Journal of the Korean Statistical Society (2002), **31**: *1, pp 1-16*

Nonparametric Bayesian Multiple Change Point Problems

Chansoo Kim[1] and Younshik Chung[2]

ABSTRACT

Since changepoint identification is important in many data analysis problem, we wish to make inference about the locations of one or more changepoints of the sequence. We consider the Bayesian nonparameteric inference for multiple changepoint problem using a Bayesian segmentation procedure proposed by Yang and Kuo (2000). A mixture of products of Dirichlet process is used as a prior distribution. To decide whether there exists a single change or not, our approach depends on nonparametric Bayesian Schwartz information criterion at each step. We discuss how to choose the precision parameter(total mass parameter) in nonparametric setting and show that the discreteness of the Dirichlet process prior can have a large effect on the nonparametric Bayesian Schwartz information criterion and leads to conclusions that are very different results from reasonable parametric model. One example is proposed to show this effect.

Keywords: Mixture of Dirichlet process, multiple changepoint, nonparametric Bayesian Schwartz information criterion.

1. Introduction

Changepoint identification is important in many data analysis problem, such as signal processing, industrial system, economics, medicine etc. Several approaches to the changepoint problem have been published including nonparametric and parametric approaches. These are Hinkley (1971), Pettitt (1980) and Zacks (1983). The related works of Bayesian approach are presented by Chernoff and Zacks (1964), Smith (1975). A Gibbs sampling approach to Bayesian inference for single changepoint problem was presented by Carlin *et al.* (1992) and Mira and Petrone (1996). However, for long data sequences, the problem of

[1]Research Institute of Computer, Information and Communication, Pusan National University, Pusan 609-735, Korea

[2]Department of Statistics, Pusan National University, Pusan 609-735, Korea

multiple changepoint has not been considered by many authors, in part because of the difficulty in handling the computations.

The aim of this paper is a fully nonparametric Bayesian approach and we wish to make inference about the location of one or more changepoints of the data sequence. Let $\mathbf{x} = (x_1, \cdots, x_n)$ be a realization of the sequence of random variables $\mathbf{X} = (X_1, \cdots, X_n)$ of length n. The random variables have a distribution function with no change

$$P_r\left(X_1 \leq x_1, \cdots, X_n \leq x_n | F\right) = \prod_{i=1}^{n} F(x_i). \tag{1.1}$$

If we may have only a vague opinion on the parametric form of the distribution function F, then the nonparametric model might be more appealing. Therefore, we approach the nonparametric extension by assigning, conditional on θ, a Dirichlet process prior on the family $\mathcal{F}$ of all probability distribution in such a way that the conditional mean of F given θ is $F_0(\cdot|\theta)$. Therefore, the support of the prior distribution is large, including all distributions on the real line. This larger support allows a wide range of shapes for F and produces a more flexible estimator.

If the change-point is occurred at any position c, then the model is

$$P_r\left(X_1 \leq x_1, \cdots, X_n \leq x_n | c, F_1, F_2\right) = \prod_{i=1}^{c} F_1(x_i) \prod_{i=c+1}^{n} F_2(x_i), \tag{1.2}$$

where the vector of random probability measures (F_1, F_2) is distributed as a mixture of products of Dirichlet processes with parameters $\alpha_1(\cdot;\theta_1)$ and $\alpha_2(\cdot;\theta_2)$.

To decide whether there exists a single change or not, this approach depends on nonparametric Bayesian estimates of Schwartz information criterion(NBSIC). The original Schwartz information criterion(SIC) (1978) is defined as $-2\ln L(\hat{\theta}|D)$ $+p\ln n$ where $L(\hat{\theta})$ is the likelihood function for the model, $\hat{\theta}$ is a ML estimator of unknown parameters θ, p is the number of unknown parameters to be estimated and n is the sample size. In order to define NBSIC, our measure depends on the Bayesian estimate of Schwartz information criterion(BSIC),

$$\text{BSIC} = \int (-2\ln L(\theta|x_1, \cdots, x_n) + p\ln n)[\theta|x_1, \cdots, x_n]d\theta, \tag{1.3}$$

where $L(\theta|D)$ is a likelihood function and p is the number of unknown parameters to be estimate. Using Bayesian binary segmentation procedure, Yang and Kuo

(2000) consider the multiple change positions for homogeneous Poisson process with BSIC. We will denote NBSIC under the model (1.1) and (1.2) as follows

$$\begin{aligned}\text{NBSIC}(0|x_1,\cdots,x_n) &= \int(-2\ln\left\{\frac{1}{\alpha(\mathbf{R};\theta)^{[n]}}\prod_{i=1}^{n^*}\alpha(x_i;\theta)(m_i(x_i:\theta)+1)_{[n_i-1]}\right\} \\ &\qquad \times[\theta|x_1,\cdots,x_n]d\theta+p_1\ln n, \qquad (1.4)\end{aligned}$$

where p_1 is the number of unknown parameters to be estimate. And for a possible change, $j=1,\cdots,n-1$, in the model (1.2)

