The Korean Communications in Statistics **Vol. 14** No. 2, 2007, pp. 443–457

A Nonparametric Bootstrap Test and Estimation for Change*

Jaehee Kim[1)]

Abstract

This paper deals with the problem of testing the existence of change in mean and estimating the change-point using nonparametric bootstrap technique. A test statistic using Gombay and Horvath (1990)'s functional form is applied to derive a test statistic and nonparametric change-point estimator with bootstrapping idea. Achieved significance level of the test is calculated for the proposed test to show the evidence against the null hypothesis. MSE and percentiles of the bootstrap change-point estimators are given to show the distribution of the proposed estimator in simulation.

Keywords: Achieved significance level; change-point; nonparametric bootstrap test; Ornstein-Uhlenbeck process.

1. Introduction

The bootstrap method, formulated by Efron and Tibshirani (1979) has been a widely applicable method in testing problems. Bootstrap methods are simulation methods for assessing frequency properties of statistical analyses, the simulation model fitted to the actual data. Often the simulation involves Monte Carlo evaluation rather than the theoretical work.

Both bootstrap and randomization methods are the same in that rejection of a null hypothesis occurs when the test statistic is large. However the approaches differ in that critical values are determined by distinct resampling methods to get the critical values from the null distribution.

The change-point problem can be considered to be one of the central problems of statistical inference, linking together statistical control theory, the theory

* This research is supported by Korea Research Foundation (R04-2004-000-10138-0).

1) Professor, Department of Statistics, Duksung Women's University, Seoul 132-714, Korea.
E-mail : jaehee@duksung.ac.kr

of estimation and testing hypotheses, classical and Bayesian approaches, and fixed sample and sequential procedures. The first publications of Page already appeared in the 1950s in connection with industrial quality control. The change-point problems was then intensively investigated in the works of Lorden (1971), Hinkley (1970), Zacks (1983) and others. Nowadays, this field of statistical research looks like a large family of mathematical problems reflecting different approaches to the main questions such as is the sample of observations homogeneous in a statistical sense? After testing for change, if the type of change is abrupt, the change-point estimation problem occurs. Therefore testing and estimating are core procedures in change analysis.

Chernoff and Zacks (1964) considered a Bayes test for mean change for the normal observations. The result of Chernoff and Zacks was later generalized by Kander and Zacks (1966) to the case of a one-parameter exponential family, and by Gardner (1969) to the case of the unknown amount of change.

The next important step was made by Hinkley (1970). He investigated in detail the maximum likelihood estimates of a change-point. Also for an arbitrary density of observed independent random variables, the asymptotic distribution of the maximum likelihood estimate of the change-point was obtained.

As a nonparametric approach, the test statistic proposed by Pettitt (1979) was based on Mann-Whitney statistic. Darkhovsh (1976) suggested the change-point estimator based on Mann-Whitney statistics. Lorden (1971) considered the sequential detection procedure from a non-Bayesian approach and proved that the well-known CUSUM procedures of Page are asymptotically minimax.

Bhattacharyya and Johnson (1968) approached the testing problem based on sign and score functions in a non-parametric fashion. Carlstein (1988) considered a test using the idea of pre- and post empirical cumulative distributions and their difference. And there are nonparametric bootstrap studies for change analysis which are introduced in the following section.

This paper is organized as follows. In section 2 a mean change-point model is defined, bootstrap researches for change problems are surveyed and the proposed bootstrap method is derived with its statistical properties. Section 3 presents some numerical results in simulation with test statistics and change-point estimators. Finally section 4 concludes the paper with a discussion of change-point problem and bootstrap technique.

2. Bootstrap Applications to Change Problems

2.1. Model of mean change

The statistical problem considered has the following form. Given a sample $X_1, X_2, \ldots, X_n$ of S -valued random variables, we wish to test the null hypothesis H_0 that the unknown probability distribution on S generating data belongs to a certain class Ω_0 against the alternative class Ω_1 with the change-point τ.

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent random variables

$$\begin{aligned} &X_1, X_2, \ldots, X_\tau \sim iid\, F(x), \\ &X_{\tau+1}, X_{\tau+2}, \ldots, X_n \sim iid\, G(x), \ \tau \in \{1, \ldots, n-1\}, \\ &F(x) \neq G(x), \end{aligned}$$

where τ is the change-point at which time the underlying distribution has changed form F to G.

