Change-point Problems
IMS Lecture Notes – Monograph Series (Volume 23, 1994)

MULTIDIMENSIONAL CHANGE-POINT PROBLEMS AND BOUNDARY ESTIMATION

BY A. B. TSYBAKOV
Université Catholique de Louvain
and Russian Academy of Sciences

We consider a multivariate extension of the change-point problem where one has to estimate a change curve (or surface). Three versions of this problem are considered

1) the regression-type model of image segmentation,
2) the estimation of a discontinuity curve in an unknown density, and
3) the estimation of the edge of Poission forest.

For these problems we give two approaches to the construction of estimators, study the rates of convergence of the proposed estimators, and show their optimality.

1. Introduction. Multidimensional change-point problems are the problems of estimating boundaries of regions of certain homogeneity in images. The first example of such a problem arises in image reconstruction: considering an image as a regression function with jump discontinuity between an object and the background, and given the noisy observations of the image, estimate the discontinuity curve. In image analysis this curve is called edge, and the problem is called *edge estimation*. In most of applications one cannot assume a parametric structure of the edge curve. However, it is often possible to postulate some general nonparametric features of the curve, such as continuity, smoothness, convexity, etc. Thus, a nonparametric estimation of edge curves is interesting. This problem was studied recently by Tsybakov (1989, 1991), Korostelev (1991), Korostelev and Tsybakov[1] (1991, 1992a,b 1993), Rudemo, Skovgaard and Stryhn (1990), Rudemo and Stryhn (1991), Carlstein and Krishnamoorthy (1992), Mammen and Tsybakov[2] (1992), Müller and Song (1992

AMS 1991 Subject Classification: Primary 62G05

Key words and phrases: Boundary fragment, edge, image analysis, optimal convergence rates, Poisson forest.

[1] Later referred to as KT.

[2] Later referred to as MT.

a,b), Qiu (1992).

The second multidimensional change-point problem is *estimation of support of a multivariate probability density*, given a sample from this density. The change-point character of this problem appears if the density is greater than some positive constant on its support, and thus it has a discontinuity at the boundary. This problem has various applications in reliability theory (Devroye and Wise (1981)), cluster analysis (Hartigan (1987)), and econometrics (Deprins, Simar and Tulkens (1984)). The study of density support estimation was started by Geffroy (1964) and Rényi and Sulanke (1964) (see also Chevalier (1976), Devroye and Wise (1981), KT (1992c), MT (1992)).

The third problem that we consider here is *estimation of the edge of Poisson forest*. This problem, posed by D. Kendall and first studied by Ripley and Rasson (1977) is the following: given a realization of a multidimensional Poisson point process with an unknown positive intensity function in some compact set G, and zero intensity outside, estimate G. Again, this is a change-point problem if the intensity function has a discontinuity at the boundary of its support G (e.g. the intensity is constant on G, which is the case considered by Ripley and Rasson (1977)). For further results on Poisson forest problem and its generalizations see Jacob and Abbar (1989).

In this paper the three above mentioned multidimensional change-point problems are considered together since the methods of estimation and the results are quite similar for them. Two possible approaches to boundary estimation are discussed, and the convergence rates of the estimators are found. The minimax lower bounds are obtained which show that the proposed estimators have optimal convergence rates.

2. Statistical Models. Assume that one wants to estimate the boundary ∂G of a closed subset G of the N-dimensional cube $[0,1]^N$, using the observations generated by one of the following models.

Model 1. (Edge estimation) One observes $(X_1, Y_1), \dots, (X_n, Y_n)$ where $X_i \in [0,1]^N$, and

$$Y_i = f(X_i) I\{X_i \in G\} + \xi_i, \quad i = 1, \dots, n.$$

Here $f : [0,1]^N \to [0,1]$ is an unknown function, satisfying $f(x) \geq a > 0$, ξ_i are independent zero-mean random variables, such that

$$\sup_x \sup_{i \geq 1} E(\exp t\xi_i \mid X_i = x) \leq c$$

for some $t, c > 0$, and X_i are independent random points in $[0,1]^N$ such that one of the following assumptions holds.

