Bernoulli **7**(3), 2001, 487–506

Analysis of change-point estimators under the null hypothesis

DIETMAR FERGER

Department of Mathematics, Dresden University of Technology, Mommsenstr. 13, D-01062 Dresden, Germany. E-mail: dietmar.ferger@math.tu-dresden.de

We consider estimators for the change-point in a sequence of independent observations. These are defined as the maximizing points of weighted U-statistic type processes. Our investigations focus on the behaviour of the estimators in the case of independent and identically distributed random variables (null hypothesis of no change), but contiguous alternatives in the sense of Oosterhoff and van Zwet are also taken into account. If the weight functions belong to the Chibisov–O'Reilly class we derive convergence in distribution, including a special Berry–Esseen result. The limit variable is the almost sure unique maximizing point of a weighted (standard or reflected) Brownian bridge with drift. For general weight functions the limiting null distribution is analytically not known. However, in the special case where no weight functions are involved it is known that the maximizer of a standard Brownian bridge is uniformly distributed on the unit interval. A corresponding result for the reflected Brownian bridge seems to be unknown in the literature. In this paper we fill this gap and actually compute the common density of the maximum and its location for a reflected Brownian bridge. From this one can find the density of the maximizer, which analytically can be expressed in terms of a series. In a special case even the finite sample size distribution of our estimator is established. Besides distributional results, we also determine the almost sure set of cluster points.

Keywords: Berry–Esseen estimates; change-point estimation; contiguous alternatives; limiting null distribution; maximizer of weighted Brownian bridges; sets of cluster points

1. Introduction

We consider a triangular array $X_{1n}, \ldots, X_{nn}$, $n \geqslant 2$, of rowwise independent random elements defined on a probability space $(\Omega, \mathscr{A}, P)$ with values in a measurable space $(\mathscr{X}, \mathscr{F})$. Suppose that the underlying distribution $\mathscr{L}(X_{in})$ of X_{in} changes at an unknown point $\tau = [n\theta]$ from ν_{1n} to some $\nu_{2n} \neq \nu_{1n}$, where $\theta \in (0, 1]$. Thus it is $\mathscr{L}(X_{in}) = 1_{\{i \leqslant \tau\}}\nu_{1n} + 1_{\{i > \tau\}}\nu_{2n}$ for $1 \leqslant i \leqslant n$ and $n \in \mathbb{N}$. Knowing nothing about ν_{1n} and ν_{2n}, we wish to estimate the change-point θ. The analysis of change-point estimators in a nonparametric framework has been of increasing interest in the last decade. A comprehensive review is given in the monographs of Brodsky and Darkhovsky (1993) and Csörgő and Horváth (1997). Commonly, the results for estimators of θ are concerned with the case of an actual change $(0 < \theta < 1)$, whereas the case of no change $(\theta = 1)$ has hardly been investigated. Indeed, hitherto only a few contributions have addressed this problem: see Ferger (1996), Gombay and Horváth (1996), Hušková (1996) or Lombard and Hart (1994). They prove convergence in distribution to a non-degenerate limit variable.

1350–7265

Note that we need to consider double-indexed random variables X_{in} in order to enable asymptotic investigations in the case of the change-point alternative $H_1: \theta < 1$. Since in statistics the null and alternative hypotheses usually constitute a common model, we prefer also in the case of no change (null hypothesis $H_0: \theta = 1$) to state our results in terms of arrays. Moreover, it is the nature of some change-point problems which requires an array as an appropriate stochastic model. By way of illustration, consider the following examples taken from Bhattacharya and Brockwell (1976) and Bhattacharya and Frierson (1981).

Example 1.1. A machine produces items, and the process continues uninterrupted until a problem occurs. The machine is assumed to be adjusted at regular intervals. Between two successive adjustments we therefore take random samples $X_{1n}, \ldots, X_{k_n n}$ of size $k_n \in \mathbb{N}$. Then X_{in} represents the value of the ith observation after the nth adjustment and $\theta = 1$ ($\theta < 1$) describes a production process which is under control (out of control). Note that due to the adjustments it is reasonable to assume that $(X_{1n}, \ldots, X_{k_n n})$, $n \in \mathbb{N}$, is a sequence of independent vectors. In this paper we only consider $k_n = n$, but the extension to the general case merely requires changes in the notation.

Example 1.2. One observes random variables $X_{1n}, \ldots, X_{nn}$, where X_{in} is the sum of a noise component Y_{in} and a possible signal a_{in}. Suppose the noises $Y_{1n}, \ldots, Y_{nn}$ form a sequence of independent and identically distributed (i.i.d.) random variables with distribution ν and that $a_{in} = 0$ for $1 \leqslant i \leqslant [n\theta]$ and $a_{in} = r_n$ for $[n\theta] < i \leqslant n$, where $r_n \to 0$ as $n \to \infty$. So, up to time $[n\theta]$ no signals have been sent, whereas after $[n\theta]$ we have received faint signals.

In this paper we study the estimators

$$\theta_n = n^{-1} \operatorname*{arg\,max}_{1 \leqslant k < n} w\left(\frac{k}{n}\right) \left| \sum_{i=k+1}^{n} \sum_{j=1}^{k} K(X_{in}, X_{jn}) \right|$$

and

$$\theta_n^+ = n^{-1} \operatorname*{arg\,max}_{1 \leqslant k < n} w\left(\frac{k}{n}\right) \sum_{i=k+1}^{n} \sum_{j=1}^{k} K(X_{in}, X_{jn}),$$

under the null hypothesis H_0 of no change ($\theta = 1$). Here $w: (0, 1) \to (0, \infty)$ is a positive weight function and $K: \mathscr{X}^2 \to \mathbb{R}$ is a measurable and antisymmetric mapping (kernel). By convention $\operatorname{arg\,max}_{t \in T} f(t)$ denotes the smallest maximizer of a function $f: T \to \mathbb{R}$, $T \subseteq \mathbb{R}$, with existing $\max f(t)$. For asymptotic properties of these estimators under the alternative $0 < \theta < 1$, see Ferger (1994a; 2001).

The paper is organized as follows. In section 2 we prove convergence in distribution of (θ_n) and (θ_n^+) provided w is a Chibisov–O'Reilly function. The distributions of the limit variables $\tau_w = \operatorname{arg\,max}_{0<t<1} w(t)|B_0(t)|$ and $\tau_w^+ = \operatorname{arg\,max}_{0<t<1} w(t)B_0(t)$, where B_0 denotes a Brownian bridge, are not known for general w. But for the special weight function $w = 1$ one can identify the limit distributions. Indeed, the maximizer $\tau_1^+ = \operatorname{arg\,max}_{0<t<1} B_0(t)$ of standard Brownian bridge is known to be uniformly distributed on (0, 1); see Ferger (1995) or Csörgő and Horváth (1997). Unlike the standard case, the distribution of the maximizing point $\tau_1 = \arg\max_{0<t<1}|B_0(t)|$ of a reflected Brownian bridge seems to be unknown in the literature. By determining the common density of $\arg\max_{0<t<1}|B_0(t)|$ and $\max_{0<t<1}|B_0(t)|$ we fill this gap. In particular, the density of τ_1 admits an explicit representation in terms of a series. For general w we at least know that τ_w and τ_w^+ are continuous random variables. For the weight function $w(t) = (t(1-t))^{-1/2}$, which is extreme in so far as w is not a Chibisov–O'Reilly function, there is still convergence in distribution, but the limit now is a *Bernoulli*$(\frac{1}{2})$ variable. Roughly speaking, this means that the estimator also under the null hypothesis of no change correctly indicates the i.i.d. situation, which can be described through $\theta = 1$ as well as $\theta = 0$. This result is actually due to Csörgő and Horváth (1997) and Lombard and Hart (1994), but for the sake of completeness a more elaborate proof is given.

