

Non-Parametric Weighted Tests for Change in Distribution Function

Abd-Elnaser S. Abd-Rabou*, Ahmed M. Gad

Statistics Department, Faculty of Economics and Political Science, Cairo University, Cairo, Egypt

Abstract For a given data set it may be required to discover if a change has been occurred. This is can be conducted using change-point analysis. Let X_1, X_2, \dots, X_n be independent random variables with respective continuous distribution functions F_1, F_2, \dots, F_n such that $F_i(0)=0$ for all i . We consider the problem of testing the null hypothesis that $F_1 = F_2 = \dots = F_n$ against the alternative of r -changes in the distribution functions of this sequence at unknown times $1 < [n\tau_1] < [n\tau_2] < \dots < [n\tau_r] < n$, where $[y]$ is the integer part of y . We study the asymptotic theory of change-point processes which are defined in terms of the empirical process. We propose and study new weighted non-parametric change-point test statistics for a possible change in distribution function of a data set.

Keywords Change Point Problem, Empirical Processes, Kiefer Process, Limit Theorems, Monte Carlo Simulation

1. Introduction

In many applications it is necessary to discover if a data set comes from a single distribution or there is a change in the distribution function. The change-point inference is an effective and powerful statistical tool for determining if and when a change in a set of data has occurred. Let X_1, X_2, \dots, X_n be independent random variables with continuous distribution functions (DF) F_1, F_2, \dots, F_n , respectively such that $F_i(0)=0$ for all i . We are interested in testing the null hypothesis of no change

$$H_0: F_1 = F_2 = \dots = F_n = F, \quad (1)$$

where F is unknown against the alternative of at most r -changes (AMRC),

$$H_1^{(r)}: F_i = \begin{cases} G_1, & i \leq [n\tau_1] \\ G_2, & [n\tau_1] < i < [n\tau_2] \\ \vdots & \vdots \\ G_{r+1}, & [n\tau_r] < i \leq n, \end{cases} \quad (2)$$

where $r \geq 1$ is specified, the distribution functions $G_i, i = 1, \dots, r+1$ and the change-point positions $\tau_i, i = 1, \dots, r$ are unknown $G_i \neq G_{i+1}, i = 1, \dots, r$. Note that $[\tau]$ denotes the integer part of τ .

The aim of this article is to introduce weighted non-parametric tests for distributional change in a data set. The asymptotic distributions of these test statistics are derived. These tests also can be applied to the mean change in a data set.

The paper is organized as follows. In Section 2 we will consider the above multiple change-point problem in the case of at most two change-points (AMTC), i.e. $r=2$. In Section 3 we generalize the AMTC results to the case of $r \geq 2$. The proposed new change-point test statistics are presented in Section 4. Also, the asymptotic distributions of the proposed test statistics are derived in Section 4. In Section 5, we propose new test statistics for the case of at most one change point. We study the applicability of the proposed tests through a Monte Carlo study in section 6.

2. The Case of at Most Two Change Points (AMTC)

In this section we treat the case of at most 2-changes (AMTC), i.e. testing H_0 of (1) against the alternative of (2), when $r = 2$.

Many authors have discussed the change-point problem, testing (detection), and estimation using both Bayesian and non-Bayesian approaches. Most of the work done in the change-point analysis is concerned with the case of at most one change (AMOC). Csörgő and Horváth[1] gave a very excellent and extensive treatment and review for the related work. For the AMTC, recent works are done by [2,3]. They propose weighted CUSUM tests for the multiple changes in the variance of a sequence of independent random variables. By using the general form of U-statistic, they studied some CUSUM test statistics for the AMTC in variance. Hawkins[4] developed a dynamic program to test for a multiple change point in the parameters of the general exponential family using repeated maximum likelihood algorithm. His algorithm involves the application of the at most one change-point maximum likelihood detector for every

* Corresponding author:

abdrfeps@feps.edu.eg (Abd-Elnaser Abd-Rabou)

Published online at <http://journal.sapub.org/ajms>

Copyright © 2013 Scientific & Academic Publishing. All Rights Reserved

segmented group of data available.

Szyzykiewicz[5]studied weighted approximations for various versions of the empirical processes under the null and continuous alternatives of the AMOC. She approximated these processes by appropriate Gaussian processes.

Let the smoothed two-parameter empirical process be given by

$$\alpha_n(s, t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[(n+1)t_1]} \{F(X_i) \leq s\} - s, \quad 0 \leq s, t \leq 1. \quad (3)$$

Let $\underline{t} = (t_1, t_2)$, where $0 \leq t_1 \leq t_2 < 1$. For $0 \leq t_1 \leq t_2 < 1, 0 \leq s \leq 1, n \geq 1$, define

$$\beta_n(s, \underline{t}) = \left\{ 1 - \frac{[(n+1)t_1]}{n} \right\} \alpha_n(s, t_2) + \frac{[(n+1)t_2]}{n} \{ \alpha_n(s, t_1) - \alpha_n(s, 1) \}, \quad (4)$$

and

$$\prod_n(s, \underline{t}) =$$

$$\{ (1 - t_1)K(s, nt_2) + t_2K(s, nt_1) - t_2K(s, n) \} / \sqrt{n}, \quad (5)$$

where $\{K(s, y); 0 \leq s \leq 1, 0 \leq y < \infty\}$ is a Kiefer process, i.e., a mean zero two-parameters Gaussian process with

$$E(K(s_1, y_1)K(s_2, y_2)) = (s_1 \wedge s_2 - s_1s_2)(y_1 \wedge y_2). \quad (6)$$

The following result follows from Theorem 8.2.1 of[5].