(2.1) $n = n_1^N$ where n_1 is an integer, and X_i is uniformly distributed in a pixel Δ_i, where the pixels $\Delta_1, \ldots, \Delta_n$ are cubes with the side of length $1/n_1$, such that $\cup_{i=1}^n \Delta_i = [0,1]^N$,

(2.2) X_i are i.i.d. random points, distributed with the density $\mu(x)$ on $[0,1]^N$, such that $0 < \mu_1 \leq \mu(x) \leq \mu_2 < \infty$, $x \in [0,1]^N$.

If (2.1) holds, then Model 1 is the typical noisy image model. The values Y_i are interpreted as image grey levels at pixels Δ_i. The set G is the image domain, or *object*, and its complement, $\overline{G} = [0,1]^N \setminus G$ is the *background*.

If (2.2) holds, then we have a different model, where the concentration of points X_i may be higher in some regions and lower in other regions. This model is interesting e.g. in forestry, where X_i's are interpreted as the locations of trees, and Y_i's as their heights.

Model 2. (Density support estimation) One observes $X_1, \ldots, X_n$, the i.i.d. random points, where X_i is distributed with the unknown Lebesgue density $\mu(x) = \mu_G(x)$ such that

$$\mu_G(x) = \frac{Q(x) I\{x \in G\}}{\int_G Q(x)dx}$$

where $Q(x)$ is a continuous function strictly positive on $[0,1]^N$.

Model 3. (Estimation of the edge of Poisson forest) One observes $X_1, \ldots, X_{\nu_n}$, a realization of a Poisson point process with an unknown intensity function $n\lambda(x)$ such that

$$\lambda(x) = \lambda_G(x) = Q(x) I\{x \in G\}$$

(where $Q(x)$ is as in Model 2). Here ν_n is a random variable having the Poisson distribution with mean $n \int_G \lambda(x)dx$, and conditional upon ν_n the random points $X_1, \ldots, X_{\nu_n}$ are i.i.d., with distribution density $\mu(x) = \lambda(x)/(\int_G \lambda(x)dx)$.

3. Assumptions on the Set G. Consider two types of assumptions on the set G.

Special case: boundary fragments. A boundary fragment is a set of the form

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

where $g : [0,1]^{N-1} \to [0,1]$.

For $\gamma, L > 0$ and integer k denote by $\Sigma_k(\gamma, L)$ the class of all functions $g : [0,1]^k \to [0,1]$ having continuous partial derivatives up to the order $\ell = \lfloor \gamma \rfloor$

(i.e. ℓ is the maximal integer satisfying $\ell < \gamma$), and such that absolute value of the difference between $g(z)$ and the Taylor polynomial of g of order ℓ at point $x \in [0,1]^k$ does not exceed $L|z-x|^\gamma$ for all $z \in [0,1]^k$.

For $0 < h < 1/2$ let

$$\Sigma_k(\gamma, L, h) = \Sigma_k(\gamma, L) \cap \{g(x) : h \le g(x) \le 1-h, x \in [0,1]^k\}.$$

Denote by $\mathcal{G}_{0,\gamma} = \mathcal{G}_{0,\gamma}(N, L, h)$ the set of all boundary fragments G, such that $g \in \Sigma_{N-1}(\gamma, L, h)$.

General case: compact classes of sets. In general we suppose that G is an element of some compact (in a metric d) class $\mathcal{G} = \{G\}$ of closed subsets of $[0,1]^N$. Denote by $\mathcal{H}(\varepsilon, \mathcal{G}, d)$ the ε-entropy of $\mathcal{G}$ with respect to the metric d. We consider two examples of d: the Lebesgue measure of symmetric difference,

$$d_1(G, G') = \mathrm{mes}(G \Delta G'),$$

and the Hausdorff metric:

$$d_\infty(G, G') = \max\{\max_{x \in G} \rho(x, G'), \max_{y \in G'} \rho(y, G)\}$$

where $\rho(x, G)$ is the Euclidean distance between x and the set G. We always assume that $\mathcal{G}$ is such that for $G \in \mathcal{G}$ the metric d_∞ is stronger than d_1.

The important examples of compact classes $\mathcal{G}$ are:

1. *Finite classes*, with card $\mathcal{G} = M$, $1 < M < \infty$ (these are e.g. the "classes of candidates" considered by Carlstein and Krishnamoorthy (1992)). For such classes $\mathcal{H}(\varepsilon, \mathcal{G}, d_\infty) \le \log M$.

