Minimax Confidence Intervals for Pointwise Nonparametric Regression Estimation

Anatoli Judisky, Alexander Nazin, Lennart Ljung, Arkadi Nemirovski
Division of Automatic Control
E-mail: Anatoli.Juditski@imag.fr, nazin@ipu.rssi.ru, ljung@isy.liu.se, nemirovs@isye.gatech.edu@isy.liu.se

13th May 2009

Report no.: LiTH-ISY-R-2908
Accepted for publication in 15th IFAC Symposium on System Identification, Saint-Malo, France

Address:
Department of Electrical Engineering
Linköpings universitet
SE-581 83 Linköping, Sweden

WWW: http://www.control.isy.liu.se
Abstract

We address a problem of estimation of an unknown regression function $f$ at a given point $x_0$ from noisy observations $y_i = f(x_i) + e_i, i = 1, ..., n$. Here $x_i \in \mathbb{R}^k$ are observable regressors and $(e_i)$ are normal i.i.d. (unobservable) disturbances. The problem is analyzed in the minimax framework, namely, we suppose that $f$ belongs to some functional class $F$, such that its finite-dimensional cut $F_n = \{ f(x_i), f \in F, i = 0, ..., n, \}$ is a convex compact set. For an arbitrary fixed regression plan $X_n = (x_1; ..., x_n)$ we study minimax on $F_n$ confidence intervals of affine estimators and construct an estimator which attains the minimax performance on the class of arbitrary estimators when the confidence level approaches 1.

Keywords: non-parametric regression, linear estimation, non-parametric identification, convex programming, identification algorithms
Minimax Confidence Intervals for Pointwise Nonparametric Regression Estimation

Anatoli Juditsky* Alexander Nazin** Lennart Ljung***
Arkadi Nemirovski****

* LIK, BP 53, F-38041 Grenoble Cedex 9, France
Email: anatoli.juditsky@imag.fr
** Institute of Control Sciences, Russian Acad. Sci., Prosfoyuznaya str., 65, 117997 Moscow, Russia
Email: nazine@ipu.rssi.ru
*** Div. of Automatic Control, Linkoping University SE-58183 Linkoping, Sweden
Email: ljung@isy.liu.se
**** School of ISyE, Georgia Institute of Technology Atlanta, Georgia 30332, USA
Email: nemirov@isye.gatech.edu

Abstract: We address a problem of estimation of an unknown regression function $f$ at a given point $x_0$ from noisy observations $y_i = f(x_i) + e_i, i = 1, ..., n$. Here $x_i \in \mathbb{R}^k$ are observable regressors and $(e_i)$ are normal i.i.d. (unobservable) disturbances. The problem is analyzed in the minimax framework, namely, we suppose that $f$ belongs to some functional class $\mathcal{F}$, such that its finite-dimensional cut $\mathcal{F}_n = \{ f(x_i), f \in \mathcal{F}, i = 0, ..., n, \}$ is a convex compact set. For an arbitrary fixed regression plan $X_n = (x_1; ..., x_n)$ we study minimax on $\mathcal{F}_n$ confidence intervals of affine estimators and construct an estimator which attains the minimax performance on the class of arbitrary estimators when the confidence level approaches 1.

Keywords: non-parametric regression, linear estimation, non-parametric identification, convex programming, identification algorithms

1. INTRODUCTION

Efficient non-parametric estimation of a regression function $f : \mathbb{R}^k \to \mathbb{R}$ based on its noisy observations

$$y_i = f(x_i) + e_i, \quad i = 1, ..., n$$

at some given design points $x_i \in \mathbb{R}^k$ is one of the basic problems for many applications including non-linear system identification, see, e.g., Ljung [1999]. Here the random noises $e_i$ are supposed to be normal i.i.d. with $E e_i = 0, E e_i^2 = \sigma^2, \sigma > 0$. We consider the problem of estimating $f$ at a given point $x_0 \in \mathbb{R}^k$.

Recently, the Direct Weight Optimization (DWO) method has been proposed to solve a problem of such kind and its properties has been studied in the case when the unknown function $f$ is continuously differentiable with Lipschitz continuous derivative having Roll et al. [2002, 2005]. Then in Juditsky et al. [2004] an adaptive version of the DWO procedure has been proposed which uses a data-driven choice of the corresponding Lipschitz constant.

