APPROXIMATELY INTEGRABLE LINEAR STATISTICAL MODELS IN NON-PARAMETRIC ESTIMATION

by

B. Ya. Levit
University of Maryland

Technical Report #90-37C

Department of Statistics
Purdue University

August 1990

Approximately Integrable Linear Statistical Models
in Non-Parametric Estimation

B. Ya. Levit

Summary

The notion of approximately integrable linear statistical models is introduced to analyze the higher order optimality properties of some common nonparametric estimators. The approximately integrable models suggest a useful approach to a unified treatment of both regular and irregular non-parametric problems. It is shown that with such models any rate of improvement ranging from $(\log\ n)^\alpha/n^2$ to $1/(n(\log\ldots\log\ n)^\alpha)$, $\alpha > 0$, of the classical non-parametric procedures can be anticipated. Both an example of a first order asymptotically optimal estimator with the unusual rate $n^{-1}\log\ n$ and an estimator with an extremely slow unimprovable rate of convergence $1/(\log\ldots\log\ n)^\alpha$ are presented.

The research was supported by the Department of Statistics, Purdue University under grants NSF DMS–8923071 and ONR Contract N00014–89–K–0170.

Approximately Integrable Linear Statistical Models in Non-Parametric Estimation

B. Ya. Levit

1. Introduction

The aim of the present report is to develop the notion of approximately integrable linear (a.i.l.) statistical models related to the study of the "next" order optimality in non-parametric estimation. It appears consistent to keep the exposition at present at the least technical level restricted so far to quadratic losses and scalar valued functionals. At the same time, the reader will probably notice a number of generalizations readily suggesting themselves, some of these to be reported elsewhere.

A useful lower bound for a local minimax risk in estimating such functionals will be derived (Section 3). Based on this bound it will be demonstrated that with a.i.l. models any rate of the "next" order improvement of the (first order) asymptotically optimal estimators may be anticipated ranging from $(\log\ n)^\alpha/n^2$ to $1/(\underbrace{\log\ldots\log}_{k}\ n)^\alpha n$, for $k=1,2,\ldots,\alpha>0$.

Clearly when this is the case the next order improvement may well challenge the asymptotic optimality of a given first order efficient estimator. At the same time, with the a.i.l. models one easily discloses nonparametric problems with first order efficient estimators converging at rates (e.g. $n^{-1}\log\ n$) different from the common one ($1/n$).

Another highlighting point is that the a.i.l. models appear to be rather well tailored to incorporate both *regular* (as e.g. cdf estimation) and *irregular* problems (such as estimation of the derivatives of cdf). Both types can be treated then, along similar lines using the above mentioned lower bound. With this approach one discovers a close relation between the optimal rates of improving the standard estimators of the regular functionals and the optimal estimability rates for the irregular ones.

We introduce a.i.l. models after presenting some prerequisites.

2. Some Preliminaries and Definitions

Let $X_1,\ldots,X_n$ be an independent sample in a measurable space $(\mathcal{X},\mathcal{A})$ with a common distribution F ranging in a given subset $\mathcal{F}$ of distributions defined on $\mathcal{A}$. It will prove convenient to supply $\mathcal{F}$ with a relevant topology $\mathcal{T}$. While different competing measures of closeness on $\mathcal{F}$ are readily available, at this stage it appears difficult to argue conclusively in favor of any particular one. Still mainly for its clear statistical meaning we will make use in the sequel of the topology $\mathcal{T} = \mathcal{T}_1$ on $\mathcal{F}$ induced by the distance in variation just to fix a workable and relatively simple one.

Given a real valued function $\Psi(F)$, $F \in \mathcal{F}$, we address below optimal rates of estimability and, provided first order efficient estimators exist, higher order optimality properties in estimating the unknown value $\Psi(F)$ based on given observations.

Let $\Psi_n = \Psi_n(X_1, \ldots, X_n)$ be an arbitrary estimator of $\Psi(F)$ and

$$R_n(\Psi_n, F) = E_F(\Psi_n - \Psi(F))^2. \tag{2.1}$$

While there are plenty of loss functions one can choose from, the particular one in (2.1) serves well the purposes of this presentation. By an estimator Ψ_n we mean below any sequence of estimators Ψ_n, $n \geq 1$.

Let us recall next some asymptotic properties a reasonable estimator Ψ_n of $\Psi(F)$ is expected to share. The underlying common idea behind the different definitions to be used below is that a "nice" estimator should exhibit reasonable global consistency properties while being locally unimprovable. To this point we present the following definitions keeping in mind their reference to a given underlying set of distributions $\mathcal{F}$.

Definition 1. The function $\Psi(F)$ is called

a) $\rho(n)$-rate estimable if there exists an estimator Ψ_n such that, locally uniformly in F,

$$R_n(\Psi_n, F) = O(\rho(n)), \qquad (n \to \infty);$$

b) exactly $\rho(n)$-rate estimable if it is $\rho(n)$-rate estimable and moreover for any sequence $\rho'(n)$, $\rho'(n)/\rho(n) \to 0$, and any non-empty vicinity $V \in \mathcal{T}$ no estimator Ψ_n satisfies the relation

$$R_n(\Psi_n, F) = O(\rho'(n))$$

uniformly on V.

Assume that $\Psi(F)$ is exactly $\rho(n)$-rate estimable. The next definition refers to the first order asymptotically optimal properties in estimating $\Psi(F)$.

Definition 2. An estimator Ψ_n is called locally asymptotically unimprovable or first order asymptotically optimal if for any non-empty vicinity V and a positive number R there exists n_0 such that for $n > n_0$ no estimator Ψ'_n satisfies the inequality

$$R_n(\Psi'_n, F) \leq R_n(\Psi_n, F) - R\rho(n), \qquad F \in V.$$

Let now Ψ_n be a first order asymptotically optimal estimator. The following definition refers to the "next" order properties of Ψ_n.

Definition 3. The estimator Ψ_n is called

a) $\rho_1(n)$-rate improvable if there exists a non-empty vicinity V, positive number R and an estimator Ψ'_n such that locally uniformly in F

$$R_n(\Psi'_n, F) \leq R_n(\Psi_n, F) - R\rho_1(n)\mathbf{1}(F \in V) + o(\rho_1(n)).$$

Otherwise Ψ_n is called $\rho_1(n)$-rate unimprovable on $\mathcal{F}$ (here $\rho_1(n)/\rho(n) = o(1)$, $n \to \infty$);
b) exactly $\rho_1(n)$-rate improvable if it is $\rho_1(n)$-rate improvable and moreover for any non-empty vicinity V and any sequence $\rho'_1(n)$, $\rho'_1(n)/\rho_1(n) \to \infty$, $n \to \infty$, Ψ_n is $\rho'_1(n)$-rate unimprovable on V.

