Change point detection in length-biased lognormal distribution

Mei Li, Suthakaran Ratnasingam, Yubin Tian & Wei Ning

COMMUNICATIONS IN STATISTICS - SIMULATION AND COMPUTATION®
https://doi.org/10.1080/03610918.2024.2386561

Taylor & Francis
Taylor & Francis Group

Check for updates

Change point detection in length-biased lognormal distribution

Mei Li[a], Suthakaran Ratnasingam[b], Yubin Tian[c], and Wei Ning[d]

[a]Faculty of Science, Center for Applied Statistics, Kunming University of Science and Technology, Key Laboratory of Applied Statistics and Data Analysis of Department of Education of Yunnan Province, Kunming, China; [b]Department of Mathematics, CA State University, San Bernardino, CA, USA; [c]School of Mathematics and Statistics, Beijing Institute of Technology, Beijing, China; [d]Department of Mathematics and Statistics, Bowling Green State University, OH, USA

ABSTRACT

In this paper, we develop two procedures for identifying the dynamic trends in the parameters of length-biased lognormal distribution based on likelihood ratio and modified information criterion. These methods mainly consider the test of the existence of the change point and provide the maximum likelihood estimation of the change point when the change point exists. In addition, the asymptotic distribution of test statistic based on likelihood ratio is derived as an extreme value distribution, while the asymptotic distribution of test statistic based on modified information criterion is derived as a chi-square distribution. And the consistency of parameter estimation is proved. Simulations are conducted to study the performance of the proposed method in terms of power, coverage probabilities and average sizes of confidence sets. The proposed methods are applied to a real data to illustrate the detecting procedures.

ARTICLE HISTORY
Received 16 February 2023
Accepted 18 July 2024

KEYWORDS
Change point problem; Length-biased lognormal distribution; Likelihood ratio; Modified information criterion; Confidence distribution

1. Introduction

It is an important subject of life science research and reliability engineering research to estimate the parameters in the life model of biological or electronic equipment and predict the remaining useful life (Cnaan 1985; Dey and Liu 1990). In practical problems, the survival time of lung cancer patients and the service life of electronic equipment in reliability engineering are life data. General biased data such as truncation and censoring often occur in life data. Length-biased data is a kind of special biased data, and also a kind of special left truncated data (Ma and Zhou 2017). In left truncated data, if the probability of starting events occurring at any time is equal, or the left truncated variables are uniformly distributed, such left truncated data is called length-biased data. Length-biased data often appears in the fields of disease screening tests and genetics (Zelen and Feinleib 1969; de Uña-Álvarez 2004). In these studies, the observed individuals are not randomly selected from the population but are selected according to the length of the individual, that is, the probability of the individual being observed is proportional to its length (Vardi 1985, 1989). During statistical analysis, ignoring the influence of left truncating variables on the observed outcomes can lead to a large bias in the statistical inference of survival times. Thus, length-biased distribution was proposed to model this type of data. Let X be a positive random

CONTACT Yubin Tian tybbit@163.com School of Mathematics and Statistics, Beijing Institute of Technology, Beijing, China

variable, that is, true life data, with common distribution function $F(x)$ and density function $f(x)$, then the length-biased distribution function of X is

$$F_{LB}(x) = \frac{1}{E(X)} \int_0^x t dF(t), \qquad x > 0, \tag{1}$$

The corresponding probability density function (pdf) is

$$f_{LB}(x) = \frac{x \cdot f(x)}{E(X)}, \tag{2}$$

where $E(X) = \int_0^\infty t dF(t) < \infty$. More details can be referred to Qin (2017).

The length-biased lognormal distribution is a kind of length-biased distribution based on lognormal distribution. Sansgiry and Akman (2000, 2006) used it for product life modeling and deduced its corresponding reliability function. Ratnaparkhi and Naik-Nimbalkar (2012) studied the estimation problems of length-biased lognormal distribution and applied it to the analysis of oil field exploration data. The pdf of the length-biased lognormal distribution is defined as follows,

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}\left[\log x - (\mu + \sigma^2)\right]^2\right\}, \tag{3}$$

where $x > 0, -\infty < \mu < \infty, \sigma > 0$, and the length-biased lognormal distribution corresponding to formula (3) is briefly denoted as $LBLN(\mu, \sigma)$. Because life data often contains differences between different treatment schemes, ages or genders. Therefore, the study of parameter changes test based on the life model established by the length-biased lognormal distribution is of great significance in detecting the mutation of a patient's life data.

The change point problem was proposed by Page (1954), who first established a procedure to detect a single change in a parameter for continuous inspection schemes. Since then, the change point problem has been widely concerned and has become increasingly important in the fields of economy, quality control, medical research and product degradation. Many papers have studied the change point problem of life model and the lognormal distribution. For example, Tian and Yang (2022) studied the change point problem of the weighted exponential distribution. Rowlands et al. (1983) proposed lognormal distribution for modeling the quality control data from a radioimmunoassay of progesterone, and the cumulative sum (CUSUM) techniques were considered to test the change point. Azmani et al. (2009) used a von Mises distribution and a lognormal distribution to model anemometer sensor data and derived an online recursive estimate of the change point in the wind speed and direction processes. Jiang (2010) defined the aging change point of the lognormal distribution as the components service life and concluded that the life defined by the aging change point is more appropriate than the mean life or a fixed fractile life. Jandhyala and Fotopoulos (2007) proposed a change point methodology for identifying dynamic trends in the parameters of a two-parameter lognormal distribution based on the loglikelihood ratio.

The existing literature mainly studies the modeling of lognormal distribution in different fields and a few pieces of literature studied the change point problem of lognormal distribution based on the assumption that change exists. There is almost no literature on the change point problem of length-biased lognormal distribution. In this paper, we will focus on the change point test of parameters based on complete data for the life model with length-biased lognormal distribution, and give the asymptotic distribution of correlation test statistics. The effectiveness of the method is further illustrated through simulation and practical application.

The rest of this paper is organized as follows. In Sec. 2, we establish the parameters change point test model for length-biased lognormal distribution. In Sec. 3, we propose the change point test methods based on likelihood ratio and modified information criterion, and study the asymptotic property of the test statistics. Then, based on profile likelihood and deviation function, we construct the confidence distribution of change point location estimation to evaluate the uncertainty of change location estimation. A simulation study is carried to indicate the performance of the detecting procedures in Sec. 4. In Sec. 5, we illustrate our method on a survival data set which was collected from 228 patients with advanced lung cancer from the North Central Cancer Treatment Group. Some discussion is provided in Sec. 6.

2. Change point model for length-biased lognormal distribution

Assume $X_1, X_2, ..., X_n$ is a sequence of independent variables from length-biased lognormal distribution, that is, $X_i \sim LBLN(\mu_i, \sigma_i)$. Let $\boldsymbol{\theta}_i = (\mu_i, \sigma_i)$, we consider the change point problem of $\boldsymbol{\theta}_i$. Due to the multiple change points problem can be dealt with the binary segmentation method (Vostrikova 1981), we assume there is at most one change in the data. Then the change point test problem can be stated as the following hypothesis test,

$$H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = ... = \boldsymbol{\theta}_k = \boldsymbol{\theta}_{k+1} = ... = \boldsymbol{\theta}_n = \boldsymbol{\theta}, \tag{4}$$

$$H_1 : \boldsymbol{\theta}_L = \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = ... = \boldsymbol{\theta}_k \neq \boldsymbol{\theta}_{k+1} = ... = \boldsymbol{\theta}_n = \boldsymbol{\theta}_R, \tag{5}$$

where $k \in [1, 2, ..., n-1]$ is the unknown possible change point location, $\boldsymbol{\theta}_L$ and $\boldsymbol{\theta}_R$ are the parameters on the left and right sides of the change point location k.

