Volume 15, No. 2 (2006), pp. 224–231 Allerton Press, Inc.

MATHEMATICAL METHODS OF STATISTICS

THE CHANGE-SET PROBLEM FOR VAPNIK–ČERVONENKIS CLASSES

E. Khmaladze[1,3], R. Mnatsakanov[2,3], and N. Toronjadze[4]

[1]School of Mathematics, Statistics and Computer Science, Victoria Univ. Wellington
P.O. Box 600, Wellington, New Zealand
E-mail: Estate.Khmaladze@mcs.vuw.ac.nz

[2]Dept. Statistics, West Viginia Univ.
P.O. Box 6330, Morgantown, WV 26506, USA
E-mail: rmnatsak@stat.wvu.edu

[3]Dept. Probab. Theory and Math. Statist., A. Razmadze Mathematical Institute
1 M. Alexidze St., 0193 Tbilisi, Georgia

[4]Financial Monitoring Service of Georgia, NBG
3/5 Leonidze St., 0105, Tbilisi, Georgia

We consider the following change-set problem in $\mathbb{R}^d$: for a pair (X, Y) of "location" $X \in \mathbb{R}^d$ and "mark" Y, suppose that $\mathbf{P}\{Y \in \cdot \mid X\} = P_1\{\cdot\}$ if X is outside certain region $G \subset \mathbb{R}^d$ and $\mathbf{P}\{Y \in \cdot \mid X\} = P_2\{\cdot\}$ if $X \in G$. This region, which we call a change-set and which could also be called an "image", is the parameter of interest.

Suppose $\mathcal{C}$ is a sub-class of Borel subsets of $\mathbb{R}^d$, totally bounded with respect to the pseudo-metric $d(G, G') = F(G \triangle G')$ induced by the distribution F of X_i's and let $\widehat{G}$ be a Maximum estimator (M-estimator) of G based on n i.i.d. pairs $(X_i, Y_i)_{i=1}^n$ under assumption that $G \in \mathcal{C}$.

In this paper $\mathcal{C}$ is a Vapnik–Červonenkis class ($V\check{C}$-class), or, in particular, a class indexed by a finite-dimensional parameter. We show that the rate of convergence of $d(\widehat{G}, G)$ for these classes is $1/n$ or $\log n/n$ and describe these cases in terms of the local covering numbers.

Key words: change-point problem, local covering numbers, $V\check{C}$-classes, M-estimators for sets.

2000 Mathematics Subject Classification: Primary 62G05, 62G20; secondary 62H35.

1. Introduction

In this paper we study the rate of convergence in the change-set problem when the possible change-set G belongs, a priori, to a $V\check{C}$-class or, in particular, a class indexed by a finite-dimensional parameter.

More specifically, following the version of this problem given, e.g., in Mammen and Tsybakov [1], Section 3, we consider an i.i.d. sequence (X_i, Y_i), $i = 1, \ldots, n$, of "locations" X_i and corresponding "marks" Y_i and assume that the locations are all identically distributed random variables in $\mathbb{R}^d$ with common distribution F, while the marks are random variables taking values in some measurable space $\{\mathbb{E}, \mathcal{E}\}$. Further, suppose on $\{\mathbb{E}, \mathcal{E}\}$ there are two distributions P_1 and P_2 and there is a Borel subset G in $\mathbb{R}^d$ such that the conditional distribution of Y_i given X_i is

$$\mathbf{P}\{\cdot \mid X_i\} = P_1\{\cdot\} I_{\{X_i \notin G\}} + P_2\{\cdot\} I_{\{X_i \in G\}} = (P_1\{\cdot\})^{I_{\{X_i \notin G\}}} (P_2\{\cdot\})^{I_{\{X_i \in G\}}}.$$

The set G is the parameter of interest in the change-set problem and we denote $\widehat{G}_n$ its estimator based on the sequence (X_i, Y_i), $i = 1, \ldots, n$.