$$\begin{aligned}\text{NBSIC}(j|D) &= \int -2\ln\frac{1}{\alpha_1(\mathbf{R};\theta_1)^{[j]}}\prod_{i=1}^{j^*}\alpha_1'(x_i;\theta_1)\left(m_i(x_i:\theta_1)+1\right)_{[n_i-1]} \\ &\qquad \times\frac{1}{\alpha_2(\mathbf{R};\theta_2)^{[n-j]}}\prod_{i=j+1}^{n^*}\alpha_2'(x_i;\theta_2)\left(m_i(x_i:\theta_2)+1\right)_{[n_i-1]} \\ &\qquad \times d[\theta_1,\theta_2|D]+p_2\ln n, \qquad (1.5)\end{aligned}$$

where $D=(x_1,\cdots,x_n)$ and p_2 is the number of unknown parameters to be estimated. $(m_i(x_i:\theta_j)+1)_{[n_i-1]}=\alpha_j'(x_i;\theta_j)$ if x_i is an atom of α_j, 0 otherwise and n_i is the number of observations equal to x_i and $\alpha^{[k]}=\alpha(\alpha-1)\cdots(\alpha+k-1)$ and the * indicates that the product is taken over distinct values only. And $[\theta|D]$ denote a posterior distribution.

If $\text{NBSIC}(0|D)\le\min_{1\le j\le n-1}\text{NBSIC}(j|D)$, we conclude that there is no changepoint. We only estimate the distribution function and parameters in the model (1.1). If $\text{NBSIC}(0|D)>\min_{1\le j\le n-1}\text{NBSIC}(j|D)$, for some j, we conclude that there is the single changepoint in the data sequence and estimate the location of changepoint in the model (1.2). For the next step, we divide the data sequence into two subsequences according to the changepoint position. We apply the proposed method to each subsequences $D_c=(x_1,\cdots,x_c)$ and $\bar{D}_c=(x_{c+1},\cdots,x_n)$ and then continue the process until no more changes are found in any of the subsequences. In order to overcome the computational difficulties, Markov chain Monte Carlo method is used. In this paper, our objective is to find the locations of the changepoints based on the nonparametric Bayesian approach and discuss the discreteness of the Dirichlet process prior. The paper is organized as follows. Section 2 introduces the nonparametric Bayesian formulation on the Dirichlet process prior for the changepoint problem and discusses how to choose the precision parameter α(the total mass parameter). Finally, in section 3, our methodology is applied to an artificial data.

2. Nonparametric Bayesian Formulation

2.1. Nonparametric setup

Let $\mathbf{x} = (x_1, \cdots, x_n)$ be a realization of the sequence of a random variable $\mathbf{X} = (X_1, \cdots, X_n)$ of length n. Given the observation, we assume that there is no change and unknown distribution function. Then the model has the following form.

Model 1 :

$$\begin{aligned} P_r\,(X_1 \leq x_1, \cdots, X_n \leq x_n|F) &= \prod_{i=1}^{n} F(x_i), \\ F|\theta &\sim \mathcal{D}(\alpha(\cdot|\theta)), \\ \theta &\sim H(\theta), \end{aligned} \tag{2.1}$$

where $\mathcal{D}$ is a Dirichlet process. For a fixed θ, the parameter of the process is the measure $\alpha((\infty, x]|\theta)$, $x \in R$, which, for simplicity, we will indicate by $\alpha(x|\theta)$. Likewise, note that $\alpha(R|\theta) = \alpha(\infty|\theta)$ for total mass parameters. The parameter space Θ is the d-dimensional random vector with distribution function $H(\theta)$.

If a single changepoint is occurred at any position c, then the model is as follows.

Model 2 :

$$\begin{aligned} P_r\,(X_1 \leq x_1, \cdots, X_n \leq x_n|F_1, F_2) &= \prod_{i=1}^{c} F_1(x_i) \prod_{i=c+1}^{n} F_2(x_i), \\ F_1, F_2|\theta_1, \theta_2 &\sim \mathcal{D}(\alpha_1(\cdot|\theta_1))\mathcal{D}(\alpha_2(\cdot|\theta_2)), \\ \theta_1, \theta_2 &\sim H(\theta_1, \theta_2), \end{aligned} \tag{2.2}$$

where the vector of random probability measures (F_1, F_2) is distributed as a mixture of products of Dirichlet processes with parameters $\alpha_1(\cdot;\theta_1)$, $\alpha_2(\cdot;\theta_2)$ and mixing distribution $H(\theta_1, \theta_2)$, if given $\Theta_1 = \theta_1$ and $\Theta_2 = \theta_2$, F_1 and F_2 are independently distributed as a Dirichlet process of parameter $\alpha_1(\cdot;\theta_1)$ and $\alpha_2(\cdot;\theta_2)$ respectively, written

$$[F_1, F_2|\theta_1, \theta_2] = D(\alpha_1(\cdot;\theta_1))D(\alpha_2(\cdot;\theta_2)).$$

Under Model 2, Mira and Petrone (1996) considered a hierarchical Bayesian nonparametric inference for a single changepoint c. We consider nonparametric Bayesian method for locating a multiple change in the given data sequence and show that the discreteness of the Dirichlet process prior can have a large effect on the nonparametric Bayesian Schwartz information criterion and leads to conclusions that are very different results from reasonable parametric model. To define the nonparametric Bayesian estimate of Schwartz information criterion, we need the likelihood functions of Model 1 and Model 2. Therefore, the following lemmas give the likelihood functions for each model.

