}_{p,q}(\cdot)$ converges in probability to $f(\cdot)$, where p remains fixed and q tends to infinity at a specified rate with the sample size n. The results of Section 4.3 are the stochastic analogs of some of the work of McWilliams (1969) involving the $e_n$-transform. Sufficient conditions, which are more informative than those in Theorem 4.1, are established for the convergence in probability of $\hat{f}_{1,q}(\cdot)$ to $f(\cdot)$ (as q and n tend to infinity). More importantly, $\hat{f}_{1,q}(\cdot)$ is shown to possess a certain optimality property for densities satisfying

$$\lim_{v\to\infty} \frac{\phi(v+1)}{\phi(v)} = R.$$

Finally, we point out that higher order ($p > 2$) results paralleling those for $\hat{f}_{1,q}(\cdot)$ are undoubtedly obtainable.

## 4.2 Conditions for the Consistency of $\hat{f}_{p,q}(\cdot)$

A minimal requirement of any density estimator is its convergency (in some sense) to the true density function as the sample size tends to infinity. Considerable attention in the literature has been focused upon establishing some form of consistency for

various types of density estimators. The mean square error consistency of Fourier series estimators has already been indicated in Chapter I. Parzen (1962) has proven that for suitably chosen weighting functions $K(\cdot)$, the kernel density estimator

$$f_n(x) = \frac{1}{nh}\sum_{j=1}^{n} K\left(\frac{x-X_j}{h}\right)$$

is mean square error consistent for $f(x)$ if $h = h(n)$ satisfies $\lim_{n\to\infty} h(n)=0$ and $\lim_{n\to\infty} nh(n) = \infty$. Other estimators and different convergence criteria have also been considered (see Tapia and Thompson (1978)).

In the following theorem conditions are stated under which $\hat{f}_{p,q}(\cdot)$ converges in probability to $f(\cdot)$.

Theorem 4.1 Suppose $f(\cdot)$ is a density defined on $[-\pi,\pi]$ which is continuous and of bounded variation on that interval. Based on the random sample $X_1,\ldots,X_n$ from $f(\cdot)$, let $\hat{\alpha}_j^{(p,q)}$ $(j=1,2,\ldots,p)$ be the solution of the system of equations in (3.4). For a fixed $p \geq 1$ and $q(n) = o(\sqrt{n})$ (where $\lim_{n\to\infty} q(n) = \infty$), suppose

$$\operatorname*{p-lim}_{n\to\infty} \frac{|\hat{\alpha}_j^{(p,q)}| \sum_{v=p+q-j+1}^{p+q} |\hat{\phi}(v)|}{z_{p,q}} = 0, \quad j=1,\ldots,p, \qquad (4.1)$$

where $z_{p,q} = \min_{x\in[-\pi,\pi]} |1-\hat{\alpha}_1^{(p,q)}e^{ix}-\ldots-\hat{\alpha}_p^{(p,q)}e^{ipx}|$. Then

$$\operatorname*{p-lim}_{n\to\infty} \hat{f}_{p,q}(x) = f(x) \text{ for all } x \in [-\pi,\pi] .$$

Proof: Using the relationship established in Chapter III we have (assuming the existence of each limit)

$$\operatorname*{p-lim}_{n\to\infty} \hat{f}_{p,q}(x) = \operatorname*{p-lim}_{n\to\infty} \hat{f}_{0,p+q}(x) + \operatorname*{p-lim}_{n\to\infty} \hat{g}_{p,q}(x) .$$

Since $f(\cdot)$ is continuous and of bounded variation on $[-\pi,\pi]$, $f(x)$ has a Fourier series representation, and thus

$$\text{Bias}[\hat{f}_{0,p+q}(x)] = \frac{-1}{2\pi} \sum_{|v| > p+q} \phi(v) e^{ivx} , \tag{4.2}$$

which tends to zero as $q \to \infty$. We have also

$$\begin{aligned}
\text{var}[\hat{f}_{0,p+q}(x)] &= \frac{1}{\pi^2} \text{var}[\text{Real}(\sum_{v=1}^{p+q} \hat{\phi}(v) e^{ivx})] \\
&\leq \frac{1}{\pi^2} \text{var}[\sum_{v=1}^{p+q} \hat{\phi}(v) e^{ivx}] = \frac{1}{\pi^2} \sum_{v=1}^{p+q} \text{var}(\hat{\phi}(v)) \\
&\quad + \frac{1}{\pi^2} \sum_{v=1}^{p+q} \sum_{k \neq v} \text{cov}(\hat{\phi}(v), \hat{\phi}(k)) e^{i(v-k)x} \\
&= \frac{1}{\pi^2} \frac{1}{n} \sum_{v=1}^{p+q} (1 - |\phi(v)|^2) \\
&\quad + \frac{1}{\pi^2} \frac{1}{n} \sum_{v=1}^{p+q} \sum_{k \neq v} (\phi(v-k) - \phi(v)\phi(-k)) e^{i(v-k)x} \\
&\leq \frac{1}{\pi^2} \left[ \frac{p+q}{n} + \frac{2(p+q)(p+q-1)}{n} \right] .
\end{aligned}$$

If $p$ is fixed and $q = o(\sqrt{n})$ it thus follows that

$$\lim_{n \to \infty} \text{var}[\hat{f}_{0,p+q}(x)] = 0. \tag{4.3}$$

By (4.2) and (4.3) we have (for $q = o(\sqrt{n})$ and unbounded)

$$\lim_{n \to \infty} E[\hat{f}_{0,p+q}(x) - f(x)]^2 = 0,$$

and thus

$$\text{p-lim}_{n \to \infty} \hat{f}_{0,p+q}(x) = f(x).$$

We must now show $\text{p-lim}_{n \to \infty} \hat{g}_{p,q}(x) = 0$ in order to prove the theorem.

Recall that

$$\hat{g}_{p,q}(x) = \frac{1}{\pi}\text{Real}\left[\sum_{j=1}^{p} \beta_j^{(p,q)}(x)\right] ,$$

where

$$\beta_j^{(p,q)}(x) = \frac{\hat{\alpha}_j^{(p,q)} e^{ijx} \sum_{v=p+q-j+1}^{p+q} \hat{\phi}(v) e^{ivx}}{1-\hat{\alpha}_1^{(p,q)} e^{ix} - \ldots - \hat{\alpha}_p^{(p,q)} e^{ipx}} .$$

Since

$$\frac{\left|\hat{\alpha}_j^{(p,q)}\right| \left| \sum_{v=p+q-j+1}^{p+q} \hat{\phi}(v) \right|}{z_{p,q}}$$

bounds $\left|\beta_j^{(p,q)}(x)\right|$, it follows (from condition (4.1)) that $\underset{n\to\infty}{\text{p-lim}}\ \beta_j^{(p,q)}(x) = 0$. Therefore, $\underset{n\to\infty}{\text{p-lim}} \sum_{j=1}^{p} \beta_j^{(p,q)}(x) = 0$ and consequently $\underset{n\to\infty}{\text{p-lim}}\ \hat{g}_{p,q}(x) = 0$. Since x was chosen arbitrarily the result follows.

Several comments are in order regarding Theorem 4.1. First, it should be pointed out that, since the method of estimating $\phi(v)$ is fixed, condition (4.1) is implicitly a condition on the Fourier coefficients of $f(\cdot)$. In this work, however, the problem of translating (4.1) into explicit conditions on the sequence $\{\phi(v)\}$ has not been solved. It is hoped that a satisfactory solution to this problem may be obtained after future research. For the present, though, we note that the importance of Theorem 4.1 lies in the fact that it points out where the difficulty rests in inducing convergence from $\hat{f}_{p,q}(\cdot)$. Since $\hat{f}_{0,p+q}(x)$ is consistent for $f(x)$, it is clear that conditions need only be established to insure that $\underset{n\to\infty}{\text{p-lim}}\ \hat{g}_{p,q}(x) = 0$.

Although we will not be able to substitute conditions for (4.1) which are as explicit as desired, the following observations make (4.1) more palatable. We have

$$\sum_{v=p+q-j+1}^{p+q} |\hat{\phi}(v)| = \sum_{v=p+q-j+1}^{p+q} |\hat{\phi}(v) - \phi(v) + \phi(v)|$$

$$\leq \sum_{v=p+q-j+1}^{p+q} |\hat{\phi}(v)-\phi(v)| + \sum_{v=p+q-j+1}^{p+q} |\phi(v)| ,$$

which converges in probability to zero (as $n \to \infty$) by an argument similar to that in Theorem 4.1. In addition, it is easily verified that

$$\sum_{v=p+q-j+1}^{p+q} |\hat{\phi}(v)-\phi(v)| = o_p\left(\frac{1}{\sqrt{n}}\right) ,$$

and thus a set of conditions which may replace condition (4.1) is:

(i) $$\frac{|\hat{\alpha}_j(p,q)|}{z_{p,q}} = o_p(\sqrt{n}) , \quad (j = 1,2,\ldots,p)$$

(ii) $$\sum_{v=q+1}^{p+q} |\phi(v)| = o\left(\frac{1}{\sqrt{n}}\right) .$$

Although it is still not precisely clear for which densities (i) and (ii) are valid, this set of conditions is somewhat more informative than condition (4.1).

It is important to note at this point that one member of the ARMA class, $\hat{f}_{0,q}(\cdot)$, is mean square error consistent for $f(\cdot)$ under the single condition that $f(\cdot)$ have a Fourier series representation for all $x \in [-\pi,\pi]$. This fact was proven in Theorem 4.1. Because of the consistency of $\hat{f}_{0,q}(\cdot)$, Theorem 4.1 would not be extremely

important unless it could be shown that $\hat{f}_{p,q}(\cdot)$ (for $p \geq 1$) in some sense converges more rapidly to $f(\cdot)$ than does $\hat{f}_{0,q}(\cdot)$. Establishing conditions which insure the more rapid convergence of $\hat{f}_{p,q}(\cdot)$ proves to be quite difficult in general. However, in the next section we consider the special case of $\hat{f}_{1,q}(\cdot)$ and obtain some quite satisfying results.

## 4.3 Large Sample Results Involving $\hat{f}_{1,q}(\cdot)$

Given a complex-valued sequence of partial sums $\{A_k, A_{k+1}, \ldots\}$ which converges to $A_\infty$, McWilliams (1969) has established conditions under which the following results hold:

$$\lim_{m\to\infty} e_n(A_m) = A_\infty \quad (n=1,2), \quad \lim_{m\to\infty} \frac{A_\infty - e_1(A_m)}{A_\infty - A_{m+1}} = 0,$$

and

$$\lim_{m\to\infty} \frac{A_\infty - e_2(A_{m+2})}{A_\infty - A_{m+j}} = 0 \quad \text{(for any } j\text{)}.$$

The theorems in the present section involve $e_1(\hat{F}_q(x))$ and are the stochastic analogs of the above results involving $e_1(A_m)$. Essentially the same results are obtained with lim replaced by p-lim.

A sufficient condition for $\lim_{m\to\infty} \frac{A_\infty - e_1(A_m)}{A_\infty - A_{m+1}} = 0$ is

$$\lim_{m\to\infty} \frac{a_m}{a_{m-1}} = R, \quad 0 < |R| < 1 \tag{4.4}$$

where $a_m = A_m - A_{m-1}$. Recalling Theorem 2.2 it is seen that, if

$$\frac{a_m}{a_{m-1}} = R \quad \text{for } m > m_0,$$

then

$$\frac{A_\infty - e_1(A_m)}{A_\infty - A_{m+1}} = 0 \text{ for } m \geq m_0 .$$

Condition (4.4) is thus seen to be a relaxing of the condition needed for $e_1(A_m)$ to be exact, with the result being that $e_1(A_m)$ converges more rapidly than $A_{m+1}$.

In the setting of interest here we have $a_m = \phi(m)e^{imx}$ and $\frac{a_m}{a_{m-1}} = \frac{\phi(m)}{\phi(m-1)} e^{ix}$. In this case, then, a condition equivalent to (4.4) is

$$\lim_{m\to\infty} \frac{\phi(m)}{\phi(m-1)} = R , \quad 0 < |R| < 1 . \tag{4.5}$$

If (4.5) holds we have

$$\lim_{m\to\infty} \frac{F_\infty(x) - e_1(F_m(x))}{F_\infty(x) - F_{m+1}(x)} = 0 ,$$

where $F_\infty(x) = \sum_{v=1}^{\infty} \phi(v)e^{ivx}$. This property suggests that, under (4.5), $e_1(\hat{F}_m(x))$ might converge more rapidly in some stochastic sense than does $\hat{F}_{m+1}(x)$. This possibility will be investigated later, but first it is necessary to verify that $e_1(\hat{F}_m(x))$ does indeed converge to $F_\infty(x)$ under a condition similar to (4.5). The verification of this fact is the subject of the next theorem.

Theorem 4.2 Suppose $f(\cdot)$ is a probability density function defined on $[-\pi,\pi]$ and that $f(\cdot)$ has a Fourier series representation for all $x$ in that interval. Further, suppose that (4.5) holds and that $\frac{R^q}{\phi(q)} = 0(1)$ (as $n \to \infty$). If $\hat{f}_{1,q}(\cdot)$ is based on a

random sample $X_1,\ldots,X_n$ from $f(\cdot)$, and $q = o(\ell n\ n)$ (with $q$ unbounded), then

$$\operatorname*{p-lim}_{n\to\infty} \hat{f}_{1,q}(x) = f(x) \quad \text{for all } x \in [-\pi,\pi] .$$

Proof: Since $\hat{f}_{1,q}(x) = \frac{1}{2\pi}[1 + 2\text{Real}(e_1(\hat{F}_q(x)))]$,

$$f(x) = \frac{1}{2\pi}[1 + 2\text{Real}(F_\infty(x))],$$

and

$$\operatorname*{p-lim}_{n\to\infty} \text{Real}\,(Z_n) = \text{Real}[\operatorname*{p-lim}_{n\to\infty} Z_n] \quad ,$$

it is sufficient to show that

$$\operatorname*{p-lim}_{n\to\infty} e_1(\hat{F}_q(x)) = F_\infty(x).$$

Observe that

$$e_1(\hat{F}_q(x)) = \frac{\hat{F}_q(x) - \hat{\alpha}_{(q)}e^{ix}\hat{F}_{q-1}(x)}{1 - \hat{\alpha}_{(q)}e^{ix}} \quad , \quad (\hat{\alpha}_{(q)} = \frac{\hat{\phi}(q+1)}{\hat{\phi}(q)})$$

$$= \frac{\hat{\phi}(q)e^{iqx} + \hat{F}_{q-1}(x)[1-\hat{\alpha}_{(q)}e^{ix}]}{1 - \hat{\alpha}_{(q)}e^{ix}}$$

$$= \hat{F}_{q-1}(x) + \frac{\hat{\phi}(q)e^{iqx}}{1-\hat{\alpha}_{(q)}e^{ix}} \quad .$$

