ON THE ESTIMATION OF THE PROBABILITY DENSITY, I[1]

By G. S. Watson[2] and M. R. Leadbetter

University of Toronto and Research Triangle Institute

0. Summary. Estimators of the form $\hat{f}_n(x) = (1/n) \sum_{i=1}^{n} \delta_n(x - x_i)$ of a probability density $f(x)$ are considered, where $x_1 \cdots x_n$ is a sample of n observations from $f(x)$. In Part I, the properties of such estimators are discussed on the basis of their mean integrated square errors $E[\int (f_n(x) - f(x))^2 dx]$ (M.I.S.E.). The corresponding development for discrete distributions is sketched and examples are given in both continuous and discrete cases. In Part II the properties of the estimator $\hat{f}_n(x)$ will be discussed with reference to various pointwise consistency criteria. Many of the definitions and results in both Parts I and II are analogous to those of Parzen [1] for the spectral density. Part II will appear elsewhere.

1. Introduction. Many authors have considered the problems of estimating the spectral density of a stationary time series from observations of the series throughout a time T. The corresponding problem of estimating probability densities has received less attention in the literature. The authors were prompted to make the present investigation while studying methods for estimating the hazard function $h(x) = f(x)/(1 - F(x))$, (Watson and Leadbetter [3]). Whittle [4], in estimating probability densities, considers the case where the unknown density $f(x)$ has a prior distribution whose means and covariances are known. He uses an "expected mean square error" criterion to optimize his estimators. That is to say, he considers estimators of the form $\hat{f}(x) = (1/n) \sum_{j=1}^{n} w_x(x_j)$, for a sample of n observations x_j, where $w_x(y)$ is chosen to make

$$E_p E_s[\hat{f}(x) - f(x)]^2, \tag{1.1}$$

a minimum. Here E_p and E_s denote averages over the prior distribution for $f(x)$, and the sampling fluctuations respectively.

Parzen [1] considers estimates $f_T(\omega)$, having a certain form, of a spectral density $f(\omega)$, from observations on a stationary time series, in time T. He uses various criteria in judging the suitability of the estimate $f_T(\omega)$ and in discussing its asymptotic properties as $T \to \infty$. In this paper we use these criteria in obtaining classes of estimators of a probability density. The estimators have the form considered by Whittle, but the criteria employed by Parzen will be used instead of the expectation (over the prior distributions) of the sampling mean square error. Should prior probabilities be available, many of the general results obtained will remain true for the mean taken also over the prior distribution,

---

Received October 30, 1961; revised June 11, 1962.

[1] Research sponsored by the Office of Naval Research under Contract Nonr-3423(00) with the Research Triangle Institute.

[2] Now at the Johns Hopkins University.

provided the quantities occurring are replaced by their means with respect to the prior distribution. Most of our results are analogous to results obtained by Parzen in the frequency spectra case. The correspondence is very close for the theorems in this paper. In Part II, which will be concerned with "pointwise consistency" the results are still similar in general form to those for frequency spectra, but the differences in detail are much more marked, due to the essential differences in the form of the estimators from those used in spectra estimation.

A recent paper by Parzen [2] on estimation of a probability density and mode has also some points of contact with the present work. These are mentioned below.

Explicitly the unknown probability density $f(x)$ is to be estimated by estimators of the form

$$\hat{f}_n(x) = (1/n) \sum_1^n \delta_n(x - x_i), \tag{1.2}$$

from a sample of n observations $x_1, \cdots, x_n$. $\delta_n(\cdot)$ is assumed to be a square integrable function, as is $f(x)$.

In terms of this notation we give the following definitions which are due to Parzen ([1], Section 4, etc.), and which give criteria on which judgements concerning the suitability of particular classes of estimators, may be based.

The *mean integrated square error* (M.I.S.E.) J_n, of the estimator $\hat{f}_n(x)$ is defined by

$$J_n = E\left(\int (\hat{f}_n(x) - f(x))^2 \, dx\right). \tag{1.3}$$

Integrals where no limits are written are to be taken over the entire real line. The symbol E denotes expectation. Let $H(n) \to \infty$ as $n \to \infty$.

The estimator $\hat{f}_n(x)$ is said to be *integratedly consistent of order* $H(n)$ if

(1.4) $H(n)J_n \to$ a limit which is finite and non-zero, as $n \to \infty$.

In the following sections various types of estimators are discussed and their orders of consistency investigated. The type of estimator appropriate in a given situation depends largely on the behavior of the characteristic function $\Phi_f(t)$ of the probability density f, for large t. Two main classes of probability densities are considered, consisting of those densities whose characteristic functions decrease in an "algebraic" and an "exponential" way as t increases. Corresponding to these classes, the appropriate estimators are of "algebraic type" and "exponential type" respectively. These terms are defined precisely in later sections.