Under the null hypothesis, the likelihood function is as follows,

$$L_0(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i; \boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{1}{x_i \sigma \sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}\left(\log x_i - (\mu + \sigma^2)\right)^2\right\}. \tag{6}$$

Consequently, the loglikelihood function is given by

$$l_0(\boldsymbol{\theta}) = -\frac{n}{2}\log(2\pi) - \sum_{i=1}^{n}\log x_i - n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(\log x_i - (\mu + \sigma^2)\right)^2. \tag{7}$$

Further, the estimates of $\boldsymbol{\theta} = (\mu, \sigma)$ are given as

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\log X_i - \hat{\sigma}^2, \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\log X_i - \frac{1}{n}\sum_{i=1}^{n}\log X_i\right)^2. \tag{8}$$

Similarly, the loglikelihood function under alternative hypothesis is given by

$$\begin{aligned} l_1(\boldsymbol{\theta}_L, \boldsymbol{\theta}_R) = {} & -\frac{n}{2}\log(2\pi) - \sum_{i=1}^{n}\log x_i - k\log\sigma_L - \frac{1}{2\sigma_L^2}\sum_{i=1}^{k}\left(\log x_i - (\mu_L + \sigma_L^2)\right)^2 \\ & -(n-k)\log\sigma_R - \frac{1}{2\sigma_R^2}\sum_{i=k+1}^{n}\left(\log x_i - (\mu_R + \sigma_R^2)\right)^2. \end{aligned} \tag{9}$$

The corresponding parameters are estimated as

$$\hat{\mu}_L = \frac{1}{k}\sum_{i=1}^{k}\log X_i - \hat{\sigma}_L^2, \qquad \hat{\sigma}_L^2 = \frac{1}{k}\sum_{i=1}^{k}\left(\log X_i - \frac{1}{k}\sum_{i=1}^{k}\log X_i\right)^2, \tag{10}$$

$$\hat{\mu}_R = \frac{1}{n-k}\sum_{i=k+1}^{n}\log X_i - \hat{\sigma}_R^2, \qquad \hat{\sigma}_R^2 = \frac{1}{n-k}\sum_{i=k+1}^{n}\left(\log X_i - \frac{1}{n-k}\sum_{i=k+1}^{n}\log X_i\right)^2. \tag{11}$$

3. Methodology

3.1. Likelihood ratio procedure

Likelihood ratio test (LRT) method is to construct test statistics based on likelihood ratio, then compare between the value of test statistic and critical value, and further decide to reject or accept the null hypothesis. The test statistic based on likelihood ratio is defined as follows,

$$\begin{aligned} Z_n &= \max_{k_0 \le k \le n-k_0} -2\log\Lambda_k = \max_{k_0 \le k \le n-k_0} -2\log\frac{L_0(\widehat{\boldsymbol{\theta}})}{L_1(\widehat{\boldsymbol{\theta}}_L, \widehat{\boldsymbol{\theta}}_R)} \\ &= \max_{k_0 \le k \le n-k_0}\left\{n\log\hat{\sigma}^2 - k\log\hat{\sigma}_L^2 - (n-k)\log\hat{\sigma}_R^2\right\} \end{aligned} \tag{12}$$

where $k_0 = 2[\log n]$ to ensure sufficient data for parameter estimation. For giving any significance level α, we reject H_0 if $Z_n > C_{n,\alpha}$, and the estimated change location is

$$\hat{k} = \operatorname*{argmax}_{k_0 \le k \le n-k_0} Z_n, \tag{13}$$

$C_{n,\alpha}$ is the critical value. To obtain $C_{n,\alpha}$, we derive the asymptotic distribution of Z_n under the null hypothesis, as specified in Theorem 3.1.

Theorem 3.1. *Under the null hypothesis H_0, when $n \to \infty$,*

$$\lim_{n\to\infty} P\left(A(u(n))Z_n^{\frac{1}{2}} - B(u(n)) \le x\right) = \exp\left(-2e^{-x}\right)$$

for $x \in \mathcal{R}$, where $u(n) = \log n$, $A(u(n)) = (2\log u(n))^{\frac{1}{2}}$, $B(u(n)) = 2\log u(n) + \log\log u(n)$.

The proof of the Theorem 3.1 is similar to the normal distribution with mean and variance change in Chen and Gupta (2011), so it is omitted here.

3.2. Modified information criterion procedure

In fact, change point problem can be regarded as model selection problem. That is, we select a better one between the null hypothesis with no change and the alternative hypothesis with at least one change. Therefore, the traditional model selection methods such as Akaike information criterion (AIC) and Schwarz information criterion (SIC) can be used in change point problem (Chen and Gupta 1997; Wang, Wang, and Zi 2020). However, as Chen, Gupta, and Pan (2006) and Basalamah et al. (2021) pointed out that, the parameter of change point position may lead to redundancy in parameter space, especially when it occurs near the beginning or end of the process. Consequently, Chen et al. proposed a modified information criterion (MIC) based on SIC by considering the location of the change point into the model complexity. Under the null model, the modified information criterion is defined as follows,

$$\begin{aligned} \mathrm{MIC}(n) &= -2\log L_0(\widehat{\boldsymbol{\theta}}) + 2\log n \\ &= n\log 2\pi + 2n\log\hat{\sigma} + 2\sum_{i=1}^{n}\log X_i + \frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}\left(\log X_i - (\hat{\mu} + \hat{\sigma}^2)\right)^2 + 2\log n. \end{aligned} \tag{14}$$

Under the alternative model, the modified information criterion is defined as

$$\begin{aligned}\mathrm{MIC}(k) &= -2\log L_1(\widehat{\boldsymbol{\theta}}_L, \widehat{\boldsymbol{\theta}}_R) + \left(4 + \left(\frac{2k}{n} - 1\right)^2\right)\log n \\ &= n\log 2\pi + 2\sum_{i=1}^{n}\log X_i + 2k\log\hat{\sigma}_L + \frac{1}{\hat{\sigma}_L^2}\sum_{i=1}^{k}\left(\log X_i - (\hat{\mu}_L + \hat{\sigma}_L^2)\right)^2 \\ &\quad +2(n-k)\log\hat{\sigma}_R + \frac{1}{\hat{\sigma}_R^2}\sum_{i=k+1}^{n}\left(\log X_i - (\hat{\mu}_R + \hat{\sigma}_R^2)\right)^2 + \left(4 + \left(\frac{2k}{n} - 1\right)^2\right)\log n,\end{aligned} \tag{15}$$

where k is the possible change location, its range is same as the likelihood ratio procedure, that is, $k \in [k_0, n-k_0]$, $k_0 = 2[\log n]$.

For the penalty item in (3.2), when $k \to 1$ or $k \to n$, it means that the change appears at the beginning or the end of the data, the penalty in $\mathrm{MIC}(k)$ is close to $5\log n$. When $k \to n/2$, the penalty is close to $4\log n$, which means the complexity of the model is minimal. This is because when the change closes to end, there is no enough data to support the estimation of parameters, resulting in a large variance of parameter estimation. In this case, a large penalty should be set. When the change point nears the middle of the data, there is enough data to estimate the parameters. The variance of parameter estimation is small, so a smaller penalty can be set. That is to say, when the suspected change points occur at the beginning or in the end of data, more powerful evidence is needed to prove the existence of such changes. Therefore, when k tends to 1 or n, a larger penalty needs to be set.

Based on the principal of minimum information criterion, we accept the null model if

$$\mathrm{MIC}(n) \le \min_{k_0 \le k \le n-k_0} \mathrm{MIC}(k),$$

which means there is no change point. And reject the null model for some k if

$$\mathrm{MIC}(n) > \min_{k_0 \le k \le n-k_0} \mathrm{MIC}(k),$$

which indicates that there is at least one change point. Hence, the estimate of the location of change point is obtained as follows,

$$\hat{k} = \operatorname*{argmin}_{k_0 \le k \le n-k_0} \mathrm{MIC}(k) \tag{16}$$

However, $\mathrm{MIC}(n)$ and $\mathrm{MIC}(k)$ could be very close sometimes, which might be caused by the fluctuation of data. Consequently, it may be improper to make a conclusion whether to reject the null hypothesis. To make the conclusion more statistically significant, we construct a test statistic M_n which defined as follows,

$$M_n = \mathrm{MIC}(n) - \min_{k_0 \le k \le n-k_0} \mathrm{MIC}(k) + 2\log n, \tag{17}$$

where $2\log n$ is to eliminate the difference of the constant terms in $\mathrm{MIC}(n)$ and $\mathrm{MIC}(k)$. We reject the null hypothesis when M_n is larger than critical value C_α. To obtain n C_α, we derive the asymptotic distribution of M_n under Wald conditions and regular conditions which are listed in Appendix A.

Theorem 3.2. *Under the null hypothesis H_0,*

$$M_n \to \chi^2(2)$$

in distribution as $n \to \infty$, where M_n is defined in Eq. (17).

The proof of the Theorem 3.2 is in Appendix C. To prove the theorem, we first give the following two lemmas.

Lemma 3.3. *Under the null hypothesis H_0,*

$$(\widehat{\boldsymbol{\theta}}_L, \widehat{\boldsymbol{\theta}}_R) \to (\boldsymbol{\theta}_0, \boldsymbol{\theta}_0)$$

in probability as $n \to \infty$, where $\boldsymbol{\theta}_0 = (\mu_0, \sigma_0)$ is the true parameter of $\boldsymbol{\theta} = (\mu, \sigma)$.

