ELSEVIER

Journal of Statistical Planning and Inference 53 (1996) 297-310

journal of statistical planning and inference

# The change-point problem for dependent observations

Liudas Giraitis [a,1], Remigijus Leipus [b,*], Donatas Surgailis [c]

[a] Heidelberg University, Im Neuenheimer Feld 294, 69120 Heidelberg, Germany
[b] Department of Mathematics, Vilnius University, Naugarduko 24, 2006 Vilnius, Lithuania
[c] Institute of Mathematics and Informatics, Akademijos 4, 2600 Vilnius, Lithuania

Received 2 February 1995; revised 6 September 1995

**Abstract**

We consider the change-point problem for the marginal distribution function of a strictly stationary time series. Asymptotic behavior of Kolmogorov–Smirnov type tests and estimators of the change point is studied under the null hypothesis and converging alternatives. The discussion is based on a general empirical process' approach which enables a unified treatment of both short-memory (weakly dependent) and long-memory time series. In particular, the case of long-memory moving-average process $X_j = \sum_{s\leqslant j} b_{j-s}\xi_s$ is studied, using the recent results of Giraitis and Surgailis (1994).

*AMS Subject Classification*: Primary 60F05; secondary 62G05; 62G10; 62M10

*Keywords:* Change-point problem; Empirical processes; Long-memory; Kolmogorov–Smirnov test

## 1. Introduction

Detection of changes in the distribution parameters of a random sequence is important for many applications; see e.g. the recent books Brodsky and Darkhovsky (1993) and Basseville and Nikiforov (1993) and the references therein.

This paper deals with nonparametric situation, which usually arises when the form of the distribution is unknown a priori. Like many statistical problems, the change-point problem allows two different formulations — a posteriori and sequential, also called off-line and on-line, respectively. In the first case, the decision about stochastic homogeneity of a random sequence (the abscence of change) is made after observing a sample of a fixed length. In the sequential formulation, the decision must be made 'on line' with the observations.

* Corresponding author.

[1] Research supported by the Alexander von Humboldt Foundation. Permanent address: Institute of Mathematics and Informatics, Akademijos 4, 2600 Vilnius, Lithuania.


SSDI 0378-3758(95)00148-4

The change-point problem (both a posteriori and sequential approaches) is well-studied in the case of independent observations (see e.g. Csörgő and Horváth, 1988; Brodsky and Darkhovsky, 1993), in which case one is interested in the detection of change of the cumulative distribution function. A natural statistic for testing the null-hypothesis is a Kolmogorov–Smirnov (K-S)-type statistic, used by several authors (Picard, 1985; Deshayes and Picard, 1986; Hawkins, 1988; Leipus, 1988; Szyszkowicz, 1994; Csörgő and Szyszkowicz, 1994).

In this paper, we develop a general asymptotic approach to the change -point problem of the marginal distribution function $F(x) = P\{X_i \leqslant x\}$ for (dependent) stationary observations $X_i, i \in \mathbb{Z}$, based on the asymptotics of the two-parameter empirical process

$$W_N(t,x) = [Nt]\big(F_{[Nt]}(x) - F(x)\big), \quad (t,x) \in [0,1] \times \mathbb{R}, \tag{1.1}$$

where

$$F_N(x) = \frac{1}{N}\sum_{j=1}^{N} \mathbf{1}\{X_j \leqslant x\} \tag{1.2}$$

is the empirical distribution function. Assuming that, for some normalizing constants $d_N \to \infty$, $d_N^{-1} W_N(t,x)$ converge weakly in the Skorokhod space $D([0,1]\times[-\infty,+\infty])$ to some (nontrivial) limit $W(t,x)$, one obtains the convergence of Type I error probabilities of rejecting the null hypothesis to the probability expressed in terms of the limit random field $W(t,x)$ (Propositions 2.1 and 2.2). The same approach applies to testing *converging* change-point alternatives introduced in Giraitis and Leipus (1992) for linear models, as a substitute for the more usual *contiguous* alternatives (Proposition 2.3). Section 3 discusses some nonparametric estimators of the change point $\theta$ itself, under the assumptions of Proposition 2.3. Finally, Section 4 discusses the change-point problem for long memory moving averages, including fractional ARIMA processes, $X_j = \sum_{s \leqslant j} b_{j-s}\xi_s, j \in \mathbb{Z}$, where $\{\xi_s\}_{s\in\mathbb{Z}}$ is an i.i.d. sequence, and the weights $b_j$ decay slowly hyperbolically as $j \to \infty$.

## 2. Testing the change-point hypotheses

Let us introduce some notation. Let $0 \leqslant \theta \leqslant 1$, and $F^{(1)}(x)$, $F^{(2)}(x)$ be two distribution functions. A random vector $\boldsymbol{X}_N = (X_1, \dots, X_N) \in \Psi_N(\theta, F^{(1)}, F^{(2)})$ if

$$X_j = \begin{cases} X_j^{(1)}, & 1 \leqslant j \leqslant [N\theta], \\ X_j^{(2)}, & [N\theta] < j \leqslant N, \end{cases} \tag{2.1}$$

where $(X_j^{(i)})_{j\in\mathbb{Z}} \equiv \boldsymbol{X}^{(i)}$ is a strictly stationary process with $P\{X_j^{(i)} \leqslant x\} = F^{(i)}(x)$, $i = 1,2$. Here, $k_N = [N\theta] + 1$ is the change-point of the marginal distribution of the sample $X_1, \dots, X_N$. Note that no assumptions about the joint distribution of the two processes $\boldsymbol{X}^{(1)}$, $\boldsymbol{X}^{(2)}$ are made. The class $\Psi_N(F) =: \Psi_N(1,F)$ refers to all vectors $(X_1, \dots, X_N)$ having the same marginal distribution $F$.

