ASYMPTOTICAL MINIMAX RESULTS IN IMAGE ANALYSIS FOR SETS WITH SMOOTH BOUNDARIES

by

E. Mammen* **and A.B. Tsybakov****

November 1992

Abstract

In this paper optimal rates of convergence are derived for estimates in image processing for N-dimensional black and white images under smoothness conditions. It is assumed that the boundaries of the regions of black pixels have a "smooth" parametrization, i.e., that the boundaries are given by "smooth" functions from the sphere S^{N-1} into $\mathbb{R}^N$. Furthermore, classes of convex regions are considered. The results are applied to edge estimation in binary images and to nonparametric estimation of the support of a density.

AMS 1991 Subject Classification 62 G05, 62 G20

Key words: image analysis, rates of convergence, ε-entropy

* Institut für Angewandte Mathematik, Universität Heidelberg, Germany

** CORE, Université Catholique de Louvain, Louvain-la-Neuve, Belgium and Institute for Problems of Information Transmission, Academy of Sciences of Russia, Moscow.

1. Introduction

Optimal rates of convergence in nonparametric estimation have been studied by Bretagnolle and Huber (1979), Ibragimov and Khasminskii (1980,1981), Stone (1980,1982), Birgé (1983) and others. These results are mainly related to nonparametric regression and density estimation. In this paper we discuss optimal convergence rates of estimates in image processing for N-dimensional black and white images. We consider estimates of the regions of black pixels under smoothness conditions. It is assumed that the boundaries of these regions have a "smooth" parametrisation, i.e., that the boundaries are given by "smooth" functions from the sphere S^{N-1} into $\mathbb{R}^N$. In particular, this includes models of black regions with "smooth" boundaries, which have been considered in Korostelev (1991), Tsybakov (1991), Korostelev and Tsybakov (1992,1993). However, smooth parametrization of the boundary does not imply that the boundary itself is smooth (see Figure 1). Our assumptions are very simple, and also for regions with smooth boundaries and for convex sets our approach gives the correct rate of convergence. We apply our results to edge estimation in binary images. Edge estimation often is a first step in image reconstruction and image segmentation (see e.g. Pratt (1979)). Furthermore, we consider nonparametric estimation of support of a density. This is related to cluster analysis, to the detection of abnormal behavior (see Devroye and Wise (1980)), and to the estimation of level sets of high - dimensional nonparametric densities.

2. Smoothness classes of regions

Let us remind the definition of smoothness classes of sets introduced by Dudley (1974). These classes are uniquely determined by constants $\gamma \geq 1, L > 0$, integers $N \geq 2, J \geq 1$ and by $\{(F_j, V_j),\ j = 1, \ldots, J\}$ where $V_1, \ldots, V_J$ are open sets on the sphere

$$S^{N-1} = \{x = (x_1, \ldots, x_N) \in \mathbb{R}^N : x_1^2 + \ldots + x_N^2 = 1\},$$

such that $\bigcup_j V_j = S^{N-1}$, and F_j are infinitely many times differentiable isomorphisms

$$F_j : \{x : x_1^2 + \ldots + x_{N-1}^2 < 1\} \to V_j.$$

The functions F_j establish a relation between the surface S^{N-1} and the covering system $\{V_j\}$. In the following we asume that these functions and the system $\{V_j\}$

.

are fixed, so that our definition of classes of sets will have only γ, L, and N as free parameters.

Denote

$$\mathcal{F}_{\gamma,L} = \{g = (g_1, \ldots, g_N) | g_i : S^{N-1} \to [0,1]$$

$$\text{with} \sup_{x,y} \frac{|\varphi^{(k)}(x) - \varphi^{(k)}(y)|}{|x-y|^{\gamma-k}} \leq L \text{ for } \varphi(x) = g_i(F_j(x)),\ j = 1, \ldots, J,\ i = 1, \ldots, N\}.$$

Here k is $\max\{m \in \mathbb{N} : m < \gamma\}$, $|\cdot|$ denotes the Euclidean norm, and $\varphi^{(k)}(x)$ is the vector of all partial derivatives of order k of a function $\varphi(x)$.

A function $g = (g_1, \ldots, g_N) \in \mathcal{F}_{\gamma,L}$ defines a surface in $\mathbb{R}^N$ with "smoothness parameter" γ. As in Dudley (1974) for each $g \in \mathcal{F}_{\gamma,L}$ we introduce the set $G(g)$ which has this surface as boundary in certain sense. The set $G(g)$ consists of all points $x \in \mathbb{R}^N$ such that either

(i) $x \in$ range (g)

or

(ii) $x \notin$ range (g) and in $\mathbb{R}^N \setminus \{x\}$ the function g is not homotopic to any constant map of S^{N-1} into a point $t \in \mathbb{R}^N \setminus \{x\}$.

Here range (g) denotes $\{g(s) | s \in S^{N-1}\}$. Three examples for $N = 2$ are given in Figure 1. These examples show that the boundary $\partial G(g) \subset$ range (g) is not necessarily smooth although the curve range (g) is smooth. Finally, the smoothness class of sets is defined as:

$$\mathcal{G}_{\gamma,L} = \{G(g) \cap [0,1]^N | g \in \mathcal{F}_{\gamma,L}\}.$$

3. Statistical Models

We consider the problem of estimating a set $G \in \mathcal{G}_{\gamma,L}$ for the following two models.

MODEL 1 (EDGE ESTIMATION). *One observes $X_i \in [0,1]^N$ and $Y_i \in \{-1, 1\}$ $(i = 1, \ldots, n)$. (Interpretation: X_i is a point, and Y_i is the image grey level at this point, $Y_i = -1$ for white, $Y_i = 1$ for black). We assume that the observations Y_i reproduce a black region G correctly with probability $1/2 + a_n$:*

$$Y_i = (2I(X_i \in G) - 1)\xi_i, \quad i = 1, \ldots, n, \tag{3.1}$$

where $\xi_1, \dots, \xi_n$, are i.i.d. random variables with

$$\xi_i = \xi_{i,n} = \begin{cases} 1 \text{ with probability } 1/2 + a_n, \\ -1 \text{ with probability} 1/2 - a_n, \end{cases}$$

for a sequence a_n with $0 < a_n < 1/2$. Here $I(\cdot)$ is the indicator function. It is assumed that $\xi_1, \dots, \xi_n$ are independent of $X_1, \dots, X_n$.

Assumptions on the distribution of $(X_1, \dots, X_n)$ will be given below. The model (3.1) can be written as

$$Y_i = 2a_n f_G(X_i) + \zeta_{in}, \quad i = 1, \dots, n, \tag{3.2}$$

where f_G is an image function,

$$f_G(X_i) = 2I\{X_i \in G\} - 1,$$

and ζ_{in} are some independent bounded zero mean random variables.

In view of (3.2) a_n is the edge step size of a binary image. The interesting cases are (i) $a_n \equiv a$ with $0 < a < 1/2$ and (ii) $a_n \to 0$. The case (i) has been considered in Korostelev and Tsybakov (1992, 1993) for smoothness classes of boundary fragments (sets G with boundaries defined by smooth functions of the first $N - 1$ coordinates). The case (ii) models the pictures with high noise. In general, this allows to treat images with grey levels varying in time and may be useful for some applications.

MODEL 2. *One observes i.i.d. random variables $X_i \in [0,1]^N$ with Lebesgue density*

$$f_G(x) = [(1/2 + a_n)I(x \in G) + (1/2 - a_n)I(x \in \overline{G})]/[2a_n\lambda(G) + 1/2 - a_n],$$

where a_n is a sequence with $0 < a_n \le 1/2$, $\overline{G} = [0,1]^N \setminus G$ and λ is the Lebesgue measure. It is assumed that $2a_n\lambda(G) + 1/2 - a_n > 0$.

If $a_n \equiv 1/2$, then G is the support of the density f_G, and we get the support estimation problem as first considered in the papers of Geffroy (1964) and Rényi and Sulanke (1963,1964). Geffroy (1964) studied a piecewise-constant support estimator for two-dimensional boundary fragments. Rényi and Sulanke (1963, 1964) investigated the case of a convex two-dimensional support G, and they

proposed to use the convex hull of sample points as estimator. Their results were generalized to the multidimensional case and extended in several ways by Efron (1965), Bosq (1971), and Chevalier (1976).

Support estimation is useful for several applications. We mention here two of them. First, as proposed by Devroye and Wise (1980), one can apply support estimation for detecting the abnormal behaviour of a system, plant or machine. If $\hat{G}_n$, the estimator of the support G, is given then one can check whether a new observation X is driven from a rather different distribution as $X_1, \ldots, X_n$. The system is decided to behave abnormally if $X \notin \hat{G}_n$.

The second application of support estimation is cluster analysis. One can define clusters as level sets of densities, i.e. the sets where the density $f(x)$ of X_i's is larger than some constant $\kappa > 0$ (see Hartigan (1975), Müller and Sawitzki (1990, 1991)). In the case of uniform density f all level sets $V_\kappa = \{f(x) \geq \kappa\}$ coincide with the support G for κ small enough. Note that we restrict ourselves to two-valued densities to make the relation to edge estimation more transparent. In general, the accuracy of level set estimation is influenced by the smoothness properties of f. The interesting problem of optimal rates in level sets estimation is not yet solved for general smoothness classes. Some upper bounds on the risks are given by Hartigan (1987) and Polonik (1991).

In cluster analysis the assumption $a_n \to 0$ may be appropriate. It accounts for situations where not all observed sample points are lying in the cluster domain.

