Non-parametric lower bounds and information functions

Novak S.Y.
MDX University London, UK

Proceedings ISNPS-3 Conf.. Springer: Avignon, 2016.

Abstract

We argue that common features of non-parametric estimation appear in parametric cases as well if there is a deviation from the classical regularity condition. Namely, in many non-parametric estimation problems (as well as some parametric cases) unbiased finite-variance estimators do not exist; neither estimator converges locally uniformly with the optimal rate; there are no asymptotically unbiased with the optimal rate estimators; etc..

We argue that these features naturally arise in particular parametric subfamilies of non-parametric classes of distributions. We generalize the notion of regularity of a family of distributions and present a general regularity condition, which leads to the notions of the information index and the information function.

We argue that the typical structure of a continuity modulus explains why unbiased finite-variance estimators cannot exist if the information index is larger than two, while in typical non-parametric situations neither estimator converges locally uniformly with the optimal rate. We present a new result on impossibility of locally uniform convergence with the optimal rate.

Key words: non-parametric estimation, lower bounds, information index.

AMS Subject Classification: 62G32.

1 Introduction

It was observed by a number of authors that in many non-parametric estimation problems the accuracy of estimation is worse than in the case of a regular parametric family of distributions, estimators depend on extra tuning "parameters", unbiased estimators are not available, the weak convergence of normalized estimators to the limiting distribution is not uniform at the optimal rate, no estimator is uniformly consistent in the considered class of distributions. These features have been observed, e.g., in the problems of non-parametric density, regression curve, and tail index estimation (cf. [10], ch. 13, and references therein).

Our aim in this paper is to develop a rigorous treatment of these features through a generalization of the notion of regularity of a family of probability distributions. We argue that features mentioned above (which might have been considered accidental drawbacks of particular estimation procedures) in reality are inevitable consequences of the "richness" of the non-parametric class of distributions under consideration.

We argue that the degree of "richness" of the class of distributions determines the accuracy of estimation. The interplay between the degree of "richness" and the accuracy of estimation can be revealed via the non-parametric lower bounds. In some situations the lower bound to the accuracy of estimation is bounded away from zero, meaning consistent estimation is impossible.

2 Regularity conditions and lower bounds

In a typical estimation problem one wants to estimate a quantity of interest a_P from a sample $X_1, ..., X_n$ of independent and identically distributed (i.i.d.) observations, where the unknown distribution $P = \mathcal{L}(X_1)$ belongs to a particular class $\mathcal{P}$.

If there are reasons to assume that the unknown distribution belongs to a *parametric* family $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$, $\Theta \subset \mathcal{X}$, where $\mathcal{X}$ is $\mathbb{R}^m$ or a Hilbert space, then it is natural to choose $a_{P_\theta} = \theta$. Other examples include $a_P = f_P$, the density of P with respect to a given measure μ (assuming every $P \in \mathcal{P}$ has a density with respect to μ), the tail index of a distribution form the class of regularly varying distributions, etc.

Let
$$d_{_H}, \ d_{_\chi} \ \text{ and } \ d_{_{TV}}$$
denote Hellinger, $\chi^{\mathbf{2}}$ and the total variation distances, respectively.

In the case of a parametric family of probability distributions a typical regularity condition states/implies that

$$d_{_H}^2(P_\theta; P_{\theta+h}) \sim \|h\|^2 I_\theta/8 \quad \text{or} \quad d_{_\chi}^2(P_\theta; P_{\theta+h}) \sim \|h\|^2 I_\theta \tag{1}$$

as $h \to 0$, $\theta \in \Theta, \theta + h \in \Theta$, where I_θ is "Fisher's information". If one of regularity conditions (1) holds, estimator $\hat\theta$ is unbiased, and function $\theta \to \mathbb{E}_\theta\|\hat\theta - \theta\|^2$ is continuous, then
$$\mathbb{E}_\theta\|\hat\theta - \theta\|^2 \geq 1/nI_\theta \qquad (\forall \theta \in \Theta). \tag{2}$$
This is the celebrated Fréchet–Rao–Cramér inequality. Thus, if an unbiased estimator with a finite second moments exists, then the optimal unbiased estimator is the one that turns lower bound (2) into equality.

However, the assumption of existence of unbiased estimators may be unrealistic even in parametric estimation problems. For instance, Barankin [1] gives an example of a parametric estimation problem where an unbiased estimator with a finite second moment does not exist.

Below we suggest a generalisation of the regularity condition for a family of probability distributions, and introduce the notion of an information index. We then present a non-parametric generalisation of the Fréchet–Rao–Cramér inequality. We give reasons why in typical non-parametric estimation problems (as well as in certain parametric ones) unbiased estimators with a finite second moment do not exist.

Notation. Below $a_n \sim b_n$ means $a_n = b_n(1+o(1))$ as $n \to \infty$. We write

$$a_n \gtrsim b_n \tag{$*$}$$

if $a_n \geq b_n(1+o(1))$ as $n \to \infty$.

