Exact Asymptotics of Minimax Bahadur Risk in Lipschitz Regression

Korostelev, A.P. *
Institute for System Analysis,
Prospekt 60-let Oktjabrja, 9, Moscow 117312, Russia, and
Department of Mathematics, Wayne State University
Detroit, MI 48202

Spokoiny, V.G.
Weierstrass Institute for Applied Analysis and Stochastics,
Mohrenstr. 39, 10117 Berlin, Germany

August 31, 1995

1991 Mathematics Subject Classification. 62G07; Secondary 62G20
Key words and phrases. Nonparametric regression, Bahadur efficiency, Chernoff function.
Abbreviated title. Exact Asymptotics of Minimax Bahadur Risk.

Abstract

The estimation problem for a Lipschitz regression at a point is studied. The exact limiting performance of the Bahadur risk is found in the minimax sense, the asymptotics being presented in the explicit form in terms of the Chernoff function.

*The research described in this paper was made possible in part by Grant MCF000 from the International Scientific Foundation

1 Introduction

Consider a nonparametric regression model with observations

$$Y_{in} = f\left(\frac{i}{n}\right) + \xi_{in}\,, \qquad i = \ldots, -1, 0, 1, \ldots; \;\; n = 1, 2, \ldots. \tag{1}$$

The regression function $f(t),\;\; t \in R^1$, belongs a priori to a class of the Lipschitz functions, i.e. $f \in \Sigma(L)$,

$$\Sigma(L) = \{f : |f(t_1) - f(t_2)| \le L|t_1 - t_2|\}$$

where L is a given positive. For each n the random variables ξ_{in} are i.i.d. with a known probability density $p(x)$. Our goal is to estimate the value $f(0)$ of the regression function at the origin from the observations Y_{in} in (1). Let $\hat{f}_n$ be an estimator, i.e. an arbitrary function of the observation Y_{in} in (1). We want to find an estimator which minimizes the probability $P_f\left(|\hat{f}_n - f(0)| > c\right)$ for a fixed positive c. Here $P_f = P_f^{(n)}$ denotes the probability of the observations Y_{in} corresponding to the true regression f. Further on we omit the superscript n for the sake of brevity.

We follow Bahadur(1960,1967) whose approach is modified in the spirit of the minimax theory (see Ibragimov and Khasminskii, 1981, Ch.1). Introduce the minimax Bahadur-type risk by

$$\beta_n(c) = \inf_{\hat{f}_n} \sup_{f \in \Sigma(L)} \frac{1}{n} \log P_f\left(|\hat{f}_n - f(0)| > c\right). \tag{2}$$

Assumption 1. *The density $p(x)$ is such that the function*

$$H(x) = -\log p(x)$$

is strictly convex and finite for any x, $\quad x \in R^1$.

Define the following function

$$G(\theta, s) = \log\left[\int_{-\infty}^{\infty} p^s(x - \theta)\, p^{1-s}(x + \theta)\, dx\right], \qquad \theta \in R^1,\; 0 \le s \le 1,$$

and introduce the logarithm of the Chernoff function (see Chernoff, 1952, Sievers, 1978)

$$S(\theta) = \min_{0 \le s \le 1} G(\theta, s), \qquad \theta \in R^1. \tag{3}$$

Under Assumption 1 the function $G(\theta, s)$ is strictly convex in s, and $G(\theta, 0) = G(\theta, 1) = 0$ which implies that the definition (3) is correct and $S(\theta)$ is negative for any $\theta \neq 0$. Note that $S(\theta)$ and $G(\theta, s)$ are symmetric in θ. It is easy to show that in the case of the symmetric density when $p(x) = p(-x)$ the infimum in (3) attains at $s = 1/2$ and $S(\theta) = G(\theta, 1/2) = \log \int \sqrt{p(x-\theta)\, p(x+\theta)}\, dx$.

The main result of this paper is in the explicit representation of the limiting performance of $\beta_n(c)$:

$$\lim_{n\to\infty} \beta_n(c) = \min_{0\le s\le 1} \frac{2}{L} \int_0^c G(\theta, s)\, d\theta. \tag{4}$$

Example 1. If ξ_{in} are $(0, \sigma^2)$-Gaussian, then $G(\theta, s) = 2s(s-1)\theta^2/\sigma^2$ and $S(\theta) = -\theta^2/(2\sigma^2)$. In this case

$$\lim_{n\to\infty} \beta_n(c) = \frac{2}{L} \int_0^c G(\theta, 1/2)\, d\theta = -\frac{c^3}{3L\sigma^2}.$$

This coincides with Korostelev, 1993.

In parallel to (1) consider a location parameter model with a sample of i.i.d. observations $X_1, \ldots, X_n$ corresponding to the density $p(x-\theta), \quad \theta \in R^1$. Let $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ be an arbitrary estimator of the location parameter θ.