To be more precise, the DWO estimator $\hat{f}_n$ is a linear estimator of the form

$$\hat{f}_n(x_0) = \sum_{i=1}^{n} \varphi_i y_i,$$

where the vector $\varphi = (\varphi_1; ..., \varphi_n)$ of weights is computed to minimize the Maximal Mean Square Error

$$\mathcal{R}_n(f_n, \mathcal{F}) = \sup_{f \in \mathcal{F}} MSE(\hat{f}_n, f)$$

$$= \sup_{f \in \mathcal{F}} E_f \left[ (\hat{f}_n(x_0) - f(x_0))^2 | X_n \right]$$

over a given functional class $\mathcal{F}$; $X_n = (x_1, ..., x_n)$.

Observe that the value $f(x_0)$ to be estimated can be seen as a linear functional of unknown “signal” $f$. The classical problem of estimation of linear functionals has received much attention in the statistical literature – some important references here are, for instance, Ibragimov, Khas’minskij [1984], Donoho, Liu [1987], Donoho, Low [1992], Donoho [1995], Cai, Low [2003]. Some important results for this problem have been obtained in the white-noise model. In particular, Donoho [1995] has shown that when it is known a priori that unknown signal belongs to a convex compact class, for several risk measures, the minimax affine estimator attains the risk bounds which coincide (up to an absolute constant) with the minimax risk. Recently in Juditsky, Nemirovski [2008] analogous results for different observation models were obtained. Our objective here is to study the implications of the theory of minimax affine estimation in the regression problem (1).

In the current paper the precision of an estimator $\hat{f}$ of $f(x_0)$ is measured with the length of the confidence interval for $f(x_0)$ at a given level of confidence. I.e., given a tolerance level $\epsilon \in (0, 1)$, we define the $\epsilon$-loss of such an
estimate at $f$ by
\[ \ell_{\epsilon}(\hat{f}; f) = \inf \left\{ \delta : P_{f} \left\{ |\hat{f} - f(x_0)| > \delta \right\} < \epsilon \right\}, \]
where, with some abuse of notation, we denote by $P_{f}$ the distribution of observations $y = (y_1, ..., y_n)$ associated with signal $f$ and regressors $X_n = (x_1, ..., x_n)$.

We consider the estimation problem in the minimax context, i.e., we suppose that the a priori information about unknown $f$ amounts to the fact that $f$ belongs to some functional class $\mathcal{F}$.

The maximal risk of the estimator $\hat{f}$ is given with
\[ \mathcal{R}_{\epsilon}(\hat{f}; \mathcal{F}) = \inf \left\{ \delta : \sup_{f \in \mathcal{F}} \left\{ |\hat{f} - f(x_0)| > \delta \right\} < \epsilon \right\}. \]
Then the minimax optimal $\epsilon$-risk is
\[ \mathcal{R}_{\epsilon}^{*}(\mathcal{F}) = \inf_{\hat{f}} \mathcal{R}_{\epsilon}(\hat{f}; \mathcal{F}), \]
where the infimum on the RHS is taken over all measurable estimation functions $f = f(X_n, y)$. We say that an estimator $\hat{f} = \hat{f}_{\varphi}$ is affine if it is of the form
\[ \hat{f}_{\varphi} = \sum_{i=1}^{n} \varphi_i y_i + \varphi_0 = \varphi^T y + \varphi_0, \]
where $\varphi = (\varphi_1; ..., \varphi_n)$. We denote with $\mathcal{R}_{A_{\epsilon}}$ the best $\epsilon$-risk which can be attained by an affine estimator:
\[ \mathcal{R}_{A_{\epsilon}}(\mathcal{F}) = \inf_{\varphi \in \mathcal{A}_{\epsilon}} \mathcal{R}_{\epsilon}(\hat{f}_{\varphi}; \mathcal{F}). \]
As we shall see later (cf Section 4), in the case when the functional class is symmetric with respect to the origin, the affine estimator reduces to the linear one.

