The Annals of Statistics
1988, Vol. 16, No. 2, 842–855

AN ARMA TYPE PROBABILITY DENSITY ESTIMATOR[1]

BY JEFFREY D. HART

Texas A & M University

Properties of a probability density estimator having the rational form of an ARMA spectrum are investigated. Under various conditions on the underlying density's Fourier coefficients, the ARMA estimator is shown to have asymptotically smaller mean integrated squared error (MISE) than the best tapered Fourier series estimator. The most interesting cases are those in which the Fourier coefficients ϕ_j are asymptotic to $Kj^{-\rho}$ as $j \to \infty$, where $\rho > 1/2$. For example, when $\rho = 2$ the asymptotic MISE of a certain ARMA estimator is only about 63% of that for the optimum series estimator. For a density f with support in $[0, \pi]$, the condition $\rho = 2$ occurs whenever $f'(0+) \neq 0$, $f'(\pi-) = 0$ and f'' is square integrable.

1. Introduction. Suppose $X_1, \ldots, X_n$ are independent observations from a density f with support contained in $[0, \pi]$. We consider estimating $f(x)$ by a quantity having the rational form

$$\hat{f}(x) = \left(\hat{\beta}_0 + 2\sum_{j=1}^{k} \hat{\beta}_j \cos jx\right) \Big/ |1 - \alpha e^{ix}|^2, \tag{1.1}$$

where $|\alpha| < 1$ and the $\hat{\beta}_j$'s are explicit functions of $X_1, \ldots, X_n$. An estimator as in (1.1) bears an obvious resemblance to autoregressive moving average (or ARMA) spectra and will thus be referred to as an ARMA estimator.

Apparently, ARMA type probability density estimators have not previously been considered in the statistical literature. Parzen (1979) and Carmichael (1984) have, however, proposed autoregressive (or AR) type estimators of the form

$$\hat{f}(x; p) = \hat{c}|1 - \hat{\alpha}_1 e^{ix} - \cdots - \hat{\alpha}_p e^{ipx}|^{-2}.$$

Carmichael (1984) obtains a consistency result for $\hat{f}(\cdot; p)$ by allowing the AR order p to tend to infinity at a certain rate with the sample size n. In the current paper, the AR order is fixed at 1 and the MA order tends to infinity with n. The motivation for the latter scheme is based on the numerical analytic device known as the e_1-transform, which in turn is related to the notion of a generalized jackknife.

Hart and Gray (1985) studied the use of ARMA representations in approximating (rather than estimating) density functions. Their results characterize the integrated squared bias of (1.1) in a number of different situations. The current paper greatly generalizes the results of Hart and Gray and also considers the mean integrated squared error (MISE) of (1.1). It is shown that there exist quite

Received October 1986; revised June 1987.
[1]Research supported in part by ONR Contract N00014-85-K-0723.
AMS 1980 *subject classifications.* Primary 62G05, 62G20; secondary 62P10.
Key words and phrases. Probability density estimation, Fourier series, generalized jackknife, regularly varying function.

general conditions under which an estimator of the form (1.1) has asymptotically smaller MISE than does *any* tapered Fourier series estimator. For example, this optimality property holds when $f'(0+) \neq 0$, $f'(\pi-) = 0$ and f'' is square integrable.

The paper will be ordered in the following way. In Section 2, the estimator to be studied is defined and a number of motivations for its use are given. The asymptotic MISE of the estimator is studied in Section 3. Here it is assumed that the Fourier coefficients ϕ_j of f are asymptotic to either Kj^{-p} or $K(-1)^j j^{-p}$ as $j \to \infty$. Cross-validated smoothing of ARMA estimates is addressed in Section 4, where the results of a simulation study are presented. ARMA estimators with AR order greater than 1 are discussed briefly in Section 5. In particular, an asymptotic MISE expression for the ARMA(2, m) estimator is given.

2. The proposed estimator. Let $X_1, \dots, X_n$ be a random sample from a density f with support contained in $[0, \pi]$. We shall assume that f has the Fourier series

$$f(x) \sim \pi^{-1}\left(1 + 2\sum_{j=1}^{\infty} \phi_j \cos jx\right), \qquad 0 \le x \le \pi, \tag{2.1}$$

where $\phi_j = \int_0^\pi \cos jx f(x)\,dx$. Cencov (1962), Kronmal and Tarter (1968), Hall (1983a) and others have investigated density estimators of the form

$$\hat{f}_n(x; m) = \pi^{-1}\left(1 + 2\sum_{j=1}^{m} \hat{\phi}_j \cos jx\right), \qquad 0 \le x \le \pi,$$

where

$$\hat{\phi}_j = n^{-1}\sum_{k=1}^{n} \cos jX_k.$$

The estimators to be studied here are

$$\hat{f}_n(x; m, \alpha) = \hat{f}_n(x; m) + \left(\frac{2}{\pi}\right)\text{Real}\left[\frac{\hat{\phi}_m \alpha \exp(i(m+1)x)}{1 - \alpha\exp(ix)}\right], \tag{2.2}$$

where $-1 < \alpha < 1$. The pair (m, α) is the smoothing parameter of the estimator and can be chosen from the data by cross-validation. We will return to this point in Section 4.

Before further discussion of (2.2) we should justify using the cosine basis as opposed to a basis with both cosine and sine functions. This study was in part motivated by the problem of estimating animal abundance using the line transect method [see Gates and Smith (1980) and Crain, Burnham, Anderson, and Laake (1979)]. The density function estimated in this setting is typically assumed to be monotone decreasing on $(0, \pi)$. Since $f(0) \neq f(\pi)$, the periodic extension of f is discontinuous at 0 and π, and hence a cosine–sine Fourier series estimator will perform poorly near these two points. As shown by Hall (1983a), though, the cosine series estimator is not adversely affected by the condition $f(0) \neq f(\pi)$. In kernel estimation, the analogous means of correcting boundary problems is the symmetrization device studied by Schuster (1985) and Cline and Hart (1986).