We consider the weak convergence, denoted by $\overset{D(I)}{\Longrightarrow}$, of random elements taking values in the space $D(I)$, $I = [0,1] \times [-\infty, +\infty]$ equipped with the Skorokhod $J_1$ topology (see Bickel and Wichura, 1971). Write $\Longrightarrow$ for the weak convergence of finite dimensional distributions, and $\overset{\mathscr{X}}{=}$ for the equality in distribution of random elements with values in a measurable space $\mathscr{X}$.

## 2.1. Testing the null hypothesis when $F(x)$ is known

Consider the pair $(\mathrm{H}_0, \mathrm{H}_1)$ of alternative hypotheses about the distribution of a given random sample $\boldsymbol{X}_N = (X_1, \ldots, X_N)$:

$$\mathrm{H}_0 : \{\boldsymbol{X}_N \in \Psi_N(F)\},$$
$$\mathrm{H}_1 : \{\exists \theta \in (0,1)\ \exists F_1 \neq F \text{ such that } \boldsymbol{X}_N \in \Psi_N(\theta, F, F_1)\}.$$

The testing procedure for the pair $(\mathrm{H}_0, \mathrm{H}_1)$ is based on the process

$$W_N^*(t,x) = (N - [Nt])(F_{N-[Nt]}^*(x) - F(x)),$$

where

$$F_{N-k}^*(x) = \frac{1}{N-k} \sum_{j=k+1}^{N} \mathbf{1}\{Y_j \leqslant x\}$$

is the empirical distribution function based on the partial sample $X_{k+1}, \ldots, X_N$. Namely, we reject the null hypothesis $\mathrm{H}_0$ when

$$T_N := d_N^{-1} \sup_{(t,x) \in I} |W_N^*(t,x)| > c,$$

where $c$, $d_N \to \infty$ are some constants.

**Proposition 2.1.** *Let the hypothesis* $\mathrm{H}_0$ *be true, i.e.* $(X_1, \ldots, X_N) = \boldsymbol{X}_N$ *is a sample from a strictly stationary process* $\boldsymbol{X}$, *with known marginal distribution function* $F(x) = P\{X_0 \leqslant x\}$. *Let, moreover,*

$$d_N^{-1} W_N(t,x) \overset{D(I)}{\Longrightarrow} W(t,x). \tag{2.2}$$

*Then for a.e.* $c > 0$

$$\lim_{N \to \infty} P\{T_N > c\} = P\{\sup_{(t,x) \in I} |W(t,x)| > c\}. \tag{2.3}$$

**Proof.** Follows from stationarity of $\boldsymbol{X}$ and of the increments $W_N(t,x) - W_N(s,x)$, the convergence (2.2), and the fact that $\sup_I |w(t,x)|$ is a continuous functional on $D(I)$. □

**Remark 2.1.** It is well-known (Lamperti, 1962) that the normalizing constants are necessarily of the form $d_N = N^{\kappa} L(N)$, with some $\kappa > 0$ and $L(\cdot)$ a slowly varying function. The limit random field $W(t,x)$ extends to a random element on $I_\infty = [0, +\infty) \times [-\infty, +\infty]$, denoted by the same letter, and taking values in the Skorokhod space $D(I_\infty)$, which is $\kappa$-*self-similar*, i.e. for any $a > 0$

$$W(at,x) \overset{D(I_\infty)}{=} a^\kappa W(t,x).$$

**Remark 2.2.** For *independent* observations $X_1,\dots,X_N$ the convergence (2.2) is well-known and the limit process is the *Kiefer process* $W(t,x) = K(t,x)$, i.e. a zero mean Gaussian process with the covariance

$$EK(t,x)K(t',x') = t \wedge t'\big(F(x \wedge x') - F(x)F(x')\big). \tag{2.4}$$

Berkes and Philipp (1977) and others obtained the convergence (2.2) for weakly dependent stationary processes $X_j, j \in \mathbb{Z}$ satisfying certain mixing conditions, to a zero mean Gaussian field $W(t,x)$ with the covariance

$$EW(t,x)W(t',x') = t \wedge t\,\sigma(x,x'), \tag{2.5}$$

where

$$\sigma(x,x') = \sum_j \big(P\{X_0 \leqslant x, X_j \leqslant x'\} - P\{X_0 \leqslant x\}P\{X_j \leqslant x'\}\big). \tag{2.6}$$

**Remark 2.3.** The empirical process of long-memory sequences of the form $X_j = H(Y_j)$, $j \in \mathbb{Z}$, where $H(\cdot)$ is a (measurable) function, and $Y_j, j \in \mathbb{Z}$ is a *Gaussian* process with zero mean and slowly decreasing covariance function: $\mathrm{Cov}(Y_0, Y_j) \sim j^{-D}$ $(j \to \infty,\ D \in (0,1))$, was studied in Dehling and Taqqu (1989). Further examples of the empirical process' convergence (2.2) are discussed in Section 4. Apparently, at the present time the change-point problem provides the most important statistical application of such a convergence.