In Section 2 we derive expressions for the $\delta_n(x)$ which minimizes the M.I.S.E. J_n. We sketch the corresponding derivations for the discrete case, where the problem is to estimate the quantities p_r associated with a discrete random variable X for which Prob $(X = r) = p_r$, $(r = 0, \pm 1, \pm 2, \cdots)$. Examples for both discrete and continuous random variables are given.

In Section 3 the case where the characteristic function of f decreases "algebraically" is considered. The minimum M.I.S.E. is evaluated for this case, and estimates of "algebraic type" discussed. It is shown that there is an estimate of algebraic type which has the same (M.I.S.E.) consistency properties as the "minimum M.I.S.E." estimate.

Section 4 contains a similar discussion for the case where the characteristic function of f has an "exponential rate of decrease" and estimates of "exponential type" are used.

2. Minimum M.I.S.E. estimation. Let $\hat{f}_n(x)$ be the estimator as defined in (1.2), of the square integrable frequency function $f(x)$. For any function $g(x)$, either integrable or square integrable, write $\Phi_g(t)$ for the Fourier transform $\Phi_g(t) = \int e^{ixt} g(x)\, dx$, where the integral has the usual interpretation for square integrable functions $g(x)$. From definition (1.3) and Parseval's formula,

$$J_n = E\left[\int (\hat{f}_n(x) - f(x))^2\, dx\right]; \tag{2.1}$$

$$= (1/2\pi)\, E\left[\int |\Phi_{\hat{f}_n}(t) - \Phi_f(t)|^2\, dt\right]. \tag{2.2}$$

We can proceed to minimize either (2.1) or (2.2), obtaining of course equivalent results.

Since

$$\Phi_{\hat{f}_n}(t) = \Phi_{\delta_n}(t)(1/n)\sum_{r=1}^{n} e^{ix_r t},$$

(2.2) becomes

$$2\pi J_n = E\left[\int \left| (\Phi_{\delta_n}(t)/n) \sum_{r=1}^{n} e^{ix_r t} - \Phi_f(t) \right|^2 dt\right],$$

$$= \int \Big[(1/n)\, |\Phi_{\delta_n}(t)|^2 (1 - |\Phi_f(t)|^2) + |\Phi_f(t)|^2 (|1 - \Phi_{\delta_n}(t)|^2)\Big]\, dt.$$

This may be re-written as

$$2\pi J_n = \int \left(\frac{1}{n} + \frac{n-1}{n} |\Phi_f(t)|^2\right)\left(\left|\Phi_{\delta_n}(t) - \frac{|\Phi_f(t)|^2}{(1/n) + [(n-1)/n]\,|\Phi_f(t)|^2}\right|^2\right) dt + \int \frac{|\Phi_f(t)|^2 (1 - |\Phi_f(t)|^2)}{1 + (n-1)\,|\Phi_f(t)|^2}\, dt \tag{2.3}$$

and so is minimized by taking $\Phi_{\delta_n}(t) = \Phi^*_{\delta_n}(t)$, where

$$\Phi^*_{\delta_n}(t) = \frac{|\Phi_f(t)|^2}{(1/n) + [(n-1)/n]\,|\Phi_f(t)|^2}. \tag{2.4}$$

The minimum M.I.S.E. J_n^*, corresponding to $\Phi_{\delta_n}^*$ is, from (2.3), given by

$$2\pi J_n^* = \int \frac{|\Phi_f(t)|^2 (1 - |\Phi_f(t)|^2)}{1 + (n-1)|\Phi_f(t)|^2} dt, \tag{2.5}$$

$$J_n^* = \frac{\delta_n^*(0)}{n} - \frac{1}{2\pi} \int \frac{|\Phi_f(t)|^4}{1 + (n-1)|\Phi_f(t)|^2} dt,$$

or

$$J_n^* = \delta_n^*(0)/n - O(1/n), \tag{2.6}$$

since the second term in (2.5) is dominated by $[1/2\pi(n-1)] \int |\Phi_f(t)|^2 dt$.

By similar arguments it may be shown that $\delta_n^*(x)$ satisfies the integral equation $(1/n)\delta_n^*(t) - \int f(x)f(x-t)\,dx + [(n-1)/n] \int\int f(x-s)f(x-t)\delta_n^*(x)\,ds\,dx = 0$, or writing $g(t) = \int f(x)f(x-t)\,dx$,

$$(1/n)\,\delta_n^*(t) + [(n-1)/n] \int g(t-s)\,\delta_n^*(s)\,ds = g(t). \tag{2.7}$$

Taking Fourier transforms of the quantities in this equation yields (2.4) again.