Lemma 3.4. *Under the null hypothesis H_0,*

$$\frac{\hat{k}}{n} \to \frac{1}{2}$$

in probability as $n \to \infty$.

The proofs of Lemma 3.3 and Lemma 3.4 are in Appendix B.

3.3. Profile loglikelihood and deviance function

The confidence distribution can be regarded as a distribution that depends on samples, which can be used to study the interval estimation and point estimation of the parameter of interest (Singh, Xie, and Strawderman 2007). In particular, it can provide the confidence interval of the parameter of interest at any nominal level through the confidence curve. In change point problem, the change location is a discrete variable, and conducting uncertainty analysis for the estimated change location based on observation is a challenging task. Similar to reference Cunen, Hermansen, and Hjort (2017), we establish the confidence curve of change point location based on profile loglikelihood and deviation function to analyze the uncertainty of change location estimation.

Assume $\{X_1, X_2, ..., X_k\}$ is a sample from the population density function $f(x, \Theta_L)$, while $\{X_{k+1}, ..., X_n\}$ comes from the population density function $f(x, \Theta_R)$. Then the loglikelihood function of sample $\{X_1, X_2, ..., X_n\}$ with change point k is defined as,

$$l(k, \Theta_L, \Theta_R) = \sum_{i=1}^{k} \log f(x_i, \Theta_L) + \sum_{i=k+1}^{n} \log f(x_i, \Theta_R). \tag{18}$$

For a given change point k, the profile loglikelihood function can be obtained by maximizing the loglikelihood function (18), where the profile loglikelihood function is defined as follows,

$$l_{prof}(k) = \max_{\Theta_L, \Theta_R} l(k, \Theta_L, \Theta_R) = l(k, \hat{\Theta}_L, \hat{\Theta}_R), \tag{19}$$

where $\hat{\Theta}_L, \hat{\Theta}_R$ are MLEs of Θ_L and Θ_R for a given k. Then $\hat{k}$ can be obtained by $\hat{k} = \arg l_{prof}(\hat{k}) = \operatorname{argmax}_k l_{prof}(k)$.

The deviance function of k is given by

$$D(k, X) = 2\left\{l_{prof}(\hat{k}) - l_{prof}(k)\right\}, \tag{20}$$

where $X = (X_1, X_2, ..., X_n)$. To construct k's confidence curve based on deviance function, we consider the distribution of $D(k, X)$ at k, which denoted as $R_k(t) = P_{k, \hat{\Theta}_L, \hat{\Theta}_R}\{D(k, X) \le t\}$. However, the location of change point is discrete, $R_k(t)$ does not satisfy Wilks Theorem (Wilks 1938). Therefore, we compute $R_k(t)$ through simulation. And the confidence curve of k is defined by

$$cc(k, x_{obs}) = R_k(D(k, x_{obs})) = P_{k, \hat{\Theta}_L, \hat{\Theta}_R}\{D(k, X) \le D(k, x_{obs})\}, \tag{21}$$

$cc(k, x_{obs}) = 1 - p$, p is p-value that tests the null hypothesis with no change point. $cc(k, x_{obs})$ can be obtained by

$$cc(k, x_{obs}) = \frac{1}{B}\sum_{j=1}^{B} I\left\{D(k, X_j^*) < D(k, x_{obs})\right\}, \tag{22}$$

where B is a large number and $B = 1000$ typically. X_j^* is a sample from $f(x, \hat{\Theta}_L)$ and $f(x, \hat{\Theta}_R)$ with a given k.

Since the test statistics used for the change point test in this paper are based on the likelihood ratio and MIC criterion respectively, we estimate the location of change point through (13) and (16) for computing the deviation function. The specific simulation steps are as follows,

Step 1. Given sample size n and change point k_{true}, generate a group of random samples with change point k_{true} based on parameters Θ_L and Θ_L, record them as x_{obs}, and calculate the deviation $D(k, x_{obs})$ at each possible change point based on x_{obs}.
Step 2. Compute $(\hat{k}, \hat{\Theta}_L, \hat{\Theta}_R)$ based on Step 1.
Step 3. Generate $B = 1000$ random samples $X_j^*(j = 1, ..., B)$, where $x_{j1}^*, ..., x_{jk}^*$ are from $f(x, \hat{\Theta}_L)$, $x_{j,k+1}^*, ..., x_{jn}^*$ are from $f(x, \hat{\Theta}_R)$. And compute $D(k, X_j^*)$. Then the confidence curve $cc(k, x_{obs})$ under the possible change locations is the frequencies of $D(k, X_j^*) < D(k, x_{obs})$.
Step 4. Repeat Step 1 to Step 3 for $N = 1000$ times to obtain N confidence curves about the change point k.
Step 5. Given significance level α, the corresponding confidence set of each confidence curve is $K_{set} = \left\{k : cc(k, x_{obs}) < 1 - \alpha\right\}$.
Step 6. Coverage probabilities of confidence sets. Based on the confidence set K_{set} obtained from each confidence curve, then the frequency of k_{true} in K_{set} is the coverage probability of the confidence set at the corresponding confidence level $1 - \alpha$.

$$cp = \frac{1}{N}\sum_{i=1}^{N} I\{k_{true} \in K_{set}\}.$$

Step 7. Average size of confidence sets. Since the confidence set is the set of estimates of the change location, it is a set of discrete points. Thus, unlike the calculation method for continuous interval length, the number of elements in the confidence set is the size of the confidence set. Noting the number of elements in each confidence set K_{set} as $m_i(i = 1, ..., N)$, then the average size of confidence sets is obtained by

$$l_{set} = \frac{1}{N}\sum_{i=1}^{N} m_i.$$

In conclusion, when cp is closer to $1 - \alpha$, the smaller l_{set} means the corresponding estimate method is better.

4. Simulation

4.1. Critical value and probability of type I error

First of all, we compute critical values and probability of type I errors based on the proposed procedure. For convenience, denoting C_{LRT} and C_{MIC} as the critical value of the methods, $P(type\ I\ error)_{LRT}$ and $P(type\ I\ error)_{MIC}$ are probability of type I error. The sample size is $n = \{50, 100, 150\}$, parameters are $(\mu, \sigma) = (0, 2)$, significance level is 0.05. Then the results are listed in Table 1.

Table 1. Critical values and probabilities of type I error.

n	C_{LRT}	$P(type\ I\ error)_{LRT}$	C_{MIC}	$P(type\ I\ error)_{MIC}$
50	12.14135	0.0524	11.15856	0.0518
100	12.32210	0.0526	10.82753	0.0529
150	12.53782	0.0474	11.01736	0.0495

Table 2. Power comparison between MIC and LRT for $\alpha = 0.05$.

n	k	Method	$(-2,3)$	$(-3,3.5)$	$(2,1)$	$(3,0.5)$	$(2,2)$	$(3,2)$	$(0,2.5)$	$(0,3)$	$(1,2.5)$	$(2,3)$
50	12	MIC	0.743	0.997	0.727	0.999	0.484	0.893	0.506	0.998	0.880	1.000
		LRT	0.733	0.997	0.731	0.999	0.484	0.893	0.502	0.998	0.878	1.000
	25	MIC	0.940	1.000	0.883	1.000	0.712	0.982	0.762	1.000	0.982	1.000
		LRT	0.932	1.000	0.853	1.000	0.663	0.973	0.722	1.000	0.976	1.000
	38	MIC	0.833	0.999	0.545	0.992	0.481	0.883	0.605	0.996	0.901	1.000
		LRT	0.835	0.999	0.548	0.992	0.475	0.881	0.595	0.997	0.900	1.000
100	25	MIC	0.998	1.000	0.986	1.000	0.898	0.999	0.915	1.000	0.999	1.000
		LRT	0.997	1.000	0.986	1.000	0.883	0.999	0.907	1.000	0.998	1.000
	50	MIC	1.000	1.000	0.999	1.000	0.982	1.000	0.988	1.000	1.000	1.000
		LRT	1.000	1.000	0.999	1.000	0.971	1.000	0.979	1.000	1.000	1.000
	75	MIC	0.996	1.000	0.982	1.000	0.902	0.999	0.929	1.000	0.998	1.000
		LRT	0.996	1.000	0.979	1.000	0.891	0.998	0.923	1.000	0.998	1.000
150	38	MIC	1.000	1.000	1.000	1.000	0.986	1.000	0.993	1.000	1.000	1.000
		LRT	1.000	1.000	1.000	1.000	0.983	1.000	0.992	1.000	1.000	1.000
	75	MIC	1.000	1.000	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000
		LRT	1.000	1.000	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000
	112	MIC	1.000	1.000	1.000	1.000	0.990	1.000	0.997	1.000	1.000	1.000
		LRT	1.000	1.000	1.000	1.000	0.989	1.000	0.996	1.000	1.000	1.000

From Table 1, it could be seen that $P(type\ I\ error)_{LRT}$ and $P(type\ I\ error)_{MIC}$ fluctuated around 0.05 under different sample sizes, which indicating that both the two procedures could effectively control the probability of type I error.

