

Piotr Majerski

**RATES OF CONVERGENCE
FOR THE MAXIMUM LIKELIHOOD ESTIMATOR
IN THE CONVOLUTION MODEL**

Abstract. Rates of convergence for the maximum likelihood estimator in the convolution model, obtained recently by S. van de Geer, are reconsidered and corrected.

Keywords: maximum likelihood estimator, entropy, Hellinger distance, rate of convergence.

Mathematics Subject Classification: 62G07.

1. INTRODUCTION

Consider independent, identically distributed random variables X_1, X_2, \dots, X_n in a measurable space $(\mathcal{X}, \mathcal{A})$ with distribution P . Suppose that

$$f_0 = \frac{dP}{d\mu} \in \mathcal{F},$$

where μ is a dominating, σ -finite measure, and \mathcal{F} is a given class of densities with respect to μ . Throughout the whole paper, \hat{f}_n will denote the maximum likelihood estimator (MLE) of f_0 and the accuracy of the estimation will be measured in the Hellinger distance defined as

$$h(\hat{f}_n, f_0) = \left(\frac{1}{2} \int \left(\sqrt{\hat{f}_n} - \sqrt{f_0} \right)^2 d\mu \right)^{\frac{1}{2}}.$$

Our interest will be focused on upper bounds for the convergence rates, when \mathcal{F} is a class of convolution densities.

The paper is organized as follows. In this section, basic notations are introduced and some technical results are formulated. In Sections 2 and 3, the rates of convergence, given in [3] and [2] for two special convolution models, are reconsidered and corrected.

For a class \mathcal{K} of functions on $(\mathcal{X}, \mathcal{A})$, let $\text{conv}(\mathcal{K})$ be the convex hull of \mathcal{K} , and $\overline{\text{conv}}(\mathcal{K})$ be its closure in the pointwise convergence topology.

For a measure Q on $(\mathcal{X}, \mathcal{A})$ and $\delta > 0$, we denote by $N(\delta, \mathcal{K}, Q)$ the δ -covering number and by $H(\delta, \mathcal{K}, Q)$ the δ -entropy of \mathcal{K} with respect to the $L_2(Q)$ -norm. Formally, for $\mathcal{K} \subset L_2(Q)$, the δ -covering number $N(\delta, \mathcal{K}, Q)$ is defined as the number of $L_2(Q)$ -balls with radius δ , necessary to cover \mathcal{K} . The δ -entropy of \mathcal{K} is $H(\delta, \mathcal{K}, Q) = \log N(\delta, \mathcal{K}, Q)$.

The following theorem, proved in [3], is an example of a relatively simple tool for obtaining the rate of convergence for the Hellinger distance between f_0 and \hat{f}_n in case f_0 belongs to a convex class of densities. For a set of indices \mathcal{Y} and some fixed $k_0(\cdot, \cdot)$, let $\mathcal{K} = \{k_0(\cdot, y) : y \in \mathcal{Y}\}$ be a class of densities on $(\mathcal{X}, \mathcal{A})$ with the envelope function $K := \sup_{k \in \mathcal{K}} k$, and let $f_0 \in \mathcal{F} = \overline{\text{conv}}(\mathcal{K})$. For $\sigma_n \downarrow 0$, let us define the class of functions

$$\tilde{\mathcal{K}}_n = \left\{ \left(\frac{k_0(\cdot, y)}{f_0} \right) \mathbf{1}\{f_0 > \sigma_n\} : y \in \mathcal{Y} \right\},$$

and moreover, let us denote by P_n , the empirical measure based on observations X_1, \dots, X_n (i.e., $P_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$).

