The Annals of Statistics
1995, Vol. 23, No. 2, 502–524

ASYMPTOTICAL MINIMAX RECOVERY OF SETS WITH SMOOTH BOUNDARIES

BY E. MAMMEN AND A. B. TSYBAKOV

Humboldt-Universität zu Berlin and Université Pierre et Marie Curie

In this paper optimal rates of convergence are derived for estimates of sets in N-dimensional "black and white" pictures under smoothness conditions. It is assumed that the boundaries of the "black" regions have a smooth parameterisation, that is, 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. Two models are studied: edge estimation models motivated by image segmentation problems and density support estimation.

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. Recently some work has been done on the nonparametric estimation of sets in images, in particular, optimal rates of convergence were studied by Korostelev (1991), Tsybakov (1989, 1991) and Korostelev and Tsybakov (1992, 1993a, b).

In this paper we derive optimal convergence rates for estimates of sets in N-dimensional "black-and-white" pictures. We consider estimation of "black" regions under smoothness conditions. It is assumed that the boundaries of these regions have a smooth parametrisation, so that the regions belong to the Dudley classes of sets [Dudley (1974)]. (As a special case, this includes the model of "boundary fragments" studied in the above-mentioned papers of Korostelev and Tsybakov.) The variety of possible sets covered by our assumptions is rather large since smooth parameterisation of the boundary does not imply that the boundary itself is smooth (see Figure 1). Also for convex sets our approach gives the correct rate of convergence.

We study the problem of estimation of sets for two different setups: edge estimation and density support estimation. The study of edge estimation is motivated by applications in image analysis. Often as a first step in image processing the original picture is transformed to a binary black-and-white picture which is then used to recover edges, that is, boundaries of black regions [see, e.g., Pratt (1978)]. Density support estimation is related to cluster analysis and quality control. The aim of this paper is not to add some new practical recipe to the variety of existing ones, but to propose, under

Received April 1992; revised March 1994.
AMS 1991 *subject classifications*. 62G05, 62G20.
Key words and phases. Boundary estimation, binary pictures, density support estimation, optimal rates of convergence, ε-entropy.