Theorem 1.

Assume that H_0 of Eq. (1) holds true. Then, there exists a Kiefer process $K(\dots)$ such that as $n \rightarrow \infty$,

$$\sup_{0 \leq s \leq 1} \sup_{0 < t_1 \leq t_2 < 1} \left| \beta_n(s, \underline{t}) - \prod_n(s, \underline{t}) \right| = o_p(1),$$

where $\beta_n(\dots)$ is as in Eq. (4) and $\prod_n(\dots)$ is as in Eq. (5).

Let $\alpha_n(\dots)$ be as in Eq. (3) and $\beta_n(s, \underline{t})$ be as in Eq. (4).

Define

$$A_n(s, y) = \left(\alpha_n(s, y) - \frac{[(n+1)y]}{n} \alpha_n(s, 1) \right). \quad (7)$$

Note that

$$\beta_n(s, \underline{t}) = \left(1 - \frac{[(n+1)t_1]}{n} \right) A_n(s, t_2) + \frac{[(n+1)t_2]}{n} A_n(s, t_1).$$

Let Q be the class of positive functions on $(0, 1)$, which are non-decreasing in a neighbourhood of zero and non-increasing in the neighbourhood of one. A function q defined on $(0, 1)$ is called positive if $\inf_{\delta \leq t \leq 1-\delta} q(t) > 0, \forall \delta \in (0, 1/2)$.

Let $q_1, q_2 \in Q$, the AMTC weighted-process, $\beta_n^w(\dots)$ is defined as

$$\beta_n^w(s, t) = \left(1 - \frac{[(n+1)t_1]}{n} \right) \frac{A_n(s, t_2)}{q_2(t_2)} + \left(\frac{[(n+1)t_2]}{n} \right) \frac{A_n(s, t_1)}{q_1(t_1)}. \quad (8)$$

Let $K(\dots)$ be a Kiefer process. Define

$$\Gamma_n(s, t) = \frac{K(s, nt) - tK(s, n)}{\sqrt{n}}, \quad 0 \leq s, t \leq 1$$

and

$$\Gamma_n^w(s, \underline{t}) = (1 - t_1) \frac{\Gamma_n(s, t_2)}{q_2(t_2)} + t_2 \frac{\Gamma_n(s, t_1)}{q_1(t_1)}. \quad (9)$$

Note that

$$\Gamma_n(s, t) \triangleq \Gamma(s, t) = K(s, t) - tK(s, 1) \triangleq B_1(s)B_2(t), \quad 0 \leq s, t \leq 1, \quad (10)$$

where $B_1(\dots)$ and $B_2(\dots)$ are two independent Brownian bridge.

Now we give the main theorem of this section.

Theorem 2.

Under H_0 of (1), there exists a Kiefer process $K(\dots)$ such that as $n \rightarrow \infty$,

1. if for $i=1,2$

$$\lim_{t \downarrow 0} \frac{t(1-t) \log \log \{ (t(1-t))^{-1/2} \}}{q_i(t)} = 0,$$

and

$$\lim_{t \uparrow 0} \frac{t(1-t) \log \log \{ (t(1-t))^{-1/2} \}}{q_i(t)} = 0,$$

then

$$\sup_{0 \leq s \leq 1} \sup_{0 \leq t_1, t_2 < 1} |\beta_n^w(s, \underline{t}) - \Gamma_n^w(s, \underline{t})| = o_p(1)$$

2. if for $i=1,2$

$$\lim_{t \downarrow 0} \frac{t(1-t) \log \log \{ (t(1-t))^{-1/2} \}}{q_i(t)} < \infty,$$

$$\lim_{t \uparrow 1} \frac{t(1-t) \log \log \{ (t(1-t))^{-1/2} \}}{q_i(t)} < \infty,$$

then

$$\sup_{0 \leq s \leq 1} \sup_{0 \leq t_1, t_2 < 1} |\beta_n^w(s, \underline{t}) - \Gamma_n^w(s, \underline{t})| = O_p(1)$$

and

$$\sup_{0 \leq s \leq 1} \sup_{0 \leq t_1, t_2 < 1} |\beta_n^w(s, \underline{t})| \xrightarrow{D} \sup_{0 \leq s \leq 1} \sup_{0 \leq t_1, t_2 < 1} |\Gamma_n^w(s, \underline{t})|.$$

Proof of Theorem 2 (Sketch)

First we can easily notice that

$$A_n(s, t_i) := \bar{\alpha}_n(s, t_i), \quad 0 \leq s, t_i \leq 1, \quad i = 1, 2 \quad (11)$$

where the right-hand side is the two-time parameter empirical process of[5].

Second if we put $q_i(t_i) = 1, i = 1, 2$ in (6.1.13) of[5], we get as $n \rightarrow \infty$

$$\sup_{0 \leq t_i \leq 1} \left| t_i - \frac{[(n+1)t_i]}{n} \right| = o(1). \quad (12)$$

Using the definition of the processes of Eq. (8) and Eq. (9), the statements of Eq. (11), Eq. (12) and Theorem 8.3.1 of[5], we complete the proof of this theorem.