The simplest setting of this problem is that of known distributions P_1 and P_2. However, whether or not this is true in any given circumstances is not really important for our purposes. Neither is it important what the nature of the space $\{\mathbb{E}, \mathcal{E}\}$ is. What we need for the estimation of G (see (1.1) below) is only the existence of some bounded "score function" ξ of a mark, such that

$$\alpha_2 := \int \xi(y)\, dP_2(y) > 0 > \int \xi(y)\, dP_1(y) := \alpha_1.$$

We equip the class of Borel subsets of $\mathbb{R}^d$ with the pseudo-metric $d(G, G') = F(G \triangle G')$ and suppose that $\mathcal{C}$ is a totally bounded sub-class of these subsets with respect to $d(\cdot, \cdot)$. Denote $\mathcal{N}_\delta$ the (finite) minimal δ-net of $\mathcal{C}$. Its cardinality, $N_\delta = \operatorname{card} \mathcal{N}_\delta$, is called the covering number of $\mathcal{C}$. Furthermore, for each $G \in \mathcal{C}$ let $\mathcal{O}(t, G)$ be the neighborhood of G of radius t and let

$$N_\delta(t, G) = \operatorname{card} \mathcal{N}_\delta \cap \mathcal{O}(t, G).$$

This function of t and δ (and G) is called the local covering number. As was pointed out in Khmaladze *et al.* [2], the local covering number in connection with the change-set problem was introduced and studied in Khmaladze *et al.* [3]. However, this concept appeared much earlier. One can see comments on this, e.g., in the Introduction of [3]. A comprehensive information is available in the recent paper of Birgé [4].

With chosen score function ξ consider the process in G' defined by

$$L_n(G', G) := \sum_1^n \big[I_{\{X_i \in G'\}} - I_{\{X_i \in G\}} \big] \xi(Y_i).$$

As the M-estimator $\widehat{G}_n$ of G we consider

$$\widehat{G}_n = \widehat{G}_n(\delta) := \arg \max_{G' \in \mathcal{N}_\delta} L_n(G', G), \tag{1.1}$$

with $\delta = \delta_n \to 0$ as $n \to \infty$. The rate of convergence of $d(\widehat{G}_n, G)$ to 0 when it is known that G belongs to a $V\check{C}$-class $\mathcal{C}$ is, as we said, the subject of this paper.

The classes of this type often occurred in the change-point problems on the line and in space. For example, see Brodsky and Darkhovsky [5], Chapter 2, for the problems on the real line with multiple change-points. For the rate $1/n$ for one change-point see, e.g., Dümbgen [6], Dempfle and Stute [7], and also Hariz and Wylie [8] for the case of long-range dependent data. Change-sets of the form of half-spaces in $\mathbb{R}^d$ were considered, e.g., in Pollard [9]. Notwithstanding previous work on the rate of convergence in such problems, we will see, however, that it is possible to give relatively elementary proof in relatively broad conditions on the local covering number of these classes. We will also see that the rate of convergence is often, but not always, $1/n$ and that it can also be $\log n/n$ due to different behavior of the local covering number.

In Section 2 we formulate the condition on $N_\delta(t, G)$ and the main theorem, which gives an upper bound for the probability $\mathbf{P}\{nd(\widehat{G}, G) \geq x\}$. In Section 3 we consider some examples.

We conclude the Introduction with a remark. If the class $\mathcal{C}$ is indexed by a finite-dimensional parameter, as is often the case, then our problem should be equivalent to the estimation problem of this parameter. If so, from the theory of estimation in finite-dimensional spaces, see, e.g., Ibragimov and Khas'minskii [10], for regular parametric families one should expect the rate $1/\sqrt{n}$, while higher rates can only be obtained in non-regular families and the corresponding estimators are usually called super-efficient estimators. For example, for the typically non-regular family of uniform distributions on the interval $[0, \theta]$, $\theta > 0$, the rate of convergence of the estimator $\hat{\theta}$ is indeed $1/n$. The reason why the super-efficient estimation is possible in our change-set problem can be intuitively explained as follows: although the conditional distribution of marks $\mathbf{P}\{\cdot \mid X_i\}$ in many statistical problems, like regression, may be a continuous function of X_i, for the change-set problem it is essentially a discontinuos function — actually, it is a step-function with two values in the set $\mathcal{P}$ of distributions on $\{\mathbb{E}, \mathcal{E}\}$. It would be possible and interesting to consider similarly a function taking several values in $\mathcal{P}$. This would correspond to several "changes" or several "levels of change" on disjoint "change-sets." Still, in these cases the statistical problem will remain non-regular.

