Communications of the Korean Statistical Society Vol. 15, No. 3, 2008, pp. 421–427

Test and Estimation for Exponential Mean Change†

Jaehee Kim[1)]

Abstract

This paper deals with the problem of testing for the existence of change in mean and estimating the change-point when the data are from the exponential distributions. The likelihood ratio test statistic and Gombay and Horvath (1990) test statistic are compared in a power study when there exists one change-point in the exponential means. Also the change-point estimator using the likelihood ratio and the change-point estimators based on Gombay and Horvath (1990) statistic are compared for their detecting capability via simulation.

Keywords: Change-point; likelihood ratio; mean change; Ornstein-Uhlenbeck process; quantile process.

1. Introduction

The identification of changes in the trend is an important issue in the analysis of incidence data. Changes in distribution at an unknown time point arises in industrial life testing and medical studies on diseases. Many studies about change have been done and still of more interest in the statistical analysis of change-point detection and estimation.

In almost all classic statistical inference is based upon the assumption that there exists a fixed probabilistic mechanism of data generation. Unlike classic statistical inference, the parametric change analysis of data about the complex objects is considered. The existence of more than one data generation process is the most important characteristic of complex system. When the hypotheses of statistical homogeneity holds true, that is, there exists only one mechanism of data generation, the law of large numbers are applied to make an inference. However if there exists change in the data generation, the probabilistic law should be applied differently. In this case all data obtained should be sorted in subsamples generated by different probabilistic mechanisms. After this classification the correct inferences can be made. It is important to detect possible changes of data generation process and the appropriate statistical analysis of such data must begin with testing and decisions about possible change. There are some shifts of mean survival *etc.*

Gombay and Horvath (1990) considered the maximum likelihood tests for change in the mean of independent random variables and proved the limit distribution as a double exponential distribution. Chen and Gupta (2000) considered the parametric change analysis including normal, exponential, Poisson and binomial distributions. Ramanayake and

† This research is supported by 2007 Duksung Women's University Research Fund.
1) Associate Professor, Department of Statistics, Duksung Women's University, Seoul 132-714, Korea. E-mail: jaehee@duksung.ac.kr

Gupta (2003) proposed a likelihood ratio type test statistic for independent exponential random variables and applied the test for epidemic change. Zou *et al.* (2007) proposed the nonparametric method based on the empirical likelihood to detect change-point.

Exponential distributions are applied in the survival and reliability analysis. In this paper, we assume that the data are from exponential distributions. We consider tests for mean change and change-point estimation for when the data are from exponential distributions.

This paper is organized as follows. In section 2 an exponential mean change-point model is defined and the proposed method is derived with its statistical properties. Section 3 presents some numerical results including simulation with test statistics and change-point estimators. Finally section 4 concludes the paper with a discussion of exponential change-point problems.

2. Exponential Mean Change Model

An exponential model is useful and appropriate in many experimental sciences. Let $X_1, X_2, \ldots, X_n$ be independent exponential random variables with parameters $\lambda_1, \lambda_2, \ldots,$ λ_n respectively.

2.1. Mean change

The hypotheses of interest are defined as

$$H_0 : \lambda_1 = \lambda_2 = \cdots = \lambda_n = \lambda \quad vs. \quad H_1 : \lambda_1 = \cdots = \lambda_k \neq \lambda_{k+1} = \cdots = \lambda_n, \tag{2.1}$$

where k is the unknown location of the single change-point. That is,

$$H_0 : X_1, X_2, \ldots, X_n \quad iid \quad f(x : \lambda) = \lambda e^{-\lambda x}$$

and

$$\begin{aligned} H_1 : X_1, X_2, \ldots, X_k \quad & iid \quad f(x; \lambda_1) = \lambda_1 e^{-\lambda_1 x}, \\ X_{k+1}, X_{k+2}, \ldots, X_n \quad & iid \quad f(x; \lambda_n) = \lambda_n e^{-\lambda_n x}, \quad 1 \leq k \leq n-1. \end{aligned}$$

2.2. Test for mean change

Under H_0, the likelihood function is

$$L_0(\lambda) = \prod_{i=1}^{n} f(X_i; \lambda) = \lambda^n e^{-\lambda \sum_{i=1}^{n} X_i}$$

and the maximum likelihood estimator(mle) of λ is $\hat{\lambda} = 1/\bar{X} = n/\sum_{i=1}^{n} X_i$. Under H_1, the likelihood function is

$$L_1(\lambda) = \prod_{i=1}^{n} f(X_i; \lambda_i) = \lambda_1^k \lambda_n^{n-k} e^{-\lambda_1 \sum_{i=1}^{k} X_i - \lambda_n \sum_{i=k+1}^{n} X_i}$$

and the *mle*'s of λ_1 and λ_n are respectively,

$$\hat{\lambda}_1 = \frac{1}{\bar{X}_k} = \frac{k}{\sum_{i=1}^{k} X_i}, \qquad \hat{\lambda}_n = \frac{1}{\bar{X}_{n-k}} = \frac{n-k}{\sum_{i=k+1}^{n} X_i}. \tag{2.2}$$

