Detecting Change-Point via Saddlepoint Approximations

Zhaoyuan LI

Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing 100872, China; Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong, China

E-mail: zyli12@hku.hk

Maozai TIAN*

Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing 100872, China; School of Statistics, Lanzhou University of Finance and Economics, Lanzhou 730101, China; School of Statistics and Information, Xinjiang University of Finance and Economics, Urumqi 830012, China

E-mail: mztian@ruc.edu.cn

Abstract It's well-known that change-point problem is an important part of model statistical analysis. Most of the existing methods are not robust to criteria of the evaluation of change-point problem. In this article, we consider “mean-shift” problem in change-point studies. A quantile test of single quantile is proposed based on saddlepoint approximation method. In order to utilize the information at different quantile of the sequence, we further construct a “composite quantile test” to calculate the probability of every location of the sequence to be a change-point. The location of change-point can be pinpointed rather than estimated within a interval. The proposed tests make no assumptions about the functional forms of the sequence distribution and work sensitively on both large and small size samples, the case of change-point in the tails, and multiple change-points situation. The good performances of the tests are confirmed by simulations and real data analysis. The saddlepoint approximation based distribution of the test statistic that is developed in the paper is of independent interest and appealing. This finding may be of independent interest to the readers in this research area.

Keywords change-point; mean shift; saddlepoint approximations; composite quantile; bootstrap

1 Introduction

Consider a sequence of random variables $X_1, X_2, \dots, X_n$, then the sequence is said to have a change-point at m if X_t for $t = 1, 2, \dots, m$ have a common distribution function $F_1(x)$ and X_t for $t = m + 1, m + 2, \dots, n$ have a common distribution function $F_2(x)$, and $F_1(x) \neq F_2(x)$ ^[1]. Change-point analysis is very important and applied in many fields, including macroeconomics^[2], environment^[3, 4], quality control^[5–7], medicine^[8], geology^[9], and etc.

Received December 4, 2015, accepted January 7, 2016

*Corresponding author

Theoretically, various authors concentrate on the change-point problems. Basically, studies can be divided into two focuses, “change-point in distribution” and “change-point in regression function”. “Change-point in distribution” refers that a pre-change distribution and a post-change distribution exist in the sequence^[10]. Studies on this issue include Pettitt^[1, 11], Hinkley^[12, 13], Talwar and Gentle^[14], Carlstein^[15], Inclan and Tiao^[16], Hawkins and Zamba^[35], Kawahara and Sugiyama^[18] and etc. “Change-point in regression function”, on the other side, refers the changed regression models, like parameter instability and structural change^[19]. Previous studies include Muller^[20] and Loader^[21], who adapt kernel smoothers, and Davis, et al.^[22], Huskova, et al.^[23] and Gombay, et al.^[24], considering autoregressive models.

The existing methods have several limitations. 1) Large-sample properties have been deduced in most of previous works, while few developments are made on small-sample properties. Thus, these methods usually fail when sample that could be obtained or observed is very small. 2) The computations for existing methods are very complicated and time-consuming. 3) In the case that change-point occurs in the lower and upper tails of the sequence, most of the approaches become ineffective. 4) Even in the common conditions without change-point in the tails, existing methods can only give an estimated interval where the change-point may be located. The limitations mentioned above hinder the applications of these change-point methods in practice.

To settle these problems, we first propose a quantile test of single τ based on the theory of saddlepoint approximations. Further, a “composite quantile test” is constructed to collect the information on each τ of the sequence and give the probability of each location to be a change-point. This detection scheme could effectively solve the problems before, and has several good properties. 1) No assumptions are made about the functional form of the sequence distribution of independent random variables $X_1, X_2, \dots, X_n$. 2) The computation of the “composite quantile test” is quick and simple. In addition, 3) the detection scheme is very sensitive to the varying information of the sequence. Thus, not only can it give the probable interval to be a change-point, but also can sensitively pinpoint the location. 4) This approach can solve the change-point problem with much smaller sample size than other existing methods. 5) The scheme works well when the change-point occurs in the tails of the sequence. 6) Multiple change-point problems can also be successfully detected.

“Mean shift” problem is an important issue in studies of no matter change-point in distribution or in regression function. Pre-shift and post-shift means are to be estimated concurrently with the change point. Many former studies focus on this problem, like Kokoszka and Leipus^[26], Ferger^[27] and Shi, et al.^[28, 29]. This paper presents the performances of “composite quantile test” in the “mean shift” problem as a representative attempt. Other change-point models are possible to be derived based on this idea.

Suppose $X_1, X_2, \dots, X_m$ are independent with mean μ , and $X_{m+1}, X_{m+2}, \dots, X_n$ are independent with mean $\mu + \delta$. These two sequences are also independent with each other, where m is an unknown positive integer. m is the unknown location of a change point X_m . Mean change-point model is formulated as:

$$X_i = \begin{cases} Z_i, & \text{if } i \leq m, \\ Z_i + \delta, & \text{if } i > m, \end{cases} \quad (1.1)$$

where $\{Z_i, i = 1, 2, \dots, n\}$ is a sequence of random variables with constant mean μ . δ is the unknown amount of change in mean. If $\delta = 0$, there is no change-point at the point of m , otherwise, change-point exists. Under the null hypothesis $X_1, X_2, \dots, X_n$ is an i.i.d. sequence, whereas under the alternative two sequences exist with $\delta < 0$, indicating the mean of $X_1, X_2, \dots, X_m$ greater than the mean of $X_{m+1}, X_{m+2}, \dots, X_n$. In practice, this problem is usually considered as the detection that whether the quality of articles decline^[7, 30], and so on. Similarly, $\delta > 0$ is meant that the mean of quantity or quality increases by δ . The objective is to determine whether $\delta \neq 0$ at m and pinpoint the location of change-point.

2 Quantile Test for Single Quantile

The hypothesis test used here is

- Null hypothesis H_0 : $\delta = 0$, that is, mean shift does not happened in the location of m , $m = 2, 3, \dots, n - 1$, denoted as $\eta \in \Theta_0$;
- Alternative hypothesis H_1 : $\delta < 0$, and change-point occurs and the mean of the sequence decreases, denoted as $\eta \in \Theta_0^c$.

For the case of mean of a sequence increasing, it's to inverse the order to be the alternative hypothesis above. For the null hypothesis of "no-change" H_0 , the test statistic T_τ is constructed as the number of observations among $X_1, X_2, \dots, X_m$ which are greater than the τ quantile of the whole sequence $X_1, X_2, \dots, X_n$, that is,

$$T_\tau = \sum_{i=1}^m I(X_i > Q_\tau(X)), \quad m = 2, 3, \dots, n - 1,$$

where $I(\cdot)$ is the indicative function, which is 1 for $X_i > Q_\tau(X)$ and 0 for $X_i \leq Q_\tau(X)$. $Q_\tau(X)$ is the τ quantile of the whole sequence.

Remark 1 In practice, we don't know the location of the change point. However, we can do it step by step. For more details about homogeneous testing, we refer to Tian and Chan^[50, 51] and the references therein. Specifically,

1. Initialization

As a starting point, consider the initial state in which $m = 2$. To this end, the data set could be divided into two parts: $\{X_1, X_2\}$ and $\{X_3, X_4, \dots, X_n\}$. Calculate the overall type I error, etc.

2. Enlarging homogenous region

Based on the previous step, we consider $m = 3$. Consequently, a test for the homogeneity hypothesis (i.e., no-change point) can be conducted. For this case, the data set is accordingly divided into two parts: $\{X_1, X_2, X_3\}$, and $\{X_4, X_5, \dots, X_n\}$. Then calculating the overall type I error based on the two updated parts will be done.

3. Iteration

Following the previous two steps, repeat, till $m = n$.

In order to speed up the procedure, we may try to use, for example, $m = 2k$, $k = 1, 2, \dots$ instead of $m = 2, 3, \dots$. Of course, there many other ways to deal with such problem, such as the inverse procedure, in which the whole data set is split into two parts for the first step. The sample size of each are not necessarily equal. Then testing for the homogeneity hypothesis (i.e., no-change point) for each parts could be conducted. Repeat till all change-points are detected out.

The distribution of T_τ is mainly considered. The cumulative density distribution of T_τ is

$$F_{T_\tau}(t) = P(T_\tau \leq t). \quad (2.1)$$

Proposition 1 Assume that the sequence $X_1, X_2, \dots, X_n$ are i.i.d. with a positive definite variance $\sigma^2(X)$, and $X \in \mathbb{R}$. The probability space is $(\Omega, \mathcal{A}, P)$. Applying saddlepoint approximation method, the first-order saddlepoint approximation of formula (2.1) is given by

$$\begin{aligned} & P(T_\tau \leq t) \\ & \approx 1 - \{n - \tau(n+1)\}^{n-\tau(n+1)+\frac{1}{2}} \{\tau(n+1)\}^{\tau(n+1)+\frac{1}{2}} (n-m)^{n-m+\frac{1}{2}} m^{m+\frac{1}{2}} \\ & \quad \times \{(t+1)(\tau(n+1)+t-m+1)\} \mathcal{B}_0(\hat{\lambda}) \times \left\{ n^{n+\frac{1}{2}} \{n-\tau(n+1)-t-1\}^{n-\tau(n+1)-t-\frac{1}{2}} \right. \\ & \quad \left. \{\tau(n+1)+t-m+1\}^{\tau(n+1)+t-m+\frac{3}{2}} (t+1)^{t+\frac{3}{2}} (m-t-1)^{m-t-\frac{1}{2}} \right. \\ & \quad \left. \{n(t+1)-nm+\tau(n+1)m\} \right\}^{-1}, \end{aligned} \quad (2.2)$$

where

$$\begin{aligned} \hat{\lambda} &= \log \left(\frac{(t+1)(\tau(n+1)+t-m+1)}{(m-t-1)(n-\tau(n+1)-t-1)} \right) \\ & \quad \times \left\{ \frac{(t+1)(n-\tau(n+1)-t-1)(m-t-1)(\tau(n+1)+t-m+1)}{(t+1)(n-\tau(n+1)-t-1)\tau(n+1)+(m-t-1)} \right\}^{1/2} \\ & \quad \times \{(\tau(n+1)+t-m+1)(n-\tau(n+1))\}^{-1/2}, \end{aligned}$$

and $\mathcal{B}_0(\hat{\lambda})$ is Esscher function, given by

$$\mathcal{B}_0(\hat{\lambda}) = \hat{\lambda} \exp \left(\frac{\hat{\lambda}^2}{2} \right) [1 - \mathfrak{N}(\hat{\lambda})],$$

where $\mathfrak{N}(x)$ is standard normal distribution function,

$$\mathfrak{N}(y) = \int_{-\infty}^y \mathfrak{n}(s) ds = \int_{-\infty}^y \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{s^2}{2} \right) ds.$$

When $t \geq \frac{m[n-\tau(n+1)]}{n}$, the formula (2.2) is valid. In order to compute the cumulative probability density, $P(T_\tau \leq t)$, let $m \times \tau \geq \frac{2n}{n+1}$.