Introduce the minimax Bahadur-type risk

$$r_n(c) = \inf_{\hat{\theta}_n} \sup_{\theta\in R^1} \frac{1}{n} \log P_\theta \left(|\hat{\theta}_n - \theta| > c \right), \qquad c > 0,$$

where $P_\theta = P_\theta^{(n)}$ is the probability of $X_1, \ldots, X_n$.

Let θ_n^* be the Pitman estimator of θ corresponding to the loss function $I(|\theta_n^* - \theta| > c)$ where $I(\cdot)$ denotes the indicator function. Under Assumption 1 this estimator can be defined as the unique solution of the equation

$$\sum_{i=1}^n [H(X_i - \theta_n^* - c) - H(X_i - \theta_n^* + c)] = 0.$$

The Pitman estimator of the location parameter is minimax and

$$\lim_{n\to\infty} r_n(c) = \lim_{n\to\infty} \frac{1}{n} \log P_0 \left(|\theta_n^*| > c \right) = S(c)$$

(see Chernoff,1952, Lehmann, 1959, Sievers,1978, Ibragimov and Khasminskii, 1981, Rubin and Rukhin, 1983).

An estimator θ_n^* which attains this limiting constant is called *asymptotically efficient* in the sense of Bahadur.

The efficiency in the sense of Bahadur is tightly linked with the theory of large deviations in estimation and hypothesis testing. It worthy mentioning that the maximum likelihood estimators were studied intensively from this point of view both for the moderate deviations (Ibragimov and Radavichyus, 1981, Radavichyus, 1983) and for the large deviations (Borovkov and Mogulskii, 1992). But the maximum likelihood estimator is not, generally speaking, efficient in the sense of Bahadur.

In Section 2 we give a direct proof of the lower bound in (5). Then we extend it to the case of Lipschitz regression. The asymptotics in (5) is well-known (see Sievers, 1978) and our proof of the lower bound serves to illustrate the main idea which is similar in the parametric and nonparametric case. Section 3 presents the construction of an efficient estimator for the Lipschitz regression at a point. Some technical results are postponed to Section 4.

2 Lower Bounds

Proposition 1. *If Assumption 1 holds, then the following lower bound is true for any $c > 0$:*

$$r_n(c) \geq S(c). \tag{5}$$

Proof Let ε be an arbitrary small positive. Consider the following two values of θ: $\theta = \pm c$. Note that

$$\begin{aligned}
&\sup_\theta P_\theta \left(|\hat{\theta}_n - \theta| > c - \varepsilon \right) \geq \\
&\geq \frac{1}{2} P_c \left(|\hat{\theta}_n - c| > c - \varepsilon \right) + \frac{1}{2} P_{-c} \left(|\hat{\theta}_n + c| > c - \varepsilon \right) = \\
&= \frac{1}{2} E^{(\pi)} \left[\frac{dP_c}{dP^{(\pi)}} I \left(|\hat{\theta}_n - c| > c - \varepsilon \right) + \frac{dP_{-c}}{dP^{(\pi)}} I \left(|\hat{\theta}_n + c| > c - \varepsilon \right) \right]
\end{aligned}$$

where the probability $P^{(\pi)}$ corresponds to some density $\pi = \pi(x)$; $E^{(\pi)}$ is the expectation w.r.t. $P^{(\pi)}$.

Let for $\theta = c$ the minimal value of the right-hand side of (3) attain at $s = \alpha = \alpha(c)$ which is unique under Assumption 1, $0 < \alpha < 1$, and satisfies

$$\int_{-\infty}^{\infty} p_-^\alpha \, p_+^{1-\alpha} \left(\log p_- - \log p_+ \right) \, dx = 0$$

where

$$p_\pm = p_\pm(x) = p(x \pm c).$$

Choose

$$\pi = \exp\left(-S(c)\right) \, p_-^\alpha \, p_+^{1-\alpha}$$

and note that under this choice

$$
\begin{aligned}
E^{(\pi)}\left[\log\left(\frac{p_+(X_i)}{\pi(X_i)}\right)\right] &= \int \pi \, \log \frac{p_+}{\pi} = \\
&= S(c) + \alpha \exp\left(-S(c)\right) \int p_-^{\alpha} \, p_+^{1-\alpha} \left(\log p_- - \log p_+\right) = \\
&= S(c).
\end{aligned}
$$

Similarly,

$$
E^{(\pi)}\left[\log\left(\frac{p_-(X_i)}{\pi(X_i)}\right)\right] = S(c).
$$

Denote by

$$
\Delta_{\pm} = \sum_{i=1}^{n} \left(\log \frac{p_{\pm}(X_i)}{\pi(X_i)} - S(c)\right).
$$

