Ann. Inst. Statist. Math.
Vol. 46, No. 1, 67–75 (1994)

ESTIMATION OF DENSITY FUNCTIONALS

RUDOLF GRÜBEL

*FB 17 (Mathematik-Informatik), Universität-GH Paderborn,
33095 Paderborn, Germany*

(Received March 30, 1992; revised February 23, 1993)

Abstract. Given a sequence of independent random variables with density f we estimate quantities θ of the form $\theta = \int \phi(f(x))dx$, ϕ a known function, by inserting histograms and kernel density estimators for the unknown f. We obtain conditions for consistency and asymptotic normality and discuss the choice of cell size and bandwidth.

Key words and phrases: Density functionals, empirical central limit theorem, estimation.

1. Introduction

Let $X_1, X_2, \dots$ be a sequence of independent and identically distributed random variables with density f. We wish to estimate a real-valued functional of f that can be written as the integral of a known function ϕ of f, i.e., we are interested in quantities of the form

$$\theta = \int \phi(f(x))dx. \tag{1.1}$$

We assume that f is bounded and satisfies a weighted Lipschitz condition (see (2.1) below) and that the known function $\phi : [0, \infty) \to \mathbb{R}$ is continuous and has a continuous and bounded second derivative on $(0, \infty)$; we need $\phi(0) = 0$ for the integral to exist.

Problems of this and related type appear in a variety of situations. They are discussed in Section 4.4 of Prakasa Rao (1983) and in Section 6.5.2 of Silverman (1986). Following Prakasa Rao ((1983), p. 266) the case with $\phi(x) = x^2$ is 'of extreme importance in nonparametric inference problems'; see also Bhattacharyya and Roussas (1969), Schuster (1974), Ahmad (1976) and Jones and Sheather (1991) in connection with this functional. Other related references are Schweder (1975) and van Es (1992).

Most authors deal with the estimator for θ that arises if a kernel density estimate is inserted for the unknown density in (1.1). In some situations the resulting estimator turns out to be $\sqrt{n}$-consistent. This good asymptotic behaviour may seem surprising on first sight, given that the estimate is based on a density estimator which is known to converge at a slower rate. Obviously, via the integral, random terms are added and some sort of 'stochastic cancellation' occurs.

In the present paper we continue to consider estimators of θ that are of the above 'plug-in' type, but here the emphasis is on estimators that are based on histograms. Most papers on the case $\phi(x) = x^2$ exploit the special form of ϕ algebraically. We will show that analytic properties of ϕ, such as those given above, are sufficient to obtain rate of convergence results and, moreover, asymptotic normality (Theorem 2.1). We also consider kernel type estimators (Theorem 2.2) where we need slightly stronger assumptions on the local behaviour of f. It is one of the main conclusions of this investigation that the choice of cell size (if histograms are used) or bandwidth (if we use kernel estimators) is much less important than in the density estimation context, a fact which has been observed by other authors too. Further, it appears that the asymptotics for histogram based and kernel based estimators are identical. A possible practical consequence of our findings therefore is that, at least to the order considered here, there is no advantage in using the kernel method. Histogram-based estimators are, of course, easier to calculate.

These results are given in Section 2. In the third section we briefly discuss our assumptions on ϕ and f.

The methods of this paper can be extended to functionals of the form

$$\theta = \int \phi(f(x))\psi(F(x))\eta(x)dx$$

where F denotes the underlying distribution function and ϕ, ψ and η are known; see Ahmad and Lin (1983) for applications and consistency of kernel type estimators of such quantities. Functionals that involve derivatives of f seem to require a different approach and will not be discussed here; see Chapter 4 of Prakasa Rao (1983) and Hall and Marron (1987), Bickel and Ritov (1988).

