Science in China Series A: Mathematics
Jul., 2009, Vol. 52, No. 7, 1525–1538
www.scichina.com math.scichina.com
www.springerlink.com

© 2009 SCIENCE IN CHINA PRESS
Springer

On the rate of convergence of the maximum likelihood estimator of a *k*-monotone density

GAO FuChang[1†] & WELLNER Jon A[2]

[1] Department of Mathematics, University of Idaho, Moscow, ID 83844-1103, USA
[2] Department of Statistics, University of Washington, Seattle, WA 98195-4322, USA
(email: fuchang@uidaho.edu, jaw@stat.washington.edu)

Abstract Bounds for the bracketing entropy of the classes of bounded k-monotone functions on $[0, A]$ are obtained under both the Hellinger distance and the $L^p(Q)$ distance, where $1 \leqslant p < \infty$ and Q is a probability measure on $[0, A]$. The result is then applied to obtain the rate of convergence of the maximum likelihood estimator of a k-monotone density.

Keywords: **bracketing metric entropy, maximum likelihood estimator, Hellinger distance**

MSC(2000): **62G05, 46B50**

1 Introduction

A function on $(0, \infty)$ is called k-monotone if $(-1)^j f^{(j)}(x)$ is non-negative, non-increasing, and convex for $0 \leqslant j \leqslant k-2$ if $k \geqslant 2$, and f is non-negative, non-increasing if $k = 1$. These functions fill the gap between monotone functions and completely monotone functions. They appear very commonly in nonparametric estimation, such as the Maximum Likelihood Estimator (MLE) in statistics via renewal theory and mixing of uniform distributions. Indeed, k-monotone functions have been studied since at least the 1950s; for example, Williamson[1] gave a characterization of m-monotone functions on $(0, \infty)$ in 1956. In recent years, there has been some interest in statistics regarding this class of functions. We refer to [2] and the references therein for recent results and their statistical applications.

Note that a k-monotone function may not be bounded near $t = 0$. In order to study the metric entropy, we restrict ourselves to the subclass that consists of only the functions that are continuous at $t = 0$. We refer to this subclass as the class of k-monotone functions on $[0, \infty)$. We denote by $\mathcal{M}_k(I)$ the class of k-monotone functions on I, and by $\mathcal{F}_k(I)$ the class of probability densities on I that are k-monotone.

For statistical applications, we wish to estimate the bracketing entropy of $\mathcal{F}_k(\mathbb{R}^+)$ and $\mathcal{M}_k(\mathbb{R}^+)$ under all $L^p(Q)$ distances, where $1 \leqslant p < \infty$ and Q is any probability measure on $\mathbb{R}^+$, and under the Hellinger distance h which is defined by

$$h(f, g) = \left(\int [\sqrt{f(x)} - \sqrt{g(x)}]^2 dx \right)^{1/2}. \tag{1}$$

Received October 15, 2008; accepted April 22, 2009
DOI: 10.1007/s11425-009-0102-y
† Corresponding author
This work was supported by National Science Foundation of USA (Grant No. DMS-0405855, DMS-0804587)

Citation: Gao F C, Wellner J A. On the rate of convergence of the maximum likelihood estimator of a k-monotone density. Sci China Ser A, 2009, 52(7): 1525–1538, DOI: 10.1007/s11425-009-0102-y

Recall that the bracketing metric entropy of a function class $\mathcal{F}$ under distance ρ is defined as $\log N_{[\cdot]}(\varepsilon, \mathcal{F}, \rho)$, where $N_{[\cdot]}(\varepsilon, \mathcal{F}, \rho)$ is defined by

$$N_{[\cdot]}(\varepsilon, \mathcal{F}, \rho) := \min\left\{n : \exists \underline{f}_1, \overline{f}_1, \dots, \underline{f}_n, \overline{f}_n \text{ s.t. } \rho(\overline{f}_k, \underline{f}_k) \leqslant \varepsilon, \mathcal{F} \subset \bigcup_{k=1}^{n} [\underline{f}_k, \overline{f}_k]\right\},$$

where

$$[\underline{f}_k, \overline{f}_k] = \{g \in \mathcal{F} : \underline{f}_k \leqslant g \leqslant \overline{f}_k\}.$$

It is easy to see that both $\mathcal{F}_k(\mathbb{R}^+)$ and $\mathcal{M}_k(\mathbb{R}^+)$ are bounded under the Hellinger distance. However, they are not compact. Indeed, for any $\delta < \sqrt{2}$, we can find infinitely many functions in $\mathcal{F}_k(\mathbb{R}^+)$ with mutual Hellinger distance at least δ. In fact, for any $\alpha > 0$, the functions $p_n(t) = 2^{n\alpha} e^{-2^{n\alpha} t}$ are clearly in $\mathcal{F}_k(\mathbb{R}^+)$. For $m > n$,

$$\begin{aligned}\int_0^\infty [\sqrt{p_n(t)} - \sqrt{p_m(t)}]^2 dt &= 2 - \int_0^\infty 2\sqrt{2^{n\alpha+m\alpha}} e^{-(2^{n\alpha}+2^{m\alpha})t/2} dt \\ &= 2 - 4\frac{\sqrt{2^{n\alpha+m\alpha}}}{2^{n\alpha}+2^{m\alpha}} = 2 - 4\frac{2^{-(m-n)\alpha/2}}{1+2^{-(m-n)\alpha}} \\ &\geqslant 2 - 4\frac{2^{-\alpha/2}}{1+2^{-\alpha}} = \delta^2\end{aligned}$$

for

$$\alpha = 2\log_2\left(\frac{2+\delta\sqrt{4-\delta^2}}{2-\delta^2}\right).$$

Note that the sequence $\{p_n(t)\}_{n\geqslant 1}$ is unbounded near the origin. This suggests us that for the Hellinger distance, we need to restrict ourselves to k-monotone functions whose values are bounded near the origin. However, the sequence $\{p_n(t)\}_{n<0}$ is uniformly bounded by 1. Thus, this example also indicates that the non-compactness of $\mathcal{F}_k(\mathbb{R}^+)$ under the Hellinger distance is partly due to the fact that the interval is unbounded. Hence, we should also restrict ourselves to bounded intervals. Therefore, in what follows, we consider the subclasses $\mathcal{F}_k^B([0, A])$ and $\mathcal{M}_k^B([0, A])$ instead, where $\mathcal{F}_k^B(I)$ and $\mathcal{M}_k^B(I)$ denote the classes of functions that are bounded by B and belong to $\mathcal{F}_k(I)$ and $\mathcal{M}_k(I)$ respectively.

By changing variables, it is easy to see that

$$\begin{aligned}N_{[\cdot]}(\varepsilon, \mathcal{F}_k^B([0, A]), h) &= N_{[\cdot]}(\varepsilon, \mathcal{F}_k^{AB}([0, 1]), h), \\ N_{[\cdot]}(\varepsilon, \mathcal{M}_k^B([0, A]), h) &= N_{[\cdot]}(\varepsilon, \mathcal{M}_k^{AB}([0, 1]), h).\end{aligned}$$

Hence, we only need to consider the case $A = 1$.

Let us remark that when $k = 1$, the problem has been studied by Van de Geer[3] based on an earlier work of Birman and Solomjak[4]. For example, it was proved that (see also [5]; Theorem 2.7.5)

$$\log N_{[\cdot]}(\varepsilon, \mathcal{M}_1^B([0, 1]), h) \leqslant CB\varepsilon^{-1} \tag{2}$$

for some absolute constant $C > 0$, and

$$\log N_{[\cdot]}(\varepsilon, \mathcal{M}_1^B(\mathbb{R}^+), \|\cdot\|_{p,Q}) \leqslant C_p B\varepsilon^{-1}, \tag{3}$$

for some positive constant C_p depending only on p, where $1 \leqslant p < \infty$, and Q is any probability measure on $\mathbb{R}^+$. For simpler proofs, see also [6, 7]. In particular, the iteration method used in [6] is useful in our argument in this paper.

