Journal of the Korean
Statistical Society
Vol. 25, No. 4, 1996

Bayes Factor for Change-point Problem with Conjugate Prior

Younshik Chung[1] and Dipak K. Dey[2]

Abstract

The Bayes factor provides a possible hierarchical Bayesian approach for studying the change point problems. A hypothesis for testing change versus no change is considered using predictive distributions. When the underlying distribution is in one-parameter exponential family with conjugate priors, Bayes factors are investigated to the hypothesis above. Finally one example is provided.

Key Words : Bayes factor; One-parameter exponential family; Conjugate prior; Change-point;

1. INTRODUCTION

Given a sequence of random variables, suppose at some unknown point in the sequence the process governing their probability distribution changes, and consider the problem of detecting the change inference concerning the change point. There is enormous literature on this and related problems for different types of statistical models both in frequentist and Bayesian literature.

[1]Department of Statistics, Pusan National University, Pusan, 609-735, Korea.
[2]Department of Statistics, University of Connecticut, Storrs, C.T. 06269 U.S.A.

Smith(1975) presents the Bayesian formulation for a finite sequence of independent observations and Carlin et. al(1992) present a hierarchical Bayesian analysis to this problem using Gibbs sampler. Frequentist approaches are studied in Hinkley(1970), Siegmund(1986) and Wolfe and Schechtman(1984). Zacks(1983) gives an extremely useful review on classical and Bayesian approach on this topic. Our goal is not only on a Bayesian parametric approach but also considering it as a model selection problem. It is considered that problem of comparing the change point model with one where there is no change. This approach was first considered in Raftery and Akman(1986) in the context of Poisson process. In this paper we consider the use of Bayes factor to detect the changes. We assume that the underlying probability function arises from one-parameter exponential family. It follows that the Bayes factors are expressible in terms of convex combination of Bayes factor for the independent problem of deciding change at any given sample point versus no change at that point. Thus, a graphical display of Bayes factor at each sample point captures whether there is any change or not. As a consequence an estimate of the change point is obtained where the individual Bayes factor attains maximum which is compared with Pettit and Young(1990)'s measure k_d in (4.1). In section 2, we present a semi-hierarchical Bayesian formulation of the problem using predictive distribution. Secton 3 is devoted the Bayes factor calculation for one-parameter exponential family models. Section 4 investigates measuring the effect of observations on Bayes factor in our form. Finally, in section 5 the method is applied to determine the change point of British Coal-mining disaster data.

2. HIERARCHICAL BAYESIAN FORMULATION

In this section, we consider that the underlying distribution is in the one-parameter exponential family as follows;

$$f(x|\theta) = exp\{\theta x - M(\theta)\}t(x). \tag{2.1}$$

Suppose that $X_1, \ldots, X_n$ is a random sample from the distribution (2.1). If there is any change in the distributions of X_i's, we assume that for some r with $1 \leq r < n$, $X_1, \cdots, X_r$ is a random sample from $f(x|\theta_1)$ and $X_{r+1}, \cdots, X_n$ is a random sample from $f(x|\theta_2)$ where $\theta_1 \neq \theta_2$. On the other hand if there is no change in distributions of X_i's, then $X_1, \cdots, X_n$ is a random sample from $f(x|\theta_1)$. Thus, the parameter space under consideration is

$$\{(\theta_1, \theta_2, r) : 1 \le r \le n-1\} \cup \{(\theta_1, n)\}.$$

Assume that given a fixed $r \le n-1$, θ_1 and θ_2 are independent and having prior distribution conjugate to (2.1) and for $r = 1$, θ_1 has a conjugate prior. Following Diaconis and Ylvisaker(1979), the form of the prior conjugate to (2.1) is given as

$$\pi_{\alpha,\beta}(\theta) = C(\alpha, \beta) exp \left[\frac{\beta\theta}{\alpha} - \frac{1-\alpha}{\alpha} M(\theta) \right] \tag{2.2}$$

where $M(\theta)$ is a known function.

