Ann. Inst. Statist. Math.
34 (1982), Part A, 39-53

NONPARAMETRIC ESTIMATION OF THE LOCATION AND SCALE PARAMETERS BASED ON DENSITY ESTIMATION

IBRAHIM A. AHMAD

(Received Mar. 24, 1981)

Summary

By representing the location and scale parameters of an absolutely continuous distribution as functionals of the usually unknown probability density function, it is possible to provide estimates of these parameters in terms of estimates of the unknown functionals.

Using the properties of well-known methods of density estimates, it is shown that the proposed estimates possess nice large sample properties and it is indicated that they are also robust against dependence in the sample. The estimates perform well against other estimates of location and scale parameters.

1. Introduction and estimation

Let $f(x, \mu, \sigma)$ and $g(x)$ be two probability density functions (p.d.f.'s) such that

$$(1) \qquad f(x, \mu, \sigma)=\sigma^{-1}g((x-\mu)/\sigma) .$$

Note that it is easily seen that

$$(2) \qquad \sigma=\int_{-\infty}^{\infty} g^2(x)dx\Big/\int_{-\infty}^{\infty} f^2(x, \mu, \sigma)dx ,$$

and

$$(3) \qquad \mu=\left[\int_{-\infty}^{\infty} xf^2(x, \mu, \sigma)dx-\int_{-\infty}^{\infty} xg^2(x)dx\right]\Big/\int_{-\infty}^{\infty} f^2(x, \mu, \sigma)dx .$$

Since, generally speaking, μ and σ are unknown, the quantities $I=\int_{-\infty}^{\infty} f^2(x, \mu, \sigma)dx$ and $J=\int_{-\infty}^{\infty} xf^2(x, \mu, \sigma)dx$ are unknown. Thus one way

AMS 1980 subject classification: Primary 62G05; Secondary 60F05.
Key words and phrases: Location and scale parameters, nonparametric density estimation, kernel methods, orthogonal series methods, empirical distributions, large sample theory, stationary mixing.

to estimate μ and σ is to estimate I and J and then plug these estimates into (2) and (3), respectively. Note that it is assumed that $\int_{-\infty}^{\infty} g^2(x)dx$ and $\int_{-\infty}^{\infty} xg^2(x)dx$ are known. Since I and J are functionals of the unknown p.d.f. a self-suggesting technique to estimate I and J is to non-parametrically estimate f and plug this estimate into I and J. An equivalent representation of I and J suggests a simpler but equivalent method. Write $I=\int_{-\infty}^{\infty} f(x, \mu, \sigma)dF(x, \mu, \sigma)$ and $J=\int_{-\infty}^{\infty} xf(x, \mu, \sigma)dF(x, \mu, \sigma)$. Thus we propose to estimate I and J by:

$$I=\int_{-\infty}^{\infty} \hat{f}(x)dF_n(x) , \tag{4}$$

and

$$J=\int_{-\infty}^{\infty} x\hat{f}(x)dF_n(x) \tag{5}$$

where $\hat{f}(x)$ is an estimate of $f(x)$ and $F_n(y)=\frac{1}{n}\sum_{j=1}^{n} I(X_j\leqq y)$ is the empirical distribution function (d.f.) of the d.f. $F(x)$. Note that both I and J are special cases of a more general functional M given by:

$$M=\int_{-\infty}^{\infty} \gamma(x)f(x, \mu, \sigma)dF(x, \mu, \sigma) , \tag{6}$$

where $\gamma(x)$ is a measurable function. Thus we may estimate M by:

$$M=\int_{-\infty}^{\infty} \gamma(x)\hat{f}(x)dF_n(x) . \tag{7}$$

Unless otherwise specified all undefined integrals extend over the entire real line. So all is remained is to find a suitable way to estimate $f(x, \mu, \sigma)$. The literature is rich with different ways of estimation p.d.f.'s from nonparametric viewpoint. A recent monograph on the subject is by Tapia and Thompson [15]. One technique for nonparametric density estimation that finds wide application is the so-called "kernel method" originated by Rosenblatt [11] and later developed by many authors, Parzen [10] and Murthy [9], among others. Let $X_1, \cdots, X_n$ be a random sample from $f(x, \mu, \sigma)$, and let $\{a_n\}$ be a sequence of real numbers such that $a_n \to 0$ as $n\to\infty$ (further conditions on $\{a_n\}$ will be stated in the sequel). Further let $k(\cdot)$ be a known Borel measurable function satisfying the following conditions:

$$k(u)\geqq 0 , \quad \int k(u)du=1 , \tag{8}$$
$$|u|k(u)\to 0 \text{ as } |u|\to\infty, \quad \text{and} \quad \sup_u k(u)<\infty .$$

The kernel estimation of $f(x, \mu, \sigma)$ is given by:

$$(9) \qquad \hat{f}_1(x)=a_n^{-1}\int k[(x-u)/a_n]dF_n(u)=(na_n)^{-1}\sum_{j=1}^{n} k[(x-X_j)/a_n] .$$

Using the estimate (9) of $f(x)$ in (7) we get

$$(10) \qquad \hat{M}_1=a_n^{-1}\int\int \gamma(x)k[(x-u)/a_n]dF_n(x)dF_n(u) =(n^2a_n)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} \gamma(X_i)k[(X_i-X_j)/a_n] .$$

Note that the estimate (10) is equivalent to the estimate W_n proposed and studied by Higgins and Tichenor [7] where,

$$(11) \qquad W_n=[n(n-1)a_n]^{-1}\sum\sum_{i\neq j} a(X_i, X_j)k[(X_i-X_j)/a_n] .$$

Thus all properties derived in the present paper do apply to W_n and thus the conditions imposed by Higgins and Tichenor [7], see their Theorem 3, for the asymptotic normality can be weakened, see our Theorem 2.3, and we can demonstrate that W_n is weakly and strongly consistent under the conditions of Theorems 2.1 and 2.2 to follow. The large sample results of $\hat{M}_1$ are presented in Section 2.