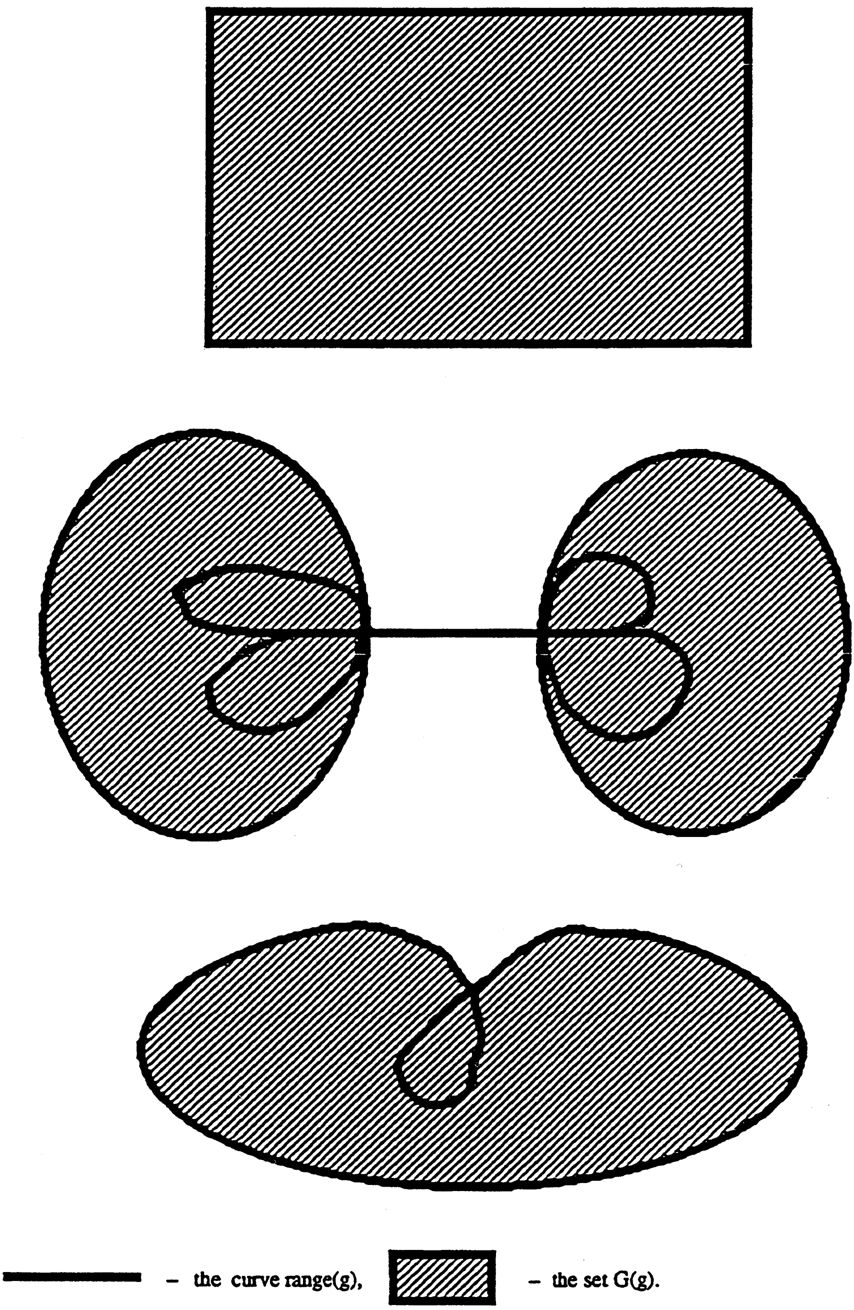

FIG. 1. *Three regions with "unsmooth" boundary and "smooth" parametrisation g.*

certain idealisations, a tool for comparing the statistical properties of different estimation methods. Such a tool, in our opinion, is provided by the knowledge of the optimal convergence rates.

2. Smoothness classes of regions. Let us recall 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 $\{(F_j, V_j),\ j = 1,\dots, J\}$, where $V_1,\dots, V_J$ are open sets on the sphere

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

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

$$F_j\colon \{x | x_1^2 + \cdots + x_{N-1}^2 < 1\} \to V_j.$$

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

$$\mathscr{F}_{\gamma, L} = \Bigg\{ g = (g_1,\dots, g_N) | g_i\colon S^{N-1} \to [0,1] \text{ with } \sup_{x,y} \frac{|\varphi^{(k)}(x) - \varphi^{(k)}(y)|}{|x-y|^{\gamma-k}} \leq L \text{ for } \varphi(x) = g_i(F_j(x)),\ j = 1,\dots, J,\ i = 1,\dots, N \Bigg\}.$$

Here k is $\max\{m \in \mathbb{N}\colon m < \gamma\}$, $|\cdot|$ 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 $g = (g_1,\dots, g_N) \in \mathscr{F}_{\gamma, L}$ defines a surface in $\mathbb{R}^N$ with "smoothness parameter" γ. As in Dudley (1974), for each $g \in \mathscr{F}_{\gamma, L}$ we introduce the set $G(g)$ which has this surface as boundary in a certain sense. The set $G(g)$ consists of all points $x \in \mathbb{R}^N$ such that either:

1. $x \in \text{range}(g)$

or

2. $x \notin \text{range}(g)$ 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. The last two examples show that the boundary $\partial G(g) \subset$ range(g) is not necessarily smooth if the curve range(g) is smooth. Finally, the smoothness class of sets is defined as

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

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

MODEL 1 (Edge estimation). One observes $X_i \in [0,1]^N$ and $Y_i \in \{-1, 1\}$ ($i = 1, \dots, n$). (Interpretation: X_i is a point and Y_i is the image 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 $1/2 + a_n$:

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

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

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

where f_G is a function taking the values ± 1,

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

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

In view of (3.2), a_n may be interpreted as the edge step size of a binary picture. 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, 1993b) for smoothness classes of boundary fragments (sets G with boundaries defined by smooth functions of the first $N - 1$ coordinates). The case (ii) models pictures with high noise. Note that for Model 1 the estimates of G considered below do not depend on a_n.

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

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

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 $\lambda(G) > 0$.

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 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 $X_1, \dots, X_n$. The system is decided to behave abnormally if $X \notin \hat{G}_n$.

Another application of support estimation is cluster analysis. One can define clusters as level sets of densities, that is, the sets, where the density $f(x)$ of X_i's is larger than some constant $\kappa > 0$ [see Hartigan (1975) and Müller and Sawitzki (1991a, b)]. In the case of a uniform density f, all level sets $V_\kappa = \{f(x) \geq \kappa\}$ coincide with the support G for κ small enough. 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).

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 (1991a, b) and Polonik (1991)]. For $\kappa > 0$ the excess mass is defined as $\int \max\{f(x) - \kappa, 0\}\, dx$. The silhouette is the relation $\kappa \to \{x\colon f(x) > \kappa\}$. 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 related 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). Other types of domains were considered by Jacob (1984), Jacob and Abbar (1989) and Abbar (1990).

4. Upper risk bounds. For two closed sets G_1 and G_2 define the distance $d_1(G_1, G_2)$ as the Lebesgue measure of their symmetric difference:

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

Let E_G denote the expectation with respect to the distribution of the sample when the underlying set is G.

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

THEOREM 4.1. *Let $\mathscr{G}_n \subseteq \mathscr{G}_{\gamma, L}$. 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* $G \in \mathscr{G}_n$, *or* $\lim_{n\to\infty} a_n = 0$ *and all sets* $G \in \mathscr{G}_n$ *are contained in a set* $\mathscr{D} \subset [0,1]^N$ *with*

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

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

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

Here $C_1 > 0$ *is a constant which depends only on* q, N, γ *and* L.

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, 1993b), where for Model 1 some special subclasses of $\mathscr{G}_{\gamma,L}$ (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 $\mathscr{G}$ be a compact set of domains G endowed with a metric d. Denote by $\mathscr{N}(\mathscr{G}, d, \varepsilon) \subseteq \mathscr{G}$ a minimal ε-net on $\mathscr{G}$ in d-metric. The estimator G_n^* of Theorem 4.1 is the maximum likelihood estimator (MLE) on the ε-net $\mathscr{N}_\varepsilon = \mathscr{N}(\mathscr{G}_n, d_1, \varepsilon)$. 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 $\mathscr{N}_\varepsilon$ are of the form

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

(for Model 1)

and

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

where

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

We do not know if for Model 2 there exists also an estimate which does not require knowledge of a_n and achieves the optimal rates of Theorem 4.1.