We further assume a prior distribution is given by

$$f[r = j | \mathbf{P}] = \prod_{j=1}^{n} P_j^{I[r=j]} \tag{2.3}$$

where $\mathbf{P} = (P_1, \ldots, P_n)$ and $P_i \ge 0$, $\sum_{i=1}^{n} P_i = 1$.

By hierarchical structure, assume that at the third stage, the prior of $\mathbf{P}$ is distributed as Dirichlet as follows;

$$f(\mathbf{P}) = m(l_1, \ldots, l_n) \prod_{i=1}^{n} P_i^{l_i - 1} \tag{2.4}$$

where $m(l_1, \ldots, l_n)^{-1} = \int \prod_{i=1}^{n} P_i^{l_i-1} d\mathbf{P}$.

We also assume that the prior on r is independent of that of θ_1 and θ_2. Thus, the likelihood and prior specification reduces to

$$\begin{aligned} f(x_1, \ldots, x_n | \theta_1, \theta_2, r) &= exp \left\{ \theta_1 \sum_{i=1}^{r} x_i - rM(\theta_1) \right. \\ &\quad \left. + \theta_2 \sum_{i=r+1}^{n} x_i - (n-r)M(\theta_2) \right\} \prod_{i=1}^{n} t(x_i) \end{aligned}$$

and

$$\begin{aligned} \pi(\theta_1, \theta_2, r) &= \pi_{\alpha_1,\beta_1}(\theta_1)\pi_{\alpha_2,\beta_2}(\theta_2) \int P_r \cdot m(l_1, \ldots, l_n) \prod_{i=1}^{n} P_i^{l_i-1} d\mathbf{P} \\ &= \frac{\pi_{\alpha_1,\beta_1}(\theta_1)\pi_{\alpha_2,\beta_2}(\theta_2) m(l_1, \ldots, l_n)}{m(l_1, \ldots, l_{r-1}, l_r + 1, l_{r+1}, \ldots, l_n)} \\ &= \pi_{\alpha_1,\beta_1}(\theta_1)\pi_{\alpha_2,\beta_2}(\theta_2) m_r \end{aligned}$$

where $m_r = \frac{m(l_1,\ldots,l_n)}{m(l_1,\ldots,l_{r-1},l_r+1,l_{r+1},\ldots,l_n)}$, for $r = n$,

$$f(x_1,\ldots,x_n|\theta_1, r=n) = exp\left[\theta_1\sum_{i=1}^{n} x_i - nM(\theta_1)\right]\prod_{i=1}^{n} t(x_i)$$

and

$$\pi(\theta_1, n) = \pi_{\alpha_1,\beta_1}(\theta_1)\cdot m_n \ .$$

3. BAYESIAN MODEL CHOICE

Suppose that we are interested in comparing two models M_0 and M_1. The formal Bayesian model choice procedure goes as follows. Let w_i be the prior probability of M_i, $i = 0, 1$ and let $f(x|M_i)$ be the predictive distribution for model M_i, i.e.

$$f(x|M_i) = \int f(x|\theta_i, M_i)\pi(\theta_i|M_i)d\theta_i \ .$$

If x is the observed data, then we choose the model yielding the larger $w_i f(x|M_i)$. Often we set $w_i = \frac{1}{2}$ and compute the Bayes factor (or M_0 with respect to M_1)

$$BF = \frac{f(x|M_0)}{f(x|M_1)} \ . \tag{3.1}$$

Jeffreys'(1961) suggests interpretive ranges for the Bayes factor as follows:

Table 1. Scale of evidence for assessing BF

Range	Evidence
$1 < BF$	Supports M_0
$10^{-\frac{1}{2}} < BF < 1$	Slight evidence against M_0
$10^{-1} < BF < 10^{-\frac{1}{2}}$	Moderate evidence against M_0
$10^{-2} < BF < 10^{-1}$	Strong evidence against M_0
$BF < 10^{-2}$	Decisive evidence against M_0