The rest of the paper is organized as follows: we formulate the problem statement in the next section, then we present the main result on the properties of affine regression estimators in Section 3 and give the construction of the minimax affine estimator in Section 4. Finally, the theorem proof is provided in the appendix.

2. PROBLEM STATEMENT

Observe that our estimation problem is essentially $(n + 1)$-dimensional. Indeed, given the fixed design $X_n = (x_1, ..., x_n) \in \mathbb{R}^{k \times n}$ and $x_0 \in \mathbb{R}^k$, denote
\[ u = u(f, X_n) \equiv (f(x_1); ..., f(x_n)) \]
and
\[ u = u(f, X_n, x_0) \equiv (f(x_0); u). \]
The following assumption is of primary importance here:

**Assumption 1.** We suppose that the functional class $\mathcal{F}$ is such that its cross-section
\[ \mathcal{F}_n = \{ \hat{u} \in \mathcal{E} \mid f \in \mathcal{F} \} \]
is a convex and compact set in $\mathcal{E}$, $\mathcal{E}$ being some finite-dimensional Euclidean space.

Note that Assumption 1 holds true for a wide range of nonparametric regression problems. For instance, it clearly holds for the functional classes in the examples below.

**Holder class $\mathcal{F}_{\sigma}(\ell + 1, L)$.** Let $\ell \in \mathbb{N}$, and $\mathcal{F}$ be the set of function $f : \mathbb{R}^k \to \mathbb{R}$ such that for any $x \in \mathbb{R}^k$,
\[ |f(x) - P_{(0)}(x_0, x)| \leq \frac{L}{(\ell + 1)!} |x - x_0|^{\ell + 1}, \]
where $P_{(0)}(x_0, \cdot)$ is the Taylor polynomial of order $\ell$ (here $|a|$ stands for the Euclidean norm of $a$), with the extra constraints
\[ f^{(j)}(x_0) \in S_{L}^{(j)}, \ j = 0, ..., \ell, \]
where $S_{L}^{(j)}$ are some convex compact sets of $\mathbb{R}^k$.

The simplest example of such a class is the Lipschitz class $\mathcal{F}_{x_0}(1, L)$ such that
\[ |f(x) - f(x_0)| \leq L|x - x_0|, \ |f(x_0)| \leq L. \]

Another example is supplied by the class $\mathcal{F}_{\sigma}(2, L)$ of bounded regular functions with Lipschitz-continuous derivative:
\[ |f(x) - f(x_0) - f'(x_0)(x - x_0)| \leq \frac{L}{2} |x - x_0|^2, \]
when $|f(x)| \leq L, \ |f'(x_0)| \leq L$.

It is obvious that any cross-section $\mathcal{F}_n$ with finite $x_i$, $i = 1, ..., n$, is a convex compact set of $\mathbb{R}^{n+1}$. Further, this set is efficiently computationally tractable, conditioned that so are the sets $S_{L}^{(j)}$, $j = 0, ..., \ell$.

**Finite-dimensional Sobolev class $W_{M}(s, L)$.** Suppose that for some family of $M$ bounded functions $\psi_0, ..., \psi_M : \mathbb{R}^k \to \mathbb{R}$, $M < \infty$, and any $x \in \mathbb{R}^k$ we have the representation
\[ f(x) = \sum_{j=0}^{M} \psi_j(x) \beta_j, \]
where $\beta = (\beta_0; ...; \beta_M) \in B_M$, $B_M$ being a convex and compact subset of $\mathbb{R}^{M+1}$. Then, clearly,
\[ \mathcal{F}_n = \{ \hat{u} \in \mathbb{R}^{n+1} \mid u_i = \sum_{j=0}^{M} \psi_j(x_i) \beta_j, \ \beta \in B_M, \ i = 0, ..., n \} \]
is a convex, compact set.

**Class $M(L)$ of bounded monotone functions.** Let $\mathcal{F}$ be a class of functions which are monotone increasing on $\mathbb{R}$ such that $|f(x)| \leq L$ for any $x \in \mathbb{R}$. Let $x_0(1); ..., x_n$ be an increasing rearrangement of $X_n$. Then
\[ \mathcal{F}_n = \{ u \in \mathbb{R}^{n+1} \mid u_0 \leq u_1 \leq ... \leq u(n); \ |u_i| \leq L \} \]
is a compact convex set.