The procedure of how Proposition 1 is derived is presented in the following. Let $\varpi(T_\tau)$ be the statistics relative to T_τ , which helps approximate the distribution of T_τ . The Laplace transform function of $\varpi(T_\tau)$ is defined as

$$\varphi(\theta) = E \exp(\theta \cdot \varpi(T_\tau)) = \sum_{\varpi(T_\tau)} \exp(\theta \cdot \varpi(T_\tau)) P(\varpi(T_\tau)).$$

A cumulant transform function is defined as

$$\kappa(\theta) = \log \varphi(\theta),$$

and a characteristic function is defined as

$$\phi(t) = E \exp(it \cdot \varpi(T_\tau)),$$

where $\theta \in \mathbb{R}$. The domain of Laplace transform is $\Theta_\theta = \{\theta \in \mathbb{R} : \varphi(\theta) < \infty\}$. The positive definite variance of $X_1, X_2, \dots, X_n$ enables the cumulant transform $\kappa(\theta)$ strictly convex on Θ_θ ^[25].

Applications of the saddlepoint to approximate the probability density of a statistic have much more development. Jensen^[32] proposed many methods and ideas to produce a very accurate approximation and we use one of them in our deviation. Let $\bar{\varpi}$ be the mean of $\varpi(T_\tau)$. The saddlepoint approximation to the density of $\bar{\varpi}$ is

$$f_{\bar{\varpi}}(w) \approx \frac{\varphi(\hat{\theta}) \exp(-n\hat{\theta} \cdot w) n^0}{(2\pi \Sigma(\hat{\theta}))^{1/2}} \left\{ 1 + \frac{1}{n} \left[\frac{\kappa_4(\hat{\theta})}{8\Sigma^2(\hat{\theta})} - \frac{5\kappa_3(\hat{\theta})^2}{24\Sigma^3(\hat{\theta})} \right] \right\}, \quad (2.3)$$

where $\kappa_3(\cdot)$ and $\kappa_4(\cdot)$ are the third and forth deviations of $\kappa(\cdot)$, $\Sigma(\hat{\theta}) = \sigma^2(\hat{\theta})$, and $\hat{\theta}$ is called the saddlepoint satisfying the saddlepoint equation,

$$\kappa'(\hat{\theta}) = w.$$

The equation is also available in (1.5) of Butler^[25], and the density approximation formula (2.3) was derived from Jensen^[32].

For continuous random variables, an saddlepoint approximation to the cumulative density function(CDF) of $\bar{\varpi}$ is equally straight forward,

$$P(\bar{\varpi} \geq w) \approx \frac{\varphi(\hat{\theta})^n \exp(-n\hat{\theta} \cdot w)}{\sqrt{n}|\hat{\theta}|\sigma(\hat{\theta})} \left\{ \mathcal{B}_0(\lambda) + \frac{\text{sgn}(\hat{\theta})\kappa_3(\hat{\theta})}{6\sqrt{n}\Sigma^{3/2}(\hat{\theta})} \mathcal{B}_3(\lambda) \right. \\ \left. + \frac{1}{n} \left[\frac{\kappa_4(\hat{\theta})}{24\Sigma^2(\hat{\theta})} \mathcal{B}_4(\lambda) + \frac{\kappa_3(\hat{\theta})^2}{72\Sigma^3(\hat{\theta})\mathcal{B}_6(\lambda)} \right] \right\}, \quad (2.4)$$

where

$$\begin{aligned} \Sigma(\hat{\theta}) &= \sigma^2(\hat{\theta}) = \kappa''(\hat{\theta}), \\ \lambda &= \sqrt{n}|\hat{\theta}|\sigma(\hat{\theta}), \\ \mathcal{B}_3 &= -\{\lambda^3 \mathcal{B}_0(\lambda) - (\lambda^3 - \lambda)(2\pi)^{-1/2}\}, \\ \mathcal{B}_4 &= \lambda^4 \mathcal{B}_0(\lambda) - (\lambda^4 - \lambda^2)(2\pi)^{-1/2}, \\ \mathcal{B}_6 &= \lambda^6 \mathcal{B}_0(\lambda) - (\lambda^6 - \lambda^4 + 3\lambda^2)(2\pi)^{-1/2}. \end{aligned}$$

In this case, the saddlepoint $\hat{\theta}$ is obtained by solving $\kappa'(\hat{\theta}) = \iota(w)$, i.e., $\epsilon = 0$.

For lattice random variables, use the expanded formula similar to (2.4)

$$\begin{aligned} P(\bar{\varpi} \geq w) &\approx \frac{\varphi(\hat{\theta})^n \exp(-n\hat{\theta} \cdot w)}{\sqrt{n}\sigma(\hat{\theta})\{1 - \exp(-|\hat{\theta}|)\}} \\ &\times \left\{ \mathcal{B}_0(\lambda) + (\sqrt{n})^{-1} \left[(\sigma(\hat{\theta}))^{-1} \left((|\hat{\theta}|)^{-1} - \frac{\exp(-|\hat{\theta}|)}{1 - \exp(-|\hat{\theta}|)} \right) \mathcal{B}_1(\lambda) \right. \right. \\ &\left. \left. + \frac{\kappa_3(\hat{\theta})}{6\Sigma^{3/2}(\hat{\theta})} \text{sgn}(\hat{\theta}) \mathcal{B}_3(\lambda) \right] + \frac{1}{n} \left[\left(\frac{\exp(-|\hat{\theta}|)}{1 - \exp(-|\hat{\theta}|)} \right) \left(\frac{1}{2} - (|\hat{\theta}|)^{-1} \right) \right] \right\} \end{aligned}$$

$$\begin{aligned}
 & + \left[\frac{\exp(-|\hat{\theta}|)}{1 - \exp(-|\hat{\theta}|)} \right]^2 \frac{\mathcal{B}_2(\lambda)}{\Sigma(\hat{\theta})} + \frac{1}{\sigma(\hat{\theta})} \left((|\hat{\theta}|)^{-1} - \frac{\exp(-|\hat{\theta}|)}{1 - \exp(-|\hat{\theta}|)} \right) \\
 & \times \left[\frac{\kappa_3(\hat{\theta})}{6\Sigma^{3/2}(\hat{\theta})} \text{sgn}(\hat{\theta}) \mathcal{B}_4(\lambda) + \frac{\kappa_4(\hat{\theta})}{24\Sigma^2(\hat{\theta})} \mathcal{B}_4(\lambda) + \frac{\kappa_3(\hat{\theta})^2}{72\Sigma^3(\hat{\theta})} \mathcal{B}_6(\lambda) \right] \Bigg\}, \quad (2.5)
 \end{aligned}$$

where

$$\begin{aligned}
 \mathcal{B}_1 &= -\lambda \{ \mathcal{B}_0(\lambda) - (2\pi)^{-1/2} \}, \\
 \mathcal{B}_2 &= \lambda^2 \{ \mathcal{B}_0(\lambda) - (2\pi)^{-1/2} \}.
 \end{aligned}$$

The saddlepoint $\hat{\theta}$ in this case is obtained by solving $\kappa'(\hat{\theta}) = \iota(w) - \epsilon/2$, i.e., $\epsilon \neq 0$. Without loss of generality, let $\epsilon = 1/n$.

The distribution of T_τ can be represented as a conditional distribution. Let $\vartheta_i = 0$ for $i \leq \tau(n+1)$ and $\vartheta_i = 1$ for $i > \tau(n+1)$. The cumulative density function of T_τ is

$$F_{T_\tau}(t) = P \left(\sum_1^n \vartheta_i U_i \leq t \mid \sum_1^n U_i = m \right), \quad t = 0, 1, \dots, m, \quad (2.6)$$

where $U_1, U_2, \dots, U_n$ are i.i.d. with $P(U_i = 1) = P(U_i = 0) = \frac{1}{2}$. Therefore, the statistic $\varpi(T_\tau) = (\sum_1^n \vartheta_i U_i, \sum_1^n U_i)$.

In order to approximate the above formula in which $\theta \in \mathbb{R}^2$, it is suitable for formula (2.5) to be expanded to the formula (2.2.15) in Jensen^[32]

$$\begin{aligned}
 & P(\bar{\varpi}_1(T_\tau) \geq w_1 \mid \bar{\varpi}_2(T_\tau) = w_2) \\
 & = \exp \left(-\frac{r^2}{2} \right) \left(\frac{|\Sigma_{22}(\tilde{\theta})|}{n|\Sigma(\tilde{\theta})|} \right)^{1/2} \frac{1}{|\hat{\theta}_1|} \left\{ \mathcal{B}_0(\hat{\lambda}) + O \left(\frac{B_0(\hat{\lambda})}{1 + \hat{\lambda}} n^{-1/2} \right) \right\}, \quad (2.7)
 \end{aligned}$$

where $r^2 = 2n \{ (\hat{\theta} - \tilde{\theta}) \cdot s - \kappa(\hat{\theta}) + \kappa(\tilde{\theta}) \}$ and ϖ_1 and ϖ_2 are defined as in the Appendix. The procedure of Proposition 1 is given. And details of the deviation from formula (2.6) to formula (2.2) for our case are in the Appendix.

According to the Proposition 1, the probability distribution of T_τ can be obtained via $P(T_\tau = t) = P(T_\tau \leq t) - P(T_\tau \leq t-1)$, which is denoted as $f_{T_\tau}(t)$ for $t = 0, 1, \dots, m$.