EXAMPLES. In each case $\delta_n^*(\cdot)$ was found by inverting (2.4).

(1) Cauchy distribution, $f(x) = 1/(\pi(1 + x^2))$, $\Phi_f(t) = e^{-|t|}$

$$\delta_n^*(x) = \frac{n}{n-1} \frac{e^{-\pi x/2}}{1 - e^{-\pi x}} \sin \frac{x \log (n-1)}{2} + \frac{2n}{\pi} \sum_{r=1}^{\infty} \left(\frac{-1}{n-1}\right)^{r+1} \frac{r}{x^2 + 4r^2}.$$

It follows from (2.6) that, $(n/\log n)J_n^* \to (1/2\pi)$. This is a particular case of a general theorem to be proved later (4.3), and shows that f_n^* is here integratedly consistent of order $(n/\log n)$, where f_n^* corresponds to δ_n^*.

(2) Gamma distribution, $f(x) = x^{p-1}e^{-x}/\Gamma(p)$ for p a positive integer, $\Phi_f(t) = (1 - it)^{-p}$. For $p = 1$, $\delta_n^*(x) = (n^{\frac{1}{2}}/2) \exp(-|x|n^{\frac{1}{2}})$, and for $p = 2$,

$$\delta_n^*(x) = [n^{\frac{3}{4}}/2(n-1)^{\frac{1}{2}}] \exp(-xn^{\frac{1}{4}} \sin \alpha/2) \cos(|x|n^{\frac{1}{4}} \cos \alpha/2 - \alpha/2)$$

where $\alpha = \pi - \tan^{-1}[(n-1)^{\frac{1}{2}}]$. For p a general integer, an exact result for $\delta_n^*(x)$ is difficult, but we have approximately that

$$\delta_n^*(x) = \frac{n}{p(n-1)^{1-1/2p}} \sum_{l=0}^{m} \exp\left\{-x(n-1)^{1/2p} \sin \frac{2l+1}{2p}\pi\right\} \sin\left\{x(n-1)^{1/2p} \cos \frac{2l+1}{2p}\pi + \frac{2l+1}{2p}\pi\right\}$$

where $m = p/2 - 1$ if p is even, and $(p-1)/2$ if p is odd. This result is obtained by approximating the zeros of $[(1 + t^2)^p + n - 1]$ by the points

$$t_l = (n-1)^{1/2p} e^{i\pi(2l+1)/2p}, \qquad l = 0, 1, \cdots, 2p - 1.$$

In each of these cases we see from (2.6) that $n^{1-1/2p} J_n^* \to (1/2\pi) \int_{-\infty}^{\infty} [dt/(1 + t^{2p})]$. These are again special cases of a general result (3.3) which is proved later.

(3) Consider the distribution given by $f(x) = (1 - \cos x)/(\pi x^2)$ with $\Phi_f(t) = 1 - |t|$ if $|t| \leqq 1$ and $\Phi_f(t) = 0$ otherwise.

$$\delta_n^*(x) = \frac{n}{(n-1)\pi}\left[\frac{\sin x}{x} - \int_0^1 \frac{\cos xt\, dt}{1+(n-1)(1-t)^2}\right].$$

Thus from (2.6), $nJ_n^* \to 1/\pi$ and the estimator given by δ_n^* is thus integratedly consistent of order n.

Since, from (2.5), for any $\Phi_f(t)$, $nJ_n^* \geqq \int |\Phi_f(t)|^2(1 - |\Phi_f(t)|^2)\, dt$, it follows that in no case can the order of integrated consistency be greater than n, i.e. J_n^* cannot decrease faster than $1/n$. The last example considered where $\Phi_f(t)$ is non zero only in a finite t range (i.e., f has "finite bandwidth") is a case where the order of integrated consistency of the optimum estimate is actually n. In fact it is n in all cases where the density $f(x)$ has a finite bandwidth since if $\Phi_f(t) = 0$ for $|t| \geqq L$, from (2.5), $2\pi n J_n^* \to \int_{-L}^{L} (1 - |\Phi_f(t)|^2)\chi_f(t)\, dt$ where $\chi_f(t) = 1$ if $|\Phi_f(t)| > 0$ and $\chi_f(t) = 0$ otherwise.

THE DISCRETE CASE. We now consider the corresponding minimum "M.I.S.E." problem in the discrete case. Let X be a discrete random variable such that Prob $(X = r) = p_r$, $r = (0, \pm 1 \pm 2 \cdots)$.