Let ψ: $x \to R^1$ be a measurable function. It appears the linear functionals of the form

$$\Psi(F) = \int \psi(x)dF(x) \tag{2.2}$$

provide useful approximations to a variety of meaningful nonparametric functionals both regular and irregular.

Let

$$\hat{\Psi}_n = \frac{1}{n}\sum_{i=1}^{n} \psi(X_i).$$

Conditions for asymptotic optimality of the estimator $\hat{\Psi}_n$ were found in Levit (1974); see also Koshevnik and Levit (1976). We summarize below the corresponding result for the sake of reference.

Theorem 2.1. Assume the set $\mathcal{F}$ satisfies the following conditions:

1) $\int \psi^2(x)dF$ is locally bounded in F,

2) for any $F \in \mathcal{F}$ there exist a sequence of functions $\psi_k(x)$ and positive numbers a_k such that $\mathcal{F}$ contains, for any k, the exponential family of distributions G_c defined by the relation

$$\frac{dG_c}{dF}(x) = \exp\{c\psi_k(x) - b(c)\}, \qquad |c| < a_k, \tag{2.3}$$

and

$$\lim_{k\to\infty} \int (\psi_k(x) - \psi(x))^2 dF = 0. \tag{2.4}$$

Then $\hat{\Psi}_n$ is first order asymptotically optimal estimator of $\Psi(F)$.

Conditions (2.2)–(2.4) represent corresponding linearity and integrability properties of the functional $\Psi(F)$; the full meaning of the later term to be explained in subsequent publications. We elaborate further on approximately linear integrable models in the next section.

3. Approximately Integrable Linear Models: Lower Bounds

Let $\Psi(F)$ be a given functional to be estimated from the sample $X_1, \ldots, X_n$ and $\psi_n(X)$, $x \in \mathcal{X}$ — a sequence of real valued measurable functions. Denote

$$\begin{aligned} B_{n,F} &= E_F \psi_n(X) - \Psi(F) \\ \sigma^2_{n,F} &= \operatorname{Var}_F \psi_n(X) \end{aligned} \tag{3.0}$$

where X is distributed according to F.

In the particular case of the functional (2.2) with

$$\sigma^2_F = \operatorname{Var}_F \psi(X) < \infty$$

denote also

$$\Delta_{n,F} = \sigma^2_F - \sigma^2_{n,F} \tag{3.0$'$}$$

Approximately integrable linear (a.i.l.) models to be considered below can be described by the following two assumptions.

Assumption AL (approximate linearity). Locally uniformly in F

$$n^{-1}\sigma^2_{n,F} + B^2_{n,F} = o(1), \qquad n \to \infty.$$

Assumption AI (approximate integrability). For every $F \in \mathcal{F}$ and any of its vicinities V there exists positive a_n such that the exponential family of distributions $G_{n,c}$ defined by the relations

$$dG_{n,c} = g(x,c)dF = \exp\{c\psi_n(x) - b_n(c)\}dF, \tag{3.1}$$

$$b_n(c) = \log \int \exp\{c\psi_n(x)\}dF \tag{3.1$'$}$$

exists and belongs to V for $|c| < a_n$.

Due to the assumption AL,

$$\hat{\Psi}_n = \frac{1}{n}\sum_{i=1}^{n} \psi_n(X_i)$$

is a consistent estimator of $\Psi(F)$ while $\sigma^2_{n,F}$ and $B_{n,F}$ provide an upper bound in estimating $\Psi(F)$ as

$$R_n(\hat{\Psi}_n, F) = n^{-1}\sigma^2_{n,F} + B^2_{n,F}. \tag{3.1''}$$

It is to be shown next that they provide a useful lower bound in estimating $\Psi(F)$ as well.

Whenever the family (3.1) is involved we will denote

$$B_{n,F}(c) = B_{n,G_{n,c}}, \quad \sigma^2_{n,F}(c) = \sigma^2_{n,G_{n,c}}, \quad \Delta_{n,F}(c) = \Delta_{n,G_{n,c}} \tag{3.2}$$

etc.

Let $\phi(a)$ denote the class of continuously differentiable probability densities λ vanishing outside the interval $(-a, a)$ with

$$I(\lambda) = \int_{-a}^{a} \frac{(\lambda'(c))^2}{\lambda(c)}\, dc < \infty.$$

Theorem 3.1. Assume that the family of distributions defined by (3.1) satisfies assumption AI w.r.t. a vicinity V of a given $F_0 \in \mathcal{F}$ and $\lambda_n(\cdot) \in \phi(a_n)$. Then the following inequalities obtain

$$\inf_{\Psi_n} \sup_{F\in V} (R_n(\Psi_n, F) - R_n(\hat{\Psi}_n, F)) \geq -\int_{-a_n}^{a_n} \left(\frac{\lambda_n'(c)}{n\lambda_n(c)} - B_{n,F_0}(c)\right)^2 \lambda_n(c)dc \tag{3.3}$$

$$\inf_{\Psi_n} \sup_{F\in V} (R_n(\Psi_n, F) - n^{-1}\sigma^2_{n,F}) \geq \int_{-a_n}^{a_n} (2n^{-1}\lambda_n'(c)B_{n,F_0}(c) - n^{-2}\frac{(\lambda_n'(c))^2}{\lambda_n(c)})dc \tag{3.4}$$

and in the case of the functional (2.2) with locally bounded $\sigma^2_F = \mathrm{Var}_F\psi(X)$

$$\inf_{\Psi_n} \sup_{F\in V} (R_n(\Psi_n, F) - n^{-1}\sigma^2_F) \geq \int_{-a_n}^{a_n} (-n^{-1}\Delta_{n,F_0}(c) +$$

$$+ 2n^{-1}\lambda_n'(c)B_{n,F_0}(c) - n^{-2}\frac{(\lambda_n'(c))^2}{\lambda_n(c)})dc. \tag{3.5}$$

In applications below we will choose ψ_n appearing in Theorem 3.1 so as to bring the upper bound (3.1″) and lower bounds (3.3)–(3.5) as close as possible in their rates of decrease. It seems tempting to optimize these lower bounds by a particular choice of $\lambda_n(\cdot)$; a task which however we won't pursue here.

