Rates of Convergence of Some Estimators

in a Class of Deconvolution Problems

by

Leonard A. Stefanski[1]
Department of Statistics
North Carolina State University
Raleigh, NC 27695-8203

Abstract

This paper studies the problem of estimating the density of U when only independent copies of X = U + Z is observable where Z is an independent measurement error. Convergence rates of a family of deconvolved Kernel density estimators are obtained under different assumptions on the density of Z.

Key Words and Phrases: Deconvolution, density estimation, mean squared error, measurement error, rates of convergence, uniform convergence.

[1] Work supported by NSF Grant DMS-86-13681. AMS 1980 subject classifications. Primary 62G05; secondary 62G-99.

1. Introduction

Let $\{(U_i, Z_i, X_i)\ i=1,\ldots,n\}$ be independent and identically distributed random 3-dimensional vectors such that $U=U_1$ has unknown density g; $Z=Z_1$ is independent of U and has known density h; and $X=U+Z$ has density $f=g*h$. The problem is to estimate g nonparametrically given only observations on $X_1, \ldots, X_n$. Recent papers addressing this problem include Stefanski and Carroll (1987), Carroll and Hall (1987), Liu and Taylor (1987a,b), Fan (1988) and Devroye (1989). The first three papers contain further references on statistical applications of deconvolution; see, for example, Mendelsohn and Rice (1982) and Crump and Seinfeld (1982).

This note investigates the convergence rates of the class of kernel estimators studied by Stefanski and Carroll (1987) for a particular class of measurement error densities. Let φ_g, φ_f and φ_h denote the characteristic functions of g, f, and h respectively. Also, let $\hat{\varphi}_f$ be the empirical characteristic function of $X_1,\ldots, X_n$. The estimator considered has the representation

$$\hat{g}(x) = (1/2\pi)\int_{-1/\lambda}^{1/\lambda} e^{-itx}\hat{\varphi}_f(t)\varphi_k(\lambda t)/\varphi_h(t)\ dt \tag{1.1}$$

where φ_k is a characteristic function which vanishes outside [-1,1] and $\lambda=\lambda_n$ is a sequence of positive constants converging to zero.

Rates of convergence of the mean integrated squared error of $\hat{g}$ and the sup norm of $(\hat{g}-g)$ are derived and shown to be obtainable by known sequences of bandwidths.

2. Asymptotic Results

Define $\mathrm{MISE}_n(\lambda) = E\int(\hat{g} - g)^2$ and $\Gamma_n(\lambda) = \sup_x|\hat{g}(x) - E\hat{g}(x)|$. In this section properties of the sequences $\{\mathrm{MISE}_n(\theta_n)\}$ and $\{\Gamma_n(\theta_n)\}$ are studied for certain sequences, $\{\theta_n\}$, such that $\theta_n \to 0$.

2.1. Mean Integrated Squared Error

By the Parseval relation

$$\mathrm{MISE}_n(\lambda) = (2\pi n)^{-1}\int \frac{|\varphi_k(\lambda t)|^2}{|\varphi_h(t)|^2}\{1-|\varphi_f(t)|^2\}dt$$

$$+ (2\pi)^{-1}\int |\varphi_g(t)|^2|1 - \varphi_k(\lambda t)|^2 dt \ . \tag{2.1}$$

Throughout the paper the following conditions are imposed on φ_k and φ_h:

(A1) φ_k is even, real valued, bounded, nonnegative and nonincreasing on $[0,\infty)$ with $\varphi_k(0) = 1$ and $\varphi_k(1) = 0$; φ_k has $\nu=\max(r,s) + 1$ bounded derivatives such that $\varphi_k(1) = \ldots = \varphi_k^{(r-1)}(1) = 0$, $\varphi_k^{(r)}(1) \neq 0$ and $\varphi_k^{(1)}(0) = \ldots = \varphi_k^{(s-1)}(0)$, $\varphi_k^{(s)}(0) \neq 0$;

(A2) φ_h is real valued and nonvanishing, i.e., $|\varphi_h(t)| > 0$ for all real t; as $t \to \infty$, $1/\varphi_h(t) \sim \alpha t^\beta \exp(\gamma t^\delta)$ where α, δ, $\gamma > 0$, $|\beta| < \infty$.

Under (A1) and (A2) $\mathrm{MISE}_n(\theta_n)$ converges to zero only if $\theta_n \to 0$ and hence only the behavior of $\mathrm{MISE}_n(\lambda)$ for small λ is of interest. This is given in the following theorem.