4.2. Power comparison

To investigate the performance of the test procedure proposed in Sec. 3, we conduct simulations for the powers of Z_n and M_n, and make a comparison between these two methods. In order to ensure the fairness of the comparison, two methods are carried out under the same settings. The parameters before the change point are set to $(\mu_L, \sigma_L) = (0, 2)$, and the parameters after the change point are set to $\mu_R = \{-3, -2, -1, 0, 1, 2, 3\}$, $\sigma_R = \{0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5\}$ respectively. We consider three sample sizes $n = 50, n = 100$ and $n = 150$. Under all sample sizes, we set the possible positions of change points to $\left\{\frac{1}{4}n, \frac{2}{4}n, \frac{3}{4}n\right\}$. The significance level is $\alpha = 0.05$. We compute the powers of all the combinations of (μ_R, σ_R). To save pages, we only show representative results, which are listed in Table 2.

From Table 2, we firstly observe that the powers of MIC are higher than LRT in general, which means that the method proposed based on MIC has good performance because of considering the influence of change point location on model complexity. Then, the power of the test increases with the increase of sample size due to the more accurate estimation of parameters. The power increases as the increase of the parameters. It also indicates that the performance of both methods are better when the change point is closer to the middle of the data.

4.3. Coverage probabilities and average sizes of confidence sets

To investigate the performance of our proposed method in estimating the change location, we study the coverage probabilities and average sizes of confidence sets based on the estimator $\hat{k}$ by

Table 3. The comparisons of coverage probabilities.

		$n=50$				$n=100$			
		$k=12$		$k=25$		$k=25$		$k=50$	
(μ_n, σ_n)	α	LRT	MIC	LRT	MIC	LRT	MIC	LRT	MIC
(-3.0,3.5)	0.90	0.888	0.888	0.887	0.887	0.886	0.886	0.884	0.884
	0.95	0.946	0.946	0.945	0.945	0.941	0.941	0.934	0.934
	0.99	0.988	0.988	0.989	0.989	0.981	0.981	0.985	0.985
(1.0,2.5)	0.90	0.863	0.865	0.884	0.884	0.898	0.898	0.884	0.884
	0.95	0.924	0.925	0.937	0.937	0.945	0.946	0.934	0.934
	0.99	0.983	0.983	0.981	0.981	0.978	0.978	0.985	0.985
(2.0,3.0)	0.90	0.886	0.887	0.896	0.896	0.896	0.895	0.906	0.906
	0.95	0.939	0.939	0.945	0.945	0.945	0.945	0.949	0.949
	0.99	0.985	0.985	0.985	0.985	0.986	0.986	0.990	0.990
(3.0,3.5)	0.90	0.954	0.956	0.948	0.948	0.962	0.963	0.954	0.954
	0.95	0.962	0.964	0.954	0.954	0.963	0.964	0.957	0.957
	0.99	0.989	0.989	0.984	0.984	0.988	0.988	0.987	0.987

Table 4. The comparisons of average sizes of confidence sets.

		$n=50$				$n=100$			
		$k=12$		$k=25$		$k=25$		$k=50$	
(μ_n, σ_n)	α	LRT	MIC	LRT	MIC	LRT	MIC	LRT	MIC
(-3.0,3.5)	0.90	2.93	2.92	2.78	2.78	2.70	2.70	2.48	2.48
	0.95	3.99	3.98	3.78	3.78	3.56	3.56	3.24	3.24
	0.99	6.87	6.87	6.25	6.25	5.60	5.59	5.22	5.22
(1.0,2.5)	0.90	6.51	6.44	5.62	5.62	5.32	5.31	2.48	2.48
	0.95	9.15	9.08	7.87	7.88	7.26	7.25	3.24	3.24
	0.99	15.43	15.38	13.50	13.51	12.09	12.07	5.22	5.22
(2.0,3.0)	0.90	1.32	1.32	1.27	1.27	1.24	1.24	1.26	1.26
	0.95	1.74	1.74	1.64	1.64	1.56	1.56	1.58	1.58
	0.99	3.04	3.04	2.65	2.65	2.52	2.52	2.49	2.49
(3.0,3.5)	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.95	1.03	1.03	1.03	1.03	1.01	1.01	1.01	1.01
	0.99	1.23	1.23	1.25	1.25	1.17	1.17	1.18	1.18

simulations.To save space, we only consider four types of settings for parameters after the change, that is, $(\mu_R, \sigma_R) = \{(-3, 3.5), (1, 2.5), (2, 3), (3, 3.5)\}$. The true location of the change are set as $k = \{\frac{1}{4}n, \frac{1}{2}n\}$ for symmetry. And the confidence level are set as $\{0.90, 0.95, 0.99\}$.

The results are presented in Tables 3 and 4. From Table 3, the MIC performs slightly better than the LRT method in general. As the differences among parameters increase, the two methods perform similarly and the coverage probabilities get closer to the prescribed confidence levels.

From Table 4, we can observe that the average sizes of confidence sets under the two methods are basically the same when the change occurs at the middle. When the change point is close to the endpoint, the average sizes of the confidence sets under the MIC procedure method are shorter. We also observe that, as the increases in differences among parameters, both two methods obtain similar average sizes of confidence sets. In addition, the average sizes of confidence sets under both methods become smaller with the increase in the sample size. In a word, when the change point is close to the endpoint, the test method based on MIC is better. While the performance of the two methods is equivalent when the change is close to the middle of data. There is no significant difference in confidence set coverage under different sample sizes, but the average sizes of confidence sets decrease as the increase of sample size. It indicates that the larger the sample size, the more accurate the estimation of change location.

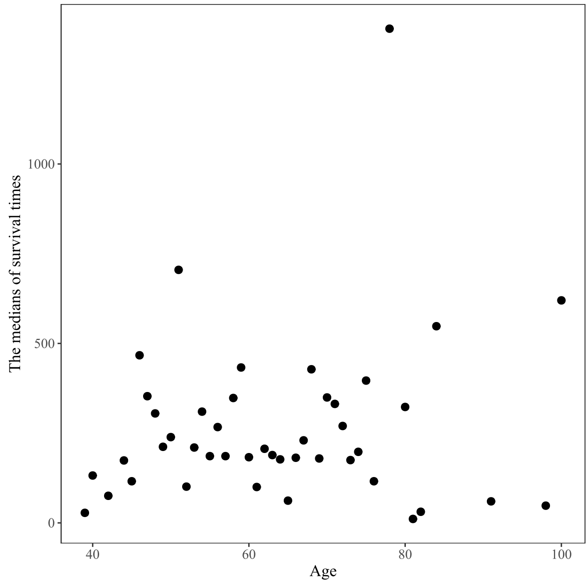

Figure 1. Scatter plot of median survival times of 44 groups of patients.

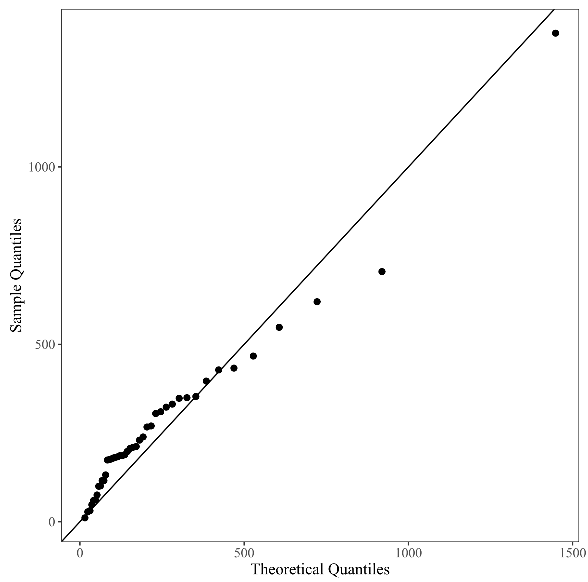

Figure 2. Q-Q Plot for testing the length-biased lognormal distribution.

