J. Austral. Math. Soc. 23 (Series A) (1977), 166–171.

A FOURIER INVERSION METHOD FOR THE ESTIMATION OF A DENSITY AND ITS DERIVATIVES

J. BLUM AND V. SUSARLA

(Received 10 June 1975; revised 13 April 1976)

1. Introduction and summary

During the last two decades a variety of methods have been developed for the problem of estimation of unknown density f wrt Lebesgue measure and its νth derivative $g(=f^{(\nu)})$ using i.i.d random variables $X_1,\cdots,X_n$ when $X_1\sim f$. For example, see Wegman (1972). In almost all the papers on the estimation of $f(x)$ or $g(x)$, various authors assumed the existence of derivatives of f of order $r(>\nu)$ at x to obtain rates for the mean-square convergences and other desirable properties for their estimators. Here it is shown that if

(A1) $\qquad g$ is Lipschitz (from left) of order $\alpha(>0)$ at x,

then estimators $\hat{g}(x)$ can be constructed for which $E[(\hat{g}(x)-g(x))^2]=O(n^{-(2\alpha-\delta)/(2\alpha+2\nu+1)})$ for any given $\delta>0$. Similar statements hold for almost sure convergence of $\hat{g}(x)$. It can also be shown that $(\hat{g}(x_1), \hat{g}(x_2))$ is asymptotically bivariate normal under certain conditions for $x_1\neq x_2$. If (A1) is satisfied with $\alpha\leqq 1$, then our estimators have all the desirable properties while other methods are not applicable in this situation since they require differentiability conditions on g. (For example, see Susarla and Kumar (1975) and its references.) Our estimators are defined by using the inversion theorem for some absolutely integrable characteristic functions. The motivation for our estimators is given in O'Bryan and Susarla (1975, 76) and Susarla and O'Bryan (1975).

Unless otherwise stated, limits are taken as $n\to\infty$. Let $R=(-\infty,\infty)$ and $i^2=-1$. We provide only the proofs for the asymptotic unbiasedness and the mean square consistency as the proof of asymptotic normality is similar to that in Susarla and Kumar (1975).

2. Definition of $\hat{g}$

Let $Y_1, \cdots, Y_n, \cdots$ be i.i.d. random variables with density function $w(\cdot)$ vanishing off (0,1). We assume throughout that X_i and Y_j are independent for any i and j. Let.

$$\phi(\cdot) = E[e^{iY_1}] \in \Phi_k = \left\{ h \,\middle|\, \int_R |t|^k |h(t)| dt < \infty \right\}. \tag{2.1}$$

The estimator for $g(x)$ is defined by

$$\hat{g}(x) = \frac{1}{2\pi} \operatorname{Re}\left[\int_R e^{-itx}(-it)^\nu k\left(\frac{t}{M_n}\right) \phi(\mu_n t) \psi_n(t) dt \right] \tag{2.2}$$

where ψ_n is the sample characteristic function of $X_1, \cdots, X_n$, $k(u) = I_{[-1,1]}(u)$, $M_n \uparrow \infty$, and $\mu_n \downarrow 0$. We show below that, by varying $M_n(\uparrow \infty)$ and $\mu_n(\downarrow 0)$ in an appropriate way, one can obtain various asymptotic results about $\hat{g}(x)$ under some conditions including (A1). For ease of reference later on, let

$$\Delta_\lambda(x) = \int_0^1 |g(x - \lambda u) - g(x)| \, w(u) du. \tag{2.3}$$

3a. Asymptotic unbiasedness of $\hat{g}(x)$

The following theorem gives necessary and sufficient (with and without rates) conditions for the asymptotic unbiasedness of $\hat{g}(x)$. Let the bias $E[\hat{g}(x)] - g(x)$ be denoted by $B_n(x)$.

