*Bernoulli* **5**(6), 1999, 1099–1118

# The asymptotic minimax constant for sup-norm loss in nonparametric density estimation

ALEXANDER KOROSTELEV[1] and MICHAEL NUSSBAUM[2]

[1]*Department of Mathematics, Wayne State University, Detroit MI 48202, USA*
[2]*Weierstrass Institute, Mohrenstr. 39, D-10117 Berlin, Germany*

We develop the exact constant of the risk asymptotics in the uniform norm for density estimation. This constant has already been found for nonparametric regression and for signal estimation in Gaussian white noise. Hölder classes for arbitrary smoothness index $\beta > 0$ on the unit interval are considered. The constant involves the value of an optimal recovery problem as in the white noise case, but in addition it depends on the maximum of densities in the function class.

*Keywords:* density estimation; exact constant; optimal recovery; uniform norm risk; white noise

## 1. Introduction and main result

Recently, in Korostelev (1993), an asymptotically minimax exact constant was found for loss in the uniform norm, for Gaussian nonparametric regression when the parameter set is a Hölder function class. This risk bound represents an analogue of the now classical $L_2$-minimax constant of Pinsker (1980) valid for a Sobolev function class. Donoho (1994) extended Korostelev's (1993) result to signal estimation in Gaussian white noise and showed it to be related to non-stochastic optimal recovery.

Here we consider density estimation from independently and identically distributed data with a sup-norm loss. Consider a sample $X_1, \ldots, X_n$ of i.i.d. observations having a probability density $f = f(x)$ in the interval $0 \leqslant x \leqslant 1$. Let $\beta$, $L$ be some positive constants, and let $\Sigma(\beta, L)$ be the class of densities

$$\Sigma(\beta, L) = \left\{ g : \int_0^1 g = 1,\ g \geqslant 0, \text{ and } |g^{\lfloor\beta\rfloor}(x_1) - g^{\lfloor\beta\rfloor}(x_2)| \leqslant L|x_1 - x_2|^{\beta - \lfloor\beta\rfloor},\ 0 \leqslant x_1, x_2 \leqslant 1 \right\}$$

where $\lfloor\beta\rfloor$ is the greatest integer strictly less than $\beta$. Assume that the density $f$ belongs a priori to $\Sigma(\beta, L)$. Consider an arbitrary estimator $\hat{f}_n = \hat{f}_n(x)$ measurable with respect to the observations $X_1, \ldots, X_n$. We define the discrepancy of $\hat{f}_n(x)$ and the true density $f(x)$ by the sup–norm $\|\hat{f}_n - f\|_\infty$, where

$$\|f\|_\infty = \sup_{0 \leqslant x \leqslant 1} |f(x)|.$$

1350–7265 

Denote by $P_f^{(n)}$ the probability distribution of the observations $X_1, \ldots, X_n$, and by $\mathrm{E}_f^{(n)}$ the expectation with respect to $P_f^{(n)}$. Let $w(u)$, $u \geq 0$, be a continuous increasing function which admits a polynomial majorant $w(u) \leq W_0(1+u^\gamma)$ with some positive constants $W_0$, $\gamma$, and such that $w(0)=0$. Introduce the minimax risk

$$r_n = r_n(w(\cdot); \beta, L, b) = \inf_{\hat{f}_n} \sup_{f\in\Sigma(\beta,L,b)} \mathrm{E}_f^{(n)} w(\psi_n^{-1}\|\hat{f}_n - f\|_\infty), \tag{1}$$

where $\psi_n = ((\log n)/n)^{\beta/(2\beta+1)}$ is the optimal rate of convergence (cf. Khasminskii 1978; Stone 1982; Ibragimov and Khasminskii 1982). The main goal of this paper is to find the exact asymptotics of the risk (1). To do this we need two additional definitions. First, note that the densities in $\Sigma(\beta, L)$ are uniformly bounded, i.e.

$$B_* = B_*(\beta, L) = \max_{f\in\Sigma(\beta,L)} \max_{0\leq x\leq 1} f(x) < +\infty. \tag{2}$$

An argument for this, based on embedding theorems, as well as further information on the value of $B_*$ is given in the Appendix. Second, denote by $\Sigma_0(\beta, L)$ an auxiliary class of functions on the whole real line:

$$\Sigma_0(\beta, L) = \{f : |f^{\lfloor\beta\rfloor}(x_1) - f^{\lfloor\beta\rfloor}(x_2)| \leq L|x_1 - x_2|^{\beta-\lfloor\beta\rfloor},\ x_1, x_2 \in \mathbb{R}\}.$$

Let $\|g\|_2$ denote the $L_2$-norm of $g$. Define the constant

$$A_\beta = \max\{g(0) | \|g\|_2 \leq 1,\ g \in \Sigma_0(\beta, 1)\}. \tag{3}$$

**Theorem.** *For any $\beta > 0$, $L > 0$, and for any loss function $w(u)$, the minimax risk (1) satisfies*

$$\lim_{n\to\infty} r_n = w(C),$$

*where*

$$C = C(\beta, L, B_*) = A_\beta \left(\frac{2B_* L^{1/\beta}}{2\beta+1}\right)^{\beta/(2\beta+1)},$$

*and the constants $B_* = B_*(\beta, L)$ and $A_\beta$ are defined by (2) and (3), respectively.*

The proof of the corresponding upper and lower asymptotic risk bounds is developed in Sections 2 and 3. A more concise argument based on asymptotic equivalence of experiments in the LeCam sense is possible (cf. Nussbaum 1996), but only in the case $\beta > \frac{1}{2}$, and under an additional assumption that the densities are uniformly bounded away from 0. While asymptotic equivalence is known to fail for $\beta \leq \frac{1}{2}$ (cf. Brown and Zhang 1998), our method here yields the sup-norm constant for density estimation for all $\beta > 0$. The proof via asymptotic equivalence can be found in Korostelev and Nussbaum (1995).

## 2. Upper asymptotic bound

Let $g$ be a solution of the extremal problem in (3), $g \in \Sigma_0(\beta, 1)$. The correctness of this definition follows from Micchelli and Rivlin (1977), and, as shown by Leonov (1997), $g$ has a compact support. Consider also the solution $g_1 \in \Sigma_0(\beta, 1)$ of the dual extremal problem

$$\min\{\|g_1\|_2 | g_1(0) = 1,\ g_1 \in \Sigma_0(\beta, 1)\}. \tag{4}$$

If $g$ is the solution of (3) then $g_1(u) = A_\beta^{-1} g(A_\beta^{1/\beta} u)$ (cf. Section 2.2 of Donoho 1994); hence $\|g_1\|_2 = A_\beta^{-(2\beta+1)/2\beta}$. Since $g$ is of compact support, so is $g_1$; let $S$ be a constant such that $g_1(u) = 0$ for $|u| > S$. Put $K(u) = g_1(u)/\int g_1$, $u \in \mathbb{R}$, and choose the bandwidth $h_n = (C\psi_n/L)^{1/\beta}$. For an arbitrary small fixed $\epsilon > 0$ define regular grid points in the interval [0, 1] by

$$x_k = \epsilon k h_n,\ k = 0, \ldots, M,$$

where $M = M(n, \epsilon) = (\epsilon h_n)^{-1}$ is assumed integer. Put $M_0 = [S/\epsilon] + 1$, and introduce the kernel estimator $f_n^*$ at the inner grid-points

$$f_n^*(x_k) = (nh_n)^{-1} \sum_{i=1}^{n} K((X_i - x_k)/h_n), \qquad k = M_0, \ldots, M - M_0.$$

**Lemma 1.** *There exists a constant $p_0 > 0$ such that for any $\alpha > 0$ the inequality*

$$\sup_{f \in \Sigma(\beta, L)} P_f^{(n)}\Big(\max_{M_0 \leqslant k \leqslant M - M_0} |f_n^*(x_k) - f(x_k)| \geqslant (1+\alpha)C\psi_n\Big) \leqslant p_0 M^{-\alpha}$$