Proof. Denote $\theta(c) = \Psi(G_{n,c})$, $\sigma^2(c) = \sigma^2_{n,F_0}(c)$, $B(c) = B_{n,F_0}(c)$, $b(c) = b_n(c)$ and $g^{(n)}(\underline{x}, c) = \prod_{i=1}^{n} g(x_i, c) = \exp\{n(c\hat{\Psi}_n - b(c))\}$. By (3.1′)

$$\theta(c) = b'(c) - B(c). \tag{3.6}$$

Hence

$$\frac{d}{dc}\log\ g^{(n)}(\underline{x}, c) = n(\hat{\Psi}_n - b'(c)) = n(\hat{\Psi}_n - \theta(c) - B(c)). \tag{3.7}$$

Consider the Bayes estimator $\tilde{\Psi}_\lambda$ of $\Psi(F)$ w.r.t. risk function $R_n(\Psi_n, F)$ and the prior distribution induced on the subfamily $G_{n,c} \in V$, $|c| < a_n$, by the $\lambda_n(c)$:

$$\tilde{\Psi}_\lambda = \frac{\int_{-a_n}^{a_n} g^{(n)}(\underline{x}, c)\theta(c)\lambda_n(c)dc}{\int_{-a_n}^{a_n} g^{(n)}(\underline{x}, c)\lambda_n(c)dc} = \hat{\psi}_n + \mu_n(\hat{\psi}_n)$$

where

$$\mu_n(\hat{\Psi}_n) = \frac{\int_{-a_n}^{a_n} g^{(n)}(\underline{x}, c)(\theta(c) - \hat{\psi}_n)\lambda_n(c)dc}{\int_{-a_n}^{a_n} g^{(n)}(\underline{x}, c)\lambda_n(c)dc}.$$

One obtains

$$\begin{aligned} R_n(\tilde{\Psi}_\lambda, \lambda) &= \int_{-a_n}^{a_n} R_n(\tilde{\Psi}_\lambda, G_{n,c})\lambda(c)dc \\ &= \int_{-a_n}^{a_n}\int_{\mathcal{X}^n} (\hat{\Psi}_n + \mu_n(\hat{\Psi}_n) - \theta(c))^2 dG^{(n)}_{n,c}(\underline{x})\lambda_n(c)dc = I_1 + I_2 + I_3, \end{aligned} \tag{3.8}$$

where

$$
\begin{aligned}
I_1 &= \int_{-a_n}^{a_n} \lambda_n(c)dc \int_{\mathcal{X}^n} (\hat{\Psi}_n - \theta(c))^2 dG_{n,c}^{(n)}(\underline{x}) \\
&= \int_{-a_n}^{a_n} (n^{-1}\sigma^2(c) + B^2(c))\lambda_n(c)dc, \qquad (3.9) \\
I_2 &= \int_{\mathcal{X}^n} \mu_n^2(\hat{\Psi}_n)dF^{(n)}(\underline{x}) \int_{-a_n}^{a_n} g^{(n)}(\underline{x},c)\lambda_n(c)dc \\
&= \int_{\mathcal{X}^n} \frac{(\int_{-a_n}^{a_n} g^{(n)}(\underline{x},c)(\theta(c)-\hat{\Psi}_n)\lambda_n(c)dc)^2}{\int_{-a_n}^{a_n} g^{(n)}(\underline{x},c)\lambda_n(c)dc} dF^{(n)}(\underline{x}), \\
I_3 &= 2\int_{\mathcal{X}^n} \mu_n(\hat{\Psi}_n)dF^{(n)}(\underline{x}) \int (\hat{\Psi}_n - \theta(c))g^{(n)}(\underline{x},c)\lambda_n(c)dc \\
&= -2I_2.
\end{aligned}
$$

Thus

$$
R_n(\tilde{\Psi}_\lambda, \lambda) = I_1 - I_2. \qquad (3.10)
$$

One then obtains further that

$$
\begin{aligned}
0 &\le \int_{-a_n}^{a_n} \int_{\mathcal{X}^n} \left(\mu_n(\hat{\Psi}_n) - \left(\frac{\lambda_n'(c)}{n\lambda_n(c)} - B(c)\right)\right) dG_{n,c}^{(n)}(\underline{x})\lambda_n(c)dc \\
&= I_2 + I_4 + I_5, \qquad (3.11)
\end{aligned}
$$

where

$$
I_4 = \int_{-a_n}^{a_n} \left(\frac{\lambda_n'(c)}{n\lambda_n(c)} - B(c)\right)^2 \lambda_n(c)dc \qquad (3.12)
$$

and

$$
\begin{aligned}
I_5 &= -2\int_{\mathcal{X}_n} \mu_n(\hat{\Psi}_n)dF^{(n)}(\underline{x}) \int_{-a_n}^{a_n} g^{(n)}(\underline{x},c)(n^{-1}\lambda_n'(c) - B(c)\lambda_n(c))dc \\
&= -2\int_{\mathcal{X}^n} \mu_n(\hat{\Psi}_n)dF^{(n)}(\underline{x}) \int_{-a_n}^{a_n} (-n^{-1}\frac{dg^{(n)}(x,c)}{dc} - g^{(n)}(x,c)B(c))\lambda_n(c)dc \\
&= -2\int_{\mathcal{X}^n} \mu_n(\hat{\Psi}_n) \int_{-a_n}^{a_n} g^{(n)}(\underline{x},c)(\theta(c)-\hat{\Psi}_n)\lambda_n(c)dcdF^{(n)}(\underline{x}) \\
&= -2I_2
\end{aligned}
$$

where integration by parts and relation (3.7) were used to obtain correspondingly the second and third equalities. Thus $I_4 \geq I_2$ and (3.8)–(3.12) result in

$$R_n(\tilde{\Psi}_\lambda, \lambda) \geq I_1 - I_4 =$$

$$\int_{-a_n}^{a_n} (R_n(\hat{\Psi}_n, G_{n,c}) - \left(\frac{\lambda_n'(c)}{n\lambda_n(c)} - B(c)\right)^2)\lambda_n(c)dc =$$

$$\int_{-a_n}^{a_n} (n^{-1}\sigma^2(c)\lambda_n(c) + (2n^{-1}\lambda_n'(c)B(c) - n^{-2}\frac{(\lambda_n'(c))^2}{\lambda_n(c)}))dc$$

wherefrom the theorem follows.

Our next goal is two-fold. First it will be shown by the use of Theorem 3.1 that any rate of the higher order improvement of first order asymptotically optimal estimators may be anticipated ranging from $(\log\ n)^\alpha n^{-2}$ to $(\underbrace{\log\ \log \ldots \log}_{k}\ n)^{-\alpha} n^{-1}$, $k = 1, 2, \ldots, \alpha > 0$, for approximately integrable models. Second a close resemblance will be exhibited between next order optimal rates of improvement for such estimators and optimal rates of estimability of some non-regular functionals, the common ground for a combined treatment of these rather different problems being furnished by the notion of a.i.l. models.

