The Annals of Statistics
1995, Vol. 23, No. 5, 1848–1861

NONPARAMETRIC TESTS FOR NONSTANDARD CHANGE-POINT PROBLEMS

BY D. FERGER

University of Giessen

We consider independent random elements $X_1, \dots, X_n$, $n \in \mathbb{N}$, with values in a measurable space $(\mathscr{X}, \mathscr{B})$ so that $X_1, \dots, X_{[n\theta]}$ have a common distribution ν_1 and the remaining $X_{[n\theta]+1}, \dots, X_n$ have a common distribution $\nu_2 \neq \nu_1$, for some $\theta \in (0,1)$. The *change point* θ as well as the distributions are unknown. A family of tests is introduced for the *nonstandard* change-point problem H_0: $\theta \in \Theta_0$ versus H_1: $\theta \notin \Theta_0$, where Θ_0 is an arbitrary subset of (0, 1). The tests are shown to be asymptotic level-α tests and to be consistent on a large class of alternatives. The same holds for the corresponding *bootstrap* versions of the tests. Moreover, we present a detailed investigation of the local power.

0. Introduction. Suppose we observe independent random elements $X_1, \dots, X_n$ in a sample space $(\mathscr{X}, \mathscr{B})$ so that $X_1, \dots, X_{[n\theta]}$ have a common distribution ν_1 and the remaining $X_{[n\theta]+1}, \dots, X_n$ have a common distribution $\nu_2 \neq \nu_1$, for some $\theta \in (0,1)$. The *change point* θ as well as the distributions are unknown. Many situations can be modeled in this way. Well-known and illustrative examples are the Nile data [cf. Cobb (1978)], the coal mines data [cf. Maguire, Pearson and Wynn (1952)] and the Lindisfarne scribes data [cf. Smith (1980)]. One basic question is whether there is a change at all. In other words, we are concerned with the test problem H_0^*: $\theta = 0$ versus H_1^*: $\theta \in (0,1)$. There is a host of papers dealing with this *standard* change-point problem. For example, see Bhattacharya and Brockwell (1976), Bhattacharya and Frierson (1981), Csörgő and Horváth (1987, 1988a, b), Deshayes and Picard (1981), Lombard (1987), Lorden (1971), Page (1954, 1955), Pettitt (1979) or Worsley (1986). In principle there exist two different approaches.

As to the first, recall that $\theta \in (0,1)$, divides the data into two *different subsamples*. Since θ is unknown, we consider, for each *possible* change point $t \in (0,1)$, the subsamples $X_1, \dots, X_{[nt]}$ and $X_{[nt]+1}, \dots, X_n$ and their corresponding empirical measures

$$^tP_n := \frac{1}{[nt]} \sum_{i=1}^{[nt]} \delta_{X_i} \quad \text{and} \quad P_n^t := \frac{1}{n-[nt]} \sum_{i=[nt]+1}^{n} \delta_{X_i},$$

where δ_x denotes the Dirac measure at $x \in \mathscr{X}$. Next a "distance" $d(^tP_n, P_n^t)$ between these two empirical measures is determined. For example, Dümbgen

Received November 1991; revised January 1995.
AMS 1991 *subject classifications*. Primary 62G10, 62G20, 62G09; secondary 60F05.
Key words and phrases. Tests for change-point problems, maximizer of a two-sided random walk, consistency, local power, bootstrap.

(1991) suggests $d({}^tP_n, P_n^t) = \sup_{D \in \mathscr{D}} |{}^tP_n(D) - P_n^t(D)|$, where $\mathscr{D} \subseteq \mathscr{B}$ is some given subclass of measurable sets. In the case $\mathscr{X} = \mathbb{R}$ and $\mathscr{D} = \{(-\infty, x]: x \in \mathbb{R}\}$ this leads to the familiar Kolmogorov–Smirnov distance between the two empirical distribution functions of the two subsamples [cf. Deshayes and Picard (1981)]. The underlying philosophy is that the distance $d({}^tP_n, P_n^t)$ is small if H_0^* holds and that there are significant deviations otherwise. Hence these tests reject H_0^* in favor of H_1^* if $\max_{t \in (0,1)} d({}^tP_n, P_n^t)$ is significantly large. Csörgő and Horváth (1988b) deal with the "distance" $d({}^tP_n, P_n^t) = \int K \, d\, {}^tP_n \otimes P_n^t$, where K is a given kernel. For $K(x, y) = \operatorname{sign}(x - y)$ we obtain a test proposed by Pettitt (1979). Of course the distance $d({}^tP_n, P_n^t)$ can also be used to define *estimators* for the unknown changepoint θ; namely, since we expect the distance d, as a function of $t \in (0, 1)$, to be maximal at the point $t = \theta$ we are led to set

$$\theta_n = \arg\max_{t \in (0,1)} d\left({}^tP_n, P_n^t\right).$$

Special versions of this estimator have been suggested by Bhattacharya and Brockwell (1976), Carlstein (1988), Hinkley (1970), Darkhovskh (1976) or Dümbgen (1991).

For the second approach assume that, for $0 < t < 1$, $S_n(t)$ is a statistic designed for testing whether the two subsamples differ in distribution or not. If, for example, an upper two-sample test is used, H_0^* is rejected if $\max_{t \in (0,1)} S_n(t)$ is too large. Similarly, $\theta_n = \arg\max_{t \in (0,1)} S_n(t)$ may serve as an estimator of θ. See Darkhovskh (1976), whose method is based on the Mann–Whitney statistic.