Now, let $\beta_n^w(\dots)$ be the process defined in Eq. (8), then by Theorem (2) and the relations in Eq. (9) and Eq. (10), we have

$$\beta_n^w(s, \underline{t}) \xrightarrow{D} \Gamma_n^w(s, \underline{t}),$$

where

$$\beta_n^w(s, \underline{t}) = B_1(s) \left\{ \frac{(1-t_1)B_2(t_2)}{q_2(t_2)} + \frac{t_2B_2(t_1)}{q_1(t_1)} \right\}, \quad (13)$$

and $B_i(\dots), i = 1, 2$ are as in Eq. (10).

As in Pouliot[3], we may use the one weight function

$$w(\underline{t}) = \frac{q_1(t_1) + q_2(t_2)}{2}, \quad 0 < t_1 \leq t_2 < 1,$$

for the whole process. It is very easy to see that Theorem 2 remains true under the one-weight function $w(\underline{t})$. In this case the corresponding limiting process of Eq. (13) becomes

$$\Gamma^w(s, \underline{t}) = \frac{1}{w(\underline{t})} B_1(s) B_2^*(\underline{t}),$$

where

$$B_2^*(\underline{t}) = (1 - t_1)B_2(t_2) + t_2B_2(t_1),$$

and $B_i(\cdot), i = 1, 2$ are independent Brownian bridge. It is clear that $\Gamma^w(s, \underline{t})$ is a mean zero Gaussian process with covariance function

$$C(\underline{s}, \underline{t}, \underline{y}) = E \{ \Gamma^w(s_1, \underline{t}) \Gamma^w(s_2, \underline{y}) \} = \frac{1}{w(\underline{t})w(\underline{y})} (s_1 \wedge s_2 - s_1 s_2) C_B(\underline{t}, \underline{y}),$$

where

$$C_B(\underline{t}, \underline{y}) = (1 - t_1)[(1 - r_1)(t_2 \wedge y_2) + y_2(t_2 \wedge y_1)] + t_2[(1 - r_1)(t_1 \wedge y_2) + y_2(t_1 \wedge y_1)] - t_2 y_2.$$

3. The Case of at Most r Change Points (AMRC)

We consider here the general case of $r \geq 1$. Following the definition of the change-point processes in (6.5) of Poulriot (2001), we define the weighted r -change point empirical process and its corresponding weighted Gaussian process as follows. Assume that $q_j(\cdot) \in Q, j = 1, \dots, r$, satisfy the two assumptions of part (1) of Theorem 2, we define the weighted r -time parameter empirical process

$$M_n^w(s, \underline{t}) := n^{-3/2} \left\{ \sum_{m=1}^{r-1} \frac{((n+1)t_{m+1}] - [(n+1)t_{m-1}]}{q_m(t_m)} A_n(s, t_m) + (1 - tr - 1)(1 - tr)[(n+1)tr] q_r(tr) \alpha_n(s, 1) \right\}, \quad (14)$$

where $0 = t_0 \leq t_1 \leq t_2 \leq \dots \leq t_r \leq t_{r+1} = 1$ and the processes $\alpha_n(\cdot, \cdot)$ and $A_n(\cdot, \cdot)$ are defined by Eq. (3) and Eq. (7) respectively. We also define the weighted r -time parameter limiting Gaussian process as follows;

$$\Lambda_n^w(s, \underline{t}) := \sum_{m=1}^{r-1} \left\{ (t_{m+1} - t_{m-1}) \frac{K^*(s, nt_m)}{\sqrt{n}} \right\} + 1 - tr - 1 - trtr q_r(tr) K(s, n), \quad (15)$$

where $K(\cdot, \cdot)$ is the Kiefer process of Eq. (6) and

$$K^*(s, nt) = K(s, nt) - tK(s, n), \quad 0 \leq s, t \leq 1.$$

Now, following the steps of the proof of Theorem 2, we can state the general weighted-sup metric approximation for the r -time parameter empirical process of Eq. (14).

Theorem 3.

Under the null hypothesis H_0 of (1), there exists a Kiefer process $K(\cdot, \cdot)$ such that with the sequence of processes $M_n^w(\cdot, \underline{t})$ and $\Lambda_n^w(\cdot, \underline{t})$ of (14) and (15) respectively, we have $asn \rightarrow \infty$

$$\sup_{0 \leq s \leq 1} \sup_{0 < t_1 \leq t_2 \leq \dots \leq t_r < 1} |M_n^w(s, \underline{t}) - \Lambda_n^w(s, \underline{t})| = o_p(1).$$

Under the conditions of Theorem 3, we obtain

$$\sup_{0 \leq s \leq 1} \sup_{0 < t_1 \leq t_2 \leq \dots \leq t_r < 1} |M_n^w(s, \underline{t})| \xrightarrow{D} |\Lambda^w(s, \underline{t})|,$$

where $\Lambda^w(s, \underline{t})$ is the process in Eq. (15) when $n=1$.

