Journal of Statistical Planning and Inference 35 (1993) 335-347
North-Holland

Nonparametric estimation in change-point models

Wen-Tao Huang

Institute of Statistical Science, Academia Sinica, Taipei, Taiwan, R.O.C.

Yi-Ping Chang

Graduate Institute of Statistics, National Central University, Chung-li, Taiwan, R.O.C.

Received 6 September 1991; revised manuscript received 28 May 1992

Abstract: We consider a new change-point model with some smooth mixture intervention. Some least-squares type estimators for the parameters are proposed and some large sample properties for these estimators are shown. A Monte Carlo study shows that the proposed estimators behave satisfactorily for the normal case.

AMS Subject Classification: Primary 62G05; secondary 62G20.

Key words and phrases: Change-point; smooth mixture intervention; least-squares type methods; consistency; convergence rate; asymptotic distribution.

1. Introduction

Recently, the problem of detecting a change in the mean of a sequence of random variables has attracted attention for study. Many authors have contributed various results to this problem. Generally, it is written by

$$\begin{aligned} y_i &= \theta_1 + e_i \quad (i = 1, \dots, \tau), \\ y_i &= \theta_2 + e_i \quad (i = \tau + 1, \dots, n), \end{aligned} \tag{1.1}$$

where $\{e_i\}$ is assumed to be independently identically distributed with mean 0 and constant variance σ^2, and τ is called a change-point. In this model, Page (1955) considered a test of no change ($\theta_1 = \theta_2$) against a change ($\theta_1 \neq \theta_2$). Hinkley (1970) studied some behaviors of the maximum likelihood estimator for the change-point

Correspondence to: Dr. Wen-Tao Huang, Institute of Statistical Science, Academia Sinica, Taipei 11528, Taiwan, Republic of China.

0378-3758/93/$06.00 © 1993—Elsevier Science Publishers B.V. All rights reserved

τ and also considered some related testing problem. Sen and Srivastava (1975) proposed a distribution free test and Worsley (1986) investigated confidence regions and tests for a change-point using maximum likelihood methods. Hawkins (1986) considered testing and estimation problems using least squares method. For a Bayesian approach Chernoff and Zacks (1964) considered some related estimation problem.

However, in many practical situations, some important models can not be presented in the form of (1.1). Lombard (1987) considered a testing problem in a smooth change model which is an extension of model (1.1). In many situations, it is more realistic to assume that a change occurs following some intervention over a period of time rather than an abrupt change. For example, suppose a society in which a person's income belongs to one of two different categories (for its simplicity) with respective distribution $F_1(y)$ and $F_2(y)$. When there is a force (intervention), say, some economic policy for instance, applying to the society, the income structure starts to change after some period of time. Initially, the structure is in a stable state, and its income distribution is given by $p_I F_1(y)+(1-p_I)F_2(y)$ where p_I is the ratio of numbers that belong to the first category to the numbers that belong to the second category. After the length of time τ_1, the structure starts to change and its ratio of numbers that belong to respective categories also changes with time. Hence, the income is distributed according to $p_i F_1(y)+(1-p_i)F_2(y)$ at time $i(\mathrm{i}>\tau_1)$. For its simplicity, we consider only that p_i is monotone. The changes continue until at time τ_2, the structure finally comes to a stable state. Thus the income distribution is given by $p_s F_1(y)+(1-p_s)F_2(y)$ where $p_i=p_s(i\geqslant\tau_2)$. Let the mean of the income at the initial and the final stable states be denoted respectively by θ_1 and θ_2. The change occurs when the state is stabilized after an intervening period. The phenomenon during the intervening period between τ_1 and τ_2 is described in terms of mixture distributions instead of its means. Such a model also occurs in some type of evolution in a biological environment, chemical engineering (liquid that contains more than one component, for instance), economics and social science (marketing and Gallop survey, for examples). The number of different categories may, in fact, extend to any finite number.

If the intervening period between τ_1 and τ_2 is not our main concern, it is reasonable to consider that the time span $\tau_2-\tau_1$ should be negligible comparing to the remaining time span when total time span is large enough. As we have pointed out in the previous example, the 'real' change occurs at time τ_2 though there are two change parameters of τ_1 and τ_2, where τ_1 indicates the time that the initial stable state breaks. Therefore, this model is completely different from the classical finite change points model.

Now, we introduce a change-point model with intervening period which is expressed as follows.

$$y_i \mid x_i=\theta_1 x_i+\theta_2(1-x_i)+e_i \quad (i=1,\ldots,n), \tag{1.2}$$

where $\{e_i\}$ is a sequence of independent error terms with mean 0 and finite variance $\mathrm{Var}(e_i)=\sigma_1^2 I(i\leqslant\tau_1)+\sigma_3^2 I(\tau_1<i<\tau_2)+\sigma_2^2 I(\tau_2\leqslant i\leqslant n)$, and the dummy variables x_i taking values 0 or 1, which can be interpreted as an unobservable change-factor, are independent random variables with $\mathrm{P}(x_i=1)=I(i\leqslant\tau_1)+p_i I(\tau_1<i<\tau_2)$, where $p_i=(\tau_2-i)/(\tau_2-\tau_1)$ and $I(A)$ is the indicator function of the set A. The parameters $\tau_1,\tau_2,\theta_1,\theta_2,\sigma_1^2,\sigma_2^2$ and σ_3^2 are generally all unknown, and τ_1, τ_2 are usually called the change-point parameters. That is, the first part of the independent observations $y_1,\ldots,y_{\tau_1}$ are from the first population with mean θ_1, the last part of the independent observations $y_{\tau_2},\ldots,y_n$ are from the second population with mean θ_2 and the middle part of the independent observations $y_i(\tau_1<i<\tau_2)$ are from a mixture of the first and the second population with mixing probability p_i. Therefore, the model (1.1) can be easily seen as an extreme case of the model (1.2) when $\tau_2=\tau_1+1$.

