The Library of the Department of Statistics
North Carolina State University

NONPARAMETRIC CHANGE-POINT ESTIMATION
FOR DATA FROM AN ERGODIC SEQUENCE

E. Carlstein,* University of North Carolina at Chapel Hill

Subhash Lele,** Johns Hopkins University

* Research supported by National Science Foundation Grant #DMS-8902973.

** Visiting the University of North Carolina at Chapel Hill, August 1989.

1. INTRODUCTION

Suppose we observe a sequence of random variables $X_1^n, X_2^n, \dots, X_n^n$, where

X_i^n has marginal distribution F for $1 \leq i \leq [\theta n]$,

X_i^n has marginal distribution G for $[\theta n] + 1 \leq i \leq n$;

here $[y]$ denotes the greatest integer not exceeding y . The unknown parameter $\theta \in (0,1)$ is the *change-point* to be estimated. The purpose of this note is simply to show that θ can be consistently estimated in a fully nonparametric scenario:

- (i) No knowledge of F or G is required. There are *no parametric assumptions* (e.g., Normality) and *no regularity conditions* (e.g., continuity) on F or G .
- (ii) No restrictions are imposed on the *strength of serial dependence* in $\{X_i^n: 1 \leq i \leq n\}$, beyond the minimal condition of ergodicity. There are *no assumptions about the dependence mechanism* (e.g., autoregression), and there are *no “mixing” conditions*.
- (iii) No prior restrictions on θ are needed.

In fact, a whole *class* of such fully nonparametric estimators will be presented.

A huge amount of work has been done on change-point estimation in general (see [13] for an annotated bibliography and [11] for a recent extensive review); the importance of the *nonparametric* approach in particular is well established (see [5] for a review). Most of the nonparametric methods assume *independence* in $\{X_i^n: 1 \leq i \leq n\}$, and most of them assume *prior knowledge* about how F and G differ (e.g., in their means, medians, or other measure of level). However, there has recently been a trend towards eliminating prior assumptions on F and G , and towards allowing for serial dependence in $\{X_i^n: 1 \leq i \leq n\}$ (see [1] – [10] and [12]).

Within the nonparametric context, it is certainly desirable to minimize the assumptions on F and G . Since change-point data is inherently time-sequential, it is natural and practical to allow for serial dependence. It is desirable to minimize any restriction on the *strength* of this serial dependence, because: The *joint* distribution structure of $\{X_i^n: 1 \leq i \leq n\}$ is much more obscure than the *marginal* distribution structure (i.e., F and G), so it is unrealistic to

assume knowledge about the former when the latter is completely unknown. Moreover, the usual “mixing” assumptions are impossible to check for a given set of data. Our main consistency result (in Section 3) shows how far we can relax the restrictions on F , G , and the dependence structure.

Let us briefly contrast our fully nonparametric approach (i.e., (i), (ii), (iii) above) with the related works ([1] – [10] and [12]) which have recently appeared in the literature. In [4] and [5], it is assumed that the X_i^n s are *independent* and that F and G are *continuous*. In [2], [6], [8], and [9], it is assumed that F and G are both *discrete* with *finite support*, and that the sequence of X_i^n s satisfies a *strong-mixing* condition; it is also assumed that θ is in the *known* interval $[\alpha, \beta]$, where $0 < \alpha < \beta < 1$. In [7], either one of two possible scenarios is required: F and G are both *discrete* with *finite support*, and the X_i^n s are *strong-mixing*; or, F and G are both *continuous* with *support* in $[0, 1]$ and satisfy a *Lipschitz condition*, and the X_i^n s are *ψ -mixing*. In [1], it is assumed that the X_i^n s are *independent* (or possibly *m-dependent*) and that F and G are *continuous*; it is also assumed that θ is in the *known* interval $[\alpha, \beta]$, $0 < \alpha < \beta < 1$. In [3] and [10], the X_i^n s are again assumed to be *independent*. In [12], it is assumed that the X_i^n s arise from a *Gaussian process*, with F and G sharing the *same mean*. We shall impose *none* of these assumptions.