When $f(x)$ is square integrable (i.e. $M<\infty$ when $\gamma(x)=1$) then another method for density estimation may be used in estimating M, namely we can use an orthogonal series expansion, since whenever $f(x)$ is square integrable then

$$(12) \qquad \hat{f}(x)=\sum_{j=0}^{\infty} \hat{\theta}_j\phi_j(x) ,$$

where $\hat{\theta}_j=\int f(x)\phi_j(x)dx$, $j\geqq 0$ and $\{\phi_j(\cdot)\}$ is an orthonormal basis of $f(x)$. We estimate $f(x)$ by (see Schwartz [13], Cencov [6], and Kronmal and Tarter [8]):

$$(13) \qquad \hat{f}_2(x)=\sum_{j=0}^{q(n)} \hat{\theta}_j\phi_j(x) ,$$

where $\hat{\theta}_j=n^{-1}\sum_{i=1}^{n}\phi_j(X_i)$, $j\geqq 0$, and $q(n)$ is an integer-valued function such that $q(n)\to\infty$ as $n\to\infty$. We assume that $|\phi_j(x)|<C(j+1)^{-\gamma}$, for some $\gamma\geqq 0$ and a positive constant C. Note that a popular choice of $\phi_j(x)$ is the normalized Hermite polynomials:

$$(14) \qquad \phi_j(x)=(2^j j!\sqrt{\pi})^{-1/2}e^{-x^2/2}H_j(x) ,$$

where $H_j(x)=(-1)^j e^{x^2/2}(d^j/dx^j)e^{-x^2/2}$, $j\geqq 0$. For this choice $|\phi_j(x)|<C_1(j+1)^{-1/4}$ if $f(x)$ has compact support and $|\phi_j(x)|<C(j+1)^{-1/12}$ in general

(see Szego [14]). We can write $f(x)$ as follows:

$$\hat{f}_2(x)=\frac{1}{n}\sum_{i=1}^{n}\Phi_n(x, X_i) , \tag{15}$$

where $\Phi_n(x, y)=\sum_{j=0}^{q(n)}\phi_j(x)\phi_j(y)$. Thus another estimate of M is given by:

$$\hat{M}_2=\int\gamma(x)\hat{f}_2(x)dF_n(x) . \tag{16}$$

The large sample theory of $\hat{M}_2$ is presented in Section 3. Some concluding remarks are given in Section 4 indicating how the large sample properties of $\hat{M}_1$ and $\hat{M}_2$ may be preserved if the observations $X_1,\cdots, X_n$ are no longer independent and also the performance of the estimates $\hat{M}_1$ and $\hat{M}_2$ relative to other methods is indicated.

2. Large-sample theory of $\hat{M}_1$

THEOREM 2.1. *Assume that* $\{a_n\}$ *is such that* $na_n^2\to\infty$ *as* $n\to\infty$ *and suppose that* $\mathrm{E}\,|\gamma(X_1)|^2<\infty$. *Then*

$$\mathrm{E}\,|\hat{M}_1-M|\to 0 \qquad as\ n\to\infty . \tag{17}$$

Remark. Note that under the condition of Theorem 2.1, $\hat{M}_1\to M$ in probability, as $n\to\infty$.

PROOF. Note that

$$\begin{aligned}
\mathrm{E}\,|\hat{M}_1-M| \quad & \\
=\mathrm{E}\,&\left|a_n^{-1}\int\int\gamma(x)k[(x-u)/a_n]dF_n(x)dF_n(u)-\int\gamma(x)f(x)dF(x)\right| \\
\leqq\mathrm{E}\,&\left|a_n^{-1}\int\int\gamma(x)k[(x-u)/a_n]dF_n(x)dF_n(u)\right. \\
&\left.-a_n^{-1}\int\int\gamma(x)k[(x-u)/a_n]dF_n(x)dF(u)\right| \\
+\mathrm{E}\,&\left|a_n^{-1}\int\int\gamma(x)k[(x-u)/a_n]dF_n(x)dF(u)\right. \\
&\left.-a_n^{-1}\int\int\gamma(x)k[(x-u)/a_n]dF(x)dF(u)\right| \\
+&\left|\int\gamma(x)f(x)dF(x)-a_n^{-1}\int\int\gamma(x)k[(x-u)/a_n]dF(x)dF(u)\right| \\
=\mathrm{E}\,&I_1+\mathrm{E}\,I_2+I_3 , \qquad \text{say} .
\end{aligned} \tag{18}$$

Thus to prove the theorem we need to prove that $\mathrm{E}\,I_1\to 0$, $\mathrm{E}\,I_2\to 0$, and $I_3\to 0$ as $n\to\infty$.

(19) $$I_3=\left|\int \gamma(x)[\mathrm{E}\,\hat{f}_1(x)-f(x)]dF(x)\right|\to 0 \quad \text{as } n\to\infty\,,$$

since $\mathrm{E}\,\hat{f}_1(x)-f(x)\to 0$ as $n\to\infty$ at all continuity points x of f whenever $na_n\to\infty$ (Theorem 1A of Parzen [10]), and since $\gamma(x)[\mathrm{E}\,\hat{f}_1(x)-f(x)]\cdot f(x)\leqq\gamma(x)[\mathrm{E}\,\hat{f}_1(x)+f(x)]f(x)$ which is integrable and converges to an integrable function $2\gamma(x)f^2(x)$, implies that the Lebesgue dominated convergence theorem (LDCT) applies (Royden [12], p. 89) and (12) holds. Next, with $g_n(x)=\gamma(x)\,\mathrm{E}\,\hat{f}_1(x)$, we have