For $k > 1$, Gao[8] also established the following metric entropy bound for $\mathcal{M}_k^B([0,1])$:

$$C_1 B^{1/k} \varepsilon^{-1/k} \leqslant \log N(\varepsilon, \mathcal{M}_k^B([0,1]), \|\cdot\|_2) \leqslant C_2 B^{1/k} \varepsilon^{-1/k}. \tag{4}$$

The method revealed a nice connection between the metric entropy of these function classes and the small ball probability of k-times integrated Brownian motions. However, because for $k > 1$ the square root of a k-monotone function may not be k-monotone, the metric entropy estimate under L^2 distance does not yield an estimate under the Hellinger distance. Furthermore, that method cannot produce any result on bracketing metric entropy. Thus, it cannot be readily used to determine the convergence rate of the MLE of a k-monotone density.

In this paper, we directly estimate the bracketing metric entropy of these function classes under the Hellinger distance and under all $L^p(Q)$ distances, where $1 \leqslant p < \infty$, and apply these estimates to statistical settings.

Our main tool is the following lemma, which provides a useful method to estimate bracketing entropy. An extension to more general integral operators will appear elsewhere.

Lemma 1. *Let $\mathcal{F}$ be a class of functions on $[0,1]$, and $\mathcal{G}$ be the class of function on $[0,1]$ defined by $\mathcal{G} = \{\int_0^{\alpha(x)} f(t)dt : f \in \mathcal{F}\}$, where $0 \leqslant \alpha(x) \leqslant 1$ is any increasing function on $[0,1]$. If $\log N(\varepsilon, \mathcal{F}, \|\cdot\|_1) \leqslant \phi(\varepsilon)$, where $\|\cdot\|_1$ stands for the L^1 distance under the Lebesgue measure on $[0,1]$, then*

(i) *There exists a constant C depending only on p, such that for any probability measure Q on $[0,1]$*

$$\log N_{[\cdot]}\left(\frac{\varepsilon}{\phi(\varepsilon)}, \mathcal{G}, \|\cdot\|_{p,Q}\right) \leqslant C\phi(\varepsilon),$$

where $\|\cdot\|_{p,Q}$ is defined by

$$\|f-g\|_{p,Q} = \left(\int_0^1 |f(x)-g(x)|^p dQ(x)\right)^{1/p}.$$

(ii) *If we further assume that for all functions in $g \in \mathcal{G}$, $g(x) \geqslant \delta$, then there exists a constant C, such that*

$$\log N_{[\cdot]}\left(\frac{\varepsilon}{2\sqrt{\delta}\phi(\varepsilon)}, \mathcal{G}, h\right) \leqslant C\phi(\varepsilon),$$

where $h(f,g)$ is defined by (1).

Proof. Let $\{f_i\}$, $1 \leqslant i \leqslant e^{\phi(\varepsilon)}$, be an ε-net for $\mathcal{F}$ in the L^1 distance under the Lebesgue measure on $[0,1]$. For each i, and $f \in \mathcal{F}$, we can write

$$\int_0^{\alpha(x)} f(t)dt = \int_0^{\alpha(x)} (f(t)-f_i(t))_+ dt - \int_0^{\alpha(x)} (f_i(t)-f(t))_+ dt + \int_0^{\alpha(x)} f_i(t)dt.$$

Thus, if we define

$$\mathcal{G}_i^+ = \left\{g : g(x) = \int_0^{\alpha(x)} (f(t)-f_i(t))_+ dt, f \in \mathcal{F}, \|f-f_i\|_1 \leqslant \varepsilon\right\},$$

$$\mathcal{G}_i^- = \left\{ g : g(x) = \int_0^{\alpha(x)} (f_i(t) - f(t))_+ dt, f \in \mathcal{F}, \|f - f_i\|_1 \leqslant \varepsilon \right\},$$

we have

$$\mathcal{G} \subset \bigcup_i \left(\mathcal{G}_i^+ - \mathcal{G}_i^- + \int_0^{\alpha(x)} f_i(t) dt \right).$$

Note that $\mathcal{G}_i^+$ and $\mathcal{G}_i^-$ both consist of non-negative increasing functions bounded by ε. Thus, by (3) we can find $e^{c\phi(\varepsilon)}$ many $\varepsilon/\phi(\varepsilon)$-brackets (with respect to $\|\cdot\|_{p,Q}$) that cover $\mathcal{G}_i^+ - \mathcal{G}_i^-$. Say these brackets are $[\underline{h}_{i,j}, \overline{h}_{i,j}]$, $1 \leqslant j \leqslant e^{c\phi(\varepsilon)}$. Then clearly the brackets

$$\left[\int_0^{\alpha(x)} f_i(t)dt + \underline{h}_{i,j}, \int_0^{\alpha(x)} f_i(t)dt + \overline{h}_{i,j} \right], \quad 1 \leqslant i \leqslant e^{\phi(\varepsilon)}, \quad 1 \leqslant j \leqslant e^{c\phi(\varepsilon)}$$

cover $\mathcal{G}$. Statement (i) follows by noticing that these are $\varepsilon/\phi(\varepsilon)$-brackets under the $\|\cdot\|_{p,Q}$ distance.

To prove Statement (ii), we notice that with the additional assumption $g \geqslant \delta$, we have for any $g_1, g_2 \in \mathcal{G}$,

$$h(g_1, g_2) = \|\sqrt{g_1} - \sqrt{g_2}\|_2 \leqslant \frac{1}{2\sqrt{\delta}} \|g_1 - g_2\|_2.$$

Hence,

$$N_{[\cdot]}\left(\frac{\varepsilon}{2\sqrt{\delta}\phi(\varepsilon)}, \mathcal{G}, h \right) \leqslant N_{[\cdot]}\left(\frac{\varepsilon}{\phi(\varepsilon)}, \mathcal{G}, \|\cdot\|_2 \right).$$

Thus, Statement (ii) follows from Statement (i).

2 Under Hellinger distance

Note that by scaling arguments, we can easily show that

$$N_{[\cdot]}(\varepsilon, \mathcal{M}_k^B([0, A]), h) = N_{[\cdot]}(\varepsilon, \mathcal{M}_k^{AB}([0, 1]), h) = N_{[\cdot]}(\varepsilon/\sqrt{AB}, \mathcal{M}_k^1([0, 1]), h),$$
$$N_{[\cdot]}(\varepsilon, \mathcal{F}_k^B([0, A]), h) = N_{[\cdot]}(\varepsilon, \mathcal{F}_k^{AB}([0, 1]), h).$$

Hence, we only need to consider the classes $\mathcal{M}_k^1([0, 1])$ and $\mathcal{F}_k^B([0, 1])$.