Recall the definitions of the Hellinger distance d_H and the χ^2-distance d_χ. If the distributions P_1 and P_2 have densities f_1 and f_2 with respect to a measure μ, then

$$\begin{aligned} d_H^2(P_1;P_2) &= \frac{1}{2}\int \left(f_1^{1/2} - f_2^{1/2}\right)^2 d\mu = 1 - \int \sqrt{f_1 f_2}\, d\mu\,, \\ d_\chi^2(P_1;P_2) &= \int (f_2/f_1 - 1)^2\, dP_1\,, \end{aligned}$$

In the definition of d_χ we presume that $\mathrm{supp}P_1 \supseteq \mathrm{supp}P_2$.

Definition 1. *We say the parametric family* $\mathcal{P} = \{P_\theta,\ \theta\in\Theta\}$, $\Theta\subset\mathcal{X}$, *obeys the regularity condition* $(R_{t,H})$ *if there exist* $\nu>0$ *and* $I_{t,H}>0$ *such that*

$$d_H^2(P_t;P_{t+h}) \sim I_{t,H}\|h\|^\nu \tag{$R_{t,H}$}$$

as $h \to 0$, $t\in\Theta$, $t+h\in\Theta$.

Family $\mathcal{P}$ *obeys the regularity condition* (R_H) *if there exist* $\nu>0$ *and function* $I_{\cdot,H}>0$ *such that* $(R_{t,H})$ *holds for every* $t\in\Theta$.

Definition 2. *We say family* $\mathcal{P}$ *obeys the regularity condition* $(R_{t,\chi})$ *if there exist* $\nu>0$ *and* $I_{t,\chi}>0$ *such that*

$$d_\chi^2(P_t;P_{t+h}) \sim I_{t,\chi}\|h\|^\nu \tag{$R_{t,\chi}$}$$

as $h \to 0$, $t\in\Theta, t+h\in\Theta$.

Family $\mathcal{P}$ *obeys the regularity condition* (R_χ) *if there exist* $\nu>0$ *and function* $I_{\cdot,\chi}>0$ *such that* $(R_{t,\chi})$ *holds for every* $t\in\Theta$.

Definitions 1 and 2 extend the notion of regularity of a parametric family of distributions.

A variant of these definitions has $\sim$ replaced with $\leq$.

We are not aware of natural examples where dependence of $d_H^2(P_t;P_{t+h})$ or $d_\chi^2(P_t;P_{t+h})$ on h is more complex. However, if such examples appear, then $(R_{t,H})$ and $(R_{t,\chi})$ can be generalised by replacing $\|h\|^\nu$ in the right-hand sides with $\psi(h)$ for a certain function ψ.

Definition 3. *If (R_H) or (R_χ) holds, then we call ν the "information index" and $I_{\cdot,H}$ and/or $I_{\cdot,\chi}$ the "information functions".*

It is known (see, e.g., [13] or [10], ch. 14) that

$$d_H^2 \leq d_{TV} \leq \sqrt{2} d_H \leq d_\chi. \tag{3}$$

If both (R_H) and (R_χ) are in force, then inequality $2d_H^2 \leq d_\chi^2$ entails

$$2I_{t,H} \leq I_{t,\chi}.$$

In Example 1 below $I_{t,\chi} = 2I_{t,H}$. In the case of a family $\{P_t = \mathcal{N}(t;1), t \in \mathbb{R}\}$ of normal random variables (r.v.s) one has

$$d_H^2(P_0;P_t) = 1 - e^{-t^2/8}, \; d_\chi^2(P_0;P_t) = e^{t^2} - 1,$$

hence $I_{t,\chi} = 8I_{t,H}$ (cf. [10], ch. 14.4).

Information index ν indicates how "rich" or "poor" the class $\mathcal{P}$ is. In the case of a regular parametric family of distributions (i.e., a family obeying (1)) one has

$$\nu = 2.$$

"Irregular" parametric families of distributions may obey (R_H) and (R_χ) with $\nu < 2$ (cf. Example 1 and [10], ch. 13).

Example 1. Let $\mathcal{P} = \{P_t, t > 0\}$, where $P_t = \mathbf{U}[0;t]$ is the uniform distribution on $[0;t]$. Then

$$\begin{aligned} d_H^2(P_{t+h};P_t) &= 1 - (1+|h|/t)^{-1/2} \sim h/2t && (t \geq h \searrow 0), \\ d_\chi^2(P_{t+h};P_t) &= h/t, \quad d_{TV}(P_{t+h};P_t) = h/(t+h) && (t \geq h > 0). \end{aligned}$$

Hence family $\mathcal{P}$ is not regular in the traditional sense (cf. (1)). Yet (R_H) and (R_χ) hold with

$$\nu = 1, \; I_{t,H} = 1/2t, \; I_{t,\chi} = 1/t.$$

The optimal estimator $t_n^* = \max\{X_1, ..., X_n\}(n+1)/n$ is unbiased, and

$$\mathbb{E}_t(t_n^* - t)^2 = t^2/n(n+2).$$

Parametric subfamilies of non-parametric classes typically obey (R_H) and (R_χ) with $\nu > 2$ (cf. Example 3 and [10], ch. 13).