(20) $$\begin{aligned}\mathrm{E}\,I_2&=\mathrm{E}\left|\int \gamma(x)\,\mathrm{E}\,\hat{f}_1(x)dF_n(x)-\int\gamma(x)\,\mathrm{E}\,\hat{f}_1(x)dF(x)\right|\\&\leqq[n^{-1}\,\mathrm{Var}\,g_n(X_1)]^{1/2}\,,\end{aligned}$$

which converges to 0 as $n\to\infty$, since $\mathrm{Var}\,g_n(X_1)=\mathrm{E}\,g_n^2(X_1)-[\mathrm{E}\,g_n(X_1)]^2$ and since for any $r\geqq 1$, $\mathrm{E}\,g_n^r(X_1)=\int g_n^r(x)f(x)dx\to\int\gamma^r(x)f^2(x)dx$ as $n\to\infty$, from Theorem 1A of Parzen [10] and the LDCT (assuming that $\int\gamma^r(x)\cdot f^2(x)dx<\infty$). Finally, we have

(21) $$\begin{aligned}\mathrm{E}\,I_1&=\mathrm{E}\left|\int\gamma(x)\hat{f}_1(x)dF_n(x)-\int\gamma(x)\,\mathrm{E}\,\hat{f}_1(x)dF_n(x)\right|\\&\leqq\mathrm{E}\left[\{\sup_x|\hat{f}_1(x)-\mathrm{E}\,\hat{f}_1(x)|\}\int|\gamma(x)|dF_n(x)\right]\\&\leqq\mathrm{E}^{1/2}\,\{\sup_x|\hat{f}_1(x)-\mathrm{E}\,\hat{f}_1(x)|\}^2\,\mathrm{E}^{1/2}\left\{n^{-1}\sum_{i=1}^n|\gamma(X_i)|\right\}^2,\end{aligned}$$

which converges to 0 as $n\to\infty$, from Theorem 3A of Parzen [10]. This completes the proof. QED

Our next result provides for the strong consistency of the estimate $\hat{M}_1$.

THEOREM 2.2. *Assume that* $\{a_n\}$ *are such that for any* $\varepsilon>0$, $\sum_{n=1}^{\infty}\exp(-\varepsilon na_n^2)<\infty$ *and that* k *is a function of bounded variation. Suppose that* $\mathrm{E}\,|\gamma(X_1)|<\infty$. *Then*

(22) $$\hat{M}_1\to M \text{ with probability one } (w.p.\ 1) \text{ as } n\to\infty\,.$$

PROOF. From the proof of Theorem 2.1 we need only to show that $I_i\to 0$ w.p. 1, $i=1,2$.

(23) $$\begin{aligned}I_1&\leqq\sup_x|\hat{f}_1(x)-\mathrm{E}\,\hat{f}_1(x)|\int|\gamma(x)|dF_n(x)\\&\leqq(\mu/a_n)\sup_x|F_n(x)-F(x)|\int|\gamma(x)|dF_n(x)\,,\end{aligned}$$

where the last inequality is obtained by integration by parts with μ the total variation of $k(\cdot)$. Since $\int|\gamma(x)|dF_n(x)\to\int|\gamma(x)|dF(x)$ w.p. 1 as $n\to\infty$ and note that for any $\varepsilon>0$

$$\sum_{n=1}^{\infty} \mathrm{P}\,[\sup_x |F_n(x)-F(x)|>(\varepsilon a_n/\mu)]\leqq 2\sum_{n=1}^{\infty}\exp(-n\varepsilon^2 a_n/\mu^2)<\infty\ . \tag{24}$$

Thus $I_1\to 0$ w.p. 1 as $n\to\infty$. Next, note that if $g_n(x)=\gamma(x)\,\mathrm{E}\,\hat{f}_1(x)$, then

$$I_2=\left|n^{-1}\sum_{j=1}^{n} g_n(X_j)-\mathrm{E}\,g_n(X_1)\right|=\left|n^{-1}\sum_{j=1}^{n} Y_{nj}-\mathrm{E}\,Y_{nj}\right|\ ,\qquad \text{say}\ , \tag{25}$$

which converges to 0 w.p. 1 since $\mathrm{E}\,Y_{nj}\to\mathrm{E}\,\gamma(X_1)$ as $n\to\infty$ and $Y_{nj},\cdots,Y_{nn}$ are independent identically distributed random variables, thus the strong law of large number applies. QED

Remark 2.1. Note that the condition $\mathrm{E}\,\gamma^2(X_1)<\infty$ $(\mathrm{E}\,|\gamma(X_1)|<\infty)$ required in the proof of Theorem 2.1 (Theorem 2.2) is not satisfied in the Cauchy density case and thus a different proof is needed when estimating the location and scale parameters of the Cauchy density.

Remark 2.2. An iterated logarithm law-type result is possible to derive from Theorem 2.2 as follows: let $\Delta_n=\sup_x|F_n(x)-F(x)|$. Then it is easily seen that with probability one $I_1\leqq Ca_n\Delta_n\leqq Ca_n n^{1/2}(\log\log n)^{-1/2}$, also with probability one, $I_2\leqq\left|n^{-1}\sum_{j=1}^{n}\gamma(X_j)-\mathrm{E}\,\gamma(X_1)\right|\leqq Cn^{1/2}(\log\log n)^{-1/2}$, and finally if $f(x)$ has bounded first derivative and $\int|u|k(u)du<\infty$, then $I_3\leqq Ca_n$. Collecting terms we see that with probability one $|\hat{M}_1-M|\leqq C\max(n^{1/2}a_n(\log\log n)^{-1/2},\,a_n)$.