Due to the LLN the random event $\mathcal{A}_n = \{|\Delta_+| < \varepsilon n\} \cap \{|\Delta_-| < \varepsilon n\}$ satisfies

$$
P^{(\pi)}\left(\mathcal{A}_n\right) \to 1 \quad \text{as} \quad n \to \infty. \tag{6}
$$

The triangular inequality guarantees that

$$
P^{(\pi)}\left(\left\{|\hat{\theta}_n - c| > c - \varepsilon\right\} \cup \left\{|\hat{\theta}_n + c| > c - \varepsilon\right\}\right) = 1.
$$

Thus we finally have

$$
\begin{aligned}
\sup_{\theta \in R^1} & P_\theta\left(|\hat{\theta}_n - \theta| > c - \varepsilon\right) \geq \\
\geq \; & \frac{1}{2} \exp\left(nS(c)\right) E^{(\pi)}\Big[\exp(\Delta_+) I\left(|\hat{\theta}_n - c| > c - \varepsilon\right) + \\
& + \exp(\Delta_-) I\left(|\hat{\theta}_n + c| > c - \varepsilon\right)\Big] \\
\geq \; & \frac{1}{2} \exp\left(n\left(S(c) - \varepsilon\right)\right) P^{(\pi)}\Big(\left\{\mathcal{A}_n, \, |\hat{\theta}_n - c| > c - \varepsilon\right\} \cup \\
& \cup \left\{\mathcal{A}_n, \, |\hat{\theta}_n + c| > c - \varepsilon\right\}\Big) \\
\geq \; & \frac{1}{2} \exp\left(n\left(S(c) - \varepsilon\right)\right) P^{(\pi)}\left(\mathcal{A}_n\right).
\end{aligned}
$$

It follows that for any $\hat{\theta}_n$ the inequality is true

$$
\sup_{\theta \in R^1} \frac{1}{n} \log P_\theta\left(|\hat{\theta}_n - \theta| > c - \varepsilon\right) \geq \left(S(c) - \varepsilon\right) + \frac{1}{n} \log\left(\frac{1}{2} P^{(\pi)}\left(\mathcal{A}_n\right)\right).
$$

Hence applying (6) we come to the inequality

$$\liminf_{n\to\infty} \sup_{\theta\in R^1} \frac{1}{n} \log P_\theta \left(|\hat{\theta}_n - \theta| > c - \varepsilon \right) \geq S(c) - \varepsilon$$

and the inequality (5) follows. ¶

Remark The density π in the proof of Proposition 1 which is the "least equidistant" from p_+ and p_- does not belong in general to the family $p(\cdot - \theta)$, $\theta \in R^1$. The Gaussian case in Example 1 is an exception: here $\pi(x) = p(x)$.

Now we turn to the equality (4). As traditional in the minimax theory, we split this result into the two parts, starting with the lower bound

$$\liminf_{n\to\infty} \beta_n(c) \geq \min_{0\leq s\leq 1} \frac{2}{L} \int_0^c G(\theta, s)\, d\theta. \tag{7}$$

Theorem 1. *If Assumption 1 is satisfied then the lower bound (7) holds for the minimax Bahadur risk (2).*

Proof Note that for any estimator $\hat{f}_n$ and for an arbitrary small $\varepsilon > 0$

$$\sup_{f\in\Sigma(L)} P_f \left(|\hat{f}_n - f(0)| > c - \varepsilon \right) \geq$$
$$\geq \frac{1}{2} P_{f_+} \left(|\hat{f}_n - c| > c - \varepsilon \right) + \frac{1}{2} P_{f_-} \left(|\hat{f}_n + c| > c - \varepsilon \right)$$

where

$$f_\pm = f_\pm(t) = \begin{cases} \pm c(1 - L|t|/c) & \text{if} \quad |t| \leq c/L \\ 0 & \text{otherwise} \end{cases}.$$

Let the minimal value in s, $0 \leq s \leq 1$, of the sum $\sum_{i=-\infty}^{\infty} G\left(f_+(i/n), s\right)$ attain at $s = \alpha_n$. Note that there are finitely many non-zero summands in this sum. Let Y_{in}'s be independent and Y_{in} have the density

$$\pi_i(x) = \exp\left(-G\left(f_+(i/n), \alpha_n\right)\right) p^{\alpha_n}\left(x - f_+(i/n)\right) p^{1-\alpha_n}\left(x - f_-(i/n)\right).$$