In particular, for the case $a_n = 1/2$ the minimisation in (4.4) reduces to

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

REMARK 1. In 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. 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 models of type (3.2), where ζ_{in} are independent zero mean random variables with

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

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

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

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 when based on a global search on $\mathscr{N}_\varepsilon$. 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, 1993b) algorithms are discussed for boundary fragments and for the classes of "star-shaped" two dimensional sets G [see also Rudemo, Skovgaard and Stryhn (1990) and Rudemo and Stryhn (1994)].

PROOF OF THEOREM 4.1. Write for brevity $\mathscr{G} = \mathscr{G}_n$. We apply Lemma A.1 of the Appendix. Consider first Model 1. Let us check (A.3). In our case $\Theta = \mathscr{G}$, $\theta = G$ 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 the 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 sets $\theta = G$, $\tilde{\theta} = \tilde{G}$ and $\theta' = G'$ with $d_1(G, \tilde{G}) \le \varepsilon$, $d_1(G, G') \ge \varepsilon t$, for some $t \ge t_0$. Then

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

where

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

Note that

$$\begin{aligned} E_G\{S\} &= 2na_n\left[-d_1(G, G') + d_1(G, \tilde{G})\right] \\ &\le 2na_n\left[-d_1(G, G') + \varepsilon\right]. \end{aligned} \tag{4.5}$$

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

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

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

$$P_G\{S \ge 0\} \le \exp\left(-\tfrac{1}{6}na_n^2 \varepsilon t\right),$$

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

Now, let us check condition (A.4). Dudley (1974) proved that there exists a constant $C_{\mathscr{H}} > 0$ depending on γ, L such that the ε-entropy [Kolmogrov and Tikhomirov (1961)] satisfies

$$\mathscr{H}(\mathscr{G}_{\gamma, L}, d_1, \varepsilon) = \log\left(\operatorname{card} \mathscr{N}(\mathscr{G}_{\gamma, L}, d_1, \varepsilon)\right) \le C_{\mathscr{H}} \varepsilon^{-(N-1)/\gamma}.$$

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

$$\mathscr{H}(\mathscr{G}_n, d_1, \varepsilon) \le C_{\mathscr{H}} \varepsilon^{-(N-1)/\gamma}. \tag{4.6}$$

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

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

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

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

If we substitute here the expression for ε, we get (4.2) for Model 1.

Consider now Model 2. We apply Lemma A.1 with $Z = (X_1, \dots, X_n)$, 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 1 of the Appendix. For showing (A.1) we have to consider

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

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

$$J = \frac{\bar{\lambda}(G \cap G') + d_1(G,G')\left[\sqrt{(1/4) - a_n^2} - ((1/2) - a_n)\right]}{\sqrt{\bar{\lambda}(G)\bar{\lambda}(G')}}. \tag{4.8}$$

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

$$\begin{aligned} J &= \frac{\lambda(G \cap G')}{\sqrt{\lambda(G)\lambda(G')}} \leq \sqrt{\frac{\lambda^2(G \cap G')}{\lambda^2(G \cap G') + \lambda(G \cap G') d_1(G,G')}} \\ &\leq (1 - d_1(G,G'))^{1/2} \leq \exp\left(\frac{-d_1(G,G')}{2}\right). \end{aligned}$$

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

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

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

$$\begin{aligned} J &\leq \left(\frac{\bar{\lambda}^2 + 2(a_n - a_n^2) d_1 \bar{\lambda} + d_1^2 (a_n - a_n^2)^2}{\bar{\lambda}(G)\bar{\lambda}(G')}\right)^{1/2} \\ &\leq \left(1 - \frac{2a_n^2 d_1}{\bar{\lambda} + 2a_n d_1} + \frac{d_1^2 a_n^2}{\left(\frac{1}{2} - a_n\right)^2}\right)^{1/2}. \end{aligned} \tag{4.11}$$

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

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

$$J \leq \left(1 - \frac{2a_n^2 d_1}{1/2 + a_n} + \frac{(1 - \delta_0) d_1 a_n^2}{(1/2 - a_n)^2}\right)^{1/2} = \left(1 - a_n^2 d_1(G, G') \kappa_n\right)^{1/2},$$

where $\kappa_n = [2/(1/2 + a_n) - (1 - \delta_0)/(1/2 - a_n)^2] \geq 4[\delta_0 + O(a_n)] \geq 2\delta_0$ for n large enough. Hence, for n large enough, we have

$$J \leq \left(1 - 2\delta_0 a_n^2 d_1(G, G')\right)^{1/2} \leq \exp\left(-\delta_0 a_n^2 d_1(G, G')\right). \tag{4.12}$$

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

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

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

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

$$P_G\left\{c_{p1} p(Z, G) \geq p(Z, \tilde{G})\right\} \leq c_{p2}. \tag{4.13}$$

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 $G \in \mathscr{G}_{\gamma, L}$ and $\tilde{G} \subseteq [0, 1]^N$ we have $d_\infty(G, \tilde{G}) \leq \varepsilon$ entails that $d_1(G, \tilde{G}) \leq \overline{C}\varepsilon$, 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 (since we assume everywhere that $\gamma \geq 1$) and thus $\mathrm{mes}(\partial G)$ is uniformly bounded for $G \in \mathscr{G}_{\gamma, L}$.