Remark 2.3. Note that no conditions on f (except the moments conditions in Theorems 2.1 and 2.2) are imposed which means that $\hat{M}_1$ leads to competing estimates of the location and scale estimates to those already available in the literature that are consistent under mild conditions and are asymptotically normal (see Theorem 2.3 to follow). The assumptions imposed on the kernel $k(\cdot)$ and the sequence $\{a_n\}$ are common and are satisfied by a large number of kernels and sequences, respectively, see Tapia and Thompson [15] for details.

Remark 2.4. It is possible to improve on the result of Theorem 2.2 and obtain a rate if we allow for some stronger conditions. Precisely if in addition to the assumptions in Theorem 2.2; $k(u)$ is such that $\int u^r k(u)du=0$, $r=1,2,\cdots,m-1$ and $\int u^m k(u)du<\infty$, f has bounded derivative of order m, and let $\{a_n\}$ and $\{b_n\}$ be two sequences of real numbers such that $a_n^m b_n = o(1)$ and for any $\varepsilon > 0$, $\sum_{n=1}^{\infty} \exp(-\varepsilon n a_n^2/b_n^2) < \infty$, then

$$b_n |\hat{M}_1 - M| \to 0 \qquad \text{w.p. 1 as } n \to \infty .$$

PROOF. Proceeding as in Theorem 2.2 we have with probability one

$$\begin{aligned} b_n|\hat{M}_1 - M| &\leqq (\mu b_n/a_n) \sup_x |F_n(x) - F(x)| \int |\gamma(x)| dF_n(x) \\ &\quad + b_n \sup_x |\mathrm{E}\hat{f}_1(x) - f(x)| \left[\int |\gamma(x)| dF_n(x) + \int |\gamma(x)| dF(x) \right] \\ &\leqq \{(\mu b_n/a_n) \sup_x |F_n(x) - F(x)| + 2b_n \sup_x |\mathrm{E}\hat{f}_1(x) - f(x)|\} \\ &\quad \cdot \int |\gamma(x)| dF(x) . \end{aligned}$$

But since $b_n \sup_x |\mathrm{E}\, f(x) - f(x)| = O(b_n a_n^m) = o(1)$, which follows by using Taylor's expansion with the fact that $\int u^r k(u) du = 0$, $r = 1, \cdots, m-1$ and $\int |u^m| k(u) du < \infty$, and since for any $\varepsilon > 0$

$$\sum_{n=1}^{\infty} \mathrm{P}\,[\sup_x |F_n(x) - F(x)| \geqq (\varepsilon a_n/\mu b_n)] \leqq 2 \sum_{n=1}^{\infty} \exp\{-(\varepsilon^2 a_n^2/\mu^2 b_n^2)\} < \infty , \tag{26}$$

the remark is proved. QED

Remark 2.5. Another possible way to estimate of M can be given as follows

$$\tilde{M}_1 = \int \gamma(x) \hat{f}_1^2(x) dx . \tag{27}$$

The estimate is of nature similar to that of Bhattacharyya and Roussas [5]. It is possible to show that $\mathrm{E}\,|\tilde{M}_1 - M| \to 0$ under the conditions of Theorem 2.1 and that $|\tilde{M}_1 - M| \to 0$ w.p. 1 under the condition of Theorem 2.2. The proof of the former follows exactly as in Bhattacharyya and Roussas [5] and is left to the interested reader, while the second follows from the following;

$$\begin{aligned} |\tilde{M}_1 - M| &\leqq \sup_x |\hat{f}_1(x) - \mathrm{E}\hat{f}_1(x)| \int |\gamma(x)| [\mathrm{E}\hat{f}_1(x) + \hat{f}_1(x)] dx \\ &\quad + \int |\gamma(x)| |(\mathrm{E}\hat{f}_1(x))^2 - (f(x))^2| dx , \end{aligned}$$

which converges to 0 with probability one as $n \to \infty$ as in Theorem 2.2.

Note also that the result of Remark 2.1 continues to hold for $\tilde{M}_1$. One major drawback of the estimate $\tilde{M}_1$ is its asymptotic distribution. While as we shall demonstrate in Theorem 2.3, $\hat{M}_1$ is asymptotically normal under mild conditions, the asymptotic distribution of $\tilde{M}_1$ is not easily obtainable.

THEOREM 2.3. *If $na_n^2 \to \infty$ and $na_n^{2m} \to 0$ as $n \to \infty$ for some integer $m \geqq 2$, if σ^2 (given by (34)) is positive and finite, if f has m derivatives with bounded mth derivative, and if $\int u^r k(u)du = 0$, $r = 1, \cdots, m-1$ and $\int |u|^m k(u)du < \infty$, then*

$n^{1/2}(\hat{M}_1 - \hat{M})$ is asymptotically normal with mean 0 and variance $2\sigma^2$.

PROOF. We divide the argument into two parts, first we prove that $n^{1/2}(\hat{M}_1 - \mathrm{E}\,\hat{M}_1)$ is asymptotically normal with one and 0 and variance $2\sigma^2$ and then we show that $n^{1/2}(\mathrm{E}\,\hat{M}_1 - M) \to 0$ as $n \to \infty$, now,

$$(28) \quad \begin{aligned} \hat{M}_1 - \mathrm{E}\,\hat{M}_1 &= a_n^{-1} \int\int \gamma(x) k[(x-u)/a_n] dF_n(x) dF_n(u) \\ &\quad - a_n^{-1} \int\int \gamma(x) k[(x-u)/a_n] dF(x) dF(u) \\ &\quad - (na_n)^{-1} k(0)\, \mathrm{E}\,\gamma(X_1) \\ &= a_n^{-1} \int\int \gamma(x) k[(x-u)/a_n] d[F_n(x) - F(x)] d[F_n(u) - F(u)] \\ &\quad + \Big[a_n^{-1} \int\int \gamma(x) k[(x-u)/a_n] dF_n(x) dF(u) \\ &\quad\quad + a_n^{-1} \int\int \gamma(x) k[(x-u)/a_n] dF(x) dF_n(u) \\ &\quad\quad - 2a_n^{-1} \int\int \gamma(x) k[(x-u)/a_n] dF(x) dF(u) \Big] \\ &\quad - (na_n)^{-1} k(0)\, \mathrm{E}\,\gamma(X_1) = A_n + B_n + C_n, \qquad \text{say}. \end{aligned}$$