Then the characteristic function $\Phi_P(\theta)$ of X is given by $\Phi_P(\theta) = \sum_{r=-\infty}^{\infty} p_r e^{ir\theta}$. The quantities p_r are to be estimated from a sample of n drawn from this distribution. Denote by n_r the number of times the value $X = r$ occurs in the sample. Corresponding to (1.2), estimators of p_r will be considered with the form

$$\hat{p}_r = (1/n) \sum_{s=-\infty}^{\infty} w_n(r - s) n_s\,, \tag{2.8}$$

omitting the dependence of $\hat{p}_r$ on n, where $w_n(r)$ is such that $\sum_r w_n(r) = 1$. It is natural to define the "M.I.S.E." in this case by

$$J_n = E \sum (\hat{p}_r - p_r)^2 = (1/2\pi)\int_0^{2\pi} |\Phi_{\hat{p}}(\theta) - \Phi_P(\theta)|^2\, d\theta, \tag{2.9}$$

where $\Phi_{\hat{P}}(\theta) = \sum_{r=-\infty}^{\infty} \hat{p}_r e^{ir\theta}$. Writing $\Phi_w(\theta) = \sum_{r=-\infty}^{\infty} w_n(r) e^{i\theta r}$, it may be shown that J_n is minimized by $\Phi_w(\theta) = \Phi_{w^*}(\theta)$ such that

$$\Phi_{w^*}(\theta) = \frac{|\Phi_P(\theta)|^2}{(1/n) + [(n-1)/n]\,|\Phi_P(\theta)|^2}, \tag{2.10}$$

as in the continuous case (2.4). The minimum M.I.S.E. J_n^* is given by

$$2\pi J_n^* = \int_0^{2\pi} \frac{|\Phi_P(\theta)|^2 (1 - |\Phi_P(\theta)|^2)}{1 + (n-1)\,|\Phi_P(\theta)|^2}\, d\theta. \tag{2.11}$$

The order of (integrated) consistency of the minimizing estimate does not depend on the form of $\Phi_P(\theta)$. This is in sharp contrast with the continuous case for which the order of (integrated) consistency depends strongly on the form of the distribution as noted in the examples, and as discussed more fully in Sections 3 and 4. In the discrete case we have the result that the minimizing estimate determined by $\Phi_{w^*}(\theta)$ is integratedly consistent of order n, since from (2.11)

$$nJ_n^* \to (1/2\pi) \int_0^{2\pi} \chi_P(\theta)(1 - |\Phi_P(\theta)|^2)\, d\theta \tag{2.12}$$

by the dominated convergence theorem, where $\chi_P(\theta) = 1$ if $|\Phi_P(\theta)| > 0$, and $\chi_P(\theta) = 0$ otherwise. That is $nJ_n^* \to l/2\pi - \sum_{r=-\infty}^{\infty} p_r^2$ as $n \to \infty$ where l is the measure of the subset of $(0, 2\pi)$ for which $|\Phi_f(t)| > 0$.

EXAMPLES.

(1) *Geometric distribution* $p_r = qp^r$, $r = 0, 1, 2 \cdots$.

$$w_n(r) = (1 + 4p/nq^2)^{-\frac{1}{2}} z_1^{|r|}$$

where z_1 is the smaller root of the equation $pz^2 - z(nq^2 + 2p) + p = 0$, from which it follows at once that $J_n^* \sim 2p/[n(1 + p)]$ as $n \to \infty$, verifying (2.12) in this case.

(2) *Poisson distribution* $p_r = e^{-m}m^r/r!$ $r = 0, 1, 2 \cdots$.

$$w_n(r) = (n/n - 1) \sum_{s=0}^{\infty} [(-1)^s/(n - 1)^s] e^{2ms} I_{|r|}(-2ms),$$

where $I_r(x)$ is the Bessel function of order r with imaginary argument.

(3) *Distribution for which* $p_0 = \alpha/(2\pi)$

$$p_r = (1 - \cos \alpha r)/(\pi \alpha r^2), \qquad r = \pm 1, \pm 2, \cdots, 0 < \alpha \leqq \pi.$$

We have, similarly to Example (3) in the continuous case,

$$w_n(r) = \frac{n}{(n-1)\pi}\left[\frac{\sin r\alpha}{r} - \int_0^{\alpha} \frac{\cos r\theta\, d\theta}{1 + (n-1)\alpha^2(\alpha - \theta)^2}\right].$$

