The Annals of Statistics
1997, Vol. 25, No. 3, 948–969

ON NONPARAMETRIC ESTIMATION OF DENSITY LEVEL SETS

BY A. B. TSYBAKOV

Université Paris VI and Institute for Problems of Information Transmission, Moscow

Let $X_1, \dots, X_n$ be independent identically distributed observations from an unknown probability density $f(\cdot)$. Consider the problem of estimating the level set $G = G_f(\lambda) = \{x \in \mathbb{R}^2: f(x) \geq \lambda\}$ from the sample $X_1, \dots, X_n$, under the assumption that the boundary of G has a certain smoothness. We propose piecewise-polynomial estimators of G based on the maximization of local empirical excess masses. We show that the estimators have optimal rates of convergence in the asymptotically minimax sense within the studied classes of densities. We find also the optimal convergence rates for estimation of convex level sets. A generalization to the N-dimensional case, where $N > 2$, is given.

1. Problem statement. Let $X_1, \dots, X_n$ be independent identically distributed observations from an unknown probability density $f(x)$, $x \in \mathbb{R}^N$, and let $\lambda > 0$ be a fixed number. Consider the λ-level set of the density f:

$$G = G_f(\lambda) = \{x \in \mathbb{R}^N: f(x) \geq \lambda\}.$$

The problem is to estimate $G = G_f(\lambda)$ given $X_1, \dots, X_n$. By an estimator of G we mean an arbitrary closed set $\hat{G}_n$ in $\mathbb{R}^N$, measurable with respect to $X_1, \dots, X_n$.

Estimation of level sets is useful in various situations. For example, the problem of cluster analysis may be reduced to that of estimation of density level sets [Hartigan (1975)]. The approach based on level sets can be applied to define tests of multimodality, as proposed by Müller and Sawitzki (1991) [see also Müller (1993) and Polonik (1995)]. Another interesting field, where the level sets estimation may be implemented, is nonlinear extension of principal component analysis. By comparing the shapes of a suitable number of level sets, related to different levels, one can look for a hidden common structure in them, which could be interpreted in terms of nonlinear (in general) "principal" curves. Estimation of support of a density f (the "0-level" set of f) has been suggested in certain statistical problems as well. Devroye and Wise (1980), Grenander (1981), Cuevas (1990), Cuevas and Fraiman (1993) used density support estimation for pattern recognition and for detection of the abnormal behavior of a system. Also in econometrics one finds some interesting connections to support estimation [see Korostelev and Tsy-

Received May 1994; revised September 1996.

AMS 1991 *subject classifications.* 62G05, 62G20.

Key words and phrases. Density level set, excess mass, shape function, optimal rate of convergence, piecewise-polynomial estimator.

bakov (1993b), Chapter 7 for further references and discussion]. In all these examples the estimators of general level sets are of interest as well. In fact, one looks for sets with high concentration of sample points, and it is sometimes better to define them as level sets (and not as supports), to avoid the effect of possible outliers.

Assume that an estimator $\hat{f}_n(x)$ of the density $f(x)$ is available. Then a straightforward estimator of level set $G_f(\lambda)$ is $\hat{G}_n = \{x \in \mathbb{R}^N: \hat{f}_n(x) \geq \lambda\}$, the *plug-in estimator*. If the density f has no "flat parts" at the level λ (for example, $|\nabla f|$ is bounded away from 0 around the level λ, where ∇f is the gradient of f), then $\hat{G}_n$ is a consistent estimator of G, and it inherits in a certain sense the convergence rate of $\hat{f}_n$ [Molchanov (1990, 1993), Cuevas and Fraiman (1993)]. Thus, the plug-in approach relates the properties of level set estimators to the smoothness of density f, and does not care about a specific shape of a level set. However, it is sometimes convenient to introduce certain restrictions on level sets directly. For example, one could be interested in estimating ellipsoidal level sets [Nolan (1991)], or convex level sets [Hartigan (1987)], or the level sets satisfying certain smoothness conditions on the boundary [Polonik (1995)]. In these examples the behavior of the density f is not very important; it may be even discontinuous, and the plug-in approach does not work. Hartigan (1987) and Müller and Sawitzki (1991) proposed another approach to level set estimation based on *excess mass*. By definition, the excess mass of a measurable set G in $\mathbb{R}^N$ is the value $M(G) = \int_G f(x)\,dx - \lambda \operatorname{mes}(G)$, where $\operatorname{mes}(G)$ denotes the Lebesgue measure of G. It is clear that $M(G_f(\lambda)) \geq M(G)$ for any G. The empirical excess mass is defined as $M_n(G) = (1/n)\sum_{i=1}^n I\{X_i \in G\} - \lambda \operatorname{mes}(G)$. A natural estimator of $G_f(\lambda)$ is the maximizer of $M_n(G)$ over a given class of level sets G. Such estimators, for different classes of level sets, were studied by Hartigan (1987), Müller and Sawitzki (1991), Nolan (1991), Müller (1993) and Polonik (1995), who proved consistency and found certain rates of convergence of the estimators. An interesting question is whether the excess mass approach gives estimators with optimal convergence rates. In this paper we show that in some cases the answer to this question is positive. However, to get results on optimal rates in more generality, we need to introduce a new class of estimators, based on maximization of *local empirical excess masses*, instead of the usual global maximizers. These estimators have a piecewise-polynomial structure, and they are studied by a technique related to that of Korostelev and Tsybakov (1993b).

Let us introduce some notation. Let $w(t)$ be a *loss function*, that is, the function defined for nonnegative t and having the following properties: $w(t)$ is nonnegative, nondecreasing, $w(t) \not\equiv 0$, $w(0) = 0$, and

$$w(t) \leq 1 + t^q, \qquad t \geq 0,$$

for some $q > 0$.

Define the risk of any estimator $\hat{G}_n$ as

$$r_n\big(G_f(\lambda), \hat{G}_n, \psi_n\big) = E_f\Big[w\Big(\psi_n^{-1} d\big(G_f(\lambda), \hat{G}_n\big)\Big)\Big],$$

where d is a distance between G and $\hat{G}_n$, $E_f(\cdot)$ denotes the expectation with respect to the probability measure P_f of $X_1, \dots, X_n$, when the underlying density is f and ψ_n is a normalizing sequence of positive numbers.

We use the following two distances between closed sets G_1, G_2: Lebesgue measure of symmetric difference $d = d_1$, with

$$d_1(G_1, G_2) = \operatorname{mes}(G_1 \Delta G_2),$$

and the Hausdorff metric $d = d_\infty$, with

$$d_\infty(G_1, G_2) = \max\left\{\max_{x \in G_1} \rho(x, G_2); \max_{x \in G_2} \rho(x, G_1)\right\},$$

where $\rho(x, G) = \min_{y \in G} |x - y|$ is the Euclidean distance between a point x and a closed set G.

We study the problem of estimating the level set $G_f(\lambda)$ in the nonparametric minimax setting, in the spirit of Ibragimov and Khasminskii (1981) and Stone (1980). Namely, we introduce a class $\mathscr{F}$ of densities f such that their level sets belong to some class of sets $\mathscr{G}$. We consider the maximal risk of $\hat{G}_n$ over a class $\mathscr{F}$:

$$\mathscr{R}_n\left(\mathscr{F}, \hat{G}_n, \psi_n\right) = \sup_{f \in \mathscr{F}} r_n\left(G_f(\lambda), \hat{G}_n, \psi_n\right).$$

Our aim is to find an estimator G_n^* that converges to G as $n \to \infty$ with the best possible rate, that is, the estimator G_n^* satisfying

$$\limsup_{n \to \infty} \mathscr{R}_n(\mathscr{F}, G_n^*, \psi_n) < \infty, \tag{1}$$

where the sequence $\psi_n = \psi_n(\mathscr{F}, d)$ is such that

$$\liminf_{n \to \infty} \inf_{\hat{G}_n} \mathscr{R}_n\left(\mathscr{F}, \hat{G}_n, \psi_n\right) > 0 \tag{2}$$

($\inf_{\hat{G}_n}$ denotes the infimum over all estimators).

If (1) and (2) hold, we call ψ_n the *optimal rate of convergence*, and we say that G_n^* has optimal rate of convergence on the class of densities $\mathscr{F}$ (in d-metric).

In this paper we find the optimal rates of convergence in d_1 and d_∞-metrics and the optimal level set estimators G_n^* for certain classes of densities $\mathscr{F}$. We mainly consider the case $N = 2$, since the proofs for $N > 2$ are quite similar, but they require more notation. The results for general N are stated in Section 5; elsewhere we assume that $N = 2$.

2. Classes of sets and classes of densities. Let $N = 2$, and let γ and L be positive constants. Denote by $\Sigma(\gamma, L)$ the class of functions $g(t)$, $t \in \mathbb{R}^1$, such that the derivatives of g up to the order k exist, are 2π-periodic, and the kth derivative satisfies the Hölder condition

$$\left|g^{(k)}(t) - g^{(k)}(t')\right| \le L|t - t'|^{\gamma - k} \quad \forall\, t, t' \in \mathbb{R}^1,$$

where $k = \lfloor \gamma \rfloor$ (i.e., k is the largest integer which is strictly less than γ).

We call the set $G \subseteq \mathbb{R}^2$ *star-shaped* about the origin if in polar coordinates (r, φ), with $0 \le r < \infty$, $0 \le \varphi < 2\pi$, it has the form

$$(3) \qquad G = \{x = (r, \varphi): 0 \le r \le g(\varphi), 0 \le \varphi < 2\pi\},$$

where $g(\cdot)$ is a 2π-periodic continuous function on $\mathbb{R}^1$ called the *shape function.*

Denote by $\mathscr{G}(\gamma, L, h)$ the class of all star-shaped sets of the form (3) such that $g \in \Sigma(\gamma, L)$ and $g(\varphi) \ge h$, $\forall\, \varphi \in [0, 2\pi)$, where $h > 0$ is a given number. We assume that the level set $G_f(\lambda)$ belongs to the class $\mathscr{G}(\gamma, L, h)$, and denote by $g_\lambda(\varphi)$ the shape function that corresponds to $G_f(\lambda)$.

