

# Middlesex University Research Repository

An open access repository of

Middlesex University research

<http://eprints.mdx.ac.uk>

Novak, Serguei (2018) Non-parametric lower bounds and information functions. In: ISNPS-Third Conference of the International Society for Nonparametric Statistics (ISNPS), 11-16 June 2016, Avignon, France.

First submitted uncorrected version (with author's formatting)

This version is available at: <http://eprints.mdx.ac.uk/20799/>

## Copyright:

Middlesex University Research Repository makes the University's research available electronically.

Copyright and moral rights to this work are retained by the author and/or other copyright owners unless otherwise stated. The work is supplied on the understanding that any use for commercial gain is strictly forbidden. A copy may be downloaded for personal, non-commercial, research or study without prior permission and without charge.

Works, including theses and research projects, may not be reproduced in any format or medium, or extensive quotations taken from them, or their content changed in any way, without first obtaining permission in writing from the copyright holder(s). They may not be sold or exploited commercially in any format or medium without the prior written permission of the copyright holder(s).

Full bibliographic details must be given when referring to, or quoting from full items including the author's name, the title of the work, publication details where relevant (place, publisher, date), pagination, and for theses or dissertations the awarding institution, the degree type awarded, and the date of the award.

If you believe that any material held in the repository infringes copyright law, please contact the Repository Team at Middlesex University via the following email address:

[eprints@mdx.ac.uk](mailto:eprints@mdx.ac.uk)

The item will be removed from the repository while any claim is being investigated.

See also repository copyright: re-use policy: <http://eprints.mdx.ac.uk/policies.html#copy>

# Non-parametric lower bounds and unbiased estimators

S.Y.Novak

MDX University London, SST, The Burroughs NW44BT, UK

June 2016

## Abstract

We introduce the notion of the *information index* and present a non-parametric generalisation of the Rao–Cramér inequality.

We show that unbiased estimators do not exist if the information index is larger than two.

For a typical non-parametric class  $\mathcal{P}$  of distributions **neither** estimator is asymptotically normal with the optimal rate *uniformly* over  $\mathcal{P}$ .

*Key words:* non-parametric lower bounds, information index, information function, uniform convergence.

## 1 Introduction

Typical estimation problem: given a sample  $X_1, \dots, X_n$  of i.i.d. observations from an unknown distribution  $P \in \mathcal{P}$ , estimate a **quantity of interest**  $a_P$ .

Hellinger distance:  $d_H^2$ ,  $\chi^2$  distances:  $d_\chi^2$ .

A typical **regularity condition**:

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

as  $h \rightarrow 0$  for every  $\theta \in \Theta, \theta+h \in \Theta$ , where  $I_\theta$  is ‘‘Fisher’s information’’.

If (1) holds and estimator  $\hat{\theta}_n$  is unbiased, then

$$\sup_{\theta \in \Theta} I_\theta \mathbb{E}_\theta \|\hat{\theta}_n - \theta\|^2 \geq 1/n. \quad (2)$$

This is the celebrated **Fréchet–Rao–Cramér inequality**.

If unbiased estimators with a finite second moments exist, then the optimal unbiased estimator is the one that turns a lower bound into equality.

Barankin [1]: a parametric estimation problem where **NO** unbiased estimator with  $\mathbb{E}_\theta \|\hat{\theta}_n - \theta\|^2 < \infty$ .

We argue: in typical **non-parametric situations** – **NO** unbiased estimators with a finite 2nd moment.

## 2 Information index

We extend the notion of regularity of a parametric family  $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$  of distributions.

**Definition.** *Parametric family  $\mathcal{P}$  obeys the **regularity condition**  $(R_H)$  if there exists number  $\nu > 0$*

and function  $I_{\cdot,H} > 0$  such that as  $h \rightarrow 0$ ,

$$d_H^2(P_t; P_{t+h}) \sim I_{t,H} \|h\|^\nu \quad (t \in \Theta, t+h \in \Theta). \quad (R_H)$$

Similarly we define  $(R_\chi)$ –regular parametric family.

We call  $\nu$  the “*information*” index.

We call  $I_{\cdot,H}$  the “*information*” function.

Information index  $\nu$  indicates how “rich” or “poor” the class  $\mathcal{P}$  is.

**Regular** parametric family of distributions:  $\nu = 2$ .

$(R_H)$ –**regular** parametric families:  $\nu < 2$ .