4. The Proposed AMRC Test Statistics

To introduce our proposed multiple change-point test statistics, we need the following integrated processes. Let $\alpha_n(\cdot, \cdot)$ be the smoothed two-parameters empirical process defined by Eq. (3). Then integrating over $0 \leq s \leq 1$, we get $\alpha_n^*(t) = \int_0^1 \alpha_n(s, t) ds = \frac{1}{\sqrt{n}} \left\{ \frac{[(n+1)t]}{2} - \sum_{i=1}^{[(n+1)t]} F(X_i) \right\}$ (16) and define the integrated empirical process difference of Eq. (4) as;

$$A_n^*(t_i) = \alpha_n^*(t_i) - \frac{[(n+1)t_i]}{n} \alpha_n^*(1), \quad i = 1, 2, \dots, r. \quad (17)$$

The generalized test statistics' integrated processes in the case of $r \geq 1$, AMRC, and $\underline{t} = (t_0, t_1, \dots, t_r, t_{r+1})$, such that $0 = t_0 < t_1 \leq \dots \leq t_r < t_{r+1} = 1$ are given by

$$M_n^*(\underline{t}) = \int_0^1 M_n(s, \underline{t}) ds = n^{-\frac{3}{2}} \left\{ \sum_{m=1}^{r-1} \left([(n+1)t_{m+1}] - [(n+1)t_{m-1}] \right) A_n^*(t_m) + (1 - [(n+1)t_{r-1}]) (1 - [(n+1)t_r]) \times [(n+1)t_r] \alpha_n^*(1) \right\}$$

where $\alpha_n^*(\cdot)$ is defined by Eq. (16) and its corresponding limiting Gaussian process is

$$\Lambda^*(\underline{t}) = \int_0^1 \Lambda(s, \underline{t}) ds = \int_0^1 B_1(s) ds \sum_{m=1}^r (t_{m+1} - t_{m-1}) B_2(t_m),$$

where $\Lambda(\cdot, \underline{t})$ is defined by Eq. (15) and $B(\cdot)$ is a standard Brownian bridge defined on the same probability space.

Next, we define the weighted processes $M_n^w(\underline{t})$ and $\Lambda^w(\underline{t})$, that are needed to construct the AMRC test statistics.

For $\underline{t} = (t_0, t_1, \dots, t_r, t_{r+1})$, such that $0 = t_0 < t_1 \leq \dots \leq t_r < t_{r+1} = 1$, we define the following weighted processes.

$$M_n^*(\underline{t}) = n^{-\frac{3}{2}} \left\{ \sum_{m=1}^{r-1} \frac{((n+1)t_{m+1}] - [(n+1)t_{m-1}]}{q_m(t_m)} A_n^*(t_m) + (1 - [(n+1)t_{r-1}]) (1 - [(n+1)t_r]) \times \frac{[(n+1)t_r]}{q_r(t_r)} \alpha_n^*(1) \right\} \quad (18)$$

and

$$\Lambda^w(\underline{t}) := \int_0^1 B_1(s) ds \sum_{m=1}^r \frac{(t_{m+1} - t_{m-1})}{q_m(t_m)} B_2(t_m), \quad (19)$$

where $\alpha_n^*(\cdot)$ and $A_n^*(\cdot)$ are given by Eq. (16) and Eq. (17) respectively and $q_j(\cdot) \in Q, j = 1, 2, \dots, r$ are the weight functions of Theorem 3.1

Theorem 4.

Let $B(\cdot)$ be a Brownian bridge and assume that H_0 of (1) holds. Then as $n \rightarrow \infty$

$$M_n^w(\underline{t}) \xrightarrow{D} \Lambda_n^w(\underline{t}),$$

where $\underline{t} = (t_0, t_1, \dots, t_r, t_{r+1})$, such that $0 = t_0 < t_1 \leq \dots \leq t_r < t_{r+1} = 1$ and $M_n^w(\underline{t})$ and $\Lambda_n^w(\underline{t})$ are the processes given by (18) and (19) respectively.

The proof of this theorem can be deduced easily from that of Theorem 3.

Corollary 1.

By the continuous Mapping Theorem and for $r \geq 1$, we have

$$T_{n1}(r) = \sup_{\underline{t}} |M_n^w(\underline{t})| \xrightarrow{D} \sup_{\underline{t}} |A^w(\underline{t})| = T_1$$

and

$$T_{n2}(r) = \int_{\underline{t}} \dots \int [M_n^w(\underline{t})]^2 = \int_{\underline{t}} \dots \int [A^w(\underline{t})]^2$$

The asymptotic distribution of $T_{n1}(r)$ and $T_{n2}(r)$ are, up to our knowledge, are unknown. For this reason we present the special case; at most one change point test statistics. Then we study the applicability of the proposed tests through a Monte Carlo simulation study.

5. The at Most One Change (AMOC) Test Statistics

First, we present the two-weight function test statistics for the CDF on change-point change.

For $0 \leq y \leq 1$, consider the following weight functions;

$$W_1(y) = y(1 - y), \tag{20}$$

and

$$W_2(y) = y(1 - y) \log \left\{ \log \left[\frac{1}{y(1-y)} \right] \right\}. \tag{21}$$

Let $\widehat{\alpha}_n(\cdot)$ is the empirical counterpart of the process in Eq. (16), defined by replacing the CDF, $F(\cdot)$, by its sample one $F_n(\cdot)$. We define the test process $\{A_n(k); 1 \leq k \leq n\}$, by

$$\begin{aligned} \widehat{A}_n(k) &:= \widehat{\alpha}_n(k) - \frac{k}{n} \widehat{\alpha}_n(n), = \\ &- \frac{1}{\sqrt{n}} \left\{ \sum_{i=1}^k F_n(X_i) - \frac{k}{n} \sum_{i=1}^n F_n(X_i) \right\} \end{aligned} \tag{22}$$

where $F_n(t) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq t)$, $t \in R$, is the sample empirical distribution function. The above test process is the natural candidate in case of testing for a change in the CDF of a sequence of independent random variables.

