ARMA MODELS, PREWHITENING, AND MINIMUM CROSS ENTROPY

by

Dimitris N. Politis

Purdue University

Technical Report #91-55

Department of Statistics
Purdue University

September 1991

ARMA Models, Prewhitening, and Minimum Cross Entropy

Dimitris N. Politis

Department of Statistics, Purdue University
West Lafayette, IN 47907-1399

Abstract

The problem of spectral estimation on the basis of observations from a finite stretch of a stationary time series is considered, in connection with knowledge of a prior estimate of the spectral density. In general, the data are not exactly compatible with the prior. For example, the first p sample autocovariances might be significantly different from the the first p Fourier coefficients of the prior spectral density.

A reasonable 'posterior' spectral density estimate would be the density that is closest to the prior according to some measure of divergence, while at the same time being compatible with the data. The cross entropy (relative entropy, Kullback-Leibler number) has often been proposed in the past to serve as such a measure of divergence.

A connection of the original Minimum Cross Entropy Spectral Analysis method to traditional prewhitening techniques and to ARMA models is pointed out. In view of this connection, a fast approximate solution of the Minimum Cross Entropy problem is also proposed. The solution is in a standard multiplicative form, that is, the posterior is equal to the prior multiplied by a 'correction' factor.

I. Introduction

Suppose $\{X_n, n \in \mathbf{Z}\}$ is a Gaussian stationary stochastic process with mean zero, and autocovariances $\gamma(k) = EX_tX_{t+k} = \int_{-\pi}^{\pi} f(w)\cos(wk)dw$, for $k \in \mathbf{Z}$, where $f(w)$ is the spectral density function. The entropy rate of the process $\{X_n, n \in \mathbf{Z}\}$ is given by (cf. [20])

$$H(f) = \frac{1}{2}\log(2\pi e) + \frac{1}{4\pi}\int_{-\pi}^{\pi}\log f(w)dw \tag{1}$$

It is well known (cf. Burg [4]) that the process with maximum entropy rate among Gaussian stationary processes satisfying the constraints $\gamma(i) = c_i, i = 0, 1, \ldots, p$ is the mean zero autoregressive AR(p) Gaussian process that satisfies these constraints. Actually, the same AR(p) Gaussian process is found to possess maximum entropy rate among all processes satisfying the same constraints [10].

The maximum entropy property of the AR(p) Gaussian process gave rise to Burg's Maximum Entropy Method (MEM) for spectral estimation which goes as follows. Observations $X_1, \ldots, X_N$ are recorded, and reliable estimates of the first p autocovariances ($p << N$) are formed by the standard formula

$$\hat{\gamma}(k) = \frac{1}{N}\sum_{i=1}^{N-k} X_iX_{i+k} \tag{2}$$

for $k = 0, 1, \ldots, p$; the unbiased version of (2) could also be used. The MEM spectral density estimate $\hat{f}(w)$ is then given by

$$\hat{f}(w) = \frac{1}{|\sum_{t=0}^{p} a_t e^{iwt}|^2} \tag{3}$$

where $a_0, a_1, \ldots, a_p$ are the autoregressive coefficients of the Gaussian AR(p) process that has up to order p, autocovariances equal to the measured ones obtained in equation (2). The success of Burg's MEM method in practical applications is to a great extent owed to the existence of a fast algorithm for the computation of $\hat{f}(w)$, which is basically a variant of the Durbin-Levinson algorithm (cf. [2], [20]).

Related to the Maximum Entropy Method is Shore's [23] Minimum Cross Entropy method of Spectral Analysis (MCESA), which was proposed in order to be able to incorporate a prior estimate $\phi(w)$ of $f(w)$ into the analysis. Shore's method hinges on the fact that (cf. [21])

the cross entropy rate (relative entropy, Kullback-Leibler divergence, Itakura-Saito distortion) between two stationary Gaussian processes $\{X_n, n \in \mathbf{Z}\}$ and $\{Y_n, n \in \mathbf{Z}\}$, with respective spectral densities $f(w)$ and $\phi(w)$ such that the ratio $f(w)/\phi(w)$ is bounded above, is equal to

$$D(f||\phi) = \frac{1}{4\pi} \int_{-\pi}^{\pi} [\frac{f(w)}{\phi(w)} - 1 - \log \frac{f(w)}{\phi(w)}] dw \tag{4}$$

The practical application of MCESA parallels that of MEM, except that the additional prior estimate $\phi(w)$ is considered known. Again observations $X_1, \ldots, X_N$ are recorded, and estimates of the first p autocovariances are calculated using (2). Then the MCESA 'posterior' estimate of $f(w)$ is the function that minimizes the cross entropy rate $D(f||\phi)$, subject to satisfying $\int_{-\pi}^{\pi} f(w) \cos(wk) dw = \hat{\gamma}(k)$, for $k = 0, 1, \ldots, p$. The MCESA estimate is given by (cf. [23])

$$\bar{f}(w) = \frac{\phi(w)}{1 + u(w)\phi(w)} \tag{5}$$

where $u(w) = \sum_{t=0}^{p} \lambda_t \cos wt$, and the constants $\lambda_t, t = 0, 1, \ldots, p$ should be chosen such that $\int_{-\pi}^{\pi} \bar{f}(w) \cos(wk) dw = \hat{\gamma}(k)$, for $k = 0, 1, \ldots, p$. The actual computation of the λ_t is rather difficult, involving some kind of iterative nonlinear numerical procedure (cf. [7], [23]).

It is worth noting that if the prior $\phi(w)$ is flat, or if it is of the form $\phi(w) = 1/|\sum_{t=0}^{q} b_t e^{iwt}|^2$, with $q \leq p$, then the MCESA solution coincides with the MEM solution. In other words, the MCESA procedure 'sees' a low order AR prior spectrum as flat.

More recently [24] another form of Minimum Cross Entropy Spectral Estimation (MCE2) was also proposed (see also [16]), based on minimizing directly the Kullback-Leibler divergence

$$K(f||\phi) = \int_{-\pi}^{\pi} f(w) \log \frac{f(w)}{\phi(w)} dw \tag{6}$$

between the normalized (to have total area equal to one) spectral densities $f(w)$ and $\phi(w)$. The minimization is carried out subject to the constraints $\int_{-\pi}^{\pi} f(w) \cos(wk) dw = \hat{\rho}(k)$, for $k = 0, 1, \ldots, p$, where $\hat{\rho}(k) = \hat{\gamma}(k)/\hat{\gamma}(0)$ are the sample autocorrelations. The MCE2 estimate is given by (cf. [24])

$$\bar{f}^{(2)}(w) = \phi(w) \exp\{-1 - \sum_{t=0}^{p} \lambda_t \cos wt\} \tag{7}$$

where the constants $\lambda_t, t = 0, 1, \ldots, p$ should be chosen such that $\int_{-\pi}^{\pi} \bar{f}^{(2)}(w) \cos(wk) dw = \hat{\gamma}(k)$, for $k = 0, 1, \ldots, p$.

Although the MCE2 original solution involved an iterative nonlinear algorithm, a different approach without nonlinear equations was suggested [6] that makes use of the cepstral representation. In addition, a resemblance of MCE2 with traditional prewhitening methods is apparent from the multiplicative form of equation (7), (see also [5]).

A closely related variant of MCE2 was suggested in [8]. The variant (which will be denoted as MCE3) involves finding the spectrum $f(w)$ that minimizes the functional

$$K^*(f||\phi) = \int_{-\pi}^{\pi} \phi(w) \log \frac{\phi(w)}{f(w)} dw \tag{8}$$

subject to the same autocorrelation constraints. The solution of MCE3 is given by [8]

$$\bar{f}^{(3)}(w) = \frac{\phi(w)}{|\sum_{t=0}^{p} \lambda_t e^{-iwt}|^2} \tag{9}$$

where the constants $\lambda_t, t = 0, 1, \ldots, p$ should be chosen such that $\int_{-\pi}^{\pi} \bar{f}^{(3)}(w) \cos(wk) dw = \hat{\rho}(k)$, for $k = 0, 1, \ldots, p$.

Note that although $K^*(f||\phi) = K(\phi||f)$, the MCE2 and MCE3 solutions are in general different. In classical statistical applications however (cf. [17]) the minimization of $K(f||\phi)$ was prefered, because of its immediate connection with the original maximum entropy method of Jaynes [15] (when the prior ϕ is flat), and the theory of maximum likelihood estimation.