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

$$\mathscr{O}_\varepsilon(G) = \left\{x \in [0, 1]^N : \rho(x, G) \leq \varepsilon\right\}.$$

Clearly, the set

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

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

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

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 inequallity 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 $c_{p1} = (1 - 2\varepsilon/\lambda_0)^n$ 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

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

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

$$\begin{aligned} \log\big(p(Z,\theta)/p(Z,\tilde{\theta}) \big) &= n \log\big[\bar{\lambda}(\tilde{G}) / \bar{\lambda}(G) \big] \\ &\quad - R \log\big[(1/2 + a_n)/(1/2 - a_n) \big], \end{aligned} \tag{4.15}$$

where $R = \sum_{i=1}^n I(X_i \in \tilde{G} \setminus G)$ is a binomial random variable with parameter $(1/2 - a_n)\lambda(\tilde{G} \setminus G)/\bar{\lambda}(G)$. 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 $9 n a_n^2 \varepsilon$. 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. Note that for the ε-net with one-sided bracketing $\mathscr{N}_\varepsilon$ constructed above, we have

$$\log(\operatorname{card} \mathscr{N}_\varepsilon) = \log\big(\operatorname{card} \mathscr{N}^0_{\varepsilon/2\bar{C}}\big) \leq \mathscr{H}\left(\mathscr{G}_{\gamma, L}, d_\infty, \frac{\varepsilon}{2\bar{C}} \right) \leq C'_{\mathscr{H}} \varepsilon^{-(N-1)/\gamma},$$

where $C'_{\mathscr{H}} > 0$ is a constant [the last inequality follows from Dudley (1974)]. Using this and arguing as in the case of Model 1, we get that it is sufficient to find $\varepsilon > 0$ such that $-\frac{1}{2} \log c_{p1} + C'_{\mathscr{H}} \varepsilon^{-(N-1)/\gamma} \leq \frac{1}{2} c_{d2} \varepsilon t$. If $a_n \equiv 1/2$, this is equivalent to

$$-\frac{n}{2} \log\left(1 - \frac{2\varepsilon}{\lambda_0} \right) + C'_{\mathscr{H}} \varepsilon^{-(N-1)/\gamma} \leq \frac{n}{4} \varepsilon t \tag{4.16}$$

since $c_{d2} = n/2$ in view of (4.9). Now $-\log(1 - 2\varepsilon/\lambda_0) \leq 4\varepsilon/\lambda_0$ for ε small enough.

The inequality (4.16) is satisfied if

$$\frac{4 n \varepsilon}{\lambda_0} + C'_{\mathscr{H}} \varepsilon^{-(N-1)/\gamma} \leq \frac{n}{4} \varepsilon t . \tag{4.17}$$

If we choose $\varepsilon = \text{const}\, n^{-\gamma/(\gamma+N-1)}$, then there exists such $t_1 > 1$ that (4.17) holds for $t \geq t_1$. This proves (A.4).

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

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

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

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

PROOF. We proceed as in the proof of Theorem 4.1, but we use another entropy bound. Instead of (4.6) we apply [see Dudley (1974)]

$$\mathscr{H}(\mathscr{G}_{\text{conv}}, d_1, \varepsilon) = O(\varepsilon^{-(N-1)/2}). \qquad \square$$

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 $G \in \mathscr{G}_{\text{conv}}$, $N = 2$ and $a_n \equiv 1/2$. The maximum likelihood estimator is then the convex hull of $X_1, \dots, X_n$:

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

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

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

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

$$E_G\big(d_1\big(G_n^{\text{MLE}}, G\big)\big) = O(n^{-2/3}), \qquad n \to \infty.$$

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 $\mathscr{G}_{\text{conv}}$ for $N = 2$.

The generalisation of this result for $N > 2$ can be also obtained, as noticed by Schneider (1988). For convenience, we formulate it as a proposition.

PROPOSITION 4.3. *Assume Model* 2 *with $a_n \equiv 1/2$. Then*

$$\sup_{G \in \mathscr{G}_{\text{conv}}} E_G\big(d_1\big(G_n^{\text{MLE}}, G\big)\big) \le C(N) n^{-2/(N+1)} \tag{4.20}$$

for all n, where $C(N)$ is a constant which depends on N only.