Now, clearly if $na_n^2 \to \infty$, then $n^{1/2} C_n \to 0$ as $n \to \infty$. Next, let $d[I(y \leqq x)] = I(y \leqq x + dx) - I(y \leqq x)$ and set $W_j(x) = dI(X_j \leqq x)$ and $W_j(u) = dI(X_j \leqq u)$ so that $\mathrm{P}[W_j(x) = 1] = dF(x)$ and $\mathrm{P}[W_l(u) = 1] = dF(u)$, $j = 1, \cdots, n$. Hence

$$(29) \quad \begin{aligned} & n^2 |\mathrm{E}\, d[F_n(x) - F(x)] d[F_n(u) - F(u)]| \\ &= \left| \mathrm{E} \left\{ \sum_{j=1}^n d[I(X_j \leqq x) - F(x)] \right\} \left\{ \sum_{l=1}^n d[I(X_l \leqq u) - F(u)] \right\} \right| \\ &\leqq \sum_{j=1}^n \sum_{l=1}^n |\mathrm{E}\,(W_j(x) - F(x))(W_l(u) - F(u))| \\ &= \sum_{j=1}^n |\mathrm{E}\,(W_j(x) - F(x))(W_j(u) - F(u))| \end{aligned}$$

$$=4n(dF(x))(1-dF(x))(dF(u))(1-dF(u))\leqq 4ndF(x)dF(u)\ .$$

Hence using Fubini's theorem we get

$$n^{1/2}\,\mathrm{E}\,A_n\leqq 4n^{-1/2}a_n^{-1}\int\int \gamma(x)k[(x-u)/a_n]dF(x)dF(u)=4n^{-1/2}\,\mathrm{E}\,\hat{M}_1\ ,$$

which converges to 0 as $n\to\infty$, since $\mathrm{E}\,\hat{M}_1\to M$ as $n\to\infty$. Finally let us evaluate B_n. Write $B_n=B_{1n}+B_{2n}$, where

$$\begin{aligned}B_{1n}&=a_n^{-1}\int\int \gamma(x)k[(x-u)/a_n]dF(x)dF_n(u)\\&\quad -a_n^{-1}\int\int \gamma(x)k[(x-u)/a_n]dF(x)dF(u)\\B_{2n}&=a_n^{-1}\int\int \gamma(x)k[(x-u)/a_n]dF_n(x)dF(u)\\&\quad -a_n^{-1}\int\int \gamma(x)k[(x-u)/a_n]dF(x)dF(u)\ .\end{aligned}$$

We shall prove that B_{in}, $i=1,2$ are asymptotically normal. Let $g_n(u)=a_n^{-1}\int\int \gamma(x)k[(x-u)/a_n]dF(x)$. Thus $B_{1n}=\frac{1}{n}\sum_{j=1}^{n} g_n(X_j)-\mathrm{E}\,g_n(X_1)=\frac{1}{n}\sum_{j=1}^{n} V_{nj}$, say. Note that $V_{nj},\cdots,V_{nn}$ are independent identically distributed random variables such that $\mathrm{E}\,V_{nj}=0$, $\mathrm{E}\,V_{nj}^2\to\sigma^2$ as given in (34), to see this we note that

$$\begin{aligned}(30)\qquad \mathrm{E}\,g_n^2(X_j)&=\int g_n^2(u)f(u)du\\&=a_n^{-2}\int\int\int \gamma(x)k[(x-u)/a_n]f(x)\gamma(y)k[(y-u)/a_n]\\&\qquad\cdot f(y)f(u)dxdydu\\&=\int\int\int k(w)k(z)\gamma(a_nw+u)f(a_nw+u)\gamma(a_nz+u)f(a_nz+u)\\&\qquad\cdot f(u)dudwdz\\&\leqq\int\int k(w)k(z)\Big[\int f^3(a_nw+u)\gamma^2(a_nw+u)du\Big]^{1/3}\\&\qquad\cdot\Big[\int f^3(a_nz+u)\gamma^2(a_nz+u)du\Big]^{1/3}\\&\qquad\cdot\Big[\int f^3(u)\gamma(a_nw+u)\gamma(a_nz+u)du\Big]^{1/3}dwdz\ ,\end{aligned}$$

where the inequality is obtained by Holder's inequality. Thus

$$(31)\qquad \limsup_n \mathrm{E}\,g_n^2(X_1)\leqq\int f^3(x)\gamma^2(x)dx\ .$$

On the other hand since $g_n^2(u)\to f^2(u)\gamma^2(u)$ we have by Fatou's lemma that

$$\liminf_n \mathrm{E}\, g_n^2(X_1) \geqq \int f^3(x)\gamma^2(x)dx \;. \tag{32}$$

Hence it follows from (31) and (32) that $\mathrm{E}\, g_n^2(X_1) \to \int f^3(x)\gamma^2(x)dx$. Thus $\mathrm{E}\, V_{nj}^2 \to \int \gamma(x) f^3(x)dx - M^2 = \sigma^2$ (say). Using similar argument we can also show that

$$\mathrm{E}\,|V_{nj}|^3 \to \int f(x)|\gamma(x)f(x)-M|^3 dx \qquad \text{as } n\to\infty \;. \tag{33}$$