In view of equation (9), the connection of MCE3 to traditional prewhitening methods, as well as to ARMA processes [8], is apparent. Cases where the MEM method yields a solution of ARMA type due to additional constraints (e.g. constraints on the cepstral coefficients, constraints on the impulse response coefficients, etc.) have also been presented [12], [14], [18], [22].

It is the purpose of this report to show the connection of MCESA to prewhitening methods and to ARMA processes. Having a spectral estimate of ARMA type is desirable because it additionally specifies a time-domain representation of the estimated process. In addition, a representation of MCESA using prewhitening yields a solution in a standard multiplicative form, that is, the posterior is equal to the prior multiplied by a 'correction' factor. This representation provides a formal analogy to Bayesian posteriors, and is intuitively appealing in view of the interpretation of certain minimum relative entropy solutions as the limits of conditional probabilities (cf. [9], [11], [25]).

As a by-product, a fast approximate solution of MCESA (the Prewhitened MCESA) will be proposed as an alternative to the nonlinear exact procedure. Some numerical examples will be given in order to illustrate the applicability of the methodology.

II. Minimum Cross Entropy Spectral Analysis and Prewhitening

The MCE2 and MCE3 solutions in equations (7) and (9) have the prior $\phi(w)$ entering in a simple multiplicative form as it happens in classical prewhitening procedures [3]. However, the MCESA solution as given in equation (5) does not possess this property. Shore [23] gave the following linear filtering interpretation of equation (5). A Gaussian process with spectral density $\phi(w)$ is passed through a linear filter with magnitude-squared transfer function equal to $\frac{1}{1+u(w)\phi(w)}$; the resulting output is the process with minimum cross entropy rate relative to the input.

Nevertheless, a connection of MCESA to traditional prewhitening methods *can* be established in view of equation (4). Note that $D(f||\phi) = D(g||1)$, where $g(w) = f(w)/\phi(w)$; in other words, the cross entropy rate between the two stationary Gaussian processes $\{X_n, n \in \mathbf{Z}\}$ and $\{Y_n, n \in \mathbf{Z}\}$, with respective spectral densities $f(w)$ and $\phi(w)$, is equal to the cross entropy rate between a process $\{Z_n, n \in \mathbf{Z}\}$ with spectral density $g(w)$ and a white noise process.

Based on this observation, the following prewhitening procedure is suggested. Suppose we are given $\phi(w)$ as a prior estimate of $f(w)$. Also suppose that $\phi(w)$ is a smooth function of w (with continuous first derivative) in order to be able to factor it as $\phi(w) = \psi(w)\psi(-w)$, (cf. [19]). Then pass the process $\{X_n, n \in \mathbf{Z}\}$ through a linear filter with transfer function $1/\psi(w)$. The output of the filter will be the process $\{Z_n, n \in \mathbf{Z}\}$ with spectral density $g(w) = f(w)/\phi(w)$, i.e. a 'whitened' version of $\{X_n\}$. In practical applications, the prior $\phi(w)$ would likely be of the rational (ARMA) type, i.e. of the form $\phi(w) = |\sum_{t=0}^{r} c_t e^{iwt}|^2 / |\sum_{t=0}^{q} b_t e^{iwt}|^2$. In that case, a factorization such that the whitening linear filter is causal can also be found.

Since $D(f||\phi) = D(g||1)$, the minimization of $D(f||\phi)$ is exactly equivalent to the minimization of the cross entropy rate between $\{Z_n\}$ and white noise. However, the original MCESA problem involves the minimization of $D(f||\phi)$ subject to constraints on the autocovariance of the $\{X_n\}$ process. This can not be immediately incorporated in the prewhitened MCESA procedure, and we propose to substitute the constraints on the autocovariance of the $\{X_n\}$ process with the corresponding constraints on the autocovariance of the 'whitened' $\{Z_n\}$. The solution to the proposed prewhitened MCESA procedure will then be an approximation to the solution of the original MCESA method. In the next section, it will be apparent that the prewhitened MCESA procedure possesses the additional advantage of being implementable by a fast algorithm, which in fact is identical to Burg's maximum entropy algorithm.

III. The Prewhitened Minimum Cross Entropy Spectral Analysis Procedure, Maximum Entropy, and ARMA Models

In the practical spectral estimation problem observations $X_1, \ldots, X_N$ are recorded, and the prior spectrum $\phi(w) = \psi(w)\psi(-w)$ is given. The observations $X_1, \ldots, X_N$ are then passed through the filter with transfer function $1/\psi(w)$, and the 'new' data $Z_1, \ldots, Z_N$ are obtained. Due to transients, the first few of the Z_t's can be thought as not representative of the true $\{Z_n\}$ process, which would be the result of passing the *whole* doubly-infinite sequence $\{X_n, n \in \mathbf{Z}\}$ through the linear filter with transfer function $1/\psi(w)$.

To make this precise, suppose that the prior can be put in the moving average (MA) form, i.e. for some positive integer r, $\phi(w) = |\sum_{t=0}^{r} h_t e^{iwt}|^2$. Then it is easy to see that Z_t can be expressed as $Z_t = \sum_{k=0}^{r} h_k X_{t-k}$; hence the observations $Z_1, \ldots, Z_r$ are not representative of the true $\{Z_n\}$ process, since for their exact calculation, the values of X_t for negative t's are required. So the first r of the Z_t's can be dropped, and the remaining $Z_{r+1}, Z_{r+2}, \ldots, Z_N$ observations are kept.

Even in the case the prior can not be put exactly in the moving average (MA) form, a trigonometric polynomial approximation is readily available. For example, it suffices that $\phi(w)$ is continuous in order to be able to find (cf. [2]) a positive integer r, and constants $h_0, h_1, \ldots, h_r$ such that

$$\sup_{w \in [-\pi, \pi]} |\phi(w) - |\sum_{t=0}^{r} h_t e^{iwt}|^2| < \epsilon$$

for any given $\epsilon > 0$. In this case too, observations $Z_{r+1}, Z_{r+2}, \ldots, Z_N$ are considered representative of the true $\{Z_n\}$ process and kept for further analysis.

Estimates of the first p autocovariances ($p << N$) of the $\{Z_n\}$ process can be computed as

$$\hat{\beta}(k) = \frac{1}{N-r} \sum_{i=r+1}^{N-k} Z_i Z_{i+k} \tag{10}$$

for $k = 0, 1, \ldots, p$, where $\beta(k) = E Z_1 Z_{1+k}$. The prewhitened MCESA estimate of $f(w)$ is then

$$\tilde{f}(w) = \phi(w)\hat{g}(w) \tag{11}$$

where $\hat{g}(w)$ is the minimizer of $D(g||1)$ subject to the constraints $\int_{-\pi}^{\pi} g(w)\cos(wk)dw = \hat{\beta}(k)$, for $k = 0, 1, \ldots, p$.

Observe however that the constraint for $k = 0$ reads $\int_{-\pi}^{\pi} g(w)dw = \hat{\beta}(0)$, i.e. the total area under g is fixed in the minimization. But from equations (1) and (4) it is obvious that $D(g||1) = -H(g) + \frac{1}{4\pi}\int_{-\pi}^{\pi} g(w)dw + constant$. Hence, because $\int_{-\pi}^{\pi} g(w)dw$ is fixed, minimization of $D(g||1)$ subject to the constraints is *equivalent* to maximization of the entropy $H(g)$.

In other words, $\hat{g}(w)$ is just the maximum entropy (MEM) autoregressive spectrum of the $\{Z_n\}$ process subject to the usual autocovariance constraints $\int_{-\pi}^{\pi} \hat{g}(w)\cos(wk)dw = \hat{\beta}(k)$, for $k = 0, 1, \ldots, p$. Consequently, the prewhitened MCESA estimate $\tilde{f}(w)$ can be obtained immediately after the MEM spectrum $\hat{g}(w)$ is computed using Burg's algorithm (or any other fast algorithm for autoregressive model fitting).

It is apparent that if the prior $\phi(w)$ is of rational (ARMA) type, the prewhitened MCESA estimate $\tilde{f}(w)$ is also of ARMA type, since $\hat{g}(w)$ is the spectral density of an AR(p) process. Hence the prewhitened MCESA procedure can be used for updating an originally estimated ARMA model in view of further data. It can also be used as a simple and fast way of fitting an ARMA model to the data. To do that, one can start by fitting a moving average (MA) model (a fast algorithm for MA model fitting can be found in [2]). The spectral density of the fitted MA model can then be used as the prior $\phi(w)$, and the prewhitened MCESA procedure can be applied to give a spectral density estimate $\tilde{f}(w)$ of ARMA type.