2. *Parametric classes* $\mathcal{G} = \mathcal{G}^\theta$, where θ is a parameter from a compact subset Θ of $\mathbb{R}^K$. For such classes one usually has

$$\mathcal{H}(\varepsilon, \mathcal{G}, d_\infty) \le \text{const } |\log \varepsilon|. \tag{3.1}$$

3. *Vapnik–Červonenkis classes.* These are (possibly, nonparametric) classes of sets $\mathcal{G}$ satisfying (3.1).

4. *Dudley's classes* $\mathcal{G}_\gamma$, or classes of sets with smooth boundaries (see Dudley (1974)). They are generalizations of boundary fragment classes $\mathcal{G}_{0,\gamma}$, in particular, $\gamma \ge 1$ is the smoothness parameter of the boundary. As shown in Dudley (1974)

$$\mathcal{H}(\varepsilon, \mathcal{G}_\gamma, d_\infty) \le \text{const } \left(\frac{1}{\varepsilon}\right)^{(N-1)/\gamma} \tag{3.2}$$

5. *Classes of sets with monotonicity or convexity restrictions.* Let $\mathcal{G}_{\mathrm{mon}}$ be the class of all boundary fragments G such that the function g is monotone in every coordinate. Clearly, $\mathcal{G}_{\mathrm{mon}}$ is compact, and

$$\mathcal{H}(\varepsilon, \mathcal{G}_{\mathrm{mon}}, d_1) \leq \mathrm{const}\ \frac{1}{\varepsilon^{N-1}}. \tag{3.3}$$

The class of all convex subset of $[0,1]^N$:

$$\mathcal{G}_{\mathrm{conv}} = \{G \subset [0,1]^N : G \text{ is convex }\}$$

is also compact and, as shown in Dudley (1974),

$$\mathcal{H}(\varepsilon, \mathcal{G}_{\mathrm{conv}}, d_\infty) \leq \mathrm{const}\ \left(\frac{1}{\varepsilon}\right)^{(N-1)/2} \tag{3.4}$$

4. The Estimators and Their Convergence Rates (the Case of Fragments). It will be more convenient for us to deal with the more general problem of estimating the set G instead of estimating the boundary ∂G. The following two approaches to this problem will be considered:

1. Extraction of boundary fragments from the original picture and estimation of functions g in each fragment by some smoothing procedure.
2. Maximum likelihood estimation on ε-nets.

Let us describe the first of these approaches. For the extraction of boundary fragments procedure we refer to KT (1991, 1992a, 1993) where Model 1 was considered. Here we only give the details of the boundary estimation procedure, assuming that the fragments are already given. It is sufficient to describe the procedure for one fragment, i.e. to assume that $G \in \mathcal{G}_{0,\gamma}$. Consider a partition of $[0,1]^{N-1}$ into the disjoint cubes Q_j, $j = 1, \ldots, M$, $M = m^{N-1}$, with edges of length $1/m$, where m is an integer. Define the slices:

$$A_j = \{x = (x_1, \ldots, x_N) \in [0,1]^N : \quad (x_1, \ldots, x_{N-1}) \in Q_j\}.$$

Let $p(x_1, \ldots, x_{N-1})$ be a polynomial of order $\ell = \lfloor \gamma \rfloor$. The set of polynomials $\mathcal{P}_{\ell,j} = \{p : h \leq p(x_1, \ldots, x_{N-1}) \leq 1 - h,\ (x_1, \ldots, x_{N-1}) \in Q_j\}$ is compact for every j. Denote by $\mathcal{P}_{m,\ell,j}$ a finite subset of $\mathcal{P}_{\ell,j}$ which consists of polynomials with discretized coefficients, the discretization step being $m^{-\gamma}$ for every coefficient. Define the "set under p":

$$B_j(p) = \{x \in A_j : x_N \leq p(x_1, \ldots, x_{N-1})\}.$$

The estimator $\hat{g}$ of the function g is defined as piecewise-polynomial function with respect to the partition $\{Q_j\}$. First consider Model 1. Then $\hat{g}$ in each Q_j is defined as a polynomial which is a solution of the minimization problem

$$\min_{p\in\mathcal{P}_{m,\ell,j}} \sum_{i=1}^{n}(Y_i' - I\{X_i \in B_j(p)\})^2, \tag{4.1}$$

where $Y_i' = I\{Y_i \geq a_1\}$ and a_1 is some number between a and 0 (it is assumed that either a or a_1 are known).