There are a number of ways of characterizing the estimator $\hat{f}_n(\cdot;\, m, \alpha)$. First, it is clear that it may be written

$$\hat{f}_n(x;\, m, \alpha) = \sum_{j=-(m+1)}^{m+1} \hat{\beta}_j e^{ijx}/|1 - \alpha e^{ix}|^2, \qquad 0 \le x \le \pi, \tag{2.3}$$

where $\hat{\beta}_j = \hat{\beta}_{-j}$ and the $\hat{\beta}_j$'s depend only on α and the $\hat{\phi}_j$'s. Aside from the issue of positivity, (2.3) has the form of an ARMA(1, $m + 1$) spectrum, hence the name ARMA probability density estimator. The form (2.3) suggests that such ARMA estimators are well suited for densities with large "power" at either 0 or π (but not both). While this is so, it will be seen that such an observation somewhat understates the value of $\hat{f}_n(\cdot;\, m, \alpha)$.

Perhaps a more interesting way of characterizing $\hat{f}_n(\cdot,\, m, \alpha)$ is in terms of the generalized jackknife and the numerical analytic device known as the e_1-transform. Using (2.2), it is easy to show that, for $m \ge 1$,

$$\hat{f}_n(x;\, m, \alpha) = \pi^{-1}\left[1 + 2\,\mathrm{Real}(G_{m,\alpha}(x))\right], \tag{2.4}$$

where $G_{m,\alpha}(x) = (\hat{F}_m(x) - \alpha e^{ix}\hat{F}_{m-1}(x))/(1 - \alpha e^{ix})$, $\hat{F}_k(x) = \sum_{j=1}^{k} \hat{\phi}_j e^{ijx}$, $k = 1, 2, \dots$, and $\hat{F}_0 \equiv 0$. The quantity $G_{m,\alpha}(x)$ is a generalized jackknife estimator [as defined by Schucany, Gray and Owen (1971)] of the function $F(x) = \sum_{j=1}^{\infty} \phi_j e^{ijx}$. The jackknife, of course, is well known as a means of reducing estimator bias.

Defining $F_k(x) = E(\hat{F}_k(x))$, we have

$$E(G_{m,\alpha}(x)) = \left(F_m(x) - \alpha e^{ix} F_{m-1}(x)\right)/(1 - \alpha e^{ix}).$$

If α is taken to be ϕ_{m+1}/ϕ_m, then $\{E(G_{m,\alpha}(x)) \colon m = 1, 2, \dots\}$ is the e_1-transform of the sequence $\{F_m(x)\}$. This transform, which dates at least to Aitken (1926), is a numerical analytic tool used for accelerating the convergence of a sequence to its limit. The work of Shanks (1955) gives a thorough account of e_1 and the more general e_n-transform. The interesting paper of Gray (1985) demonstrates the close connection between many numerical analytic methods (including the e_n-transform) and the statistical notion of jackknifing to reduce bias. For more uses of the e_n-transform in statistical problems, see Gray, Watkins and Adams (1972) and Morton and Gray (1984).

Also of interest is the Fourier series of $\hat{f}_n(\cdot;\, m, \alpha)$. From (2.2),

$$\hat{f}_n(x;\, m, \alpha) = \hat{f}_n(x;\, m) + (2/\pi) \sum_{j=m+1}^{\infty} \hat{\phi}_m \alpha^{j-m} \cos jx.$$

The Fourier coefficients $\hat{\phi}_j(m, \alpha)$ of the ARMA estimator are thus

$$\begin{aligned} \hat{\phi}_j(m, \alpha) &= \hat{\phi}_j, \qquad j = 0, 1, \dots, m, \\ &= \hat{\phi}_m \alpha^{j-m}, \qquad j = m + 1, \dots. \end{aligned} \tag{2.5}$$

It is of interest to contrast the tapering scheme (2.5) with that of traditional series type estimators [see Wahba (1981)]. The Fourier coefficients $\tilde{\phi}_j$ of such estimators are

$$\tilde{\phi}_j = w_n(j)\hat{\phi}_j, \qquad j \geq 0, \tag{2.6}$$

where $w_n(j)$ does not depend on the data. Wahba (1981) considers, for example, $w_n(j) = (1 + \lambda j^{2k})^{-1}$, with λ and k being parameters chosen by the user. The scheme (2.5) differs from (2.6) in that, for $j \geq m + 1$, $\hat{\phi}_j(m, \alpha) = \hat{\phi}_m \alpha^{j-m}$ rather than $\hat{\phi}_j \alpha^{j-m}$. Surprisingly, this simple difference between the two tapering schemes can result in a profound difference in the biases of the corresponding estimators. It will be shown that in a variety of situations the MISE of (2.2) is asymptotically smaller than that of any estimator with Fourier coefficients as in (2.6).

3. Mean integrated squared error results. Define the mean integrated squared error of the estimator $\hat{f}$ by $J(\hat{f}, f) = E\int_0^\pi (\hat{f}(x) - f(x))^2\,dx$. The MISE will be used as a basis for comparing $\hat{f}_n(\cdot; m, \alpha)$ with other Fourier series estimators. It is straightforward to show that

$$\begin{aligned} J(\hat{f}_n(\cdot; m, \alpha), f) = (2/\pi)\Bigg[& n^{-1} \sum_{j=1}^{m} \operatorname{var}(\cos jX_1) \\ & + \alpha^2(1 - \alpha^2)^{-1} \operatorname{var}(\cos mX_1)/n \Bigg] \\ & + (2/\pi) \sum_{j=m+1}^{\infty} \left(\phi_j - \phi_m \alpha^{j-m}\right)^2, \end{aligned} \tag{3.1}$$

where $\operatorname{var}(\cos jX_1) = (1 + \phi_{2j})/2 - \phi_j^2$. Note that taking $\alpha = 0$ in (3.1) yields $J(\hat{f}_n(\cdot; m), f)$. This is important since it makes clear that one may always choose α so that the MISE of an ARMA estimator is no bigger than that of $\hat{f}_n(\cdot; m)$.

For future reference we define

$$J_{n,\text{opt}} = \left(\frac{2}{\pi}\right) \sum_{j=1}^{\infty} \frac{\phi_j^2 \operatorname{var}(\cos jX_1)}{\operatorname{var}(\cos jX_1) + n\phi_j^2}.$$

Watson (1969) showed that $J_{n,\text{opt}}$ is the minimum MISE among estimators of the form

$$\hat{f}_w(x) = (1/\pi)\left(1 + 2 \sum_{j=1}^{\infty} w_n(j)\hat{\phi}_j \cos jx\right). \tag{3.2}$$

The estimator $\hat{f}_n(\cdot; m)$ and Wahba's (1981) estimator are, of course, special cases of (3.2). The singular integral estimators of Hall (1983a) are also of this form.