To prove this proposition observe that by Groemer's (1974) inequality we have

$$E_G\big(d_1\big(G_n^{\mathrm{MLE}}, G\big)\big) \le E_B\big(d_1\big(G_n^{\mathrm{MLE}}, B\big)\big) \quad \text{for all } G \in \mathscr{G}_{\mathrm{conv}},$$

where B is the Euclidean ball such that $\lambda(B) = \lambda(G)$. Hence,

$$\sup_{G \in \mathscr{G}_{\mathrm{conv}}} E_G\big(d_1\big(G_n^{\mathrm{MLE}}, G\big)\big) \le E_{B_0}\big(d_1\big(G_n^{\mathrm{MLE}}, B_0\big)\big), \tag{4.21}$$

where B_0 is the ball such that $\lambda(B_0) = 1$. Now the right-hand side of (4.21) does not exceed the right-hand side of (4.20), where $C(N)$ is some constant [see Bárány (1992) and the references therein].

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 (1993b), Chapter 8].

REMARK 5. Note that we have the same optimal rates for $\mathscr{G}_{2,L}$ and $\mathscr{G}_{\mathrm{conv}}$. This is not surprising: one comes to a similar conclusion in nonparametric density estimation and regression; see Mammen (1991a, b) and 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 on 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 $\mathscr{X} = (X_1, \dots, X_n)$ in Model 1 we assume in this section only that $\mathscr{X}$ is an arbitrary collection of random points in $[0,1]^N$ and that $\mathscr{X}$ is independent of $\xi_1, \dots, \xi_n$.

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

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

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 $\mathscr{X}$.

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

The basic idea of the proof is to use the fact that the minimax risk is bounded below by the Bayes risk of any prior that concentrates in the class

$\mathscr{G}_{\gamma, L}$. We construct such a prior which places its mass equally on the vertices of a hypercube.

Consider first Model 1. Let $0 \le \eta(t) \le 1$ be an infinitely many times continuously 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 λ_0, $\omega = (\omega_1, \dots, \omega_M)$ and for an integer r with $0 < \lambda_0 < 1$, $\omega_j \in \{0, 1\}$, $j = 1, \dots, M$ and $0 \le r \le m^\gamma - 1$, we introduce the family of sets

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

where

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

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

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

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

For $k \in \{0,1\}$, $j = 1, \dots, M$, define $\omega_0^{j,k}$ as the vector ω with $\omega_i = 0$ for $i \ne j$ and $\omega_j = k$, that is,

$$\omega_0^{j,k} = \big(0, \dots, 0, \underbrace{k}_{j}, 0, \dots, 0\big).$$

We write

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

and

$$\mathscr{T}_{jr} = \left\{x \in [0,1]^N : g_r\big(x_1, \dots, x_{N-1}, \omega_0^{j,0}\big) \le x_N \le g_r\big(x_1, \dots, x_{N-1}, \omega_0^{j,1}\big)\right\}.$$

Assume that the ω_j's are i.i.d. Bernoulli random variables. Denote by E_ω the expectation with respect to this distribution and by $E_{\mathscr{X}}$ the expectation with respect to $\mathscr{X} = (X_1, \dots, X_n)$. We now lower bound the minimax risk by the Bayes risk for the chosen prior:

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

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

$$\begin{aligned}(5.4)\qquad &E_\omega E_{G^r(\omega)}\Big(d_1\big(G^r(\omega)\cap\mathscr{T}_{jr},\hat{G}_n\cap\mathscr{T}_{jr}\big)|\mathscr{X}\Big)\\ &\quad= E_{\omega^j}\Big[E_{\omega_j}E_{G^r(\omega)}\Big(d_1\big(G^r(\omega)\cap\mathscr{T}_{jr},\hat{G}_n\cap\mathscr{T}_{jr}\big)|\mathscr{X},\omega^j\Big)\Big].\end{aligned}$$

Here

$$\begin{aligned}&E_{\omega_j}\Big[E_{G^r(\omega)}\Big(d_1\big(G^r(\omega)\cap\mathscr{T}_{jr},\hat{G}_n\cap\mathscr{T}_{jr}\big)|\mathscr{X},\omega^j\Big)\Big]\\ &\quad=\frac{1}{2}\Big[E_{G^r(\omega^{j,0})}\Big(d_1\big(G^r(\omega^{j,0})\cap\mathscr{T}_{jr},\hat{G}_n\cap\mathscr{T}_{jr}\big)|\mathscr{X},\omega^j\Big)\\ (5.5)\qquad &\qquad\quad+E_{G^r(\omega^{j,1})}\Big(d_1\big(G^r(\omega^{j,1})\cap\mathscr{T}_{jr},\hat{G}_n\cap\mathscr{T}_{jr}\big)|\mathscr{X},\omega^j\Big)\Big]\\ &\quad\ge\frac{\bar{d}_1}{2}\int\min\Big(\prod_{i\in U_{jr}}dP_{i,G^r(\omega^{j,1})},\prod_{i\in U_{jr}}dP_{i,G^r(\omega^{j,0})}\Big)\prod_{i\notin U_{jr}}dP_{i,G^r(\omega^{j,0})},\end{aligned}$$

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

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

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), page 226] one gets

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

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

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

It follows from (5.4)–(5.7) that

$$\begin{aligned}&E_\omega\Big[E_{G^r(\omega)}\Big(d_1\big(G^r(\omega)\cap\mathscr{T}_{jr},\hat{G}_n\cap\mathscr{T}_{jr}\big)|\mathscr{X}\Big)\Big]\\ (5.8)\qquad &\quad\ge\frac{\lambda(\mathscr{T}_{jr})}{2}E_\omega\left[\prod_{i\in U_{jr}}\left(1-\frac{H^2\big(P_{i,G^r}(\omega^{j,1}),P_{i,G^r}(\omega^{j,0})\big)}{2}\right)^2\Bigg|\mathscr{X}\right].\end{aligned}$$