*holds.*

***Proof.*** Define the bias and stochastic terms by

$$b_{nk} = \mathrm{E}_f^{(n)}[f_n^*(x_k)] - f(x_k)$$

and

$$z_{nk} = f_n^*(x_k) - \mathrm{E}_f^{(n)}[f_n^*(x_k)].$$

For any $\alpha > 0$ the following inequalities are true:

$$\begin{aligned}
&P_f^{(n)}\Big(\max_{M_0 \leqslant k \leqslant M - M_0} (f_n^*(x_k) - f(x_k)) \geqslant (1+\alpha)C\psi_n\Big) \\
&\quad = P_f^{(n)}\Big(\max_{M_0 \leqslant k \leqslant M - M_0} (z_{nk} + b_{nk}) \geqslant (1+\alpha)C\psi_n\Big) \\
&\quad \leqslant P_f^{(n)}\Big(\max_{M_0 \leqslant k \leqslant M - M_0} z_{nk} \geqslant (1+\alpha)C\psi_n - \max_{M_0 \leqslant k \leqslant M - M_0} |b_{nk}|\Big) \\
&\quad \leqslant \sum_{k=M_0}^{M-M_0} P_f^{(n)}\left(z_{nk} \geqslant (1+\alpha)C\psi_n - \sup_{f \in \Sigma_0(\beta, L), f(0)=0} \left|\int_{-\infty}^{\infty} h_n^{-1} K(u/h_n) f(u)\,\mathrm{d}u\right|\right) \\
&\quad \leqslant \sum_{k=M_0}^{M-M_0} P_f^{(n)}\left(z_{nk} \geqslant (1+\alpha)C\psi_n\left(1 - \sup_{f \in \Sigma_0(\beta, 1), f(0)=0} \left|\int_{-\infty}^{\infty} K(u) f(u)\,\mathrm{d}u\right|\right)\right),
\end{aligned}$$

where the standard renormalization technique applies (see Donoho 1994). Define

$$K_\delta(u) = \delta^{-2/(2\beta+1)} g(\delta^{-2/(2\beta+1)} u) \Big/ \int g$$

for any $\delta > 0$, where $g$ is again the solution of (3). The optimal recovery identity (Micchelli and Rivlin 1977; Donoho 1994) implies that

$$\sup_{f \in \Sigma_0(\beta,1)} \sup_{\|z\|_2 \leqslant 1} \left| \int_{-\infty}^{\infty} K_\delta(u) f(u)\,\mathrm{d}u - f(0) + \delta \int_{-\infty}^{\infty} K_\delta(u) z(u)\,\mathrm{d}u \right| = \delta^{2\beta/(2\beta+1)} A_\beta,$$

hence

$$\sup_{f \in \Sigma_0(\beta,1), f(0)=0} \left| \int_{-\infty}^{\infty} K_\delta(u) f(u)\,\mathrm{d}u \right| + \delta \|K_\delta\|_2 = \delta^{2\beta/(2\beta+1)} A_\beta.$$

A choice $\delta = A_\beta^{-(2\beta+1)/2\beta}$ yields

$$K_\delta(u) = A_\beta^{1/\beta} g(A_\beta^{1/\beta} u) \Big/ \int g = g_1(u) \Big/ \int g_1 = K(u),$$

and hence

$$1 - \sup_{f \in \Sigma_0(\beta,1), f(0)=0} \left| \int_{-\infty}^{\infty} K(u) f(u)\,\mathrm{d}u \right| = A_\beta^{-(2\beta+1)/2\beta} \|K\|_2.$$

By further calculation we obtain

$$\sqrt{nh_n/(B_* \|K\|_2^2)}\, C\psi_n\, A_\beta^{-(2\beta+1)/2\beta} \|K\|_2 = \left( \frac{2}{2\beta+1} \log n \right)^{1/2}$$

and that, for any $\epsilon < 1$ and any $n$ satisfying

$$\log n > (2\beta+1)(\log \epsilon^{-1} + \beta^{-1} \log(L/C)),$$

we have

$$\left( \frac{2}{2\beta+1} \log n \right)^{1/2} \geqslant \sqrt{2 \log M}.$$

Thus, the latter sum of probabilities can be estimated from above by

$$\sum_{k=M_0}^{M-M_0} P_f^{(n)} \left( \sqrt{nh_n/(B_* \|K\|_2^2)} z_{nk} \geqslant (1+\alpha)\sqrt{2 \log M} \right).$$

Note that

$$\sqrt{nh_n/(B_* \|K\|_2^2)} z_{nk} = n^{-1/2} \sum_{i=1}^{n} \xi_{ik},$$

where

$$\xi_{ik} = \sqrt{h_n/(B_*\|K\|_2^2)}(h_n^{-1}K((X_i - x_k)/h_n) - \mathrm{E}_f^{(n)}[h_n^{-1}K((X_i - x_k)/h_n)]),$$

$$i = 1, \ldots, n, \quad k = M_0, \ldots, M - M_0.$$

The random variables $\xi_{ik}$, $i = 1, \ldots, n$, are independent for any fixed $k$, and

$$\mathrm{E}_f^{(n)}[\xi_{ik}] = 0, \qquad \mathrm{var}_f^{(n)}[\xi_{ik}] = B_*^{-1}f(x_k) + o_n(1) \leqslant 1 + o_n(1), \tag{5}$$

where $o_n(1) \to 0$ as $n \to \infty$ uniformly in $i$, $k$, and $f \in \Sigma(\beta, L)$. Moreover, for any integer $m \geqslant 3$, the following bounds hold:

$$\mathrm{E}_f^{(n)}|\xi_{ik}|^m \leqslant (h_n/(B_*\|K\|_2^2))^{m/2}(2^{m+1}B_*SH_*^m/h_n^{m-1}) = 2B_*Sh_n(\lambda/\sqrt{h_n})^m, \tag{6}$$

where $H_* = \max_{u\in\mathbb{R}}|K(u)|$ and $\lambda = 2H_*/\sqrt{B_*\|K\|_2^2}$. The Chebyshev exponential inequality, known as Chernoff's upper bound, yields

$$\begin{aligned} &P_f^{(n)}\Big(\max_{M_0\leqslant k\leqslant M-M_0}(f_n^*(x_k) - f(x_k)) \geqslant (1+\alpha)C\psi_n\Big) \\ &\leqslant \sum_{k=M_0}^{M-M_0} P_f^{(n)}\left(n^{-1/2}\sum_{i=1}^n \xi_{in} \geqslant (1+\alpha)\sqrt{2\log M}\right) \\ &\leqslant M\exp(-c(1+\alpha)\sqrt{2\log M})(\mathrm{E}_f^{(n)}[\exp(c\xi_{in}/\sqrt{n})])^n. \end{aligned}$$

Using (5) and (6), we can estimate the moment generating function as follows:

$$\begin{aligned} \mathrm{E}_f^{(n)}[\exp(c\xi_{in}/\sqrt{n})] &\leqslant 1 + \frac{c^2}{2n}\mathrm{var}_f^{(n)}[\xi_{ik}] + \sum_{m\geqslant 3}\frac{1}{m!}\left(\frac{c}{\sqrt{n}}\right)^m \mathrm{E}_f^{(n)}|\xi_{ik}|^m \\ &\leqslant 1 + \frac{c^2}{2n}(1+o_n(1)) + \frac{2B_*S\lambda^3c^3}{n\sqrt{nh_n}}\sum_{m\geqslant 3}\frac{1}{m!}\left(\frac{\lambda c}{\sqrt{nh_n}}\right)^{m-3} \\ &\leqslant 1 + \frac{c^2}{2n}\left(1 + \mathrm{o}_n(1) + \frac{4B_*S\lambda^3 c}{\sqrt{nh_n}}\exp(\lambda c/\sqrt{nh_n})\right) \\ &\leqslant \exp\left(\frac{c^2}{2n}(1+o_n(1))\right). \end{aligned} \tag{7}$$

The latter inequality is true for any $c = o_n(\sqrt{nh_n})$ as $n \to \infty$. If we choose $c = \sqrt{2\log M}$, then (7) implies that

$$\begin{aligned} &P_f^{(n)}\Big(\max_{M_0\leqslant k\leqslant M-M_0}(f_n^*(x_k) - f(x_k)) \geqslant (1+\alpha)C\psi_n\Big) \\ &\leqslant M\exp(-2(1+\alpha)\log M)\exp((1+o_n(1))\log M) \\ &\leqslant M\exp(-(1+\alpha)\log M) = M^{-\alpha} \end{aligned}$$

for any $n$ large enough. The probability of the random event

$$\left\{\min_{M_0 \leqslant k \leqslant M - M_0} (f_n^*(x_k) - f(x_k)) \leqslant -(1+\alpha)C\psi_n\right\}$$

admits the same upper bound, and this proves the lemma. □

To extend the definition of $f_n^*(x_k)$ to the grid-points $x_k$ which are close to the endpoints of the interval [0, 1], we take a kernel $K_0(u)$ with the support [0, 1] satisfying the orthogonality conditions

$$\int_0^1 K_0 = 1 \qquad \text{and} \qquad \int_0^1 u^j K_0 = 0, \quad j = 1, \ldots, \lfloor \beta \rfloor.$$

Put

$$f_n^*(x_k) = (n\kappa h_n)^{-1} \sum_{i=1}^n K_0((X_i - x_k)/(\kappa h_n)), \qquad k = 0, \ldots, M_0 - 1, \tag{8}$$

where a small positive constant $\kappa$ is chosen in Lemma 2 below. For the grid-points $x_k \in [1 - Sh_n, 1]$ we define

$$f_n^*(x_k) = (n\kappa h_n)^{-1} \sum_{i=1}^n K_0((x_k - X_i)/(\kappa h_n)), \qquad k = M - M_0 + 1, \ldots, M.$$

Put $\mathscr{M} = \{0, \ldots, M_0 - 1\} \cup \{M - M_0 + 1, \ldots, M\}$.