In Section 3 we consider the one-sided estimator θ_n^+ pertaining to $w = 1$. As explained above, $\theta_n^+ \xrightarrow{\mathscr{L}} U(0, 1)$, which by Pólya's theorem is equivalent to

$$\sup_{0\leqslant x\leqslant 1} |P(\theta_n^+ \leqslant x) - x| \to 0, \qquad n \to \infty. \tag{1.1}$$

Under a uniform moment condition on K, rates of convergence in (1.1) are established. In Section 4 we determine the almost sure set of cluster points of the sequences (θ_n) and (θ_n^+). Finally, in Section 5 a necessary and sufficient condition is presented under which (a slight modification of) θ_n^+ induced by $w = 1$ is uniformly distributed on the grid $\{kn^{-1} : 0 \leqslant k \leqslant n-1\}$ for finite sample size $n \in \mathbb{N}$. Section 6 contains two technical results which are needed in the proofs.

2. Convergence in distribution

In this section we show that, under the null hypothesis of no change, θ_n and θ_n^+ converge in distribution to the almost surely (a.s.) unique maximizer of a weighted reflected Brownian bridge and a weighted standard Brownian bridge, respectively. For the general weights under consideration one can approximate the limit distribution by the Monte Carlo method. In the special case of no weights ($w = 1$) it is possible to give explicit analytical expressions.

Our first result deals with the special case $\nu_{1n} = \nu$ for all $n \in \mathbb{N}$; that is, here the common distribution of $X_{1n}, \ldots, X_{nn}$ may not depend on n. However, if the sample space $\mathscr{X}$ is equal to the real line $\mathbb{R}$, then we can get rid of this restriction as long as ν_{1n} is contiguous to some ν in the sense of Oosterhoff and van Zwet (1979).

In the following let $\mathscr{W}$ denote the class of continuous functions $w: (0, 1) \to (0, \infty)$, which are monotone decreasing in a neighbourhood of zero and monotone increasing in a neighbourhood of one.

Theorem 2.1. *For each $n \in \mathbb{N}$, let $X_{1n}, \ldots, X_{nn}$ be i.i.d. random elements in $(\mathscr{X}, \mathscr{F})$ with common distribution ν. Assume K is antisymmetric with*

$$\int |K|^p \, \mathrm{d}\nu \otimes \nu < \infty, \quad p > 2, \qquad \sigma^2 = \int \left[\int K(x, y)\nu(\mathrm{d}y) \right]^2 \nu(\mathrm{d}x) > 0. \tag{2.1}$$

If $w \in \mathscr{W}$ satisfies the Chibisov–O'Reilly condition

$$\int_0^1 (t(1-t))^{-1} \exp(-cw(t)^{-2}(t(1-t))^{-1}) \mathrm{d}t < \infty, \qquad \forall c > 0, \tag{2.2}$$

then

$$\theta_n \xrightarrow{\mathscr{L}} \tau_w = \operatorname*{arg\,max}_{0<t<1} w(t)|B_0(t)| \tag{2.3}$$

and

$$\theta_n^+ \xrightarrow{\mathscr{L}} \tau_w^+ = \operatorname*{arg\,max}_{0<t<1} w(t)B_0(t), \tag{2.4}$$

where B_0 denotes a Brownian bridge. The maximizing points τ_w and *τ_w^+ are a.s. unique.*

Proof. Put $S_n(t) = \sum_{i=[nt]+1}^{n} \sum_{j=1}^{[nt]} K(X_{in}, X_{jn})$, $0 \leqslant t \leqslant 1$, and

$$\Gamma_n(t) = \begin{cases} \sigma^{-1} w(t) n^{-3/2} S_n(t), & 0 < t < 1, \\ 0, & t \in \{0, 1\}. \end{cases}$$

By (2.1) and (2.2) we can apply Corollary 4.1 of Csörgő and Horváth (1988), which says that there exists a sequence $(B_0^{(n)})$ of Brownian bridges such that

$$\sup_{0<t<1} w(t)|\sigma^{-1} n^{-3/2} S_n(t) - B_0^{(n)}(t)| = o_P(1). \tag{2.5}$$

Since $w \in \mathscr{W}$ meets requirement (2.2), Corollary 1.2 of Csörgő and Horváth (1993, p. 189) states that there is an $\Omega_0 \in \mathscr{A}$ with $P(\Omega_0) = 1$ such that

$$\lim_{t \to 0} w(t) B_0^{(n)}(t) = \lim_{t \to 1} w(t) B_0^{(n)} = 0, \qquad \forall n \in \mathbb{N}, \forall \omega \in \Omega_0. \tag{2.6}$$

For those $\omega \in \Omega_0$ we define the continuous process

$$\Gamma_0^{(n)}(t) = \begin{cases} w(t) B_0^{(n)}(t), & 0 < t < 1, \\ 0, & t \in \{0, 1\}. \end{cases}$$

By construction the stochastic processes Γ_n and $\Gamma_0^{(n)}$ are random elements in the Skorokhod space $D = D[0, 1]$ endowed with the Skorokhod metric s. According to (2.5),

$$\sup_{0 \leqslant t \leqslant 1} |\Gamma_n(t) - \Gamma_0^{(n)}(t)| = \sup_{0<t<1} |\Gamma_n(t) - \Gamma_0^{(n)}(t)| = o_P(1).$$

Thus, by Slutsky's theorem, we obtain

$$\Gamma_n \xrightarrow{\mathscr{L}} \Gamma_0^{(1)} \qquad \text{as } n \to \infty \text{ in } D[0, 1]. \tag{2.7}$$

The mapping $\psi_n : D \to [0, 1]$ defined by

$$\psi_n(f) = \underset{1 \leqslant k < n-1}{\arg\max} f\left(\frac{k}{n}\right), \qquad f \in D,$$

entails the representations

$$\theta_n = \psi_n(|\Gamma_n|), \qquad \theta_n^+ = \psi_n(\Gamma_n). \tag{2.8}$$

In view of the extended continuous mapping theorem (Billingsley, 1968, Theorem 5.5) we wish to extend the 'argmax functional' to the space D in a suitable way. Since $f \in D$ possibly does not possess a maximizing value, we introduce for all $f \in D$ the set

$$S(f) = \left\{ 0 \leqslant u \leqslant 1 : f(u) = \sup_{0 \leqslant s \leqslant 1} f(s) \text{ or } f(u-) = \sup_{0 \leqslant s \leqslant 1} f(s) \right\}.$$

By Lemma 6.1, $S(f)$ is a non-empty, closed set, whence the mapping $\psi : D \to [0, 1]$ with

$$\psi(f) = \min S(f), \qquad f \in D,$$

is well defined. Clearly, if f is continuous then $S(f)$ is the set of maximizers, so that $\psi(f)$ is the smallest maximizer. In particular, because $\Gamma_0^{(1)}$ is a.s. continuous by (2.6), we have

$$\tau_w = \psi(|\Gamma_0^{(1)}|) \text{ a.s.} \quad \text{and} \quad \tau_w^+ = \psi(\Gamma_0^{(1)}) \text{ a.s.} \tag{2.9}$$

To complete the proof, consider

$$E = \{ f \in D : \exists (f_n) \subseteq D, f_n \to_s f, \psi_n(f_n) \not\to \psi(f) \}.$$

By Lemma 6.1, the set $\hat{C}$ of continuous functions on [0, 1] with unique maximizers is contained in the complement of E in D: $\hat{C} \subseteq D \backslash E$. Example 2.7 of Ferger (1999) shows that $P(|\Gamma_0^{(1)}| \in \hat{C}) = 1 = P(\Gamma_0^{(1)} \in \hat{C})$. Consequently, the extended continuous mapping theorem is applicable, which by (2.7)–(2.9) gives the desired result. □

Remark 2.2. (i) An essential tool in the above proof is Corollary 4.1 of Csörgő and Horváth (1988). However, their results are only formulated for sequences $X_1, X_2, \ldots$ of i.i.d. real-valued random variables rather than for arrays $X_{1n}, \ldots, X_{nn}$, $n \in \mathbb{N}$, of rowwise i.i.d. $\mathscr{X}$-valued random elements. But checking the proofs shows that all statements of their Sections 2 and 4 remain valid in the general case.

(ii) If $w \in \mathscr{W}$ is bounded then it suffices to require the existence of the second moment ($p = 2$) in (2.1). Indeed, then one can apply Theorem 4.1 of Csörgő and Horváth (1988) to show that $\Gamma_n \xrightarrow{\mathscr{L}} wB_0$ in $D[0, 1]$. The rest of the proof remains the same.