Thus $n^{1/2}B_{1n}$ is asymptotically normal with mean 0 and variance σ^2 given by:

$$\sigma^2 = \int \gamma^2(x)f^3(x)dx - M^2 \;. \tag{34}$$

By an entirely similar argument we get that if $h_n(x) = a_n^{-1}\int k[(x-u)/a_n] \cdot dF(u) = \mathrm{E}\, f(x)$, then $B_{2n} = \frac{1}{n}\sum_{j=1}^{n} \gamma(X_j)h_n(X_j) - \mathrm{E}\,\gamma(X_1)h_n(X_1) = n^{-1}\sum_{j=1}^{n} W_{nj}$, where $W_{nj}, \cdots, W_{nn}$ are independent identically distributed random variables such that $\mathrm{E}\, V_{n1} = 0$ and $\mathrm{E}\, V_{n1}^2 \to \sigma^2$ as $n\to\infty$, for

$$\begin{aligned}
\mathrm{E}\,\gamma^2(X_1)h_n^2(X_1) &\\
&= \int \gamma^2(x)h_n^2(x)f(x)dx \\
&= a_n^{-2}\int\int\int \gamma^2(x)k[(x-y)/a_n]k[(x-z)/a_n]f(y)f(z)f(x)dxdydz \\
&= \int\int\int k(w)k(v)f(x-a_nw)f(x-a_nv)\gamma^2(x)f(x)dxdwdz \\
&\leqq \int\int k(w)k(v)\left[\int f^3(x-a_nw)\gamma^2(x)dx\right]^{1/3}\left[\int f^3(x-a_nv)\gamma^2(x)dx\right]^{1/3} \\
&\qquad \cdot \left[\int f^3(x)\gamma^2(x)dx\right]^{1/3} dwdz \;.
\end{aligned} \tag{35}$$

Thus

$$\limsup_n \mathrm{E}\,\gamma^2(X_1)h_n^2(X_1) \leqq \int f^3(x)\gamma^2(x)dx \;. \tag{36}$$

But also by Fatou's lemma $\liminf_n \mathrm{E}\,\gamma^2(X_1)h_n(X_1) \geqq \int f^3(x)\gamma^2(x)dx$. The balance of the proof proceeds as in B_{1n}. Hence B_n is asymptotically normal with zero mean and variance $2\sigma^2$. To finish the proof we want to show that $n^{1/2}(\mathrm{E}\,\hat{M}_1 - M) = o(1)$ but this follows from

$$\text{(37)} \qquad n^{1/2}|\mathrm{E}\,\hat{M}_1-M|=n^{1/2}\left|\int \gamma(x)\,\mathrm{E}\,\hat{f}_1(x)f(x)dx-\int \gamma(x)f(x)dx\right|$$
$$\leqq n^{1/2}\int |\gamma(x)||\mathrm{E}\,\hat{f}_1(x)-f(x)|f(x)dx$$
$$\leqq C(na_n^{2m})^{1/2}\sup_x |f^{(m)}(x)|\int |\gamma(x)|f(x)dx=o(1)\ .$$

since $\int u^r k(u)du=0,\ r=1,\cdots,m-1$ and $\mathrm{E}\,|\gamma(X_1)|<\infty$. QED

Note that if $na_n^4\to 0$ and if $f''(x)$ is bounded for all x with $\int uk(u)du=0$ and $\int u^2k(u)du<\infty$, then Theorem 2.3 holds. Note also that if one is interested in an asymptotically normal confidence interval for M, a consistent estimate of $\sigma^2=\int f(x)[\gamma(x)f(x)-M]^2dx$ is needed. One such estimate would be

$$\text{(38)} \qquad \hat{\sigma}_1^2=\int [\gamma(x)\hat{f}_1(x)-\hat{M}_1]^2dF_n(x)\ .$$

It can be shown in a straightforward but tedious way that $\hat{\sigma}_1^2$ is a consistent estimate of σ^2. We leave the details to the interested reader. Thus an asymptotically normal confidence interval of M is given by:

$$\text{(39)} \qquad \hat{M}_1\pm(\hat{\sigma}_1/n^{1/2})z_{\alpha/2}\ ,$$

where $z_{\alpha/2}$ is the standard normal variate such that $\mathrm{P}\,[|Z|\leqq z_{\alpha/2}]=1-\alpha$, where Z is the standard normal random variable.

3. Large sample theory of $\hat{M}_2$

THEOREM 3.1. *Assume that $f(x)$ is uniformly continuous of bounded variations, and square integrable, further assume that $\mathrm{E}\,|\gamma(X_1)|^2<\infty$, then*

$$\text{(40)} \qquad \mathrm{E}\,|\hat{M}_2-M|\to 0 \qquad as\ n\to\infty\ .$$

PROOF. Note that

$$\text{(41)} \qquad \mathrm{E}\,|\hat{M}_2-M|=\mathrm{E}\left|\int\int \gamma(x)\Phi_n(x,y)dF_n(x)dF_n(y)-\int \gamma(x)f(x)dF(x)\right|$$
$$\leqq \mathrm{E}\left|\int\int \gamma(x)\Phi_n(x,y)dF_n(x)dF_n(y)\right.$$
$$\left.-\int\int \gamma(x)\Phi_n(x,y)dF_n(x)dF(y)\right|$$
$$+\mathrm{E}\left|\int\int \gamma(x)\Phi_n(x,y)dF_n(x)dF(y)\right.$$