Theorem 1. *Assume that for some non-decreasing sequence $\rho_n \geq 1$*

$$\int_{f_0 > \sigma_n} \frac{K^2}{f_0} d\mu \leq \rho_n^2, \quad n = 1, 2, \dots,$$

and

$$\lim_{C \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left(\sup_{0 < \delta < \delta_0} \left(\frac{\delta}{\rho_n} \right)^w N(\delta, \tilde{\mathcal{K}}_n, P_n) > C \right) = 0,$$

for some $0 < w < \infty$ and $\delta_0 > 0$. Then, for

$$\begin{aligned} \tau_n^2 &\geq \int_{f_0 \leq \sigma_n} f_0 d\mu, \quad n = 1, 2, \dots, \\ \tau_n &\geq n^{-(2+w)/(4+4w)} \rho_n^{w/(2+2w)}, \quad n = 1, 2, \dots, \end{aligned}$$

there is

$$h(\hat{f}_n, f_0) = O_P(\tau_n).$$

The following lemma will be used in entropy calculations. Although it is proved in [1], we present another proof, along the lines suggested in [3], because the technique applied will be useful in the next section.

Lemma 1. *Let*

$$\mathcal{G} = \{g : [0, \infty) \rightarrow [0, 1], g \text{ non-increasing}\}.$$

Then there exists a constant C such that for each probabilistic measure Q on $[0, \infty)$,

$$H(\delta, \mathcal{G}, Q) \leq C\delta^{-1}, \text{ for all } \delta > 0.$$

Proof. It is easy to see that

$$\mathcal{G} \subset \overline{\text{conv}}(\mathcal{K}), \tag{1}$$

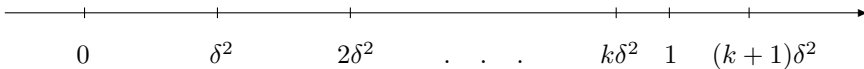
where $\mathcal{K} = \{\mathbf{1}_{[0,y)} : y \in [0, \infty)\}$. It is a consequence of the fact that $\text{conv}(\mathcal{K})$ consists of functions $\sum_{i=1}^n w_i \mathbf{1}_{[y_{i-1}, y_i)}$, where $0 = y_0 < \dots < y_n < \infty$, $1 \geq w_1 > \dots > w_n > 0$ and $n \in \mathbb{N}$, and that any function $g \in \mathcal{G}$ can be approximated by a sequence of functions from $\text{conv}(\mathcal{K})$.

Inclusion (1) implies that $H(\delta, \mathcal{G}, Q) \leq H(\delta, \overline{\text{conv}}(\mathcal{K}), Q)$. Therefore, by the Ball and Pajor Theorem (see, e.g., [4]), it suffices to show that there exists a constant C_1 such that for each probabilistic measure Q

$$N(\delta, \mathcal{K}, Q) \leq C_1 \delta^{-2}.$$

Note that \mathcal{G} is a subset of the ball of radius 1 centered at zero. Hence, for $\delta \geq 1$ the entropy equals 0 and the statement of the lemma holds. Therefore, it is enough to consider $\delta \in (0, 1)$.

If Q has no atoms, i.e., $Q[0, x)$ is a continuous function of x , the δ -covering may be constructed as follows. Take $0 < \delta < 1$ and divide the interval $(0, 1)$ as in the following figure,



where $k\delta^2$ is the maximal multiplicity of δ^2 , which is less than 1.

Therefore,

$$k = \begin{cases} \left\lfloor \frac{1}{\delta^2} \right\rfloor & \text{for } \left\lfloor \frac{1}{\delta^2} \right\rfloor \neq \frac{1}{\delta^2}, \\ \left\lfloor \frac{1}{\delta^2} \right\rfloor - 1 & \text{for } \left\lfloor \frac{1}{\delta^2} \right\rfloor = \frac{1}{\delta^2}, \end{cases}$$

where $\lfloor \cdot \rfloor$ is the floor function. Then we select a set of $k + 2$ points and a set of k functions in the following way

$$\begin{aligned} x_0 &= 0, \\ x_1 : Q[0, x_1) &= \delta^2, \quad f_1(x) := \mathbf{1}_{[0, x_1)}(x), \\ &\vdots \\ x_k : Q[0, x_k) &= k\delta^2, \quad f_k(x) := \mathbf{1}_{[0, x_k)}(x), \\ x_{k+1} &= \infty. \end{aligned}$$