Denote by $P^{(\pi)}$ the joint distribution of Y_{in}'s. As in the proof of Proposition 1 we obtain the inequality

$$\sup_{f\in\Sigma(L)} \frac{1}{n} \log P_f \left(|\hat{f}_n - f(0)| > c - \varepsilon \right) \geq$$
$$\geq \frac{1}{2} \exp\left\{ \sum_{i=-\infty}^{\infty} G\left(f_+(i/n), \alpha_n\right) - n\varepsilon \right\} P^{(\pi)} \left\{ \mathcal{A}_n \right\}$$

where

$$\mathcal{A}_n = \{|\Delta_+| < \varepsilon n\} \cap \{|\Delta_-| < \varepsilon n\}$$

with the random events defined by the following zero-mean random variables w.r.t. $P^{(\pi)}$:

$$\Delta_{\pm} = \sum_{i=-\infty}^{\infty} \left[\log \left(\frac{p\,(Y_{in} - f_{\pm}(i/n))}{\pi_i(Y_{in})} \right) - G\left(f_+\left(i/n\right), \alpha_n\right) \right].$$

To complete the proof it suffices to note that the sum $\frac{1}{n}\sum_{i=-\infty}^{\infty} G\left(f_+(i/n), s\right)$ converges to $\int_{-\infty}^{\infty} G(f_+(t), s)\, dt$ uniformly in $s,\ 0 \le s \le 1$, which implies that

$$\begin{aligned}\sum_{i=-\infty}^{\infty} G\left(f_+(i/n), \alpha_n\right) &= n(1+o(1)) \min_{0\le s\le 1} \int_{-c/L}^{c/L} G\left(c\left(1 - \frac{L|t|}{c}\right), s\right) dt = \\ &= n(1+o(1)) \min_{0\le s\le 1} \frac{2}{L} \int_0^c G(\theta, s)\ d\theta\end{aligned}$$

where $o(1) \to 0$ as $n \to \infty$. ¶

3 Efficient Estimator for Lipschitz regression

The Pitman estimator $\theta_n^* = \theta_n^*(X_1, \ldots, X_n)$ of the location parameter θ which is efficient in the sense of Bahadur can be defined as the center of the interval

$$\left\{ \theta : \frac{1}{n} \sum_{i=1}^{n} H(X_i - \theta) \le \lambda_* \right\}$$

where $\lambda_* = \lambda_*(c)$ is chosen such that the length of this interval equals $2c$. Thus in this case θ_n^* might be called *interval-median estimator*.

Now we extend this definition to the case of the Lipschitz regression. Put $N = [cn/L]$ and define the log-likelihood function $\mathcal{L}_N(\vartheta)$ of $(2N+1)$-dimensional argument $\vartheta = (\vartheta_{-N}, \ldots \vartheta_0, \ldots \vartheta_N)$ by

$$\mathcal{L}_N(\vartheta) = -\frac{1}{2N+1} \sum_{i=-N}^{N} \log\, p(Y_{in} - \vartheta_i) = \frac{1}{2N+1} \sum_{i=-N}^{N} H(Y_{in} - \vartheta_i).$$

Define a set $B_0 \subset R^{2N+1}$ as "traces of the Lipschitz functions":

$$B_0 = \{\vartheta : |\vartheta_i - \vartheta_j| \le L|i-j|/n, \quad |i|, |j| \le N\}.$$

Let

$$B(\lambda) = \{\vartheta : \mathcal{L}_N(\vartheta) \le \lambda\} \cap B_0\ , \quad \lambda \in R^1,$$

and let

$$b_+(\lambda) = \max_{\vartheta \in B(\lambda)} \vartheta_0\,, \qquad b_-(\lambda) = \min_{\vartheta \in B(\lambda)} \vartheta_0\,.$$

This definition is correct since $B(\lambda)$ is a convex set (if it is non-empty). As in the case of location parameter, choose $\lambda = \lambda_* = \lambda_*(c)$ such that $b_+(\lambda_*) - b_-(\lambda_*) = 2c$, and define the interval-median estimator

$$f_n^* = \frac{1}{2}\left[b_+(\lambda_*) + b_-(\lambda_*)\right].$$

Assumption 2. *The function $H(x)$ is continuously differentiable and*

$$\lim_{x\to\infty} \frac{H'(x)}{H(x)} = 0.$$

Lemma 1. *If Assumption 1 and 2 are fulfilled, then for an arbitrary large constant $S_0 > 0$ there exists $A = A(S_0)$ such that the following inequality holds for all n large enough*

$$\sup_{f\in\Sigma(L)} P_f\left(|f_n^* - f(0)| > A\right) \le \exp(-nS_0). \tag{8}$$

Denote by $w_N(\delta, \vartheta)$ the modulus of continuity

$$w_N(\delta, \vartheta) = \max_{\vartheta' :\, |\vartheta_i - \vartheta'_i| \le \delta,\, |i| \le N} |\mathcal{L}_N(\vartheta') - \mathcal{L}_N(\vartheta)|$$

where δ is a fixed positive.

Lemma 2. *Under the assumptions of Lemma 1 for any $S_0 > 0$, $A_0 > 0$ there exists $C_0 = C_0(S_0, A_0)$ such that for all n large enough the inequality holds*

$$\sup_{f\in\Sigma(L)} P_f\left(w_N(\delta, \vartheta) > \delta C_0\right) \le \exp(-nS_0) \tag{9}$$

uniformly in ϑ such that $|\vartheta_0 - f(0)| \le A_0$.