One final point of comparison between our approach and the works cited above: In each of the references [2], [4] – [10], and [12], a certain “norm” is used to calculate the basic statistic (this notion of a “norm” will be made precise in Section 2). Specifically, a *Kolmogorov-Smirnov norm* is used in [4], [5], [7], [10], [12]; and a *Cramér-von Mises norm* is used in [2], [6], [8], [9], [12]. Our method is based on a general “*Mean-Dominant norm*” (defined in Section 2), which includes as special cases the Kolmogorov-Smirnov norm, the Cramér-von Mises norm, as well as many other norms. Thus, our approach actually provides a whole *class* of consistent nonparametric estimators; moreover, our general formulation allows us to simultaneously analyze estimators based on *all* these norms with one unified theoretical argument (see Sections 2 and 4).

2. THE ESTIMATOR

The estimator is constructed from the *pre-t empirical c.d.f.*:

$$_t h^n(x) := \sum_{j=1}^{nt} \mathbf{I}\{X_j^n \leq x\} / nt,$$

and the *post-t empirical c.d.f.*:

$$h_t^n(x) := \sum_{j=nt+1}^n \mathbf{I}\{X_j^n \leq x\} / n(1-t).$$

The index t corresponds to *possible* values of the change-point estimator; we will only need to consider $t \in T_n$, where

$$T_n := \{1/n, 2/n, \dots, (n-1)/n\} \cap [\alpha_n, 1-\alpha_n],$$

and $\{\alpha_n: n \geq 1\}$ is any deterministic sequence satisfying $\alpha_n \downarrow 0$ and $n\alpha_n \uparrow \infty$ as $n \rightarrow \infty$.

For fixed $t \in T_n$, compute the differences between the pre- t and post- t empirical c.d.f.s at the sample observations, i.e.,

$$d_{ni}^t := |_t h^n(X_i^n) - h_t^n(X_i^n)|, \quad 1 \leq i \leq n.$$

These n differences are now combined via a “Mean-Dominant norm” $S_n: \mathbf{R}^n \mapsto \mathbf{R}$, yielding the criterion function

$$D_n(t) := t(1-t) S_n(d_{n1}^t, d_{n2}^t, \dots, d_{nn}^t).$$

Our fully nonparametric estimator θ_n is then defined to be:

$$\theta_n \in T_n \text{ such that } D_n(\theta_n) = \max_{t \in T_n} D_n(t).$$

The “Mean-Dominant norm” $S_n(\cdot, \cdot, \dots, \cdot)$ is any function satisfying the following natural conditions (whenever the arguments d_i and d'_i are all nonnegative):

- (1) [Symmetry] $S_n(\cdot, \cdot, \dots, \cdot)$ is symmetric in its n arguments.
- (2) [Homogeneity] $S_n(cd_1, cd_2, \dots, cd_n) = cS_n(d_1, d_2, \dots, d_n)$ whenever $c \geq 0$.
- (3) [Triangle Inequality] $S_n(d_1 + d'_1, d_2 + d'_2, \dots, d_n + d'_n) \leq S_n(d_1, d_2, \dots, d_n) + S_n(d'_1, d'_2, \dots, d'_n)$.
- (4) [Identity] $S_n(1, 1, \dots, 1) = 1$.
- (5) [Monotonicity] $S_n(d_1, d_2, \dots, d_n) \leq S_n(d'_1, d'_2, \dots, d'_n)$ whenever $d_i \leq d'_i \forall i$.
- (6) [Mean-Dominance] $S_n(d_1, d_2, \dots, d_n) \geq \sum_{1 \leq i \leq n} d_i / n$.