### 2.2. *Testing the null hypothesis when $F(x)$ is unknown*

Consider now the pair $(\tilde{\mathrm{H}}_0, \tilde{\mathrm{H}}_1)$ of alternative hypotheses defined by

$$\tilde{\mathrm{H}}_0 : \{\exists F \text{ such that } \boldsymbol{X}_N \in \Psi_N(F)\},$$

$$\tilde{\mathrm{H}}_1 : \{\exists \theta \in (0,1)\ \exists F_1 \neq F_2 \text{ such that } \boldsymbol{X}_N \in \Psi_N(\theta, F_1, F_2)\}.$$

To test $(\tilde{\mathrm{H}}_0, \tilde{\mathrm{H}}_1)$, we use the statistic

$$\tilde{T}_N := d_N^{-1} \sup_{(t,x)\in I} |V_N(t,x)|,$$

where

$$V_N(t,x) = \frac{[Nt](N-[Nt])}{N}\big(F_{[Nt]}(x) - F^*_{N-[Nt]}(x)\big). \tag{2.7}$$

**Proposition 2.2.** *Assume the hypothesis* $\tilde{\mathrm{H}}_0$ *is true, and the convergence* (2.2) *holds again. Then for a.e.* $c > 0$

$$\lim_{N\to\infty} P\{\tilde{T}_N > c\} = P\{\sup_{(t,x)\in I} |W(t,x) - tW(1,x)| > c\}. \tag{2.8}$$

**Proof.** Write

$$V_N(t,x) = \frac{N-[Nt]}{N} W_N(t,x) - \frac{[Nt]}{N} W_N^*(t,x)$$
$$= (1-t_N)W_N(t,x) - t_N(W_N(1,x) - W_N(t,x)) = W_N(t,x) - t_N W_N(1,x),$$

where $t_N = [Nt]/N \to t$. Hence, the convergence (2.8) follows by the same argument as in the previous proposition. □

### 2.3. *Testing converging alternatives*

The asymptotics of Type II error probability is usually discussed in the context of *contiguous* models (contiguous alternatives). For K-S type statistics (2.3), (2.6), and independent observations $X_1,\dots,X_N$, the limits of the probabilities $P\{T_N > c\}, P\{\tilde{T}_N > c\}$ under contiguous alternatives were studied by Leipus (1988), Szyszkowicz (1994), Csörgő and Szyszkowicz (1994). Khmaladze and Parjanadze (1986), Pardzhanadze and Khmaladze (1986) considered asymptotically most powerful rank tests based on sequential ranks and obtained the weak limit under contiguous alternatives for the uniform empirical process simultaneously with the limits of the corresponding empirical rank processes.

However, the contiguity assumption (on the observations before and after the change point) is rather difficult to verify for certain dependent models, in particular, for the linear model studied in Section 4. Therefore, we introduce a related notion of *converging alternatives*, which is formulated in terms of the joint asymptotics of the corresponding pair of empirical processes, and which was first studied in Giraitis and Leipus (1992) in the context of the empirical spectral process of a moving average process.

Let $\mathscr{X}_2$ be a class of *bivariate* strictly stationary processes $(\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}) = (X_j^{(1)}, X_j^{(2)})_{j\in\mathbb{Z}}$. Write $\mathscr{X}_1 = \{\boldsymbol{X} : \exists \boldsymbol{X}' \text{ such that } (\boldsymbol{X}, \boldsymbol{X}') \in \mathscr{X}_2\}$ for the corresponding class of *univariate* stationary processes. Introduce the class $\Psi_N(\theta; \mathscr{X}_2)$ of all vectors $\boldsymbol{X}_N = (X_1,\dots,X_N)$ such that (2.1) holds with $(\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}) \in \mathscr{X}_2$ and $F^{(i)}(x) = P\{X_j^{(i)} \leqslant x\}$, $i = 1,2$. Let $\Psi_N(\mathscr{X}_1) \equiv \Psi_N(1; \mathscr{X}_2)$ be defined analogously.

In the following definition, the class $\mathscr{X}_2 = \mathscr{X}_2^{(N)}$ and the alternative distribution functions $F^{(i)}(x) = F^{(i,N)}(x)$, $i = 1,2$ depend on the sample size $N$. Put

$$\tilde{\mathrm{H}}_0^{(N)} = \{\boldsymbol{X}_N \in \Psi_N(\mathscr{X}_1^{(N)})\},$$
$$\tilde{\mathrm{H}}_1^{(N)} = \{\exists \theta \in (0,1) \text{ such that } \boldsymbol{X}_N \in \Psi_N(\theta; \mathscr{X}_2^{(N)})\}$$

and

$$W_N^{(i,N)}(t,x) \equiv W_N^{(i,N)}(t,x; \boldsymbol{X}^{(i,N)}) = [Nt]\big(F_{[Nt]}^{(i,N)}(x) - F^{(i,N)}(x)\big),$$

$i = 1,2$, where

$$F_N^{(i,N)}(x) = \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}(X_j^{(i,N)} \leqslant x)$$

is the corresponding empirical distribution function.

**Definition 2.1.** The pair $(\tilde{\mathrm{H}}_0^{(N)}, \tilde{\mathrm{H}}_1^{(N)})$ of alternative hypotheses is said to be converging if there exist $d_N \to \infty$, a real function $G(x)$, $x \in \mathbb{R}$ from the Skorokhod space $D[-\infty, +\infty]$, and a random field $(W^{(1)}(t,x), W^{(2)}(t,x)), (t,x) \in I$ such that, for any pair $(\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}) \in \mathscr{X}_2^{(N)}$,

$$d_N^{-1}(W_N^{(1,N)}(t,x), W_N^{(2,N)}(t,x)) \overset{D(I)}{\Longrightarrow} (W^{(1)}(t,x), W^{(2)}(t,x)) \tag{2.9}$$

and

$$\lim_{N\to\infty} d_N^{-1} N\big(F^{(1,N)}(x) - F^{(2,N)}(x)\big) = G(x) \quad (\text{in } D[-\infty, +\infty]), \tag{2.10}$$

Denote $\tilde{\mathrm{H}}_1^{(N)}(\theta)$ the alternative $\tilde{\mathrm{H}}_1^{(N)}$ when $\theta \in (0,1)$ is fixed.