Typically, the estimator of the change-point is the maximizer of the (log) likelihood function and its distribution if F and G are known. When the forms of F and G are unknown, for nonparametric models, Darkhovskh (1976) and Carlstein (1988) obtained the consistency results of certain nonparametric estimators of the change-point.

2.2. Survey of bootstrap change-point estimation

Hinkley and Schechtman (1987) showed the conditional bootstrap analysis of mean-shift in Nile river flow data and compared with the parametric and semiparametric likelihood analyses.

Boukai (1993) applied a bootstrap resampling scheme to get a nonparametric estimator of the change-point and his estimator is based on the Kolmogorov-Smirnov norm as proposed by Carlstein (1988). Boukai (1993)'s bootstrapping Kolmogorov-Smirnov method is as follows.

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent random variables with the *cdf* $G_{n,i}$. In the context of the change-point problem, the model is assumed as for some unknown $\tau \in (0, 1)$

$$G_{n,i} = \begin{cases} G^0, \ i \leq [n\tau], \\ G^1, \ i > [n\tau], \end{cases}$$

where G^0 and G^1 are two continuous *cdfs* ($G^0 \neq G^1$) and $[y]$ denotes the greatest integer not exceeding y. Here $m = [n\tau]$ is the change-point of the sequence

$X_1, X_2, \ldots, X_n$ at which time the underlying distribution has changed from G^0 to G^1. Let n be fixed. For each $k, k = 1, 2, \ldots, n-1$, let F_k and $\tilde{F}_k$ denote the empirical cdfs of the first k and the last $n-k$ observations respectively. So that

$$F_k(x) = F_{k,n}(x) = \frac{1}{k}\sum_{i=1}^{k} I(X_i \leq x),$$

$$\tilde{F}_k(x) = \tilde{F_{k,n}}(x) = \frac{1}{n-k}\sum_{i=k+1}^{n} I(X_i \leq x)$$

with $I(A)$ being the indicator function of the set A . F_k and $\tilde{F}_k$ are unbiased estimators of $\bar{G}_k(x) = 1/k\sum_{i=1}^{k} G_i$, $\tilde{G}_k(x) = 1/(n-k)\sum_{i=k+1}^{n} G_i$ respectively. The K-S estimator of the change-point m is based on the Kolmogorov-Smirnov norm

$$D_n = c_n(k)||F_k - \tilde{F}_k||$$

with $c_n(k) = (k(n-k))^{1/2}/n$ and where $||f||$ denotes the usual supremum norm of $|f|$. The estimator $\hat{m}$ of m is taken to be the maximizer of $D_n(k)$ over the set $\{1, 2, \ldots, n-1\}$ so that

$$D_n(\hat{m}) = \max_{1\leq k\leq n-1} D_n(k).$$

Boukai (1993) nonparametric bootstrapping estimation procedure for the change-point is based on the following resampling scheme. Given the n observations $X = \{X_1, X_2, \ldots, X_n\}$ and the observed value of $\hat{m}$, the resampled observations $X^* = \{X_1^*, X_2^*, \ldots, X_n^*\}$ from X are obtained in such a way that X_i^* has a *cdf* $V_{\hat{m},i}$
where

$$V_{\hat{m},i} = \begin{cases} F_{\hat{m}} & i \leq \hat{m}, \\ \tilde{F}_{\hat{m}} & i > \hat{m}. \end{cases}$$

That is, given χ and $\hat{m}$,

$$X_1^*, X_2^*, \ldots, X_{\hat{m}}^* \sim F_{\hat{m}} \qquad \text{and} \qquad X_{\hat{m}+1} \ldots, X_n^* \sim \tilde{F}_{\hat{m}}$$

are conditionally independent and identically distributed and the resample X^* has a change-point at $\hat{m}$. Based on the resample X^* , the bootstrap estimate $\hat{m}^*$ of $\hat{m}$ is the maximizer of

$$D_n^*(k) = c_n(k)||F_k^* - \tilde{F}_k^*|| \qquad \text{over} \qquad \{1, 2, \ldots, n-1\}$$

where F_k^*, $\tilde{F}_k^*$ are the bootstrapped versions of F_k, $\tilde{F}_k$ as

