Statistics and Probability Letters 138 (2018) 183–189

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Lower bounds in estimation at a point under multi-index constraint☆

Nora Serdyukova

University of Concepción, Chile

ARTICLE INFO

Article history:
Received 13 July 2017
Received in revised form 18 February 2018
Accepted 24 February 2018
Available online 10 March 2018

MSC:
62G05
62G08
62G20

Keywords:
Nonparametric function estimation
Multi-index model
Minimax

ABSTRACT

In the framework of nonparametric anisotropic multivariate function estimation we study the lower bounds for the minimax risk. It is shown that the multi-index assumption leads to new minimax lower bounds containing a logarithmic factor.

1. Introduction

The lower bounds for the minimax risk, apart from being a challenging mathematical problem, serve as a benchmark for the best obtainable quality of an arbitrary estimator.

Let $\mathcal{D} = [-1, 1]^d$. We observe a path $\{Y_n(x),\ x \in \mathcal{D}\}$ satisfying the stochastic differential equation

$$Y_n(\mathrm{d}x) = F(x)\mathrm{d}x + \frac{1}{\sqrt{n}} W(\mathrm{d}x), \tag{1}$$

where W is a Brownian sheet, $1/\sqrt{n}$, $n \in \mathbb{N}$, is the deviation parameter and $F \in L_2(\mathcal{D})$. The aim is to estimate the value $F(t)$, $t \in [-1/2, 1/2]^d$, from the path $\{Y_n(x)\}$ in the Gaussian white noise (GWN) model (1).

Suppose that one seeks an estimator of the value $F(t)$, of a function $F : \mathbb{R}^d \to \mathbb{R}$, under a *structural constraint* that there exist an unknown function $f : \mathbb{R}^m \to \mathbb{R}$, $m < d$, and some unknown linearly independent unit vectors $\theta_k \in \mathbb{S}^{d-1}$, $k = 1, \ldots, m$, such that

$$F(x) = f\left(\theta_1^\top x, \ldots, \theta_m^\top x\right). \tag{2}$$

Here $\mathbb{S}^{d-1}$ is a unit sphere of $\mathbb{R}^d$. This model assumption is called "multi-index" and appears, for instance, in semiparametric estimation and dimension reduction problems (see Stone, 1985 and Hristache et al., 2001). Constraint (2) gives advantages of dimension reduction but is more flexible than the single-index model (see, e.g., Delecroix et al., 2006, Goldenshluger and Lepski, 2008and references therein). Moreover, the structural assumptions are very useful in the case of functional data where the nonparametric methods provide slow convergence rate (see Chen et al., 2011).

☆ Funding of the CONICYT-FONDECYT 11140570 is acknowledged.
E-mail address: nserdyukova@udec.cl.

https://doi.org/10.1016/j.spl.2018.02.062
0167-7152/

In what follows, we consider a problem of nonparametric estimation at a given point. The corresponding risk of some estimator $\hat{F}(t)$ of $F(t)$ is given by

$$\mathcal{R}^{(n)}_{r,t}\left(\hat{F}, F\right) = \left(\mathbb{E}^{(n)}_F |\hat{F}(t) - F(t)|^r\right)^{1/r}, \qquad t \in [-1/2, 1/2]^d. \tag{3}$$

One aims at obtaining as great as possible lower bound $\psi_n(\mathbb{F})$ on the minimax risk, usually called lower rate of convergence or minimax lower bound,

$$\inf_{\tilde{F}} \sup_{F \in \mathbb{F}} \mathcal{R}^{(n)}_{r,t}\left(\tilde{F}, F\right) \gtrsim \psi_n(\mathbb{F}), \qquad n \to \infty,$$

where $\mathbb{F}$ is some class of functions and infimum is over all possible estimators. The latter inequality says that, on the class $\mathbb{F}$, the estimators $\tilde{F}(t)$ of $F(t)$ cannot converge to $F(t)$ faster than $\psi_n(\mathbb{F})$.

What this paper is really about. It was shown in the prominent work (Lepskii, 1990) on the pointwise adaptation that for estimating at a given point, that is, under the losses determined by the pointwise seminorm (3), in the Gaussian white noise (GWN) model there exist no rate optimal adaptive estimators over the Hölder classes. The latter means that the upper bound on the risk of adaptive estimators diverges from the minimax rate of convergence (see the definition below) by a logarithmic factor. Later, similar results were obtained for the density and regression models (see Brown and Low, 1996 and Gaïffas, 2007, respectively). The fact about the unavoidable "payment" for the pointwise adaptation has got to be a somewhat of common knowledge.