Before we process with the detailed calculation, we make some observations that can simplify the later arguments. Firstly, because k-monotone C^∞ functions are dense in $\mathcal{M}_k^1([0, 1])$ (cf. [8]), we can and will assume that all the densities in $\mathcal{M}_k^1([0, 1])$ are continuously k-times differentiable. Secondly, if for $I = [a, b] \subset [0, 1]$ we define

$$\mathcal{H}_k^B(I) = \{f : f(u) = g(b - u), g \in \mathcal{M}_k^B(I)\}.$$

then for every $f \in \mathcal{H}_k^1([0, 1])$,

$$\frac{d^j f(u)}{du^j} = \frac{d^j}{du^j}(g(1 - u)) = (-1)^j g^{(j)}(1 - u) \geqslant 0$$

for all $0 \leqslant j \leqslant k$, and for all $u \in [0, 1]$. For $f \in \mathcal{H}_k^1([0, 1])$, we can write

$$\begin{aligned} f(u) = & f(0) + f'(0)u + \cdots + \frac{f^{(k-2)}(0)}{(k-2)!} u^{k-2} \\ & + \int_0^u \int_0^{t_{k-2}} \cdots \int_0^{t_1} f^{(k-1)}(s) ds dt_1 \cdots dt_{k-2}. \end{aligned} \tag{5}$$

All the terms on the right-hand side are non-negative. The sum of the first $k-1$ terms is a polynomial of u with degree $k-2$, and non-negative coefficients.

2.1 Bounded k-monotone functions

For $f \in \mathcal{H}_k^1([0,1])$, because $f(1) \leqslant 1$, we have

$$\sum_{k=0}^{k-2} f^{(k)}(0) \leqslant f(1) \leqslant 1.$$

Denote

$$\mathcal{P}_k = \left\{ a_0 + a_1 u + \cdots + a_{k-2} u^{k-2} : a_0, a_1, \ldots, a_{k-2} \geqslant 0; \sum_{i=0}^{k-2} a_i \leqslant 1 \right\},$$

and denote by $\widetilde{\mathcal{H}}_k^B(I)$ the class of functions g on $I \subset [0,1]$ that satisfy $0 \leqslant g \leqslant 1$ and are of the form

$$g(u) = \int_0^u \int_0^{t_{k-2}} \cdots \int_0^{t_1} f(s) ds dt_1 \cdots dt_{k-2},$$

where f are non-negative increasing functions on I. By (5), we have $\mathcal{H}_k^1([0,1]) \subset \mathcal{P}_k + \widetilde{\mathcal{H}}_k^1([0,1])$, and thus

$$N_{[\cdot]}(\varepsilon, \mathcal{H}_k^1([0,1]), h) \leqslant N_{[\cdot]}(\varepsilon/2, \mathcal{P}_k, h) \cdot N_{[\cdot]}(\varepsilon/2, \widetilde{\mathcal{H}}_k^1([0,1]), h). \tag{6}$$

Note that the set

$$\left\{ a_0 + a_1 u + \cdots + a_{k-2} u^{k-2} : a_i \in \left\{ \frac{1}{N}, \frac{2}{N}, \ldots, \frac{N}{N} \right\}, 0 \leqslant i \leqslant k-2 \right\}$$

forms a $\sqrt{k/N}$-net for $\mathcal{P}_k$ under the Hellinger distance. Indeed, for any $p = a_0 + a_1 u + \cdots + a_{k-2} u^{k-2} \in \mathcal{P}_m$, choose

$$\tilde{p} = \frac{\lceil a_0 N \rceil}{N} + \frac{\lceil a_1 N \rceil}{N} u + \cdots + \frac{\lceil a_{k-2} N \rceil}{N} u^{k-2}.$$

Then

$$|\sqrt{p} - \sqrt{\tilde{p}}|^2 \leqslant |p - \tilde{p}| \leqslant k/N,$$

which implies that $h(p, \tilde{p}) \leqslant \sqrt{k/N}$. Note that there are no more than N^{k-1} elements in this set. By choosing $N = \lceil 4k\varepsilon^{-2} \rceil$, we obtain

$$\log N_{[\cdot]}(\varepsilon/2, \mathcal{P}_k, h) \leqslant k \log(1 + 4k/\varepsilon^2). \tag{7}$$

Of course, because $\|p - \tilde{p}\|_2 \leqslant k/N$, we also have

$$\log N_{[\cdot]}(\varepsilon/2, \mathcal{P}_k, \|\cdot\|_2) \leqslant k \log(1 + 2k/\varepsilon). \tag{8}$$

Our next goal is to estimate $N_{[\cdot]}(\varepsilon/2, \widetilde{\mathcal{H}}_k^1([0,1]), h)$. To this end, we first consider

$$N_{[\cdot]}(\varepsilon/2, \widetilde{\mathcal{H}}_k^1([0,1]), \|\cdot\|_2).$$

For $g \in \widetilde{\mathcal{H}}_k^1([0,1])$, we have

$$\begin{aligned} g(1) &= \int_0^1 \int_0^{t_{k-2}} \cdots \int_0^{t_1} f(s) ds dt_1 \cdots dt_{k-2} du \\ &\geqslant \int_{1/2}^1 \int_{1/2}^{t_{k-2}} \cdots \int_{1/2}^{t_1} f(s) ds dt_1 \cdots dt_{k-2} du \\ &\geqslant f(1/2) \int_{1/2}^1 \int_{1/2}^{t_{k-2}} \cdots \int_{1/2}^{t_1} ds dt_1 \cdots dt_{k-2} du \\ &= f(1/2) \frac{1}{2^{k-1}(k-1)!}. \end{aligned}$$

Because $g(1)$ is bounded by 1, and f is increasing, for all $0 \leqslant u \leqslant 1/2$ we have

$$f(u) \leqslant f(1/2) \leqslant 2^{k-1}(k-1)!.$$

Define $f_1 = \min\{f, 2^{k-1}(k-1)!\}$ and $f_2 = f - f_1$. Then, by the above argument we see that f_2 is non-negative and increasing, and is supported on $[1/2, 1]$. For $g \in \widetilde{\mathcal{H}}_k^1([0,1])$ and $0 \leqslant u \leqslant 1$, we can write

$$\begin{aligned} g(u) &= \int_0^u \int_0^{t_{k-2}} \cdots \int_0^{t_1} [f_1(s) + f_2(s)] ds dt_1 \cdots dt_{k-2} \\ &= \int_0^u \cdots \int_0^{t_1} f_1(s) ds dt_1 \cdots dt_{k-2} + 1_{[1/2,1]}(u) \int_0^{2u-1} \cdots \int_0^{t_1} \frac{1}{2} f_2(1/2 + s/2) ds \cdots dt_{k-2}. \end{aligned}$$

We construct two function classes:

$$\begin{aligned} \mathcal{U}_k &= \left\{ \int_0^u \int_0^{t_{k-2}} \cdots \int_0^{t_1} f(s) ds dt_1 \cdots dt_{k-2} : 0 \leqslant f \leqslant 2^{k-1}(k-1)!,\ f \text{ increases} \right\}, \\ \mathcal{V}_k &= \{ l(u) 1_{[1/2,1]}(u) : l(u) = h(2u-1), h \in \widetilde{\mathcal{H}}_k^1([0,1]) \}. \end{aligned}$$

Then the decomposition above gives $\widetilde{\mathcal{H}}_k^1([0,1]) \subset \mathcal{U}_k + \mathcal{V}_k$, and thus we have for any $0 < \theta < 1$,

$$N_{[\cdot]}(\varepsilon, \widetilde{\mathcal{H}}_k^1([0,1]), \|\cdot\|_2) \leqslant N_{[\cdot]}((1-\theta)\varepsilon, \mathcal{U}_k, \|\cdot\|_2) \cdot N_{[\cdot]}(\theta\varepsilon, \mathcal{V}_k, \|\cdot\|_2). \tag{9}$$

We first claim that

$$N_{[\cdot]}(\varepsilon, \mathcal{U}_k, \|\cdot\|_2) \leqslant \exp(C\varepsilon^{-\frac{1}{k}}) \tag{10}$$

for some constant C depending only on k. Indeed, the claim is clearly true for $k = 1$, because in this case $\mathcal{U}_k$ consists of monotone functions that are bounded by 1. Thus, when $k = 1$, the inequality (10) is the special case of (3).