Besides the question whether a change has occurred or not it is in many situations more natural to ask if a change has taken place within a certain time period or not. As an example, consider an environmental system which is exposed to an external pollution during a given time period. Then we want to know whether the pollution has effects on a certain population living in this system. Formally, we are faced with the test situation

$$H_0: \theta \in \Theta_0 \quad \text{versus} \quad H_1: \theta \notin \Theta_0,$$

where Θ_0 is a given subset of the open unit interval. If the time interval degenerates to a single point, that is, $\Theta_0 = \{\theta_0\}$, one can proceed as follows: take any estimator θ_n of the unknown change point θ for which an (asymptotic) distribution theory is available. In many cases we have that $n(\theta_n - \theta) \to_{\mathscr{L}} \xi$, where ξ is the maximizer of a two-sided random walk on $\mathbb{Z}$. Results of that kind have been proved by Hinkley (1970) in a parametric setup and by Dümbgen (1991) and Ferger (1994a–c) in a nonparametric framework. The decision rule

$$\tau_n = 1_{\{n|\theta_n - \theta_0| > c\}}$$

gives an asymptotic level-α test, provided c satisfies the equality $1 - F(c) = \alpha$, with F denoting the distribution function of the limit variable $|\xi|$. Moreover, it is easy to see that τ_n is consistent, that is, $\lim_{n \to \infty} P_\theta(\tau_n = 1) = 1$ for each $\theta \neq \theta_0$. In general F is unknown. Hence we cannot determine the critical value c. It can be approximated, however, via bootstrap methods as carried out by Dümbgen (1991) and Ferger (1994a). This method does not apply when Θ_0 contains more than one element, for example, $\Theta_0 = [a, b]$ with $0 < a < b < 1$. To the best of our knowledge we are not aware of any contribution to this *nonstandard* test situation. It will be the subject of the present paper.

1. The tests and their large-sample properties. In this section we will present a large class of nonparametric tests for H_0 versus H_1. To facilitate the presentation of the large-sample results, we prefer to work with a triangular array $X_{1n}, \dots, X_{nn}$ of random elements rather than a sequence. Now, consider the modified empirical measures of the two subsamples generated by $t \in (0,1)$:

$$ {}^t\mu_n := n^{-1} \sum_{1 \le i \le nt} \delta_{X_{in}} \quad \text{and} \quad \mu_n^t := n^{-1} \sum_{nt < i \le n} \delta_{X_{in}}. $$

If, for example, $\mathscr{X}$ is a separable metric space a variation of arguments of Varadarajan (1958) yields that ${}^t\mu_n$ and μ_n^t converge in the weak topology with probability 1 to the measures

$$ {}^t\mu = 1_{\{t \le \theta\}} t\nu_1 + 1_{\{t > \theta\}}\big(\theta\nu_1 + (t - \theta)\nu_2\big) $$

and

$$ \mu^t = 1_{\{t \le \theta\}}\big((\theta - t)\nu_1 + (1 - \theta)\nu_2\big) + 1_{\{t > \theta\}}(1 - t)\nu_2, $$

respectively. Let $K: \mathscr{X}^2 \to \mathbb{R}$ be a measurable mapping (kernel). Set

$$ r(t) = \int K \, d\mu^t \otimes {}^t\mu = \begin{cases} t\mu(\theta - t) + t\lambda(1 - \theta), & 0 \le t \le \theta, \\ (1 - t)\lambda\theta + (1 - t)\tau(t - \theta), & \theta < t \le 1, \end{cases} $$

with $\mu = \int K \, d\nu_1 \otimes \nu_1$, $\tau = \int K \, d\nu_2 \otimes \nu_2$ and $\lambda = \int K \, d\nu_2 \otimes \nu_1$. Obviously r is continuous on $[0,1]$ and differentiable at $t \neq \theta$. Let $w(t)$, $0 < t < 1$, be a *weight function* of the type

$$ w(t) = t^{-a}(1 - t)^{-b}, \qquad 0 \le a, b < 1. $$

For the sake of simplicity let us assume, for a moment, that K is antisymmetric, that is, $K(x, y) = -K(y, x)$ for all $x, y \in \mathscr{X}$. In that case $\mu = \tau = 0$ and the function $r(t)$ is a simple polygonal line through the points $(0,0)$, $(\theta, \lambda\theta(1 - \theta))$ and $(1, 0)$. Thus θ is the unique maximizer or minimizer of $r(t)$ according as λ is positive or negative. The weighted function

$$ \rho(t) = w(t) r(t), \qquad 0 \le t \le 1, $$

has roughly the same shape as $r(t)$. Consequently, we define our estimator θ_n of θ to be the maximizer of the empirical analogue of $|\rho|$:

$$ \theta_n = \operatorname*{arg\,max}_{t \in G_n} w(t) |r_n(t)| \tag{1.1} $$

with

$$r_n(t) = \int K \, d\mu_n^t \otimes {}^t\mu_n = n^{-2} \sum_{nt < i \le n} \sum_{1 \le j \le nt} K(X_{in}, X_{jn}), \qquad 0 \le t \le 1,$$

and

$$G_n = \{kn^{-1} : 1 \le k \le n - 1\}.$$

In the general case we need the following assumption (P) on the shape of the function ρ.

(P) There exists a positive constant L such that either

$$(1.2) \quad \rho(\theta) - \rho(t) \ge L|t - \theta| \quad \forall\, t \in [0,1] \quad \text{and} \quad -\inf_{0 \le t \le 1} \rho(t) < \rho(\theta)$$

or

$$(1.3) \quad \rho(t) - \rho(\theta) \ge L|t - \theta| \quad \forall\, t \in [0,1] \quad \text{and} \quad \sup_{0 \le t \le 1} \rho(t) < -\rho(\theta).$$

Geometrically (P) means that the graph of $|\rho(t)|$ has a unique peak at θ. Of course (P) is an implicit assumption on the underlying distributions ν_1 and ν_2. Note that, for an antisymmetric K, (P) reduces to the simple requirement that λ is nonzero. Especially the quantities L or λ may be interpreted as a means to measure the distance between ν_1 and ν_2.