Since $q = o(\ell n\ n)$ it follows from the proof of Theorem 4.1 that $\operatorname*{p-lim}_{n\to\infty} \hat{F}_{q-1}(x) = F_\infty(x)$. Thus, if

$$\operatorname*{p-lim}_{n\to\infty} \frac{\hat{\phi}(q)e^{iqx}}{1-\hat{\alpha}_{(q)}e^{ix}} = 0$$

the result is proven. We have

$$\operatorname*{p-lim}_{n\to\infty} \hat{\phi}(q) = \lim_{n\to\infty} \phi(q) + \operatorname*{p-lim}_{n\to\infty}(\hat{\phi}(q) - \phi(q))$$

$$= \operatorname*{p-lim}_{n\to\infty} (\hat{\phi}(q) - \phi(q)).$$

Since $E[\hat{\phi}(q) - \phi(q)] = 0$ for each q, and $\operatorname{var}[\hat{\phi}(q) - \phi(q)] = \frac{1}{n}(1-|\phi(q)|^2) \to 0$ as $n \to \infty$, it follows that $\operatorname*{p-lim}_{n\to\infty}(\hat{\phi}(q) - \phi(q)) = 0$ and consequently $\operatorname*{p-lim}_{n\to\infty} \hat{\phi}(q) = 0$. As $|\hat{\phi}(q)e^{iqx}| = |\hat{\phi}(q)|$, we also have

$$\operatorname*{p-lim}_{n\to\infty}|\hat{\phi}(q)e^{iqx}| = 0. \tag{4.6}$$

Now consider

$$\operatorname*{p-lim}_{n\to\infty} \hat{\alpha}_{(q)} = \operatorname*{p-lim}_{n\to\infty} \frac{\frac{\phi(q+1)}{\phi(q)} + \frac{1}{\phi(q)}(\hat{\phi}(q+1)-\phi(q+1))}{1 + \frac{1}{\phi(q)}(\hat{\phi}(q) - \phi(q))}.$$

By the above $\hat{\phi}(q+j) - \phi(q+j) = 0_p(\frac{1}{\sqrt{n}})$ $(j = 0,1)$, and thus if $\lim_{n\to\infty} \frac{1}{\phi(q)} \cdot \frac{1}{\sqrt{n}} = 0$ we have $\operatorname*{p-lim}_{n\to\infty} \hat{\alpha}_{(q)} = R$. Now,

$$\lim_{n\to\infty} \frac{1}{\phi(q)} \cdot \frac{1}{\sqrt{n}} = \lim_{n\to\infty} \frac{R^q}{\phi(q)} \cdot \frac{1}{R^q\sqrt{n}}.$$

By hypothesis $\left| \frac{R^q}{\phi(q)} \right|$ is bounded, and so it is sufficient to show

$$\lim_{n\to\infty} \left|\frac{1}{R^q\sqrt{n}}\right| = \lim_{n\to\infty} \frac{1}{r^q\sqrt{n}} = 0 \quad (r = |R|).$$

Now, $\ell n(r^q\sqrt{n}) = q\,\ell n\, r + \frac{1}{2}\,\ell n\, n = \ell n\, n \{\frac{q\,\ell n\, r}{\ell n\, n} + \frac{1}{2}\} \to \infty$ as $n \to \infty$ since $q = o(\ell n\, n)$. Since $\ell n(r^q\sqrt{n}) \to \infty$ we have $r^q\sqrt{n} \to \infty$, and thus $\lim_{n\to\infty} \frac{1}{r^q\sqrt{n}} = 0$. As stated above this implies that $\operatorname*{p-lim}_{n\to\infty} \hat{\alpha}_{(q)} = R$ and consequently $\operatorname*{p-lim}_{n\to\infty} 1-\hat{\alpha}_{(q)}e^{ix} = 1-Re^{ix} \neq 0$. Along with (4.6), it then follows that

$$\operatorname*{p-lim}_{n\to\infty} \frac{\hat{\phi}(q)e^{iqx}}{1-\hat{\alpha}_{(q)}e^{ix}} = 0, \quad \text{and the proof is complete.}$$

The assumption that q tends to $\infty$ at a rate slower than

$\ell n$ n is undoubtedly more severe than is needed to induce $\hat{f}_{1,q}(\cdot)$ to converge. From the above proof it is clear that if

$$\underset{n\to\infty}{\text{p-lim}}\ (1-\hat{\alpha}_{(q)}e^{ix}) = Z(x),$$

where $Z(x)$ satisfies $P[Z(x)\neq 0]=1$, then the result of Theorem 4.2 follows. However, the assumption that $q=o(\ell n\ n)$ proves to be advantageous since this assumption will be necessary in order to prove subsequent more rapid convergence results.

In the next theorem we establish a more rapid convergence property of $e_1(\hat{F}_q(x))$. This result and its proof closely parallel the result and proof of McWilliams in the deterministic setting.

Theorem 4.3 Under the conditions of Theorem 4.2 we have

$$\underset{n\to\infty}{\text{p-lim}}\ \frac{F_\infty(x) - e_1(\hat{F}_q(x))}{F_\infty(x)-\hat{F}_{q+1}(x)} = 0.$$

Proof: Clearly

$$\frac{F_\infty(x) - e_1(\hat{F}_q(x))}{F_\infty(x) - \hat{F}_{q+1}(x)} = \frac{F_\infty(x) - \dfrac{\hat{F}_q(x)-\hat{\alpha}_{(q)}e^{ix}\hat{F}_{q-1}(x)}{1-\hat{\alpha}_{(q)}e^{ix}}}{F_\infty(x) - \hat{F}_{q+1}(x)}$$

$$= \frac{F_\infty(x) - \hat{F}_q(x) - \dfrac{\hat{\phi}_{(q+1)}e^{i(q+1)x}}{1-\hat{\alpha}_{(q)}e^{ix}}}{F_\infty(x) - \hat{F}_{q+1}(x)}$$

$$= \frac{F_\infty(x)-\hat{F}_{q+1}(x) - \dfrac{\hat{\phi}_{(q+1)}\hat{\alpha}_{(q)}e^{i(q+2)x}}{1 - \hat{\alpha}_{(q)}e^{ix}}}{F_\infty(x) - \hat{F}_{q+1}(x)}$$

$$= 1 - \frac{\dfrac{\hat{\phi}_{(q+1)}\hat{\alpha}_{(q)}e^{i(q+2)x}}{1 - \hat{\alpha}_{(q)}e^{ix}}}{F_\infty(x) - \hat{F}_{q+1}(x)}.$$

Therefore,

$$\operatorname*{p-lim}_{n\to\infty} \frac{F_\infty(x) - e_1(\hat{F}_q(x))}{F_\infty(x) - \hat{F}_{q+1}(x)}$$

$$= 1 - \frac{R}{1-Re^{ix}} \operatorname*{p-lim}_{n\to\infty} \frac{\hat{\alpha}_{(q+1)}^{-1}\hat{\phi}(q+2)e^{i(q+2)x}}{F_\infty(x) - \hat{F}_{q+1}(x)}$$

$$= 1 - \frac{1}{1-Re^{ix}} \operatorname*{p-lim}_{n\to\infty} \frac{\frac{\hat{\phi}(q+2)}{\phi(q+2)}}{[\phi(q+2)e^{i(q+2)x}]^{-1}[F_{q+1}(x) - \hat{F}_{q+1}(x) + \sum_{v=q+2}^{\infty} \phi(v)e^{ivx}]}.$$

Using arguments similar to that in Theorem 4.2, it is easily verified that

$$\operatorname*{p-lim}_{n\to\infty} \frac{\hat{\phi}(q+2)}{\phi(q+2)} = 1 \quad \text{and} \quad \operatorname*{p-lim}_{n\to\infty} \frac{1}{\phi(q)} [F_{q+1}(x) - \hat{F}_{q+1}(x)] = 0.$$

In addition, McWilliams has shown that

$$\lim_{n\to\infty} [\phi(q+2)e^{i(q+2)x}]^{-1} \sum_{v=q+2}^{\infty} \phi(v)e^{ivx} = \frac{1}{1-Re^{ix}}$$

under the assumption that $\lim_{m\to\infty} \frac{\phi(m)}{\phi(m-1)} = R$. It therefore follows that

$$\operatorname*{p-lim}_{n\to\infty} \frac{F_\infty(x) - e_1(\hat{F}_q(x))}{F(x) - F_{q+1}(x)} = 1 - \frac{1}{1-Re^{ix}} \cdot 1 - Re^{ix}$$

$$= 0.$$

Two points should be made about the result in Theorem 4.3. First, the theorem should be regarded as an optimality property of $\hat{f}_{1,q}(\cdot)$ but not as proof that $\hat{f}_{1,q}(\cdot)$ converges more rapidly to $f(\cdot)$ than $\hat{f}_{0,q+1}(\cdot)$. In order to prove this result it is

necessary to show

$$\underset{n\to\infty}{\text{p-lim}} \quad \frac{\text{Real}(F_\infty(x) - e_1(\hat{F}_q(x)))}{\text{Real}(F_\infty(x) - \hat{F}_{q+1}(x))} = 0 ,$$

which of course is not an immediate consequence of Theorem 4.3. However, since f(x) is completely determined by $F_\infty(x)$, an estimator of $F_\infty(x)$ which has good properties is of considerable importance. The second point to be made regarding Theorem 4.3 involves the rate at which $q \to \infty$. In order to insure the convergence of $\hat{f}_{1,q}(\cdot)$ it was seen in Theorem 4.2 that we must have $q = o(\ell n\ n)$. However, restricting q in this way in a comparison of $e_1(\hat{F}_q(x))$ and $\hat{F}_{q+1}(x)$ is, in a sense, unfair, since $\hat{F}_m(x)$ is consistent for $F_\infty(x)$ even when $m = o(\sqrt{n})$. Ideally, a comparison of the rate of convergence of $e_1(\hat{F}_q(x))$ and $\hat{F}_m(x)$ should be made with $q = o(\ell n\ n)$ and $m = o(\sqrt{n})$. By modifying the conditions of Theorem 4.3 it is possible to show that

$$\underset{n\to\infty}{\text{p-lim}} \quad \frac{F_\infty(x) - e_1(\hat{F}_q(x))}{F_\infty(x) - \hat{F}_m(x)} = 0 ,$$

where $q = o(\ell n\ n)$ and $m = o(\sqrt{n})$. This fact will be proven in Theorem 4.4.

Before moving to Theorem 4.5 we shall examine the effect which an assumption like (4.5) has on $\text{var}[\hat{F}_m(x)]$. This is important, as the rate at which $\hat{F}_m(x)$ converges to $F_\infty(x)$ is directly affected by $\text{var}[\hat{F}_m(x)]$. Now, since $\frac{\phi(m)}{\phi(m-1)}$ cannot converge to R any faster than in the case where

$$\frac{\phi(m)}{\phi(m-1)} = R \quad \text{for } m > m_0 , \tag{4.7}$$

we will assume (4.7) and then calculate $\text{var}[\hat{F}_m(x)]$. Under (4.7) we have

$$\phi(m) - R\phi(m-1) = 0 \text{ for } m > m_0,$$

which implies that $\phi(m_0+k) = \phi(m_0)R^k$, $k = 1,2,\ldots$ . Using the formula for $\text{var}[\hat{F}_m(x)]$ obtained in Theorem 4.1, we have

$$\text{var}[\hat{F}_m(x)] = \frac{1}{n}\sum_{v=1}^{m}(1-|\phi(v)|^2) + \frac{1}{n}\sum_{k=1}^{m-1}\sum_{v=k+1}^{m}(\phi(v-k)-\phi(v)\phi(-k))e^{i(v-k)x}$$

$$+ \frac{1}{n}\sum_{v=1}^{m-1}\sum_{k=v+1}^{m}(\phi(v-k)-\phi(v)\phi(-k))e^{i(v-k)x} \quad .$$

The first term in this expression is $0(\frac{m}{n})$ regardless of what assumption is made about $\phi(v)$, and thus only the convariance terms need to be considered in investigating the rate of convergence of $\text{var}[\hat{F}_m(x)]$. The second covariance term is simply the complex conjugate of the first, and so we consider only the first term. Under assumption (4.7) this term is (for $m > m_0 + 1$)

$$\frac{1}{n}\sum_{k=1}^{m-1}\sum_{v=k+1}^{m}\phi(v-k)e^{i(v-k)x} - \frac{1}{n}\sum_{k=1}^{m_0}\phi(-k)e^{-ikx}\sum_{v=k+1}^{m}\phi(v)e^{ivx} \qquad (4.8)$$

$$- \frac{|\phi(m_0)|^2}{n}\sum_{k=m_0+1}^{m-1}(\bar{R})^{k-m_0}e^{-ikx}\sum_{v=k+1}^{m}R^{v-m_0}e^{ivx} \quad .$$

It is easily verified that the last term in (4.8) is

$$\frac{|\phi(m_0)|^2}{n}\cdot\frac{|R|^2Re^{ix}}{1-Re^{ix}}\left\{\frac{1-|R|^{2(m-m_0-1)}}{1-|R|^2} - \frac{(Re^{ix})^{m-m_0-1}[1-(\bar{R}e^{-ix})^{m-m_0-1}]}{1-\bar{R}e^{-ix}}\right\}.$$

The second term in (4.8) is

$$\frac{1}{n}\sum_{k=1}^{m_0}\phi(-k)e^{-ikx}\left\{\sum_{v=k+1}^{m_0+1}\phi(v)e^{ivx}+\phi(m_0)\sum_{v=m_0+2}^{m}R^{v-m_0}e^{ivx}\right\}$$

$$=\frac{1}{n}\sum_{k=1}^{m_0}\sum_{v=k+1}^{m_0+1}\phi(-k)\phi(v)e^{i(v-k)x}$$

$$+\frac{\phi(m_0)}{nR^{m_0}}\cdot\frac{(Re^{ix})^{m_0+2}-(Re^{ix})^{m+1}}{1-Re^{ix}}\sum_{k=1}^{m_0}\phi(-k)e^{-ikx}\quad,$$

and the first term is

$$\frac{1}{n}\sum_{k=1}^{m-1}\sum_{v=1}^{m-k}\phi(v)e^{ivx}=\frac{1}{n}\sum_{k=1}^{m-1}(m-k)\phi(k)e^{ikx}$$

$$=\frac{m}{n}\left\{\sum_{k=1}^{m_0}\phi(k)e^{ikx}+\frac{\phi(m_0)m-1}{R^{m_0}}\sum_{k=m_0+1}(Re^{ix})^k\right\}$$

$$-\frac{1}{n}\left\{\sum_{k=1}^{m_0}k\phi(k)e^{ikx}+\frac{\phi(m_0)}{R^{m_0}}\sum_{k=m_0+1}^{m-1}k(Re^{ix})^k\right\}$$

$$=\frac{m}{n}\sum_{k=1}^{m_0}\phi(k)e^{ikx}-\frac{1}{n}\sum_{k=1}^{m_0}k\phi(k)e^{ikx}$$

$$+\frac{m}{n}\cdot\frac{\phi(m_0)}{R^{m_0}}\left[\frac{(Re^{ix})^{m_0+1}-(Re^{ix})^{m}}{1-Re^{ix}}\right]$$

$$-\frac{1}{n}\cdot\frac{\phi(m_0)}{R^{m_0}}\left[\frac{(Re^{ix})^{m_0+2}-(Re^{ix})^{m+1}}{(1-Re^{ix})^2}\right.$$

$$\left.+\frac{(m_0+1)(Re^{ix})^{m_0+1}-m(Re^{ix})^{m}}{1-Re^{ix}}\right]\quad.$$

The important thing to note about the three terms which make up (4.8) is that they are all $0(\frac{m}{n})$. Since $\frac{1}{n}\sum_{k=1}^{m}(1-|\phi(v)|^2)$ is

also $0(\frac{m}{n})$, this implies that $\text{var}[\hat{F}_m(x)] = 0(\frac{m}{n})$. Under condition (4.7) it thus follows that the truncation point m may be allowed to become large more quickly than generally stated. Specifically, $\hat{F}_m(x)$ is consistent for $F_\infty(x)$ so long as $m = o(n)$.

We are now in a position to establish our most important large sample result involving $e_1(\hat{F}_q(x))$.