Let f be a probability density, with a star-shaped level set $G_f(\lambda)$, and let $0 \le \alpha < \infty$. For $\alpha > 0$, we say that the density f is *α-regular around the level* λ, if there exist constants $b_2 > b_1 > 0$, $\delta_0 > 0$ such that

$$(4) \qquad b_1 \le \frac{|f(x) - \lambda|}{|r - g_\lambda(\varphi)|^\alpha} \le b_2,$$

for all $x = (r, \varphi)$ such that $|f(x) - \lambda| \le \delta_0$. If the set $\{x: |f(x) - \lambda| \le \delta_0\}$ is empty, we put $\alpha = 0$, and we say that f is *0-regular around the level* λ. This corresponds to the case where the density f has a jump around the level λ. In general, condition (4) excludes the possibility for a density f to have flat parts at level λ. The behavior of the usual densities, such as the Gaussian one, corresponds to $\alpha = 1$. If condition (4) holds only with $\alpha > 1$, it may indicate that there is a local extremum or inflexion point at the level λ.

Fix positive numbers $\gamma, L, h, \lambda, b_1, \delta_0$ and $b_2 > b_1$, $f_{\max} > \lambda$, $\alpha \ge 0$. Denote by $\mathscr{F}_{\gamma, \alpha}$ the class of all probability densities f on $\mathbb{R}^2$ satisfying simultaneously the following conditions:

C1. f is uniformly bounded by $f_{\max}$ on $\mathbb{R}^2$;
C2. f is α-regular around the level λ;
C3. the level set $G_f(\lambda) \in \mathscr{G}(\gamma, L, h)$.

For brevity, the dependence of $\mathscr{F}_{\gamma, \alpha}$ on $\lambda, L, h, f_{\max}, b_1, b_2, \delta_0$ is not shown explicitly in the notation (these parameters do not influence the optimal rate of convergence ψ_n). It will be always assumed that δ_0 is sufficiently small (at least, $\delta_0 < \lambda$, $\delta_0 < f_{\max} - \lambda$), without specifying explicitly how small it is. One can easily derive the necessary restrictions on δ_0 in terms of λ, h, α, b_1, b_2 and $f_{\max}$, but they are cumbersome and of no interest for the results.

REMARK 1. There exists $R > 0$ that depends only on γ, L and λ, such that $G_f(\lambda) \subseteq B(0, R)$, $\forall\, f \in \mathscr{F}_{\gamma, \alpha}$, where $B(0, R)$ is the circle of radius R centered at 0. In fact, since f is a density, we have

$$1 \ge \lambda \operatorname{mes}(G_f(\lambda)) = \lambda \int_0^{2\pi} \int_0^{g_\lambda(\varphi)} r\, dr\, d\varphi = (\lambda/2) \int_0^{2\pi} g_\lambda^2(\varphi)\, d\varphi.$$

Thus, g_λ is bounded in $L_2[0, 2\pi]$, uniformly over $\mathscr{F}_{\gamma, \alpha}$. Since also $g_\lambda \in \Sigma(\gamma, L)$, it follows that there exists a constant $R = R(\gamma, L, \lambda)$ such that $\sup_\varphi g_\lambda(\varphi) \le R$.

EXAMPLE 1 (A family of densities that belong to $\mathscr{F}_{\gamma,1}$). This example describes densities that have homothetic star-shaped level sets for all levels. As a particular case, one gets elliptically contoured densities, such as the multivariate normal.

Take a monotone decreasing differentiable function $f_*: [0,\infty) \to [0,\infty)$ and a shape function $g_* \in \mathscr{G}(\gamma, L, h)$. Define the bivariate density (in polar coordinates):

$$f(r,\varphi) = \kappa f_*(r/g_*(\varphi)),$$

where the constant $\kappa > 0$ is chosen so that $f(r,\varphi)$ integrates to 1 over $\mathbb{R}^2$. Let $\lambda < \kappa f_*(0)$ be given. Denote by t_λ the positive number such that $\kappa f_*(t_\lambda) = \lambda$. Clearly, the shape function of the λ-level set of $f(r,\varphi)$ is $g_\lambda(\varphi) = t_\lambda g_*(\varphi)$. As $\kappa f'_*(t_\lambda) \neq 0$, there exist positive constants b'_1, b'_2 such that

$$b'_1 \le \frac{|\kappa f_*(t) - \lambda|}{|t - t_\lambda|} \le b'_2,$$

for all t such that $|\kappa f_*(t) - \lambda|$ is small enough. Substituting here $t = r/g_*(\varphi)$, $g_\lambda(\varphi) = t_\lambda g_*(\varphi)$, one obtains (4) with $\alpha = 1$. Hence, $f \in \mathscr{F}_{\gamma,1}$.

Along with $\mathscr{F}_{\gamma,\alpha}$ consider a class of densities with convex level sets. Denote

$$\mathscr{G}_{\text{conv}} = \left\{G \subseteq B(0,R): G \text{ is a closed convex set, } G \supseteq B\left(0, (\mu_0/\pi)^{1/2}\right)\right\},$$

where $0 < \mu_0 < \lambda^{-1}$ and $R > (\mu_0/\pi)^{1/2}$ are fixed constants. Clearly, any set $G \in \mathscr{G}_{\text{conv}}$ is star-shaped about the origin.

Define $\mathscr{F}_{\text{conv},\alpha}$ as the class of all probability densities f on $\mathbb{R}^2$ that satisfy C1, C2 and the following condition.

C4. The level set $G_f(\lambda) \in \mathscr{G}_{\text{conv}}$.

Again, we keep in the notation only the dependence of $\mathscr{F}$ on α, since the values of $\lambda, \mu_0, f_{\max}, b_1, b_2, \delta_0$ do not influence the optimal rate of convergence.

3. Level set estimators for $\mathscr{F}_{\gamma,\alpha}$. In this section we define the estimators of level sets $G_f(\lambda)$ that have optimal convergence rates on the classes of densities $\mathscr{F}_{\gamma,\alpha}$ in d-metrics, with $d = d_1$ or $d = d_\infty$. The definition of the estimator depends on γ, α and on d.

Let

$$\delta_n = \begin{cases} (1/n)^{1/((2\alpha+1)\gamma+1)}, & \text{if } d = d_1, \\ (\log n/n)^{1/((2\alpha+1)\gamma+1)}, & \text{if } d = d_\infty. \end{cases}$$

Denote $M = M_n = 1/\delta_n$ and assume without loss of generality that M is an integer. Put $\varphi_l = 2\pi l \delta_n$, $l = 0, 1, \dots, M$, and define the sectors

$$S_l = \{(r,\varphi): 0 \le r \le R,\ \varphi_{l-1} \le \varphi < \varphi_l\}, \qquad l = 1, \dots, M.$$

In each sector S_l, define the parametric family of subsets

$$B_l(\theta) = \{(r, \varphi) \in S_l : 0 \le r \le \theta_0 + \theta_1(\varphi - \varphi_{l-1}) + \cdots + \theta_k(\varphi - \varphi_{l-1})^k\},$$

where $k = \lfloor \gamma \rfloor$, and $\theta = (\theta_0, \dots, \theta_k)$ is a vector of real parameters. Denote

$$\Theta^* = \{\theta : h \le \theta_0 + \theta_1(\varphi - \varphi_{l-1}) + \cdots + \theta_k(\varphi - \varphi_{l-1})^k \le R, \\ \forall \varphi : \varphi_{l-1} \le \varphi < \varphi_l\},$$

where R is the radius of the circle containing the level sets (see Remark 1). The condition $\theta \in \Theta^*$ guarantees that $B_l(\theta) \subseteq S_l$. Note that Θ^* does not depend on l. Consider the following discrete subset of Θ^*:

$$\Theta_n = \{\theta = (m_0 \delta_n^\gamma, m_1 \delta_n^{\gamma-1}, \dots, m_k \delta_n^{\gamma-k})\} \cap \Theta^*,$$

where $m = (m_0, m_1, \dots, m_k)$ is a $(k+1)$tuple of integers.

Define the piecewise-polynomial estimator of the shape function $g_\lambda(\varphi)$:

$$\hat{g}(\varphi) = \hat{\theta}_{0l} + \hat{\theta}_{1l}(\varphi - \varphi_{l-1}) + \cdots + \hat{\theta}_{kl}(\varphi - \varphi_{l-1})^k, \\ \varphi_{l-1} \le \varphi < \varphi_l,\ l = 1, \dots, M,$$

where the vector $(\hat{\theta}_{0l}, \dots, \hat{\theta}_{kl}) = \hat{\theta}_l \in \Theta_n$ is a solution of the following maximization problem:

$$\hat{\theta}_l = \arg\max_{\theta \in \Theta_n} J_l(\theta),$$

where $J_l(\theta)$ is the empirical excess mass corresponding to the parametric set $B_l(\theta)$; that is,

$$J_l(\theta) = \frac{1}{n} \sum_{i=1}^{n} I\{X_i \in B_l(\theta)\} - \lambda \operatorname{mes}(B_l(\theta)).$$

The set $\{(r, \varphi): 0 \le r \le \hat{g}(\varphi), 0 \le \varphi < 2\pi\}$ is not necessarily closed. Let us take as an estimator of $G = G_f(\lambda)$ the closure of this set:

$$G_n^* = \mathrm{Cl}\{(r, \varphi) : 0 \le r \le \hat{g}(\varphi), 0 \le \varphi < 2\pi\}. \tag{5}$$

THEOREM 1. *For any loss function w the λ-level set estimator G_n^* has optimal rate of convergence on the class of densities $\mathscr{F}_{\gamma,\alpha}$ in d_1- and d_∞-metrics respectively, depending on the choice of δ_n. The optimal rate has the following form:*

$$\begin{aligned} \psi_n(\mathscr{F}_{\gamma,\alpha}, d_1) &= n^{-\gamma/((2\alpha+1)\gamma+1)}, \\ \psi_n(\mathscr{F}_{\gamma,\alpha}, d_\infty) &= (n/\log n)^{-\gamma/((2\alpha+1)\gamma+1)}. \end{aligned} \tag{6}$$

The proof of Theorem 1 is given in Section 6.