We present now lower bounds to the accuracy of estimation when (R_H) or (R_χ) holds. Theorem 1 below indicates that the accuracy of estimation is determined by the information index and the information function.

Definition 4. *We say that set* Θ *obeys property* (A_ε) *if for every* $t \in \Theta$ *there exists* $t' \in \Theta$ *such that* $\|t' - t\| = \varepsilon$. *Property (A) holds if* (A_ε) *is in force for all small enough* $\varepsilon > 0$.

We say that estimator $\hat{\theta}$ *with a finite first moment has "regular" bias if for every* $t \in \Theta$ *there exists* $c_t > 0$ *such that*

$$\|\mathbb{E}_{t+h}\hat{\theta} - \mathbb{E}_t\hat{\theta}\| \sim c_t\|h\| \qquad (h \to 0). \tag{4}$$

An unbiased estimator obeys (4) with $c_t \equiv 1$. If Θ is an interval, then (A) trivially holds.

Theorem 1 [10] *Assume property (A), and suppose that estimator* $\hat{t}_n$ *obeys (4).*

If (R_χ) *holds with* $\nu \in (0;2)$, *then, as* $n \to \infty$,

$$\sup_{t\in\Theta}(nI_{t,\chi})^{2/\nu}\mathbb{E}_t\|\hat{t}_n - t\|^2/c_t^2 \gtrsim y_\nu^{2/\nu}/\,(e^{y_\nu} - 1)\,, \tag{5}$$

where y_ν *is the positive root of the equation* $2(1-e^{-y}) = \nu y$.

If the function $t \to \mathbb{E}_t\|\hat{t} - t\|^2$ *is continuous, then, as* $n \to \infty$,

$$(nI_{t,\chi})^{2/\nu}\mathbb{E}_t\|\hat{t}_n - t\|^2/c_t^2 \gtrsim y_\nu^{2/\nu}/\,(e^{y_\nu} - 1) \qquad (\forall t \in \Theta). \tag{5*}$$

If (R_χ) *holds with* $\nu > 2$, *then* $\mathbb{E}_t\|\hat{t}_n\|^2 = \infty$ $(\exists t \in \Theta)$.

The result holds with (R_χ) *replaced by* (R_H) *if* $I_{t,\chi}$ *is replaced with* $I_{t,H}$ *and the right-hand side of (5) is replaced with* $(\ln 4/3)^{2/\nu}/4$.

According to (5), the rate of the accuracy of estimation for estimators with regular bias cannot be better than $n^{-1/\nu}$. Moreover, (5) establishes that the natural normalizing sequence for $\hat{t}_n - t$ depends in a specific way on n, ν, and the information function.

Theorem 1 supplements the Fréchet–Rao–Cramér inequality that deals with the case $\nu = 2$. Note that (5*) formally extends to the case $\nu = 2$ with $y_2 := 0$ and the right-hand side of (5*) treated as $\lim_{y\to 0} y/\,(e^y - 1) = 1$.

According to Theorem 1, an estimator $\hat{t}_n$ cannot be unbiased or have a regular bias if (R_χ) or (R_H) holds with $\nu > 2$ and $\mathbb{E}_t\|\hat{t}_n\|^2 < \infty$ for every $t \in \Theta$.

Lower bounds involving continuity moduli are presented in the next section.

3 Lower bounds based on continuity moduli

We consider now a general situation where one cannot expect regularity conditions to hold (cf. Example 3).

Let $\mathcal{P}$ be an arbitrary class of probability distributions, and let the quantity of interest a_P be an element of a metric space $(\mathcal{X}, d)$. Given $\varepsilon > 0$, we denote by

$$\mathcal{P}_{H}(P,\varepsilon) = \{Q \in \mathcal{P} : d_{H}(P;Q) \le \varepsilon\}$$

the neighborhood of distribution $P \in \mathcal{P}$. We call

$$w_{H}(P,\varepsilon) = \sup_{Q\in\mathcal{P}_H(P,\varepsilon)} d(a_Q;a_P)/2 \quad \text{and} \quad w_{H}(\varepsilon) = \sup_{P\in\mathcal{P}} w_H(P,\varepsilon)$$

the moduli of continuity.

For instance, if $\mathcal{P} = \{P_t, t\in\Theta\}$, $a_{P_t} = t$ and $d(x;y) = |x-y|$, then

$$2w_{H}(P_t,\varepsilon) = \sup\{|h| : d_{H}(P_t;P_{t+h}) \le \varepsilon\}$$

and $w_{H}(\varepsilon) = \sup_t w_{H}(P_t,\varepsilon)$.