THEOREM 2.1. If $\int t^{2s}|\varphi_g(t)|^2 dt < \infty$, then

$$\mathrm{MISE}_n(\lambda) = \{A_r n^{-1}\exp(2\gamma\lambda^{-\delta})\lambda^{(2r+1)\delta-2\beta-1} + B_s\lambda^{2s}\}\{1 + c_n(\lambda)\}, \tag{2.2}$$

where $c_n(\lambda) \to 0$ as $\lambda \to 0$ uniformly in n and

$$B_s = \frac{\{\varphi_k^{(s)}(0)\}^2}{2\pi(s!)^2}\int t^{2s}|\varphi_g(t)|^2 dt, \quad A_r = \binom{2r}{r}\frac{\{\varphi_k^{(r)}(1)\}^2\alpha^2}{\pi(2\gamma\delta)^{2r+1}} \ .$$

COROLLARY 2.1.1. If $\{\theta_n\}$ is any sequence of positive constants for which $\theta_n \to 0$ and $\mathrm{MISE}_n(\theta_n)$ stays bounded then $\liminf \theta_n^\delta \,\ell og\, n \geq 2\gamma$.

PROOF. The fact that $(2\pi)^{-1} \int |\varphi_g(t)|^2 |1 - \varphi_k(\lambda t)|^2 dt \sim B_s \lambda^{2s}$ is a standard result in kernel density estimation.

Since $|\varphi_f|/|\varphi_h| = |\varphi_g|$,

$$(2\pi)^{-1} \int \frac{|\varphi_k(\lambda t)|^2}{|\varphi_h(t)|^2} \{1 - |\varphi_f(t)|^2\} dt \sim (2\pi)^{-1} \int \frac{|\varphi_k(\lambda t)|^2}{|\varphi_h(t)|^2} \, dt$$

$$\sim A_r \exp(2\gamma\lambda^{-\delta}) \lambda^{(2r+1)\delta - 2\beta - 1}$$

The last relation follows from Lemma 3.1.

The corollary follows directly from Lemma 3.2. ////

REMARK. The condition that $t^{2s}|\varphi_g(t)|^2$ is integrable can be derived from the assumption that g has s continuous derivatives and $g^{(s)}$ is square integrable.

THEOREM 2.2. If λ_n minimizes $\mathrm{MISE}_n(\lambda)$, then

(i) $\lambda_n^\delta \,\ell og\, n \to 2\gamma$;

(ii) $(\ell og\, n)^{2s/\delta} \,\mathrm{MISE}_n(\lambda_n) \to B_s(2\gamma)^{2s/\delta}$.

In addition, if $\{\theta_n\}$ is any sequence of positive constants, then $\{(\ell og\, n)^{2s/\delta}\} \,\mathrm{MISE}_n(\theta_n) \to B_s(2\gamma)^{2s/\delta}$ if and only if $\theta_n^\delta = 2\gamma/(\ell og\, n\tau_n)$ where $\{\tau_n\}$ is such that $(\ell og\, \tau_n)/\ell og\, n \to 0$ and $\tau_n(\ell og\, n)^{(2\beta+2s+1)/\delta - 2r-1} \to 0$.

PROOF OF THEOREM 2.2. These results follow directly from Lemmas 3.3 and 3.4 after noting that it is sufficient to study (2.2) with $c_n(\lambda)$ set equal to zero. ////

Remarks: 1) For the problem of estimating f, both the optimal bandwidth and the optimal mean integrated squared error decrease algebraically when φ_k is a proper characteristic function. The theorem indicates that for estimation of g, the corresponding rates are logarithmic. Consequently optimal estimators of g are not obtained by deconvolving optimal estimators of f. This lack of invariance was noted by Rice (1986) in a similar context and is related to the restrictions imposed on the kernels employed. Stefanski and Carroll (1987) have shown that the optimal bandwidths for estimating g and f are in general identical when the improper kernel, $(\pi x)^{-1}\sin x$, is employed. The corresponding transform, $1(|t| \le 1)$, does not satisfy (A1).

2) In contrast to standard kernel density estimation the optimal rate of decrease of $MISE_n(\lambda_n)$ is obtained by *known* deterministic sequences. However, this is a small consolation in light of the lethargic behavior of $MISE_n(\lambda_n)$.

2.2. Strong Uniform Consistency

Using a recent result of Csorgo (1985), a bound on the rate of convergence of Γ_n to zero is easily obtained under tail conditions on f.