Let $I_n^{(1)}(a, b) = (na, nb)$ and $I_n^{(2)}(a, b) = [1, na] \cup [nb, n]$ where $(a, b) = (0.071033\dots, 0.928966\dots)$, see[3]. Now, we propose the following AMOC test statistics;

$$T_1(n) := \max_{1 \leq k \leq n} |\widehat{A}_n(k)|, \tag{23}$$

$$T_2(n) := \max_{1 \leq k \leq n} \frac{|\widehat{A}_n(k)|}{\left[W_2\left(\frac{k}{n}\right) \right]^{1/2}} \tag{24}$$

$$T_3(n) := \max_{1 \leq k \leq n} \begin{cases} |\widehat{A}_n(k)|, k \in I_n^{(1)}(a, b) \\ \frac{|\widehat{A}_n(k)|}{\left[W_2\left(\frac{k}{n}\right) \right]^{1/2}}, k \in I_n^{(2)}(a, b), \end{cases} \tag{25}$$

$$T_4(n) := \begin{cases} \frac{|\widehat{A}_n(k)|}{\left[W_1\left(\frac{k}{n}\right) \right]^{1/2}}, k \in I_n^{(1)}(a, b) \\ \frac{|\widehat{A}_n(k)|}{\left[W_2\left(\frac{k}{n}\right) \right]^{1/2}}, k \in I_n^{(2)}(a, b), \end{cases} \tag{26}$$

$$T_5(n) := \frac{1}{n} \left\{ \sum_{I_n^{(1)}(a,b)} (|\widehat{A}_n(k)|)^2 + \sum_{I_n^{(2)}(a,b)} \left(\frac{|\widehat{A}_n(k)|}{\left[W_2\left(\frac{k}{n}\right) \right]^{1/2}} \right)^2 \right\} \tag{27}$$

and

$$T_6(n) := \frac{1}{n} \left\{ \sum_{I_n^{(1)}(a,b)} \left(\frac{|\widehat{A}_n(k)|}{\left[W_1\left(\frac{k}{n}\right) \right]^{1/2}} \right)^2 + \sum_{I_n^{(2)}(a,b)} \left(\frac{|\widehat{A}_n(k)|}{\left[W_2\left(\frac{k}{n}\right) \right]^{1/2}} \right)^2 \right\} \tag{28}$$

where $W_1(\cdot)$, $W_2(\cdot)$ and $A_n(\cdot)$ are given by Eq. (20), Eq. (21) and Eq. (22) respectively.

Note that the first four test statistics are CDF change point versions analogues to Pouliot (2001). The last two $T_5(n)$ and $T_6(n)$ are new proposed test statistics. The limiting distributions of the above test statistics are unknown in literature. Thus we conduct a Monte Carlo study to determine the performance of these test statistics.

6. Estimated Critical Values and Powers

The critical values of the proposed tests in (23)-(28) have been evaluated via simulation. Also, the power of the proposed tests have been estimated. These estimation tasks are conducted using three simulation studies.

6.1. Simulation 1

The aim of this simulation is two-folds. First, to estimate the critical values of each test at different sample sizes under different distributions. Second, to show that the critical values are stable. The upper 5% critical values for each test of the six tests given in (23)-(28) have been obtained via simulation study. A sample of each distribution has been simulated. The sample sizes are fixed at 15, 20, 25, 30, 40, 50, 100, 200, 500, and 1000. The underlying distributions are the normal distribution, the chi-square distribution, the exponential distribution and the uniform distribution. Each test value is evaluated for each sample. This process is replicated 10000 times. Then, each test values are sorted and the upper 95% percentile is obtained.

The simulation results are displayed in Table 1 and Table 2. The other distributions critical values have a similar behaviour. From these results we can notice that the second test has a higher critical values followed by the third test, for all the four distributions. The fifth test has the lowest critical values across all the four distributions. The critical values of the first, fourth and sixth tests are close.

Generally, the critical values of all tests starts at a higher (lower) level for smaller sample size; $n = 15$, but they converge reasonably as the sample size increase, see the table 1. This convergence appears from a sample size as large as 100.

Table 1. The estimated upper 5% critical values of the normal and chi-square distributions

The normal distribution						
<i>n</i>	<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	<i>T6</i>
15	0.3443	1.1783	0.4780	0.4780	0.1933	0.3217
20	0.3522	1.1863	0.4617	0.4617	0.1390	0.2935
25	0.3600	1.2215	0.5721	0.4507	0.1782	0.3201
30	0.3651	1.2331	0.6035	0.4425	0.1803	0.3223
40	0.3656	1.2359	0.6135	0.4089	0.1732	0.3141
50	0.3649	1.2360	0.6432	0.4057	0.1881	0.3364
100	0.3765	1.2787	0.6900	0.3860	0.1877	0.3336
200	0.3797	1.2810	0.7243	0.3667	0.1870	0.3328
500	0.3847	1.3049	0.7571	0.3517	0.1909	0.3343
1000	0.3900	1.3170	0.7738	0.3415	0.1880	0.3376
The chi-square distribution						
<i>n</i>	<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	<i>T6</i>
15	0.3443	1.1783	0.4780	0.4780	0.3182	0.1933
20	0.3578	1.2161	0.4617	0.4617	0.1402	0.2890
25	0.3600	1.2143	0.5721	0.4507	0.1753	0.3197
30	0.3621	1.2290	0.6035	0.4425	0.1801	0.3274
40	0.3676	1.2500	0.6014	0.4089	0.1710	0.3161
50	0.3719	1.2585	0.6432	0.4057	0.1973	0.3410
100	0.3735	1.2633	0.6900	0.3860	0.1878	0.3343
200	0.3811	1.2933	0.7243	0.3706	0.1853	0.3420
500	0.3863	1.3113	0.7514	0.3517	0.1889	0.3335
1000	0.3890	1.3065	0.7647	0.3429	0.1855	0.3361