An additional feature of the estimator $\tilde{f}(w)$ is its strong consistency and asymptotic normality as the sample size N goes to infinity. This is the subject of the following theorem which is true under the common assumption that $\{X_n\}$ is a linear time series. The assumption is satisfied in many interesting cases, including the case where $f(w)$ is the spectral density of an ARMA process.

Theorem. *Suppose that $\{X_n, n \in \mathbf{Z}\}$ is a (not necessarily Gaussian) strictly stationary stochastic process with mean zero, and autocovariances $\gamma(k) = EX_tX_{t+k} = \int_{-\pi}^{\pi} f(w)\cos(wk)dw$, for $k \in \mathbf{Z}$, where $f(w)$ is the spectral density function.*

Also suppose that

(i) $\{X_n, n \in \mathbf{Z}\}$ is a linear time series, i.e. it satisfies $X_t = \sum_{k=0}^{\infty} d_k W_{t-k}$, where $\{W_n, n \in \mathbf{Z}\}$ is an i.i.d. sequence of random variables with finite fourth moments, and $d_0, d_1, \ldots$ is a sequence satisfying $d_0 = 1$ and $\sum_{t=0}^{\infty} |d_t| < \infty$;

(ii) the prior $\phi(w)$ *is of moving average (MA) form, i.e. for some positive integer* r*,* $\phi(w) = |\sum_{t=0}^{r} h_t e^{iwt}|^2$*;*

(iii) both $f(w)$ *and* $\phi(w)$ *are bounded away from zero, uniformly in* w*;*

(iv) the order $p = p_N \to \infty$ *as the sample size* $N \to \infty$*, but* $p_N = O(N^{1/3}/[\log N(\log\log N)^{1+\delta}])$*, for some* $\delta > 0$*;*

then it is true that as $N \to \infty$*,*

$$\sup_{w\in[-\pi,\pi]} |\tilde{f}(w) - f(w)| \to 0$$

with probability one.

If in addition it is assumed that

(v) as $N \to \infty$*,* $N^{1/2}\sum_{t=N}^{\infty} |d_t| \to 0$*;*

then it is also true that the joint asymptotic distribution of

$$\sqrt{N/p_N}(\tilde{f}(0)-f(0)), \sqrt{N/p_N}(\tilde{f}(w_1)-f(w_1)), \ldots, \sqrt{N/p_N}(\tilde{f}(w_n)-f(w_n)), \sqrt{N/p_N}(\tilde{f}(\pi)-f(\pi))$$

is independent, normal, mean zero, with respective variances

$$4f^2(0), 2f^2(w_1), \ldots, 2f^2(w_n), 4f^2(\pi)$$

for any points satisfying $0 < w_1 < \cdots < w_n < \pi$*.*

The proof amounts to verifying that the conditions of Theorem 5 in [13] and Theorem 6 in [1] are true as applied to the 'whitened' sequence $\{Z_{r+1}, Z_{r+2}, \ldots\}$. This however is immediate, since under conditions (i) and (ii), we can write $Z_t = \sum_{k=0}^{r} h_k X_{t-k}$, for $t = r+1, r+2, \ldots, N$. Hence the sequence $\{Z_{r+1}, Z_{r+2}, \ldots\}$ is strictly stationary with spectral density equal to $f(w)/\phi(w)$, and is also a linear time series, since it is obvious that Z_t, for $t > r$, can be expressed as $Z_t = \sum_{k=0}^{\infty} g_k W_{t-k}$, where the sequence $g_0, g_1, \ldots$ is absolutely summable.

It is interesting to note that the asymptotic distribution of $\tilde{f}(w)$ is identical to the asymptotic distribution of the standard (MEM) autoregressive estimate $\hat{f}(w)$. In other words, the availability or not of the prior $\phi(w)$ makes a difference in the spectral density estimator only in finite samples.

IV. Some numerical examples

A small simulation experiment was conducted in order to illustrate the applicability of the proposed prewhitened MCESA method. For the examples, two short time series (each of length 120) were generated through two parametric Gaussian ARMA models. The general ARMA model is of the form:

$$X_t + a_1 X_{t-1} + \ldots + a_p X_{t-p} = W_t + b_1 W_{t-1} + \ldots + b_q W_{t-q} \qquad (12)$$

where the sequence $\{W_n\}$ is a sequence of i.i.d. normal random variables.

Example a. For the first example, time series $X_t, t = 1, \ldots, 120$, was generated using $b_1 = -0.9$ and $b_2 = 0.81$, (and the other b's zero), and $a_1 = -1.352, a_2 = 1.338, a_3 = -0.662$, and $a_4 = 0.240$, (and the other a's zero). The true spectral density of $\{X_n\}$ is then given by

$$f_a(w) = \frac{1}{2\pi} \frac{|\sum_{t=0}^{2} b_t e^{iwt}|^2}{|\sum_{t=0}^{4} a_t e^{iwt}|^2}$$

(with $a_0 = 1 = b_0$), and is shown in Figure 0. In this and all subsequent spectral density plots, the interval $(0, 2\pi]$ was discretized; for example, Figure 0 actually shows a plot of the sequence $f_a(w_j), j = 1, 2, \ldots, 100$, where $w_j = 2\pi j/100$. Note that due to symmetry $f_a(0) = f_a(2\pi)$.

Three different prior estimates of the spectral density were considered. All three are of the form $|\sum_{t=0}^{2} b_t e^{iwt}|^2$, i.e. they represent spectral densities of moving average (MA) processes. More specifically, the three priors were

$$\phi_1(w) = |1 - 0.9e^{iw} + 0.81e^{2iw}|^2$$

$$\phi_2(w) = |1 - e^{iw} + e^{2iw}|^2$$

$$\phi_3(w) = |1 - 0.5e^{iw} + 0.2e^{2iw}|^2$$

and are pictured in Figures 1, 2, and 3 respectively. Note that $\phi_1(w)$ is the 'best' prior, since it exactly identifies the numerator of the true spectral density $f_a(w)$. Accordingly, $\phi_2(w)$ is a 'good' prior, and $\phi_3(w)$ is a 'bad' prior.

The observed time series $X_t, t = 1, \ldots, 120$, was prewhitened using each of the three priors. In each case the first 20 observations of the 'whitened' process were discarded to avoid the effects of transients. From the remaining 100 'whitened' observations, autocovariance estimates were calculated and an autoregressive (AR) model of order p was fit using the first p sample autocovariances. The choice of the order p was made using the AIC criterion [2], that is, autoregressive models of order $1, \ldots, 10$ were fit to the data, and the order p was chosen to minimize the AIC information criterion. In all cases considered, minimization of AIC also turned out to minimize the related BIC criterion.

The 'posterior' estimates of $f_a(w)$ corresponding to the three priors were found to be

$$f_a^{(1)}(w) = \frac{1}{2\pi}\phi_1(w)/|1 - 1.301e^{iw} + 1.077e^{2iw} - 0.383e^{3iw}|^2$$

$$f_a^{(2)}(w) = \frac{1}{2\pi}\phi_2(w)/|1 - 1.390e^{iw} + 1.313e^{2iw} - 0.410e^{3iw}|^2$$

$$f_a^{(3)}(w) = \frac{1}{2\pi}\phi_3(w)/|1 - 0.715e^{iw}|^2$$

The three posteriors are pictured in Figures 4, 5, and 6. It is apparent that posteriors $f_a^{(1)}(w)$ and $f_a^{(2)}(w)$ are relatively good estimates of $f_a(w)$, while $f_a^{(3)}(w)$ is not, missing both the troughs around the points $w = 0$ and $w = 1.07$ (that is, for w_j, with $j = 0$ and $j = 17$).

Example b. For the second example, time series $X_t, t = 1, \ldots, 120$, was generated using the same b coefficients in equation (12), i.e. $b_1 = -0.9$ and $b_2 = 0.81$, (and the other b's zero), but with $a_1 = -2.760, a_2 = 3.809, a_3 = -2.654$, and $a_4 = 0.924$, (and the other a's zero). The true spectral density of $\{X_n\}$ is again given by

$$f_b(w) = \frac{1}{2\pi}\frac{|\sum_{t=0}^{2} b_t e^{iwt}|^2}{|\sum_{t=0}^{4} a_t e^{iwt}|^2}$$

(with $a_0 = 1 = b_0$), and is shown in Figure 7.