In this paper, we consider the problem of testing for a change as one comparing the change point model(M_0) with one(M_1) which assumes no change. We introduce the following notations:

$$S_r = \sum_{i=1}^{r} X_i \ , \quad T_r = \sum_{i=r+1}^{n} X_i \quad \text{for} 1 \le r \le n \ .$$

Obviously for $r = n$, $T_n = 0$. First note that

$$\begin{aligned}
f(x|M_0) &= f(x_1, x_2, \cdots, x_n|\text{change}) \\
&= \frac{f(x_1, x_2, \cdots, x_n, 1 \le r < n)}{f(1 \le r < n)} \\
&= \frac{\int\int f(x_1, x_2, \cdots, x_n|\theta_1, \theta_2, 1 \le r < n)\pi(\theta_1, \theta_2, 1 \le r < n)d\theta_1 d\theta_2}{f(1 \le r < n)} \\
&= \frac{\sum_{r=1}^{n-1} \int\int f(x_1, x_2, \cdots, x_n|\theta_1, \theta_2, r)\pi(\theta_1, \theta_2, r)\ \ d\theta_1\ d\theta_2}{f(1 \le r < n-1)} \\
&= \Bigg(C(\alpha_1, \beta_1)C(\alpha_2, \beta_2) \sum_{r=1}^{n-1} m_r \int\int \Bigg[exp\left(S_r + \frac{\beta_1}{\alpha_1}\right)\theta_1 \\
&\quad - \left(\frac{1-\alpha_1}{\alpha_1} + r\right) M(\theta_1) + \left(T_r + \frac{\beta_2}{\alpha_2}\right)\theta_2 \\
&\quad - \left(\frac{1-\alpha_2}{\alpha_2} + (n-r)\right) M(\theta_2)\Bigg] \\
&\quad \times \prod_{i=1}^{n} t(x_i) d\theta_1 d\theta_2\Bigg) \left(\sum_{j=1}^{n-1} m_j\right)^{-1}. \qquad (3.2)
\end{aligned}$$

Similarly, we can get

$$\begin{aligned}
f(x|M_1) &= f(x_1, \cdots, x_n|\text{no} \quad \text{change}) \\
&= C(\alpha_1, \beta_1) \int exp\Bigg\{\left(S_n + \frac{\beta_1}{\alpha_1}\right)\theta_1 \\
&\quad - \left(\frac{1-\alpha_1}{\alpha_1} + n\right) M(\theta_1)\Bigg\} \prod_{i=1}^{n} t(x_i)\ \ d\theta_1. \qquad (3.3)
\end{aligned}$$

Combining (3.2) and (3.3), we get the following expression for Bayes factor for change versus no changes as

$$BF = \sum_{r=1}^{n-1} \frac{m_r}{\sum_{i=1}^{n-1} m_i} BF(r), \qquad (3.4)$$

where

$$\begin{aligned}
BF(r) = \Bigg(\int\int exp\Bigg\{(S_r + \frac{\beta_1}{\alpha_1})\theta_1 - (\frac{1-\alpha_1}{\alpha_1} + r)M(\theta_1) + \left(T_r + \frac{\beta_2}{\alpha_2}\right)\theta_2 \\
- \quad \left(\frac{1-\alpha_2}{\alpha_2} + (n-r)\right) M(\theta_2)\Bigg\} d\theta_1\ d\theta_2\Bigg) \Bigg(\int\int exp\Bigg\{\left(S_n + \frac{\beta_1}{\alpha_1}\right)\theta_1
\end{aligned}$$

$$- \left(\frac{1-\alpha_1}{\alpha_1}+n\right) M(\theta_1) + \frac{\beta_2}{\alpha_2}\theta_2 - \frac{1-\alpha_2}{\alpha_2}M(\theta_2)\Big\} \, d\theta_1 \, d\theta_2\Big)^{-1}$$

since

$$C(\alpha_2,\beta_2)^{-1} = \int exp\left(\frac{\beta_2}{\alpha_2}\theta_2 - \frac{1-\alpha_2}{\alpha_2}M(\theta_2)\right) \; d\theta_2.$$

Note that from (3.3) $BF(r)$ itself is the Bayes factor for the independent problem of deciding "change at r" versus "no change at r". Also the overall Bayes factor is a convex combination of such individual Bayes factors. This Bayes factor indicates only whether the change happens or not.

