

10. *Novikov A. A.* Consistency of LS-estimates in regression models with martingale errors. — In: *Statistics and Control of Stochastic Processes*. Berlin-Heidelberg: Springer-Verlag, 1985, p. 389-409.

Поступила в редакцию  
26.VIII.1993

© 1995 r.

GOLDSTEIN L.\*, KHAS'MINSKII R.\*\*

ON EFFICIENT ESTIMATION  
OF SMOOTH FUNCTIONALS

The problem of the estimation of smooth functionals  $\Lambda$  defined on a set of densities  $\mathcal{F}$  is considered. A simple «plug-in» estimator  $\Lambda(\hat{f}_n)$  is shown to be asymptotically efficient in the sense of Levit [5], [6], where  $\hat{f}_n$  is an «undersmoothed» kernel estimate of the density  $f$ . The approach is compared to others in the literature.

*Key words and phrases:* independent observations, a «plug-in» estimator, a kernel estimator, a locally asymptotically minimax estimator, a smooth functional, asymptotic efficiency.

**Introduction.** Given independent observations from an unknown density function  $f$  the question of estimating the value of a smooth functional  $\Lambda$  at  $f$  has been considered in a number of papers, for instance, [6], [3], [1].

In this note we restrict ourselves to the problem of determining what conditions are necessary to insure the asymptotic efficiency, in the sense of Levit [5] and Ibragimov and Khas'minskii [4], of the simple «plug-in» estimator  $\Lambda(\hat{f}_n)$ , where  $\hat{f}_n$  is an «undersmoothed» kernel estimator of  $f$ . Our approach is similar to [6] and [2]. As in [2], the bandwidth  $b$  used in the estimate  $\hat{f}_n$  is not optimal for the estimation of  $f$ ; rather, we take advantage of the smooth, integral nature of the derivative of the functional  $\Lambda$  and undersmooth. That is, as a certain variance may be bounded independently of  $b$ , by choosing a small bandwidth, that is, by undersmoothing, the bias term becomes negligible and the estimator's behavior is determined by the average of a sum of independent and identically distributed random variables, which, when scaled, converge to normal with an appropriate constant.

**1. Assumptions and notation.** Given  $X_1, X_2, \dots, X_n$  independent random variables with density  $f$ , we consider the estimation of  $\Lambda(f)$ , where  $\Lambda$  is a functional defined on a set of densities  $\mathcal{F}$  in which  $f$  is known to lie. We use  $\|h\|_q$  to denote the usual  $L_q$ -norm of the function  $h$  and  $\|h\|_f^2$  to denote  $\int h^2 f$ . For positive constants  $r$  and  $L$  with  $r = p + \gamma$ ,  $p$  an integer and  $0 < \gamma \leq 1$ , let

$$W_q^r(L) = \left\{ f \in C^p: \left\| f^{(p)}(\cdot + y) - f^{(p)}(\cdot) \right\|_q < L|y|^\gamma \right\}.$$

We use  $X$  to denote a random variable with density function  $f$  and  $C$  will denote a positive constant, not necessarily the same at each occurrence. Our result requires the following assumptions.

\*Department of Mathematics, University of Southern California, Los Angeles, California, USA.

\*\*Institute for Problems of Information Transmission, Moscow, Russia.

Research was supported in part by National Science Foundation grant DMS 90-05833.

A1. The functional  $\Lambda$  is Fréchet differentiable in  $L_2$  with derivative  $T(f, x)$ ; that is, with  $h, f \in L_2$  and with

$$T_f g = \int T(f, x) g(x) dx,$$

we have

$$\Lambda(h) - \Lambda(f) - T_f(h - f) = o(\|h - f\|_2).$$

Furthermore,  $T(f, x)$  satisfies a Hölder condition of index  $\alpha$ :

$$\|T(h, \cdot) - T(f, \cdot)\|_2 \leq C \|h - f\|_2^\alpha.$$

A2. There exist positive constants  $r$  and  $L$  such that  $\mathcal{F} \subset W_q^r(L)$  for  $q \in \{2, \infty\}$ .

A3. The Hölder index  $\alpha$  satisfies  $\alpha > 1/(2r - 1)$ .