$$F_k^*(x) = \frac{1}{k}\sum_{i=1}^{k} I(X_i^* \leq x),$$

$$\tilde{F}_k^*(x) = \frac{1}{n-k}\sum_{i=k+1}^{n} I(X_i^* \leq x).$$

According to the procedure the bootstrap estimator $\hat{m}^*$ satisfies

$$D_n^*(\hat{m}^*) = \max_{1\leq k\leq n-1} D_n^*(k).$$

The empirical bootstrap cdfs F^* and $\tilde{F}^*$ are unbiased estimators of

$$\bar{V}_k(x) = \frac{1}{k}\sum_{i=1}^{k} V_{\hat{m},i}, \ \tilde{V}_k(x) = \frac{1}{n-k}\sum_{i=k+1}^{n} V_{\hat{m},i}$$

respectively.

Antoch and Huskova (1995) considered a robust M-estimators of the change-point in general location models and studied the consistency and the limiting distribution of the estimator. The limiting distribution of the normalized change-point estimator was shown to that of the location of the supremum of a two-sided Wiener process with drift. Let $\{X_1, X_2, \ldots, X_n\}$ be a sequence of independent random variables such that for a unique value of $m \in \{1, 2, \ldots, n-1\}$, the following model holds

$$X_t = \begin{cases} \theta_0 + \epsilon_t, & t = 1, 2, \ldots, \tau, \\ \theta_1 + \epsilon_t = \theta_0 + \delta + \epsilon_t, & t = \tau + 1, \ldots, n. \end{cases}$$

The random variables $\{\epsilon_1, \ldots, \epsilon_n\}$ are *iid* with zero mean and unknown variance σ^2 . The location parameter θ_0 , the change amount δ as well as the change-point m are unknown. They suggested the change-point estimator $\hat{m}(\gamma)$, $0 \leq \gamma \leq 1/2$, defined as

$$\hat{m}(\gamma) = \text{argmax}\left\{|U_k(\gamma)|;\ k = 1, 2, \ldots, n-1\right\},$$

where

$$U_k(\gamma) = \left(\frac{n}{k(n-k)}\right)^{\gamma} \sum_{i=1}^{k}(X_i - \bar{X}_n)$$

and for $n = 1, 2, \ldots,\ \bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i$. Note that if the error terms ϵ_i have a normal $N(0, \sigma^2)$, then $\hat{m}(1/2)$ is the maximum likelihood estimator of m.

Let $\hat{\theta}_0$ and $\hat{\delta}$ be the estimators for θ_0 and δ respectively. Define the estimated residuals

$$e_t^* = \begin{cases} X_t - \hat{\theta}_0, & t = 1, \ldots, \hat{m}(\gamma), \\ X_t - \hat{\theta}_0 - \hat{\delta}, & t = \hat{m}(\gamma) + 1, \ldots, n \end{cases}$$

and centered residuals

$$\tilde{e_t} = e_t^* - \frac{1}{n}\sum_{i=1}^{n} e_i^*, \qquad t = 1, \ldots, n.$$

Take $e_1^*, \ldots, e_n^*$ iid from the empirical distribution function of $\tilde{e_1}, \ldots, \tilde{e_n}$ and consider the bootstrap observations

$$X_t^* = \begin{cases} \hat{\theta}_0 + e_t^*, & t = 1, 2, \ldots, \hat{m}(\gamma), \\ \hat{\theta}_0 + \hat{\delta} + e_t^*, & t = \hat{m}(\gamma) + 1, \ldots, n. \end{cases}$$

For simplicity, consider the case with $\gamma = 1/2$. The bootstrap estimator corresponding to $\hat{m}(1/2)$ is

$$\hat{m^*} = \text{argmax}\left\{|U_k^*(\gamma)|;\ k = 1, 2, \ldots, n-1\right\},$$

where

$$U_k^* = \sqrt{\frac{n}{k(n-k)}}\sum_{i=1}^{k}(X_i^* - \bar{X}_n^*) \quad \text{and} \quad \bar{X}_n^* = n^{-1}\sum_{i=1}^{n} X_i^*.$$

2.3. Proposed test for change and change-point estimation with Bootstrap

We consider the one change-point model as

$$X_t = \begin{cases} \theta_0 + \epsilon_t, & t = 1, 2, \ldots, \tau, \\ \theta_1 + \epsilon_t, & t = \tau + 1, \ldots, n, \end{cases}$$

where the errors ϵ_t are *iid* from the continuous distribution θ. θ is a continuous mean function with the unknown change-point τ and $\theta_1 = \theta_0 + \Delta$.

