The Annals of Statistics
2021, Vol. 49, No. 1, 612–613
https://doi.org/10.1214/20-AOS1993
Main article: https://doi.org/10.1214/12-AOS1053

CORRECTION NOTE: "OPTIMAL TWO-STAGE PROCEDURES FOR ESTIMATING LOCATION AND SIZE OF THE MAXIMUM OF A MULTIVARIATE REGRESSION FUNCTION" *Ann. Statist.* 40 (2012) 2850–2876

BY EDUARD BELITSER[1,*] SUBHASHIS GHOSAL[2] AND HARRY VAN ZANTEN[1,†]

[1]*Department of Mathematics, VU Amsterdam,* [*]*e.n.belitser@vu.nl;* [†]*j.h.van.zanten@vu.nl*
[2]*Department of Statistics, North Carolina State University, sghosal@stat.ncsu.edu*

We rectify a wrongly stated fact in the paper of Belitser, Ghosal and van Zanten (*Ann. Statist.* **40** (2012) 2850–2876).

In the paper of Belitser, Ghosal and van Zanten (*Ann. Statist.* **40** (2012) 2850–2876), on page 2851 it is stated that "the minimax rate for estimating the maximum value of the function ranging over an α-smooth nonparametric class (e.g., isotropic Hölder class defined below) is $n^{-\alpha/(2\alpha+d)}$" (the rate $n^{-\alpha/(2\alpha+d)}$ is also mentioned in abstract, on pp. 2553 and 2859). This is not correct: instead of "$n^{-\alpha/(2\alpha+d)}$" one should read "$n^{-\alpha/(2\alpha+d)}$ up to a log factor."

Recall the model $Y_k = f(\boldsymbol{x}_k) + \xi_k$, $\boldsymbol{x}_k \in D \subset \mathbb{R}^d$, $k = 1, \dots, n$, the assumptions and notation (like $\asymp$, $\lesssim$, etc.) from Belitser et al. (2012). Consider this model now under a fixed equidistant design $\boldsymbol{x}_k \in D \subset \mathbb{R}^d$, $k = 1, \dots, n$. Define

$$R\big(\mathcal{H}_d(\alpha, D)\big) = \inf_{\hat{M}} \sup_{f \in \mathcal{H}_d(\alpha, D)} \mathbb{E}_f |\hat{M} - M_f|, \qquad M_f = \sup_{\boldsymbol{x} \in D} f(\boldsymbol{x}),$$

the infimum is taken over all estimators. The results of [1–3] suggest that

$$R\big(\mathcal{H}_d(\alpha, D)\big) \asymp (n/\log n)^{-\alpha/(2\alpha+d)}. \tag{1}$$

We call (1) *conjecture* for now, because, despite our extensive search in the literature, we were unable to find an exact reference that establishes (or implies) (1). The results [1] and [3] are obtained only for the one dimensional case in the white noise model, so formally they do not provide a conclusive proof of (1). Below we provide a proof for the upper bound. The lower bound should proceed in the same way as for the problem of estimating the function in the sup-norm. In essence, the difficulty of the problem of estimating the maximal values of a function is the same as that of the problem of estimating the function in the sup-norm.

Here, we provide a short argument for the logarithmic sandwich bound

$$n^{-\alpha/(2\alpha+d)} \lesssim R\big(\mathcal{H}_d(\alpha, D)\big) \lesssim (n/\log n)^{-\alpha/(2\alpha+d)}. \tag{2}$$

The first inequality of (2) can be argued by comparing the problem of estimating the maximal values of a function to the problem of estimating a function value at a fixed point $\boldsymbol{x}_0 \in D$, the so-called pointwise estimation problem. The pointwise minimax rate is known to be

$$R_{\text{pw}}\big(\mathcal{H}_d(\alpha, D)\big) = \inf_{\hat{f}(\boldsymbol{x}_0)} \sup_{f \in \mathcal{H}_d(\alpha, D)} \mathbb{E}_f \big|\hat{f}(\boldsymbol{x}_0) - f(\boldsymbol{x}_0)\big| \asymp n^{-\alpha/(2\alpha+d)}.$$

Clearly, the former problem is not easier than the latter, which implies the first relation in (2): $n^{-\alpha/(2\alpha+d)} \lesssim R_{\text{pw}}(\mathcal{H}_d(\alpha, D)) \lesssim R(\mathcal{H}_d(\alpha, D))$.

Received June 2020.
MSC2020 subject classifications. Primary 62L12; secondary 62G05, 62H12, 62L05.
Key words and phrases. Two-stage procedure, optimal rate, sequential design, multistage procedure, adaptive estimation.

The second inequality of (2) can be argued by comparing to the problem of estimating a function in the sup-norm. It is known that

$$R_{\sup}(\mathcal{H}_d(\alpha, D)) = \inf_{\hat{f}} \sup_{f \in \mathcal{H}_d(\alpha, D)} \mathbb{E}_f \sup_{\boldsymbol{x} \in D} \left|\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right| \asymp \left(\frac{n}{\log n}\right)^{-\alpha/(2\alpha+d)}.$$

Let $\hat{f}$ be a minimax estimator for the above problem. The second relation in (2) follows:

$$\begin{aligned} R(\mathcal{H}_d(\alpha, D)) &\leq \sup_{f \in \mathcal{H}_d(\alpha, D)} \mathbb{E}_f \left|\sup_{\boldsymbol{x} \in D} \hat{f}(\boldsymbol{x}) - \sup_{\boldsymbol{x} \in D} f(\boldsymbol{x})\right| \\ &\leq \sup_{f \in \mathcal{H}_d(\alpha, D)} \mathbb{E}_f \sup_{\boldsymbol{x} \in D} \left|\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right| \lesssim R_{\sup}(\mathcal{H}_d(\alpha, D)) \lesssim \left(\frac{n}{\log n}\right)^{-\alpha/(2\alpha+d)}. \end{aligned}$$

Acknowledgments. The authors are grateful to Ran Chen for asking a question, Alexandre Tsybakov for pointing out (1) and providing the reference [1] and to Oleg Lepski [2] for providing a useful insight into (1).

REFERENCES

[1] IBRAGIMOV, I. A. and HASMINSKI, R. Z. (1982). Estimation of the maximal signal value in Gaussian white noise. *Mat. Zametki* **32** 529–536. Translation: *Math. Notes* **32** (1982) 746–750. MR0679245

[2] LEPSKI, O. V. (2020). Personal communication.

[3] LEPSKI, O. V. (1993). Estimation of the maximum of a nonparametric signal up to a constant. *Teor. Veroyatn. Primen.* **38** 187–194. Translation: *Theory Probab. Appl.* **38** (1993) 152–158. MR1317793 https://doi.org/10.1137/1138013