5. Application: survival times of lung cancer data

To illustrate our testing method, we apply the test procedure based on LRT and MIC to detect possible changes in the survival time data set of Canadian lung cancer patients. The data set consists of 228 patients with advanced lung cancer from the North Central Cancer Treatment Group (Loprinzi et al. 1994), which contains the basic information of patients, such as age, sex, institution code and some biological indicators tracked and measured, such as survival time, survival status, ECOG performance score as rated by the physician. Because only patients diagnosed with lung cancer and who survived until the beginning of the survey can enter the survey and be observed. Therefore, this is a length-biased dataset. We select 165 patients from this dataset,

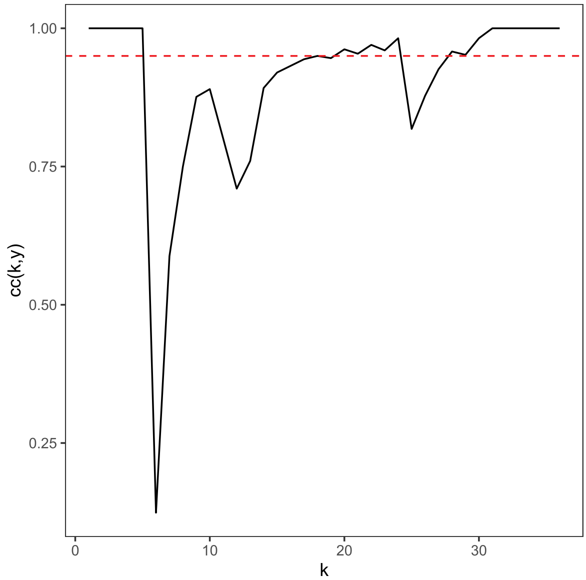

Figure 3. The confidence curve of $\hat{k} = 6$.

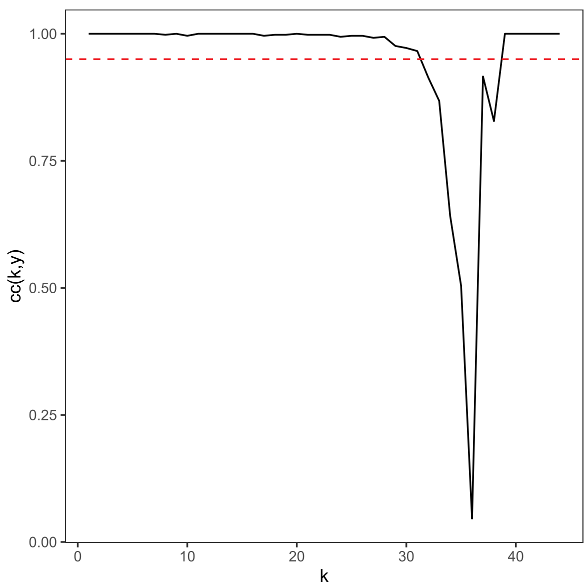

Figure 4. The confidence curve of $\hat{k} = 36$.

divide into 44 groups by age, and compute the median of survival times for these groups, which is shown in Figure 1. The Q-Q plot of the 44 medians survival times and length-biased lognormal distribution is shown in Figure 2.

From Figure 2, we observe that almost all points fall near the line $y = x$, which indicates that the medians of survival times follow the length-biased lognormal distribution. Therefore, we apply the proposed methods of length-biased lognormal distribution to detect whether there are change points in the medians survival times of these 44 groups.

Based on likelihood ratio procedure, we obtain $Z_n = 17.3524$, which is larger than the critical value $C_{LRT} = 16.4194$. Similarly, $M_n = 15.8120 > \chi^2_{0.95} = 5.9915$. Thus, we reject the null hypothesis, that is, there is a change point and the location is 36. Combining with binary segmentation procedure, we detect another change. Therefore, there are two change points $\{6, 36\}$ in the median survival time dataset, corresponding to the ages of 46 and 75 in the data. And the estimations of parameters are $(\hat{\mu}_1, \hat{\sigma}_1) = (4.0456, 0.8490)$, $(\hat{\mu}_2, \hat{\sigma}_2) = (5.1802, 0.4920)$, $(\hat{\mu}_3, \hat{\sigma}_3) = (2.3662, 1.6050)$.

We also plot the confidence curves of $\hat{k}$, which are shown in Figures 3 and 4. The abscissas in Figures 3 and 4 represent the possible value of the change points, the ordinate is $cc(k, y)$, the red line is the baseline with a confidence level of 0.95, and the k corresponding to the lowest peak is the estimated location of the change.

6. Conclusions

In this paper, we study the change point testing problem of life model based on the length-biased lognormal distribution. Firstly, we propose two approaches based on likelihood ratio and MIC criterion. Then we prove that the asymptotic distribution of the test statistic based on likelihood ratio is an extreme value distribution, and the asymptotic distribution of the test statistics based on MIC criterion is a chi-square distribution with 2 degrees of freedom. To investigate the performance of the proposed methods, some simulations are conducted in terms of power, coverage probabilities and average sizes of confidence sets. The results show that the method based on MIC performs better than the one based on likelihood ratio in general. Finally, an application to the survival time data sets of lung cancer patients illustrates the detecting procedure.

In view of the wide use of length-biased lognormal distribution, this paper only considers the change point test under complete data. However, the collected data are censored sometimes in many practical applications due to some random causes. The proposed method is not applicable in this situation. Therefore, we will modify the proposed method to detect the change point in censored data.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by National Natural Science Foundation of China under Grant [12131001].

References

Azmani, M., S. Reboul, J. B. Choquel, and M. Benjelloun. 2009. A recursive change point estimate of the wind speed and direction. *Annals of Mathematics* 169 (2):449–90.

Basalamah, D., K. K. Said, W. Ning, and Y. Tian. 2021. Modified information criterion for linear regression change-point model with its applications. *Communications in Statistics – Simulation and Computation* 50 (1): 180–97. doi: 10.1080/03610918.2018.1554109.

Chen, J., A. K. Gupta, and J. Pan. 2006. Information criterion and change point problem for regular models. *Sankhya* 68:252–82.

Chen, J., and A. K. Gupta. 1997. Testing and locating variance changepoints with application to stock prices. *Journal of the American Statistical Association* 92 (438):739–47. doi: 10.2307/2965722.

Chen, J., and A. K. Gupta. 2011. *Parametric statistical change point analysis: With applications to genetics, medicine, and finance*. 2 ed. New York: Birkhäuser Boston.

Cnaan, A. 1985. Survival models with two phases and length biased sampling. *Communications in Statistics - Theory and Methods* 14 (4):861–86. doi: 10.1080/03610928508828955.

Cunen, C., G. Hermansen, and N. L. Hjort. 2017. Confidence distributions for change-points and regime shifts. *Journal of Statistical Planning and Inference* 195:14–34. doi: 10.1016/j.jspi.2017.09.009.

de Uña-Álvarez, J. 2004. Nonparametric estimation under length-biased sampling and type i censoring: A moment based approach. *Annals of the Institute of Statistical Mathematics* 56 (4):667–81. doi: 10.1007/BF02506482.

Dey, D. K., and P. L. Liu. 1990. Estimation of parameters and reliability from generalized life models. *Communications in Statistics – Theory and Methods* 19 (3):1073–99. doi: 10.1080/03610929008830248.

Jandhyala, V. K., and S. B. Fotopoulos. 2007. Estimating the unknown change point in the parameters of the log-normal distribution. *Environmetrics* 18 (2):141–55. doi: 10.1002/env.808.

Jiang, R. 2010. Aging change point of lognormal distribution and its implications. *Advanced Materials Research* 118-120 (120):439–43. doi: 10.4028/www.scientific.net/AMR.118-120.439.

Loprinzi, C. L., J. A. Laurie, H. S. Wieand, J. E. Krook, P. J. Novotny, J. W. Kugler, J. Bartel, M. Law, M. Bateman, and N. E. Klatt. 1994. Prospective evaluation of prognostic variables from patient-completed questionnaires. North central cancer treatment group. *Journal of Clinical Oncology* 12 (3):601–7. doi: 10.1200/JCO.1994.12.3.601.

Ma, H., and Y. Zhou. 2017. Pseudo-likelihood for case–cohort studies under length-biased sampling. *Communications in Statistics – Theory and Methods* 46 (1):28–48. doi: 10.1080/03610926.2014.983613.

Page, E. S. 1954. Continuous inspection schemes. *Biometrika* 41 (1-2):100–15. doi: 10.2307/2333009.

Qin, J. 2017. *Examples and basic theories for length biased sampling problems*. Singapore: Springer Singapore.

Ratnaparkhi, M. V., and U. Naik-Nimbalkar. 2012. The length-biased lognormal distribution and its application in the analysis of data from oil field exploration studies. *Journal of Modern Applied Statistical Methods* 11 (1):255–60. doi: 10.22237/jmasm/1335846060.