Several statistical problems are related to support estimation. First, we mention the estimation of the silhouette and of the excess mass (Hartigan (1987), Müller and Sawitzki (1990, 1991), Polonik (1991)). For $\kappa > 0$ the excess mass is defined as $\int \max\{f(x) - \kappa, 0\} dx$. The silhouette is the relation $\kappa \to \{x : f(x) > \kappa\}$. For the uniform density on G one gets $\max\{1 - \kappa\lambda(G), 0\}$, thus excess mass estimation reduces to the estimation of the area of the support G. Another realted problem is finding the edge of a Poisson forest. This problem posed by D.G. Kendall is the following: given the realization of a planar Poisson point process with unknown positive intensity within a compact set G and zero intensity outside G, estimate G. For convex G this problem was studied by Ripley and Rasson (1977) and Moore (1984), and for some other types of domains by Jacob (1984), Jacob and Abbar (1989) and Abbar (1990).

4. Upper risk bounds

The next theorem gives upper bounds for the risks of estimators in the smoothness class $\mathcal{G}_{\gamma,L}$.

THEOREM 4.1. *Let $\mathcal{G}_n \subseteq \mathcal{G}_{\gamma,L}$. Suppose that either*
(i) Model 1 holds, and X_i are independent and uniformly distributed on $[0,1]^N$, or

(ii) Model 2 holds, and
either : $a_n \equiv 1/2$ and $\lambda(G) \geq \lambda_0$ for some fixed $\lambda_0 > 0$ for all sets $G \in \mathcal{G}_n$,
or : $\lim_{n\to\infty} a_n = 0$, all sets $G \in \mathcal{G}_n$ are contained in a set $\mathcal{D} \subset [0,1]^N$
with

$$\lambda(\mathcal{D}) < 1. \tag{4.1}$$

Then there exists an estimate G_n^* such that for every $q > 0$

$$\sup_{G\in\mathcal{G}_n} E_G(d_1^q(G_n^*, G)) \leq C_1(na_n^2)^{-q\gamma/(\gamma+N-1)}. \tag{4.2}$$

Here $C_1 > 0$ is a constant which depends only on q, N, γ, and L; E_G denotes the expectation with respect to the distribution which corresponds to a set G according to Model 1 or 2, respectively. For two sets G_1 and G_2 the distance $d_1(G_1, G_2)$ is the Lebesgue measure of the symmetric difference:

$$d_1(G_1, G_2) = \lambda(G_1 \Delta G_2).$$

It will be shown in Section 5 that the rates of G_n^* given in Theorem 4.1 are optimal if $na_n^2 \to \infty$. Theorem 4.1 generalizes a result in Korostelev and Tsybakov (1992, 1993) and Tsybakov (1991) where for Model 1 some special subclasses of $\mathcal{G}_{\gamma,L}$ (boundary fragments) were considered, with $a_n \equiv$const.

Let us now define the estimator G_n^* that we use in the proof of Theorem 4.1. For doing this we need some notation. Let $\mathcal{G}$ be a compact set of domains G endowed with a metric d. Denote by $\mathcal{N}(\mathcal{G}, d, \varepsilon)$ a minimal ε-net on $\mathcal{G}$ in d-metric. The estimator G_n^* of Theorem 4.1 is the maximum likelihood estimator (MLE) on the ε-net $\mathcal{N}_\varepsilon = \mathcal{N}(\mathcal{G}_n, d_1, \varepsilon)$. The value of ε depends on n, γ, and L (see the

proof of Theorem 4.1). For Models 1 and 2 the maximum likelihood estimators on ε-nets $\mathcal{N}_\varepsilon$ are of the form:

$$G_n^* = \arg\min_{G\in\mathcal{N}_\varepsilon} \sum_{i=1}^{n} [I\{X_i \in G\}\ I\{Y_i = -1\} + I\{X_i \in \overline{G}\}\{Y_i = 1\}] \tag{4.3}$$

(for Model 1),

and

$$G_n^* = \arg\min_{G\in\mathcal{N}_\varepsilon} [2a_n\lambda(G) + 1/2 - a_n][1/2 + a_n]^{-\lambda_n(G)}[1/2 - a_n]^{-1+\lambda_n(G)}, \tag{4.4}$$

where $\lambda_n(G) = n^{-1}\#\{i : X_i \in G\}$ (for Model 2).

In particular, for the case $a_n = 1/2$ the minimisation in (4.4) reduces to the sets G with $\lambda_n(G) = 1$, and thus

$$G_n^* = \arg\min_{G\in\mathcal{N}_\varepsilon} \{\lambda(G) : G \supseteq \{X_1, \ldots, X_n\}\}.$$

REMARK 1. In the part (i) of Theorem 4.1 one can consider also other assumptions on the distribution of the X_i's. In particular, one can assume that the X_i's are independent and uniformly distributed on pairwise disjoint cubes I_i's. More explicitly: for a subsequence of n such that $r = n^{1/N} \in \mathbb{N}$, construct a grid on $[0,1]^N$ of n cubes with side length r^{-1}. Denote the cubes by $I_1, \ldots, I_n$. Now assume that the X_i's are independent and that X_i has a uniform distribution on I_i. Then one can show as in the proof of Theorem 4.1 that there exists an estimate G_n^* with rate $(na_n^2)^{-\gamma/(\gamma+N-1)}$. This rate can in general not be achieved if the X_i's are nonrandom and equal to the midpoints of the I_i's. This can be shown as in Section 3.1 of Korostelev and Tsybakov (1992).

REMARK 2. For Model 1 Theorem 4.1 remains valid for a more general class of error distributions. In fact, the theorem holds for more general models of type (3.2) where ζ_{in} are independent zero mean random variables with

$$\sup_{n\geq 1, 1\leq i\leq n} E(\exp(t\zeta_{in})) < \infty \quad \text{for } t \text{ small enough.}$$

For such models G_n^* can then be chosen as the least squares estimate on an ε-net $\mathcal{N}_\varepsilon$,

$$G_n^* = \arg\min_{G\in\mathcal{N}_\varepsilon} \sum_{i=1}^{n} (Y_i - 2a_n f_G(X_i))^2,$$

where f_G is defined as in (3.2). In our model this coincides with the estimate given in (4.3).

REMARK 3. The estimator G_n^* given by (4.3) or (4.4) is primarily of theoretical value. Algorithms for the numerical calculation of this estimator would have exponential complexity (if they would be based on a global search on $\mathcal{N}_\varepsilon$). Estimators which can be calculated easier and which achieve the optimal rates exist for several submodels. For instance, for sets with Lipschitz boundaries ($\gamma = 1$), edge estimators with optimal rates can be based on a simple classification (Tsybakov (1989,1991)). We conjecture also that for Model 1 the algorithm of Nagao and Matsuyama (1979) has an optimal rate if $\gamma = 1$. For smoother boundaries ($\gamma > 1$) the calculation of optimal estimators seems to be a hard numerical problem. In Korostelev and Tsybakov (1992, 1993) algorithms are discussed for boundary fragments. Another example is given by the classes of "star-shaped" two-dimensional sets G. A set G is called *star-shaped* (with smoothness of order γ) if its arithmetical center $x^* = \lambda(G)^{-1} \int_G x dx$ is an inner point of G, and the boundary $\Gamma = \partial G$ can be represented as

$$\Gamma = \{x \in [0,1]^2 : x = x^* + \Delta,\ \Delta = (g(\phi)\cos\phi, g(\phi)\sin\phi), 0 \le \phi < 2\pi\},$$

where $g(\phi) > 0$ is a function having the smoothness parameter γ. In the polar coordinate system centered at x^* one can represent Γ as a curve $r = g(\phi)$ where r is the radius and ϕ is the polar angle. In this coordinate system we define the new data $X_i' = (r_i, \phi_i)$ where ϕ_i is the polar angle corresponding to X_i and $r_i = g(\phi)$. Clearly, if G is contained in $[0,1]^2$ then $0 \le r_i < \sqrt{2}$ and $0 \le \phi_i < 2\pi$. Thus, for x^* known, the estimation of G is reduced to edge estimation for boundary fragments on $K = [0, \sqrt{2}] \times [0, 2\pi]$, and the optimal procedures of Korostelev and Tsybakov (1992, 1993), Korostelev (1991), and Tsybakov (1991) can be applied with minor changes. The estimation of star-shaped sets with known center x^* was studied by Rudemo, Skovgaard and Stryhn (1990) who proved consistency of a piecewise-constant estimator of G for $\gamma = 1$ and discussed some interesting applications (see also Rudemo and Stryhn (1991)).

If the center x^* is unknown one has to use an estimator $\widehat{x^*}$ of x^*. For instance, put

$$\widehat{x^*} = (\lambda(G_n))^{-1} \int_{G_n} x dx$$

where G_n is a consistent estimator of the domain G. Such an estimate G_n can be found in Tsybakov (1989), see also Korostelev and Tsybakov (1993). It has the

rate of convergence $\mathcal{O}((\log n/n)^{1/N})$ which is insufficient since also the estimator $\widehat{x^*}$ and the final edge estimator inherit the same rate. To avoid this effect let us make a reasonable change in the definition of star-shaped sets. Assume that there exists a fixed open subset G^* of G, with x^* as an inner point, and that for any $x_0 \in G^*$ the boundary Γ can be represented as

$$\Gamma = \{x \in [0,1]^2 : x = x_0 + \Delta,\ \Delta = (g(x_0,\phi)\cos\phi, g(x_0,\phi)\sin\phi), 0 \le \phi < 2\pi\},$$

where $g(x_0,\phi)$ has the smoothness parameter γ for all $x_0 \in G^*$. Now, if G_n is just a consistent estimator of G in d_1-metric then for n large enough $\widehat{x^*} \in G^*$ with probability close to 1. Hence, the problem can be reduced to boundary fragments as described earlier, with $\widehat{x^*}$ as a central point.