A4. There exist positive constants  $c_1, c_2$  such that  $c_1 < \text{Var} T(f, X) < c_2$  for all  $f \in \mathcal{F}$ .

Furthermore,

$$\begin{aligned} \sup_{f \in \mathcal{F}} \|T(f, \cdot + y) - T(f, \cdot)\|_f^2 &\rightarrow 0 \text{ as } y \rightarrow 0, \\ \sup_{f \in \mathcal{F}} \|T(f, \cdot)\|_f < \infty \text{ and } \sup_{f \in \mathcal{F}} \|T(f, \cdot)\|_2 < \infty. \end{aligned}$$

Let  $\hat{f}_n$  be the kernel estimator of  $f$  and bandwidth  $b$  based on the independent observations  $X_1, X_2, \dots, X_n$  with density  $f$ ;

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{b} K\left(\frac{X_i - x}{b}\right). \tag{1}$$

We take the kernel  $K$  to have support  $[-1, 1]$ , be bounded, and satisfy

$$\int_{-1}^1 K(x) dx = 1, \quad \int_{-1}^1 x^j K(x) dx = 0, \quad j = 1, 2, \dots, p. \tag{2}$$

With  $c$  an arbitrary positive constant, we take the bandwidth  $b = b_n$  as

$$b_n = cn^{-\beta}, \quad \text{where } \frac{1}{2r} < \beta < \frac{\alpha}{\alpha + 1}. \tag{3}$$

2. Main result  
Theorem 1. Let  
of the form (1) with  
uniformly for  $f \in \mathcal{F}$

where  $\sigma_f^2 = \text{Var}_f T(f, X)$

Corollary 1. If  
minimax (LAM) wh  
exists no estimator  
of (5).

Proof. It fol

Hence, both assertio

Lemma 1. Let  
A4, uniformly for  $f$

and

Lemma 2. Let

For the proof of

into bias and error t

and

First we establish an  
Proposition 1.  
A2 there exists a co  
Proof. By e

$$\begin{aligned} B(x) &= \frac{1}{(p-1)} \\ &= \frac{1}{(p-1)} \end{aligned}$$

2. Main result.

**Theorem 1.** *Let assumptions A1, A2, A3, and A4 be satisfied, and suppose  $\hat{f}_n$  is of the form (1) with kernel  $K$  and bandwidth  $b$  satisfying (2) and (3) respectively. Then, uniformly for  $f \in \mathcal{F}$*

$$\mathcal{L}\left(\sqrt{n}\left(\Lambda(\hat{f}_n) - \Lambda(f)\right)\right) \Rightarrow \mathcal{N}(0, \sigma_f^2) \text{ as } n \rightarrow \infty, \tag{4}$$

where  $\sigma_f^2 = \text{Var}_f T(f, X)$ . Furthermore

$$\mathbb{E}_f \left( \sqrt{n} \left( \Lambda(\hat{f}_n) - \Lambda(f) \right) \right)^2 \rightarrow \sigma_f^2. \tag{5}$$

**Corollary 1.** *It follows from [5] that the estimator  $\Lambda(\hat{f}_n)$  is locally asymptotically minimax (LAM) when the conditions of the theorem are fulfilled. In particular, there exists no estimator that uniformly achieves a smaller constant on the right-hand side of (5).*

**P r o o f.** It follows from assumption A1 that

$$\left| \Lambda(\hat{f}_n) - \Lambda(f) - T_f(\hat{f}_n - f) \right| < C \|\hat{f}_n - f\|_2^{1+\alpha}.$$

Hence, both assertions of the theorem follow from the two lemmas below.