**Non–parametric** classes:  $(R_H)$  with  $\nu > 2$ .

**Example 1.** Let  $P_t = \mathbf{U}[0; t]$ ,  $\mathcal{P} = \{P_t, t > 0\}$ . Then

$$d_H^2(P_{t+h}; P_t) \sim h/2t \quad (t \geq h \searrow 0).$$

Family  $\mathcal{P}$  is not regular in the traditional sense (cf. (1)).

Yet  $(R_H)$  holds with

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

*Non–uniform lower bound:* for any estimator  $\hat{t}_n$

$$\sup_{t>0} t^{-1} \mathbb{E}_t^{1/2}(\hat{t}_n - t)^2 \geq 0.8/(n-1.6) \quad (3)$$

as  $n \geq 2$ , while the *uniform* bound is

$$\sup_t \mathbb{E}_t^{1/2}(\hat{t}_n - t)^2 = \infty.$$

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

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

Lower bound indicates: the accuracy of estimation is determined by the *information index* and the *information function*.

Any *unbiased estimators* with finite second moment if  $(R_H)$  holds with  $\nu > 2$ ?

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

Estimator  $\hat{\theta}$  has “regular” bias if for every  $t \in \Theta$  there exists  $c_t > 0$  such that

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

We write  $a_n \gtrsim b_n$  if  $a_n \geq b_n(1+o(1))$  as  $n \rightarrow \infty$ .

**Theorem 1** Assume  $(R_\chi)$  and  $(A)$ , and suppose that estimator  $\hat{t}_n$  has “regular” bias [obeys (4)].

If  $\nu \in (0; 2)$ , then

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

as  $n \rightarrow \infty$ , where  $y_\nu$  is the positive root of the equation  $\nu y = 2(1 - e^{-y})$ .

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

Thus, if  $\nu \in (0; 2)$ , then the accuracy of estimation for regular-bias estimators is  $n^{-1/\nu}$ .

**Example 2.** Parametric family  $\mathcal{P}$  with densities

$$f_\theta(x) = \varphi(x - \theta)/2 + \varphi(x + \theta)/2,$$

where  $\varphi$  is the standard normal density;  $a_{P_\theta} = \theta$ ,

$$d_H(P_0; P_h) \sim h^2/4.$$

Thus,  $(R_H)$  holds with

$$\nu = 4, I_{t, H} = 1/16;$$

the accuracy of estimation cannot be better than  $n^{-1/4}$ .

*General problem:* estimate a quantity of interest  $a_P$ .

**Corollary 2** *If  $(R_H)$  or  $(R_\chi)$  holds with  $\nu > 2$  and  $\sup_{P \in \mathcal{P}} \mathbb{E}_P \|\hat{a}_n - a_P\|^2 < \infty$ , then estimator  $\hat{a}_n$  is **biased**.*

### 3 Continuity moduli

Let  $a_P$  be an element of a metric space  $(\mathcal{X}, d)$ . For any  $\varepsilon > 0$  we denote by

$$\mathcal{P}_H(P, \varepsilon) = \{Q \in \mathcal{P} : d_H(P; Q) \leq \varepsilon\}$$

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

$$\begin{aligned} w_H(P, \varepsilon) &= \sup_{Q \in \mathcal{P}_H(P, \varepsilon)} d(a_P; a_Q)/2, \\ w_H(\varepsilon) &= \sup_{P \in \mathcal{P}} w_H(P, \varepsilon) \end{aligned}$$

the *moduli of continuity* of  $\{a_P : P \in \mathcal{P}\}$ .

Similarly we define  $\mathcal{P}_\chi(P, \varepsilon)$ ,  $\mathcal{P}_{TV}(P, \varepsilon)$ ,  $w_\chi(\cdot)$ ,  $w_{TV}(\cdot)$ .

Continuity moduli describe how the ‘‘closeness’’ of  $a_Q$  to  $a_P$  reflects the ‘‘closeness’’ of  $Q$  to  $P$ .

The ‘‘richer’’ class  $\mathcal{P}$ , the poorer the accuracy of estimation.

**Lemma 3** *Assume that for any  $c > 0$  there exists  $C \in (0; \infty)$  such that  $w.(c\varepsilon) \leq Cw.(\varepsilon)$ . For any estimator  $\hat{a}_n$  and every  $P_0 \in \mathcal{P}$ ,*

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

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

For example, (6) and Chebyshev’s inequality yield

$$\sup_{P \in \mathcal{P}_H(P_0, \varepsilon)} \mathbb{E}_P d(\hat{a}_n; a_P) \geq w_H(P_0, \varepsilon)(1 - \varepsilon^2)^n/2. \quad (7)$$

Maximize  $w_H(P, \varepsilon)(1 - \varepsilon^2)^n$  in  $\varepsilon$ .