Table 2. The estimated upper 5% critical values of the exponential and uniform distributions

The exponential distribution						
<i>n</i>	<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	<i>T6</i>
15	0.3443	1.1783	0.4780	0.4780	0.1920	0.3253
20	0.3578	1.2129	0.4617	0.4617	0.1381	0.2913
25	0.3600	1.2105	0.5721	0.4507	0.1771	0.3195
30	0.3621	1.2290	0.6035	0.4425	0.1810	0.3252
40	0.3656	1.2486	0.6014	0.4089	0.1692	0.3235
50	0.3705	1.2585	0.6494	0.4057	0.1924	0.3289
100	0.3740	1.2681	0.6901	0.3860	0.1879	0.3389
200	0.3794	1.2869	0.7271	0.3667	0.1860	0.3299
500	0.3853	1.3076	0.7657	0.3503	0.1902	0.3330
1000	0.3924	1.3276	0.7837	0.3429	0.1922	0.3439
The uniform distribution						
<i>n</i>	<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	<i>T6</i>
15	0.3443	1.2002	0.4780	0.4780	0.1928	0.3257
20	0.3578	1.2161	0.4617	0.4617	0.1398	0.2883
25	0.3600	1.2306	0.5721	0.4507	0.1793	0.3129
30	0.3621	1.2246	0.6035	0.4425	0.1814	0.3251
40	0.3656	1.2362	0.6135	0.4310	0.1733	0.3213
50	0.3691	1.2470	0.6432	0.4057	0.1915	0.3406
100	0.3770	1.2804	0.6979	0.3860	0.1887	0.3294
200	0.3774	1.2803	0.7269	0.3667	0.1877	0.3359
500	0.3839	1.2997	0.7515	0.3517	0.1872	0.3350
1000	0.3864	1.3092	0.7705	0.3429	0.1874	0.3514

6.2. Simulation 2

The aim of this simulation study is to estimate the power of the tests (23) - (28) assuming there is one change point in the mean. A sample of fixed size is generated from each distribution. The sample sizes are fixed at 20, 50, and 100 units, to cover small, moderate and large sample size. The samples are generated from the normal distribution, the chi-square distribution, the exponential distribution and the uniform distribution. The change point positions are fixed at the first tail (15%*n*) of the sample and in the middle (50%*n*) of the sample. Different change shifts have been used, namely $\Delta=0.5$, $\Delta=1.0$ and $\Delta=1.5$. The replications number is 10000. The percentage of times the test statistic exceeds the estimated critical values is reported for each change, test statistic, and sample size. The results are displayed in Table 3 and Table 4.

The simulation results show that the estimated power of each test under each distribution increase as the change position moves to the middle of the sample. The estimated powers of all tests increase as the change shift increases and the sample size increases.

From the results we can see that the fourth test has the highest power followed by the sixth test for all distributions in the different settings. The third test has the lowest power in the different setting. The estimated powers of the first and the second tests are comparable in the different settings. The powers under the uniform distribution has the highest values, whereas the lowest powers are under the chi-square distribution. This is not surprising because any change in the mean of the uniform random variable affect the distribution boundaries too. However, the chi-square distribution will change its shape very slowly with such minor location changes.

Table 3. The estimated powers of the normal and chi-square distributions (change in the mean)