REMARK 2. For the typical case where $\alpha = 1$, one gets the optimal rate of convergence $n^{-\gamma/(3\gamma+1)}$ in d_1, and by a log-factor worse rate in d_∞. For the extremal case $\alpha = 0$, Theorem 1 gives the optimal rate $n^{-\gamma/(\gamma+1)}$, the same as for estimation of support of a uniform density [Korostelev and Tsybakov (1993a, b)]. This coincidence is natural, since in both situations one deals with the jump behavior of a density at the contour representing the boundary of either a level set or of a support.

REMARK 3. To prove that (6) gives optimal rates of convergence, one needs to show the relation (1) (called *upper bound*) and (2) (called *lower bound*). Here we prove that the upper bound is attained by the estimator (5). In the special case $d = d_1$, $\gamma > 1$ the correct upper bound of order $n^{-\gamma/((2\alpha+1)\gamma+1)}$ is attained also on global excess mass estimates [Polonik (1995)].

4. Estimators of convex level sets. In this section we show that certain estimators of convex level sets have optimal rates of convergence. We consider the classes of densities $\mathscr{F}_{\text{conv},\alpha}$. The case of convex level sets appears to be quite different from that studied in Section 3. For the estimation of convex sets, the dependence of optimal rates ψ_n on the metric d is more important: one loses a polynomial factor rather than a logarithmic one when passing from d_1- to d_∞-metric.

Hartigan (1987) introduced the following estimator of convex level sets based on the maximization of the empirical excess mass:

$$G_n^H = \arg\max_{G \text{ convex}} \left[\frac{1}{n} \sum_{i=1}^{n} I\{X_i \in G\} - \lambda \operatorname{mes}(G) \right].$$

He proposed a computational algorithm for G_n^H and conjectured that G_n^H converges to $G_f(\lambda)$ with the rate $(n/\log n)^{-2/7}$ in the Hausdorff metric d_∞. We will show that this conjecture is not true for general convex level sets $G_f(\lambda)$. But first we consider the behavior of G_n^H with respect to the distance d_1.

THEOREM 2. *For any loss function w the λ-level set estimator G_n^H has optimal rate of convergence on the class of densities $\mathscr{F}_{\text{conv},\alpha}$ in d_1-metric. The optimal rate is*

$$\psi_n(\mathscr{F}_{\text{conv},\alpha}, d_1) = n^{-2/(4\alpha+3)}. \tag{7}$$

The proof of Theorem 2 is given in Section 6. In fact, we will prove only the lower bound for the maximal risks. Proof of the upper bound, that is, of the fact that G_n^H attains the rate (7), is an easy consequence of a result of Polonik (1995), and it is therefore omitted.

Note that for the typical case $\alpha = 1$, Hartigan's estimator G_n^H converges with the rate $n^{-2/7}$ in d_1-metric, and this is the optimal rate. On the other hand, in the extreme case $\alpha = 0$ (the density f has a jump at the level λ), Theorem 2 gives $\psi_n = n^{-2/3}$, which is shown by Mammen and Tsybakov (1995) to be the optimal rate in estimation of a convex support of a uniform density in $\mathbb{R}^2$ (cf. Remark 2 in Section 3).

The behavior of convex level sets estimators in the Hausdorff metric is substantially worse, which is shown by the following lower bound.

THEOREM 3. *Let* $\psi_n = n^{-1/(2\alpha+2)}$ *and let* $d = d_\infty$. *Then for any loss function* w *the following lower bound holds*:

$$\liminf_{n\to\infty} \inf_{\hat{G}_n} \mathscr{R}_n\big(\mathscr{F}_{\mathrm{conv},\,\alpha}, \hat{G}_n, \psi_n\big) > 0. \tag{8}$$

The proof of Theorem 3 is given in Section 6.

The rate $\psi_n = n^{-1/(2\alpha+2)}$ is optimal for $\mathscr{F}_{\mathrm{conv},\,\alpha}$ in d_∞-metric, up to a log-factor. This follows from Theorems 1 and 3. In fact, if $f \in \mathscr{F}_{\mathrm{conv},\,\alpha}$, then $f \in \mathscr{F}_{1,\,\alpha}$, since the boundary of a convex set is Lipschitz (it is not hard to find the expression for the Lipschitz constant L in terms of μ_0). Hence, one can use the estimator G_n^* with $\delta_n = (n/\log n)^{-1/(2\alpha+2)}$ (corresponding to $\mathscr{F}_{1,\,\alpha}$ and $d = d_\infty$), and apply the upper bound of Theorem 1. We find that G_n^* has the convergence rate $(n/\log n)^{-1/(2\alpha+2)}$. (Note, however, that G_n^*, which is thus used to estimate a convex level set, is not a convex set itself.) This argument and Theorem 3 lead to the following corollary.

COROLLARY. *For any loss function* w *the optimal convergence rate on the class of densities* $\mathscr{F}_{\mathrm{conv},\,\alpha}$ *in* d_∞-*metric satisfies*

$$n^{-1/(2\alpha+2)} \le \psi_n(\mathscr{F}_{\mathrm{conv},\,\alpha}, d_\infty) \le (n/\log n)^{-1/(2\alpha+2)}.$$

Comparing this with Theorems 1 and 2, we find that, qualitatively, the behavior of the best convex level set estimators in d_1-metric is similar to those of smooth level sets, with smoothness $\gamma = 2$, while the behavior in d_∞-metric corresponds roughly to smoothness $\gamma = 1$. This effect is analogous to that discovered by Korostelev and Tsybakov (1995) in the problem of estimating the convex support of a uniform density. Seemingly, these results are particular cases of a general breakdown effect in optimal convergence rates [cf. Nemirovskii (1985), Donoho, Johnstone, Kerkyacharian and Picard (1995), who discuss similar phenomena in the classical nonparametric regression and density estimation problems].

5. Extensions. Let us mention some extensions of the results of Sections 3 and 4.

E1. Condition (4) can be given in a slightly different form (which is easier to interpret, but harder to handle analytically):

$$b_1 \le \frac{|f(x) - \lambda|}{\rho\big(x, \partial G_f(\lambda)\big)^{\alpha}} \le b_2, \tag{4'}$$

where $\partial G_f(\lambda)$ denotes the boundary of $G_f(\lambda)$. All the results remain valid if (4′) is replaced by (4). This condition can be weakened even further if one is interested in the rates of convergence in d_1-metric only. In fact, for getting the correct upper bounds on the risks in d_1-metric (up to a log-factor) one can apply the method of Polonik (1995) that works under the following condition:

$$\operatorname{mes}\{x: |f(x) - \lambda| \le \delta\} \le b\delta^{1/\alpha}, \qquad 0 < \delta < \delta_0,$$

where $b > 0$, $\delta_0 > 0$ and $\alpha > 0$ are constants. This condition is a weaker analog of the left inequalities in (4) and (4').

E2. The results can be made uniform in λ (i.e., one can add $\sup_{\lambda \in \Lambda}$ in the definition of the maximal risk $\mathscr{R}_n$), if it is assumed that (4) or (4') holds uniformly in $\lambda \in \Lambda$, where Λ is a compact interval.

E3. Theorem 1 can be extended to the dimension $N > 2$ in the following way. Let $\Sigma_N(\gamma, L)$ be the Hölder class of functions on the unit sphere in $\mathbb{R}^N$, and let (r, Φ) be the polar coordinates in $\mathbb{R}^N$, where Φ is an $(N-1)$-vector of polar angles. Denote by $\mathscr{G}_N(\gamma, L, h)$ the class of all star-shaped sets of the form $G = \{x = (r, \Phi): 0 \le r < g(\Phi),\ \Phi \in [0, 2\pi)^{N-1}\}$, such that $g \in \Sigma_N(\gamma, L)$ and $g(\Phi) \ge h$, $h > 0$.

Define the class $\mathscr{F}_{\gamma, \alpha, N}$ of all probability densities f on $\mathbb{R}^N$ bounded by $f_{\max}$, satisfying (4') and such that $G_f(\lambda) \in \mathscr{G}_N(\gamma, L, h)$. As in the two-dimensional case, there exists $R > 0$ such that $G_f(\lambda) \subseteq B(0, R)$, where $B(0, R)$ is a Euclidean ball of radius R centered at 0 (cf. Remark 1).

Consider the partition of $[0, 2\pi)^{N-1}$ into $M = s_n^{N-1}$ cubes $Q_1, \dots, Q_M$ with equal edges of length $2\pi/s_n$, where

$$s_n = \begin{cases} \left[n^{1/((2\alpha+1)\gamma + N - 1)} \right], & \text{if } d = d_1, \\ \left[(n/\log n)^{1/((2\alpha+1)\gamma+N-1)} \right], & \text{if } d = d_\infty. \end{cases}$$

Let $p(\Phi)$ be a polynomial (in dimension $N-1$) of order $k = \lfloor \gamma \rfloor$, and let θ be the vector of its coefficients. We write then $p(\Phi) = p(\Phi; \theta)$. Denote by $\mathfrak{N}_{nl}$ a minimal $s_n^{-\gamma}$-net in Euclidean metric on the set of coefficients

$$\{\theta: h \le p(\Phi; \theta) \le R, \forall\ \Phi \in Q_l\}.$$

Let $J_l(\theta)$ be the empirical excess mass of the set

$$B_l(\theta) = \{x = (r, \Phi): 0 \le r < p(\Phi; \theta), \Phi \in Q_l\},$$

and let $\hat{\theta}_l = \arg\max_{\theta \in \mathfrak{N}_{nl}} J_l(\theta)$. The estimator $G^*_{n,N}$ (N-dimensional analog to G^*_n) is defined as

$$G^*_{n,N} = \operatorname{Cl}\left\{ (r, \Phi): 0 \le r \le \sum_{l=1}^{M} p(\Phi; \hat{\theta}_l) I\{\Phi \in Q_l\} \right\}.$$

The following result is the N-dimensional analog to Theorem 1.