If change-point does not exist at the location m of the sequence, the distribution of T_τ is exactly approximated by saddlepoint approximation method, the formula (2.2). In order to exactly compute its value, formula (2.2) is transformed to

$$\begin{aligned}
 & P(T_\tau \leq t) \\
 & \approx 1 - \left\{ \frac{n - \tau(n+1)}{n - \tau(n+1) - t - 1} \right\}^{n - \tau(n+1) - t - \frac{1}{2}} \times \left\{ \frac{\tau(n+1)}{\tau(n+1) + t - m + 1} \right\}^{\tau(n+1) + \frac{1}{2}} \\
 & \times \left\{ \frac{n - \tau(n+1)}{\tau(n+1) + t - m + 1} \right\}^{t+1} \times \left\{ \frac{n - m}{n} \right\}^{n + \frac{1}{2}} \\
 & \times \left\{ \frac{\tau(n+1) + t - m + 1}{n - m} \right\}^m \times \left\{ \frac{m}{m - t - 1} \right\}^{m + \frac{1}{2}} \times \left\{ \frac{m - t - 1}{t + 1} \right\}^{t+1} \\
 & \times (t+1)^{-\frac{1}{2}} \times \frac{\{(t+1)[\tau(n+1) + t - m + 1] \mathcal{B}_0(\hat{\lambda})\}}{n(t+1) - nm + \tau(n+1)m}. \quad (2.8)
 \end{aligned}$$

The distribution of T_τ approximated by saddlepoint approximation is under the null hypothesis of “no-change”.

For each location m , the significant level is set to be $\alpha = 0.05$. The criterion for hypothesis test is the value of T_τ , named as $t_0(\tau)$ when the accumulative density distribution of T_τ equals 0.95. That is

$$t_0(\tau) = \arg \left\{ \sum_{t=0}^{t_0(\tau)} f_{T_\tau}(t) = 0.95 \right\}.$$

Here, $t_0(\tau)$ is not an integer necessarily, and

$$\sum_{t=0}^{\text{int}(t_0(\tau))} f_{T_\tau}(t) \leq 0.95 \leq \sum_{t=0}^{\text{int}(t_0(\tau))+1} f_{T_\tau}(t),$$

where $\text{int}(t)$ denotes the largest integer smaller than t . The plots of T_τ when $m = 10$, $m = 50$ and $m = 90$ are presented in Figure 1. The curve of T_τ intersects 0.95 at $t_0(\tau)$. When k number of τ are considered, for each location $m = 4, 5, \dots, n-1$, the threshold matrix of $t_0(\tau)$ by $\{\tau_j, j = 1, 2, \dots, k\}$ are plotted in Figure 2.

Figure 1 F_{T_τ} : (a) describes F_{T_τ} when $m = 10$ which intersects line 0.95 at $t_0(\tau)$; (b) describes F_{T_τ} when $m = 50$ which intersects line 0.95 at $t_0(\tau)$; (c) describes F_{T_τ} when $m = 90$ which intersects line 0.95 at $t_0(\tau)$

Figure 2 The matrix of $t_0(\tau)$: (a) describes the surface of $t_0(\tau)$ in the large sample case ($n = 100$); (b) is the surface of $t_0(\tau)$ in the small sample case ($n = 20$)

When “mean shift” happens, the distribution of T_τ of data would largely deviate from the distribution under the null hypothesis. If m is given, the T_τ for target sequence is calculated as $t_1(\tau)$ and compared with $t_0(\tau)$. If $t_1(\tau) > t_0(\tau)$, the null hypothesis is rejected and change point exists at m location of the sequence $X_1, X_2, \dots, X_n$. If $t_1(\tau) < t_0(\tau)$ instead, null hypothesis cannot be rejected and the change-point cannot be determined. By shifting the value of m from 2 to $n - 1$, we can detect whether change point exists in the target sequence and pinpoint the location. Therefore, a quantile test that rejects H_0 if $t_{1_X}(\tau) > t_0(\tau)$, is a test with rejection region $R, \{t_X : t_{1_X}(\tau) > t_0(\tau)\}$.

To compute the power of hypothesis test, under the alternative hypothesis, the distribution of T_τ is obtained by applying the bootstrap method^[33]. The basic steps proceed as follows. Assuming that the random sample of size n is drawn from an unspecified probability distribution, Ξ .

- Step 1. Placing a probability of $1/n$ at each point, $x_1, x_2, \dots, x_n$. Each sample's element has the same probability to be drawn.
- Step 2. From the sample $x_1, x_2, \dots, x_n$, draw a random sample of size d with replacement.
- Step 3. Calculate the statistic of interest, T_τ . For resample, yield $T_\tau^* = T_\tau \times \frac{n}{d}$.
- Step 4. Repeat Steps 2 and 3 N ($N > 1000$) times.
- Step 5. Construct the relative frequency through the N number of T_τ^* by placing a probability of $1/N$ at each point, $T_{\tau 1}^*, T_{\tau 2}^*, \dots, T_{\tau N}^*$. The distribution obtained is the bootstrapped estimate of the sampling distribution of T_τ .

The bootstrapped sampling distribution of T_τ is denoted as $f_{T_\tau}^{\text{boot}}(t)$. The power of hypothesis test, denoted as $\beta_\tau(\eta)$, can be calculated by

$$\beta_\tau(\eta) = \begin{cases} \sum_{t=\lceil \text{int}(t_0(\tau))+1 \rceil}^m f_{T_\tau}(t), & \text{if } \eta \in \Theta_0, \\ 1 - \sum_{t=0}^{\text{int}(t_0(\tau))} f_{T_\tau}^{\text{boot}}(t), & \text{if } \eta \in \Theta_0^c, \end{cases} \quad (2.9)$$

where, $\sum_{t=\lceil \text{int}(t_0(\tau))+1 \rceil}^m f_{T_\tau}(t)$ is the probability of Type I Error the test makes. $\sum_{t=0}^{\text{int}(t_0(\tau))} f_{T_\tau}^{\text{boot}}(t)$ is the probability of Type II Error. A good hypothesis test has power near 1 for most $\eta \in \Theta_0^c$ and near 0 for most $\eta \in \Theta_0$. Due to

$$\sup_{\eta \in \Theta_0} \beta_\tau(\eta) = \alpha,$$

quantile test is a size α test.

3 Composite Quantile Test

Instead of adopting one τ quantile (median or other τ) for change-point detection, which will drop much of the valuable information, we try to collect information from several τ of the sequence $X_1, X_2, \dots, X_n$. A complete detection scheme is constructed to compute the probability of each location m to be a change point. The information across k uniformly spaced

quantiles $\{\tau_j, j = 1, 2, \dots, k\}$ is aggregated to maximise the efficiency of making inference, where

$$\tau_j = j/(k+1), \quad j = 1, 2, \dots, k.$$

The probability of one location m to be a change-point is calculated by

$$P(m) = \frac{\Lambda_m}{\sum_{m'=2}^{n-1} \Lambda_{m'}}, \quad (3.1)$$

where

$$\Lambda_m = \Delta_m - \min\{\Delta_{m'}, m' \in \{2, 3, \dots, n-1\}\},$$

and

$$\Delta_m = \sum_{j=1}^k \text{sgn}(t_1(\tau_j) - t_0(\tau_j)) \times (t_1(\tau_j) - t_0(\tau_j))^2,$$

where

$$\text{sgn}(t_1(\tau_j) - t_0(\tau_j)) = \begin{cases} 1, & \text{if } t_1(\tau_j) \geq t_0(\tau_j), \\ -1, & \text{if } t_1(\tau_j) < t_0(\tau_j), \end{cases}$$

$t_0(\tau_j)$ is the theoretical criterion for hypothesis test at significant level 95% when m and τ_j are given. $t_1(\tau_j)$ denotes the number of points in series $x_1, x_2, \dots, x_m$ which are larger than τ_j quantile of the whole sequence $x_1, x_2, \dots, x_n$. The distance between $t_1(\tau_j)$ and $t_0(\tau_j)$ indicates the deviation of the distribution of proposed statistics in the target sequence from theoretical criterion in the null hypothesis. Euclidean distance is used here.

The proposed “composite quantile regression” can be considered as a special procedure to combine the results of each single quantile test together by a varying weight. For given j , the distance between $t_1(\tau_j)$ and $t_0(\tau_j)$ determines the weight of the information from τ_j quantile test. In this regard, “composite quantile regression” is a series of hypothesis tests to the greatest extent collecting the valuable information from the whole sequence, instead of limited data points. This is a more sensitive way to detect the change-point location without restricting the forms of the sequence distribution.

4 Simulation

Monte Carlo simulation is applied to present the performances of the proposed single quantile test and composite quantile test. Three other detecting methods—single tests, CUSUM^[5, 34], Student-t statistic^[35, 36] and Mann-Whitney statistics^[37, 38] are adopted as benchmarks. Three scenarios with different degree of “mean shift” are considered in both large and small sample cases. Multiple change-points situation is also tested.

4.1 Single Quantile Test and Power

We analyze the situations of large size sample $n = 100$ and small size sample $n = 20$. Given the simulated sequences with a change-point located at $m = 4, 5, \dots, n-1$, the rate of successful detection in the Monte Carlo simulation is defined as $S(\tau) = \frac{N^*}{N}$, where N^* is the number of successful detection, and N ($N = 10^4$) denotes the times of Monte Carlo experiments. The generated sequence X_t for $t = 1, 2, \dots, m$ is distributed by $Norm(\mu_1, \sigma)$ and sequence X_t for

$t = m + 1, \dots, n$ is distributed by $Norm(\mu_2, \sigma)$, where $\mu_1 = \mu_2 + \delta \cdot \sigma$. In the previous empirical studies, the mean shift is usually larger than one σ . Thus, δ is set as -1 , -1.5 and -2 in the scenarios respectively indicating the degree of “mean shift”. For simplicity, $\sigma = 1$. The results of $S(\tau)$ are presented in Figure 3. Most of location m show great success rate $S(\tau)$ except for extreme conditions of $m \geq n - 3$. This is a very good performance in comparison with studies before. Quantile test possesses the advantage that it can find the location of change point exactly.