Lemma 2.1. *Assume that $\theta \in R^d$, $\alpha(\cdot;\theta)$ is absolutely continuous with respect to Lebesgue measure with density $\alpha'(\cdot;\theta)$. Then the likelihood function of $x_1, \cdots, x_n$ given θ is*

$$[x_1, \cdots, x_n|\theta] = \frac{1}{\alpha(\mathbf{R};\theta)^{[n]}} \prod_{i=1}^{n^*} \alpha'(x_i;\theta)\,(m_i(x_i;\theta)+1)_{[n_i-1]}, \qquad (2.3)$$

*where $(m_i(x_i:\theta)+1)_{[n_i-1]} = \alpha'(x_i;\theta)$ if x_i is an atom of α, 0 otherwise and n_i is the number of observations equal to x_i and $\alpha^{[k]} = \alpha(\alpha-1)\cdots(\alpha+k-1)$ and the * indicates that the product is taken over distinct values only.*

The derivation is based on Lemma 1 of Antoniak (1974). From Lemma 2.1, it follows that the posterior distribution of θ given $x_1, \cdots, x_n$, is proportion to

$$[\theta|x_1, \cdots, x_n] \propto [x_1, \cdots, x_n|\theta]dH(\theta). \qquad (2.4)$$

We assume that a single changepoint is occurred under the model (2.1). From the model (2.2), we get a likelihood function of θ_1 and θ_2 given $x_1, \cdots, x_n$.

Lemma 2.2. *Assume that θ_1 and $\theta_2 \in R^d$ and let $\alpha_1(\cdot;\theta_1)$ and $\alpha_2(\cdot;\theta_2)$ be absolutely continuous with respect to Lebesgue measure with densities $\alpha_1'(\cdot;\theta_1)$ and $\alpha_2'(\cdot;\theta_2)$, respectively. Then the likelihood of c, θ_1, θ_2 given $(x_1, \cdots, x_n)$ is*

$$\begin{aligned}[x_1, \cdots, x_n|c, \theta_1, \theta_2] &= \frac{1}{\alpha_1(\mathbf{R};\theta_1)^{[c]}} \prod_{i=1}^{c^*} \alpha_1'(x_i;\theta_1)\,(m_i(x_i:\theta_1)+1)_{[n_i-1]} \\ &\quad \times \frac{1}{\alpha_2(\mathbf{R};\theta_2)^{[n-c]}} \prod_{i=c+1}^{n^*} \alpha_2'(x_i;\theta_2)\,(m_i(x_i:\theta_2)+1)_{[n_i-1]},\end{aligned}$$

*where $\alpha^{[k]} = \alpha(\alpha-1)\cdots(\alpha+k-1)$ and the * indicates that the product is taken over distinct values only.*

Lemma 2.2 is a sequence of Lemma 1 of Antoniak (1974) and is equivalent to the Proposition 1 of Mira and Petrone (1996). To identify and to estimate how many changepoints are occurred, we need to define the nonparametric Bayesian estimate of Schwartz information criterion(NBSIC). This measure depends on the Bayesian estimate of Schwartz information criterion(BSIC). This measure is given in (1.3). If the likelihood $L(\theta|D)$ has no changepoints, then from (2.3) and (1.3), the nonparametric Bayesian estimate of SIC(NBSIC) is defined by the following form

$$\begin{aligned} \text{NBSIC}(0|x_1,\cdots,x_n) &= \int -2\ln\left\{\frac{1}{\alpha(\mathbf{R};\theta)^{[n]}}\prod_{i=1}^{n^*}\alpha(x_i;\theta)(m_i(x_i:\theta)+1)_{[n_i-1]}\right\} \\ &\quad \times [\theta|x_1,\cdots,x_n]d\theta \;+\; p_1\ln n, \end{aligned} \tag{2.5}$$

where p_1 is the number of unknown parameter.

For a given changepoint, we consider the j^{th} intermediary likelihood of c,θ_1,θ_2 given $(x_1,\cdots,x_n)$, which is obtained as, for $j=1,\cdots,n-1$,

$$\begin{aligned} &[x_1,\cdots,x_n|c=j,\theta_1,\theta_2] \\ &= \frac{1}{\alpha_1(\mathbf{R};\theta_1)^{[j]}}\prod_{i=1}^{j^*}\alpha_1'(x_i;\theta_1)\,(m_i(x_i:\theta_1)+1)_{[n_i-1]} \\ &\quad \times\frac{1}{\alpha_2(\mathbf{R};\theta_2)^{[n-j]}}\prod_{i=j+1}^{n^*}\alpha_2'(x_i;\theta_2)\,(m_i(x_i:\theta_2)+1)_{[n_i-1]}\,. \end{aligned} \tag{2.6}$$

According to the definition of (1.3), we define $\text{NBSIC}(j|x_1,\cdots,x_n)$ as

$$\text{NBSIC}(j|D) = \int -2\ln[x_1,\cdots,x_n|c=j,\theta_1,\theta_2][\theta_1,\theta_2|D]d\theta_1 d\theta_2 + p_2\ln n, \tag{2.7}$$

where $D=(x_1,\cdots,x_n)$ and p_2 is the dimensions of (θ_1,θ_2). $[x_1,\cdots,x_n|c=j,\theta_1,\theta_2]$ is given in (2.6).