THEOREM 4. *For any loss function w the λ-level set estimator $G^*_{n,N}$ has optimal rate of convergence on the class of densities $\mathscr{F}_{\gamma,\alpha,N}$ in d_1- and d_∞-metrics, respectively, depending on the choice of s_n. The optimal rate has* *the following form*:

$$\psi_n(\mathscr{F}_{\gamma,\alpha,N}, d_1) = n^{-\gamma/((2\alpha+1)\gamma+N-1)},$$

$$\psi_n(\mathscr{F}_{\gamma,\alpha,N}, d_\infty) = (n/\log n)^{-\gamma/((2\alpha+1)\gamma+N-1)}.$$

Theorems 2 and 3 are generalized to the N-dimensional case similarly. In particular, $\psi_n(\mathscr{F}_{\mathrm{conv},\alpha,N}, d_1) = n^{-2/(4\alpha+N+1)}$, and (up to a log-factor) $\psi_n(\mathscr{F}_{\mathrm{conv},\alpha,N}, d_\infty) = n^{-1/(2\alpha+N)}$, where $\mathscr{F}_{\mathrm{conv},\alpha,N}$ is the N-dimensional analog to $\mathscr{F}_{\mathrm{conv},\alpha}$.

E4. In this paper we assume that $\lambda > 0$. As $\lambda \to 0$, we get the problem of density support estimation which is quite different. In fact, let, by definition, $G_f(0) = \mathrm{Cl}\{x \in \mathbb{R}^N: f(x) > 0\}$, and assume that $G_f(0) \in \mathscr{G}_N(\gamma, L, h)$ and C1 and C2 of Section 3 hold, with the necessary modifications [α-regularity around the level 0 means that (4$'$) is satisfied, with $\lambda = 0$ for $x \in G_f(0)$ such that $|f(x)| \le \delta_0$]. Then there exist estimators of $G_f(0)$ that have the d_1 rate of convergence $n^{-\gamma/((\alpha+1)\gamma+N-1)}$ if $N \ge 2$; this rate is optimal [Härdle, Park and Tsybakov (1995)]. We see that, except for the case $\alpha = 0$ (which corresponds to the jump behavior of a density near the boundary of its support or level set), the estimators of support converge faster than the estimators of level sets. This fact is easy to understand by comparing the likelihood ratios (or Hellinger distances) for the "worst hypotheses" appearing in the lower bounds. If $\lambda > 0$, the squared Hellinger distance between a density and its perturbation by a "small" function $\tilde{f} \ge 0$ around the level λ is roughly $\int [(\lambda + \tilde{f}(x))^{1/2} - \lambda^{1/2}]^2\,dx \sim \int \tilde{f}^2(x)\,dx$, while for $\lambda = 0$ this distance is proportional to $\int \tilde{f}(x)\,dx$. Thus, if $\lambda = 0$, the perturbations of smaller order than for $\lambda > 0$ can be distinguished with the same accuracy.

6. Proofs.

PROOF OF THEOREM 1. We introduce some notation. By definition, there is a one-to-one correspondence between the vectors $\theta = (\theta_0, \dots, \theta_k) \in \Theta_n$ and the vectors of integers $m = (m_0, \dots, m_k)$. It is therefore convenient to write $\theta = \theta(m)$.

DEFINITION 1. Denote by $\hat{m}_l$ the vector of integers such that $\hat{\theta}_l = \theta(\hat{m}_l)$.

DEFINITION 2 (Intersection of the slice S_l with the level set $G_f(\lambda)$).

$$D_l = G \cap S_l, \qquad l = 1, \dots, M.$$

For brevity we write

$$g(\varphi) = g_\lambda(\varphi).$$

DEFINITION 3 (The Taylor approximation of the set D_l).

$$B_l^{(0)} = \{(r, \varphi) \in S_l : 0 \le r \le g_l^{(0)}(\varphi)\},$$

where $g_l^{(0)}(\varphi) = \theta_0^{(0)} + \theta_1^{(0)}(\varphi - \varphi_{l-1}) + \cdots + \theta_k^{(0)}(\varphi - \varphi_{l-1})^k$, $\theta_j^{(0)} = m^{(0,j)}\delta_n^{\gamma-j}$, $m^{(0,j)} = [g^{(j)}(\varphi_{l-1})/(j!\delta_n^{\gamma-j})]$, $j = 0, \dots, k$. Here $[\cdot]$ is the integer part of a real number. Denote

$$m_l^{(0)} = \left(m^{(0,0)}, m^{(0,1)}, \dots, m^{(0,k)} \right).$$

For any vector of integers $m = (m_0, \dots, m_k)$ set

$$\|m\| = \max_{0 \le j \le k} |m_j|.$$

Since $g \in \Sigma(\gamma, L)$, we have

$$(9) \quad \left| g(\varphi) - g_l^{(0)}(\varphi) \right| \le \delta_n^\gamma L/k! + (k+1)\delta_n^\gamma = c_* \delta_n^\gamma, \qquad \varphi_{l-1} \le \varphi < \varphi_l,$$

where the summand $(k+1)\delta_n^\gamma$ is due to the discretization, and $c_* = L/k! + (k+1)$.

LEMMA 1. *For any Lebesgue measurable set $B \subseteq S_l$ we have*

$$\int (f(x) - \lambda) I\{x \in B\}\, dx$$
$$= \int (f(x) - \lambda) I\{x \in D_l\}\, dx - \int |f(x) - \lambda| I\{x \in B \Delta D_l\}\, dx.$$

PROOF.

$$\int (f(x) - \lambda) I\{x \in B\}\, dx - \int (f(x) - \lambda) I\{x \in D_l\}\, dx$$
$$= \int (f(x) - \lambda) I\{x \in B \setminus D_l\}\, dx - \int (f(x) - \lambda) I\{x \in D_l \setminus B\}\, dx$$
$$= -\int |f(x) - \lambda| I\{x \in B \Delta D_l\}\, dx. \qquad \square$$

LEMMA 2. *There exist positive constants c_1 and c_2, such that for any vector of integers m with $\theta(m) \in \Theta_n$ we have*

$$(10) \qquad \operatorname{mes}\left(D_l \Delta B_l^{(0)}\right) \le c_1 \delta_n^{\gamma+1},$$

$$(11) \qquad c_2 \delta_n^{\gamma+1} \|m - m_l^{(0)}\| \le \operatorname{mes}\left(B_l(\theta(m)) \Delta B_l^{(0)}\right) \le c_1 \delta_n^{\gamma+1} \|m - m_l^{(0)}\|,$$

$$(12) \qquad \operatorname{mes}(B_l(\theta(m)) \Delta D_l) \ge \delta_n^{\gamma+1}\left(c_2 \|m - m_l^{(0)}\| - c_1\right), \qquad l = 1, \dots, M.$$

The constants c_1 and c_2 do not depend on l, m or on the density $f \in \mathscr{F}_{\gamma,\alpha}$.

PROOF. Using (9) and the inequality $g(\varphi) \le R$, we get

$$\operatorname{mes}\left(D_l \Delta B_l^{(0)}\right) = \iint_{\varphi_{l-1}\ \min(g, g_l^{(0)})}^{\varphi_l\ \max(g, g_l^{(0)})} r\, dr\, d\varphi$$
$$\le (R + c_* \delta_n^\gamma) \int_{\varphi_{l-1}}^{\varphi_l} \left| g(\varphi) - g_l^{(0)}(\varphi) \right| d\varphi \le c_1 \delta_n^{\gamma+1},$$

which gives (10). Similarly, if $\theta(m) \in \Theta_n$,

$$\begin{aligned}
&\operatorname{mes}\big(B_l(\theta(m))\Delta B_l^{(0)}\big)\\
&\quad\le (R + c_*\delta_n^\gamma)\int_{\varphi_{l-1}}^{\varphi_l} \delta_n^\gamma \left|\sum_{j=0}^{k}\big(m_j - m^{(0,j)}\big)\big((\varphi-\varphi_{l-1})/\delta_n\big)^j\right| d\varphi\\
&\quad= (R + c_*\delta_n^\gamma)\,\delta_n^{\gamma+1}\mathscr{J},
\end{aligned}$$

where

$$\mathscr{J} = \int_0^1 \left|\sum_{j=0}^{k}\big(m_j - m^{(0,j)}\big)y^j\right| dy$$

and

$$\operatorname{mes}\big(B_l(\theta(m))\Delta B_l^{(0)}\big) \ge (h - c_*\delta_n^\gamma)\,\delta_n^{\gamma+1}\mathscr{J}.$$

Since all the norms in the space of polynomials of order k are equivalent, the integral $\mathscr{J}$ is bounded from above and from below by the expressions of the form $C\|m - m_l^{(0)}\|$, where $C > 0$ is a constant depending on k only. This proves (11). Now, (12) follows from (10) and (11) by use of the triangle inequality. □

Denote

$$\zeta_{il} = I\{X_i \in B_l(\theta(m))\} - I\{X_i \in B_l^{(0)}\} - \lambda\big[\operatorname{mes}(B_l(\theta(m))) - \operatorname{mes}(B_l^{(0)})\big].$$