Theorem 4.4 Under the conditions of Theorem 4.2 and the additional assumptions that $\frac{\phi(m)}{R^m} = 0(\sqrt{m})$ and

$$1 - \frac{\phi(q)}{\phi(q-1)} e^{ix} \left( \sum_{v=0}^{\infty} \frac{\phi(q+v)}{\phi(q)} e^{ivx} \right)^{-1} = o\left( \frac{\sqrt{m}}{\sqrt{n}} R^{-q} \right) \qquad (4.9)$$

we have

$$\underset{n\to\infty}{\text{p-lim}} \quad \frac{F_\infty(x) - e_1(\hat{F}_q(x))}{F_\infty(x) - \hat{F}_m(x)} = 0 \quad \text{for all } x \in [-\pi, \pi] ,$$

where $m = [n^\alpha]$, $0 < \alpha < \frac{1}{2}$ .

Proof:
$$\frac{F_\infty(x) - e_1(\hat{F}_q(x))}{F_\infty(x) - \hat{F}_m(x)} = \frac{F_\infty(x) - \hat{F}_{q-1}(x) - \frac{\hat{\phi}(q)e^{iqx}}{1-\hat{\alpha}_{(q)}e^{ix}}}{F_\infty(x) - \hat{F}_m(x)}$$

$$= \frac{F_{q-1}(x) - \hat{F}_{q-1}(x) + \sum_{v=q}^{\infty} \phi(v)e^{ivx} - \frac{\hat{\phi}(q)e^{iqx}}{1-\hat{\alpha}_{(q)}e^{ix}}}{F_m(x) - \hat{F}_m(x) + \sum_{v=m+1}^{\infty} \phi(v)e^{ivx}}$$

$$= \frac{\frac{\sqrt{n}}{\sqrt{m}}(F_{q-1}(x) - \hat{F}_{q-1}(x)) + \frac{\sqrt{n}}{\sqrt{m}}\phi(q)e^{iqx}\left\{ \sum_{v=0}^{\infty} \frac{\phi(v+q)}{\phi(q)} e^{ivx} - \frac{\frac{\hat{\phi}(q)}{\phi(q)}}{1-\hat{\alpha}_{(q)}e^{ix}} \right\}}{\frac{\sqrt{n}}{\sqrt{m}}(F_m(x) - \hat{F}_m(x)) + \frac{\sqrt{n}}{\sqrt{m}}\phi(m+1)e^{i(m+1)x} \sum_{v=0}^{\infty} \frac{\phi(m+1+v)}{\phi(m+1)} e^{ivx}} . \qquad (4.10)$$

We have

$$\lim_{n\to\infty} \frac{\sqrt{n}}{\sqrt{m}}\, \phi(m+1) = \lim_{n\to\infty} \frac{\sqrt{n}}{\sqrt{m}}\, \frac{\phi(m+1)}{R^{m+1}}\, R^{m+1} = 0$$

since

$$\frac{\phi(m+1)}{R^{m+1}} = 0(\sqrt{m}) \quad \text{and} \quad m = [n^{\alpha}] .$$

Also recall that

$$\lim_{m\to\infty} \sum_{v=0}^{\infty} \frac{\phi(m+1+v)}{\phi(m+1)} e^{ivx} = \frac{1}{1-Re^{ix}} .$$

The previous considerations concerning $\mathrm{var}[\hat{F}_m(x)]$ indicate that one of the following holds

$$\mathrm{var}[\frac{\sqrt{n}}{\sqrt{m}}(F_m(x) - \hat{F}_m(x))] \to c_1 \neq 0$$

or

$$\mathrm{var}[\frac{\sqrt{n}}{\sqrt{m}}(F_m(x) - \hat{F}_m(x))] \to \infty \quad \text{as } n \to \infty.$$

It thus follows that the denominator of (4.10) converges in probability to a random variable $Z(x)$ which satisfies $P[Z(x) \neq 0] = 1$. Therefore, in order to prove the result it is sufficient to show that the numerator of (4.10) converges in probability to zero. Clearly (since $q = o(\ell n\ n)$ and $m = [n^{\alpha}]$) the first term of the numerator goes to zero in probability as $n \to \infty$. The second term is (with $a_q(x) = \sum_{v=0}^{\infty} \frac{\phi(q+v)}{\phi(q)} e^{ivx}$)

$$\frac{\sqrt{n}}{\sqrt{m}}\, \phi(q)e^{iqx}\left\{a_q(x) - \frac{\frac{\hat{\phi}(q)}{\phi(q)}}{1-\hat{\alpha}_{(q)}e^{ix}}\right\} = \frac{\sqrt{n}}{\sqrt{m}}\phi(q)e^{iqx}\left\{\frac{a_q(x)(1-\hat{\alpha}_{(q)}e^{ix}) - \frac{\hat{\phi}(q)}{\phi(q)}}{1-\hat{\alpha}_{(q)}e^{ix}}\right\}$$

$$= \frac{\sqrt{n}}{\sqrt{m}}\,\phi(q)e^{iqx}\left\{\frac{a_q(x)\left(\frac{\hat{\phi}(q)}{\phi(q)} - \frac{\hat{\phi}(q+1)}{\phi(q)}\,e^{ix}\right) - \left(\frac{\hat{\phi}(q)}{\phi(q)}\right)^2}{\left(\frac{\hat{\phi}(q)}{\phi(q)} - \frac{\hat{\phi}(q+1)}{\phi(q)}\,e^{ix}\right)}\right\}.$$

The denominator of the expression in brackets converges in probability to $(1-Re^{ix})$. The numerator may be expressed as

$$a_q(x)\left[1 - \frac{\phi(q+1)}{\phi(q)}e^{ix} + \frac{0_p(\frac{1}{\sqrt{n}})}{\phi(q)}(1-e^{ix})\right] - \left(1 + \frac{0_p(\frac{1}{\sqrt{n}})}{\phi(q)}\right)^2$$

$$= a_q(x)\left(1 - \frac{\phi(q+1)}{\phi(q)}e^{ix}\right) - 1 + \frac{0_p(\frac{1}{\sqrt{n}})}{\phi(q)}\left[a_q(x)(1-e^{ix}) - 2 - \frac{0_p(\frac{1}{\sqrt{n}})}{\phi(q)}\right].$$

The limit in probability of the numerator of (4.10) is thus

$$\frac{1}{1-Re^{ix}}\ \underset{n\to\infty}{\text{p-lim}}\ \frac{\sqrt{n}}{\sqrt{m}}0_p(\frac{1}{\sqrt{n}})e^{iqx}[a_q(x)(1-e^{ix}) - 2 - \frac{0_p(\frac{1}{\sqrt{n}})}{\phi(q)}]$$

$$+ \frac{1}{1-Re^{ix}}\ \lim_{n\to\infty}\ \frac{\sqrt{n}}{\sqrt{m}}\,\phi(q)e^{iqx}[a_q(x)(1 - \frac{\phi(q+1)}{\phi(q)}\,e^{ix}) - 1]\,. \qquad (4.11)$$

Clearly $\underset{n\to\infty}{\text{p-lim}}\ \frac{\sqrt{n}}{\sqrt{m}}0_p(\frac{1}{\sqrt{n}}) = 0$, and, as noted in Theorem 4.2,

$$\underset{n\to\infty}{\text{p-lim}}\ \frac{0_p(\frac{1}{\sqrt{n}})}{\phi(q)} = 0 \text{ and } \lim_{n\to\infty} a_q(x) = \frac{1}{1-Re^{ix}}$$

Thus, (4.11) is simply

$$\frac{1}{(1-Re^{ix})^2}\ \lim_{n\to\infty}\ \frac{\sqrt{n}}{\sqrt{m}}\,\phi(q)e^{iqx}[(1 - \frac{\phi(q+1)}{\phi(q)}\,e^{ix}) - a_q^{-1}(x)]\,,$$

which by hypothesis is zero, and the proof is complete.

In light of previous considerations we point out that, if instead of assuming (4.9) we assume that (4.7) holds and $m = [n^{\alpha}]$, $0 < \alpha < 1$, then the preceding proof remains valid.

Assumption (4.9), which is crucial to the proof of Theorem 4.4, has to do with how quickly the ratio $\frac{\phi(q+1)}{\phi(q)}$ approaches R. Note that since

$$\lim_{n\to\infty} \frac{\phi(q+1)}{\phi(q)} = R, \ (1 - \frac{\phi(q+1)}{\phi(q)} e^{ix}) - a_q^{-1}(x) \to 0$$

as $n \to \infty$. However, in order for (4.9) to be satisfied $\frac{\phi(q+1)}{\phi(q)}$ must converge to R rapidly enough to compensate for the fact that $\frac{\sqrt{n}}{\sqrt{m}}|R^q| \to \infty$. The convergence of $\frac{\phi(q+1)}{\phi(q)}$ to R is the more rapid of the two, and (4.9) is obviously satisfied, in the case where (4.7) holds. Condition (4.9) may thus be regarded as an indication of how far the Fourier coefficients of $f(\cdot)$ may depart from the model "$\phi(m)-R\phi(m-1)=0$ for $m > m_0$" while still maintaining the property

$$\operatorname*{p-lim}_{n\to\infty} \frac{F_\infty(x) - e_1(\hat{F}_q(x))}{F_\infty(x) - \hat{F}_m(x)} = 0 .$$

We would indeed be remiss if the current section was concluded without a discussion of how the previous results involving $\hat{f}_{1,q}(\cdot)$ should properly fit into a general approach to density estimation. First, if the Fourier series estimator $\hat{f}_{0,m}(\cdot)$ is employed to estimate $f(\cdot)$ then the results of this section provide a justification for at least considering $\hat{f}_{1,q}(\cdot)$ as an alternative to $\hat{f}_{0,m}(\cdot)$. The nonparametric nature of the density estimation problem does not allow us to make specific assumptions (such as those in Theorems 4.2 and 4.4) about the underlying density, but this fact should not blind us from the realization that one estimator may perform better than another in certain situations. In Theorem 4.4 conditions were established under which $\hat{f}_{1,q}(\cdot)$ would reasonably be expected to perform better

than $\hat{f}_{0,m}(\cdot)$. With these thoughts in mind, the only question which remains is the following. For a given data set, how does one recognize if the situation calls for the use of $\hat{f}_{1,q}(\cdot)$ (for some q) rather than $\hat{f}_{0,m}(\cdot)$? Two possible answers to this question will be offered in Chapter V.

## 4.4 Extending Results Involving $\hat{f}_{1,q}(\cdot)$ to Higher Orders

As mentioned in the previous section, McWilliams has investigated certain properties of the $e_2$-transform. Since

$$\hat{f}_{2,q}(x) = \frac{1}{2\pi}\,[1 + 2\text{Real}(e_2(\hat{F}_q(x))] \quad (\text{for } q \geq 2),$$

some insight into when $\hat{f}_{2,q}(\cdot)$ may be of value as an estimator of $f(\cdot)$ can be gained by considering the theorems of McWilliams involving $e_2(A_m)$. The following two theorems, stated without proof, have been proven by McWilliams (1969).

Theorem 4.5 If $A_m \to A_\infty$, $\dfrac{a_{m+1}}{a_m} = R_m \to R \neq 1$, and

$$\lim_{m\to\infty} \frac{R_{m+1} - R_m}{R_{m+2} - R_{m+1}} = Q \neq R, \text{ then } e_2(A_m) \to A_\infty .$$

Theorem 4.6 If the conditions of Theorem 4.5 are satisfied, and if further $R \neq 0$ and

$$\frac{A_\infty - A_{m+1}}{A_\infty - A_m} \to R ,$$

then

$$\lim_{m\to\infty} \frac{A_\infty - e_2(A_{m+2})}{A_\infty - A_{m+j}} = 0 \quad \text{for any } j .$$

Using the results of Theorems 4.5 and 4.6 and proceeding as in Section 4.3 one could undoubtedly establish results for $e_2(\hat{F}_q(x))$ paralleling those of the previous section. The proofs of these results would be somewhat more tedious than those in the first order case (due to the increased complexity of the $e_2$-transform), and perhaps not altogether necessary. It seems that the proven worth of $e_2(A_m)$ (as evidenced in Theorems 4.5 and 4.6) in the deterministic setting is alone a motivation for considering $\hat{f}_{2,q}(\cdot)$ to be, in certain situations, a viable competitor of the Fourier series estimator.

Except for the situation in which $\{a_m\} \in L(n,\Delta)$, conditions insuring the convergence of $e_n(A_m)$ and the more rapid convergence of $e_n(A_m)$ than $A_{n+m}$ have not been established for the cases where $n \geq 3$. However, the importance of the exactness result obtained in Theorem 2.2 should not be overlooked. Even when the assumption $\{a_m\} \varepsilon L(n,\Delta)$ is only approximately satisfied, $e_n(A_m)$ can be expected to be of considerable value. This fact was demonstrated in the examples of Chapter II. Likewise using Theorem 2.2 as a justification for considering $\hat{f}_{p,q}(\cdot)$ to be a candidate estimator of $f(\cdot)$, there remains the problem of how to select p and q. This problem is the subject of the next chapter.

# CHAPTER V

# THE PROBLEM OF SELECTING p AND q

## 5.1 Introduction

Up to this point, the primary objective of this work has been to illustrate and discuss the various reasons why the class of ARMA estimators are of value in the density estimation problem. Armed with a suitable class of estimators, we are left, however, with the practical problem of choosing an appropriate estimator (based on data $X_1,\ldots,X_n$) from this class. Given a realization $x_1,\ldots,x_n$ from $f(\cdot)$, the class of ARMA estimates of the density function is indexed only by p and q, and so choosing an appropriate estimate is equivalent to choosing appropriate values of p and q.

All density estimation methods have a problem similar to the one described above. Typically a class of estimates is indexed by a parameter, often referred to as a smoothing parameter, and a suitable value of this parameter must be chosen in order to arrive at a final estimate of $f(\cdot)$. For example, when employing a kernel density estimator

$$f_n(x) = \frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{x-X_j}{h}\right),$$

a suitable choice of the smoothing parameter (or window width) h must be made. Duin (1976) and Hermans and Habbema (1976) have proposed a modified maximum-likelihood approach to the problem of choosing h. In addition, Silverman (1980) has suggested a

method for choosing the window width based on the so called test graph theorem. The problem of choosing a smoothing parameter also arises in a method proposed by Wahba (1977). Wahba's estimator is (for $x \in [0,1]$ and n even)

$$\hat{f}(x) = \sum_{v=-n/2}^{n/2} \frac{\hat{\phi}(2\pi v)}{1+\lambda(2\pi v)^4} e^{2\pi i v x} ,$$

which is seen to be a Fourier series estimator to which a low-pass filter has been applied. The smoothing of $\hat{f}(\cdot)$ is accomplished by varying the parameter $\lambda$ rather than the truncation point of the Fourier series as in the method of Kronmal and Tarter. Wahba (1978) chooses $\lambda$ so that the estimated MISE of $\hat{f}(\cdot)$ is a minimum. The problem we have discussed is shared even by the primitive histogram estimator, whose smoothing parameters are the number and size of its class intervals.

In ARMA density estimation, the pair of values (p,q) may be regarded as the smoothing parameter. In the remainder of this chapter two different methods for choosing this parameter will be presented. In the first method we propose the use of the S-array for choosing (p,q), since the ARMA (p,q) representation for $f(\cdot)$ is equivalent to assuming that $\{\phi(v)\} \in L(p,\Delta)$ for $v > q$. In the second method (p,q) is chosen in such a way that the estimated MISE of $\hat{f}_{p,q}(\cdot)$ is minimized over a suitably restricted subclass of ARMA estimates.