To estimate $\text{NBSIC}(0|D)$ in (2.5) and $\text{NBSIC}(j|D)$ in (2.7), we can use a Gibbs sampler. Let t denote the value in the t^{th} iteration for $t=1,\cdots,T$ in the Markov chain Monte Carlo. A Monte Carlo estimates of $\text{NBSIC}(0|D)$ and $\text{NBSIC}(j|D)$, for $j=1,\cdots,n-1$, with direct substitution from the Gibbs sampler are

$$\widehat{\text{NBSIC}}(0|D) = \frac{1}{T}\sum_{t=1}^{T} -2\ln[x_1,\cdots,x_n|\theta] + p_1\ln n, \tag{2.8}$$

and for $j = 1, \cdots, n-1$,

$$\widehat{\text{NBSIC}}(j|D) = \frac{1}{T}\sum_{t=1}^{T} -2\ln[x_1, \cdots, x_n | c = j, \theta_1, \theta_2] + p_2 \ln n, \tag{2.9}$$

where $[x_1, \cdots, x_n|\theta]$ and $[x_1, \cdots, x_n|c = j, \theta_1, \theta_2]$ are given in (2.3) and (2.6), respectively.

If $\widehat{\text{NBSIC}}(0|D) \leq \min_{1\leq j\leq n-1}\widehat{\text{NBSIC}}(j|D)$, we conclude that there is no changepoint in the original sequence, we just only estimate the distribution function F and stop the procedure. If $\widehat{\text{NBSIC}}(0|D) > \min_{1\leq j\leq n-1}\widehat{\text{NBSIC}}(j|D)$, for some j, we conclude that there is the single changepoint in the original sequence, estimate the changepoint, $\hat{c} = k^{th}$ position of data which satisfies $\widehat{\text{NBSIC}}(k|D) = \min_{1\leq j\leq n-1}\widehat{\text{NBSIC}}(j|D)$.

For the next step, we divide the original sequence into two subsequences : $D_c = (x_1, \cdots, x_c)$ and $\bar{D}_c = (x_{c+1}, \cdots, x_n)$. For the two subsequences D_c and $\bar{D}_c$, changepoint, as like in the original sequence, and continue the process until no more changes are founded in any of the subsequences.

The detailed algorithm is as follows.

1. The first step :

(1a) We consider that there does not exit a changepoint, $D = (x_1, \cdots, x_n)$. From the model 1, we generate θ and F using Gibbs sampler. The full conditional distribution(FCD) are

$$[F|\theta, D] \sim \mathcal{D}\left(\alpha(\cdot;\theta) + \sum_{i=1}^{n}\delta_{x_i}(\cdot)\right),$$

$$[\theta|F, D] \propto \frac{1}{\alpha(\mathbf{R};\theta)^{[n]}}\prod_{i=1}^{n^*}\alpha'(x_i;\theta)\left(m_i(x_i:\theta)+1\right)_{[n_i-1]}h(\theta),$$

where $h(\theta)$ is a density function of θ.

$$\widehat{\text{NBSIC}}(0|D) = \frac{1}{T}\sum_{t=1}^{T} -2\ln[x_1, \cdots, x_n|\theta^{(t)}] + p_1 \ln n.$$

(1b) We assume that a single changepoint is occurred at any position, $j =$

$1, \cdots, n-1$. From the model (2.2), the FCD are

$$[F_1|\theta_1, \theta_2, F_2, D] \sim \mathcal{D}\left(\alpha_1(\cdot;\theta_1) + \sum_{i=1}^{j} \delta_{x_i}(\cdot)\right),$$

$$[F_2|\theta_1, \theta_2, F_1, D] \sim \mathcal{D}\left(\alpha_2(\cdot;\theta_2) + \sum_{i=j+1}^{n} \delta_{x_i}(\cdot)\right),$$

$$[\Theta|F_1, F_2, D] \propto \frac{1}{\alpha_1(\mathbf{R};\theta_1)^{[j]}} \prod_{i=1}^{j^*} \alpha_1'(x_i;\theta_1)\,(m_i(x_i:\theta_1)+1)_{[n_i-1]}$$
$$\times \frac{1}{\alpha_2(\mathbf{R};\theta_2)^{[n-j]}} \prod_{i=j+1}^{n^*} \alpha_2'(x_i;\theta_2)\,(m_i(x_i:\theta_2)+1)_{[n_i-1]}\, h(\theta_1, \theta_2),$$

where $\Theta = (\theta_1, \theta_2)$. Therefore for $j = 1, \cdots, n-1$,

$$\widehat{\text{NBSIC}}(j|D) = \frac{1}{T}\sum_{t=1}^{T} -2\ln[x_1, \cdots, x_n|c = j, \theta_1^{(t)}, \theta_2^{(t)}] + p_2 \ln n.$$

(1c) If $\widehat{\text{NBSIC}}(0|D) \leq \min_{1\leq j\leq n-1}\widehat{\text{NBSIC}}(j|D)$, we conclude that there is no changepoint in the original sequence, we just only estimate the distribution of F and the parameter θ and then stop the procedure. Otherwise, we conclude that there is the single changepoint and go to the second step.