LEMMA 3. *There exist positive constants c_3 and c_4, such that for any vector of integers m with $\theta(m) \in \Theta_n$ we have*

$$E_f(\zeta_{1l}) \le -\delta_0\,\delta_n^{\gamma(\alpha+1)+1}\big(c_2\|m - m_l^{(0)}\| - c_3\big),$$

$$\operatorname{Var}_f(\zeta_{1l}) = E_f\big([E_f(\zeta_{1l}) - \zeta_{1l}]^2\big) \le c_4\delta_n^{\gamma+1}\|m - m_l^{(0)}\|, \qquad l = 1,\dots,M,$$

where $\delta_0 > 0$ is the constant used in the definition (4) *of α-regularity, and the constants c_3 and c_4 do not depend on l, m or on the density $f \in \mathscr{F}_{\gamma,\alpha}$.*

PROOF. Denote for brevity $B = B_l(\theta(m))$, $B^0 = B_l^{(0)}$. Using Lemma 1 we get

$$\begin{aligned}
E_f(\zeta_{1l}) &= \int (f(x) - \lambda)I\{x \in B\}\,dx - \int (f(x) - \lambda)I\{x \in B^0\}\,dx\\
&= \int |f(x) - \lambda|\big\{I\{x \in B^0\Delta D_l\} - I\{x \in B\Delta D_l\}\big\}\,dx.
\end{aligned} \tag{13}$$

By virtue of (4), (9) and (10) we have

$$\begin{aligned}
\int |f(x) - \lambda| I\{x \in B^0\Delta D_l\}\,dx &\le \operatorname{mes}(B^0\Delta D_l) \sup_{\substack{x=(r,\varphi):\\ |r-g(\varphi)|\le c_*\delta_n^\gamma}} |f(x) - \lambda|\\
&\le c_1 b_2 c_*^\alpha\,\delta_n^{\gamma(\alpha+1)+1}
\end{aligned} \tag{14}$$

if n is large enough. Let $\delta_0 > 0$ be the constant used in the definition (4) of α-regularity. For any Lebesgue measurable set $A \subseteq S_l$,

$$\begin{aligned}
&\int |f(x) - \lambda| I\{x \in A\}\, dx \\
(15)\qquad &\geq \int |f(x) - \lambda| I\{x \in A, |f(x) - \lambda| \geq \delta_0 \delta_n^{\alpha\gamma}\}\, dx \\
&\geq \delta_0 \delta_n^{\alpha\gamma} \operatorname{mes}\{A \cap \{x \in S_l : |f(x) - \lambda| \geq \delta_0 \delta_n^{\alpha\gamma}\}\} \\
&\geq \delta_0 \delta_n^{\alpha\gamma} \left(\operatorname{mes}(A) - \operatorname{mes}\{x \in S_l : |f(x) - \lambda| \leq \delta_0 \delta_n^{\alpha\gamma}\}\right).
\end{aligned}$$

Let $\alpha > 0$. Then, in view of (4),

$$\begin{aligned}
\operatorname{mes}\{x \in S_l : |f(x) - \lambda| \leq \delta_0 \delta_n^{\alpha\gamma}\} &\leq \operatorname{mes}\{(r, \varphi) \in S_l : b_1 |r - g(\varphi)|^\alpha \leq \delta_0 \delta_n^{\alpha\gamma}\} \\
&\leq 4\pi R (\delta_0 / b_1)^{1/\alpha} \delta_n^{\gamma+1}.
\end{aligned}$$

Thus, the last expression in (15) is bounded from below by

$$\delta_0 \delta_n^{\alpha\gamma} \left(\operatorname{mes}(A) - 4\pi R (\delta_0 / b_1)^{1/\alpha} \delta_n^{\gamma+1}\right).$$

If $\alpha = 0$, then by the definition of α-regularity (see Section 2) the set $\{x: |f(x) - \lambda| \leq \delta_0\}$ is empty. Using this, we get

$$\operatorname{mes}\{x \in S_l : |f(x) - \lambda| \leq \delta_0 \delta_n^{\alpha\gamma}\} = 0,$$

and thus, if $\alpha = 0$, the last expression in (15) is bounded from below by $\delta_0 \delta_n^{\alpha\gamma} \operatorname{mes}(A)$. Hence,

$$\int |f(x) - \lambda| I\{x \in A\}\, dx \geq \delta_0 \delta_n^{\alpha\gamma} \left(\operatorname{mes}(A) - b \delta_n^{\gamma+1}\right),$$

where $b = 4\pi R (\delta_0 / b_1)^{1/\alpha} I\{\alpha > 0\}$. Now, set $A = B \Delta D_l$, and use the last inequality and (12). We get

$$\int |f(x) - \lambda| I\{x \in B \Delta D_l\}\, dx \geq \delta_0 \delta_n^{\gamma(\alpha+1)+1} \left(c_2 \|m - m_l^{(0)}\| - c_1 - b\right),$$

which, together with (13) and (14), yields the first inequality of Lemma 3.

To prove the second inequality, note that

$$\begin{aligned}
\operatorname{Var}_f(\zeta_{1l}) &= \operatorname{Var}_f\left(I\{X_1 \in B\} - I\{X_1 \in B^0\}\right) \\
&\leq E_f\left[\left(I\{X_1 \in B\} - I\{X_1 \in B^0\}\right)^2\right] \\
&= E_f\left[I\{X_1 \in B \Delta B^0\}\right] \leq f_{\max} \operatorname{mes}(B \Delta B^0) \\
&\leq f_{\max} c_1 \delta_n^{\gamma+1} \|m - m_l^{(0)}\|,
\end{aligned}$$

where we used (11). □

PROOF OF THE UPPER BOUND OF THEOREM 1. For any $t > 0$, by definition of $\hat{\theta}_l$ in Section 3 and Definition 1 of $\hat{m}_l$ in terms of $\hat{\theta}_l$, we have

$$
\begin{aligned}
P_f\big(\|\hat{m}_l - m_l^{(0)}\| > t\big) &\le P_f\Big(\max_{\|m - m_l^{(0)}\| > t} \big(J_l(\theta(m)) - J_l\big(\theta\big(m_l^{(0)}\big)\big)\big) > 0\Big) \\
&\le \sum_{r=1}^{\infty} P_f\Big(\max_{m \in \mathscr{L}(r)} \big(J_l(\theta(m)) - J_l\big(\theta\big(m_l^{(0)}\big)\big)\big) > 0\Big) \qquad (16)\\
&\le \sum_{r=1}^{\infty} \operatorname{card} \mathscr{L}(r) \max_{m \in \mathscr{L}(r)} P_f\left(\frac{1}{n}\sum_{i=1}^{n} \zeta_{il} > 0\right),
\end{aligned}
$$

where $\mathscr{L}(r)$ is the following set of vectors m with integer coordinates:

$$\mathscr{L}(r) = \big\{m : tr < \|m - m_l^{(0)}\| \le t(r+1)\big\} \cap \{m : \theta(m) \in \Theta_n\}, \quad r = 1, 2, \dots,$$

and $m_l^{(0)}$ is introduced in Definition 3.

Assume that $t > 2c_3/c_2$. Consequently, $c_2\|m - m_l^{(0)}\| - c_3 \ge c_2\|m - m_l^{(0)}\|/2$, for $m \in \mathscr{L}(r)$, $r \ge 1$. Note that, for l fixed, ζ_{il} are i.i.d. bounded random variables $|\zeta_{il}| \le 2(1 + \pi R^2) = c_5$. Using the Bernstein inequality and Lemma 3, we find

$$
\begin{aligned}
&P_f\left(\frac{1}{n}\sum_{i=1}^{n} \zeta_{il} > 0\right) \\
&\quad\le P_f\left(\frac{1}{n}\sum_{i=1}^{n} \big(\zeta_{il} - E_f(\zeta_{1l})\big) > \delta_0 \delta_n^{\gamma(\alpha+1)+1} c_2 \|m - m_l^{(0)}\|/2\right) \\
&\quad\le 2\exp\left(-\frac{n\big[\delta_0 \delta_n^{\gamma(\alpha+1)+1} c_2 \|m - m_l^{(0)}\|/2\big]^2}{2\operatorname{Var}_f(\zeta_{1l}) + \frac{1}{3} c_5 \delta_n^{\gamma(\alpha+1)+1} c_2 \|m - m_l^{(0)}\|}\right) \qquad (17)\\
&\quad\le 2\exp\left(-\frac{n \delta_n^{2\gamma(\alpha+1)+2} \delta_0^2 \|m - m_l^{(0)}\| (c_2/2)^2}{2c_4 \delta_n^{\gamma+1} + \frac{1}{3} c_5 \delta_n^{\gamma(\alpha+1)+1} c_2}\right) \\
&\quad\le 2\exp\big(-c_6 n \delta_n^{\gamma(2\alpha+1)+1} r t\big),
\end{aligned}
$$

for all $m \in \mathscr{L}(r)$, where $c_6 > 0$ is a constant.

The definition of $\mathscr{L}(r)$ implies $\operatorname{card} \mathscr{L}(r) \le c_7 t^{k+1} r^k$, where $c_7 > 0$ depends only on k. Using this and (16), (17), we obtain

$$
\begin{aligned}
P_f\big(\|\hat{m}_l - m_l^{(0)}\| > t\big) &\le 2c_7 \sum_{r=1}^{\infty} t^{k+1} r^k \exp\big(-c_6 n \delta_n^{\gamma(2\alpha+1)+1} r t\big) \qquad (18)\\
&\le c_8 \exp\big(-c_9 n \delta_n^{\gamma(2\alpha+1)+1} t\big) \quad \text{for } t > 2c_3/c_2,
\end{aligned}
$$

where the constants $c_8, c_9 > 0$ do not depend on f, l and n.