**Lemma 2.** *There exist constants $p_0$ and $p_1$ such that, for any $n$ and for any $\alpha > 0$, the inequality*

$$\sup_{f \in \Sigma(\beta, L)} P_f^{(n)}\left(\max_{k \in \mathscr{M}} |f_n^*(x_k) - f(x_k)| \geqslant (1+\alpha)C\psi_n\right) \leqslant p_0 M^{-\alpha p_1} \tag{9}$$

*holds.*

***Proof.*** To prove (9), it suffices to derive the upper bound for the probability

$$P_f^{(n)}\left(\max_{0 \leqslant k < M_0} (f_n^*(x_k) - f(x_k)) \geqslant (1+\alpha)C\psi_n\right) \leqslant p_0 M^{-\alpha p_1}. \tag{10}$$

The bias $b_{nk}$ of the estimator (8) at any point $x_k$ is $O((\kappa h_n)^\beta)$ as $n \to \infty$ (see Devroye and Györfi 1985). Choose $\kappa$ so small that

$$|b_{nk}| \leqslant C\psi_n/2, \qquad k = 0, \ldots, M_0 - 1.$$

Taking into account our choice of $\kappa$, and following the lines of the proof of Lemma 1, for all $n$ large enough we have the inequalities

$$
\begin{aligned}
&P_f^{(n)}\Big(\max_{0\leqslant k<M_0}(f_n^*(x_k)-f(x_k))\geqslant(1+\alpha)C\psi_n\Big)\\
&\quad\leqslant\sum_{k=0}^{M_0-1}P_f^{(n)}(z_{nk}\geqslant(1+\alpha)C\psi_n-C\psi_n/2)\\
&\quad\leqslant\sum_{k=0}^{M_0-1}P_f^{(n)}\Big(\sqrt{n\psi_n^{1/\beta}}z_{nk}\geqslant(\tfrac{1}{2}+\alpha)C\sqrt{\log n}\Big)\\
&\quad\leqslant\sum_{k=0}^{M_0-1}P_f^{(n)}\Bigg(n^{-1/2}\sum_{i=1}^{n}\xi'_{ik}\geqslant(\tfrac{1}{2}+\alpha)\sqrt{\log M}\Bigg),
\end{aligned}
$$

where

$$
\xi'_{ik}=\sqrt{\psi_n^{1/\beta}/C^2(2\beta+1)}\bigg(\frac{1}{\kappa h_n}K_0\Big(\frac{X_i-x_k}{\kappa h_n}\Big)-\mathrm{E}_f^{(n)}\Big[\frac{1}{\kappa h_n}K_0\Big(\frac{X_i-x_k}{\kappa h_n}\Big)\Big]\bigg).
$$

Similarly to (7), we obtain the inequality

$$
\mathrm{E}_f^{(n)}[\exp(c\xi'_{in}/\sqrt{n})]\leqslant\exp\bigg(\frac{c^2}{2n}\,\mathrm{var}_f^{(n)}[\xi'_{in}](1+o_n(1))\bigg),
$$

with the only difference that the variance $\mathrm{var}_f^{(n)}[\xi'_{in}]\leqslant\sigma_0^2$ is bounded by some constant $\sigma_0^2\geqslant 0$ which is not necessarily 1, as in (7). Note that $M_0$ is independent of $n$. Applying Chebyshev's exponential inequality, we have that, uniformly in $f\in\Sigma(\beta, L)$,

$$
\begin{aligned}
&P_f^{(n)}\Big(\max_{0\leqslant k<M_0}(f_n^*(x_k)-f(x_k))\geqslant(1+\alpha)C\psi_n\Big)\\
&\leqslant M_0\exp\Big(-c(\tfrac{1}{2}+\alpha)\sqrt{\log M}\Big)\exp\bigg(\frac{c^2\sigma_0^2}{2}(1+o_n(1))\bigg).
\end{aligned}
$$

Under the choice $c=\sqrt{\log M}/\sigma_0^2$, the latter formula yields the upper bound

$$
M_0\exp\bigg(-\frac{\alpha}{2\sigma_0^2}\log M\bigg)\leqslant M_0M^{-\alpha/(2\sigma_0^2)}.
$$

This completes the proof of (10), and the lemma follows. □

The derivatives $f^{(m)}(x)$, $m=1,\ldots,\lfloor\beta\rfloor$, of a density $f\in\Sigma(\beta, L)$, can be estimated in the sup-norm with the minimax rate $O(h_n^{\beta-m})$ as $n\to\infty$. We need the following version of the upper bound.

**Lemma 3.** *For any $m$, $m=1,\ldots,\lfloor\beta\rfloor$, there exist an estimator $f_n^{(m)}$ and positive constants $p_0$, $p_1$ and $C_1$ such that, for any $n$ and for any $\alpha>0$, the inequality*

$$
\sup_{f\in\Sigma(\beta,L)}P_f^{(n)}\Big(\max_{0\leqslant k\leqslant M}|f_n^{(m)}(x_k)-f^{(m)}(x_k)|\geqslant(1+\alpha)C_1h_n^{\beta-m}\Big)\leqslant p_0M^{-\alpha p_1}
$$

*holds.*

***Proof.*** Note that the upper bound in this lemma is crude since $C_1$ is not necessarily optimal. Choose the kernel $K_0(u)$ as in Lemma 2, i.e. $K_0(u)$ has support in [0, 1] and satisfies the orthogonality conditions. Assume that $K_0$ has $\lfloor\beta\rfloor + 1$ continuous derivatives. For a fixed $m$, $m \leqslant \lfloor\beta\rfloor$, put

$$f_n^{(m)}(x_k) = \begin{cases} \dfrac{(-1)^m}{h_n^{1+m}} \displaystyle\sum_{i=1}^n K_0^{(m)}\left(\frac{X_i - x_k}{h_n}\right) & \text{if } 0 \leqslant x_k \leqslant \frac{1}{2}, \\[2ex] \dfrac{1}{h_n^{1+m}} \displaystyle\sum_{i=1}^n K_0^{(m)}\left(\frac{x_k - X_i}{h_n}\right) & \text{if } \frac{1}{2} < x_k \leqslant 1, \end{cases}$$

where $K_0^{(m)}$ is the $m$th derivative of $K_0$. Standard arguments show that at each point the bias term is bounded from above by $C_2 h_n^{\beta-m}$ with a positive constant $C_2$ uniformly in $f \in \Sigma(\beta, L)$ and $x_k \in [0, 1]$. Take $C_1 > 2C_2$. Then

$$P_f^{(n)}\left(\max_{0\leqslant k\leqslant M} |f_n^{(m)}(x_k) - f^{(m)}(x_k)| \geqslant (1+\alpha)C_1 h_n^{\beta-m}\right) \leqslant P_f^{(n)}\left(\max_{0\leqslant k\leqslant M} |z_{nk}^{(m)}| \geqslant (\tfrac{1}{2}+\alpha)C_1 h_n^{\beta-m}\right),$$

where $z_{nk}^{(m)} = f_n^{(m)}(x_k) - \mathrm{E}_f^{(n)}[f_n^{(m)}(x_k)]$ are zero-mean random variables. Following the lines of the proof of Lemma 2, we find that, for all $n$ large enough, the latter probability is bounded from above by

$$2M \exp\left(-c(\tfrac{1}{2}+\alpha)\, C_1 \sqrt{\log M}\right) \exp\left(\frac{c^2 \sigma_m^2}{2}\right),$$

with an arbitrary positive $c$ and a constant $\sigma_m^2 > 0$ independent of $n$. Choose $C_1 > \sqrt{8\sigma_m^2}$, and put $c = (\frac{1}{2}+\alpha)C_1\sqrt{\log M}/\sigma_m^2$. Direct calculations show that the latter bound turns into

$$2M \exp\left(-\frac{1}{2\sigma_m^2}\left(\tfrac{1}{2}+\alpha\right)^2 C_1^2 \log M\right) \leqslant 2M^{1-4(1/2+\alpha)^2} \leqslant 2M^{-4\alpha}$$

which proves the lemma. □

***Proof of Theorem: upper risk bound.*** Take the estimators $f_n^*$ and $f_n^{(m)}$ as in Lemmas 1–3. For any $x \in [x_k, x_{k+1})$, let $f_n^*$ be the polynomial approximation

$$f_n^*(x) = f_n^*(x_k) + \sum_{m=1}^{\lfloor\beta\rfloor} \frac{1}{m!} f_n^{(m)}(x_k)(x-x_k)^m, \qquad x_k \leqslant x < x_{k+1},\ k = 0, \ldots, M-1.$$