2. *The second step :* We have estimated the changepoint, $\hat{c} = k$, which satisfies $\widehat{\text{NBSIC}}(k|D) = \min_{1\leq j\leq n-1}\widehat{\text{NBSIC}}(j|D)$. According to $\hat{c}$, we divide the original data into two subsequences, say $D_k = (x_1, \cdots, x_k)$ and $\bar{D}_k = (x_{k+1}, \cdots, x_n)$.
($2a_1$) For the first subsequence $D_k = (x_1, \cdots, x_k)$, applying the step (1a) in the first step is applied again. Therefore, the FCD are

$$[F|\theta, D] \sim \mathcal{D}\left(\alpha(\cdot;\theta) + \sum_{i=1}^{k} \delta_{x_i}(\cdot)\right),$$

$$[\theta|F, D] \propto \frac{1}{\alpha(\mathbf{R};\theta)^{[k]}} \prod_{i=1}^{k^*} \alpha'(x_i;\theta)\,(m_i(x_i:\theta)+1)_{[n_i-1]}\, h(\theta),$$

where k^* is the number of distinct values on 1,2, $\cdots$, k. Then the estimate of NBSIC$(0|D)$ is obtained in similar way by

$$\widehat{\text{NBSIC}}(0|D) = \frac{1}{T}\sum_{t=1}^{T} -2\ln[x_1, \cdots, x_n|\theta^{(t)}] + p_1 \ln n.$$

$(2a_2)$ We assume that a single changepoint is occurred at any position j, $j = 1, \cdots, k-1$. By the similar way in the step (1b), the FCD are, for $j = 1, \cdots, k-1$,

$$[F_1|\theta_1, \theta_2, F_2, D_c] \sim \mathcal{D}\left(\alpha_1(\cdot;\theta_1) + \sum_{i=1}^{j} \delta_{x_i}(\cdot)\right),$$

$$[F_2|\theta_1, \theta_2, F_1, D_c] \sim \mathcal{D}\left(\alpha_2(\cdot;\theta_2) + \sum_{i=j+1}^{k} \delta_{x_i}(\cdot)\right),$$

$$[\Theta|F_1, F_2, D_c] \propto \frac{1}{\alpha_1(\mathbf{R};\theta_1)^{[j]}} \prod_{i=1}^{j^*} \alpha_1'(x_i;\theta_1)\,(m_i(x_i:\theta_1)+1)_{[n_i-1]}$$
$$\times \frac{1}{\alpha_2(\mathbf{R};\theta_2)^{[k-j]}} \prod_{i=j+1}^{k^*} \alpha_2'(x_i;\theta_2)\,(m_i(x_i:\theta_2)+1)_{[n_i-1]}\, h(\theta_1, \theta_2),$$

$$\mathrm{N\widehat{BSI}C}(j|D_c) = \frac{1}{T}\sum_{t=1}^{T} -2\ln[x_1, \cdots, x_n|c = j, \theta_1^{(t)}, \theta_2^{(t)}] + p_2 \ln n.$$

$(2a_3)$ we conclude as like (1c) in the first step.
$(2b_1)$ $\bar{D}_k = (x_{k+1}, \cdots, x_n)$, the steps $(2a_1)$-$(2a_3)$ are applied again to the data $\bar{D}_k$.

3. The third step : We continue the above processes until no more changes are found in a way of each subsequences.

2.2. Choosing the total mass α

This section shows that the discreteness of the Dirichlet process prior can have a large effect on the nonparametric Bayesian Schwartz information criterion and leads to conclusion that are very different results from reasonable parametric models.

First, let $\alpha'(x_i;\theta)$ in (2.5) and $\alpha_1'(x_i;\theta_1)$ and $\alpha_2'(x_i;\theta_2)$ in (2.6) be $\alpha f_0(x_i|\theta)$, $\alpha_1 f_1(x_i|\theta_1)$ and $\alpha_2 f_2(x_i|\theta_2)$, respectively, where $\alpha = \alpha(R|\theta)$, $\alpha_1 = \alpha_1(R|\theta)$, $\alpha_2 = \alpha_2(R|\theta)$ for the total mass parameters and f_0, f_1 and f_2 in model (2.1) and (2.2) are the baseline distributions. Also, we consider that all observations are distinct values for simplicity. Then the $NBSIC(0|D)$ and $NBSIC(j|D)$ are expressed respectively by

$$N\widehat{BSI}C(0|D) = \frac{1}{T}\sum_{t=1}^{T}\left(-2\ln\left\{\frac{\alpha^n}{\alpha^{[n]}}\prod_{i=1}^{n} f_0(x_i|\theta^{(t)})\right\}\right) + p_1 \ln n, \qquad (2.10)$$

and for $j = 1, \cdots, n-1$,

$$\widehat{NBSIC}(j|D) = \frac{1}{T}\sum_{t=1}^{T}\left(-2\ln\left\{\frac{\alpha_1^j}{\alpha_1^{[j]}}\prod_{i=1}^{j} f_1(x_i|\theta_1^{(t)})\frac{\alpha_2^{n-j}}{\alpha_2^{[n-j]}}\prod_{i=j+1}^{n} f_2(x_i|\theta_2^{(t)})\right\}\right) + p_2 \ln n, \qquad (2.11)$$

where $\theta^{(t)}$, $\theta_1^{(t)}$ and $\theta_2^{(t)}$ are drawn from the posterior distributions and T is the required iteration number.