4. A.i.l. Models: First Applications

Without loss of generality we can restrict ourselves, within the scope of the paper, to estimating the simplest function

$$\Psi(F) = \int_{R^1} x dF.$$

As is well known the tail behavior of the distributions $F \in \mathcal{F}$ is of primary importance in assessing the asymptotic properties of the sample mean

$$\hat{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

We will proceed examining the crucial role of this tail behavior in assessing the best rates of improving upon $\hat{X}_n$. In particular the quantity

$$\xi_F^{(2)}(a) = \int_{|x|>a} x^2 dF$$

will matter as may be inferred from the next lemma.

Lemma 4.1. Assume that for a given space of distributions $(\mathcal{F}, \mathcal{T})$ there exist a) a continuous decreasing to zero function $\xi(\nu)$, $\nu > 0$, a positive γ and for any $F \in \mathcal{F}$ such $\gamma_1(F)$, $\gamma_2(F)$, $\gamma_3(F)$, $0 < \gamma_1(F) \le \gamma_2(F) \le \gamma$, $\gamma_3(\cdot)$ locally bounded in F, that

$$\gamma_1(F)\xi(\nu) \le \xi_F^{(2)}(\nu) \le \gamma_2(F)\xi(\nu), \qquad \nu > \gamma_3(F), \tag{4.1}$$

and b) for any $F \in \mathcal{F}$ and any its vicinity $V \subset F$, a $\delta = \delta(F, V) > 0$ exists, such that the family of distributions G_c of the form

$$\frac{dG_c}{dF} = \exp\{cx^{(\nu)} - b(c)\}, \qquad |c| < \delta\nu^{-1}$$

belongs to V for all sufficiently large ν, where

$$x^{(\nu)} = x\ \mathbf{1}_{(|x|<\nu)}(x).$$

Define $\nu = \nu_n$ by the relation

$$\xi(\nu) = \frac{\nu^2}{2\gamma n}. \tag{4.2}$$

Let ν_n' be any sequence such that for $n \to \infty$

$$\frac{\nu_n'}{\nu_n} = 1 + o(1), \qquad \frac{\xi(\nu_n')}{\xi(\nu_n)} = 1 + o(1). \tag{4.3}$$

Denote $\psi_n(x) = x^{(\nu_n')}$ and let

$$\hat{\Psi}_n = \frac{1}{n}\sum_{i=1}^{n} \psi_n(X_i). \tag{4.3'}$$

Then 1) $\hat{X}_n$ is exactly $(\frac{\nu_n}{n})^2$-rate improvable and 2) locally uniformly in F,

$$R_n(\hat{\Psi}_n, F) \le \frac{\sigma_F^2}{n} - \frac{\gamma_1(F)}{4\gamma}(\frac{\nu_n}{n})^2(1 + o(1)), n \to \infty.$$

Proof. Let

$$\xi_F^{(2)}(\nu, c) = \xi_{G_c}^{(2)}(\nu).$$

With notations (3.0), (3.0′) one obtains

$$|B_{n,F}| \le (\nu_n')^{-1}\xi_F^{(2)}(\nu_n'), \tag{4.4}$$

$$\Delta_{n,F} = \sigma_F^2 - \sigma_{n,F}^2 = \xi_F^{(2)}(\nu_n') - (2\Psi(F) - B_{n,F})B_{n,F}$$
$$= \xi_F^{(2)}(\nu_n')(1+o(1)), \quad (n \to \infty) \tag{4.5}$$

locally uniformly in F, and

$$\xi_F^{(2)}(\nu_n', c) \le e^{2\delta} \xi_F^{(2)}(\nu_n'). \tag{4.6}$$

Now applying Theorem 3.1 with $\psi_n(x) = x^{(\nu_n')}$, $a_n = \delta(\nu_n')^{-1}$, $\lambda_n(c) = \lambda(a_n c)$ where $\lambda(\cdot) \in \phi(1)$, $\lambda_1 = \int_{-1}^{1} |\lambda'(c)| dc$, $\lambda_2 = \int \frac{(\lambda'(c))^2}{\lambda(c)} dc$, $\lambda_1^2 \le \lambda_2 < \infty$, one obtains from (3.5), using (4.2)–(4.6),

$$\inf_{\Psi_n} \sup_{F \in V} (R_n(\Psi_n, F) - n^{-1}\sigma_F^2) \ge$$
$$-n^{-1} \sup_{|c|<a_n} \xi_F^{(2)}(\nu_n', c)(1+\delta^{-1}\lambda_1)(1+o(1)) - \lambda_2 \delta^{-2} (\frac{\nu_n'}{n})^2 \ge$$
$$\ge -n^{-1} e^{2\delta} \gamma_2(F) \xi(\nu_n')(1+\delta^{-1}\lambda_1)(1+o(1)) - \lambda_2\delta^{-2}(\frac{\nu_n'}{\nu})^2 =$$
$$= -\left(\frac{e^{2\delta}\gamma_2(F)}{2\gamma}(1+\delta^{-1}\lambda_1) + \lambda_2\delta^{-2}\right)(\frac{\nu_n}{n})^2(1+o(1)), \quad (n \to \infty),$$

proving assertion 1).

Applying once again (4.2)–(4.6) one obtains further locally uniformly in F:

$$R_n(\hat{\Psi}_n, F) = n^{-1}\sigma_{n,F}^2 + B_{n,F}^2$$
$$\le n^{-1}\sigma_F^2 + (-n^{-1}\xi_F^{(2)}(\nu_n') + (\nu_n')^{-2}(\xi_F^{(2)}(\nu_n'))^2)(1+o(1))$$
$$\le n^{-1}\sigma_F^2 - \frac{1}{2n}\xi_F^{(2)}(\nu_n')(1+o(1))$$
$$\le n^{-1}\sigma_F^2 - \frac{\gamma_1(F)}{2n}\xi(\nu_n)(1+o(1))$$
$$= n^{-1}\sigma_F^2 - \frac{\gamma_1(F)}{4\gamma}(\frac{\nu_n}{n})^2(1+o(1)),$$

proving assertion 2).

We present next a few examples in which Lemma 4.1 can be effectively used to define rates of improvement of the sample mean $\hat{X}_n = n^{-1} \sum_{i=1}^{n} X_i$. Notice that in examples 1–3 $\hat{X}_n$ is a first order asymptotically optimal estimator of $\Psi(F) = E_F X$ due to Theorem 2.1.

Below $\overline{F}(x)$ stands for $1 - F(x) + F(-x)$, $(x > 0)$.