Similarly we define $\mathcal{P}_\chi(P,\varepsilon)$, $\mathcal{P}_{TV}(P,\varepsilon)$, $w_\chi(\cdot)$ and $w_{TV}(\cdot)$ using the $\chi^{\mathbf{2}}$-distance d_χ and the total variation distance d_{TV}. For instance, if $a_P \in \mathbb{R}$ and $d(x;y) = |x-y|$, then

$$w_{TV}(P,\varepsilon) = \sup_{Q\in\mathcal{P}_{TV}(P,\varepsilon)} |a_Q - a_P|/2.$$

The notion of continuity moduli has been available in the literature on non-parametric estimation for a while (cf. Donoho & Liu [3] and Pfanzagl [11, 12]). It helps to quantify the interplay between the degree of "richness" of class $\mathcal{P}$ and the accuracy of estimation.

Lemma 2 [10] *For any estimator* $\hat{a}$ *and every* $P_0 \in \mathcal{P}$,

$$\sup_{P\in\mathcal{P}_H(P_0,\varepsilon)} P(d(\hat{a}_n;a_P) \ge w_{H}(P_0,\varepsilon)) \ge (1-\varepsilon^2)^{2n}/4, \tag{6}$$

$$\sup_{P\in\mathcal{P}_\chi(P_0,\varepsilon)} P(d(\hat{a}_n;a_P) \ge w_{\chi}(P_0,\varepsilon)) \ge [1+(1+\varepsilon^2)^{n/2}]^{-2}\,. \tag{7}$$

Let R be a loss function. Lemma 2 and Chebyshev's inequality yield a lower bound to $\sup_{P\in\mathcal{P}_H(P_0,\varepsilon)} \mathbb{E}_P R(d(\hat{a}_n;a_P))$. For example, (6) with $R(x) = x^2$ yields

$$\sup_{P\in\mathcal{P}_H(P_0,\varepsilon)} \mathbb{E}_P^{1/2} d^2(\hat{a}_n;a_P) \ge w_{H}(P_0,\varepsilon)(1-\varepsilon^2)^n/2\,. \tag{8}$$

An (8)-type result for asymptotically unbiased estimators has been presented by Pfanzagl [12]. Note that Lemma 2 does not impose any extra assumptions.

The best possible rate of estimation can be found by maximizing the right-hand side of (8) in ε. For instance, if

$$w_{_H}(P,\varepsilon) \gtrsim J_{_{H,P}}\varepsilon^{2r} \qquad (\varepsilon\to 0) \tag{9}$$

for some $J_{_{H,P}}>0$, then the rate of the accuracy of estimation cannot be better than n^{-r}.

If $(R_{_H})$ and/or $(R_{_\chi})$ hold for a parametric subfamily of $\mathcal{P}$, then

$$2w_{_H}(P_t,\varepsilon) \sim (\varepsilon^2/I_{t,H})^{1/\nu} \quad \text{and/or} \quad 2w_{_\chi}(P_t,\varepsilon) \sim (\varepsilon^2/I_{t,\chi})^{1/\nu}, \tag{10}$$

yielding (9) with $r = 1/\nu$. Hence the best possible rate of the accuracy of estimation is $n^{-1/\nu}$.

The drawback of this approach is the difficulty of calculating the continuity moduli.

Example 2. Consider the parametric family $\mathcal{P}$ of distributions P_θ with densities

$$f_\theta(x) = \varphi(x-\theta)/2 + \varphi(x+\theta)/2 \qquad (\theta\in \mathbb{R}),$$

where φ is the standard normal density. Set

$$a_{P_\theta} = \theta,\ d(\theta_1;\theta_2) = |\theta_1 - \theta_2|.$$

Then

$$d_{_H}(P_{\rm o};P_h) \sim h^2/4\,.$$

Thus, $(R_{_{0,H}})$ holds with

$$\nu = 4,\ I_{{\rm o},H} = 1/16,$$

$w_{_H}(P_{\rm o},\varepsilon) \sim \sqrt{\varepsilon}$ as $\varepsilon \to 0$; there is no asymptotically unbiased with the optimal rate finite-variance estimator; the rate of the accuracy of estimation in a neighborhood of the standard normal distribution $P_{\rm o}$ cannot be better than $n^{-1/4}$ (cf. Liu and Brown [6]). An application of (13.8) in [10] yields

$$\sup_{0\le\theta\le\varepsilon} \mathbb{E}_{P_\theta}|\hat\theta_n - \theta|^2 \gtrsim 1/2\sqrt{en} \qquad (n\to\infty) \tag{11}$$

for an arbitrary estimator $\hat\theta_n$ and any $\varepsilon > 0$.