Rowlands, R. J., K. Griffiths, K. W. Kemp, A. B. Nix, G. Richards, and D. W. Wilson. 1983. Application of cusum techniques to the routine monitoring of analytical performance in clinical laboratories. *Statistics in Medicine* 2 (2):141–5. doi: 10.1002/sim.4780020207.

Sansgiry, P., and O. Akman. 2000. Transformations of the lognormal distribution as a selection model. *The American Statistician* 54 (4):307–9. doi: 10.2307/2685783.

Sansgiry, P., and O. Akman. 2006. Reliability estimation via length-biased transformation. *Communications in Statistics* 30 (11):2473–9. doi: 10.1081/STA-100107699.

Singh, K., M. Xie, and W. E. Strawderman. 2007. Complex datasets and inverse problems: Tomography, networks and beyond—Confidence distribution (cd): distribution estimator of a parameter. *Lecture Notes Monograph* 54: 132–50.

Tian, W., and Y. Yang. 2022. Change point analysis for weighted exponential distribution. *Communications in Statistics – Simulation and Computation* 53 (1):367–79. doi: 10.1080/03610918.2021.2020288.

Vardi, Y. 1985. Empirical distributions in selection bias models. *Annals of Statistics* 13:178–203.

Vardi, Y. 1989. Multiplicative censoring, renewal processes, deconvolution and decreasing density: Nonparametric estimation. *Biometrika* 76 (4):751–61. doi: 10.1093/biomet/76.4.751.

Vostrikova, L. J. 1981. Detecting 'disorder' in multidimensional random process. *Soviet Math Doklady* 24:55–9.

Wald, A. 1949. Note on the consistency of the maximum likelihood estimate. *The Annals of Mathematical Statistics* 20 (4):595–601. doi: 10.1214/aoms/1177729952.

Wang, Y., Z. Wang, and X. Zi. 2020. Rank-based multiple change-point detection. *Communications in Statistics – Theory and Methods* 49 (14):3438–54. doi: 10.1080/03610926.2019.1589515.

Wilks, S. S. 1938. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics* 9 (1):60–2. doi: 10.1214/aoms/1177732360.

Zelen, M., and M. Feinleib. 1969. On the theory of screening for chronic diseases. *Biometrika* 56 (3):601–14. doi: 10.2307/2334668.

Appendix A: Wald conditions and regularity conditions

W1. The distribution of X is either discrete for (μ, σ) or is absolutely continuous for (μ, σ).

W2. For sufficiently small δ and for sufficiently large ρ, the expected values

$$E\left[\log \sup_{|\mu'-\mu|<\delta, |\sigma'-\sigma|<\delta} f(X; \mu', \sigma')\right]^2 < \infty,$$

$$E\left[\log \sup_{|\mu'-\mu|>\rho, |\sigma'-\sigma|>\rho} f(X; \mu', \sigma')\right]^2 < \infty.$$

W3. The density function $f(x; \mu, \sigma)$ is continuous in (μ, σ) for all x.

W4. The cumulative distribution function $F(x; \mu, \sigma)$ is identifiable.

W5. $\lim_{||\boldsymbol{\theta}|| \to \infty} f(x; \boldsymbol{\theta}) = 0$ for all x.

W6. The parameter space Θ is a closed subset of the 2-dimensional Cartesian space.

W7. $f(x; \mu, \sigma, \delta) = \sup_{|\mu'-\mu|<\delta, |\sigma'-\sigma|<\delta} f(X; \mu', \sigma')$ is a measurable function of x for any fixed (μ, σ) and δ.

R1. For each $\boldsymbol{\theta} = (\mu, \sigma) \in \Theta$, the following derivatives are existing for all x,

$$\frac{\partial \log f(x; \mu, \sigma)}{\partial \boldsymbol{\theta}}, \frac{\partial^2 \log f(x; \mu, \sigma)}{\partial \boldsymbol{\theta}^2}, \frac{\partial^3 \log f(x; \mu, \sigma)}{\partial \boldsymbol{\theta}^3}.$$

R2. For $\boldsymbol{\theta}$ in the neighborhood $N(\boldsymbol{\theta}_0)$, there exist function $g(x)$ and $H(x)$ such that the following relations hold for all x,

$$\left|\frac{\partial f(x, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right| \le g(x), \left|\frac{\partial^2 f(x, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2}\right| \le g(x),$$

$$\left|\frac{\partial^2 \log f(x, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2}\right|^2 \le H(x), \left|\frac{\partial^3 \log f(x, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^3}\right| \le H(x),$$

and

$$\int g(x)dx < \infty, \qquad E_{\boldsymbol{\theta}}(H(X)) < \infty.$$

R3. For each $\boldsymbol{\theta} \in \Theta$,

$$0 < E_{\boldsymbol{\theta}}\left[\left(\frac{\partial \log f(X; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^2\right] < \infty, \qquad E_{\boldsymbol{\theta}}\left[\left|\frac{\partial \log f(X; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|^3\right] < \infty$$

Appendix B: Proofs for lemmas

Proof of Lemma 3.3. Similar to the proof in Wald (1949), we only need to show that, let $(\tilde{\mu}_L, \tilde{\sigma}_L, \tilde{\mu}_R, \tilde{\sigma}_R) \ne (\mu_0, \sigma_0, \mu_0, \sigma_0)$, for $\forall \epsilon > 0$, there exists a small neighborhood N_ϵ of $(\tilde{\mu}_L, \tilde{\sigma}_L, \tilde{\mu}_R, \tilde{\sigma}_R)$,

$$N_\epsilon = \left\{(\mu_L, \sigma_L, \mu_R, \sigma_R) : |\mu_L - \tilde{\mu}_L| < \delta_\epsilon, |\sigma_L - \tilde{\sigma}_L| < \delta_\epsilon, |\mu_R - \tilde{\mu}_R| < \delta_\epsilon, |\sigma_R - \tilde{\sigma}_R| < \delta_\epsilon\right\} \tag{B1}$$

such that

$$\sup_{(\mu_L, \sigma_L, \mu_R, \sigma_R) \in N_\epsilon} [l_1(\mu_L, \sigma_L, \mu_R, \sigma_R) - l_0(\mu_0, \sigma_0)] < 0. \tag{B2}$$

Let

$$Y_i = \log f(X_i; \tilde{\mu}_L, \tilde{\sigma}_L) - \log f(X_i; \mu_0, \sigma_0)$$

$$Z_i = \log f(X_i; \tilde{\mu}_R, \tilde{\sigma}_R) - \log f(X_i; \mu_0, \sigma_0)$$

When $(\mu_L, \sigma_L, \mu_R, \sigma_R) \in N_\epsilon$, $(\mu_L, \sigma_L, \mu_R, \sigma_R) \ne (\mu_0, \sigma_0, \mu_0, \sigma_0)$, by using Jensen's Inequality for Y_i and Z_i,

$$E(Y_i) = E\left(\log \frac{f(X_i; \tilde{\mu}_L, \tilde{\sigma}_L)}{f(X_i; \mu_0, \sigma_0)}\right) \le \log E\left(\frac{f(X_i; \tilde{\mu}_L, \tilde{\sigma}_L)}{f(X_i; \mu_0, \sigma_0)}\right) = 0, \tag{B3}$$

$$E(Z_i) = E\left(\log \frac{f(X_i; \tilde{\mu}_R, \tilde{\sigma}_R)}{f(X_i; \mu_0, \sigma_0)}\right) \le \log E\left(\frac{f(X_i; \tilde{\mu}_R, \tilde{\sigma}_R)}{f(X_i; \mu_0, \sigma_0)}\right) = 0 \tag{B4}$$

Then we have,

$$\begin{aligned}
&\sup[l_1(\mu_L,\sigma_L,\mu_R,\sigma_R)-l_0(\mu_0,\sigma_0)] \\
&\quad=\sup\{[l_1(\mu_L,\sigma_L,\mu_R,\sigma_R)-l_0(\ \tilde\mu_L,\tilde\sigma_L,\tilde\mu_R,\tilde\sigma_R)]+[l_1(\tilde\mu_L,\tilde\sigma_L,\tilde\mu_R,\tilde\sigma_R)-l_0(\mu_0,\sigma_0)]\} \\
&\quad\le 2\sum_{i=1}^{k}(Y_i-E(Y_i))+2\sum_{i=k+1}^{n}(Z_i-E(Z_i))+2\sum_{i=1}^{k}E(Y_i)+2\sum_{i=k+1}^{n}E(Z_i) \\
&\quad=2\sum_{i=1}^{k}(Y_i-E(Y_i))+2\sum_{i=k+1}^{n}(Z_i-E(Z_i))+2kE(Y_1)+2(n-k)E(Z_1) \\
&\quad\le 2\min(k,n-k)(E(Y_1)+E(Z_1))+O_p\left(\sqrt{n}\right)
\end{aligned} \tag{B5}$$