3.1. Normal Distributions

Assume that $X_1,\cdots,X_r$ is a random sample from Normal distribution with unknown mean θ_1 and known variance σ^2 and $X_{r+1},\cdots,X_n$ is random sample from Normal distribution with unknown mean θ_2 and known variance σ^2. Without loss of generality assume that $\sigma^2 = 1$. Also under conjugacy assume $\theta_i \sim N(\mu_i, \tau_i^2)$ independently for $i = 1,2$ and all hyperparameters are known. Then defining

$$A(r) = (\frac{\tau_1^2+n}{\tau_1^2+r})^{\frac{1}{2}} \exp\{\frac{1}{2}(\frac{(S_r+\frac{\mu_1}{\tau_1^2})^2}{\tau_1^2+r} - \frac{(S_n+\frac{\mu_1}{\tau_1^2})^2}{\tau_1^2+n})\}$$

and

$$B(r) = (\frac{\tau_2^2}{\tau_2^2+n-r})^{\frac{1}{2}} \exp\{\frac{1}{2}(\frac{(T_r+\frac{\mu_2}{\tau_2^2})^2}{\tau_2^2+n-r} - \frac{\mu_2^2}{\tau_2^2})\},$$

it can be shown that the Bayes factor for change versus no change is given as

$$BF = \sum_{r=1}^{n-1} \frac{m_r}{\sum_{i=1}^{n-1} m_i} A(r)B(r). \tag{3.5}$$

Again the Bayes factor for change at r versus no change is given $BF(r) = A(r)B(r)$ and the overall Bayes factor is convex combination of $BF(r)$.

3.2. Poisson Distribution

Assume that $X_1,\cdots,X_r$ is a random sample from Poisson with mean λ_1 and $X_{r+1},\cdots,X_n$ is random sample from Poisson with mean λ_2. Also assume $\lambda_i \sim Gam(a_i, b_i)$ independently for $i = 1,2$. That is, $p(\lambda) = \frac{1}{\Gamma(a)b^a}\lambda^{a-1}\exp\{-\frac{\lambda}{b}\}$. Let $\theta_i = \log\lambda_i$ for $i = 1,2$. Then

$$BF = \sum_{r=1}^{n-1} \frac{m_r}{\sum_{i=1}^{n-1} m_i} BF(r) \tag{3.6}$$

where

$$BF(r) = \left(\frac{b_2}{b_1}\right)^{T_r} \frac{\Gamma(S_r + a_1)\Gamma(T_r + a_2)(nb_1 + 1)^{S_n + a_1}}{\Gamma(a_2)\Gamma(S_n + a_1)(rb_1 + 1)^{S_r + a_1}((n-r)b_2 + 1)^{T_r + a_2}}. \tag{3.7}$$

4. INFLUENCE ON BAYES FACTOR

If the change is happened, we are interested in which observation is seriously effected to that change. So, to measure the effect on the Bayes factor of observation d, Pettit and Young(1990) suggested the quantity k_d defined by

$$k_d = log_{10} BF - log_{10} BF^{(d)} \ , \tag{4.1}$$

where $BF^{(d)}$ is the Bayes factor excluding observation d. In different way,

$$\begin{aligned} k_d &= \log \frac{f(X|M_0)}{f(X|M_1)} - \log \frac{f(X_{(d)}|M_0)}{f(X_{(d)}|M_1)} \\ &= \log \frac{f(X|M_0)}{f(X_{(d)}|M_0)} - \log \frac{f(X|M_1)}{f(X_{(d)}|M_1)} \end{aligned} \tag{4.2}$$

where X is all the data and $X_{(d)}$ is the all the data excluding observation d.