Gombay and Horvath (1990) consider the test statistic based on

$$Z_k = 2\left\{kg(\bar{X}_k) + (n-k)g(\bar{X}_{n-k}) - ng(\bar{X}_n)\right\},$$

where g is a given continuous function,

$$\bar{X}_k = \frac{1}{k}\sum_{i=1}^{k} X_i, \quad \bar{X}_{n-k} = \frac{1}{n-k}\sum_{i=k+1}^{n} X_i \quad \text{and} \quad \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Their test rejects H_0 in favor of H_1 for large values of

$$Z(i,j) = \max_{i<m<j} \frac{Z_m}{g^{(2)}(\mu)},$$

where $g^{(2)}$ is the second derivative of g and for suitably chosen i and j. They proved that under H_0 , for all $0 < \lambda_1 \le 1 - \lambda_2 < 1$ as $n \to \infty$, $Z(m_1, m_2)/\sigma^2 \to \sup_{0\le s\le\Lambda} |V(s)|^2$ in distribution where $m_1 = n\lambda_1$, $m_2 = n(1-\lambda_2)$, $\Lambda = (1/2)\log\{(1-\lambda_1)(1-\lambda_2)/(\lambda_1\lambda_2)\}$ and $\{V(s), -\infty < s < \infty\}$ be an Ornstein-Uhlenbeck process, i.e. a Gaussian process with mean zero and covariance $\exp(-|t-s|)$. Or the limiting distribution can be expressed as

$$\max_{m_1\le k\le m_2} Z_k \to \sigma^2 \sup_{\lambda_1\le k\le 1-\lambda_2} \{W(t) - tW(1)\}^2 / \{t(1-t)\}$$

in distribution.

We propose a bootstrap test of Gombay and Horvath's functional form. Since their test is derived from the maximum likelihood ratio test, it is suitable for any distribution.

Nonparametric bootstrap estimation procedure is in the followings.

(i) Let n be fixed. For each k, $k = 1, 2, \ldots, n-1$.

For the sample $\{X_1, X_2, \ldots, X_n\}$, calculate the test statistic with the bootstrapped sample and get the estimators $\hat{\tau}, \hat{\theta}_0, \hat{\Delta}$.

(ii) Let $\hat{\theta}_0$ and $\hat{\Delta}$ be the estimators for θ_0 and δ respectively. Define the estimated residuals

$$e_t^* = \begin{cases} X_t - \hat{\theta}_0, & t = 1, \ldots, \hat{\tau}, \\ X_t - \hat{\theta}_0 - \hat{\Delta}, & t = \hat{\tau}+1, \ldots, n \end{cases}$$

and centered residuals

$$\tilde{e}_t = e_t^* - \frac{1}{n}\sum_{i=1}^{n} e_i^*, \quad t = 1, \ldots, n.$$

Take $e_1^*, \ldots, e_n^*$ iid from the empirical distribution function of $\tilde{e}_1, \ldots, \tilde{e}_n$ and consider the bootstrap observations

$$X_t^* = \begin{cases} \hat{\theta}_0 + e_t^*, & t = 1, 2, \ldots, \hat{\tau}, \\ \hat{\theta}_0 + \hat{\Delta} + e_t^*, & t = \hat{\tau} + 1, \ldots, n. \end{cases}$$

(iii) The bootstrap estimator corresponding to Z_k is

$$Z_k^* = 2\left\{kg(\bar{X}_k^*) + (n-k)g(\bar{X}_{n-k}^*) - ng(\bar{X}_n^*)\right\},$$

where g is a given continuous function,

$$\bar{X}_k^* = \frac{1}{k}\sum_{i=1}^{k} X_i^*, \quad \bar{X}_{n-k}^* = \frac{1}{n-k}\sum_{i=k+1}^{n} X_i^* \quad \text{and} \quad \bar{X}_n^* = \frac{1}{n}\sum_{i=1}^{n} X_i^*.$$

Calculate the following test statistic as

$$Z(i,j)^* = \max_{i<m<j} \frac{Z_k^*}{g^{(2)}(\theta)},$$

where g^2 is the second derivative of g and for suitably chosen i and j.