Suppose that (10) is true for $k = r$. That is, $N_{[\cdot]}(\varepsilon, \mathcal{U}_r, \|\cdot\|_2) \leqslant \exp(C\varepsilon^{-1/r})$. Because

$$N(\varepsilon, \mathcal{U}_r, \|\cdot\|_1) \leqslant N_{[\cdot]}(\varepsilon, \mathcal{U}_r, \|\cdot\|_2),$$

by applying Lemma 1 for L^2 norm under Lebesgue measure, we have

$$N_{[\cdot]}(\varepsilon^{1+\frac{1}{r}}, \mathcal{U}_{r+1}, \|\cdot\|_2) \leqslant e^{C\varepsilon^{-1/r}}, \tag{11}$$

which implies that $N_{[\cdot]}(\varepsilon,\mathcal{U}_{r+1},\|\cdot\|_2)\leqslant\exp(C\varepsilon^{-1/(r+1)})$ with a different constant C. Hence (10) holds for all $k\geqslant 1$. Therefore,

$$N_{[\cdot]}((1-\theta)\varepsilon,\mathcal{U}_k,\|\cdot\|_2)\leqslant\exp(C[(1-\theta)\varepsilon]^{-1/k}). \tag{12}$$

Next, we prove that

$$N_{[\cdot]}(\theta\varepsilon,\mathcal{V}_k,\|\cdot\|_2)\leqslant N_{[\cdot]}(\sqrt{2}\theta\varepsilon,\widetilde{\mathcal{H}}_k^1([0,1]),\|\cdot\|_2). \tag{13}$$

Indeed, if $[\underline{h}_i,\overline{h}_i]$, $1\leqslant i\leqslant N$ are $\sqrt{2}\theta\varepsilon$-brackets under the $\|\cdot\|_2$ distance that cover $\widetilde{\mathcal{H}}_k^1([0,1])$, then the brackets

$$[\underline{l}_i(x),\overline{l}_i(x)]=:[1_{[1/2,1]}(x)\overline{h}_i(2x-1),1_{[1/2,1]}(x)\underline{h}_i(2x-1)],$$

$1\leqslant i\leqslant N$, clearly cover $\mathcal{V}_k$. To see that they are $\theta\varepsilon$-brackets, we notice that

$$\begin{aligned}\|\underline{l}_i-\overline{l}_i\|_2&=\left\{\int_{1/2}^1|\overline{h}_i(2x-1)-\underline{h}_i(2x-1)|^2dx\right\}^{1/2}\\&=\frac{1}{\sqrt{2}}\left\{\int_0^1|\overline{h}_i(u)-\underline{h}_i(u)|^2du\right\}^{1/2}\\&\leqslant\theta\varepsilon.\end{aligned}$$

Applying (12) and (13) to (9), we obtain

$$N_{[\cdot]}(\varepsilon,\widetilde{\mathcal{H}}_k^1([0,1]),\|\cdot\|_2)\leqslant\exp(C[(1-\theta)\varepsilon]^{-1/k})N_{[\cdot]}(\sqrt{2}\theta\varepsilon,\widetilde{\mathcal{H}}_k^1([0,1]),\|\cdot\|_2).$$

Choosing $\frac{1}{\sqrt{2}}<\theta<1$ and by iteration, we obtain

$$N_{[\cdot]}(\varepsilon,\widetilde{\mathcal{H}}_k^1([0,1]),\|\cdot\|_2)\leqslant\exp(C'\varepsilon^{-1/k}) \tag{14}$$

for a different constant C depending only on k. Plugging (14) and (8) into (6), we obtain

$$N_{[\cdot]}(\varepsilon,\mathcal{M}_k^1([0,1]),\|\cdot\|_2)=N_{[\cdot]}(\varepsilon,\mathcal{H}_k^1([0,1]),\|\cdot\|_2)\leqslant\exp(C\varepsilon^{-1/k}). \tag{15}$$

If we let

$$\widetilde{\mathcal{M}}_k^\delta(I)=\{g\in\mathcal{M}_k^1(I):g\geqslant\delta\},$$

then because for all $g_1,g_2\in\widetilde{\mathcal{M}}_k^\delta([0,1])$ we have $h(g_1,g_2)\leqslant\frac{1}{\sqrt{2\delta}}\|g_1-g_2\|_2$, we obtain

$$N_{[\cdot]}(\varepsilon,\widetilde{\mathcal{M}}_k^\delta([0,1]),h)\leqslant N_{[\cdot]}(\sqrt{2\delta}\varepsilon,\mathcal{M}_k^1([0,1]),\|\cdot\|_2)\leqslant\exp(C\delta^{-\frac{1}{2k}}\varepsilon^{-\frac{1}{k}}). \tag{16}$$

Back to our goal of estimating $N_{[\cdot]}(\varepsilon,\widetilde{\mathcal{H}}_k^1([0,1]),h)$. We define

$$\mathcal{A}=\{g\in\widetilde{\mathcal{H}}_k^1([0,1]);g(1/2)<\delta\}\quad\text{and}\quad\mathcal{B}=\{g\in\widetilde{\mathcal{H}}_k^1([0,1]);g(1/2)\geqslant\delta\}.$$

Then

$$\mathcal{A}\subset\delta\widetilde{\mathcal{H}}_k^1([0,1/2])+\mathcal{H}_k^1([1/2,1]),\quad\mathcal{B}\subset\widetilde{\mathcal{H}}_k^1([0,1/2])+\widetilde{\mathcal{M}}_k^\delta([1/2,1]).$$

Therefore

$$N_{[\cdot]}(\varepsilon,\mathcal{A},h)\leqslant N_{[\cdot]}(\theta\varepsilon,\delta\widetilde{\mathcal{H}}_k^1([0,1/2]),h)\cdot N_{[\cdot]}((1-\theta)\varepsilon,\mathcal{H}_k^1([1/2,1]),h), \tag{17}$$

$$N_{[\cdot]}(\varepsilon,\mathcal{B},h)\leqslant N_{[\cdot]}((1-\eta)\varepsilon,\widetilde{\mathcal{H}}_k^1([0,1/2]),h)\cdot N_{[\cdot]}(\eta\varepsilon,\widetilde{\mathcal{M}}_k^\delta([1/2,1]),h). \tag{18}$$

Note that

$$N_{[\cdot]}(\theta\varepsilon, \delta\widetilde{\mathcal{H}}^1_k([0,1/2]), h) \leqslant N_{[\cdot]}(\theta\varepsilon/\sqrt{\delta}, \widetilde{\mathcal{H}}^1_k([0,1]), h).$$

Also note that

$$\begin{aligned} N_{[\cdot]}((1-\theta)\varepsilon, \mathcal{H}^1_k([1/2,1]), h) &= N_{[\cdot]}((1-\theta)\varepsilon, \mathcal{M}^1_k([0,1/2]), h) \\ &= N_{[\cdot]}(\sqrt{2}(1-\theta)\varepsilon, \mathcal{M}^1_k([0,1]), h) \\ &\leqslant N_{[\cdot]}(\sqrt{2}(1-\theta)^2\varepsilon, \widetilde{\mathcal{H}}^1_k([0,1]), h) \cdot N_{[\cdot]}(\sqrt{2}(1-\theta)\theta\varepsilon, \mathcal{P}_k, h) \\ &\leqslant N_{[\cdot]}(\sqrt{2}(1-\theta)^2\varepsilon, \widetilde{\mathcal{H}}^1_k([0,1]), h) \cdot \left(1 + \frac{2k}{(1-\theta)^2\theta^2\varepsilon^2}\right)^k, \end{aligned}$$

where the last inequality follows from (7). We choose θ and δ so that $\sqrt{2}(1-\theta)^2 = 1.4$ and $\theta/\sqrt{\delta} = L$, where L is a large number to be fixed later. Thus by plugging the two bounds above into (17) we obtain

$$N_{[\cdot]}(\varepsilon, \mathcal{A}, h) \leqslant N_{[\cdot]}(1.4\varepsilon, \widetilde{\mathcal{H}}^1_k([0,1]), h) \cdot N_{[\cdot]}(L\varepsilon, \widetilde{\mathcal{H}}^1_k([0,1]), h) \cdot \left(1 + \frac{C_1}{\varepsilon^2}\right)^k \tag{19}$$

for some constant C_1.