REMARK. Observe that θ_n and ρ depend on K. We write $\theta_n = \theta_n[K]$ and $\rho = \rho[K]$. Since $\theta_n[K] = \theta_n[-K]$ and $\rho[-K] = -\rho[K]$ we can theoretically assume that, for example, (1.2) holds. Ferger and Stute (1992) proved (for bounded kernels and $w = 1$) that

$$|\theta_n - \theta| = O(n^{-1} \log n) \quad P_\theta\text{-a.s.}, \forall\, \theta \in (0,1).$$

Here and in the sequel the notation P_θ is used to stress the fact that θ is the true change point. Now, it is plausible to *reject H_0 if the distance of θ_n and Θ_0 is too large*. Formally,

$$\tau_n = 1_{\{\theta_n \notin \Theta_0^{\varepsilon/n}\}},$$

where for $A \subseteq [0,1]$ and $\varepsilon > 0$, A^ε denotes the ε-neighborhood of A: $A^\varepsilon = \{x \in [0,1] : d(x, A) < \varepsilon\}$ with $d(x, A) = \inf\{|x - a| : a \in A\}$. Furthermore, c is defined by

$$(1.4) \qquad c = \min\{k \in \mathbb{N} : 1 - F_{\max}(k - 1) + F_{\min}(-k) \le \alpha\},$$

with $F_{\max}$ and $F_{\min}$ denoting the distribution functions of the $\mathbb{Z}$-valued random variables $T_{\max}$ and $T_{\min}$ as defined in Section 4. In applications, $\Theta_0 = [a, b]$, $0 < a < b < 1$, so that $\theta_0^\varepsilon = (a - \varepsilon, b + \varepsilon)$. Besides (P) we need the following *second moment condition*:

$$(\mathrm{M}) \qquad \int |K|^2 \, d\nu_i \otimes \nu_j < \infty \qquad \forall\, 1 \le j \le i \le 2.$$

THEOREM 1.1. *Let* $(\mathscr{X}, \mathscr{B})$ *be a measurable space and assume that* (P) *and* (M) *hold. Then, for all* $\alpha \in (0,1)$,

$$\limsup_{n\to\infty} P_\theta(\tau_n \text{ rejects } H_0) \le \alpha \qquad \forall\, \theta \in \Theta_0 \tag{1.5}$$

and

$$\lim_{n\to\infty} P_\theta(\tau_n \text{ rejects } H_0) = 1 \qquad \forall\, \theta \notin \overline{\Theta}_0, \tag{1.6}$$

where $\overline{\Theta}_0$ *denotes the topological closure of* Θ_0.

The theorem states that τ_n is a consistent and asymptotic level-α test. Of course, the consistency (1.6) of the test τ_n is a *global power* property. In contrast the next theorem describes the *local power* of τ_n. Here, we let the change point $\theta = \bar{\theta}_n \notin \Theta_0$ tend to Θ_0 as the sample size n tends to infinity. More precisely, let $(\bar{\theta}_n) \subseteq (0,1) \setminus \Theta_0$ be a convergent sequence such that the distance

$$\varepsilon_n = d(\bar{\theta}_n, \Theta_0)$$

between $\bar{\theta}_n$ and Θ_0 converges to zero. Without loss of generality, we can and do assume that $\overline{\Theta}_n \in G_n$ for all $n \in \mathbb{N}$. Observe that the limit $\bar{\theta} = \lim_{n\to\infty} \theta_n$ necessarily is a boundary point of Θ_0.

THEOREM 1.2. *Assume the assumptions of Theorem* 1.1 *hold and that* $\overline{\Theta}_0 \subset (0,1)$.

If $\varepsilon_n = \eta^{-1} c n^{-1}$, *where* $\eta \in (0,1)$ *is arbitrary, then*

$$\liminf_{n\to\infty} P_{\bar{\theta}_n}(\tau_n \text{ rejects } H_0) \ge F_{\max}([\tilde{\eta} c]) - F_{\min}(-[\tilde{\eta} c] - 1), \tag{1.7}$$

where $\tilde{\eta} = \eta^{-1}(1-\eta)$. *Note that the lower bound in* (1.7) *increases to* 1 *as* $\eta \downarrow 0$. *Moreover, the rate* n^{-1} *is exact in the following sense*: *if* ε_n *tends to zero at a rate slower than* n^{-1}, *that is*, $n\varepsilon_n \to \infty$, *then*

$$\lim_{n\to\infty} P_{\bar{\theta}_n}(\tau_n \text{ rejects } H_0) = 1; \tag{1.8}$$

otherwise, that is, if $n\varepsilon_n \to 0$, *then*

$$\limsup_{n\to\infty} P_{\bar{\theta}_n}(\tau_n \text{ rejects } H_0) \le \alpha. \tag{1.9}$$

A detailed discussion concerning an appropriate choice of the kernel K follows in Section 3. The critical value c is in general *unknown*. As a way out we will show the validity of a bootstrap approximation in the next section.

2. Bootstrap approximations. If the *uniqueness condition*

$$F_{\max} = F_{\min} \tag{U}$$

is satisfied, the unknown critical value c can be rewritten as

$$c = F^{-1}(1-\alpha) + 1,$$

with F denoting the distribution function of $|T_{\min}|$. By Theorem 4.1,

$$\xi_n = n\theta_n - [n\theta]$$

converges in distribution to $T_{\min}$. Therefore we approximate the unknown distribution $\mathscr{L}(|T_{\min}|)$ by the conditional distribution

$$\mathscr{L}(|\xi_n^*| \,|X_{1n}, \dots, X_{nn}),$$

where

$$\xi_n^* = n\theta_n^* - n\theta_n$$

and θ_n^* is the change point estimator (1.1) pertaining to a bootstrap sample $X_{1n}^*, \dots, X_{nn}^*$. To be precise, given $\mathbf{X}_n = (X_{1n}, \dots, X_{nn})$ the variables X_{in}^* are independent with distribution $\nu_{1,n}$ for $1 \le i \le n\theta_n$ and $\nu_{2,n}$ otherwise. Here $\nu_{1,n}$ and $\nu_{2,n}$ denote the empirical measures of the two subsamples generated by θ_n:

$$\nu_{1,n} = \frac{1}{n\theta_n} \sum_{1 \le i \le n\theta_n} \delta_{X_{in}} \quad \text{and} \quad \nu_{2,n} = \frac{1}{n - n\theta_n} \sum_{n\theta_n < i \le n} \delta_{X_{in}}.$$

The existence of a probability space $(\Omega, \mathscr{A}, P)$ which is rich enough to carry all random variables is ensured by a canonical construction [cf. Ferger (1994a)]. Now, let F_n^* be the conditional distribution function of $|\xi_n^*|$ given $\mathbf{X}_n$, that is,

$$F_n^*(x) = P(|\xi_n^*| \le x | \mathbf{X}_n), \qquad x \in \mathbb{R}.$$

We define the bootstrap estimate of c by

$$c_n^* = c_n^*(\mathbf{X}_n) = (F_n^*)^{-1}(1 - \alpha) + 1.$$

Then the *bootstrap test* is given by

$$\tau_n^* = 1_{\{\theta_n^* \notin \Theta_0^{c_n^*/n}\}}.$$

We can prove the following counterparts of Theorem 1.1 and Theorem 1.2.