For Models 2,3 the estimator $\hat{g}$ in Q_j is defined as a polynomial solving the problem

$$\min_{\substack{p\in\mathcal{P}_{m,\ell,j}:\\ B_j(p)\supseteq\{X_i:X_i\in A_j\}}} \text{mes}\,(B_j(p)) \tag{4.2}$$

where mes (B_j) is the Lebesgue measure of B_j.

The estimator G_n^* of the set G is then defined as the closure of the set

$$\{(x_1,\ldots,x_N)\in[0,1]^N : 0\leq x_N\leq \hat{g}(x_1,\ldots,x_{N-1})\}.$$

THEOREM 1. *For Models 1,2 and 3 we have*

$$\sup_{G\in\mathcal{G}_{0,\gamma}} E(d^q(G_n^*,G)) = \mathcal{O}(\psi_n^q), \quad \forall\,\gamma,q>0, \tag{4.3}$$

as $n\to\infty$, where $d=d_1$ or $d=d_\infty$, and

(i) $\psi_n = n^{-\gamma/(\gamma+N-1)}$ if $d=d_1$ and G_n^* is the estimator defined so that in (4.1), (4.2) $m=[n^{1/(\gamma+N-1)}]$.

(ii) $\psi_n = (n/\log n)^{-\gamma/(\gamma+N-1)}$ if $d=d_\infty$ and G_n^* is the estimator defined so that in (4.1), (4.2) $m=[(n/\log n)^{1/(\gamma+N-1)}]$.

This theorem generalizes other known results on estimation of boundary fragments. For Model 1 under various types of assumptions the result of Theorem 1 was shown by Korostelev (1991), Tsybakov (1991), KT (1991, 1992a,b,c, 1993). For Model 2 in case $0<\gamma\leq 1$, $N=2$ a result close to Theorem 1 (i) was shown by Geffroy (1964). For Model 2 with uniform density $\mu(x) = I\{x\in G\}/(\text{mes } G)$ Theorem 1 is proved in KT (1992c, 1993). In general, the proof techniques for Model 1 carry over to Models 2 and 3. For Model 3 the results on convergence rates seem not to be available in the literature.

5. The Estimators for General Classes of Sets and Their Convergence Rates. For general classes of sets we apply the second approach, i.e. the maximum likelihood estimation on ε-nets. Let $\mathcal{G}$ be a compact class of subsets of $[0,1]^N$, and denote by $\mathcal{N}_\varepsilon$ a minimal ε-net on $\mathcal{G}$ with respect to the metric d_1.

Let $p_G(Z_n)$ be the joint density of observations $Z_n = \{X_1, \ldots, X_n,\ Y_1, \ldots, Y_n\}$ in Model 1, or $Z_n = \{X_1, \ldots, X_n\}$ in Model 2, or $Z_n = \{\nu_n, X_1, \ldots, X_{\nu_n}\}$ in Model 3 when the underlying set is G. Define the maximum likelihood estimator on $\mathcal{N}_\varepsilon$:

$$G_n^{\mathrm{MLE}} = \arg\max_{G \in \mathcal{N}_\varepsilon} p_G(Z_n).$$

For Model 1 we assume in this section that $f(x)$ is a known function, and for Models 2, 3 we assume that $\mu(x)$ and $\lambda(x)$ are uniform densities on G. This, ensures that G_n^{MLE} can be calculated, given the data.