(iv) The bootstrap change-point estimate is obtained as

$$\hat{\tau}^* = \text{argmax}\left\{|Z(i,j)^*|,\ 1 < i < j < n\right\}.$$

With B times repetitions, the bootstrap critical values are obtained from the bootstrap distribution. Therefore the bootstrap test can be done with the bootstrap critical values.

The bootstrap change-point estimate can be obtained as the mean of the bootstrap estimates

$$\hat{\tau}_b^* = \bar{\hat{\tau}}^* = \frac{1}{B}\sum_{i=1}^{B} \hat{\tau}^*$$

and the bootstrap variance is estimated as

$$\hat{\text{Var}}^* = \frac{1}{B-1}\sum_{i=1}^{B} (\hat{\tau}^* - \hat{\tau}_b^*)^2.$$

Having observed $\hat{Z}$, the achieved significance level (ASL) of the test is defined to be the probability of observing at least that large a value when the null hypothesis is true,

$$\text{ASL} = P(Z \geq \hat{Z}).$$

The smaller the value of ASL, the stronger is the evidence against H_0 . The quantity $\hat{Z}$ is fixed at its observed value, the random variable Z has the null hypothesis distribution, the distribution of $\hat{Z}$ if H_0 is true.

The following theorems show the asymptotic properties of the proposed estimators.

Lemma 2.1 *With the bootstrapped sample* $\{X_1^*, X_2^*, \ldots, X_n^*\}$

$$\bar{X}_k^* \to \bar{X}_k, \quad \bar{X}_{n-k}^* \to \bar{X}_{n-k}, \quad \bar{X}_n^* \to \bar{X}_n \quad \text{in probability.}$$

Proof: For $k \leq \tau$, we have

$$\begin{aligned} E\left(\bar{X}_k\right) &= \frac{1}{k}\sum_{i=1}^{k} EX_i = \theta_0 \;, \\ E\left(\bar{X}_k^*\right) &= \frac{1}{k}\sum_{i=1}^{k} EX_i^* = E(X_1^*) = \theta_0. \end{aligned}$$

Also

$$\begin{aligned} E\left(\bar{X}_{n-k}\right) &= \frac{1}{n-k}\left\{\sum_{i=k+1}^{n} EX_i\right\} = \frac{1}{n-k}\left\{\sum_{i=k+1}^{\tau} EX_i + \sum_{i=\tau+1}^{n} EX_i\right\} \\ &= \frac{1}{n-k}\left\{(n-k)\theta_0 + (n-\tau)\Delta\right\} \\ &= \theta_0 + \frac{n-\tau}{n-k}\Delta. \end{aligned}$$

Let $a = \sum_{i=k+1}^{n} I(X_i^* \in \{X_1, \ldots, X_\tau\})$ be the number of X_i^*'s in $\{X_1, \ldots, X_\tau\}$, the sample before the change-point. Then

$$\begin{aligned} E\left(\bar{X}_{n-k}^*\right) &= \frac{1}{n-k}\left\{\sum_{i=k+1}^{n} EX_i^*\right\} = \frac{1}{n-k}\left\{E(a\cdot\theta_0) + E[(n-k-a)(\theta_0+\Delta)]\right\} \\ &= \frac{1}{n-k}\left\{(n-k)\theta_0 + (n-\tau)\Delta\right\} \\ &= \theta_0 + \frac{n-\tau}{n-k}\Delta. \end{aligned}$$

Since

$$E(a) = E\left(\sum_{i=k+1}^{n} I(X_i^* \in \{X_1, \ldots, X_\tau\})\right) = \frac{n-\tau}{n-k} \times (n-k) = n-\tau$$

$$\begin{aligned} E\left(\bar{X}_n\right) &= \frac{1}{n}\left\{\sum_{i=1}^{n} EX_i\right\} = \frac{1}{n}\left\{\sum_{i=1}^{\tau} EX_i + \sum_{i=\tau+1}^{n} EX_i\right\} \\ &= \frac{1}{n}\left\{\tau\theta_0 + (n-\tau)(\theta_0+\Delta)\right\} \\ &= \theta_0 + \frac{n-\tau}{n}\Delta, \\ E\left(\bar{X}_n^*\right) &= \frac{1}{n}\left\{\sum_{i=1}^{n} EX_i^*\right\} = \frac{1}{n}E[(k+a)\theta_0 + (n-k-a)(\theta_0+\Delta)] \\ &= \theta_0 + \frac{n-\tau}{n}\Delta \end{aligned}$$