3. Characteristic functions which decrease algebraically. Discussion of the asymptotic properties of J_n^*, as $n \to \infty$, requires some knowledge of the behaviour of $\Phi_f(t)$ as $|t| \to \infty$. Following Parzen we make the following definitions: $\Phi_f(t)$, the characteristic function of f, *decreases algebraically of degree* $p > 0$ if

$$\lim_{|t| \to \infty} |t|^p |\Phi_f(t)| = K^{\frac{1}{2}} > 0. \tag{3.1}$$

An estimator $\hat{f}_n(x) = (1/n) \sum_1^n \delta_n(x - x_i)$ has *algebraic form* if $\Phi_{\delta_n}(t)$ can be written as $h(A_n t)$, where $h(t)$ is a bounded even square integrable function and $A_n \to 0$ as $n \to \infty$. This is equivalent to requiring

$$\delta_n(x) = A_n^{-1} k(A_n^{-1} x) \quad \text{with} \quad h(t) = \Phi_k(t). \tag{3.2}$$

Parzen [2] has considered this type of estimate and shown that they are consistent and asymptotically normal, as well as considering their pointwise mean square error.

In this section $\Phi_f(t)$ is assumed to decrease algebraically of degree p. The section is divided into three parts. In 3A, the order of consistency of the optimum estimate, based on δ_n^*, is investigated. In 3B, the same is done for estimates of algebraic type and, in 3C, an estimate of algebraic type is constructed with the same consistency properties as that based on δ_n^*.

3A. *Consistency properties of the optimum estimate.*

THEOREM. *Let $\Phi_f(t)$ decrease algebraically of degree $p > \frac{1}{2}$, K as in* (3.1), *then J_n^*, the minimum* M.I.S.E., *satisfies*

$$n^{1-1/2p}J_n^* \to (1/2\pi)K^{1/2p} \int [dt/(1 + |t|^{2p})] \quad \text{as} \quad n \to \infty . \tag{3.3}$$

PROOF. Choose T such that, for $|t| \geqq T$, $||t|^{-2p}|\Phi_f(t)|^{-2} - K^{-1}| < \epsilon$ where $\epsilon > 0$ is fixed.

$$n^{1-1/2p}\int_{-\infty}^{\infty} \frac{|\Phi_f(t)|^2}{1 + (n-1)|\Phi_f(t)|^2}\,dt = n^{1-1/2p}\int_{-T}^{T} \frac{|\Phi_f(t)|^2}{1 + (n-1)|\Phi_f(t)|^2}\,dt$$

$$+n^{1-1/2p}\int_{-\infty}^{\infty} \frac{dt}{(n-1) + |t|^{2p}K^{-1}} - n^{1-1/2p}\int_{-T}^{T} \frac{dt}{(n-1) + |t|^{2p}K^{-1}}$$

$$+ n^{1-1/2p}\int_{|t|>T} \left[\frac{1}{(n-1) + \dfrac{1}{|\Phi_f(t)|^2|t|^{2p}}|t|^{2p}} - \frac{1}{(n-1) + |t|^{2p}K^{-1}} \right] dt.$$

The first and third terms are dominated by $2Tn^{1-1/2p}/(n-1) \to 0$ as $n \to \infty$. The second term tends to a finite limit, namely,

$$\lim \left(\frac{n}{n-1}\right)^{1-1/2p} \int \frac{ds}{1 + |s|^{2p}K^{-1}} = K^{1/2p}\int \frac{dt}{1 + |t|^{2p}}.$$

The fourth term is dominated by

$$P\left[n^{1-1/2p}\int \frac{ds}{(n-1) + |t|^{2p}K^{-1}}\right]\epsilon,$$

where $|\Phi_f(t)|^2|t|^{2p} \leqq P$, say, a bound which exists since $|\Phi_f(t)|^2|t|^{2p}$ is continuous and tends to a limit as $|t| \to \infty$. Thus the fourth term is arbitrarily small (uniformly in n) for sufficiently large T and an application of (2.4) and (2.6) yields the desired result. Hence the estimate $\hat{f}_n(x)$ formed from δ_n^* is, when $\Phi_f(t)$ decreases algebraically of degree $p > 0$, integratedly consistent of order $n^{1-1/2p}$.

3B. *Consistency properties of an estimate of algebraic type.*

THEOREM. *Let $\hat{f}_n(x)$ be an estimate of algebraic type, and $\Phi_f(t)$ decrease algebraically of degree $p > 0$. If $\int [(1 - h(t))^2/|t|^{2p}]\,dt$ exists (where $\Phi_{\delta_n}(t) = h(A_n t)$) and if $A_n = Dn^{-1/2p}$, then*

$$n^{1-1/2p}J_n \to \frac{1}{2\pi D}\int h^2(t)\,dt + \frac{K}{2\pi}D^{2p-1}\int \frac{(1-h(t))^2}{|t|^{2p}}\,dt \tag{3.4}$$

where J_n is the M.I.S.E. *corresponding to δ_n and $K = \lim_{t\to\infty} (|\Phi_f(t)|^2|t|^{2p})$.*