It is clear from (2.5) that $\hat{f}_n(\cdot; m, \alpha)$ will be the most efficient when the Fourier coefficients of f decay geometrically. Hart (1986) deals with the situation in which $\phi_j = \exp(-aj)R(j)$ and R is regularly varying at infinity. In particular, suppose that $R(j) = \beta + \gamma_j \exp(-\delta j)$, where $\beta \neq 0$, $\delta > 0$ and $\gamma_j \to \gamma$ as $j \to \infty (|\gamma| < \infty)$. Hart (1986) shows in this case that

$\lim_{n\to\infty} J(\hat{f}_n(\cdot; m_n, e^{-a}), f)/J_{n,\text{opt}} = a/(a+\delta)$, where m_n minimizes $J(\hat{f}_n(\cdot; m, e^{-a}), f)$. An example of the situation just described is when f has Fourier coefficients $\phi_j = 1/\cosh(aj)$, $j \geq 0$. In this case $a/(a+\delta)$ is 1/3, and so there is an ARMA estimator whose asymptotic MISE is only 1/3 of $J_{n,\text{opt}}$.

In the remainder of this section, we shall deal with cases in which either $\phi_j \sim Kj^{-\rho}$ or $\phi_j \sim (-1)^j Kj^{-\rho}$, where $K \neq 0$ and $\rho > 1/2$. Here it is not clear from (2.5) that ARMA estimators should be more efficient than traditional Fourier series estimators. It turns out, though, that if ρ is 1, 2 or 3 or is at least 3.26, then the best ARMA estimator is more efficient than any estimator of the form (3.2). Before stating a theorem to this effect, Lemma 1 on the integrated squared bias of $\hat{f}_n(\cdot; m, \alpha)$ is given. The lemma is actually more general than is needed for the subsequent MISE theorems.

LEMMA 1. *Suppose that $\phi_j = R_{-\rho}(j)$, $j \geq 1$, where $R_{-\rho}$ is continuous and regularly varying at infinity, i.e., $R_{-\rho}$ satisfies $\lim_{x\to\infty} R_{-\rho}(\lambda x)/R_{-\rho}(x) = \lambda^{-\rho}$ for each $\lambda > 0$. If, in addition, $m(1-\alpha_m) \to c > 0$ as $m \to \infty$, then*

$$\sum_{j=m+1}^{\infty} \left(\phi_j - \phi_m \alpha_m^{j-m}\right)^2 = m\phi_m^2 \int_0^{\infty} \left((1+y)^{-\rho} - e^{-cy}\right)^2 dy + o\left(m\phi_m^2\right).$$

PROOF. Since the ϕ_j's are regularly varying, $\sum_{j=m+1}^{\infty} \phi_j^2/(m\phi_m^2) \to \int_0^{\infty}(1+y)^{-2\rho}\, dy$ as $m \to \infty$. Also,

$$\phi_m^2 \sum_{j=m+1}^{\infty} \alpha_m^{2(j-m)}/\left(m\phi_m^2\right) = \alpha_m^2 m^{-1}\left(1-\alpha_m^2\right)^{-1} \to 1/(2c) = \int_0^{\infty} e^{-2cy}\, dy.$$

There exist [see Seneta (1976), pages 19–20] functions $\underline{\phi}$ and $\bar{\phi}$, defined on $(0, \infty)$, with the properties $\underline{\phi}(t) \sim \bar{\phi}(t)$ as $t \to \infty$ and $\underline{\phi}(t) \leq \phi_{[t]+1} \leq \bar{\phi}(t)$ for all $t > 0$. By assumption,

$$1 - (1+\varepsilon)c/m \leq \alpha_m \leq 1 - (1-\varepsilon)c/m$$

for $0 < \varepsilon < 1$ and all m sufficiently large. It follows that

$$\begin{aligned} \limsup_{m\to\infty} & \sum_{j=m+1}^{\infty} \phi_j \alpha_m^{j-m}/(m\phi_m) \\ &\leq \lim_{m\to\infty} \int_m^{\infty} (m\phi_m)^{-1} \bar{\phi}(t)(1-(1-\varepsilon)c/m)^{t-m}\, dt \\ &= \lim_{m\to\infty} \int_1^{\infty} \left(\bar{\phi}(mu)/\phi_m\right)(1-(1-\varepsilon)c/m)^{m(u-1)}\, du. \end{aligned}$$

Making use of dominated convergence, the last limit is

$$\int_1^{\infty} u^{-\rho} e^{-(1-\varepsilon)c(u-1)}\, du.$$

Obtaining a similar lower bound and using the fact that ε may be taken arbitrarily small, we have

$$\lim_{m\to\infty} \sum_{j=m+1}^{\infty} \phi_j \alpha_m^{j-m}/(m\phi_m) = \int_0^{\infty} (1+y)^{-\rho} e^{-cy}\, dy.$$

The lemma now follows upon combining the previous results. □

Of interest is whether the quantity c in Lemma 1 can be chosen so that the integrated squared bias of the ARMA estimator is less than that of $\hat{f}_n(\cdot; m)$. Define $B(\hat{f}, f)$ to be the integrated squared bias of the density estimator $\hat{f}$. Now, if α_m is such that $m(1-\alpha_m) \to \rho$ as $m \to \infty$, then under the conditions of Lemma 1, it follows that

$$(3.3)\qquad \begin{aligned} &B(\hat{f}_n(\cdot; m, \alpha_m), f)/B(\hat{f}_n(\cdot; m), f) \\ &\quad \to (2\rho - 1)\int_0^\infty \left((1+y)^{-\rho} - e^{-\rho y}\right)^2 dy < 1. \end{aligned}$$

It is worth noting that the e_1-transform choice for α_m, i.e., $\alpha_m = \phi_{m+1}/\phi_m$, satisfies $m(1-\alpha_m) \to \rho$ when $\phi_m = Km^{-\rho} + o(m^{-(1+\rho)})$ $(K \neq 0)$.