Put $\varepsilon^2 = c^2/n$ in (8). Then

$$\sup_{P\in\mathcal{P}_H(P_{\rm o},\varepsilon)} \mathbb{E}_P^{1/2} d(\hat a_n;a_P)^2 \gtrsim e^{-c^2} w_{_H}(P_{\rm o},c/\sqrt{n}\,)/2\,. \tag{8*}$$

Thus, the rate of the accuracy of estimation of a_P in a neighborhood of $P_{\rm o}$ cannot be better than that of

$$w_{_H}(P_{\rm o},1/\sqrt{n}\,)$$

(cf. Donoho & Liu [3]). More specifically, if (9) holds, then

$$\sup_{P\in\mathcal{P}_H(P_o,\varepsilon)} \mathbb{E}_P^{1/2} d^2(\hat{a}_n; a_P) \gtrsim e^{-c^2} J_{H,P_o} c^{2r} n^{-r}/2\,. \tag{12}$$

If $J_{H,\cdot}$ is uniformly continuous on $\mathcal{P}$, then (12) with $c^2 = r$ yields the *non-uniform* lower bound

$$\sup_{P\in\mathcal{P}} J_{H,P}^{-1} \mathbb{E}_P^{1/2} d^2(\hat{a}_n; a_P) \gtrsim (r/e)^r n^{-r}/2\,. \tag{13}$$

Lower bound (13) is non-uniform because of the presence of the term depending on P in the left-hand side of (13). Note that the traditional approach would be to deal with $\sup_{P\in\mathcal{P}} \mathbb{E}_P\, d^2(\hat{a}_n; a_P)$ (cf. [13]); the latter can in some cases be meaningless while $\sup_{P\in\mathcal{P}} J_{H,P}^{-1} \mathbb{E}_P\, d^2(\hat{a}_n; a_P)$ is finite (cf. (14)).

Example 1 (continued). Let $a_{P_t} = t$, $d(t;s) = |t-s|$. Then

$$w_H(P_t, \varepsilon) = t\varepsilon^2(1-\varepsilon^2/2)/(1-\varepsilon^2)^2 \ge t\varepsilon^2\,,$$

and (9) holds with $r = 1, J_{H,P_t} = t$. According to (8) with $\varepsilon^2 = 1/n$,

$$\sup_{P_s\in\mathcal{P}_H(P_t,\varepsilon)} \mathbb{E}_s^{1/2}|\hat{t}_n - s|^2 \ge t/2en$$

for any estimator $\hat{t}_n$. Hence $\sup_{t>0} \mathbb{E}_t|\hat{t}_n - t|^2 = \infty$, while the non-uniform bound is

$$\sup_{t>0} \mathbb{E}_t^{1/2}|\hat{t}_n/t - 1|^2 \ge 1/2en\,(1+2/n)\,. \tag{14}$$

Remark. In typical non-parametric situations the rate of the accuracy of estimation is worse than $n^{-1/2}$. However, an interesting fact is that if we choose $a_P = P$ and $d = d_H$, then $w_H(P,\varepsilon) = \varepsilon/2$ for all P, (9) holds with $r=1/2$, $J_{H,P} = 1/2$, hence

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P\, d_H^2(\hat{a}_n; a_P) \gtrsim 1/32en. \tag{15}$$

4 On unbiased estimation

It is not difficult to notice that in most estimation problems concerning non-parametric classes of distributions the available estimators are biased. The topic was studied by a number of authors (see Pfanzagl [12] and references therein). Examples include non-parametric density, regression curve, hazard function (failure rate) and tail index estimation.

We notice in [7] that the sample autocorrelation is a non-negatively biased estimator of the autocorrelation function (the bias is positive unless the distribution of the sample elements is symmetric).

Theorem 1 suggests a way of showing that there are no unbiased finite-variance estimators for a given class $\mathcal{P}$ of distributions if the class contains a parametric family of distributions obeying the regularity condition $(R_{_H})$ or $(R_{_\chi})$ with $\nu > 2$.

Example 3. Let $\mathcal{P}_b$, where $b > 0$, be the class of distributions P such that

$$\sup_{0<x\le 1} \left| x^{-\alpha_P} P(X<x) - 1 \right| x^{-b\alpha_P} < \infty \qquad (\exists \alpha_{_P} > 0)$$

(the Hall class). Note that $F(x) \equiv P(X < x) = x^\alpha(1+O(x^{b\alpha}))$ as $x \to 0$ if $P \in \mathcal{P}_b$. We consider the problem of estimating index $\alpha \equiv \alpha_{_P}$ from a sample of independent observations when the unknown distribution belongs to $\mathcal{P}_b$.

Let $P_{\alpha,0}$ and $P_{\alpha,\gamma}$ be the distributions with distribution functions (d.f.s)

$$\begin{aligned} F_{\alpha,0}(y) &= y^\alpha \mathbb{I}\{0 < y \le 1\}, \\ F_{\alpha,\gamma}(y) &= \delta^{-\gamma} y^{\alpha+\gamma} \mathbb{I}\{0<y\le\delta\} + y^\alpha \mathbb{I}\{\delta<y\le 1\}, \end{aligned}$$

where $\delta = \gamma^{1/b\alpha}$, $\gamma \in (0;1)$. One can check that

$$d_{_H}^2(P_{\alpha,0};P_{\alpha,\gamma}) = \gamma^{1/b}\Big[1 - \sqrt{1+\gamma/\alpha}\big/(1+\gamma/2\alpha)\Big] \le \gamma^{1/r}/8\alpha^2\,, \tag{16}$$

$$d_{_\chi}^2(P_{\alpha,0};P_{\alpha,\gamma}) = \gamma^{1/r}\alpha^{-2}(1+\gamma/2\alpha)^{-1} \le \gamma^{1/r}/\alpha^2\,, \tag{17}$$

where $r = b/(1+2b)$. Thus, $(R_{t,H})$ and $(R_{t,\chi})$ hold with $\nu = 2+1/b$.