PROOF. From (2.3), with $\Phi_{\delta_n}(t) = h(A_n t)$,

$$
\begin{aligned}
2\pi n^{1-1/2p} J_n &= n^{-1/2p} \int h^2(A_n t)(1 - |\Phi_f(t)|^2)\, dt \\
&\quad + n^{1-1/2p} \int |\Phi_f(t)|^2 (1 - h(A_n t))^2\, dt.
\end{aligned} \tag{3.5}
$$

The first term on the right-hand side is, for $A_n = Dn^{-1/2p}$,

$$D^{-1} \int h^2(t)\, dt - n^{-1/2p} \int h^2(A_n t) |\Phi_f(t)|^2\, dt \to D^{-1} \int h^2(t)\, dt,$$

since $h(t)$ is bounded. The second term on the right-hand side of (3.5) becomes

$$KD^{-1} \int \frac{(1 - h(t))^2}{|t|^{2p} D^{-2p}}\, dt + D^{-1} \int \frac{(1 - h(t))^2}{|t|^{2p}} \left[\left\{ n^{\frac{1}{2}} |\Phi_f(D^{-1} n^{1/2p} t)|\, |t|^p \right\}^2 - \frac{K}{D^{-2p}} \right] dt,$$

where the second integral tends to zero by the dominated convergence theorem. The right-hand side of (3.5) therefore converges, as $n \to \infty$, to $D^{-1} \int h^2(t)\, dt + KD^{2p-1} \int [(1 - h(t))^2/|t|^{2p}]\, dt$, which completes the proof.

This theorem shows that an estimate of algebraic type is integratedly consistent of order $n^{1-1/2p}$, if Φ_f decreases algebraically of degree p. It corresponds to part of Parzen's result (5.28) [1].

The remainder of the result corresponding to Parzen's (5.28) may be stated as follows:

THEOREM. *Under the conditions of the previous theorem, except that* $A_n = Dn^{-1/2r}$, *then, for* $0 < r < p$, $n^{1-1/2r} J_n \to (1/2\pi D) \int h^2(t)\, dt$.

This is proved in the same way as the previous theorem and states that f_n is integratedly consistent of order $n^{1-1/2r}$ for $r < p$.

3C. *An estimator of algebraic type, with the asymptotic optimum property.* If $\Phi_f(t)$ is known to decrease algebraically, of degree p, an estimator can be found whose M.I.S.E. J_n satisfies the same (M.I.S.E.) consistency relation (3.3) as does J_n^*. This means that if the asymptotic efficiency of the estimator is defined as $\lim_{n\to\infty} [J_n^*/J_n]$ then this estimate has asymptotic efficiency 1. The estimator is of algebraic type with $h(t) = (1 + |t|^{2p})^{-1}$, $A_n = Dn^{-1/2p}$. By (3.4) it follows that

$$n^{1-1/2p} J_n \to \frac{1}{2\pi D} \int \frac{1 + KD^{2p}|t|^{2p}}{(1 + |t|^{2p})^2}\, dt$$

and $D = K^{-1/2p}$ yields (3.3).

4. Characteristic functions which decrease exponentially. The following definitions for this case corresponding to those in Section 3, are again very similar to those used by Parzen.

The characteristic function $\Phi_f(t)$ is said to decrease exponentially, of coefficient $\rho > 0$ if

(4.1a) $$|\Phi_f(t)| \leqq Ae^{-\rho|t|}, \quad \text{for some constant } A \text{ and all } t.$$

(4.1b) $$\lim_{v\to\infty} \int_0^1 [1 + e^{2\rho v}|\Phi_f(vt)|^2]^{-1}\, dt = 0.$$

This definition includes a wide variety of characteristic functions which would usually be considered as having an exponential rate of decrease.

An estimate $\hat{f}_n(x)$ formed from $\delta_n(x)$ is said to have exponential form if

(4.2) $$\Phi_{\delta_n}(t) = h(A_n e^{\alpha|t|}) \quad \text{where } A_n \to 0 \text{ as } n \to \infty, \qquad \alpha > 0,$$

and h is as before, bounded and square integrable. The remainder of this section is divided into four parts, the first three corresponding to those in Section 3. The fourth part deals with an extension, which includes the case of normal random variables, to a generalized type of exponential decrease.