Figure 3 $S(\tau)$: $S(\tau)$ in cases of $n = 100$ and $n = 20$ with $\delta = -1, -1.5, -2$, respectively

Sequences without change-points are generated to test the probability of Type I Error. The probability of Type I Error is defined as $W(\tau) = \frac{\hat{N}}{N}$, where $\hat{N}$ is the times of Monte Carlo experiments which make Type I Error. As reported in Figure 4, some large $W(\tau)$ are at the lower tail of sequence when small τ is adopted. At other location $W(\tau)$ are close to zero. Even in the extreme case of small sample ($n = 20$), for more than half of locations, $W(\tau)$ still show very small values. However, small quantile test fail when sample size is extreme small. For large sample size, the Type I Errors of CUSUM, Student-t and Mann-Whitney are 0.9185, 0.0522 and 0.0556, respectively. The corresponding values are 0.9908, 0.0482 and 0.051 for extreme small sample size. CUSUM fails, while Student-t and Mann-Whitney still perform well. The Type I Errors of Student-t and Mann-Whitney are close to the means of Type I Errors of 0.5, 0.6 and 0.7 quantiles obtained from Figure 4, which are 0.0580, 0.0549 and 0.0538 for large sample size, and 0.0639, 0.490 and 0.0573 for extreme small size.

Figure 4 $W(\tau)$: $W(\tau)$ in cases of $n = 100$ and $n = 20$

The power of quantile test is defined as Formula (2.9). Bootstrap resampling is introduced with repeating times $N = 10^4$. Figure 5 presents the results of β_τ in three scenarios $\delta = -1, -1.5, -2$ with $n = 100$ and $n = 20$. The mean of β_τ is reported in Table 1. β_τ , increasing with higher degree of “mean shift”, shows great performance of the quantile test by single τ . Figure 6 shows the powers of CUSUM, Student-t, Mann-Whitney, and single quantile tests using 0.5, 0.6 and 0.7 quantiles. In the simulation, the sequences are generated with one change point at location m . Single quantile test can exactly find this location, while CUSUM, Student-t and Mann-Whitney can only detect the existence of change point but cannot always give the exact location. In the calculation of power for them in Figure 6, we do not consider whether they find the exact location. The high power of CUSUM is accompanied by extremely high Type I Error. Therefore, it makes little sense to compare the power of CUSUM with other methods. The powers of Student-t and Mann-Whitney are close. For both large and small sample size, the power of single quantile tests is almost always higher than these two methods. Single quantile tests are more powerful than other methods when change point existing in the first half part of sequence. As showed in Table 2, the powers of Student-t and Mann-Whitney are very weak when we consider whether they find the exact location of change point. The power of CUSUM can reach 67 percent and the powers of Student-t and Mann-Whitney are

smaller than 20 percent. They are far smaller than that of single quantile test as showed in Table 1.

Figure 5 β_τ : β_τ in cases of $n = 100$ and $n = 20$ under three scenarios $\delta = -1, -1.5, -2$, respectively

Figure 6 Power comparison of CUSUM, Student-t, Mann-Whitney and single quantile tests in cases of $n = 100$ and $n = 20$ under three scenarios $\delta = -1, -1.5, -2$, respectively

Table 1 $\bar{\beta}_\tau$ in three scenarios for both $n = 100$ and $n = 20$

δ	-1	-1.5	-2
$n = 100$	0.9088	0.9763	0.9895
$n = 20$	0.7941	0.9575	0.9577

Table 2 Summary of statistics about the powers of three methods under the requirement of finding exact location of change point

$n = 100$	$\delta = -1$	Max	Min	Median	Mean
	CUSUM	0.2983	0	0.2245	0.2006
	Student-t	0.1113	0	0.1014	0.0895
	Mann-Whitney	0.1160	0	0.1013	0.0884
	$\delta = -1.5$	Max	Min	Median	Mean
	CUSUM	0.4908	0	0.3737	0.3353
	Student-t	0.1319	0	0.1246	0.1174
	Mann-Whitney	0.1496	0	0.1218	0.1143
	$\delta = -2$	Max	Min	Median	Mean
	CUSUM	0.6538	0.0087	0.5085	0.4559
	Student-t	0.1205	0	0.1109	0.1090
	Mann-Whitney	0.1616	0	0.1129	0.1071
$n = 20$	$\delta = -1$	Max	Min	Median	Mean
	CUSUM	0.3281	0	0.2783	0.2468
	Student-t	0.0451	0	0.0374	0.0325
	Mann-Whitney	0.0526	0	0.0414	0.0337
	$\delta = -1.5$	Max	Min	Median	Mean
	CUSUM	0.5153	0	0.4413	0.3906
	Student-t	0.0906	0	0.0776	0.0667
	Mann-Whitney	0.1051	0	0.0793	0.0655
	$\delta = -2$	Max	Min	Median	Mean
	CUSUM	0.6726	0.2430	0.5747	0.5373
	Student-t	0.1079	0	0.0948	0.0869
	Mann-Whitney	0.1365	0	0.0941	0.0845

4.2 Composite Quantile Test and Multiple Change-Points Situation

Based on quantile test of single τ , “composite quantile test” is constructed as Formula (3.1). In this part, single change-point is set at location $\hat{m}$ of the generated sequence. $\hat{m}$ shifts from 4 to $n - 1$. For each given $\hat{m}$, the probability $P(m)$ of each location m to be a change-point is calculated by collecting the information of each single quantile test. The results are presented in Figure 7 both for $n = 100$ and $n = 20$. A series of peaks can be observed on the diagonal of

the surface, where the simulated change-points are actually located. That means, by applying “composite quantile test”, the highest probabilities pinpoint the actual change-point locations. Similar patterns can also be observed for small sample cases. The ridge-like surfaces of Figure 7 reflect the sensitive performance of “composite quantile test”. Even in the extreme conditions that generated change-point occurs in the tails of the sequence, the proposed method still responds sensitively and gives the highest probability to the right locations.

Figure 7 $P(m)$: $P(m)$ in cases of $n = 100$ and $n = 20$ under three scenarios $\delta = -1, -1.5, -2$, respectively

Multiple change-points cases are further investigated to confirm the advantage of the proposed method. For two change-points condition, the sequences which have two change-points are generated. The two change-points at first are on the two tails of the sequence, one on the lower and the other on the upper. The two points approach to each other and eventually join together. The degree of “mean shift” is set as $\delta = -1$. The length of sequence is set as $n = 150$. It is the same with three change-points condition, except for a third change-point which is constantly in the middle of a $n = 200$ length sequence. The results of two and three change-points situations are plotted in Figure 8.

Two ridges are observed in Figure 7(a) and (b), exactly positioning the locations of change-points. Similar patterns can be found for three change-points situation, except that the change-point in the middle is more emphasized by the method when two other points are on the tails.

Multiple change-points situation indicates that the “composite quantile test” works robustly in dealing with unfavorable data conditions.

Figure 8 $P(m)$: $P(m)$ for two and three change-point $\delta = -1$

5 Real Data

Two sets of data are analyzed. The first one is a much analysed^[39–43] data set of intervals between British coal-mining disasters which resulted in 10 or more fatalities during the 112-year period 1851–1962, which was gathered by Maguire et al.^[44], extended and corrected by Jarrett^[45]. And Carlin, et al.^[46, 47] converted Jarrett’s table into yearly intervals. The second data is a 60-year (1912–1971) annual series of temperatures from Libby, Montana. The documented change point is 1938 (resulting from a change in latitude and elevation)^[48]. The plots of these two time series are showed in Figure 9(a) and (b). As presented in Figure 9(c) and (d), the probability of each location to be a change point is computed by “composite quantile test”.

The aim is to find a location m^* cutting off the sequence making $F_1(x) \neq F_2(x)$. $F_1(x)$ and $F_2(x)$ are common distribution function for $X_t, t = 1, 2, \dots, m$ and $X_t, t = m + 1, m + 2, \dots, n$, respectively. For the first data, two peaks with very near locations are identified at $m = 36$ and

$m = 39$. The probabilities of these two points are also very close. The standard deviation is $\sigma_L = 1.6475$, which is the empirical basis to choose the value of δ in Monte Carlo experiments. In detail, at location $m = 36$, $\bar{\mu}_1^{36} = 3.250$ is the mean of the sequence $X_1, X_2, \dots, X_m$ and $\bar{\mu}_2^{36} = 0.9737$ is the mean of the sequence $X_{m+1}, X_{m+2}, \dots, X_n$. Thus the mean difference is $\bar{\mu}_1^{36} - \bar{\mu}_2^{36} = 1.382\sigma_L$. At location $m = 39$, on the other hand, $\bar{\mu}_1^{39} = 3.1538$ and $\bar{\mu}_2^{39} = 0.9315$. And $\bar{\mu}_1^{39} - \bar{\mu}_2^{39} = 1.349\sigma_L$. It should be noticed that the mean difference at m does not necessarily lead to a high $P(m)$. Change-point m^* makes the sample points before and after m^* coming from different distributions. According to the result of composite quantile test, two special points are pinpointed and more likely to be change-points. In previous studies, a probable interval is usually given without exact location to be a change-point. For example, Carlin, et al.^[49] gave the probable interval of change-point 1889–1892, which is $m = 39$ to $m = 42$ in our context. The point $m = 36$ is not detected and the interval $m = 39$ to $m = 42$ is also too wide to be accurate. Student-t detect only one change point at $m = 36$ and also Mann-Whitney detect a change point at $m = 41$. In this regard, “composite quantile test” responds more sensitively to change “tendency” of the sequence than methods before.

Figure 9 Real data: (a) and (b) describe the time series of dataset 1 and 2. (c) and (d) describe $P(m)$, the probability of being change-point

In the second data, one peak is observed at $m = 19$, which is the year of 1930. The standard

deviation is $\sigma_S = 0.8769$. The mean of points before this peak is $\bar{\mu}_1 = 6.5626$ and the mean of points after that point is $\bar{\mu}_2 = 7.3521$. The mean difference is $\bar{\mu}_2 - \bar{\mu}_1 = 0.9\sigma_S$. The proposed method senses a tendency that the distribution of sample is going to change. Student-t and Mann-Whitney detect the same change point location. This data is also studies by Reeves, et al^[48]. They applied eight methods, where the standard normal homogeneity (SNH) test and nonparametric SNH procedure detect a change point in 1930, Sawa's Bayes criteria, two-phase regression with a common trend and modified vincent's method detect a change point in 1947, and Akaike's information criteria, two-phase regression (LR) and a modified LR method detect a change point in 1945. Our method detect the same change point location as SNH test and nonparametric SNH procedure. This location differs from the documented change point. This discrepancy maybe caused by other undocumented change points in the series or changed information happening before the location of true change point.