THEOREM 2.1. *Let $\mathscr{X}$ be a separable metric space and let K be bounded and uniformly continuous. If* (P) *holds, then, for all $\alpha \in (0, 1)$,*

$$\limsup_{n \to \infty} P_\theta(\tau_n^* \textit{ rejects } H_0) \le \alpha \qquad \forall\, \theta \in \operatorname{Int}(\Theta_0), \tag{2.1}$$

where $\operatorname{Int}(\Theta_0)$ *denotes the interior of* Θ_0. *Moreover,*

$$\lim_{n \to \infty} P_\theta(\tau_n^* \textit{ rejects } H_0) = 1 \qquad \forall\, \theta \notin \bar{\theta}_0. \tag{2.2}$$

THEOREM 2.2. *Suppose the assumptions of Theorem* 2.1 *and* (U) *hold and that $\Theta_0 \subset (0, 1)$. If $\varepsilon_n = \eta^{-1} c n^{-1}$, where $\eta \in (0, 1)$ is arbitrary, then*

$$\liminf_{n \to \infty} P_{\bar{\theta}_n}(\tau_n^* \textit{ rejects } H_0) \ge 2F\left(\left[\tfrac{1}{2}\tilde{\eta} c\right]\right) - 1, \tag{2.3}$$

where $\tilde{\eta} = \eta^{-1}(1 - \eta)$. Note that the lower bound converges to 1 *as $\eta \to 0$. Moreover, the rate n^{-1} is exact in the following sense: if ε_n tends to zero at a* *rate slower than* n^{-1}, *that is,* $n\varepsilon_n \to \infty$, *then*

$$\lim_{n\to\infty} P_{\bar\theta_n}(\tau_n^* \text{ rejects } H_0) = 1; \tag{2.4}$$

otherwise, that is, if $n\varepsilon_n \to 0$, *then, for all* $0 < p < 1$,

$$\limsup_{n\to\infty} P_{\bar\theta_n}(\tau_n^* \text{ rejects } H_0) \le 2 - \big(F([pc] - 1) + F([(1-p)c] - 1)\big). \tag{2.5}$$

REMARKS.

(i) Dümbgen (1991) used the same resampling method for constructing bootstrap confidence sets. In practice one can approximate the critical values c_n^* via Monte Carlo simulation.

(ii) If it is known in advance whether (1.2) or (1.3) holds, it is advantageous to deal with the one-sided variant of our test. In view of the remark in Section 1 we can w.l.o.g. assume that (1.2) holds. Then we need to replace θ_n by

$$\theta_n^+ = \arg\max_{t\in G_n} w(t) r_n(t).$$

3. Choice of kernels. As we pointed out in Section 1, antisymmetric kernels K play a particular role. So, before a set of data can be analyzed, an antisymmetric kernel K needs to be chosen such that, hopefully, $\lambda \equiv \lambda[K, \nu_1, \nu_2] \ne 0$. In many cases where some prior information about the type of change is available, we are able to find such kernels. At first let $\mathscr{X}$ be the real line. To detect a *change in location* we may take $K(x, y) = f(x) - f(y)$ with f strictly monotone. Indeed, if, for example, there is a positive shift after the time $[n\theta]$, then a strictly increasing f induces a positive λ. For a *change in scale* we may take the same f and set $K(x, y) = f(x^2) - f(y^2)$, obtaining a positive λ for a scaling factor with abolute value greater than 1. These kernels also work if ν_2 is an *ε-contaminated* ν_1, that is, $\nu_2 = (1-\varepsilon)\nu_1 + \varepsilon\nu_3$. In this case $\lambda[K, \nu_1, \nu_2] = \varepsilon\lambda[K, \nu_1, \nu_3]$, so that the above conclusions apply to ν_1 and ν_3, if ν_3 results from ν_1 by a change in location or scale. If we want to detect a *change in the kth (absolute) moment*, then $K(x, y) = x^k - y^k (= |x|^k - |y|^k)$ yields a nonzero λ. Finally, to detect a *change from* ν_1 *to a stochastically larger* ν_2, $K(x, y) = \operatorname{sign}(x - y)$ is suitable. We are also able to deal with models involving general sample spaces. Let $\mathscr{X}$ be any normal linear space. Take, for instance, $\mathscr{X} = \mathbb{R}^d$, $d \ge 1$, or $\mathscr{X} = C[0,1]$. Consider the *general location-scale model*:

$$X_{[n\theta]+1} =_{\mathscr{L}} T(X_{[n\theta]}) + x_0,$$

where $T\colon \mathscr{X} \to \mathscr{X}$ is some linear mapping and $x_0 \in \mathscr{X}$. For T equal to the identity map Id and $x_0 \ne 0$, this reduces to the location model and, for $T \ne Id$ and $x_0 = 0$, to the scale model. Suppose the expectation $\eta = EX_{[n\theta]} \in \mathscr{X}$ (in the sense of Pettis integral) exists. Let $f \in \mathscr{X}^*$, the dual space of $\mathscr{X}$, and define $K(x, y) = f(x - y)$. Then $\lambda = f(y_0)$ with $y_0 = (T - Id)(\eta) + x_0$. Provided that $y_0 \ne 0$ (which is always true in the location model as well as in the scale model if η is not an eigenvector of the eigenvalue 1), there exists, as a consequence of the Hahn–Banach theorem, an $f \in \mathscr{X}^*$ such that $\lambda = f(y_0) = \|y_0\| > 0$. For $\mathscr{X} = \mathbb{R}^d$, $d \geq 1$, we can construct such an f explicitly: let $y_0 = (y_1, \dots, y_d)$ and $D = \{1 \leq i \leq d\colon y_i \neq 0\} \neq \emptyset$. For each $i \in D$, take some $\lambda_i > 0$ and define $a_i = \lambda_i y_i^{-1}$. Then for $f(x) := \sum_{i \in D} a_i x_i$, $x = (x_1, \dots, x_d)$, one has $\lambda = f(y_0) = \sum_{i \in D} \lambda_i > 0$.