In this model, we consider σ_1^2, σ_2^2 and σ_3^2 as our nuisance parameters. We are interested in estimating the parameters τ_1, τ_2, θ_1 and θ_2.

In Section 2 we derive some least-squares type estimators for all these unknown parameters. In Section 3 we show some large sample properties of these estimators, and some Monte Carlo results are studied in Section 4.

2. Some least-squares type methods

Under the model (1.2), the error sum of squares is proportional to

$$L_n(\tau_1,\tau_2,\theta_1,\theta_2)=\frac{1}{n}\left(\sum_{i=1}^{\tau_1}(y_i-\theta_1)^2+\sum_{i=\tau_1+1}^{\tau_2-1}[y_i-\{p_i\theta_1+(1-p_i)\theta_2\}]^2\right.$$
$$\left.+\sum_{i=\tau_2}^{n}(y_i-\theta_2)^2\right). \tag{2.1}$$

We are desirable to derive some explicit forms that are convenient for numerical computation.

When τ_1 and τ_2 are known, we can use some estimators $\bar{y}(1,\tau_1)$ and $\bar{y}(\tau_2,n)$ to estimate θ_1 and θ_2, respectively, where $\bar{y}(a,b)$ denotes the quantity $\sum_{i=a}^{b} y_i/(b-a+1)$. Then, standard arguments show that these estimators are unbiased. If τ_1 and τ_2 are unknown, let $\hat{\tau}_1$ and $\hat{\tau}_2$ denote respectively the values of τ_1 and τ_2 which minimize the sum of squares given by

$$\bar{L}_n(\tau_1,\tau_2)=\frac{1}{n}\left(\sum_{i=1}^{\tau_1}\{y_i-\bar{y}(1,\tau_1)\}^2\right.$$
$$+\sum_{i=\tau_1+1}^{\tau_2-1}[y_i-\{p_i\bar{y}(1,\tau_1)+(1-p_i)\bar{y}(\tau_2,n)\}]^2$$
$$\left.+\sum_{i=\tau_2}^{n}\{y_i-\bar{y}(\tau_2,n)\}^2\right), \tag{2.2}$$

which is obtained from (2.1) by plugging in some pseudo estimators of θ_1 and θ_2. We propose estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ as follows.

$$\hat{\theta}_1 = \bar{y}(1, \hat{\tau}_1) \quad \text{and} \quad \hat{\theta}_2 = \bar{y}(\hat{\tau}_2, n), \tag{2.3}$$

where $\hat{\tau}_1$ and $\hat{\tau}_2$ are defined by (2.2).

3. Consistency and asymptotic distribution

In order to show that some large sample properties of the proposed least-squares type estimators given by (2.2) and (2.3) are to be held, we need the following assumptions.

Assumption 1. For convenience, suppose τ_{10}, τ_{20}, θ_{10}, θ_{20} and p_{i0} be the true values of τ_1, τ_2, θ_1, θ_2 and p_i, respectively, and we assume $\theta_{10} \neq \theta_{20}$ and $\tau_{10} < \tau_{20}$.

Assumption 2. When $n \to \infty$, we assume $\tau_{10} \to \infty$, $\tau_{20} \to \infty$, $\tau_{20} - \tau_{10} \to \infty$ and satisfying $[n\xi] < \tau_{10} < \tau_{20} < [n(1-\xi)]$ for some $\xi (0 < \xi < \frac{1}{2})$.

Theorem 3.1. *Under Assumptions* 1 *and* 2, $n^{-1}(\hat{\tau}_1 - \tau_{10}) \to 0$ *and* $n^{-1}(\hat{\tau}_2 - \tau_{20}) \to 0$ *with probability one as* $n \to \infty$.

Proof. For any $\varepsilon > 0$, let

$$\Lambda_{n,i}(\varepsilon) := \{(\tau_1, \tau_2) : n^{-1}|\tau_i - \tau_{i0}| > \varepsilon, [n\xi] < \tau_1 < \tau_2 < [n(1-\xi)]\},$$

and $\Lambda_n(\varepsilon) = \bigcup_{i=1}^{2} \Lambda_{n,i}(\varepsilon)$. Then, under Assumptions 1 and 2, through a tedious but straightforward computation, it can be verified that, applying Kolmogorov's theorem, if $(\tau_1, \tau_2) \in \Lambda_n(\varepsilon)$, then for n sufficiently large,

$$\bar{L}_n(\tau_1, \tau_2) - \bar{L}_n(\tau_{10}, \tau_{20}) > 0 \quad \text{with probability one.}$$

By the definitions of $\hat{\tau}_1$ and $\hat{\tau}_2$, we have

$$\mathrm{P}\{(\hat{\tau}_1, \hat{\tau}_2) \notin \Lambda_n(\varepsilon) \quad \text{for } n \text{ sufficiently large}\} = 1. \qquad \square$$

For its own mathematical interest, we can conclude the following results.

Corollary 3.1. *Under Assumptions* 1 *and* 2, *and in addition, if* $\tau_{10}/n \to \lambda_{10}$, $\tau_{20}/n \to \lambda_{20}$ *as* $n \to \infty$, *and* $\lambda_{10} < \lambda_{20}$, *then* $\hat{\tau}_1/n \to \lambda_{10}$ *and* $\hat{\tau}_2/n \to \lambda_{20}$ *with probability one as* $n \to \infty$.

In the model we are interested in, the intervening period of $(\tau_{20} - \tau_{10})$ is naturally supposed to be short comparing to other period in some sense. Therefore, we have the following which is of our main interest.