The normal distribution								
<i>n</i>	Δ	<i>position</i>	<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	<i>T6</i>
20	0.5	0.15	6.3	6.7	0.4	66.4	6.6	7.0
		0.50	14.6	15.9	1.7	71.4	8.4	12.7
	1.0	0.15	9.2	9.2	0.6	78.8	10.4	13.1
		0.50	44.6	47.7	10.5	90.0	19.4	38.6
	1.5	0.15	13.9	13.6	1.0	90.5	15.2	21.1
		0.50	79.6	82.1	35.3	98.6	38.1	69.7
50	0.5	0.15	10.8	10.7	6.8	92.9	7.7	11.9
		0.50	33.7	35.1	7.1	95.9	13.2	26.0
	1.0	0.15	30.6	27.9	15.4	98.8	20.2	35.6
		0.50	87.0	88.5	18.1	99.8	41.9	76.7
	1.5	0.15	63.7	58.6	31.5	99.9	40.1	67.3
		0.50	99.8	99.8	60.5	100.0	78.9	98.5
100	0.5	0.15	16.0	14.8	14.2	98.0	16.3	21.6
		0.50	59.1	60.7	10.9	99.3	24.7	49.3
	1.0	0.15	57.8	52.9	45.7	99.9	53.3	69.1
		0.50	99.3	99.4	56.2	100.0	81.3	97.8
	1.5	0.15	94.7	92.6	81.9	100.0	87.0	95.8
		0.50	100.0	100.0	98.8	100.0	99.6	100.0
The chi-square distribution								
<i>n</i>	Δ	<i>position</i>	<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	<i>T6</i>
20	0.5	0.15	4.3	4.4	0.2	61.6	4.9	5.4
		0.50	4.7	5.1	0.3	61.8	5.1	5.2
	1.0	0.15	4.9	5.1	0.3	63.0	5.7	6.3
		0.50	6.8	7.2	0.7	63.2	5.6	7.3
	1.5	0.15	4.8	5.0	0.3	64.1	6.3	6.9
		0.50	9.4	9.9	1.1	67.0	6.5	9.5
50	0.5	0.15	5.5	5.4	5.5	89.0	4.5	5.1
		0.50	6.3	6.4	5.0	89.0	4.6	5.7
	1.0	0.15	5.8	5.5	5.3	89.7	5.6	6.7
		0.50	10.7	11.2	4.8	90.8	6.1	9.4
	1.5	0.15	7.8	7.4	6.8	91.9	7.0	9.4
		0.50	18.7	19.5	5.8	92.7	7.9	14.9
100	0.5	0.15	5.8	5.7	5.8	94.4	6.6	6.0
		0.50	8.1	8.1	5.4	94.3	6.6	7.5
	1.0	0.15	7.5	7.4	7.9	95.6	8.9	8.9
		0.50	18.5	19.3	6.3	96.4	10.0	15.7
	1.5	0.15	10.9	10.4	11.1	96.7	12.6	14.4
		0.50	35.4	37.0	8.3	98.0	15.5	28.5

Table 4. The estimated powers of the exponential and the uniform distributions (change in the mean)

The exponential distribution								
<i>n</i>	Δ	<i>position</i>	<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	<i>T6</i>
20	0.5	0.15	6.9	7.0	0.5	72.6	10.1	12.4
		0.50	25.5	27.2	4.5	80.9	12.5	24.4
	1.0	0.15	11.1	10.7	0.8	82.8	16.2	22.2
		0.50	62.2	65.5	24.0	95.6	29.0	56.4
	1.5	0.15	13.7	13.4	1.2	90.6	20.5	31.1
		0.50	85.8	88.0	49.3	99.4	46.2	80.3
50	0.5	0.15	18.3	16.2	16.7	96.8	19.2	25.9
		0.50	60.0	61.2	9.8	98.4	24.8	50.9
	1.0	0.15	46.9	42.3	36.3	99.5	41.9	57.1
		0.50	96.7	97.2	35.3	100.0	65.9	93.3
	1.5	0.15	71.5	65.8	54.0	99.9	61.4	77.6
		0.50	100.0	100.0	76.5	100.0	90.6	99.6
100	0.5	0.15	34.0	31.6	33.2	99.5	34.5	42.1
		0.50	89.2	9.1	23.5	100.0	49.4	80.2
	1.0	0.15	19.7	77.1	11.0	100.0	71.6	83.4
		0.50	100.0	100.0	88.6	100.0	96.3	99.9
	1.5	0.15	96.2	95.2	86.6	100.0	90.1	96.5
		0.50	100.0	100.0	99.8	100.0	100.0	100.0
The uniform distribution								
<i>n</i>	Δ	<i>position</i>	<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	<i>T6</i>
20	0.5	0.15	13.2	12.2	1.0	93.4	16.7	25.0
		0.50	85.3	87.2	43.2	99.7	42.2	77.8
	1.0	0.15	21.4	19.4	1.7	100.0	28.1	48.2
		0.50	100.0	100.0	100.0	100.0	81.1	100.0
	1.5	0.15	21.4	19.5	1.7	100.0	28.4	48.8
		0.50	100.0	100.0	100.0	100.0	81.7	100.0
50	0.5	0.15	69.7	63.8	32.8	100.0	45.4	73.3
		0.50	99.9	100.0	73.2	100.0	85.9	99.4
	1.0	0.15	100.0	100.0	98.5	100.0	96.6	100.0
		0.50	100.0	100.0	100.0	100.0	100.0	100.0
	1.5	0.15	100.0	100.0	98.2	100.0	96.3	100.0
		0.50	100.0	100.0	100.0	100.0	100.0	100.0
100	0.5	0.15	98.2	97.0	86.0	100.0	92.9	98.3
		0.50	100.0	100.0	99.7	100.0	99.9	100.0
	1.0	0.15	100.0	100.0	100.0	100.0	100.0	100.0
		0.50	100.0	100.0	100.0	100.0	100.0	100.0
	1.5	0.15	100.0	100.0	100.0	100.0	100.0	100.0
		0.50	100.0	100.0	100.0	100.0	100.0	100.0

6.3. Simulation 3

The aim of this simulation is to estimate the power of the proposed test assuming that there is change point in the distribution rather than only the distribution mean. A sub-sample is simulated from a given distribution augmented by another sub-sample of another distribution. This means that there is a change in distribution. The change positions are the first 15% point of the sample and the middle of the sample. The sample sizes are fixed at $n = 20$, $n = 50$ and $n = 100$. Four different distributions have been used; namely the normal distribution, the chi-square distribution, the exponential distribution and the uniform distribution. The results in Table 6 show the estimated power of the tests for the normal distribution against the other three distributions (the chi-square distribution, the exponential distribution and the uniform distribution). Also, the table shows the estimated power of the tests assuming the chi-square distribution

against the exponential and the uniform distribution in addition to the exponential distribution against the uniform distribution.