6 Conclusion

"Mean shift" problem in change-point studies is discussed in this paper. Saddlepoint approximation method is applied to approximate the cumulative density function of statistic T_τ for null hypothesis of "no-change", and a quantile test of single τ is proposed. In order to capture the overall tendency information of the sequence, we construct "composite quantile test" to calculate the probability of each location to be a change-point. As showed in Monte Carlo experiments and real data analysis, the proposed detection scheme performs better than existing methods. The types of functional form of sequence distribution are not restrained. The sample sizes ($n = 100, 20$) considered in our context are much smaller than previous studies. In the cases that change-points at lower or upper tails of sequence, the proposed test still works well. In addition, the robustness of this scheme is confirmed by multiple change-points conditions. The ideas behind this paper are extendable to other complex issues in change-point studies. In addition, the extension of this testing to two-sided is left as a future research topic.

Acknowledgements

The work was partially supported by the major research projects of philosophy and social science of the Chinese Ministry of Education (15JZD015), National Natural Science Foundation of China (11271368), Project supported by the Major Program of Beijing Philosophy and Social Science Foundation of China (15ZDA17), Project of Ministry of Education supported by the Specialized Research Fund for the Doctoral Program of Higher Education of China (20130004110007), the Key Program of National Philosophy and Social Science Foundation Grant (13AZD064). The work was partially supported by the Major Project of Humanities Social Science Foundation of Ministry of Education (15JJD910001), the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China (15XNL008), the Project of Flying Apsaras Scholar of Lanzhou University of Finance & Economics, and the Project of Tianshan Mountain Scholar of Xinjiang University of Finance & Economics.

References

- [1] Pettitt A N. A non-parametric approach to the change-point problem. *Applied Statistics*, 1979, 28(2): 126–135.
- [2] Kim C J, Nelson C R. Has the US economy become more stable? A bayesian approach based on a Markov-switching model of the business cycle. *Review of Economics & Statistics*, 1999, 81(4): 608–616.
- [3] Solow A. Testing for climate change — An application of the two-phase regression model. *Journal of Climate & Applied Meteorology*, 1987, 26(10): 1401–1405.
- [4] Maguire R O, Sims J T. Measuring agronomic and environmental soil phosphorus saturation and predicting phosphorus leaching with mehlich 3. *Soil Science Society of America Journal*, 2002, 66(6): 2033–2039.
- [5] Page E S. Continuous inspection schemes. *Biometrika*, 1954, 41(1/2): 100–115.
- [6] Lai T L. Sequential changepoint detection in quality control and dynamical systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1995, 57(4): 613–658.
- [7] Mahmoud M A, Parker P A, Woodall W H, et al. A change point method for linear profile data. *Quality & Reliability Engineering International*, 2007, 23(2): 247–268.
- [8] Hall C B, Lipton R B, Sliwinski M, et al. A change point model for estimating the onset of cognitive decline in preclinical alzheimer’s disease. *Statistics in Medicine*, 2000, 19(11–12): 1555–1566.
- [9] Leone D, Nott D J, Dunsmuir W T M. Smoothing and change point detection for gamma ray count data. *Mathematical Geology*, 2003, 35(2): 175–194.
- [10] Mei Y. Sequential change-point detection when unknown parameters are present in the pre-change distribution. *The Annals of Statistics*, 2006, 34(1): 92–122.
- [11] Pettitt A N. A simple cumulative sum type statistic for the change-point problem with zero-one observations. *Biometrika*, 1980, 67(1): 79–84.
- [12] Hinkley D V. Inference about the change-point in a sequence of random variables. *Biometrika*, 1970, 57(1): 1–17.
- [13] Hinkley D V. Inference about the change-point from cumulative sum tests. *Biometrika*, 1971, 58(3): 509–523.
- [14] Talwar P P, Gentle J E. Detecting a scale shift in a random sequence at an unknown time point. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 1981, 30(3): 301–304.
- [15] Carlstein E. Nonparametric change-point estimation. *The Annals of Statistics*, 1988, 16(1): 188–197.
- [16] Inclin C, Tiao G C. Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association*, 1994, 89(427): 913–923.
- [17] Hawkins D M, Zamba K D. A change-point model for a shift in variance. *Journal of Quality Technology*, 2005, 37(1): 21–31.
- [18] Kawahara Y, Sugiyama M. Sequential change-point detection based on direct density-ratio estimation. *Statistical Analysis & Data Mining*, 2012, 5(2): 114–127.
- [19] Andrews D W K. Tests for parameter instability and structural change with unknown change point. *Econometrica*, 1993, 61(4): 821–856.
- [20] Muller H G. Change-points in nonparametric regression analysis. *The Annals of Statistics*, 1992, 20(2): 737–761.
- [21] Loader C R. Change point estimation using nonparametric regression. *The Annals of Statistics*, 1996, 24(4): 1667–1678.
- [22] Davis R A, Huang D, Yao Y C. Testing for a change in the parameter values and order of an autoregressive model. *The Annals of Statistics*, 1995, 23(1): 282–304.
- [23] Huskova M, Praskova Z, Steinebach J. On the detection of changes in autoregressive time series I. Asymptotics. *Journal of Statistical Planning & Inference*, 2007, 137(4): 1243–1259.
- [24] Gombay E, Serban D. Monitoring parameter change in time series models. *Journal of Multivariate Analysis*, 2009, 100(4): 715–725.
- [25] Bai J. Least squares estimation of a shift in linear processes. *Journal of Time Series Analysis*, 1994, 15(5): 453–472.
- [26] Kokoszka P, Leipus R. Change-point in the mean of dependent observations. *Statistics & Probability Letters*, 1998, 40(4): 385–393.
- [27] Ferger D. Exponential and polynomial tailbounds for change-point estimators. *Journal of Statistical Planning & Inference*, 2001, 92(1–2): 73–109.

- [28] Shi X, Wu Y, Miao B. A note on the convergence rate of the kernel density estimator of the mode. *Statistics & Probability Letter*, 2009, 79(17): 1866–1871.
- [29] Shi X, Wu Y, Miao B. Strong convergence rate of estimators of change point and its application. *Computational Statistics & Data Analysis*, 2009, 53(4): 990–998.
- [30] Samuel T R, Pignatiello Jr J J, Calvin J A. Identifying the time of a step change with X control charts. *Quality Engineering*, 1998, 10(3): 521–527.
- [31] Barndorff-Nielsen O. Hyperbolic distributions and distributions on hyperbolae. *Scandinavian Journal of Statistics*, 1978, 5(3): 151–157.
- [32] Jensen J L. Saddlepoint approximations. Clarendon Press, Oxford, 1995.
- [33] Varian H. Bootstrap tutorial. *Mathematica Journal*, 2005, 9(4): 768–775.
- [34] Csorgo M, Horvath L. Limit theorems in change-point analysis. Wiley, New York, 1997.
- [35] Hawkins D M, Qiu P, Chang W K. The changepoint model for statistical process control. *Journal of Quality Technology*, 2003, 35(4): 355–366.
- [36] Hawkins D M, Zamba K D. Statistical process control for shifts in mean or variance using a changepoint formulation. *Technometrics*, 2005, 47(2): 164–173.
- [37] Hawkins D M, Deng Q. A nonparametric change-point control chart. *Journal of Quality Technology*, 2010, 42(2): 165–173.
- [38] Ross G J, Tasoulis D K, Adams N M. Nonparametric monitoring of data streams for changes in location and scale. *Technometrics*, 2011, 53(4): 379–389.
- [39] Worsley K J. Confidence regions and tests for a change-point in a sequence of exponential family random variables. *Biometrika*, 1986, 73(1): 91–104.
- [40] Raftery A E, Akman V E. Bayesian analysis of a poisson process with a changepoint. *Biometrika*, 1986, 73(1): 85–89.
- [41] Siegmund D. Boundary crossing probabilities and statistical applications. *The Annals of Statistics*, 1986, 14(2): 361–404.
- [42] Siegmund D. Confidence sets in change-point problems. *International Statistical Review*, 1998, 56(1): 31–48.
- [43] Hand D J, Daly F, Lunn A D, et al. A handbook of small data sets, volume 1. CRC Press, US, 1994.
- [44] Maguire B A, Pearson E S, Wynn A H A. The time intervals between industrial accidents. *Biometrika*, 1952, 39(1/2): 168–180.
- [45] Jarrett R G. A note on the intervals between coal-mining disasters. *Biometrika*, 1979, 66(1): 191–193.
- [46] Carlin B P, Polson N G, Stoffer D S. A monte carlo approach to nonnormal and nonlinear state-space modelling. *Journal of the American Statistical Association*, 1992, 41(2): 493–500.
- [47] Carlin B P, Chib S. Bayesian model choice via markov chain monte carlo methods. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1995, 57(3): 473–484.
- [48] Reeves J, Chen J, Wang X L, et al. A review and comparison of changepoint detection techniques for climate data. *Journal of Applied Meteorology & Climatology*, 2007, 46(6): 900–915.
- [49] Carlin B P, Gelfand A E, Smith A F M. Hierarchical bayesian analysis of changepoint problems. *Applied Statistics*, 1992, 87(418): 389–405.
- [50] Tian M Z, Chan N H. Saddle point approximation and volatility estimation of value-at-risk. *Statistica Sinica*, 2010, 20: 1239–1256.
- [51] Tian M Z, Chan N H. Adaptive quantile regression with precise risk bounds. *Statistica Sinica*, to appear.
- [52] Butler R W. Saddlepoint approximations with applications. Cambridge University Press, UK, 2007.