Denote

$$\psi_n(c) = \inf\{\varepsilon > 0 : \mathcal{H}(\varepsilon, \mathcal{G}, d_1) \le cn\varepsilon\},$$

$$\mathcal{H}(\varepsilon, \mathcal{G}, d_1) = \log(\mathrm{card}\, \mathcal{N}_\varepsilon).$$

THEOREM 2. *Assume that either*

(i) Model 1 holds, where ξ_i are i.i.d. random variables with density p, and there exist $\Delta_1, \Delta_2 > 0$ such that

$$\int \sqrt{p(y)p(y-u)}\,dy \le 1 - \Delta_1 \quad \text{for} \quad |u| \ge a, \tag{5.1}$$

$$\int p(y)\sqrt{\frac{p(y)}{p(y+u)}}\,dy \le \Delta_2 \quad \text{for} \quad |u| \le 1. \tag{5.2}$$

or

(ii) Model 2 or Model 3 holds, with $\mu(x) \equiv (\mathrm{mes}\,(G))^{-1}$, $x \in G$, or $\lambda(x) \equiv \lambda = \mathrm{const}$, $x \in G$, respectively, and $\inf_{G \in \mathcal{G}} \mathrm{mes}\,(G) \ge m_0 > 0$.

Let

$$\liminf_n \psi_n(c) n / (\log n) > 0, \quad \forall\, c > 0. \tag{5.3}$$

Then for every $q > 0$ there exists a constant $c_q > 0$ such that

$$\sup_{G \in \mathcal{G}} E(d_1^q(G_n^{\mathrm{MLE}}, G)) = \mathcal{O}(\psi_n^q), \quad \text{as} \quad n \to \infty, \tag{5.4}$$

where $\psi_n = \psi_n(c_q)$, and G_n^{MLE} is MLE on $\mathcal{N}_{\psi_n}$.

REMARKS.

1) Since $\mathcal{H}(\varepsilon, \mathcal{G}, d_1)$ is monotone nonincreasing in ε, and positive for ε small enough, the value $\psi_n(c)$ is well-defined.

2) For the case (ii) of Theorem 2 (Models 2,3) the estimators have a simple form
$$G_n^{\mathrm{MLE}} = \mathop{\arg\min}_{\substack{G\in\mathcal{N}_\varepsilon:\\ G\supset\{X_1,\dots,X_n\}}} \ \mathrm{mes}\,(G).$$

3) Condition (5.3) was introduced only to guarantee the convergence of moments (5.4). If (5.3) is omitted we get, however, the convergence in probability:
$$\lim_{t\to\infty}\limsup_{n\to\infty}\sup_{G\in\mathcal{G}} P\{d_1(G_n^{\mathrm{MLE}},G)\ge\psi_n t\}=0. \tag{5.5}$$

Proof of Theorem 2. Consider Model 1 first, and apply Lemma A.1 from MT(1992). It suffices to verify the condition (A3) of this lemma, with $d=d_1$, $\alpha=1$, which, in turn, is implied by
$$\sup_{\substack{G',\tilde G: d_1(G,\tilde G)\le\varepsilon\\ d_1(G,G')\ge\varepsilon t}} \int p_G(Z_n)\sqrt{\frac{p_{G'}(Z_n)}{p_{\tilde G}(Z_n)}}dZ_n \le \exp(-\bar c n\varepsilon t), \tag{5.6}$$

for $t>0$ large and some $\bar c>0$. Let (2.2) be satisfied. To prove (5.6) note that the integral in (5.6) is the n-product of the integrals
$$J=\int p_G(y|x)\sqrt{\frac{p_{G'}(y|x)}{p_{\tilde G}(y|x)}}\mu(x)dxdy$$

where $p_G(y|x)=p_1(y|x)I\{x\in G\}+p_2(y|x)I\{x\notin G\}$, $p_1(y\mid x)=p(y-f(x))$, $p_2(y|x)=p(y)$. Direct calculations show that
$$\begin{aligned}J=\ &\mathrm{mes}\,(G\cup\tilde G\cup G')+\left(\int p_1\sqrt{\frac{p_1}{p_2}}\mu-1\right)\mathrm{mes}\,\left((G\cap G')\setminus\tilde G\right)\\ &+\left(\int p_2\sqrt{\frac{p_2}{p_1}}\mu-1\right)\mathrm{mes}\,\left(\tilde G\setminus(G\cup G')\right)\\ &+\left(\int\sqrt{p_1p_2}\mu-1\right)\left[\mathrm{mes}\,\left((G\cap\tilde G)\setminus G'\right)+\mathrm{mes}\,\left(G'\setminus(G\cup\tilde G)\right)\right].\end{aligned}$$