Uniformly in $f \in \Sigma(\beta, L)$ we have the inequality

$$\|f_n^* - f\|_\infty \leqslant L(\varepsilon h_n)^\beta / \lfloor\beta\rfloor! + \max_{0\leqslant k\leqslant M} |f_n^*(x_k) - f(x_k)| + \sum_{m=1}^{\lfloor\beta\rfloor} \frac{1}{m!}(\varepsilon h_n)^m \max_{0\leqslant k\leqslant M} |f_n^{(m)}(x_k) - f^{(m)}(x_k)|,$$

where the first term on the right-hand side appears from the Taylor expansion of the density functions $f \in \Sigma(\beta, L)$. When the events complementary to those in Lemmas 1–3 hold, then

$$\|f_n^* - f\|_\infty \leqslant (1+\alpha)(C + C_2\varepsilon)\psi_n,$$

with a positive constant $C_2$ independent of $n$, $\alpha$ and $\varepsilon$. Applying Lemmas 1–3, we have

$$\sup_{f \in \Sigma(\beta, L)} P_f^{(n)}(\|f_n^* - f\|_\infty \geqslant (1+\alpha)(C + C_2\varepsilon)\psi_n) \leqslant p_2 M^{-\alpha p_3}, \tag{11}$$

where $p_2 = (1 + \lfloor\beta\rfloor)p_0$ and $p_3 = \min[1;\, p_1]$. Take an arbitrary small $\alpha_0$, and put

$$\alpha_j = j\alpha_0,\ u_j = (C + C_2\varepsilon)(1 + \alpha_j), \qquad j = 1, 2, \ldots.$$

Finally, for any continuous loss functions $w(u)$ with the polynomial majorant, we obtain from (11) that

$$\sup_{f \in \Sigma(\beta, L)} \mathrm{E}_f^{(n)} w(\psi_n^{-1}\|f_n^* - f\|_\infty) \leqslant w((1+\alpha_0)(C + C_2\varepsilon)) + W_0 \sum_{j=1}^{\infty}(1 + u_{j+1}^\gamma) p_2 M^{-j\alpha_0 p_3}.$$

Since the latter sum is vanishing as $n \to \infty$, and $\alpha_0$, $\varepsilon$ are arbitrary small, the upper bound follows. $\square$

# 3. Lower asymptotic bound

We first formulate a lemma in a general framework. For each $j = 1, \ldots, M$, let $Q_{j,\vartheta}$, $\vartheta \in [-1, 1]$, be a dominated family of distributions on some measurable space $(\mathscr{X}_j, \mathscr{F}_j)$. Let $R = [-1, 1]^M$, $\theta \in R$, and let $Q_\theta = \bigotimes_{j=1}^M Q_{j,\theta_j}$, $\theta \in R$, be the family of product measures indexed by $\theta = (\theta_1, \ldots, \theta_M)$. Define $\|\theta\|_M = \max_{1 \leqslant j \leqslant M}|\theta_j|$.

**Lemma 4.** *Let $\pi_j$ be discrete prior distributions with finite support on* $[-1, 1]$, *and consider the Bayes risks*

$$r_{j,T}(\pi_j) = \inf_{\hat\vartheta_j} \int_{[-1,1]} Q_{j,\vartheta}(|\hat\vartheta_j - \vartheta| > T)\pi_j(d\vartheta), \qquad j = 1, \ldots, M, \tag{12}$$

*where the infimum is taken over non-randomized estimators $\hat\vartheta_j$ of $\vartheta$ depending only on data from $\mathscr{X}_j$. Let $\hat\theta$ denote non-randomized estimators of $\theta$ depending on the whole data vector $x = (x_j)_{j=1,\ldots,M}$, $x_j \in \mathscr{X}_j$, let $\pi = \bigotimes_{j=1}^M \pi_j$ and consider the Bayes risk*

$$r_T(\pi) = \inf_{\hat\theta} \int Q_\theta(\|\hat\theta - \theta\|_M > T)\pi(\mathrm{d}\theta).$$

*Then, for any $T > 0$,*

$$r_T(\pi) = 1 - \prod_{j=1}^M (1 - r_{j,T}(\pi_j)).$$

***Proof.*** The $j$th Bayes risk $r_{j,T}(\pi_j)$ with data $x_j$ from $\mathscr{X}_j$ can be found as follows. Let $Q_{j,x_j}$ be the posterior distribution for $\vartheta$ and $Q_j$ be the marginal distribution for $x_j$; then

$$\int_{[-1,1]} Q_{j,\vartheta}(|\hat{\vartheta}_j - \vartheta| > T)\pi_j(\mathrm{d}\vartheta) = 1 - \int g_{j,T}(x_j,\, \hat{\vartheta}_j(x_j))Q_j(\mathrm{d}x_j),$$

where $g_{j,T}$ is the posterior gain

$$g_{j,T}(x_j,\, t) = Q_{j,x_j}(|t - \vartheta| \leqslant T).$$

If $S_j$ is the finite support of $\pi_j$ then $Q_{j,x_j}$ is concentrated on $S_j \subset [-1,\, 1]$. For any $t \in [-1,\, 1]$, we have

$$g_{j,T}(x_j,\, t) = \sum_{\vartheta \in S_j : |t-\vartheta| \leqslant T} Q_{j,x_j}(\{\vartheta\}).$$

This function of $t$ takes only finitely many values, and a maximum in $t$ is attained on some closed interval $t \in [t_{\min}(x_j),\, t_{\max}(x_j)]$. For uniqueness, take $\hat{\vartheta}^*_j(x_j) = t_{\max}(x_j)$ as a Bayes estimator. We then have

$$\max_{t \in [-1,1]} g_{j,T}(x_j,\, t) = g_{j,T}(x_j,\, \hat{\vartheta}^*_j(x_j)), \tag{13}$$

$$r_{j,T}(\pi_j) = 1 - \int g_{j,T}(x_j,\, \hat{\vartheta}^*_j(x_j))Q_j(\mathrm{d}x_j). \tag{14}$$

Consider now the global problem: we have

$$\begin{aligned} r_T(\pi) &= \inf_{\hat{\theta}} \int Q_\theta\Big(\|\hat{\theta} - \theta\|_M > T\Big)\pi(\mathrm{d}\theta) \\ &= \inf_{\hat{\theta}} \int \left(1 - \int \left(\prod_{j=1}^M \chi_{[-T,T]}(\hat{\theta}_j - \theta_j)\right) Q_\theta(\mathrm{d}x)\right)\pi(\mathrm{d}\theta) \\ &= 1 - \sup_{\hat{\theta}} \int g_T(x,\, \hat{\theta}(x)) \prod_{j=1}^M Q_j(\mathrm{d}x_j), \end{aligned} \tag{15}$$

where $g_T(x,\, u)$ is the posterior gain (for $u = (u_j)_{j=1,\ldots,M}$):

$$g_T(x,\, u) = \prod_{j=1}^M Q_{j,x_j}(|u_j - \vartheta| \leqslant T) = \prod_{j=1}^M g_{j,T}(x_j,\, u_j).$$

Then (13) implies

$$\max_{u \in R} g(x,\, u) = \prod_{j=1}^M \max_{t \in [-1,1]} g_{j,T}(x_j,\, t) = \prod_{j=1}^M g_{j,T}(x_j,\, \hat{\vartheta}^*_j(x_j)).$$