The proofs of these lemmas are postponed to the next section.

Theorem 2. *If Assumptions 1 and 2 are satisfied then the following upper bound is true uniformly in $f \in \Sigma(L)$:*

$$\limsup_{n\to\infty} \frac{1}{n} \log P_f\left(|f_n^* - f(0)| > c\right) \le \min_{0\le s\le 1} \frac{2}{L} \int_0^c G(\theta, s)\, d\theta.$$

Proof Take $S_0 = 1 + \frac{2}{L}\int_0^c S(\theta)\, d\theta$ and choose $A = A(S_0)$ due to Lemma 1. Assume that $f_n^* > f(0) + c$. This means that

$$\lambda_* - \mathcal{L}_N\left(f\left(-\frac{N}{n}\right), \dots, f\left(\frac{N}{n}\right)\right) \le 0.$$

By definition, there exists a random point $\tilde{\vartheta} = (\tilde{\vartheta}_{-N}, \dots, \tilde{\vartheta}_N)$ such that $\tilde{\vartheta} \in B_0$, $\tilde{\vartheta}_0 = b_+(\lambda_*)$, and $\mathcal{L}_N(\tilde{\vartheta}) = \lambda_*$, i.e.

$$\mathcal{L}_N(\tilde{\vartheta}) - \mathcal{L}_N\left(f\left(-\frac{N}{n}\right), \dots, f\left(\frac{N}{n}\right)\right) \le 0.$$

Unfortunately, the random point $\tilde{\vartheta}$ cannot be substituted in this inequality by a deterministic one. For this reason we approximate $\tilde{\vartheta}$ by a point from a finite set of deterministic points. Let the random event $\{|f_n^* - f(0)| \le A\}$ hold. In this case $\tilde{\vartheta} \in B_1$ where

$$B_1 = B_0 \cap \{\vartheta : |\vartheta_0 - f(0)| \le A + c\}.$$

Let δ be a small positive. Choose a finite set $\Psi = \Psi(\delta)$ of points $\psi^{(k)} = \left(\psi_{-N}^{(k)}, \dots, \psi_N^{(k)}\right)$, $k = 1, \dots, M$, such that

$$\left|\psi_i^{(k)} - \psi_j^{(k)}\right| \le \delta + L|i - j|/n, \quad |i|, |j| \le N.$$

For any $\vartheta \in B_1$ there exists $\psi^{(k)} = \psi^{(k)}(\vartheta) \in \Psi$ satisfying

$$\left|\psi_i^{(k)} - \vartheta_i\right| \le \delta, \quad |i| \le N,$$

and the cardinality $card\,\Psi = M = M(\delta)$ is independent of n and $f \in \Sigma(L)$.

The set Ψ can be obtained from the discrete piecewise approximation of the Lipschitz functions $\psi(t)$ with $|\psi(0) - f(0)| \le A + c$.

Since $\tilde{\vartheta} \in B_1$, there exists $\tilde{\psi} \in \Psi$ such that

$$\left|\tilde{\vartheta}_i - \tilde{\psi}_i\right| \le \delta \quad \text{for} \quad |i| \le N.$$

Hence

$$\tilde{\psi}_0 \ge \tilde{\vartheta}_0 - \delta = b_+(\lambda_*) - \delta = b_-(\lambda_*) + 2c - \delta > f(0) + 2c - \delta,$$

i.e.

$$\tilde{\psi}_0 - f(0) > 2c - \delta.$$

Put $A_0 = A + 2c$ and choose $C_0 = C_0(S_0, A_0)$ in accordance with Lemma 2. The inequalities (8) and (9) guarantee that uniformly in $f \in \Sigma(L)$ for all n large enough we have

$$\begin{aligned} &P_f\left(f_n^* > f(0) + c\right) \le \\ &\quad \le \; P_f\left(f_n^* > f(0) + c; \quad |f_n^* - f(0)| \le A; \quad w_N(\delta, \psi^{(k)}) \le \delta C_0\right) + \\ &\qquad +(M+1)\exp(-nS_0) \le \\ &\quad \le \sum_{k:\,|\psi_0^{(k)} - f(0)| > 2c - \delta} \\ &\qquad P_f\left\{\mathcal{L}_N\left(\psi_{-N}^{(k)}, \dots, \psi_N^{(k)}\right) - \mathcal{L}_N\left(f\left(-\frac{N}{n}\right), \dots, f\left(\frac{N}{n}\right)\right) \le \delta C_0\right\} + \\ &\qquad + (M+1)\exp(-nS_0). \end{aligned} \tag{10}$$