On the other hand, by choosing η so that $\sqrt{2}(1-\eta) = 1.4$, we have

$$N_{[\cdot]}((1-\eta)\varepsilon, \widetilde{\mathcal{H}}^1_k([0,1/2]), h) \leqslant N_{[\cdot]}(1.4\varepsilon, \widetilde{\mathcal{H}}^1_k([0,1]), h).$$

Recall that by (16) we have

$$N_{[\cdot]}(\eta\varepsilon, \widetilde{\mathcal{M}}^\delta_k([1/2,1]), h) \leqslant \exp(C_2\varepsilon^{-1/k}),$$

with a constant C_2 depending on L. Plugging into (18), we obtain

$$N_{[\cdot]}(\varepsilon, \mathcal{B}, h) \leqslant N_{[\cdot]}(1.4\varepsilon, \widetilde{\mathcal{H}}^1_k([0,1]), h) \cdot \exp(C_2\varepsilon^{-1/k}). \tag{20}$$

But $\widetilde{\mathcal{H}}^1_k([0,1]) \subset \mathcal{A} \cup \mathcal{B}$, (19) and (20) imply that

$$\begin{aligned} \log N_{[\cdot]}(\varepsilon, \widetilde{\mathcal{H}}^1_k([0,1]), h) \leqslant{} & \log N_{[\cdot]}(1.4\varepsilon, \widetilde{\mathcal{H}}^1_k([0,1]), h) + \log N_{[\cdot]}(L\varepsilon, \widetilde{\mathcal{H}}^1_k([0,1]), h) \\ & + C_2\varepsilon^{-1/k} + k\log\left(1 + \frac{C_1}{\varepsilon^2}\right) + 2. \end{aligned}$$

Let $Z(\varepsilon) = \varepsilon^{1/k}\log N_{[\cdot]}(\varepsilon, \widetilde{\mathcal{H}}^1_k([0,1]), h)$. Then the inequality above implies that

$$Z(\varepsilon) \leqslant 1.4^{-1/k}Z(1.4\varepsilon) + L^{-1/k}Z(L\varepsilon) + C$$

for some constant C, which further implies that

$$\sup_{\eta \geqslant \varepsilon} Z(\eta) \leqslant (1.4^{-1/k} + L^{-1/k}) \sup_{\eta \geqslant \varepsilon} Z(\eta) + C.$$

By choosing L large so that $1.4^{-1/k} + L^{-1/k} < 1$, we immediate obtain

$$\log N_{[\cdot]}(\varepsilon, \widetilde{\mathcal{H}}^1_k([0,1]), h) \leqslant C\varepsilon^{-1/k}, \tag{21}$$

for some constant C. Together with (7), we have

$$\log N_{[\cdot]}(\varepsilon, \mathcal{H}_k^1([0,1]), h) \leqslant C\varepsilon^{-1/k}.$$

Summarizing, we obtain

Theorem 2. *There exists a constant C depending only on k, such that*

$$\log N_{[\cdot]}(\varepsilon, \mathcal{M}_k^B([0,A]), h) \leqslant C(AB)^{\frac{1}{2k}} \varepsilon^{-\frac{1}{k}}. \tag{22}$$

2.2 Bounded k-monotone densities

Now we consider $N_{[\cdot]}(\varepsilon, \mathcal{F}_k^B([0,1]), h)$. Because $\mathcal{F}_k^B([0,1]) \subset \mathcal{M}_k^B([0,1])$, the result of Theorem 2 applies to $\mathcal{F}_k^B([0,1])$. Our goal is to improve the constant $CB^{\frac{1}{2k}}$ by using the the extra fact that $\int_0^1 g(u)du = 1$ for $g \in \mathcal{F}_k^B([0,1])$. For convenience, we will actually relax the condition $\int_0^1 g(u)du = 1$ in the definition of $\mathcal{F}_k^B([0,1])$ to the condition $\int_0^1 g(u)du \leqslant 1$.

Assume $B > 2$. For $1/B < \delta < 1$, we define

$$X_\delta = \left\{ f \in \mathcal{F}_k^B([0,1]) : \int_0^{1/2} f(u)du < \delta \right\} \quad \text{and} \quad Y_\delta = \left\{ f \in \mathcal{F}_k^B([0,1]) : \int_0^{1/2} f(u)du \geqslant \delta \right\}.$$

Then, $\mathcal{F}_k^B([0,1]) = X_\delta \cup Y_\delta$. For any $f \in \mathcal{F}_k^B([0,1])$, because f increases and $\int_{1/2}^1 f(u)du \leqslant 1$, we have $f(u) \leqslant 2$ for all $u \in [0,1/2]$. Hence, we can decompose X_δ and Y_δ into two classes of functions with disjoint supports:

$$\begin{aligned} X_\delta &\subset \delta\mathcal{F}_k^{2/\delta}([0,1/2]) + \mathcal{F}_k^B([1/2,1]), \\ Y_\delta &\subset \mathcal{F}_k^2([0,1/2]) + (1-\delta)\mathcal{F}_k^{B/(1-\delta)}([1/2,1]). \end{aligned}$$

Therefore, we have

$$\begin{aligned} N_{[\cdot]}(\varepsilon, \mathcal{F}_k^B([0,1]), h) &\leqslant N_{[\cdot]}(2\sqrt{\delta}\varepsilon, \delta\mathcal{F}_k^{2/\delta}([0,1/2]), h) \cdot N_{[\cdot]}(\sqrt{1-4\delta}\varepsilon, \mathcal{F}_k^B([1/2,1]), h) \\ &\quad + N_{[\cdot]}(\sqrt{\delta}\varepsilon, \mathcal{F}_k^2([0,1/2]), h) \cdot N_{[\cdot]}(\sqrt{1-\delta}\varepsilon, (1-\delta)\mathcal{F}_k^{B/(1-\delta)}([1/2,1]), h) \\ &= N_{[\cdot]}(2\varepsilon, \mathcal{F}_k^{2/\delta}([0,1/2]), h) \cdot N_{[\cdot]}(\sqrt{1-4\delta}\varepsilon, \mathcal{F}_k^B([1/2,1]), h) \\ &\quad + N_{[\cdot]}(\sqrt{\delta}\varepsilon, \mathcal{F}_k^2([0,1/2]), h) \cdot N_{[\cdot]}(\varepsilon, \mathcal{F}_k^{B/(1-\delta)}([1/2,1]), h) \\ &\leqslant N_{[\cdot]}(2\varepsilon, \mathcal{F}_k^B([0,1]), h) \cdot N_{[\cdot]}(\sqrt{1-4\delta}\varepsilon, \mathcal{F}_k^{B/2}([0,1]), h) \\ &\quad + \exp(C\delta^{-\frac{1}{2k}}\varepsilon^{-\frac{1}{k}}) \cdot N_{[\cdot]}(\varepsilon, \mathcal{F}_k^{B/(2-2\delta)}([0,1]), h) \\ &\leqslant [N_{[\cdot]}(2\varepsilon, \mathcal{F}_k^B([0,1]), h) + \exp(C\delta^{-\frac{1}{2k}}\varepsilon^{-\frac{1}{k}})] \\ &\quad \cdot N_{[\cdot]}(\sqrt{1-4\delta}\varepsilon, \mathcal{F}_k^{B/(2-2\delta)}([0,1]), h). \end{aligned}$$