(iii) Note that the first part of condition (2.1) is fulfilled for all distributions ν whenever a bounded kernel is chosen. Observe that most frequently K is of the type $K(x, y) = a(x) - a(y)$ with some mapping $a : \mathscr{X} \to \mathbb{R}$. Then $\sigma^2 = \text{var}\{a(X_{11})\}$ and the second part of (2.1) excludes the degenerate case that our transformed observations $a(X_{in})$ are all constant with probability one. Many examples for an appropriate choice of K are given in Ferger (1994a; 1994c) and Ferger and Stute (1992).

If $\mathscr{X} = \mathbb{R}$ then the statements of Theorem 2.1 can be extended to arrays, where for all $n \in \mathbb{R}$ the nth row $X_{1n}, \ldots, X_{nn}$ of the array consists of independent random variables X_{in}

with distribution function (df) F_{in} which may depend on n and even on i. Here we have to assume that all F_{in}, $1 \leqslant i \leqslant n$, $n \in \mathbb{N}$, are absolutely continuous with respect to an arbitrary df F ($F_{in} \ll F$), and that the densities $\mathrm{d}F_{in}/\mathrm{d}F$ are determined by

$$\left\{\frac{\mathrm{d}F_{in}}{\mathrm{d}F}(F^{-1}(u))\right\}^{1/2} = 1 + \frac{1}{2\sqrt{n}} g\left(\frac{i}{n}, u\right) + a_{in}, \qquad 0 < u < 1, \tag{2.10}$$

where $g \in L_2([0, 1]^2)$ is bounded and

$$\int_0^1 g(t, u)\mathrm{d}u = 0, \qquad \forall t \in [0, 1].$$

Choosing

$$a_{in} = \left\{1 - \frac{1}{4n}\int_0^1 g^2\left(\frac{i}{n}, u\right)\mathrm{d}u\right\}^{1/2} - 1, \qquad 1 \leqslant i \leqslant n,$$

ensures that $\mathrm{d}F_{in}/\mathrm{d}F$ is a probability density; moreover, we see that

$$\max_{1 \leqslant i \leqslant n} |a_{in}| = O(n^{-1}).$$

Using the formula for the change of variable one obtains

$$F_{in}(x) = F(x) + \frac{1}{\sqrt{n}}\int_0^{F(x)} g\left(\frac{i}{n}, u\right)\mathrm{d}u + O(n^{-1}), \qquad x \in \mathbb{R}. \tag{2.11}$$

An important special case is given by g of the type

$$g(t, u) = h(u), \qquad 0 \leqslant t, u \leqslant 1,$$

with bounded $h \in L_2([0, 1])$ satisfying $\int_0^1 h(u)\mathrm{d}u = 0$. This corresponds to $X_{1n}, \ldots, X_{nn}$ being i.i.d. with common df F_n determined by

$$\left\{\frac{\mathrm{d}F_n}{\mathrm{d}F}(F^{-1}(u))\right\}^{1/2} = 1 + \frac{1}{2\sqrt{n}} h(u) + O(n^{-1}), \qquad 0 < u < 1, \tag{2.12}$$

which according to (2.11) entails the representation

$$F_n(x) = F(x) + \frac{1}{\sqrt{n}}\int_0^{F(x)} h(u)\mathrm{d}u + O(n^{-1}), \qquad x \in \mathbb{R}. \tag{2.13}$$

Notice that by Theorem 1 of Oosterhoff and van Zwet (1979) the sequence $F_{1n} \otimes \ldots \otimes F_{nn}$ is contiguous with respect to $F \otimes \ldots \otimes F$. The next result extends Theorem 2.1 to the case of real-valued observations.

Theorem 2.3. (i) *For all* $n \in \mathbb{N}$, *let* $X_{1n}, \ldots, X_{nn}$ *be independent real-valued random variables, where* X_{in} *have the respective dfs* F_{in} *determined by* (2.10). *Assume that* K *is antisymmetric with* (2.1) *and*

$$\sigma^2 = \int \left[\int K(x, y) F(\mathrm{d}y) \right]^2 F(\mathrm{d}x) \in (0, \infty).$$

If $w \in \mathscr{W}$ *satisfies the Chibisov–O'Reilly condition* (2.2)*, then*

$$\theta_n \xrightarrow{\mathscr{L}} \tau_{w,b} = \underset{0<t<1}{\arg\max}\, w(t)|\sigma B_0(t) + b(t)| \tag{2.14}$$

and

$$\theta_n^+ \xrightarrow{\mathscr{L}} \tau_{w,b}^+ = \underset{0<t<1}{\arg\max}\, w(t)[\sigma B_0(t) + b(t)], \tag{2.15}$$

where

$$b(t) = G(t) - tG(1), \qquad 0 \leqslant t \leqslant 1,$$

with

$$G(t) = \int_0^t \int_0^1 \int_{-\infty}^{\infty} g(s, u) K(x, F^{-1}(u)) F(\mathrm{d}x) \mathrm{d}u\, \mathrm{d}s.$$

The maximizers $\tau_{w,b}$ *and* $\tau_{w,b}^+$ *are a.s. unique.*

(ii) *If, in addition,* $X_{1n}, \ldots, X_{nn}$ *are i.i.d. with common df* F_n *determined by* (2.12)*, then*

$$\theta_n \xrightarrow{\mathscr{L}} \tau_w, \qquad \theta_n^+ \xrightarrow{\mathscr{L}} \tau_w^+. \tag{2.16}$$

Proof. Using the notation of the above proof, Theorem 3.4(b) of Szyszkowicz (1991) states that

$$\Gamma_n \xrightarrow{\mathscr{L}} \Gamma_0 := w[\sigma B_0 + b] \qquad \text{in } D[0, 1].$$

The same arguments following (2.7) prove (2.14) and (2.15) upon noticing that $P(|\Gamma| \in \hat{C}) = P(\Gamma_0 \in \hat{C}) = 1$ by Theorems 2.2 and 2.4 of Ferger (1999). As to (2.16), check that $b(t) = 0$ for all $t \in [0, 1]$ if $g(t, u) = h(u)$. □

Theorem 2.3 immediately also gives the asymptotic behaviour of θ_n and θ_n^+ under the alternative $H_1: \theta < 1$ with contiguous distributions. We make this more precise in the following:

Corollary 2.4. *For all* $n \in \mathbb{N}$*, let* $X_{1n}, \ldots, X_{nn}$ *be independent random variables such that, for some* $\theta \in (0, 1)$*,* X_{in} *has df* F *or* F_n*, respectively, according to* $i \leqslant [n\theta]$ *or* $i > [n\theta]$*. If* F_n *is determined by* (2.12)*, then*

$$\theta_n \xrightarrow{\mathscr{L}} \underset{0<t<1}{\arg\max}\, w(t)|\sigma B_0(t) + b^*(t)|$$

and

$$\theta_n^+ \xrightarrow{\mathscr{L}} \underset{0<t<1}{\arg\max}\, w(t)[\sigma B_0(t) + b^*(t)],$$

with

$$b^*(t) = \alpha \begin{cases} (1-\theta)t, & 0 \leq t \leq \theta, \\ \theta(1-t), & \theta \leq t \leq 1, \end{cases}$$

and

$$\alpha = \int_0^1 \int_{-\infty}^{\infty} h(u) K(F^{-1}(u), x) F(\mathrm{d}x)\mathrm{d}u.$$

Proof. Apply Theorem 2.3 to $g(t, u) = 1_{\{t>\theta\}} h(u)$. □

Note that in the above corollary the post-change df F_n converges to the pre-change df F with rate $n^{-1/2}$. If this rate in (2.12) is replaced by any slower rate $r_n \to 0$ – that is, if $r_n n^{1/2} \to \infty$ – then the asymptotic behaviour of θ_n becomes completely different. Indeed in this case, Theorem 1.1 of Ferger (1994a) states that under some regularity conditions,

$$nr_n^2(\theta_n - \theta) \xrightarrow{\mathscr{L}} \alpha^{-2} T, \tag{2.17}$$

provided $\alpha \neq 0$. Here, the limit variable T is the a.s. unique maximizer of a two-sided Brownian motion on $\mathbb{R}$ with a linear downward drift. This theorem also provides the Lebesgue density of T. The reader will find further limit theorems of the type (2.17), for instance, in Antoch and Hušková (1999), Bhattacharya (1987), Bhattacharya and Brockwell (1976) or Dümbgen (1991).