PROOF OF THEOREM 4.1. Write for brevity $\mathcal{G} = \mathcal{G}_n$. We apply Lemma A.1 of the Appendix. Consider first Model 1. Let us check (A.3). In our case $\Theta = \mathcal{G},\ \theta = G$, and

$$p(Z,\theta) = \prod_{i=1}^{n} p(X_i, Y_i, G),$$
$$p(X_i, Y_i, G) = \begin{cases} 1 + 2a_n \text{ if } Y_i = 1,\ X_i \in G \text{ or } Y_i = -1,\ X_i \in \overline{G} \\ 1 - 2a_n \text{ if } Y_i = -1,\ X_i \in G \text{ or } Y_i = 1,\ X_i \in \overline{G} \end{cases}$$

Here $Z = ((X_1,Y_1),\dots,(X_n,Y_n))$, and $p(Z,G)$ is the density of the distribution of Z with respect to dominating measure $\nu = (\nu_0 \times \nu_1)^n$ where ν_0 is the Lebesgue measure on $[0,1]^N$ and ν_1 is the measure concentrated at points $1, -1$, with $\nu_1\{1\} = \nu_1\{-1\} = 1/2$.

We prove (A.3) with $\alpha = 1, t_0 = 2$. For this we have to consider the sets $\theta = G,\quad \tilde{\theta} = \tilde{G}$ and $\theta' = G'$ such that $d_1(G,\tilde{G}) \le \varepsilon,\ d_1(G,G') \ge \varepsilon t$, for some $t \ge t_0$. Then

$$P_\theta\{p(Z,\theta') \ge p(Z,\tilde{\theta})\} = P_G\{(\frac{1}{2} + a_n)^S (\frac{1}{2} - a_n)^{-S} \ge 1\} = P_G\{S \ge 0\}$$

where
$$S = \sum_{i=1}^{n} (I\{Y_i = 1,\ X_i \in G'\} + I\{Y_i = -1,\ X_i \notin G'\}) -$$
$$- \sum_{i=1}^{n} \left(I\{Y_i = 1,\ X_i \in \tilde{G}\} + I\{Y_i = -1,\ X_i \notin \tilde{G}\} \right).$$

Note that

$$
\begin{aligned}
(4.5) \qquad E_G\{S\} =& 2na_n[\lambda(G \cap (G' \setminus \tilde{G})) - \\
& - \lambda(\overline{G} \cap (G' \setminus \tilde{G})) - \lambda(G \cap (\tilde{G} \setminus G')) + \\
& + \lambda(\overline{G} \cap (\tilde{G} \setminus G'))] = 2na_n[-d_1(G, G') + d_1(G, \tilde{G})] \leq \\
& \leq 2na_n[-d_1(G, G') + \varepsilon].
\end{aligned}
$$

Next, note that $S = \sum_{i=1}^{n} \eta_i$, where η_i's are independent random variables, $|\eta_i| \leq 1$, and Var $\{\eta_i\} \leq E\{\eta_i^2\} = d_1(\tilde{G}, G') \leq d_1(G, G') + d_1(G, \tilde{G}) \leq d_1(G, G') + \varepsilon$. Applying the Bernstein inequality and using (4.5) we find

$$
\begin{aligned}
& P_G\{S \geq 0\} \leq P_G\{\frac{1}{n} \sum_{i=1}^{n} (\eta_i - E\{\eta_i\}) \geq 2a_n(d_1(G, G') - \varepsilon)\} \\
& \leq \exp\left(-\frac{n[2a_n(d_1(G, G') - \varepsilon)]^2}{4(d_1(G, G') + \varepsilon)}\right).
\end{aligned}
$$

Put $t_0 = 2$. Then for $d_1(G, G') \geq t\varepsilon,\ t \geq t_0$, we have

$$
P_G\{S \geq 0\} \leq \exp\left(-\frac{1}{6} n a_n^2 \varepsilon t\right)
$$

which proves (A.3) with $c_{d1} = 1,\ c_{d2} = na_n^2/6,\ \alpha = 1$.

Now, let us check the condition (A.4). Dudley (1974) proved that there exists a constant $C_{\mathcal{H}} > 0$ depending on γ, L such that the ε-entropy

$$
\mathcal{H}(\mathcal{G}_{\gamma,L}, d_1, \varepsilon) = \log\ (\text{card}\ \mathcal{N}(\mathcal{G}_{\gamma,L}, d_1, \varepsilon))
$$

satisfies

$$
(4.6) \qquad \mathcal{H}(\mathcal{G}_{\gamma,L}, d_1, \varepsilon) \leq C_{\mathcal{H}} \varepsilon^{-(N-1)/\gamma}.
$$

Hence, for any $\mathcal{G}_n \subseteq \mathcal{G}_{\gamma,L}$

$$
\mathcal{H}(\mathcal{G}_n, d_1, \varepsilon) \leq C_{\mathcal{H}} \varepsilon^{-(N-1)/\gamma}.
$$

Choose $\varepsilon = (C_{\mathcal{H}}/na_n^2)^{\frac{\gamma}{\gamma+N-1}}$. Then

$$
\mathcal{H}(\mathcal{G}_n, d_1, \varepsilon) \leq C_{\mathcal{H}} \varepsilon^{-(N-1)/\gamma} = na_n^2 \varepsilon = \frac{1}{2} c_{d2} \varepsilon,
$$

which proves (A.4) with $c_{p1} = 1$ and $c_{p2} = 0$. Hence we can apply Lemma A.1. In our case $D = 1,\ \alpha = 1$, and (A.5) implies

$$\sup_{G\in\mathcal{G}_n} E_G(\varepsilon^{-q} d_1^q(G_n^*, G)) \le \max(t_0, t_1) + \varepsilon^{-q} \exp(-\frac{1}{12} n a_n^2 \varepsilon).$$

If we substitute here the expression for ε we get

$$\sup_{G\in\mathcal{G}_n} E_G(d_1^q(G_n^*, G)) \le C_1 (n a_n^2)^{-q\gamma/(\gamma+N-1)},$$

where C_1 is a positive constant. This proves (4.2) for Model 1.

Consider now Model 2. We apply Lemma A.1 with $Z = (X_1, \ldots, X_n)$, ν being the N-product of Lebesgue measures on $[0,1]^N$, and

$$p(Z, G) = \prod_{i=1}^{n} f_G(X_i).$$

Let us check Assumption (I) of the Appendix. For showing (A1) we have to consider

$$\int \sqrt{p(Z,G)p(Z,G')}d\nu = [\int \sqrt{f_G(x) f_{G'}(x)}dx]^n. \tag{4.7}$$

Now, denoting $\overline{\lambda}(G) = 2a_n\lambda(G) + \frac{1}{2} - a_n$ we get

$$\begin{aligned} &\int \sqrt{f_{G'}(x) f_G(x)}dx \\ &= \frac{1}{\sqrt{\overline{\lambda}(G)\overline{\lambda}(G')}}[\lambda(G\cap G')(\frac{1}{2} + a_n) \\ &+ (\lambda(G \setminus G') + \lambda(G' \setminus G))\sqrt{\frac{1}{4} - a_n^2} + \lambda([0,1]^N \setminus (G \cup G'))(\frac{1}{2} - a_n)] \\ &= \frac{\overline{\lambda}(G\cap G') + d_1(G, G')[\sqrt{\frac{1}{4} - a_n^2} - (\frac{1}{2} - a_n)]}{\sqrt{\overline{\lambda}(G)\overline{\lambda}(G')}}. \end{aligned} \tag{4.8}$$

Consider first the case $a_n = 1/2$. If $\lambda(G \cap G') = 0$, then

$$\int \sqrt{f_G(x) f_{G'}(x)}dx = 0,$$

and (A.1) holds. If $\lambda(G \cap G') \neq 0$, then

$$\int \sqrt{f_G(x) f_{G'}(x)} dx = \frac{\lambda(G \cap G')}{\sqrt{\lambda(G)\lambda(G')}}$$

$$\leq \sqrt{\frac{\lambda^2(G \cap G')}{\lambda^2(G \cap G') + \lambda(G \cap G') d_1(G, G')}}$$

$$= (1 - \frac{d_1(G, G')}{\lambda(G \cap G') + d_1(G, G')})^{1/2}$$

$$\leq (1 - d_1(G, G'))^{1/2} \leq \exp(-d_1(G, G')/2).$$

Hence, for $a_n = 1/2$, we have

$$\int \sqrt{p(Z, G) p(Z, G')} d\nu \leq \exp(-n d_1(G, G')/2). \tag{4.9}$$

Assume now that $\lim_{n \to \infty} a_n = 0$. Then

$$\overline{\lambda}(G \cap G') \geq \frac{1}{2} - a_n > 0 \text{ for } n \text{ large enough.}$$

Note that

$$\overline{\lambda}(G)\overline{\lambda}(G') \geq \overline{\lambda}^2(G \cap G') + 2a_n \overline{\lambda}(G \cap G') d_1(G, G'). \tag{4.10}$$

Denote for brevity $\overline{\lambda} = \overline{\lambda}(G \cap G'),\ d_1 = d_1(G, G')$. From (4.8) and (4.10) we get

$$\int \sqrt{f_G(x) f_{G'}(x)} \leq \left(\frac{\overline{\lambda}^2 + 2(a_n - a_n^2) d_1 \overline{\lambda} + d_1^2 (a_n - a_n^2)^2}{\overline{\lambda}(G)\overline{\lambda}(G')} \right)^{1/2}$$

$$\leq \left(\frac{\overline{\lambda}^2 + 2(a_n - a_n^2) d_1 \overline{\lambda}}{\overline{\lambda}^2 + 2a_n \overline{\lambda} d_1} + \frac{d_1^2 a_n^2}{(\frac{1}{2} - a_n)^2} \right)^{1/2}$$

$$= \left(1 - \frac{2a_n^2 d_1}{\overline{\lambda} + 2a_n d_1} + \frac{d_1^2 a_n^2}{(\frac{1}{2} - a_n)^2} \right)^{1/2}.$$

Clearly,

$$\overline{\lambda}(G \cap G') + 2a_n d_1(G, G') = a_n [2\lambda(G \cup G') - 1] + \frac{1}{2} \leq a_n + \frac{1}{2}.$$

Also, by (4.1), we have for a $\delta_0 > 0$ that $d_1(G, G') \le \lambda(\mathcal{D}) \le 1 - \delta_0$. Together with (4.11) this entails

$$\int \sqrt{f_G(x) f_{G'}(x)} dx \le \left(1 - \frac{2a_n^2 d_1}{\frac{1}{2} + a_n} + \frac{(1-\delta_0) d_1 a_n^2}{(\frac{1}{2} - a_n)^2}\right)^{1/2} = (1 - a_n^2 d_1(G, G')\kappa_n)^{1/2}$$

where $\kappa_n = \left[\frac{2}{1/2 + a_n} - \frac{1-\delta_0}{(1/2 - a_n)^2}\right] \ge 4[\delta_0 + \mathcal{O}(a_n)] \ge 2\delta_0$ for n large enough.