However, let us recall that the aforementioned notion of optimality originates from the classical minimax theory (see, for instance, Ibragimov and Has'minskii, 1981). This approach is rather objective in the sense that it allows to judge the accuracy of arbitrary estimators, but it contains two subjective "tuning" components: the way of measuring the accuracy of estimation—the loss function taken to be the pointwise seminorm in this case, and the class of functions $\mathbb{F}$ for which the maximal risk is considered. If

$$\mathcal{R}^{(n)}_{r,t}\left(\hat{F}, \mathbb{F}\right) \asymp \psi_n(\mathbb{F}), \qquad n \to \infty,$$

one says that the estimator $\hat{F}(t)$ of $F(t)$ is rate optimal over the class $\mathbb{F}$.

For instance, for $F : \mathbb{R}^d \to \mathbb{R}$ the well known examples of $\mathbb{F}$ are given by the isotropic Hölder classes $\mathbb{H}_d(\beta, L)$, $\beta > 0, L > 0$ or, more generally, by the anisotropic Hölder classes $\mathbb{H}_d(\boldsymbol{\beta}, L)$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)$, (see Definition 1). The minimax rate on these classes in the standard statistical models is very well known (see, e.g. Tsybakov, 2009, Kerkyacharian et al., 2008):

$$\psi_n(\beta, L) = L^{d/(2\beta+d)} n^{-\beta/(2\beta+d)}, \tag{4}$$

$$\psi_n(\gamma, L) = L^{1/(2\gamma+1)} n^{-\gamma/(2\gamma+1)}, \quad \gamma^{-1} = \sum_{k=1}^{d} \beta_k^{-1}, \tag{5}$$

for $\mathbb{H}_d(\beta, L)$ and $\mathbb{H}_d(\boldsymbol{\beta}, L)$, respectively. In the anisotropic case (5) the dimension d is hidden in the harmonic mean γ (see Hoffmann and Lepski, 2002, Klutchnikoff, 2005), also referred to as "exponent of global smoothness" (see Birgé, 1986).

Usually, the class $\mathbb{F}$ is determined by some multi-parameter, say, α: $\mathbb{F} = \mathbb{F}_\alpha$. In the previous example $\alpha = (\beta, L) \in \mathbb{R}^2_+$, or $\alpha = (\boldsymbol{\beta}, L) \in \mathbb{R}^{d+1}_+$. Suppose now that one managed to construct an estimator $F^*(t)$ of $F(t)$ such that *for any* α taking values in some sufficiently large compact set $\mathcal{A}$

$$\mathcal{R}^{(n)}_{r,t}\left(F^*, \mathbb{F}_\alpha\right) \asymp \psi_n(\mathbb{F}), \qquad n \to \infty.$$

Then $F^*(t)$ is optimally rate adaptive over $\cup_{\alpha \in \mathcal{A}} \mathbb{F}_\alpha$. As already mentioned, in Lepskii (1990) is shown that in the case $\mathbb{F}_\alpha = \mathbb{H}_1(\boldsymbol{\beta}, L)$ no optimally rate adaptive estimators available. In the density and regression models similar results were obtained in Brown and Low (1996) and Gaïffas (2007), respectively.

The lack of the rate optimal adaptivity means that the adaptive rate for estimating in the standard models is worse than the one of (4) or (5) by some multiplicative factor, usually logarithmic. Denote $F^*(t)$ an estimator independent of the knowledge of β or $\boldsymbol{\beta}$. Then for any $\beta \le \beta_{max}$, for some $\beta_{max} > 0$ or, respectively, $\boldsymbol{\beta} \le \boldsymbol{\beta}_{max}$, for some $\boldsymbol{\beta}_{max} = (\beta_{1,max}, \ldots, \beta_{d,max})$,

$$\mathcal{R}^{(n)}_{r,t}\left(F^*, \mathbb{H}(\beta, L)\right) \lesssim L^{d/(2\beta+d)} \left(\frac{\ln(n)}{n}\right)^{\beta/(2\beta+d)}, \tag{6}$$

$$\mathcal{R}^{(n)}_{r,t}\left(F^*, \mathbb{H}(\boldsymbol{\beta}, L)\right) \lesssim L^{1/(2\gamma+1)} \left(\frac{\ln(n)}{n}\right)^{\gamma/(2\gamma+1)}, \quad n \to \infty. \tag{7}$$

Nevertheless, the choice of $\mathbb{F} = \mathbb{H}(\beta, L)$ or $\mathbb{F} = \mathbb{H}(\boldsymbol{\beta}, L)$ is rather subjective. Suppose that in a standard statistical model one seeks an estimator of the value $F(t)$ of a function $F : \mathbb{R}^d \to \mathbb{R}$ under the *structural constraint* (2). Then a natural question is if the rate appearing in the lower bounds on the minimax risk for the smoothness classes of such "structured" functions (see the definition below) coincides with the rates in (4) and (5)? In what follows, it will be shown that the lower bounds contain an additional logarithmic factor.