$$-\int\int \gamma(x)\Phi_n(x, y)dF(x)dF(y)\Big|$$

$$+\Big|\int\int \gamma(x)\Phi_n(x)dF(x)dF(y)-\int \gamma(x)f(x)dF(x)\Big|$$

$$=\mathrm{E}\, I_1+\mathrm{E}\, I_2+I_3\,, \qquad \text{say}\,.$$

But as in Theorem 2.1, $I_3=\int \gamma(x)[\mathrm{E}\,\hat{f}_2(x)-f(x)]dF(x)\to 0$ as $n\to\infty$ using LDCT and since (see Ahmad [1], Theorem 2.2) $\mathrm{E}\,\hat{f}_2(x)\to f(x)$ as $n\to\infty$ whenever $q^2(n)/n\to 0$, as $n\to\infty$ for all x. Next, again $\mathrm{E}\, I_2=\mathrm{E}\Big|\int \gamma(x)\cdot \mathrm{E}\,\hat{f}_2(x)dF_n(x)-\int \gamma(x)\,\mathrm{E}\,\hat{f}_2(x)dF(x)\Big|\to 0$, as $n\to\infty$ provided that $\mathrm{E}\,|\gamma(X_1)|^2<\infty$. Finally, we have

$$\mathrm{E}\, I_1\leqq \mathrm{E}^{1/2}\,\{\sup|\hat{f}_2(x)-\mathrm{E}\,\hat{f}_2(x)|\}^2\,\mathrm{E}^{1/2}\Big\{n^{-1}\sum_{i=1}^{n}|\gamma(X_i)|\Big\}^2\,, \tag{42}$$

which converges to 0 as $n\to\infty$ in view of Theorem 2.2 of Ahmad [1]. QED

THEOREM 3.2. *Let $f(x)$ be as in Theorem* 2.1 *and assume that* $\mathrm{E}\,|\gamma(X_1)|<\infty$ *and that for any* $\varepsilon>0$, $\sum_{n=1}^{\infty}\exp[-\varepsilon n/q^2(n)]<\infty$, *then*

$$\hat{M}_2\to M \qquad w.p.\ 1\ as\ n\to\infty\,.$$

PROOF. In the proof of Theorem 2.2 we proceed to get

$$I_1\leqq\sup_x|\hat{f}_2(x)-\mathrm{E}\,\hat{f}_2(x)|\int|\gamma(x)|dF_n(x)\,, \tag{43}$$

which converges to 0 and $n\to\infty$ since by Theorem 3.1 of Ahmad [2], $\sup_x|\hat{f}_2(x)-\mathrm{E}\,\hat{f}_2(x)|\to 0$ w.p. 1 as $n\to\infty$ and $\int \gamma(x)dF_n(x)\to\int|\gamma(x)|dF(x)$ w.p. 1 as $n\to\infty$ by the strong law of large numbers. $I_2\to 0$ w.p. 1 as in Theorem 2.2 and $I_3\to 0$ as in Theorem 3.1 above. QED

To discuss the asymptotic normality of $\hat{M}_2$ we recall that

$$\begin{aligned}\hat{M}_2-\mathrm{E}\,\hat{M}_2=&\int\int \gamma(x)\Phi_n(x, y)d[F_n(x)-F(x)]d[F_n(y)-F(y)]\\ &+\Big\{\int\int \gamma(x)\Phi_n(x, y)dF_n(x)dF(y)\\ &+\int\int \gamma(x)\Phi_n(x, y)dF(x)dF_n(y)\\ &-2\int\int \gamma(x)\Phi_n(x, y)dF(x)dF(y)\Big\}\\ =&A_n+B_n\,, \qquad \text{say}\,.\end{aligned} \tag{44}$$

But as in Theorem 2.3,

$$(45) \qquad n^{1/2}\,\mathrm{E}\,A_n \leqq 4n^{-1/2}\iint \gamma(x)\Phi_n(x, y)dF(x)dF(y) = 4n^{-1/2}\,\mathrm{E}\,\hat{M}_2\,,$$

which converges to 0 as $n\to\infty$ since $\mathrm{E}\,\hat{M}_2 \to M$, as $n\to\infty$. Finally write

$$(46) \qquad \begin{aligned} B_n &= \left\{\iint \gamma(x)\Phi_n(x, y)dF_n(x)dF(y) - \iint \gamma(x)\Phi_n(x, y)dF(x)dF(y)\right\} \\ &\quad + \left\{\iint \gamma(x)\Phi_n(x, y)dF(x)dF_n(y) - \iint \gamma(x)\Phi_n(x, y)dF(x)dF(y)\right\} \\ &= B_{1n} + B_{2n}\,, \qquad \text{say}\,. \end{aligned}$$

Let $g_n(y)=\int \gamma(x)\Phi_n(x, y)dF(x)$, then $B_{1n}=n^{-1}\sum_{i=1}^{n} g_n(X_i)$ and $B_{2n}=n^{-1}\sum_{i=1}^{n} g_n(X_i)$. Thus proceeding as in Theorem 2.3 above with minor modification it is possible to show that B_{in} is asymptotically normal with mean 0 and variance σ^2 provided that $|\theta_j|\leqq C_r(j+1)^{-r/2}$ for some integer $r\geqq 3$, and $(n/q(n))^{r+2r-2}\to 0$ as $n\to\infty$, since (see Ahmad [1]) in this case $\mathrm{E}\,\hat{f}_2(x)-f(x)=O((q(n))^{-r/2-r+1})$. Thus if we collect this argument we arrive at:

THEOREM 3.3. *If $f(x)$ is uniformly continuous of bounded variation, and is square integrable. If $n/q^2(n)\to\infty$ as $n\to\infty$ and $n/q(n)^{r-2+2r}\to 0$ as $n\to\infty$ for some $r\geqq 3$ and if $|\theta_j|\leqq C(j+1)^{-r/2}$, then $n^{1/2}(\hat{M}_2-M)$ is asymptotically normal with mean 0 and variance $2\sigma^2$.*

We conclude this section by writing that it is possible to estimate M alternatively by $\tilde{M}_2=\int \gamma(x)\hat{f}_2^2(x)dx$ and demonstrate the mean and strong consistency of M_2 as those of M_2 above. The asymptotic normality is difficult to obtain.