Corollary 3.2. *Under Assumptions* 1 *and* 2, *and in addition, if* $\tau_{10}/n \to \lambda_0, \tau_{20}/n \to \lambda_0$ *and* $n^{-1}(\tau_{20}-\tau_{10}) = Kh_n$ *where* $h_n \to 0$, $nh_n \to \infty$ *as* $n \to \infty$ *and K is a positive number, then* $\hat{\tau}_1/n \to \lambda_0$ *and* $\hat{\tau}_2/n \to \lambda_0$ *with probability one as* $n \to \infty$.

The length of the intervening period is related to the rate of convergence of each change-point estimator which is shown in the following.

Theorem 3.2. *If assumptions of Corollary* 3.2 *hold, and in addition,* $h_n^{-(1-\varrho)}(\tau_{10}/n-\lambda_0) \to 0$, *and* $h_n^{-(1-\varrho)}(\tau_{20}/n-\lambda_0) \to 0$ *as* $n \to \infty$ *for some* $0<\varrho<1$, *then* $h_n^{-(1-\varrho)}(\hat{\tau}_1/n-\lambda_0) \to 0$ *and* $h_n^{-(1-\varrho)}(\hat{\tau}_2/n-\lambda_0) \to 0$ *in probability as* $n \to \infty$.

Proof. Apply the analogous argument given in Yao and Au (1989). For any $\varepsilon>0$, let

$$\Lambda^*_{n,i,k}(\varepsilon) := \{(\tau_1,\tau_2) : h_n^{\upsilon_{k+1}}\varepsilon < n^{-1}|\tau_i-\tau_{i0}| \leqslant h_n^{\upsilon_k}\varepsilon, [n\xi]<\tau_1<\tau_2<[n(1-\xi)]\},$$

$$\Lambda^*_{n,i}(\varepsilon) := \{(\tau_1,\tau_2) : h_n^{1-\varrho}\varepsilon < n^{-1}|\tau_i-\tau_{i0}| \leqslant \varepsilon, [n\xi]<\tau_1<\tau_2<[n(1-\xi)]\},$$

$$\bar{\Lambda}^*_{n,k}(\varepsilon) = \{\textstyle\bigcup_{i=1}^2 \Lambda_{n,i}(\varepsilon)\}^c \cap \{\bigcup_{i=1}^2 \Lambda^*_{n,i,k}(\varepsilon)\},$$

and

$$\bar{\Lambda}^*_n(\varepsilon) = \{\textstyle\bigcup_{i=1}^2 \Lambda_{n,i}(\varepsilon)\}^c \cap \{\bigcup_{i=1}^2 \Lambda^*_{n,i}(\varepsilon)\},$$

where $\upsilon_{k+1} = \frac{1}{3}(2\upsilon_k+1)$ and $\upsilon_0=0$. It is clear that $\upsilon_k \uparrow 1$ as $k\to\infty$ and $\bar{\Lambda}^*_n(\varepsilon) \subseteq \bigcup_{k=0}^\infty \bar{\Lambda}^*_{n,k}(\varepsilon)$. Therefore, for any $0<\varrho<1$, there exists a unique finite M such that $\upsilon_M \leqslant 1-\varrho < \upsilon_{M+1}$ and $\bar{\Lambda}^*_n(\varepsilon) \subseteq \bigcup_{k=0}^M \bar{\Lambda}^*_{n,k}(\varepsilon)$. Let

$$\bar{y}_i^*(\tau_1,\tau_2) = \bar{y}(1,\tau_1)I(i\leqslant\tau_1) + \{p_i\bar{y}(1,\tau_1)+(1-p_i)\bar{y}(\tau_2,n)\}I(\tau_1<i<\tau_2) + \bar{y}(\tau_2,n)I(\tau_2\leqslant i), \tag{3.1}$$

where $p_i = (\tau_2-i)/(\tau_2-\tau_1)$. Through a straightforward computation, under assumptions of Theorem 3.2, it can be shown that

$$\sum_{i=1}^n \max_{(\tau_1,\tau_2)\in\bigcup_{k=0}^M \bar{\Lambda}^*_{n,k}(\varepsilon)} \frac{\{E\bar{y}_i^*(\tau_1,\tau_2)-E\bar{y}_i^*(\tau_{10},\tau_{20})\}^2}{\sum_{i=1}^n \{E\bar{y}_i^*(\tau_1,\tau_2)-E\bar{y}_i^*(\tau_{10},\tau_{20})\}^2} = \mathrm{O}(1),$$

and if $(\tau_1,\tau_2)\in\Lambda^*_{n,j,k}(\varepsilon)$ for any fixed j and k, then

$$\sum_{i=1}^n \{E\bar{y}_i^*(\tau_1,\tau_2)-E\bar{y}_i^*(\tau_{10},\tau_{20})\}^2 \geqslant Cnh_n,$$

where C is a positive real number. For example, for any fixed k, if $(\tau_1,\tau_2)\in\Lambda^*_{n,1,k}(\varepsilon)$ with $\tau_1<\tau_{10}\leqslant\tau_2\leqslant\tau_{20}$, then

$$\sum_{i=1}^n \{E\bar{y}_i^*(\tau_1,\tau_2)-E\bar{y}_i^*(\tau_{10},\tau_{20})\}^2$$

$$\geqslant \sum_{i=\tau_1+1}^{\tau_{10}} (1-p_i)^2\{E\bar{y}(\tau_2,n)-\theta_{10}\}^2$$

$$= \frac{(\tau_{10}-\tau_1)(\tau_{10}-\tau_1+1)\{2(\tau_{10}-\tau_1)+1\}}{6(\tau_2-\tau_1)^2}\{E\bar{y}(\tau_2,n)-\theta_{10}\}^2$$