Obviously, for $n = 0, \dots, k$, there is $Q[x_n, x_{n+1}] \leq \delta^2$. Take any $y \in [0, \infty)$. Then, for some $n \in \{0, \dots, k\}$, there is $y \in [x_n, x_{n+1})$, and

$$\|\mathbf{1}_{[0,y)} - f_n\|_{L_2(Q)}^2 = \int \mathbf{1}_{[x_n,y)}^2 dQ = Q[x_n, y] \leq Q[x_n, x_{n+1}] \leq \delta^2.$$

In other words, the $L_2(Q)$ -balls of radius δ , centered at f_1, \dots, f_k cover the class \mathcal{K} , therefore

$$N(\delta, \mathcal{K}, Q) \leq k \leq \delta^{-2}.$$

Now let us consider the general case, when Q is any probabilistic measure. For an arbitrarily chosen δ , we construct, as previously, the sequence of centers, but if for some n there exists no such x that $Q[0, x] = n\delta^2$, then instead of x_n , we take x such, that $Q[0, x] < n\delta^2 < Q[0, x]$. For the chosen points x_1, \dots, x_l , there is $l \leq k$ and $Q(x_n, x_{n+1}) \leq \delta^2$. Let us take $y \in [0, \infty)$. If for some $n \in \{1, \dots, l\}$ $y = x_n$, then $\|\mathbf{1}_{[0,y)} - \mathbf{1}_{[0,x_n)}\|_{L_2(Q)}^2 = 0$. Otherwise, if $y \in (x_n, x_{n+1})$ for some n , then

$$\|\mathbf{1}_{[0,y)} - \mathbf{1}_{[0,x_{n+1})}\|_{L_2(Q)}^2 = \int \mathbf{1}_{[y,x_{n+1})} dQ \leq Q(x_n, x_{n+1}) \leq \delta^2,$$

and, since $l \leq k$, there is $k \leq \delta^2$. □

2. CONVOLUTION MODEL WITH A MONOTONIC KERNEL

Let Y and Z be independent random variables on $[0, 1]$. Suppose that Z has a given density k_0 with respect to the Lebesgue measure. The distribution θ of Y is unknown. We observe independent copies X_1, \dots, X_n of $X = Z + Y$. Therefore,

$$f_0 \in \mathcal{F} = \left\{ \int_0^1 k_0(\cdot - y) d\theta(y) : \theta \in \Theta \right\},$$

where Θ is the class of all probabilistic measures on $[0, 1]$. If we put $\mathcal{K} = \{k_0(\cdot - y) : y \in [0, 1]\}$, then $\mathcal{F} = \overline{\text{conv}}(\mathcal{K})$ (see [3]). In this section, the special case of a monotonic kernel $k_0(x) = 2x\mathbf{1}\{0 \leq x \leq 1\}$ will be handled. As in [3], in order to simplify the analysis of the shape of f_0 , we assume that θ is the uniform distribution (a more general case, when θ has a density bounded away from zero and infinity gives similar results).

We want to apply Theorem 1, so we need to calculate the covering number of $\tilde{\mathcal{K}}_n = \{(k/f_0)\mathbf{1}\{f_0 > \sigma_n\} : k \in \mathcal{K}\}$. With θ being the uniform distribution, one obtains

$$f_0(x) = \begin{cases} x^2 & \text{for } 0 \leq x \leq 1; \\ x(2-x) & \text{for } 1 \leq x \leq 2; \\ 0 & \text{otherwise} \end{cases}$$

and it is convenient to obtain the covering numbers in two steps.

Define

$$\tilde{\mathcal{K}}_n^{(1)} = \left\{ \tilde{k} \mathbf{1}_{[0,1]} : \tilde{k} \in \tilde{\mathcal{K}}_n \right\}, \quad \tilde{\mathcal{K}}_n^{(2)} = \left\{ \tilde{k} \mathbf{1}_{[1,2]} : \tilde{k} \in \tilde{\mathcal{K}}_n \right\}.$$

Lemma 2 (see [3]). *There exists a constant A_1 such that*

$$N(\delta, \tilde{\mathcal{K}}_n^{(1)}, P_n) \leq A_1 \delta^{-1}, \quad \text{for all } \delta \in (0, 1) \text{ a.s.,}$$

for each n sufficiently large.