**Proposition 2.3.** *Assume the pair* $(\tilde{\mathrm{H}}_0^{(N)}, \tilde{\mathrm{H}}_1^{(N)})$ *is converging and* $\theta \in (0,1)$. *Then under the alternative* $\tilde{\mathrm{H}}_1^{(N)}(\theta)$ *for a.e.* $c > 0$

$$\lim_{N\to\infty} P\{\tilde{T}_N > c\} = P\{\sup_{(t,x)\in I} |Z(t,x)| > c\}, \tag{2.11}$$

*where*

$$\begin{aligned} Z(t,x) = {} & (1-t)\Big(W^{(1)}(t\wedge\theta,x) - W^{(2)}(t\wedge\theta,x) + W^{(2)}(t,x)\Big) \\ & -t\Big(W^{(1)}(t\vee\theta,x) - W^{(2)}(t\vee\theta,x) + W^{(2)}(1,x) \\ & -W^{(1)}(t,x)\Big) + (t\wedge\theta - t\theta)G(x). \end{aligned} \tag{2.12}$$

**Proof.** Similar to the proof of Proposition 2.2, write

$$\begin{aligned} V_N(t,x) = {} & (1-t_N)\Big(W_N^{(1,N)}(t\wedge\theta,x) - W_N^{(2,N)}(t\wedge\theta,x) + W_N^{(2,N)}(t,x)\Big) \\ & -t_N\Big(W_N^{(1,N)}(t\vee\theta,x) - W_N^{(2,N)}(t\vee\theta,x) + W_N^{(2,N)}(1,x) - W_N^{(1,N)}(t,x)\Big) \\ & +(t_N\wedge\theta_N - t_N\theta_N)N(F^{(1,N)}(x) - F^{(2,N)}(x)). \end{aligned} \tag{2.13}$$

Hence, the convergence (2.11) follows from Definition 2.1, similarly as in Propositions 2.1 and 2.2. □

**Remark 2.4.** *Independent* observations satisfy the conditions of Proposition 2.3 (Definition 1.1) if $F^{(i,N)}(x)$ weakly converge to a distribution function $F(x)$, $i = 1,2$, and

$$\lim_{N\to\infty} \sqrt{N}(F^{(1,N)}(x) - F^{(2,N)}(x)) = G(x)$$

at each continuity point of $G(x)$, where $G(x)$ has bounded variation. The limit empirical process in (2.9) (with $d_N = \sqrt{N}$ ) is $(W^{(1)}(t,x), W^{(2)}(t,x)) = (K^{(1)}(t,x), K^{(2)}(t,x))$, where $K^{(i)}(t,x)$, $i = 1,2$ are independent Kiefer processes with the same covariance (2.4). Converging alternatives for moving-average observation processes are discussed in Section 4.

**Remark 2.5.** In a similar way, one can discuss testing change-point alternatives, using the *empirical characteristic function*

$$C_N(z) = \int_{\mathbb{R}} e^{izx} dF_N(x) = \frac{1}{N} \sum_{j=1}^{N} e^{izX_j}.$$

The natural counterparts of the statistics $W_N(t,x)$, $W_N^*(t,x)$, $V_N(t,x)$ are their Fourier–Stieltjes transforms; in particular, the null-hypothesis $\tilde{H}_1$ can be tested using

$$U_N(t,z) = \int_{\mathbb{R}} e^{izx} dV_N(t,x) = \frac{[Nt](N-[Nt])}{N}(C_{[Nt]}(z) - C^*_{N-[Nt]}(z)),$$

where $C^*_{N-k}(z) = (N-k)^{-1}\sum_{j=k+1}^{N} e^{izX_j}$. The convergence of the one-parameter process $U_N(1,z)$ for independent and weakly dependent observations was discussed in Feuerverger and Mureika (1977), Csörgő (1981) and Feuerverger (1990). Beran and Ghosh (1990, 1991) consider the above convergence for strongly dependent Gaussian variables.

## 3. Estimation of the change point

In order to be able to consistently estimate the change point $k_N = [N\theta] + 1$, or the parameter $\theta$, we need that the alternative distribution functions $F^{(i,N)}(x)$, $i = 1,2$ converge *more slowly* than the empirical processes $W_N^{(i,N)}(t,x)$, $i = 1,2$. This leads to the following

**Definition 3.1.** A pair $(\tilde{H}_0^{(N)}, \tilde{H}_1^{(N)})$ of converging alternatives is said slowly converging if the convergence (2.9), (2.10) holds with $(W^{(1)}(t,x), W^{(2)}(t,x)) \equiv 0$ and $G(x) \not\equiv 0$.

We consider two types of estimators of $\theta$ based on the uniform distance and the $L^2$-distance between distribution functions, respectively. Put

$$||V_N||_\infty(t) = \sup_x |V_N(t,x)| \tag{3.1}$$

and

$$\widehat{\theta}_{N,\infty} = \arg\max\{||V_N||_\infty(t) : t \in [0,1]\}, \tag{3.2}$$

where $V_N(t,x)$ is defined by (2.7). Also, let

$$\widehat{\rho}_{N,\infty} = \frac{||V_N||_\infty(\widehat{\theta}_{N,\infty})}{\widehat{\theta}_{N,\infty}(1-\widehat{\theta}_{N,\infty})}. \tag{3.3}$$

**Theorem 3.1.** *Let the pair $(\tilde{H}_0^{(N)}, \tilde{H}_1^{(N)})$ be slowly converging. Then under the hypothesis $\tilde{H}_1^{(N)}$ for any $\theta \in (0,1)$*

$$(\widehat{\theta}_{N,\infty}, d_N^{-1}\widehat{\rho}_{N,\infty}) \Longrightarrow (\theta, ||G||_\infty), \quad N \to \infty, \tag{3.4}$$