Since the numerator of $f_b(w)$ is the same as that of $f_a(w)$, the same three priors $\phi_1(w), \phi_2(w)$, $\phi_3(w)$ are used for this example too. The prewhitened procedure was performed in the same manner as in the first example, and the 'posterior' estimates of $f_b(w)$ were found to be

$$f_b^{(1)}(w) = \frac{1}{2\pi}\phi_1(w)/|1 - 2.344e^{iw} + 2.773e^{2iw} - 1.620e^{3iw} + 0.504e^{4iw}|^2$$

$$f_b^{(2)}(w) = \frac{1}{2\pi}\phi_2(w)/|1 - 1.253e^{iw} + 0.918e^{2iw}|^2$$

$$f_b^{(3)}(w) = \frac{1}{2\pi}\phi_3(w)/|1 - 1.980e^{iw} + 1.697e^{2iw} - 0.4713e^{3iw}|^2$$

The three posteriors are pictured in Figures 8, 9, and 10. Posterior $f_b^{(1)}(w)$ is a relatively accurate estimate of $f_b(w)$, although it does not have the resolution required to identify the two peaks in the graph of $f_b(w)$. Posterior $f_b^{(2)}(w)$ suffers from the same lack of resolution, with the added disadvantage that it puts more spectral 'mass' at the location of the smaller of the two peaks of $f_b(w)$. Lastly, posterior $f_b^{(3)}(w)$ completely misses the presence of the smaller of the two peaks of $f_b(w)$.

As a final note, some discussion on the use of the AIC (or BIC) criterion for the choice of p is in order. The minimization of AIC is proposed based on the principle of parsimony. In other words, to justify an increase of the order of the model (and the number of parameters to be estimated from the data), a significant improvement in the fit must ensue. In the autoregressive case, this means a significant reduction in the prediction error.

However, suppose one is not willing to abide by this principle. For our two examples, posterior estimates of the spectral density were also calculated using the 'good' prior $\phi_2(w)$, and a pre-chosen order of $p = 8$. The two posteriors are pictured in Figures 11 and 12. It is obvious that they are not patently any better than the posteriors calculated using the AIC criterion minimization. In particular, the posterior in Figure 11 shows some 'spurious' details not existing in the true spectrum, and the posterior in Figure 12 is almost identical to the posterior of Figure 10, where the 'bad' prior was used. □

V. Conclusions

A new method of updating a prior estimate of the spectral density of a stationary time series in light of further data was proposed. The method is termed the Prewhitened Minimum Cross Entropy Spectral Analysis (MCESA), and yields an approximate solution to the problem of finding a spectral density estimate that is closest (in the cross entropy measure of divergence) to the prior, while at the same time being compatible with the data. The Prewhitened MCESA method is an asymptotically consistent estimation method, and its implementation can be carried out by a fast algorithm, which is a variant of the Durbin-Levinson algorithm.