Hence, for any k and a small neighborhood N_ϵ of $(\tilde\mu_L,\tilde\sigma_L,\tilde\mu_R,\tilde\sigma_R)$, we have

$$\sup_{(\mu_L,\sigma_L,\mu_R,\sigma_R)\in N_\epsilon}[l_1(\mu_L,\sigma_L,\mu_R,\sigma_R)-l_0(\mu_0,\sigma_0)]<0. \tag{B6}$$

■

Proof of Lemma 3.4. Based on modified information criterion, we have MIC(k) as follows,

$$\mathrm{MIC}(\boldsymbol\theta_L,\boldsymbol\theta_R;k)=-2l_1(\boldsymbol\theta_L,\boldsymbol\theta_R)+\left\{4+\left(\frac{2k}{n}-1\right)^2\right\}\log n.$$

It is obvious that

$$\mathrm{MIC}(k)=\mathrm{MIC}(\widehat{\boldsymbol\theta}_L,\widehat{\boldsymbol\theta}_R;k)\le\mathrm{MIC}(\boldsymbol\theta_L,\boldsymbol\theta_R;k)$$

for any $(\boldsymbol\theta_L,\boldsymbol\theta_R)$. For any $\epsilon>0$, we define $\Delta_\epsilon=\left\{k:|\frac{k}{n}-\frac{1}{2}|<\epsilon\right\}$. When the location of change point is $k=\frac{n}{2}$, there is no penalty for the change point position in MIC(k), and $\log l_1\left(\boldsymbol\theta_0,\boldsymbol\theta_0;\frac{n}{2}\right)=\log l_0(\boldsymbol\theta_0;n)$, then we can obtain that

$$\begin{aligned}
P(\hat k\notin\Delta_\epsilon)\quad&\le P\left\{\min_{k\notin\Delta_\epsilon}\mathrm{MIC}(k)\le\mathrm{MIC}\left(\boldsymbol\theta_0;\frac{n}{2}\right)\right\} \\
&=P\left\{\min_{k\notin\Delta_\epsilon}-2l_1(\widehat{\boldsymbol\theta}_L,\widehat{\boldsymbol\theta}_R;k)+\left(\frac{2k}{n}-1\right)^2\log n\le-2l_1\left(\boldsymbol\theta_0;\frac{n}{2}\right)\right\} \\
&=P\left\{\max_{k\notin\Delta_\epsilon}2l_1(\widehat{\boldsymbol\theta}_L,\widehat{\boldsymbol\theta}_R;k)-\left(\frac{2k}{n}-1\right)^2\log n\ge 2l_1\left(\boldsymbol\theta_0;\frac{n}{2}\right)\right\} \\
&\le P\left\{\max_{k\notin\Delta_\epsilon}l_1(\widehat{\boldsymbol\theta}_L,\widehat{\boldsymbol\theta}_R;k)-l_1\left(\boldsymbol\theta_0;\frac{n}{2}\right)\ge\frac{1}{2}\epsilon^2\log n\right\},
\end{aligned} \tag{B7}$$

where $\boldsymbol\theta_0=(\mu_0,\sigma_0)$, $\boldsymbol\theta_L=(\mu_L,\sigma_L)$, $\boldsymbol\theta_R=(\mu_R,\sigma_R)$, $(\widehat{\boldsymbol\theta}_L,\widehat{\boldsymbol\theta}_R)$ is the MLE of $(\boldsymbol\theta_L,\boldsymbol\theta_R)$. From Lemma 1 in Chen, Gupta, and Pan (2006), we can naturally come to the conclusion that

$$\max_{k\notin\Delta_\epsilon}\left[l_1(\widehat{\boldsymbol\theta}_L,\widehat{\boldsymbol\theta}_R;k)-l_1\left(\boldsymbol\theta_0,\boldsymbol\theta_0;\frac{n}{2}\right)\right]\le o_p(\log\log n),$$

thus $P(\hat k\notin\Delta_\epsilon)\to 0$. This proves Lemma 3.4. ■

Appendix C: Proof of theorem

Proof of Theorem 3.2. Lemma 3.4 implies that the range of $\frac{k}{n}$ can be restricted to an arbitrarily small neighborhood of $\frac{1}{2}$. If k is in such a neighborhood, we have $\min\frac{(k,n-k)}{\sqrt{n}}\to\infty$. According

Lemma 3.3, we consider $\boldsymbol{\theta}_L = (\mu_L, \sigma_L)$, $\boldsymbol{\theta}_R = (\mu_R, \sigma_R)$, $\boldsymbol{\theta} = (\mu, \sigma)$ which are in a small neighborhood of true value $\boldsymbol{\theta}_0 = (\mu_0, \sigma_0)$, that is, $\mathcal{M}_\delta = \{\boldsymbol{\theta} : |\boldsymbol{\theta} - \boldsymbol{\theta}_0| < \delta\}$, where $\delta > 0$. Then the test statistic M_n can be exprressed as

$$\begin{aligned} M_n &= \mathrm{MIC}(n) - \min_{k\in\Delta_\epsilon}\mathrm{MIC}(k) + 2\log n \\ &= \max_{k\in\Delta_\epsilon}\{\mathrm{MIC}(n) - \mathrm{MIC}(k)\} + 2\log n\} \\ &= \max_{k\in\Delta_\epsilon}\left\{2l_1(\boldsymbol{\theta}_L, \boldsymbol{\theta}_R) - 2l_0(\boldsymbol{\theta}) - \left(\frac{2k}{n} - 1\right)^2 \log n\right\} \\ &\leq 2\max_{k\in\Delta_\epsilon}\left\{\sup_{\boldsymbol{\theta}_L\in\mathcal{M}_\delta}\sum_{i=1}^{k}\log f(X_i;\boldsymbol{\theta}_L) + \sup_{\boldsymbol{\theta}_R\in\mathcal{M}_\delta}\sum_{i=k+1}^{n}\log f(X_i;\boldsymbol{\theta}_R)\right. \\ &\quad \left. - \sup_{\boldsymbol{\theta}\in\mathcal{M}_\delta}\sum_{i=1}^{n}\log f(X_i;\boldsymbol{\theta})\right\} + o_p(1). \end{aligned} \tag{C1}$$

where

$$f(X_i;\boldsymbol{\theta}) = \frac{1}{X_i\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2}\left[\log X_i - (\mu+\sigma^2)\right]^2\right\}.$$

Then by Taylor expansion for $\log f(X_i;\boldsymbol{\theta})$ at $\boldsymbol{\theta}_0 = (\mu_0, \sigma_0)$, we have

$$\log f(X_i;\boldsymbol{\theta}) - \log f(X_i;\boldsymbol{\theta}_0) = S(X_i;\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T + \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)H(\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T + \frac{1}{6}M\delta^3 + o_p(1),$$

where $(\boldsymbol{\theta}-\boldsymbol{\theta}_0) = (\mu-\mu_0, \sigma-\sigma_0)$, $S(X_i;\boldsymbol{\theta}) = \left(\frac{\partial \log f(X_i;\boldsymbol{\theta})}{\partial\mu}, \frac{\partial \log f(X_i;\boldsymbol{\theta})}{\partial\sigma}\right)$ is the score function of $f(X_i;\boldsymbol{\theta})$ and

$$\begin{cases} \dfrac{\partial \log f(X_i;\boldsymbol{\theta})}{\partial\mu} = \dfrac{1}{\sigma^2}\left(\log X_i - (\mu+\sigma^2)\right), \\ \dfrac{\partial \log f(X_i;\boldsymbol{\theta})}{\partial\sigma} = -\dfrac{1}{\sigma} + \dfrac{1}{\sigma^3}\left(\log X_i - (\mu+\sigma^2)\right)^2 + \dfrac{2}{\sigma}\left(\log X_i - (\mu+\sigma^2)\right), \end{cases}$$

$$H(\boldsymbol{\theta}) = \begin{pmatrix} \dfrac{\partial^2 \log f(X_i;\theta)}{\partial\mu^2} & \dfrac{\partial^2 \log f(X_i;\boldsymbol{\theta})}{\partial\mu\partial\sigma} \\ \dfrac{\partial^2 \log f(X_i;\boldsymbol{\theta})}{\partial\sigma\partial\mu} & \dfrac{\partial^2 \log f(X_i;\boldsymbol{\theta})}{\partial\sigma^2} \end{pmatrix},$$

the elements of $H(\boldsymbol{\theta})$ are listed as follows,

$$\begin{cases} \dfrac{\partial^2 \log f(X_i;\boldsymbol{\theta})}{\partial\mu^2} = -\dfrac{1}{\sigma^2}, \\ \dfrac{\partial^2 \log f(X_i;\boldsymbol{\theta})}{\partial\mu\partial\sigma} = -\dfrac{2}{\sigma^3}\left[\log X_i - (\mu+\sigma^2)\right] - \dfrac{2}{\sigma}, \\ \dfrac{\partial^2 \log f(X_i;\boldsymbol{\theta})}{\partial\sigma\partial\mu} = -\dfrac{2}{\sigma^3}\left[\log X_i - (\mu+\sigma^2)\right] - \dfrac{2}{\sigma}, \\ \dfrac{\partial^2 \log f(X_i;\boldsymbol{\theta})}{\partial\sigma^2} = \dfrac{1}{\sigma^2} - \dfrac{3}{\sigma^4}\left[\log X_i - (\mu+\sigma^2)\right]^2 - \dfrac{6}{\sigma^2}\left[\log X_i - (\mu+\sigma^2)\right] - 4. \end{cases}$$