$$\geqslant \frac{(nh_n^{\upsilon_{k+1}}\varepsilon)^3}{3(nh_n^{\upsilon_k})^2}\{E\bar{y}(\tau_2,n)-\theta_{10}\}^2$$

$$= Cnh_n.$$

From these properties and a tedious but straightforward computation, it can be shown that

$$\min_{(\tau_1,\tau_2)\in\bigcup_{k=0}^M \bar{A}_{n,k}^*(\varepsilon)} \bar{L}_n(\tau_1,\tau_2)-\bar{L}_n(\tau_{10},\tau_{20})$$

$$\geqslant \min_{(\tau_1,\tau_2)\in\bigcup_{k=0}^M \bar{A}_{n,k}^*(\varepsilon)} \frac{1}{n}\sum_{i=1}^n \{E\bar{y}_i^*(\tau_1,\tau_2)-E\bar{y}_i^*(\tau_{10},\tau_{20})\}^2(1+o_p(1))$$

$$\geqslant Ch_n + o_p(h_n),$$

where C is a positive real number. Therefore

$$\lim_{n\to\infty} \mathrm{P}\left[\min_{(\tau_1,\tau_2)\in\bigcup_{k=0}^M \bar{A}_{n,k}^*(\varepsilon)} \bar{L}_n(\tau_1,\tau_2)-\bar{L}_n(\tau_{10},\tau_{20})>0\right]=1.$$

Hence, by the definitions of $\hat{\tau}_1$ and $\hat{\tau}_2$, we have

$$\lim_{n\to\infty} \mathrm{P}\left\{(\hat{\tau}_1,\hat{\tau}_2)\notin\bigcup_{k=0}^M \bar{A}_{n,k}^*(\varepsilon)\right\}=1.$$

Together with Corollary 3.2 and the fact that $h_n^{-(1-\varrho)}(\tau_{10}/n-\lambda_0)\to 0$ and $h_n^{-(1-\varrho)}(\tau_{20}/n-\lambda_0)\to 0$ as $n\to\infty$, the proof is complete. □

Theorem 3.3. *If assumptions of Corollary 3.1 hold, and in addition,* $\sup_i E(|e_i|^{2+\delta})<\infty$ *for some* $\delta>0$, *then* $\hat{\theta}_1\to\theta_{10}$ *and* $\hat{\theta}_2\to\theta_{20}$ *with probability one as* $n\to\infty$.

Proof. It suffices to show that $\hat{\theta}_1\to\theta_{10}$ with probability one as $n\to\infty$. Proofs are analogous for remaining parts and thus are omitted.

For an arbitrarily given $\varepsilon>0$, let η be a fixed number such that $\eta<\lambda_{10}\varepsilon/2|\theta_{20}-\theta_{10}|$. Then, we have

$$\mathrm{P}\left[\bigcup_{n=m}^{\infty}\{|\bar{y}(1,\hat{\tau}_1)-\theta_{10}|\geqslant\varepsilon\}\right]$$

$$\leqslant \mathrm{P}\left(\bigcup_{n=m}^{\infty}[\{|\bar{y}(1,\hat{\tau}_1)-\theta_{10}|\geqslant\varepsilon\},(n^{-1}|\hat{\tau}_1-\tau_{10}|<\eta)]\right)$$

$$+\mathrm{P}\left\{\bigcup_{n=m}^{\infty}(n^{-1}|\hat{\tau}_1-\tau_{10}|\geqslant\eta)\right\}. \tag{3.2}$$

By Theorem 3.1, the second term on the right-hand side of (3.2) tends to zero as $m \to \infty$, and the first term is less than or equal to

$$\sum_{n=m}^{\infty} \mathrm{P}\left(\sup_{A_n \leqslant k \leqslant B_n} \left| k^{-1} \sum_{i=1}^{k} y_i - \theta_{10} \right| \geqslant \varepsilon\right)$$
$$\leqslant \sum_{n=m}^{\infty} \mathrm{P}\left[\sup_{A_n \leqslant k \leqslant B_n} \left| k^{-1} \sum_{i=1}^{k} \{y_i - E(y_i)\} \right| + \sup_{A_n \leqslant k \leqslant B_n} \left| k^{-1} \sum_{i=1}^{k} E(y_i) - \theta_{10} \right| \geqslant \varepsilon\right], \tag{3.3}$$

where $A_n = [\tau_{10} - n\eta]$ and $B_n = [\tau_{10} + n\eta]$.

It is easy to see that the event

$$\sup_{A_n \leqslant k \leqslant B_n} \left| k^{-1} \sum_{i=1}^{k} \{y_i - E(y_i)\} \right| + \sup_{A_n \leqslant k \leqslant B_n} \left| k^{-1} \sum_{i=1}^{k} E(y_i) - \theta_{10} \right| \geqslant \varepsilon,$$

implies that

$$\begin{aligned}
\sup_{A_n \leqslant k \leqslant B_n} \left| k^{-1} \sum_{i=1}^{k} \{y_i - E(y_i)\} \right| &\geqslant \varepsilon - \sup_{\tau_{10} \leqslant k \leqslant B_n} \left| k^{-1} \sum_{i=1}^{k} E(y_i) - \theta_{10} \right| \\
&> \varepsilon - (B_n - \tau_{10}) |\theta_{20} - \theta_{10}| / \tau_{10} \\
&\to \varepsilon - \eta |\theta_{20} - \theta_{10}| / \lambda_{10} \quad \text{as } n \to \infty \\
&> \tfrac{1}{2}\varepsilon \equiv \varepsilon', \quad \text{say.}
\end{aligned}$$