Finally, assume that we observe continuous random functions $X_i = (X_i(t)\colon 0 \leq t \leq 1)$ on $[0,1]$ such that the change has been caused by a *transformation in time* $t \in [0,1]$:

$$X_{[n\theta]+1} =_{\mathscr{L}} X_{[n\theta]} \circ \varphi,$$

where $\varphi\colon [0,1] \to [0,1]$ is continuous. For example, our data could be the height growth curves of children observed over a fixed period of time [see, e.g., Müller (1988)]. Recently there has been an increasing interest in the statistical literature on so-called functional data analysis. Here we are faced with data for which the ith observation is a real function rather than a point in the Euclidean space $\mathbb{R}^d$ [cf., e.g., Rice and Silverman (1991)]. If $EX_{[n\theta]}^2(t) = h(t)$, $0 \leq t \leq 1$, and $\int_0^1 (h(\varphi(t)) - h(t))\, dt \neq 0$, one can take $K(x,y) = \int_0^1 (x^2(t) - y^2(t))\, dt$, since then $\lambda = \int_0^1 (h(\varphi(t)) - h(t))\, dt \neq 0$.

4. Proofs. The proofs of our theorems basically rely on Theorem 4.1. For its formulation we need to generalize our original model slightly. Let $X_{1n}, \dots, X_{nn}$, $n \in \mathbb{N}$, be a triangular array of rowwise independent random elements defined on a probability space $(\Omega, \mathscr{A}, P)$ with values in $(\mathscr{X}, \mathscr{B})$. Suppose there exist sequences $(\nu_{1,n})$ and $(\nu_{2,n})$ of distributions and a sequence $(\bar{\theta}_n)$ in $(0,1)$ such that X_{in} has distribution $\nu_{1,n}$ for $1 \leq i \leq n\bar{\theta}_n$ and distribution $\nu_{2,n}$ for $n\bar{\theta}_n < i \leq n$. If necessary we will write $P = P_{(\bar{\theta}_n)}$ or simply $P = P_{\bar{\theta}_n}$ in order to stress the fact that P depends on the sequence $(\bar{\theta}_n)$. We assume that the following *stability condition* holds:

$$\text{(S)} \qquad \nu_{1,n} \to_w \nu_1, \qquad \nu_{2,n} \to_w \nu_2, \qquad \bar{\theta}_n \to \theta \in (0,1),$$

where $\to_w$ denotes convergence in the weak topology on the space of probability measures on $(\mathscr{X}, \mathscr{B})$. Moreover, let Z be a two-sided random walk on $\mathbb{Z}$ defined by

$$Z(k) = \begin{cases} \sum_{j=1}^{k} H(Y_j), & k \geq 0, \\ -\sum_{j=k+1}^{0} H(X_j), & k < 0. \end{cases}$$

Here, $(X_i)_{i \in \mathbb{Z}}$ and $(Y_i)_{i \in \mathbb{Z}}$ are two independent sequences of i.i.d. random elements with $X_i \sim \nu_1$ and $Y_i \sim \nu_2$. The mapping H is given by

$$H(z) = (1-\theta)\int K(x,z)\,\nu_2(dx) - \theta \int K(z,y)\,\nu_1(dy), \qquad z \in \mathscr{X}.$$

Finally, let

$$Y(k) = w(\theta)[Z(k) - m(k)] + \Delta(k), \qquad k \in \mathbb{Z},$$

where

$$m(k) = \begin{cases} [(1-\theta)\tau - \theta\lambda]k, & k \geq 0, \\ [(1-\theta)\lambda - \theta\mu]k, & k < 0, \end{cases}$$

is the expectation $EZ(k)$ and

$$\Delta(k) = w'(\theta)\lambda\theta(1-\theta)k + w(\theta)m(k), \qquad k \in \mathbb{Z},$$

is a drift function. The following sign variable is only introduced for a convenient presentation of our results:

$$\sigma = \begin{cases} 1, & \text{if (1.2) holds,} \\ -1, & \text{if (1.3) holds.} \end{cases}$$

Notice that the gradient of the drift function Δ coincides with the right derivative $\rho'(\theta+)$ or the left derivative $\rho'(\theta-)$ according as k is positive or negative. Thus, by (P), σY is a two-sided centered random walk with a linear negative drift. By the strong law of large numbers,

$$\lim_{k \to \pm\infty} \sigma Y(k) = -\infty \quad \text{with probability 1,}$$

whence the following random variables are well defined almost surely:

$$T_{\min} = \min\{k_0 \in \mathbb{Z}: \sigma Y(k) \leq \sigma Y(k_0)\ \forall\, k \in \mathbb{Z}\}$$

and

$$T_{\max} = \max\{k_0 \in \mathbb{Z}: \sigma Y(k) \leq \sigma Y(k_0)\ \forall\, k \in \mathbb{Z}\},$$

that is, $T_{\min}$ and $T_{\max}$ are the minimal and the maximal maximizers of σY.

THEOREM 4.1. *Let $\mathscr{X}$ be a separable metric space and let K be bounded and uniformly continuous. Suppose* (S) *and* (P) *hold. Then, for all $z \in \mathbb{Z}$,*

$$\begin{aligned} P(T_{\max} \leq z) &\leq \liminf_{n\to\infty} P\left(n\theta_n - \left[n\bar{\theta}_n\right] \leq z\right) \\ &\leq \limsup P\left(n\theta_n - \left[n\bar{\theta}_n\right] \leq z\right) \\ &\leq P(T_{\min} \leq z). \end{aligned} \tag{4.1}$$

If $\nu_{1,n} = \nu_1$, $\nu_{2,n} = \nu_2$ and $\bar{\theta}_n = \theta$ for all $n \in \mathbb{N}$, then the conditions can be weakened to the following: $(\mathscr{X}, \mathscr{B})$ *is a measurable space and* (M) *and* (P) *hold. Moreover, if in addition* (U) *holds, then*

$$\xi_n = n\theta_n - \left[n\bar{\theta}_n\right] \to_{\mathscr{L}} T_{\min}. \tag{4.2}$$

REMARKS.