*where $||G||_\infty = \sup_x |G(x)|$.*

**Proof.** Denote $G(t,x) = (t \wedge \theta - t\theta)G(x)$, then

$$\sup_{(t,x)\in I} |G(t,x)| = \theta(1-\theta)||G||_\infty. \tag{3.5}$$

From the proof of Proposition 2.3, under $\mathrm{H}_1^{(N)}$, we have

$$d_N^{-1} V_N(t,x) \overset{D(I)}{\Longrightarrow} G(t,x),$$

which implies

$$d_N^{-1} \sup_{(t,x)\in I} |V_N(t,x)| \equiv d_N^{-1} ||V_N||_\infty(\widehat{\theta}_{N,\infty}) \Longrightarrow \theta(1-\theta)||G||_\infty. \tag{3.6}$$

Moreover, using the representation (2.13) and the convergence $d_N^{-1} \sup_{(t,x)\in I} |W_N^{(i,N)}(t,x)| \Longrightarrow 0$, we obtain

$$\tau(\widehat{\theta}_{N,\infty}, \theta) N d_N^{-1} ||F^{(1,N)} - F^{(2,N)}||_\infty \Longrightarrow \theta(1-\theta)||G||_\infty$$

or

$$\tau(\widehat{\theta}_{N,\infty}, \theta) \Longrightarrow \theta(1-\theta) \equiv \tau(\theta,\theta),$$

where $\tau(t,\theta) := t \wedge \theta - t\theta$. Consequently, by the inequality $\tau(\theta,\theta) - \tau(t,\theta) \geqslant |t-\theta|((1-\theta)\wedge\theta)$, $t \in (0,1)$ we obtain

$$\widehat{\theta}_{N,\infty} \Longrightarrow \theta. \tag{3.7}$$

Now, (3.4) follows from (3.6), (3.7). □

Apart from the Kolmogorov–Smirnov-type statistics $T_N$, $\tilde{T}_N$, other statistics can be applied in the change-point problem, in particular the Cramér–von Mises-type statistics based on the $L_2$-distance, such as

$$\sup_{0\leqslant t\leqslant 1} \int_{\mathbb{R}} V_N^2(t,x)\,\mathrm{d}x \tag{3.8}$$

or

$$\int_0^1 \int_{\mathbb{R}} V_N^2(t,x)\,\mathrm{d}t\,\mathrm{d}x. \tag{3.9}$$

The asymptotics of the integrals (3.8), (3.9) under the hypothesis of Sect. 2 can be obtained from Propositions 2.1–2.3.

Consider estimation of $\theta$ and of the distance $||F^{(1,N)} - F^{(2,N)}||_2 := \left(\int_{\mathbb{R}} |F^{(1,N)}(x) - F^{(2,N)}(x)|^2 \mathrm{d}x\right)^{1/2}$ using the statistic (3.8). Put

$$||V_N||_2(t) = \left(\int_{\mathbb{R}} V_N^2(t,x)\,\mathrm{d}x\right)^{1/2}$$

and define

$$\widehat{\theta}_{N,2} = \arg\max\{||V_N||_2(t) : t \in [0,1]\},$$
$$\widehat{\rho}_{N,2} = \frac{||V_N||_2(\widehat{\theta}_{N,2})}{\widehat{\theta}_{N,2}(1-\widehat{\theta}_{N,2})}.$$

**Theorem 3.2.** *Let the assumptions of Theorem* 3.1 *be satisfied, and, moreover,* $d_N^{-1}N||F^{(N,1)} - F^{(N,2)}||_2 \to ||G||_2$. *Then*

$$(\widehat{\theta}_{N,2}, d_N^{-1}\widehat{\rho}_{N,2}) \Longrightarrow (\theta, ||G||_2), \quad N \to \infty.$$

**Proof.** The convergence (3.6) implies

$$d_N^{-2} \sup_{t\in[0,1]} \int_{\mathbb{R}} V_N^2(t,x)\mathrm{d}x = d_N^{-2}||V_N||_2^2(\widehat{\theta}_{N,2}) \Longrightarrow \theta^2(1-\theta)^2||G||_2^2,$$

from which the convergences $\widehat{\theta}_{N,2} \Longrightarrow \theta$ and $d_N^{-1}\widehat{\rho}_{N,2} \Longrightarrow ||G||_2$ follow, similarly as in the proof of Theorem 3.1. □

**Remark 3.1.** In the case of independent observations, Dümbgen (1991) studied the estimator $\widehat{\theta}_{N,D}$ which maximizes

$$s_N\left(\frac{1}{[Nt]}\sum_{1\leqslant j\leqslant[Nt]}\delta_{X_j} - \frac{1}{N-[Nt]}\sum_{[Nt]<j\leqslant N}\delta_{X_j}\right),$$

where $s_N(\cdot)$ is a seminorm on the $N$-dimensional Euclidean space and $\delta_x$ is Dirac's measure. In particular, he showed that if $\gamma_N > 0$ satisfy $s_N(F^{(1,N)} - F^{(2,N)}) \geqslant C_0\gamma_N^{-1}$ for some $C_0 > 0$ and all sufficiently large $N$ and $\gamma_N^2 = \mathrm{o}(N/\log\log N)$, then

$$\widehat{\theta}_{N,D} = \theta + \mathrm{O}_P(\gamma_N^2/N).$$

The estimators of Carlstein (1988) and Darkhovsky (1976) are special cases of Dümbgen's estimator $\widehat{\theta}_{N,D}$. See also Carlstein and Lele (1993) and Ferger (1994b) for recent results on this estimator.