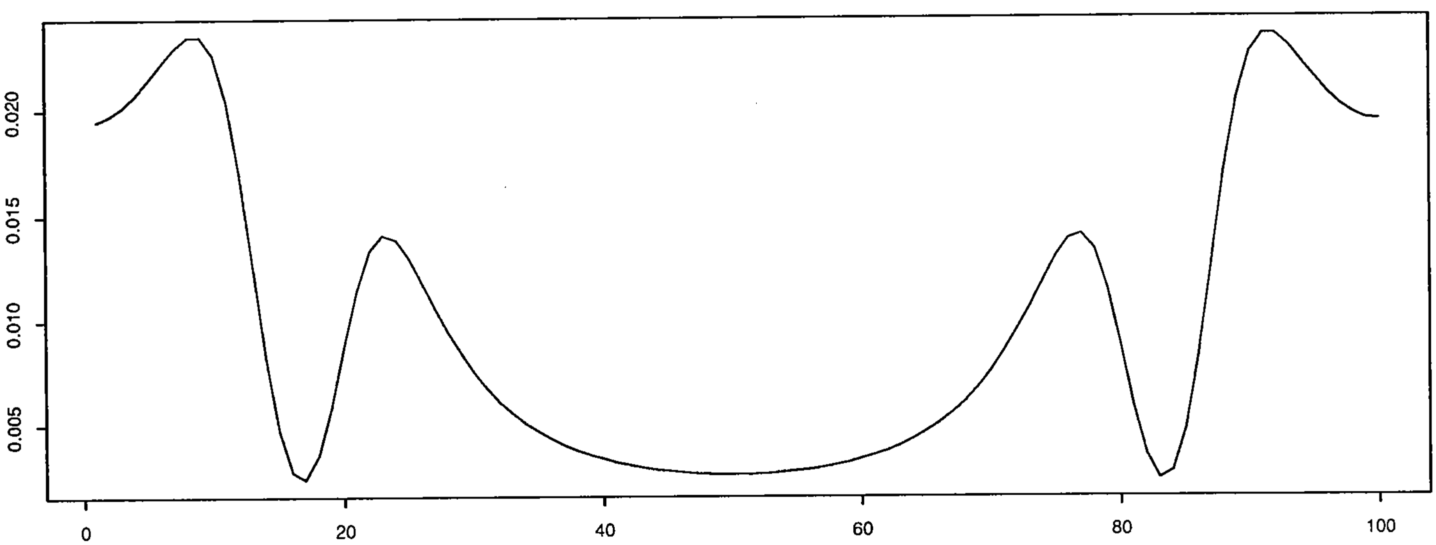

Figure 0. True spectral density in the first example

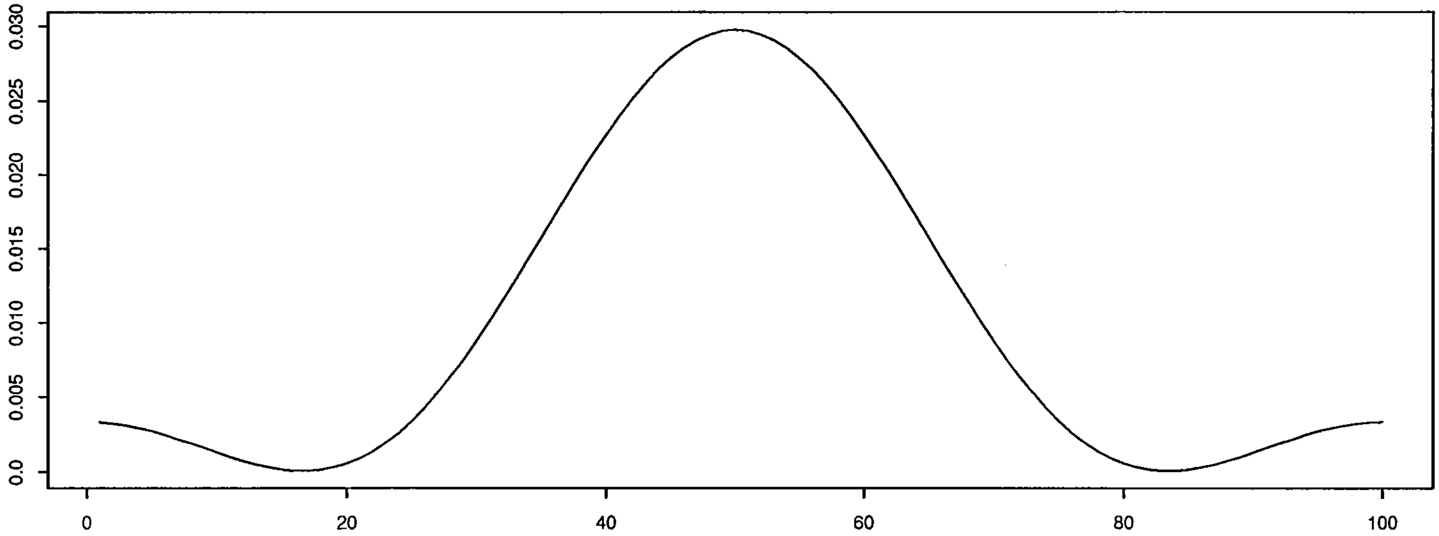

Figure 1. 'Best' prior spectral density

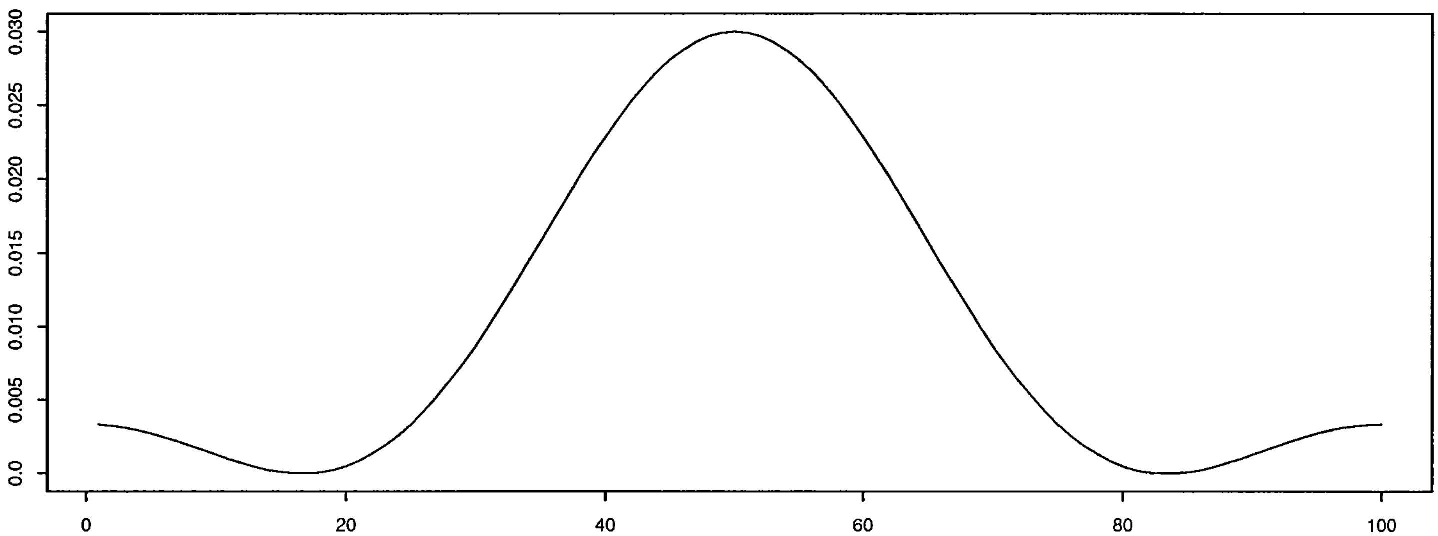

Figure 2. 'Good' prior spectral density

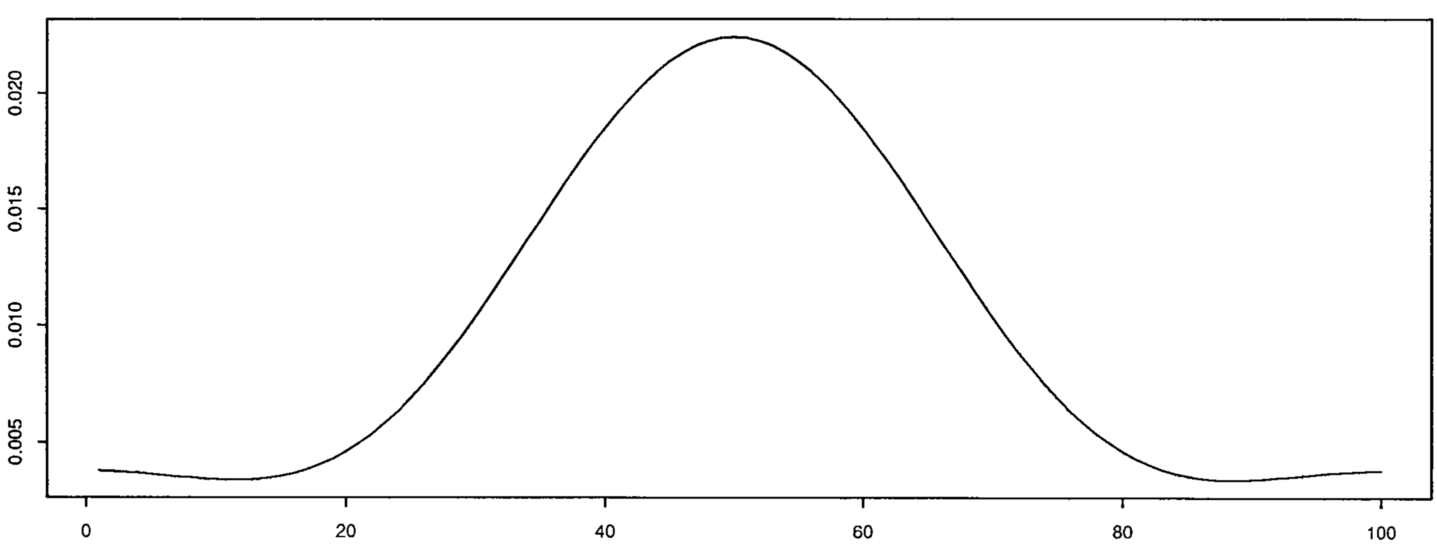

Figure 3. 'Bad' prior spectral density

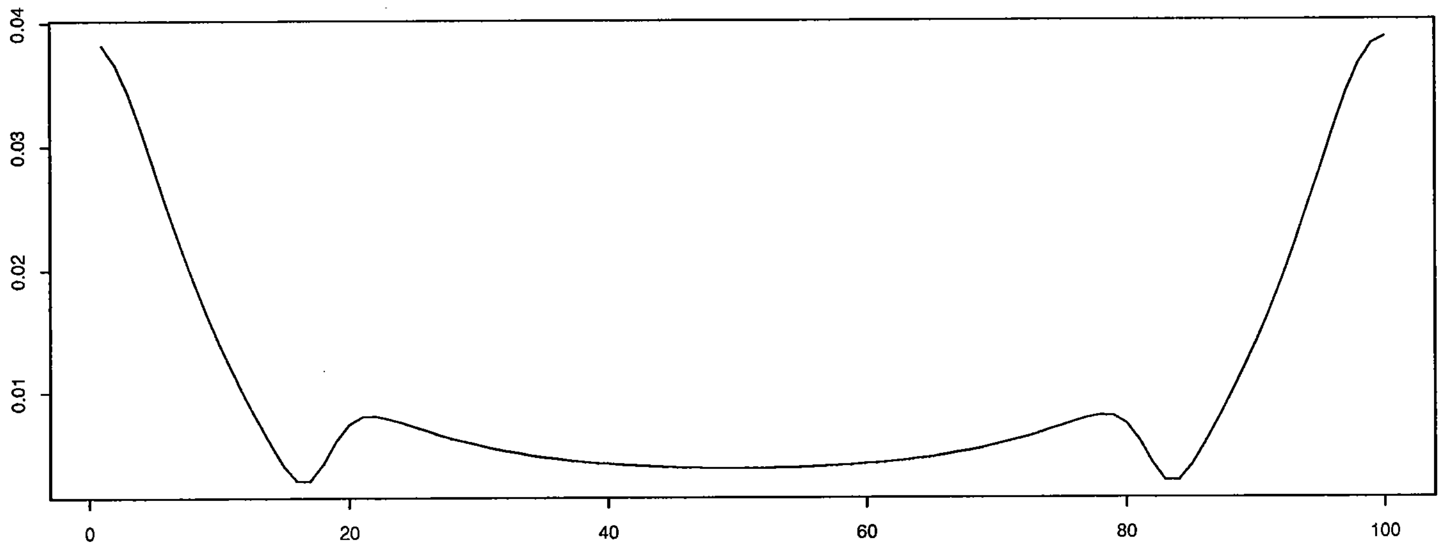

Figure 4. 'Posterior' estimate using the 'best' prior in the first example

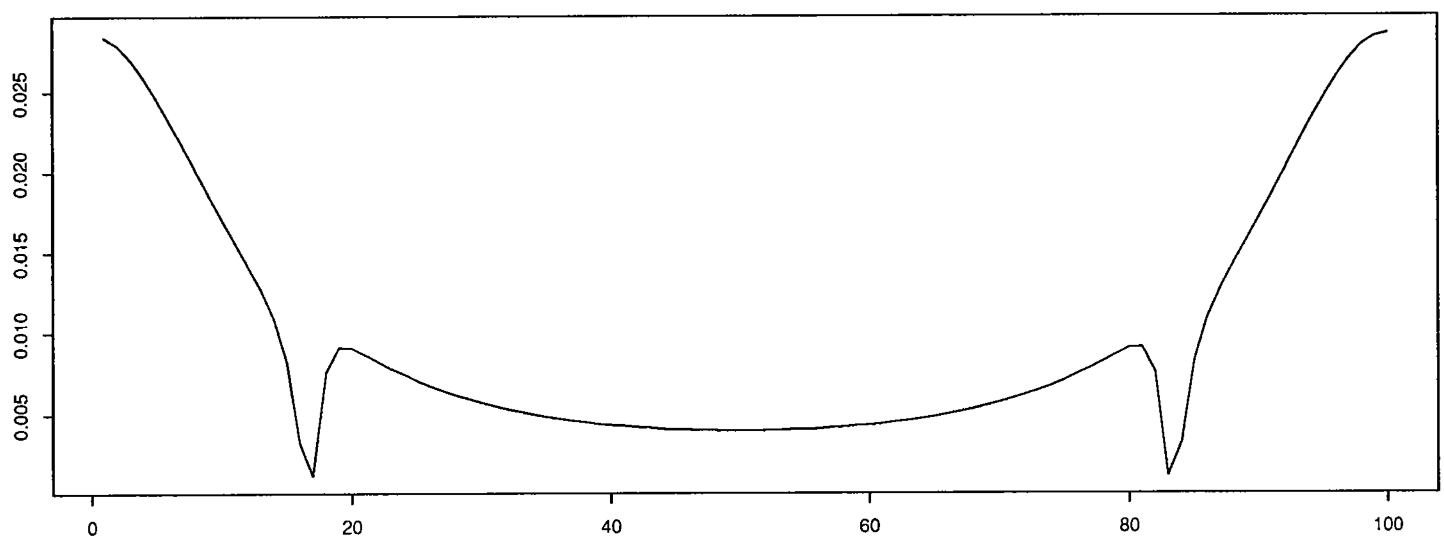

Figure 5. 'Posterior' estimate using the 'good' prior in the first example

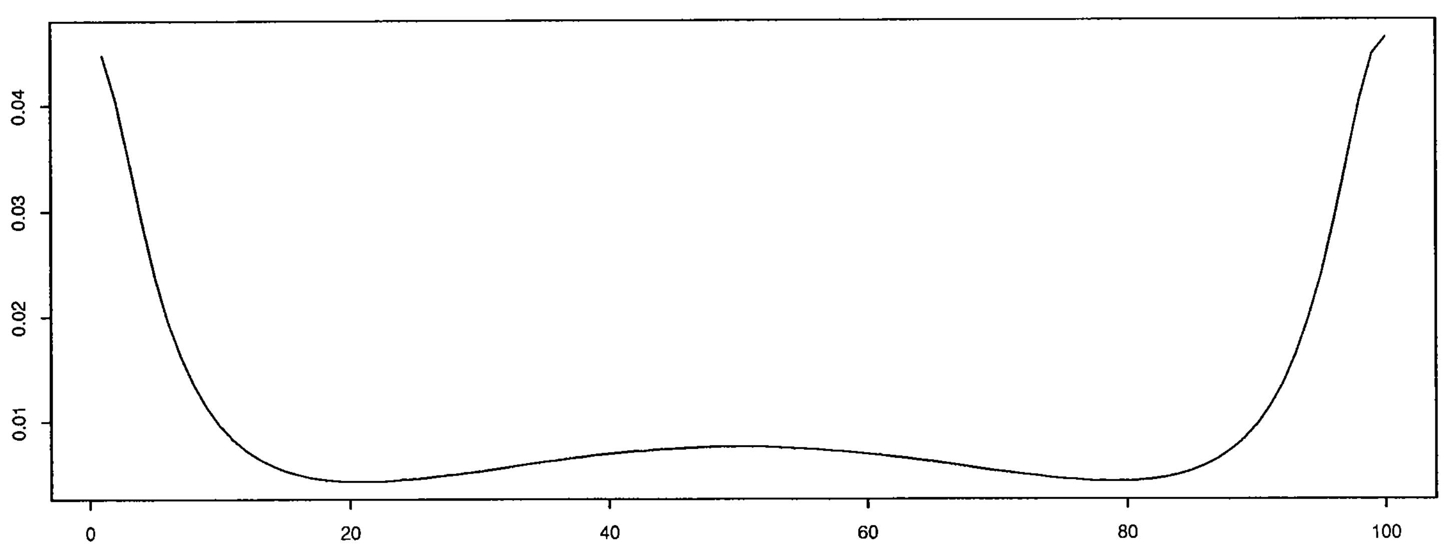

Figure 6. 'Posterior' estimate using the 'bad' prior in the first example

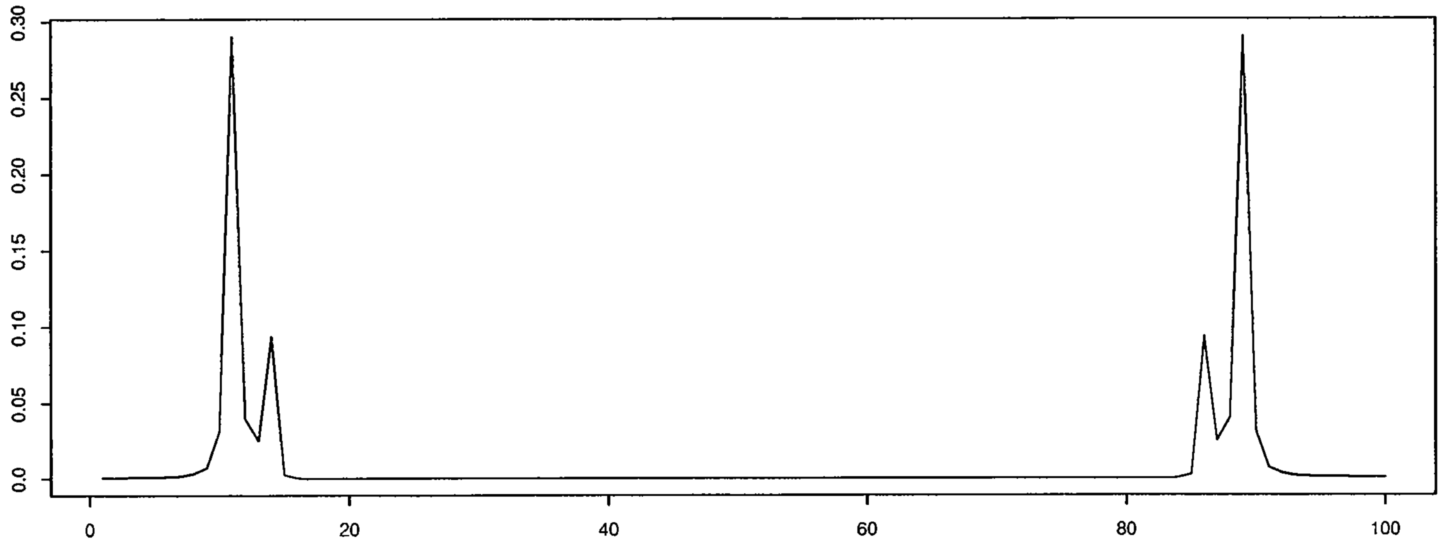

Figure 7. True spectral density in the second example

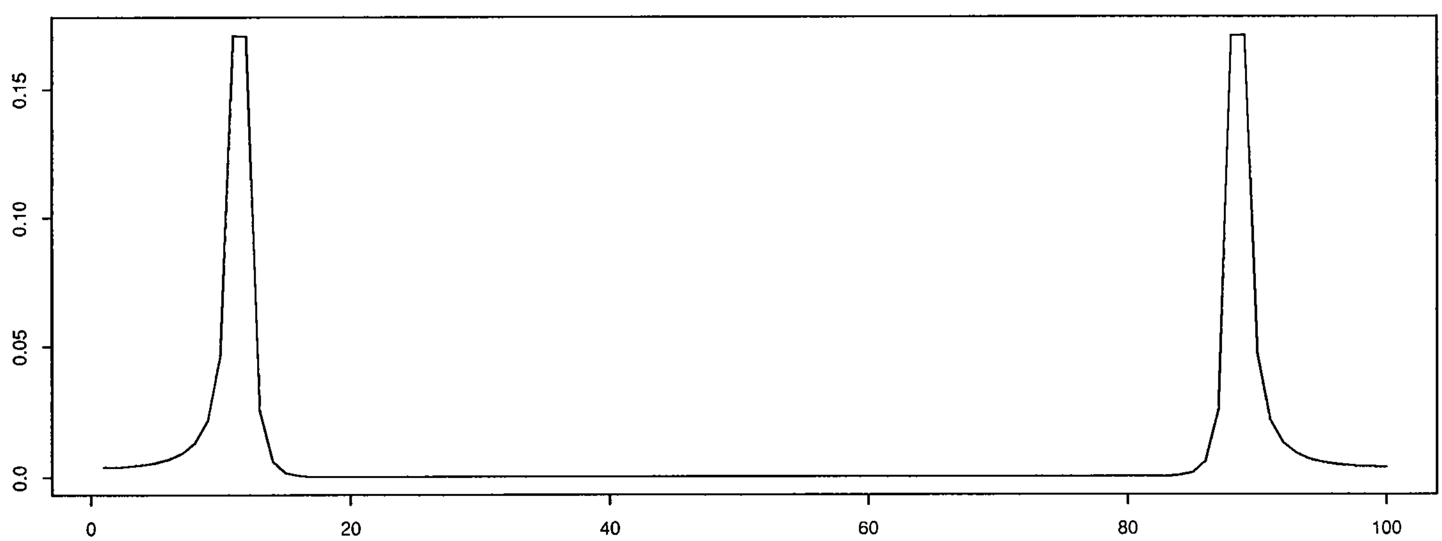

Figure 8. 'Posterior' estimate using the 'best' prior in the second example

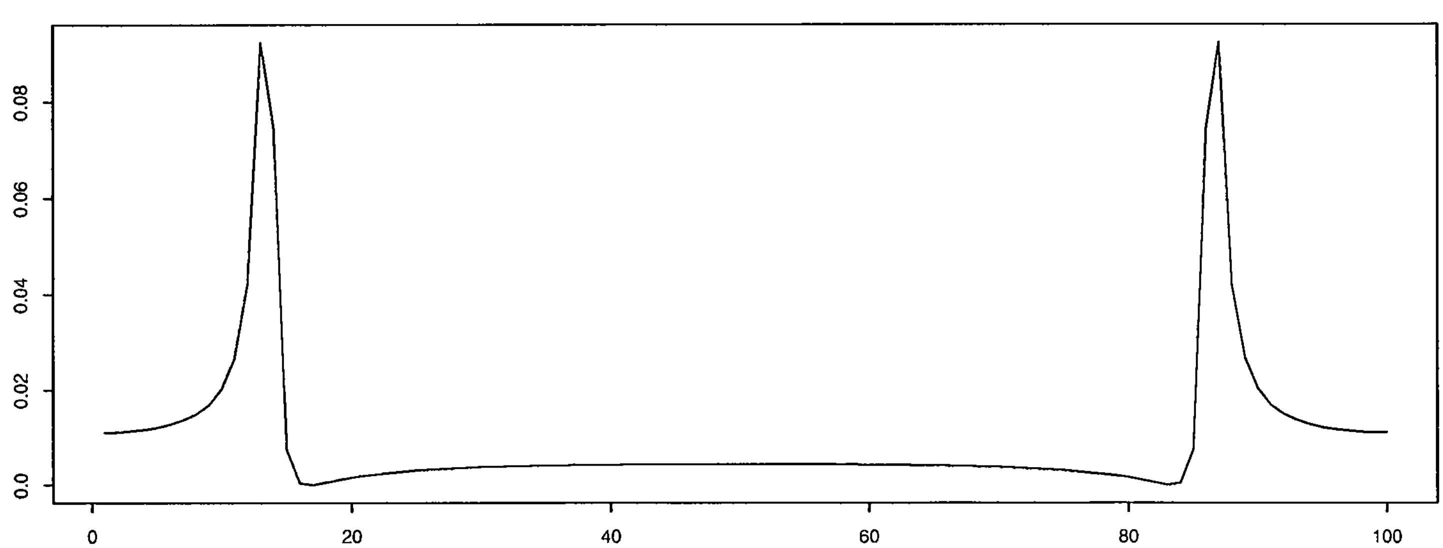

Figure 9. 'Posterior' estimate using the 'good' prior in the second example

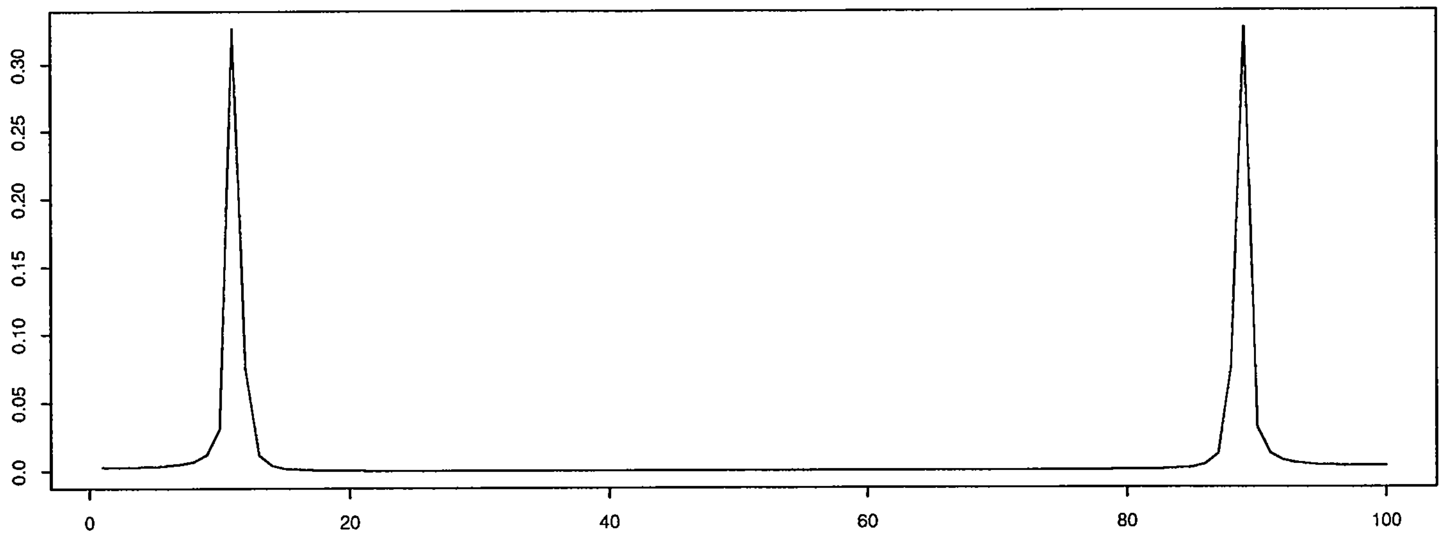

Figure 10. 'Posterior' estimate using the 'bad' prior in the second example

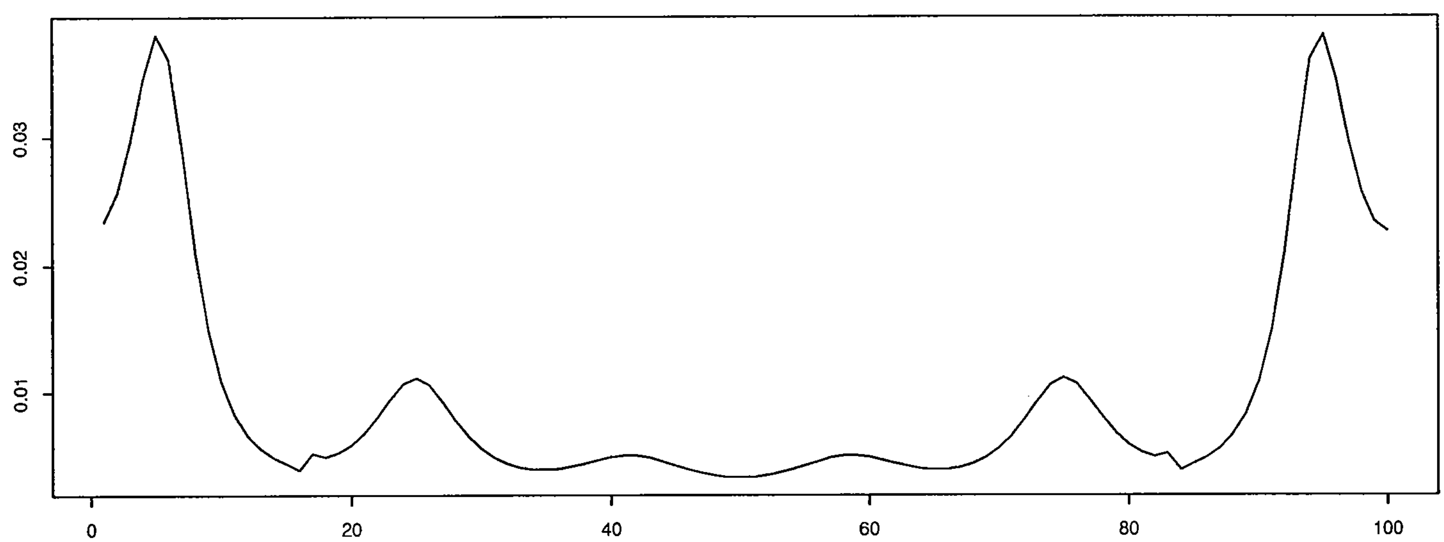