(i) A proof of Theorem 4.1 is given in Ferger (1994b).

(ii) It is easy to see that the uniqueness condition (U) is fulfilled if $\nu_1 \circ H^{-1}$ and $\nu_2 \circ H^{-1}$ have no atoms. This is true in the location, the scale, the contamination and the moment models provided ν_1 is atomless (as well as ν_2 in the case of the moment model) and K is such as described in Section 3. Also note that (U) is equivalent with $T_{\max} = T_{\min}$ almost surely.

PROOF OF THEOREMS 1.1 AND 2.1. Let $\theta \in \Theta_0$. Note that

$$\{\tau_n \text{ rejects } H_0\} \subseteq \{|\theta_n - a| \ge c/n\} \qquad \forall\, a \in \Theta_0.$$

Since $\theta \in \Theta_0$ and c is an integer, it follows that

$$P_\theta(\tau_n \text{ rejects } H_0) \le P_\theta(|n\theta_n - [n\theta]| \ge c).$$

Hence (1.5) follows from (1.4) and (4.1). To prove (1.6), take an arbitrary $\theta \notin \overline{\Theta}_0$ and define $\varepsilon_0 = d(\theta, \overline{\Theta}_0)$, so that $\varepsilon_0 > 0$. Observe that

$$\begin{aligned} d(\theta_n, \Theta_0) &\ge d\left(\frac{[n\theta]}{n}, \overline{\Theta}_0\right) - \left|\theta_n - \frac{[n\theta]}{n}\right| \quad (\forall\, n \in \mathbb{N}) \\ &\ge \frac{2}{3}\varepsilon_0 - \left|\theta_n - \frac{[n\theta]}{n}\right|, \end{aligned}$$

for all n exceeding some $n_0(\varepsilon_0) \in \mathbb{N}$, by continuity of $d(\cdot, \overline{\Theta}_0)$. Thus, for all $n \ge n_0(\varepsilon_0)$ with $cn^{-1} \le \frac{1}{6}\varepsilon_0$,

$$\left\{\left|\theta_n - \frac{[n\theta]}{n}\right| < \frac{1}{2}\varepsilon_0\right\} \subseteq \{\theta_n \notin \Theta_0^{c/n}\}. \tag{4.3}$$

By Ferger [(1994b), Theorem 2.2],

$$\theta_n - \frac{[n\theta]}{n} = o(1) \quad \text{with } P_\theta\text{-probability } 1. \tag{4.4}$$

Therefore (1.6) follows from (4.3) and (4.4). As to the proof of Theorem 2.1, note that

$$\{\tau_n^* \text{ rejects } H_0\} \subseteq \{|\theta_n^* - \theta_n| \ge c_n^*/n\} \cup \{\theta_n \notin \Theta_0\}.$$

Consequently,

$$\begin{aligned} &\limsup_{n\to\infty} P_\theta(\tau_n^* \text{ rejects } H_0) \\ &\quad \le \limsup_{n\to\infty} P_\theta(|\theta_n^* - \theta_n| \ge c_n^*/n) + \limsup_{n\to\infty} P_\theta(\theta_n \notin \Theta_0) = \mathrm{I} + \mathrm{II}. \end{aligned}$$

From Fatou's lemma and (4.4) we can conclude that II is equal to zero if $\theta \in \Theta_0$. Moreover, by the definition of c_n^*,

$$I = \limsup_{n\to\infty} E(P_0(|\xi_n^*| \ge c_n^* | \mathbf{X}_n)) \le \alpha,$$

which proves (2.1). To prove (2.2), let $\theta \notin \overline{\Theta}_0$. Recall that $\varepsilon_0 = d(\theta, \overline{\Theta}_0) > 0$. Similarly as above we have that, for all $n \ge n_0(\varepsilon_0)$,

$$\left\{|\theta_n^* - \theta_n| \le \frac{1}{4}\varepsilon_0, \left|\theta_n - \frac{[n\theta]}{n}\right| \le \frac{1}{4}\varepsilon_0, \frac{c_n^*}{n} \le \frac{1}{6}\varepsilon_0\right\} \subseteq \{\theta_n^* \notin \Theta_0^{c_n^*/n}\}.$$

Since $\{c_n^*/n \le \frac{1}{6}\varepsilon_0\} = \{F_n^*(\frac{1}{6}n\varepsilon_0 - 1) \ge 1 - \alpha\}$ we may infer with Fatou's lemma and (4.4) that

$$\begin{aligned} &\liminf_{n\to\infty} P_\theta(\tau_n^* \text{ rejects } H_0) \\ &\quad \ge \liminf_{n\to\infty} P_\theta\big(|\theta_n^* - \theta_n| \le \tfrac{1}{4}\varepsilon_0\big) \\ &\qquad + P_\theta\Big(\liminf_{n\to\infty}\big\{F_n^*\big(\tfrac{1}{6}n\varepsilon_0 - 1\big) \ge 1 - \alpha\big\}\Big) - 1. \end{aligned} \tag{4.5}$$

Set $\Omega_0 = \{\theta_n \to \theta,\ \nu_{1,n} \to \nu_1,\ \nu_{2,n} \to \nu_2 \text{ as } n \to \infty\}$. Then, by (4.4) and Ferger [(1994a), Lemma 3.2], $P_\theta(\Omega_0) = 1$. By (4.1),

$$\liminf_{n\to\infty} F_n^*(z) \ge F_{\max}(z) \qquad \forall\, z \in \mathbb{Z}, \forall\, \omega \in \Omega_0.$$