In order to calculate the δ -covering number for the class $\tilde{\mathcal{K}}_n^{(2)}$, let us deal with the class \mathcal{K} first. It is asserted in [3] that there exists a constant C such that for any probabilistic measure on $[0, 2]$ there is $N(\delta, \mathcal{K}, Q) \leq C\delta^{-1}$. The suggested line of the proof is, however, incorrect (it is asserted that such an inequality holds true for the δ -covering number in the supremum norm. However, it cannot be true, because for any $k_1 \neq k_2 \in \mathcal{K}$, there is $\|k_1 - k_2\|_\infty = 2$ and, hence, for $\delta < 1$, there follows $N_\infty(\delta, \mathcal{K}) = \infty$).

The following lemma gives a corrected upper bound for the covering number.

Lemma 3. *There exists a constant A_0 such that for any probabilistic measure Q on $[0, 2]$,*

$$N(\delta, \mathcal{K}, Q) \leq A_0 \delta^{-2}, \quad \text{for all } \delta \in (0, 1). \tag{2}$$

Proof. Take $\delta \in (0, 1)$ and define $\tilde{k}_0(x) := 2x \mathbf{1}\{0 \leq x < 1\}$. Let y_i , $i = 1, \dots, N$, be points chosen in such a way that $Q(1 + y_{i-1}, 1 + y_i) \leq \delta^2$, $i = 2, \dots, N$ (the proof of Lemma 1 implies that $N < 1/\delta^2$). Moreover, let $y_{N+k} := k\delta$, for $k = 1, \dots, \lfloor 1/\delta \rfloor$, $y_0 := 0$, and $y_{N+\lfloor 1/\delta \rfloor+1} := 1$. For simplicity, we assume that the points y_i are arranged increasingly. Obviously, for $i = 1, \dots, N + \lfloor 1/\delta \rfloor + 1$,

$$y_i - y_{i-1} \leq \delta \quad \text{and} \quad Q(1 + y_{i-1}, 1 + y_i) \leq \delta^2. \tag{3}$$

As the centers of the balls for the δ -covering of the class \mathcal{K} , we take $\tilde{k}_0(\cdot - y_i)$ and $k_0(\cdot - y_i)$, for $i = 0, \dots, N + \lfloor 1/\delta \rfloor + 1$. Since

$$2 \left(N + \left\lfloor \frac{1}{\delta} \right\rfloor + 2 \right) \leq \frac{8}{\delta^2} \quad \text{for } \delta \in (0, 1), \tag{4}$$

it suffices to show that the balls cover \mathcal{K} . Take $y \in [0, 1]$ such that $y \neq y_i$ for all i (otherwise, $k_0(\cdot - y)$ is one of the chosen centers). Since $y \in (y_{i-1}, y_i)$ for some $i \in \{1, \dots, N + \lfloor 1/\delta \rfloor + 1\}$, there is

$$\begin{aligned} \int_{[0,2]} \left[k_0(x - y) - \tilde{k}_0(x - y_i) \right]^2 dQ(x) &\leq \\ &\leq \int_{[y, 1+y]} 4(y_i - y_{i-1})^2 dQ(x) + \int_{(1+y, 1+y_i)} 4 dQ(x) \leq \\ &\leq 4(y_i - y_{i-1})^2 + 4Q(1 + y_{i-1}, 1 + y_i) \leq 8\delta^2, \end{aligned}$$

for this i , because of (3). In view of inequality (4), it follows that $N(\sqrt{8}\delta, \mathcal{K}, Q) \leq 8\delta^{-2}$ for $\delta \in (0, 1)$. Hence, $N(\delta, \mathcal{K}, Q) \leq 64\delta^{-2}$ for $\delta \in (0, 1)$. \square