Thus a Bayes estimator of $\theta$ is

$$\hat{\theta}^*(x) = (\hat{\vartheta}^*_j(x_j))_{j=1,\ldots,M},$$

and from (15) and (14) we obtain

$$r_T(\pi) = 1 - \int g_T(x, \hat{\theta}^*(x)) \prod_{j=1}^{M} Q_j(\mathrm{d}x_j)$$

$$= 1 - \prod_{j=1}^{M} \int g_{j,T}(x_j, \hat{\vartheta}_j^*(x_j)) Q_j(\mathrm{d}x_j)$$

$$= 1 - \prod_{j=1}^{M} (1 - r_{j,T}(\pi_j)).$$

□

Back in our density problem, take a small value $\epsilon = \epsilon(\alpha) \in (0, 1)$; the final choice of $\epsilon$ will be made below. Let $f_* \in \Sigma(\beta, L)$ be such that $f_*^{\lfloor \beta \rfloor}(x)$ is constant in an interval $x \in [t_1, t_2]$, $t_2 - t_1 \leqslant \epsilon$, and $f_*(x) \geqslant B_*/(1+\epsilon)$ for $x \in [t_1, t_2]$ (cf. Lemma A.3 below). Set $f_0 = f_*(t_1)$; then $f_0 \geqslant B_*/(1+\epsilon)$. Consider again the solution $g_1 \in \Sigma_0(\beta, 1)$ of the extremal problem (4); recall $\|g_1\|_2 = A_\beta^{-(2\beta+1)/2\beta}$ and that $S$ is such that $g_1(u) = 0$ for $|u| > S$. Define

$$g_\epsilon(u) = g_1(u - S) - \epsilon g_1(\epsilon(u - 2S(1 + \epsilon^{-1}))), \qquad u \in R.$$

As is easily seen, $\int g_\epsilon = 0$, $\int g_\epsilon^2 = (1+\epsilon)\|g_1\|_2^2$ and $g_\epsilon \in \Sigma_0(\beta, 1)$ for $\epsilon$ sufficiently small. Set $l_n = h_n 2S(1 + 1/\epsilon)$ and redefine $M = M(n, \epsilon)$ from Section 2 as $M = [n^{1/((2\beta+1)(1+\epsilon))}]$. Introduce a family of functions

$$f(x; \theta) = f(x; \theta_1, \ldots, \theta_M) = f_*(x) + Lh_n^\beta \sum_{j=1}^{M} \theta_j g_\epsilon(h_n^{-1}(x - a_j)), \qquad 0 \leqslant x \leqslant 1, \tag{16}$$

where $a_1 = t_1$, $a_{j+1} - a_j = l_n$, $j = 1, \ldots, M$, $\theta = (\theta_1, \ldots, \theta_M) \in R$. The density $f(x; \theta)$ differs from $f_*(x)$ only in the interval $[t_1, t_1 + Ml_n] \subseteq [t_1, t_2]$ for $n$ large since $Mh_n \to 0$ as $n \to \infty$ for any fixed $\epsilon$. Since $f_*^{\lfloor \beta \rfloor}$ is constant on $x \in [t_1, t_2]$, we obtain that, for $\epsilon$ sufficiently small and $n$ sufficiently large, $f(x; \theta) \in \Sigma(\beta, L)$ for $\theta \in R$. Write, for brevity's sake, $P_{f(\cdot;\theta)}^{(n)} = P_\theta^{(n)}$ and $\mathrm{E}_{f(\cdot;\theta)}^{(n)} = \mathrm{E}_\theta^{(n)}$.

Define intervals $J_j = [a_j, a_j + l_n)$, $j = 1, \ldots, M$, and let $P_{j,\theta_j}$ be the conditional distribution of $X_1$ given that $X_1 \in J_j$ when $\theta$ obtains. Let $\kappa(\cdot, \cdot)$ be the Kullback–Leibler information number: for laws $P_1$, $P_2$ such that $P_1 \ll P_2$,

$$\kappa(P_1, P_2) = \int \log \frac{\mathrm{d}P_1}{\mathrm{d}P_2} \, \mathrm{d}P_1.$$

Consider also

$$\kappa_2^2(P_1, P_2) = \int \left( \log \frac{\mathrm{d}P_1}{\mathrm{d}P_2} \right)^2 \mathrm{d}P_1,$$

$$\kappa_\infty(P_1, P_2) = \operatorname*{ess\,sup}_{P_1} \left| \log \frac{\mathrm{d}P_1}{\mathrm{d}P_2} \right|.$$

**Lemma 5.** *Let* $\vartheta \in [0, 1]$ *and consider the quantities* $\kappa = \kappa(P_1, P_2)$, $\kappa_2 = \kappa_2(P_1, P_2)$ and $\kappa_\infty = \kappa_{\infty(P_1, P_2)}$ *for measures* $P_1 = P_{j,\vartheta}$, $P_2 = P_{j,-\vartheta}$ *and* $j = 1, \ldots, M$. *Set*

$$\mu = 2(1+\epsilon)^2/(2\beta+1), \qquad n_0 = nl_n f_0. \tag{17}$$

*Then, uniformly over* $j = 1, \ldots, M$, *as* $n \to \infty$:

(i) $\kappa = 2\vartheta^2 \mu_0 n_0^{-1} \log n\,(1 + o(1))$ *for some positive constant* $\mu_0 = \mu_0(\beta, L, \epsilon)$, $\mu_0 \leqslant \mu$;
(ii) $\kappa_2^2 = 2\kappa(1 + o(1))$;
(iv) $\kappa_\infty^2 = O(n_0^{-1} \log n)$.

***Proof.*** Define

$$\eta_j = l_n^{-1} \int_{J_j} f_*(x)\,\mathrm{d}x.$$

The distribution $P_{j,\vartheta}$ has density

$$f_j(x;\vartheta) = (f_*(x) + \vartheta L h_n^\beta g_\epsilon(h_n^{-1}(x - a_j)))/l_n \eta_j, \qquad x \in J_j.$$

Observe that $f_*(x) = f_0 + o(1)$ and $\eta_j = f_0 + o(1)$ uniformly in $j$ and $x$. Let us write $o^*(1)$, $O^*(1)$ for quantities which are $o(1)$ or $O(1)$ as $n \to \infty$ uniformly over $x \in J_j$ and $j = 1, \ldots, M$. Recall $f_0 \geqslant B_*/(1+\epsilon)$. Define further

$$z_j(x) = L h_n^\beta g_\epsilon(h_n^{-1}(x - a_j))/f_*(x);$$

we then obtain

$$f_j(x;\vartheta) = l_n^{-1}(1 + \vartheta z_j(x))(1 + o^*(1)), \qquad x \in J_j. \tag{18}$$

Now $\int g_\epsilon = 0$ entails

$$\int z_j(x) f_*(x)\,\mathrm{d}x = 0,$$

and as a consequence

$$\int z_j(x) f_j(x;\vartheta)\,\mathrm{d}x = \vartheta l_n^{-1} \left( \int z_j^2(x)\,\mathrm{d}x \right)(1 + o^*(1)). \tag{19}$$

Note the following relation: for $0 < z \to 0$,

$$\log \frac{1+z}{1-z} = 2z + O(z^3). \tag{20}$$

Note also

$$z_j^2(x) = O^*(h_n^{2\beta}) \tag{21}$$

and the following equalities of order of magnitude (denoted $\asymp$), which are immediate consequences of our definitions:

$$h_n^{2\beta} \asymp (\log n/n)^{2\beta/(2\beta+1)} \asymp n_0^{-1} \log n. \tag{22}$$

*Proof of (i)*. We have

$$\kappa = \int \log \frac{1 + \vartheta z_j(x)}{1 - \vartheta z_j(x)} f_j(x;\, \vartheta)\, \mathrm{d}x;$$

consequently, in view of (19) and (20),

$$\begin{aligned}\kappa &= 2\vartheta \int z_j(x) f_j(x;\, \vartheta)\, \mathrm{d}x + O(|z_j(x)|^3)\\ &= 2\vartheta^2 l_n^{-1} \left( \int z_j^2(x)\, \mathrm{d}x \right)(1 + o^*(1)) + O^*((n_0^{-1} \log n)^{3/2}).\end{aligned} \tag{23}$$

Note that

$$l_n^{-1} \int z_j^2(x)\, \mathrm{d}x = l_n^{-1} f_0^{-2} L^2 h_n^{(2\beta+1)} (1 + \epsilon) \| g_1 \|_2^2 (1 + o^*(1)).$$

Recall $\| g_1 \|_2^2 = A_\beta^{-(2\beta+1)/\beta}$; an evaluation of the right-hand side above yields

$$l_n^{-1} \int z_j^2(x)\, \mathrm{d}x = (B_*/f_0(1 + \epsilon))\mu n_0^{-1} \log n (1 + o^*(1)). \tag{24}$$

Set $\mu_0 = (B_*/f_0(1 + \epsilon))\mu$; then $\mu_0$ depends on $\epsilon$, $\beta$, $B_* = B_*(\beta, L)$ and $f_0 = f_*(t_1)$, and the function $f_*$ can be selected to depend only on $\beta$ and $L$ (cf. Lemma A.3). The inequality $f_0 \geqslant B_*/(1 + \epsilon)$ now completes the proof of (i).