For general weight functions $w \in \mathscr{W}$, the distributions of τ_w and τ_w^+ are not known, but we have the following features.

Lemma 2.5. *The random variables τ_w and τ_w^+ are continuous. If $w \in \mathscr{W}$ is symmetric about $\frac{1}{2}$ then τ_w and τ_w^+ are symmetrically distributed about $\frac{1}{2}$.*

Proof. The first assertion is shown in Example 2.7 of Ferger (1999). Notice, furthermore, that $\arg\max_{0\leq t\leq 1} f(1-t) = 1 - \arg\max_{0\leq t\leq 1} f(t)$ for all continuous f with unique maximizing point. Recall that $\{B_0(t) : 0 \leq t \leq 1\} \overset{\mathscr{L}}{=} \{B_0(1-t) : 0 \leq t \leq 1\}$ to get the desired symmetry, upon noticing again that $\Gamma_0^{(1)}$ and $|\Gamma_0^{(1)}| \in \hat{C}$ a.s. □

If $w = 1$ then $\tau_1 = T_0 := \arg\max_{0\leq t\leq 1} |B_0(t)|$ and $\tau_1^+ = T_0^+ := \arg\max_{0\leq t\leq 1} B_0(t)$. In this case the distributions are completely known. Indeed, Ferger (1995) proves that T_0^+ is uniformly distributed on (0, 1). The distribution of T_0 as determined in the next theorem is much more complicated.

Theorem 2.6. *Let B_0 denote a Brownian bridge and put $M_0 = \max_{0\leq t\leq 1}|B_0(t)|$. Then we have:*

(i) The random vector (M_0, T_0) has Lebesgue density $f_{(M_0,T_0)}$ given by

$$f_{(M_0,T_0)}(u, v) = \sqrt{\frac{8}{\pi}} f(u, v) f(u, 1 - v), \tag{2.18}$$

where $0 \leqslant u < \infty$, $0 < v < 1$ *and*

$$f(u, v) := uv^{-3/2} \sum_{j=0}^{\infty} (-1)^j (2j+1) \exp\left\{ -\frac{(2j+1)^2}{2} \frac{u^2}{v} \right\}.$$

(ii) The random variable T_0 *has Lebesgue density*

$$f_{T_0}(x) = 2 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} (-1)^{i+j} \alpha_i \alpha_j \{\alpha_i^2 (1-x) + \alpha_j^2 x\}^{-3/2}, \qquad 0 < x < 1, \tag{2.19}$$

where $\alpha_i = 2i + 1$.

Proof. (i) Let φ denote the standard normal density and let B denote a Brownian motion. Put $M = \max_{0 \leqslant t \leqslant 1} |B(t)|$ and $T = \arg\max_{0 \leqslant t \leqslant 1} |B(t)|$. By (11.34) in Billingsley (1968), for all $0 \leqslant x < \infty$ and $0 < y < 1$, we have

$$\begin{aligned}
P(M_0 \leqslant x, T_0 \leqslant y) &= \lim_{\varepsilon \downarrow 0} P(M \leqslant x, T \leqslant y \,|\, |B(1)| \leqslant \varepsilon) \\
&= \lim_{\varepsilon \downarrow 0} \frac{P(M \leqslant x, T \leqslant y, |B(1)| \leqslant \varepsilon)}{P(|B(1)| \leqslant \varepsilon)} \\
&= \lim_{\varepsilon \downarrow 0} \frac{\frac{1}{2} \int_0^x \int_0^y \left[\frac{1}{\varepsilon} \int_0^\varepsilon d(s, t, u) \mathrm{d}u \right] \mathrm{d}t \, \mathrm{d}s}{\frac{1}{2\varepsilon} \int_{-\varepsilon}^{\varepsilon} \varphi(u) \mathrm{d}u} \\
&= \frac{1}{2} \frac{\int_0^x \int_0^y d(s, t, 0) \mathrm{d}t \, \mathrm{d}s}{\varphi(0)} \\
&= \sqrt{\frac{\pi}{2}} \int_0^x \int_0^y d(s, t, 0) \mathrm{d}t \, \mathrm{d}s,
\end{aligned}$$

where d denotes the common density of $(M, T, |B(1)|)$. Thus

$$f_{(M_0,T_0)}(u, v) = \sqrt{\frac{\pi}{2}} d(u, v, 0), \qquad 0 \leqslant u \leqslant \infty, \; 0 < v < 1.$$

By (1.13.8) in Borodin and Salminen (1996, p. 258),

$$d(u, v, 0) \frac{4}{\pi} f(u, v) f(u, 1 - v),$$

which yields (2.18).

(ii) By (i), for all $0 < x < 1$, we have

$$f_{T_0}(x) = \sqrt{\frac{8}{\pi}} \int_0^\infty f(u, x) f(u, 1-x) \mathrm{d}u.$$

Put $a_i := (-1)^i \alpha_i$ and $c_i := \alpha_i^2$ and recall that

$$\int_0^\infty u^2 \exp\left\{-a\frac{u^2}{2}\right\} \mathrm{d}u = \sqrt{\frac{\pi}{2}} a^{-3/2}, \qquad \forall a > 0.$$

Since, for all $0 \leqslant u < \infty$, $0 < x < 1$,

$$f(u, x) f(u, 1-x) = (x(1-x))^{-3/2} u^2 \sum_{i,j \geqslant 0} a_i a_j \exp\left\{-\frac{u^2}{2}\left[\frac{c_i}{x} + \frac{c_j}{1-x}\right]\right\},$$

interchanging integration and summation gives (2.19). □

Observe that $w(t) = (t(1-t))^{-a} \in \mathscr{W}$ for all $0 \leqslant a < \frac{1}{2}$, but $w \notin \mathscr{W}$ if $a = \frac{1}{2}$. This extreme case is treated in the following:

Theorem 2.7. *Under the assumptions of Theorem* 2.1,

$$\theta_{n,1/2} := \frac{1}{n} \arg\max_{1 \leqslant k < n} \frac{\left| \sum_{i=k+1}^{n} \sum_{j=1}^{k} K(X_{in}, X_{jn}) \right|}{\sqrt{k(n-k)}} \xrightarrow{\mathscr{L}} Z, \qquad n \to \infty, \tag{2.20}$$

where $P(Z=0) = P(Z=1) = \frac{1}{2}$.

Proof. Put $k_n = (\log n)^{-2}$, $I_n = [k_n, 1-k_n]$, $G_n = \{kn^{-1} : 1 \leqslant k < n\}$ and recall the definition of $S_n(t)$ in the proof of Theorem 2.1. Then

$$\theta_{n,1/2} = \arg\max_{t \in G_n} (t(1-t))^{-1/2} n^{-3/2} |S_n(t)|.$$

Next define

$$Y_n = \sup_{t \in G_n \cap I_n} (t(1-t))^{-1/2} n^{-3/2} |S_n(t)|,$$

$$Z_n = \sup_{t \in G_n \setminus I_n} (t(1-t))^{-1/2} n^{-3/2} |S_n(t)|$$

and

$$V_n = \sup_{t \in G_n} (t(1-t))^{-1/2} n^{-3/2} |S_n(t)| = \max(Y_n, Z_n).$$

From Theorem 4.3 of Csörgő and Horváth (1988) it follows (recall Remark 2.2(i)) that

$$\frac{V_n}{(2 \log\log n)^{1/2}} \xrightarrow{P} 1, \qquad n \to \infty. \tag{2.21}$$

The derivation of (2.44) and (2.45) in Csörgő and Horváth (1988) shows that

$$Y_n = O_P((\log\log\log n)^{1/2}),$$

which implies

$$\frac{Y_n}{(2\log\log n)^{1/2}} \xrightarrow{P} 0, \qquad n \to \infty. \tag{2.22}$$