Note that (2) holds true for all finite (not necessarily probabilistic) measures and apply Lemma 3 with $dQ = ((1/f_0^2)\mathbf{1}\{f_0 > \sigma_n\}\mathbf{1}_{[1,2]}dP_n)/A^2\rho_n^2$, to obtain

$$N(\delta, \tilde{\mathcal{K}}_n^{(2)}, P_n) \leq A^2 A_0 \left(\frac{\rho_n}{\delta}\right)^2, \quad \text{for all } \delta \in (0, 1), \tag{5}$$

on the set

$$\left\{ \int_{f_0 > \sigma_n} \frac{1}{f_0^2} \mathbf{1}_{[1,2]} dP_n \leq A^2 \rho_n^2 \right\}.$$

So, for

$$\int_{f_0 > \sigma_n} \frac{1}{f_0} \mathbf{1}_{[1,2]} dx \leq \rho_n^2, \tag{6}$$

there is

$$\begin{aligned} \limsup_{n \rightarrow \infty} P \left(\sup_{0 < \delta < 1} \left(\frac{\delta}{\rho_n}\right)^2 N(\delta, \tilde{\mathcal{K}}_n^{(2)}, P_n) > A_0 A^2 \right) &\leq \\ &\leq \limsup_{n \rightarrow \infty} P \left(\int \frac{1}{f_0^2 \rho_n^2} \mathbf{1}\{f_0 > \sigma_n\} \mathbf{1}_{[1,2]} dP_n > A^2 \right) \longrightarrow 0, \quad \text{as } A \rightarrow \infty. \end{aligned}$$

Because of Lemma 2, if (6) holds, we can write

$$\lim_{A \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left(\sup_{0 < \delta < 1} \left(\frac{\delta}{\rho_n}\right)^2 N(\delta, \tilde{\mathcal{K}}_n^{(i)}, P_n) > A \right) = 0,$$

for $i = 1, 2$.

Some effort is needed to see that the above remains true for the whole class \mathcal{K} . To this end, it will be shown that

$$N(\delta, \tilde{\mathcal{K}}_n, Q) \leq N(\delta, \tilde{\mathcal{K}}_n^{(1)}, Q) + N(\delta, \tilde{\mathcal{K}}_n^{(2)}, Q), \quad \text{for all } \delta \in (0, 1). \tag{7}$$

Notice that the functions from $\tilde{\mathcal{K}}_n$ are continuous at $x = 1$ and can be obtained as 'junctions' of the functions from $\tilde{\mathcal{K}}_n^{(1)}$ and $\tilde{\mathcal{K}}_n^{(2)}$. It is not hard to verify that the balls covering the classes $\tilde{\mathcal{K}}_n^{(1)}$ and $\tilde{\mathcal{K}}_n^{(2)}$ can be represented in \mathbb{R}^2 as sets, bounded by two functions from the corresponding class.

Therefore, if we construct the centers of the balls for the covering of $\tilde{\mathcal{K}}_n$ as 'junctions' of the centers of the balls from the coverings of $\tilde{\mathcal{K}}_n^{(1)}$ and $\tilde{\mathcal{K}}_n^{(2)}$, it is sufficient to choose those pairs of centers only for which the representations of the corresponding balls do touch each other at $x = 1$. The number of such pairs is less then the sum of the numbers of balls covering the sets $\tilde{\mathcal{K}}_n^{(1)}$ and $\tilde{\mathcal{K}}_n^{(2)}$, so that (7) holds true (see Fig. 1).

From that, for the whole class $\tilde{\mathcal{K}}_n$, we obtain

$$\lim_{A \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left(\sup_{0 < \delta < 1} \left(\frac{\delta}{\rho_n}\right)^2 N(\delta, \tilde{\mathcal{K}}_n, P_n) > A \right) = 0.$$