2. Main results

We first introduce some notation. For all $n \in \mathbb{N}$, $l \in \mathbb{Z}$ and $h > 0$ let

$$I_{hl} := ((l-0.5)h, (l+0.5)h], \qquad \xi_{nhl} := \sum_{i=l}^{n} 1_{I_{hl}}(X_i), \qquad p_{hl} := P(X_1 \in I_{hl});$$

here 1_A denotes the indicator function of the set A. We define the estimator $\hat{\theta}_{nh}$ of θ by

$$\hat{\theta}_{nh} := \int \phi(\hat{f}_{nh}(x))dx \qquad \text{where} \qquad \hat{f}_{nh} := \sum_{l\in\mathbb{Z}} \frac{1}{nh}\xi_{nhl}1_{I_{hl}}.$$

In words: ξ_{nhl} counts the number of X-values among the first n that land in I_{hl} and $\hat{f}_{nh}$ is the histogram based on the partition $\{I_{hl} : l \in \mathbb{Z}\}$. Note that only finitely many cells are non-empty; in particular, the integral exists. The value $\hat{\theta}_{nh}$ is obtained on using a histogram with cells of length h as an estimator of the unknown density. In our results we let n tend to infinity and take h to depend on n.

For our first result we assume that f satisfies the following condition: there exists a $\delta > 0$ and a bounded function $g : \mathbb{R} \to \mathbb{R}$ with $\int g(x)^2 dx < \infty$ such that

$$|f(x+y) - f(x)| \leq |y| g(x) \qquad \text{for all } x \in \mathbb{R}, |y| < \delta. \tag{2.1}$$

This will be satisfied if f has as bounded derivative f' with $f'(x) = o(|x|^{-\gamma})$ as $|x| \to \infty$ for some $\gamma > 1/2$.

The following theorem gives conditions on the asymptotic behaviour of the sequence of cell sizes $\{h_n\}$ which imply asymptotic normality of the histogram-based estimator of the density functional. An interesting aspect of this result is the fact that rate and limit variance are independent of the choice of the sequence $\{h_n\}$, provided that the sequence satisfies the asymptotic restrictions mentioned earlier.

The theorem will be proved on using the theory of empirical processes. Let P be the distribution of X_1 and let P_n denote the empirical (probability) measure that assigns mass $1/n$ to each of $X_1, \ldots, X_n$. Following empirical process conventions we write $Q(g)$ for $\int g(x) Q(dx)$. Let $Z_n := \sqrt{n}(P_n - P)$ be the empirical process, indexed by (bounded, measurable) functions g via $Z_n(g) = \sqrt{n}(P_n(g) - P(g))$. We write '$\to_{\text{distr}}$' for convergence in distribution and denote the normal distribution with mean μ and variance σ^2 by $N(\mu, \sigma^2)$. Note that $\phi'(f(X_1))$ is a bounded random variable.

THEOREM 2.1. *Under the above assumptions, if $h_n^4 = o(n^{-1})$ and $h_n^{-2} = o(n)$ as $n \to \infty$ then*

$$\sqrt{n}(\hat{\theta}_{nh_n} - \theta) \to_{\text{distr}} N(0, \sigma^2) \qquad \textit{with} \qquad \sigma^2 = \text{var}(\phi'(f(X_1))).$$

PROOF. Fix $n \in \mathbb{N}$ and $l \in \mathbb{Z}$. Write h for h_n to avoid a flood of double indices and assume $h \leq \delta$. Expand ϕ at p_{hl}/h:

$$\phi(y) = \phi\left(\frac{p_{hl}}{h}\right) + \left(y - \frac{p_{hl}}{h}\right)\phi'\left(\frac{p_{hl}}{h}\right) + \frac{1}{2}\left(y - \frac{p_{hl}}{h}\right)^2 \phi''(z_{hl}(y)),$$

where $z_{hl}(y)$ is some number between y and p_{hl}/h. Insert $\hat{f}_{nh}(x)$ for y and integrate over I_{hl}: as $\hat{f}_{nh}(x)$ is constant on I_{hl} we get

$$\int_{I_{hl}} \phi(\hat{f}_{nh}(x))dx = h\Bigg(\phi\left(\frac{p_{hl}}{h}\right) + \left(\frac{1}{nh}\xi_{nhl} - \frac{p_{hl}}{h}\right)\phi'\left(\frac{p_{hl}}{h}\right) \\ + \frac{1}{2}\left(\frac{1}{nh}\xi_{nhl} - \frac{p_{hl}}{h}\right)^2 \phi''(\zeta_{nhl})\Bigg)$$