While (3.3) is encouraging, our real concern is with the MISE of ARMA estimators. Hence, our first theorem gives approximations to the MISE of $\hat{f}_n(\cdot; m, \alpha_m)$ and of $\hat{f}_n(\cdot; m)$.

Theorem 1. *Let the Fourier coefficients of the density f satisfy either $j^\rho \phi_j \to K \neq 0$ or $(-1)^j j^\rho \phi_j \to K \neq 0$ as $j \to \infty$, where $\rho > 1/2$. In the former case let $m(1-\alpha_m) \to c > 0$ as $m \to \infty$, while in the latter let $m(1+\alpha_m) \to c > 0$. Then, defining $I_{\rho,c} = \int_0^\infty ((1+y)^{-\rho} - e^{-cy})^2\, dy$,*

$$(3.4)\qquad \begin{aligned} J(\hat{f}_n(\cdot; m, \alpha_m), f) &= \pi^{-1}\left[(m/n)\left(1 + (2c)^{-1}\right) + 2K^2 m^{1-2\rho} I_{\rho,c}\right] \\ &\quad + o\left(m/n + m^{1-2\rho}\right) \end{aligned}$$

and

$$J(\hat{f}_n(\cdot; m), f) = \pi^{-1}\left[m/n + 2K^2 m^{1-2\rho}(2\rho-1)^{-1}\right] + o\left(m^{1/2}/n + m^{1-2\rho}\right).$$

Furthermore, if m_n and $\tilde{m}_n$ are the respective minimizers of $J(\hat{f}_n(\cdot; m), f)$ and $J(\hat{f}_n(\cdot; m, \alpha_m), f)$ and if $J_{FS,n}$ and $J_{A,n}$ are the respective minima, then as $n \to \infty$,

$$(3.5)\qquad m_n \sim (2K^2)^{1/(2\rho)} n^{1/(2\rho)},$$

$$(3.6)\qquad \tilde{m}_n \sim \left[2K^2(2\rho-1)I_{\rho,c}/\left(1 + (2c)^{-1}\right)\right]^{1/(2\rho)} n^{1/(2\rho)},$$

$$(3.7)\qquad J_{n,FS} \sim \pi^{-1}(2K^2)^{1/(2\rho)} 2\rho(2\rho-1)^{-1} n^{1/(2\rho)-1} = A_\rho n^{1/(2\rho)-1}$$

and

$$(3.8)\qquad J_{n,A} \sim A_\rho\left(1 + (2c)^{-1}\right)^{1-1/(2\rho)}\left[(2\rho-1)I_{\rho,c}\right]^{1/(2\rho)} n^{1/(2\rho)-1}.$$

Proof. Using the conditions $\rho > 1/2$ and $j^\rho|\phi_j| \to |K|$, we have $\sum_{j=1}^m \operatorname{var}(\cos jX_1) = m/2 + o(m^{1/2})$. The integrated variance approximation for $\hat{f}_n(\cdot; m)$ follows from this, and so does that for $\hat{f}_n(\cdot; m, \alpha_m)$ after also using the fact that $\alpha_m^2(1-\alpha_m^2)^{-1} = m/(2c) + o(m)$. When $\phi_j \sim Kj^{-\rho}$, the integrated squared bias approximation for $\hat{f}_n(\cdot; m, \alpha_m)$ is immediate using Lemma 1. When

TABLE 1
Asymptotic MISE *of* ARMA *and Fourier series estimators for densities with algebraically decaying Fourier coefficients*[a]

ρ	c	Type of estimator ARMA(ρ)	ARMA(c)	Optimum Fourier series	Fourier series with 0–1 weights
0.55	0.30	10.123	9.711	10.137	11
0.75	0.30	2.265	1.951	2.418	3
1	0.50	1.358	1.111	1.571	2
2	1.39	0.853	0.699	1.111	4/3
3	2.30	0.797	0.679	1.047	6/5
4	3.33	0.789	0.691	1.026	8/7
10	9.30	0.834	0.783	1.004	20/19
20	20	0.883	0.883	1.001	40/39

[a] It is assumed that $j^\rho\phi_j \to K \neq 0$ or that $(-1)^j j^\rho \phi_j \to K \neq 0$. For a given ρ and an estimator, a table value in columns 3–6 is the limit of $\pi(2K^2)^{-1/(2\rho)}n^{1-1/(2\rho)}$ MISE. Hence, ratios of values in the same row are asymptotic relative efficiencies. ARMA(η) denotes an ARMA estimator for which $m(1-|\alpha_m|) \to \eta$ as $m \to \infty$ (see Theorem 1). The values of c are approximate solutions of (3.9). Details on how the ARMA efficiencies were calculated are available from the author.

$(-1)^j\phi_j \sim Kj^{-\rho}$ and $m(1+\alpha_m) \to c$, note that

$$\sum_{j=m+1}^{\infty}\left(\phi_j - \phi_m\alpha_m^{j-m}\right)^2 = \sum_{j=m+1}^{\infty}\left((-1)^j\phi_j - (-1)^m\phi_m(-\alpha_m)^{j-m}\right)^2$$

and use Lemma 1. The integrated squared bias approximation for $\hat{f}_n(\cdot; m)$ follows upon observing that $\sum_{j=m+1}^{\infty}\phi_j^2/(m\phi_m^2) \to (2\rho-1)^{-1}$. □

One may show that the derivative of (3.8) with respect to c is 0 when

$$(3.9) \qquad (4c^2-\rho) = 2c\left(2c^2+(2\rho-1)c-\rho\right)e^c E_\rho(c),$$

where $E_\rho(c) = \int_1^\infty e^{-cy}y^{-\rho}\,dy$ is the so-called exponential integral. Using (3.9) and the tabled values of $E_\rho(c)$ in Abramowitz and Stegun (1972), approximately optimal values of c have been determined for selected ρ. These are given in Table 1.

The next theorem shows that when $c=\rho$, the expression (3.8) is less than (3.7) for each $\rho > 1/2$. More importantly, the theorem and subsequent numerical results give conditions under which $\lim_{n\to\infty} J_{n,A}/J_{n,\mathrm{opt}} < 1$.