By iteration, we obtain

$$\begin{aligned} N_{[\cdot]}(\varepsilon, \mathcal{F}_k^B([0,1]), h) &\leqslant [N_{[\cdot]}(2^m\varepsilon, \mathcal{F}_k^B([0,1]), h) + \exp(C\delta^{-\frac{1}{2k}}\varepsilon^{-\frac{1}{k}})] \\ &\quad \cdot N_{[\cdot]}((\sqrt{1-4\delta})^m\varepsilon, \mathcal{F}_k^{B/(2-2\delta)^m}([0,1]), h). \end{aligned}$$

Let $\delta = 1/\log_2 B$ and choose $m = \lceil \log_2 B \rceil$, we have

$$N_{[\cdot]}(\varepsilon, \mathcal{F}_k^B([0,1]), h) \leqslant \exp(C(\log B)^{-\frac{1}{2k}}\varepsilon^{-\frac{1}{k}}).$$

We summarize this bound in the following theorem:

Theorem 3. *Let $\mathcal{F}_k^B([0,A])$ be the class of k-monotone densities on $[0,A]$ that are bounded by B. Then*

$$\log N_{[\cdot]}(\varepsilon, \mathcal{F}_k^B([0,A]), h) \leqslant C|\log AB|^{\frac{1}{2k}}\varepsilon^{-\frac{1}{k}},$$

where C is a constant depending only on k.

3 Under $L^p(Q)$ Distances

In this section we will consider the bracketing entropy of $\mathcal{M}_k^B([0,A])$ and $\mathcal{F}_k^B([0,A])$ under the $L^p(Q)$ distance, where $1 \leqslant p < \infty$ and Q is any probability measure on $[0,A]$. We will prove the following theorem:

Theorem 4. (i) *There exists a constant C depending only on p and k, such that for any probability measure Q on $[0,A]$ that is absolutely continuous with respect to Lebesgue measure with bounded density q,*

$$\log N_{[\cdot]}(\varepsilon, \mathcal{M}_k^B([0,A]), \|\cdot\|_{p,Q}) \leqslant C(\|q\|_\infty AB^p)^{1/(pk)}\varepsilon^{-1/k}, \tag{23}$$

$$\log N_{[\cdot]}(\varepsilon, \mathcal{F}_k^B([0,A]), \|\cdot\|_{p,Q}) \leqslant C[\log(\|q\|_\infty AB^p)]^{1/(pk)}\varepsilon^{-1/k}. \tag{24}$$

(ii) *Let $\mathcal{G}_k^B([0,A])$ consist of k-monotone functions g on $[0,A]$, such that $g'(0) \geqslant -B$, and let $\widetilde{\mathcal{G}}_k^B([0,A])$ consist of probability densities that belong to $\mathcal{G}_k^B([0,A])$. Then there exists a constant C depending only on p and k, such that for any probability measure Q on $[0,A]$,*

$$\log N_{[\cdot]}(\varepsilon, \mathcal{G}_k^B([0,A]), \|\cdot\|_{p,Q}) \leqslant CA^{1/(pk)}B^{1/k}\varepsilon^{-1/k}, \tag{25}$$

$$\log N_{[\cdot]}(\varepsilon, \widetilde{\mathcal{G}}_k^B([0,A]), \|\cdot\|_{p,Q}) \leqslant C[\log(AB^p)]^{1/(pk)}\varepsilon^{-1/k}. \tag{26}$$

Remark 5. *In view of the result proved in* [8] *the rate $\varepsilon^{-1/k}$ is sharp when $p=2$.*

Proof. The result is known for the case $k=1$. Thus, we only need to consider the case $k>1$.

To prove the first inequality in the statement (i), we note that

$$N_{[\cdot]}(\varepsilon, \mathcal{M}_k^B([0,A]), \|\cdot\|_{p,Q}) \leqslant N_{[\cdot]}\left(\frac{\varepsilon}{\|q\|_\infty^{1/p}A^{1/p}B}, \mathcal{M}_k^1([0,1]), \|\cdot\|_p\right),$$

where $\|\cdot\|_p$ is the L^p distance under Lebesgue measure. Thus, it suffices to prove that for any $\eta>0$, $N_{[\cdot]}(\eta, \mathcal{M}_k^1([0,1]), \|\cdot\|_p) \leqslant C\eta^{-1/k}$. However, this follows from the same argument as in the proof of Theorem 2. Indeed, the only change needed is to replace (11) by

$$N_{[\cdot]}(\varepsilon^{1+\frac{1}{k-1}}, \mathcal{U}_k, \|\cdot\|_p) \leqslant e^{C\varepsilon^{-1/(k-1)}},$$

which leads to $N_{[\cdot]}(\varepsilon, \mathcal{U}_k, \|\cdot\|_p) \leqslant e^{C\varepsilon^{-1/k}}$. The first inequality in the statement (i) then follows by iteration. The proof of the second inequality is also similar to that of Theorem 3.

To prove the first inequality in the statement (ii), we note that by the change of variables $x=Au$, we have

$$N_{[\cdot]}(\varepsilon, \mathcal{G}_k^B([0,A]), \|\cdot\|_{p,Q}) = N_{[\cdot]}(\varepsilon, \mathcal{G}_k^{A^{1/p}B}([0,1]), \|\cdot\|_{p,P}),$$

where P is the probability measure on $[0,1]$ defined by $P([0,u]) = Q([0,Au])$. Thus, it suffices to consider the case $AB^p=1$. Furthermore, by approximation, we can assume that P is absolutely continuously with respect to the Lebesgue measure on $[0,1]$. Let α be the inverse function of $P([0,x])$. By the change of variable $u = P([0,x])$, it is easy to see that

$$N_{[\cdot]}(\varepsilon, \mathcal{G}_k^1([0,1]), \|\cdot\|_{p,P}) = N_{[\cdot]}(\varepsilon, \mathcal{G}_{k,\alpha}^1([0,1]), \|\cdot\|_p),$$

where

$$\mathcal{G}_{k,\alpha}^1([0,1]) = \{g(\alpha(u)) : g \in \mathcal{G}_k^1([0,1])\} = \left\{ \int_0^{\alpha(u)} f(t)dt : f \in \mathcal{M}_{k-1}^1([0,1]) \right\}.$$

By (23) we have

$$N_{[\cdot]}(\varepsilon, \mathcal{M}_{k-1}^1([0,1]), \|\cdot\|_1) \leqslant \exp(C\varepsilon^{\frac{1}{k-1}}).$$

Applying Lemma 1, we obtain

$$N_{[\cdot]}(\varepsilon^{1+\frac{1}{k-1}}, \mathcal{G}_{k,\alpha}^1([0,1]), \|\cdot\|_p) \leqslant \exp(C\varepsilon^{\frac{1}{k-1}})).$$

which leads the inequality (25). The proof of (26) follows the same argument as Theorem 3, and is thus omitted.