**Lemma 1.** *Let  $\hat{f}_n$  be given by (1), (2), and (3). Then, under assumptions A2 and A4, uniformly for  $f \in \mathcal{F}$ ,*

$$\mathcal{L}\left(\sqrt{n}T_f(\hat{f}_n - f)\right) \Rightarrow \mathcal{N}(0, \sigma_f^2),$$

and

$$\mathbb{E}\left(\sqrt{n}T_f(\hat{f}_n - f)\right)^2 \rightarrow \sigma_f^2.$$

**Lemma 2.** *Let  $\hat{f}_n$  be given by (1), (2), and (3). Then, under conditions A2 and A3*

$$n \mathbb{E}_f \|\hat{f}_n - f\|^{2+2\alpha} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

For the proof of these lemmas, we will use the decomposition

$$\hat{f}_n(x) - f(x) = B(x) + Z(x)$$

into bias and error term, where

$$B(x) = \int_{x-b}^{x+b} \frac{1}{b} K\left(\frac{t-x}{b}\right) f(t) dt - f(x) \tag{6}$$

and

$$Z(x) = \hat{f}_n(x) - \mathbb{E}_f \hat{f}_n(x).$$

First we establish an upper bound on the bias  $B$ .

**Proposition 1.** *Let  $\hat{f}_n$  be given as in (1) with  $K$  as in (2). Then, under assumption A2 there exists a constant  $C$  such that  $\|B\|_q \leq Cb^r$  for  $q \in \{2, \infty\}$ .*

**P r o o f.** By equation (6), a change of variable, Taylor's formula and (2), we find

$$\begin{aligned} B(x) &= \frac{1}{(p-1)!} \int_{-1}^1 du \int_0^{bu} f^{(p)}(x+t) (bu-t)^{p-1} K(u) dt \\ &= \frac{1}{(p-1)!} \int_{-1}^1 du \int_0^{bu} \left( f^{(p)}(x+t) - f^{(p)}(x) \right) (bu-t)^{p-1} K(u) dt. \end{aligned}$$

Therefore, by A2 with  $q = \infty$  we have for all  $x$

$$|B(x)| \leq \frac{1}{(p-1)!} \int_{-1}^1 |K(u)| |bu|^{p-1} du \int_{-b}^b |f^{(p)}(x+t) - f^{(p)}| dt \leq C b^{p-1} L b^{\gamma+1} \leq C b^r;$$

hence  $\|B\|_\infty \leq C b^r$ . Analogously,  $\|B\|_2 \leq C b^r$ .

Proof of Lemma 1. Since  $T_f$  is linear,  $T_f(\hat{f}_n - f) = T_f B + T_f Z$ . Firstly we have

$$|T_f B| \leq \|T(f, \cdot)\|_2 \|B\|_2 \leq C b^r$$

so that  $\sqrt{n}|T_f B| < C n^{1/2-r\beta} \rightarrow 0$  uniformly in  $f$ .

Consider  $T_f Z$ . Let

$$\varepsilon_i(x) = \frac{1}{b} K\left(\frac{X_i - x}{b}\right) - \int \frac{1}{b} K\left(\frac{t-x}{b}\right) f(t) dt.$$

Then

$$T_f Z = \frac{1}{n} \sum_{i=1}^n \int T(f, x) \varepsilon_i(x) dx = \frac{1}{n} \sum_{i=1}^n \eta_i,$$

the average of  $n$  independent and identically distributed mean zero random variables.

We have

$$\begin{aligned} & \left| \mathbf{E}_f \int_{-1}^1 \frac{1}{b} K\left(\frac{X-x}{b}\right) T(f, x) dx - \mathbf{E}_f T(f, X) \right| \\ &= \left| \int \int_{-1}^1 \frac{1}{b} K\left(\frac{t-x}{b}\right) f(t) T(f, x) dx dt - \int T(f, X) f(x) dx \right| \\ &= \left| \int \int_{-1}^1 K(u) (f(x+bu) - f(x)) T(f, x) du dx \right| \\ &\leq \int_{-1}^1 \int |K(u)| |f(x+bu) - f(x)| |T(f, x)| dx du \leq C b^\gamma \|T(f, \cdot)\|_2 \rightarrow 0 \end{aligned}$$

uniformly for  $f \in \mathcal{F}$ . Hence,

$$\begin{aligned} \text{Var } \eta_i &= \text{Var} \left( \int T(f, x) \frac{1}{b} K\left(\frac{X-x}{b}\right) dx \right) \\ &= \int \left[ \int T(f, x) \frac{1}{b} K\left(\frac{t-x}{b}\right) dx \right]^2 f(t) dt - \left( \mathbf{E}_f \int \frac{1}{b} K\left(\frac{X-x}{b}\right) T(f, x) dx \right)^2 \\ &= \int \left[ \int T(f, x) \frac{1}{b} K\left(\frac{t-x}{b}\right) dx \right]^2 f(t) dt - \left( \mathbf{E}_f T(f, X) \right)^2 + o(1), \end{aligned}$$