4A. *Consistency properties of the optimum estimate.*

THEOREM. *Let $\Phi_f(t)$ decrease exponentially of coefficient ρ. Then J_n^*, the minimum* M.I.S.E. *satisfies*

(4.3) $$\lim_{n\to\infty} (n/\log n) J_n^* = 1/(2\pi\rho).$$

PROOF. We have

(4.4) $$\left| \int \frac{|\Phi_f(t)|^2}{1 + (n-1)|\Phi_f(t)|^2}\, dt - \int \frac{e^{-2\rho|t|}}{1 + (n-1)e^{-2\rho|t|}}\, dt \right| \leq 2(1 + A) \int_0^\infty \frac{e^{-2\rho t}}{[1 + (n-1)|\Phi_f(t)|^2][1 + (n-1)e^{-2\rho t}]}\, dt,$$

using (4.1a), since $|\Phi_f(t)|^2$ is even. But this is not greater than

$$2(1 + A)\left[\frac{1}{n-1}\int_0^{\frac{\log(n-1)}{2\rho}} \frac{dt}{1 + (n-1)|\Phi_f(t)|^2} + \int_{\frac{\log(n-1)}{2\rho}}^\infty e^{-2\rho t}\, dt\right],$$

$$= 2(1 + A)\left[\frac{1}{n-1}\frac{\log(n-1)}{2\rho}\int_0^1 \frac{ds}{1 + e^{2\rho v}|\Phi_f(vs)|^2} + \frac{1}{2\rho(n-1)}\right],$$

(where $v = \log(n-1)/(2\rho)$, $t = vs$), and this expression tends to zero as $n \to \infty$. We have also that

(4.5) $$\frac{n}{\log n}\int \frac{e^{-2\rho|t|}}{1 + (n-1)e^{-2\rho|t|}}\, dt \to \frac{1}{\rho},$$

and hence by (4.4)

(4.6) $$\frac{n}{\log n}\int \frac{|\Phi_f(t)|^2}{1 + (n-1)|\Phi_f(t)|^2}\, dt \to \frac{1}{\rho} \quad \text{as} \quad n \to \infty.$$

Finally, using (2.5), the result follows.

Hence if $\Phi_f(t)$ decreases exponentially of coefficient ρ, then J_n^* is integratedly consistent of order $n/\log n$.

4B. *Consistency properties of an estimate of exponential type.* Let $\hat{f}_n(x)$ be an estimate of exponential type. That is, $\Phi_{\delta_n}(t) = h(A_n e^{\alpha|t|})$. In addition to being bounded (by the bound B), and square integrable, suppose that $h(t)$ satisfies the following condition:

$$(4.7) \qquad |1 - h(t)| \leqq B_1|t| \quad \text{for} \quad |t| \leqq 1.$$

Then the following result holds.

THEOREM. *Let $\Phi_f(t)$ satisfy* (4.1a), *and let $\hat{f}_n(x)$ be an estimate of exponential type such that $h(t)$ satisfies* (4.7). *Let $A_n = Dn^{-b}$ for $b > \frac{1}{2}$ and $\alpha \leqq 2\rho b$. Then the* M.I.S.E. *J_n corresponding to $\hat{f}_n(x)$ satisfies*

$$(4.8) \qquad \lim_{n\to\infty} [(n/\log n)J_n] = (1/2\pi)(2b/\alpha).$$

To prove this we establish first two lemmas.

LEMMA 1. *Under the conditions of the theorem,*

$$(4.9) \qquad (1/\log n) \int_{Dn^{-b}}^{\infty} (h^2(t)/t)\, dt \to b \quad \text{as} \quad n \to \infty.$$

For this integral may be written as

$$\frac{1}{\log n}\int_{Dn^{-b}}^{1} \frac{dt}{t} - \frac{1}{\log n}\int_{Dn^{-b}}^{1} \frac{1 - h^2(t)}{t}\, dt + \frac{1}{\log n}\int_{1}^{\infty} \frac{h^2(t)}{t}\, dt.$$

The first term is $b - \log D/\log n \to b$ as $n \to \infty$. Using (4.7) it follows that the second and third terms tend to zero. Hence, the result follows.

LEMMA 2. *Again under the conditions of the theorem,*

$$(4.10) \qquad (n/\log n) \int |\Phi_f(t)|^2(1 - \Phi_{\delta_n}(t))^2\, dt \to 0 \quad \text{as} \quad n \to \infty.$$

For this expression is equal to

$$\frac{2n}{\log n}\int_{Dn^{-b}}^{\infty} \left|\Phi_f\left(\frac{1}{\alpha}\log\left(\frac{n^b}{D}x\right)\right)\right|^2 \frac{(1 - h(x))^2}{\alpha x}\, dx,$$

$$\leqq \frac{2n}{\log n}\int_{Dn^{-b}}^{1} A \exp\left[-2\frac{\rho}{\alpha}\log\left(\frac{n^b x}{D}\right)\right] B_1^2 \frac{x}{\alpha}\, dx$$

$$+ \frac{2n}{\log n}\int_{1}^{\infty} (1 + B)^2 A \exp\left[-2\frac{\rho}{\alpha}\log\left(\frac{n^b x}{D}\right)\right]\frac{dx}{\alpha x}.$$

Each term tends to zero by virtue of the assumptions $\alpha \leqq 2\rho b$, $b \geqq \frac{1}{2}$, which completes the proof of Lemma 2.