For each summand in the latter sum the following inequality holds (Wentzell, Ch.3, 1990, Freidlin and Wentzell, Sec.5.1, 1983):

$$\frac{1}{n}\log P_f\left\{\mathcal{L}_N\left(\psi_{-N}^{(k)},\dots,\psi_N^{(k)}\right)-\mathcal{L}_N\left(f\left(-\frac{N}{n}\right),\dots,f\left(\frac{N}{n}\right)\right)\le\delta C_0\right\}\le$$

$$\le\frac{1}{n}\min_{0\le s\le 1}\sum_{i=-N}^{N}G\left(\frac{1}{2}\left(\psi_i^{(k)}-f(i/n)\right),s\right)+h \tag{11}$$

where h is an arbitrary positive; n and $1/\delta$ are large enough. Some comments are pertinent concerning the inequality (11). The probability in the left-hand side is close to $P_f\left(\sum_{i=-N}^{N}\eta_{in}\le 0\right)$ with the random variables $\eta_{in}=H(Y_{in}-\phi_i^{(k)})-H(Y_{in}-f(i/n))=H(\xi_{in}-(\psi_i^{(k)}-f(i/n)))-H(\xi_{in})$ satisfying

$$\log E_f\left[\exp(s\eta_{in})\right]=G\left(\frac{1}{2}\left(\psi_i^{(k)}-f(i/n)\right),s\right).$$

Let $\max_{0\le s\le 1}\left[us-\sum_{i=-N}^{N}G\left(\frac{1}{2}\left(\psi_i^{(k)}-f(i/n)\right),s\right)\right]$ be the Legendre transform of the latter sum in s. Its value at the origin $u=0$ is equal to

$$\max_{0\le s\le 1}\left[-\sum_{i=-N}^{N}G\left(\frac{1}{2}\left(\psi_i^{(k)}-f(i/n)\right),s\right)\right]=$$

$$-\min_{0\le s\le 1}\sum_{i=-N}^{N}G\left(\frac{1}{2}\left(\psi_i^{(k)}-f(i/n)\right),s\right).$$

This quantity governs the log-asymptotics of the probability $P_f\left(\sum_{i=-N}^{N}\eta_{in}\le 0\right)$ as indicated in (11).

Since $\psi_0^{(k)}-f(0)>2c-\delta$ we have for any i with $|i|\le N$

$$\begin{aligned}\frac{1}{2}\left(\psi_i^{(k)}-f(\frac{i}{n})\right) &\ge \frac{1}{2}\left(\psi_0^{(k)}-f(0)-|\psi_i^{(k)}-\psi_0^{(k)}|-|f(i/n)-f(0)|\right)\ge\\ &\ge \frac{1}{2}\left(2c-\delta-(L|i/n|+\delta)-L|i/n|\right)=\\ &= c-\delta-L|i/n|.\end{aligned}$$

The function $G(\theta,s)$ is decreasing in θ for each s. Hence

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=-N}^{N}G\left(\frac{1}{2}\left(\psi_i^{(k)}-f(\frac{i}{n})\right),s\right)\le\int_{-c/L}^{c/L}G(c-\delta-L|t|,s)\,dt=$$

$$= \frac{2}{L}\int_0^{c-\delta} G(\theta,s)\,d\theta. \tag{12}$$

The number of summands in the right-hand side of (10) does not increase with n; δ and h are arbitrary small. Therefore (10)-(12) imply the inequality

$$\limsup_{n\to\infty}\frac{1}{n}\log P_f\left(f_n^* > f(0)+c\right) \le \min_{0\le s\le 1}\frac{2}{L}\int_0^{c} G(\theta,s)\,d\theta$$

uniformly in $f \in \Sigma(L)$. The similar inequality can be obtained for the probability $P_f\left(f_n^* < f(0)-c\right)$ following the same lines. ¶

4 Proof of Lemmas

Proof of Lemma 1. To prove this lemma we verify that with probability exponentially close to 1 the function $\mathcal{L}_N(\vartheta)$ is smaller than some constant λ_0 on the cube

$$K_0 = \{\vartheta : |\vartheta_i - f(0)| \le 2c\}$$

and the minimal values of this function over the cubes $K_\pm = \{\vartheta : |\vartheta_i - f(0) \mp A_1| \le 2c\}$ exceed $2\lambda_0$ for A_1 large enough. It means that $\lambda_*(c) \le \lambda_0$ and $|b_\pm(\lambda_*) - f(0)| \le A_1 + 2c$ which implies the lemma with $A = A_1 + 3c$. To do this, we first check the values of $\mathcal{L}_N(\vartheta)$ along the diagonal $\vartheta_{-N} = \ldots = \vartheta_N = \theta$ at the point $\theta = f(0)$ and $\theta = f(0) \pm A_1$. Then we use convexity of $H(x)$ to show that the oscillation of $\mathcal{L}_N(\vartheta)$ on the cubes K_0 and $K_\pm$ is finite.