Figure 11. 'Posterior' estimate in the first example, using p=8, and the 'good' prior

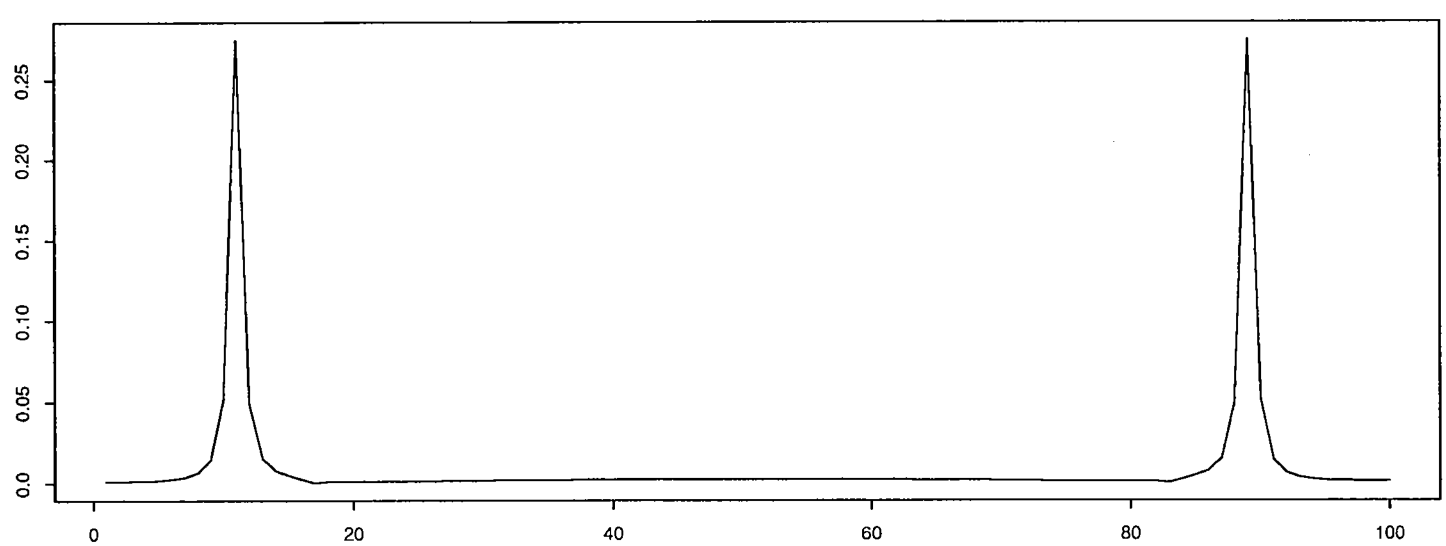

Figure 12. 'Posterior' estimate in the second example, using p=8, and the 'good' prior

References

[1] Berk, K.N. (1974), Consistent autoregressive spectral estimates, *Ann. Statist.*, 489-502.

[2] Brockwell, P.J. and Davis, R.A. (1991), *Time Series: Theory and Methods*, Springer Verlag.

[3] Blackman, R.B. and Tukey, J.W. (1958), *The Measurement of Power Spectra*, Dover, New York.

[4] Burg, J.P. (1967), Maximum Entropy Spectral Analysis, *Proc. 37th Ann. Int. Mtg. Soc. Explor. Geophys.*, Oklahoma City.

[5] Burr, R.L. (1987), Minimum cross entropy spectral analysis and prewhitening methods, *IEEE Trans. Acoust. Speech, Signal Proc.*, 35, 1473.

[6] Burr, R.L. and Lytle, D. W. (1986), Comments on 'A general method of minimum cross entropy spectral estimation', *IEEE Trans. Acoust. Speech, Signal Proc.*, 34, 1324-1326.

[7] Burr, R.L. and Lytle, D. W. (1988), Minimum cross entropy spectral analysis of time-varying biological signals, *IEEE Trans. Systems, Man, Cybern.*, 18, 802-807.

[8] Choi, B.S. (1990), An Information Theoretic Spectral Density, *IEEE Trans. Acoust. Speech, Signal Proc.*, 38, 717-721.