So, for m sufficiently large, the right-hand side of (3.3) is less than

$$\begin{aligned}
&\sum_{n=m}^{\infty} \mathrm{P}\left[\sup_{A_n \leqslant k \leqslant B_n} \left| k^{-1} \sum_{i=1}^{k} \{y_i - E(y_i)\} \right| > \varepsilon'\right] \\
&\leqslant \sum_{n=m}^{\infty} \mathrm{P}\left[\sup_{A_n \leqslant k \leqslant B_n} \left| \sum_{i=1}^{k} \{y_i - E(y_i)\} \right| > \varepsilon' A_n\right] \\
&\leqslant \sum_{n=m}^{\infty} E\left[\sup_{A_n \leqslant k \leqslant B_n} \left| \sum_{i=1}^{k} \{y_i - E(y_i)\} \right|^{2+\delta}\right] / (\varepsilon' A_n)^{2+\delta}.
\end{aligned} \tag{3.4}$$

Applying the Burkholder's inequality (Burkholder, 1973) and the Jensen's inequality, the right-hand side of (3.4) is less then or equal to

$$\begin{aligned}
&\sum_{n=m}^{\infty} CE\left(\left[\sum_{i=1}^{B_n} \{y_i - E(y_i)\}^2\right]^{(2+\delta)/2}\right) / A_n^{2+\delta} \\
&\leqslant \sum_{n=m}^{\infty} C' B_n^{(2+\delta)/2-1} \sum_{i=1}^{B_n} E([\{y_i - E(y_i)\}^2]^{(2+\delta)/2}) / A_n^{2+\delta} \\
&\leqslant \sum_{n=m}^{\infty} C^* n^{-(1+\delta/2)},
\end{aligned}$$

which tends to zero as $m \to \infty$, where C, C' and C^* are some constants. The proof is thus complete. □

Theorem 3.4. *If assumptions of Corollary* 3.2 *hold, and in addition,* $\sup_i E(|e_i|^{2+\delta}) < \infty$ *for some* $\delta > 0$, *then* $\hat\theta_1 \to \theta_{10}$ *and* $\hat\theta_2 \to \theta_{20}$ *with probability one as* $n \to \infty$.

In the following of this section, we derive the asymptotic distributions of the proposed least-squares type estimators.

Under assumptions of Corollary 3.2, and in addition, $\sup_i E(|e_i|^{2+\delta}) < \infty$ for some $\delta > 0$, the usual multivariate central limit theorem shows that the random vector

$$n^{1/2}\left[\frac{1}{\tau_{10}}\sum_{i=1}^{\tau_{10}}\{y_i - E(y_i)\}, \frac{1}{n-\tau_{20}+1}\sum_{i=\tau_{20}}^{n}\{y_i - E(y_i)\}\right]'$$

converges in distribution to a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix

$$V = \begin{bmatrix} \sigma_1^2/\lambda_0 & 0 \\ 0 & \sigma_2^2/(1-\lambda_0) \end{bmatrix}. \tag{3.5}$$

Applying the same technique in the proof of the Anscombe's Theorem (see, e.g., Chow and Teicher, 1988) we can show that the random vector

$$n^{1/2}\left[\frac{1}{\hat\tau_1}\sum_{i=1}^{\hat\tau_1}\{y_i - E(y_i)\}, \frac{1}{\mathrm{n}-\hat\tau_2+1}\sum_{i=\hat\tau_2}^{n}\{y_i - E(y_i)\}\right]'$$

converges in distribution to the same multivariate normal distribution with mean $\mathbf{0}$ and covariance V defined by (3.5). In order to obtain the asymptotic distributions of $\hat\theta_1$ and $\hat\theta_2$, we need the following lemma.

Lemma 3.1. *Under assumptions of Theorem* 3.2, *and in addition,* $nh_n^{2(1-\varrho)} = \mathrm{O}(1)$ *for some* $\varrho > 0$, *the following convergence hold in probability as* $n \to \infty$.

(i) $n^{-1/2}\sum_{i=1}^{\hat\tau_1}\{E(y_i) - \theta_{10}\} \to 0$; *and*

(ii) $n^{-1/2}\sum_{i=\hat\tau_2}^{n}\{E(y_i) - \theta_{20}\} \to 0$.

Proof. We suffice to prove only part (i). For any $\varepsilon > 0$, let η be a fixed number such that $\eta < \varepsilon/|\theta_{20} - \theta_{10}|$, we have

$$\mathrm{P}\left[n^{-1/2}\left|\sum_{i=1}^{\hat\tau_1}\{E(y_i) - \theta_{10}\}\right| \geq \varepsilon\right]$$

$$\leq \mathrm{P}\left[n^{-1/2}\left|\sum_{i=1}^{\hat\tau_1}\{E(y_i) - \theta_{10}\}\right| \geq \varepsilon, n^{-1}|\hat\tau_1 - \tau_{10}| \leq n^{-1/2}\eta\right]$$

$$+\mathrm{P}(n^{-1}|\hat{\tau}_1-\tau_{10}|>n^{-1/2}\eta)$$

$$\leqslant \mathrm{P}\left[n^{-1/2}\sum_{i=1}^{\hat{\tau}_{10}+n^{1/2}\eta}|\{E(y_i)-\theta_{10}\}|\geqslant\varepsilon\right]$$

$$+\mathrm{P}(n^{-1}|\hat{\tau}_1-\tau_{10}|>n^{-1/2}\eta). \tag{3.6}$$

The first term on the right-hand side of (3.6) tends to zero as $n\to\infty$. By Theorem 3.2 and $nh_n^{2(1-\varrho)}=\mathrm{O}(1)$, we have $n^{-1/2}(\hat{\tau}_1-\tau_{10})=\mathrm{o}_p(1)$, hence the second term on the right-hand side of (3.6) tends to zero as n tends to infinity. □

Using the Slutsky's Theorem, we can conclude the following asymptotic normality of $\hat{\theta}_1$ and $\hat{\theta}_2$.