Then the Eq. (C1) can be further expressed as

$$\begin{aligned} M_n \quad &\leq \max_{k\in\Delta_\epsilon}\Bigg\{\sup_{\boldsymbol{\theta}_L\in\mathcal{M}_\delta}\left[2\sum_{i=1}^{k}S(X_i;\boldsymbol{\theta}_0)(\boldsymbol{\theta}_L-\boldsymbol{\theta}_0)^T+(\boldsymbol{\theta}_L-\boldsymbol{\theta}_0)\sum_{i=1}^{k}H(\boldsymbol{\theta}_0)(\boldsymbol{\theta}_L-\boldsymbol{\theta}_0)^T\right] \\ &+\sup_{\boldsymbol{\theta}_R\in\mathcal{M}_\delta}\left[2\sum_{i=k+1}^{n}S(X_i;\boldsymbol{\theta}_0)(\boldsymbol{\theta}_R-\boldsymbol{\theta}_0)^T+(\boldsymbol{\theta}_R-\boldsymbol{\theta}_0)\sum_{i=k+1}^{n}H(\boldsymbol{\theta}_0)(\boldsymbol{\theta}_R-\boldsymbol{\theta}_0)^T\right] \\ &-\sup_{\boldsymbol{\theta}\in\mathcal{M}_\delta}\left[2\sum_{i=1}^{n}S(X_i;\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T+(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\sum_{i=1}^{n}H(\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T\right]\Bigg\}+o_p(1). \end{aligned}$$

Let $I(\boldsymbol{\theta})=-E(\frac{\partial S(X_i;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}})$, and $I(\boldsymbol{\theta})$ is the Fisher information matrix of length-biased lognormal distribution, then

$$I(\boldsymbol{\theta})=-E\begin{pmatrix}\dfrac{\partial^2\log f(X_i;\boldsymbol{\theta})}{\partial\mu^2} & \dfrac{\partial^2\log f(X_i;\boldsymbol{\theta})}{\partial\mu\partial\sigma} \\ \dfrac{\partial^2\log f(X_i;\boldsymbol{\theta})}{\partial\sigma\partial\mu} & \dfrac{\partial^2\log f(X_i;\boldsymbol{\theta})}{\partial\sigma^2}\end{pmatrix}=\begin{pmatrix}\dfrac{1}{\sigma^2} & \dfrac{2}{\sigma} \\ \dfrac{2}{\sigma} & \dfrac{2}{\sigma^2}+4\end{pmatrix}.$$

It is obvious that $I(\boldsymbol{\theta})$ is nonsingular. Using the positive definite matrix $I(\boldsymbol{\theta}_0)$ and the property of the quadratic function, then we have

$$\begin{aligned} M_n \quad &\leq \max_{k\in\Delta_\epsilon}\Bigg\{\left(\frac{1}{\sqrt{k}}\sum_{i=1}^{k}S(X_i;\boldsymbol{\theta}_0)\right)I^{-1}(\boldsymbol{\theta}_0)\left(\frac{1}{\sqrt{k}}\sum_{i=1}^{k}S(X_i;\boldsymbol{\theta}_0)^T\right) \\ &+\left(\frac{1}{\sqrt{n-k}}\sum_{i=k+1}^{n}S(X_i;\boldsymbol{\theta}_0)\right)I^{-1}(\boldsymbol{\theta}_0)\left(\frac{1}{\sqrt{n-k}}\sum_{i=k+1}^{n}S(X_i;\boldsymbol{\theta}_0)^T\right) \\ &-\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}S(X_i;\boldsymbol{\theta}_0)\right)I^{-1}(\boldsymbol{\theta}_0)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}S(X_i;\boldsymbol{\theta}_0)^T\right)\Bigg\}+o_p(1). \end{aligned}$$

Let $W_k=\sum_{i=1}^{k}I^{-\frac{1}{2}}(\boldsymbol{\theta}_0)S(X_i;\boldsymbol{\theta}_0)^T$, $W_n=\sum_{i=1}^{n}I^{-\frac{1}{2}}(\boldsymbol{\theta}_0)S(X_i;\boldsymbol{\theta}_0)^T$, then

$$\begin{aligned} M_n &\leq \max_{k\in\Delta_\epsilon}\Big\{\frac{1}{\sqrt{k}}W_k^TW_k\frac{1}{\sqrt{k}}+\frac{1}{\sqrt{n-k}}(W_n-W_k)^T(W_n-W_k)\frac{1}{\sqrt{n-k}} \\ &-\frac{1}{\sqrt{n}}W_n^TW_n\frac{1}{\sqrt{n}}\Big\}+o_p(1) \\ &=\max_{k\in\Delta_\epsilon}\left\{\frac{n(n-k)W_k^TW_k+nk(W_n-W_k)^T(W_n-W_k)-k(n-k)W_n^TW_n}{nk(n-k)}\right\}+o_p(1) \\ &=\max_{k\in\Delta_\epsilon}\left\{\left[n\cdot\frac{k}{n}\left(1-\frac{k}{n}\right)\right]^{-1}\left(W_k-\frac{k}{n}W_k\right)^T\left(W_k-\frac{k}{n}W_k\right)\left[n\cdot\frac{k}{n}\left(1-\frac{k}{n}\right)\right]^{-1}\right\} \\ &+o_p(1) \\ &\leq \max_{k\in\Delta_\epsilon}T_n(t)T_n^T(t)+o_p(1), \end{aligned} \tag{C2}$$

where

$$\begin{aligned} T_n(t)&=\left(\frac{[nt]}{n}\left(1-\frac{[nt]}{n}\right)\right)^{-\frac{1}{2}}n^{\frac{1}{2}}\left\{W_{[nt]}^T+(nt-[nt])S(X_{[nt]+1};\boldsymbol{\theta}_0)-\frac{[nt]}{n}W_n^T\right\} \\ &=(T_{n1}(t),T_{n2}(t)) \end{aligned}$$

From Donsker's theorem, for $t\in[\frac{1}{2}-\epsilon,\frac{1}{2}+\epsilon]$, as $n\to\infty$,

$$T_{ni}(t)\xrightarrow{D}[t(1-t)]^{-\frac{1}{2}}B_0(t),\qquad i=1,2, \tag{C3}$$

where $B_0(t)$ is a Brownian bridge. Consequently,

$$\begin{aligned} M_n &\le \sup_{|t-\frac{1}{2}|<\epsilon} T_{n1}^2(t) + T_{n2}^2(t) + o_p(1) \\ &\to \sup_{|t-\frac{1}{2}|<\epsilon} [t(1-t)]^{-1} B_0^2(t) + [t(1-t)]^{-1} B_0^2(t), \end{aligned} \tag{C4}$$

As $\epsilon \to 0$, $\sup_{|t-\frac{1}{2}|<\epsilon} \left| B_0(t) - B_0\left(\frac{1}{2}\right) \right| \to 0$, then for arbitrary small $\epsilon > 0$,

$$\left[\frac{1}{2}\left(1-\frac{1}{2}\right)\right]^{-1} B_0^2\left(\frac{1}{2}\right) \sim \chi^2(1).$$

Thus,

$$P(M_n \le s) \ge P(\chi^2(2) \le s - o_p(1)) \Rightarrow \underline{\lim}_{n\to\infty} P(M_n \le s) \ge P(\chi^2(2) \le s).$$

And because $S_n \ge \text{MIC}(n) - \text{MIC}\left(\frac{n}{2}\right) + 2\log n$, that is,

$$\overline{\lim_{n\to\infty}} P(M_n \le s) \le P(\chi^2(2) \le s),$$

Therefore, $M_n \xrightarrow{D} \chi^2(2)$ as $n \to \infty$. ■