THEOREM 2. *Let the conditions of Theorem* 1 *hold with* $c = \rho$. *If* $J_{n,A}$ *and* $J_{n,FS}$ *are the minima of* $J(\hat{f}_n(\cdot; m, \alpha_m), f)$ *and* $J(\hat{f}_n(\cdot; m), f)$, *then for each* $\rho > 1/2$,

$$\lim_{n\to\infty} J_{n,A}/J_{n,FS} = B_\rho = \left[(2\rho-1)I_{\rho,\rho}\right]^{1/(2\rho)}\left(1+(2\rho)^{-1}\right)^{1-1/(2\rho)} < 1.$$

Furthermore, for each $\rho > 1/2$,

$$\lim_{n\to\infty} J_{n,A}/J_{n,\mathrm{opt}} = C_\rho = \left[(2\rho - 1)I_{\rho,\rho}\right]^{1/(2\rho)}\left(1 + (2\rho)^{-1}\right)^{1-1/(2\rho)}2\rho \times (2\rho - 1)^{-1}\Big/\int_0^\infty (1 + y^{2\rho})^{-1}\,dy,$$

where $C_\rho < 1$ *for each* $\rho \geq 3.26$. (*For other values of* ρ *see Table* 1.)

PROOF. The first of the two limits follows immediately from Theorem 1. To establish that $B_\rho < 1$, first note that $(2\rho - 1)I_{\rho,\rho} < (2\rho)^{-1}$ as $(1 + y)^{-\rho}e^{-\rho y} > e^{-2\rho y}$ for $y > 0$. Hence, $B_\rho < (2\rho)^{-1/(2\rho)}(1 + (2\rho)^{-1})^{1-1/(2\rho)}$, which is less than or equal to 1 for $\rho > 1/2$ since $(x + 1) \geq (1 + x^{-1})^x$ for $x \geq 1$.

A proof essentially the same as that of the first theorem in Section 3 of Watson and Leadbetter (1963) shows that (under the conditions of our Theorem 2) $J_{n,\mathrm{opt}} \sim n^{1/(2\rho)-1}\pi^{-1}(2K^2)^{1/(2\rho)}\int_0^\infty(1 + y^{2\rho})^{-1}\,dy$. Combining this with (3.8) gives the limit C_ρ. Now, $\int_0^\infty(1 + y^{2\rho})^{-1}\,dy = \pi[2\rho\sin(\pi/(2\rho))]^{-1}$ [see, e.g., Bronshtein and Semendyayev (1985), page 60], which is larger than 1 for each $\rho > 1/2$. To show that $C_\rho < 1$, it is thus sufficient to show that $C_\rho\int_0^\infty(1 + y^{2\rho})^{-1}\,dy \leq 1$. For $\rho \geq 3.26$ this is done as in the proof of the first inequality by using the fact that $(x + 1) \geq (1 + x^{-1})^x(1 - x^{-1})^{-x}$ for $x \geq 6.52$. □

This section is concluded with several remarks about the results in Theorems 1 and 2.

REMARK 1. Theorem 2 shows that an improvement in MISE is possible if one uses an ARMA estimator instead of a traditional series estimator. Table 1 indicates that the resulting improvement can be quite dramatic. For example, if $\rho = 2$, the asymptotic MISE of the best ARMA estimator is only about 63% of that for the optimum series estimator. An example of an (asymptotically) optimum series estimator for a given $\rho > 1/2$ is Wahba's estimator $\hat{f}_{\lambda,\rho}$ with taper $w_n(j) = (1 + \lambda j^{2\rho})^{-1}$. Of course, in practice ρ is unknown and will have to be estimated from the data and so $\hat{f}_{\lambda,\rho}$, like $\hat{f}_n(\cdot;\, m, \alpha)$, has two tuning constants. In fairness to Wahba's method, we should point out that $\hat{f}_{\lambda,\rho}$ is guaranteed to be optimum among estimators of the form (3.2) when $j^\rho|\phi_j| \to K > 0$. This condition is weaker than the ones in Theorems 1 and 2.

REMARK 2. Although Theorem 2 establishes that $C_\rho < 1$ only for $\rho \geq 3.26$, the numerical evidence in Table 1 suggests the inequality holds for each $\rho > 1/2$. Furthermore, one can show that B_ρ/C_ρ has limit 1 as $\rho \to 1/2$.

REMARK 3. Qualitative conditions on f under which the conditions of Theorems 1 and 2 hold can be obtained by expanding ϕ_j via integration by parts. Suppose f has $2k$ derivatives ($k \geq 1$) on $[0, \pi]$ with $f^{(2k)}$ square integrable. If $f^{(r)}(0+) = f^{(r)}(\pi-) = 0$ for $r = 1, 3, \ldots, 2k - 3$, $f^{(2k-1)}(\pi-) = 0$ and

$f^{(2k-1)}(0+) \neq 0$, then $j^{2k}\phi_j \to (-1)^k f^{(2k-1)}(0+)$, and Theorems 1 and 2 hold with $\rho = 2k$. In particular, when $k = 1$ we have $f'(0+) \neq 0$ and $f'(\pi-) = 0$. Exponential-like densities are one example of such behavior. The poor performance of the series estimator $\hat{f}_n(0; m)$ when $f'(0) \neq 0$ has been noted by Buckland (1985) in the setting of line transect sampling. Table 1 and the example in the next section show that the ARMA estimator can yield a large improvement over $\hat{f}_n(\cdot; m)$ when $f'(0+) \neq 0$.

REMARK 4. Hart and Gray (1985) claimed that ARMA estimators are often more parsimonious than traditional series estimators. Evidence of this is seen by examining (3.5)–(3.8) and Theorem 2. When c is chosen optimally, we have $\lim_{n \to \infty} \tilde{m}_n / m_n < 1$ for each $\rho > 1/2$. For example, when $\rho = 2$, the best ARMA estimator uses (in the limit) only about 39% as many Fourier coefficients as does $\hat{f}_n(\cdot; m)$; and yet the MISE of $\hat{f}_n(\cdot; m)$ is about twice that of the ARMA estimator.

REMARK 5. Finally, we note that our results can be generalized to allow for regularly varying ϕ_j's [see Hart (1986)]. In this case $\phi_j = j^{-\rho}L(j)$, where L is any slowly varying function. An example of a slowly varying characteristic function is $L(t) = (1 + \log(1 + |t|))^{-1}$.