We have also $\mathrm{mes}\,\left((G\cap G')\setminus\tilde G\right), \mathrm{mes}\,\left(\tilde G\setminus(G\cup G')\right)\le d_1(G,\tilde G)\le\varepsilon$, $\left|\mathrm{mes}\,\left((G\cap\tilde G)\setminus G'\right)-\mathrm{mes}\,(G\setminus G')\right|\le d_1(G,\tilde G)\le\varepsilon$ $\left|\mathrm{mes}\,\left(G'\setminus(G\cup\tilde G)\right)-\mathrm{mes}\,(G'\setminus G)\right|\le d_1(G,\tilde G)\le\varepsilon$. Using these inequalities, (5.1), (5.2), and the fact that $a\le f(x)\le 1$, we get
$$J\le 1+2\Delta_2\varepsilon-\Delta_1(d_1(G,G')-2\varepsilon)\le 1-\varepsilon(t-2\Delta_1-2\Delta_2).$$

This entails that J^n does not exceed the RHS of (5.6) for $t > 0$ large, and proves the theorem for Model 1 under (2.2). The case (2.1) of Model 1 is treated similarly. For Model 2 and $\mathcal{G} = \mathcal{G}_\gamma$ the theorem is proved in Theorem 4.1 of MT(1992), and its extension to general $\mathcal{G}$ is straightforward in view of Lemma A.1 of MT (1992). The proof for Model 3 follows the same lines as for Model 2 since $X_1, \ldots, X_{\nu_n}$ for fixed ν_n are uniformly distributed in G.

Examples. 1). Let $\mathcal{G} = \mathcal{G}^\theta$ be a parametric class of subsets of $[0,1]^N$. Using Theorem 2 and (3.1) we have that the rate of convergence of G_n^{MLE} is $\psi_n = \mathcal{O}((\log n)/n)$. The same conclusion holds for G_n^{MLE} on Vapnik-Červonenkis classes $\mathcal{G}$.

2). For finite classes of sets one gets the rate $\psi_n = \mathcal{O}(1/n)$ if M is fixed. This ψ_n does not satisfy (5.3), so we can claim only (5.5), not (5.4). If $M = M_n \to \infty$ as $n \to \infty$, then $\psi_n = \mathcal{O}((\log M_n)/n)$. In particular, for polynomially increasing M_n one gets the rate $\psi_n = \mathcal{O}((\log n)/n)$, and for the case $M_n = (\sqrt{n})^{\sqrt{n}}$ considered by Carlstein and Krishnamoorthy (1992) one gets $\psi_n = \mathcal{O}\left((\log n)/\sqrt{n}\right)$ which gives an improvement of their result.

3). For Dudley's classes $\mathcal{G}_\gamma$ in view of (3.2) one gets $\psi_n = \mathcal{O}\left(n^{-\frac{\gamma}{\gamma+N-1}}\right)$, (cf. MT (1992)), which coincides with the rate of convergence for boundary fragments as given in Theorem 1.

4). For the class of monotone fragments $\mathcal{G}_{\mathrm{mon}}$ and for the class of convex sets $\mathcal{G}_{\mathrm{conv}}$ we use (3.3) and (3.4), which leads to $\psi_n = \mathcal{O}(n^{-1/N})$, and $\psi_n = \mathcal{O}(n^{-2/(N+1)})$ respectively (see also MT (1992), KT (1993)).

6. Optimality of the Estimators. The proposed estimators have optimal rates of convergence in the examples considered above. This means that

$$\liminf_{n\to\infty} \inf_{\hat{G}_n} \sup_{G\in\mathcal{G}} E(d^q(\hat{G}_n, G))\psi_n^{-q} > 0, \quad \forall\, q > 0, \tag{6.1}$$