Example 1. For given α, β, γ and a real μ consider the class $\mathcal{F}$ of distribution functions $F(x)$, $(x \in R)$ such that for some $\gamma_1(F)$, $\gamma_2(F)$, $0 < \gamma_1(F) \leq \gamma_2(F) < \gamma$ and locally bounded function $\gamma_3(F)$

$$\gamma_1(F) \leq (x^\mu e^{-\beta x^\alpha})^{-1}\overline{F}(x) \leq \gamma_2(F), \qquad x > \gamma_3(F).$$

Let

$$\nu_n' = \frac{1}{\beta}(\log\ 2\gamma n + \frac{\mu}{\alpha}\log\ \log\ 2\gamma n - \frac{\log\ \beta}{\alpha})^{\frac{1}{\alpha}}$$

and $\hat{\Psi}_n$ be defined as in (4.3′).

Proposition 4.1 a) $\hat{X}_n$ is exactly $\frac{(\log\ n)^{\frac{2}{\alpha}}}{n^2}$-rate improvable on $\mathcal{F}$ and b) locally uniformly in $F \in \mathcal{F}$

$$R_n(\hat{\Psi}_n, F) \leq \frac{\sigma^2(F)}{n} - \frac{\gamma_1(F)}{4\gamma}\frac{(\log\ n)^{\frac{2}{\alpha}}}{n^2}(1 + o(1)), \quad n \to \infty.$$

Notice that the smaller is α, i.e. the heavier are the tails of F, the larger is the improvement rate of the sample mean.

Proof. Using the relations

$$\xi_F^{(2)}(\nu) = \int_\nu^\infty x^2 d\overline{F}(x) = 2\int_\nu^\infty x\ \overline{F}(x)dx + \nu^2\overline{F}(\nu) \tag{4.7}$$

and

$$\int_\nu^\infty x^\mu e^{-\beta x^\alpha} dx = \frac{1}{\alpha\beta}\nu^{\mu+1-\alpha}e^{-\beta\nu^\alpha}(1 + o(1)), \quad (\nu \to \infty)$$

one readily verifies the relation (4.1) with $\xi(\nu) = \frac{1}{\alpha\beta}\nu^{\mu+2-\alpha}e^{-\beta\nu^\alpha}$. Clearly the assumption b) of Lemma 4.1 holds with some δ, $0 < \delta \leq \frac{1}{2}\log\frac{\gamma}{\gamma_2(F)}$. Thus the proposition follows immediately from Lemma 4.1.

Example 2. Assume that for some $\alpha > 2$, $\gamma > 0$

$$\mathcal{F} = \{F | \gamma_1(F) \leq \overline{F}(x)x^\alpha \leq \gamma_2(F), \quad x > \gamma_3(F)\}$$

where $0 < \gamma_1(F) \leq \gamma_2(F) < \gamma$, $\gamma_3(F)$ is locally bounded.

Let

$$\nu_n' = \left(\frac{2\alpha\gamma}{\alpha-2}n\right)^{\frac{1}{\alpha}}$$

and $\hat{\Psi}_n$ be defined as above in (4.3′).

Proposition 4.2 a) $\hat{X}_n$ is exactly $n^{-\frac{2(\alpha-1)}{\alpha}}$-rate improvable on $\mathcal{F}$ and b) locally uniformly in F

$$R_n(\hat{\Psi}_n, F) \leq \frac{\sigma^2(F)}{n} - (\frac{\alpha}{\alpha-2})^{\frac{2}{\alpha}} \frac{\gamma_1(F)}{\alpha^{\frac{2(\alpha-2)}{2}} \gamma^{\frac{\alpha-2}{\alpha}}} n^{-\frac{2(\alpha-1)}{\alpha}}(1+o(1)), \ (n \to \infty).$$

Notice that again the smaller is α the higher is the improvement rate of the sample mean.

Proof. By (4.7) the relation (4.1) holds with $\xi(\nu) = \frac{\alpha}{\alpha-2}\nu^{2-\alpha}$. Thus the Proposition 4.2 is implied again by Lemma 4.1 along the argument already used in proving Proposition 4.1.

Denote

$$\log_k x = \underbrace{\log\ \log \ldots \log}_{k \text{ times}} x$$

Example 3. Assume that for some $\alpha > 1$, $k = 1, 2, \ldots$

$$\mathcal{F} = \{F | \gamma_1(F) \leq \prod_{i=1}^{k-1} \log_i x (\log_k x)^\alpha x^2 \overline{F}(x) \leq \gamma_2(F), \ x > \gamma_3(F)\}$$

for some $0 < \gamma_1(F) \leq \gamma_2(F) < \infty$ and a locally bounded $\gamma_3(F)$.

The example exhibits the following peculiar properties. First the attainable rate of improvement of $\hat{X}_n$ is very high, namely $((\log_k n)^{\alpha-1} n)^{-1}$, which is practically comparable to the order n^{-1} of the leading term of the risk $R_n(\hat{X}_n, F)$ for most sample sizes. This apparently suggests that in a still larger class of nonparametric problems the first order asymptotic optimality of a given estimator cannot be taken as a guard against its improvability in some reasonable applications by appealing to higher order properties.

Second in distinction to the former examples 1, 2 the improving estimator we present below is even second order unimprovable, or second order admissible. This sort of conclusion, which can be drawn, with the help of Theorem 3.1, whenever the bias and variance terms don't match each other, doesn't seem to be excessive, whence the higher order terms of the risk expansion fall close to the leading one.

Let

$$\nu_n' = \begin{cases} \left(\frac{n}{\log\ n}\right)^{1/2}, & \text{if } k > 1, \\ \left(\frac{n}{(\log\ n)^{\alpha}}\right)^{1/2}, & \text{if } k \neq 1. \end{cases}$$

Proposition 4.3 a) $\hat{X}_n$ is exactly $((\log_k n)^{\alpha-1}n)^{-1}$-rate improvable on $\mathcal{F}$; b) locally uniformly in F

$$R_n(\hat{\Psi}_n, F) \leq n^{-1}\sigma_F^2 - 2\gamma_1(F)((\log_k \sqrt{n})^{\alpha-1}n)^{-1}(1+o(1)), \quad (n \to \infty)$$

and c) $\hat{\Psi}_n$ is second order admissible, or $((\log_k n)^{\alpha-1}n)^{-1}$-rate unimprovable on $\mathcal{F}$.