## 5.2 S-Array Method of Selecting p and q

As proven in Chapter II, the assumption that a function $f(\cdot)$

has an ARMA (p,q) representation is essentially equivalent to the assumption that

$$\{\phi(v)\} \in L(p,\Delta) \quad \text{for } v > q.$$

Therefore, given estimated Fourier coefficients $\hat{\phi}(1), \hat{\phi}(2),\ldots,\hat{\phi}(M)$, a natural way of selecting (p,q) is to examine an S-array (see Gray, Kelley, and McIntire (1978)) composed of values $S_n(\hat{\phi}(m)e^{imx})$ $(x\in[-\pi,\pi])$.

$$S_{p'}(\hat{\phi}(m)e^{imx}) \simeq c_1 \quad , \quad m \geq q' \quad \text{and}$$

$$S_{p'}(\hat{\phi}(m)e^{imx}) \simeq c_2 \quad , \quad m \leq -q' - 1$$

supports the choice of (p',q') for the smoothing parameter (p,q) in the sense that such a pattern supports the existence of a similar pattern in the S-array based on $S_n(\phi(m)e^{imx})$.

Some experience with simulated data has shown that even when a good constancy pattern exists in the parametric S-array, the sample S-array tends to be more noisy than arrays encountered in time series applications. The method of selecting (p,q) discussed above must undoubtedly, then, involve a good deal of subjectivity. For this reason the S-array should be regarded as a tool for pointing out a restricted class of candidate ARMA estimates. Additional analysis may be performed on the restricted class of estimates to determine a final estimate of $f(\cdot)$.

One possibility for arriving at a final estimate would be to perform some sort of smoothing in the S-array columns where constancy patterns are apparent. Tukey (1978) has suggested the use of his 3RSSS smoothing procedure as a means of making noisy patterns in the S-array more informative. After smoothing competing columns

a choice for (p,q) may become obvious. A second possibility for obtaining an estimate would be to estimate the MISE for each candidate in the restricted class of ARMA estimates, and then choose that estimate which minimizes the estimated MISE. A procedure for estimating MISE is discussed in the next section.

In Chapter VI, the S-array method of selecting (p,q) will be exemplified in the analysis of two different data sets. The smoothing procedure discussed previously has not been investigated, but we do examine the estimated MISE criterion.

## 5.3 MISE Criterion for Selecting p and q

Ideally we would like to choose an estimator from the ARMA class which satisfies some optimality criterion with respect to $f(\cdot)$. A criterion which is common in the estimation of probability density functions is to seek an estimator which minimizes the MISE. In ARMA density estimation this entails choosing (p,q) such that

$$\text{MISE}(\hat{f}_{p,q}) = E[\int_{-\pi}^{\pi}(\hat{f}_{p,q}(x) - f(x))^2 dx]$$

is minimized. This, however, is an impossible task since the optimal value of (p,q) depends on $f(\cdot)$, the function which is to be estimated. Therefore, given data $x_1,\ldots,x_n$, our approach will be to choose as our estimate of $f(\cdot)$ that $\hat{f}_{p,q}(\cdot)$ for which an estimated $\text{MISE}(\hat{f}_{p,q})$ is minimized. In the remainder of this section we discuss the problem of estimating $\text{MISE}(\hat{f}_{p,q})$.

Consider

$$\int_{-\pi}^{\pi}(\hat{f}_{p,q}(x)-f(x))^2 dx = \int_{-\pi}^{\pi}\hat{f}^2_{p,q}(x)dx - 2\int_{-\pi}^{\pi}\hat{f}_{p,q}(x)f(x)dx + \int_{-\pi}^{\pi}f^2(x)dx.$$

From this expression it is clear that the value of (p,q) which minimizes

$$J(\hat{f}_{p,q}) = E[\int_{-\pi}^{\pi} \hat{f}^2_{p,q}(x)dx - 2\int_{-\pi}^{\pi} \hat{f}_{p,q}(x)f(x)dx]$$

also minimizes MISE $(\hat{f}_{p,q})$. It is therefore sufficient to consider only the estimation of $J(\hat{f}_{p,q})$. This observation greatly simplifies our problem.

Recalling relationship (3.6) we have

$$J(\hat{f}_{p,q}) = E\Big[\int_{-\pi}^{\pi} \hat{f}^2_{0,p+q}(x)dx + 2\int_{-\pi}^{\pi} \hat{f}_{0,p+q}(x)\hat{g}_{p,q}(x)dx$$

$$+ \int_{-\pi}^{\pi} \hat{g}^2_{p,q}(x)dx - 2\int_{-\pi}^{\pi} \hat{f}_{0,p+q}(x)f(x)dx - 2\int_{-\pi}^{\pi} \hat{g}_{p,q}(x)f(x)dx\Big]$$

$$= J(\hat{f}_{0,p+q}) + J(\hat{g}_{p,q}) + 2E[\int_{-\pi}^{\pi} \hat{f}_{0,p+q}(x)\hat{g}_{p,q}(x)dx]$$

Now $J(\hat{f}_{0,p+q})$ has a particularly simple form in terms of $\phi(1),\ldots,$ $\phi(p+q)$ for which there is an unbiased estimator. We have

$$J(\hat{f}_{0,p+q}) = E\Big[\frac{1}{2\pi}\Big(1+2\sum_{v=1}^{p+q}|\hat{\phi}(v)|^2\Big)\Big] - 2\int_{-\pi}^{\pi} f_{0,p+q}(x)f(x)dx$$

$$= E\Big[\frac{1}{2\pi}\Big(1+2\sum_{v=1}^{p+q}|\hat{\phi}(v)|^2\Big)\Big] - \frac{1}{\pi}\Big(1+2\sum_{v=1}^{p+q}|\phi(v)|^2\Big).$$

Since $E(\frac{n}{n-1}|\hat{\phi}(v)|^2 - \frac{1}{n-1}) = |\phi(v)|^2$, it follows that an unbiased estimate of $J(\hat{f}_{0,p+q})$ is

$$\hat{J}(\hat{f}_{0,p+q}) = \frac{1}{2\pi}\Big(1 + 2\sum_{v=1}^{p+q}|\hat{\phi}(v)|^2\Big) - \frac{1}{\pi}\Big[1 + 2\sum_{v=1}^{p+q}\Big(\frac{n}{n-1}|\hat{\phi}(v)|^2 - \frac{1}{n-1}\Big)\Big]$$

$$= \frac{-1}{2\pi}\Big(1 + \frac{2(n+1)}{(n-1)}\sum_{v=1}^{p+q}|\hat{\phi}(v)|^2\Big) + \frac{2(p+q)}{\pi(n-1)} \quad .$$

Interestingly, it is seen that the first term of $\hat{J}(\hat{f}_{0,p+q})$ decreases as p+q increases, but that the second term penalizes an increase

in $p + q$. Thus, $\hat{J}(\hat{f}_{0,p+q})$ is sensitive to both the fidelity and stability of $\hat{f}_{0,p+q}(\cdot)$. In addition, it is easily verified that $\hat{J}(\hat{f}_{0,m+1}) - \hat{J}(\hat{f}_{0,m})$ is the same estimate of MISE $(\hat{f}_{0,m+1})$ - MISE$(\hat{f}_{0,m})$ as that derived by Kronmal and Tarter (1968). This fact points out a correspondence between the optimal stopping rule of Kronmal and Tarter and $\hat{J}(\hat{f}_{0,m})$.

An unbiased estimator of the last term of $\hat{J}(\hat{f}_{p,q})$ is

$$2\int_{-\pi}^{\pi} \hat{f}_{0,p+q}(x)\hat{g}_{p,q}(x)dx,$$

which is zero whenever the estimate $\hat{f}_{p,q}(\cdot)$ satisfies condition S. This leaves us with the problem of estimating

$$J(\hat{g}_{p,q}) = E[\int_{-\pi}^{\pi}\hat{g}^2_{p,q}(x)dx] - 2E[\int_{-\pi}^{\pi}\hat{g}_{p,q}(x)dF(x)].$$

Since an unbiased estimator of $E[\int_{-\pi}^{\pi}\hat{g}^2_{p,q}(x)dx]$ is $\int_{-\pi}^{\pi}\hat{g}^2_{p,q}(x)dx$, we focus our attention on $E[\int_{-\pi}^{\pi}\hat{g}_{p,q}(x)dF(x)]$. To estimate this quantity we propose the use of the bootstrap mechanism of Efron (1979).

In order to illustrate the bootstrap in this setting, let $X = (X_1,\ldots,X_n)$ denote a random sample from $f(\cdot)$. Further, let $\hat{g}_{p,q,X}(\cdot)$ indicate that $\hat{g}_{p,q}(\cdot)$ is based on the sample X, and write

$$R(X,F) = \int_{-\pi}^{\pi}\hat{g}_{p,q,X}(x)dF(x).$$

If $(x_1,\ldots,x_n)$ is a realization of X with corresponding empirical cdf $F_n(\cdot)$, then the bootstrap estimate of $E[R(X,F)]$ is

$$E[R(X^*,F_n)] = E[\frac{1}{n}\sum_{j=1}^{n}\hat{g}_{p,q,X^*}(x_j)],$$

where $X^*$ is a random sample from $F_n(\cdot)$. This estimate is seen to be Fisher consistent, or in other words, the estimate is equal to the parameter it estimates when $F_n(\cdot) = F(\cdot)$.

As it is not possible to analytically evaluate $E[R(X^*,F_n)]$, Efron suggests that numerous samples $X^*$ be generated from $F_n(\cdot)$ in order to empirically evaluate the expectation to a close approximation. In our application this procedure would be prohibitive since $E[R(X^*,F_n)]$ must be evaluated for numerous different candidate estimators, $\hat{f}_{p,q}(\cdot)$. Fortunately, Efron also derives a second order approximation to $E[R(X^*,F_n)]$ by expanding $R(P^*) = R(X^*,F_n)$ in a Taylor series about $\frac{1}{n}(1,1,\ldots,1)$ where $P^* = (P_1^*,\ldots,P_n^*)$ and $P_i^* = \frac{1}{n}$(number of $X_j^*$'s which equal $x_i$). ($R(X^*,F_n)$ depends on $X^*$ only through $P^*$ since $R$ is symmetric in the $X_i^*$'s.)

Wong (1979) has derived an explicit expression for Efron's approximation to $E[R(X^*,F_n)]$ which he shows, in fact, to be a jackknife approximation to the bootstrap. We have

$$E[R(X^*,F_n)] \simeq \sum_{j=1}^{n} R(x_{(j)},F_n) - (n-1)R(x,F_n),$$

where $x_{(j)} = (x_1,\ldots,x_{j-1},x_{j+1},\ldots,x_n)$. In our problem this implies that

$$E[R(X^*,F_n)] \simeq \sum_{j=1}^{n} \{\frac{1}{n}\sum_{i=1}^{n} \hat{g}_{p,q,x_{(j)}}(x_i)\} - \frac{(n-1)}{n}\sum_{i=1}^{n} \hat{g}_{p,q,x}(x_i).$$

An even simpler approximation is possible by noting that (for n reasonably large) $\hat{g}_{p,q,x_{(j)}}(t) \simeq \hat{g}_{p,q,x}(t)$ except possibly for $t$ in a neighborhood of $x_j$. This observation gives

$$E[R(X^*,F_n)] \simeq \sum_{j=1}^{n} \{ \frac{1}{n} \hat{g}_{p,q,x_{(j)}}(x_j) + \frac{1}{n} \sum_{i=1}^{n} \hat{g}_{p,q,x}(x_i) - \frac{1}{n} \hat{g}_{p,q,x}(x_j) \} - \frac{n-1}{n} \sum_{i=1}^{n} \hat{g}_{p,q,x}(x_i)$$

$$= \frac{1}{n} \sum_{j=1}^{n} \hat{g}_{p,q,x_{(j)}}(x_j).$$

Wong (1979) uses this second type of approximation to show that the modified maximum likelihood method of choosing a smoothing parameter is related to the bootstrap.

If we take

$$\hat{J}(\hat{g}_{p,q}) = \int_{-\pi}^{\pi} \hat{g}^2_{p,q}(x)dx - \frac{2}{n} \sum_{j=1}^{n} \hat{g}_{p,q,X_{(j)}}(X_j),$$

then our estimator of $J(\hat{f}_{p,q})$ is

$$\hat{J}(\hat{f}_{p,q}) = \hat{J}(\hat{f}_{0,p+q}) + \hat{J}(\hat{g}_{p,q}) + 2\int_{-\pi}^{\pi} \hat{f}_{0,p+q}(x)\hat{g}_{p,q}(x)dx.$$

Now, if $A$ is some subclass of the class of all ARMA estimates, then we define $\hat{f}_{p',q'}(\cdot)$ to be the $A$ ARMA estimate of $f(\cdot)$ if and only if $\hat{f}_{p',q'}(\cdot) \in A$ and

$$\hat{J}(\hat{f}_{p',q'}) \leq \hat{J}(\hat{f}_{p,q}) \quad \text{for all } \hat{f}_{p,q}(\cdot) \in A.$$

The class $A$ should be large enough to insure that an adequate estimate of $f(\cdot)$ is obtained, but not so large that the computing time required to identify $\hat{f}_{p',q'}(\cdot)$ is excessive. One method of restricting the size of $A$ is suggested in the previous section through the use of the S-array. One might also consider the class $A_M$ of all ARMA estimators $\hat{f}_{p,q}(\cdot)$ satisfying $p+q \leq M$.

The considerations of Chapter IV indicate that, if M is chosen to be a function of the sample size n, it would be reasonable to have $M = o(\sqrt{n})$.

In the next chapter, we will investigate the Á ARMA estimator by means of simulated data.

# CHAPTER VI

# ARMA DENSITY ESTIMATION IN PRACTICE, AND A SUMMARY

## 6.1 Introduction

In this final chapter the use of ARMA representations in density estimation is exemplified with the aid of both real and simulated data. In Sections 6.2 and 6.3 two data sets which have appeared previously in the literature are considered. The LRL data of Good and Gaskins (1980) and the Maguire data of Maguire, Pearson, and Wynn (1952) are analyzed, and density estimates are obtained using the results of Chapter V. The effectiveness of the estimated MISE criterion for choosing p and q is evaluated in Section 6.4 by means of simulated data. The results of the simulation study show the criterion to be quite effective in distinguishing between density estimates which have important differences in ISE.

Section 6.5 is devoted to summarizing the density estimation results obtained in this work. In addition, some areas for future research in ARMA density estimation are indicated.

## 6.2 The LRL Data

Good and Gaskins (1980) have analyzed a data set, which they call the LRL (Lawrence Radiation Laboratory) data, consisting of "n = 25,752 events from a scattering reaction". The data are recorded in the paper of Good and Gaskins in the form

of a frequency table made up 172 bins of width 10 MeV each. The ith bin includes $n_i$ events, and the bins are centered at the values (in MeV)

$$y_i = 285 + 10(i-1), \quad i = 1,2,\ldots,172.$$

In the analysis to follow we consider the transformed data

$$x_i = \frac{\pi}{1720}(2y_i - 2280) \ , \ i = 1,2,\ldots,172.$$

The Fourier coefficients $\phi(v)$, $v = 1,2,\ldots$, associated with the density $f(\cdot)$ of the transformed data are estimated by

$$\hat{\phi}(v) = \frac{1}{n} \sum_{j=1}^{172} n_j e^{-ivx_j} \ , \quad v = 1,2,\ldots \ . \tag{6.1}$$

The aim of Good and Gaskins in analyzing the LRL data was to obtain an estimate of the underlying probability density by using their maximum penalized-likelihood method (see Good and Gaskins (1971)), and to then describe a procedure for assessing the likelihood that a bump found in the estimate is also present in the underlying density. Our purpose in analyzing the LRL data is to

(i) illustrate the cogent information contained in the S-array about the type of ARMA estimate which should be fit, and

(ii) to obtain an estimate comparable to that of Good and Gaskins.