Hence, the maximum likelihood-ratio procedure test statistic is

$$\Lambda = \frac{L_0(\hat{\lambda})}{L_1(\hat{\lambda}_1, \hat{\lambda}_n)} \tag{2.3}$$

and

$$2\log\Lambda = 2\log\left[\left(\frac{k}{n}\right)^{-k}\left(\frac{\sum_{i=1}^{k} X_i}{\sum_{i=1}^{n} X_i}\right)^{k}\left(1-\frac{k}{n}\right)^{-(n-k)}\cdot\left(1-\frac{\sum_{i=1}^{k} X_i}{\sum_{i=1}^{n} X_i}\right)^{n-k}\right].$$

Let

$$\beta_n(x;k) = \frac{\sum_{i=1}^{k} X_i}{\sum_{i=1}^{n} X_i} \quad \text{and} \quad r_n(k) = \frac{k}{n}. \tag{2.4}$$

And it is easy to see that $\beta_n(x;k) \sim \text{beta}(k, n-k)$, beta distribution. Let $U_n(k)$ be the k^{th} order statistic of a random sample of n from $u(0,1)$. Then the *pdf* of $U_n(k)$ is

$$f_{U_n(k)}(u) = \frac{n!}{(k-1)!(n-k)!}u^{k-1}(1-u)^{n-k}, \quad 0 < u < 1,$$

which is the *pdf* of beta$(k, n-k)$ distribution under H_0. Therefore $\beta_n(x;k) \overset{d}{=} dU_n(k)$. Then, consider the log-likelihood as

$$2\log\Lambda = 2\log\left[r_n^{-k}(k)\beta_n^k(x;k)\{1-\beta_n(x;k)\}^{n-k}\{1-r_n(k)\}^{-(n-k)}\right].$$

Let

$$\begin{aligned} f_n(x;k) &= -2\log\Lambda \\ &= 2n\left[-r_n(k)\log\left[\frac{\beta_n(x;k)}{r_n(k)}\right] - \{1-r_n(k)\}\log\left[\frac{1-\beta_n(x;k)}{1-r_n(k)}\right]\right]. \end{aligned}$$

Chen and Gupta (2000) showed that by the second order Taylor expansion of $f_n(x;k)$ at the point $r_n(k)$ while viewing $f_n(x;k)$ as a function of $\beta_n(x;k)$ and $r_n(k)$ converges in distribution to $\tilde{f}_n(x;k)$.

The likelihood based test statistic given in Chen and Gupta (2000) is then given by

$$V = \max_{1\le k<n} f_n(x;k) \overset{d}{=} \max_{0\le y\le 1} \tilde{f}_n(y). \tag{2.5}$$

Here

$$\tilde{f}_n(y) = \left\{ \frac{n^{\frac{1}{2}}\left(U_n(y) - \frac{k}{n}\right)}{\frac{k}{n}\left(1-\frac{k}{n}\right)} \right\}^2 \{1 + R_n(y)\}, \quad 0 \le y \le 1,$$

where $R_n(y)$ is the remainder term in Taylor expansion and

$$U_n(y) = \begin{cases} U_n(k), & \text{for } \frac{k-1}{n} < y \le \frac{k}{n}, \\ 0, & \text{for } y = 0. \end{cases}$$

The likelihood test rejects H_0 if $V > c$. Note that the uniform quantile process is defined as

$$\tilde{U}_n(y) = n^{\frac{1}{2}}\{U_n(y) - y\}, \quad 0 \le y \le 1 \tag{2.6}$$

and the asymptotic distribution of $f_n(x;k)$ can be derived with the limit theorems concerning uniform quantile process.

2.3. Proposed change-point estimation for mean change

Based on the likelihood, the change-point can be estimated as

$$\hat{k}_{LRT} = \arg\max_{1\le i<n} f_n(x;i). \tag{2.7}$$

Gombay and Horvath (1990) test is developed as the function of *mle*'s. They consider the test statistic based on

$$Z_k = 2\left\{kg(\bar{X}_k) + (n-k)g(\bar{X}_{n-k}) - ng(\bar{X}_n)\right\}, \tag{2.8}$$

where

$$\bar{X}_k = \frac{1}{\hat{\lambda}_1} = \frac{1}{k}\sum_{i=1}^{k} X_i, \qquad \bar{X}_{n-k} = \frac{1}{\hat{\lambda}_n} = \frac{1}{n-k}\sum_{i=k+1}^{n} X_i.$$

Here the function g is a convex or concave function with which $g^{(2)}(\mu)$ is not near zero. For the hypotheses (1), their test rejects H_0 in favor of H_1 for large values of

$$T_{GH} = Z(i,j) = \max_{i<m<j} \frac{|Z_m|}{g^{(2)}(\mu)}, \tag{2.9}$$

where $g^{(2)}$ is the second derivative of g and for suitably chosen i and j.