*Proof of (ii)*. We have

$$\begin{aligned}\kappa_2^2 &= \int \left( \log \frac{1 + \vartheta z_j(x)}{1 - \vartheta z_j(x)} \right)^2 f_j(x;\, \vartheta)\, \mathrm{d}x\\ &= \int (2\vartheta z_j(x) + O^*((n_0^{-1} \log n)^{3/2}))^2 f_j(x;\, \vartheta)\, \mathrm{d}x\\ &= 4\vartheta^2 \left( \int z_j^2(x) f_j(x;\, \vartheta)\, \mathrm{d}x \right) + O^*((n_0^{-1} \log n)^2)\\ &= 4\vartheta^2 l_n^{-1} \left( \int z_j^2(x)\, \mathrm{d}x \right)(1 + o^*(1)) + O^*((n_0^{-1} \log n)^{3/2}),\end{aligned}$$

so that (ii) follows from (23) and (24).

*Proof of (iii)*. This is an immediate consequence of (18), (21) and (22). □

Let us state a result on large deviations for sums of i.i.d. random variables. Let $Z$, $Z_1$, $Z_2, \ldots$ be a sequence of independent real random variables with common law $Q$.

**Lemma 6.** *Assume the following:*
(i) $\mathrm{E}_Q Z = 0$, $\mathrm{var}_Q Z = 1$;
(ii) *there exists a positive constant $C$ such that $|Z| \leqslant C$ $Q$-almost surely.*

*Let $x_n$ be a sequence such that $x_n \to \infty$, $x_n = o(n^{1/2})$. Then, for every $\delta > 0$, we have*

$$\mathrm{Pr}_Q\left(n^{-1/2}\sum_{i=1}^{n} Z_i > x_n\right) \geqslant \exp(-x_n^2(1+\delta)/2)(1+o(1)), \qquad n \to \infty,$$

*uniformly over all $Q$ fulfilling (i) and (ii) for a given constant $C$.*

***Proof.*** For the moment generating function of $Z$ we have an expansion

$$\mathrm{E}\exp(tZ) = 1 + t^2/2 + \phi,$$

with a remainder term satisfying

$$|\phi| \leqslant |t|^3 C^3 \mathrm{e}^C/3!$$

uniformly over the class of distributions fulfilling (i) and (ii). Hence uniformly over $Q$ the following lower bound holds:

$$\lim_{n\to\infty}(x_n^{-2})\log \mathrm{Pr}_Q((Z_1 + \ldots + Z_n)/(x_n\sqrt{n}) > 1) \geqslant -\tfrac{1}{2}$$

(see Wentzell 1990, Theorem 4.4.1; or Freidlin and Wentzell 1984, Section 5.1, Example 4.) Thus, for all $n$ large uniformly over $Q$ satisfying (i), (ii) we have

$$\log \mathrm{Pr}_Q((Z_1 + \ldots + Z_n)/\sqrt{n} > x_n) \geqslant (-\tfrac{1}{2} - \delta)x_n^2$$

and the lemma follows. □

For measures $P_1$, $P_2$ and $P_0 = P_1 + P_2$, let $\Pi(P_1, P_2)$ be the testing affinity between $P_1$ and $P_2$:

$$\Pi(P_1, P_2) = \int \min(\mathrm{d}P_1/\mathrm{d}P_0, \mathrm{d}P_2/\mathrm{d}P_0)\,\mathrm{d}P_0.$$

Let $\nu$ be natural and consider the $\nu$-fold product measure $P_{j,\vartheta}^{\otimes\nu}$ of $P_{j,\vartheta}$ with itself, for fixed $\vartheta \in [0, 1]$ and for $-\vartheta$, and $j = 1, \ldots, M$.

**Lemma 7.** *Let* $\vartheta \in [0, 1]$ *and assume that*

$$n_0(1-\epsilon) \leqslant \nu \leqslant n_0(1+\epsilon).$$

*Then, if $\epsilon$ is sufficiently small,*

$$\Pi(P_{j,\vartheta}^{\otimes\nu}, P_{j,-\vartheta}^{\otimes\nu}) \geqslant 2n^{-\vartheta^2\mu'}(1+o(1))$$

*uniformly over* $j = 1, \ldots, M$, *where*

$$\mu' = (1+\epsilon)^6/(2\beta+1).$$

***Proof.*** It is well known that if $P_1 \ll P_2$ and $P_2 \ll P_1$ then

$$\Pi(P_1, P_2) = P_1(\mathrm{d}P_2/\mathrm{d}P_1 \geqslant 1) + P_2(\mathrm{d}P_1/\mathrm{d}P_2 > 1).$$

Set $P_1 = P_{j,\vartheta}^{\otimes \nu}$, $P_2 = P_{j,-\vartheta}^{\otimes \nu}$ and consider i.i.d. random variables $\lambda_1, \ldots, \lambda_\nu$ having the law of

$$\lambda = \log(\mathrm{d}P_{j,-\vartheta}/\mathrm{d}P_{j,\vartheta})$$

under $P_{j,\vartheta}$. Then

$$P_1(\mathrm{d}P_2/\mathrm{d}P_1 \geqslant 1) = P_{j,\vartheta}^{\otimes \nu}\left(\sum_{i=1}^{\nu} \lambda_i \geqslant 0\right). \tag{25}$$

Note that

$$\begin{aligned}
\mathrm{E}\lambda &= -\kappa(P_{j,\vartheta}, P_{j,-\vartheta}),\\
\operatorname{var}\lambda &= \kappa_2^2(P_{j,\vartheta}, P_{j,-\vartheta}) - \kappa^2(P_{j,\vartheta}, P_{j,-\vartheta})\\
&= 2\kappa(P_{j,\vartheta}, P_{j,-\vartheta})(1+o^*(1)),
\end{aligned}$$

according to Lemma 5. Set $\lambda_i^* = (\lambda_i - \mathrm{E}\lambda)/(\operatorname{var}\lambda)^{1/2}$, $i = 1, \ldots, \nu$; then (25) takes the form

$$P_1(\mathrm{d}P_2/\mathrm{d}P_1 \geqslant 1) = P_{j,\vartheta}^{\otimes \nu}\left(\nu^{-1/2}\sum_{i=1}^{\nu} \lambda_i^* \geqslant -\nu^{1/2}\mathrm{E}\lambda/(\operatorname{var}\lambda)^{1/2}\right).$$

We use Lemma 6 for a lower bound to this large-deviation probability. Note that $\operatorname{var}\lambda_1^* = 1$ and

$$|\lambda_1^*| = |\lambda - \mathrm{E}\lambda|/(\operatorname{var}\lambda)^{1/2} \leqslant (\kappa_2^2(P_{j,\vartheta}, P_{j,-\vartheta}))^{-1/2} 2\kappa_\infty(P_{j,\vartheta}, P_{j,-\vartheta}),$$

which, according to Lemma 5, is uniformly bounded for all sufficiently large $n$. This lemma also yields

$$\begin{aligned}
-\nu^{1/2}\mathrm{E}\lambda/(\operatorname{var}\lambda)^{1/2} &\leqslant (1+\epsilon)^{1/2} n_0^{1/2} 2^{-1/2}(\kappa(P_{j,\vartheta}, P_{j,-\vartheta}))^{1/2}(1+o^*(1))\\
&\leqslant (1+\epsilon)\vartheta\mu^{1/2}(\log n)^{1/2}
\end{aligned} \tag{26}$$

for sufficiently large $n$. Moreover, since (cf. (22))

$$\nu \asymp n_0 \asymp n^{2\beta/(2\beta+1)}(\log n)^{1/(2\beta+1)},$$

it follows that the right-hand side of (26) is of order $(\log \nu)^{1/2}$, hence $o(\nu^{1/2})$. Thus Lemma 6 is applicable for $x_n = (1+\epsilon)\vartheta\mu^{1/2}(\log n)^{1/2}$: for every $\delta > 0$,

$$P_1(\mathrm{d}P_2/\mathrm{d}P_1 \geqslant 1) \geqslant \exp(-\tfrac{1}{2}x_n^2(1+\delta))(1+o^*(1)).$$

Selecting $\delta = \epsilon$, we obtain

$$
\begin{aligned}
P_1(\mathrm{d}P_2/\mathrm{d}P_1 \geqslant 1) &\geqslant n^{-\vartheta^2(1+\epsilon)^4\mu/2}(1+o^*(1)) \\
&= n^{-\vartheta^2\mu'}(1+o^*(1)).
\end{aligned}
$$

For $P_2(\mathrm{d}P_1/\mathrm{d}P_2 \geqslant 1)$ this lower bound is proved analogously. □

Define numbers

$$\nu_j = \sum_{i=1}^{n} \chi_{J_j}(X_i), \qquad j = 1, \ldots, M. \tag{27}$$

The joint distribution of $\nu = (\nu_1, \ldots, \nu_M)$ under $P_\theta^{(n)}$ does not depend on $\theta$; call it $P^{(n)\nu}$.