Clearly,

$$\begin{aligned}(5.9)\qquad &H^2\big(P_{i,G^r(\omega^{j,1})},P_{i,G^r(\omega^{j,0})}\big)\le\Big(\sqrt{\tfrac{1}{2}+a_n}-\sqrt{\tfrac{1}{2}-a_n}\Big)^2,\\ &\lambda(\mathscr{T}_{jr})\ge c_1(Mm^\gamma)^{-1}.\end{aligned}$$

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

$$E_\omega\Big[E_{G^r(\omega)}\Big(d_1\Big(G^r(\omega) \cap \mathscr{T}_{jr}, \hat{G}_n \cap \mathscr{T}_{jr}\Big)|\mathscr{X}\Big)\Big]$$

$$\geq c_2(Mm^\gamma)^{-1}\Big(\tfrac{1}{2} + \sqrt{\tfrac{1}{4} - a_n^2}\Big)^{2\,\mathrm{card}\,U_{jr}} \tag{5.10}$$

$$\geq c_2(Mm^\gamma)^{-1}\big(1 - 2a_n^2\big)^{2\,\mathrm{card}\,U_{jr}}.$$

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

$$\mathrm{card}\,\mathscr{D}_{1m} \geq c_3 m^\gamma,$$

$$\mathrm{card}\,\Psi_r \geq M/2 \quad \text{for } r \in \mathscr{D}_{1m}, \tag{5.11}$$

$$\mathrm{card}\,U_{jr} \leq c_4 n/(m^\gamma M) \leq c_4/a_n^2, \quad \text{for } j \in \Psi_r \text{ and } r \in \mathscr{D}_{1m},$$

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

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

$$\sup_{G \in \mathscr{G}_{\gamma, L}} E_G\big(d_1(\hat{G}_n, G)\big) \geq \frac{1}{m^\gamma} E_{\mathscr{X}}\left[E_\omega\left\{\sum_{r \in \mathscr{D}_{1m}} \sum_{j \in \Psi_r} \frac{c_2}{(Mm^\gamma)}\big(1 - 2a_n^2\big)^{2\,\mathrm{card}\,U_{jr}}\right\}\right]$$

$$\geq \frac{c_3 M}{2} \frac{c_2}{(Mm^\gamma)}\big(1 - 2a_n^2\big)^{2c_4/a_n^2}.$$

Since $a_n \leq 1/2$, then $(1 - 2a_n^2)^{2c_4/a_n^2}$ is larger than some constant. The assumption that $na_n^2 \geq 1$ entails the inequality $m^{-\gamma} = ([(na_n^2)^{1/(\gamma+N-1)}] + 1)^{-\gamma} \geq 2^{-\gamma}(na_n^2)^{-\gamma/(\gamma+N-1)}$. 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

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

and the densities

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

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

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

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

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

where

$$
\mathscr{T}_j=\left\{x\in[0,1]^N: g\left(x_1,\dots,x_{N-1},\omega_0^{j,0}\right)\le x_N\le g\left(x_1,\dots,x_{N-1},\omega_0^{j,1}\right)\right\}
$$

and

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

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

$$
\begin{aligned}
\bar J\equiv\int\sqrt{f_{G(\omega^{j,1})}(x)f_{G(\omega^{j,0})}(x)}\,dx &= \frac{\bar\lambda+d_1\Big[\sqrt{(1/4)-a_n^2}-((1/2)-a_n)\Big]}{\sqrt{\bar\lambda^2+2a_n\bar\lambda d_1}}\\
&\ge\left(\frac{\bar\lambda^2+2\bar\lambda d_1\Big[\sqrt{(1/4)-a_n^2}-((1/2)-a_n)\Big]}{\bar\lambda^2+2a_n\bar\lambda d_1}\right)^{1/2}\\
&\ge 1-4d_1\left[\frac{1}{2}-\sqrt{\frac{1}{4}-a_n^2}\right]\ge 1-8a_n^2 d_1,
\end{aligned}\tag{5.14}
$$

where $\bar\lambda=\bar\lambda(G(\omega^{j,1}))$ and $d_1=d_1(G(\omega^{j,0}),G(\omega^{j,1}))$. From (5.14) we obtain

$$
\begin{aligned}
&H^2\big(P_{i,G(\omega^{j,1})},P_{i,G(\omega^{j,0})}\big)\\
&\quad=2(1-\bar J)\le 16a_n^2 d_1\big(G(\omega^{j,0}),G(\omega^{j,1})\big)\\
&\quad=16c_6a_n^2(Mm^\gamma)^{-1}\le c_7 n^{-1}.
\end{aligned}\tag{5.15}
$$

From (5.13) and (5.15) we get

$$
\sup_{G\in\mathscr{G}_{\gamma,L}} E_G\big(d_1(G,\hat G_n)\big)\ge c_5 m^{-\gamma}\big(1-(c_7/2)n^{-1}\big)^{2n}.
$$