A reader can easily produce other examples of functional classes which ensures that the corresponding $\mathcal{F}_n$ are convex and compact subsets of $\mathbb{R}^n$ for any finite $(n + 1)$-uple $x_0, ..., x_n$. Observe also that more sets of the sort can be obtained as intersections of such sets, etc.

3. MAIN RESULT

We have the following result:

**Theorem 1.** Suppose that Assumption 1 holds. Then for any $\epsilon \in (0, 1/2]$, one can point out an affine estimator $\hat{f}$ satisfying the relation

\footnote{For details on computational tractability and complexity issues in Convex Optimization, see, e.g., [Ben-Tal, Nemirovski, 2001, Chapter 4]. A reader not familiar with this area will not lose much when interpreting a computationally tractable convex set as a set given by a finite system of inequalities $p_i(x) \leq 0$, $i = 1, ..., m$, where $p_i(x)$ are convex polynomials.}
\[
\mathcal{R}_{\epsilon}(\hat{f}_\epsilon; \mathcal{F}) \leq \frac{\text{erfinv}(\epsilon/2)}{\text{erfinv}(\epsilon)} \mathcal{R}^*_\epsilon(\mathcal{F}),
\]
where erfinv(\epsilon) is the inverse of the error function
\[
\text{erf}(s) = \frac{1}{\sqrt{2\pi}} \int_{-s}^{\infty} \exp\{-t^2/2\}dt.
\]
Therefore, erfinv(\epsilon/2)/erfinv(\epsilon) → 1 as \epsilon → +0.

In other words, for small tolerance levels \epsilon, the risk of the best affine estimator becomes arbitrarily close to the exact minimax risk in our setup:
\[
\lim_{\epsilon \to 0} \frac{\mathcal{R}_{\epsilon}(\hat{f}_\epsilon; \mathcal{F})}{\mathcal{R}^*_\epsilon(\mathcal{F})} = 1.
\]

4. CONSTRUCTION OF THE MINIMAX AFFINE ESTIMATOR

In order to compute the affine estimator \( \hat{f}_\epsilon = \varphi^T y + \varphi_0 \) which attains the bound (3) one can act as follows:

1) Let \( \tilde{u} = (u_0; u) \) and \( \tilde{v} = (v_0; v) \). Solve the optimization problem
\[
\max_{a \in \mathbb{R}^d, \nu \geq 0} \frac{1}{2}(v_0 - u_0) \quad \text{subject to} \quad \nu g(u,v) = |u-v| - 2\sigma \text{erfinv}(\epsilon/2) \leq 0.
\]
By theory of convex programming an optimal solution \( (\tilde{u}, \tilde{v}) \) to this problem can be augmented by Lagrange multipliers \( \nu \geq 0 \) such that the vectors
\[
e_{\tilde{u}} = \frac{\partial}{\partial u} \left[ \frac{1}{2}(u_0 - v_0) + \nu g(u,v) \right] = \frac{1}{2} \nu \frac{u-v}{|u-v|},
\]
\[
e_{\tilde{v}} = \frac{\partial}{\partial v} \left[ \frac{1}{2}(u_0 - v_0) + \nu g(u,v) \right] = \frac{1}{2} \nu \frac{u-v}{|u-v|},
\]
belong to normal cones of \( \mathcal{F}_n \) at the points \( \tilde{u}, \tilde{v} \), respectively: for any \( \hat{u}, \hat{v} \in \mathcal{F}_n \)
\[
e_{\tilde{u}}^T (\hat{u} - \tilde{u}) \geq 0, \quad e_{\tilde{v}}^T (\hat{v} - \tilde{v}) \geq 0,
\]
and
\[
\nu g(\tilde{u}, \tilde{v}) = 0.
\]
2) There are two possible cases: (a) \( \nu = 0 \) and (b) \( \nu > 0 \). In the case of (a), (6) implies that \( (\tilde{u}, \tilde{v}) \) is an optimal solution to the problem obtained from (5) by eliminating the constraint \( g(\hat{u}, \hat{v}) \leq 0 \), that is, \( \hat{v}_0 \) is the maximum, and \( \tilde{u}_0 \) is the minimum of \( u_0 \) on \( \mathcal{F}_n \). In this case, an estimate which reproduces \( u_0, \tilde{u} \in \mathcal{F}_n \), with the risk \( \frac{1}{2}|\tilde{v}_0 - \tilde{u}_0| \) is trivial — this is the constant estimate
\[
\hat{f}_\epsilon = \varphi_0 = \frac{1}{2}|\tilde{v}_0 + \tilde{u}_0|.
\]