Special cases of the “Mean-Dominant norm” include the *Kolmogorov-Smirnov* norm

$$S_n^{KS}(d_1, d_2, \dots, d_n) := \sup_{1 \leq i \leq n} \{d_i\},$$

the *Cramér-von Mises* norm

$$S_n^{Cv}(d_1, d_2, \dots, d_n) := (\sum_{1 \leq i \leq n} d_i^2/n)^{1/2},$$

and the *arithmetic-mean* norm

$$S_n^{am}(d_1, d_2, \dots, d_n) := \sum_{1 \leq i \leq n} d_i/n.$$

The intuition behind our method is as follows. The c.d.f.s ${}_t h^n(x)$ and $h_t^n(x)$ are empirical approximations of the unknown distributions

$${}_t h(x) := \mathbb{I}\{t \leq \theta\} F(x) + \mathbb{I}\{t > \theta\} (\theta F(x) + (t - \theta) G(x)) / t$$

and

$$h_t(x) := \mathbb{I}\{t \leq \theta\} ((\theta - t) F(x) + (1 - \theta) G(x)) / (1 - t) + \mathbb{I}\{t > \theta\} G(x),$$

respectively. Therefore the differences d_{ni}^t are empirical approximations of

$$\begin{aligned} \delta_{ni}^t &:= |{}_t h(X_i^n) - h_t(X_i^n)| \\ &= \delta_{ni}^\theta (\mathbb{I}\{t \leq \theta\} (1 - \theta) / (1 - t) + \mathbb{I}\{t > \theta\} \theta / t), \quad 1 \leq i \leq n. \end{aligned}$$

And, the criterion function $D_n(t)$ is an empirical approximation of the corresponding function

$$\Delta_n(t) := t(1 - t) S_n(\delta_{n1}^t, \delta_{n2}^t, \dots, \delta_{nn}^t).$$

Now, by S_n 's homogeneity, we have

$$\Delta_n(t) = \rho(t) S_n(\delta_{n1}^\theta, \delta_{n2}^\theta, \dots, \delta_{nn}^\theta),$$

where

$$\rho(t) := \mathbb{I}\{t \leq \theta\} t(1 - \theta) + \mathbb{I}\{t > \theta\} (1 - t)\theta.$$

Notice that the maximizer of $\Delta_n(t)$ over $t \in (0, 1)$ is *precisely* at $t = \theta$. Thus, the maximizer of the analogous sample-based criterion function $D_n(t)$ is a reasonable estimator of θ . This logic applies for *any* “Mean-Dominant norm” S_n .

3. MAIN RESULT

The data $\{X_i^n: 1 \leq i \leq n\}$ are embedded in a stationary ergodic sequence

$$\{(U_i, V_i): -\infty < i < +\infty\},$$

where U_i has marginal distribution F , and V_i has marginal distribution G . Specifically, the data arise as

$$X_i^n = U_i \mathbb{I}\{1 \leq i \leq [\theta n]\} + V_i \mathbb{I}\{[\theta n] + 1 \leq i \leq n\}.$$

There will be no further constraints on the serial dependence structure of $\{X_i^n: 1 \leq i \leq n\}$. The only assumption on the unknown marginal distributions is simply that $F \neq G$. The unknown change-point parameter $\theta \in (0, 1)$ is unrestricted. In this scenario, our fully nonparametric estimator θ_n (defined in Section 2) is consistent.

THEOREM: $\theta_n \xrightarrow{\mathbf{P}} \theta \quad \text{as } n \rightarrow \infty.$

See Section 4 for a proof of this result.

4. PROOF

We begin by presenting three preliminary Lemmas which will be used in proving the Theorem.

LEMMA 1: Denote $\mu_F := \int_{-\infty}^{\infty} |F(x) - G(x)| dF(x)$, $\mu_G := \int_{-\infty}^{\infty} |F(x) - G(x)| dG(x)$, and $\mu := \theta \mu_F + (1 - \theta) \mu_G$. Then $\mu > 0$.

Proof: By assumption we have $\Lambda := \{x \in \mathbb{R}: |F(x) - G(x)| > 0\} \neq \emptyset$. It suffices to show that either $\int_{\Lambda} dF(x) > 0$ or $\int_{\Lambda} dG(x) > 0$. The case where Λ contains a discontinuity point of F or G is trivial, so we will now presume that F and G are continuous at each $x \in \Lambda$.