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 set G on classes $\mathscr{G}_{\gamma,L}$. 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*

$$\liminf_n \inf_{\hat{G}_n} \sup_{G\in\mathscr{G}_{\text{conv}}} \left(na_n^2\right)^{2q/(N+1)} E_G\left(d_1^q\left(\hat{G}_n, G\right)\right) \geq C_0'. \tag{5.16}$$

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

PROOF. 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 $t = (x_1,\dots,x_{N-1})$. The sets

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

are convex if λ_0 is small enough. In fact

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

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 $\mathscr{X}$ be fixed. Introduce for $r = 0,1,\dots,m^\gamma - 1$ the curved strips

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

It is clear that there exist at least $3m^\gamma/4$ strips V_r such that the number of points X_i in each strip is less than $4n/m^\gamma$ (otherwise the total number of points would exceed n). This entails that there exist not less than $m^\gamma/4$ 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 $\mathscr{D}_{1m}$. Then

$$\operatorname{card} \mathscr{D}_{1m} \geq m^\gamma/4. \tag{5.17}$$

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

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

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

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

where

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

and $I_r = \{j\colon b_j \in Q_r\}$. 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 $r \in \mathscr{D}_{1m}$ there exist at least $\beta M/2$ pieces of the form

$$\big\{x \in V_r\colon |x_k - b_j| \leq 1/(2m),\, k = 1,\dots,N-1\big\}, \qquad j = 1,\dots,M_r,$$

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

$$\mathscr{T}_{jr1} = \Big\{x \in [0,1]^N\colon \tilde{g}_{r1}\big(x_1,\dots,x_{N-1},\omega_0^{j,0}\big) \leq x_N \leq \tilde{g}_{r1}\big(x_1,\dots,x_{N-1},\omega_0^{j,1}\big)\Big\}.$$

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

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

where $U_{jr1} = \{i\colon X_i \in \mathscr{T}_{jr1}\}$.

Note that (5.17) and (5.18) are of the same form as (5.11) and that $\lambda(\mathscr{T}_{jr1}) = \lambda(\mathscr{T}_{jr})$. 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 $G(\omega) = \{x \in \{0,1]^N\colon 0 \leq x_N \leq g_1(t,\omega)\}$. $\square$

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 $(\mathscr{Z},\mathscr{F})$ and indexed by parameter $\theta \in \Theta$. Assume that there exists a measure ν on $(\mathscr{Z},\mathscr{F})$ which dominates the measures P_θ. We define for $Z \in \mathscr{Z}$,

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

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

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

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

Assume either of the following assumptions:

ASSUMPTION 1. For $\theta, \theta' \in \mathscr{N}_\varepsilon$ 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 $\theta \in \mathcal{N}_\varepsilon$ such that $d(\theta, \tilde{\theta}) \le \varepsilon$ and

$$P_\theta\left\{c_{p1} p(Z, \theta) \ge p(Z, \tilde{\theta})\right\} \le c_{p2}, \tag{A.2}$$

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

ASSUMPTION 2. There exist $c_{d1}, c_{d2}, \alpha > 0$, $t_0 > 1$, such that

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

for $t \ge t_0$ and for every $\varepsilon > 0$.

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

LEMMA A.1. *Assume Assumption* 1 *or* 2 *and let there exist an* $\varepsilon > 0$ *such that*

$$-\tfrac{1}{2} \log c_{p1} + \log(\operatorname{card} \mathcal{N}_\varepsilon) \le \tfrac{1}{2} c_{d2} \varepsilon^\alpha t^\alpha, \qquad t \ge t_1, \tag{A.4}$$

for some $t_1 > 0$. *Here we set, by definition,* $c_{p1} = 1$ *and* $c_{p2} = 0$ *if Assumption* 2 *holds. Then*

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

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 $t > 1$,

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

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 1 holds. Choose $\tilde{\theta}$ such that (A.2) holds. Then

$$
\begin{aligned}
P_\theta\Bigl(\max_{\substack{\theta' \in \mathscr{N}_\varepsilon:\\ d(\theta,\theta') \geq \varepsilon t}} p(Z,\theta') \geq p(Z,\tilde{\theta})\Bigr)
&\leq P_\theta\Bigl(\max_{\substack{\theta' \in \mathscr{N}_\varepsilon:\\ d(\theta,\theta') \geq \varepsilon t}} p(Z,\theta') \geq c_{p1} p(Z,\theta)\Bigr) + c_{p2}\\
&\leq (\operatorname{card} \mathscr{N}_\varepsilon) \max_{\substack{\theta' \in \mathscr{N}_\varepsilon:\\ d(\theta,\theta') \geq \varepsilon t}} P_\theta\Biggl(\sqrt{\frac{p(Z,\theta')}{p(Z,\theta)}} \geq \sqrt{c_{p1}},\ p(Z,\theta) \neq 0\Biggr) + c_{p2} \qquad \text{(A.7)}\\
&\leq c_{p1}^{-1/2} (\operatorname{card} \mathscr{N}_\varepsilon) \max_{\substack{\theta' \in \mathscr{N}_\varepsilon:\\ d(\theta,\theta') \geq \varepsilon t}} \int \sqrt{p(Z,\theta) p(Z,\theta')}\, d\nu + c_{p2}\\
&\leq c_{p1}^{-1/2} c_{d1} (\operatorname{card} \mathscr{N}_\varepsilon) \exp(-c_{d2} \varepsilon^\alpha t^\alpha) + c_{p2}.
\end{aligned}
$$

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

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

Assume now that Assumption 2 holds. Then (A.3), (A.4) and (A.6) entail

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

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

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

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

Acknowledgments. We would like to thank L. Devroye and P. Jacob for references related to support estimation. We would like to thank the referees and an Associate Editor for many helpful remarks.