where ζ_{nhl} is a random variable with values between $\xi_{nhl}/(nh)$ and p_{hl}/h. Now insert $f(x)$ for y in the above Taylor expansion and use $\int_{I_{hl}}(f(x) - p_{hl}/h)dx = 0$:

$$\int_{I_{hl}} \phi(f(x))dx = h\phi\left(\frac{p_{hl}}{h}\right) + \frac{1}{2}\int_{I_{hl}}\left(f(x) - \frac{p_{hl}}{h}\right)^2 \phi''(z_{hl}(f(x)))dx.$$

We can now express the difference between estimator and true parameter value as follows,

$$
\begin{aligned}
(2.2) \qquad \hat{\theta}_{nh} - \theta = & \sum_{l\in\mathbb{Z}} \left(\frac{1}{n}\xi_{nhl} - p_{hl}\right) \phi'\left(\frac{p_{hl}}{h}\right) \\
& + \frac{h}{2}\sum_{l\in\mathbb{Z}} \left(\frac{1}{nh}\xi_{nhl} - \frac{p_{hl}}{h}\right)^2 \phi''(\zeta_{nhl}) \\
& - \frac{1}{2}\sum_{l\in\mathbb{Z}} \int_{I_{hl}} \left(f(x) - \frac{p_{hl}}{h}\right)^2 \phi''(z_{hl}(f(x)))dx.
\end{aligned}
$$

Some care is needed with the interpretation of the sums as l ranges over an infinite set. For the first term it is enough to note that only finitely many of the ξ_{nhl}'s are not equal to 0, that ϕ' is bounded on compact intervals and that the p_{hl}'s sum to a finite value. Below we will show that the second term is of order $o_P(n^{-1/2})$; the arguments used for this imply L^1-summability of the second series for any fixed n and h. Similarly, the bound given below for its individual terms implies absolute summability of the third series.

Since ξ_{nhl} has a binomial distribution with parameters n and p_{hl} we obtain for the second term on the right hand side of (2.2), using $\sum_l p_{hl} = 1$,

$$
\begin{aligned}
E\sum_{l\in\mathbb{Z}} \left| \left(\frac{1}{nh}\xi_{nhl} - \frac{p_{hl}}{h}\right)^2 \phi''(\zeta_{nhl}) \right| & \le \sup_{x>0} |\phi''(x)| \frac{1}{n^2h^2} \sum_{l\in\mathbb{Z}} \operatorname{var} \xi_{nhl} \\
& = O\left(\frac{1}{nh^2}\right).
\end{aligned}
$$

From $h_n^{-1} = o(n^{-1/2})$ and Chebychev's inequality we obtain the desired $o_P(n^{-1/2})$-behaviour.

Note that the third term on the right hand side of (2.2) depends on n through h only; also, no random terms appear in that term. Using (2.1) we get

$$
\begin{aligned}
\int_{I_{hl}} \left(f(x) - \frac{p_{hl}}{h}\right)^2 dx & = h^{-2}\int_{I_{hl}} \left(\int_{I_{hl}} (f(x) - f(y))dy\right)^2 dx \\
& \le h^2 \int_{I_{hl}} g(x)^2 dx
\end{aligned}
$$

so that

$$
\begin{aligned}
& \left| \frac{1}{2}\sum_{l\in\mathbb{Z}} \int_{I_{hl}} \left(f(x) - \frac{p_{hl}}{h}\right)^2 \phi''(z_{hl}(x))dx \right| \\
& \qquad \le \frac{1}{2}\sup_{x>0} |\phi''(x)| h^2 \int g(x)^2 dx = O(h^2).
\end{aligned}
$$

Using the other bound on h, $h_n = o(n^{-1/4})$, we see that this term too contributes an asymptotically negligible amount to $\sqrt{n}(\hat{\theta}_{nh_n} - \theta)$.