Appendix

The Proof of Formula (2.2)

Let $\varpi_1(T_\tau) = \sum_1^n \vartheta_i U_i$ and $\varpi_2(T_\tau) = \sum_1^n U_i$, then the Laplace transform $\varphi(\theta_1, \theta_2)$ of $(\varpi_1(T_\tau), \varpi_2(T_\tau))$ is

$$\varphi(\theta_1, \theta_2) = \left(\frac{1 + \exp(\theta_2)}{2} \right)^{\tau(n+1)} \left(\frac{1 + \exp(\theta_1 + \theta_2)}{2} \right)^{n - \tau(n+1)}.$$

The cumulant transform $\kappa(\theta_1, \theta_2)$ is

$$\begin{aligned}\kappa(\theta_1, \theta_2) &= \log \varphi(\theta_1, \theta_2) \\ &= \tau(n+1) \left(\frac{1 + \exp(\theta_2)}{2} \right) + (n - \tau(n+1)) \left(\frac{1 + \exp(\theta_1 + \theta_2)}{2} \right).\end{aligned}$$

Through

$$\begin{aligned}\kappa'(\theta_1, \theta_2) &= \frac{1}{n} \frac{\partial \kappa(\theta_1, \theta_2)}{\partial (\theta_1, \theta_2)} \\ &= \left(\frac{1-c}{1 + \exp(-\theta_1 - \theta_2)}, \frac{1 + c \exp(-\theta_1 - \theta_2) + (1-c) \exp(-\theta_2)}{(1 + \exp(-\theta_2))(1 + \exp(-\theta_1 - \theta_2))} \right),\end{aligned}$$

we get the saddlepoints

$$\hat{\theta}_1 = \log \left(\frac{b_1(c - b_2 + b_1)}{(b_2 - b_1)(1 - c - b_1)} \right), \quad \hat{\theta}_2 = -\log \left(\frac{c - b_2 + b_1}{b_2 - b_1} \right),$$

which is the solution $\kappa'_n(\hat{\theta}_1, \hat{\theta}_2) = (b_1, b_2)$, where b_1 and b_2 are the maximum likelihood estimates of means of $\varpi_1(T_\tau)$ and $\varpi_2(T_\tau)$, respectively. Here, $b_1 = \frac{t}{n}$, $b_2 = \frac{m}{n}$ and $c = \tau(1 + \frac{1}{n})$. Meanwhile,

$$\begin{aligned}\Sigma(\theta_1, \theta_2) &= \frac{1}{n} \frac{\partial^2 \kappa(\theta_1, \theta_2)}{\partial (\theta_1, \theta_2) \partial (\theta_1, \theta_2)'}, \\ \Sigma_{22}(\tilde{\theta}) &= \frac{c \exp(-\theta_2)}{\{1 + \exp(-\theta_2)\}^2} + \frac{(1-c) \exp(-\theta_1 - \theta_2)}{\{1 + \exp(-\theta_1 - \theta_2)\}^2}.\end{aligned}$$

In fact, $\hat{\theta}_1$ is equivalent to $\hat{\theta}$ of (2.7), and $\hat{\theta}_2$ is equivalent to $\tilde{\theta}$ in (2.7). Put these results into the formula (2.7), formula (2.2) follows.

The Proof of Proposition 1

In the more general situation as (3.2), the saddle-point approximation to the density (3.1) as follows:

The sample $X_1, X_2, \dots, X_n$ are i.i.d. with a positive definite variance. Let $\varphi(\theta) = E \exp(\theta \cdot X_1)$ be the Laplace transform defined for $\theta \in \Theta$.

Let f_n be the density of $\bar{X} = (X_1 + X_2 + \dots + X_n)/n$. The first d_1 coordinates $X_i^{(1)}$ of X_i are continuous variables and the remaining $d_2 = d - d_1$ coordinates $X_i^{(2)}$ belong to the lattice $\mathbb{Z}^{d_2}$, $\mathbb{Z}^{d_2}$ being the minimal span for $X^{(2)}$. For any $\theta \in \Theta$, have

$$f_n(x) = \varphi(\theta)^n \exp(-n\theta \cdot x) n^{(d_1 - d_2)/2} f_{n,\theta}(0), \quad (\text{A.1})$$

where $f_{n,\theta}$ is $n^{d_2/2}$ times the density of $\{\sqrt{n}(\bar{X}^{(1)} - x^{(1)}), \bar{X}^{(2)} - x^{(2)}\}$ under the measure P_θ .

$\gamma(t) = E \exp(itX)$ is the characteristic function of X . Define $\mu_k = EX^k$ and $\bar{\mu}_k = E|X|^k$, then $|\mu_j| \leq \bar{\mu}_j \leq (\bar{\mu}_{k+1})^{j/(k+1)}$ for $j = 1, 2, \dots, k$. According the inequalities

$$\begin{aligned}\left| \exp(ix) - \sum_{j=0}^k (ix)^j / j! \right| &\leq |x|^{k+1} / (k+1)!, \\ \left| -\log(1-z) - \sum_{j=1}^k z^j / j \right| &\leq |z|^{k+1} / \{(k+1)(1-|z|)\}, \quad \text{for } z = 1 - \gamma(t), |z| < 1.\end{aligned}$$

we have

$$\left| \gamma(t) - \sum_{j=0}^k \frac{mu_j}{j!} (it)^j \right| \leq \frac{\bar{\mu}_{k+1}}{(k+1)!} |t|^{k+1}, \quad (\text{A.2})$$

so, give above that we can find constants $c_1(\bar{\mu}_{k+1}), c_2(\bar{\mu}_{k+1})$ depending on $\bar{\mu}_{k+1}$ such that

$$\left| \log \gamma(t) - \sum_{j=1}^k \frac{\kappa_j}{j!} (it)^j \right| \leq c_2(\bar{\mu}_{k+1}) |t|^{k+1} \quad \text{for } |t| \leq c_1(\bar{\mu}_{k+1}), \quad (\text{A.3})$$

here the cumulants κ_j of X are defined by

$$\kappa_j = \frac{1}{i^j} \frac{d^j}{dt^j} \ln \gamma(t)|_{t=0}. \quad (\text{A.4})$$

The characteristic function of $Y = \sqrt{n}(\bar{X} - \mu_1)$, by (A.3),

$$\begin{aligned} E \exp(itY) &= \gamma\left(\frac{t}{\sqrt{n}}\right)^n \exp(-i\sqrt{n}t\mu_1) \\ &= \exp\left\{-\frac{\kappa_2}{2}t^2 - i\frac{\kappa_3}{6\sqrt{n}}t^3 + \cdots + \frac{\kappa_k}{k!n^{(k-2)/2}}(it)^k + \omega \frac{c_2(\bar{\mu}_{k+1})}{n^{(k-1)/2}}|t|^{k+1}\right\} \end{aligned} \quad (\text{A.5})$$

for $|t| \leq \sqrt{n}c_1(\bar{\mu}_{k+1})$. Consider $E \exp(itY) \in L^1$ for $n > \xi$, the variable Y has a bounded density g_n given

$$\begin{aligned} g_n(y) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-ity) E\{\exp(itY)\} dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{\kappa_2}{2}t^2\right) \left\{1 + \sum_{j=3}^k n^{-(j-2)/2} P_j(it)\right\} \exp(-ity) dt \\ &\quad - \frac{1}{2\pi} \int_{|t| > \sqrt{n}c_3} \exp\left(-\frac{\kappa_2}{2}t^2\right) \left\{1 + \sum_{j=3}^k n^{-(j-2)/2} P_j(it)\right\} \exp(ity) dt \\ &\quad + \omega_1 \frac{c_4(\kappa_2, \bar{\mu}_{k+1})}{n^{(k-1)/2}} \frac{1}{2\pi} \int_{|t| < \sqrt{n}c_3} \exp\left(-\frac{\kappa_2}{4}t^2\right) \left\{|t|^{k+1} + |t|^{3(k-1)}\right\} dt \\ &\quad + \frac{1}{2\pi} \int_{|t| > \sqrt{n}c_3} \gamma\left(\frac{t}{\sqrt{n}}\right)^n \exp(-i\sqrt{n}t\mu_1) \exp(-ity) dt. \end{aligned} \quad (\text{A.6})$$

Define

$$\delta(c) = \sum_{|t| > c} |\gamma(t)|. \quad (\text{A.7})$$

According to that if X has a density w.r.t Lebesgue measure then $\gamma(t) \rightarrow 0$ for $|t| \rightarrow \infty$, $\gamma(t)^n$ tends to zero for $t \rightarrow \infty$, then estimate the last term by

$$\sqrt{n}\delta(c_3)^{n-\xi} \frac{1}{2\pi} \int_{-\infty}^{\infty} |\gamma(t)|^\xi dt = \frac{1}{2\pi} \sqrt{n}\delta(c_3)^{n-\xi} (\|\gamma\|_\xi)^\xi. \quad (\text{A.8})$$

Then the expansion is

$$\begin{aligned} &\left| g_n(y) - \left\{ \mathbf{n}(y; \kappa_2) + \sum_{j=3}^k n^{-(j-2)/2} \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left(\frac{-\kappa_2}{2}t^2\right) P_j(it) \exp(-ity) dt \right\} \right| \\ &\leq \frac{c_5(\kappa_2, \bar{\mu}_{k+1})}{n^{(k-1)/2}} + \frac{1}{2\pi} \sqrt{n}\delta(c_3(\kappa_2, \bar{\mu}_{k+1}))^{n-\xi} (\|\gamma\|_\xi)^\xi, \end{aligned} \quad (\text{A.9})$$

where $\mathbf{n}(y; \kappa_2) = (2\pi\kappa_2)^{-1/2} \exp\left(-\frac{y^2}{2\kappa_2}\right)$ is the density of a normally distributed variable with variance κ_2 .

Use simple form of $\mathbf{n}$ as

$$\mathbf{n}(y) = (2\pi)^{-1} \int \exp\left(-\frac{t^2}{2}\right) \exp(-ity) dt,$$

with $\mathbf{n}(y) = (y; 1)$, then

$$\frac{d^m}{dy^m} \mathbf{n}(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} (-it)^m \exp\left(-\frac{t^2}{2}\right) \exp(-ity) dt.$$

Define the Hermite polynomial $H_m(y)$ as $H_m(y) = (-1)^m \mathbf{n}(y)^{-1} \frac{d^m}{dy^m} \mathbf{n}(y)$, so

$$\begin{aligned} \frac{1}{2\pi} \int_{-\infty}^{\infty} (it)^m \exp\left(-\frac{\kappa_2 t^2}{2}\right) \exp(-ity) dt &= H_m\left(\frac{y}{\sqrt{\kappa_2}}\right) \mathbf{n}\left(\frac{y}{\sqrt{\kappa_2}}\right) \kappa_2^{-(m+1)/2} \\ &= \kappa_2^{-m/2} H_m\left(\frac{y}{\sqrt{\kappa_2}}\right) \mathbf{n}(y; \kappa_2). \end{aligned} \quad (\text{A.10})$$

Combining (A.9) and (A.10) the expansion with $k = 4$ becomes

$$g_n(y) \approx \mathbf{n}(y; \kappa_2) \left\{ 1 + \frac{\zeta}{6\sqrt{n}} H_3\left(-\frac{y}{\sqrt{\kappa_2}}\right) + \frac{1}{n} \left[\frac{\zeta_4}{24} H_4\left(\frac{y}{\sqrt{\kappa_2}}\right) + \frac{\zeta_3^2}{72} H_6\left(\frac{y}{\sqrt{\kappa_2}}\right) \right] \right\}, \quad (\text{A.11})$$

where

$$\zeta_m = \kappa_m / \kappa_2^{m/2}, \quad m = 3, 4, \dots,$$

are the standardized cumulants. (A.11) is the Edge-worth expansion of the density $g_n(y)$.