by the above. Therefore

$$\begin{aligned} |\text{Var } \eta_i - \sigma_f^2| &= \int \left\{ \left[ \int T(f, x) \frac{1}{b} K\left(\frac{t-x}{b}\right) dx \right]^2 - T^2(f, t) \right\} f(t) dt + o(1) \\ &\leq \left\| \int T(f, x) \frac{1}{b} K\left(\frac{\cdot-x}{b}\right) dx - T(f, \cdot) \right\|_f \end{aligned}$$

uniformly for  $f \in \mathcal{F}$ . the lemma follow.

Proof of Lemma 2.  $n \rightarrow \infty$ . Further we

$$n \|\hat{f}_n - f\|_2^{2+2\alpha} \leq$$

Notice that

$$\|\varepsilon_i\|_2^2 \leq \frac{2}{b^2}$$

Now,

$$\mathbf{E} \|Z\|_2^4 =$$

The terms are mean contributions are the terms where  $i_1 = i_3 = i_4 = i_2 = i_3 = i_4$ . He

$$\frac{1}{n^2} \left( \int \mathbf{E} \varepsilon_1^2(x) dx \right)$$

using the Cauchy-Schwarz

since  $b$  is chosen in a

### 3. Discussion.

[6], [3]. Goldstein and N functionals of regress They show efficiency Hölder index.

$$\begin{aligned} & \times \left\| \int T(f, x) \frac{1}{b} K\left(\frac{\cdot - x}{b}\right) dx + T(f, \cdot) \right\|_f + o(1) \\ & \leq C \left\| \int T(f, x) \frac{1}{b} K\left(\frac{\cdot - x}{b}\right) dx - T(f, \cdot) \right\|_f + o(1) \\ & \leq C \left\{ \int_{-\infty}^{\infty} f(t) dt \int_{-1}^1 |T(f, t - bu) - T(f, t)|^2 du \right\}^{1/2} \\ & \quad + o(1) \rightarrow 0 \text{ as } b \rightarrow 0 \end{aligned}$$

uniformly for  $f \in \mathcal{F}$ . From this assertion and the central limit theorem both assertions of the lemma follow.

**Proof of Lemma 2.** It follows from the proposition that  $\sqrt{n}\|B\|_2 \rightarrow 0$  as  $n \rightarrow \infty$ . Further we have

$$n\|\hat{f}_n - f\|_2^{2+2\alpha} \leq n \left( \mathbf{E} \|\hat{f}_n - f\|^4 \right)^{(1+\alpha)/2} \leq n 8^{(1+\alpha)/2} \left( \mathbf{E} \|Z\|_2^4 + \|B\|_2^4 \right)^{(1+\alpha)/2}.$$

Notice that

$$\|\varepsilon_i\|_2^2 \leq \frac{2}{b^2} \left\{ \int K^2\left(\frac{X_i - x}{b}\right) dx + \int \left[ \mathbf{E} K\left(\frac{X_i - x}{b}\right) \right]^2 dx \right\} \leq \frac{4\|K\|_2^2}{b}.$$

Now,

$$\begin{aligned} \mathbf{E} \|Z\|_2^4 &= \mathbf{E} \left\{ \int \left[ \frac{1}{n} \sum_{i=1}^n \varepsilon_i(x) \right]^2 dx \right\}^2 \\ &= \frac{1}{n^4} \sum_{i_1, i_2, i_3, i_4} \int \int \mathbf{E} (\varepsilon_{i_1}(x) \varepsilon_{i_2}(x) \varepsilon_{i_3}(y) \varepsilon_{i_4}(y)) dx dy. \end{aligned}$$