Since

$$P(\theta_{n,1/2} \in I_n) \leqslant P(Z_n \leqslant Y_n) = P(V_n = Y_n),$$

we can conclude from (2.21) and (2.22) that

$$P(\theta_{n,1/2} \in I_n) \to 0, \qquad n \to \infty. \tag{2.23}$$

Write $S_n(t) = S_n(t;\, X_{1n}, \ldots, X_{nn})$ to stress the dependence on the observations X_{in}. Then by antisymmetry of K,

$$S_n(t;\, X_{nn}, \ldots, X_{1n}) = -S_n(1-t;\, X_{1n}, \ldots, X_{nn}),$$

for all observations $X_{1n}, \ldots, X_{nn}$ and for all $t \in G_n$. Clearly $(X_{1n}, \ldots, X_{nn}) \stackrel{\mathscr{L}}{=} \mathrm{P}(X_{nn}, \ldots, X_{1n})$, whence

$$\{|S_n(t)| : t \in G_n\} \stackrel{\mathscr{L}}{=} \{|S_n(1-t)| : t \in G_n\},$$

and therefore $\theta_{n,1/2} \stackrel{\mathscr{L}}{=} 1 - \theta_{n,1/2}$ for all $n \in \mathbb{N}$. Combine this with (2.23) to see that

$$P(\theta_{n,1/2} < (\log n)^{-2}) \to \tfrac{1}{2}, \qquad n \to \infty,$$

which gives the desired result. □

Remark 2.8. The above proof is due to Lajos Horváth (private communication). A short sketch of the proof is given by Lombard and Hart (1994, p. 205) and Csörgő and Horváth (1997, p. 135).

3. Berry–Esseen estimates

In this section we confine ourselves to the one-sided estimator θ_n^+ with $w = 1$ – that is, to

$$\theta_n^+ = n^{-1} \underset{1 \leqslant k < n}{\arg\max} \sum_{i=k+1}^{n} \sum_{j=1}^{k} K(X_{in}, X_{jn}). \tag{3.1}$$

From Theorem 2.1 and Remark 2.2(ii) we can infer under the second-moment condition (2.1) with $p = 2$ that $\theta_n^+ \xrightarrow{\mathscr{L}} \tau_1^+$, where τ_1^+ is uniformly distributed on (0, 1). If higher moments exist, then it is possible to establish rates of convergence in law. This is the subject of the next theorem.

Theorem 3.1. *For all $n \in \mathbb{N}$, let $X_{1n}, \ldots, X_{nn}$ be i.i.d. random elements in $(\mathscr{X}, \mathscr{F})$ with common distribution ν_n. Assume that*

$$\sup_{n\geqslant 1}\int |K|^p \mathrm{d}\nu_n\otimes\nu_n<\infty, \qquad \textit{for some } p>2, \tag{3.2}$$

and that $\sigma_n^2=\int[\int K(x,\,y)\nu_n(\mathrm{d}y)]^2\nu_n(\mathrm{d}x)$ *satisfies*

$$\liminf_{n\to\infty}\sigma_n^2>0. \tag{3.3}$$

Then we have:

$$\sup_{x\in[0,1]}|P(\theta_n^+\leqslant x)-x|=\begin{cases} O\Big(n^{-\frac{p-2}{2(p+1)}}\Big), & 2<p<5,\\ O\Big(n^{-1/4}(\log n)^{3/4}\Big), & p\geqslant 5.\end{cases}$$

If the array (X_{in}) *arises in the usual way from a sequence* $X_1, X_2, \ldots$ *i.i.d. with* $X_1\sim\nu$ – *that is,* $X_{in}=X_i$, $1\leqslant i\leqslant n$, $n\in\mathbb{N}$ – *then under* (2.1) *we actually have*

$$\sup_{x\in[0,1]}|P(\theta_n^+\leqslant x)-x|=\begin{cases} o\Big(n^{-\frac{p-2}{2(p+1)}}\Big), & 2<p<\infty,\\ O\Big(n^{-1/2}\log n\Big), & K \textit{ bounded}.\end{cases}$$

Proof. Let ξ_n be the continuous random polygonal line with vertices at the points $(kn^{-1},\,\sigma n^{-3/2}S_n(kn^{-1}))$, $0\leqslant k\leqslant n$. Thus we have

$$\theta_n^+=\psi(\xi_n).$$

For $x\in[0,\,1)$ we define the mapping $T_x\colon C[0,\,1]\to\mathbb{R}$ by

$$T_x(f)=\sup_{x\leqslant t\leqslant 1}f(t)-\sup_{0\leqslant t\leqslant x}f(t), \qquad f\in C[0,\,1],$$

where as usual $C[0,\,1]$ is the set of continuous functions on $[0,\,1]$. Recall that $\psi(B_0)=T_0^+$ is uniformly distributed on $(0,\,1)$. Here, without loss of generality, we can assume that B_0 is defined on the same probability space $(\Omega,\,\mathscr{A},\,P)$ that carries our random elements X_{in}. This is possible because the following arguments only involve the distribution of B_0. So for all $x\in[0,\,1)$,

$$\begin{aligned}|P(\theta_n^+\leqslant x)-x| &= |P(\psi(\xi_n)\leqslant x)-P(\psi(B_0)\leqslant x)| \\ &= |P(T_x(\xi_n)\leqslant 0)-P(T_x(B_0)\leqslant 0)|.\end{aligned} \tag{3.4}$$

Our goal is to apply Lemma 6.2. For that purpose, note that T_x is Lipschitz continuous:

$$|T_x(f)-T_x(g)|\leqslant 2\sup_{0\leqslant t\leqslant 1}|f(t)-g(t)| \qquad \forall f,\,g\in C[0,\,1]. \tag{3.5}$$

Furthermore, by the Markov property of B_0 the df H_x of $T_x(B_0)$ can be written as

$$
\begin{aligned}
H_x(\lambda) &= P(T_x(B_0) \leqslant \lambda) \\
&= P\left(\sup_{x \leqslant t \leqslant 1} B_0(t) \leqslant \lambda + \sup_{0 \leqslant t \leqslant x} B_0(t) \right) \\
&= \iint P(M'_x \leqslant \lambda + y | B_0(x) = u) m_{x,u}(\mathrm{d}y) \mathscr{L}(B_0(x))(\mathrm{d}u)
\end{aligned}
$$

for all $\lambda \in \mathbb{R}$, where

$$
M'_x = \sup_{x \leqslant t \leqslant 1} B_0(t)
$$

and

$$
m_{x,u}(y) = P\left(\sup_{0 \leqslant t \leqslant x} B_0(t) \leqslant y | B_0(x) = u \right).
$$

From (17) in Shorack and Wellner (1986, p. 38) it follows that

$$
m_{x,u}(y) = \begin{cases} 1 - \exp\left\{ -\dfrac{2}{x} y(y-u) \right\}, & y > \max(0, u), \\ 0, & \text{otherwise,} \end{cases}
$$

and

$$
P(M'_x \leqslant \lambda + y | B_0(x) = u) = \begin{cases} 1 - \exp\left\{ -\dfrac{2}{1-x} (\lambda + y - u)(\lambda + y) \right\}, & y > \max(u - \lambda, -\lambda), \\ 0, & \text{otherwise.} \end{cases}
$$

Herewith we obtain, for $\lambda \geqslant 0$:

$$
\begin{aligned}
H_x(\lambda) = &\int_{-\infty}^{0} \int_{0}^{\infty} \left[1 - \exp\left\{ -\frac{2}{1-x} (\lambda + y - u)(\lambda + y) \right\} \right] m_{x,u}(\mathrm{d}y) \mathscr{L}(B_0(x))(\mathrm{d}u) \\
&+ \int_{0}^{\infty} \int_{u}^{\infty} \left[1 - \exp\left\{ -\frac{2}{1-x} (\lambda + y - u)(\lambda + y) \right\} \right] m_{x,u}(\mathrm{d}y) \mathscr{L}(B_0(x))(\mathrm{d}u).
\end{aligned}
$$