If $\frac{\alpha^{[n]}}{\alpha^n}$ and $\frac{\alpha_1^j \alpha_2^{n-j}}{\alpha_1^{[j]}\alpha_2^{[n-j]}}$ are equal to 1, then the equations (2.10) and (2.11) are equal to the parametric Bayesian Schwartz information criterion. These terms give effect on NBSIC. Therefore, we can see that these terms play an important role in Bayesian nonparametric setup using DPP.

First, let $\alpha = \alpha_1 = \alpha_2$. Then $\frac{\alpha_1^j \alpha_2^{n-j}}{\alpha_1^{[j]}\alpha_2^{[n-j]}}$ in (2.11) is equal to $\frac{\alpha^n}{\alpha^{[j]}\alpha^{[n-j]}}$ and is given in Figure 2.2.

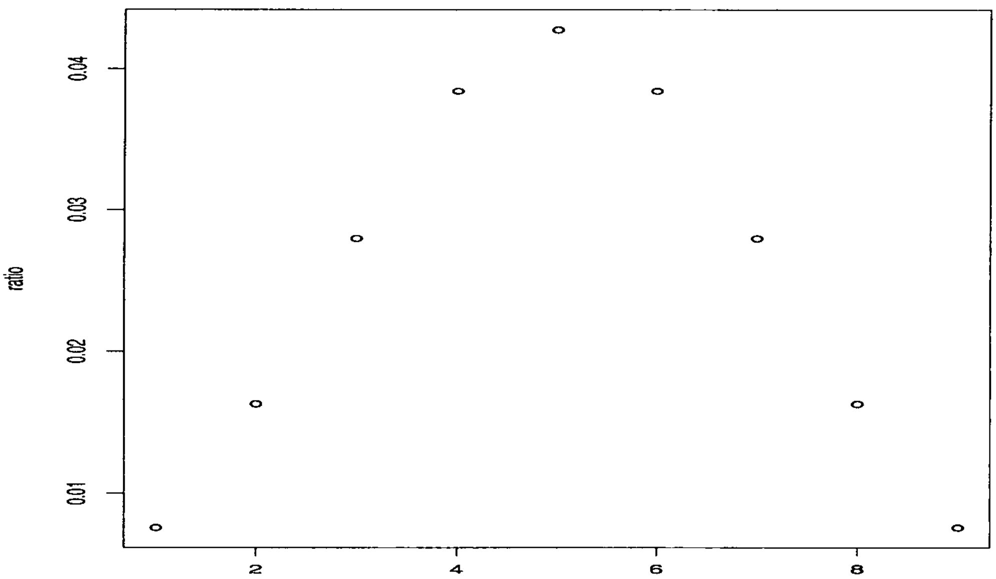

FIGURE 2.1 $\frac{\alpha^n}{\alpha^{[j]}\alpha^{[n-j]}}$ *for* $\alpha = 5$, $n = 10$

Figure 2.1 shows that this term is much bigger as close to $\frac{n}{2}$ and this behavior is more marked if n is larger. As α goes to infinity, it converges to the parametric result. As seen above, the difference of the BSIC and nonparametric BSIC based on DPP is whether these terms of the equation (2.11) and (2.12) exist or not. Therefore, the effect term is very important to explain the discreteness of DPP and leads misleading results against nonparametric Bayesian approach.

Next, let $\alpha \neq \alpha_1 = \alpha_2$. Then it is given in Figure 2.2 for various α_1, given $\alpha = 5$ and $n = 20$. In Figure 2.2, the horizontal line is $-2\ln\frac{\alpha^n}{\alpha^{[n]}}$ and the dotted

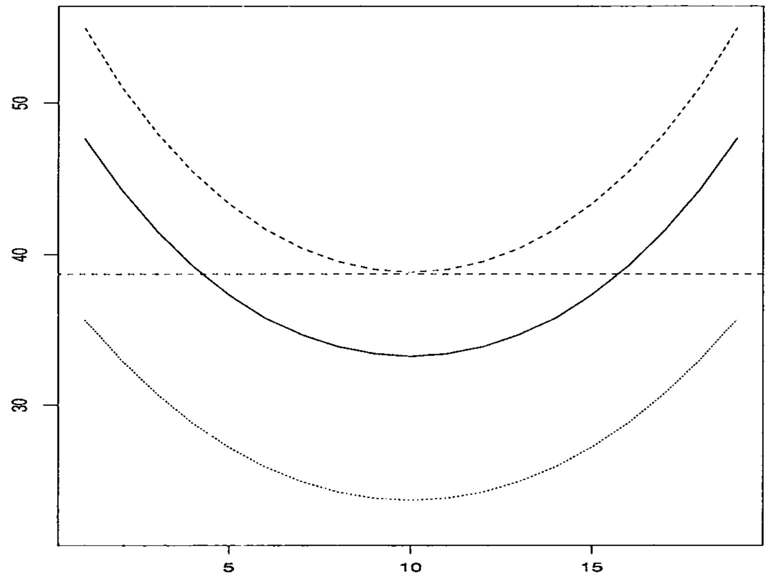

FIGURE 2.2 $-2\ln\frac{\alpha^n}{\alpha^{[n]}}$ *and* $-2\ln\frac{\alpha_1^n}{\alpha_1^{[j]}\alpha_1^{[n-j]}}$ *for* $\alpha = 5$, $\alpha_1 = 2.18, 3, 5$ *and* $n = 20$

line, straight line and dashed line denote the values of $-2\ln\frac{\alpha_1^n}{\alpha_1^{[j]}\alpha_1^{[n-j]}}$ with $\alpha_1 = 5$, $\alpha_1 = 3$ and $\alpha_1 = 2.18$, respectively.