If $d = d_1$, then $\delta_n = n^{-1/(\gamma(2\alpha+1)+1)}$, and (18) reads as

$$P_f(\xi_l > t) \le c_8 \exp(-c_9 t), \qquad t > 2c_3/c_2, \tag{19}$$

where $\xi_l = \|\hat{m}_l - m_l^{(0)}\|$. Next, if $d = d_\infty$, then $\delta_n = (n/\log n)^{-1/(\gamma(2\alpha+1)+1)}$, and (18) reads as

$$P_f(\xi_l > t) \le c_8 n^{-c_9 t}, \qquad t > 2c_3/c_2. \tag{20}$$

Using (19) and (20) and arguing in a standard way [see, e.g., pages 115–117 of Korostelev and Tsybakov (1993b)], one gets, both for $d = d_1$ and $d = d_\infty$,

$$(21) \qquad \sup_{f \in \mathscr{F}_{\gamma,\alpha}} E_f\left(|\delta_n^{-\gamma} d(G, G_n^*)|^q\right) < \infty \qquad \forall\, q > 0,$$

which proves the upper bound (1) in Theorem 1.

PROOF OF THE LOWER BOUND OF THEOREM 1. Consider only the case where $\alpha > 0$, since for $\alpha = 0$ the lower bound is proved in the same way as in the problem of estimating the support of a uniform density [Korostelev and Tsybakov (1993b), Chapter 7].

We will estimate from below the maximal risk over the class of densities $\mathscr{F}_{\gamma,\alpha}$ by the maximal risk over a finite family $\mathscr{N} \subset \mathscr{F}_{\gamma,\alpha}$, and then apply the Fano lemma. This approach is due to Ibragimov and Khasminskii (1981) [see also Birgé (1983)]. To derive the results in the shortest way we use Lemma A.1 of Tsybakov (1982) which we state below for convenience, in the form adapted to our purposes.

LEMMA 4. *Let d be a pseudometric. Assume that there exists a subset $\mathscr{N} \subset \mathscr{F}$ and numbers $\varepsilon > 0$, $0 < \theta < 1$, such that*

$$(22) \quad 2 \le \operatorname{card}(\mathscr{N}) = s < \infty, \qquad d\big(G_f(\lambda), G_h(\lambda)\big) \ge \varepsilon \qquad \forall\, f, h \in \mathscr{N}, f \neq h,$$

and

$$(23) \qquad \max_{f, h \in \mathscr{N}} I(P_f, P_h)/(\log s) \le 1/2,$$

where $I(P_f, P_h)$ is the Kullback information of P_f with respect to P_h. Then, for any loss function w, we have

$$\inf_{\hat{G}_n} \sup_{f \in \mathscr{F}} E_f\left(w\left(\frac{u d\big(G_f(\lambda), \hat{G}_n\big)}{\varepsilon}\right)\right) \ge \nu w\left(\frac{u}{2}\right) \qquad \forall\, u > 0,$$

where $\nu > 0$ is an absolute constant.

Since w is a loss function, there exists $u > 0$ such that $w(u/2) > 0$. Thus, to prove the lower bound we need to construct a family $\mathscr{N} \subset \mathscr{F}_{\gamma,\alpha}$, such that (22) and (23) are satisfied, with $\varepsilon \sim \psi_n(\mathscr{F}_{\gamma,\alpha}, d)$, for n large enough.

CONSTRUCTION OF THE FAMILY $\mathscr{N}$. Let $\eta(t)$ be an infinitely many times differentiable function on $\mathbb{R}^1$, such that $\operatorname{supp} \eta = (-\pi, \pi)$, $\eta(t) \ge 0$, and $\max_t \eta(t) = 1$. Let M be an integer, and let $0 < \tau < 1$,

$$r_0 = \left(\frac{2}{\pi(f_{\max} + \lambda)}\right)^{1/2}, \qquad \Delta = \left(\frac{2\delta_0}{b_1 + b_2}\right)^{1/\alpha}.$$

Note that $\lambda < (\pi r_0^2)^{-1} < f_{\max}$.

Consider the function

$$f_0(x) = \lambda_{\max} I\{r \le r_0 - \Delta\} + \left(\lambda - \frac{b_1 + b_2}{2}|r - r_0|^\alpha \operatorname{sign}(r - r_0)\right)$$
$$\times I\{-\Delta < r - r_0 \le \Delta\}, \qquad x = (r, \varphi),$$

where $\lambda_{\max}$ is a positive constant chosen so that f_0 is a probability density and $f_0 \in \mathscr{F}_{\gamma,\alpha}$. This is possible if δ_0 is small enough. In fact, the equality $\int f_0(x)\,dx = 1$ reads as $\lambda_{\max} = (\pi r_0^2)^{-1} + O(\Delta)$, and thus $\lambda < \lambda_{\max} < f_{\max}$ for δ_0 small enough. Therefore, $G_{f_0}(\lambda) = B(0, r_0)$, $g_\lambda(\varphi) \equiv r_0$, and, if $|f_0(x) - \lambda| \le \delta_0$, we have

$$|f_0(x) - \lambda| = \frac{b_1 + b_2}{2}|r - r_0|^\alpha = \frac{b_1 + b_2}{2}\left|r - g_\lambda(\varphi)\right|^\alpha.$$

For $j = 0, 1, \dots, M$, denote $\varphi_j = 2\pi j/M$,

$$\eta_j(\varphi) = \tau M^{-\gamma}\eta(M(\varphi - (2j-1)\pi/M)),$$

$$r_{*j}(\varphi) = r_0 - \eta_j(\varphi)\frac{(2b_2)^{1/\alpha}}{(2b_2)^{1/\alpha} - (b_1 + b_2)^{1/\alpha}},$$

$$r_j^*(\varphi) = r_0 + \eta_j(\varphi)\frac{(2b_1)^{1/\alpha}}{(b_1 + b_2)^{1/\alpha} - (2b_1)^{1/\alpha}}.$$

We assume that τ is small enough to guarantee that $r_0 - \Delta < r_{*j}(\varphi)$, $r_j^*(\varphi) < r_0 + \Delta$. Let $\omega = (\omega_1, \dots, \omega_M)$ be a vector with binary entries $\omega_j \in \{0, 1\}$. Define the function

$$\begin{aligned} f_1(x, \omega) = \sum_{j=1}^{M} \omega_j \Bigg\{ &\left(b_2\left|r - r_0 + \eta_j(\varphi)\right|^\alpha - \frac{b_1 + b_2}{2}|r - r_0|^\alpha\right) \\ &\times I\left\{r_{*j}(\varphi) < r \le r_0 - \eta_j(\varphi),\ \varphi_{j-1} \le \varphi < \varphi_j\right\} \\ &+ \left(\frac{b_1 + b_2}{2}|r - r_0|^\alpha \operatorname{sign}(r - r_0) - b_1\left|r - r_0 + \eta_j(\varphi)\right|^\alpha\right) \\ &\times I\left\{r_0 - \eta_j(\varphi) < r \le r_j^*(\varphi),\ \varphi_{j-1} \le \varphi < \varphi_j\right\}\Bigg\}, \end{aligned}$$

where $x = (r, \varphi)$. Clearly, $f_1(x, \omega) \le 0$, $\forall\ x, \omega$. Let

$$\Delta_j = \left\{x = (r, \varphi) : r_{*j}(\varphi) < r \le r_j^*(\varphi),\ \varphi_{j-1} \le \varphi < \varphi_j\right\}.$$

LEMMA 5. *The function $f_1(x, \omega)$ and the set Δ_j satisfy*

$$\max_{x,\omega}|f_1(x, \omega)| = \max_{\omega}\max_{x \in \Delta_j}|f_1(x, \omega)| = O(\tau^\alpha M^{-\alpha\gamma}), \tag{24}$$

$$\operatorname{mes}(\Delta_j) = \operatorname{mes}(\Delta_1) = O(\tau M^{-\gamma-1}), \tag{25}$$

$$\int f_1(x, \omega)\,dx = -C_\eta \tau^{\alpha+1}\sum_{j=1}^{M}\omega_j M^{-(\alpha+1)\gamma-1}, \tag{26}$$

where the constant $C_\eta > 0$ depends only on $\eta(\cdot)$.

PROOF. The definition of $r_{*j}(\varphi)$ and $r_j^*(\varphi)$ implies

$$\max_{\omega} \max_{x \in \Delta_j} |f_1(x, \omega)|$$

$$\leq \frac{b_1 + b_2}{2} \max_{\varphi_{j-1} \leq \varphi < \varphi_j} \left(|r_{*j}(\varphi) - r_0|^{\alpha} + |r_j^*(\varphi) - r_0 + \eta_j(\varphi)|^{\alpha} \right)$$

$$= O\left(\max_{\varphi} |\eta_j(\varphi)|^{\alpha} \right) = O(\tau^{\alpha} M^{-\alpha\gamma}).$$

Next, $\mathrm{mes}(\Delta_1) \leq 2\pi M^{-1} \max_{\varphi} \eta_j(\varphi) = O(\tau M^{-\gamma-1})$. Combining (24) and (25), we get $\int f_1(x, \omega)\, dx = O(\tau^{\alpha+1} \sum_{j=1}^{M} \omega_j M^{-(\alpha+1)\gamma-1})$, and it is easy to show that the constant in the $O(\cdot)$ depends on $\eta(\cdot)$ only. $\square$

For any vector ω consider the function $f(x, \omega) = f_0(x) + f_1(x, \omega) + f_2(x, \omega)$, where

$$f_2(x, \omega) = \frac{1}{\pi(r_0 - \Delta)^2} C_{\eta} \tau^{\alpha+1} \sum_{j=1}^{M} \omega_j M^{-(\alpha+1)\gamma-1} I\{r \leq r_0 - \Delta\}.$$

Since $\int (f_1 + f_2) = 0$, $\forall\ \omega$, the function $f(x, \omega)$ is a probability density. Moreover, for M large enough, $\max_x f(x) < f_{\max}$, and the λ-level set of f is that of $f_0 + f_1$, that is, the set

$$G_{f(\cdot, \omega)}(\lambda) = \left\{ x = (r, \varphi) : 0 \leq r \leq r_0 - \sum_{j=1}^{M} \omega_j \eta_j(\varphi) \right\}.$$

For τ small enough the shape function $g(\varphi) = r_0 - \sum_{j=1}^{M} \omega_j \eta_j(\varphi) \in \Sigma(\gamma, L)$, and hence $G_{f(\cdot, \omega)}(\lambda) \in \mathscr{G}(\gamma, L, h)$. The definition of f_0 and f_1 implies that $f_0 + f_1$ is α-regular around the level λ, and thus f is α-regular.