Select $x_0 \in \Lambda$ with (say) $F(x_0) > G(x_0)$. Then $\sigma := \{y \in (-\infty, x_0): F(x) > G(x) \ \forall x \in (y, x_0]\}$

is nonempty, by continuity. Denote $y_0 := \inf\{y \in \sigma\}$. If $y_0 = -\infty$, then $(-\infty, x_0] \subseteq \Lambda$ and therefore $\int_A dF(x) \geq F(x_0) > G(x_0) \geq 0$. If $y_0 > -\infty$, then $F(y_0) \leq G(y_0)$ and also $(y_0, x_0] \subseteq \Lambda$, yielding $\int_A dF(x) \geq F(x_0) - F(y_0) \geq F(x_0) - G(y_0) > G(x_0) - G(y_0) \geq 0$. $\square$

LEMMA 2: Let $\{Y_i: -\infty < i < +\infty\}$ be a stationary ergodic sequence, where Q is the marginal distribution of Y_i . Let $\{a_n: n \geq 1\}$ be a deterministic sequence satisfying $a_n \uparrow \infty$ as $n \rightarrow \infty$, and let $\{N_n: n \geq 1\}$ be integer-valued random variables for which $N_n \geq a_n \forall n$. Define for $m_2 \geq m_1 + 1$,

$$W_Q(m_1, m_2) := \sup_{-\infty < y < \infty} \left| \sum_{i=m_1+1}^{m_2} \mathbb{I}\{Y_i \leq y\} / (m_2 - m_1) - Q(y) \right|.$$

Then $W_Q(0, N_n) \xrightarrow{a.s.} 0$ as $n \rightarrow \infty$.

Proof: Follows directly from the Theorem in [14]. $\square$

LEMMA 3: Denote $A_n := T_n \cap \{[\alpha_n, \theta - \alpha_n] \cup (\theta + \alpha_n, 1 - \alpha_n]\}$. Let $\{R_n: n \geq 1\}$ be a (possibly random) sequence with $R_n \equiv r_n((U_1, V_1), (U_2, V_2), \dots, (U_n, V_n))$ for a deterministic function $r_n: \mathbb{R}^{2n} \mapsto A_n$. Then $|\Delta_n(R_n) - D_n(R_n)| \xrightarrow{P} 0$ as $n \rightarrow \infty$.

Proof: Denote $H_{ni} := |h_{R_n}^n(X_i^n) - h_{R_n}(X_i^n)|$, ${}_n H := |R_n h^n(X_i^n) - R_n h(X_i^n)|$, and $e_{ni} := {}_n H + H_{ni}$, so that $d_{ni}^{R_n} \leq e_{ni} + \delta_{ni}^{R_n}$ and therefore:

$$\Gamma_n := D_n(R_n) - \Delta_n(R_n) \leq S_n(e_{n1}, e_{n2}, \dots, e_{nn}) \leq S_n({}_n H, {}_n H, \dots, {}_n H) + S_n(H_{n1}, H_{n2}, \dots, H_{nn})$$

by (5) & (3). The same bound holds for $-\Gamma_n$ and hence for $|\Gamma_n|$. Now,

$$\begin{aligned} H_{ni} = & \left| \mathbb{I}\{R_n > \theta\} \left(\sum_{j=nR_n+1}^n \frac{\mathbb{I}\{X_j^n \leq X_i^n\}}{n - nR_n} - G(X_i^n) \right) + \right. \\ & \mathbb{I}\{R_n \leq \theta\} \left(\left(\sum_{j=nR_n+1}^{[\theta n]} \frac{\mathbb{I}\{X_j^n \leq X_i^n\}}{[\theta n] - nR_n} - F(X_i^n) \right) \frac{[\theta n] - nR_n}{n - nR_n} + F(X_i^n) \left(\frac{[\theta n] - nR_n}{n - nR_n} - \frac{\theta - R_n}{1 - R_n} \right) + \right. \\ & \left. \left. \left(\sum_{j=[\theta n]+1}^n \frac{\mathbb{I}\{X_j^n \leq X_i^n\}}{n - [\theta n]} - G(X_i^n) \right) \frac{n - [\theta n]}{n - nR_n} + G(X_i^n) \left(\frac{n - [\theta n]}{n - nR_n} - \frac{1 - \theta}{1 - R_n} \right) \right) \right| \end{aligned}$$