Note that the maximum occurs at the change-point when there is a change-point. Therefore we consider the change-point estimation based on Gombay and Horvath (1990) test which has a functional form of the maximum likelihood as follows:

$$\hat{k}_{GH} = \arg\max_{1\le i<m<j<n} \frac{|Z_m|}{g^{(2)}(\mu)} \tag{2.10}$$

where $\hat{k}_{GH1}$ with $g_1(t) = t^2$ and $\hat{k}_{GH2}$ with $g_2(t) = \exp(t)$ for T_{GH} in (10).

Gombay and Horvath (1990) showed that the limiting distribution of their test is

$$\frac{Z(m_1, m_2)}{\sigma^2} \rightarrow \sup_{0 \leq s \leq \Lambda} |V(s)|, \tag{2.11}$$

in distribution, where $0 < \lambda_1 \leq 1 - \lambda_2 < 1$ as $n \rightarrow \infty$, $m_1 = n\lambda_1$, $m_2 = n(1 - \lambda_2)$, $\Lambda = 1/2\{\log(1 - \lambda_1)(1 - \lambda_2)/\lambda_1\lambda_2\}$ and $\{V(s), -\infty < s < \infty\}$ is an Ornstein-Uhlenbeck process, *i.e.* a Gaussian process with mean zero and covariance $\exp(-|t - s|)$. Therefore the distribution of the change-point can be shown as

$$\text{argmax}_{m_1, m_2} \frac{Z(m_1, m_2)}{\sigma^2} \rightarrow \text{argmax} \left\{ s : \sup_{0 \leq s \leq \Lambda} |V(s)| \right\}, \quad \text{as} \quad n \rightarrow \infty. \tag{2.12}$$

3. Simulation

To assess and compare the performance of the procedure, we simulated samples of data sets. A simulation study is conducted to see the power of the likelihood ratio test with level according to the sample size, the amount of change and the location of change. The parametric test is compared with the Gombay and Horvath (1990) test as the function of *mle*'s. For the change tests, the LRT based test U_{LRT} and the Gombay and Horvath (1990) tests T_{GH1} and T_{GH2} with $g_1(t) = t^2$, $g_2(t) = \exp(t)$ are compared in power study. Several g functions were tried but g_1, g_2 gave better results. For the change-point estimation, the ability to detect the change-point is studied with calculation of bias and MSE(mean square error). A random sample $X_1, X_2, \ldots, X_n$ are generated from the exponential distributions with the parameter $\lambda_1, \lambda_2, \ldots, \lambda_n$. The exponential mean level change model with one change-point is as follows:

$$\lambda_i = \begin{cases} \lambda_1, & i = 1, \ldots, k, \\ \lambda_n = \lambda_1 + \Delta, & i = k + 1, \ldots, n, \end{cases} \tag{3.1}$$

where $\lambda_1 = 1$ without loss of generality. The amount of change $\Delta = -0.5, 0, 0.5, 1, 1.5$ for power study and $\Delta = -0.5, 0.5, 1, 1.5$ for estimator comparison study, the sample size $n = 50$ and the location of change at $k/n = 0.3, 0.5, 0.8$ are considered. The repetition $r = 1,000$ were used in this simulation. The range of the points is restricted from 5^{th} to 45^{th} point due to boundary consideration. For the power study, critical values were evaluated from the empirical distribution with 10,000 repetitions and used. Table 3.1 gives the simulation results as the power of test for the mean change. The LRT, U_{LRT} has more power than T_{GH1}, T_{GH2} when the exponential rate is decreasing *i.e.* the mean is increasing after the change-point and U_{LRT} has less power than T_{GH1}, T_{GH2} in the increasing exponential rate.