Integration yields

$$
\begin{aligned}
H_x(\lambda) = 1 &+ 2\sqrt{2x(1-x)/\pi}\, \lambda \exp\left\{ -\frac{2}{1-x} \lambda^2 \right\} \\
&+ 2[x(4\lambda^2 - 1) + 1] \left[\Phi\left(2\sqrt{\frac{x}{1-x}} \lambda \right) - 1 \right] \exp\{-2\lambda^2\},
\end{aligned} \tag{3.6}
$$

for all $\lambda \geqslant 0$ and $0 \leqslant x < 1$, where Φ is the standard normal df. Differentiation gives

$$\frac{\mathrm{d}H_x(\lambda)}{\mathrm{d}\lambda} = 4\exp\left\{-\frac{2}{1-x}\lambda^2\right\}(1-2\lambda^2)\sqrt{2x(1-x)/\pi}$$
$$+ 8\lambda\exp\{-2\lambda^2\}[x(4\lambda^2-3)+1]\left(1-\Phi\left(2\sqrt{\frac{x}{1-x}}\lambda\right)\right).$$

Consider this expression separately for the three cases $0 \leqslant \lambda \leqslant 1/\sqrt{2}$, $1/\sqrt{2} < \lambda \leqslant \frac{1}{2}\sqrt{3}$ and $\lambda > \frac{1}{2}\sqrt{3}$ to infer that

$$L_0 := \sup_{\substack{0\leqslant x<1\\ \lambda\geqslant 0}} \frac{\mathrm{d}H_x(\lambda)}{\mathrm{d}\lambda} < \infty.$$

Consequently, by the mean value theorem, H_x is Lipschitz continuous on $[0, \infty)$ with a Lipschitz constant L_0 that does not depend on x:

$$|H_x(\lambda+h) - H_x(\lambda)| \leqslant L_0 h, \qquad \forall h \in \mathbb{R}, \forall\lambda \geqslant 0, \forall 0 \leqslant x < 1.$$

Since $B_0(\cdot) \overset{\mathscr{L}}{=} B_0(1-\cdot)$, we have $T_x(B_0) \overset{\mathscr{L}}{=} -T_{1-x}(B_0)$, whence the above inequality holds for all $\lambda \in \mathbb{R}$ and for all $0 < x < 1$. By (3.5) we are in a position to apply Lemma 6.2 to the term on the right-hand side of (3.4) and obtain

$$\sup_{0\leqslant x\leqslant 1} |P(\theta_n^+ \leqslant x) - x| = \sup_{0<x<1} |P(\theta_n^+ \leqslant x) - x| \leqslant (4L_0+1)\rho(P\circ\Gamma_n^{-1}, P\circ B_0^{-1}),$$

where ρ denotes the Prokhorov metric. Now, in the case of (3.2) and (3.3), Theorems 1.1 and 1.2 of Ferger (1994b) state that

$$\rho(P\circ\Gamma_n^{-1}, P\circ B_0^{-1}) = \begin{cases} O\left(n^{-\frac{p-2}{2(p+1)}}\right), & \text{if } 2<p<5,\\ O\left(n^{1/4}(\log n)^{3/4}\right), & \text{if } p \geqslant 5, \end{cases} \tag{3.7}$$

whence the first part of the theorem follows. If the X_{in} arise from a single sequence, then the rates in (3.7) can be sharpened. This is the message of Theorems 1.3 and 1.4 of Ferger (1994b), which say that now

$$\rho(P\circ\Gamma_n^{-1}, P\circ B_0^{-1} = \begin{cases} o\left(n - \dfrac{p-2}{2(p+1)}\right), & \text{if } 2<p<\infty,\\ O(n^{-1/2})\log n, & \text{if } K \text{ is bounded.} \end{cases}$$

This completes the proof. □

Remark 3.2. Note that

$$H_x(0) = P(T_x(B_0) \leqslant 0) = P\left(\operatorname*{arg\,max}_{0\leqslant t\leqslant 1} B_0(t) \leqslant x\right)$$

and that

$$H_0(\lambda) = P\left(\sup_{0\leqslant t\leqslant 1} B_0(t) \leqslant \lambda\right).$$

So formula (3.6) for $\lambda = 0$ again yields that the maximizer $\psi(B_0)$ of a Brownian bridge is uniformly distributed on (0, 1). For $x = 0$ it reduces to the well-known boundary non-crossing probability $P(\sup_{0 \leqslant t \leqslant 1} B_0(t) \leqslant \lambda) = 1 - \exp(-2\lambda^2)$, $\lambda \geqslant 0$.

4. Almost sure divergence

We return to the case of general $w \in \mathscr{W}$. First it will be shown that our estimators are a.s. divergent under the null hypothesis. Here a specification of the probabilistic relation between the rows of the array (X_{in}) is required.

Theorem 4.1. *Assume that the rows* $(X_{1n}, \ldots, X_{nn}), n \in \mathbb{N}$, *are independent* $\mathscr{X}^n$*-valued random variables. Then under the assumptions of Theorem* 2.1 *the estimators* (θ_n), (θ_n^+) *and* $\theta_{n,1/2}$ *are divergent with probability one. If* $\mathscr{X} = \mathbb{R}$ *the almost sure divergence still holds for* (θ_n) *and* (θ_n^+) *under the weaker assumption of Theorem* 2.3(ii).

Proof. Let θ_n^* denote any of the three estimators θ_n, θ_n^+, $\theta_{n,1/2}$. By Kolmogorov's zero–one law we know that

$$P((\theta_n^*) \text{ is divergent}) \in \{0, 1\}.$$

Assume the above probability is equal to zero, which means θ_n^* converges a.s. to some limit ξ^*. According to Lemma 1.16.6 in Gänssler and Stute (1977), this ξ^* is a.s. constant. Since almost sure convergence implies convergence in distribution, we can conclude from Theorems 2.1, 2.3 or 2.7, respectively, that ξ^* is equal in distribution to the limit variables τ_w, τ_w^+ or Z, respectively. However, τ_w and τ_w^+ are actually continuous by Lemma 2.5 and Z is a *Bernoulli*$(\frac{1}{2})$ variable, so that in all three cases we arrive at a contradiction. This completes the proof. □

Recall Example 1.1 to see that the independence assumption on the rows can have a quite reasonable statistical justification. In our next result we make the statement on divergence more precise by specifying exactly the almost sure set of cluster points.

Theorem 4.2. *Under the assumptions of Theorem* 2.1 *or Theorem* 2.3(ii), *assume that the rows* $(X_{1n}, \ldots, X_{nn})$, $n \in \mathbb{N}$, *are pairwise independent. If the distribution functions* H *and* H^+ *of* τ_w *and* τ_w^+, *respectively, are strictly monotone, then the sets* C *and* C^+ *of cluster points of* (θ_n) *and* (θ_n^+), *respectively, coincide with probability one with the closed unit interval:*

$$C = [0, 1] = C^+ \ a.s. \tag{4.1}$$

The set $C_{1/2}$ *of cluster points of* $(\theta_{n,1/2})$ *contains the two boundary points of* [0, 1]*:*

$$\{0, 1\} \subseteq C_{1/2} \ a.s. \tag{4.2}$$

Proof. Consider an arbitrary fix $x \in (0, 1)$ and let $\varepsilon_0 > 0$ be such that $0 < x - \varepsilon < x + \varepsilon < 1$ for all $0 < \varepsilon < \varepsilon_0$. By (2.3) or (2.16), we have that

$$\lim_{n\to\infty} P(|\theta_n - x| \leqslant \varepsilon) \geqslant H(x+\varepsilon) - H(x-\varepsilon) > 0.$$

Therefore the series

$$\sum_{n\geqslant 1} P(|\theta_n - x| \leqslant \varepsilon)$$

is divergent, whence by Theorem 4.2.5 of Chung (1974), $P(\limsup_{n\to\infty}\{|\theta_n - x| \leqslant \varepsilon\}) = 1$ for all $0 < \varepsilon \leqslant \varepsilon_0$. Since

$$\bigcap\nolimits_{0<\varepsilon\in\mathbb{Q}} \limsup_{n\to\infty}\{|\theta_n - x| \leqslant \varepsilon\} \subseteq \{x \in C\},$$

this shows that $x \in C$ a.s. It follows that $Q := (0, 1) \cap \mathbb{Q} \subseteq C$ a.s. Let $\bar{A}$ denote the topological closure of a set $A \subset \mathbb{R}$. Since C is closed and Q is dense in $I = [0, 1]$, we can further conclude that $I \subseteq \overline{Q} \subseteq \overline{C} = C \subseteq I$, so $C = I$ a.s. The second equality in (4.1) follows analogously.