Now, $\frac{1}{6}n\varepsilon_0 - 1 \to \infty$ as $n \to \infty$, so that

$$P_\theta\Big(\liminf_{n\to\infty}\big\{F_n^*\big(\tfrac{1}{6}n\varepsilon_0 - 1\big) \ge 1 - \alpha\big\}\Big) = 1. \tag{4.6}$$

Another application of Fatou's lemma and (4.1) yields

$$\begin{aligned} &\liminf_{n\to\infty} P_\theta\big(|\theta_n^* - \theta_n| \le \tfrac{1}{4}\varepsilon_0\big) \\ &\quad \ge \int_{\Omega_0} \liminf_{n\to\infty} P_\theta\big(n|\theta_n^* - \theta_n| \le \tfrac{1}{4}\varepsilon_0 n \big| \mathbf{X}_n\big)\, dP_\theta \ge P_\theta(\Omega_0) = 1. \end{aligned} \tag{4.7}$$

Combine (4.5)–(4.7) to get (2.2). □

PROOF OF THEOREMS 1.2 AND 2.2. It is well known that the function $d(\cdot, \Theta_0)$ is Lipschitz continuous, which ensures that

$$d(\theta_n, \Theta_0) \ge d\big(\bar\theta_n, \Theta_0\big) - |\theta_n - \bar\theta_n| = \varepsilon_n - |\theta_n - \bar\theta_n|.$$

Therefore,

$$\begin{aligned} &\big\{|\theta_n - \bar\theta_n| \le (1 - \eta)\varepsilon_n\big\} \\ &\quad \subseteq \big\{d(\theta_n, \Theta_0) \ge \eta\varepsilon_n\big\} \quad [\forall\, \eta \in (0,1)] \\ &\quad = \big\{\tau_n \text{ rejects } H_0\big\} \quad \text{by definition of } \varepsilon_n. \end{aligned} \tag{4.8}$$

By (4.1) it follows that

$$\begin{aligned} \liminf_{n\to\infty} P_{\bar\theta_n}(\tau_n \text{ rejects } H_0) &\ge \liminf_{n\to\infty} P_{\bar\theta_n}\big(|\theta_n - \bar\theta_n| \le (1 - \eta)\varepsilon_n\big) \\ &= \liminf_{n\to\infty} P_{\bar\theta_n}\big(n\theta_n - n\bar\theta_n \le [\tilde\eta c]\big) \\ &\quad - \limsup_{n\to\infty} P_{\bar\theta_n}\big(n\theta_n - n\bar\theta_n \le -[\tilde\eta c] - 1\big) \\ &\ge F_{\max}([\tilde\eta c]) - F_{\min}(-[\tilde\eta c] - 1), \end{aligned}$$

which shows (1.7). For the proof of (1.8) set

$$\eta_n = cn^{-1}\varepsilon_n^{-1}.$$

Since $n\varepsilon_n \to \infty$ by assumption we have that $\eta_n \in (0,1)$ eventually. By (4.8),

$$P_{\bar{\theta}_n}(\tau_n \text{ rejects } H_0) \le P\big(|n\theta_n - n\bar{\theta}_n| \le (1-\eta_n)n\varepsilon_n\big).$$

However $(1-\eta_n)n\varepsilon_n$ converges to infinity, so that a further application of (4.1) yields (1.8). Recall the Lipschitz continuity of $d(\cdot,\Theta_0)$ to see that

$$d(\theta_n,\Theta_0) \le |\theta_n - \bar{\theta}_n| + d\big(\bar{\theta}_n,\Theta_0\big) = |\theta_n - \bar{\theta}_n| + \varepsilon_n.$$

Thus

$$P_{\bar{\theta}_n}(\tau_n \text{ rejects } H_0) \le P\big(|n\theta_n - n\bar{\theta}_n| \ge c - n\varepsilon_n\big),$$

which yields (1.9) upon noticing (4.1) and $n\varepsilon_n \to 0$. This completes the proof of Theorem 1.2. As to the proof of Theorem 2.2, set $\Omega_0 = \{\theta_n \to \theta, \nu_{1,n} \to \nu_2, \nu_{2,n} \to \nu_2\}$. Note that using Theorem 2.2 of Ferger (1994b) and similar arguments as in the proof of Lemma 3.2 of Ferger (1994a), one shows that

$$P_{(\bar{\theta}_n)}(\Omega_0) = 1,$$

if (S) holds. Hence, by (4.2), F_n^* converges weakly to F with $P_{(\bar{\theta}_n)}$-probability 1. By the lemma of Chung [(1974), page 133], the convergence even holds uniformly on $\mathbb{R}$. Especially, we can deduce that

$$c_n^* \to c \quad \text{with } P_{(\bar{\theta}_n)}\text{-probability } 1. \tag{4.9}$$

Since

$$d(\theta_n^*,\Theta_0) \ge \varepsilon_n - |\theta_n^* - \theta_n| - |\theta_n - \bar{\theta}_n|,$$

it holds that, for all $\eta \in (0,1)$,

$$\left\{|\theta_n^* - \theta_n| \le \frac{1}{2}(1-\eta)\varepsilon_n, |\theta_n - \bar{\theta}_n| \le \frac{1}{2}(1-\eta)\varepsilon_n\right\} \tag{4.10}$$

$$\subseteq \left\{d(\theta_n^*,\Theta_0) \ge \frac{c}{n}\right\}$$

$$\subseteq \{\tau_n^* \text{ rejects } H_0\} + \{c < c_n^*\}. \tag{4.11}$$

Note that c_n^* and c are integer-valued so that, by Fatou's lemma and (4.9),

$$\lim_{n\to\infty} P_{(\bar{\theta}_n)}(c \le c_n^*) = 0.$$

Therefore (4.10) implies that

$$\begin{aligned}
&\liminf_{n\to\infty} P(\tau_n^* \text{ rejects } H_0)\\
&\quad\ge \liminf_{n\to\infty} P\big(n|\theta_n^* - \theta_n| \le \tfrac{1}{2}\tilde{\eta}c\big) + \liminf_{n\to\infty} P\big(n|\theta_n - \bar{\theta}_n| \le \tfrac{1}{2}\tilde{\eta}c\big) - 1\\
&\quad\ge \int_{\Omega_0} \liminf_{n\to\infty} P\big(n|\theta_n^* - \theta_n| \le \tfrac{1}{2}\tilde{\eta}c|\mathbf{X}_n\big)\,dP\\
&\qquad + \liminf_{n\to\infty} P\big(n|\theta_n - \bar{\theta}_n| \le \tfrac{1}{2}\tilde{\eta}c\big) - 1\\
&\quad= 2F\big(\big[\tfrac{1}{2}\tilde{\eta}c\big]\big) - 1.
\end{aligned}$$