In this example, the estimated MISE criterion for choosing (p,q)

is considered only for estimates with p = 0, as some modification of the jackknife approximation to the bootstrap is needed for grouped data.

Table 6.1 shows a portion of the S-array for the sequence $\{(-1)^m\hat{\phi}(m)\}$, where the $\hat{\phi}(m)$ are as defined in (6.1). The array based on $\{(-1)^m\hat{\phi}(m)\}$ has been tabled since it shows a much clearer constancy pattern than does the array based on $\{\hat{\phi}(m)\}$. This behavior is caused by the fact that, as will be seen shortly, estimates of the density $f(\cdot)$ have considerably more "power" near $x = 0$ than near $x = \pi$.

The constancy apparent in the first two columns of the array in Table 6.1 gives clear preference to ARMA estimates with p = 1 or p = 2. However, a fuller understanding of the information contained in Table 6.1 can be gained by initially considering the estimates $\hat{f}_{1,0}(\cdot)$ and $\hat{f}_{2,0}(\cdot)$, which are plotted in Figures 6.1 and 6.2 respectively. From these two figures it is clear that the constancy in the first column of the S-array corresponds to a bump (in the terminology of Good and Gaskins) at about x = -1.40, and the constancy in the second column corresponds to this same bump and another smaller bump at about x = .55. Interestingly, $\hat{f}_{2,0}(\cdot)$ is virtually the same estimate as that obtained by Good and Gaskins except for the presence of 11 additional, very small bumps in their estimate.

An area for future research is establishing a method of transforming the original sequence of estimated Fourier coeffi-

TABLE 6.1

LRL DATA S-ARRAY FOR $\{(-1)^{v}\hat{\phi}(v)\}$

| m/n | 1 | | 2 | | 3 | | 4 | |
|---|---|---|---|---|---|---|---|---|
| -7 | -1.3247 | -1.2329 | 2.2230 | 5.3957 | -3.0545 | .8639 | 2.6744 | - .5604 |
| -6 | - .9090 | -1.1078 | 4.9327 | 1.7863 | -2.9125 | .0762 | -23.3285 | 18.5090 |
| -5 | -1.2293 | -1.9280 | 4.3365 | 2.2353 | -11.1300 | -13.0432 | 2.2525 | 8.6208 |
| -4 | -1.7303 | -1.0765 | 3.5323 | 1.1760 | .3243 | 3.5225 | 1.3842 | -1.8984 |
| -3 | -1.1405 | - .9692 | 3.0521 | 2.9844 | - 1.2961 | 3.0222 | -12.3920 | 9.1277 |
| -2 | - .6838 | -1.6675 | 4.5754 | 2.2416 | - 5.7802 | -.9417 | 7.9142 | 11.8507 |
| -1 | -1.9582 | -1.4539 | 4.7452 | 1.9676 | 3.8462 | 56.4148 | 16.3735 | 6.7288 |
| 0 | -1.3160 | - .4795 | 1.7975 | .6885 | - 1.7911 | -.6242 | 2.0011 | .7121 |
| 1 | - .8902 | - .5789 | 1.8128 | .7952 | - 6.8675 | 4.0155 | 1.6265 | -.1811 |
| 2 | -1.1465 | -1.0106 | 2.1618 | 1.0597 | - 1.1163 | .6193 | .9432 | -.2865 |
| 3 | -1.4316 | - .6361 | 2.3888 | .4422 | - 1.1480 | .2373 | -1.4954 | 1.9320 |
| 4 | -1.0608 | - .5114 | 1.8618 | .7690 | - 2.1166 | -1.2363 | 1.4764 | .9570 |
| 5 | - .9264 | - .8966 | 1.6925 | .6826 | .2784 | 1.9385 | .6057 | -1.5580 |
| 6 | -1.1998 | - .7585 | 1.4623 | 1.1284 | - .5614 | 1.7155 | -9.1205 | - .6029 |

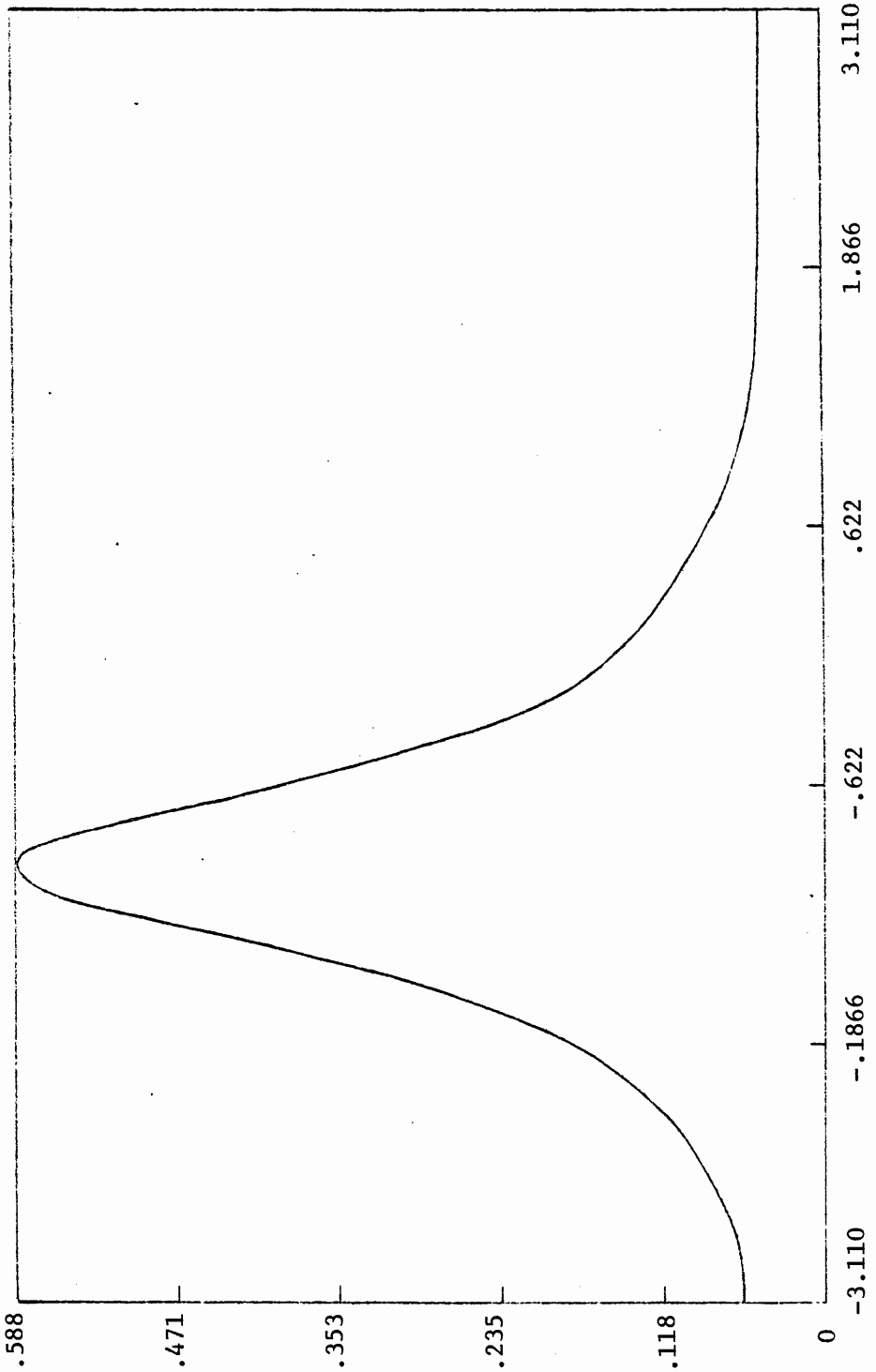


FIGURE 6.1

The Estimate $\hat{f}_{1,0}(\cdot)$ for the LRL Data

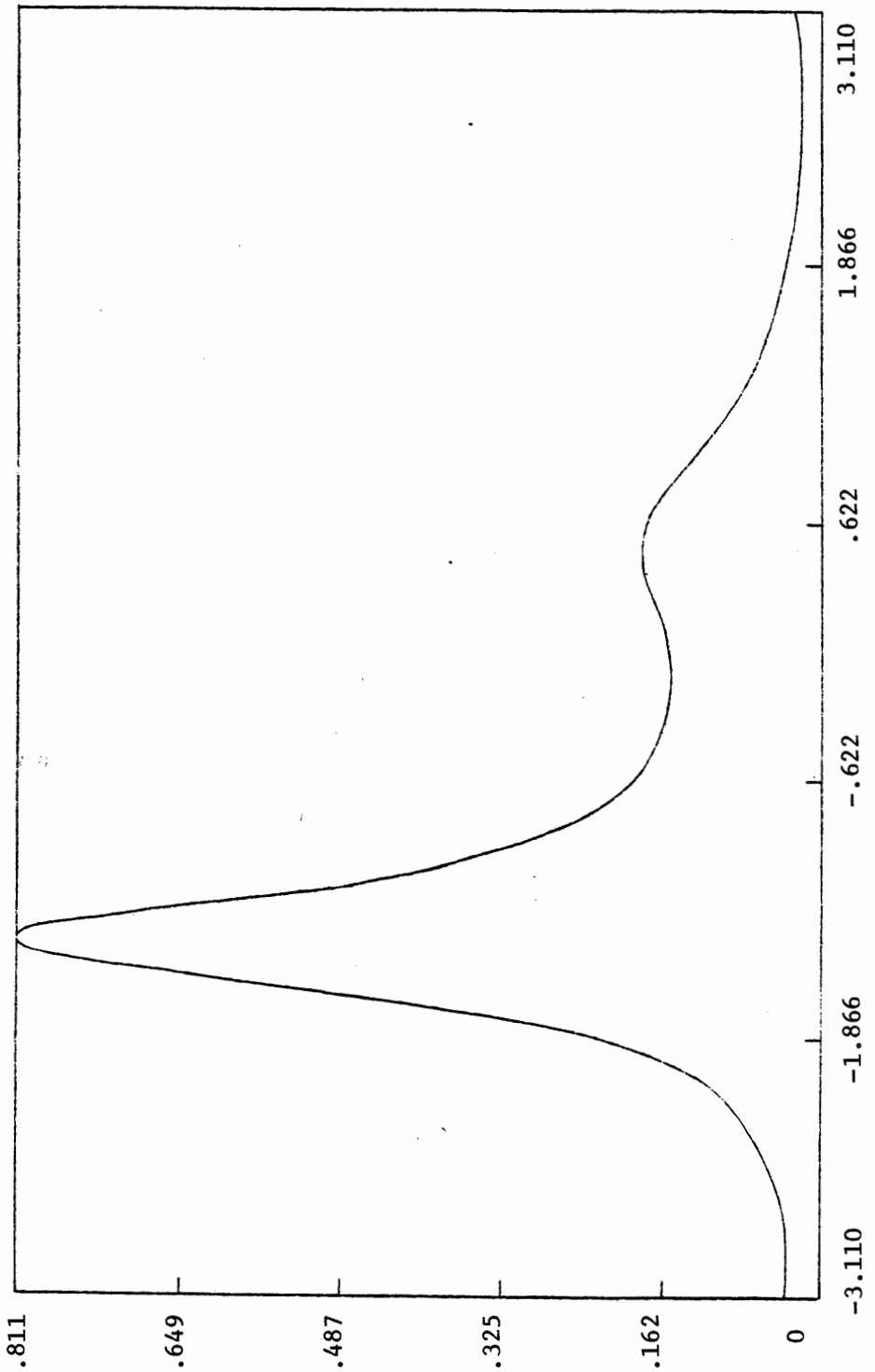


FIGURE 6.2

The Estimate $\hat{f}_{2,0}$ (•) for the LRL Data

cients in such a way that the dominating effect of major peaks is filtered out. In the current example, such a method would allow us to remove the effect of the two major peaks (seen in Fig. 6.2) so that the possible presence of smaller peaks could be carefully investigated.

In the absence of a suitable filtering technique, we arrive at a final estimate of $f(\cdot)$ by choosing a Fourier series estimate $\hat{f}_{0,m}(\cdot)$ which satisfies $m \in \{1,2,\ldots,50\}$ and

$$\hat{J}(\hat{f}_{0,m}) \leq \hat{J}(\hat{f}_{0,k}), \quad \text{for } k = 1,2,\ldots,50.$$

Note that 50 is certainly not too large a truncation point to consider in this case because of the extremely large size of the sample.

The minimum value of $\hat{J}(\hat{f}_{0,k})$ (for $k = 1,2,\ldots,50$) occurs at $k = 42$. The estimate $\hat{f}_{0,42}(\cdot)$ is plotted in Figure 6.3 and nine of the thirteen bumps of Good and Gaskins are identified (using their numbering scheme). Our much simpler analysis seems to have arrived at essentially the same results as those of Good and Gaskins, although collaboration with a subject matter expert would be essential to correctly interpret differences in the estimates.

As a final observation concerning the LRL data we point out the similarity of $\hat{f}_{2,0}(\cdot)$ and $\hat{f}_{0,42}(\cdot)$, which is striking when one considers that $\hat{f}_{2,0}(\cdot)$ is based only on $\hat{\phi}(1)$ and $\hat{\phi}(2)$. The paucity of parameters required for $\hat{f}_{2,0}(\cdot)$ to correctly describe the major features of the LRL data becomes important in smaller samples. This fact will be illustrated in Section 6.4.