THEOREM 3.1. *Let $k \geqq \nu$ and $\Delta_\lambda(x) \to 0$ as $\lambda \downarrow 0$. Then there exist μ_n and M_n such that $B_n(x) \to 0$. On the other hand, if $B_n(x + t) - B_n(x) \to 0$ as $n \to \infty$ and $t \downarrow 0$, then $\Delta_\lambda(x) \to 0$. If $k > \nu$ (see (2.1)), then $|B_n(x)| = O(M_n^{-\alpha(k-\nu)/(\alpha+k+1)})$ with $M_n \uparrow \infty$ and $\mu_n^{\alpha+k+1} = M_n^{\nu-k}$.*

PROOF. By a triangle inequality,

$$|B_n(x)| \leqq |g_{\mu_n, M_n}(x) - g_{\mu_n}(x)| + |g_{\mu_n}(x) - g(x)|$$

where

$$2\pi g_{\mu_n}(x) = \int_R (-it)^\nu e^{-itx} \phi(\mu_n t) \psi(t) dt = 2\pi \int_0^1 g(x - \mu_n u) w(u) du$$

with ψ denoting the characteristic function of X_1 (the equality here follows from inversion theorem and differentiation under integral sign), and

$$g_{\mu_n, M_n}(x) (= E[\hat{g}(x)]) = (2\pi)^{-1} \int_{-M_n}^{M_n} (-it)^\nu e^{-itx} \phi(\mu_n t) \psi(t) \, dt.$$

https://doi.org/10.1017/S1446788700018164 Published online by Cambridge University Press

The first term in the rhs of the above inequality is seen to be exceeded by

$$\mu_n^{-(\nu+1)} \int_{|t| \geqq M_n \mu_n} |t|^\nu | \, |\phi(t)| \, dt$$

by a change of variable while the second term is exceeded by $\Delta_{\mu_n}(x)$ (see (2.2)). Hence

$$|B_n(x)| \leqq \mu_n^{-(\nu+1)} \int_{|t| \geqq M_n \mu_n} |t|^\nu | \phi(t) | dt + \Delta_{\mu_n}(x). \tag{3.1}$$

First let $\mu_n \downarrow 0$ and then let $M_n \uparrow \infty$ so that the rhs of the above inequality goes to zero. This is possible since $\phi \in \Phi_k$ (see (2.1)) with $k \geqq \nu$. This completes the proof of the first part.

The second result follows by a triangle inequality and the condition $|B_n(x+t) - B_n(x)| \to 0$ as $n \uparrow \infty$ and $t \downarrow 0$. The third result follows from (3.1) and the facts that $\Delta_{\mu_n}(x) \leqq c\mu_n^\alpha$ for some constant c by (A1) and that the first term of rhs of (3.1) is bounded $(\mu_n^{k+1} M_n^{k-\nu})^{-1} \int |t|^k | \phi(t)|dt$. Note that $\int |t|^k |\phi(t)| dt < \infty$ since $\phi \in \Phi_k$ by assumption.

REMARK 3.1. The condition placed on g, namely (A1) or $\Delta_\lambda(x) \to 0$ as $\lambda \downarrow 0$, is the weakest assumption among all results similar to theorem 3.1.

3b. Mean square consistency

The mean square consistency results (with and without rates) for $\hat{g}$ are obtained in this section. It can be shown very easily from the independence of (X_j, Y_j), $j = 1, \cdots, n$ that

LEMMA 3.1. $n(\nu+1)^2 \, var(\hat{g}(x)) \leqq 4M_n^{2(\nu+1)}$.

An improvement (in terms of rate) of Lemma 3.1 *is*

LEMMA 3.2. *Let* $\phi \in \Phi_k$ *(see* (2.1)) *with* $k \geqq \nu$, *and let* $\Delta_\lambda(x)$ *(see* (2.3)) $\to 0$ *as* $\lambda \downarrow 0$ *and f be continuous at x. Then for* μ_n *and* M_n *satisfying the conclusion of the first part of Theorem* 3.1,

$$\left| nM_n^{-(2\nu+1)} \mathrm{var}(\hat{g}(x)) - \pi^{-2} f(x) \int_R \left(\int_0^1 u^\nu \cos su \, du \right)^2 ds \right| \to 0$$

for even ν. *For odd* ν, *replace cos by sin.*

PROOF. We consider the case of even ν only, the other case being similar. In this proof, let $K(s) = \int_0^1 u^\nu \cos su \, du$. Then

$$\int_R K^2(s) ds < \infty \quad \text{and} \quad |sK^2(s)| \to 0 \quad \text{as} \quad |s| \to \infty.$$