Proof. It follows from (4.7) that locally uniformly in F

$$2\gamma_1(F)(\log_k \nu)^{1-\alpha}(1+o(1)) \leq \xi_F^{(2)}(\nu) \leq 2\gamma_2(F)(\log_k \nu)^{1-\alpha}(1+o(1)), \quad (\nu \to \infty). \tag{4.8}$$

Using relations

$$|B_{n,F}| \leq \xi_F^{(1)}(\nu) = \int_{\nu}^{\infty} x d\overline{F}(x) = \int_{\nu}^{\infty} \overline{F}(x)dx + \nu\overline{F}(\nu) \tag{4.9}$$

one obtains similarly

$$|B_{n,F}| \leq \begin{cases} 2\gamma_2(F)(\nu \log \nu)^{-1}(1+o(1)), & k > 1, \\ 2\gamma_2(F)(\nu(\log \nu)^{\alpha})^{-1}(1+o(1)), & k = 1. \end{cases} \tag{4.10}$$

Thus (3.1″), (4.5) result in the following:

$$\begin{aligned} R_n(\hat{\Psi}_n, F) &= n^{-1}\sigma_F^2 - n^{-1}\xi_F^{(2)}(\nu_n')(1+o(1)) + B_{n,F}^2 \\ &\leq n^{-1}\sigma_F^2 - 2\gamma_1(F)((\log_k \sqrt{n})^{\alpha-1}n)^{-1}(1+o(1)), \end{aligned}$$

proving the second assertion of the proposition.

To prove the first and last statements notice that for any non-void vicinity V of F there exists $\delta > 0$ such that the family $G_{n,c}$ defined by (3.1) with $\psi_n(x) = x^{(\nu_n')}$, $|c| < a_n = \delta(\nu_n')^{-1}$, belongs to V. Now using the inequality (3.3) with $\lambda_n(c)$ as in Lemma 4.1 one obtains from (4.9)

$$\begin{aligned} &\inf_{\Psi_n} \sup_{F \in V}(R_n(\Psi_n, F) - R_n(\hat{\Psi}_n, F)) \geq \\ &\geq -\sup_{|c|<a_n} B_{n,F}^2(c) - \frac{\lambda_1}{a_n n} \sup_{|c|<a_n} |B_{n,F}(c)| - \frac{\lambda_2}{(a_n n)^2} \\ &= \begin{cases} O(n \log\ n)^{-1}, & k > 1 \\ O(n(\log\ n)^{\alpha})^{-1}, & k = 1 \end{cases} = o(n(\log_k n)^{\alpha-1})^{-1}, n \to \infty. \end{aligned} \tag{4.11}$$

Notice that the logarithmic term incorporated in ν_n' is essential only in deriving the lower bound (4.11), while a simpler estimator $\hat{\Psi}_n$ with $\psi_n(x) = x^{(\sqrt{n})}$ satisfies both the assertions b), c) of Proposition 4.3.

So far we have analyzed higher order asymptotic properties $\hat{X}_n$ under progressively heavier tail behavior of the underlying distribution $F \in \mathcal{F}$. It is all but natural to inquire further what happens with this estimator while F ranges over the class

$$\mathcal{F} = \{F\colon\ \gamma_1(F) \le x^2\overline{F}(x) \le \gamma_2(F), \quad x > \gamma_3(F)\}$$

where $0 < \gamma_1(F) \le \gamma_2(F) < \infty$ and $\gamma_3(F)$ is locally bounded.

Notice that $\hat{X}_n$ is no longer first order asymptotically optimal or even risk finite in that case. Still Theorem 3.1 allows us to arrive at a meaningful result and moreover is exhibiting a new kind of phenomena. We shall see that there still exists an asymptotically optimal estimator $\hat{\Psi}_n$ of the mean $E_F X$ which however is in that case only $\log(n)/n$-rate consistent and moreover the normalized risk $\frac{n}{\log\ n} R_n(\hat{\Psi}_n, F)$ does not need to converge.

Define

$$\eta_F^{(2)}(\nu) = \int_0^\nu x^2 d\overline{F}(x)$$

and let

$$\nu_n = \sqrt{n}, \quad \psi_n(x) = x^{(\nu_n)}, \quad \hat{\Psi}_n = \frac{1}{n}\sum_{i=1}^n \psi_n(X_i)$$

Proposition 4.4. a) The functional $\Psi(F) = E_F X$ is exactly $\log(n)/n$-rate estimable on $\mathcal{F}$; b)

$$R_n(\hat{\Psi}_n, F) \le \frac{2\gamma_2(F)\log\ n}{n}(1 + o(1)), \quad n \to \infty,$$

locally uniform in F and c) $\hat{\Psi}_n$ is first order asymptotically optimal and exactly n^{-1}-rate improvable on $\mathcal{F}$.

Proof. The inequality (4.9) when applied to $F \in \mathcal{F}$ gives locally uniformly in F

$$|B_{n,F}| \le \frac{2\gamma_2(F)}{\nu}(1 + o(1)).$$

On the other hand

$$\eta_F^{(2)}(\nu) = \int_0^\nu x^2 dF = \int_0^\nu 2x\overline{F}(x)dx - \nu^2 F(\nu) \le 2\gamma_2(F)\log\ \nu(1 + o(1)),\ \nu \to \infty.$$

Thus

$$R_n(\hat{\Psi}_n, F) = n^{-1} \text{ Var } X^{(\nu)} + B^2_{n,F} \le n^{-1}\eta^{(2)}_F(\nu) + B^2_{n,F} \le$$
$$\le \frac{2\gamma_2(F)\log\, n}{n}(1+o(1)).$$

Now the same argument leading to (4.11) applies with $a_n = \delta\nu_n^{-1}$ and

$$\sup_{|c|<a_n} B^2_{n,F}(c) + \frac{1}{a_n n}\sup_{|c|<a_n}|B_{n,F}(c)| + \frac{1}{(\nu_n n)^2} = O(n^{-1})$$

implying that $\hat{\Psi}_n$ is at most n^{-1}-rate improvable. That it is indeed that rate improvable can be easily demonstrated by considering the estimator $n^{-1}\sum_{i=1}^{n} X_i^{(\delta\nu_n)}$ with δ sufficiently small.

Remark 4.1. The way we have defined the class $\mathcal{F}$ is essential for the crucial assumption AI of Theorem 3.1 to be fulfilled while the very definition of $\mathcal{F}$ allows for the oscillations in the normed risk behavior of $\hat{\Psi}_n$.

Proceeding further with heavy tailed distributions F one is led to considering the nonregular linear functionals still covered by Theorem 3.1 which will allow optimal rates conclusions to be derived for such functionals.

Example 5. Let, for some $\alpha, \gamma, 1<\alpha<2, 0<\gamma<\infty$,

$$\mathcal{F} = \{F|\gamma_1(F) \le x^{-\alpha}F(x) \le \gamma_2(F),\ x > \gamma_3(F)\}$$

where $0<\gamma_1(F)\le\gamma_2(F)<\gamma$, $\gamma_3(F)$ being locally bounded.