$$\begin{aligned} &\leq \mathbf{I}\{R_n > \theta\} W_G(nR_n, n) + \mathbf{I}\{R_n \leq \theta\} (W_F(nR_n, [\theta n]) + W_G([\theta n], n) + 2/n(1-R_n)) \\ &\leq W_G(\mathbf{I}\{R_n > \theta\} nR_n + \mathbf{I}\{R_n \leq \theta\} [\theta n], n) + W_F(\mathbf{I}\{R_n \leq \theta\} nR_n, [\theta n]) + 2/n(1-\theta) =: H_n. \end{aligned}$$

(In $W_F(\cdot)$, U_i plays the role of Y_i ; in $W_G(\cdot)$, V_i plays the role of Y_i .) An analogous bound (say, ${}_nH$) can be obtained for ${}_nH$. Since H_n and ${}_nH$ do not depend on i , we have $|\Gamma_n| \leq {}_nH + H_n$, by (5), (2), & (4).

We will now show that $H_n \xrightarrow{\mathbf{P}} 0$; a similar argument applies to ${}_nH$. We deal explicitly with the second summand in H_n ; the first summand can be handled analogously to the second, and the third summand goes to zero deterministically. Denote $Z_i := (U_i, V_i)$ and $\mathcal{Z}_n^k := (Z_{k+1}, Z_{k+2}, \dots, Z_{k+n})$. Note that:

$$\begin{aligned} W_F(\mathbf{I}\{r_n(\mathcal{Z}_n^0) \leq \theta\} nr_n(\mathcal{Z}_n^0), [\theta n]) &\stackrel{\mathfrak{D}}{=} W_F(\mathbf{I}\{r_n(\mathcal{Z}_n^{-[\theta n]-1}) \leq \theta\} nr_n(\mathcal{Z}_n^{-[\theta n]-1}) - [\theta n] - 1, -1) \\ &= W_F(0, [\theta n] - \mathbf{I}\{r_n(\mathcal{Z}_n^{-[\theta n]-1}) \leq \theta\} nr_n(\mathcal{Z}_n^{-[\theta n]-1})), \end{aligned}$$

where the last expression uses $Y_i \equiv U_{-i}$. We can now apply Lemma 2, since $r_n(\cdot) \in A_n$ and so:

$$[\theta n] - \mathbf{I}\{r_n(\cdot) \leq \theta\} nr_n(\cdot) \geq [\theta n] - n(\theta - \alpha_n) \geq (n\alpha_n - 1) \uparrow \infty.$$

This establishes $W_F(\mathbf{I}\{R_n \leq \theta\} nR_n, [\theta n]) \xrightarrow{\mathbf{P}} 0$. $\square$

The following notation and definitions will be needed to prove the Theorem. The random variable $\tilde{\theta}_n$ is defined as follows:

if $\theta_n \in A_n$, then let $\tilde{\theta}_n = \theta_n$;

if $\theta_n \notin A_n$, then let $\tilde{\theta}_n \in A_n$ satisfy $D_n(\tilde{\theta}_n) = \max_{t \in A_n} D_n(t)$.

Note that, in either case, we have $\tilde{\theta}_n \in A_n$ and $D_n(\tilde{\theta}_n) = \max_{t \in A_n} D_n(t)$. Also define the nonrandom entity

$$\begin{aligned} t_n := & \mathbf{I}\left\{\rho\left([\theta - \alpha_n]n/n\right) \geq \rho\left([\theta + \alpha_n]n/n + 1/n\right)\right\} [\theta - \alpha_n]n/n + \\ & + \mathbf{I}\left\{\rho\left([\theta - \alpha_n]n/n\right) < \rho\left([\theta + \alpha_n]n/n + 1/n\right)\right\} ([\theta + \alpha_n]n/n + 1/n), \end{aligned}$$

which satisfies $t_n \in A_n$ and $\Delta_n(t_n) = \max_{t \in A_n} \Delta_n(t)$.