In the case we considered, (A.11) becomes

$$g_n(y) \approx \mathbf{n}(y; \kappa_2) \left\{ 1 + \frac{1}{6\sqrt{n}} H_3^d(y; \{\kappa\}) + \frac{1}{n} \left[\frac{1}{24} H_4^d(y; \{\kappa\}) + \frac{1}{72} H_{3,3}^d(y; \{\kappa\}) \right] \right\}, \quad (\text{A.12})$$

where

$$H_m^d(y; \{\kappa\}) = \sum_{i_1, \dots, i_m=1}^d \kappa_{m, (i_1, \dots, i_m=1)} H_{m, (i_1, \dots, i_m=1)}^d(y; \kappa_2)$$

and

$$H_{l,m}^d(y; \{\kappa\}) = \sum_{i_1, \dots, i_m=1}^d \sum_{j_1, \dots, j_l=1}^d \kappa_{m, (i_1, \dots, i_m=1)} \kappa_{l, (j_1, \dots, j_l)} H_{m+l, (i_1, \dots, i_m, j_1, \dots, j_l)}^d(y; \kappa_2).$$

The error term of the approximation is, with $k = 4$,

$$\frac{c_2(\kappa_2, \bar{\mu}_{k+1})}{n^{(k-1)/2}} + \frac{n^{d/2}}{(2\pi)^d} \delta(c_1(\kappa_2, \bar{\mu}_{k+1}))^{n-\xi} (\|\gamma\|_\xi)^\xi, \quad (\text{A.13})$$

where

$$\delta(c) = \sup_{|t| > c, t \in \mathbb{R}^{d_1} \times [-\pi, \pi]^{d_2}} |\gamma(t)|$$

and $\|\gamma\|_\xi$ is the $L^\xi(\mathbb{R}^{d_1} \times [-\pi, \pi]^{d_2})$ -norm.

With

$$\varphi_\theta(t) = \frac{\varphi(\theta + it)}{\varphi(\theta)} \exp(-it \cdot \mu(\theta)),$$

(A.1) can be written as

$$f_{n,\theta}(0) = (2\pi)^{-d} \int_{\mathbb{R}^{d_1} \times [-\sqrt{n}\pi, \sqrt{n}\pi]^{d_2}} \varphi_\theta\left(\frac{t}{\sqrt{n}}\right)^n \exp\{i\sqrt{n}t \cdot (\mu(\theta) - x)\} dt. \quad (\text{A.14})$$

Then, according TLF, expand the inversion integral

$$\frac{1}{(2\pi)^d} \int_{J(\theta)\sqrt{n}} \bar{\varphi}_\theta \left(\frac{it}{\sqrt{n}} \right)^n dt,$$

the main term of the expansion is $\mathbf{n}(0; I) = (2\pi)^{-d/2}$ and the error term is according to (A.13), with $k = 5$, of the form

$$\frac{a_2(E_\theta | X_\theta|^6)}{n62} + (2\pi)^{-d} n^{d/2} \delta \{a_1(E_\theta | X_\theta|^6)\}^{n-\xi} (\|\bar{\varphi}_\theta\|_\xi)^\xi,$$

where a_1 and a_2 are constants depending on the specified argument. We can obtain uniformly for $\theta \in \Theta_0$ the error term of the form $O(n^{-2})$. Due to

$$H_{m,(i_1,\dots,i_m)}^d(0; \kappa_2) = 0 \text{ for } m \text{ odd and for any } (i_1, \dots, i_m),$$

the term of order $n^{-3/2}$ that has been included in the expansion is zero. Therefore, combining (A.1), (A.12) and (A.13), we obtain

$$\begin{aligned} f_n(x) &= \varphi(\theta)^n \exp(-n\theta \cdot x) n^{(d_1-d_2)/2} \mathbf{n}(0; \Sigma(\theta)) \\ &\times \left\{ 1 + \frac{1}{n} \left[\frac{1}{24} H_4^d(0, \{\kappa(\theta)\}) + \frac{1}{72} H_{3,3}^d(0, \{\kappa(\theta)\}) \right] + O(n^{-2}) \right\}, \end{aligned} \quad (\text{A.15})$$

where $\{\kappa(\theta)\}$ are cumulants $\kappa_{m,(i_1,\dots,i_m)}(\theta)$ of the form

$$\kappa_{m,(i_1,\dots,i_m)}(\theta) = \frac{\partial^m \ln \varphi(\theta)}{\partial \theta_{i_1} \dots \partial \theta_{i_m}}.$$

Now give the expansion

$$\begin{aligned} &\left| E \exp(it \cdot Y) - \exp\left(-\frac{1}{2} t \kappa_2 t^*\right) \left\{ 1 + \sum_{j=3}^k n^{-(j-2)/2} P_j(\{\kappa(it)\}) \right\} \right| \\ &\leq \frac{c_2(\kappa_2, \bar{\mu}_{k+1})}{n^{(k-1)/2}} \exp\left(-\frac{|t|^2}{4} \lambda(\kappa_2)\right) \{|t|^{k+1} + |t|^{3k-1}\} \end{aligned} \quad (\text{A.16})$$

for $|t| < \sqrt{n} c_1(\kappa_2, \bar{\mu}_{k+1})$, where $\lambda(\kappa_2)$ is the smallest eigenvalue of κ_2 . With $\kappa_2 = I$ and $\kappa_{m,(i_1,\dots,i_m)}$ is replaced by $\zeta_{m,(i_1,\dots,i_m)}$:

$$\zeta_{m,(i_1,\dots,i_m)}(\theta) = \frac{\partial^m}{\partial \xi_{i_1} \dots \partial \xi_{i_m}} \log \varphi(\theta + \xi \Theta^\dagger(\theta)) \big|_{\xi=0}.$$

According to

$$\begin{aligned} &\int_{\mathbb{R}^d} \exp\left(-\frac{1}{2} t \Theta t^*\right) (it_{l_1}) \dots (it_{l_m}) \frac{(it_1)^{a_1}}{1 + it_1/\beta} dt \\ &= \int_{\mathbb{R}^d} \exp\left\{-\frac{1}{2} (t^{(2)} - t_1 \gamma) \Sigma_{22} (t^{(2)} - t_1 \gamma)^* - \frac{\sigma_1^2}{2} t_1^2\right\} (it_{l_1}) \dots (it_{l_m}) \frac{(it_1)^{a_1}}{1 + it_1/\beta} dt \\ &= \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2} u \Sigma_{22} u^*\right) \prod_{j=1}^m (iu_{l_j} + it_1 \gamma_{l_j}) \exp\left(-\frac{\sigma_1^2}{2} t_1^2\right) \frac{(it_1)^{a_1}}{1 + it_1/\beta} du dt_1 \\ &= \sum_{s \in I(m)} \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2} u \Sigma_{22} u^*\right) (iu_{l_{s_1}}) \dots (iu_{l_{s_{|s|}}}) \delta_s \exp\left(-\frac{\sigma_1^2}{2} t_1^2\right) \frac{(it_1)^{a_1+b_s}}{1 + it_1/\beta} du dt_1, \end{aligned} \quad (\text{A.17})$$

where $\delta_s = \prod_{j \in \{1,2,\dots,m\} \setminus s} q_j$, with $q_j = \gamma_{l_j}$. For $\theta_1 \geq 0$, the expansion is

$$\int_{x_1}^\infty f_n(y, x_2, \dots, x_d) dy = \frac{\varphi(\theta)^n \exp(-n\theta \cdot x)}{n|\theta_1| \sigma_1(\theta)} n^{(d_1-d_2)/2} \mathbf{n}(0; \Sigma_{22}(\theta))$$

$$\begin{aligned}
& \times \left\{ B_0(\lambda) + \frac{\text{sgn}(\theta_1)}{\sqrt{n}} [\tau_1(\theta)B_1(\lambda) + \tau_3(\theta)B_3(\lambda)] \right. \\
& + \frac{1}{n} [\tau_0(\theta)B_0(\lambda) + \tau_2(\theta)B_2(\lambda) + \tau_4(\theta)B_4(\lambda) + \tau_6(\theta)B_6(\lambda)] \\
& \left. + O\left(\frac{B_0(\lambda)}{n^{3/2}}\right) \right\}, \tag{A.18}
\end{aligned}$$

where $\lambda = \sqrt{n}|\theta_1|\sigma_1(\theta)$. The coefficients $\tau_0(\theta), \dots, \tau_6(\theta)$ are

$$\begin{aligned}
\tau_1(\theta) &= -\frac{1}{2} \sum_{j=2}^d \zeta_{(1jj)}, \quad \tau_3(\theta) = \frac{1}{6} \zeta_{(111)}, \\
\tau_0(\theta) &= \frac{1}{8} \sum_{j,k=2}^d \zeta_{(jjkk)} - \sum_{j,k,l=2}^d \left[\frac{1}{12} \zeta_{(jkl)}^2 + \frac{1}{8} \zeta_{(jjk)} \zeta_{(llk)} \right], \\
\tau_2(\theta) &= -\frac{1}{4} \sum_{j=2}^d \zeta_{(11jj)} + \sum_{j,k=2}^d \left[\frac{1}{4} \zeta_{(11j)} \zeta_{(jkk)} + \frac{1}{8} \zeta_{1jj} \zeta_{(1kk)} + \frac{1}{4} \zeta_{(1jk)}^2 \right], \\
\tau_4 &= \frac{1}{24} \zeta_{(1111)} - \sum_{j=2}^d \left\{ \frac{1}{12} \zeta_{(111)} \zeta_{(1jj)} + \frac{1}{8} \zeta_{(11j)}^2 \right\}, \\
\tau_6(\theta) &= \frac{1}{72} \zeta_{(111)}^2.
\end{aligned}$$