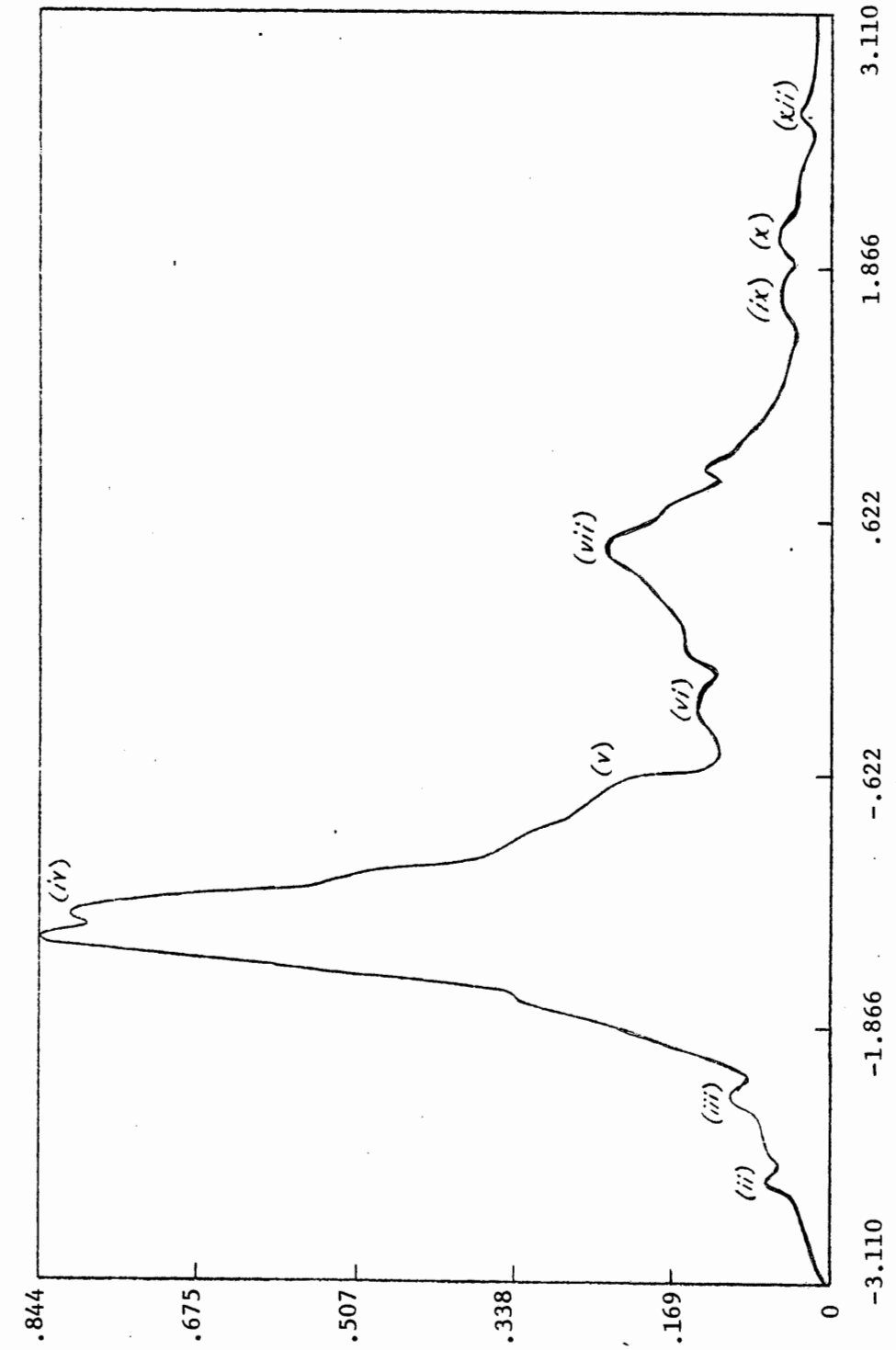


FIGURE 6.3
The Estimate $\hat{f}_{0,42}(\cdot)$ for the LRL Data

## 6.3 The Maguire Data

The data set to be analyzed in this section appears in Carmichael (1976) and has been studied by Maguire, Pearson and Wynn (1952), Boneva, Kendall and Stefanov (1971), and, in a density estimation context, by Carmichael (1976). The data, which we shall refer to as the Maguire data, consists of 109 "time intervals in days between explosions in mines involving more than ten men killed, from December 6, 1875 to May 29, 1951." For our purposes, the 109 values will be regarded as independent realizations of a random variable whose density function we wish to estimate.

An initial look at a histogram of the Maguire data indicates the possibility of an underlying exponential type density. Therefore, since Fourier series approximation methods work best for functions whose tails are similar, we have employed what Carmichael refers to as symmetrization. Symmetrization entails transforming the original data to the interval $[0,\pi]$, and then estimating the density $f^*(\cdot)$ of the transformed data by first estimating

$$f(x) = \frac{1}{2}f^*(|x|), \quad x \in [-\pi,\pi].$$

The evenness of $f(\cdot)$ implies that

$$\phi(v) = \int_{-\pi}^{\pi} (\cos vx - i\sin vx)\, f(x)\,dx$$

$$= 2\int_{0}^{\pi} \cos vx\, f(x)\,dx$$

$$= \int_{0}^{\pi} \cos vx\, f^*(x)\,dx, \quad |v| = 0,1,2,\ldots .$$

Given the previous expression for $\phi(v)$, we form estimates

$$\hat{\phi}(v) = \frac{1}{109} \sum_{j=1}^{109} \cos vx_j, \quad |v| = 0,1,2,\ldots ,$$

of $\phi(v)$, where $x_j$, $j = 1,\ldots,109$, is the Maguire data rescaled to the interval $[0,\pi]$. (The transformation $x_j = \frac{\pi}{1630} y_j$ was used since 1630 was the longest number of days between explosions.) Estimates

$$\hat{f}^*_{p,q}(x) = 2\hat{f}_{p,q}(x) , \quad x \in [0,\pi],$$

of the density $f^*(\cdot)$ may then be constructed, where

$$\hat{f}_{p,q}(x) = \frac{1}{2\pi} [1 + 2\text{Real}\{e_p(\hat{F}_q(x)) - \hat{F}_0(x)\}]$$

and

$$\hat{F}_j(x) = \sum_{v=k}^{j} \hat{\phi}(v) e^{ivx}.$$

A portion of the S-array for $\{(-1)^v \hat{\phi}(v)\}$ is given in Table 6.2. Note that, since in this case $\hat{\phi}(v)$ is real-valued, the S-array is also real-valued. Based on Table 6.2, $\hat{f}_{1,0}(\cdot)$ and $\hat{f}_{1,2}(\cdot)$ seem to be the best supported estimates of $f(\cdot)$, and certainly no estimate $\hat{f}_{p,q}(\cdot)$ for $p > 1$ is supported. For purposes of the MISE critierion of Chapter V, this pattern in the S-array suggests that, among estimates $\hat{f}_{p,q}(\cdot)$ ( $p > 0$), we limit consideration to estimates with $p = 1$.

In order to objectively choose an estimate of $f^*(\cdot)$, we shall calculate $\hat{J}(\hat{f}^*_{p,q})$ for each estimate in the class

$$A = \{\hat{f}^*_{p,q}(\cdot): p = 0,1 \text{ and } p + q \leq 20\}.$$

Since $\phi(v)$ is estimated differently here than in previous chapters, the estimate $\hat{J}(\hat{f}_{p,q})$ derived in Chapter V must be modified

TABLE 6.2

MAGUIRE DATA S-ARRAY FOR $\{(-1)^{v}\hat{\phi}(v)\}$

| m/n | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| -6 | -2.3442 | 4.1873 | 7.6413 | .6243 | 31.5796 | .5846 |
| -5 | -2.6173 | 2.2878 | -1.1070 | -24.0702 | 25.8274 | -3.2688 |
| -4 | -2.5935 | -1.9669 | 54.7387 | -27.1260 | 82.1131 | -4.2798 |
| -3 | -2.6195 | -57.7808 | 106.6615 | -13.2873 | -1.5643 | 6.9471 |
| -2 | -2.2463 | 2.3738 | 1.9639 | -4.6324 | -6.4719 | 4.8715 |
| -1 | -2.2317 | 81.9580 | 3.3973 | -9.9555 | -8.5664 | 50.7821 |
| 0 | -1.8119 | 1.8528 | -1.1989 | 1.0701 | -1.2228 | 1.2530 |
| 1 | -1.8024 | 30.1227 | - .7049 | 2.6439 | -1.0416 | -1.3733 |
| 2 | -1.6175 | 1.5569 | -1.5230 | 1.7872 | - .4512 | 6.5955 |
| 3 | -1.6275 | .6980 | 12.5592 | 4.2834 | -5.1103 | 3.7425 |
| 4 | -1.6183 | -12.1080 | 4.3602 | 3.4221 | 68.0605 | 2.8217 |
| 5 | -1.7439 | 4.6512 | -3.5795 | -4.6306 | 19.4732 | -.4310 |

slightly. We have

$$\text{MISE}(\hat{f}^*_{p,q}) = E[\int_0^\pi (\hat{f}^*_{p,q}(x) - f^*(x))^2 dx]$$

and

$$J(\hat{f}^*_{p,q}) = E\{\int_0^\pi [\hat{f}^*_{p,q}(x)]^2 dx - 2\int_0^\pi \hat{f}^*_{p,q}(x) dF^*(x)\}.$$

Proceeding as in Chapter V, it may be verified that the appropriate estimate of $J(\hat{f}^*_{p,q})$ is

$$\hat{J}(\hat{f}^*_{p,q}) = -\frac{1}{\pi}[1 + \frac{2(n+1)}{(n-1)}\sum_{v=1}^{p+q}\hat{\phi}^2(v)] + \frac{2}{\pi(n-1)}\sum_{v=1}^{p+q}(1+\hat{\phi}(2v)), p=0,$$

and

$$\hat{J}(\hat{f}^*_{p,q}) = \hat{J}(\hat{f}^*_{0,p+q}) + 4\int_0^\pi \hat{g}^2_{p,q}(x)dx - \frac{4}{n}\sum_{j=1}^{n}\hat{g}_{p,q,x_{(j)}}(x_j)$$

$$+ 8\int_0^\pi \hat{f}_{0,p+q}(x)\hat{g}_{p,q}(x)dx, \; p > 0 .$$

Table 6.3 contains the value of $\hat{J}(\hat{f}^*_{p,q})$ for each of the estimates in A, and shows the minimum of $\hat{J}(\hat{f}^*_{p,q})$ to occur at $\hat{f}^*_{0,12}(\cdot)$. This estimate is plotted in Figure 6.4, and, for the sake of comparison, the estimate $\hat{f}^*_{1,10}(\cdot)$ (at which $\hat{J}(\hat{f}^*_{1,k})$ is minimized) is plotted in Figure 6.5. The estimate $\hat{f}^*_{1,10}(\cdot)$ is seen to be smoother in the tail than is $\hat{f}^*_{0,12}(\cdot)$, a feature which we noted to be a characteristic of ARMA approximators. Whether or not the extra smoothing done by $\hat{f}^*_{1,10}(\cdot)$ is warranted might best be judged by someone knowledgeable with the physical situation which generated this data.

An interesting aspect of Table 6.3 is the magnitude of $\hat{J}(\hat{f}^*_{1,0})$ relative to the minimum value of $\hat{J}(\hat{f}^*_{p,q})$. A comparison of these two numbers confirms that a low frequency component is

TABLE 6.3

VALUES OF $\hat{J}(\hat{f}^*_{p,q})$ FOR THE MAGUIRE DATA

| k | $\hat{J}(\hat{f}^*_{0,k})$ | $\hat{J}(\hat{f}^*_{1,k-1})$ |
|---|---|---|
| 1 | -.7360 | -1.1907 |
| 2 | -1.0037 | -1.1939 |
| 3 | -1.1021 | -1.1807 |
| 4 | -1.1363 | -1.1762 |
| 5 | -1.1449 | -1.1687 |
| 6 | -1.1468 | -1.1828 |
| 7 | -1.1502 | - .9564 |
| 8 | -1.1695 | -1.1538 |
| 9 | -1.1805 | -1.2262 |
| 10 | -1.2075 | -1.1659 |
| 11 | -1.2179 | -1.2266 |
| 12 | -1.2291 | .2571 |
| 13 | -1.2250 | -1.2215 |
| 14 | -1.2205 | -1.1965 |
| 15 | -1.2141 | -1.2155 |
| 16 | -1.2085 | -1.2176 |
| 17 | -1.2039 | -1.2222 |
| 18 | -1.2057 | -1.2138 |
| 19 | -1.2091 | -1.1927 |
| 20 | -1.2106 | -1.1490 |