The asymptotic distribution of $\sqrt{n}(\hat{\theta}_{nh_n} - \theta)$ is therefore determined by the first term in (2.2): we can write

$$\sqrt{n}(\hat{\theta}_{nh_n} - \theta) = Z_n(\psi_n) + o_P(1)$$

where $\psi_n = \sum_{l\in\mathbb{Z}} \phi'(p_{hl}/h_n)1_{I_{hl}}$. Let $\psi_\infty := \phi' \circ f$ and let $\mathcal{F}$ be the set of functions ψ_n, $n \in \mathbb{N} \cup \{\infty\}$. These are bounded and measurable functions on $\mathbb{R}$. For such functions ψ the following norms are finite,

$$\|\psi\|_\infty := \sup_{x\in\mathbb{R}} |\psi(x)|, \qquad \|\psi\|_{2,P} := (P(\psi^2))^{1/2}, \qquad \|\psi\|_{2,n} := (P_n(\psi^2))^{1/2}.$$

We obviously have $\|\psi\|_{2,P} \le \|\psi\|_\infty$, $\|\psi\|_{2,n} \le \|\psi\|_\infty$. In particular, $\mathcal{F}$ is a subset of the L^2-spaces associated with P and P_n.

We regard Z_n as a random quantity with values in the space of bounded functions on $\mathcal{F}$ and we plan to show that Z_n converges in distribution as $n \to \infty$. This will follow from the central limit theorem for empirical processes as given in Pollard ((1984), p. 157), once we have shown that $\mathcal{F}$ is a pointwise bounded, totally bounded, permissible subset of $L^2(P)$ and that the sequence $\{Z_n\}$ is uniformly stochastically equicontinuous, i.e.

$$\forall \eta > 0 \forall \epsilon > 0 \exists \delta > 0 : \limsup_{n\to\infty} \mathbb{P}\left(\sup_{(f,g)\in F(\delta)} |Z_n(f) - Z_n(g)| > \eta \right) < \epsilon \tag{2.3}$$

where $F(\delta) = \{(f,g) : f,g \in \mathcal{F}, \|f-g\|_{2,P} < \delta\}$ and $\mathbb{P}$ denotes the probability on the background probability space on which the random variables X_n, $n \in \mathbb{N}$, are defined.

From our assumptions it follows that ϕ' is bounded on compact intervals, also, f is bounded. This implies $\|\psi_\infty\|_\infty < \infty$ and further, using

$$|\psi_n(x)| \le \sup_{l\in\mathbb{Z}} |\phi'(p_{hl}/h)|, \qquad p_{hl}/h \le \|f\|_\infty,$$

that $\sup\{|\psi_n(x)| : x \in \mathbb{R}, n \in \mathbb{N}\} < \infty$; hence $\mathcal{F}$ is pointwise bounded. From the mean value theorem and boundedness of g in (2.1) we obtain the existence of some $c < \infty$ such that

$$\|\psi_n - \psi_\infty\|_\infty \le ch_n \qquad \text{for all } n \in \mathbb{N}. \tag{2.4}$$

This shows that $\mathcal{F}$ consists of an $L^2(P)$-converging sequence, together with its limit. As a consequence $\mathcal{F}$ is compact and therefore totally bounded.

Permissibility is needed to obtain measurability of quantities such as $\sup_{f\in\mathcal{F}} |Z_n(f)|$. In our case $\mathcal{F}$ is countable and all such quantities are automatically measurable. Alternatively, it is straightforward to check the conditions listed in the definition of permissibility in Pollard ((1984), p. 196).