This proves (2.3). Using (4.10) the proof of (2.4) is similar to that of (1.8). Finally, since

$$d(\theta_n^*, \Theta_0) \le \varepsilon_n + |\theta_n^* - \theta_n| + |\theta_n - \bar{\theta}_n|,$$

it follows that, for all $0 < p < 1$,

$$\begin{aligned}
&\limsup_{n\to\infty} P(\tau_n^* \text{ rejects } H_0)\\
&\quad\le \limsup_{n\to\infty} P\big(n|\theta_n^* - \theta_n| \ge p(c_n^* - n\varepsilon_n)\big)\\
&\qquad + \limsup_{n\to\infty} P\big(n|\theta_n - \bar{\theta}_n| \ge (1-p)(c_n^* - n\varepsilon_n)\big)\\
&\quad\le \int_{\Omega_0} \limsup_{n\to\infty} P\big(n|\theta_n^* - \theta_n| \ge p(c_n^* - n\varepsilon_n)|\mathbf{X}_n\big)\, dP\\
&\qquad + \limsup_{n\to\infty} P\big(n|\theta_n - \bar{\theta}_n| \ge (1-p)(c_n^* - n\varepsilon_n)\big).
\end{aligned}$$

An application of (4.2) yields (2.5) upon noticing that $c_n^* \to c$ on Ω_0 and $n\varepsilon_n \to 0$. □

REFERENCES

BHATTACHARYA, P. K. and BROCKWELL, P. J. (1976). The minimum of an additive process with applications to signal estimation and storage theory. *Z. Wahrsch. Verw. Gebiete* **37** 51–75.

BHATTACHARYA, P. K. and FRIERSON, D., JR. (1981). A nonparametric control chart for detecting small disorders. *Ann. Statist.* **9** 544–554.

CARLSTEIN, E. (1988). Nonparametric change-point estimation. *Ann. Statist.* **16** 188–197.

CHUNG, K. L. (1974). *A Course in Probability Theory.* Academic Press, San Diego.

COBB, W. G. (1978). The problem of the Nile: conditional solution to a change-point problem. *Biometrika* **65** 243–251.

CSÖRGŐ, M. and HORVÁTH, L. (1987). Nonparametric tests for the change-point problem. *J. Statist. Plann. Inference* **17** 1–9.

CSÖRGŐ, M. and HORVÁTH, L. (1988a). Nonparametric methods for change-point problems. In *Handbook of Statistics* **7**. *Quality Control and Reliability* (P. R. Krishnaiah and C. R. Rao, eds.) 403–425. North-Holland, Amsterdam.

CSÖRGŐ, M. and HORVÁTH, L. (1988b). Invariance principles for change-point problems. *J. Multivariate Anal.* **27** 151–168.

DARKHOVSKH, B. S. (1976). A nonparametric method for the a posteriori detection of the "disorder" time of a sequence of independent random variables. *Theory Probab. Appl.* **21** 178–183.

DESHAYES, J. and PICARD, D. (1981). Convergence de processus a double indices: application aux tests de rupture dans un modèle. *C. R. Acad. Sci. Paris* **292** 449–452.

DÜMBGEN, L. (1991). The asymptotic behavior of some nonparametric change-point estimators. *Ann. Statist.* **19** 1471–1495.

FERGER, D. (1994a). Asymptotic distribution theory of change-point estimators and confidence intervals based on bootstrap approximation. *Mathematical Methods of Statistics* **3** 362–378.

FERGER, D. (1994b). Change-point estimators based on weighted empirical processes. Unpublished manuscript. (Preprint, Univ. Giessen.)

FERGER, D. (1994c). Change-point estimators in case of small disorders. *J. Statist. Plann. Inference* **40** 33–49.

FERGER, D. and STUTE, W. (1992). Convergence of change-point estimators. *Stochastic Process. Appl.* **42** 345–351.

HINKLEY, D. V. (1970). Inference about the changepoint in a sequence of random variables. *Biometrika* **57** 1–17.

LOMBARD, F. (1987). Rank tests for change-point problems. *Biometrika* **74** 615–624.

LORDEN, G. (1971). Procedures for reacting to a change in distribution. *Ann. Math. Statist.* **42** 1897–1908.

MAGUIRE, B. A., PEARSON, E. S. and WYNN, A. H. A. (1952). The time intervals between industrial accidents. *Biometrika* **38** 168–180.

MÜLLER, H. G. (1988). *Nonparametric Regression Analysis of Longitudinal Data. Lecture Notes in Statist.* **46**. Springer, Berlin.

PAGE, E. (1954). Continuous inspection schemes. *Biometrika* **42** 100–115.

PAGE, E. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika* **42** 523–526.

PETTITT, A. N. (1979). A nonparametric approach to the change-point problem. *J. Roy. Statist. Soc. Ser. C* **28** 126–135.

RICE, J. A. and SILVERMAN, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc. Ser. B* **53** 233–243.

SMITH, A. F. M. (1980). Change-point problems: approaches and applications. *Trabajos de Estadistica y de Investigación Operativa* **31** 83–98.

VARADARAJAN, V. S. (1958). Weak convergence of measures on separable metric spaces. *Sankhyā* **19** 15–22.

WORSLEY, K. J. (1986). Confidence regions and tests for a changepoint in a sequence of exponential family random variables. *Biometrika* **73** 91–104.

MATHEMATISCHES INSTITUT
JUSTUS-LIEBIG-UNIVERSITÄT
GIESSEN
ARNDTSTR. 2
D-35392 GIESSEN
GERMANY