4. Concluding remarks

1. Higgins and Tichenor [7] discussed the asymptotic relative efficiency of M_1 to the maximum likelihood estimates in the Cauchy case (note that the asymptotic normality of $\hat{M}_1$ and $\hat{M}_2$ uses conditions that apply to the Cauchy case) and shown that it is equal to 1 for both the location and scale parameters. It is easy to check the asymptotic relative efficiency of μ and σ using $\hat{M}_i$ and using other competitor nonparametric methods, such as rank statistics estimates, that are asymptotically normal. If we take the robust estimates L or M (see Andrews, et al. [4]), then we can observe at once that our estimates approach

the approximate normality under far less conditions and for certain choices of the functions needed in the L and M estimates are better and almost always at least as good as the L or M estimate.

2. Besides good large sample properties and the apparent good performance as robust estimates of μ and σ, $\hat{M}_i$ is also robust if the observations are no longer independent. Suppose that $\{X_n\}$ is a strictly stationary strong mixing process, i.e., if $\sigma(a, b)=\sigma(X_{a+1}, \cdots, X_{a+b})$ is the σ-field generated by $X_{a+1}, \cdots, X_{a+b}$ and if A and B are two measurable events such that $A \in \sigma(0, m)$ and $B \in \sigma(m+n, \infty)$, then

$$|\mathrm{P}(AB)-\mathrm{P}(A)\,\mathrm{P}(B)| \leqq \alpha(n)\ , \tag{47}$$

where $\alpha(n)$ is an integer-valued function such that $\alpha(n) \downarrow 0$ as $n \to \infty$. If $\hat{f}_i(x)$ is based on the first n observations of such process then the large sample theorems continue to hold under suitable condition, e.g., if we consider $f_2(x)$, then Theorem 3.1 continues to hold if $\sum_{n=1}^{\infty} \alpha(n) < \infty$, while Theorem 3.2 continues to hold if for any $\varepsilon > 0$, $\sum_{n=1}^{\infty} \{\exp[-\varepsilon p(n)/q^2(n)]\}\{1+k\alpha(m(n))\}p(n) < \infty$, where $p(n)$ and $m(n)$ are integer valued factors, e.g., $p(n)=[n^{(1-\lambda)}]$ and $m(n)=[n^{\lambda}]$ for some $\lambda > 0$, and finally Theorem 3.3 holds in this case whenever $\sum_{n=1}^{\infty} \alpha(n) < \infty$. For details of these results we use the work of Ahmad [2] and [1]. If one uses the results of Ahmad [3] concerning $\hat{f}_1(x)$ when $\{X_n\}$ is strictly stationary strong mixing then it is also possible to establish extensions of Theorems 2.1–2.3 in this case.

3. Observe here that our estimates are much simpler to compute than the standard robust estimates and they are fairly robust themselves since this is indeed an inherent property of density estimation. In addition to this, the rates of convergence in all three large sample results are possible to establish and we established an iterated logarithm result in Remark 2.4. The rate in the central limit theorem is possible to establish, of course under some extra conditions, we could prove that $n^{1/2}(\hat{M}_1 - M)$ approaches the normal at the rate a_n while $n^{1/2}(\hat{M}_2 - M)$ has the rate $(q(n))^{-1}$. We shall leave the details out.

UNIVERSITY OF PETROLEUM AND MINERALS

REFERENCES

[1] Ahmad, I. A. (1980). Integrated mean square properties of density estimation by orthogonal series methods for dependent variables, *Ann. Inst. Statist. Math.*, A, to appear.

[2] Ahmad, I. A. (1979). Strong consistency of density estimation by orthogonal series

methods for dependent variables, *Ann. Inst. Statist. Math.*, **31**, A, 279-288.

[3] Ahmad, I. A. (1973). Nonparametric estimation of p.d.f.'s for dependent variables with application, Ph.D. dissertation, Florida State University.

[4] Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H. and Tukey, J. W. (1972). *Robust Estimates of Location*, Princeton University Press, Princeton, N.J.

[5] Bhattacharyya, G. K. and Roussas, G. G. (1969). Estimation of a certain functional of probability density function, *Skand. Aktuarietidskr.*, **52**, 203-206.

[6] Cencov, N. N. (1962). Estimation of an unknown density function from observations, *Dokl. Akad. Nauk SSSR*, **147**, 45-48 (in Russian).

[7] Higgins, J. J. and Tichenor, D. M. (1977). Window estimates of location and scale with applications to the Cauchy distribution, *Applied Mathematics and Computations*, **3**, 113-126.

[8] Kronmal, R. and Tarter, M. (1968). The estimation of probability densities and cumulatives by Fourier series methods, *J. Amer. Statist. Ass.*, **63**, 925-952.

[9] Murthy, V. K. (1965). Estimation of a probability density, *Ann. Math. Statist.*, **36**, 1027-1031.

[10] Parzen, E. (1962). On the estimation of a probability density function and mode, *Ann. Math. Statist.*, **33**, 1065-1076.

[11] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function, *Ann. Math. Statist.*, **27**, 832-837.

[12] Royden, H. L. (1968). *Real Analysis* (Second Edition), McMillan, New York.

[13] Schwartz, S. C. (1967). Estimation of a probability density by an orthogonal series, *Ann. Math. Statist.*, **38**, 1262-1265.

[14] Szego, G. (1959). *Orthogonal Polynomial*, Amer. Math. Soc. Coll. Publ. 32, Providence, R.I.

[15] Tapia, R. A. and Thompson, T. R. (1978). *Nonparametric Probability Density Estimation*, Johns Hopkins Univ. Press, Baltimore, Maryland.