Theorem 3.5. *Under assumptions of Lemma* 3.1, *and in addition,* $\sup_i E(|e_i|^{2+\delta})<\infty$ *for some* $\delta>0$, *as* $n\to\infty$, $n^{1/2}(\hat{\theta}_1-\theta_{10})$ *and* $n^{1/2}(\hat{\theta}_2-\theta_{20})$ *are asymptotically independent and normally distributed with common mean* 0 *and variance* σ_1^2/λ_0 *and* $\sigma_2^2/(1-\lambda_0)$, *respectively.*

To obtain the asymptotic confidence regions for θ_{10} and θ_{20}, we propose estimators for σ_1^2 and σ_2^2 as follows.

$$\hat{\sigma}_1^2=\frac{1}{\hat{\tau}_1}\sum_{i=1}^{\hat{\tau}_1}\{y_i-\bar{y}(1,\hat{\tau}_1)\}^2,$$

and

$$\hat{\sigma}_2^2=\frac{1}{n-\hat{\tau}_2+1}\sum_{i=\hat{\tau}_2}^{n}\{y_i-\bar{y}(\hat{\tau}_2,n)\}^2.$$

Using analogous arguments, we can also obtain the following theorem.

Theorem 3.6. *Under assumptions of Corollary* 3.2, *and in addition, if* $\sup_i E(|e_i|^{4+\delta})<\infty$ *for some* $\delta>0$, *then* $\hat{\sigma}_1^2\to\sigma_1^2$ *and* $\hat{\sigma}_2^2\to\sigma_2^2$ *with probability one as* $n\to\infty$.

If we know a priori that $\sigma_1^2=\sigma_2^2=\sigma_3^2=\sigma^2$, then we construct a pool estimator $\hat{\sigma}^2$ given by

$$\hat{\sigma}^2=\bar{L}_n(\hat{\tau}_1,\hat{\tau}_2)-\frac{1}{n}\sum_{i=\hat{\tau}_1+1}^{\hat{\tau}_2-1}\hat{p}_i(1-\hat{p}_i)(\hat{\theta}_2-\hat{\theta}_1)^2,$$

where $\hat{\tau}_i$ and $\hat{\theta}_i (i=1,2)$ are defined by (2.2) and (2.3), and $\hat{p}_i=(\hat{\tau}_2-i)/(\hat{\tau}_2-\hat{\tau}_1)$.

Theorem 3.7. *Under assumptions of Corollary* 3.2, *and in addition, if* $\sup_i E(|e_i|^{4+\delta})<\infty$ *for some* $\delta>0$, *then* $\hat{\sigma}^2\to\sigma^2$ *with probability one as* $n\to\infty$.

4. Monte Carlo results

It is desirable to see how the proposed estimators perform in small samples. In this section we suppose that $\sigma_1^2=\sigma_2^2=\sigma_3^2=\sigma^2$. Associated with different intervening periods, Tables 1, 2 and 3 show, respectively, some mean absolute error and bias of the proposed estimators given in Section 2 based on simulated data for respective sample sizes $n=50$, 100 and 200. In each case, $y_1,\dots,y_{[n\lambda_{10}]}$ are generated from the normal $N(\theta_{10},1)$, $y_{[n\lambda_{20}]},\dots,y_n$ are generated from the normal $N(\theta_{20},1)$ and $y_i([n\lambda_{10}]+1\leqslant i\leqslant[n\lambda_{20}]-1)$ are generated from the mixture $p_{i0}N(\theta_{10},1)+$

Table 1
Mean absolute error of four estimators $\hat{\tau}_2/n$, $\hat{\theta}_1$, $\hat{\theta}_2$, and $\hat{\sigma}^2$ computed from 1000 samples of size n from Normal distribution with $\sigma^2=1$, and $\theta_{10}=0$, where $\lambda_{20}=\lambda_0$ and $\lambda_{10}=\lambda_{20}-(\log n)^{-1}$. Corresponding bias are given in parentheses