Hence, for n large enough, we have

$$\int \sqrt{f_G(x) f_{G'}(x)} dx \le (1 - 2\delta_0 a_n^2 d_1(G, G'))^{1/2} \le \exp(-\delta_0 a_n^2 d_1(G, G')). \tag{4.12}$$

It follows from (4.7), (4.9), (4.12) that for both cases $a_n \equiv 1/2$, and $\lim_{n\to\infty} a_n = 0$, we have

$$\int \sqrt{p(Z, G) p(Z, G')} d\nu \le \exp(-c_{d2} d_1(G, G'))$$

where $c_{d2} = n a_n^2 \min(\delta_0, 1/2)$. This gives (A.1).

Now we have to prove (A.2), namely, that for every $G \in \mathcal{G}_n$ there exists $\tilde{G} \in \mathcal{N}_\varepsilon$ with $d_1(G, \tilde{G}) \le \varepsilon$ and

$$P_G\{c_{p1} p(Z, G) \ge p(Z, \tilde{G})\} \le c_{p2}. \tag{4.13}$$

Denote by $d_\infty(G, \tilde{G})$ the Hausdorff distance between the sets G and $\tilde{G}$:

$$d_\infty(G, \tilde{G}) = \max\left\{\max_{x \in \tilde{G}} \rho(x, G), \max_{x \in G} \rho(x, \tilde{G})\right\}$$

where $\rho(x, G)$ is the Euclidean distance between the point x and the set G.

Note that for $G \in \mathcal{G}_{\gamma, L}$ we have: $d_\infty(G, \tilde{G}) \le \varepsilon$ entails that $d_1(G, \tilde{G}) \le \overline{C}\varepsilon$, where $\overline{C} > 0$ does not depend on G and $\tilde{G}$. This follows from the fact that the boundary of G is piecewise Lipschitzian under our assumptions (we assume everywhere that $\gamma \ge 1$).

To prove (A.2) we use a special construction of ε-net $\mathcal{N}_\varepsilon$ which could be called "ε-net with one-sided bracketing". Consider some minimal $\varepsilon/2\overline{C}$-net $\mathcal{N}^0_{\varepsilon/2\overline{C}}$ on $\mathcal{G}_n$ in Hausdorff metric d_∞. For every set $G \subset [0,1]^N$ define its ε-neighbourhood as

$$\mathcal{O}_\varepsilon(G) = \{x \in [0,1]^N : \rho(x, G) \le \varepsilon\}.$$

Clearly, the set

$$\mathcal{N}_\varepsilon = \{\tilde{G} = \mathcal{O}_{\varepsilon/2\overline{C}}(G') : G' \in \mathcal{N}^0_{\varepsilon/2\overline{C}}\}$$

is the $\varepsilon/\overline{C}$-net for $\mathcal{G}_n$ in Hausdorff metric and also for every $G \in \mathcal{G}_n$ there exists $\tilde{G} \in \mathcal{N}_\varepsilon$ which satisfies $G \subseteq \tilde{G}$, $d_\infty(G, \tilde{G}) \leq \varepsilon/\overline{C}$. Namely, one can take $\tilde{G} = \mathcal{O}_{\varepsilon/2\overline{C}}(G')$ where G' is the element of $\mathcal{N}^0_{\varepsilon/2\overline{C}}$ closest to G in Hausdorff metric. Then

$$G \subseteq \tilde{G}, \quad d_1(G, \tilde{G}) \leq \varepsilon. \tag{4.14}$$

Let us prove that for the sets $\tilde{G}$ defined for each G in such a way that (4.14) is true we have the first inequality in (4.13) with certain c_{p1}, c_{p2}. Let $a_n \equiv 1/2$. Since $G \subseteq \tilde{G}$ it is clear that (4.13) is satisfied if we have

$$c_{p1} \leq \left(\frac{\lambda(G)}{\lambda(\tilde{G})}\right)^n.$$

Note that this is true with $c_{p1} = \left(1 - \dfrac{2\varepsilon}{\lambda_0}\right)^n$ if ε is small enough. In fact, since $|\lambda(\tilde{G}) - \lambda(G)| \leq d_1(\tilde{G}, G) \leq \varepsilon,\ \lambda(G) \geq \lambda_0$, we have

$$\left(\frac{\lambda(G)}{\lambda(\tilde{G})}\right)^n = \left(1 - \frac{\lambda(\tilde{G}) - \lambda(G)}{\lambda(\tilde{G})}\right)^n \geq \left(1 - \frac{\varepsilon}{\lambda_0 - \varepsilon}\right)^n,\ \varepsilon < \lambda_0.$$

Assume now $a_n \to 0$. Then for ε and a_n small enough (4.13) holds with $c_{p1} = \exp(-10na_n^2\varepsilon)$ and $c_{p2} = \exp(-\frac{1}{5}na_n^2\varepsilon)$. To see this note that

$$\log(p(Z,\theta)/p(Z,\tilde{\theta})) = n\log[\overline{\lambda}(\tilde{G})/\overline{\lambda}(G)] - R\log[(1/2 + a_n)/(1/2 - a_n)] \tag{4.15}$$

where $R = \sum_{i=1}^{n} I(X_i \in \tilde{G} \setminus G)$ is a Binomial random variable with parameter $(1/2 - a_n)\lambda(\tilde{G} \setminus G)/\overline{\lambda}(G)$. Taylor expansions show that for ε and a_n small enough the expectation of the left hand side of (4.15) is absolutely bounded by $9na_n^2\varepsilon$. Hence (4.13) with the specified values of c_{p1} and c_{p2} follows by application of Bernstein's inequality.

It remains to find an $\varepsilon > 0$ such that (A.4) is satisfied. Using (4.5) and arguing as in the case of Model 1 we get that it is sufficient to find $\varepsilon > 0$ such that $-\dfrac{1}{2}\log c_{p1} + C_{\mathcal{H}}\varepsilon^{-(N-1)/\gamma} \leq \dfrac{1}{2}c_{d2}\varepsilon t$. If $a_n \equiv 1/2$ this is equivalent to

$$-\frac{n}{2}\log\left(1 - \frac{2\varepsilon}{\lambda_0}\right) + C_{\mathcal{H}}\varepsilon^{-(N-1)/\gamma} \leq \frac{n}{4}\varepsilon t$$

since $c_{d2} = n/2$ in view of (4.9). Now, $-\log\,(1 - \dfrac{2\varepsilon}{\lambda_0}) \le \dfrac{4\varepsilon}{\lambda_0}$ for ε small enough. The inequality (4.15) is satisfied if

$$\frac{4n\varepsilon}{\lambda_0} + C_{\mathcal{H}}\varepsilon^{-(N-1)/\gamma} \le \frac{n}{4}\varepsilon t. \tag{4.16}$$

If we choose $\varepsilon = \text{const}\ n^{-\gamma/(\gamma+N-1)}$, then there exists such $t_1 > 1$ that (4.16) holds for $t \ge t_1$. This proves (A.4).

For the case $a_n \to 0$ we prove (A.4) similarly, with $\varepsilon = \text{const}\ (na_n^2)^{-\gamma/(\gamma+N-1)}$.

■

Our approach can easily be extended to other classes $\mathcal{G}_n$ which fulfil the certain entropy conditions. Let us mention the class $\mathcal{G}_{\text{conv}}$ of all closed convex subsets of $[0,1]^N$. The following theorem gives an upper bound for the optimal convergence rate in $\mathcal{G}_{\text{conv}}$.

THEOREM 4.2. *Let $G \in \mathcal{G}_n \subseteq \mathcal{G}_{\text{conv}}$. Assume condition* (i) *or* (ii) *(resp.) of Theorem 4.1. Then for the estimate G_n^* defined by (4.3) or (4.4) (resp.) with an appropriate choice of $\varepsilon = \varepsilon_n$ we have for $q > 0$*

$$\sup_{G \in \mathcal{G}_n} E_G(d_1^q(G_n^*, G)) = \mathcal{O}((na_n^2)^{-2q/(N+1)}).$$

PROOF OF THEOREM 4.2. We proceed as in the proof of Theorem 4.1, but we use another entropy bound. Instead of (4.5) we apply (see Dudley (1974))

$$\mathcal{H}(\mathcal{G}_{\text{conv}}\,, d_1, \varepsilon) = \mathcal{O}(\varepsilon^{-(N-1)/2}). \tag{4.17}$$

■

REMARK 4. It is interesting to compare Theorem 4.2 with the results of Rényi and Sulanke (1963,1964). They studied the problem of support estimation (Model 2) assuming that $G \in \mathcal{G}_{\text{conv}}\,, N = 2$, and $a_n \equiv 1/2$. The maximum likelihood estimator is then the convex hull of $X_1, \ldots, X_n$:

$$G_n^{\text{MLE}} = \arg \min_{\substack{G \in \mathcal{G}_{\text{conv}}\,: \\ G \supseteq \{X_1, \ldots, X_n\}}} \lambda(G) = \text{conv}\ \{X_1, \ldots, X_n\}. \tag{4.18}$$

Rényi and Sulanke (1964) assumed that the edge ∂G is twice continuously differentiable and proved that for dimension $N = 2$ the estimator (4.18) satisfies

$$E_G\left(\lambda(G_n^{\text{MLE}}\,)\right) - \lambda(G) = \mathcal{O}(n^{-2/3}),\ n \to \infty. \tag{4.19}$$

Since G is convex, we have that G_n^{MLE} is contained in G. Hence (4.19) implies

$$E_G\left(d_1(G_n^{\mathrm{MLE}}, G)\right) = \mathcal{O}(n^{-2/3}),\ n \to \infty.$$

Thus, G_n^{MLE} has the same rate of convergence as our estimator G_n^* for the case $N = 2$ (see Theorem 4.2 with $q = 1$). In the next section we show that the upper bounds of Theorems 4.1 and 4.2 cannot be improved. This entails, in particular, that G_n^{MLE} has optimal rate of convergence in $\mathcal{G}_{\mathrm{conv}}$ for $N = 2$. We are not aware of a generalization of this result for $N > 2$.