**Lemma 8.** *For the event*

$$\mathcal{N}_n = \left\{ \sup_{j=1,\ldots,M} |\nu_j/n_0 - 1| < \epsilon \right\},$$

*where $n_0$ is given by (17) we have*

$$P^{(n)\nu}(\mathcal{N}_n) \to 1.$$

***Proof.*** Note that $\nu_j$ is a sum of i.i.d. Bernoulli random variables $\chi_{J_j}(X_i)$, $i = 1, \ldots, n$, with expectation $\int_{J_j} f_*$ and variance $(\int_{J_j} f_*)(1 - \int_{J_j} f_*)$. Let $n_j = n\int_{J_j} f_*$. Bennett's inequality (Shorack and Wellner 1986, Appendix A, p. 851) yields, for any $\epsilon' > 0$,

$$P^{(n)\nu}(|\nu_j - n_j| \geqslant n_j\epsilon') \leqslant \exp(-\epsilon' n_j^{1/2} C_{\epsilon'}) \tag{28}$$

for a constant $C'_\epsilon$. Observe $l_n^{-1}\int_{J_j} f_* = f_0 + o(1)$ uniformly in $j$, hence $n_j/n_0 \to 1$ uniformly. Note also

$$|\nu_j/n_0 - 1| \leqslant |\nu_j/n_j - 1|(n_j/n_0) + |n_j/n_0 - 1|.$$

Select $\epsilon' \leqslant \epsilon/3$ and $n$ sufficiently large such that $|n_j/n_0 - 1| < \epsilon'$; then (28) and $M = [n^{1/((2\beta+1)(1+\epsilon))}]$ imply the assertion. □

***Proof of Theorem: lower risk bound.*** We omit those details which are similar to the Gaussian case in Korostelev (1993). It suffices to prove that for an arbitrary estimator $\hat{f}_n$ and for any small $\alpha > 0$,

$$\liminf_{n\to\infty} \sup_{f\in\Sigma(\beta,L,b)} P_f^{(n)}(\|\hat{f}_n - f\|_\infty > (1-\alpha)C\psi_n) = 1.$$

Standard arguments show that this is implied by

$$\liminf_{n\to\infty} \sup_{\theta\in R} P_\theta^{(n)}(\|\hat{\theta}_n - \theta\|_M > 1 - \alpha) = 1, \tag{29}$$

where $\hat{\theta}_n = (\hat{\theta}_{n1}, \ldots, \hat{\theta}_{nM})$ is an arbitrary estimator of $\theta = (\theta_1, \ldots, \theta_M)$, $\|\theta\|_M = \max_{1\leqslant j\leqslant M} |\theta_j|$. For the intervals $J_j = [a_j, a_j + l_n)$ define conditional empirical distribution functions

$$\bar{F}_{nj}(t) = \nu_j^{-1} \sum_{i=1}^{n} \chi_{[a_j, a_j + tl_n)}(X_i), \qquad t \in [0, 1], j = 1, \ldots, M,$$

where $\nu_j$ are defined in (27).

Though the random variables $\bar{F}_{nj}$ under $P_\theta^{(n)}$ are dependent via the sample $X_1, \ldots, X_n$, they are conditionally independent given the number of sample points in each $J_j$. Thus for sets $D_1, \ldots, D_M$ in the appropriate sample space,

$$P_\theta^{(n)}(\bar{F}_{n1} \in D_1, \ldots, \bar{F}_{nM} \in D_M | \nu_1 = n_1, \ldots, \nu_M = n_M) = \prod_{j=1}^{M} P_\theta^{(n)}(\bar{F}_{nj} \in D_j | \nu_j = n_j). \quad (30)$$

Let $P_{j,\theta_j,\nu_j}^{(n)}$ be the conditional distribution of the process $\bar{F}_{nj}$ given $\nu_j$; define also a conditional empirical for the complement of $\bigcup_{j=1}^{M} J_j$ in [0, 1] and let $P_{0,\nu}^{(n)}$ be its conditional distribution given $\nu = (\nu_1, \ldots, \nu_M)$. Then $P_{\theta,\nu}^{(n)} = (\bigotimes_{j=1}^{M} P_{j,\theta_j,\nu_j}^{(n)}) \otimes P_{0,\nu}^{(n)}$ represents the conditional distribution of the whole sample $X_1, \ldots, X_n$ given $\nu$. Recall that $P^{(n)\nu}$ is the joint $P_\theta^{(n)}$-distribution of $\nu$, which is is independent of $\theta \in R$. Put $\mathscr{C}_n = \{\|\hat{\theta}_n - \theta\|_M > 1 - \alpha\}$. Consider a prior distribution $\pi = \bigotimes_{j=1}^{M} \pi_j$ on $R$ where each $\pi_j$ has finite support in $[-1, 1]$. Then

$$\begin{aligned} \inf_{\hat{\theta}_n} \sup_{\theta \in R} P_\theta^{(n)}(\mathscr{C}_n) &\geq \inf_{\hat{\theta}_n} \int_R \int_{\mathscr{N}_n} P_{\theta,\nu}^{(n)}(\mathscr{C}_n) P^{(n)\nu}(\mathrm{d}\nu) \pi(\mathrm{d}\theta) \\ &\geq P^{(n)\nu}(\mathscr{N}_n) \inf_{\nu \in \mathscr{N}_n} \inf_{\hat{\theta}_n} \int P_{\theta,\nu}^{(n)}(\mathscr{C}_n) \pi(\mathrm{d}\theta). \end{aligned}$$

In view of Lemma 8 it now suffices to prove

$$\inf_{\nu \in \mathscr{N}_n} \inf_{\hat{\theta}_n} \int P_{\theta,\nu}^{(n)}(\mathscr{C}_n) \pi(\mathrm{d}\theta) \geq 1 + o(1). \quad (31)$$

Applying Lemma 4, we obtain

$$\inf_{\hat{\theta}_n} \int P_{\theta,\nu}^{(n)}(\mathscr{C}_n) \pi(\mathrm{d}\theta) \geq 1 - \prod_{j=1}^{M} (1 - r_{j,1-\alpha}(\pi_j)), \quad (32)$$

where $r_{j,1-\alpha}(\pi_j)$ is the Bayes risk (12) for $Q_{j,\theta_j} = P_{j,\theta_j,\nu_j}^{(n)}$, $T = 1 - \alpha$. Now let us estimate this Bayes risk in each of the $M$ (conditionally) independent problems, for $\nu \in \mathscr{N}_n$. Note that each measure $P_{j,\theta_j,\nu_j}^{(n)}$ can be construed as coming from an i.i.d. sample of size $\nu_j$ governed by the conditional distribution of $X_1$ given $J_j$; i.e. by $P_{j,\theta_j}$. Consider a test of the hypothesis $\theta_j = \theta_j^+ = 1 - \alpha/2$ versus $\theta_j = \theta_j^- = -(1 - \alpha/2)$. Let $\pi_j$ be uniform on $\{\theta_j^+, \theta_j^-\}$; then we have (see Strasser 1985, Section 14.5 (4))

$$r_{j,1-\alpha}(\pi_j) \geq \tfrac{1}{2} \Pi(P_{j,\theta_j^+,\nu_j}^{(n)}, P_{j,\theta_j^-,\nu_j}^{(n)}).$$

Now apply Lemma 7, noting that

$$\Pi(P^{(n)}_{j,\theta_j^+,\nu_j},\ P^{(n)}_{j,\theta_j^-,\nu_j}) = \Pi(P^{\otimes\nu_j}_{j,\theta_j^+},\ P^{\otimes\nu_j}_{j,\theta_j^-})$$

and that on $\mathcal{N}_n$ we have $n_0(1-\epsilon) \leqslant \nu_j \leqslant n_0(1+\epsilon)$. We obtain

$$r_{j,1-\alpha}(\pi_j) \geqslant n^{-(1-\alpha/2)^2\mu'}, \tag{33}$$

for all $j = 1, \ldots, M$, if $n$ is large enough. Hence, for the right-hand side in (32) we obtain a lower bound of at least

$$1 - \prod_{j=1}^{M}(1 - n^{-(1-\alpha/2)^2\mu'}) \geqslant 1 - \exp(-Mn^{-(1-\alpha/2)^2\mu'}). \tag{34}$$

We obtain $Mn^{-(1-\alpha/2)^2\mu'} = (1+o(1))n^{\mu''}$ for an exponent

$$\begin{aligned}\mu'' &= 1/(2\beta+1)(1+\epsilon) - (1-\alpha/2)^2\mu' \\ &= 1/(2\beta+1)(1+\epsilon) - (1-\alpha/2)^2(1+\epsilon)^6/(2\beta+1).\end{aligned}$$

For given $\alpha > 0$, $\epsilon$ can be chosen such that $\mu'' > 0$. In that case $\exp(-Mn^{-(1-\alpha/2)^2\mu'}) \to 0$ and (34) implies (31). □

# Appendix: Analytic facts

The fact that densities of the class $\Sigma(\beta, L)$ are uniformly bounded in sup-norm follows from standard embedding theorems.