The terms are mean zero, hence there is no contribution when any index is unpaired. The contributions are therefore in four parts,  $n(n-1)$  terms where  $i_1 = i_2 \neq i_3 = i_4$ ,  $n(n-1)$  terms where  $i_1 = i_3 \neq i_2 = i_4$ ,  $n(n-1)$  terms where  $i_1 = i_4 \neq i_2 = i_3$ , and  $n$  terms where  $i_1 = i_2 = i_3 = i_4$ . Hence the sum is majorized by

$$\frac{1}{n^2} \left( \int \mathbf{E} \varepsilon_1^2(x) dx \right)^2 + \frac{2}{n^2} \int \int (\mathbf{E} \varepsilon_1(x) \varepsilon_1(y))^2 dx dy + \frac{1}{n^3} \mathbf{E} \left\{ \int \varepsilon_1^2(x) dx \right\}^2 \leq \frac{C}{n^2 b^2},$$

using the Cauchy-Schwarz inequality on the second term. Hence,

$$n \mathbf{E} \|\hat{f}_n - f\|_2^{2+2\alpha} \leq C n n^{-1-\alpha} b^{-1-\alpha} + o(1) = o(1),$$

since  $b$  is chosen in accordance with (3). This proves Lemma 2, and hence the theorem.

**3. Discussion.** It is interesting to compare Theorem 1 with analogous results in [2], [6], [3].

Goldstein and Messer [2] study the plug-in estimator above for the estimation of functionals of regression and density functions and their derivatives on a finite interval. They show efficiency in order only and under the more restrictive condition  $\alpha = 1$  on the Hölder index.

Levit [6] considers the estimation of a functional of a distribution function and its derivatives. Aside from studying an estimator whose form is different from the plug-in estimator above, Levit's results differ in two ways from those here. Firstly, Levit's assumptions are more restrictive in some cases. For example, if

$$\Lambda(f) = \int_{-\infty}^{\infty} f^{1+\alpha}(x) dx, \quad 0 < \alpha < 1,$$

the assumptions of Levit's paper [6] are never satisfied, yet the approach above is applicable. Furthermore, when specializing Levit's results to cases where the functional  $\Lambda$  depends only on the density  $f$ , Levit's functionals are of the form

$$\Lambda(f) = \int_{-\infty}^{\infty} \Phi(f, x) f(x) dx$$

for some function  $\Phi$ . In particular, his approach does not apply to functionals of the form, say

$$\Lambda(f) = \int_{-\infty}^{\infty} f(x-1) f(x+1) dx$$

or

$$\Lambda(f) = \int_{-\infty}^{\infty} \Phi \left( \int_{-\infty}^x f^2(u) du, x \right) dx$$

whereas the theorem above applies.

Khas'minskii and Ibragimov [3] consider an estimator of  $\Lambda(f)$  composed of two terms and depending on an arbitrary optimal pointwise estimator of the density  $f$ . The first term in their estimator is  $\Lambda(\hat{f}_n)$ , the same as what appears here except that  $\hat{f}_n$  is chosen to achieve the optimal rate for the estimation of  $f$  at a point; the next term is a second order correction. As they consider the estimation of  $\Lambda$  independent of the method of estimating  $f$ , undersmoothing is not considered. By using the second order term, they are able to construct an estimator which is efficient under the weaker condition on the Hölder index  $\alpha > 1/(2r)$ , yet the form of their estimator is consequently somewhat more complicated, and their result requires an extra technical condition not assumed here.

Finally, we note that under further conditions it is possible to extend the result above and show the efficiency of the plug-in estimator for a more general class of functionals as considered in [6] and [2] which depend on  $f$  and its derivatives.

**Acknowledgments.** The second author thanks the department of Mathematics at the University of Southern California for their hospitality during his visit while this paper was being written.

#### REFERENCES

1. Bickel P. J., Ritov Y. Estimating integrated squared density derivatives: sharp best order of convergence estimates. — *Sankhyā*, Ser. A, 1988, v. 50, p. 381–393.
2. Goldstein L., Messer K. Optimal plug-in estimators for nonparametric functional estimation. — *Ann. Statist.*, 1992. (To appear.)
3. Has'minskii R. Z., Ibragimov I. A. On the nonparametric estimation of functionals. — In: *Proceedings of the Second Prague Symposium on Asymptotic Statistics.* / Ed. by P. Mandl and M. Huškova. Amsterdam: North-Holland, 1979, p. 41–51.
4. Ibragimov I. A., Khas'minskii R. Z. *Statistic Estimation: Asymptotic Theory.* Berlin, New York: Springer-Verlag, 1981.
5. Levit B. On optimality of some statistical estimators. — In: *Proceedings of the Prague Symposium on Asymptotic Statistics. Vol. 2.* Prague: Charles Univ., 1974, p. 215–238.