For estimation of the area $\lambda(G)$ (which, in fact, was the original problem of Rényi and Sulanke, 1964) the estimator $\lambda(G_n^{\mathrm{MLE}})$ is not optimal. For $N = 2$ one gets here the optimal rate $n^{-5/6}$ (see Korostelev and Tsybakov (1993), Ch. 8).

REMARK 5. Note that we have the same optimal rates for $\mathcal{G}_{2,L}$ and $\mathcal{G}_{\mathrm{conv}}$. This is not surprising: one comes to a similar conclusion in nonparametric density estimation and regression, see Mammen (1991a,b), Nemirovskii, Polyak and Tsybakov (1985).

REMARK 6. Theorems 4.1 and 4.2 can be applied to the model of a Poisson forest (see above). In this model one has a Poisson random number of observations, with the distribution of each observation as in Model 2. Thus, the results of Theorems 4.1 and 4.2 hold with n replaced by the expected sample size. The same is true for the lower bounds given in the next section.

5. Lower bounds for minimax risks

In this section we show that the rates given in Theorem 4.1 and 4.2 are optimal. For the tuple of points $\mathcal{X} = (X_1, \ldots, X_n)$ in Model 1 we assume in this section only that $\mathcal{X}$ is an arbitrary collection of random points in $[0,1]^N$ and that $\mathcal{X}$ is independent of $\xi_1, \ldots, \xi_n$.

THEOREM 5.1. *Let $na_n^2 \geq 1$. Assume Model 1 or Model 2. Then for $q \geq 1$*

$$\liminf_n \inf_{\hat{G}_n} \sup_{G \in \mathcal{G}_{\gamma,L}} (na_n^2)^{q\gamma/(\gamma+N-1)} E_G(d_1^q(\hat{G}_n, G)) \geq C_0 > 0, \tag{5.1}$$

where the infimum is taken over all estimates $\hat{G}_n$ of G, and C_0 is a constant which depends on q, N, γ, and L, and which, for Model 1, does not depend on the distribution of $\mathcal{X}$.

PROOF THEOREM 5.1. We proceed similarly as in the proof of Theorem 2 in Tsybakov (1989), see also Korostelev and Tsybakov (1992, 1993). It is sufficient to prove the theorem for $q = 1$ since $E_G(d_1^q(\hat{G}_n, G)) \geq (E_G(d_1(\hat{G}_n, G)))^q$ for $q \geq 1$.

Consider first Model 1. Let $0 \leq \eta(t) \leq 1$ be an infinitely many times continously differentiable function on $\mathbb{R}^{N-1}$ with support $(-1/2, 1/2)^{N-1}$. Put $m = [(na_n^2)^{1/(\gamma+N-1)}] + 1$ and $M = m^{N-1}$. For $\lambda_0,\ \omega = (\omega_1, \ldots, \omega_M)$, and for an integer r with $0 < \lambda_0 < 1,\ \omega_j \in \{0, 1\}\ (j = 1, \ldots, M)$, and $0 \leq r \leq m^\gamma - 1$ we introduce the family of sets

$$G^r(\omega) = \{x = (x_1, \ldots, x_N) \in [0, 1]^N :\ 0 \leq x_N \leq g_r(x_1, \ldots, x_{N-1}, \omega)\}$$

where

$$g_r(t, \omega) = \frac{r}{m^\gamma} + \frac{\lambda_0}{m^\gamma} \sum_{j=1}^{M} \omega_j \eta(m(t - b_j)). \tag{5.2}$$

Here t is the vector of the first $N - 1$ coordinates of x, i.e. $t = (x_1, \ldots, x_{N-1})$. Furthermore $B_m^{N-1} = \{b_1, \ldots, b_m\}$ is the regular grid on $[0, 1]^{N-1}$, i.e.

$$B_m^{N-1} = \{t = (x_1, \ldots, x_{N-1}) : mx_j - 1/2 \in \{0, 1, \ldots, m - 1\},\ j = 1, \ldots, N - 1\}.$$

For λ_0 small enough the sets $G^r(\omega)$ belong to $\mathcal{G}_{\gamma,L}$ for all ω.

For $k \in \{0, 1\},\ j = 1, \ldots, M$, define $\omega_0^{j,k}$ as the vector ω with $\omega_i = 0$ for $i \neq j$ and $\omega_j = k$, i.e. $\omega_0^{j,k} = (0, \ldots, 0, \underbrace{k}_{j}, 0, \ldots, 0)$. We write

$$\omega^j = (\omega_1, \ldots, \omega_{j-1}, \omega_{j+1}, \ldots, \omega_M)$$

and

$$\mathcal{T}_{jr} = \{x \in [0, 1]^N : g_r(x_1, \ldots, x_{N-1}, \omega_0^{j,0}) \leq x_N \leq g_r(x_1, \ldots, x_{N-1}, \omega_0^{j,1})\}.$$

The basic idea of the proof is to introduce a probability distribution on the set of all ω. In particular, we assume that the ω_j's are i.i.d. Bernoullian random variables. Denote by E_ω the expectation with respect to this distribution and by $E_\mathcal{X}$ the expectation with respect to $\mathcal{X} = (X_1, \ldots, X_n)$. Since the maximum is greater than the expectation we have

$$\begin{aligned}
\sup_{G \in \mathcal{G}_{\gamma,L}} E_G(d_1(\hat{G}_n, G)) &\geq \max_{r,\omega} E_{G^r(\omega)}(d_1(G^r(\omega), \hat{G}_n)) \\
&\geq \frac{1}{m^\gamma} E_\mathcal{X} \left[E_\omega(\sum_{r=0}^{m^\gamma - 1} E_{G^r(\omega)}(d_1(G^r(\omega), \hat{G}_n)|\mathcal{X})) \right] \\
&\geq \frac{1}{m^\gamma} E_\mathcal{X}[E_\omega\{ \sum_{r=0}^{m^\gamma - 1} \sum_{j=1}^{M} E_{G^r(\omega)}(d_1(G^r(\omega) \cap \mathcal{T}_{jr}, \hat{G}_n \cap \mathcal{T}_{jr})|\mathcal{X}\}]
\end{aligned} \tag{5.3}$$

We show now that the expression in square brackets in (5.3) is greater than a constant that does not depend on $\mathcal{X}$. From now on, assume that $\mathcal{X}$ is fixed. We have

$$(5.4)\quad E_\omega\left[E_{G^r(\omega)}(d_1\ (G^r(\omega)\cap\mathcal{T}_{jr},\hat{G}_n\cap\mathcal{T}_{jr})|\mathcal{X})\right] = E_{\omega^j}[E_{\omega_j}E_{G^r(\omega)}(d_1(G^r(\omega)\cap\mathcal{T}_{jr},\hat{G}_n\cap\mathcal{T}_{jr})|\mathcal{X},\omega^j)].$$

Here

$$(5.5)\quad E_{\omega_j}\left[E_{G^r(\omega)}(d_1(G^r(\omega)\cap\mathcal{T}_{jr},\hat{G}_n\cap\mathcal{T}_{jr})|\mathcal{X},\omega^j)\right]$$

$$= \frac{1}{2}\left[E_{G^r(\omega^{j,0})}(d_1(G^r(\omega^{j,0})\cap\mathcal{T}_{jr},\hat{G}_n\cap\mathcal{T}_{jr})|\mathcal{X},\omega^j)\right.$$

$$\left.+E_{G^r(\omega^{j,1})}(d_1(G^r(\omega^{j,1)})\cap\mathcal{T}_{jr},\hat{G}_n\cap\mathcal{T}_{jr})|\mathcal{X},\omega^j)\right]$$

$$\geq \frac{\overline{d}_1}{2}\int\min(\prod_{i\in U_{jr}}dP_{i,G^r(\omega^{j,1})},\prod_{i\in U_{jr}}dP_{i,G^r(\omega^{j,0})})\prod_{i\notin U_{jr}}dP_{i,G^r(\omega^{j,0})}$$

where $\overline{d}_1 = d_1(G^r(\omega^{j,0})\cap\mathcal{T}_{jr},G^r(\omega^{j,1})\cap\mathcal{T}_{jr}), U_{jr}=U_{jr}(\mathcal{X})=\{i:X_i\in\mathcal{T}_{jr}\}$, $P_{i,G}$ is the distribution of Y_i satisfying (3.1) with X_i fixed, and

$$\omega^{j,0}=(\omega_1,\ldots,\omega_{j-1},0,\omega_{j+1},\ldots,\omega_M),$$
$$\omega^{j,1}=(\omega_1,\ldots,\omega_{j-1},1,\omega_{j+1},\ldots,\omega_M).$$

In (5.5) we used the fact that $P_{i,G^r(\omega^{j,0})}=P_{i,G^r(\omega^{j,1})}$ for $i\in U_{jr}$. By Le Cam's inequality (Devroye and Györfi (1985), p. 226) one gets

$$(5.6)\qquad \int\min(dP,dQ)\geq\frac{1}{2}(1-H^2(P,Q)/2)^2$$

where $H(P,Q)$ is the Hellinger distance between two probability measures P and Q. Note that

$$(5.7)\qquad d_1(G^r(\omega^{j,0})\cap\mathcal{T}_{jr},G^r(\omega^{j,1})\cap\mathcal{T}_{jr})=\lambda(\mathcal{T}_{jr}).$$