**Lemma A.1.** *For any* $L > 0$ *and* $\beta > 0$,

$$B_*(\beta, L) = \max_{f\in\Sigma(\beta,L)} \max_{0\leqslant x\leqslant 1} f(x) < +\infty.$$

***Proof.*** Apply Theorem 17.4 of Besov *et al*. (1979), using the fact that $f$ is bounded in $L_1$-norm on [0, 1]. □

For $\beta \leqslant 1$, the value of $B_*(\beta, L)$ can be found.

**Lemma A.2.** *For any* $L > 0$ *and* $0 < \beta \leqslant 1$,

$$B_*(\beta, L) = \begin{cases} ((\beta+1)/\beta)^{\beta/(\beta+1)}L^{1/(\beta+1)} & \text{if } L \geqslant (\beta+1)/\beta, \\ 1 + L/(\beta+1) & \text{if } L \leqslant (\beta+1)/\beta. \end{cases}$$

***Proof.*** It can be shown that the extremal density is

$$f(x) = \max((f(0) - Lx^\beta), 0), \qquad x \in [0, 1].$$

An easy calculation from $\int f(x)\,\mathrm{d}x = 1$ yields $f(0)$. □

**Lemma A.3.** *For any $L > 0$ and $\beta > 0$, and every $\epsilon \in (0, 1)$ there exist $t_1, t_2 \in [0, 1]$, $0 < t_2 - t_1 \leqslant \epsilon$ and a function $f_* \in \Sigma(\beta, L)$ such that, for all $x \in [t_1, t_2]$,*

$$f_*(x) \geqslant B_*(\beta, L)/(1+\epsilon), \tag{35}$$

$$f_*^{\lfloor\beta\rfloor}(x) = f_*^{\lfloor\beta\rfloor}(t_1).$$

***Proof.*** Let $f$ be a solution in $f$ of problem (2), i.e. $\|f\|_\infty = B_*(\beta, L)$. Let $\tilde{\epsilon} \in (0, \epsilon)$, and let $t_1, t_2 \in [0, 1]$, $t_2 - t_1 = \tilde{\epsilon}$ be such that $f(x) \geqslant B_*(\beta, L)/(1+\epsilon/2)$ for $x \in [t_1, t_2]$. Since $f \in \Sigma(\beta, L)$ is continuous on $[0, 1]$, such $t_1, t_2 \in [0, 1]$ exist for sufficiently small $\tilde{\epsilon}$. Let $m = \lfloor\beta\rfloor$, $\gamma = \beta - m$, and let $t_0 \in [t_1, t_2]$ be such that $f^{(m)}(t_0) \geqslant f^{(m)}(x)$, for $x \in [t_1, t_2]$. Since $f^{(m)}$ is continuous, such a $t_0$ exists. Define a function $g_0$ by

$$g_0(x) = \begin{cases} f^{(m)}(t_0) - f^{(m)}(t_1), & x \in [0, t_1), \\ f^{(m)}(t_0) - f^{(m)}(x), & x \in [t_1, t_2], \\ f^{(m)}(t_0) - f^{(m)}(t_2), & x \in (t_2, 1]. \end{cases}$$

Note that $g_0(x) \geqslant 0$, $x \in [0, 1]$ and

$$\|g_0\|_\infty \leqslant L|t_2 - t_1|^\gamma = L\tilde{\epsilon}^\gamma.$$

Let $Q$ be the integral operator $Qg(t) = \int_0^t g(u)\,\mathrm{d}u$, $t \in [0, 1]$, and define $\tilde{g} = Q^m g_0$ ($m$-fold application of $Q$). Then $\tilde{g}(x) \geqslant 0$, $x \in [0, 1]$ and

$$\|\tilde{g}\|_\infty \leqslant \|g_0\|_\infty \leqslant L\tilde{\epsilon}^\gamma. \tag{36}$$

Define $\tilde{f} = f + \tilde{g}$. Since $\tilde{f}^{(m)}(t) = f^{(m)}(t_0)$ on $[t_1, t_2]$ while $\tilde{f}^{(m)}(t) - f^{(m)}(t)$ is constant outside $(t_1, t_2)$, it follows that

$$|\tilde{f}^{(m)}(x_1) - \tilde{f}^{(m)}(x_2)| \leqslant L|x_1 - x_2|^\gamma, \qquad x_1, x_2 \in [0, 1].$$

Furthermore, $\tilde{f} \geqslant f$; and, by (36),

$$\|\tilde{f} - f\|_\infty \leqslant L\tilde{\epsilon}^\gamma.$$

Defining $f_* = \tilde{f}/\int\tilde{f}$, we see that $f_*$ is a density in $\Sigma(\beta, L)$. Moreover, $f_*(x) \geqslant B_*(\beta, L)/(1+\epsilon/2)\int\tilde{f}$ for $x \in [t_1, t_2]$. By selecting $\tilde{\epsilon}$ sufficiently small, we achieve (35). □

## Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 373, Quantification and Simulation of Economic Processes, Berlin, Germany. The research of the first author was supported by the International Science Foundation under grant MCF000. The authors gratefully acknowledge the help of Alexander Tsybakov with the concise proof of Lemma A.1.

# References

Besov, O.V., Il'in, V.P. and Nikol'skii, S.M. (1979) *Integral Representation of Functions and Imbedding Theorems*, Vol. II. Washington. DC: V.H. Winston & Sons; and New York: Wiley.

Brown, L.D. and Zhang, C.-H. (1998) Asymptotic nonequivalence of nonparametric experiments when the smoothness index is $1/2$. *Ann. Statist.*, **26**, 279–287.

Devroye, L. and Györfi, L. (1985) *Nonparametric Density Estimation: the $L_1$-View.* New York: Wiley.

Donoho, D. (1994) Asymptotic minimax risk (for sup-norm loss): solution via optimal recovery. *Probab. Theory Related Fields*, **99**, 145–170.

Freidlin, M.I., and Wentzell, A.D. (1984) *Random Perturbation of Dynamical Systems.* New York: Springer Verlag.

Ibragimov, I.A. and Khasminskii, R.Z. (1982) Bounds for the risks of non-parametric regression estimates. *Theory Probab. Appl.*, **27**, 84–99.

Khasminskii, R.Z. (1978) A lower bound on the risk of non-parametric estimates of densities in the uniform metric. *Theory Probab. Appl.*, **23**, 794–798.

Korostelev, A.P. (1993) Exact asymptotically minimax estimator for nonparametric regression in uniform norm. *Theory Probab. Appl.*, **38**, 775–782.

Korostelev, A.P. and Nussbaum, M. (1995) Density estimation in the uniform norm and white noise approximation. Preprint No. 154, Weierstrass Institute, Berlin.

Leonov, S.L (1997) On the solution of an optimal recovery problem and its applications in nonparametric regression. *Math. Methods Statist.*, **6**, 476–490.

Micchelli, C. and Rivlin, T. (1977) A survey of optimal recovery. In C. Micchelli and T. Rivlin (eds), *Optimal Estimation in Approximation Theory,* pp. 1–54. New York: Plenum Press.

Nussbaum, M. (1996) Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.*, **24**, 2399–2430.

Pinsker, M.S. (1980) Optimal filtering of square integrable signals in Gaussian white noise. *Problems Inform. Transmission*, **16**, 120–133.

Shorack, G. and Wellner, J. (1986). *Empirical Processes with Applications to Statistics*. New York: Wiley.

Stone, C. (1982) Optimal global rates of convergence for nonparametric regression. *Ann. Statist.*, **10**, 1040–1053.

Strasser, H. (1985) *Mathematical Theory of Statistics.* Berlin: Walter de Gruyter.

Wentzell, A. (1990) *Theorems on Large Deviations for Markov Stochastic Processes.* Dordrecht: Kluwer.

Received February 1996 and revised March 1999