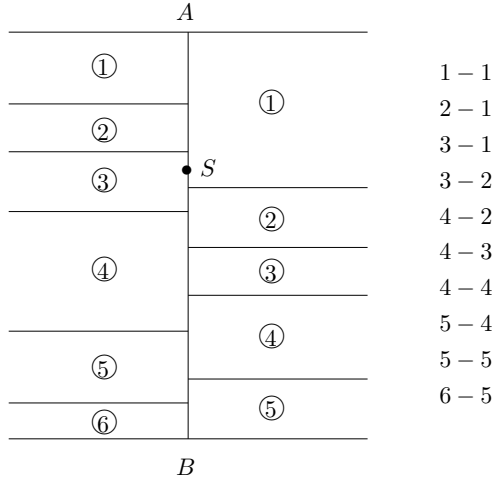


Fig. 1. Schematic representation of the balls covering $\tilde{\mathcal{K}}_n^{(1)}$ (to the left of AB , which corresponds to $x = 1$) and $\tilde{\mathcal{K}}_n^{(2)}$ (to the right of AB). In order to construct a center of the ball for the covering of $\tilde{\mathcal{K}}_n$, two centers are joined to form a (not necessarily continuous) function on $[0, 2]$: one from the covering of $\tilde{\mathcal{K}}_n^{(1)}$ and one from the covering of $\tilde{\mathcal{K}}_n^{(2)}$. For example, a (continuous) function from $\tilde{\mathcal{K}}_n$ that crosses the $x = 1$ line at the point S would belong to the ball centered at the junction 3 - 1. On the right, the list of junctions sufficient to form a covering of $\tilde{\mathcal{K}}_n$ in this particular configuration. Obviously, the covering number of $\tilde{\mathcal{K}}_n$ is not greater than $N(\delta, \tilde{\mathcal{K}}_n^{(1)}, Q) + N(\delta, \tilde{\mathcal{K}}_n^{(2)}, Q) - 1$

The envelope function of the class \mathcal{K} takes the form

$$K(x) = 2x\mathbf{1}_{[0,1)}(x) + 2\mathbf{1}_{[1,2]}(x).$$

Hence, using the specific form of f_0 ,

$$\int_{f_0 > \sigma_n} \frac{K^2}{f_0} dx = 4 - 4\sqrt{\sigma_n} + 2 \log \left| \frac{\sqrt{1 - \sigma_n} + 1}{\sqrt{1 - \sigma_n} - 1} \right| \asymp \log \frac{1}{\sigma_n}$$

and

$$\int_{f_0 \leq \sigma_n} f_0 dx = \frac{1}{3}\sigma_n^{3/2} + \frac{2}{3} - \sqrt{1 - \sigma_n} + \frac{1}{3}(1 - \sigma_n)^{3/2} = \frac{1}{3}\sigma_n^{3/2} + o(\sigma_n^{3/2}).$$

Because

$$\int_{f_0 > \sigma_n} \frac{1}{f_0} \mathbf{1}_{[1,2]} dx = \frac{1}{2} \log \left| \frac{\sqrt{1 - \sigma_n} + 1}{\sqrt{1 - \sigma_n} - 1} \right| \asymp \log \frac{1}{\sigma_n},$$

in order to satisfy condition (6) and the assumptions of Theorem 1, we need to hold

$$\rho_n^2 \geq A \log \frac{1}{\sigma_n}, \quad \tau_n^2 \geq B\sigma_n^{3/2}, \quad \text{and} \quad \tau_n \geq Cn^{-1/3} \rho_n^{1/3},$$

with suitably chosen constants. So, with the optimal $\sigma_n \asymp n^{-4/9}$, we arrive at the rate

$$h(\hat{f}_n, f_0) = O_P\left(n^{-1/3}(\log n)^{1/6}\right).$$

Note that the rate asserted in [3] was $O_P(n^{-3/8}(\log n)^{1/8})$, but that result does not seem to be correct, because of the faulty proof of Lemma 3 in [3].

3. CONVOLUTION MODEL WITH A STRICTLY CONVEX KERNEL

Let us now consider the convolution model with a strictly convex kernel

$$k_0(x) = [3 - 12x(1 - x)] \mathbf{1}_{[0,1]}(x),$$

which was studied in [2]. Again, the rate $O_P(n^{-3/8}(\log n)^{1/8})$, asserted in [2], does not seem to be correct, because the δ -covering number for the class \mathcal{K} cannot be of the order δ^{-1} (k_0 is discontinuous at 0 and 1).