From the results we can notice that the fourth test has the highest power followed by the sixth test. Generally, the third test has the lowest estimated powers. The highest powers are obtained for the chi-square distribution against the rest of the distributions.

7. Discussion and Conclusions

In this paper we presented new non-parametric weighted type test statistics for a change in the cumulative distribution function of a set of data. These proposed test statistics are based on the empirical processes. The asymptotic distributions of these test statistics are unknown and intractable to be studied theoretically.

We conducted a simulation study to estimate the critical values and powers of the proposed tests in the at most one change point. Our weighted proposed tests have good performance in all settings; different distributions, different

sample sizes and different change positions. The difficulty of tracing the limiting distributions of the proposed weighted test statistics encourage the search for a simple new weighted test statistics.

Table 5. The estimated powers of a distribution change

The normal distribution against the chi-square distribution							
<i>n</i>	<i>position</i>	<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	<i>T6</i>
20	0.15	25.3	22.3	1.9	100.0	28.2	47.0
	0.50	100.0	100.0	100.0	100.0	80.8	100.0
50	0.15	100.0	100.0	98.0	100.0	95.7	100.0
	0.50	100.0	100.0	100.0	100.0	100.0	100.0
100	0.15	100.0	100.0	100.0	100.0	100.0	100.0
	0.50	100.0	100.0	100.0	100.0	100.0	100.0
The normal distribution against the exponential distribution							
<i>n</i>	<i>position</i>	<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	<i>T6</i>
20	0.15	9.4	9.4	0.5	78.0	12.9	16.2
	0.50	46.6	49.7	11.3	91.0	20.6	38.8
50	0.15	33.0	30.3	24.1	98.4	26.6	39.0
	0.50	88.0	89.1	19.8	99.7	41.8	76.4
100	0.15	57.6	53.4	50.4	99.9	53.0	66.6
	0.50	99.4	99.5	57.3	100.0	82.2	97.6
The normal distribution against the uniform distribution							
<i>n</i>	<i>position</i>	<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	<i>T6</i>
20	0.15	7.7	7.4	0.4	71.9	13.1	14.2
	0.50	27.5	29.1	5.6	80.0	11.5	22.5
50	0.15	19.8	18.6	21.8	97.7	21.9	27.1
	0.50	55.4	57.0	9.9	98.2	20.2	44.1
100	0.15	31.3	29.0	38.4	99.4	34.7	41.0
	0.50	82.9	83.8	21.7	99.8	44.7	73.5
The chi-square distribution against the exponential distribution							
<i>n</i>	<i>position</i>	<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	<i>T6</i>
20	0.15	20.8	18.6	1.3	100.0	26.7	48.3
	0.50	100.0	100.0	99.9	100.0	79.1	100.0
50	0.15	100.0	100.0	94.7	100.0	94.1	100.0
	0.50	100.0	100.0	100.0	100.0	100.0	100.0
100	0.15	100.0	100.0	100.0	100.0	100.0	100.0
	0.50	100.0	100.0	100.0	100.0	100.0	100.0
The chi-square distribution against the uniform distribution							
<i>n</i>	<i>position</i>	<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	<i>T6</i>
20	0.15	21.1	18.8	1.6	100.0	26.9	48.0
	0.50	100.0	100.0	100.0	100.0	78.9	100.0
50	0.15	100.0	100.0	98.2	100.0	95.0	100.0
	0.50	100.0	100.0	100.0	100.0	100.0	100.0
100	0.15	100.0	100.0	100.0	100.0	100.0	100.0
	0.50	100.0	100.0	100.0	100.0	100.0	100.0
The exponential distribution against the uniform distribution							
<i>n</i>	<i>position</i>	<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	<i>T6</i>
20	0.15	5.7	6.0	0.4	67.6	9.8	9.6
	0.50	13.3	14.6	1.8	72.2	7.9	13.6
50	0.15	11.0	9.8	12.8	94.9	14.2	16.4
	0.50	31.0	31.9	7.0	95.0	12.4	25.0
100	0.15	16.7	15.3	22.1	98.3	20.8	23.4
	0.50	55.2	57.1	11.3	99.0	23.0	44.4

REFERENCES

- [1] Csörgő, M. and Horváth, L., 1979, *Limit Theorems in Change-Point Analysis*, John Wiley & Sons, New York.
- [2] Orasch, M. R., 1999, Multiple change-points with an application to financial modelling, Ph.D. thesis, Carleton University, Ottawa, Canada.
- [3] Pouliot, W. J., 2001, Non-parametric, Non-sequential change-points analysis, Ph.D. thesis, Carleton University, Ottawa, Canada.
- [4] Hawkins, D. M., 2001, Fitting multiple change-point models to data, *Computational Statistics & Data Analysis*, 37, 323-341.
- [5] Syzsykowicz, B., 1992, Weak convergence of stochastic processes in weighted metrics and their applications to contiguous change-points analysis, Ph.D. thesis, Carleton University, Ottawa, Canada.
- [6] Aly, E.-E., Abd-Rabou, A. S. and Al-Kandari, N. M., 2003, Tests for multiple change points under ordered alternatives, *Metrika*, 57, 209-221.