REFERENCES

ABBAR, H. (1990). Un estimateur spline du contour d'une répartition ponctuelle aléatoire. *Statist. Anal. Données* **15** 1–19.

BÁRÁNY, I. (1992). Random polytopes in smooth convex bodies. *Mathematika* **39** 81–92.

BIRGÉ, L. (1983). Approximation dans les espaces métriques et théorie de l'estimation. *Z. Wahrsch. Verw. Gebiete* **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. Wahrsch. Verw. Gebiete* **47** 119–137.

CHEVALIER, J. (1976). Estimation du support et du contenu du support d'une loi de probabilité. *Ann. Inst. H. Poincaré Sec. B* **12** 339–364.

DEVROYE, L. and GYÖRFI, L. (1985). *Nonparametric Density Estimation: The L_1-View*. Wiley, New York.

DEVROYE, L. and WISE, G. L. (1980). Detection of abnormal behavior via nonparametric estimation of the support. *SIAM J. Appl. 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. *Publ. Inst. Statist. Univ. Paris* **13** 191–120.

GROEMER, M. (1974). On the mean value of the volume of a random polytope in a convex set. *Arch. Math.* **25** 86–90.

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

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

IBRAGIMOV, I. A. and KHASMINSKII, R. Z. (1980). On nonparametric estimation of regression. *Dokl. Akad. Nauk SSSR* **252** 780–784.

IBRAGIMOV, I. A. and KHASMINSKII, R. Z. (1981). *Statistical Estimation: Asymptotic Theory*. Springer, New York.

JACOB, P. (1984). Estimation du contour discontinu d'un processus ponctuel sur le plan. *Publ. Inst. Statist. Univ. Paris* **29** 1–25.

JACOB, P. and ABBAR, H. (1989). Estimating the edge of Cox process area. *Cahiers Centre Études Rech. Opér.* **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 Probab. Appl.* **36** 153–159.

KOROSTELEV, A. P. and TSYBAKOV, A. B. (1992). Asymptotically minimax image reconstruction problems. In *Topics in Nonparametric Estimation* (R. Z. Khasminskii, ed.) 45–86. Amer. Math. Soc., Providence, RI.

KOROSTELEV, A. P. and TSYBAKOV, A. B. (1993a). Estimation of the density support and its functionals. *Problems Inform. Transmission* **29** 3–18.

KOROSTELEV, A. P. and TSYBAKOV, A. B. (1993b). *Minimax Theory of Image Reconstruction. Lecture Notes in Statist.* **82**. Springer, New York.

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. (1991). Using excess mass estimates to investigate the modality of a distribution. In *The Frontiers of Statistical Scientific Theory and Industrial Application* **2**. *Proceedings of the ICOSCO-I Conference* (A. Öztürk, E. C. van der Muelen, E. J. Dudewicz and P. R. Nelson, eds.) 355–382. American Science Press, Syracuse, NY.

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. *Comput. Graphics Image Process.* **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 Inform. 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 dimension, Preprint 611, SFB 123, Univ. 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. Wahrsch. Verw. Gebiete* **2** 75–84.

RÉNYI, A. and SULANKE, R. (1964). Über die knovexe Hülle von n zufällig gewählten Punkten II. *Z. Wahrsch. Verw. Gebiete* **3** 138–147.

RIPLEY, B. D. and RASSON, J. P. (1977). Finding the edge of a Poisson forest. *J. Appl. Probab.* **14** 483–491.

RUDEMO, M., SKOVGAARD, I. and STRYHN, H. (1990). Maximum likelihood estimation of curves in images, Report 90-4, Dept. Mathematics and Physics, Royal Veterinary and Agricultural Univ., Frederiksberg, Denmark.

RUDEMO, M. and STRYHN, H. (1994). Approximating the distributions of maximum likelihood contour estimators in two-region images. *Scand. J. Statist.* **21** 41–56.

SCHNEIDER, R. (1988). Random approximation of convex sets. *Journal of Microscopy* **151** (Pt. 3) 211–227.

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 Inform. 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.

INSTITUT FÜR STOCHASTIK
FACHBEREICH MATHEMATIK
HUMBOLDT-UNIVERSITÄT ZU BERLIN
UNTER DEN LINDEN 6
10099 BERLIN
GERMANY

LABORATOIRE DE STATISTIQUE THÉORIQUE ET APPLIQÉE
UNIVERSITÉ PIERRE ET MARIE CURIE
4 PLACE JUSSIEU
75252 PARIS CEDEX 05
FRANCE