6. Levit B. Ja. A. Inform. Trans.

© 1995 r.

LINEAR

We consider based on the Gaussian distributions. A Gaussian measure  $E(Y | X)$  is a radonifying operator.

*Key words:* conditional Gaussian

1. Introduction. model  $X = GY + N$  in separable Banach and  $G: C \rightarrow B$  is an estimator for a broad

provided  $R_N$  is invertible; however, the linear subspace of  $B$  of measurable linear estimators are measurable. These are called measurable

In Section 2 we consider  $L: B \rightarrow C$  with respect to a Hilbert space  $H$  and the measure. More precisely, in Section 3 we show that if  $X$  and  $Y$  are jointly Gaussian, the possibility of a rigorous measurable linear estimator (1.1).

This paper is devoted to separable spaces  $B = C$  and

\*Institut für Mathematik, Münster, Federal Republic of Germany. Research was supported by the Deutsche Forschungsgemeinschaft.

6. Levit B. Ja. Asymptotically efficient estimation of nonlinear functionals. — Problems Inform. Transmission, 1978, v. 14, № 3, p. 204–209.

Поступила в редакцию  
20.IV.1992

© 1995 г.

LUSCHGY H.\*

## LINEAR ESTIMATORS AND RADONIFYING OPERATORS

We consider the problem of estimating a signal  $Y$  with values in a Banach space based on the observation  $X$  with values in another Banach space given their joint Gaussian distribution. Linear estimators are defined to be measurable linear transformations. A characterization of measurable linear transformations with respect to a Gaussian measure by radonifying operators is established. The Bayes estimator  $E(Y | X)$  is shown to be a measurable linear transformation and the associated radonifying operator is derived.

*Key words and phrases:* radonifying operator, measurable linear transformation, conditional Gaussian distribution.

1. **Introduction.** Consider the problem of estimating a signal  $Y$  in the observation model  $X = GY + N$ , where the noise  $N$  and  $Y$  are independent Gaussian random vectors in separable Banach spaces  $B$  and  $C$  with covariance operators  $R_N$  and  $R_Y$ , respectively, and  $G: C \rightarrow B$  is a continuous linear operator. In case  $B$  is finite-dimensional, the Bayes estimator for a broad class of loss functions is the linear estimator

$$LX = E(Y | X) = R_Y G^* (G R_Y G^* + R_N)^{-1} X \quad (1.1)$$

provided  $R_N$  is invertible and the means are zero. In the infinite-dimensional model, however, the linear operator on the right-hand side of (1.1) is only defined on a dense subspace of  $B$  of measure zero and typically unbounded. Thus one is led to a notion of a linear estimator which is weaker than that of a continuous linear operator. Here linear estimators are measurable transformations which are linear on a subspace of measure one. These are called measurable linear transformations.

In Section 2 we establish a correspondence between measurable linear transformations  $L: B \rightarrow C$  with respect to a Gaussian measure and radonifying operators  $H: C \rightarrow C$  for a Hilbert space  $H$  which is up to an isometry the reproducing kernel Hilbert space of the measure. Moreover, a Fourier type expansion of such transformations is derived. In Section 3 we show that  $E(Y | X)$  is a measurable linear transformation provided that  $X$  and  $Y$  are jointly Gaussian (but without any other model assumption). This implies the possibility of a rigorous interpretation of (1.1). Namely,  $LX = E(Y | X)$  is the unique measurable linear transformation which extends the operator on the right-hand side of (1.1).

This paper is motivated by the article of Mandelbaum [9]. He deals with Hilbert spaces  $B = C$  and Hilbert-Schmidt operators  $H = C \rightarrow C$  which are known to be the

\*Institut für Mathematische Statistik, Universität Münster, Einsteinstr. 62, D-4400 Münster, Federal Republic of Germany.

Research was supported by a Heisenberg grant of the Deutsche Forschungsgemeinschaft.