2. The Main Theorem

Consider a $V\check{C}$-class $\mathcal{C}$ of subsets of $\mathbb{R}^d$. The definition of the $V\check{C}$-classes is of combinatorial nature (they "shatter" at least some collection of v points in $\mathbb{R}^d$; for the precise definition see, e.g., Pollard [9]). It is well known that this definition implies that the covering number of a $V\check{C}$-class in the pseudo-metric $d(G, G')$ satisfies, uniformly in F, the inequality

$$N_\delta \leq K(4e)^v \left(\frac{1}{\delta}\right)^{v-1}, \tag{2.1}$$

where K is a universal constant and v is the index of the $V\check{C}$-class. In the present section it is this last property which is of interest and importance and not the combinatorial property *per se*. So, we will consider here classes with the covering number satisfying inequality (2.1). In Khmaladze *et al.* [2], Section 2, the quasi-optimal resolution level is defined as the value δ_n which minimizes the product $N_\delta e^{nc_1\delta}$, where the constant c_1 depends on the first two moments of the score function ξ. If we have only an upper bound for N_δ, we can use it instead. Although formally this entails some loss of accuracy, the inequalities below prove quite high rate of convergence anyway. Using the bound (2.1) for N_δ one can easily find that the optimal δ_n is

$$\delta_n = \frac{v-1}{nc_1}. \tag{2.2}$$

Define the sequence z_n through the equation $N_{\delta_n} e^{nc_1\delta_n} e^{-nc_1 z} = 1$. Again using the upper bound (2.1) we obtain

$$z_n = \frac{v-1}{nc_1} \log \frac{nc_1}{v-1} + \frac{1}{nc_1} K_0 \tag{2.3}$$

with $K_0 = \log K(4e)^v + v - 1$. One can think intuitively of z_n as the upper bound in probability for $d(\widehat{G}_n, G)$ if one does not use local covering numbers. (The most recent paper we know about, where const $\cdot z_n$ is used as an a.s. bound, is Ferger [11], see inequality (2.9) there. We repeat this result in Remark 2.1 with const = 1.)

Let us assume now that the local covering number of the class $\mathcal{C}$ satisfies the following inequality: for $t \leq 2z_n$

$$N_\delta(t) \leq c_2 \frac{t^m}{\delta^{m_1}}, \qquad 1 \leq m \leq m_1 \leq v-1. \tag{2.4}$$

The case of $m = m_1 = v-1$ is certainly most straightforward and describes "usual" cases when the class $\mathcal{C}$ is equally dense at any G. This is the case considered in Example 1 in Section 3 below. The case, where $m = m_1 < v-1$, corresponds to the situation when the class is "poorer" or "narrower" than average around given G, which is illustrated by Example 2. The cases when $m < m_1$ are, in our opinion, very instructive and are illustrated by Example 3. It is only in these cases that $d(\widehat{G}_n, G)$ becomes of order $\log n/n$.

Now we can formulate the statement concerning the rate of $d(\widehat{G}_n, G_0)$ in probability.

Theorem 2.1. *Under the assumption* (2.4) *we have for* $\varepsilon_n > 3\delta_n/2$

$$\mathbf{P}\left\{d(\widehat{G}_n, G_0) > \varepsilon_n\right\} \leq c_2 \frac{e^{v-1}}{(v-1)^{m_1}} (nc_1)^{m_1-m} \int_{nc_1\varepsilon_n}^{\infty} e^{-\tau}\tau^m \, d\tau + \frac{1}{N_{\delta_n}}. \tag{2.5}$$

Corollary 2.1. *Suppose* $\varepsilon_n > 3\delta_n/2$ *and condition* (2.4) *is satisfied.*

(i) *If* $m = m_1$, *then*

$$\mathbf{P}\left\{d(\widehat{G}_n, G_0) > \frac{L}{nc_1}\right\} \leq c_2 \frac{e^{v-1} m!}{(v-1)^m} \left\{1 - F_{m+1}(L)\right\}, \tag{2.6}$$

where F_m *is Gamma distribution function with the shape parameter* m.

(ii) *If* $m < m_1$, *then*

$$\mathbf{P}\left\{d(\widehat{G}_n, G_0) > \varepsilon_n(p)\right\} \leq p \tag{2.7}$$

for

$$\varepsilon_n(p) = \frac{1}{nc_1}\big\{(m_1 - m)\log nc_1 + m\log\log nc_1 + L + o(1)\big\}, \tag{2.8}$$

where the constant $L = L(p)$ *satisfies the equation*

$$c_2 \frac{e^{v-1}}{(v-1)^{m_1}}(m_1 - m)^m e^{-L} = p.$$

Remark 2.1. Clearly, from (2.5) it follows that

$$\mathbf{P}\big\{d(\widehat{G}_n, G_0) \geq z_n\big\} \to 0, \qquad n \to \infty.$$

However, one can say more:

$$\mathbf{P}\big\{\limsup_{n\to\infty}[d(\widehat{G}_n, G_0)/z_n] > 1\big\} = 0.$$

Remark 2.2. Although for $m < m_1$ the upper bound z_n is of the same order of magnitude as $\varepsilon_n(1)$, the difference is still essential: whereas $d(\widehat{G}_n, G_0)$ exceeds z_n only with probability tending to 0, it may exceed $\varepsilon_n(1) + L/nc_1$ with sufficiently large probability.