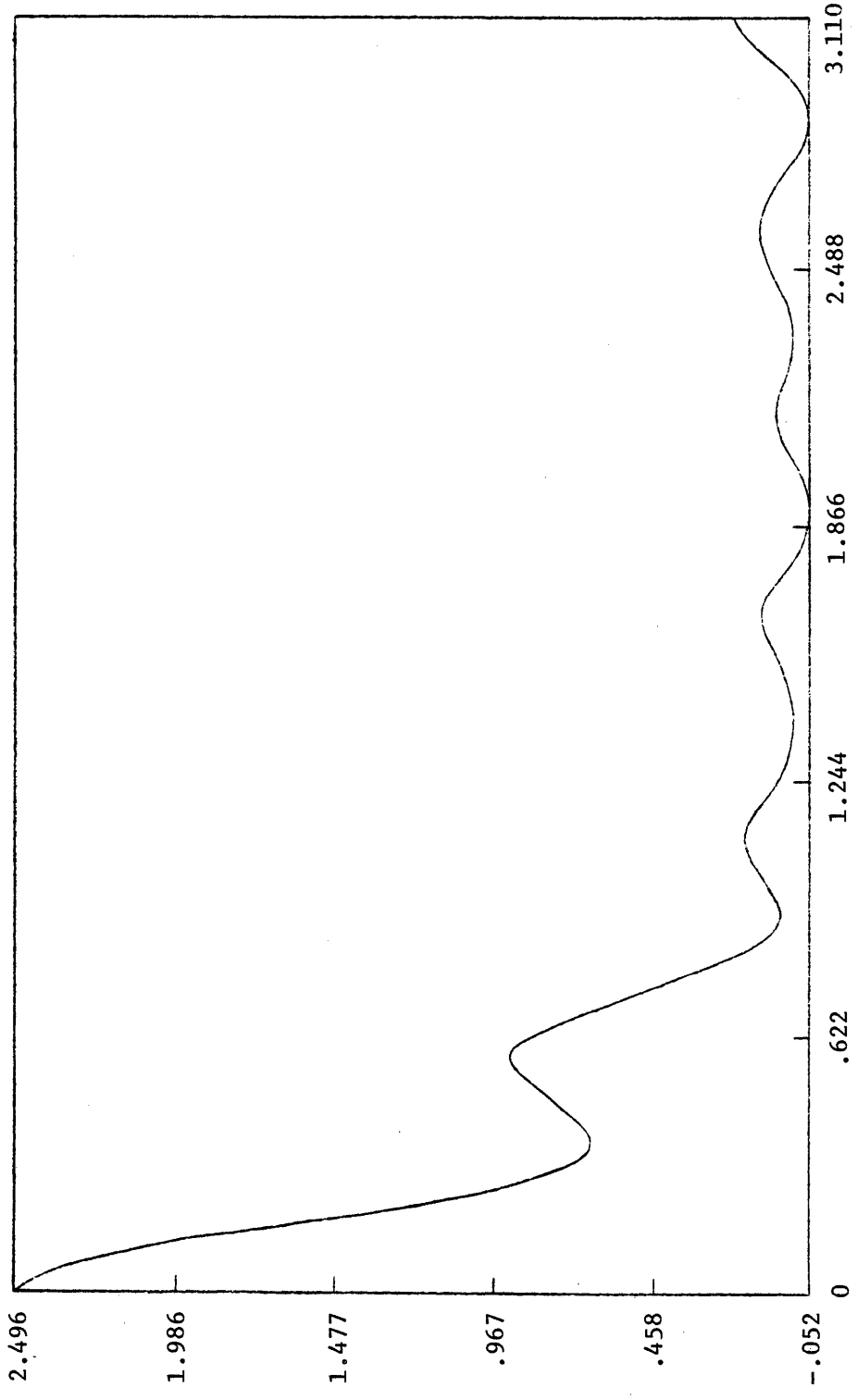


FIGURE 6.4

The Estimate $\hat{f}^{*}_{0,12}(\cdot)$ for the Maguire Data

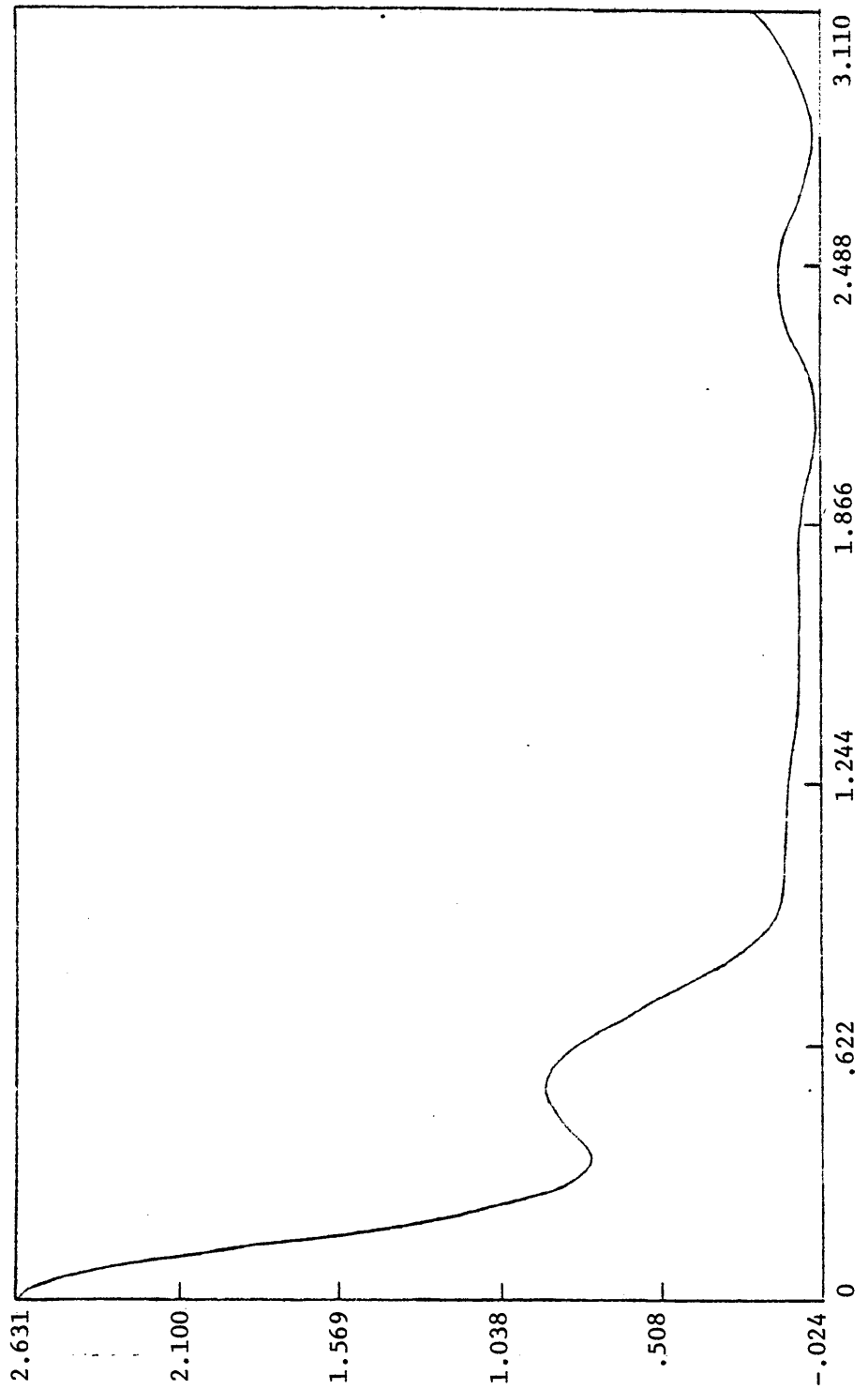


FIGURE 6.5

The Estimate $\hat{f}^*_{1,10}(\cdot)$ for the Maguire Data

the dominant feature of the Maguire data, a fact which was predicted by the first column of the S-array.

A final comment about this data set concerns the ability of $\hat{J}(\hat{f}^*_{p,q})$ to distinguish between differing estimates. Figure 6.6 shows a plot of the estimate $\hat{f}^*_{1,6}(\cdot)$, which has the third largest value of $\hat{J}(\hat{f}^*_{p,q})$ in Table 6.3. Note that $\hat{f}^*_{1,6}(\cdot)$ does not have its maximum at zero, which is true of only one of the other estimates considered. The other estimate of which this is true, $\hat{f}^*_{1,11}(\cdot)$, maximizes $\hat{J}(\hat{f}^*_{p,q})$ and is such that $\hat{f}^*_{1,11}(0) \simeq -11.87$. This is evidence that the criterion $\hat{J}(\hat{f}^*_{p,q})$ is able to identify the poorer estimates of $f^*(\cdot)$.

## 6.4 A Simulation Study

The purpose of this section is to investigate

(i) the effectiveness of the MISE criterion for selecting (p,q) in distinguishing between estimates which have important differences in ISE, and

(ii) the possible savings in ISE which may be attained by using ARMA, rather than Fourier series, density estimation.

In order to accomplish the above, simulations (which will hereafter be referred to as Simulations 1, 2, and 3) involving three different density functions have been carried out. A description of these simulations follows.

Simulation 1. In this study, 25 independent random samples, each of size 100, were generated from the Beta (12,3) distribution using the IMSL subroutine GGBTR. The data

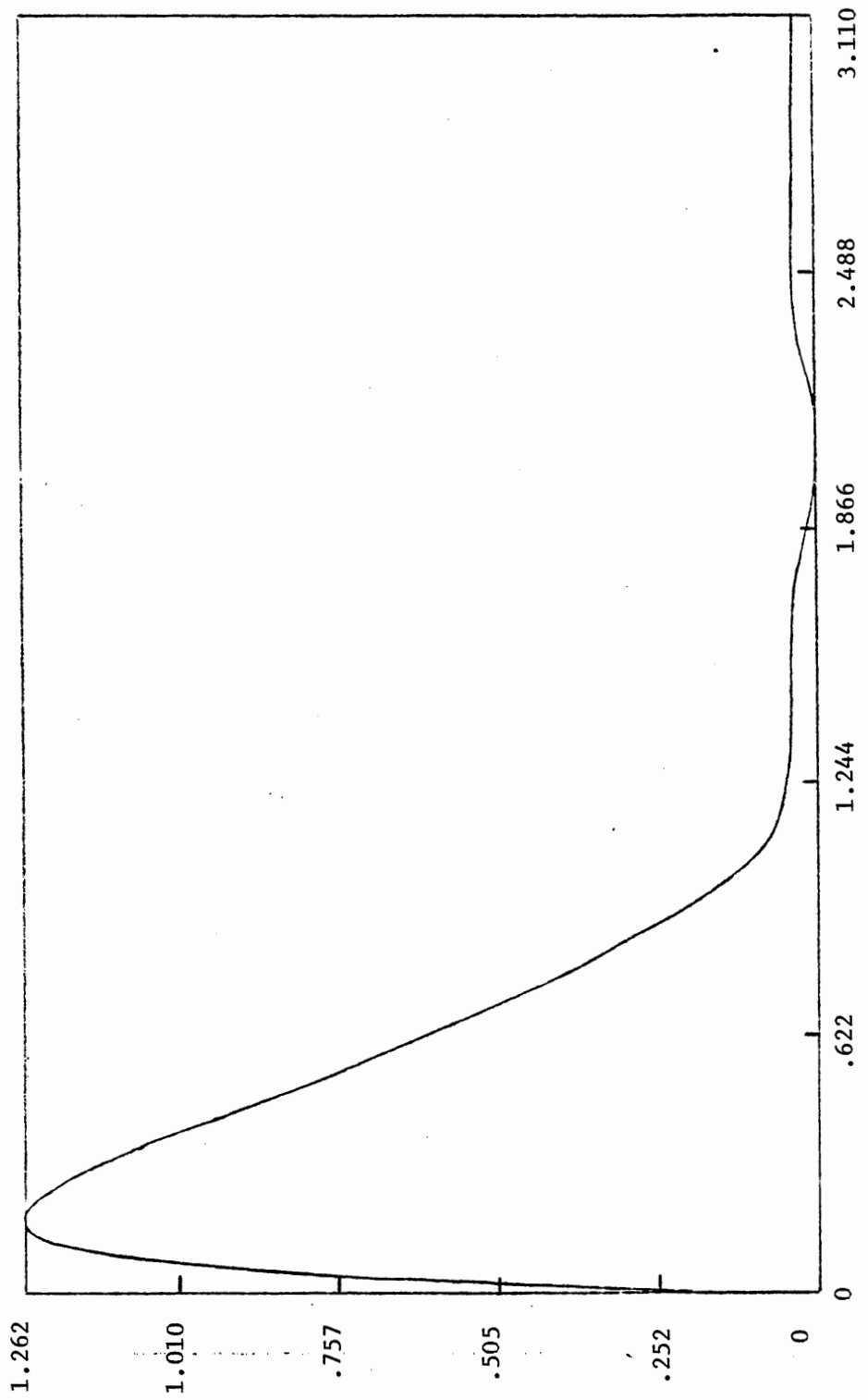


FIGURE 6.6

The Estimate $\hat{f}^*_{1,6}(\cdot)$ for the Maguire Data

$y_i$, $i = 1,2,\ldots,100$, in each sample was transformed to $x_i = \pi(2y_i - 1)$, $i = 1,2,\ldots,100$ (so that the estimated density, $f'(\cdot)$, is the Beta (12,3) density shifted and rescaled to fill $[-\pi,\pi]$). Then, for each sample, $\hat{J}(\hat{f}'_{p,q})$ was calculated for each estimate in the class

$$A_{11} = \{\hat{f}'_{p,q}(\cdot): p + q \leq 10\};$$

and $A_{11}$, $A_{12}$, and $A_{13}$ ARMA estimates were identified, where

$$A_{12} = \{\hat{f}'_{0,q}(\cdot): q = 1,2,\ldots,10\}, \text{ and}$$

$$A_{13} = \{\hat{f}'_{p,0}(\cdot): p = 1,2,\ldots,10\}.$$

Finally, for each of these three estimates, $\hat{f}'_{\hat{p},\hat{q}}(\cdot)$ say,

$$ISE(\hat{f}'_{\hat{p},\hat{q}}) = \int_{-\pi}^{\pi} (\hat{f}'_{\hat{p},\hat{q}}(x) - f'(x))^2 dx$$

was evaluated.

Simulation 2. In this simulation, 25 independent random samples of size 50 were generated from the density function

$$f''(x) = \frac{2e^{-4|x|}}{(1-e^{-4\pi})} I_{[-\pi,\pi]}(x),$$

a truncated Laplace density. This was done by first generating samples from the U(0,1) distribution (using IMSL subroutine GGUBS), and then employing the probability integral transformation. For each sample, $\hat{J}(\hat{f}''_{p,q})$ was calculated for each estimate in

$$A_{21} = \{\hat{f}''_{p,q}(\cdot): p = 0,1 \text{ and } p + q \leq 10\},$$

and $A_{21}$, $A_{22}$, and $A_{23}$ ARMA estimates were identified, where

$$A_{22} = \{\hat{f}''_{0,q}(\cdot): q = 1,2,\ldots,10\}$$

and

$$A_{23} = A_{22} \cup \{\hat{f}''_{1,q}(\cdot): q = 0,1,\ldots,4\} .$$

For each of these three estimates, $ISE(\hat{f}''_{p,q})$ was evaluated.

Simulation 3. In this final simulation, 15 independent random samples of size 100 were generated from the density (pictured in Figure 6.7)

$$f'''(x) = \frac{1}{2\pi}[1 + 2\text{Real}\{e_2(F_4(x))\}]\ I_{[-\pi,\pi]}(x),$$

where

$$F_4(x) = \sum_{v=1}^{4} \hat{\phi}(v)e^{ivx} \text{ and } \{\hat{\phi}(v)\}$$

is the sequence of estimated Fourier coefficients from the LRL data. The samples were generated as in Simulation 2, although in this case values of the inverse cdf had to be evaluated numerically. For each sample, $\hat{J}(\hat{f}'''_{p,q})$ was calculated for each estimate in

$$A_{31} = \{\hat{f}'''_{p,q}(\cdot): p = 0,1,2 \text{ and } p + q \leq 10\},$$

and the $A_{31}$ and $A_{32}$ ARMA estimates were identified, where

$$A_{32} = \{\hat{f}'''_{0,q}(\cdot): q = 1,2,\ldots,10\} .$$

(Reasons for considering the classes of estimates defined in this and the previous simulation are discussed below.) Finally, $ISE(\hat{f}'''_{p,q})$ was calculated for these two estimates.

The results of Simulations 1,2, and 3 are summarized in Tables 6.4 - 6.6. In order to define some descriptive measures

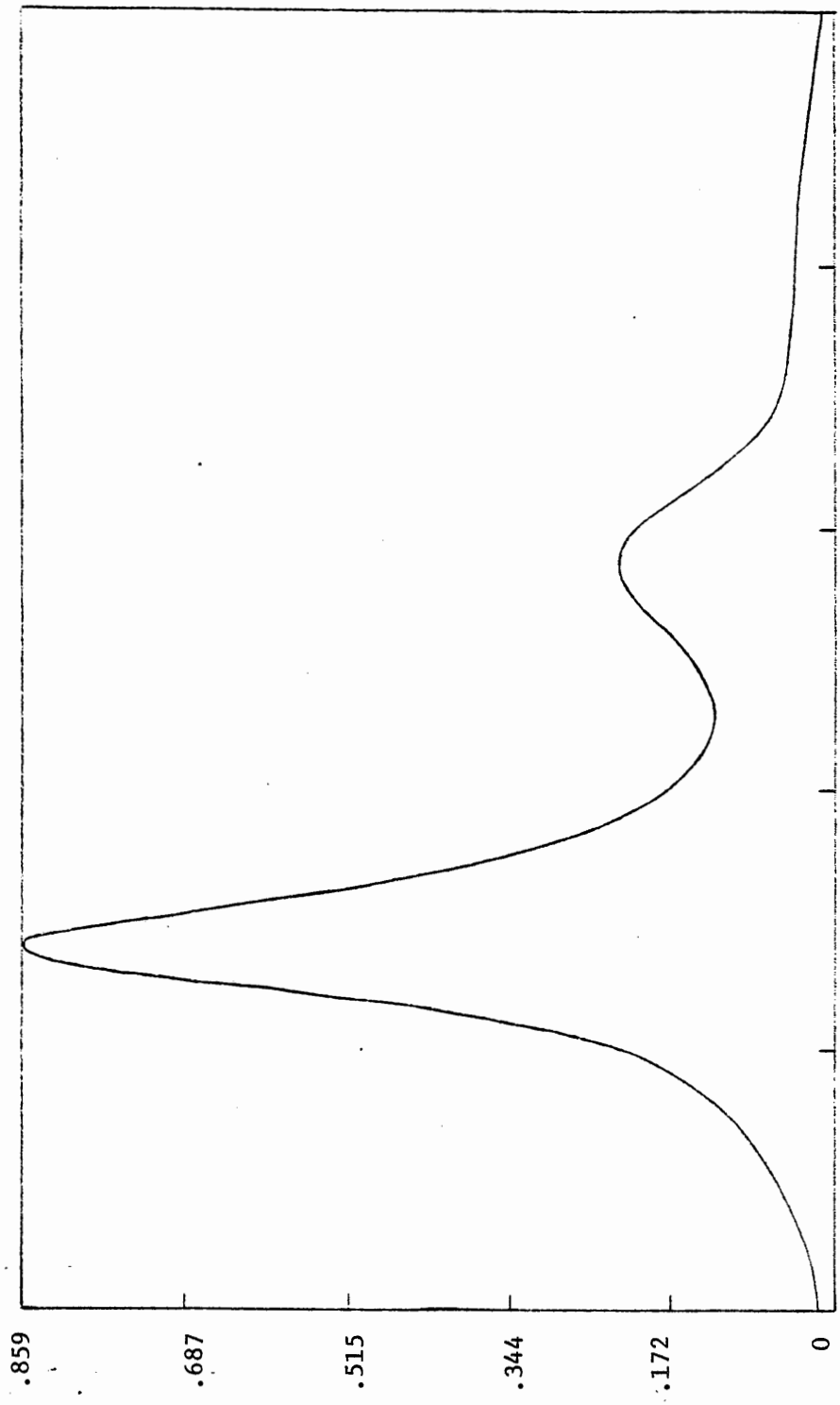


FIGURE 6.7
The Density Function f'''(·)

which appear in Table 6.4, let $\hat{f}_{ijk}(\cdot)$ be the $A_{ij}$ ARMA estimate in the kth repetition of Simulation i. Then

$$\overline{ISE}_{ij} = \frac{1}{m_i} \sum_{k=1}^{m_i} ISE(\hat{f}_{ijk})$$

and

$$[SD(ISE)_{ij}]^2 = \frac{1}{m_i - 1} \sum_{k=1}^{m_i} [ISE(\hat{f}_{ijk}) - \overline{ISE}_{ij}]^2,$$

where $m_1 = m_2 = 25$ and $m_3 = 15$. The quantity $F_{ij}$ is the proportion of repetitions in the ith simulation for which $\hat{f}_{ijk}(\cdot) \equiv \hat{f}_{i2k}(\cdot)$, i.e., $F_{ij}$ is the proportion of cases in which the $A_{ij}$ ARMA estimate is a Fourier series estimate. Finally, for $j \neq 2$, let

$$y_{ijk} = \ell n \left[ \frac{ISE(\hat{f}_{ijk})}{ISE(\hat{f}_{i2k})} \right], \quad k \in S_{ij},$$

where $S_{ij}$ is the set of $m_i(1-F_{ij})$ repetition indices k for which $ISE(\hat{f}_{ijk}) \neq ISE(\hat{f}_{i2k})$. Then, based on the Wilcoxon signed-rank statistic for the $y_{ijk}$, the Hodges-Lehmann estimate of the median of the $y_{ijk}$ distribution is denoted by $\hat{\theta}_{ij}$. Hence, the quantity $\hat{R}_{ij} = e^{\hat{\theta}_{ij}}$ is an estimate of the median, $R_{ij}$, of the conditional distribution of

$$\frac{ISE(\hat{f}_{ijk})}{ISE(\hat{f}_{i2k})}$$

given that $\dfrac{ISE(\hat{f}_{ijk})}{ISE(\hat{f}_{i2k})} \neq 1.$

Table 6.5 contains the results of three tests of hypothesis which address the question of whether or not a