For $y_1 < y_2$, one has

$$\begin{aligned} & \int (k_0(\cdot - y_2) - k_0(\cdot - y_1))^2 dQ = \\ &= \int_{[y_2, 1+y_1]} (k_0(\cdot - y_2) - k_0(\cdot - y_1))^2 dQ + \int_{[y_1, y_2]} k_0^2(\cdot - y_1) dQ + \int_{(1+y_1, 1+y_2]} k_0^2(\cdot - y_2) dQ \leq \\ & \leq [36(y_2 - y_1)]^2 + 9Q[y_1, y_2] + 9Q(1 + y_1, 1 + y_2]. \end{aligned}$$

Hence, reasoning as in the proof of Lemma 3, one can easily see that, for some constant A and for any probabilistic measure Q on $[0, 2]$,

$$N(\delta, \mathcal{K}, Q) \leq A\delta^{-2} \quad \text{for all } \delta \in (0, 1). \quad (8)$$

Let us assume that θ has a density g_0 with respect to the Lebesgue measure, and that, for some constant $c_1 > 0$,

$$\frac{1}{c_1} \leq g_0(y) \leq c_1, \quad \text{for all } y \in [0, 1]. \quad (9)$$

Then,

$$\int_{f_0 > \sigma_n} \frac{K^2(x)}{f_0(x)} dx = 9 \int_{f_0 > \sigma_n} \frac{1}{f_0(x)} dx \asymp c_2 \log\left(\frac{1}{\sigma_n}\right), \quad (10)$$

and

$$\int_{f_0 \leq \sigma_n} f_0(x) dx \geq c_3 \sigma_n^2, \quad (11)$$

for some suitable, strictly positive constants c_2 and c_3 depending on c_1 .

Using (8) with $dQ = dP_n(1/f_0^2)\mathbf{1}\{f_0 > \sigma_n\}/(C\rho_n^2)$, one obtains

$$N(\delta, \tilde{\mathcal{K}}_n, P_n) \leq AC \left(\frac{\rho_n}{\delta}\right)^2 \quad \text{for all } \delta \in (0, 1),$$

on the set

$$\left\{ \int_{f_0 > \sigma_n} \frac{1}{f_0^2} dP_n \leq C\rho_n^2 \right\}.$$

So, for

$$\int_{f_0 > \sigma_n} \frac{1}{f_0} dx \leq \rho_n^2, \tag{12}$$

there is

$$\lim_{C \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left(\sup_{0 < \delta < 1} \left(\frac{\delta}{\rho_n}\right)^2 N(\delta, \tilde{\mathcal{K}}_n, P_n) > C \right) = 0.$$

In view of (10), (11) and (12), the following inequalities must hold, if we want to apply Theorem 1

$$\rho_n^2 \geq c_4 \log \frac{1}{\sigma_n}, \quad \tau_n^2 \geq c_3 \sigma_n^2, \quad \text{and} \quad \tau_n \geq n^{-1/3} \rho_n^{1/3}.$$

Hence, again, we arrive at the rate

$$h(\hat{f}_n, f_0) = O_P \left(n^{-1/3} (\log n)^{1/6} \right),$$

this time with the optimal $\sigma_n \asymp n^{-1/3}$.

Acknowledgments

The author is greatly indebted to Z. Szkutnik for his valuable remarks.

REFERENCES

[1] Birman M. S., Solomjak A., *Piecewise-polynomial approximations of functions of the classes W_p^α* , Mat. Sb. **73** (1967), 295–317.
 [2] Geer S. Van De, *Empirical Process in M-Estimation*, Cambridge: Cambridge University Press, 2000.
 [3] Geer S. Van De, *Rates of convergence for the maximum likelihood estimator in mixture models*, Nonparametric Statistics **6** (1996), 293–310.
 [4] Vaart A.W. Van der, Wellner J. A., *Weak Convergence & Empirical Process*. New York, Springer 1996.

Piotr Majerski
majerski@uci.agh.edu.pl

AGH University of Science and Technology
Faculty of Applied Mathematics
al. Mickiewicza 30, 30-059 Kraków, Poland

Received: February 16, 2005.