4. Choice of smoothing parameter by cross-validation. In the practice of density estimation, one must usually select smoothing parameters via some data-driven method. As a result, data-based density estimators are not as efficient (at least in small samples) as theory would suggest. Hence, it is not realistic to expect that the MISE improvement discussed in previous sections is fully attainable in practice, except perhaps in very large samples. Some improvement in small samples seems likely, though, if the ARMA estimator's smoothing parameter is reasonably chosen. This is borne out in the simulation study of this section.

One means of choosing the smoothing parameter (m, α) of the ARMA estimator is the cross-validatory method introduced by Rudemo (1982) and Bowman (1984). This method chooses the pair $(\hat{m}, \hat{\alpha})$ that minimizes

$$\hat{R}(m, \alpha) = \int_0^{\pi} \hat{f}_n^2(x; m, \alpha)\, dx - (2/n) \sum_{i=1}^{n} \hat{f}_{n,i}(X_i; m, \alpha), \tag{4.1}$$

where $\hat{f}_{n,i}$ indicates the estimator calculated by deleting the data value X_i. Rudemo (1982) showed that $\hat{R}(m, \alpha)$ is an unbiased estimator of the risk

$$R(m, \alpha) = J(\hat{f}_n(\cdot; m, \alpha), f) - \int_0^{\pi} f^2(x)\, dx.$$

A number of results now exist showing that density estimates chosen by cross-validation are asymptotically efficient [see, for example, Hall (1983b, 1987), Stone (1984) and Hall and Marron (1987)].

To investigate the behavior of cross-validated ARMA estimates, a small simulation study was conducted. The density considered was

$$f(x) = 2e^{-2x}(1 + e^{-4(\pi - x)})(1 - e^{-4\pi})^{-1}, \qquad 0 \le x \le \pi, \tag{4.2}$$

which has Fourier coefficients

$$\phi_j = \left(1 + (j/2)^2\right)^{-1}, \qquad j = 1, 2, \dots .$$

This density satisfies the conditions of Theorem 2 with $\rho = 2$.

For (4.2) and $n = 50$, the minimum MISE among ARMA estimators is 0.00633. This minimum occurs at $(m, \alpha) = (1, 0.64)$. Among Fourier series estimators $\hat{f}_n(\cdot; m)$, the optimum m and MISE are 5 and 0.04160. Since 0.00633/0.0416 = 0.152, we see that the asymptotic relative efficiency of 0.52 from Table 1 understates the improved efficiency of the optimum ARMA estimator at $n = 50$.

Twenty independent random samples of size $n = 50$ were generated from the density (4.2). This was done by generating values from the exponential density $g(x) = 2e^{-2x} I_{(0,\infty)}(x)$, and using the fact that, if Y has density g, then

$$X = Y I_{(0,\pi)}(Y) + \sum_{j=1}^{\infty} |Y - 2j\pi| I_{((2j-1)\pi,(2j+1)\pi)}(Y)$$

has density (4.2). Since little of the mass of g is larger than π, the graphs of f and g on $(0, \pi)$ are virtually indistinguishable.

For each of the 20 data sets, the minimizer of $\hat{R}$ [defined by (4.1)] for $0 \le \alpha < 1$, $1 \le m \le 20$ was approximated. Also, the minimizer of $\hat{R}(m, 0)$ for $1 \le m \le 20$ was determined. This latter value of m is simply a cross-validatory choice of the smoothing parameter of $\hat{f}_n(\cdot; m)$ [see Hart (1985) and Diggle and Hall (1986) for more on this subject]. The integrated squared errors of the cross-validated ARMA and Fourier series estimates were determined for each data set. Denote these two ISEs I_A and I_F, respectively.

The results of the simulation are summarized in Table 2. The fact that average I_A was a bit larger than average I_F is misleading. In 16 of the 20 cases, the ratio I_A/I_F was between 0.151 and 0.690. Note the trimmed mean and the confidence interval for the median of I_A/I_F in Table 2. These more accurately reflect the overall performance of the two cross-validated density estimates. From the 16 cases in which I_A was less than I_F, a typical comparison of the two

TABLE 2
Summary of simulation study[a]

	Type of estimate	
	ARMA	**Fourier series**
Average ISE	0.0542	0.0522
Median ISE	0.0183	0.0354
Trimmed mean ISE	0.0287	0.0390

[a]The trimmed means exclude the three (out of 20) largest values of ISE. A 95% confidence interval for the median of I_A/I_F is (0.3607, 0.6581), where I_A and I_F are, respectively, the ISEs of cross-validated ARMA and Fourier series estimates.