According to Theorem 1, there are no unbiased finite-variance estimators of index α.

Note that

$$d_{_H}^2(P_{\alpha,0};P_{\alpha,h}) \sim h^{2+1/b}/8\alpha^2 \qquad (h \to 0)$$

for the parametric family $\{P_{\alpha,h}, 0 \le h < 1\} \subset \mathcal{P}_b$, while

$$d_{_H}^2(P_{\alpha,0};P_{\alpha+h,0}) \sim h^2/8\alpha^2 \qquad (h \to 0)$$

for the parametric family $\{P_{\alpha+h,0}, 0 \le h < 1\} \subset \mathcal{P}_b$ (cf. [10], p. 293).

The next theorem shows that in typical non-parametric situations there are no asymptotically unbiased with the rate estimators.

Let $\{\mathcal{P}_n, n \ge 1\}$ be a non-increasing sequence of neighborhoods of a particular distribution P_0, and let $\{z_n\}$ be a sequence of positive numbers. Pfanzagl [12] calls estimator $\{\hat{a}_n\}$ *asymptotically unbiased* uniformly in $\mathcal{P}_n$ with the rate $\{z_n\}$ if

$$\limsup_{u\to\infty} \limsup_{n\to\infty} \sup_{P\in\mathcal{P}_n} |\mathbb{E}_P K_u((\hat{a}_n - a_P)/z_n)| = 0,$$

where $K_u(x) = x\mathbb{I}\{|x| \le u\}$.

Denote $\mathcal{P}_{n,\varepsilon} = \mathcal{P}_{_\chi}(P_0, \varepsilon/\sqrt{n})$, where $\varepsilon > 0$.

Theorem 3 [12] *Suppose that*

$$\limsup_{\varepsilon\to 0} \varepsilon^{-1} \liminf_{n\to\infty} w_\chi(P_0, \varepsilon/\sqrt{n})/z_n = \infty, \tag{18}$$

$$\lim_{u\to\infty} \liminf_{n\to\infty} P_0(|\hat{a}_n - a_{P_0}|/z_n \le u) > 0, \tag{19}$$

$$\lim_{u\to\infty} \liminf_{n\to\infty} \mathbb{E}_{P_0} K_u^2((\hat{a}_n - a_{P_0})/z_n) < \infty. \tag{20}$$

Then estimator $\{\hat{a}_n\}$ *cannot be asymptotically unbiased with the rate* $\{z_n\}$ *uniformly in* $\mathcal{P}_{n,\varepsilon}$ *for some* $\varepsilon > 0$.

Pfanzagl [12] showed that in a number of particular non-parametric estimation problems

$$\inf_{\varepsilon>0} \varepsilon^{-c} \liminf_{n\to\infty} w_\chi(P_0, \varepsilon/\sqrt{n}\,)/z_n > 0 \qquad (\exists\ c \in (0;1)) \tag{21}$$

(cf. (10)). Note that (21) entails (18).

5 On consistent estimation

The rate of the accuracy of estimation can be poor if the class $\mathcal{P}$ of distributions is "rich". In utmost cases the lower bound is bounded away from zero meaning neither estimator is consistent uniformly in $\mathcal{P}$. We present below few such examples.

Example 4. Let $\mathcal{F}$ be a class of distributions with absolutely continuous distribution functions on $\mathbb{R}$ such that $\int |f(x+y) - f(x)|dx \le |y|$. Ibragimov & Khasminskiy [5] have shown that

$$\sup_{f\in\mathcal{F}} \mathbb{E}_f \int |\hat{f}_n - f| \ge 2^{-9} \qquad (n \ge 1)$$

for any estimator $\hat{f}_n$ of density f (see Devroye [2] for a related result).

Example 5. Consider the problem of non-parametric regression curve estimation. Given a sample of i.i.d. pairs $(X_1, Y_1), \dots, (X_n, Y_n)$, one wants to estimate the regression function

$$\psi(x) = \mathbb{E}\{Y|X = x\}.$$

There is no uniformly consistent estimator if the only assumption about $\mathcal{L}(X, Y)$ is that function ψ is continuous.

Let $\mathcal{P}$ be a class of distributions of random pairs (X, Y) taking values in $\mathbb{R}^2$ such that function $\psi(\cdot) = \mathbb{E}\{Y|X = \cdot\}$ is continuous. Set

$$f_0(x,y) = 1\!\!\mathrm{I}\{|x| \vee |y| \le 1/2\}\,, \;\; f_1(x,y) = f_0(x,y) + hg(xh^{-c})g(y)\,,$$

where $c > 0$, $h \in (0;1)$ and $g(x) = \sin(2\pi x) 1\!\!\mathrm{I}\{|x| \le 1/2\}$. These are the densities of two distributions of a random pair (X, Y).