This k_d is expressed as the difference in the logarithms of the conditional predictive ordinates(CPO) for the two models. Pettit(1990) mentioned that the CPO is a measure to detect surprising observations. Thus large values of $|k_d|$ indicate that such observation d has a large influence on the Bayes factor. In order to calculate k_d in (4.1), first compute $BF^{(d)}$ as follows; For convenience, let $BF_i^{(d)}(0) = 0$,

$$\begin{aligned} BF^{(d)} &= \frac{f(X_{(d)}|M_0)}{f(X_{(d)}|M_1)} \\ &= \sum_{r<d} \frac{m_r}{\sum_{k=1\ k\neq d}^{n-1} m_k} BF_1^{(d)}(r) + \sum_{r>d}^{n-1} \frac{m_r}{\sum_{k=1\ k\neq d}^{n-1} m_k} BF_2^{(d)}(r) \end{aligned} \tag{4.3}$$

where let

$$S_{r_1} = \sum_{i=1}^{r} X_i \ , \quad T_{r_1} = \sum_{i=r+1}^{n} X_i - X_d$$

$$S_{r_2} = \sum_{i=1}^{r} X_i - X_d \ , \quad T_{r_2} = \sum_{i=r+1}^{n} X_i, \quad S_n^{'} = \sum_{i=1}^{n} X_i - X_d$$

and

$$BF_i^{(d)}(r) = \frac{BF_{iN}^{(d)}}{BF_D^{(d)}}$$

for $i = 1, 2$,

$$\begin{aligned} BF_{iN}^{(d)}(r) &= \int\int exp(S_{r_i} + \frac{\beta_1}{\alpha_1})\theta_1 + (T_{r_i} + \frac{\beta_2}{\alpha_2})\theta_2 - (\frac{1-\alpha_1}{\alpha_1} + r)M(\theta_1) \\ &\quad -(\frac{1-\alpha_2}{\alpha_2} + (n-1-r))M(\theta_2) \ d\theta_1 \, d\theta_2 \end{aligned} \tag{4.4}$$

and

$$\begin{aligned} BF_D^{(d)}(r) &= \int\int exp\left\{(S_n^{'} + \frac{\beta_1}{\alpha_1})\theta_1 - (\frac{1-\alpha_1}{\alpha_1} + n - 1)M(\theta_1) \right. \\ &\quad \left. + \frac{\beta_2}{\alpha_2}\theta_2 - \frac{1-\alpha_2}{\alpha_2}M(\theta_2)\right\} \ d\theta_1 \, d\theta_2 \ . \end{aligned}$$

4.1. Normal Distributions

Assume that $X_1, \cdots, X_r$ is a random sample from Normal distribution with unknown mean θ_1 and known variance σ^2 and $X_{r+1}, \cdots, X_n$ is random sample from Normal distribution with unknown mean θ_2 and known variance σ^2. Without loss of generality assume that $\sigma^2 = 1$. Also under conjugacy assume $\theta_i \sim N(\mu_i, \tau_i^2)$ independently for $i = 1, 2$ and all hyperparameters are known. Then it can be shown that the Bayes factor for change versus no change is given as

$$\begin{aligned} BF^{(d)} &= \frac{f(X_{(d)}|M_0)}{f(X_{(d)}|M_1)} \\ &= \sum_{r<d} \frac{m_r}{\sum_{k=1 \ k\neq d}^{n-1} m_k} A_1(r)B_1(r) + \sum_{r>d}^{n-1} \frac{m_r}{\sum_{k=1 \ k\neq d}^{n-1} m_k} A_2(r)B_2(r) \end{aligned} \tag{4.5}$$

where

$$A_i(r) = \left(\frac{\tau_1^2 + (n-1)}{\tau_1^2 + (r+1-i)} \right)^{\frac{1}{2}} \exp\left\{ \frac{1}{2} \left(\frac{(S_{r_i} + \frac{\mu_1}{\tau_1^2})^2}{\tau_1^2 + (r+1-i)} - \frac{(S_n' + \frac{\mu_1}{\tau_1^2})^2}{\tau_1^2 + n - 1} \right) \right\}$$

and

$$B_i(r) = \left(\frac{\tau_2^2}{\tau_2^2 + n + 1 - i - r} \right)^{\frac{1}{2}} \exp\left\{ \frac{1}{2} \left(\frac{(T_{r_i} + \frac{\mu_2}{\tau_2^2})^2}{\tau_2^2 + n + 1 - i - r} - \frac{\mu_2^2}{\tau_2^2} \right) \right\}.$$