n	θ_{20}	λ_0	$\hat{\tau}_2/n$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\sigma}^2$
50	1	0.3	0.190(0.083)	0.554(0.136)	0.183(−0.002)	0.168(−0.057)
		0.4	0.135(−0.010)	0.374(−0.027)	0.160(0.029)	0.172(−0.060)
		0.5	0.131(−0.044)	0.311(−0.058)	0.175(0.001)	0.179(−0.074)
		0.6	0.124(−0.068)	0.265(−0.070)	0.174(0.021)	0.191(−0.086)
		0.7	0.130(−0.090)	0.225(−0.055)	0.198(0.000)	0.182(−0.085)
	4	0.3	0.071(−0.047)	0.777(−0.532)	0.139(−0.031)	0.404(0.232)
		0.4	0.065(−0.042)	0.391(0.156)	0.151(−0.028)	0.410(0.134)
		0.5	0.068(−0.037)	0.287(0.075)	0.178(−0.053)	0.409(0.063)
		0.6	0.075(−0.054)	0.219(0.034)	0.186(−0.052)	0.371(0.037)
		0.7	0.080(−0.061)	0.196(0.036)	0.225(−0.084)	0.361(0.052)
100	1	0.3	0.109(−0.033)	0.314(0.015)	0.110(0.018)	0.112(−0.022)
		0.4	0.104(−0.042)	0.227(−0.017)	0.107(0.010)	0.128(−0.042)
		0.5	0.101(−0.044)	0.181(−0.033)	0.116(0.013)	0.131(−0.052)
		0.6	0.101(−0.055)	0.161(−0.030)	0.126(0.009)	0.128(−0.039)
		0.7	0.102(−0.064)	0.140(−0.032)	0.146(0.003)	0.130(−0.046)
	4	0.3	0.057(−0.046)	0.405(0.197)	0.106(−0.019)	0.300(0.163)
		0.4	0.056(−0.032)	0.247(0.061)	0.113(−0.029)	0.311(0.066)
		0.5	0.052(−0.027)	0.183(0.040)	0.118(−0.028)	0.298(0.047)
		0.6	0.057(−0.039)	0.150(0.032)	0.140(−0.042)	0.281(0.076)
		0.7	0.056(−0.041)	0.126(0.026)	0.167(−0.061)	0.274(0.079)
200	1	0.3	0.082(−0.048)	0.189(0.016)	0.070(0.003)	0.088(−0.016)
		0.4	0.078(−0.037)	0.153(−0.001)	0.076(0.006)	0.088(−0.008)
		0.5	0.083(−0.030)	0.125(−0.013)	0.085(0.013)	0.093(−0.030)
		0.6	0.080(−0.045)	0.104(−0.013)	0.088(−0.001)	0.087(−0.020)
		0.7	0.084(−0.045)	0.092(−0.007)	0.104(0.005)	0.087(−0.019)
	4	0.3	0.036(−0.024)	0.231(0.074)	0.073(−0.013)	0.205(0.057)
		0.4	0.035(−0.017)	0.164(0.045)	0.077(−0.012)	0.199(0.064)
		0.5	0.038(−0.017)	0.118(0.024)	0.089(−0.013)	0.195(0.042)
		0.6	0.039(−0.025)	0.100(0.022)	0.096(−0.028)	0.191(0.067)
		0.7	0.040(−0.024)	0.089(0.017)	0.113(−0.031)	0.188(0.054)

$(1-p_{i0})N(\theta_{20}, 1)$ where $p_{i0}=([n\lambda_{20}]-i)/([n\lambda_{20}]-[n\lambda_{10}])$. In each table, we take $\lambda_{20}=\lambda_0$. In Table 1, we take $\lambda_{10}=\lambda_{20}-(\log n)^{-1}$, in Table 2, we take $\lambda_{10}=\lambda_{20}-n^{-1/2}$ and in Table 3, we take $\lambda_{10}=\lambda_{20}-\log n/n$. Here we take $\xi=0.05$ and compute the estimators defined in Section 2 using the generated data, and λ_0 is estimated by $\hat{\tau}_2/n$. In each table, we repeat 1000 times for each case of size $n=50$, 100 and 200. For random number generation, we use the subroutines DRNNOA and RNBIN of the IMSL Package. It is seen that when n is 200 (large), the simulation results seem better when the intervening period is shorter, and this is explained in Theorem 3.2 in some sense.

Table 2
Mean absolute error of four estimators $\hat{\tau}_2/n$, $\hat{\theta}_1$, $\hat{\theta}_2$, and $\hat{\sigma}^2$ computed from 1000 samples of size n from Normal distribution with $\sigma^2=1$, and $\theta_{10}=0$, where $\lambda_{20}=\lambda_0$ and $\lambda_{10}=\lambda_{20}-n^{-1/2}$. Corresponding bias are given in parentheses

n	θ_{20}	λ_0	$\hat{\tau}_2/n$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\sigma}^2$
50	1	0.3	0.131(0.050)	0.437(−0.003)	0.165(−0.004)	0.176(−0.077)
		0.4	0.111(0.019)	0.329(−0.050)	0.165(0.024)	0.183(−0.090)
		0.5	0.096(−0.012)	0.269(−0.080)	0.172(0.033)	0.184(−0.085)
		0.6	0.100(−0.029)	0.245(−0.079)	0.183(0.013)	0.184(−0.091)
		0.7	0.103(−0.053)	0.230(−0.066)	0.221(−0.003)	0.182(−0.087)
	4	0.3	0.042(−0.026)	0.365(0.099)	0.137(−0.015)	0.276(−0.050)
		0.4	0.042(−0.013)	0.261(0.031)	0.154(−0.024)	0.299(−0.058)
		0.5	0.043(−0.011)	0.211(0.018)	0.167(−0.011)	0.300(−0.070)
		0.6	0.044(−0.024)	0.181(0.016)	0.180(−0.030)	0.287(−0.090)
		0.7	0.045(−0.028)	0.162(0.006)	0.222(−0.051)	0.273(−0.078)
100	1	0.3	0.072(0.000)	0.239(−0.040)	0.098(0.018)	0.124(−0.052)
		0.4	0.061(0.002)	0.185(−0.043)	0.107(0.016)	0.122(−0.047)
		0.5	0.066(0.003)	0.162(−0.044)	0.122(0.013)	0.132(−0.058)
		0.6	0.064(−0.009)	0.138(−0.033)	0.139(0.021)	0.129(−0.059)
		0.7	0.060(−0.020)	0.120(−0.027)	0.147(0.015)	0.120(−0.047)
	4	0.3	0.029(−0.019)	0.211(0.029)	0.094(−0.007)	0.182(−0.031)
		0.4	0.027(−0.008)	0.157(0.018)	0.103(−0.008)	0.179(−0.024)
		0.5	0.026(−0.007)	0.133(0.012)	0.117(−0.015)	0.183(−0.034)
		0.6	0.027(−0.017)	0.112(0.011)	0.129(−0.012)	0.173(−0.034)
		0.7	0.030(−0.019)	0.102(0.009)	0.147(−0.023)	0.172(−0.014)
200	1	0.3	0.043(−0.003)	0.144(−0.022)	0.069(0.009)	0.086(−0.026)
		0.4	0.048(0.003)	0.119(−0.026)	0.077(0.011)	0.090(−0.035)
		0.5	0.044(0.000)	0.100(−0.014)	0.085(0.008)	0.087(−0.024)
		0.6	0.042(−0.005)	0.083(−0.008)	0.098(0.014)	0.088(−0.028)
		0.7	0.040(−0.009)	0.073(−0.007)	0.104(0.006)	0.089(−0.030)
	4	0.3	0.019(−0.011)	0.133(0.020)	0.069(−0.005)	0.117(0.000)
		0.4	0.019(−0.008)	0.105(0.015)	0.075(−0.009)	0.119(−0.003)
		0.5	0.019(−0.008)	0.094(0.014)	0.083(−0.012)	0.120(0.008)
		0.6	0.019(−0.012)	0.079(0.013)	0.096(−0.006)	0.116(−0.002)
		0.7	0.020(−0.012)	0.071(0.009)	0.107(−0.021)	0.121(−0.004)