Proof of Theorem 2.1. By Theorem 2.2 of Khmaladze *et al.* [2] we have the inequality: if $\varepsilon_n > \frac{3}{2}\delta_n > 0$, then

$$\mathbf{P}\big\{d(\widehat{G}_n, G_0) > \varepsilon_n\big\} \leq \int_{\varepsilon_n}^1 e^{-nc_1(t-\delta_n)} N_{\delta_n}(dt), \tag{2.9}$$

for any behavior of the covering number $N_{\delta_n}(t)$. Integration by parts in the right-hand side of (2.9) yields

$$\big[N_{\delta_n} - N_{\delta_n}(\varepsilon)\big]e^{-nc_1(1-\delta_n)} + nc_1\int_{\varepsilon_n}^1 e^{-nc_1(t-\delta_n)}\big[N_{\delta_n}(t) - N_{\delta_n}(\varepsilon_n)\big]\,dt.$$

Here the first summand is of order $O((nc_1)^{v-1}e^{-nc_1})$, while using the definition of z_n one finds that

$$\begin{aligned} nc_1\int_{2z_n}^1 & e^{-nc_1(t-\delta_n)}\big[N_{\delta_n}(t) - N_{\delta_n}(\varepsilon_n)\big]\,dt \leq N_{\delta_n}e^{-nc_1(2z_n-\delta_n)} \\ & \leq e^{-nc_1 z_n} = (\frac{v-1}{nc_1})^{v-1}e^{-(v-1)-\log(K(4e)^v)} \leq \frac{1}{N_{\delta_n}}. \end{aligned}$$

At the same time, using condition (2.4) one obtains

$$\begin{aligned} nc_1 & \int_{\varepsilon_n}^{2z_n} e^{-nc_1(t-\delta_n)}\big[N_{\delta_n}(t) - N_{\delta_n}(\varepsilon_n)\big]\,dt \\ & \leq c_2(nc_1)^{m_1-m}\frac{e^{v-1}}{(v-1)^{m_1}}\int_{nc_1\varepsilon_n}^{\infty} e^{-\tau}\tau^m\,d\tau. \end{aligned}$$

The last two inequalities combined prove (2.5). □

Proof of Corollary 2.1. (i) For $m = m_1$ the leading term on the right-hand side of (2.5) becomes

$$c_2 \frac{e^{v-1}\, m!}{(v-1)^m} \big\{1 - F_{m+1}(nc_1\varepsilon_n)\big\}.$$

Then the choice $\varepsilon_n = L/nc_1 \geq (3/2)\delta_n$ gives inequality (2.6).

(ii) If $m < m_1$, we have $\tau_n = nc_1\varepsilon_n \to \infty$. Then

$$1 - F_{m+1}(\tau_n) \sim \frac{1}{m!}\tau_n^m e^{-\tau_n}$$

and the solution of the equation

$$c_2 \frac{e^{v-1}}{(v-1)^{m_1}} \tau_n^m e^{-\tau_n} = \frac{p}{(nc_1)^{m_1-m}}$$

has asymptotic representation

$$\tau_n = (m_1 - m)\log nc_1 + m \log\log nc_1 + L + o(1),$$

which yields (2.7). □

3. Some Examples

Example 1. Let $X_i \in \mathbb{R}^d$ and for $\theta \in \mathbb{R}^d$ let $G = [\theta, \infty) = \{x\colon x \geq \ \theta\}$. Let us take $\mathcal{C} = \{[\theta, \infty),\, \theta \in \mathbb{R}^d\}$ as the collection of possible change-sets (cf. Brodsky and Darkhovsky [5], Chapter 6). The covering number of this $\mathcal{C}$ in the pseudo-metric $d(G, G') = F(G \triangle G')$ is never larger than const $\cdot\delta^{-d}$, though for specific F it can be essentially smaller. To insure that δ^{-d} is the right order, suppose F is absolutely continuous with positive density. Then there is no loss of generality in reducing the rectangles above to $[\theta, \mathbf{1}]$, $\mathbf{1} = (1, \ldots, 1)$, $\theta \in [0, 1]^d$, and assuming that F has uniform marginal distributions. Let us assume, mostly for notational convenience, that F is simply uniform on $[0, 1]^d$ and put $d = 2$. Then, since