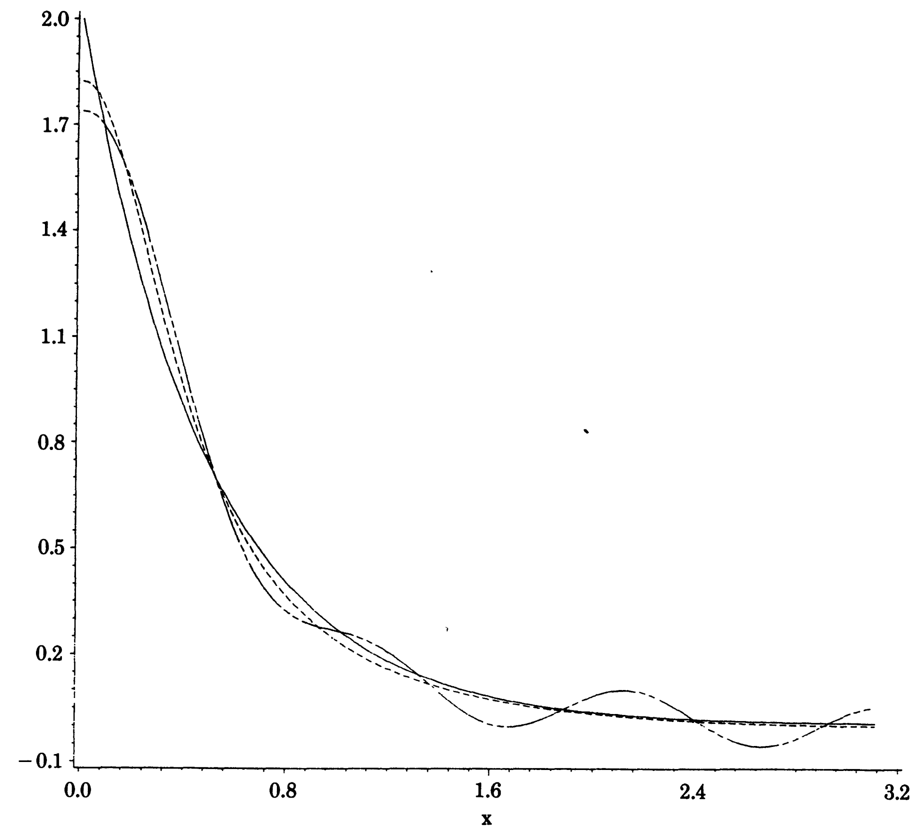

FIG. 1. *Wrapped exponential density, ARMA estimate and Fourier series estimate: The solid curve is the density* (4.2). *The ARMA estimate has the higher value at* 0 *and no spurious bumps. The two estimates were calculated from the same set of data, each being fitted by cross-validation.*

estimates is given in Figure 1. The qualitative improvements obtained with the ARMA estimate are a better estimate of $f(0)$ and an absence of spurious bumps. See Hart and Gray (1985) for further discussion of the qualitative properties of ARMA approximators.

It is also important to point out what happened in the four cases where $I_A > I_F$. In these cases, cross-validation chose too large an α and/or too large an m. When α is too near 1, the ARMA estimate tends to be too large near 0, thus inflating the estimate's integrated squared error. That cross-validation would occasionally choose too rough an ARMA estimate is not unexpected. It is well known from other settings that in small samples cross-validation tends to drastically undersmooth around 5–20% of the time [see, e.g., Hart (1985)]. Scott and Terrell (1987) give an explanation for this phenomenon.

Clearly, further experimentation is needed to determine how efficiently cross-validation smooths ARMA estimates. Recent work of Scott and Terrell (1987) shows that a biased version of cross-validation provides more efficient kernel estimators in moderate and large samples. Such an idea is also worth pursuing in the setting of ARMA estimators. For example, a modification of $\hat{R}(m, \alpha)$ that places a more severe penalty on large values of $|\alpha|$ would discourage the occasional undersmoothing observed in the simulation study.

5. More general ARMA estimators. To define an ARMA(k, m) estimator ($k \geq 2$), one may appeal to the notion of a higher order jackknife as in Schucany, Gray and Owen (1971). For simplicity, we shall only discuss the ARMA$(2, m)$ estimator here. Define for each $m \geq 2$,

$$\hat{f}_n(x; m, \boldsymbol{\alpha}) = \pi^{-1}\left[1 + 2\,\mathrm{Real}\left(\left(G_m(x; \boldsymbol{\alpha})\right)\right)\right], \tag{5.1}$$

where $G_m(x; \boldsymbol{\alpha}) = (\hat{F}_m(x) - \alpha_1 e^{ix}\hat{F}_{m-1}(x) - \alpha_2 e^{2ix}\hat{F}_{m-2}(x))/(1 - \alpha_1 e^{ix} - \alpha_2 e^{2ix})$. The real numbers α_1 and α_2 are such that the zeros of $h(z) = 1 - \alpha_1 z - \alpha_2 z^2$ are outside the unit circle in the complex plane. The Fourier coefficients $\hat{\phi}_j(m, \boldsymbol{\alpha})$ of the estimator (5.1) satisfy the second order difference equation $a_j - \alpha_1 a_{j-1} - \alpha_2 a_{j-2} = 0$ for $j > m + 2$. To develop results analogous to those in the previous section, it is thus natural to consider the model

$$\phi_j = \left(Kj^{-\rho} + o(j^{-\rho})\right)\cos(j\theta), \tag{5.2}$$

where $\rho > 1/2$ and $0 < \theta < \pi$. (The cases $\theta = 0$ and $\theta = \pi$ are covered by Theorem 1.)

Under (5.2), the MISE of $\hat{f}_n(\cdot; m)$ is

$$J\left(\hat{f}_n(\cdot; m), f\right) = \pi^{-1}\left[m/n + K^2(2\rho - 1)^{-1}m^{1-2\rho}\right] + o\left(m^{1/2}/n + m^{1-2\rho}\right),$$

and so the efficiencies in Table 1 are still applicable. (The integrated squared bias approximation is established by arguing as in Theorem 1 and using Abel's lemma.) Now, let $\boldsymbol{\alpha}_m = (\alpha_{1m}, \alpha_{2m})$, where $\alpha_{1m} = 2r_m\cos\theta$, $\alpha_{2m} = -r_m^2$ and $0 \leq r_m < 1$. If $m(1 - r_m) \to c > 0$ and (5.2) holds, then

$$\begin{aligned} &J\left(\hat{f}_n(\cdot; m, \boldsymbol{\alpha}_m), f\right) \\ &\quad = \pi^{-1}\left[mn^{-1}\left(1 + \frac{2(1 - \phi_1\cos\theta)}{c|1 - e^{2i\theta}|^2}\right) + K^2 I_{\rho,c} m^{1-2\rho}\right] \\ &\qquad + o\left(\frac{m}{n} + m^{1-2\rho}\right). \end{aligned} \tag{5.3}$$

Notice that the bias term in (5.3) is essentially the same as that in (3.4), but the variance terms are the same only when $\theta = \pi/2$. If $\theta = \pi/2$, the asymptotic efficiency of the ARMA$(2, m)$ estimator can be obtained from Table 1. We see that if θ is near 0 or π, the variance of $\hat{f}_n(\cdot; m, \boldsymbol{\alpha}_m)$ becomes large. This can be counteracted by taking c to be larger than it is in Table 1. Increasing c, however, will make the integrated squared bias of the ARMA estimator larger, although one can show that $I_{\rho,c} < (2\rho - 1)^{-1}$ for each $c \geq \rho$. In any event, it appears that if $j^\rho\phi_j$ or $(-1)^j j^\rho\phi_j$ oscillate slowly about zero, then any increased efficiency obtained with an ARMA estimator will be small in comparison to the increases seen in Table 1. If, on the other hand, $j^\rho\phi_j$ or $(-1)^j j^\rho\phi_j$ oscillate slowly about a nonzero constant, then the ARMA$(1, m)$ estimator can be efficient. This follows from Remark 5 of Section 4 and the fact that slowly varying functions can be infinitely oscillating [see Seneta (1976), page 49].