**Remark 3.2.** Ferger and Stute (1992) studied an U-statistic-type estimator $\widehat{\theta}_{N,U}$ for the parameter $\theta$, when the observations are independent and the alternatives are fixed, i.e. $F^{(i,N)} = F^{(i)}$, $i = 1,2$ for all $N$, and showed that $\widehat{\theta}_{N,U} - \theta = \mathrm{O}(\ln N/N)$ with probability 1. In the recent paper, Ferger (1994a) obtained the asymptotic distribution of a related class of max-type estimators of $\theta$ when the alternatives approach each other in a certain sense.

## 4. The empirical process of long-memory sequences

Conditions of Propositions 2.1–2.3 are well-studied for independent or, at least what concerns the basic convergence (2.2), for weakly dependent observations $X_j$, $j \in \mathbb{Z}$. In recent years, there is a considerable interest in statistical inference for long-memory time series, including the behavior of empirical processes (Beran, 1992; Dehling and Taqqu, 1989). One of the basic models is the *moving-average* process

$$X_j = \sum_{s \leqslant j} b_{j-s}\xi_s, \tag{4.1}$$

where $b_s$, $s \in \mathbb{Z}_+ = \{0, 1, ...\}$ are (non-random) weights such that $\sum b_s^2 < \infty$, and $\xi_s$, $s \in \mathbb{Z}$ is a (noise) sequence of i.i.d. random variables, not necessarily Gaussian, with zero mean and variance 1. The long-memory condition is usually introduced by requiring that the weights decay slowly hyperbolically:

$$b_s = L(s)\, s^{-(1+D)/2}, \tag{4.2}$$

where $0 < D < 1$, and $L(\cdot)$ is a slowly varying function. Condition (4.2) guarantees the corresponding hyperbolic decay condition

$$\mathrm{Cov}(X_0, X_j) = \tilde{L}(j) j^{-D} \tag{4.3}$$

of the covariance, with a slowly varying function $\tilde{L}(j) \sim dL^2(j)$, where $d = \int_0^\infty (u(1+u))^{-(1+D)/2}\,\mathrm{d}u$. The series (4.1), (4.2) include *fractional ARIMA* models defined by

$$\Phi(B)(1-B)^{D/2}X_j = \Psi(B)\xi_j,$$

where $BX_j = X_{j-1}$ is the backshift operator, $(1-B)^{D/2} = \sum_{k=0}^{\infty} \binom{D/2}{k}(-B)^k$ is the fractional difference operator, and $\Phi(z)$, $\Psi(z)$ are polynomials satisfying usual conditions (Granger and Joyeux, 1980; Hosking, 1981).

Statistical analysis of (non-Gaussian) long-memory series (4.1), (4.2) is not easy since the usual techniques of Hermite expansions do not apply. Similar to the Gaussian case, even quadratic statistics may tend to a non-Gaussian limit, and the proofs of the convergence often are technically complicated, see e.g. Giraitis and Surgailis (1990). The first result on the convergence of the empirical process was recently obtained in Giraitis and Surgailis (1994) (see also Giraitis, et al., 1994). Introduce the *fractional Brownian motion* $Z_D(t)$, $t \in [0, 1]$ $(0 < D < 2)$, which is a (a.s. continuous) Gaussian process with zero mean and the covariance

$$EZ_D(t)Z_D(s) = \tfrac{1}{2}\left(|t|^{2-D} + |s|^{2-D} - |t-s|^{2-D}\right).$$

The *fractional Brownian bridge* $Z_D^{(0)}(t)$, $t \in [0, 1]$ can be defined by

$$Z_D^{(0)}(t) = Z_D(t) - tZ_D(1).$$

Put

$$d_N = d^{1/2} L(N) N^{1-D/2}. \tag{4.4}$$

**Theorem 4.1** (Giraitis and Surgailis, 1994). *Let* $0 < D < 1$, *and let for the moving-average process* $X_j$ *of* (4.1), (4.2), *the following conditions be satisfied*:

$$|E e^{iu\xi_0}| \leqslant C(1+|u|)^{-\gamma} \quad (\exists C < \infty,\ \exists \gamma > 0) \tag{4.5}$$

*and*

$$E|\xi_0|^m < \infty \quad (\forall m > 0). \tag{4.6}$$

*Then*

$$d_N^{-1} W_N(t,x) \overset{D(I)}{\Longrightarrow} c_D^{1/2} f(x) Z_D(t), \tag{4.7}$$

*where* $f(x) = F'(x) = \mathrm{d}P\{X_j \leqslant x\}/\mathrm{d}x$ *is the marginal probability density*, $Z_D(t)$, $t \in [0,1]$ *is a fractional Brownian motion, and*

$$c_D = \int_0^1 \int_0^1 |t-s|^{-D} \mathrm{d}t\,\mathrm{d}s = 2(1-D)^{-1}(2-D)^{-1}.$$