Let ψ_k , $k \in \{0;1\}$, denote the corresponding regression curves. Then

$$\psi_0 \equiv 0\,, \;\; \psi_1(x) = 2\pi^{-2}h\sin(2\pi h^{-c}x)\mathbb{1}\{|x| \le h^c/2\}\,.$$

Hence $\|\psi_0 - \psi_1\| = 2\pi^{-2}h$.

Note that $d_\chi^2(f_0;f_1) \le h^{2+c}/4$. Applying Lemma 13.1 [10], we derive

$$\max_{i\in\{0,1\}} \mathbb{P}_i\left(\|\hat\psi_n - \psi\| \ge h/\pi^2\right) \ge (1+d_\chi^2)^{-n}/4 \ge \exp(-nh^{2+c}/4)/4$$

for any regression curve estimator $\hat\psi_n$. With $c = n-2$ and $h = n^{-1/n}$, we get

$$\sup_{P\in\mathcal{P}} \mathbb{P}\left(\|\hat\psi_n - \psi\| \ge 1/9\right) \ge 1/4e^{1/4}\,. \tag{22}$$

Hence no estimator is consistent uniformly in $\mathcal{P}$.

Example 6. Consider the problem of non-parametric estimation of the distribution function of the sample maximum. No uniformly consistent estimator exists in a general situation. Indeed, it is shown in [8, 9] that for any estimator $\{\hat F_n\}$ of the distribution function of the sample maximum there exist a d.f. F such that

$$\limsup_{n\to\infty} \mathbb{P}_F\left(\|\hat F_n - F^n\| \ge 1/9\right) \ge 1/3\,.$$

Moreover, one can construct d.f.s F_0 and F_1 such that

$$\max_{i\in\{0;1\}} \mathbb{P}_{F_i}\left(\|\hat F_n - F_i^n\| \ge 1/4\right) \ge 1/4 \qquad (n \ge 1),$$

where F_0 is uniform on $[0;1]$ and $F_1 \equiv F_{1,n} \to F_0$ everywhere as $n \to \infty$.

An estimator $\tilde a_n(\cdot) \equiv \tilde a_n(\cdot, X_1, ..., X_n)$ is called *shift-invariant* if

$$\tilde a_n(x, X_1, ..., X_n) = \tilde a_n(x+c, X_1+c, ..., X_n+c)$$

for every $x \in \mathbb{R}$, $c \in \mathbb{R}$. An estimator $\tilde a_n(\cdot)$ is called *scale-invariant* if

$$\tilde a_n(x, x_1, ..., x_n) = \tilde a_n(cx, cx_1, ..., cx_n) \qquad (\forall c > 0)$$

for all $x, x_1, ..., x_n$.

Examples of shift- and scale-invariant estimators of F^n include F_n^n, where F_n is the empirical distribution function, and the "blocks" estimator

$$\tilde F_n = \Big(\sum_{i=1}^{[n/r]} \mathbb{1}\{M_{i,r} < x\}/[n/r]\Big)^n,$$

where $M_{i,r} = \max\{X_{(i-1)r+1}, ..., X_{ir}\}$ $(1 \le r \le n)$.

For any shift- or scale-invariant estimator $\{\tilde F_n\}$ of the distribution function of the sample maximum there holds

$$\mathbb{P}_{F_0}\Big(\|\tilde F_n - F_0^n\| \ge 1/4\Big) \ge 1/4 \qquad (n \ge 1). \tag{23}$$

Thus, consistent estimation of the distribution function of the sample maximum is only possible under certain assumptions on the class of unknown distributions.

6 On uniform convergence

We saw that the rate of the accuracy of estimation cannot be better than $w_H(P, 1/\sqrt{n}\,)$. According to Donoho & Liu [3], if a_P is linear and class $\mathcal{P}$ of distributions is convex, then there exists an estimator $\hat{a}_n$ attaining this rate.

We show now that in typical non-parametric situations neither estimator converges locally uniformly with the optimal rate.

Definition 5. *Let $\mathcal{P}'$ be a subclass of $\mathcal{P}$. We say that estimator $\hat{a}_n$ converges to a_P with the rate z_n uniformly in $\mathcal{P}'$ if there exists a non-defective distribution P^* such that*

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}'} |P\left((\hat{a}_n - a_P)/z_n \in A\right) - P^*(A)| = 0 \tag{24}$$

for every measurable set A with $P^*(\partial A) = 0$.

Note that for every $P \in \mathcal{P}'$ (24) yields the weak convergence $(\hat{a}_n - a_P)/v_n \Rightarrow P^*$.

The following result on impossibility of locally uniform convergence with the optimal rate is due to Pfanzagl [11]. It involves a continuity modulus based on the total variation distance.