$$d\big([\theta, \mathbf{1}], [\theta', \mathbf{1}]\big) \leq |\theta_1 - \theta_1'| + |\theta_2 - \theta_2'|$$

and

$$d\big([\theta, \mathbf{1}], [\theta', \mathbf{1}]\big) \geq |\theta_1 - \theta_1'| \max(\theta_2, \theta_2') + |\theta_2 - \theta_2'| \max(\theta_1, \theta_1'),$$

the δ-net for the class of rectangles $\mathcal{C} = \{[\theta, \mathbf{1}],\, \theta \in [0, 1]^2\}$ is formed by the collection $\mathcal{N}_\delta = \{[z_{ij}, \mathbf{1}], z_{ij} = (i\delta - \delta/2, j\delta - \delta/2),\ i, j = 1, \ldots, 1/\delta\}$ (assume $1/\delta$ is an integer). It is then clear that $N_\delta \sim 1/\delta^2$, so that $v - 1 = 2$. It is also clear that the area of those θ', for which $d([\theta, \mathbf{1}], [\theta', \mathbf{1}]) \leq t$, is no greater than t^2. Therefore we have the following inequality for the local covering number:

$$N_\delta(t) \leq \left[\frac{t^2}{\delta^2}\right] + 1 \sim \frac{t^2}{\delta^2},$$

which is the case with $m = m_1 = v - 1 = 2$. So that from (2.6) we derive the rate $\varepsilon_n(1) = L/nc_1$ with $L = 3.789046$ (the $(1 - 2/e^2)$-quantile of F_3).

Another, and, perhaps, more natural example of the two-dimensional change-point problem with possible change-sets being half-spaces (see, e.g., Pollard [9], Chapter 10) can be obtained in a similar way.

If we consider as the class of change-sets the class $\mathcal{C}$ of rectangles $[\theta, \eta] = \{x \in [0,1]^2 : \theta \le x \le \eta\}$ we will obtain that $N_\delta \sim 1/\delta^4$ and

$$N_\delta(t) \le \left[\frac{t^4}{\delta^4}\right] + 1 \sim \frac{t^4}{\delta^4},$$

which again is the case with $m = m_1 = v - 1$. Similarly, we obtain the rate $\varepsilon_n(1) = L/nc_1$ with $L = 6.76661$.

Example 2. Let $\mathcal{C}$ be the class of rectangles $[\theta, \eta] = \{x \in [0,1]^2 : \theta \le x \le \eta\}$, which is "locally restricted" in a neighborhood of rectangles of the form $[0, \eta]$: for $r > 1$ and some "small" $\Delta > 0$

$$\mathcal{C} = \{[\theta, \eta],\ \theta, \eta \in [0,1]^2 \text{ and if } 0 \le \theta_1 \le \Delta, \text{ then } 0 \le \theta_2 \le {\theta_1}^r\}.$$

This is a simple form of restrictions which may occur in real practice. Suppose, for example, we analyse data on a marked point process (like possibly polluted sites or locations of bush fires) on a certain territory (see, e.g., Baddeley and Turner [12] for analysis of bush fires in New Brunswick, Canada) and suppose we choose a "window" of a given shape (as rectangles $[\theta, \eta]$ above) as a sufficiently good approximation of a change-set. Then the possible locations for this window can be restricted by topography of the territory.

As we mentioned in Example 1, the class $\mathcal{C}$ is a $V\check{C}$-class with index $v = 5$. Moreover, in a neighborhood of any rectangle $[\theta, \eta]$ with $\theta_1 > 2\Delta$ one obtains for the local covering number

$$N_\delta(t, [\theta, \eta]) \sim \text{const}\, \frac{t^4}{\delta^4}$$

as $\delta \to 0$. However, in a neighborhood of $[0, \eta]$ the class is thinner: for $t \le 2\delta \log \frac{1}{\delta}$

$$N_\delta(t, [0, \eta]) \le \left(\left[\frac{t}{\delta}\right] + 1\right)^3 \sim \frac{t^3}{\delta^3}.$$

Indeed, for $r > 1$ and $\theta_1 \le 2\delta \log \frac{1}{\delta}$ we have $\theta_2 \le (2\delta \log \frac{1}{\delta})^r = o(\delta)$. Therefore, in the δ-net around $[0, \eta]$ we need only rectangles $[x, y]$ with the lower vertices $x = (i\delta/4, 0)$, $i = 1, 2, \ldots, [4t/\delta]$, while the upper vertices $y = (j\delta/4, k\delta/4)$ remain unrestricted. In this case $\varepsilon_n(1) = L/nc_1$ with $L = 5.556604$. A more obvious version of the example can be obtained by letting $\theta_2 = 0$ if $\theta_1 \le \Delta$.