Note that this case indeed takes place when \( \epsilon \) is small enough, that is, given the number of observations, our reliability requirement is too strong to allow for a nontrivial estimation.

Now assume that (b) is realized, i.e. \( \nu > 0 \). In this case, our estimation is
\[
\hat{f}_\epsilon = \varphi^T y + \varphi_0
\]
where
\[
\varphi = \frac{\nu}{\sigma \text{erfinv}(\epsilon/2)} (\tilde{v} - \tilde{u}) = \frac{2\nu}{|\tilde{v} - \tilde{u}|}
\]
and
\[
\varphi_0 = \frac{1}{2} \left[ \tilde{v}_0 + \tilde{u}_0 \right] - \frac{\nu}{2} \tilde{v}^T (\tilde{v} + \tilde{u}).
\]

(7) and (8) for \( \nu > 0 \).

Example: Let us consider the simple case when unknown function \( f \) belongs to the class \( \mathcal{F}_n(2, L) \) of regular functions with Lipschitz-continuous gradient (2). In this case the set \( \mathcal{F}_n \) is defined by the system of \( n + 1 \) affine and one quadratic inequalities. Indeed, \( \tilde{u} = (u_0; u_1; ..., u_n) \in \mathcal{F}_n \) iff there is \( w \in \mathbb{R}^k \) such that \( \tilde{u}, w \in \Delta \), where
\[
\Delta = \left\{ |u_i - u_0 - (x_i - x_0)w| \leq \frac{1}{2} L |x_i - x_0|^2, \right. \]
for \( i = 1, ..., n \), and \( |u_0| \leq L, \quad |w| \leq L. \)

The set \( \mathcal{F}_n \) is symmetric, thus the constant term \( \varphi_0 \) vanishes and the constraint \( \tilde{u}, \tilde{v} \in \mathcal{F}_n \) in the problem (5) is equivalent to \( d \equiv \frac{1}{2} (\tilde{u} - \tilde{v}) \in \mathcal{F}_n \). Further, in the case in question (5) is a convex conic problem with a quadratic constraint:
\[
\max d_0 \quad \text{subject to} \quad d, w \leq 2\sigma \text{erfinv}(\epsilon/2), \quad (d, w) \in \Delta.
\]

Let \( f: \mathbb{R} \to \mathbb{R} \) be the quadratic function
\[
f(x) = 4x^2 + 5x + 5.
\]
We are to recover \( f \) on the regular grid \( X = \{ 0 : 0.1 : 2 \} \subset [0, 2] \) given \( n = 20 \) observations
\[
y_i = f(x_i) + e_i, \quad i = 1, ..., n,
\]
where \( x_i \geq 0 \) for \( i = 1, ..., n \) are from the standard normal distribution, and \( e_i \sim N(0, 2) \). A typical recovery for \( \epsilon = 0.01 \) is presented on Fig. 1. On Fig. 2 we present the optimal weights \( \varphi_i, i = 1, ..., 20 \) for estimation of \( f(0) \); on Fig. 3 the corresponding weights for estimation of \( f(1) \) are plotted. As it was already noticed in Roll [2003], in this case the weights \( \varphi_i \) depend piece-wise quadratically on \( x \).