It follows from (5.4) - (5.7) that

$$(5.8)\quad E_\omega\left[E_{G^r(\omega)}\ (d_1(G^r(\omega)\cap\mathcal{T}_{jr},\hat{G}_n\cap\mathcal{T}_{jr})|\mathcal{X})\right]$$

$$\geq\frac{\lambda(\mathcal{T}_{jr})}{2}E_\omega[\prod_{i\in U_{jr}}(1-H^2(P_{i,G^r(\omega^{j,1})},P_{i,G^r(\omega^{j,0})})/2)^2|\mathcal{X}].$$

Clearly,

$$H^2(P_{i,G^r(\omega^{j,1})}, P_{i,G^r(\omega^{j,0})}) \le \left(\sqrt{\frac{1}{2}+a_n} - \sqrt{\frac{1}{2}-a_n}\right)^2, \tag{5.9}$$
$$\lambda(\mathcal{T}_{jr}) \ge c_1(Mm^\gamma)^{-1}.$$

Here and later $c_i, \quad i = 1, 2, \ldots,$ are positive constants. The first inequality in (5.9) follows from the definition of ξ_i. Now, (5.8) and (5.9) entail

$$\begin{aligned} E_\omega \left[E_{G^r(\omega)}(d_1(G^r(\omega) \cap \mathcal{T}_{jr}, \hat{G}_n \cap \mathcal{T}_{jr})|\mathcal{X})\right] \\ \ge c_2(Mm^\gamma)^{-1}(\frac{1}{2} + \sqrt{\frac{1}{4}-a_n^2})^{2\text{card } U_{jr}} \\ \ge c_2(Mm^\gamma)^{-1}(1-2a_n^2)^{2\text{card } U_{jr}}. \end{aligned} \tag{5.10}$$

An elementary argument (cf. Theorem 2 of Tsybakov (1989)) shows that for $\mathcal{X}$ fixed and for m large enough there exists a subset $\mathcal{D}_{1m} = \mathcal{D}_{1m}(\mathcal{X})$ of the set of integers $\{0 \le r \le m^\gamma - 1\}$ and subsets $\Psi_r = \Psi_r(\mathcal{X})$ of $\{1, \ldots, M\}$ such that

$$\begin{aligned} &\text{card } \mathcal{D}_{1m} \ge c_3 m^\gamma, \\ &\text{card } \Psi_r \ge M/2 \text{ for } r \in \mathcal{D}_{1m}, \\ &\text{card } U_{jr} \le c_4 n/(m^\gamma M) \le \frac{c_4}{a_n^2}, \text{ for } j \in \Psi_r \text{ and } r \in \mathcal{D}_{1m}, \end{aligned} \tag{5.11}$$

where c_3 and $c_4 > 0$ are some constants.

Using (5.3), (5.10), (5.11) we find

$$\begin{aligned} \sup_{G \in \mathcal{G}_{\gamma,L}} E_G(d_1(\hat{G}_n, G)) &\ge \frac{1}{m^\gamma} E_{\mathcal{X}}[E_\omega\{\sum_{r \in \mathcal{D}_{1m}} \sum_{j \in \Psi_r} \frac{c_2}{(Mm^\gamma)}(1-2a_n^2)^{2\text{card } U_{jr}}\}] \\ &\ge \frac{c_3 M}{2} \frac{c_2}{(Mm^\gamma)} \cdot (1-2a_n^2)^{2c_4/a_n^2}. \end{aligned}$$

Since $a_n \le 1/2$ then $(1-2a_n^2)^{2c_4/a_n^2}$ is larger than some constant. The assumption that $na_n^2 \ge 1$ entails the inequality $m^{-\gamma} = ([(na_n^2)^{\frac{1}{\gamma+N-1}}] + 1)^{-\gamma} \ge 2^{-\gamma}(na_n^2)^{-\frac{\gamma}{\gamma+N-1}}$. Thus (5.1) follows.

Consider now Model 2. Introduce the functions (5.2) with $r = (1/2)m^\gamma$. Denote these functions by $g(t, \omega)$. Define the sets

$$G(\omega) = \{x \in [0,1]^N : 0 \le x_N \le g(x_1, \ldots, x_{N-1}, \omega)\},$$

and the densities

$$f_{G(\omega)}(x) = [2a_n I(x \in G(\omega)) + \frac{1}{2} - a_n]/(2a_n\lambda(G(\omega)) + \frac{1}{2} - a_n).$$

Clearly, $G(\omega) \in \mathcal{G}_{\gamma,L}$ for λ_0 small enough, and $\inf_\omega \lambda(G(\omega)) \geq \frac{1}{2}$. Hence

$$2a_n\lambda(G(\omega)) + \frac{1}{2} - a_n \geq \frac{1}{2}. \tag{5.12}$$

Denote the distribution of X_i by $P_{i,G}$ if the underlying set is G.

Using the same argument as in (5.3) - (5.7) we get

$$\begin{aligned} \sup_{G\in\mathcal{G}_{\gamma,L}} E_G(d_1(\hat{G}_n, G)) &\geq \max_\omega E_{G(\omega)}(d_1(G(\omega), \hat{G}_n)) \\ &\geq E_\omega\{\sum_{j=1}^{M} E_{G(\omega)}(d_1(G(\omega)\cap\mathcal{T}_j, \hat{G}\cap\mathcal{T}_j))\} \\ &\geq \frac{1}{2}\sum_{j=1}^{M}\lambda(\mathcal{T}_j)E_\omega[\int \min(\prod_{i=1}^{n} dP_{i,G(\omega^{j,1})}, \prod_{i=1}^{n} dP_{i,G(\omega^{j,0})})] \\ &\geq c_5 m^{-\gamma} E_\omega[\prod_{i=1}^{n}(1 - H^2(P_{i,G(\omega^{j,1})}, P_{i,G(\omega^{j,0})})/2)^2], \end{aligned} \tag{5.13}$$

where

$$\mathcal{T}_j = \{x \in [0,1]^N : g(x_1, \ldots, x_{N-1}, \omega_0^{j,0}) \leq x_N \leq g(x_1, \ldots, x_{N-1}, \omega_0^{j,1})\}$$

and

$$\lambda(\mathcal{T}_j) = c_6(Mm^\gamma)^{-1}.$$

To complete the proof it suffices to estimate from above the Hellinger distance in (5.13). Assume that ω (and hence $\omega^{j,0}$ and $\omega^{j,1}$) are fixed. We have

$$\begin{aligned} H^2(P_{i,G(\omega^{j,1})}, P_{i,G(\omega^{j,0})}) \\ = 2(1 - \int \sqrt{f_{G(\omega^{j,1})}(x) f_{G(\omega^{j,0})}(x)}dx). \end{aligned} \tag{5.14}$$

Using (4.8) and the relations $G(\omega^{j,1}) \cap G(\omega^{j,0}) = G(\omega^{j,0})$ and $\overline{\lambda}(G(\omega^{j,1})) = \overline{\lambda}(G(\omega^{j,0})) + 2a_n d_1(G(\omega^{j,0}), G(\omega^{j,1}))$ we find:

$$\int \sqrt{f_{G(\omega^{j,1})}(x) f_{G(\omega^{j,0})}(x)}dx = \frac{\overline{\lambda} + d_1[\sqrt{\frac{1}{4} - a_n^2} - (\frac{1}{2} - a_n)]}{\sqrt{\overline{\lambda}^2 + 2a_n\overline{\lambda}d_1}}$$

where $\overline{\lambda} = \overline{\lambda}(G(\omega^{j,1}))$ and $d_1 = d_1(G(\omega^{j,0}), G(\omega^{j,1}))$.

Hence

$$(5.15)\quad \int \sqrt{f_{G(\omega^{j,1})}(x) f_{G(\omega^{j,0})}(x)}dx = \left(\frac{(\overline{\lambda} + d_1[\sqrt{\frac{1}{4} - a_n^2} - (\frac{1}{2} - a_n)])^2}{\overline{\lambda}^2 + 2a_n\overline{\lambda}d_1} \right)^{1/2}$$

$$\geq \left(\frac{\overline{\lambda}^2 + 2\overline{\lambda}d_1[\sqrt{\frac{1}{4} - a_n^2} - (\frac{1}{2} - a_n)]}{\overline{\lambda}^2 + 2a_n\overline{\lambda}d_1} \right)^{1/2} = \left(1 - \frac{2d_1[\frac{1}{2} - \sqrt{(\frac{1}{4} - a_n^2)}]}{\overline{\lambda} + 2a_n d_1} \right)^{1/2}$$

$$\geq 1 - 4d_1[\frac{1}{2} - \sqrt{\frac{1}{4} - a_n^2}] \geq 1 - 8a_n^2 d_1.$$

From (5.14) and (5.15) we obtain

$$(5.16)\qquad H^2(P_{i,G(\omega^{j,1})}, P_{i,G(\omega^{j,0})}) \leq 16a_n^2 d_1(G(\omega^{j,0}), G(\omega^{j,1}))$$

$$= 16a_n^2\lambda(\mathcal{T}_j) = 16c_6 a_n^2 (Mm^\gamma)^{-1} \leq c_7 n^{-1}.$$

From (5.13) and (5.16) we get

$$\sup_{G \in \mathcal{G}_{\gamma,L}} E_G(d_1(G, \hat{G}_n)) \geq c_5 m^{-\gamma}(1 - (c_7/2)n^{-1})^{2n}.$$

Using the definition of m we get (5.1) for Model 2.

■

Theorems 4.1 and 5.1 show that $(na_n^2)^{-\gamma/(\gamma+N-1)}$ is the optimal rate of convergence for estimates of sets G on classes $\mathcal{G}_{\gamma,L}$. The next theorem shows that the rate given for convex sets in Theorem 4.2 is optimal.