(A.18) has a similar expression

$$\begin{aligned}
& \sum_{y \geq x_1, y \in \mathbb{Z}/n} f_n(y, x_2, \dots, x_d) \\
&= \frac{\varphi(\theta)^n \exp(-n\theta \cdot x)}{\sigma_1(\theta) \{1 - \exp(-|\theta_1|)\}} n^{(d_1-d-2)/2} \mathbf{n}(0; \Sigma_{22}(\theta)) \\
& \times \left\{ B_0(\lambda) + \frac{1}{\sqrt{n}} \left[\frac{\frac{1}{|\theta_1|} - \gamma_{\theta_1}}{\sigma_1(\theta)} B_1(\lambda) + \text{sgn}(\theta_1) [\tau_1(\theta)B_1(\lambda) + \tau_3(\theta)B_3(\lambda)] \right] \right. \\
& + \frac{1}{n} \left[\left(\gamma_{\theta_1} \left(\frac{1}{2} - \frac{1}{|\theta_1|} \right) + \gamma_{\theta_1}^2 \right) \frac{B_2(\lambda)}{\sigma_1(\theta)^2} + \tau_0(\theta)B_0(\lambda) + \tau_2(\theta)B_2(\lambda) + \tau_4(\theta)B_4(\lambda) \right. \\
& + \tau_6(\theta)B_6(\lambda) + \frac{\text{sgn}(\theta_1)}{\sigma_1(\theta)} \left(\frac{1}{|\theta_1|} - \gamma_{\theta_1} \right) [\tau_1(\theta)B_2(\lambda) + \tau_3(\theta)B_4(\lambda)] \left. \right] \\
& \left. + O\left(B_0(\lambda)n^{-3/2}\right) \right\}, \tag{A.19}
\end{aligned}$$

where

$$\begin{aligned}
\lambda &= \sqrt{n}|\theta_1|\sigma_1(\theta), \quad \gamma_{\theta_1} = \exp(-|\theta_1|)/\{1 - \exp(-|\theta_1|)\}, \\
\sigma_1(\theta)^2 &= \Sigma_{11}(\theta) - \Sigma_{12}(\theta)\Sigma_{22}(\theta)^{-1}\Sigma_{21}(\theta).
\end{aligned}$$

Apply (A.15) to the marginal density of $(\bar{X}_2, \dots, \bar{X}_d)$, the following approximation:

$$\begin{aligned}
& P(\bar{X}_1 \geq x_2 \mid (\bar{X}_2, \dots, \bar{X}_d) = (x_2, \dots, x_d)) \\
&= \frac{\varphi(\hat{\theta})^n \exp(-n\hat{\theta} \cdot x) \mathbf{n}(0; \Sigma_{22}(\hat{\theta})) n^{-1/2}}{|\hat{\theta}_1| \sigma_1(\hat{\theta}) \varphi(\tilde{\theta})^n \exp(-n\tilde{\theta} \cdot x) \mathbf{n}(0, \Sigma_{22}(\tilde{\theta}))} \left\{ B_0(\hat{\lambda}) + O\left(\frac{B_0(\hat{\lambda})}{1 + \hat{\lambda}} n^{-\frac{1}{2}}\right) \right\}, \tag{A.20}
\end{aligned}$$

where $\hat{\lambda} = \sqrt{n}|\hat{\theta}_1|\sigma_1(\hat{\theta})$ and

$$\mu(\hat{\theta}) = x, \quad \tilde{\theta} = (0, \tilde{\theta}^{(2)}), \quad (\mu(\tilde{\theta})_2, \dots, \mu(\tilde{\theta})_d) = (x_2, \dots, x_d).$$

Let the exponential family generated by X and P consists of the probability measures P_θ given:

$$\frac{dP_\theta}{dP}(\omega) = \varphi(\theta)^{-1} \exp(\theta \cdot X(\omega)).$$

And let r^2 be the log likelihood ratio statistic for testing $\theta_1 = 0$ in the exponential family, obtain

$$r^2 = 2n \left\{ (\hat{\theta} - \tilde{\theta}) \cdot x - \kappa(\hat{\theta}) + \kappa(\tilde{\theta}) \right\}.$$

Therefore

$$\begin{aligned} & P(\bar{X}_1 \geq x_1 \mid (\bar{X}_2, \dots, \bar{X}_d) = (x_2, \dots, x_d)) \\ &= \exp \left(-\frac{r^2}{2} \right) \left(\frac{|\Sigma_{22}(\tilde{\theta})|}{n|\Sigma(\tilde{\theta})|} \right)^{1/2} \frac{1}{|\hat{\theta}_1|} \left\{ B_0(\hat{\lambda}) + O \left(\frac{B_0(\hat{\lambda})}{1 + \hat{\lambda}} n^{-1/2} \right) \right\}. \end{aligned} \quad (\text{A.21})$$

The Laplace transform $\varphi_n(\theta_1, \theta_2)$ of $(\sum_{i=1}^n a_i Z_i, \sum_{i=1}^n Z_i)$ is

$$\varphi_n(\theta_1, \theta_2) = \left(\frac{1 + \exp(\theta_2)}{2} \right)^k \left(\frac{1 + \exp(\theta_1 + \theta_2)}{2} \right)^{n-k}. \quad (\text{A.22})$$

With $\kappa_n(\theta_1, \theta_2) = \log \varphi_n(\theta_1, \theta_2)$ and $\alpha = k/n$ we find

$$\begin{aligned} \mu_n(\theta_1, \theta_2) &= \frac{1}{n} \frac{\partial \kappa_n(\theta_1, \theta_2)}{\partial(\theta_1, \theta_2)} \\ &= \left(\frac{1 - \alpha}{1 + \exp(-\theta_1 - \theta_2)}, \frac{\alpha}{1 + \exp(-\theta_2)} + \frac{1 - \alpha}{1 + \exp(-\theta_1 - \theta_2)} \right) \end{aligned} \quad (\text{A.23})$$

and

$$\begin{aligned} \sum_n (\theta_1, \theta_2) &= \frac{1}{n} \frac{\partial^2 \kappa_n(\theta)}{\partial \theta \partial \theta^*} \\ &= \begin{pmatrix} \frac{(1-\alpha) \exp(-\theta_1 - \theta_2)}{[1 + \exp(-\theta_1 - \theta_2)]^2} & \frac{(1-\alpha) \exp(-\theta_1 - \theta_2)}{[1 + \exp(-\theta_1 - \theta_2)]^2} \\ \frac{(1-\alpha) \exp(-\theta_1 - \theta_2)}{[1 + \exp(-\theta_1 - \theta_2)]^2} & \frac{\alpha \exp(-\theta_2)}{[1 + \exp(-\theta_1 - \theta_2)]^2} + \frac{(1-\alpha) \exp(-\theta_1 - \theta_2)}{[1 + \exp(-\theta_1 - \theta_2)]^2} \end{pmatrix}. \end{aligned} \quad (\text{A.24})$$

The solution $\hat{\theta}$ to $\mu_n(\hat{\theta}) = (x_1, x_2)$, in the range $0 < x_1 < 1 - \alpha$, $0 < x_2 - x_1 < \alpha$ is

$$\hat{\theta}_2 = -\log \left(\frac{\alpha}{x_2 - x_1} - 1 \right),$$

and

$$\hat{\theta}_1 + \hat{\theta}_2 = -\log \left(\frac{1 - \alpha}{x_1} - 1 \right),$$

and the solution $\tilde{\theta}_2$ to $\mu_n(0, \tilde{\theta}_2)_2 = x_2$ is

$$\tilde{\theta}_2 = -\log \left(\frac{1}{x_2} - 1 \right). \quad (\text{A.25})$$

For $x_1 > (1 - \alpha)x_2$ and with $x_1 = \frac{s}{n}$, $x_2 = \frac{m}{n}$, (3.1) becomes

$$P \left(\sum_1^n a_i Z_i \geq s \mid \sum_1^n Z_i = m \right)$$

$$\begin{aligned}
 & \approx \frac{\left(\frac{\alpha}{\alpha+x_1-x_2}\right)^k \left(\frac{1-\alpha}{1-\alpha-x_1}\right)^{n-k} \left(\frac{1-\alpha-x_1}{x_1}\right)^{nx_1} \left(\frac{\alpha+x_1-x_2}{x_2-x_1}\right)^{n(x_2-x_1)} \sqrt{x_2(1-x_2)}}{\sqrt{n} \left(\frac{1}{1-x_2}\right)^n \left(\frac{1-x_2}{x_2}\right)^{nx_2} \sqrt{|\sum_n(\hat{\theta})| \{1 - \exp(-\hat{\theta}_1)\}}} B_0(\hat{\lambda}) \\
 & = \frac{(n-k)^{n-k+\frac{1}{2}} k^{k+\frac{1}{2}} (n-m)^{n-m+\frac{1}{2}} m^{m+\frac{1}{2}} \left\{ \frac{s(k+s-m)}{ns-(n-k)m} \right\} B_0(\hat{\lambda})}{n^{n+\frac{1}{2}} (n-k-s)^{n-k-s+\frac{1}{2}} (k+s-m)^{k+s-m+\frac{1}{2}} s^{s+\frac{1}{2}} (m-s)^{m-s+\frac{1}{2}}}, \tag{A.26}
 \end{aligned}$$

where

$$\begin{aligned}
 \hat{\lambda} &= \sqrt{n} \hat{\theta}_1 \sigma_1(\hat{\theta}) \\
 &= \log \left(\frac{s(k+s-m)}{(m-s)(n-k-s)} \right) \sqrt{\frac{s(n-k-s)(m-s)(k+s-m)}{s(n-k-s)k + (m-s)(k+s-m)(n-k)}}, \tag{A.27}
 \end{aligned}$$

and $B_0(\hat{\lambda})$ is Esscher function,

$$B_0(\hat{\lambda}) = \hat{\lambda} \exp \left(\frac{\hat{\lambda}^2}{2} \right) [1 - \mathfrak{N}(\hat{\lambda})], \tag{A.28}$$

where $\mathfrak{N}(x)$ is standard normal distribution function,

$$\mathfrak{N}(x) = \int_{-\infty}^x \mathfrak{n}(y) dy = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{y^2}{2} \right) dy.$$