[9] Choi, B.S., Cover, T.M. and Csiszár, I. (1985), Conditional limit theorems under Markov conditioning, Technical Report No. 58, Department of Statistics, Stanford University.

[10] Choi, B.S. and Cover, T.M. (1984), An Information-Theoretic Proof of Burg's Maximum Entropy Spectrum, *Proc. IEEE*, vol. 72, 8, 1094-1095.

[11] Csiszár, I. (1985), A generalized Maximum Entropy Principle and its Bayesian justification, *Bayesian Statistics 2*, p. 83-98, Proceedings of the Second Valencia International Symposium on Bayesian Statistics, North Holland.

[12] Franke, J. (1985), ARMA processes have maximum entropy among time series with prescribed autocovariances and impulse responses, *Adv. Appl. Prob.*, 17, 810-840.

[13] Hong-Zhi, A., Zhao-Gu, C., and Hannan, E.J. (1982), Autocorrelation, autoregression, and autoregressive approximation, *Ann. Statist.*, 926-936.

[14] Ihara, S. (1984), Maximum entropy spectral analysis and ARMA processes, *IEEE Trans. Information Theory*, 30, 377-379.

[15] Jaynes, E.T. (1957), Information theory and statistical mechanics, *Phys. Rev.*, 106, 620-630.

[16] Johnson, R.W. and Shore, J.E. (1984), Which is the better entropy expression for speech processing: $-S \log S$ or $\log S$?, *IEEE Trans. Acoust. Speech, Signal Proc.*, 32, 129-136.