**Remark 4.1.** The limit random field $W(t,x) = c_D^{1/2} f(x) Z_D(t)$ in Theorem 4.1 coincides with the corresponding limit in Dehling and Taqqu (1989), Theorem 1.1 for Gaussian $X_j, j \in \mathbb{Z}$. Condition (4.6) can be relaxed in the sense that moments of $\xi_0$ of a sufficiently high order may be infinite. The proof of Theorem 4.1 is based on the following "weak uniform reduction principle" (cf. Theorem 3.1 of Dehling and Taqqu (1989)): there are constants $C(\delta), \gamma > 0$ such that for any $0 < \delta < 1$

$$P\{\sup_I d_N^{-1} |W_N(t,x) + f(x) S_{[Nt]}| > \delta\} \leqslant C(\delta) N^{-\gamma}, \tag{4.8}$$

where

$$S_N = \sum_{j=0}^{N} X_j. \tag{4.9}$$

Fix $0 < D < 1$ and a slowly varying function $L_0(\cdot)$, and consider the class $\mathscr{X}_1 = \mathscr{X}_1(D, L_0(\cdot))$ of all moving-average stationary processes $\boldsymbol{X} \equiv (X_j)_{j\in\mathbb{Z}}$ of (4.1), (4.2) with

$$\lim_{j\to\infty} L(j)/L_0(j) = 1,$$

and satisfying the conditions of Theorem 4.1. From Proposition 2.2 and Theorem 4.1 it follows

**Corollary 4.1.** *Let the hypothesis* $\tilde{\mathrm{H}}_0 = \tilde{\mathrm{H}}_0(\mathscr{X}_1)$ *be true, i.e.* $(X_1, \ldots, X_N)$ *is a sample from a stationary moving average process* $\boldsymbol{X} \in \mathscr{X}_1$. *Then for any* $c > 0$

$$P\{\tilde{T}_N > c\} = P\{d_N^{-1} \sup_{(t,x)\in I} |V_N(t,x)| > c\} = P\{\sup_{t\in[0,1]} |Z_D^{(0)}(t)| > c/(c_D^{1/2} \|f\|_\infty)\}.$$

There are several ways one can discuss the change-point problem for converging alternatives about the moving average model, by specifying an appropriate class $\mathscr{X}_2^{(N)}$. One possibility is as follows. Fix $0 < D < 1$, and a slowly varying function $L(\cdot)$. Put $b_j = L(j)j^{-(1+D)/2}$, $j \in \mathbb{Z}_+$. Let $\mathscr{X}_2^{(N)} = \mathscr{X}_2^{(N)}(D,L)$ be the class of all pairs $(\boldsymbol{X}^{(1,N)}, \boldsymbol{X}^{(2,N)})$ of moving averages of the form

$$X_j^{(i,N)} = \sum_{s\leqslant j} b_{j-s}^{(i,N)} \xi_s, \quad i = 1,2, \tag{4.10}$$

satisfying the following four conditions:

(a.1) $b_j^{(i,N)} = L^{(i,N)}(j)j^{-(1+D)/2}$, where $L^{(i,N)}(\cdot)$ varies slowly at infinity;

(a.2) $b_j^{(i,N)} = b_j(1 + o(1))$ as $N \to \infty$ uniformly in $j \geqslant 0$;

(a.3) $\xi_s$, $s \in \mathbb{Z}$ are i.i.d. and satisfy conditions (4.5), (4.6) of Theorem 4.1;

(a.4) $\lim_{N\to\infty} d_N^{-1} N(f^{(i,N)}(x) - f(x)) = g^{(i)}(x)$ uniformly on compacts and in $L^1(\mathbb{R})$, where $f^{(i,N)}(x) = dF^{(i,N)}(x)/dx$ is the marginal density of $X^{(i,N)}(j)$ (4.10), and $f(x)$ is the density of $X_j = \sum_{s\leqslant j} b_{j-s}\xi_s$.

Let $(\tilde{\mathrm{H}}_0^{(N)}, \tilde{\mathrm{H}}_1^{(N)})$ be the pair of alternative hypotheses defined in Section 2, and corresponding to the class $\mathscr{X}_2^{(2)} = \mathscr{X}_2^{(N)}(D,L)$. In particular, the alternative $\tilde{\mathrm{H}}_1^{(N)}(\theta)$ is

$$X_j = \begin{cases} \sum_{s\leqslant j} b_{j-s}^{(1,N)} \xi_s, & 1 \leqslant j \leqslant [N\theta], \\ \sum_{s\leqslant j} b_{j-s}^{(2,N)} \xi_s, & [N\theta] < j \leqslant N. \end{cases}$$

**Theorem 4.2.** *The pair* $(\tilde{\mathrm{H}}_0^{(N)}, \tilde{\mathrm{H}}_1^{(N)})$ *is converging in the sense of Definition* 2.1, *with* $d_N$ *given by* (4.4),

$$W^{(1)}(t,x) = W^{(2)}(t,x) = c_D^{1/2} f(x) Z_D(t), \quad i = 1,2,$$

$Z_D(t)$ *being the fractional Brownian motion, and*

$$G(x) = \int_{-\infty}^{x} (g^{(1)}(y) - g^{(2)}(y))\mathrm{d}y.$$

**Proof.** From the assumptions (a.1)–(a.3) analogously as in the proof of Theorem 4.1 (see Giraitis and Surgailis, 1994) one obtains the "weak uniform reduction principle":

$$P\{\sup_I d_N^{-1} |W_N^{(i,N)}(t,x) + f^{(i,N)}(x) S_{[Nt]}^{(i,N)}| > \delta\} \leqslant C(\delta) N^{-\gamma}, \quad i = 1,2,$$

where the constants $C(\delta), \gamma > 0$ do not depend on $N$. Hence, the convergence (2.9) follows from (a.4) and

$$d_N^{-1} (S_{[Nt]}^{(1,N)}, S_{[Nt]}^{(2,N)}) \overset{D[0,1]}{\Longrightarrow} c_D^{1/2} (Z_D(t), Z_D(t)). \tag{4.11}$$

With (a.1)–(a.3) in mind, the last convergence is a rather simple fact; see Giraitis and Surgailis (1994), or Taqqu (1975). In particular, the asymptotic normality of finite dimensional distributions can be directly verified by computing cumulants of the left-hand side of (4.11). Finally, (2.10) obviously follows from (a.4).