Since $(X_1, Y_1), \cdots, (X_n, Y_n)$ are i.i.d,

$$nM_n^{-(2\nu+1)} \operatorname{var}(\hat{g}(x)) = (M_n^{2\nu+1})^{-1}(E[(\hat{g}_{n,1}(x))^2] - (E[\hat{g}_{n,1}^{(x)}])^2) \tag{3.2}$$

where $2\pi\hat{g}_{n,1}(x) = \operatorname{Re}[\int_{-M_n}^{M_n}(-it)^\nu e^{it(X_1+\mu_n Y_1-x)}dt]$. Since

$$E[\hat{g}_{n,1}(x)] = E[\hat{g}(x)], E[\hat{g}_{n,1}(x)] \to g(x)$$

by first part of Theorem 3.1. So consider

$$\begin{aligned}\pi^2 E[(\hat{g}_{n,1}(x))^2] &= \int_R \left(\int_0^{M_n} t^\nu \cos t(u-x)dt\right)^2 f_{\mu_n}(u)du \\ &= M_n^{2\nu+1}\int_R K^2(s) f_{\mu_n}\left(x + \frac{s}{M_n}\right)ds \\ &= M_n^{2\nu+1}\int_R K^2(s)\int f\left(x+\frac{s}{M_n} - \mu_n u\right)w(u)du\,ds\end{aligned} \tag{3.3}$$

where $f_{\mu_n}(u) = \int_0^1 f(x-\mu_n u)w(u)du$.

Following the pattern of proof Theorem 1A of Parzen (1962), we can show that the double integral in the rhs of (3.3)$\to f(x)\int K^2(s)ds$ by using the conditions $\int_R K^2(s)ds < \infty$, $|sK^2(s)| \to 0$ and the continuity of f at x. Thus the proof follows from (3.2) and (3.3).

As a corollary to the above two results, we have

THEOREM 3.2. *Let the conditions of the first part of Theorem 3.1 hold. Then there exist μ_n and M_n such that $E[(\hat{g}(x) - g(x))^2] \to 0$. If, in addition* (A1) *holds at x and $k > \nu$, then*

$$E[(\hat{g}(x) - g(x))^2] = O(n^{-2\alpha(k-\nu)/\{(2\nu+1)(\alpha+k+1)+2(k-\nu)\alpha\}}) \tag{3.4}$$

with $M_n^{(2\nu+1)(\alpha+k+1)+2\alpha(k-\nu)} = n^{\alpha+k+1}$ and $\mu_n^{\alpha+k+1} = M_n^{\nu-k}$.

PROOF. We use the equality $E[(\hat{g}(x) - g(x))^2] = (B_n(x))^2 + \operatorname{var}(\hat{g}(x))$. The first result follows from the first part of Theorem 3.1 and Lemma 3.1. The second part follows from the rate part of Theorem 3.1 and Lemma 3.2. In the latter case, we chose M_n so that the square of the rate in Theorem 3.1 $= n^{-1}M_n^{2\nu+1}$.

REMARK 3.2. As $k \to \infty$, the rate in (3.4) goes to $O(n^{-2\alpha/(2\alpha+2\nu+1)})$. This limiting rate coincides with rate result of Theorem 2 of Wahba (1971) if $\nu = 0$. For $\alpha = 1$, our rate misses the rate of Theorem 2 of Wahba (1971) and the rate obtained by Singh (1974) for any ν by a positive quantity (in the exponent of n^{-1}) which goes to zero as $k \uparrow \infty$. If g is Lipschitz (from left) of order 1 at x (which is implied by the differentiability conditions on f or g as the case may be), neither Wahba nor Singh has any rate results while we can obtain the rate $O(n^{-2-\varepsilon)/2\nu+3)})$ for any given $\varepsilon > 0$.

3c. Asymptotic normality of $\hat{g}(x)$

The following result concerning the asymptotic normality of $(\hat{g}(x_1), \hat{g}(x_2))'$ can be proved using Lemma 3.2 and the method of proof used in Susarla and Kumar (1975) for their asymptotic normality result.