Let us consider now a simple bi-variate quadratic function \( f: \mathbb{R}^2 \to \mathbb{R} \) as follows:
\[
f(x) = 5|x|^2 + (5; 10)^T x + 15
\]
(c.f. Roll [2003], p. 98). Suppose that we want to recover \( f \) on the regular \( 6 \times 6 \) grid \( X = \{ 0 : 0.4 : 2 \} \subset [0, 2]^2 \) given \( n = 20 \) noisy observations taken at the points \( x_i \)
\[
i = 1, ..., n \) in the positive quadrant, sampled from the standard normal distribution on \( \mathbb{R}^2 \), and the independent disturbances \( e_i \) satisfy \( e_i \sim N(0, 2) \).

A typical result of minimax linear recovery tuned for \( \epsilon = 0.01 \) is presented on Fig. 4.
Fig. 1. Recovery of $f$ from $n = 20$ observations: solid line – true signal, + represent the observations, dashed line – minimax linear recovery, dotted lines – upper and lower bounds of confidence intervals for $f(x)$.

Fig. 2. Optimal weights $\tilde{\varphi}_i, i = 1, \ldots, 20$ for estimation of $f(0)$.

Fig. 3. Optimal weights $\tilde{\varphi}_i, i = 1, \ldots, 20$ for estimation of $f(1)$.

Fig. 4. Recovery of bivariate $f$ from $n = 20$ observations: ·-plot – true signal, +-plot – observations, ⋆-plot – minimax linear recovery.

REFERENCES


We present here the proof of Theorem 1.

For the sake of conciseness we use the notation \( u_0 = f(x_0) \) and \( u_i = f(x_i), \) \( i = 1, \ldots, n. \) Then for an affine estimator \( \hat{u} = \varphi^T y + \varphi_0 \) of \( u_0, \)

\[
\hat{u} - u_0 = \varphi^T y - u_0 + \varphi_0 + \sigma \varphi^T \epsilon
\]

where \( \eta \sim N(0,1), \) and we have for any \( \bar{u} = (u_0; u) \in F_n: \)

\[
P_u \left( \hat{u} - u_0 \geq \max_{u \in F_n} |\varphi^T u - u_0| + \varphi_0 + \sigma |\varphi| \text{erfinv}(\epsilon/2) \right) \leq \epsilon/2.
\]

In the same way we get

\[
P_u \left( \hat{u} - u_0 \leq \min_{u \in F_n} |\varphi^T u - u_0| + \varphi_0 - \sigma |\varphi| \text{erfinv}(\epsilon/2) \right) \leq \epsilon/2.
\]

Let us denote

\[
\Psi(\varphi; u; v; \epsilon) = \varphi^T (u - v) + (v_0 - u_0) + 2\sigma |\varphi| \text{erfinv}(\epsilon/2).
\]

If we set

\[
\varphi_0 = \frac{1}{2} \left[ \max_{u \in F_n} |v_0 - \varphi^T v| - \max_{u \in F_n} |\varphi^T u - u_0| \right],
\]

then for any \( \varphi \in \mathbb{R}^n \) the R-risk of estimation \( \hat{u}, \)

\[
\mathcal{R}_a(\hat{u}; F_n) = \inf_{\delta} \left\{ \delta : \sup_{u \in F_n} P_u \left\{ |\hat{u} - u_0| > \delta \right\} < \epsilon \right\}
\]

is bounded with \( S(\varphi; \epsilon) \) where

\[
2S(\varphi; \epsilon) = \max_{u, \hat{u} \in F_n} \Psi(\varphi; u, \hat{u}; \epsilon).
\]

Let now

\[
S(\epsilon) = \inf_{\varphi} S(\varphi; \epsilon) = \inf_{\varphi} \max_{u, \hat{u} \in F_n} \frac{1}{2} \Psi(\varphi; u, \hat{u}; \epsilon).
\]

Observe that \( \Psi(\varphi; u, \hat{u}; \epsilon) \) is linear in \( \hat{u}, \) \( \bar{v} \) and convex in \( \varphi. \) Further, \( F_n \) being compact, one can exchange min and max in (10), so that

\[
S(\epsilon) = \max_{\hat{u}, \bar{v}\in F_n} \frac{1}{2} \Psi(\varphi; \hat{u}, \bar{v}; \epsilon).
\]