For the case $\alpha{=}\alpha_1{=}\alpha_2$, the difference of $-2\ln\frac{\alpha^n}{\alpha^{[n]}}$ and $-2\ln\frac{\alpha^n}{\alpha^{[j]}\alpha^{[n-j]}}$ is

$$2\ln\frac{\alpha^{[n]}}{\alpha^{[j]}\alpha^{[n-j]}} \tag{2.12}$$

and $\alpha^{[n]}$ is always greater than $\alpha^{[j]}\alpha^{[n-j]}$. Therefore, $NBSIC(0|D)$ is always greater than $min_k NBSIC(k|D)$, for all k. Therefore, our criterion favors a model which has a change. For example, as seen in Figure 2.2, the horizontal line is $-2\ln\frac{\alpha^n}{\alpha^{[n]}}$ and the dotted line is $-2\ln\frac{\alpha^n}{\alpha^{[j]}\alpha^{[n-j]}}$ and its minimum value is obtained at $j = \frac{n}{2}$. In this case, it is seen that the equation (2.12) is always greater than 0 and this result gives effect on $NBSIC(0|D)$. Therefore, it is seem to occur a change and we conclude that $NBSIC$ always selects a model with change and leads a misleading result.

Next, we consider $\alpha \neq \alpha_1 = \alpha_2$ and $\alpha > \alpha_1$. For given α, we have different results in multiple changepoint problem according to how we choose α_1. In Figure 2.2, it is seen that $-2\ln \frac{\alpha_1^n}{\alpha_1^{[j]}\alpha_2^{[n-j]}}$ is a convex function, given n and the minimum is obtained at $j = \frac{n}{2}$. If $\alpha_1 = 2.18$, then this function is always greater than $-2\ln \frac{\alpha^n}{\alpha^{[n]}}$ and if $\alpha_1 = 3$, then this function is greater or smaller in some regions. As mentioned above, the relation of the minimum value of $-2\ln \frac{\alpha_1^n}{\alpha_1^{[j]}\alpha_2^{[n-j]}}$ and $-2\ln \frac{\alpha^n}{\alpha^{[n]}}$ is affect to identify a model selection whether there is a change or not.

3. An Illustrative Example

We consider an application to the simulation data set which is generated from normal distribution. The aim is to identify whether there is a change or not and to find whether there actually is a change in the simulated data set. That is, we handle how many changepoints are occurred in given data sequence.

Now consider the simulated data which contains the three changepoints. The simulated data set have forty observations which are generated from normal distributions with mean 2, 5, 8 and 11 respectively. For all above cases, its variance is 1. The data set is described in Figure 3.1. Figure 3.1 is seem to have two or three change positions(the exact number of changepoint is three).

To compare the parametric Bayesian and nonparametric Bayesian based on DPP approaches, we apply our algorithm to the simulated data set. According to the algorithm of the mentioned above, we find the parametric Bayesian estimate of Schwartz information criterion($\widehat{BIC}$) for given data sets. It is given in Table 3.1.

Table 3.1 gives the values of the parametric Bayesian estimate of Schwartz information criterion. In Table 3.1, the values of parenthesis denote the position of changepoint. For example, the 20th data in Step1 denotes that $N\widehat{BSIC}(20|D)$ =184.8484 is the smallest among its all corresponding values. Therefore the change happens at the 20th data among all data. It is seen that this data set has three changepoints and the changepositions are 10th, 20th and 30th data.

To calculate the $N\widehat{BSIC}$, let the baseline functions of f_0, f_1 and f_2 have normal probability density functions. And the prior distribution of θ has also

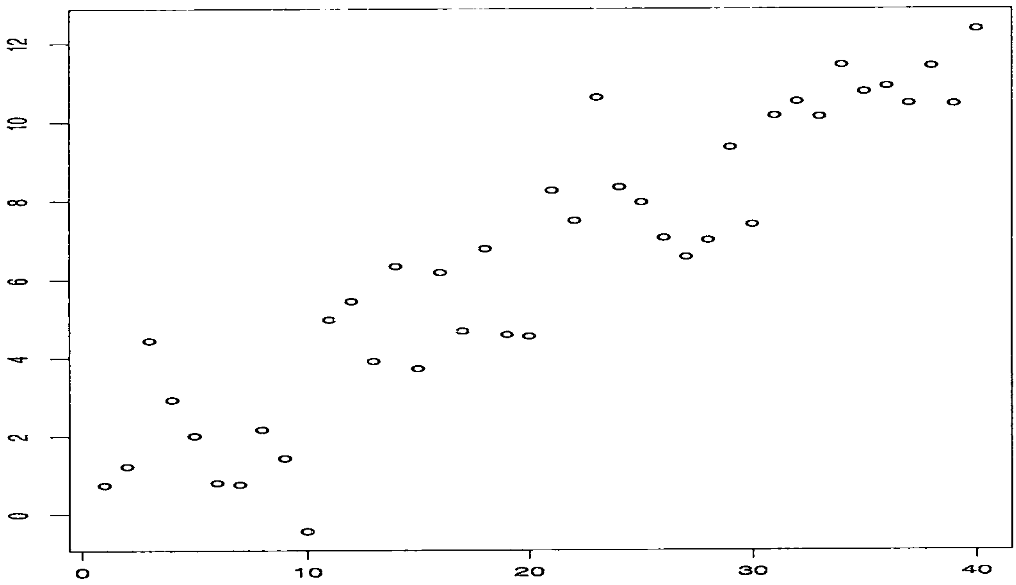

FIGURE 3.1 *The simulated data set*

TABLE 3.1 *Parametric Bayesian estimates of the Schwartz information criterion for the simulated data set*