Similarly, $\sum_{n\geqslant 1} P(\theta_{n,1/2} \leqslant \varepsilon)$ is divergent for all $\varepsilon > 0$, because $\lim_{n\to\infty} P(\theta_{n,1/2} \leqslant \varepsilon) = \frac{1}{2} > 0$ by (2.12). Therefore $0 \in C_{1/2}$ a.s. and by the same arguments $1 \in C_{1/2}$, so that $\{0, 1\} \subseteq C_{1/2}$ a.s., which finishes the proof. □

Note that the above result extends Theorem 4.1 even under a weaker assumption on the rows of the array, but on the other hand strict monotonicity of the dfs of τ_w and τ_w^+ is needed. For $w = 1$ this is ensured by Theorem 2.6.

5. An exact result

In this section we give a finite sample size result. Therefore it is no longer necessary to deal with arrays. Consider the slightly modified one-sided estimator of θ_n^+ with no weights ($w = 1$) pertaining to a sample $X_1, \ldots, X_n$, $n \in \mathbb{N}$:

$$\tilde{\theta}_n^+ = \frac{1}{n} \operatorname*{arg\,max}_{0\leqslant k<n} \sum_{i=k+1}^{n} \sum_{j=1}^{k} K(X_i, X_j).$$

The difference from the original θ_n^+ lies in the fact that the argmax now includes the point zero. Here, as usual, the summation over the empty set is defined to be zero. Our next theorem gives the exact finite sample size distribution of $\tilde{\theta}_n^+$. It suffices to require exchangeability of $X_1, \ldots, X_n$, that is, $(X_1, \ldots, X_n) \stackrel{\mathscr{L}}{=} (X_{\pi(1)}, \ldots, X_{\pi(n)})$ for all permutations π of the integers $1, \ldots, n$.

Theorem 5.1. *Let $X_1, \ldots, X_n$ be exchangeable. Then $\tilde{\theta}_n^+$ is uniformly distributed on the grid $\{kn^{-1} : 0 \leqslant k < n\}$, that is,*

$$P\left(\tilde{\theta}_n^+ = \frac{k}{n}\right) = \frac{1}{n}, \qquad \textit{for all } 0 \leqslant k \leqslant n-1,$$

if and only if

$$P\left(\sum_{i=k+1}^{n}\sum_{j=1}^{k} K(X_i, X_j) = 0\right) = 0, \qquad \textit{for all } 1 \leqslant k \leqslant n-1. \tag{5.1}$$

Proof. We consider the increments Y_k of $S_k = \sum_{i=k+1}^{n}\sum_{j=1}^{k} K(X_i, X_j)$:

$$\begin{aligned} Y_k &= \sum_{i=k+1}^{n}\sum_{j=1}^{k} K(X_i, X_j) - \sum_{i=k}^{n}\sum_{j=1}^{k-1} K(X_i, X_j) \\ &= \sum_{i=k}^{n} K(X_i, X_k) - \sum_{j=1}^{k} K(X_k, X_j) \\ &= \sum_{i=1}^{n} K(X_i, X_k), \qquad 1 \leqslant k \leqslant n, \end{aligned}$$

where the last equality holds by antisymmetry of K. By definition $T_n := n\tilde{\theta}_n^+$ is the smallest index $k \in \{0, 1, \ldots, n-1\}$ with $S_k = \max_{0 \leqslant i \leqslant n-1} S_i$. Our goal is the application of Theorem 2 of Andersen (1953). For that purpose it remains to show that $Y_1, \ldots, Y_n$ are exchangeable. Let π be an arbitrary permutation of the integers $1, \ldots, n$. Then we have

$$(Y_1, \ldots, Y_n) = \left(\sum_{i=1}^{n} K(X_i, X_k)\right)_{1 \leqslant k \leqslant n} = T(X_1, \ldots, X_n),$$

where $T : \mathscr{X}^n \to \mathbb{R}^n$ is measurable. It follows that

$$\begin{aligned} (Y_{\pi(1)}, \ldots, Y_{\pi(n)}) &= \left(\sum_{i=1}^{n} K(X_i, X_{\pi(k)})\right)_{1 \leqslant k \leqslant n} \\ &= \left(\sum_{i=1}^{n} K(X_{\pi(i)}, X_{\pi(k)})\right)_{1 \leqslant k \leqslant n} \\ &= T(X_{\pi(1)}, \ldots, X_{\pi(n)}) \\ &\stackrel{\mathscr{L}}{=} T(X_1, \ldots, X_n) = (Y_1, \ldots, Y_n). \end{aligned}$$

Choosing $C = \Omega$ in Theorem 2 of Andersen (1953) yields the assertion. □

One might expect the original estimator θ_n^+ with $w = 1$ also to be uniformly distributed on the grid $\{kn^{-1} : 1 \leqslant k \leqslant n-1\}$. But surprisingly, simulation studies strongly confirm our conjecture that this is not true. The two final examples below concern the necessary and sufficient condition (5.1).

Example 5.2. Let $X_1, \ldots, X_n$ be i.i.d. and $K(x, y) = f(x) - f(y)$ with $f : \mathscr{X} \to \mathbb{R}$ measurable. If $f(X_1)$ is continuous, then (5.1) is fulfilled for all $n \in \mathbb{N}$. Under the assumption that X_1 has no atoms, it suffices to require that the sets $\{f = r\}$ are countable for all $r \in \mathbb{R}$, which is quite a weak requirement.

Example 5.3. Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli variables with success parameter $p \in (0, 1)$. If $K(x, y) = x - y$, then (5.1) is violated for all $n \in \mathbb{N}$. For example, for $n = 4$, $p = 0.3$ with $p_k = P(\tilde{\theta}_4^+ = k/4)$, $0 \leqslant k \leqslant 3$, a Monte Carlo method (with 10^7 replicates) produced $p_0 = 0.458$, $p_1 = 0.210$, $p_2 = 0.166 = p_3$.

6. Technical lemmas

Lemma 6.1. *We use the notation of the proof of Theorem* 2.1. *Then the following statements hold:*

(i) For all $f \in D = D[0, 1]$, the set $S(f) \subseteq [0, 1]$ is non-empty and closed.
(ii) The set $\hat{C}$ of continuous functions on $[0, 1]$ *with a unique maximizing point is contained in the complement of E in D: $\hat{C} \subseteq D \backslash E$.*

Proof. (i) First we show that $S(f) \neq \varnothing$. Put $M(f) = \sup_{0 \leqslant t \leqslant 1} f(t)$. Since $f \in D$, it is bounded and thus $\xi := M(f) \in \mathbb{R}$. In particular, there exists a sequence $(u_n) \subseteq [0, 1]$ with $f(u_n) \to \xi$. By compactness, we can assume without loss of generality that (u_n) converges to some $u \in [0, 1]$. (Otherwise take a suitable subsequence.) If f is continuous at u then $f(u) = \xi$, whence $u \in S(f)$. If f has a positive jump at u, so that $f(u) > f(u-)$, then $u_n > u$ for all but a finite number of $n \in \mathbb{N}$. (Otherwise there exists a subsequence $(u_{n_k})_k$ of (u_n) with $u_{n_k} \uparrow u$, which implies $\xi = f(u-) < f(u) \leqslant \xi$, a contradiction.) Therefore, by the right continuity of f we obtain $\xi = f(u)$ and $u \in S(f)$. If f has a negative jump, so that $f(u-) < f(u)$, then $u_n < u$ for all but a finite number of $n \in \mathbb{N}$. (Otherwise there exists a subsequence $(u_{n_k})_k$ with $u_{n_k} \downarrow u$, whence $\xi = f(u) < f(u-)$ and $f(v) > \xi$ for all $v < u$ sufficiently close to u, which again is a contradiction.) Thus $\xi = f(u-)$ and $u \in S(f)$. This shows that $S(f)$ is non-empty. To prove that $S(f)$ is closed let $(u_n) \subseteq S(f)$ be a convergent sequence with limit u. We can assume without loss of generality that either $u_n \downarrow u$ or $u_n \uparrow u$. In the first case $f(u) = \lim_{n\to\infty} f(u_n) = \lim_{n\to\infty} \xi = \xi$, so that $u \in S(f)$. In the second case $f(u) \leqslant f(u-)$, because otherwise $f(u) > f(u-) = \lim_{n\to\infty} f(u_n) = \xi \geqslant f(u)$. If $f(u) = f(u-)$ then $u \in S(f)$ by continuity of f. If $f(u) < f(u-)$, then $f(u-) = \lim_{n\to\infty} f(u_n) = \xi$ and $u \in S(f)$.