For the comparison of change-point estimators, mean and mse of each estimator were calculated in Table 3.2. The change-point estimator with U_{LRT} is best when the change-point occurs in the former part or in the middle of data sequence with decreasing exponential rate (increasing mean). The change-point estimators with T_{GH1}, T_{GH2} have smaller MSE when the change-point occurs in the former part or in the middle with increasing exponential rate (decreasing mean) or when the change-point occurs in the

Table 3.1: Power comparison study of Change tests in Exponential distribution with the sample size with $n = 50$ in 1,000 repetitions

	change-pt	k=15		k=25		k=40	
	Test	α=0.10	α=0.05	α=0.10	α=0.05	α=0.10	α=0.05
$\lambda_1 = 1.0$ $\lambda_n = 0.5$	U_{LRT}	0.475	0.354	0.595	0.459	0.394	0.289
	T_{GH1}	0.827	0.690	0.828	0.728	0.604	0.482
	T_{GH2}	0.850	0.683	0.828	0.684	0.587	0.442
$\lambda_1 = 1.0$ $\lambda_n = 1.0$	U_{LRT}	0.071	0.040	0.077	0.038	0.071	0.039
	T_{GH1}	0.062	0.027	0.072	0.036	0.065	0.026
	T_{GH2}	0.055	0.021	0.054	0.017	0.045	0.019
$\lambda_1 = 1.0$ $\lambda_n = 1.5$	U_{LRT}	0.238	0.145	0.219	0.146	0.176	0.103
	T_{GH1}	0.097	0.052	0.095	0.042	0.055	0.023
	T_{GH2}	0.052	0.017	0.043	0.015	0.027	0.013
$\lambda_1 = 1.0$ $\lambda_n = 2.0$	U_{LRT}	0.504	0.387	0.561	0.449	0.338	0.215
	T_{GH1}	0.150	0.072	0.190	0.084	0.092	0.034
	T_{GH2}	0.056	0.016	0.061	0.016	0.038	0.014
$\lambda_1 = 1.0$ $\lambda_n = 2.5$	U_{LRT}	0.740	0.632	0.803	0.696	0.525	0.407
	T_{GH1}	0.201	0.108	0.260	0.126	0.125	0.048
	T_{GH2}	0.075	0.042	0.071	0.016	0.043	0.016

Table 3.2: Comparison of Change-point Estimators with $n = 50$, $k = 15$, $k = 25$ and $k = 40$ in 1,000 Repetitions in the Exponential distribution

	change-pt	k=15		k=25		k=40	
	Estimator	Mean	Mse	Mean	Mse	Mean	Mse
$\lambda_1 = 1.0$ $\lambda_n = 0.5$	$\hat{k}_{LRT}$	19.012	104.764	25.621	75.921	32.810	206.840
	$\hat{k}_{GH1}$	23.136	182.142	29.314	85.002	35.854	140.496
	$\hat{k}_{GH2}$	26.195	258.851	31.295	110.235	37.083	120.507
$\lambda_1 = 1.0$ $\lambda_n = 1.5$	$\hat{k}_{LRT}$	21.167	198.823	24.206	134.878	28.365	307.689
	$\hat{k}_{GH1}$	18.737	162.329	21.838	132.014	24.884	399.722
	$\hat{k}_{GH2}$	17.778	152.082	20.794	138.468	23.784	429.880
$\lambda_1 = 1.0$ $\lambda_n = 2.0$	$\hat{k}_{LRT}$	17.777	112.587	24.327	74.671	32.449	190.237
	$\hat{k}_{GH1}$	14.769	77.649	20.593	85.577	28.288	293.874
	$\hat{k}_{GH2}$	14.087	73.831	19.321	100.349	26.692	338.738
$\lambda_1 = 1.0$ $\lambda_n = 2.5$	$\hat{k}_{LRT}$	15.953	56.663	24.187	44.689	35.286	113.350
	$\hat{k}_{GH1}$	13.607	38.393	20.811	65.491	30.099	236.261
	$\hat{k}_{GH2}$	13.122	38.884	19.520	83.940	27.977	301.313

latter part with decreasing rate after change-point. The estimation ability depends on the location of the change-point and the direction of exponential rate change. Because exponential distributions are not symmetric about the mean, test and estimation for change are affected by the direction of change, increasing or decreasing.

4. Concluding Remarks

We considered testing and estimating problems of change-point when the observations are from exponential distributions. The numerical results lend support to the argument that the likelihood ratio test and change-point estimation are not always best even under the parametric distributional assumptions. But the function of the maximum likelihood estimator could play a role in change-point estimation. More powerful tests based on other functions is expected as a further research with effectively reflecting the properties of exponential distributions. Also the numerical results show that testing and estimating depend on the location of change-point. Therefore one possible conclusion is that one should choose a test statistic and an estimator on a subjective basis depending on where one expects a change to take place.

References

Chen, J. and Gupta, A. K. (2000). *Parametric Statistical Change Point Analysis*, Birkhäuser, Boston.

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

Ramanayake, A. and Gupta, A. K. (2003). Tests for an epidemic change in a sequence of exponentially distributed random variables, *Biometirical Journal*, **45**, 946–958.

Zou, C., Liu, Y., Qin, P. and Wang, Z. (2007). Empirical likelihood ratio test for the change-point problem, *Statistics & Probability Letters*, **77**, 374–382.

[Received September 2007, Accepted April 2008]