Therefore $\bar{X}_k^* \to \bar{X}_k$, $\bar{X}_{n-k}^* \to \bar{X}_{n-k}$, $\bar{X}_n^* \to \bar{X}_n$ in probability. □

Theorem 2.1 *With the bootstrapped sample* $\{X_1^*, X_2^*, \ldots, X_n^*\}$, $Z_k^* \to Z_k$ *in probability.*

Proof: From Lemma 2.1 and taking a continuous function g,

$$g(\bar{X}_k^*) \to g(\bar{X}_k), \quad g(\bar{X}_{n-k}^*) \to g(\bar{X}_{n-k}), \quad g(X_n^*) \to g(\bar{X}_n) \text{ in probability.}$$

□

The following theorem show that the limiting distribution of Z_k^* is the limiting distribution of Z_k .

Theorem 2.2 *The limiting distribution can be expressed as*

$$\max_{m_1 \le k \le m_2} Z_k^* \to \sigma^2 \sup_{\lambda_1 \le k \le 1-\lambda_2} \{W(t) - tW(1)\}^2 / t(1-t).$$

in distribution under H_0 *where* $0 < \lambda_1 \le 1-\lambda_2 < 1$, $m_1 = n\lambda_1$ *and* $m_2 = n(1-\lambda_2)$.

Proof: From the previous result, $Z_k^* \to Z_k$ in probability therefore $Z_k^* \to Z_k$ in distribution. Since $\max_{m_1 \le k \le m_2} Z_k \to \max_{m_1 \le k \le m_2} Z_k^*$, the result follows. □

3. Simulation

A random sample $X_1, X_2, \ldots, X_n$ are generated from the normal distribution with the mean 0 and the variance $\sigma^2 = 1$. Suppose that for a unique value of $\tau \in \{1, 2, \ldots, n-1\}$, the following mean level change model holds

$$X_t = \begin{cases} \theta_0 + \epsilon_t, & t = 1, 2, \ldots, \tau, \\ \theta_1 + \epsilon_t = \theta_0 + \Delta + \epsilon_t, & t = \tau + 1, \ldots, n, \end{cases}$$

where $\theta_0 = 0$ without loss of generality. The amount of change $\Delta = 0.5, 1, 1.5$, the sample size $n = 50$ and the location of change at $\tau/n = 0.3, 0.5, 0.8$ are considered. The bootstrap repetition $B = 1,000$ were used in this simulation. The proposed bootstrap test of Gombay and Horvath's functional form with $g(t) = t^2$ is considered in the simulation. Figure 3.1 shows the histogram and the smoothed density of the test statistic values under H_0, which gives a skewed distribution. Figure 3.2 shows the histogram and the smoothed density of bootstrapped values of the change-point when there is no change. Figure 3.3 shows the histogram of the test statistic values under H_1 with $\Delta = 1, \tau = 25$. The peak in Figure 3.3 occurs around the true change-point. Therefore the bootstrap distribution gives information about the location of the change-point. Figure 3.4 shows the histogram of bootstrapped values of the change-point estimator under H_1 with $\Delta = 1, \tau = 25$. While Figure 3.2 seems a uniform distribution except the boundaries, Figure 3.4 shows the location of the change-point. Table 3.1 shows that the sample mean (SM), the standard deviation (SD), upper percentiles and ASL of bootstrap test statistics and change-point estimators when there is one change-point in mean.

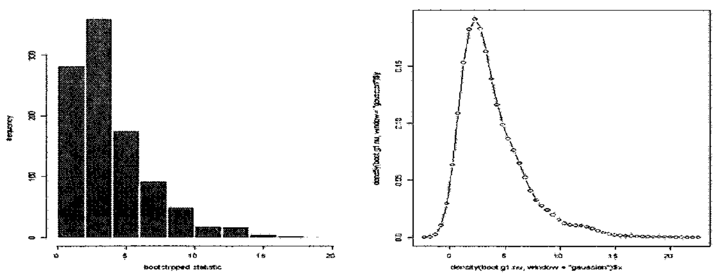

Figure 3.1: Histogram and Estimated Density of Bootstrapped Test Statistics Under H_0 with no Change in $B = 1,000$ Repetitions.