Finally, we have to check stochastic equicontinuity. By the equicontinuity lemma, see p. 150 in Pollard (1984), (2.3) will follow from the existence of an envelope to $\mathcal{F}$ in $L^2(P)$, which in turn follows from the above boundedness argument, and

$$\forall \eta > 0 \forall \epsilon > 0 \exists \gamma > 0 : \limsup_{n\to\infty} \mathbb{P}(J_2(\gamma, P_n, \mathcal{F}) > \eta) < \epsilon \tag{2.5}$$

where

$$J_2(\gamma, P_n, \mathcal{F}) = \int_0^\gamma \left(2 \log \frac{N_2(u, P_n, \mathcal{F})^2}{u} \right)^{1/2} du$$

and $N_2(u, P_n, \mathcal{F})$ denotes the number of balls of radius u in $\|\cdot\|_{2,n}$-norm that are needed to cover $\mathcal{F}$. Centering one of these balls at the limit function ψ_∞ we see that $N_2(u, P_n, \mathcal{F})$ is bounded from above by one plus the number of ψ_n's outside a ball of radius u about ψ_∞ with respect to $\|\cdot\|_\infty$. From (2.4) and $h_n = O(n^{-1/4})$ it follows that the latter is at most $O(\delta^{-4})$. This means that $J_2(\gamma, P_n, \mathcal{F})$ can be bounded by a (deterministic) function $l(\gamma)$ with $l(\gamma) \downarrow 0$ as $\gamma \downarrow 0$ (the same elementary calculations as in Pollard ((1984), p. 146) apply); in particular, (2.5) holds and the equicontinuity lemma applies.

This completes the check of the assumptions of the empirical central limit theorem. By the latter, $Z_n \to_{\text{distr}} Z$ where Z is a Gaussian process with mean 0 and covariance function

$$\begin{aligned} \operatorname{cov}(Z(f_1), Z(f_2)) &= P(f_1 - P(f_1))(f_2 - P(f_2)) \\ &\quad (= \operatorname{cov}(f_1(X_1), f_2(X_1))) \qquad \text{for all } f_1, f_2 \in \mathcal{F}; \end{aligned}$$

Z has almost all its paths continuous with respect to the $L_2(P)$-seminorm. In particular,

$$Z(\psi_n) \to Z(\psi_\infty) \qquad \text{almost surely.} \tag{2.6}$$

According to the Skorohod-Dudley representation theorem (Pollard (1984), p. 71) there exist $\tilde{Z}_n$, $\tilde{Z}$ on a suitable probability space such that $\tilde{Z}_n$ has the same distribution as Z_n and $\tilde{Z}$ has the same distribution as Z such that $\tilde{Z}_n$ converges to $\tilde{Z}$ almost surely, i.e.,

$$\sup_{\psi \in \mathcal{F}} |\tilde{Z}_n(\psi) - \tilde{Z}(\psi)| \to 0 \qquad \text{almost surely.}$$

As (2.6) also holds for $\tilde{Z}$ we obtain

$$\begin{aligned} |\tilde{Z}_n(\psi_n) - \tilde{Z}(\psi_\infty)| &\le \sup_{\psi \in \mathcal{F}} |\tilde{Z}_n(\psi) - \tilde{Z}(\psi)| + |\tilde{Z}(\psi_n) - \tilde{Z}(\psi_\infty)| \\ &\to 0 \qquad \text{almost surely as } n \to \infty. \end{aligned}$$

Now note that $\tilde{Z}_n(\psi_n)$ has the same distribution as $Z_n(\psi_n)$. As explained above the latter differs from $\sqrt{n}(\hat{\theta}_n - \theta)$ by $o_P(1)$ only. Almost sure convergence implies convergence in distribution, hence the statement of the theorem follows on noting that the asserted limit distribution is the distribution of $Z(\psi_\infty)$. □

A technical aside in connection with the above proof: it is tempting to write

$$\begin{aligned} |Z_n(\psi_n) - Z(\psi_\infty)| &\le \sqrt{n} P_n(|\psi_n - \psi_\infty|) + \sqrt{n} P(|\psi_n - \psi_\infty|) \\ &\le 2\sqrt{n} \|\psi_n - \psi_\infty\|_\infty. \end{aligned}$$

However, the rate of convergence of ψ_n to ψ_∞ is not fast enough for the last expression to be of order $o_P(1)$. Indeed, some cancellation occurs if we subtract $P(\psi_n - \psi_\infty)$ from $P_n(\psi_n - \psi_\infty)$; this effect is lost in the above crude argument.