Suppose without loss of generality that $H(x) > 0$ (otherwise a constant can be added to H without any influence on f_n^*). We can also assume without loss of generality that $E\xi_{in} = 0$. Note that for $|i| \le N$ the mean values of $H(Y_{in} - f(0))$ are bounded uniformly in $f \in \Sigma(L)$, i.e.

$$\sup_{f\in\Sigma(L)} E_f\, H(Y_{in} - f(0)) \le \mu_1 < +\infty.$$

The same is true for the variance:

$$\sup_{f\in\Sigma(L)} \mathrm{Var}_f\, H(Y_{in} - f(0)) \le \sigma_0^2 < +\infty.$$

Applying the Chernoff bound, we have for a fixed small z that for each $f \in$

$\Sigma(L)$

$$
\begin{aligned}
& P_f\left(\mathcal{L}_N(f(0),\ldots,f(0))>\lambda_1\right)= \\
&= P_f\left(\frac{1}{2N+1}\sum_{i=-N}^{N} H(Y_{in}-f(0))>\lambda_1\right)\leq \\
&\leq P_f\left(\sum_{i=-N}^{N}\left[H(Y_{in}-f(0))-E_f\,H(Y_{in}-f(0))\right]>(2N+1)(\lambda_1-\mu_1)\right)\leq \\
&\leq \exp\left(-(2N+1)z(\lambda_1-\mu_1)+(2N+1)\sigma_0^2 z^2\right)\leq \\
&\leq \exp(-4nS_0).
\end{aligned}
$$

Here the obvious relations are used for z small enough

$$
E_f\left(\exp(z\xi)\right)=1+\frac{z^2}{2}\mathrm{Var}(\xi)+o(z^2)\leq 1+z^2\mathrm{Var}(\xi)\leq\exp\left(z^2\mathrm{Var}(\xi)\right)
$$

where ξ is the zero-mean random variable

$$
\xi=\sum_{i=-N}^{N}\left[H(Y_{in}-f(0))-E_f\,H(Y_{in}-f(0))\right]
$$

with the finite moment generating function $E_f\left(\exp(z\xi)\right)$ in a neighborhood of the origin $z=0$. If we take

$$
\lambda_1=\frac{4n}{z(2N+1)}S_0+\sigma_0 z+\mu_1
$$

in the latter expression, we arrive at

$$
\sup_{f\in\Sigma(L)} P_f\left(\mathcal{L}_N(f(0),\ldots,f(0))>\lambda_1\right)\leq\exp(-4nS_0). \tag{13}
$$

Since $H(x)$ is convex, the following inequalities are true for any $f\in\Sigma(L)$, any $|i|\leq N$, and A_1 large enough :

$$
E_f\,\left[H(Y_{in}-f(0)-A_1)\right]\geq H\left(f(i/n)-f(0)-A_1\right)\geq H(c-A_1).
$$

This implies that $\mu_f(A_1)\to\infty$ as $A_1\to\infty$ where

$$
\mu_f(A_1)=E_f\left[\mathcal{L}_N(f(0)+A_1,\ldots,f(0)+A_1)\right]
$$

and $\mu_f(A_1)\geq H(c-A_1)>8\lambda_1$ for A_1 large enough uniformly in $f\in\Sigma(L)$.

On the other hand, Assumption 2 guarantees that for any θ fixed

$$
\lim_{x\to\infty}\frac{H(x+\theta)-H(x)}{H(x)}=0. \tag{14}
$$

Indeed, assume for the definiteness that $x \to +\infty$. Since $H'(x)$ is monotone we have for $\theta < 0$ and large x that

$$\frac{H(x+\theta) - H(x)}{H(x)} = \frac{\theta H'(\tilde{x})}{H(x)} \leq \theta \frac{H'(x)}{H(x)}$$

where $x + \theta < \tilde{x} < x$, and Assumption 2 applies directly. If $\theta > 0$, one has

$$\begin{aligned} \frac{H(x+\theta) - H(x)}{H(x)} &= \frac{\theta H'(\tilde{x})}{H(x+\theta) - \theta H'(\tilde{x})} \\ &\leq \theta \frac{H'(x+\theta)}{H(x+\theta)} \left(1 - \theta \frac{H'(x+\theta)}{H(x+\theta)}\right)^{-1}, \end{aligned}$$

and (19) follows. This equality yields the relation

$$\lim_{A_1 \to \infty} \frac{\operatorname{Var}_f H(Y_{in} - f(0) - A_1)}{(E_f\, H(Y_{in} - f(0) - A_1))^2} = 0 \tag{15}$$

uniformly in $f \in \Sigma(L)$ and $|i| \leq N$. Again, applying the Chernoff bound, we obtain from (15) that