**Corollary 4.2.** *Under the alternative* $\tilde{\mathrm{H}}_1^{(N)}$ $(0 < \theta < 1)$ *for any* $c > 0$

$$\lim_{N\to\infty} P\{\tilde{T}_N > c\} = P\{\sup_{(t,x)\in I} |c_D^{1/2} f(x) Z_D^{(0)}(t) + (t \wedge \theta - t\theta)G(x)| > c\}.$$

## References


Basseville, M. and I.V. Nikiforov (1993). *Detection of Abrupt Changes: Theory and Application.* Prentice Hall, Englewood Cliffs, NJ.

Beran, J.(1992). Statistical methods for data with long-range dependence (with discussions). *Statist. Sci.* **7**, 404–427.

Beran, J. and S. Ghosh (1990). Goodness of fit tests and long-range dependence. In: *Directions in Robust Statistics and Diagnostics, Part I. IMA Vol. Math. Appl.* **33**. Springer, New York, 21–33.

Beran, J. and S. Ghosh (1991). Slowly decaying correlations, testing normality, nuisance parameters. *J. Amer. Statist. Assoc.* **86**, 785–791.

Berkes, I. and W. Philipp (1977). An almost sure invariance principle for the empirical distribution function of mixing random variables. *Z. Wahrsch. Verw. Geb.* **41**, 115–137.

Bickel, P.J. and M.J. Wichura (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Statist.* **42**, 1656–1670.

Brodsky, B.E. and B.S. Darkhovsky (1993). *Nonparametric Methods in Change-Point Problems.* Mathematics and its Applications, Vol. **243**. Kluwer. Dordrecht.

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

Carlstein, E. and S. Lele (1993). Nonparametric change-point estimation for data from an ergodic sequence. *Theory Probab. Appl.* **38**, 910–917.

Csörgő, S. (1981). Limit behaviour of the empirical characteristic function. *Ann. Probab.* **9**, 130–144.

Csörgő, M. and L. Horváth (1988). Nonparametric methods for the change-point problems. In: P.R. Krishnaiah and C.R. Rao, Eds., *Handbook of Statistics, Vol. 7.* Elsevier, Amsterdam, 403–425.

Csörgő, M. and B. Szyszkowicz (1994). Applications of multitime parameter processes to change-point analysis. In: B. Grigelionis et al, Eds., *Probab. Theory and Mathematical Statistics, Proc. 6th Vilnius Conference, 1993.*, TEV - VSP, Vilnius - Utrecht, 159–222.

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

Dehling, H. and M.S. Taqqu (1989). The empirical process of some long-range dependent sequences with an application to U -statistics. *Ann. Statist.* **17**, 1767–1783.

Deshayes, J. and D. Picard (1986). Off-line statistical analysis of change-point models using non parametric and likelihood methods. In: *M. Basseville and A. Benveniste, Eds., Detection of Abrupt Changes in Signals and Dynamical Systems, Lecture Notes Cont. Inf. Sci., Vol. 77.* Springer, Berlin, 103–168.

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 the rate of almost sure convergence of Dümbgen's change-point estimators. *Statist. Probab. Lett.* **19**, 27–31.

Ferger, D. and W. Stute (1992). Convergence of changepoint estimators. *Stochastic. Proc. Appl.* **42**, 345–351.

Feuerverger, A. (1990). An efficiency result for the empirical characteristic function in stationary time-series models. *Canad. J. Statist.* **18**, 155–161.

Feuerverger, A. and R.A. Mureika (1977). The empirical characteristic function and its applications. *Ann. Statist.* **5**, 88–97.

Giraitis, L., H. Koul and D. Surgailis (1994). Asymptotic normality of regression estimators with long memory errors, *preprint.*

Giraitis, L. and R. Leipus (1992). Testing and estimating in the change-point problem for the spectral function. *Lithuanian Math. J.* **32**, 20–38.

Giraitis, L. and D. Surgailis (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle's estimate. *Probab. Theory Related Fields* **86**, 87–104.

Giraitis, L. and D. Surgailis (1994). Asymptotics of the empirical process of a long memory linear sequence. Beiträge zur Statistik, 24. Universität Heidelberg.

Granger, C.W. and R. Joyeux (1980). An introduction to long-memory time series models and fractional differencing. *J. Time Series Anal.* **1**, 15–29.

Hawkins, D.L. (1988). Retrospective and sequential tests for a change in distribution based on Kolmogorov – Smirnov-type statistics. *Seq. Anal.* **7**, 23–51.

Hosking, J.R.M. (1981). Fractional differencing. *Biometrika* **68**, 165–176.

Khmaladze, E.V. and A.M. Parjanadze (1986). Functional limit theorems for linear statistics from sequential ranks. *Probab. Theory Related Fields* **73**, 585–595.

Lamperti, J. (1962). Semi-stable stochastic processes. *Trans. Amer. Math. Soc.* **104**, 62–78.

Leipus, R. (1988). Weak convergence of two parameter empirical fields in change-point problems. *Lithuanian Math. J.* **28**, 716–723.

Pardzhanadze, A.M. and E.V. Khmaladze (1986). On the asymptotic theory of statistics from sequential ranks. *Theory Probab. Appl.* **31**, 669–682.

Picard, D. (1985). Testing and estimating change-points in time series. *Adv.Appl. Probab.* **17**, 841–867.

Szyszkowicz, B. (1994). Weak convergence of weighted empirical type processes under contiguous and changepoint alternatives. *Stochastic Process Appl.* **50**, 281–313.

Taqqu, M.S. (1975). Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z. Wahrsch. Verw. Geb.* **31**, 287–302.