PROOF OF THE THEOREM. From (2.3)

$$2\pi \frac{n}{\log n} J_n = \frac{1}{\log n}\int h^2(Dn^{-b}e^{\alpha|t|})\, dt - \frac{1}{\log n}\int h^2(Dn^{-b}e^{\alpha|t|})|\Phi_f(t)|^2\, dt$$

$$+ \frac{n}{\log n}\int |\Phi_f(t)|^2(1 - \Phi_{\delta_n}(t))^2\, dt.$$

The first term is $(2/\log n)\int_{Dn^{-b}}^{\infty} h^2(x)(dx/\alpha) \rightarrow 2b/\alpha$ by Lemma 1. The second integral tends to zero since it is dominated by $(B^2/\log n)\int |\Phi_f(t)|^2\, dt$. The third integral tends to zero by Lemma 2.

This completes the proof of the theorem which shows that the estimate $\hat{f}_n(x)$ of exponential type, is consistent of order $n/\log n$.

4C. *Estimates of exponential type with the asymptotic optimum property.* From (4.8) any estimate of exponential type, with $h(t)$ satisfying (4.7) and $A_n = Dn^{-b}$ $(b > \frac{1}{2},\ \alpha \leqq 2\rho b)$, has M.I.S.E. J_n such that

$$\lim_{n\rightarrow\infty} [(n/\log n)J_n] = (1/2\pi)(2b/\alpha).$$

Hence any such estimate with $\alpha = 2\rho b$ is an estimate with the asymptotic optimum property (4.3), a result which is again analogous to Parzen's (5.16) in the case of estimation of a spectral density.

4D. *Generalization of exponential decrease.* The definition of exponential type decrease has an extension which includes the case of normal random variables. This extension is to the case of characteristic functions which decrease exponentially of degree r and coefficient ρ. A characteristic function $\Phi_f(t)$ is said to decrease in this way if the following conditions hold:

(i) $|\Phi_f(t)| \leqq Ae^{-\rho|t|^r}$ for some constants $A > 0$, $\rho > 0$ and $0 < r \leqq 2$.

(ii) $$\int_0^1 \frac{dt}{1 + \exp(2\rho v^r)|\Phi_f(tv)|^2} \rightarrow 0 \quad \text{as} \quad v \rightarrow \infty.$$

We can generalize the definition of estimators of exponential type so that they are suitable for this case and similar results to those in Section 4 then hold. For example the following result is the extension of 4.3.

THEOREM. *Let $\Phi_f(t)$ decrease exponentially of coefficient ρ and degree r. Then J_n^*, the minimum* M.I.S.E. satisfies

$$\lim_{n\rightarrow\infty} [n/(\log n)^{1/r}]J_n^* = [1/\pi(2\rho)^{1/r}].$$

This is proved in a similar way to (4.3).

General conclusions. The form of the optimum estimate (in the minimum M.I.S.E. sense) depends heavily on the behavior of Φ_f. However, in the cases where something is known of the behavior of $\Phi_f(t)$ as t becomes large it is sometimes possible to obtain estimates with the same (M.I.S.E.) consistency properties as the optimum estimate. That this is true in the cases where Φ_f decreases algebraically and exponentially respectively is seen from Sections 3C and 4C. It is evident also that the M.I.S.E. consistency results for estimates of exponential type are not strongly dependent on the form of the function h, when Φ_f decreases exponentially. In the case of estimates of algebraic type, for Φ_f decreasing algebraically, we need to know more about the precise form of the function h to discuss the M.I.S.E. consistency relations, than we do in the exponential case.

REFERENCES

[1] PARZEN, E. (1958). On asymptotically efficient consistent estimates of the spectral density of a stationary time series. *J. Roy. Statist. Soc. Ser. B* **20** 303–322.

[2] PARZEN, E. (1962). On estimation of a probability density and mode. *Ann. Math. Statist.* **33** 1065–1076.

[3] WATSON, G. S. and LEADBETTER, M. R. (1961). Hazard analysis. Submitted to *Biometrika*.

[4] WHITTLE, P. (1958). On the smoothing of probability density functions. *J. Roy. Statist. Soc. Ser. B* **20** 334–343.