where $\inf_{\hat{G}_n}$ denotes the infinum over all estimators, $\psi_n = \psi_{n,\mathcal{G}}$ is the same sequence as in the appropriate upper bounds (4.3) or (5.4), and $d = d_1$ or d_∞.The sequence $\psi_{n,\mathcal{G}}$ depends on the class $\mathcal{G}$ and on the metric d. In particular, the rates ψ_n given in Theorem 1 are optimal for $\mathcal{G}_{0,\gamma}$ with $d = d_1$ and $d = d_\infty$ respectively, and the rates given in Examples 3), 4) of Section 5 are optimal for $\mathcal{G} = \mathcal{G}_\gamma, \mathcal{G}_{\mathrm{mon}}$ and $\mathcal{G}_{\mathrm{conv}}$, with $d = d_1$. For Examples 1), 2) of Section 5 one can use the trivial lower bound (6.1) with $\psi_n = 1/n$. This gives the optimality in parametric case ($\mathcal{G} = \mathcal{G}^\theta$) up to the log-factor only. However, the log-factor in the upper bound (5.4) for $\mathcal{G} = \mathcal{G}^\theta$ can be eliminated by modifying the proof of Lemma A.1 in MT (1992) (use the chaining argument there). For the conditions guaranteeing (6.1) in Model 1 we refer to KT (1991, 1992a, 1993). For Model 2 (6.1) holds without additional conditions (see KT (1992c, 1993), MT (1992)). Note that it is sufficient to prove (6.1) for the simplest case of uniform density μ (or constant intensity λ). The proof of (6.1) for Model 3 follows that for Model 2, and also does not require additional conditions.

7. Concluding Remarks.

1. For Model 1 we did not consider the case where $X_1, \ldots, X_n$ are on the *regular grid*. Mathematically, a rather full treatment of this case can be found in Tsybakov (1989). However, the regular grid design has the following defect: it rules out the possibility of estimating with optimal rates if the boundary is smooth enough, namely if $\gamma > 1$ (KT (1991, 1992a, 1993)). Also the assumption (2.1) on X_i seems to be more natural from the statistical viewpoint, since in applications one is never sure that the measured image values correspond exactly to the centers of pixels Δ_i, but rather to some (possibly random) points within the pixels.

2. The results of this paper are valid for the modifications of Models 1-3 with *non-zero background*. For example, in Model 1 this modification is:

$$Y_i = f_1(X_i)I\{X_i \in G\} + f_2(X_i)I\{X_i \notin G\} + \xi_i,$$

where $f_1(x) \geq a > 0$, and $f_2(x) < a$ everywhere on $[0,1]^N$ (see Tsybakov (1989, 1991), KT (1991, 1992a, 1993)). Similarly, one can modify Models 2 and 3.

3. Another extension is related to models with varying jump size. There are two interesting options here. First, the jump between the object and the background tends to 0 with some rate a_n (e.g. for Model 1 with $f_1(x) \equiv$ const , $f_2(x) = f_1 - a_n$), so that the change-point character of the model is asymptotically eliminated. Then one can show that the convergence of $G_n^*, G_n^{\mathrm{MLE}}$ to G still occurs if $na_n^2 \to \infty$ as $n \to \infty$. The rate of convergence differs from those in the case of constant step in that n should be replaced by na_n^2 (e.g. for $\mathcal{G}_\gamma, \mathcal{G}_{0,\gamma}$ the rate is $\psi_n = (na_n^2)^{-\gamma/(\gamma+N-1)}$, see MT (1992), and for "almost parametric" classes $\mathcal{G}$ the rate is $(na_n^2)^{-1}$, see Müller and Song (1992b)). The second option is related to contamined data, with $f_2 = a_n \to 0$, $f_1 \geq A > 0$. It can be shown that the contamination does not affect the rate of convergence.

4. Carlstein and Krishnamoorthy (1992), Rudemo, Skovgaard and Stryhn (1990) consider the model which is more general than Model 1 with constant f in the sense that the distribution of Y_i for fixed X_i is not necessarily of shift type. They assume that Y_i are independent for fixed $X_1, \ldots, X_n$ and $p_G(y|x) = p_1(y)I\{x \in G\} + p_2(y)I\{x \notin G\}$ where $p_1(y)$ and $p_2(y)$ are some

different densities. If we call the Model 1 with this modification Model 4 then the following result is obtained exactly as Theorem 2:

THEOREM 3. *Assume that Model 4 holds, and*

$$\int \sqrt{p_1(y)p_2(y)}dy < 1, \quad \int (p_1(y)\sqrt{\frac{p_1(y)}{p_2(y)}} + p_2(y)\sqrt{\frac{p_2(y)}{p_1(y)}})dy < \infty. \quad (7.1)$$

Then under (5.3) for every $q > 0$ there exists such $c_q > 0$ that (5.4) holds.