4.2. Poisson Distribution

Assume that $X_1, \cdots, X_r$ is a random sample from Poisson with mean λ_1 and $X_{r+1}, \cdots, X_n$ is random sample from Poisson with mean λ_2. Also assume $\lambda_i \sim Gam(a_i, b_i)$ independently for $i = 1, 2$. That is, $p(\lambda) = \frac{1}{\Gamma(a)b^a}\lambda^{a-1}\exp\{-\frac{\lambda}{b}\}$. Let $\theta_i = \log \lambda_i$ for $i = 1, 2$. Then

$$k_d = log_{10}BF - log_{10}BF^{(d)} \quad , \tag{4.6}$$

where

$$\begin{aligned} BF^{(d)} &= \frac{f(X_{(d)}|M_0)}{f(X_{(d)}|M_1)} \\ &= \sum_{r<d} \frac{m_r}{\sum_{k=1\ k\neq d}^{n-1} m_k} BF_1^{(d)}(r) + \sum_{r>d}^{n-1} \frac{m_r}{\sum_{k=1\ k\neq d}^{n-1} m_k} BF_2^{(d)}(r) \end{aligned} \tag{4.7}$$

and for $i = 1, 2$,

$$\begin{aligned} BF_i^{(d)}(r) &= \frac{\Gamma(S_{r_i} + a_1)\Gamma(T_{r_i} + a_2)((n-1)b_1 + 1)^{S_n' + a_1}}{\Gamma(a_2)\Gamma(S_n' + a_1)((r+1-i)b_1 + 1)^{S_{r_i} + a_1}} \\ &\times \frac{1}{((n+1-i-r)b_2 + 1)^{T_{r_i} + a_2}} \times \left(\frac{b_2}{b_1} \right)^{T_{r_i}}. \end{aligned} \tag{4.8}$$

5. ILLUSTRATIVE EXAMPLE

We now apply the results of the previous sections to data set consisting of intervals between coal-mining disasters given by Jarrett(1979). Rudems(1982) suggests that a change-point model may be appropriate. Also Carlin et.

al(1992) consider the hierarchical Bayesian approach to this using Gibbs sampler.

Assume that $X_1, \cdots, X_r$ is a random sample from Poisson with mean λ_1 and $X_{r+1}, \cdots, X_n$ is random sample from Poisson with mean λ_2. Also assume $\lambda_i \sim Gam(a_i, b_i)$ independently for $i = 1, 2$ and l_i is one for all i in (2.4). That is, let $\theta_i = \exp \lambda_i$ for $i = 1, 2$. Then

$$BF = \frac{1}{n-1} \sum_{r=1}^{n-1} BF(r) \tag{5.1}$$

where $BF(r)$ is in (3.7).

Then

$$BF = \frac{1}{n-1} \sum_{r=1}^{n-1} BF(r) \tag{5.2}$$

where $BF(r)$ is in (3.7).

With $a_1 = 2, a_2 = 1$ and $b_1 = b_2 = 1$, the Bayes factor for change versus no change is calculated using (5.1), and is turned out to be 6.69×10^{12}, which overwhelmingly supports the "change" model.