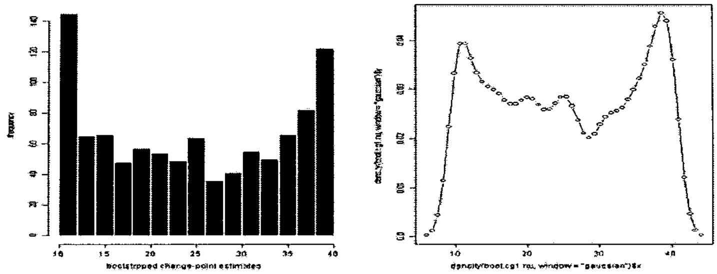

Figure 3.2: Histogram and Estimated Density of Bootstrapped Change-point Estimator Under H_0 with no Change in $B = 1,000$ Repetitions.

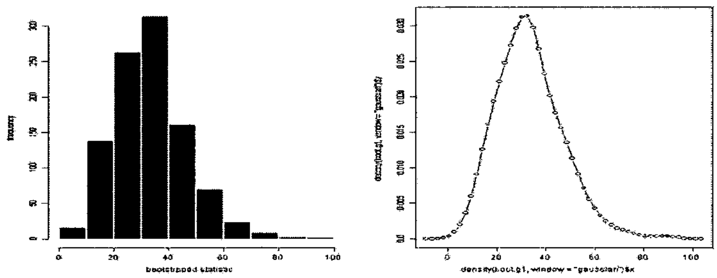

Figure 3.3: Histogram and Estimated Density of Bootstrapped Test Statistics Under $\mu_0 = 0$, $\mu_1 = 1$ and $\tau = 25$ in $B = 1,000$ Repetitions.

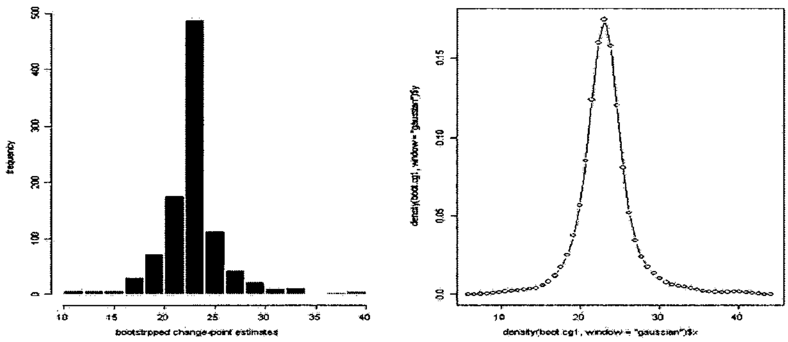

Figure 3.4: Histogram and Estimated Density of Bootstrapped Change-point Estimator when $\mu_0 = 0$, $\mu_1 = 1$ and $\tau = 25$ in $B = 1,000$ Repetitions.

Table 3.1: Simulation Results of the proposed Bootstrap statistics with upper percentiles, mean(SM), standard deviation(SD) and achieved significance level(ASL) with the Sample Size $n = 50$ in $B = 1,000$ Repetitions

Δ	τ	test	5%	10%	90%	95%	change-point estimation	5%	10%	90%	95%
0.5	15	SM 9.3400 SD 7.5243	19.5621	16.7364	3.2775	2.3948	SM 22.36 MSE 151.037	36.08	32.76	13.36	11.48
		ASL=0.2354									
	25	SM 9.5398 SD 7.2516	19.5692	16.6553	3.5496	2.6421	SM 25.52 MSE 83.124	37.34	34.80	15.94	13.12
		ASL=0.2426									
	30	SM 9.7969 SD 7.2154	20.0284	17.2748	3.6066	2.6567	SM 26.97 MSE 87.529	38.28	35.96	17.12	13.84
		ASL=0.2327									
1.0	15	SM 16.0878 SD 5.6955	30.6535	26.6530	6.9520	5.3369	SM 18.15 MSE 57.886	30.52	25.68	12.68	11.14
		ASL=0.0535									
	25	SM 18.7895 SD 5.1072	34.8241	30.4529	8.5191	6.6650	SM 24.53 MSE 39.369	33.74	30.52	18.60	15.66
		ASL=0.0468									
	30	SM 19.8511 SD 4.8599	36.2495	31.9772	9.2140	7.1895	SM 28.66 MSE 40.830	36.10	33.76	23.00	19.02
		ASL=0.0187									
1.5	15	SM 28.7718 SD 4.1532	51.0236	45.0466	14.2427	11.3064	SM 16.72 MSE 29.596	24.74	21.10	13.10	11.84
		ASL=0.0139									
	25	SM 33.7326 SD 3.7342	57.9897	51.8110	17.5264	14.2282	SM 25.11 MSE 18.036	31.48	28.86	21.36	18.94
		ASL=0.003									
	30	SM 33.3680 SD 3.7398	58.0334	51.7005	17.1525	13.9117	SM 29.42 MSE 19.072	34.98	32.98	25.68	22.74
		ASL=0.003									