TABLE 6.4

DESCRIPTIVE STATISTICS FOR SIMULATIONS 1, 2 AND 3

| | Type of Estimate | ISE | SD(ISE) | F | $\hat{R}$ |
|---|---|---|---|---|---|
| Simulation 1 | $A_{11}$(ARMA) | .01645 (.01261) | .02209 (.01117) | .12 | .76716 |
| | $A_{12}$(Fourier series) | .01644 (.01588) | .01358 (.01358) | 1.0 | --- |
| | $A_{13}$(autoregressive) | .26521 (.26073) | .10495 (.10473) | 0.0 | 19.56 |
| Simulation 2 | $A_{21}$ (ARMA) | .07513 | .05059 | .32 | .74923 |
| | $A_{22}$ (Fourier series) | .07589 | .03790 | 1.0 | --- |
| | $A_{23}$ (restricted ARMA) | .06654 | .05244 | .36 | .56839 |
| Simulation 3 | $A_{31}$ (ARMA) | .02008 | .00926 | .27 | .56867 |
| | $A_{32}$ (Fourier series) | .02831 | .01175 | 1.0 | --- |

The numbers in parentheses were calculated after deleting one repetition for which the ISE of the $A_{11}$ estimate was an outlier.

TABLE 6.5

RESULTS OF TESTS OF THE HYPOTHESES

$H_{0i}$ vs. $H_{1i}$ (i = 1,2,3)

| Hypothesis | Sample Size | P value | 95% C.I. |
|---|---|---|---|
| $H_{01}$ vs. $H_{11}$ | 22 | $.0527 < P_{11} < .0570$ | $.546 < R_{11} < 1.074$ |
| $H_{02}$ vs $H_{12}$ | 16 | $P_{23} < .005$ | $.414 < R_{23} < .793$ |
| $H_{03}$ vs $H_{13}$ | 11 | $P_{31} < .005$ | $.366 < R_{31} < .887$ |

savings in ISE results from using ARMA density estimation rather than Fourier series density estimation. The three hypotheses tested are

$$H_{01}: R_{11} = 1 \quad \text{vs.} \quad H_{11}: R_{11} < 1,$$

$$H_{02}: R_{23} = 1 \quad \text{vs.} \quad H_{12}: R_{23} < 1,$$

and

$$H_{03}: R_{31} = 1 \quad \text{vs.} \quad H_{13}: R_{31} < 1.$$

The test statistic for $H_{0i}$ vs $H_{1i}$ is the Wilcoxon signed-rank statistic

$$w_{ij} = \sum_{k \in S_{ij}} r(|y_{ijk}|) I_{(0,\infty)}(y_{ijk}),$$

where $r(\cdot)$ denotes rank. The reason for the use of the log transformation is that, since typically the distribution of ISE is skewed,

$$\ell n\left[\frac{ISE(\hat{f}_{ijk})}{ISE(\hat{f}_{i2k})}\right] = \ell n[ISE(\hat{f}_{ijk})] - \ell n[ISE(\hat{f}_{i2k})]$$

is more nearly symmetrically distributed about zero under $H_{0i}$ than is $ISE(\hat{f}_{ijk}) - ISE(\hat{f}_{i2k})$ about its mean. This is an important consideration since the Wilcoxon test is based on an assumption of symmetry. The results of the above tests are indicated in Table 6.5 by P values which are defined by $P_{ij} = P(W^{+} \leq w_{ij})$ where $W^{+}$ is a random variable having the distribution of a Wilcoxon signed-rank statistic. In addition, 95% confidence intervals for the parameters $R_{ij}$ are given.

The results in Table 6.5 address the second of the two

considerations which were the intent of our investigation at the beginning of this section. These results are strong evidence that a savings in ISE is realized if the MISE criterion of Chapter V is allowed to choose from a class of ARMA ($p>0$, $q > 0$) estimates rather than a class containing only Fourier series estimates. It is important to note, though, that the results obtained are conditional on the particular densities considered, the sample sizes used, and the classes of ARMA estimates chosen for consideration. Perhaps the most important of these three points is the choice of a class of estimates. The estimates $\hat{R}_{21}$ and $\hat{R}_{23}$ in Table 6.4 indicate that the number of ARMA estimates in the chosen class can be an important consideration. Further, it is not clear how the results of Tables 6.4 and 6.5 would have been affected if the classes $A_{21}$, $A_{23}$, and $A_{31}$ had included estimates with larger values of p. The restriction of the size of $A_{21}$ and $A_{23}$ was motivated by the fact that, in some initial repetitions of Simulation 2 (previous to those upon which Tables 6.4 and 6.5 are based), the S-array for $\{(-1)^{v}\hat{\phi}(v)\}$ showed a good constancy pattern in column 1. A similar statement is true regarding $A_{31}$ and Simulation 3, in which case constancy was apparent in the first two columns of a typical S-array.

The first of the two points which were to be investigated in this section concerned the ability of the MISE criterion to distinguish between estimates having important differences in ISE. Evidence of this ability is given in Table 6.6 . In

(1) The class of Fourier series and autoregressive density estimators is a subclass of the class of ARMA estimators.

(2) The ARMA estimator $\hat{f}_{p,q}(\cdot)$ estimates an approximator $f_{p,q}(\cdot)$ which was shown to be related to the $e_n$-transform. This relationship implies that $f_{p,q}(x)$ is often a better approximation to $f(x)$ than is $f_{0,p+q}(x)$, a Fourier series approximator.

(3) The estimator $\hat{f}_{p,q}(\cdot)$ may be expressed in terms of a quantity which was shown to be an adaptive, generalized jackknife statistic.

(4) The mixture of densities having autoregressive representations is, in general, a density which has an ARMA representation. This result implies that ARMA representations often require fewer parameters to adequately fit a density than do autoregressive representations.

(5) In a probability sense, the estimator $\hat{f}_{1,q}(\cdot)$ possesses (under certain conditions) a more rapid convergence property analogous to that possessed by $e_1$ in the deterministic setting.

(6) Two solutions were proposed to the problem of selecting an appropriate estimate from the class of ARMA estimates. One solution utilizes the S-array, and in the other solution an estimator is sought which will minimize $\mathrm{MISE}(\hat{f}_{p,q})$.

(7) Simulation studies indicate that (for the densities considered) a savings in ISE results from allowing the MISE criterion to choose from a class of ARMA ($p > 0$, $q > 0$) estimates rather than from a class of only Fourier series estimates.

TABLE 6.6

THE ABILITY OF THE MISE CRITERION TO DISTINGUISH ESTIMATES WITH DIFFERENT VALUES OF ISE

| Type of Estimate | P' | P" |
|---|---|---|
| $A_{11}$ | 88.00 | 63.64 |
| $A_{13}$ | 100.00 | 100.00 |
| $A_{21}$ | 68.00 | 58.82 |
| $A_{23}$ | 64.00 | 81.25 |
| $A_{31}$ | 73.33 | 90.91 |

Notes:

1. P' is the percentage of cases in which $\hat{J}(\hat{f}_{ijk}) \neq \hat{J}(\hat{f}_{i2k})$.
2. Among the cases satisfying $\hat{J}(\hat{f}_{ijk}) \neq \hat{J}(\hat{f}_{i2k})$, P" is the percentage of cases in which the sign of $\hat{J}(\hat{f}_{ijk}) - \hat{J}(\hat{f}_{i2k})$ and $ISE(\hat{f}_{ijk}) - ISE(\hat{f}_{i2k})$ are the same.

addition, it is noted that for all three simulations the estimates which had the larger values of $\hat{J}(\hat{f}_{p,q})$ were consistently among the poorer (in terms of ISE) estimates.

The final remarks to be made in this section concern Simulation 3. In Section 6.2 it was observed that the first two estimated Fourier coefficients of the LRL data contained essentially all the information about the main features of that data set, a fact that is not detected by Fourier series estimates, $\hat{f}_{0,q}(\cdot)$. With this in mind, one of the aims of Simulation 3 was to illustrate that, if moderate sized samples were generated from a density like that of the LRL data, a parsimonious ARMA estimate would be preferred to a Fourier series estimate. That this is the case is evidenced by Table 6.5 and the average number of Fourier coefficients, $N(A_{3j})$, used by the $A_{3j}$ ARMA estimates of Simulation 3. We have

$$N(A_{31}) = 4 \quad \text{and} \quad N(A_{32}) = 5.47,$$

and thus the considerable savings in ISE obtained using the $A_{31}$ ARMA estimate occurred even though the ARMA estimates were, on the average, based on fewer fitted parameters than the Fourier series estimates.

## 6.5 A Summary

A new class of estimators of a probability density function, referred to as the class of ARMA estimators, has been introduced in this work. The principal results obtained concerning this class of estimators may be summarized as follows.

Although some important results have been obtained in this work, there remain numerous topics for future research in ARMA density estimation. Some of these topics, such as the establishment of more general large sample properties, the routine choice of a class of estimates, and further investigation of the problem of selecting p and q, have been alluded to previously. However, perhaps the most important area for future research is a large-scale comparison of ARMA density estimation to other common methods of density estimation. Even though new, different methods of viewing an old problem are of value, it is probably desirable to be somewhat economic with regard to the number of new methods proposed. For this reason, before being recommended for widespread use each new method should be validated against existing methods. A part of this validation for ARMA density estimation has been accomplished in this work, and the results thus far obtained indicate the possibility of a valuable new method.

# LIST OF REFERENCES


Apostol, Tom M. (1973). Mathematical Analysis, Addison-Wesley.

Boneva, L. I., Kendall D., and Stefanov, I. (1971). "Spline transformations: three new diagnostic aids for the statistical data-analyst," Journal of the Royal Statistical Society Series B, 33, 1-37.

Carmichael, J.P. (1976). "The autoregressive method: a method of approximating and estimating positive functions," Ph.D. dissertation, Statistical Science Division, State University of New York at Buffalo.

Cencov, N.N. (1962). "Evaluation of an unknown distribution density from observations," Soviet Mathematics, 3, 1559-1562.

Duin, R.P.W. (1976). "On the choice of smoothing parameters for Parzen estimators of probability density functions," IEEE Transactions on Computers, C-25, 1175-1179.

Efron, Bradley (1979). "Bootstrap methods: another look at the jackknife," Annals of Statistics, 7, 1-26.

Good, I.J. and Gaskins, R.A. (1971). "Nonparametric roughness penalties for probability densities," Biometrika, 58, 255-277.

Good, I.J. and Gaskins, R.A. (1980). "Density estimation and bump-hunting by the penalized likelihood method exemplified by scattering and meteroite data," Journal of the American Statistical Association, 75, 42-56.

Gray, H. L., Houston, A.G., and Morgan, F.W. (1978). "On G-spectral estimation," in Proceedings of the 1976 Tulsa Symposium on Applied Time Series, New York: Academic Press.

Gray, H. L., Kelley, G.D., and McIntire, D.D. (1978). "A new approach to ARMA modeling," Communications in Statistics, B7, 1-78.

Gray, H. L. and Schucany, W. R. (1972). The Generalized Jackknife Statistic, New York: Marcel Dekker, Inc.

Hermans, J. and Habbema, J.D.F. (1976). Manual for the ALLOC Discriminant Analysis Programs, University of Leiden, Department of Medical Statistics.

Kronmal, R.A. and Tarter, M.E. (1968). "The estimation of probability densities and cumulative by Fourier series methods," Journal of the American Statistical Association, 63, 925-952.

Maguire, B.A., Pearson, E.S., and Wynn, A.H.A. (1952). "The time intervals between industrial accidents," Biometrika, 39, 168-180.

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER<br>156 | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE *(and Subtitle)*<br>On ARMA Probability Density Estimation | | 5. TYPE OF REPORT & PERIOD COVERED<br>TECHNICAL REPORT |
| | | 6. PERFORMING ORG. REPORT NUMBER<br>156 |
| 7. AUTHOR(s)<br>Jeffrey D. Hart and Henry L. Gray | | 8. CONTRACT OR GRANT NUMBER(s)<br>N00014-75-C-0439 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>Southern Methodist University<br>Dallas, Texas 75275 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS |
| 11. CONTROLLING OFFICE NAME AND ADDRESS<br>Office of Naval Research<br>Arlington, VA 22217 | | 12. REPORT DATE<br>December 1981<br>13. NUMBER OF PAGES<br>125 |
| 14. MONITORING AGENCY NAME & ADDRESS*(If different from Controlling Office)* | | 15. SECURITY CLASS. *(of this report)*<br>15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT *(of this Report)*

This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted for any purposes of the United States Government.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Probability density estimation; Kronmal-Tarter method; Autoregressive estimators; Generalized jackknife

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)* A new method of probability density estimation is investigated which exploits the Fourier series representation of a density function. The new method employs density estimators $\hat{f}_{p,q}(\cdot)$, $p=0,1,2,\ldots$ and $q = 0,1,2,\ldots$, which are such that $\hat{f}_{0,q}(\cdot)$ is a Fourier series (Kronmal-Tarter type) estimator and $\hat{f}_{p,0}(\cdot)$ is an autoregressive estimator. Each of the estimators $\hat{f}_{p,q}(\cdot)$ (referred to as ARMA estimators) is shown to depend upon the $e_n$-transform, thus providing a strong motivation for the use of estimators with both $p > 0$ and

$q > 0$. Small and large sample properties of ARMA density estimators are obtained and a data-based method of selecting optimal values p and q is proposed. The results of a simulation study show that, for the densities considered, a savings in integrated square error is attained by using ARMA, rather than Fourier series, density estimation.