Step	*Data Set*	$\widehat{BSIC}(0)$	$min_{1\leq j\leq N-1}\widehat{BSIC}(j\|D)$
Step1	all data	363.6413	184.8484 (20th data)
Step2	1st - 20th data	99.47418	75.41471 (10th data)
	21st - 40th data	83.86497	67.52570 (30th data)
Step3	1st - 10th data	39.63537	47.62841
	11th - 20th data	34.27936	39.45640
	21st - 30th data	35.40147	37.32937
	31st - 40th data	30.71867	32.87593

normal distribution. We obtain the nonparametric Bayesian estimate of Schwartz information criterion($\widehat{NBSIC}$) for $\alpha_1 = 2.401$. It is given in Table 3.2.

TABLE 3.2 *Nonparametric Bayesian estimates of the Schwartz information criterion for the simulated data set for $\alpha = 5$ and $\alpha_1 = \alpha_2 = 2.401$*

Step	*Data Set*	$\widehat{NBSIC}(0)$	$min_{1 \le j \le N-1} \widehat{NBSIC}(j\|D)$
Step1	all data	497.1648	300.3642 (20th data)
Step2	1th - 20th data	138.31	113.4641 (10th data)
	21h - 40th data	122.7008	105.5751 (30th data)
Step3	1th - 10th data	42.5563	44.4174
	11th - 20th data	46.1169	50.4379
	21th - 30th data	47.2389	48.4658
	31th - 40th data	42.5563	44.4174

It is seen that same conclusion is obtained via the parametric Bayesian analysis in Table 3.1. $\alpha_1 = 2.401$ is obtained by solving $-2\ln\frac{\alpha_1^n}{\alpha_1^{[\frac{n}{2}]}\alpha_2^{[n-\frac{n}{2}]}}$ is approximately equal to $-2\ln\frac{\alpha^n}{\alpha^{[n]}}$.

In conclusion, if $\alpha = \alpha_1 = \alpha_2$, then $NBSIC(0|D)$ is always greater than $min_j NBSIC(j|D)$ for all j. Our criterion favors a model with a change. Also, α_1 is much smaller than α and both are small(for example, $\alpha = 1$), then $min_j NBSIC$ $(j|D)$ is greater than $NBSIC(0|D)$. It takes a model with no change. And both cases give misleading results. But, for a proper value of α's as mentioned in Section 2.2, we could have a similar result as like a parametric Bayesian analysis.

REFERENCES

Antoniak, C. E. (1974). "Mixtures of Dirichlet processes with applications to nonparametric problems", *The Annals of Statistics*, **2**, 1152-1174.

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

Chernoff, H. and Zacks, S. (1964). "Estimating the correct mean of a normal distribution which is subjected to a change in time", *The Annals of Mathematical Statistics*, **35**, 999-1018.

Escobar, M. D. (1988). *Estimating the Means of Several Normal Populations by Nonparametric Estimation of the Distribution of the Means*, unpublished dissertation, Yale University.

Escobar, M. D. (1994). "Estimating normal means with a Dirichlet process prior", *Journal of the American Statistical Association*, **89**, 268-277.

Escobar, M. D. and West, M. (1995). "Bayesian density estimation and inference using mixtures", *Journal of the American Statistical Association*, **90**, 577-588.

Ferguson, T. S. (1973). " Bayesian analysis of some nonparametric problems", *The Annals of Statistics*, **1**, 209-230.

Gelfans, A. E. and Smith, A. F. M. (1990). "Sampling-Based approaches to calculating marginal densities", *Journal of the American Statistical Association*, **85**, 398-409.

Hinkley, D. V. (1971). "Inference about the changepoint from cumulative sum tests", *Biometrika*, **67**, 79-84.

Kuo, L. (1986). "Computations of mixtures of Dirichlet processes", *SIAM Journal of Scientific and Statistical Computing*, **7**, 60-71.

MacEachern, S. N. (1994). "Estimating normal means with a conjugate style Dirichlet process prior", *Communications in Statistics : Simulation and Computation*, **23**, 727-741.

Mira, A. and Petrone, S. (1996). "Bayesian hierarchical nonparametric inference for change-point problems", *Bayesian Statistics*, **5**, 693-703.

Petitt, A. N. (1980). "A simple cummulative sum type statistic for the change-point problem with zero-one observations", *Biometrika*, **67**, 79-84.

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

Yang, T. and Kuo, L. (2000). "Bayesian binary segmentation procedure for homogeneous Poisson process with multiple change positions", TR0002, Department of Statistics, University of Connecticutt.

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. Academic Press. New York.