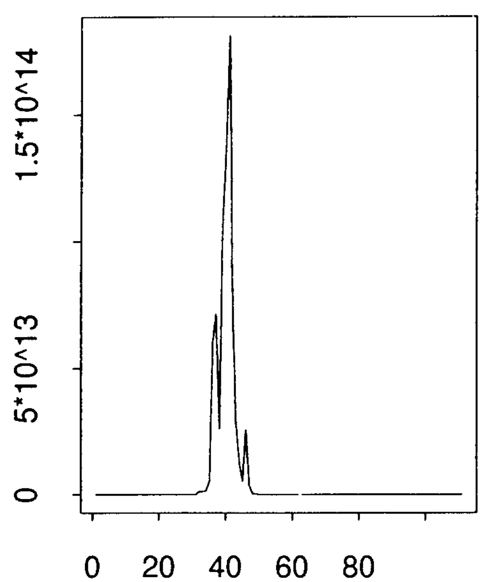

Figure 1 : Values of $BF(r)$

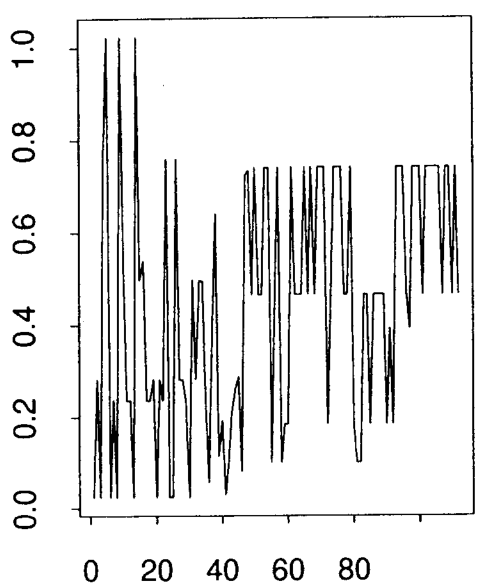

Figure 2 : Values of $|k_d|$

The individual Bayes factor $BF(r)$ are plotted in Figure 1 which indicates that the maximum change is at 41^{th} observation which corresponds to the year 1891. This result closely pararells to Carlin et. al(1992)'s result that the posterior mode of r is $r = 41$. So we can guess that the change may happen at 41^{th} observation. At the center of $41th$ observation, the first half of data' values are larger than the remaining data. And $|k_d|$'s values are plotted in Figure 2. The Figure 2 indicates that the values of $|k_5|, |k_9|$ and $|k_{14}|$ are relatively large and the corresponding value of data are all 0 in the first half of data. So those data can be outliers in the first half part.

REFERENCES

(1) Carlin, B. P., Gelfand, A. E. and Smith, A. F. M.(1992), Hierarchical Bayesian analysis of change point problems, *Applied Statistics*, **41**, 389-405.

(2) Diaconis, P. and Ylvisaker, D (1979), Conjugate priors for exponential families, *The Annals of Statistics*, **7**, 269-281.

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

(4) Jasrett, R. G.(1979), A note on the intervals between coal-mining disasters, *Biometrika*, **66**, 191-193.

(5) Jeffreys, H. (1961), *Theory of Probability,* 3rd ed, Oxford University press.

(6) Pettit, L. I. (1990), The conditional predictive ordinate for the normal distribution, *Journal of the Royal Statistical Society* B, **52**, 174-184.

(7) Pettit, L. I. and Young, K. D. S.(1990), Measuring the effect of observations on Bayes factors, *Biometrika*, **77**, 455-466.

(8) Raftery, A. E. and Akman, V. E.(1986), Bayesian analysis of Poisson Process with a change point, *Biometrika*, **73**, 85-89.

(9) Rudems, M.(1982), Empirical choice of histograms and kernel density estimators, *Scandinavian Journal of Statistics*, **9**, 65-78.

(10) Siegmund, D.(1986), Boundary crossing probabilities and statistical applications, *The Annals of Statistics*, **14**, 361-404.

(11) Smith, A. F. M.(1975), A Bayesian approach to inference about a change point in a sequence of random variables, *Biometrika*, **62**, 407-416.

(12) Wolfe, D. A. and Schechtman, E.(1984), Nonparametric statistical procedures for the change point problem, *Journal of Statistical planning and inference*, **9**, 389-396.

(13) Zacks, S.(1983), Survey of classical and Bayesian approaches to the change point problem: fixed sample and sequential procedures of testing and estimation, *In recent advances in Statistics Herman Chernoff Festschrift*, 245-269, New York, Academic Press.