Thus, the estimator G_n^{MLE} has optimal rates of convergence also in this general situation, in contrast to the estimator of Carlstein and Krishnamoorthy (1992) which has slower rate since its is based on $\sqrt{n}$-convergent rather than n-convergent statistics. However, the estimator of Carlstein and Krishnamoorthy (1992) has an important advantage of being distribution-free.

5. Some other examples of possible sets G may be considered. Note the two of them: star-shaped sets (see Rudemo, Skovgaard and Stryhn (1990)), and the sets G such that both G and $[0,1]^N \setminus G$ satisfy the "cone condition" which is equivalent to piecewise Lipschitz boundary assumption (Tsybakov (1989, 1991), KT (1991, 1992a, 1993), M ller and Song (1992a)). The star-shaped sets can be reduced to boundary fragments, and the theory for fragments applied with some modifications. For the cone-condition sets one can construct very simple estimators converging with the rate $\psi_n = \mathcal{O}(\sqrt{(\log n)/n})$ (see e.g. KT (1991, 1992a, 1993)).

REFERENCES

CARLSTEIN, E. and KRISHNAMOORTHY, C. (1992). Boundary estimation, *J. Amer. Statist. Assn.* **87**, 430–438.

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

DEPRINS, D., SIMAR, L. and TULKENS, H. (1984). Measuring labor efficiency in post offices. In: *The Performance of Public Enterprises: Concepts and Measurements*, M. Marchand, P. Pestieau and H. Tulkens, eds. Amsterdam, North-Holland, 243–267.

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.

GEFFROY, J. (1964). Sur un problème d'estimation géométrique, *Publications de l'Institut de Statistique des Universités de Paris*, **13** 191–210.

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

JACOB, P. and ABBAR, H. (1989). Estimating the edge of Cox process area, *Cahiers Centre Études Rech. Opér.* **31**, n. 3–4.

KOROSTELEV, A. P. (1991). Minimax reconstruction of two-dimensional images, *Theory Probab. Appl.* **36**, 153–159.

KOROSTELEV, A. P. and TSYBAKOV, A. B. (1991). Asymptotically minimax image reconstruction problems. Preprint No 612, SFB 123, Universit.

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

KOROSTELEV, A. P. and TSYBAKOV, A. B. (1992b). Minimax linewise algorithm for image reconstruction, CORE Discussion Paper No 9249, (To appear in *Computer Intensive Methods in Statistics*, W. Härdle and L. Simar, eds. Physica-Verlag, 1993).

KOROSTELEV, A. P. and TSYBAKOV, A. B. (1992c). Estimation of support of a probability density and estimation of support functionals, CORE Discussion Paper No 9229, (To appear in *Problems of Information Transmission*, 1993, **29**.

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

MAMMEN, E. and TSYBAKOV, A. B. (1992). Asymptotical minimax results in image analysis for sets with smooth boundaries, Institute of Statistics Discussion Paper No 9205, Université Catholique de Louvain.

MÜLLER, H.-G. and SONG, K.-S. (1992a). On the estimation of multidimensional boundaries. Unpublished manuscript.

MÜLLER, H.-G. and SONG, K.-S. (1992b). On multidimensional change-points in regression. Unpublished manuscript.

QIU, P. (1992). Nonparametric estimation of jump surfaces. Unpublished manuscript.

RÉNYI, A. and SULANKE, R. (1964). Über die konvexe Hülle von n zufällig gewählten Punkten II, *Z. Wahr. 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. of Math. and Phys., Royal Veterinary and Agricultural Univ., Frederiksberg, Denmark.

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

TSYBAKOV, A. B. (1989). Optimal estimation accuracy of nonsmooth images, *Problems of 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 DE STATISTIQUE
UNIVERSITÉ CATHOLIQUE DE LOUVAIN
34, VOIE DU ROMAN PAYS
B-1348 LOUVAIN-LA-NEUVE, BELGIUM