In order to prove the Theorem, it suffices to establish that $\tilde{\theta}_n \xrightarrow{\mathbf{P}} \theta$ as $n \rightarrow \infty$. To see that this is sufficient, write $\mathbf{P}\{|\theta_n - \theta| > \varepsilon\} = \mathbf{P}\{|\theta_n - \theta| > \varepsilon \cap \theta_n \in A_n\} + \mathbf{P}\{|\theta_n - \theta| > \varepsilon \cap \theta_n \notin A_n\}$.

The latter probability is zero for n sufficiently large; the other probability is bounded by $\mathbb{P}\{|\tilde{\theta}_n - \theta| > \varepsilon\}$.

To establish $\tilde{\theta}_n \xrightarrow{\mathbb{P}} \theta$, denote $\delta_n := \sum_{1 \leq i \leq n} \delta_{ni}^\theta / n$ and $\bar{\theta} := \min\{\theta, 1 - \theta\}$, and notice that:

$$\begin{aligned} \Psi_n &:= \Delta_n(\theta) - \Delta_n(\tilde{\theta}_n) = \left(\rho(\theta) - \rho(\tilde{\theta}_n) \right) S_n(\delta_{n1}^\theta, \delta_{n2}^\theta, \dots, \delta_{nn}^\theta) \\ &\geq \left(\mathbf{I}\{\tilde{\theta}_n \leq \theta\}(\theta - \tilde{\theta}_n)(1 - \theta) + \mathbf{I}\{\tilde{\theta}_n > \theta\}(\tilde{\theta}_n - \theta)\theta \right) \delta_n \geq |\tilde{\theta}_n - \theta| \bar{\theta} \delta_n; \end{aligned}$$

where the first inequality follows from (6). Therefore

$$\begin{aligned} \mathbb{P}\{|\tilde{\theta}_n - \theta| > \varepsilon\} &\leq \mathbb{P}\{\Psi_n > \varepsilon \bar{\theta} \delta_n \cap \delta_n \geq \nu\} + \mathbb{P}\{\Psi_n > \varepsilon \bar{\theta} \delta_n \cap \delta_n < \nu\} \\ &\leq \mathbb{P}\{\Psi_n > \varepsilon \bar{\theta} \nu\} + \mathbb{P}\{\delta_n < \nu\}, \end{aligned}$$

where $\nu := \mu/2 > 0$ by Lemma 1. To deal with the latter probability, write $\delta_n = ([\theta n] \underline{\delta}_n + (n - [\theta n]) \bar{\delta}_n) / n$, where $\underline{\delta}_n := \sum_{1 \leq i \leq [\theta n]} \delta_{ni}^\theta / [\theta n]$ and $\bar{\delta}_n := \sum_{[\theta n] + 1 \leq i \leq n} \delta_{ni}^\theta / (n - [\theta n])$. Note that $\{\delta_n < \nu\} \Rightarrow \{\nu < |\delta_n - \mu| \leq |\underline{\delta}_n - \mu_F| + |\bar{\delta}_n - \mu_G| + 2|\theta - [\theta n]/n|\}$. Since the term in the last modulus goes to zero deterministically, we only need to consider $|\bar{\delta}_n - \mu_G|$ (the remaining term is handled analogously). Now $\bar{\delta}_n^* := \sum_{1 \leq i \leq n - [\theta n]} |F(V_i) - G(V_i)| / (n - [\theta n]) \xrightarrow{\mathbb{P}} \mu_G$ by the Ergodic Theorem, and $\bar{\delta}_n \stackrel{\mathbb{Q}}{=} \bar{\delta}_n^*$, so that $|\bar{\delta}_n - \mu_G| \xrightarrow{\mathbb{P}} 0$.