4 Rates of convergence for the maximum likelihood estimator of a bounded k-monotone density

Let $\hat{p}_{n,k}$ be the MLE of a k-monotone density p_0 on $[0,A]$ based on $X_1,\dots,X_n$ i.i.d. with density $p_0 \in \mathcal{F}_k^B([0,A])$ for some $0 < A, B < \infty$. Thus p_0 is bounded and concentrated on $[0,A]$.

From [2] (see also [9]) we know that $\hat{p}_{n,k}$ is characterized by

$$\begin{aligned} 1 &\geqslant \int_0^y \frac{k}{y^k} \frac{(y-x)^{k-1}}{\hat{p}_{n,k}(x)} \, d\mathbb{F}_n(x) \qquad \text{for all} \ \ y > 0 \\ &\geqslant \frac{1}{\hat{p}_{n,k}(0)} \int_0^y \frac{k}{y^k} (y-x)^{k-1} \, d\mathbb{F}_n(x) = \frac{1}{\hat{p}_{n,k}(0)} \int_0^y \frac{k}{y} \left(1 - \frac{x}{y}\right)^{k-1} d\mathbb{F}_n(x), \end{aligned} \tag{27}$$

with equality in the inequality in (27) at points $\tau \in \mathrm{supp}(\hat{G}_{n,k}) = \{\tau_1, \tau_2, \dots, \tau_m\}$, where we may assume that $0 < \tau_1 < \cdots < \tau_m$ (with m random) and where

$$\hat{p}_{n,k}(x) = \int_0^\infty \frac{k(y-x)_+^{k-1}}{y^k} \, d\hat{G}_{n,k}(y).$$

Therefore

$$\hat{p}_{n,k}(0) \geqslant \int_0^\infty \frac{k}{y^k} (y-x)_+^{k-1} d\mathbb{F}_n(x) \qquad \text{for all} \ \ y > 0. \tag{28}$$

To apply our entropy bounds we need to show that $\hat{p}_{n,k}$ is bounded with (arbitrarily) high probability when p_0 is bounded. This is the content of the following proposition.

Proposition 6. *Suppose that $p_0 \in \mathcal{F}_k^B([0,A])$. Then the MLE $\hat{p}_{n,k}$ satisfies*

$$\hat{p}_{n,k}(0+) \leqslant k \sup_{x>0} (\mathbb{F}_n(x)/x) = O_p(1).$$

Proof. The characterization of the MLE implies that

$$\begin{aligned} 1 &\geqslant \frac{k}{y^k}\int_0^y \frac{(y-x)^{k-1}}{\hat{p}_{n,k}(x)}\, d\mathbb{F}_n(x) \qquad \text{for all } \ 0<y\leqslant \tau_1 \\ &= \frac{k}{y}\int_0^y \frac{(1-x/y)^{k-1}}{\hat{p}_{n,k}(x)}\, d\mathbb{F}_n(x) \end{aligned} \tag{29}$$

with equality at $y=\tau_1$:

$$1 = \frac{k}{\tau_1}\int_0^{\tau_1} \frac{(1-x/\tau_1)^{k-1}}{\hat{p}_{n,k}(x)}\, d\mathbb{F}_n(x). \tag{30}$$

Now note that the support of $\hat{G}_{n,k}$ is concentrated on $y\geqslant \tau_1$, so $x/y\leqslant x/\tau_1$, or $(1-x/y)\geqslant (1-x/\tau_1)$ for $y\geqslant\tau_1$ and $0\leqslant x\leqslant \tau_1$. Thus it follows that

$$\begin{aligned} \hat{p}_{n,k}(x) &= \int_0^\infty \frac{k}{y}\left(1-\frac{x}{y}\right)_+^{k-1} d\hat{G}_{n,k}(y) \\ &\geqslant \left(1-\frac{x}{\tau_1}\right)_+^{k-1}\int_0^\infty \frac{k}{y}\, d\hat{G}_{n,k}(y) = \left(1-\frac{x}{\tau_1}\right)_+^{k-1} \hat{p}_{n,k}(0). \end{aligned}$$

Combining the last two displays we find that

$$\begin{aligned} 1 &= \frac{k}{\tau_1}\int_0^{\tau_1} \frac{(1-x/\tau_1)^{k-1}}{\hat{p}_{n,k}(x)}\, d\mathbb{F}_n(x) \\ &\leqslant \frac{k}{\tau_1}\int_0^{\tau_1} \frac{(1-x/\tau_1)^{k-1}}{\hat{p}_{n,k}(0)(1-x/\tau_1)^{k-1}}\, d\mathbb{F}_n(x) \leqslant \frac{k}{\tau_1\hat{p}_{n,k}(0)}\mathbb{F}_n(\tau_1), \end{aligned}$$

which yields

$$\begin{aligned} \hat{p}_{n,k}(0) &\leqslant k\frac{\mathbb{F}_n(\tau_1)}{\tau_1} \leqslant k\sup_{t>0}\frac{\mathbb{F}_n(t)}{t} \\ &= k\sup_{t>0}\frac{\mathbb{F}_n(t)}{F_0(t)}\frac{F_0(t)}{t} \leqslant k\sup_{t>0}\frac{\mathbb{F}_n(t)}{F_0(t)}\cdot p_0(0+) = O_p(1) \end{aligned}$$

by Daniels' inequality; see [10, Theorem 9.1.2, p. 345].

Now suppose that $\mathcal{P}$ is a collection of densities with respect to a sigma-finite measure μ. The following theorem is a simplified version of [5, Theorem 3.4.4, p. 327]. Our rate theorem for the MLE $\hat{p}_{n,k}$ over the class $\mathcal{P}_k([0,A])$, the class of k-monotone densities on $[0,A]$, will be proved by combining the upper bound of this theorem with (an easy modification of) the rate results given in [5, Theorem 3.2.5, p. 289].

Theorem 7. *Suppose that $X_1,\dots,X_n$ are i.i.d. P_0 with density $p_0\in\mathcal{P}$. Let h be the Hellinger distance between densities, and m_p be defined, for $p\in\mathcal{P}$, by*

$$m_p(x) = \log\left(\frac{p(x)+p_0(x)}{2p_0(x)}\right).$$

Then

$$\mathbb{M}(p)-\mathbb{M}(p_0) \equiv P_0(m_p - m_{p_0}) \lesssim -h^2(p,p_0).$$

Furthermore, with $\mathcal{M}_\delta = \{m_p - m_{p_0} :\ h(p,p_0)\leqslant\delta,\ p\in\mathcal{P}_0\}$, we also have

$$E^*_{P_0}\|\mathbb{G}_n\|_{\mathcal{M}_\delta} \lesssim \tilde{J}_{[\cdot]}(\delta,\mathcal{P}_0,h)\left(1+\frac{\tilde{J}_{[\cdot]}(\delta,\mathcal{P}_0,h)}{\delta^2\sqrt{n}}\right) \equiv \phi_n(\delta,\mathcal{P}_0), \tag{31}$$

where

$$\tilde{J}_{[\cdot]}(\delta, \mathcal{P}_0, h) = \int_0^\delta \sqrt{1 + \log N_{[\cdot]}(\varepsilon, \mathcal{P}_0, h)}\, d\varepsilon.$$

Here is our main result of this section:

Theorem 8. *Suppose that $p_0 \in \mathcal{F}_k^B([0,A]) \subset \mathcal{P}_k([0,A])$ for some $0 < A, B < \infty$. Then the maximum likelihood estimator $\hat{p}_{n,k}$ of p_0 in $\mathcal{P}_k$ satisfies*

$$h(\hat{p}_{n,k}, p_0) = O_p(n^{-\frac{k}{2k+1}}).$$

Remark 9. *This generalizes the rate result of* [3] (*with resulting rate of convergence $n^{-1/3}$*) *to $k > 1$. For the case $k = 1$, closely related results with the Hellinger metric replaced by the L_1 metric, were obtained by* [11–13]. *The rate established in Theorem* 8 *is apparently consistent with the local rate result of $n^{-k/(2k+1)}$ established* (*up to an envelope conjecture*) *by* [2].