[17] Kullback, S. (1959), *Information theory and Statistics*, Peter Smith, (reprinted by Dover).

[18] Lagunas-Hernandez, M.A., Santamaria-Perez, M.E., and Figueiras-Vidal, A.R. (1984), ARMA model maximum entropy power spectral estimation, *IEEE Trans. Acoust. Speech, Signal Proc.*, 32, 984-991.

[19] Lamperti, J. (1977), *Stochastic processes*, Springer-Verlag.

[20] Papoulis, A. (1984), *Probability, random variables, and stochastic processes*, McGraw-Hill, New York.

[21] Pinsker, M.S. (1964), *Information and information stability of random variables and processes*, Holden-Day, San Francisco.

[22] Politis, D. N. (1992), Moving average processes and maximum entropy, to appear in *IEEE Trans. Information Theory.*

[23] Shore, J.E. (1981), Minimum cross entropy spectral analysis, *IEEE Trans. Acoust. Speech, Signal Proc.*, 29, 230-237.

[24] Tzannes, M.A., Politis, D.N., and Tzannes, N.S. (1985), A general method of minimum cross entropy spectral estimation, *IEEE Trans. Acoust. Speech, Signal Proc.*, 33, 748-752.

[25] Van Campenhout, J.M. and Cover, T.M. (1984), Maximum entropy and conditional probability, *IEEE Trans. Inform. Theory*, IT-27, 483-489.