The minimization with respect to \( \varphi \) is immediate: the infimum is \(-\infty \) if \( |u - v| > 2\sigma \text{erfinv}(\epsilon/2) \), and is \( \frac{1}{2}(v_0 - u_0) \) when \( |u - v| \leq 2\sigma \text{erfinv}(\epsilon/2) \). Thus

\[
S(\epsilon) = \max_{\hat{u}, \bar{v}\in F_n} \frac{1}{2}(v_0 - u_0)
\]

subject to

\[
g(\hat{u}, \bar{v}) \equiv |u - v| - 2\sigma \text{erfinv}(\epsilon/2) \leq 0.
\]

Note that this problem is solvable. Let us denote \( \hat{u} = (\hat{u}_0; \ldots; \hat{u}_n) \) and \( \bar{v} = (\bar{v}_0; \ldots; \bar{v}_n) \) an optimal solution to (11). The lower bound for the \( \epsilon \)-risk for this one-dimensional subproblem can be easily obtained using a simple 2-point test problem (cf., for instance, Stark [1992]). We present its proof here for the sake of completeness.

Suppose that for some \( \kappa > 0 \) and \( \epsilon \leq 1/2 \)

\[
\text{Risk}^*_\epsilon(\mathcal{F}) + \kappa \leq \frac{\text{erfinv}(\epsilon)}{\text{erfinv}(\epsilon/2)} S(\epsilon).
\]

Let \( \epsilon = 2\epsilon. \) Observe that \( S(\epsilon) \) is a positive and concave function of \( \gamma = \text{erfinv}(\epsilon/2) \) (by (9) it is a minimum of affine in \( \gamma \) functions). As \( S(1) = 0, \) for any \( 0 < \epsilon \leq 1, \)

\[
\frac{S(\epsilon)}{\text{erfinv}(\epsilon/2)} \geq \frac{S(\epsilon)}{\text{erfinv}(\epsilon/2)},
\]

and \( S(\epsilon) \geq \frac{\text{erfinv}(\epsilon)}{\text{erfinv}(\epsilon/2)} S(\epsilon). \)

Let now \( \hat{u} \) and \( \bar{v} \) be an optimal solution to (11). Note that (12) implies, by definition of \( \epsilon \)-risk, the existence of \( \epsilon' < \epsilon \) and of estimate \( \hat{u} \) such that

\[
\sup_{\hat{u} \in \{\hat{u}, \bar{v}\}} P_u ([\hat{u} - u_0] \geq \text{Risk}^*_\epsilon(\mathcal{F}) + \kappa/2) \leq \epsilon'.
\]

Consider the test

\[
\psi(y) = \mathbb{I} \left( \hat{u} \geq \frac{1}{2}\left( \hat{u}_0 + \bar{v}_0 \right) \right)
\]

of the hypothesis \( \hat{u} = \bar{u} \) against the alternative \( \hat{u} = \bar{v}. \) For this test we have

\[
P_u(\psi(y) = 1) + P_u(\psi(y) = 0) \leq 2\epsilon' < 2\epsilon.
\]

On the other hand, \( \hat{u} \) and \( \bar{v} \) satisfy

\[
|\hat{u} - \bar{v}| \leq 2\sigma \text{erfinv}(\epsilon/2).
\]

It is obvious that the optimal risk of (the sum of probabilities of errors) any test of \( \hat{u} \) against \( \bar{v} \) is not less than \( \epsilon = 2\epsilon, \) what contradicts (13).
We address a problem of estimation of an unknown regression function $f$ at a given point $x_0$ from noisy observations $y_i = f(x_i) + e_i$, $i = 1, ..., n$. Here $x_i \in \mathbb{R}^k$ are observable regressors and $(e_i)$ are normal i.i.d. (unobservable) disturbances. The problem is analyzed in the minimax framework, namely, we suppose that $f$ belongs to some functional class $F$, such that its finite-dimensional cut $F_n = \{ f(x_i), f \in F, i = 0, ..., n \}$ is a convex compact set. For an arbitrary fixed regression plan $X_n = (x_1; ..., x_n)$ we study minimax on $F_n$ confidence intervals of affine estimators and construct an estimator which attains the minimax performance on the class of arbitrary estimators when the confidence level approaches 1.