Table 3.1 gives that the proposed bootstrap test rejects H_0 when the amount of change $\Delta = 1.0$, 1.5. However, when $\Delta = 0.5$, the proposed test does not reject for no change since the small change can not be captured by bootstrap resampling due to possible perturbation. The right part of Table 3.1 shows that the upper percentiles of the proposed bootstrap change-point estimator providing confidence intervals and that the change-point estimation works better when the change-point occurs in the middle of the data sequence.

4. Conclusion

Bootstrap methods are widely applicable method by distinct resampling technique. We suggested a bootstrap test for change and a bootstrap change-point estimator using the Gombay and Horvath's functional form of statistics. The proposed method works better when the change-point occurs at the middle of the data sequence than elsewhere. Other functional form of change statistics can be applied with the bootstrap method. For example, contamination of a sampled distribution can degrade the performance of a statistical estimator in which case a version of the weighted bootstrap method can be developed with each data value being assigned a weight according to an assessment of its influence on dispersion.

References

Antoch, J. and Huskova, M. (1995). Change-point problem and bootstrap. *Journal of Nonparametric Statistics*, **5**, 123–144.

Bhattacharyya, G. K. and Johnson, R. A. (1968). Nonparametric tests for shift at an unknown time point, *Annals of Mathematical Statistics*, **39**, 1731–1743.

Boukai, B.(1993). A nonparametric bootstrapped estimate of the change-point. *Journal of Nonparametric Statistics*, **3**, 123–134.

Carlstein, E. (1988). Nonparametric change-point estimation. *The Annals of Statistics*, **16**, 188–197.

Chernoff, H. and Zacks, S. (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. *The Annals of Mathematical Statistics*, **35**, 999–1018.

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

Eastwood, V. R. (1993). Some nonparametric methods for changepoint problems. *The Canadian Journal of Statistics*, **21**, 209–222.

Efron, B, and Tibshirani, R. J. (1979). *An Introduction to the Bootstrap.* Chapman & Hall/CRC, London.

Gardner, L. A. Jr.(1969). On detecting changes in the mean of normal variates. *Annals of Mathematical Statistics*, **40**, 116–126.

Gombay, E. and Horvath, L. (1990a). On the rate of approximations for maximum likelihood tests in change-point models. *Journal of Multivariate Analysis*, **56**, 120–152.

Gombay, E. and Horvath, L. (1990b). Asymptotic distributions of maximum likelihood tests for change in the mean. *Biometrika*, **77**, 411–414.

Hinkley, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika*, **57** 1–17.

Hinkley, D. and Schechtman, E. (1987). Conditional bootstrap methods in the mean-shift model. *Biometrika*, **74**, 85–93.

Kander, Z. and Zacks, S. (1966). Test procedures for possible changes in parameters of statistical distributions occurring at unknown time points. *The Annanls of Mathematical Statisitcs*, **37**, 1196–1210.

Lorden, G. (1971). Procedures for reacting to a change in distribution. *The Annanls of Mathematical Statisitcs*, **42**, 1897–1908.

Page, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika*, **42**, 523–527.

Pettitt, A. N. (1979). A non-parametric approach to the change-point problem. *Applied Statistics*, **28**, 126–135.

Romano, J. P. (1989). Bootstrap and randomization tests of some nonparametric hypotheses. *The Annals of Statistics*, **17**, 141–159.

Zacks, S. (1983). Survey of classical and Bayesian approaches to the change-point problem: fixed sample and sequential procedures of testing and estimation. *Recent Advances in Statistics, Academic Press*, 245–269.

[Received June 2007, Accepted July 2007]