Table 3
Mean absolute error of four estimators $\hat{\tau}_2/n$, $\hat{\theta}_1$, $\hat{\theta}_2$, and $\hat{\sigma}^2$ computed from 1000 samples of size n from Normal distribution with $\sigma^2=1$, and $\theta_{10}=0$, where $\lambda_{20}=\lambda_0$ and $\lambda_{10}=\lambda_{20}-\log n/n$. Corresponding bias are given in parentheses

n	θ_{20}	λ_0	$\hat{\tau}_2/n$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\sigma}^2$
50	1	0.3	0.104(0.058)	0.388(−0.082)	0.151(0.023)	0.187(−0.091)
		0.4	0.093(0.045)	0.279(−0.063)	0.161(0.029)	0.187(−0.094)
		0.5	0.091(0.033)	0.260(−0.064)	0.178(0.034)	0.181(−0.098)
		0.6	0.072(0.005)	0.208(−0.069)	0.182(0.026)	0.179(−0.092)
		0.7	0.089(−0.024)	0.215(−0.085)	0.224(−0.003)	0.179(−0.095)
	4	0.3	0.020(0.000)	0.254(−0.016)	0.130(0.005)	0.255(−0.171)
		0.4	0.025(0.013)	0.211(−0.005)	0.148(−0.003)	0.256(−0.150)
		0.5	0.025(0.012)	0.182(0.005)	0.160(0.005)	0.249(−0.146)
		0.6	0.020(0.000)	0.159(−0.010)	0.169(0.001)	0.240(−0.158)
		0.7	0.020(0.000)	0.143(−0.002)	0.208(−0.023)	0.234(−0.156)
100	1	0.3	0.051(0.020)	0.223(−0.049)	0.103(0.018)	0.124(−0.054)
		0.4	0.056(0.028)	0.169(−0.049)	0.111(0.029)	0.122(−0.053)
		0.5	0.056(0.022)	0.151(−0.043)	0.120(0.022)	0.129(−0.060)
		0.6	0.051(0.017)	0.131(−0.027)	0.134(0.029)	0.127(−0.061)
		0.7	0.048(0.004)	0.119(−0.031)	0.148(0.005)	0.125(−0.050)
	4	0.3	0.012(−0.003)	0.172(−0.005)	0.097(0.001)	0.150(−0.084)
		0.4	0.015(0.005)	0.136(−0.005)	0.101(0.008)	0.150(−0.074)
		0.5	0.014(0.005)	0.122(−0.001)	0.113(0.000)	0.154(−0.079)
		0.6	0.013(−0.002)	0.109(0.000)	0.121(0.003)	0.151(−0.088)
		0.7	0.013(−0.003)	0.098(0.002)	0.136(−0.016)	0.145(−0.074)
200	1	0.3	0.033(0.015)	0.141(−0.033)	0.069(0.010)	0.085(−0.034)
		0.4	0.031(0.015)	0.110(−0.023)	0.075(0.009)	0.090(−0.034)
		0.5	0.031(0.015)	0.096(−0.020)	0.079(0.013)	0.087(−0.025)
		0.6	0.032(0.012)	0.084(−0.011)	0.096(0.018)	0.089(−0.031)
		0.7	0.029(0.009)	0.074(−0.011)	0.107(0.014)	0.087(−0.034)
	4	0.3	0.008(−0.003)	0.115(0.005)	0.066(0.000)	0.092(−0.041)
		0.4	0.008(0.001)	0.094(0.005)	0.072(−0.001)	0.099(−0.038)
		0.5	0.008(0.001)	0.085(0.001)	0.075(0.001)	0.094(−0.028)
		0.6	0.008(−0.003)	0.077(0.003)	0.090(0.002)	0.097(−0.039)
		0.7	0.008(−0.002)	0.069(0.001)	0.101(−0.004)	0.098(−0.044)

Acknowledgements

The authors are thankful to a referee and an associate editor for their patience to read this paper carefully and critically and very helpful comments and suggestions to clarify this presentation. They also thank the referee for pointing out some errors in a proof. This research was partially supported by Grant NSC 81-0208-M-001-84 from the National Science Council of Taiwan, R.O.C.

References

Burkholder, D.L. (1973). Distribution function inequalities for martingales. *Ann. Prob.* **1**, 19–42.

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

Chow, Y.S. and H. Teicher (1988). *Probability Theory: Independence, Interchangeability, Martingales.* Springer-Verlag, Heidelberg.

Hawkins, D.L. (1986). A simple least squares method for estimating a change in mean. *Comm. Statist.-Simula.* **15**, 655–679.

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

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

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

Sen, A. and M.S. Srivastava (1975). On tests for detecting changes in means. *Ann. Statist.* **3**, 98–108.

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

Yao, Y.C. and S.T. Au (1989). Least-squares estimation of a step function. *Sankhyā A* **51**, 370–381.