Proof. For simplicity, we write $\hat{p}_n \equiv \hat{p}_{n,k}$. Let $M > 0$ and $K > 0$. Then

$$\mathbb{P}(r_n h(\hat{p}_n, p_0) \geqslant 2^M) \leqslant \mathbb{P}(r_n h(\hat{p}_n, p_0) \geqslant 2^M,\ \hat{p}_n(0) \leqslant K) + \mathbb{P}(\hat{p}_n(0) > K) \equiv \mathrm{I}_n + \mathrm{II}_n,$$

where, by Proposition 6,

$$\mathrm{II}_n \leqslant \mathbb{P}(k\|\mathbb{F}_n/F_0\|_\infty p_0(0+) > K) \leqslant \frac{kp_0(0+)}{K}$$

by Daniels' inequality, and hence II_n can be made arbitrarily small (uniformly in n) by choosing K large. Now we essentially follow the proof of [5, Theorem 3.2.5] (with θ identified with p), but exploit the fact that $\hat{p}_n(0+) \leqslant K$. Thus, letting $\mathbb{M}_n(p) \equiv P_n m(p)$ we have for any large $\eta > 0$,

$$\begin{aligned}
&\mathbb{P}(r_n h(\hat{p}_n, p_0) > 2^M,\ \hat{p}_n(0+) \leqslant K) \\
&\quad\leqslant \sum_{j=M}^{\lfloor \log_2(\eta r_n)\rfloor} \mathbb{P}\Big(\sup_{f\in S_{j,n}} (\mathbb{M}_n(p) - \mathbb{M}_n(p_0) \geqslant 0\Big) + \mathbb{P}(2h(\hat{p}_n, p_0) \geqslant \eta) \\
&\quad\equiv I_{A,n} + I_{B,n},
\end{aligned}$$

where the shells $S_{j,n}$ are now defined with the additional restriction that $p(0+) \leqslant K$:

$$\begin{aligned}
S_{j,n} &= \{p \in \mathcal{P}_k([0,A]) :\ 2^{j-1} < r_n h(p, p_0) \leqslant 2^j,\ p(0+) \leqslant K\} \\
&= \{p \in \mathcal{F}_k^K([0,A]) :\ 2^{j-1} < r_n h(p, p_0) \leqslant 2^j\}.
\end{aligned}$$

Here the term $I_{B,n}$ can be made arbitrarily small for all large n by the consistency of $\hat{p}_n$ established by [9]. Thus the same argument as in [5] yields, with $\mathcal{P}_0 = \mathcal{F}_k^K$ in (31), and $\phi_n(\delta, \mathcal{F}_k^K) \equiv \phi_n(\delta)$,

$$I_{A,n} \lesssim \sum_{j\geqslant M} \frac{\phi_n(2^j/r_n) r_n^2}{\sqrt{n}2^{2j}}. \tag{32}$$

By (31)

$$\phi_n(\delta) \equiv \phi_n(\delta, \mathcal{F}_n^K) = \tilde{J}_{[\cdot]}(\delta, \mathcal{F}_n, h)\bigg(1 + \frac{\tilde{J}_{[\cdot]}(\delta, \mathcal{F}_n, h)}{\delta^2\sqrt{n}}\bigg).$$

A direct calculation using Theorem 3 gives

$$\tilde{J}_{[\cdot]}(\delta, \mathcal{F}_n, h) = \frac{\sqrt{CD}}{1-\frac{1}{2k}}\delta^{1-\frac{1}{2k}}.$$

for the same constant C as in Theorem 3, where $D \equiv |\log(AK)|^{1/(2k)}$. This implies,

$$\phi_n(\delta) = \frac{\sqrt{CD}}{1-\frac{1}{2k}}\delta^{1-\frac{1}{2k}}\left(1 + \frac{\sqrt{CD}}{1-\frac{1}{2k}}\frac{\delta^{1-\frac{1}{2k}}}{\delta^2\sqrt{n}}\right).$$

By taking

$$r_n \equiv \left(c_0\frac{1-\frac{1}{2k}}{\sqrt{CD}}\right)^{\frac{2k}{2k+1}} n^{\frac{k}{2k+1}}$$

where $c_0 \equiv (\sqrt{5}-1)/2$, we have $r_n^2\phi_n(1/r_n) = \sqrt{n}$. Note that the functions $\delta \mapsto \phi_n(\delta)/\delta$ are decreasing, therefore for any $j > 0$,

$$\phi_n(2^j/r_n)r_n^2 \leqslant 2^j\phi_n(1/r_n)r_n^2.$$

Hence, (32) can be estimated by

$$I_{A,n} \lesssim \sum_{j\geqslant M}\frac{\phi_n(2^j/r_n)r_n^2}{\sqrt{n}2^{2j}} \leqslant \sum_{j\geqslant M} 2^{-j} = 2^{-(M-1)}.$$

References

1 Williamson R E. Multiply monotone functions and their Laplace transforms. *Duke Math J*, **23**: 189–207 (1956)

2 Balabdaoui F, Wellner J A. Estimation of a k-monotone density: limit distribution theory and the spline connection. *Ann Statist*, **35**(6): 2536–2564 (2007)

3 van de Geer S. Hellinger-consistency of certain nonparametric maximum likelihood estimators. *Ann Statist*, **21**(1): 14–44 (1993)

4 Birman M Š, Solomjak M Z. Piecewise polynomial approximations of functions of classes $W_p{}^{\alpha}$. *Mat Sb*, **73**(115): 331–355 (1967)

5 van der Vaart A W, Wellner J A. Weak convergence and empirical processes. Springer Series in Statistics. New York: Springer-Verlag, 1996

6 Blei R, Gao F, Li W. Metric entropy of high-dimensional distributions. *Proc Amer Math Soc*, **135**(12): 4009–4018 (2007)

7 Gao F, Wellner J A. Entropy estimate for high-dimensional monotonic functions. *J Multivariate Anal*, **98**(9): 1751–1764 (2007)

8 Gao F. Entropy estimate for k-monotone functions via small ball probability of integrated Brownian motion. *Electron Commun Probab*, **13**: 121–130 (2008)

9 Balabdaoui F, Wellner J A. Estimation of a k-monotone density, part 1: characterizations, consistency, and minimax lower bounds. Technical Report, University of Washington, Department of Statistics, 2004

10 Shorack G R, Wellner J A. Empirical processes with applications to statistics. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. New York: John Wiley & Sons Inc, 1986

11 Groeneboom P. Some current developments in density estimation. *Mathematics and Computer Science* (*Amsterdam,* 1983), **1**: 163–192 (1986)

12 Birgé L. On estimating a density using Hellinger distance and some other strange facts. *Probab Theory Related Fields*, **71**(2): 271–291 (1986)

13 Birgé L. The Grenander estimator: a nonasymptotic approach. *Ann Statist*, **17**(4): 1532–1549 (1989)