(ii) Let $f \in \hat{C}$ and $(f_n) \subseteq D$ with $f_n \to_s f$. We have to show that

$$\psi_n(f_n) \to \psi(f), \qquad n \to \infty. \tag{6.1}$$

Since f is continuous we actually have $f_n \to_d f$, where d denotes the sup metric on D.

Let $\varepsilon > 0$ and put $t_0 = \psi(f)$, $t_n = [nt_0]/n$, $U_\varepsilon = (t_0 - \varepsilon, t_0 + \varepsilon)$ and $m_\varepsilon = \sup\{f(t): t \in [0, 1]\backslash U_\varepsilon\}$. Because t_0 is unique, $\delta = \delta_\varepsilon = \frac{1}{5}(f(t_0) - m_\varepsilon) > 0$. Moreover, there exists an $N_0 = N_0(\delta_\varepsilon) \in \mathbb{N}$ such that $d(f_n, f) < \delta$ and $|f([nt_0]/n) - f(t_0)| < \delta$ for all $n \geqslant N_0$. Thus for all $t \in [0, 1]\backslash U_\varepsilon$ and for all $n \geqslant N_0$, we have

$$\begin{aligned} f_n(t_n) - f_n(t) &= f(t_n) - f(t) + [f_n(t_0) - f(t_0)] + [f(t) - f_n(t)] \\ &\quad + [f(t_n) - f(t_0)] + [f_n(t_n) - f(t_n)] + [f(t_0) - f_n(t_0)] \\ &\geqslant f(t_0) - m_\varepsilon - 4d(f_n, f) - |f(t_n) - f(t_0)| \\ &> 5\delta - 4\delta - \delta > 0. \end{aligned}$$

Consequently $f_n(t_n) > f_n(t)$ for all $t \in [0, 1]\backslash U_\varepsilon$ and for all $n \geqslant N_0$, which implies

$$\psi_n(f_n) \in U_\varepsilon, \qquad \text{for all } n \geqslant N_0,$$

proving (6.1). □

The following lemma due to Borovkov (1973) is a sharpening of the continuous mapping theorem under an additional smoothness condition of Lipschitz continuity. Let (S, d) be a metric space endowed with the Borel σ-algebra $\mathscr{B}(S)$, and let ρ denote the Prokhorov metric on the set of probability measures on $\mathscr{B}(S)$.

Theorem 6.2 (Borovkov). *Let $T: S \to \mathbb{R}$ be a mapping with*

$$|T(x) - T(y)| \leqslant K_1 d(x, y), \qquad \forall x, y \in S,$$

for some constant $K_1 < \infty$, and let Q be a probability measure on $\mathscr{B}(S)$ with

$$Q(T \leqslant \lambda + h) - Q(T \leqslant \lambda) \leqslant K_2 h, \qquad \forall \lambda \in \mathbb{R}, \forall h > 0,$$

for some constant $K_2 < \infty$. Then for each random element ξ on a probability space $(\Omega, \mathscr{A}, P)$ with values in $(S, \mathscr{B}(S))$, we have

$$\sup_{\lambda \in \mathbb{R}} |P(T(\xi) \leqslant \lambda) - Q(T \leqslant \lambda)| \leqslant (2K_1 K_2 + 1)\rho(P \circ \xi^{-1}, Q).$$

References

Andersen, E.S. (1953) On the fluctuations of sums of random variables, I. *Math. Scand.*, **1**, 263–285.

Antoch, J. and Hušková, M. (1999) Estimators of changes. In S. Ghosh (ed.), *Asymptotics, Nonparametrics and Time Series. A tribute to Mandan Lal Puri*, pp. 533–578. New York: Marcel Dekker.

Bhattacharya, P.K. (1987) Maximum likelihood estimation of a change-point in the distribution of independent random variables: general multiparameter case. *J. Multivariate Anal.*, **23**, 183–208.

Bhattacharya, P.K. and Brockwell, P.J. (1976) The minimum of an additive process with applications to signal estimation and storage theory. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **37**, 51–75.

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

Billingsley, P. (1968) *Convergence of Probability Measures*. New York: Wiley.
Borodin, A.N. and Salminen, P. (1996) *Handbook of Brownian Motion – Facts and Formulae*. Basel: Birkhäuser.
Borovkov, A.A. (1973) On the rate of convergence for the invariance principle. *Theory Probab. Appl.*, **18**, 207–225.
Brodsky, B.E. and Darkhovsky, B.S. (1993) *Nonparametric Methods in Change-point Problems*. Dordrecht: Kluwer.
Chung, K.L. (1974) *A Course on Probability Theory*, 2nd edition. New York: Academic Press.
Csörgő, M. and Horváth, L. (1988) Invariance principles for change-point problems. *J. Multivariate Anal.*, **27**, 151–168.
Csörgő, M. and Horváth, L. (1993) *Weighted Approximations in Probability and Statistics*. New York: Wiley.
Csörgő, M. and Horváth, L. (1997) *Limit Theorems in Change Point Analysis*. New York: Wiley.
Dümbgen, L. (1991) The asymptotic behavior of some nonparametric change-point estimators. *Ann. Statist.*, **19**, 1471–1495.
Ferger, D. (1994a) Change-point estimators in case of small disorders. *J. Statist. Plann. Inference*, **40**, 33–49.
Ferger, D. (1994b) On exact rates of convergence in functional limit theorems for U-statistic type processes. *J. Theoret. Probab.*, **7**, 709–723.
Ferger, D. (1994c) On the power of nonparametric change-point tests. *Metrika*, **41**, 277–297.
Ferger, D. (1995) The joint distribution of the running maximum and its location of D-valued Markov processes. *J. Appl. Probab.*, **32**, 842–845.
Ferger, D. (1996) On the asymptotic behavior of change-point estimators in case of no change with applications to testing. *Statist. Decisions*, **14**, 137–143.
Ferger, D. (1999) On the uniqueness of maximizers of Markov–Gaussian processes. *Statist. Probab. Lett.*, **45**, 71–77.
Ferger, D. (2001) Exponential and polynomial tailbounds for change-point estimators. *J. Statist. Plann. Inference*, **92**, 73–109.
Ferger, D. and Stute, W. (1992) Convergence of change-point estimators. *Stochastic Process. Appl.*, **42**, 345–351.
Gänssler, P. and Stute, W. (1977) *Wahrscheinlichkeitstheorie*. Berlin: Springer-Verlag.
Gombay, E. and Horváth, L. (1996) Approximations for the time of change and the power function in change-point models. *J. Statist. Plann. Inference*, **52**, 43–66.
Hušková, M. (1996) Tests and estimators for the change-point problem based on M-statistics. *Statist. Decisions*, **14**, 115–136.
Lombard, F. and Hart, J.D. (1994) The analysis of change-point data with dependent errors. In E. Carlstein, H.G. Mueller and D. Siegmund (eds), *Change-point Problems*, IMS Lecture Notes Monogr. Ser., Vol. 23, pp. 194–209. Hayward, CA: Institute of Mathematical Statistics.
Oosterhoff, J. and van Zwet, W.R. (1979) A note on contiguity and Hellinger distance. In J. Jurečková (ed.) *Contributions on Statistics: Jaroslav Hájek Memorial Volume*, pp. 157–166. Dordrecht: Reidel.
Shorack, G.R. and Wellner, J.A. (1986) *Empirical Processes with Applications to Statistics*. New York: Wiley.
Szyszkovicz, B. (1991) Changepoint problems and contiguous alternatives. *Statist. Probab. Lett.*, **11**, 299–308.

Received June 1999 and revised July 2000