Example 3. Let $\mathcal{C}$ be the class of the unions of k rectangles (which is, of course, the same as the unions of no more than k rectangles),

$$\mathcal{C} = \Big\{ \bigcup_{i=1}^{k} [\theta^{(i)}, \eta^{(i)}],\ \theta^{(i)} < \eta^{(i)},\ \theta^{(i)}, \eta^{(i)} \in [0,1]^d \Big\}.$$

We can restrict ourselves to the case of $k = 2$ and to $d = 2$ or even $d = 1$. So, let us assume that we may have a change on a union of two subintervals of $[0, 1]$. This is certainly a $V\check{C}$-class with $v = 5$. If $G_0 = [\theta^{(1)}, \eta^{(1)}] \cup [\theta^{(2)}, \eta^{(2)}]$ with $\eta^{(1)} < \theta^{(2)}$, then $N_\delta(t, G_0)$ will be locally again of order $(t/2\delta)^4$. However, suppose $G_0 = [\theta, \eta]$. Since for $[x, y]$ "close" to G_0 and $z > \max(\eta, y)$ or $u < \min(\theta, x)$ we have

$$d(G_0, [x, y] \cup [z, u]) = |\theta - x| + |\eta - y| + |z - u|,$$

we see that there are no more than t/δ points x and y and, for each choice of z, we have also no more than t/δ points u, but the number of different z's is of order $1/\delta$ rather than t/δ. Hence

$$N_\delta(t, G_0) \sim c_2 \frac{t^3}{\delta^4}.$$

We see that this effect of different m and m_1 is created by one small "freelance" interval. With two or more such small intervals and with $d > 1$ this difference will be greater. Since in the last example $m = 3$, $m_1 = v - 1 = 4$, we conclude from (2.8) that

$$\varepsilon_n(1) = \{\log(nc_1) + 3 \log \log(nc_1) + L + o(1)\}/nc_1$$

with $L = 4 \log((e \cdot c_2)/4)$.

Acknowledgement. The authors wish to thank the referee for very careful reading and a number of good suggestions.

References

[1] E. Mammen and A. B. Tsybakov, *Asymptotic minimax recovery of sets with smooth boundaries*, Ann. Statist., 23 (1995), 502–524.

[2] E. Khmaladze, R. Mnatsakanov, and N. Toronjadze, *The change-set problem and local covering numbers*, submitted to Math. Methods Statist., 2006.

[3] E. Khmaladze, R. Mnatsakanov, and N. Toronjadze, *The Change Set Problem.* Part 1: *Estimation of the change set and the rate of convergence*, Report S97-12, Dept of Statist., UNSW, Sydney, 1997.

[4] L. Birgé, *Model selection via testing: an alternative to (penalized) maximum likelihood estimators*, Ann. Inst. H. Poincaré. Probab. Statist., 42 (2006), 273–325.

[5] B. E. Brodsky and B. S. Darkhovsky, *Non-Parametric Statistical Diagnosis: Problems and Methods*, Kluwer Academic Publisher, Dordrecht, 2000.

[6] L. Dümbgen, *The asymptotic behaviour of some nonparametric change-point estimators*, Ann. Statist., 19 (1991), 1471–1495.

[7] A. Dempfle and W. Stute, *Nonparametric estimation of a discontinuity in regression*, Statistica Neerlandica, 56 (2002), 233–242.

[8] S. B. Hariz and J. J. Wylie, *Rates of convergence for the change-point estimator for long-range dependent sequences*, Statist. Probab. Lett., 73 (2005), 155–164.

[9] D. Pollard, *Asymptopia*, Manuscript, Yale University, Dept. of Statist., New Haven, Connecticut, 2000.

[10] I. A. Ibragimov and R. Z. Has'minskii, *Statistical Estimation: Asymptotic Theory*, Springer-Verlag, New York–Berlin, 1981.

[11] D. Ferger, *Boundary estimation based on set-indexed empirical processes*, J. Nonparam. Statist., 16 (2004), 245–260.

[12] A. Baddeley and R. Turner, *Modeling spatial point patterns in R*, submitted, 2005.

[*Received April* 2004; *revised June* 2006]