$$\sup_{f \in \Sigma(L)} P_f \left(\mathcal{L}_N(f(0) + A_1, \ldots, f(0) + A_1) \leq \frac{1}{2} \mu_f(A_1) \right) \leq \exp(-4nS_0)$$

if $A_1 = A_1(S_0)$ is large enough. Thus $\mathcal{L}_N(\vartheta)$ is greater than $4\lambda_1$ at the center of the cube K_+ with probability exponentially close to 1. The same is true at the center of the cube K_-. Finally, the equality (14) and convexity of $H(x)$ entail the following property: for any fixed $d > 0$ and for any $x \in R^1$

$$\max_{|u| \leq d} |H(x+u) - H(x)| \leq h_0 + H(x) \tag{16}$$

with some constant $h_0 = h_0(d)$. The inequality (16) implies that the random function $\mathcal{L}_N(\vartheta)$ satisfies

$$\max_{\vartheta \in K_0} \mathcal{L}_N(\vartheta) \leq h_0(c) + \frac{5}{4} \mathcal{L}_N(f(0), \ldots, f(0))$$

with P_f-probability 1. This together with (13) gives us the following inequality:

$$\sup_{f \in \Sigma(L)} P_f \left(\max_{\vartheta \in K_0} \mathcal{L}_N(\vartheta) > h_0(c) + \frac{5}{4} \lambda_1 \right) \leq \exp(-4nS_0).$$

Applying (16) once again, we get P_f-almost surely that

$$\min_{\vartheta \in K_+} \mathcal{L}_N(\vartheta) \geq \frac{3}{4} \mathcal{L}_N(f(0) + A_1, \ldots, f(0) + A_1) - h_0(c)$$

and

$$\sup_{f \in \Sigma(L)} P_f \left(\min_{\vartheta \in K_+} \mathcal{L}_N(\vartheta) \leq \frac{1}{4} \mu_f(A_1) \right) \leq \exp(-4nS_0)$$

if A_1 is large enough. The analogous inequality for the cube K_- proves the lemma.

4.1 Proof of Lemma 2

Assumption 2 guarantees that $|H'(x)| \leq h_1 + H(x)$ with some constant h_1 for any $x \in R^1$. Thus, one gets that $w_N(\delta, \vartheta) \leq \delta\,(h_1 + \mathcal{L}_N(\vartheta))$ and

$$\sup_{f\in\Sigma(L)} P_f\left(w_N(\delta,\vartheta') > \delta C_0\right) \leq \sup_{f\in\Sigma(L)} P_f\left(h_1 + \mathcal{L}_N(\vartheta') > C_0\right) \leq \exp(-4nS_0)$$

if C_0 is large enough.

References

[1] Bahadur,R.(1960) On the asymptotic efficiency of tests and estimators. *Sankhya* **22** 229-252

[2] Bahadur, R.(1967) Rates of convergence of estimates and tests statistics. *Ann.Math.Stat.* **38** 303-324.

[3] Borovkov,A.A. and Mogulskii,A.A.(1992) Large deviation and testing of statistical hypothesis. *Proc. Inst. Math. Russian Acad. Sci., Siberian Division* **19** (English transl.: *Siber. Adv. Math.* **2** 1992 N.3,4. **3** 1993 N.1,2)

[4] Chernoff, H.(1952) A measure of asymptotic efficiency for tests of an hypothesis based on the sum of observations. *Ann.Math.Stat.* **23** 493-502.

[5] Freidlin,M.I. and Wentzell, A.D. (1984) *Random Perturbations of Dynamic Systems.* Springer. New York.

[6] Fu,J.C.(1982) Large sample point estimation: a large deviation theory approach. *Ann.Statist.* **10**, N.3. 762-771.

[7] Ibragimov,I.A. and Khasminskii,R.Z.(1981) *Statistical Estimation: Asymptotic Theory.* Springer. New York, etc.

[8] Ibragimov,I.A. and Radavichyus,M.E.(1981) Probability of large deviation for the maximum likelihood estimator. *Sov. Math. Dokl.* **23** N.2. 403-406.

[9] Korostelev,A.P. (1993) Minimaxity criterion in nonparametric regression based on large deviations probabilities. *Submitted to Ann.Statist.*

[10] Lehmann,E.L. (1959) *Testing Statistical Hypotheses.* Wiley, New York et al.

[11] Radavichyus,M.E.(1983) On the probability of large deviations of maximum likelihood estimators. *Sov. Math. Dokl.* **27** N.1. 127-131.

[12] Rubin,H. and Rukhin,A.L.(1983) Convergence rates of large deviations probabilities for point estimators. *Statistics and Probability Letters* **1** 197-202.

[13] Sievers,G.L.(1978) Estimation of location:A large deviation comparison. *Ann. Statist.* **6** 610-618.

[14] Wentzell, A.D. (1990) *Limit Theorems on Large Deviations for Markov Stochastic Processes.* Kluwer, Dordrecht.