THEOREM 5.2. *Let $na_n^2 \geq 1$. Assume Model 1 or Model 2. Then for any $q \geq 0$ there exists $C_0' > 0$ such that*

$$\liminf_n \inf_{\hat{G}_n} \sup_{G \in \mathcal{G}_{\text{conv}}} (na_n^2)^{\frac{2q}{N+1}} E_G(d_1^q(\hat{G}_n, G)) \geq C_0'.$$

For Model 1, the constant C_0' does not depend on the distrubution of $\mathcal{X}$.

PROOF OF THEOREM 5.2. Let $g_r(t, \omega)$ be the functions (5.2) where $\gamma = 2$, and $m = [(na_n^2)^{1/(N+1)}] + 1$. Denote

$$g_{r1}(t, \omega) = g_r(t, \omega) - |t|^2,$$

where $t = (x_1, \ldots, x_{N-1})$.

The sets

$$G_{r1}(\omega) = \{x = (x_1, \ldots, x_N) \in [0,1]^N : 0 \le x_N \le g_{r1}(x_1, \ldots, x_{N-1}, \omega)\}$$

are convex if λ_0 is small enough. In fact

$$\nabla^2 g_{r1}(t, \omega) = \nabla^2 g_r(t, \omega) - I_{N-1} = \lambda_0 \sum_{j=1}^{M} \omega_j \nabla^2 \eta(m(t - b_j)) - I_{N-1},$$

where I_{N-1} is the $(N-1) \times (N-1)$ identity matrix. The norm $\|\nabla^2 \eta\|$ is bounded. Therefore for λ_0 small enough the matrix $\nabla^2 g_{r1}(t, \omega)$ is negative definite. This shows that g_{r1} is concave, and that G_{r1} is convex.

Now let $\mathcal{X}$ be fixed. Introduce for $r = 0, 1, \ldots, m^\gamma - 1$ the curved strips

$$V_r = \{x \in [0,1]^N : \frac{r}{m^\gamma} - |t|^2 \le x_N \le \frac{r+1}{m^\gamma} - |t|^2\}.$$

It is clear that there exists at least $\frac{3}{4} m^\gamma$ strips V_r such that the number of points X_i in each strip is less than $\frac{4n}{m^\gamma}$ (otherwise the total number of points would exceed n). This entails that there exists not less than $\frac{1}{4} m^\gamma$ strips V_r with this property for $r \in [m^\gamma/4, 3m^\gamma/4]$. Denote the set of indices r corresponding to these strips by $\mathcal{D}_{1m}$. Then

$$\text{card } \mathcal{D}_{1m} \ge m^\gamma/4. \tag{5.17}$$

Note that $G_{r1}(\omega) \subset Q_r$ where

$$Q_r = \{(x_1, \ldots, x_N) \in [0,1]^N : |t|^2 < \frac{r+1}{m^\gamma}\}.$$

Hence $G_{r1}(\omega)$ is given as

$$G_{r1}(\omega) = \{x = (x_1, \ldots, x_N) \in [0,1]^N : 0 \le x_N \le \tilde{g}_{r1}(x_1, \ldots, x_{N-1}, \omega)\},$$

where

$$\tilde{g}_{r1}(t, \omega) = \lambda_0 \sum_{j \in I_r} \omega_j \eta(m(t - b_j)) - |t|^2$$

and $I_r = \{j : b_j \in Q_r\}$. Write $M_r = \text{card } I_r$.

Clearly, for $r \in [m^\gamma/4, 3m^\gamma/4]$ we have $M_r \geq \beta M$ where $0 < \beta < 1$ is a fixed number. Using the same argument as above we get that in each strip V_r with $r \in \mathcal{D}_{1m}$ there exists at least $\beta M/2$ pieces of the form

$$\{x \in V_r : |x_k - b_j| \leq 1/(2m),\ k = 1, \ldots, N-1\}, \quad j = 1, \ldots, M_r,$$

such that the number of points X_i in each piece is less than $8nm^{-\gamma}(\beta M)^{-1}$. But these pieces contain the sets

$$\mathcal{T}_{jr1} = \{x \in [0,1]^N : \tilde{g}_{r1}(x_1, \ldots, x_{N-1}, \omega_0^{j,0}) \leq x_N \leq \tilde{g}_{r1}(x_1, \ldots, x_{N-1}, \omega_0^{j,1})\}$$

Thus we have proved that for each $r \in [m^\gamma/4, 3m^\gamma/4]$ there exists a subset $\Psi_{r1} \subseteq \{1, \ldots, M_r\}$ such that

$$\begin{aligned} &\text{card } \Psi_{r1} \geq \beta M/2, \\ &\text{card } U_{jr1} \leq (8/\beta) n (m^\gamma M)^{-1} \text{ for } j \in \Psi_{r1} \text{ and } r \in \mathcal{D}_{1m}, \end{aligned} \tag{5.18}$$

where $U_{jr1} = \{i : X_i \in \mathcal{T}_{jr1}\}$.

Note that (5.17) and (5.18) are of the same form as (5.11), and that $\lambda(\mathcal{T}_{jr1}) = \lambda(\mathcal{T}_{jr})$. Therefore one can proceed as in the proof of Theorem 5.1 for Model 1.

For Model 2 the proof is similar to that in Theorem 5.1. Instead of $g(t, \omega)$ one chooses the functions $g_1(t, \omega) = g(t, \omega) - |t|^2$ with $\gamma = 2$, and with λ_0 small enough. Then one uses the sets $G(\omega) = \{x \in [0,1]^N : 0 \leq x_N \leq g_1(t, \omega)\}$.

■

Appendix. Let Θ be a compact metric space with metric d, and let ε be a positive number. Consider the family of distributions P_θ defined on a measurable space $(\mathcal{Z}, \mathcal{F})$ and indexed by parameter $\theta \in \Theta$. Assume that there exists a measure ν on $(\mathcal{Z}, \mathcal{F})$ which dominates the measures P_θ. We define for $Z \in \mathcal{Z}$

$$p(Z, \theta) = \frac{dP_\theta}{d\nu}(Z).$$

Choose a fixed ε-net $\mathcal{N}_\varepsilon = \mathcal{N}(\Theta, d, \varepsilon)$ on Θ with respect to the metric d.

For an observation Z in $\mathcal{Z}$ with density $p(\cdot, \theta)$ we define the maximum likelihood estimator on $\mathcal{N}_\varepsilon$

$$\theta^* = \arg \max_{\theta \in \mathcal{N}_\varepsilon} p(Z, \theta).$$

Assume that either

(I) For $\theta, \theta' \in \Theta$ the density $p(Z, \theta)$ satisfies

$$\int \sqrt{p(Z,\theta)p(Z,\theta')}d\nu \leq c_{d1}\exp(-c_{d2}d^{\alpha}(\theta,\theta')), \tag{A.1}$$

where $c_{d1}, c_{d2}, \alpha > 0$ are constants which do not depend on θ, θ', and for every $\theta \in \Theta$ there exists $\tilde{\theta} \in \mathcal{N}_{\varepsilon}$ such that $d(\theta, \tilde{\theta}) \leq \varepsilon$,

$$P_{\theta}\{c_{p1}p(Z,\theta) \geq p(Z,\tilde{\theta})\} \leq c_{p2}, \tag{A.2}$$

where $c_{p1} > 0$ and $c_{p2} > 0$ do not depend on $\theta, \tilde{\theta}$,
or

(II) There exist $c_{d1}, c_{d2}, \alpha > 0, t_0 > 1$, such that

$$\sup_{\substack{\theta',\tilde{\theta}: d(\theta,\tilde{\theta})\leq\varepsilon \\ d(\theta,\theta')\geq\varepsilon t}} P_{\theta}\{p(Z,\theta') \geq p(Z,\tilde{\theta})\} \leq c_{d1}\exp(-c_{d2}(\varepsilon t)^{\alpha}), \tag{A.3}$$

for $t \geq t_0$, and for every $\varepsilon > 0$.

The following lemma is a modification of a result in Birgé (1983).

Lemma A.1. *Assume (I) or (II), and let there exist an $\varepsilon > 0$ such that*

$$-\frac{1}{2}\log c_{p1} + \log(\text{card}\,\mathcal{N}_{\varepsilon}) \leq \frac{1}{2}c_{d2}\varepsilon^{\alpha}t^{\alpha}, \quad t \geq t_1, \tag{A.4}$$

for some $t_1 > 0$. Here we set, by definition, $c_{p_1} = 1, c_{p2} = 0$ if (II) holds. Then

$$\begin{aligned} \sup_{\theta\in\Theta} E_{\theta}(\varepsilon^{-q}d^{q}(\theta^*,\theta)) &\leq \max(t_0,t_1) + \\ &+ (\mathcal{D}/\varepsilon)^{q}\left[c_{d1}\exp(-\frac{1}{2}c_{d2}\varepsilon^{\alpha}) + c_{p2}\right], \end{aligned} \tag{A.5}$$

where E_{θ} is the expectation with respect to P_{θ}, θ^* is the MLE on $\mathcal{N}_{\varepsilon}$ and $\mathcal{D}$ is the diameter of Θ:

$$\mathcal{D} = \sup_{\theta,\theta'\in\Theta} d(\theta,\theta').$$

PROOF. If θ^* is the MLE on $\mathcal{N}_{\varepsilon}$, then for any $t > 1$

$$P_{\theta}(d(\theta^*,\theta) \geq \varepsilon t) \leq P_{\theta}(\max_{\substack{\theta'\in\mathcal{N}_{\varepsilon}: \\ d(\theta,\theta')\geq\varepsilon t}} p(Z,\theta') \geq p(Z,\tilde{\theta})) \tag{A.6}$$

where $\tilde{\theta}$ is an element of ε-net $\mathcal{N}_\varepsilon$ such that $d(\theta,\tilde{\theta}) \le \varepsilon$. Consider first the case when Assumption (I) holds. Choose $\tilde{\theta}$ such that (A.2) holds. Then