The ideas of this article could also be applied to other density estimators of the type

$$\hat{f}_m(x) = (1/\pi)\left[1 + 2\sum_{j=1}^{m} w_j\hat{\phi}_j\cos jx\right].$$

The form (2.4) immediately suggests jackknifed versions of $\hat{f}_m$. So long as the truncation bias of $\hat{f}_m$ is not dominated by the bias due to the w_j's, results analogous to Theorem 2 could be established for a jackknifed $\hat{f}_m$.

Acknowledgments. The author expresses his deep appreciation to H. L. Gray, whose work in numerical analysis was the stimulus for this paper. Also, the author benefitted greatly from the insight of Daren B. H. Cline, who provided a proof of Lemma 1. Finally, the comments of an Associate Editor and three referees were most helpful in leading to this much simplified version of the original manuscript.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I. A. (1972). *Handbook of Mathematical Functions*. Dover, New York.

AITKEN, A. C. (1926). On Bernoulli's numerical solution of algebraic equations. *Proc. Roy. Soc. Edinburgh Sect. A* **46** 289–305.

BOWMAN, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika* **71** 353–360.

BRONSHTEIN, I. N. and SEMENDYAYEV, K. A. (1985). *Handbook of Mathematics*. Van Nostrand Reinhold, New York.

BUCKLAND, S. T. (1985). Perpendicular distance models for line transect sampling. *Biometrics* **41** 177–195.

CARMICHAEL, J. P. (1984). Consistency of an autoregressive density estimator. *Math. Operationsforsch. Statist. Ser. Statist.* **15** 383–387.

CENCOV, N. N. (1962). Evaluation of an unknown distribution density from observations. *Soviet Math.* **3** 1559–1562.

CLINE, D. B. H. and HART, J. D. (1986). Kernel estimation of densities with discontinuities or discontinuous derivatives. Technical Report, Dept. Statistics, Texas A & M Univ.

CRAIN, B. R., BURNHAM, K. P., ANDERSON, D. R. and LAAKE, J. L. (1979). A Fourier series estimator of population density for line transect sampling. *Biometrical J.* **21** 731–748.

DAVIS, K. B. (1977). Mean integrated square error properties of density estimates. *Ann. Statist.* **5** 530–535.

DIGGLE, P. J. and HALL, P. (1986). The selection of terms in an orthogonal series density estimator. *J. Amer. Statist. Assoc.* **81** 230–233.

GATES, C. E. and SMITH, P. W. (1980). An implementation of the Burnham–Anderson distribution-free method of estimating wildlife densities from line transect data. *Biometrics* **36** 155–160.

GRAY, H. L. (1985). On a unification of bias reduction and numerical approximation. *Graybill Festschrift*. North-Holland, Amsterdam. To appear.

GRAY, H. L., WATKINS, T. A. and ADAMS, J. E. (1972). On the jackknife statistic, its extensions, and its relation to e_n-transformations. *Ann. Math. Statist.* **43** 1–30.

HALL, P. (1983a). Measuring the efficiency of trigonometric series estimates of a density. *J. Multivariate Anal.* **13** 234–256.

HALL, P. (1983b). Large sample optimality of least squares cross-validation in density estimation. *Ann. Statist.* **11** 1156–1174.

HALL, P. (1987). Cross-validation and the smoothing of orthogonal series density estimators. *J. Multivariate Anal.* **21** 189–206.

HALL, P. and MARRON, J. S. (1987). Extent to which least-squares cross-validation minimises integrated square error in nonparametric density estimation. *Probab. Theory Related Fields* **74** 567–582.

HART, J. D. (1985). On the choice of a truncation point in Fourier series density estimation. *J. Statist. Comput. Simulation* **21** 95–125.

HART, J. D. (1986). ARMA estimators of probability densities with exponential or regularly varying Fourier coefficients. ONR Technical Report No. 2, Dept. Statistics, Texas A & M Univ.

HART, J. D. and GRAY, H. L. (1985). The ARMA method of approximating probability density functions. *J. Statist. Plann. Inference* **12** 137–152.

KRONMAL, R. A. and TARTER, M. E. (1968). The estimation of probability densities and cumulatives by Fourier series methods. *J. Amer. Statist. Assoc.* **63** 925–952.

MORTON, M. J. and GRAY, H. L. (1984). The *G*-spectral estimator. *J. Amer. Statist. Assoc.* **79** 692–701.

PARZEN, E. (1979). Nonparametric statistical data modeling. *J. Amer. Statist. Assoc.* **74** 105–121.

RUDEMO, M. (1982). Empirical choice of histogram and kernel density estimators. *Scand. J. Statist.* **9** 65–78.

SCHUCANY, W. R., GRAY, H. L. and OWEN, D. B. (1971). On bias reduction in estimation. *J. Amer. Statist. Assoc.* **66** 524–533.

SCHUSTER, E. F. (1985). Incorporating support constraints into nonparametric estimators of densities. *Comm. Statist. A—Theory Methods* **14** 1123–1136.

SCOTT, D. W. and TERRELL, G. R. (1987). Biased and unbiased cross-validation in density estimation. *J. Amer. Statist. Assoc.* **82** 1131–1146.

SENETA, E. (1976). *Regularly Varying Functions. Lecture Notes in Math.* **508.** Springer, Berlin.

SHANKS, D. (1955). Nonlinear transformation of divergent and slowly convergent sequences. *J. Math. Phys. Sci.* **34** 1–42.

STONE, C. J. (1984). An asymptotically optimal window selection rule for kernel density estimates. *Ann. Statist.* **12** 1285–1297.

WAHBA, G. (1981). Data-based optimal smoothing of orthogonal series density estimates. *Ann. Statist.* **9** 146–156.

WATSON, G. S. (1969). Density estimation by orthogonal series. *Ann. Math. Statist.* **40** 1496–1498.

WATSON, G. S. and LEADBETTER, M. R. (1963). On the estimation of the probability density. I. *Ann. Math. Statist.* **34** 480–491.

DEPARTMENT OF STATISTICS
TEXAS A & M UNIVERSITY
COLLEGE STATION, TEXAS 77843-3143