We now consider the estimator of θ that arises if a kernel density estimator instead of a histogram is used as an estimator for the density. Let K, the kernel, be a probability density, symmetric about 0 and with support $[-1,1]$. For $h > 0$ let K_h be defined by $K_h(x) = h^{-1}K(h^{-1}x)$. The kernel density estimate for f with kernel function K and bandwidth h and the associated estimator of θ are then given by

$$\tilde{f}_{nh}(x) := \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i), \qquad \tilde{\theta}_{nh} = \int \phi(\tilde{f}_{nh}(x))dx.$$

The role of the cell size of histograms is taken over by the bandwidth if we use kernel density estimators. We need a stronger assumption on the local behaviour of f: we assume that f has a derivative f' that satisfies, for some $\delta > 0$,

$$|f'(x+y) - f'(x)| \le |y|g(x) \qquad \text{for all } x \in \mathbb{R},\ |y| < \delta, \tag{2.7}$$

where g is an integrable function.

THEOREM 2.2. *Under the above assumptions, if $h_n^4 = o(n^{-1})$ and $h_n^{-2} = o(n)$ as $n \to \infty$ then*

$$\sqrt{n}(\tilde{\theta}_{nh_n} - \theta) \to_{\text{distr}} N(0,\sigma^2) \qquad \textit{with} \qquad \sigma^2 = \operatorname{var}(\phi'(f(X_1))).$$

PROOF. Write K_n for K_{h_n} etc. and assume $h < \delta$. We have

$$\phi(y) = \phi(f(x)) + (y - f(x))\phi'(f(x)) + \frac{1}{2}(y - f(x))^2\phi''(\xi(x,y))$$

with some $\xi(x,y)$ between y and $f(x)$. This gives

$$\begin{aligned}\sqrt{n}(\tilde{\theta}_{nh} - \theta) &= \sqrt{n}\left(\int \tilde{f}_{nh}(x)\phi'(f(x))dx - E\int \tilde{f}_{nh}(x)\phi'(f(x))dx\right) \\ &\quad + \sqrt{n}\int (E\tilde{f}_{nh}(x) - f(x))\phi'(f(x))dx \\ &\quad + \sqrt{n}\frac{1}{2}\int (\tilde{f}_{nh}(x) - f(x))^2\phi''(\xi(x,\tilde{f}_{nh}(x))dx.\end{aligned} \tag{2.8}$$

We consider each term on the right hand side of (2.8) separately. Let

$$\psi_n(y) := \int K_n(x-y)\phi'(f(x))dx;$$

it is straightforward to deduce from our assumptions on ϕ and f that

$$\|\psi_n - \phi' \circ f\|_\infty = O(h_n).$$

With $\hat{F}_n$ denoting the empirical distribution function we have

$$\int \tilde{f}_{nh}(x)\phi'(f(x))dx = \int \psi_n(y)\hat{F}_n(dy)$$

so the first term on the right hand side of (2.8) can be written as $Z_n(\psi_n)$, where Z_n again denotes the empirical process. The reader will recognize that this is similar to the situation considered in the proof of Theorem 2.1. Indeed, the same arguments as given there apply, and it follows that the first term converges in distribution to the normal distribution given in the statement of the theorem.

It remains to obtain $o_P(1)$-behaviour for the two other terms in (2.8). We have

$$E\tilde{f}_{nh}(x) - f(x) = \int K(y)(f(x+hy) - f(x))dy.$$

Also, from (2.7),

$$|f(x+hy) - f(x) - hyf'(x)| \leq h^2y^2g(x)$$

so that, using symmetry of K,

$$\sqrt{n}\left|\int (E\tilde{f}_{nh}(x) - f(x))dx\right| \leq \sqrt{n}h^2 \int y^2K(y)dy \int g(x)dx = o(1).$$