Define

$$\begin{aligned} \psi_n(x) &= \min(|x|,\nu) \text{ sign } x \\ \sigma^2_{n,F} &= \text{Var}_F\Psi_n(X) \end{aligned} \tag{4.12}$$

with $\nu=\nu_n\to\infty$ to be defined below.

In asserting lower bounds in this and the next examples we use the following lemma.

Lemma 4.2. Let $V\subset\mathcal{F}$ be a vicinity of a given $F\in\mathcal{F}$ with $\Psi(\cdot)$ bounded on V and the family $G_{n,c}$ be defined by (3.1), (4.12). Assume that $G_{n,c}\in V$, $|c|<\delta\nu_n^{-1}$, for some $\delta>0$. Let $\lambda_n(c)=\lambda(\delta^{-1}\nu_n c)$, where $\lambda\in\phi(1)$ is a symmetric density with $\lambda_2 = \int_{-1}^{1}\frac{(\lambda'(c))^2}{\lambda(c)}dc<\infty$.

Then

$$\inf_{\Psi_n} \sup_{F \in V} (R_n(\Psi_n, F) - \sigma^2, F) \geq$$

$$\geq -\frac{2\nu_n}{\delta n} \int_{\nu_n}^{\infty} \overline{F}(x)dx \int_0^1 (e^{c\delta} - e^{-c\delta})\lambda'(c)dc(1 + o(1)) - \frac{\lambda_2 \nu_n^2}{\delta^2 n^2}, \qquad n \to \infty.$$

Proof. For a fixed δ and $|c| < \delta \nu_n^{-1}$ one obtains

$$e^{b_n(c)} = E_F e^{c\psi_n(X)} = E_F(1 + c\psi_n(X) + o(c\psi_n(X))^2) =$$
$$= 1 + O(\nu_n^{-1}) = 1 + o(1), \quad n \to \infty.$$

Next with $B_{n,F}(c) = E_{G_{n,c}}(\psi_n(X) - X)$ integration by parts results in the following relations

$$\begin{aligned} B_{n,F}(c) &= \int_{\nu}^{\infty} xd(1 - G_{n,c}(x) - G_{n,c}(-x)) + \nu(1 - G_{n,c}(\nu) - G_{n,c}(-\nu)) \\ &= -\int_{\nu}^{\infty} (1 - G_{n,c}(x) - G_{n,c}(-x))dx \\ &= -e^{-b_n(c)} \int_{\nu}^{\infty} (e^{c\nu}(1 - F(x)) - e^{-c\nu}F(-x))dx \\ &= -(1 + o(1)) \int_{\nu}^{\infty} (e^{c\nu}(1 - F(x)) - e^{-c\nu}F(-x))dx. \end{aligned} \tag{4.13}$$

Thus

$$B_{n,F}(c) - B_{n,F}(-c) = -(1 + o(1))(c^{c\nu} - e^{-c\nu}) \int_{\nu}^{\infty} \overline{F}(x)dx$$

so that

$$\int_{|c|<\delta\nu^{-1}} \lambda_n'(c) B_{n,F}(c)dc = \int_0^{\delta\nu^{-1}} \lambda_n'(c)(B_{n,F}(c) - B_{n,F}(-c))dc =$$

$$= -\frac{\nu}{\delta} \int_{\nu}^{\infty} \overline{F}(x)dx \int_0^1 (e^{c\delta} - e^{-c\delta})\lambda'(c)dc(1 + o(1)), \ n \to \infty, \ \nu \to \infty.$$

Hence using the inequality (3.4) of Theorem 3.1 with λ_n as specified gives the result in question.

Consider next an estimator of $\Psi(F) = E_F X$ of the form

$$\hat{\Psi}_n = n^{-1} \sum_{i=1}^{n} \psi_n(X_i) \tag{4.13$'$}$$

where Ψ_n is defined through (4.12) with

$$\nu_n = \rho n^{\frac{1}{\alpha}}, \qquad \rho > 0.$$

Proposition 4.5. a) $\Psi(F)$ is exactly $\frac{n^{2(1-\alpha)}}{\alpha}$-rate estimable on $\mathcal{F}$ and b) locally uniformly in F

$$R_n(\hat{\Psi}_{n,F}) \leq \left(\frac{2\gamma_2(F)}{2-\alpha} \rho^{2-\alpha} + \left(\frac{\gamma_2(F)}{\alpha - 1} \right)^2 \rho^{2(1-\alpha)} \right) n^{\frac{2(1-\alpha)}{\alpha}} (1 + o(1)), \ n \to \infty.$$

Proof. Let V be a non-empty vicinity in $\mathcal{F}$, $F \in V$. Using the family $G_{n,c}$ as in Lemma 4.2 it is easy to check that $G_{n,c} \in V$, for $|c| < \delta \nu_n^{-1}$, and sufficiently small $\delta > 0$. Thus by Lemma 4.2

$$\inf_{\Psi_n} \sup_{F \in V} R_n(\Psi_n, F) \geq \frac{2\gamma_1(F)\nu_n^{2-\alpha}}{\delta(\alpha - 1)n} \lambda_3 (1 + o(1)) - \frac{\lambda_2 \nu_n^2}{\delta^2 n^2} =$$
$$= \left(\frac{\gamma_1(F)\lambda_3}{\delta(\alpha - 1)\rho^\alpha} - \frac{\lambda_2}{\delta^2} \right) \rho^2 n^{\frac{2(1-\alpha)}{\alpha}} (1 + o(1)),$$

where

$$\lambda_3 = - \int_0^1 \lambda'(c)(e^{c\delta} - e^{-c\delta}) dc \tag{4.14}$$

can be made positive by a proper choice of $\lambda(\cdot)$ e.g. by making $\lambda'(c)$ negative for $0 < c < 1$. Choose further ρ small enough to make the bound positive ensuring the lower rate bound as stated.

To prove the last statement one obtains

$$\sigma_{n,F}^2 \leq E_F \psi_n^2(X) = \int_0^\nu x^2 dF + \nu^2 \overline{F}(\nu) =$$
$$= 2 \int_0^\nu x \overline{F}(x) dx \leq \frac{2\gamma_2(F)}{2-\alpha} \nu^{2-\alpha} (1 + o(1)) \tag{4.15}$$

and along the lines of (4.13)

$$|B_{n,F}| = |E_F \psi_n(X) - \Psi(F)| \leq \int_\nu^\infty \overline{F}(x) dx \leq \frac{\gamma_2(F)}{\alpha - 1} \nu^{1-\alpha} (1 + o(1)). \tag{4.16}$$

Thus with $\nu = \nu_n = \rho n^{\frac{1}{\alpha}}$, $\rho > 0$

$$R_n(\hat{\Psi}_n, F) = n^{-1}\sigma^2_{n,F} + B^2_{n,F} \leq$$
$$\leq \left(\frac{2\gamma_2(F)\rho^{2-\alpha}}{2-\alpha} + \left(\frac{\gamma_1(F)}{\alpha-1}\right)^2 \rho^{2(1-\alpha)} \right) n^{\frac{2(1-\alpha)}{n}} (1+o(1))$$

locally uniformly in F.