We conclude that $f(\cdot, \omega) \in \mathscr{F}_{\gamma, \alpha}$ for any ω. Define

$$\mathscr{N} = \{ f(x, \omega) : \omega \in \Omega \},$$

where Ω is an appropriately chosen set of binary vectors. The definition of this set is different for $d = d_1$ and $d = d_{\infty}$.

PROOF OF THE LOWER BOUND FOR THE CASE $d = d_1$. Set $M = [n^{1/(\gamma(2\alpha+1)+1)}] + 1$. A well-known combinatorial result [see, e.g., Korostelev and Tsybakov (1993b), Lemma 2.7.4] guarantees the existence of at least $s = [2^{M/8}]$ binary vectors $\omega^1, \dots, \omega^s$ of length M such that

$$\sum_{j=1}^{M} |\omega_j^l - \omega_j^k| \geq M/16, \qquad 1 \leq l, k \leq s,\ k \neq l, \tag{27}$$

where $\omega^l = (\omega_1^l, \dots, \omega_M^l)$, that is, ω_j^l is the jth component of ω^l.

Choose $\mathscr{N}$ as the set of densities $f(x, \omega)$, with $\Omega = \{\omega^1, \dots, \omega^s\}$ such that (27) holds. The condition (22) for the set $\mathscr{N}$ is satisfied with $\varepsilon =$

$(r_0\tau\int\eta/32)M^{-\gamma}$, since, in view of (27),

$$\begin{aligned} d_1\big(G_{f(\cdot,\omega^l)}(\lambda), G_{f(\cdot,\omega^k)}(\lambda)\big) &= \sum_{j=1}^{M} |\omega_j^l - \omega_j^k| \int_{\varphi_{j-1}}^{\varphi_j} d\varphi \int_{r_0-\eta_j(\varphi)}^{r_0} r\,dr \\ &\geq (M/16)\left(r_0\tau M^{-\gamma-1}\int \eta(u)\,du + O(M^{-2\gamma-1})\right) \\ &\geq \left(r_0\tau\int\eta\Big/32\right)M^{-\gamma}, \end{aligned}$$

for M large enough.

Now, for every pair of probability measures $P^l = P_{f(\cdot,\omega^l)}$, $P^k = P_{f(\cdot,\omega^k)}$,

$$\begin{aligned} I(P^l, P^k) &= \int \log\frac{dP^l}{dP^k}\,dP^l = n\int \log\frac{f(x,\omega^l)}{f(x,\omega^k)} f(x,\omega^l)\,dx \\ &= n\int \log\frac{f_0(x)+\tilde{f}_l(x)}{f_0(x)+\tilde{f}_k(x)}\big(f_0(x)+\tilde{f}_l(x)\big)\,dx, \end{aligned} \tag{28}$$

where $\tilde{f}_l(x) = f_1(x,\omega^l) + f_2(x,\omega^l)$, for brevity. Since $f_0+\tilde{f}_l$, $f_0+\tilde{f}_k$ are uniformly bounded from below by $\lambda - \delta_0 > 0$ on their support, it is clear that

$$\left|\log\frac{f_0+\tilde{f}_l}{f_0+\tilde{f}_k} - \frac{1}{f_0+\tilde{f}_k}\big(\tilde{f}_l-\tilde{f}_k\big)\right| \leq c_{13}\big(\tilde{f}_l-\tilde{f}_k\big)^2,$$

where $c_{13} > 0$. Since also $\int\tilde{f}_l(x)\,dx = \int\tilde{f}_k(x)\,dx = 0$, we get, with some $c_{14} > 0$,

$$\begin{aligned} I(P^l, P^k) &\leq c_{14} n\int\big(\tilde{f}_l(x)-\tilde{f}_k(x)\big)^2\,dx \\ &\leq c_{14} n\left(\sum_{j=1}^{M}|\omega_j^l-\omega_j^k|\operatorname{mes}(\Delta_j)\max_{x,\omega}|f_1(x,\omega)|^2\right. \\ &\qquad \left. + C_\eta^2\tau^{2(\alpha+1)}\frac{1}{\pi(r_0-\Delta)^2}\left(\sum_{j=1}^{M}|\omega_j^l-\omega_j^k|\right)^2 M^{-2(\alpha+1)\gamma-2}\right) \\ &= O\big(n(\tau M^{-\gamma})^{(2\alpha+1)}\big), \end{aligned} \tag{29}$$

where we used Lemma 5. Since $s = [2^{M/8}]$, we find

$$I(P^l, P^k)/(\log s) = O(n\tau^{2\alpha+1}M^{-\gamma(2\alpha+1)-1}) \leq c_{15}\tau^{2\alpha+1}.$$

Here $c_{15} > 0$ does not depend on k and l, and as τ can be chosen arbitrarily small, (23) is satisfied. This proves the lower bound if $d = d_1$. □

PROOF OF THE LOWER BOUND FOR THE CASE $d = d_\infty$. Set $M = [(n/\log n)^{1/(\gamma(2\alpha+1)+1)}] + 1$. We use again Lemma 4, but we choose a different set Ω in the definition of the family $\mathscr{N}$: $\Omega = \{\omega^{(1)},\dots,\omega^{(M)}\}$, where $\omega^{(j)}$ is the binary vector of length M whose jth component is 1 and other components are 0. The condition (22) for this family $\mathscr{N}$ is satisfied with $\varepsilon = \tau M^{-\gamma}$, since $d_\infty(G_{f(\cdot,\,\omega^{(l)})}(\lambda), G_{f(\cdot,\,\omega^{(k)})}(\lambda)) = \tau M^{-\gamma} \max_t \eta(t) = \tau M^{-\gamma}$. To check (23) we act similarly to the case of $d = d_1$. Denote $P^{(l)} = P_{f(\cdot,\,\omega^{(l)})}$, $\tilde{f}_l(x) = f_1(x, \omega^{(l)}) + f_2(x, \omega^{(l)})$. Note that $f_2(x, \omega^{(k)}) = f_2(x, \omega^{(l)})$. As in (29),

$$\begin{aligned} I(P^{(l)}, P^{(k)}) &\le c_{14} n \int \left(\tilde{f}_l(x) - \tilde{f}_k(x)\right)^2 dx \\ &= c_{14} n \int \left(f_1^2(x, \omega^{(l)}) + f_1^2(x, \omega^{(k)})\right) dx \\ &\le 2 c_{14} n \operatorname{mes}(\Delta_j) \max_{x,\,\omega} |f_1(x, \omega)|^2 \\ &= O(n \tau^{2\alpha+1} M^{-\gamma(2\alpha+1)-1}) = O(\tau^{2\alpha+1} \log n). \end{aligned}$$

Since $s = M$, (23) is satisfied for τ small enough. □

PROOF OF THEOREM 2. As noted in Section 4, we need to show only the lower bound. Also, it suffices to consider the case $\alpha > 0$, since for $\alpha = 0$ the proof of the lower bound follows the same lines as in Mammen and Tsybakov (1995).

To prove the lower bound of Theorem 2, we use the argument of Theorem 1 (case $d = d_1$). The only difference is in the choice of the functions η_j. We choose them so that the level sets are convex, namely

$$\eta_j(\varphi) = r_0 \left(1 - \frac{\cos(\pi/M)}{\cos(\varphi - (2j-1)\pi/M)}\right), \qquad \varphi_{j-1} \le \varphi < \varphi_j.$$

Then the shape function $g(\varphi) = r_0 - \sum_{j=1}^{M} \omega_j \eta_j(\varphi)$ bounds a convex set, and the function $f(x, \omega)$ belongs to $\mathscr{F}_{\mathrm{conv},\,\alpha}$ for every $\omega \in \Omega$. It is easy to show that, as $M \to \infty$, one has

$$\eta_j(\varphi) = M^{-2} \eta(M(\varphi - (2j-1)\pi/M)) + o(M^{-2}),$$

where $\eta(u) = r_0(\pi^2 - u^2) I\{|u| \le \pi\}$, and $o(\cdot)$ is uniform in φ. Thus, all the calculations are carried out similarly to the proof of the lower bound in Theorem 1, with $\gamma = 2$. However, since now $\tau = 1$, we have to choose M in a slightly different form: $M = [\tau_0 n^{1/(4\alpha+3)}] + 1$, where the factor $\tau_0 > 0$ is large enough to guarantee that the left-hand side of (29) is less than $1/2$. □

PROOF OF THEOREM 3. We use Lemma 4, with $s = 2$. Thus, it suffices to define two densities, $\bar{f}_0$ and $\bar{f}_1$, belonging to $\mathscr{F}_{\mathrm{conv},\,\alpha}$, such that

$$d_\infty\left(G_{\bar{f}_0}(\lambda), G_{\bar{f}_1}(\lambda)\right) \ge \varepsilon, \tag{30}$$

$$I\left(P_{\bar{f}_0}, P_{\bar{f}_1}\right) \le (\log 2)/2, \tag{31}$$

where $\varepsilon \sim n^{-1/(2\alpha+2)}$.

Choose $r_0 = (2/\pi(\lambda + 1/\mu_0))^{1/2}$. Then $\pi r_0^2 > \mu_0$. Let $\Delta > 0$ be small enough that $\pi(r_0 - \Delta)^2 > \mu_0$. Define $\vartheta = 2\arccos(r_0/(r_0 + \Delta))$, and define the shape function

$$\bar{g}(\varphi) = \begin{cases} r_0[\cos(\vartheta/2 - |\vartheta/2 - \varphi|)]^{-1}, & 0 \le \varphi < \vartheta, \\ r_0, & \vartheta \le \varphi \le 2\pi. \end{cases}$$

Let

$$\begin{aligned} \bar{f}_0(x) = {} & \bar{\lambda}_{\max} I\{r \le \bar{g}(\varphi) - \Delta\} \\ & + \left(\lambda - \frac{b_1 + b_2}{2}|r - \bar{g}(\varphi)|^{\alpha} \operatorname{sign}(r - \bar{g}(\varphi))\right) \\ & \times I\{-\Delta < r - \bar{g}(\varphi) \le \Delta\}, \end{aligned}$$

where $x = (r, \varphi)$ and $\bar{\lambda}_{\max}$ is a positive constant chosen so that $\bar{f}_0$ is a probability density, $\bar{f}_0$ is α-regular around the level λ and $\bar{f}_0(x) \le f_{\max}$. Arguing as in the case of f_0 (see the proof of lower bounds in Theorem 1), one easily shows that this is possible if δ_0 and Δ are small enough (but fixed). Now, the level set $G_{\bar{f}_0}(\lambda)$ is convex since it is defined by the shape function $\bar{g}$. Also, $\operatorname{mes}(G_{\bar{f}_0}(\lambda)) > \pi r_0^2 > \mu_0$. Hence $\bar{f}_0 \in \mathscr{F}_{\mathrm{conv}, \alpha}$.

Let us define $\bar{f}_1$. Denote $M = [\tau_0 n^{1/(2\alpha+2)}] + 1$, where $\tau_0 > 0$. Introduce the function

$$\begin{aligned} \bar{\eta}(\varphi) = {} & r_0\left([\cos(\vartheta/2 - |\vartheta/2 - \varphi|)]^{-1} - [\cos(\vartheta/2 - \pi/M)]^{-1}\right) \\ & \times I\{|\vartheta/2 - \varphi| \le \pi/M\}. \end{aligned}$$

Clearly, $\bar{g}(\varphi) - \bar{\eta}(\varphi)$ is the shape function of the convex set $G_{\bar{f}_0}(\lambda) \cap B(0, \bar{r})$, where $\bar{r} = r_0[\cos(\vartheta/2 - \pi/M)]^{-1}$, whose Lebesgue measure is greater than μ_0 for M large. As $M \to \infty$, one has

$$\bar{\eta}(\varphi) = M^{-1}\eta(M(\varphi - \vartheta/2)) + o(M^{-1}),$$

where $\eta(u) = c_{16}(\pi - |u|)I\{|u| \le \pi\}$, $c_{16} > 0$ and $o(\cdot)$ is uniform in φ. Now, one can define $\bar{f}_1$ similarly to $f_1(x, \omega)$. Set

$$r_*(\varphi) = \bar{g}(\varphi) - \bar{\eta}(\varphi)\frac{(2b_2)^{1/\alpha}}{(2b_2)^{1/\alpha} - (b_1 + b_2)^{1/\alpha}},$$

$$r^*(\varphi) = \bar{g}(\varphi) + \bar{\eta}(\varphi)\frac{(2b_1)^{1/\alpha}}{(b_1 + b_2)^{1/\alpha} - (2b_1)^{1/\alpha}},$$

assuming that M is large enough to guarantee that $\bar{g}(\varphi) - \Delta < r_*(\varphi)$, $r^*(\varphi) < \bar{g}(\varphi) + \Delta$, and let

$$\begin{aligned} \bar{f}_{1,1}(x) = {} & \left(b_2|r - \bar{g}(\varphi) + \bar{\eta}(\varphi)|^{\alpha} - \frac{b_1 + b_2}{2}|r - \bar{g}(\varphi)|^{\alpha}\right) \\ & \times I\{r_*(\varphi) < r \le \bar{g}(\varphi) - \bar{\eta}(\varphi)\} \\ & + \left(\frac{b_1 + b_2}{2}|r - \bar{g}(\varphi)|^{\alpha} - b_1|r - \bar{g}(\varphi) + \bar{\eta}(\varphi)|^{\alpha}\right) \\ & \times I\{\bar{g}(\varphi) - \bar{\eta}(\varphi) < r \le r^*(\varphi)\}, \qquad x = (r, \varphi). \end{aligned}$$

As for $f_1(x, \omega)$, one proves that $\bar{f}_{1,1}$ is α-regular around the level λ. Also, define the function $\bar{f}_{1,2}$ proportional to $I\{r \le \bar{g}(\varphi) - \Delta\}$ and such that $\bar{f}_{1,1} + \bar{f}_{1,2}$ integrates to 0. In particular, $\max_x \bar{f}_{1,2}(x) = O(M^{-\alpha-2})$. Finally, put $\bar{f}_1 = \bar{f}_0 + \bar{f}_{1,1} + \bar{f}_{1,2}$. Note that $\bar{f}_1$ is α-regular around the level λ and $G_{\bar{f}_1}(\lambda) = G_{\bar{f}_0}(\lambda) \cap B(0, \bar{r})$ for M large enough. Thus, $\bar{f}_1 \in \mathscr{F}_{\mathrm{conv},\alpha}$ for M large enough.

By the construction, $d_\infty(G_{\bar{f}_0}(\lambda), G_{\bar{f}_1}(\lambda)) = M^{-1}\overline{\eta}(0) = c_{16}\pi M^{-1}(1 + o(1))$, which gives the desired value of ε in (30). Next, as in (29),

$$
\begin{aligned}
&I\left(P_{\bar{f}_0}, P_{\bar{f}_1}\right)\\
&\quad \le c_{14} n \int \bar{f}_1^2(x)\,dx = c_{14} n \int \left(\bar{f}_{1,1}^2(x) + \bar{f}_{1,2}^2(x)\right) dx\\
&\quad \le 2c_{14} n \left(\mathrm{mes}\{(r,\varphi): r_*(\varphi) < r \le r^*(\varphi)\} \max_x \left|\bar{f}_{1,1}(x)\right|^2 + O(M^{-2(\alpha+2)})\right)\\
&\quad = O(nM^{-2\alpha-2}),
\end{aligned}
$$

and choosing τ_0 large enough, we get (31). □

REFERENCES

BIRGÉ, L. (1983). Approximation dans les espaces métriques et la théorie de l'éstimation. *Z. Wahrsch. Verw. Gebiete* **65** 181–238.

CUEVAS, A. (1990). On pattern analysis in the non-convex case. *Kybernetes* **19** 26–33.

CUEVAS, A. and FRAIMAN, R. (1993). Pattern analysis via nonparametric density estimation. Unpublished manuscript.

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

DONOHO, D., JOHNSTONE, I., KERKYACHARIAN, G. and PICARD, D. (1995). Wavelet shrinkage: asymptopia? (with discussion). *J. Roy. Statist. Soc. Ser. B* **57** 301–369.

GRENANDER, U. (1981). *Abstract Inference*. Wiley, New York.

HÄRDLE, W., PARK, B. U. and TSYBAKOV, A. B. (1995). Estimation of non-sharp support boundaries. *J. Multivariate Anal.* **55** 205–218.

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. (1981). *Statistical Estimation: Asymptotic Theory*. Springer, New York.

KOROSTELEV, A. P. and TSYBAKOV, A. B. (1993a). Estimation of the density support and its functionals. *Problems Inform. Transmission* **29** 1–15.

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

KOROSTELEV, A. P. and TSYBAKOV, A. B. (1995). Asymptotic efficiency in estimation of a convex set. *Problems Inform. Transmission* **30** 317–327. [Originally published in *Problemy Perdachi Informatsii* (1994) **30** 33–44. (In Russian.)]

MAMMEN, E. and TSYBAKOV, A. B. (1995). Asymptotical minimax recovery of sets with smooth boundaries. *Ann. Statist.* **23** 502–525.

MOLCHANOV, I. S. (1990). Empirical estimation of distribution quantiles of random closed sets. *Theory Probab. Appl.* **35** 594–600.

MOLCHANOV, I. S. (1993). A limit theorem for solutions of inequalities. Unpublished manuscript.

MÜLLER, D. W. (1993). The excess mass approach in statistics. *Beiträge zur Statistik* **3**. Univ. Heidelberg.

MÜLLER, D. W. and SAWITZKI, G. (1991). Excess mass estimates and tests of multimodality. *J. Amer. Statist. Assoc.* **86** 738–746.

NEMIROVSKII, A. S. (1985). On nonparametric estimation of smooth regression functions. *Izv. Acad. Nauk SSSR, Techn. Kibernetika* **3** 50–60.

NOLAN, D. (1991). The excess-mass ellipsoid. *J. Multivariate Anal.* **39** 348–371.

POLONIK, W. (1995). Measuring mass concentration and estimating density contour clusters—an excess mass approach. *Ann. Statist.* **23** 855–881.

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

TSYBAKOV, A. B. (1982). Signal estimation when there is incomplete information on the noise distribution. *Problems Inform. Transmission* **18** 116–130.

INSTITUT DE STATISTIQUE
URA 1321
UNIVERSITÉ PARIS VI
PL. JUSSIEU, 4, B.P. 158
75252 PARIS
FRANCE
E-MAIL: tsybakov@ccr.jussieu.fr