Let $\mathcal{X} = \mathbb{R}$. Denote $\mathcal{P}_{TV}^{(n)}(P_0, \varepsilon) = \{P\in\mathcal{P} : d_{TV}(P^n; P_0^n) \le \varepsilon\}$, and recall that

$$w_{TV}^{(n)}(P_0, \varepsilon) = \sup_{P\in\mathcal{P}_{TV}^{(n)}(P_0,\varepsilon)} |a_P - a_{P_0}|/2.$$

Theorem 4 [11] *Suppose that*

$$\lim_{\varepsilon\downarrow 0} \varepsilon^{-1} \limsup_{n\to\infty} w_{TV}^{(n)}(P_0, \varepsilon)/z_n = \infty. \tag{25}$$

Then neither estimator can converge to a_P with the rate z_n uniformly in $\mathcal{P}_{TV}^{(n)}(P_0, \varepsilon)$ for some $\varepsilon \in (0; 1)$.

Example 7. Let $\mathcal{P}_b^+$, where $b > 0$, be the non-parametric class of distributions on $(0; 1]$ with densities

$$f(x) = C_{\alpha,b} x^{\alpha-1}(1+r(x)),$$

where $\sup_{0<x\le 1} |r(x)|x^{-\alpha b} < \infty$. We consider the problem of estimating index α.

Denote $r = b/(1+2b)$. Pfanzagl [11] showed that

$$\varepsilon^{-2r} \liminf_{n\to\infty} n^r w_{TV}^{(n)}(P_0, \varepsilon) > 0 \qquad (\forall \varepsilon \in (0; 1)). \tag{26}$$

Since $r < 1/2$, (25) and (26) entail that neither estimator of index α can converge to α uniformly in $\mathcal{P}^{(n)}_{TV}(P_0, \varepsilon)$ with the rate $z_n = n^{-b/(1+2b)}$.

The next theorem presents a result on impossibility of locally uniform convergence with the optimal rate involving the modulus of continuity w_H based on the Hellinger distance. The Hellinger distance may be preferable to the total variation distance in identifying the optimal rate of the accuracy of estimation as there are cases where

$$d_{TV}(P_0; P_1) \gg d^2_H(P_0; P_1)$$

for "close" distributions P_0 and P_1. For instance, consider family $\mathcal{P} = \{P_{\alpha,\gamma}\}_{\gamma \geq 0}$, where distributions $\{P_{\alpha,\gamma}\}$ have been defined in Example 3. Then

$$d_{TV}(P_{\alpha,o}; P_{\alpha,\gamma}) \sim \frac{\gamma^{1/r-1}}{\alpha e} \gg d^2_H(P_{\alpha,o}; P_{\alpha,\gamma}) \sim \frac{\gamma^{1/r}}{8\alpha^2} \qquad (\gamma \to 0).$$

Theorem 5 *If (9) holds for a particular* $P \in \mathcal{P}$ *with* $r < 1/2$, *then neither estimator converges to* a_P *with the rate* n^{-r} *uniformly in* $\mathcal{P}_H(P, 1/\sqrt{n}\,)$.

Theorem 5 generalises Theorem 13.9 in [10] by relaxing the assumption that there exists a positive continuous derivative of distribution P^* with respect to the Lebesgue measure.

Acknowledgments. The author is grateful to the anonymous reviewer for helpful comments.

References

[1] Barankin E. W. (1949) Locally best unbiased estimates. Ann. Math. Statist., v. 20, 477–501.

[2] Devroye L. (1995) Another proof of a slow convergence result of Birgé. — Statist. Probab. Letters, v. 23, No 1, 63–67.

[3] Donoho D.L. & Liu R.C. (1991) Geometrizing rates of convergence II, III. — Ann. Statist., v. 19, No 2, 633–667, 668–701.

[4] Huber C. (1997) Lower bounds for function estimation. — In: Festschrift for L. Le Cam, 245–258. Springer, New York.

[5] Ibragimov I.A. and Khasminskii R.Z. (1980) Estimation of distribution density. — Zap. Nauch. Sem. LOMI, v. 98, 61–85.

[6] Liu R.C. and Brown L.D. (1993) Nonexistence of informative unbiased estimators in singular problems. — Ann. Statist., v. 21, No 1, 1–13.

[7] Novak S.Y. (2006) A new characterization of the normal law. — Statist. Probab. Letters, v. 77, No 1, 95–98.

[8] Novak S.Y. (2010) Lower bounds to the accuracy of sample maximum estimation. — Theory Stochast. Processes, v. 15(31), No 2, 156–161.

[9] Novak S.Y. (2010) Impossibility of consistent estimation of the distribution function of a sample maximum. — Statistics, v. 44, No 1, 25–30.

[10] Novak S.Y. (2011) *Extreme value methods with applications to finance.* — London: Taylor & Francis/CRC. ISBN: 978-1-43983-574-6.

[11] Pfanzagl J. (2000) On local uniformity for estimators and confidence limits. — J. Statist. Plann. Inference, v. 84, 27–53.

[12] Pfanzagl J. (2001) A nonparametric asymptotic version of the Cramér-Rao bound. — In: State of the art in probability and statistics. Inst. Math. Statist. Lecture Notes Monogr. Ser., v. 36, 499–517. Beachwood, OH.

[13] Tsybakov A.B. (2009) *Introduction to nonparametric estimation.* — Springer: New York.