To prove the Theorem, it now suffices to show that $\Psi_n \xrightarrow{\mathbb{P}} 0$. We have $\Psi_n \leq |\Delta_n(\theta) - \Delta_n(t_n)| + |\Delta_n(t_n) - D_n(\tilde{\theta}_n)| + |D_n(\tilde{\theta}_n) - \Delta_n(\tilde{\theta}_n)|$, where the first modulus is deterministically dominated (using (5) & (4)) by:

$$\rho(\theta) - \rho(t_n) = \mathbf{I}\{t_n \leq \theta\}(\theta - t_n)(1 - \theta) + \mathbf{I}\{t_n > \theta\}(t_n - \theta)\theta \leq \alpha_n + n^{-1};$$

and where the second modulus is dominated by $|D_n(t_n) - \Delta_n(t_n)| + |D_n(\tilde{\theta}_n) - \Delta_n(\tilde{\theta}_n)|$ (because either $\Delta_n(t_n) \geq D_n(\tilde{\theta}_n) \geq D_n(t_n)$ or $D_n(\tilde{\theta}_n) \geq \Delta_n(t_n) \geq \Delta_n(\tilde{\theta}_n)$ holds). Thus, applying Lemma 3 with $R_n \equiv t_n$ and with $R_n \equiv \tilde{\theta}_n$ completes our proof of the Theorem.

REFERENCES

- [1] Asatryan, D. and Safaryan, I. (1986). Nonparametric methods for detecting changes in the properties of random sequences. In *Detection of Changes in Random Processes* (L. Telksnys, ed.), 1-13. Optimization Software, New York.
- [2] Brodskii, B. and Darkhovskii, B. (1986). The a posteriori method of detecting the disruption of a random field. In *Detection of Changes in Random Processes* (L. Telksnys, ed.), 32-38. Optimization Software, New York.
- [3] Carlstein, E. (1988). Nonparametric change-point estimation. *Annals of Statistics*, 16, 188-197.
- [4] Csörgő, M. and Horváth, L. (1987). Nonparametric tests for the changepoint problem. *Journal of Statistical Planning and Inference*, 17, 1-9.
- [5] Csörgő, M. and Horváth, L. (1988). Nonparametric methods for changepoint problems. In *Handbook of Statistics, Volume 7* (P. Krishnaiah and C. R. Rao, ed.s), 403-425. Elsevier, The Netherlands.
- [6] Darkhovskii, B. (1984). On two estimation problems for times of change of the probabilistic characteristics of a random sequence. *Theory of Probability and its Applications*, 29, 478-487.
- [7] Darkhovskii, B. (1985). A nonparametric method of estimating intervals of homogeneity for a random sequence. *Theory of Probability and its Applications*, 30, 845-849.
- [8] Darkhovskii, B. (1986). A general method for estimating the instant of change in the probabilistic characteristics of a random sequence. In *Detection of Changes in Random Processes* (L. Telksnys, ed.), 47-52. Optimization Software, New York.
- [9] Darkhovskii, B. and Brodskii, B. (1980). A posteriori detection of the "disorder" time of a random sequence. *Theory of Probability and its Applications*, 25, 624-628.
- [10] Deshayes, J. and Picard, D. (1981). Convergence de processus à double indice: application aux tests de rupture dans un modèle. *C.R. Acad. Sci. Paris*, 292, 449-452.
- [11] Krishnaiah, P. and Miao, B. (1988). Review about estimation of change points. In *Handbook of Statistics, Volume 7* (P. Krishnaiah and C. R. Rao, ed.s), 375-402. Elsevier, The Netherlands.
- [12] Picard, D. (1985). Testing and estimating change-points in time series. *Advances in Applied Probability*, 17, 841-867.
- [13] Shaban, S. (1980). Change-Point Problem and Two-Phase Regression: An Annotated Bibliography. *International Statistical Review*, 48, 83-93.
- [14] Tucker, H. (1959). A generalization of the Glivenko-Cantelli theorem. *Annals of Mathematical Statistics*, 30, 828-830.