THEOREM 2.3. In addition to previously stated conditions, assume that $\Pr(|X|>x) \le Lx^{-\tau}$ for all x large enough where L and τ are arbitrary positive constants. If $\{\lambda_n\}$ is any nonincreasing sequence of constants such that for some $A > 0$ liminf $n^A \lambda_n > 0$, then limsup $R_n\Gamma_n$ is bounded almost surely where

$$R_n = (n/\ell og\ n)^{1/2}\lambda_n^{\beta+1-(r+1)\delta}\exp(-\gamma\lambda_n^{-\delta}) \ .$$

If $R_n \to \infty$, then $\Gamma_n \to 0$ almost surely.

PROOF OF THEOREM 2.3. Since

$$\hat{g}(x) - E\hat{g}(x) = (2\pi)^{-1}\int_{-1/\lambda}^{1/\lambda} e^{itx}\{\hat{\varphi}_f(t) - \varphi_f(t)\}\varphi_k(\lambda t)/\varphi_h(t)dt \ ,$$

$$\Gamma_n = \sup_x|\hat{g}(x) - E\hat{g}(x)| \le (\pi)^{-1}\ \Delta_n \int_0^{1/\lambda} \varphi_k(\lambda t)/\varphi_h(t)\ dt$$

where

$$\Delta_n = \sup_{|t|\le 1/\lambda} |\hat{\varphi}_f(t) - \varphi_f(t)| \ .$$

Example (1) of Csorgo (1985) implies that

$$\limsup \left(\frac{n}{\ell og\ n}\right)^{1/2} \Delta_n < \infty, \text{ almost surely.} \tag{2.3}$$

Combining (2.3) with the result in Lemma 3.1 on the asymptotic behavior of $\int_0^{1/\lambda} \varphi_k(\lambda t)/\varphi_h(t)dt$ shows that limsup $R_n\Gamma_n$ is bounded almost surely, hence if $R_n \to \infty$, $\Gamma_n \to 0$ almost surely. ////

Remarks: 1) If $\sup_x|\hat{E}g(x) - g(x)| \sim D\lambda^s$, then to a first-order approximation Theorem 2.3 can be used to show that there exists a constant, C, such that for all n large enough

$$\sup_x |\hat{g}(x) - g(x)| \le C\left(\frac{\ell og\ n}{n}\right)^{1/2} \lambda^{(r+1)\delta-\beta-1}\exp(\gamma\lambda^{-\delta}) + D\lambda^s, \text{ almost surely.}$$

And hence it is possible to deduce from Lemmas 3.3 and 3.4 that for $\lambda \sim (2\gamma/\ell og\ n)^{1/\delta}$, limsup $(\ell og\ n)^{s/\delta} \sup_x|\hat{g}(x) - g(x)|$ is bounded almost surely. Although the inequalities employed in proving Theorem 2.3 are crude, the resulting convergence rate of order $(\ell og\ n)^{s/\delta}$ is probably not. For example, in the case of normal measurement error (δ=2) Carroll and Hall (1987) have shown that the rate of convergence of $\hat{g}(x_0)$ to $g(x_0)$ for fixed x_0 cannot be faster than $(\ell og\ n)^{-s/2}$; thus in this case Theorem 2.3 can be used to determine a sharp bound on the rate of convergence of $\sup_x|\hat{g}(x) - g(x)|$ even though the bound on Γ_n is not sharp.

3. Some Technical Lemmas

The first lemma establishes the rates of growth of certain integrals.

LEMMA 3.1. Under the conditions specified in (A1) and (A2),

$$\int_0^{1/\lambda} \varphi_k(\lambda t)/\varphi_h(t)dt \sim \frac{(-1)^r \varphi_k^{(r)}(1)\alpha}{(\gamma\delta)^{r+1}}\ \lambda^{(r+1)\delta-\beta-1}\exp(\gamma\lambda^{-\delta}); \tag{3.1}$$

and

$$\int_0^{1/\lambda} \varphi_k^2(\lambda t)/\varphi_h^2(t)dt \sim \frac{\{\varphi_k^{(r)}(1)\}^2\alpha^2}{(2\gamma\delta)^{2r+1}}\ \binom{2r}{r}\lambda^{(2r+1)\delta-2\beta-1}\ \exp(2\gamma\lambda^{-\delta}). \tag{3.2}$$

PROOF. Let $J(\cdot)$ be a nonnegative function defined on $(0,\infty)$ such that $J(t) \sim a(t^b)\exp(ct^d)$ where a, c, d > 0.

Using L'Hopitals Rule and induction on j it can be shown that as $\lambda \to 0$

$$\int_0^{1/\lambda} (1-\lambda t)^j\ J(t)dt \sim \frac{aj!}{(cd)^{j+1}}\ \lambda^{(j+1)d-b-1}\ \exp(c\lambda^{-d}) \quad . \tag{3.3}$$

Equations (3.1) and (3.2) follow from (3.3) upon invoking the Taylor series representation

$$\varphi_k(t) = \frac{(-1)^r \varphi_k^{(r)}(1)(1-t)^r}{r!} - \frac{(-1)^r}{r!}\int_t^1 (z-t)^r \varphi_k^{(r+1)}(z)dz$$

and noting that the contribution of the remainder term is of second-order importance. ////

Now for $\lambda > 0$, define

$$J_{1,n}(\lambda) = an^{-c}(\ell og\ n)^b \lambda^d \exp(p\lambda^{-q})$$

and

$$J_n(\lambda) = J_{1,n}(\lambda) + v\lambda^w ,$$

where $0 < a, c, p, q, v, w < \infty$ and $|b|, |d| < \infty$.

The next three lemmas outline the behavior of $\{J_{1,n}(\theta_n)\}$ and $\{J_n(\theta_n)\}$ for sequences $\{\theta_n\}$ converging to zero.

LEMMA 3.2. If $\{\theta_n\}$ is any sequence of positive constants such that $\theta_n \to 0$ and $\limsup J_{1,n}(\theta_n) < \infty$, then $\liminf \theta_n^q\ \ell og\ n \geq p/c$.

PROOF. There exists a number B, $0 < B < \infty$, and an $N = N(b,c)$ such that $J_{1,n}(\theta_n) \leq B$ and $b(\ell og\ \ell og\ n) < c(\ell og\ n)$ for all $n > N$. It follows that for $n > N$,

$$\theta_n^q\ \ell og\ n > \frac{d(\theta_n^q)\ \ell og\ \theta_n + p - \theta_n^q\ \ell og\ (B/a)}{c - \frac{b(\ell og\ \ell og\ n)}{\ell og\ n}} .$$

The right hand side above converges to p/c and the result follows. ////

Note that if $\tau_n = n^{-1} \exp(\frac{p}{c}\ \theta_n^{-q})$, then

$$\limsup\ (\ell og\ \tau_n/\ell og\ n) = -1 + (p/c)/(\liminf\ \theta_n^q\ \ell og\ n) \leq -1+1 = 0. \qquad \text{(A.2)} \quad ////$$

LEMMA 3.3. For any given sequence $\{\theta_n\}$, $(\ell og\ n)^{w/q}\ J_n(\theta_n) \to v(p/c)^{w/q}$ if and only if $\theta_n = \{p/(c\ \ell og\ n\tau_n)\}^{1/q}$ where

$$\ell og\ \tau_n/\ell og\ n \to 0\ ; \qquad (3.4)$$

$$\tau_n^c\ (\ell og\ n)^{b+(w-d)/q} \to 0\ . \qquad (3.5)$$

PROOF. Suppose that $\theta_n^q = p/(c\ \ell og\ n\tau_n)$ where $\{\tau_n\}$ satisfies (3.4) and (3.5). Under (3.4) $\ell og\ n/\ell og\ n\ \tau_n \to 1$. Thus

$$(\ell og\ n)^{w/q}\ v\theta_n^w = v(p/c)^{w/q}(\ell og\ n/\ell og\ n\tau_n) \to v(p/c)^{w/q};$$

and

$$(\ell og\ n)^{w/q}\ J_{1,n}(\theta_n) = a(p/c)^{d/q}\ \tau_n^c(\ell og\ n)^{b+(w-d)/q}\ (\ell og\ n/\ell og\ n\tau_n)^{d/q} \to 0$$

under (3.5). It follows that $(\ell og\ n)^{w/q}\ J_n(\theta_n) \to v(p/c)^{w/q}$ as claimed.

Now suppose that $(\ell og\ n)^{w/q}\ J_n(\theta_n) \to v(p/c)^{w/q}$. It follows that

$$\limsup\ (\ell og\ n)^{w/q}\ v\theta_n^w \le v(p/c)^{w/q} \tag{3.6}$$

and hence that $\theta_n \to 0$ and $\limsup\ \theta_n^q \ell og\ n \le p/c$. In light of Lemma 3.2 this means that $\theta_n^q\ \ell og\ n \to p/c$. Now if $\tau_n = n^{-1}\ \exp(\frac{p}{c}\ \theta_n^{-q})$, then $\theta_n^q\ \ell og\ n = (p/c)/(1 + \ell og\ \tau_n/\ell og\ n)$ and thus it must be that $\ell og\ \tau_n/\ell og\ n \to 0$, i.e., (3.4) holds. It also follows that $(\ell og\ n)^{w/q}\ v\ \theta_n^w \to v(p/c)^{w/q}$ and thus $(\ell og\ n)^{w/q}\ J_{1,n}(\theta_n) \to 0$. But $(\ell og\ n)^{w/q}\ J_{1,n}(\theta_n) \sim a(p/c)^{d/q}\ \tau_n^c(\ell og\ n)^{b+(w-d)q}$ when (3.4) holds and thus (3.5) follows.

LEMMA 3.4. If λ_n minimizes $J_n(\lambda)$ then

$$\text{(i)}\quad \lambda_n^q\ \ell og\ n \to p/c\ ; \tag{3.7}$$

$$\text{(ii)}\quad (\ell og\ n)^{w/q} J_n(\lambda_n) \to v(p/c)^{w/q}\ . \tag{3.8}$$

PROOF. In Lemma 3.3 the existence of a sequence, $\{\theta_n\}$, was established such that $(\ell og\ n)^{w/q}\ J_n(\theta_n) \to v(p/c)^{w/q}$ and hence $J_n(\theta_n) \to 0$. Since $J_n(\lambda_n) \le J_n(\theta_n)$ it must be that $J_n(\lambda_n) \to 0$; but this can only happen if $\lambda_n \to 0$. Consequently Lemma 3.2 implies

$$\liminf \lambda_n^q \, \ell og \, n \geq p/c. \tag{3.9}$$

Suppose now that $\limsup \lambda_n^q \, \ell og \, n > p/c$. By passing to a subsequence, an $\epsilon > 0$ can be found for which $\lambda_n^q \, \ell og \, n > p/c + \epsilon$ infinitely often. This means that $(\ell og \, n)^{w/q} \, J_n(\lambda_n) \geq v(p/c + \epsilon)^{w/q}$ infinitely often which is a contradiction since $(\ell og \, n)^{w/q} \, J_n(\lambda_n) \leq (\ell og \, n)^{w/q} \, J_n(\theta_n) \to v(p/c)^{w/q}$. Thus $\limsup \lambda_n^q \, \ell og \, n \leq p/c$. This inequality and (3.9) imply (3.7). Finally (3.8) follows by noting that $(\ell og \, n)^{w/q} \, v \, \lambda_n^w \leq (\ell og \, n)^{w/q} \, J_n(\lambda_n) \leq (\ell og \, n)^{w/q} \, J_n(\theta_n)$ and both bounds converge to $v(p/c)^{w/q}$. ////

REFERENCES

Carroll, R. J. and Hall, P. (1987). Optimal rates of convergence for deconvolving a density. *J. Am. Statist. Assoc.*, **83**, 1184-1186.

Crump, J. G. and Seinfeld, J. H. (1982). A new algorithm for inversion of aerosol size distribution data. *Aerosol Science and Technology*, 1, 15-34.

Csorgo, S. (1985). Rates of uniform convergence for the empirical characteristic function. *Acta. Sci. Math.*, **48**, 97-102.

Devroye, L. (1989). A note on consistent deconvolution in density estimation. Preprint.

Fan, J. (1988). On the optimal rates of convergence for nonparametric deconvolution problem. Statistics Technical Report 157, University of California, Berkeley.

Liu, M. C. and Taylor, R. L. (1987a). A consistent nonparametric density estimator for the deconvolution problem. Statistics Technical Report 73, University of Georgia.

Liu, M. C. and Taylor, R. L. (1987b). Simulations and computations of nonparametric density estimates for the deconvolution problem. Statistics Technical Report 74, University of Georgia.

Medelsohn, J. and Rice, J. (1982). Deconvolution of micro fluorometric histograms with B-splines. *J. Am. Statist. Assoc.*, **77**, 748-753.

Rice, J. (1986). Choice of smoothing parameter in deconvolution problems. *Contemporary Math.*, **59**, 137-152.

Stefanski, L. A. and Carroll, R. J. (1987). Deconvoluting kernel density estimators. Institute of Statistics Mimeograph Series No. 1909, North Carolina State University.