$$
\begin{aligned}
(A.7)\quad & P_\theta\Big(\max_{\substack{\theta'\in\mathcal{N}_\varepsilon:\\ d(\theta,\theta')\ge\varepsilon t}} p(Z,\theta') \ge p(Z,\tilde{\theta})\Big) \le \\
& \le P_\theta\Big(\max_{\substack{\theta'\in\mathcal{N}_\varepsilon:\\ d(\theta,\theta')\ge\varepsilon t}} p(Z,\theta') \ge c_{p1}p(Z,\theta)\Big) + c_{p2} \ge \\
& \le (\operatorname{card}\mathcal{N}_\varepsilon) \max_{\substack{\theta'\in\mathcal{N}_\varepsilon:\\ d(\theta,\theta')\ge\varepsilon t}} P_\theta\left(\sqrt{\frac{p(Z,\theta')}{p(Z,\theta)}} \ge \sqrt{c_{p1}},\ p(Z,\theta)\ne 0\right) + c_{p2} \le \\
& \le c_{p1}^{-1/2}\,(\operatorname{card}\mathcal{N}_\varepsilon) \max_{\substack{\theta'\in\mathcal{N}_\varepsilon:\\ d(\theta,\theta')\ge\varepsilon t}} \int \sqrt{p(Z,\theta)p(Z,\theta')}d\nu + c_{p2} \le \\
& \le c_{p1}^{-1/2} c_{d1}\,(\operatorname{card}\mathcal{N}_\varepsilon)\ \exp(-c_{d2}\varepsilon^\alpha t^\alpha) + c_{p2}.
\end{aligned}
$$

It follows from (A.4), (A.6) and (A.7) that for $t \ge t_1$

$$
(A.8)\qquad P_\theta(d(\theta^*,\theta) \ge \varepsilon t) \le c_{d1}\exp(-c_{d2}\varepsilon^\alpha t^\alpha/2) + c_{p2}.
$$

Assume now that (II) holds. Then (A.3), (A.4) and (A.6) entail

$$
\begin{aligned}
(A.9)\qquad & P_\theta(d(\theta^*,\theta) \ge \varepsilon t) \le \\
& \le (\operatorname{card}\mathcal{N}_\varepsilon) \sup_{\substack{\theta',\tilde{\theta}: d(\theta,\tilde{\theta})\le\varepsilon,\\ d(\theta,\theta')\ge\varepsilon t}} P_\theta\{p(Z,\theta') \ge p(Z,\tilde{\theta})\} \le \\
& \le c_{d1}\exp(-c_{d2}\varepsilon^\alpha t^\alpha/2),\ t \ge \max(t_0,t_1).
\end{aligned}
$$

Using (A.8) and (A.9) we find

$$
\begin{aligned}
& E_\theta(\varepsilon^{-q}d^q(\theta^*,\theta)) = \\
& = \int_0^{\mathcal{D}'} P_\theta(d(\theta^*,\theta) \ge \varepsilon t^{1/q})dt \le \\
& \le \max(t_0,t_1) + c_{d1}\int_{\max(t_0,t_1)}^{\mathcal{D}'} \exp(-c_{d2}\varepsilon^\alpha t^{\alpha/q}/2)dt + \mathcal{D}'c_{p2} \le \\
& \le \max(t_0,t_1) + c_{d1}\mathcal{D}'\exp(-\frac{1}{2}c_{d2}\varepsilon^\alpha) + \mathcal{D}'c_{p2},
\end{aligned}
$$

where $\mathcal{D}' = (\mathcal{D}/\varepsilon)^q$.

■

Acknowledgement. We would like to thank L. Devroye and P. Jacob for the references related to support estimation.

References:

Abbar, H. (1990), Un estimateur spline du contour d'une répartition ponctuelle aléatoire, *Statistique et Analyse des Données*, 15, n.3, 1–19.

Birgé, L. (1983), Approximation dans les espaces métriques et théorie de l'estimation, *Z. Wahrscheinlichkeitstheorie und verw. Geb*, 65, 181–238.

Bosq, D. (1971), *Contribution à la théorie de l'estimation fonctionelle*, Thèse, Paris.

Bretagnolle, J. and Huber, C. (1979), Estimation des densités: risque minimax, *Z. Wahrscheinlichkeitstheorie und verw. Geb*, 47, 119–137.

Chevalier, J. (1976), Estimation du support et du contenu du suport d'une loi de probabilité, *Ann. Inst. H. Poincaré, sec. B*, 12, n. 4, 339–364.

Devroye, L. and Györfi, L. (1985), *Nonparametric Density Estimation: the L_1-view*, Willey, N.Y.

Devroye, L. and Wise, G.L. (1980), Detection of abnormal behavior via nonparametric estimation of the support, *SIAM J. Applied Math.*, 38, 480–488.

Dudley, R.M. (1974), Metric entropy of some classes of sets with differentiable boundaries, *J. Approx. Theory*, 10, 227–236.

Efron, B. (1965), The convex hull of a random set of points, *Biometrika*, 52, 331–343.

Geffroy, J. (1964), Sur un problème d'estimation géométrique, *Publications de l'Institut de Statistique des Universités de Paris*, 13, 191–120.

Hartigan, J.A. (1975), *Clustering Algorithms*, Wiley, New York.

Hartigan, J.A. (1987), Estimation of a convex density contour in two dimensions, *J. Amer. Stat. Assoc.*, 82, 267–270.

Ibragimov, I.A. and Khasminskii, R.Z. (1980), On nonparametric estimation of regression, *Doklady Acad. Nauk SSSR*, 252, 780–784.

Ibragimov, I.A. and Khasminskii, R.Z. (1981), *Statistical Estimation: Asymptotic Theory*, Springer-Verlag, New York.

Jacob, P. (1984), Estimation du contour discontinu d'un processus ponctuel sur le plan, *Publications de l'Institut de Statistique des Universités de Paris*, 29, 3–4, 1–25.

Jacob, P. and Abbar, H. (1989), Estimating the edge of Cox process area, *Cahiers du centre d'Etudes de Recherche Opérationnelle*, 31, 3–4.

Kolmogorov, A.N. and Tikhomirov, V.M. (1961), ε-entropy and ε-capacity of sets in functional spaces, *Amer. Math. Soc. Transl. (Ser. 2)*, 17, 277–364.

Korostelev, A.P. (1991), Minimax reconstruction of two-dimensional images, *Theory of Probability and its Appl.*, 36, 153–159.

Korostelev, A.P. and Tsybakov, A.B. (1992), Asymptotically minimax image reconstruction problems, in *"Topics in Nonparametric Estimation"*, Khasminskii R.Z., ed., AMS, Providence, RI, 45–86.

Korostelev, A.P. and Tsybakov, A.B. (1993), *Minimax Theory of Image Reconstruction*, Springer, NY. e.a.

Mammen, E. (1991a), Nonparametric regression under qualitative smoothness assumptions, *Ann. Statist.*, 19, 741–759.

Mammen, E. (1991b), Nonparametric curve estimation and simple curve characteristics, in *Nonparametric Functional Estimation and Related Topics. Proceedings of the NATO Advanced Study Institute, Spetses (Greece)*, (G. Roussas ed.), 133–140, Kluwer, Dordrecht.

Moore, M. (1984), On the estimation of a convex set, *Ann. Statist.*, 12, 1090-1099.

Müller, D.W. and Sawitzki, G. (1990), Using excess mass estimates to investigate the modality of a distribution, *Proceedings of the ICOSCO I Conference (First International Conference on Statistical Computing, Çesme, Izmir, Turkey 1987)*, ed. by E.C. Van der Meulen.

Müller, D.W. and Sawitzki, G. (1991), Excess mass estimates and tests for multimodality, *J. Amer. Statist. Assoc.*, 86, 738–746.

Nagao, M. and Matsuyama, T. (1979), Edge preserving smoothing, *Computer Graphics and Image Processing*, 9, 394–407.

Nemirovskii, A.S., Polyak, B.T. and Tsybakov, A.B. (1985), Rate of convergence of nonparametric estimates of maximum-likelihood type, *Problems of Information Transmission*, 21, 258–272.

Polonik, W. (1991), Excess mass estimates over classes of sets, estimation of the density contour clusters and tests for multi-modality in more than one dimensions, *Preprint* 611, SFB 123, Universität Heidelberg.

Pratt, W.K. (1978), *Digital Image Processing*, Wiley, New York.

Rényi, A. and Sulanke, R. (1963), Über die konvexe Hülle von n zufällig gewählten Punkten, *Z. Wahrscheinlichkeitstheorie und verw. Geb*, 2, 75–84.

Rényi, A. and Sulanke, R. (1964), Über die konvexe Hülle von n zufällig gewählten Punkten II, *Z. Wahrscheinlichkeitstheorie und verw. Geb*, 3, 138–147.

Ripley, B.D. and Rasson, J.P. (1977), Finding the edge of a Poisson forest, *J. Appl. Prob.*, 14, 483–491.

Rudemo, M., Skovgaard, I. and Stryhn, H. (1990), Maximum likelihood estimation of curves in images, *Report 90-4*, Dept. of Math. and Phys., Royal Veterinary and Agricultural Univ., Frederiksberg, Denmark.

Rudemo, M., and Stryhn, H. (1991), Approximating the distributions of maximum likelihood contour estimators in two - region images, *Report 91-2*, Dept. of Math. and Phys., Royal Veterinary and Agricultural Univ., Frederiksberg, Denmark.

Stone, C.J. (1980), Optimal rates of convergence for nonparametric estimators, *Ann. Statist*, 8, 1348–1360.

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

Tsybakov, A.B. (1989), Optimal estimation accuracy of nonsmooth images, *Problems of Information Transmission*, 25, 180–191.

Tsybakov, A.B. (1991), Nonparametric techniques in image estimation, in *Nonparametric Functional Estimation and Related Topics. Proceedings of the NATO Advanced Study Institute, Spetses (Greece)*(G. Roussas ed.) 669–678, Kluwer, Dordrecht.