If for some  $J_{H,P} > 0$

$$w_H(P, \varepsilon) \gtrsim J_{H,P} \varepsilon^{2r} \quad (\exists P \in \mathcal{P}) \quad (8)$$

then the rate of estimation cannot be better than  $n^{-r}$ .

If  $(R_H)$  holds for a parametric subfamily of  $\mathcal{P}$ , then

$$2w_H(P_t, \varepsilon) \sim (\varepsilon^2 / I_{t,H})^{1/\nu} \quad (9)$$

If  $(R_\chi)$  holds, then

$$2w_\chi(P_t, \varepsilon) \sim (\varepsilon^2 / I_{t,\chi})^{1/\nu}.$$

Thus,  $(R_H)$  and/or  $(R_\chi)$  yield (8) with

$$r = 1/\nu;$$

the accuracy of estimation cannot be better than  $n^{-1/\nu}$ .

If (8) holds for all small enough  $\varepsilon$  and  $J_{H,\cdot}$  is uniformly continuous on  $\mathcal{P}$ , then

$$\sup_{P \in \mathcal{P}} J_{H,P}^{-1} \mathbb{E}_P^{1/2} d(\hat{a}_n; a_P)^2 \gtrsim (r/e)^r n^{-r} / 2. \quad (10)$$

Calculating *continuity moduli* is not easy.

**Example 3.** Let  $\mathcal{P} = \{P_t, t \in \mathbb{R}\}$ , where  $P_t = \mathcal{N}(t; 1)$ , and let  $a_{P_t} = t$  and  $d(t; s) = |t - s|$ . Then

$$w_H(P_t, \varepsilon) = \sqrt{\ln(1 - \varepsilon^2)^{-2}} \geq \sqrt{2} \varepsilon$$



for every  $t$ . Hence (8) and (10) hold with  $J_{H,P} = \sqrt{2}$  and  $r = 1/2$ .

#### 4 Uniform convergence

The rate of the accuracy of estimation cannot be better than  $w_H(P, 1/\sqrt{n})$ . If  $a_P$  is linear and class  $\mathcal{P}$  of distributions is convex, then there exists an estimator  $\hat{a}_n$  attaining this rate [2].

In typical **non-parametric** situations **neither** estimator converges *locally uniformly* with the optimal rate.

More information: [2, 3, 4].

Let  $\mathcal{P}'$  be a subclass of  $\mathcal{P}$ . Estimator  $\hat{a}_n$  converges weakly to  $a_P$  with the rate  $v_n$  *uniformly* in  $\mathcal{P}'$  if there exists a non-degenerate distribution  $P_0$  such that

$$\lim_{n \rightarrow \infty} \sup_{P \in \mathcal{P}'} |P((\hat{a}_n - a_P)/v_n \in A) - P_0(A)| = 0 \quad (11)$$

for every measurable set  $A \subset \mathcal{X}$  with  $P_0(\partial A) = 0$ .

**Theorem 4** *Assume that  $\mathcal{X} = \mathbb{R}$ , and let  $P \in \mathcal{P}$ . If  $w_H(P, \varepsilon) \sim J_{H,P} \varepsilon^{2r}$ , where  $r < 1/2$ , and*

$$\sup_{P_* \in \mathcal{P}_H(P, 1/\sqrt{n})} |J_{H,P_*}/J_{H,P} - 1| \rightarrow 0$$

*as  $n \rightarrow \infty$ , then **neither** estimator converges to  $a_P$  with the rate  $n^{-r}$  uniformly in  $\mathcal{P}_H(P, 1/\sqrt{n})$ .*

## References

- [1] Barankin E. W. (1949) Locally best unbiased estimates. *Ann. Math. Statist.*, v. 20, 477–501.
- [2] Donoho D.L. & Liu R.C. (1991) Geometrizing rates of convergence II, III. — *Ann. Statist.*, v. 19, No 2, 633–667, 668–701.
- [3] Novak S.Y. (2011). *Extreme value methods with applications to finance*. — London: CRC. ISBN: 978-1-43983-574-6.
- [4] Pfanzagl J. (2000) On local uniformity for estimators and confidence limits. — *J. Statist. Plann. Inference*, v. 84, 27–53.
- [5] Pfanzagl J. (2001) A nonparametric asymptotic version of the Cramér-Rao bound. — *State of the art in probability and statistics (Leiden, 1999)*, 499–517, *IMS Lecture Notes Monogr. Ser.*, v. 36, Inst. Math. Statist., Beachwood, OH.