Finally, from the assumptions that ϕ'' is bounded and Chebychev's inequality we see that for the last term it is enough to show

$$E\int (\tilde{f}_{nh}(x) - f(x))^2dx = o(n^{-1/2}). \tag{2.9}$$

This, however, is a statement on the mean integrated squared error (MISE) of a kernel density estimator. We can now use the standard arguments that yield the well-known statement on the MISE-optimal choice of bandwidth in kernel density estimation (see Section 3.3.1 in Silverman (1984), for example): the MISE can be written as the sum of a squared bias term and a variance term. Under our assumptions on f and K, these are of order $O(h^4)$ and $O(n^{-1}h^{-1})$ respectively; hence, (2.9) follows from our assumptions on h. □

If the density itself has to be estimated then the stronger local assumptions on f, i.e. (2.7) as compared to (2.1), result in a better speed of convergence, for example, $O(n^{-4/5})$ for the kernel method as compared to $O(n^{-2/3})$ for histograms if the mean integrated squared error is considered. This is not the case in the situation here—the asymptotic rate is the same for both the histogram and the kernel based estimator of the density functional. However, both our theorems give sufficient conditions; to what extent these conditions are necessary we do not know.

3. Discussion

An interesting aspect of the results above is the following: if we take the cell length or bandwidth of the estimators to be asymptotically proportional to $n^{-\gamma}$, then the optimal rate $O_P(1/\sqrt{n})$ for the difference between estimator and parameter holds if $1/4 < \gamma < 1/2$. Note that this range does not depend on f or ϕ. If our conditions on ϕ and f are not satisfied then the range of smoothing parameters that entails $O_P(1/\sqrt{n})$-behaviour of $\hat{\theta}_{nh} - \theta$ may well depend on quantitative features of ϕ and f such as the growth rate of ϕ.

Our assumptions on ϕ and f can be classified into global and local ones. Among the four categories arising some are obviously needed, whereas others restrict the applicability of our results. It is clear, for example, that local irregularities of ϕ will be 'invisible through the data'. Radical examples would involve a discontinuous ϕ, but it is also possible to construct examples exhibiting this phenomenon with ϕ and f continuous. On the other hand, interesting functionals exist that are based on ϕ-functions with unbounded second derivative near 0: Silverman ((1986), Section 6.5.3) discusses applications to projection pursuit where $\phi(x) = x \log x$ is of interest. Estimation of entropy is also discussed in Ahmad and Lin (1976) (the proofs in this reference, however, are not correct), see also van Es (1992) and the references there.

References

Ahmad, I. A. (1976). On asymptotic properties of an estimate of a functional of a probability density, *Scand. Actuar. J.*, **3**, 176–181.

Ahmad, I. A. and Lin, P. (1976). A nonparametric estimation of the entropy for absolutely continuous distributions, *IEEE Trans. Inform. Theory*, **22**, 372–375.

Ahmad, I. A. and Lin, P. (1983). Consistency of a nonparametric estimator of a density functional, *Metrika*, **30**, 21–29.

Bhattacharyya, G. K. and Roussas, G. G. (1969). Estimation of a certain functional of a probability density function, *Skand. Aktuarietidskr.*, **3**, 201–206.

Bickel, P. J. and Ritov, Y. (1988). Estimating integrated squared density derivatives: sharp best order of convergence estimates, *Sankya Ser. B*, **50**, 381–393.

Hall, P. and Marron, J. S. (1987). Estimation of integrated squared density derivatives, *Statist. Probab. Lett.*, **6**, 109–115.

Jones, M. C. and Sheather, S. J. (1991). Using non-stochastic terms to advantage in kernel-based estimation of integrated squared derivatives, *Statist. Probab. Lett.*, **11**, 511–514.

Pollard, D. (1984). *Convergence of Stochastic Processes*, Springer, New York.

Prakasa Rao, B. L. S. (1983). *Nonparametric Functional Estimation*, Academic Press, New York.

Schuster, E. F. (1974). On the rate of convergence of an estimate of a functional of a probability density, *Scand. Actuarial J.*, **1**, 103–107.

Schweder, T. (1975). Window estimation of the asymptotic variance of rank estimators of location, *Scand. J. Statist.*, **2**, 113–126.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

van Es, B. (1992). Estimating functionals related to a density by a class of statistics based on spacings, *Scand. J. Statist.*, **19**, 61–72.