THEOREM 3.3. *If the conditions (including those on μ_n and M_n) of Lemma 3.2 are satisfied and $nM_n^{-1} \to \infty$, then*

$$(nM_n^{-(2\nu+1)})^{1/2} \begin{pmatrix} \hat{g}(x_1) - E[\hat{g}(x_1)] \\ \hat{g}(x_2) - E[\hat{g}(x_2)] \end{pmatrix}$$

converges in law to the bivariate normal distribution with zero mean vector and the covariance matrix =

$$\begin{bmatrix} \pi^{-2} f(x_1) \int_R K^2(s)ds & 0 \\ 0 & \pi^{-2} f(x_2) \int_R K^2(s)ds \end{bmatrix}$$

4. Concluding remarks

The estimator $\hat{g}(x)$ can be shown to be strongly consistent (with a rate) by using Lemma 2 of Dvoretzky, Kiefer and Wolfowitz, but the proof is rather long and involved and to some extent, goes along the lines of proof of Nadaraya (1965). All the results presented here can be shown to hold uniformly in all x belonging to a set $D \subset (-\infty, \infty)$ provided conditions uniform in D are assumed.

The results of this paper can be extended to the problem of estimation of $\bar{f}(x) = n^{-1} \Sigma_{j=1}^{n} f_j(x)$ and its partial derivatives using independent p-variate random vectors $X_1, \cdots, X_n$ whenever $X_j \sim f_j$, a density wrt the Lebesgue measure on $(R^p, \mathbb{B}^p)$. Such an extension would provide results analogous to those in Susarla and Kumar (1975).

The mean square result (Theorem 3.2) was shown to be applicable in empirical Bayes decision problems with non-identical components by Susarla and O'Bryan (1975) and O'Bryan and Susarla (1975, 76).

Acknowledgements

The research of the first author was supported by NSF Grant No. GP–25736 and that of the second author was supported by a research grant from the University of Wisconsin-Milwaukee and by the United States Army under contract No. DAAG–75–C–0024 at the Mathematics Research Center, Madison. Thanks are due to a referee for some suggestions which improved the clarity of the paper.

References

[1] A. Dvoretzky, J. Kiefer, and J. Wolfowitz (1956), 'Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator', *Ann. Math. Statist.* **24,** 642–669.

[2] E. A. Nadaraya (1965), 'On non-parametric estimates of density functions and regression curves', *Theor. Prob. Appl.* **10,** 186–190.

[3] T. O'Bryan and V. Susarla (1975), 'An empirical Bayes estimation problem with nonidentical components involving normal distributions', *Communications in Statistics* **4,** 1033–1042.

[4] T. O'Bryan and V. Susarla (1976), 'Rates in the empirical Bayes estimation problem with non-identical components. Case of normal distributions', *Ann. Inst. Statist. Math.* **28,** Part A, 389–397.

[5] E. Parzen (1962), 'On the estimation of probability density and mode', *Ann. Math. Statist.* **33,** 1065–1076.

[6] R. S. Singh (1974), 'Estimation of derivatives of the average of μ-densities and sequence-compound estimation in exponential families', RM–318, Department of Statistics and Probability, Michigan State University.

[7] V. Susarla and S. Kumar (1975), 'Estimation of partial derivatives of the average of densities belonging to a family of densities', *J. Austral. Math. Soc.* **20,** 230–241.

[8] V. Susarla and T. O'Bryan (1975), 'An empirical Bayes two action problem with nonidentical components for a translated exponential distribution', *Communications in Statistics* **4,** 767–775.

[9] G. Wahba (1971), 'A polynomial algorithm for density estimation', *Ann. Math. Statist.* **42,** 1870–1886.

[10] E. J. Wegman (1972), 'Nonparametric probability density estimation: I. A summary of available methods', *Technometrics* **14**, 533–546.

University of Wisconsin-Milwaukee,
Milwaukee, Wisconsin, U.S.A.

Mathematics Research Center,
Madison, Wisconsin, U.S.A.