A slightly different upper bound would result for the estimator $\hat{\Psi}_{n1} = n^{-1}\sum_{i=1}^{n} X_i^{(\nu_n)}$.

$$R_n(\hat{\Psi}_{n1}, F) \leq \left(\left(\frac{2\gamma_2(F)}{2-\alpha} - \gamma_1(F) \right)\rho^{2-\alpha} + \left(\frac{\alpha\gamma_1(F)}{\alpha-1}\right)^2 \rho^{2(1-\alpha)} \right) \times$$
$$\times n^{\frac{2(1-\alpha)}{\alpha}} (1+o(1)).$$

Example 6. Let for some integer $k \geq 1$ and given $\alpha > 1$, $\gamma > 0$

$$\mathcal{F} = \{F | \gamma_1(F) \leq x (\prod_{i=1}^{k-1} \log_i x)(\log_k x)^\alpha \overline{F}(x) \leq \gamma_2(F),\ x > \gamma_3(F)\}$$

with some $0 < \gamma_1(F) \leq \gamma_2(F) < \gamma$, $\gamma_3(F)$ being locally bounded.

Proposition 4.6. The functional $\Psi(F) = E_F X$ is exactly $(\log_k n)^{2(1-\alpha)}$-rate estimable on $\mathcal{F}$ and b) the estimator (4.13′), (4.12) with $\nu = \nu_n = n$ satisfies the relation

$$R_n(\hat{\Psi}_{n,F}) \leq \gamma_2^2(F)(\log_k n)^{2(1-\alpha)}(1+o(1))$$

locally uniformly in F.

Proof. Applying Lemma 4.2 in the same manner as in Proposition 4.5 with

$$\nu = \nu_n = \rho n (\log_k n)^{1-\alpha}$$

one obtains for an arbitrary vicinity $V \in \mathcal{F}$, $F \in V$,

$$\inf_{\Psi_n} \sup_{F \in V} R_n(\hat{\Psi}_{n,F}) \geq$$
$$\left(\frac{2\gamma_1(F)\rho\lambda_1(1+o(1))}{\delta(\log_k \rho n)^{\alpha-1}(\log_k n)^{\alpha-1}} - \frac{\lambda_2\rho^2}{\delta^2(\log_k n)^{2(\alpha-1)}} \right)$$

with a positive δ and $\lambda_1 > 0$. For a sufficiently small ρ this gives a lower bound

$$\inf_{\Psi_n} \sup_{F \in V} R_n(\hat{\Psi}_{n,F}) \geq c(V)(\log_k n)^{2(1-\alpha)}$$

with a positive constant $c(V)$.

Now for the estimator $\hat{\Psi}_n$ (4.15), (4.16) give with $\nu = n$:

$$\sigma^2_{n,F} \leq 2\int_0^\nu x\overline{F}(x)dx \leq 2\gamma_2(F)\int_{\sqrt{\nu}}^{\nu}(\prod_{i=1}^{k-1}\log_i x)^{-1}(\log_k x)^{-\alpha}dx(1+o(1))$$

$$\leq \begin{cases} 2\gamma_2(F)(\log\ \nu)^{-1}\nu(1+o(1)), & k>1 \\ 2\gamma_2(F)(\log\ \nu)^{-\alpha}\nu(1+o(1)), & k=1 \end{cases}, (\nu \to \infty)$$

and

$$|B_{n,F}| \leq \int_\nu^\infty \overline{F}(x)dx \leq \gamma_2(F)(\log_k \nu)^{1-\alpha}(1+o(1)), \quad \nu \to \infty,$$

wherefrom the statement b) follows.

The example just considered appears to be instructive in several aspects. First it exhibits an estimator with an extremely slow, though best attainable, speed of convergence. Next it differs from the previous ones (as well as many other estimation problems) in that the risk of the best convergence rate estimator is mainly contributed by the bias rather than the variance term. Notice that just as in the two previous examples there exists an estimator with quadratic risk tending to zero at the best rate though the sample mean clearly has no even finite second order moments.

The examples 1–6 feature the sort of results one can arrive at with the introduced notion of a.i.l. modes. Further applications to a wider class of functionals will be presented elsewhere.

References

B. Ya. Levit (1974). On optimality of some statistical estimates, *Proc. Prague Sympos. Asymptotic Statist.*, **2**, pp. 215–238.

Yu. A. Koshevnik, B. Ya. Levit (1976). On a non-parametric analogue of the information matrix, *Theor. Prob. Appl.*, **21**, **4**, 738–753.

B. Ya. Levit
The University of Maryland
College Park Campus
Department of Mathematics
College Park, MD 20742
byl@lakisis.umd.edu

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release, distribution unlimited.
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Technical Report #90-37C	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Purdue University 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Department of Statistics West Lafayette, IN 47907	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING / SPONSORING ORGANIZATION Office of Naval Research 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-89-K-0170, NSF Grant DMS-8923071
8c. ADDRESS (City, State, and ZIP Code) Arlington, VA 22217-5000	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
Approximately Integrable Linear Statistical Models in Non-Parametric Estimation

12. PERSONAL AUTHOR(S)
B. Ya. Levit

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical	FROM ____ TO ____	August 1990	21

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The notion of approximately integrable linear statistical models in introduced to analyze the higher order optimality properties of some common nonparametric estimators. The approximately integrable models suggest a useful approach to a unified treatment of both regular and irregular non-parametric problems. It is shown that with such models any rate of improvement ranging from $(\log n)^{\alpha}/n^2$ to $1/(n(\log\ldots\log n)^{\alpha})$, $\alpha > 0$, of the classical non-parametric procedures can be anticipated. Both an example of a first order asymptotically optimal estimator with the unusual rate $n^{-1}\log n$ and an estimator with an extremely slow unimprovable rate of convergence $1/(\log\ldots\log n)^{\alpha}$ are presented.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL B. Ya. Levit, author Shanti S. Gupta, Principal Investigator	22b. TELEPHONE (Include Area Code) 301-454-6284; 317-494-6031 22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted.
All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED