Statistics, Vol. 39, No. 5, October 2005, 373–387

Bayesian single change point detection in a sequence of multivariate normal observations

YOUNG SOOK SON† and SEONG W. KIM*‡

†Department of Statistics, Chonnam National University, Kwangju 500-757, Korea
‡Division of Applied Mathematics, Hanyang University,
Ansan 426-791, Korea

(Received 5 March 2004; revised 5 July 2005; in final form 5 August 2005)

A Bayesian method is used to see whether there are changes of mean, covariance, or both at an unknown time point in a sequence of independent multivariate normal observations. Noninformative priors are used for all competing models: no-change model, mean change model, covariance change model, and mean and covariance change model. We use several versions of the intrinsic Bayes factor of Berger and Pericchi (Berger, J.O. and Pericchi, L.R., 1996, The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association,* **91**, 109–122; Berger, J.O. and Pericchi, L.R., 1998, Accurate and stable Bayesian model selection: the median intrinsic Bayes factor. *Sankhya Series B*, **60**, 1–18.) to detect a change point. We demonstrate our results with some simulated datasets and a real dataset.

Keywords: Change point; Default Bayes factor; Intrinsic bayes factor; Noninformative prior; Posterior probability

1. Introduction

Change point problems originally arise from quality control or reliability. Nowadays, these problems cover almost all areas including environmental sciences, hydrology, signal processing, biology, climatology, economics, etc. There are considerable amounts of work on change point problems and related topics and these have appeared in the literature. Change point problems have generally been solved by the maximum likelihood methodology and the Bayesian procedure based on the parametric method, the nonparametric method, and the decision theoretic method. Our primary interest is restricted to the parametric Bayesian method applicable to change point problems in a sequence of independent normal observations. Suppose that there is a sequence of independent normal random variables, $X_1, X_2, \ldots, X_n$. These are observed along with time. This sequence is said to have a change at a time point r,

*Corresponding author. Email: seong@hanyang.ac.kr

Statistics
ISSN 0233-1888 print/ISSN 1029-4910 online © 2005 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/02331880500315339

often called a change point if

$$X_t \sim \begin{cases} N(\mu_1, \sigma_1^2), & t = 1, 2, \ldots, r, \\ N(\mu_2, \sigma_2^2), & t = r+1, r+2, \ldots, n, \end{cases} \tag{1}$$

provided at least $\mu_1 \neq \mu_2$ or $\sigma_1^2 \neq \sigma_2^2$. All the studies on change point problems are generally divided into two parts. The first part is to detect the existence of changes, that is, to test the no-change model.

$$X_t \sim N(\mu_1, \sigma_1^2), \quad t = 1, 2, \ldots, n,$$

against the change model (1) using the Bayes factor or the 'posterior probability'. The posterior probability will be explained in section 2. The second part is to obtain parameter estimates under the assumption that a change has occurred in a sequence. Smith [1] gave posterior probabilities for five different cases: case 1 with μ_1, μ_2, σ_1^2, and σ_2^2 known; case 2 with μ_1 known, μ_2 unknown, and $\sigma_1^2 = \sigma_2^2 = \sigma^2$ known; case 3 with μ_1 known, μ_2 unknown, and $\sigma_1^2 = \sigma_2^2 = \sigma^2$ unknown; case 4 with μ_1, μ_2, and $\sigma_1^2 = \sigma_2^2 = \sigma^2$ unknown; and case 5 with μ_1, μ_2, and $\sigma_1^2 \neq \sigma_2^2$ unknown. Note that discrete prior distributions are assumed for the change point. Lee and Heghinian [2] studied a mean change problem in a sequence of independent normal random variables with unknown common variance σ^2. They derived the posterior distribution of the change point under the uniform prior for the change point r, normal conjugate priors for means μ_1 and μ_2, and the noninformative prior for variance σ^2. A study of Menzefricke [3] is similar to that of Smith [1], except prior assumptions. As a case study on the univariate normal change model, Perreault *et al.* [4] conducted a Bayesian change point analysis under conjugate priors for a mean change model in a sequence of independent normal random variables. It was applied to hydrometeorological sequences and precipitation and runoff data series over eastern Canada and the US during the 20th century. Perreault *et al.* [5, 6] provided a Bayesian methodology under conjugate priors to test existence, types, and strength of changes in a sequence of hydrometeorological random variables. In their method, four univariate normal models were considered: a no-change model, a mean change model, a variance change model, and a simultaneous change model both in mean and variance. Bayesian model selection was performed by the posterior probability through the usual Bayes factor and parameters estimation via Gibbs sampling for the selected change models. The mean change model for a sequence of independent $p \times 1$ multivariate normal vectors, $(X_1, X_2, \ldots, X_n)$ considered by Booth and Smith [7] is

$$X_t \sim \begin{cases} N_p(\mu_1, \Sigma), & t = 1, 2, \ldots, r, \\ N_p(\mu_2, \Sigma), & t = r+1, r+2, \ldots, n, \end{cases} \tag{2}$$

where μ_1 and μ_2 ($\mu_1 \neq \mu_2$) are $p \times 1$ unknown mean vectors and Σ is a $p \times p$ unknown common covariance matrix. They derived the Bayes factor for testing the mean change model (2) with the prior $\pi(\mu_1, \mu_2, \Sigma) = |\Sigma|^{-(p+3)/2}$ against the no-change model with $\mu_1 = \mu_2$ in equation (2). Perreault *et al.* [8] applied a Bayesian change point procedure under conjugate priors for a mean change model of a sequence of multivariate normal vectors. The results were applied to hydrological time series such as streamflow data for six rivers situated in the Northern Québec Labrador region.

In our article, an objective Bayesian method is used for detecting a mean change, a covariance change, or both changes in a sequence of independent multivariate normal vectors. Noninformative priors are assumed for a set of parameters of a multivariate normal distribution and for the change point. Because we use improper priors, as is well known, the ordinary Bayes factor is not well defined. Berger and Pericchi [9] introduced the intrinsic Bayes factor

(IBF) using a data-splitting method, which makes the Bayes factor well defined. There have been several articles which use the IBF. Varshavsky [10] discussed some problems in a stationary autoregressive process. Lingham and Sivaganesan [11] conducted a test for the shape parameter of the power law process. Kim [12] analysed comparisons of two exponential means. Kim and Ibrahim [13] dealt with model selection problems in generalized linear models. In particular, we use several versions of the IBF of Berger and Pericchi [9, 14]: the arithmetic intrinsic Bayes factor (AIBF), the geometric intrinsic Bayes factor (GIBF), and the median intrinsic Bayes factor (MIBF). They are derived to test whether there are any changes. If there are any, the next interesting feature is to see what kind of changes they are.

The rest of this article is organized as follows. In the next section, the definitions of default Bayes factors are reviewed. In section 3, we compute the IBFs in order to choose one among the four models, a no-change model, a mean change model, a covariance change model, and a mean and covariance change model. In section 4, we present some numerical results through a simulation study and real data analysis. We provide a brief discussion in section 5.

2. Default Bayes factors

Let $X = (X_1, X_2, \ldots, X_n)$ be a sequence of random variates from a multivariate normal distribution. We consider four types of models to check whether there is any kind of change in the sequence. The competing models are a no-change model (M_0), a mean change model (M_1), a covariance change model (M_2), and a mean and covariance change model (M_3). Let r_j be a single change point of each change model M_j ($j = 1, 2, 3$). The models are constructed as follows:

$$\begin{cases} M_0\colon X_t \sim N_p(\mu_1, \Sigma_1), \quad t = 1, 2, \ldots, n, \\ M_{1,r_1}\colon X_t \sim \begin{cases} N_p(\mu_1, \Sigma_1), & t = 1, 2, \ldots, r_1, \\ N_p(\mu_2, \Sigma_1), & t = r_1 + 1, r_1 + 2, \ldots, n, \end{cases} \\ M_{2,r_2}\colon X_t \sim \begin{cases} N_p(\mu_1, \Sigma_1), & t = 1, 2, \ldots, r_2, \\ N_p(\mu_1, \Sigma_2), & t = r_2 + 1, r_2 + 2, \ldots, n, \end{cases} \\ M_{3,r_3}\colon X_t \sim \begin{cases} N_p(\mu_1, \Sigma_1), & t = 1, 2, \ldots, r_3, \\ N_p(\mu_2, \Sigma_2), & t = r_3 + 1, r_3 + 2, \ldots, n, \end{cases} \end{cases}$$

where X_t is a $p \times 1$ normal vector, μ_1 and $\mu_2 (\mu_1 \neq \mu_2)$ are unknown $p \times 1$ mean vectors, Σ_1 and Σ_2 ($\Sigma_1 \neq \Sigma_2$) are unknown $p \times p$ covariance matrices of multivariate normal distributions, and $r_j \in I$ with I being a set of positive integers to be constructed in order that all the parameters in the model M_{j,r_j} ($j = 1, 2, 3$) can be estimated. Now, we review an important role of Bayes factors as a useful tool for hypotheses testing or model selection in Bayesian inference. Let $x = (x_1, \ldots, x_n)$ denote an observed sequence of X. For models M_0 and M_{j,r_j} ($j = 1, 2, 3$), let $\theta_0 = (\mu_1, \Sigma_1)$, $\theta_1 = (\mu_1, \mu_2, \Sigma_1)$, $\theta_2 = (\mu_1, \Sigma_1, \Sigma_2)$, and $\theta_3 = (\mu_1, \mu_2, \Sigma_1, \Sigma_2)$. Here, $\theta_j \in \Theta_j$, where Θ_j is the parameter space for θ_j ($j = 0, 1, 2, 3$). A usual Bayes factor B_{j0} of comparing model M_{j,r_j} against model M_0 is given by

$$B_{j0}(\boldsymbol{x}|r_j) = \frac{m_j(r_j, \boldsymbol{x})}{m_0(\boldsymbol{x})}, \quad j = 1, 2, 3, \tag{3}$$

where

$$m_0(\boldsymbol{x}) = \int_{\Theta_0} \pi_0(\theta_0) l_0(\theta_0|\boldsymbol{x}) \mathrm{d}\theta_0, \tag{4}$$

and

$$m_j(r_j, \boldsymbol{x}) = \int_{\Theta_j} \pi_j(\theta_j, r_j) l_j(\theta_j, r_j|\boldsymbol{x}) \mathrm{d}\theta_j, \quad j = 1, 2, 3. \tag{5}$$

Here, $l_0(\theta_0|\boldsymbol{x})$ and $l_j(\theta_j, r_j|\boldsymbol{x})$ are the likelihood functions of model M_0 and model M_{j,r_j}, respectively, and $\pi_0(\theta_0)$ and $\pi_j(\theta_j, r_j)$ are the corresponding prior densities. Further, we call $m_0(\boldsymbol{x})$ and $m_j(r_j, \boldsymbol{x})$ the marginal densities or the predictive densities. In the beginning of Bayesian experiment without any prior information on parameters, default priors such as non-informative priors can be used. Objective Bayesian methods based on noninformative priors are called 'default' or 'automatic' Bayesian methods in comparison with the subjective Bayesian method with subjective prior information (cf. [15]). Most of the noninformative priors are improper, so caution is needed according to whether priors are proper or improper because of unspecified constants incorporated into the Bayes factor. Thus, prior specifications are quite important in Bayesian analysis, especially in hypothesis testing or model selection. The IBF of Berger and Pericchi [9] is classified as a default Bayes factor which is free from arbitrariness of noninformative improper priors. Their default Bayes factors are simpler and more automatic to use in the sense that they do not need to specify hyperparameters under conjugate priors nor to consider the imaginary constant as appeared in Spiegelhalter and Smith [16].

The IBF uses the minimal training sample to convert the improper prior to the proper posterior density. The minimal training sample represents a part of the full sample to guarantee the finiteness of marginal densities for all models. Let $\boldsymbol{x}(l)$ be the lth minimal training sample and let L denote the total number of minimal training samples. Then, the AIBF, the GIBF, and the MIBF of Berger and Pericchi [9, 14] for testing model M_{j,r_j} to model M_0 are defined as follows:

$$B_{j0}^{\mathrm{AIBF}}(r_j) = B_{j0}(\boldsymbol{x}|r_j) \cdot \frac{1}{L} \sum_{l=1}^{L} B_{0j}(\boldsymbol{x}(l)|r_j), \tag{6}$$

$$B_{j0}^{\mathrm{GIBF}}(r_j) = B_{j0}(\boldsymbol{x}|r_j) \cdot \left\{ \prod_{l=1}^{L} B_{0j}(\boldsymbol{x}(l)|r_j) \right\}^{1/L}, \tag{7}$$

and

$$B_{j0}^{\mathrm{MIBF}}(r_j) = B_{j0}(\boldsymbol{x}|r_j) \cdot \underset{1 \le l \le L}{\mathrm{Median}} \{B_{0j}(\boldsymbol{x}(l)|r_j)\}, \tag{8}$$

where $B_{0j}(\boldsymbol{x}(l)|r_j) = m_0(\boldsymbol{x}(l))/m_j(r_j, \boldsymbol{x}(l))$, and $m_j(r_j, \boldsymbol{x}(l))$ is defined by replacing the full sample $\boldsymbol{x}$ by the lth minimal training sample $\boldsymbol{x}(l)$ in equation (5). The posterior probability of each model is more useful than the Bayes factor for selecting one model among more than two models. That is, we can select the most plausible model, which has the largest posterior probability. The posterior probability of model M_{j,r_j} with a fixed change point r_j is defined by

$$P(M_{j,r_j}|\boldsymbol{x}) = \left\{ \sum_{i=0}^{3} \sum_{r_i \in I} \left\{ \frac{p_i}{p_j} \right\} \left\{ \frac{B_{i0}^*(r_i)}{B_{j0}^*(r_j)} \right\} \right\}^{-1}, \quad j = 1, 2, 3, \tag{9}$$

where B^* denotes the AIBF, the GIBF, or the MIBF, and p_0 and p_j are the prior model probability of M_0 and M_{j,r_j}, $(j = 1, 2, 3)$ respectively. Finally, the posterior probability of each change model M_j is computed by

$$P(M_j|\boldsymbol{x}) = \sum_{r_j \in I} P(M_{j,r_j}|\boldsymbol{x}), \quad j = 1, 2, 3. \tag{10}$$

3. Computation

We compute several IBFs such as the AIBF, the GIBF, and the MIBF to choose one model among the four models, M_0, M_1, M_2, and M_3. For the moment, we assume that the change point r_j is known for $j = 1, 2, 3$. We use noninformative improper priors for each model to reflect initial diffuseness. They are denoted by π_j, $j = 0, 1, 2, 3$. Then the priors are

$$\begin{cases} \pi_0(\mu_1, \Sigma_1) & = c_0|\Sigma_1|^{-(p+1)/2}, & \Sigma_1 > 0, \\ \pi_1(\mu_1, \mu_2, \Sigma_1) & = c_1|\Sigma_1|^{-(p+1)/2}, & \Sigma_1 > 0, \\ \pi_2(\mu_1, \Sigma_1, \Sigma_2) & = c_2\{|\Sigma_1| \cdot |\Sigma_2|\}^{-(p+1)/2}, & \Sigma_1 > 0, \quad \Sigma_2 > 0, \\ \pi_3(\mu_1, \mu_2, \Sigma_1, \Sigma_2) & = c_3\{|\Sigma_1| \cdot |\Sigma_2|\}^{-(p+1)/2}, & \Sigma_1 > 0, \quad \Sigma_2 > 0, \end{cases}$$

where c_j is an undefined normalizing constant. Throughout this article, we use the following notations:

$$\begin{aligned} V_d^e &= \sum_{j=d}^{e}(x_j - \bar{x}_d^e)(x_j - \bar{x}_d^e)', \\ U_d^e(\mu) &= V_d^e + (e - d + 1)(\mu - \bar{x}_d^e)(\mu - \bar{x}_d^e)', \\ W(d, e, f) &= \int \{1 + (\mu - \bar{x}_d^e)' \, (S_d^e)^{-1}(\mu - \bar{x}_d^e)\}^{-(e-d+1)/2} \\ &\quad \times \{1 + (\mu - \bar{x}_{e+1}^f)' \, (S_{e+1}^f)^{-1}(\mu - \bar{x}_{e+1}^f)\}^{-(f-e)/2} \mathrm{d}\mu, \end{aligned} \tag{11}$$

where $\bar{x}_d^e = \sum_{j=d}^{e} x_j/(e - d + 1)$, $S_d^e = V_d^e/(e - d + 1)$. The likelihood function $l_j(\cdot|\boldsymbol{x})$ under each M_0 model and M_{j,r_j} $(j = 1, 2, 3)$ model is given by

$$\begin{cases} l_0(\mu_1, \Sigma_1|\boldsymbol{x}) = (2\pi)^{-np/2}|\Sigma_1|^{-n/2} \exp\left\{-\dfrac{1}{2}tr\{\Sigma_1^{-1}U_1^n(\mu_1)\}\right\}, \\ l_1(\mu_1, \mu_2, \Sigma_1, r_1|\boldsymbol{x}) = (2\pi)^{-np/2}|\Sigma_1|^{-n/2} \exp\left\{-\dfrac{1}{2}tr\{\Sigma_1^{-1}U_1^{r_1}(\mu_1)\}\right\} \\ \qquad\qquad \times \exp\left\{-\dfrac{1}{2}tr\{\Sigma_1^{-1}U_{r_1+1}^n(\mu_2)\}\right\}, \\ l_2(\mu_1, \Sigma_1, \Sigma_2, r_2|\boldsymbol{x}) = (2\pi)^{-np/2}|\Sigma_1|^{-r_2/2}|\Sigma_2|^{-n-r_2/2} \exp\left\{-\dfrac{1}{2}tr\{\Sigma_1^{-1}U_1^{r_2}(\mu_1)\}\right\} \\ \qquad\qquad \times \exp\left\{-\frac{1}{2}tr\{\Sigma_2^{-1}U_{r_2+1}^n(\mu_1)\}\right\}, \\ l_3(\mu_1, \mu_2, \Sigma_1, \Sigma_2, r_3|\boldsymbol{x}) = (2\pi)^{-np/2}|\Sigma_1|^{-r_3/2}|\Sigma_2|^{-(n-r_3)/2} \exp\left\{-\dfrac{1}{2}tr\{\Sigma_1^{-1}U_1^{r_3}(\mu_1)\}\right\} \\ \qquad\qquad \times \exp\left\{-\dfrac{1}{2}tr\{\Sigma_2^{-1}U_{r_3+1}^n(\mu_2)\}\right\}. \end{cases}$$

PROPOSITION 1 *The marginal densities for the full sample are given by*

$$m_0(\boldsymbol{x}) = \frac{c_0 \prod_{i=1}^{p} \Gamma((n - i)/2)}{n^{p/2}\pi^{p(2n-p-1)/4}|V_1^n|^{(n-1)/2}},$$

$$m_1(r_1, \boldsymbol{x}) = c_1 \cdot \pi_1(r_1) \cdot \left\{ \frac{\prod_{i=1}^{p} \Gamma((n-i-1)/2)}{r_1^{p/2}(n-r_1)^{p/2}\pi^{p(2n-p-3)/4}|V_1^{r_1} + V_{r_1+1}^{n}|^{(n-2)/2}} \right\},$$

$$m_2(r_2, \boldsymbol{x}) = c_2 \cdot \pi_2(r_2) \cdot \left\{ \frac{W(1, r_2, n)\prod_{i=1}^{p} \Gamma((r_2-i+1)/2)\Gamma((n-r_2-i+1)/2)}{\pi^{p(n-p+1)/2}|V_1^{r_2}|^{r_2/2}|V_{r_2+1}^{n}|^{(n-r_2)/2}} \right\},$$

$$m_3(r_3, \boldsymbol{x}) = c_3 \cdot \pi_3(r_3) \cdot \left\{ \frac{\prod_{i=1}^{p} \Gamma((r_3-i)/2)\Gamma((n-r_3-i)/2)}{r_3^{p/2}(n-r_3)^{p/2}\pi^{p(n-p-1)/2}|V_1^{r_3}|^{(r_3-1)/2}|V_{r_3+1}^{n}|^{(n-r_3-1)/2}} \right\}.$$

Proof The proof is fairly simple, and thus is omitted. ■

Now, let the change point r_j be unknown. If we use the uniform priors for r_j, then the predictive density for the change model M_j is

$$m_j(\boldsymbol{x}) = \sum\nolimits_{r_j \in I} m_j(r_j, \boldsymbol{x}), \quad j = 1, 2, 3.$$

Also, the posterior probability density $f_j(r_j|\boldsymbol{x})$ of change point r_j for given data $\boldsymbol{x}$ is

$$f_j(r_j|x) = \frac{m_j(r_j, x)}{m_j(x)}, \quad j = 1, 2, 3. \tag{12}$$

PROPOSITION 2 *The size of the minimal training sample $x(l)$ is $2(p+1)$.*

Proof Note that model M_{3,r_3} is the encompassing model in the sense that other models M_0, M_{1,r_1}, and M_{2,r_2} are nested within M_{3,r_3}. So, it is sufficient to check how many training samples are needed in order for $m_3^N(r_3, \boldsymbol{x})$ to be finite. We use Result 3.3 of Johnson and Wichern [17]. If $r_3 \le p$, that is, sample size $\le$ number of variables, then $|V_1^{r_3}| = 0$. Hence, the size of a minimal training sample is $2(p+1)$. ■

Remark 1 Our minimal training sampling plan is the simple random sampling for which each $p+1$ observations are chosen at both sides centering the change point.

Remark 2 A uniform prior $\pi_j(r_j) = 1/d_{r_j}$ for the change point r_j is assumed, where $r_j (r_a \le r_j \le r_b)$ is an integer and $d_{r_j} = r_b - r_a + 1$. The support $[r_a, r_b]$ of $\pi_j(r_j)$ is restricted in order that the predictive density of the minimal training sample can be finite. Thus, we set $r_a = p+1$ and $r_b = n-p-1$ according to our minimal training sampling scheme.

We need the marginal densities $m_0(\boldsymbol{x}(l))$ and $m_j(r_j, \boldsymbol{x}(l))$ $(j = 1, 2, 3)$ to compute the IBFs in equations (6), (7), and (8). These can be easily obtained by simple replacement in $m_0(\boldsymbol{x})$ and $m_j(r_j, \boldsymbol{x})$. That is, replacing n by $2(p+1)$, r_j and $n - r_j$ by $p+1$, and $\boldsymbol{x}$ by $\boldsymbol{x}(l)$ yields predictive densities of the minimal training sample. We note that the marginal densities of M_{2,r_2} for both the full sample and the training sample do not have closed forms. We need to compute the following integral

$$W(d, e, f) = \int_{-\infty}^{\infty} g(\mu)\mathrm{d}\mu = \int_{-\infty}^{\infty} \frac{g(\mu)}{f_\mu(\mu)} \cdot f_\mu(\mu)\mathrm{d}\mu = E_{f_\mu(\mu)}\left[\frac{g(\mu)}{f_\mu(\mu)}\right], \tag{13}$$

where $g(\mu) = \prod_{k=1}^{2}(1+Q_k)^{-N_k/2}$, $Q_1 = (\mu - \bar{x}_d^e)'(S_d^e)^{-1}(\mu - \bar{x}_d^e)$, $Q_2 = (\mu - \bar{x}_{e+1}^f)'(S_{e+1}^f)^{-1}(\mu - \bar{x}_{e+1}^f)$, $N_1 = e - d + 1$, and $N_2 = f - e$. Here, μ is the random vector with

pdf $f_\mu(\mu)$, called the importance sampling density function. Then, the Monte Carlo estimate of the integral in equation (13) is

$$\hat{W} = \frac{1}{N}\sum_{i=1}^{N}\frac{g(\mu_i)}{f_\mu(\mu_i)},$$

where $\{\mu_1, \mu_2, \ldots, \mu_N\}$ is a sample of size N generated from the density $f_\mu(\mu)$.

PROPOSITION 3 *For the integral* (13) *an approximate optimal importance sampling distribution, minimizing* $var(\hat{W})$*, is set as a multivariate normal distribution with mean* μ_0 *and covariance matrix* J^{-1}*, where*

$$\mu_0 = K\bar{x}_d^e + (I-K)\bar{x}_{e+1}^f, \quad K = N_1(S_d^e J)^{-1}, \quad \textit{and} \quad J = N_1(S_d^e)^{-1} + N_2(S_{e+1}^f)^{-1}.$$

Proof It can be proved by similar arguments in the proof of Proposition 3 of Kim and Kim [18].

■

4. Numerical studies

In this section, we present some simulation results and a real data analysis to check the performance of default Bayes factors and the posterior probability density of change point appeared in section 3. We refer to simulation schemes in Marks and Dunn [19], who conducted a test on the equality of two independent multivariate normal distributions. All experiments are performed on the basis of the sequences of independent quad-variate normal random vectors with a sample size $n = 50$. We use a change point $r_j = n/2 = 25$ $(j = 1, 2, 3)$. We set $\mu_1 = 0_p$ and $\mu_2 = [\tau(1+\sqrt{\lambda})\,, 0'_{p-1}]'$, where 0_p is a p-variate column vector of zeroes. Without loss of generality, we set $\Sigma_1 = I$ and $\Sigma_2 = \Lambda$, where I is an identity matrix and Λ is a diagonal matrix with all λ's being diagonal elements. This setting is based on the simultaneous diagonalization theorem and the invariance property: suppose that $X_i(i = 1, 2, \ldots, r_3) \sim N_p(\mu_1, \Sigma_1)$ and $X_i(i = r_3+1, r_3+2, \ldots, n) \sim N_p(\mu_2, \Sigma_2)$. Consider the transformations $X_i^* = CX_i + v_1$, $i = 1, 2, \ldots, r_3$ and $X_i^* = CX_i + v_2$, $i = r_3+1, r_3+2, \ldots, n$, where C is a nonsingular matrix and v_1 and v_2 are any vectors. Then, $X_i^*(i = 1, 2, \ldots, r_3) \sim N_p(\mu_1^*, \Sigma_1^*)$ and $X_i^*(i = r_3+1, r_3+2, \ldots, n) \sim N_p(\mu_2^*, \Sigma_2^*)$, where $\mu_1^* = C\mu_1$, $\mu_2^* = C\mu_2$, $\Sigma_1^* = C\Sigma_1 C'$, and $\Sigma_2^* = C\Sigma_2 C'$. Also, there exists a nonsingular matrix C such that $C\Sigma_1 C' = I$ and $C\Sigma_2 C' = \Lambda$. Hence, $B_{j0}^{\mathrm{AIBF}}(r_j)$, $B_{j0}^{\mathrm{GIBF}}(r_j)$, and $B_{j0}^{\mathrm{MIBF}}(r_j)$ are all invariant for these transformations. Here, $\tau(\lambda)$ is called the measure of degree for separation between means (covariances) of two populations. We use different choices of $\tau = 0, 0.5, 1, 1.5, 2$ and $\lambda = 1, 2, 4, 6, 8$ to generate random variates from a multivariate normal distribution. We randomly select 30 sets of minimal training samples of each $p+1$ on both sides centering the change point by simple random sampling. We use the importance samples of size 100 in order to compute the marginal for M_{2,r_2}. We calculated the average posterior probabilities for each model on the basis of 1000 replications. They are computed under the assumption of equal prior model probability. Numerical values are given in table 1. Because of space limit, the results of the GIBF and the MIBF are omitted. However, as a substitute for comparing three IBFs, we compute the 'acceptance rate' of each model provided in figure 1. This acceptance rate is the measure for which the posterior probability supports the true model.

From table 1 and figure 1, we conjecture that (i) generally the AIBF is almost equal to or better than the GIBF or the MIBF except for the true model M_0 in the sense of high accurate

Table 1. Simulated averages and standard deviations in parentheses of posterior probabilities of each model computed through the AIBF.

τ	λ	Model	$P(M_0\|\boldsymbol{x})$	$P(M_1\|\boldsymbol{x})$	$P(M_2\|\boldsymbol{x})$	$P(M_3\|\boldsymbol{x})$
0.0	1	M_0	0.7904 (0.1691)	0.1976 (0.1664)	0.0095 (0.0332)	0.0025 (0.0155)
	2	M_2	0.7422 (0.2141)	0.2127 (0.1917)	0.0381 (0.1109)	0.0069 (0.0414)
	4	M_2	0.2092 (0.2846)	0.0867 (0.1665)	0.6839 (0.3664)	0.0203 (0.0592)
	6	M_2	0.0147 (0.0748)	0.0088 (0.0546)	0.9471 (0.1386)	0.0295 (0.0845)
	8	M_2	0.0001 (0.0016)	0.0001 (0.0021)	0.9626 (0.0987)	0.0371 (0.0987)
0.5	1	M_1	0.3617 (0.2953)	0.6305 (0.3005)	0.0062 (0.0418)	0.0016 (0.0124)
	2	M_3	0.3451 (0.2852)	0.6247 (0.2958)	0.0215 (0.0875)	0.0087 (0.0446)
	4	M_3	0.1035 (0.1922)	0.2240 (0.3142)	0.4534 (0.3676)	0.2192 (0.2851)
	6	M_3	0.0041 (0.0353)	0.0162 (0.0907)	0.6666 (0.3383)	0.3130 (0.3282)
	8	M_3	0.0001 (0.0016)	0.0012 (0.0283)	0.6787 (0.3271)	0.3200 (0.3267)
1.0	1	M_1	0.0066 (0.0330)	0.9933 (0.0333)	0.0001 (0.0010)	0.0001 (0.0016)
	2	M_3	0.0054 (0.0233)	0.9806 (0.0890)	0.0005 (0.0062)	0.0135 (0.0841)
	4	M_3	0.0028 (0.0168)	0.3726 (0.4082)	0.0280 (0.0876)	0.5966 (0.4023)
	6	M_3	0.0002 (0.0030)	0.0266 (0.1252)	0.0657 (0.1587)	0.9076 (0.1975)
	8	M_3	0.0000 (0.0000)	0.0002 (0.0037)	0.0852 (0.1876)	0.9146 (0.1876)
1.5	1	M_1	0.0000 (0.0000)	1.0000 (0.0004)	0.0000 (0.0000)	0.0000 (0.0004)
	2	M_3	0.0000 (0.0001)	0.9873 (0.0815)	0.0000 (0.0000)	0.0127 (0.0815)
	4	M_3	0.0000 (0.0000)	0.4084 (0.4237)	0.0004 (0.0032)	0.5913 (0.4234)
	6	M_3	0.0000 (0.0000)	0.0257 (0.1208)	0.0027 (0.0271)	0.9716 (0.1233)
	8	M_3	0.0000 (0.0000)	0.0015 (0.0320)	0.0022 (0.0139)	0.9963 (0.0348)
2.0	1	M_1	0.0000 (0.0000)	0.9998 (0.0038)	0.0000 (0.0000)	0.0002 (0.0038)
	2	M_3	0.0000 (0.0000)	0.9907 (0.0601)	0.0000 (0.0000)	0.0093 (0.0601)
	4	M_3	0.0000 (0.0000)	0.4042 (0.4219)	0.0000 (0.0000)	0.5958 (0.4219)
	6	M_3	0.0000 (0.0000)	0.0336 (0.1447)	0.0000 (0.0002)	0.9664 (0.1447)
	8	M_3	0.0000 (0.0000)	0.0025 (0.0453)	0.0001 (0.0008)	0.9974 (0.0453)

rate; (ii) when the degree of separation between covariances is $\lambda = 2$, three IBFs are useless in finding true models, in the sense that the model cannot detect the change of covariance; (iii) when $\lambda = 4, 6,$ and 8, the change of covariance overwhelms the change of mean in the mean and covariance model, which results in selecting the covariance model; (iv) when the degree of separation between means (covariances) is $\tau = 0.5(\lambda \neq 1)$, the acceptance rates by three IBFs are less than $1/4$ or almost equal. When the IBF is weak in detecting the true model, it could be the case that the competing models are very close together. Thus, it may be better, in terms of predictions, to act as if the simpler model is the sampling model.

Figures 2–5 are the histograms showing the proportions of the mode that maximize the posterior density in equation (12). As shown in figure 5, when the degree of change is quite large ($\lambda = 1.5$), the mean change model detects the true change point with proportion over 0.8 regardless of the change of variance. The U-type histograms in figures 3a and 4a result from the fact that the degree of change is low. However, these U-type histograms should be moved on to the degenerate histogram with the proportion 1 at the true change point as the degree of change gets larger.

Sullivan and Woodall [20] used a preliminary control chart to detect a change in the mean vector, the covariance matrix, or in both on the basis of a sequence of multivariate observations. In addition, they applied this methodology to a dataset in Holmes and Mergen [21]. This dataset consists of 56 individual bivariate observations from a European plant producing grit or gravel for which the percentages of particles (by weight) are large and medium in size.

The data with a total of 56 observations are given in the second and the third columns of table 2. Figure 6 shows a plot of whole data named as Group 0 (observations 1–56). Two elliptical regions represent 50% and 90% confidence regions. Numerical results from this analysis are presented in table 3. All Bayes factors strongly prefer a mean change model M_1.

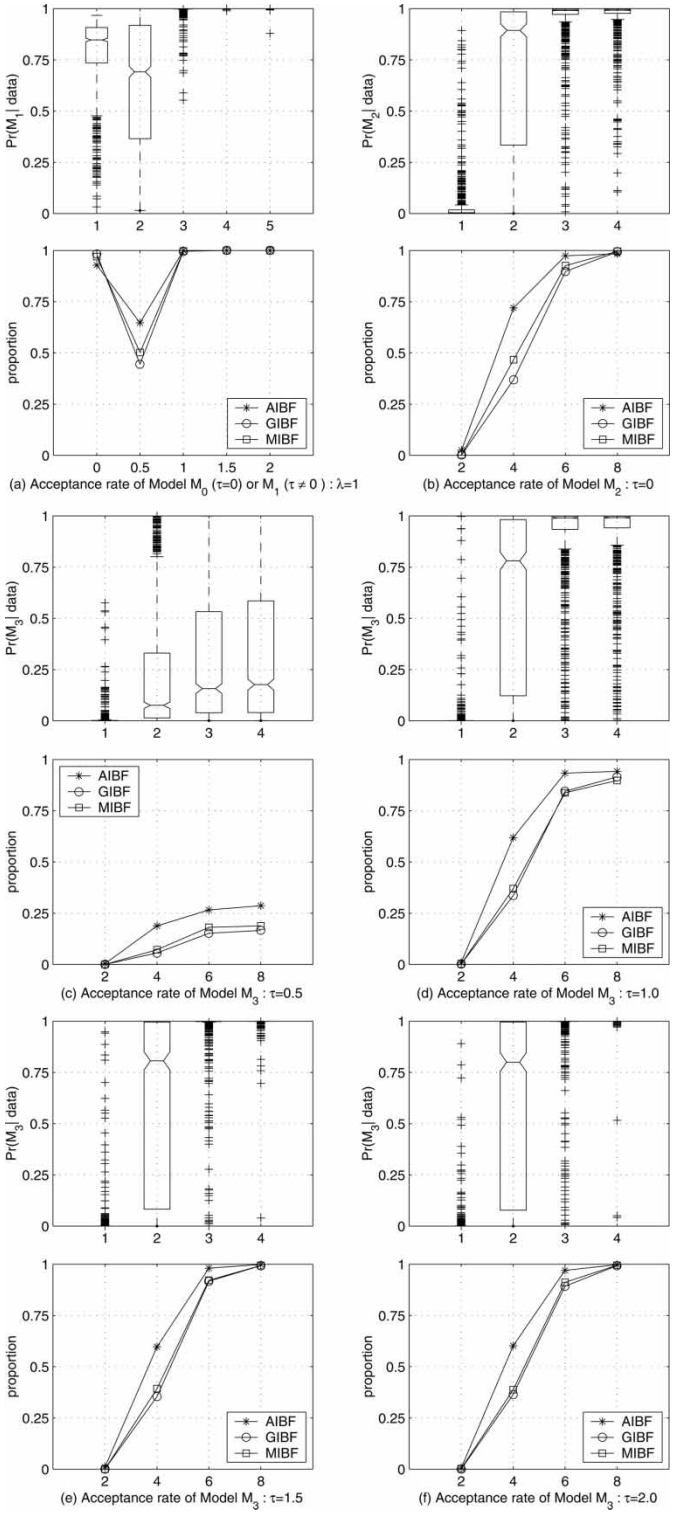

Figure 1. Box-plots of posterior probabilities for each model through the AIBF and the acceptance rate of each model by the AIBF, the GIBF, and the MIBF.

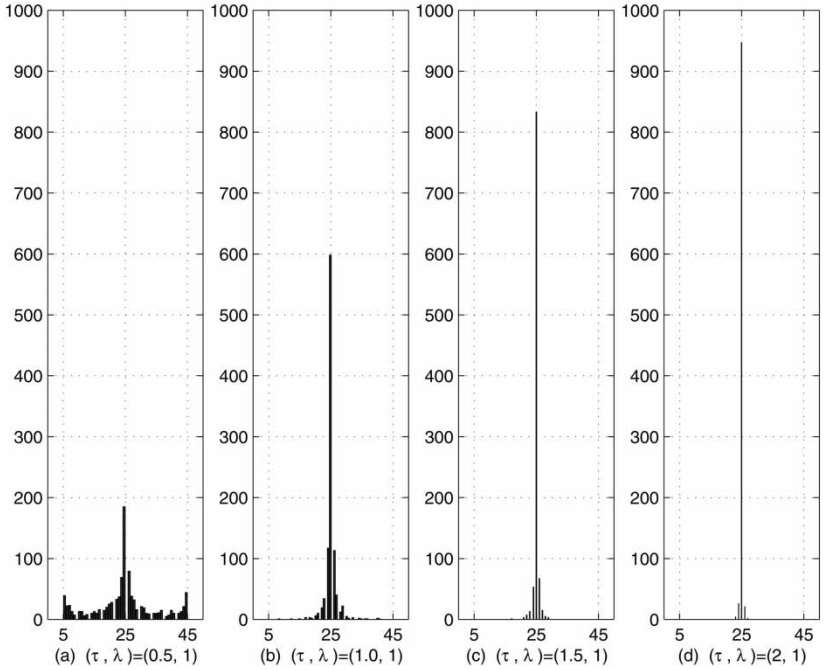

Figure 2. Histograms of proportions on 1000 replications of the mode of change point for a mean change model.

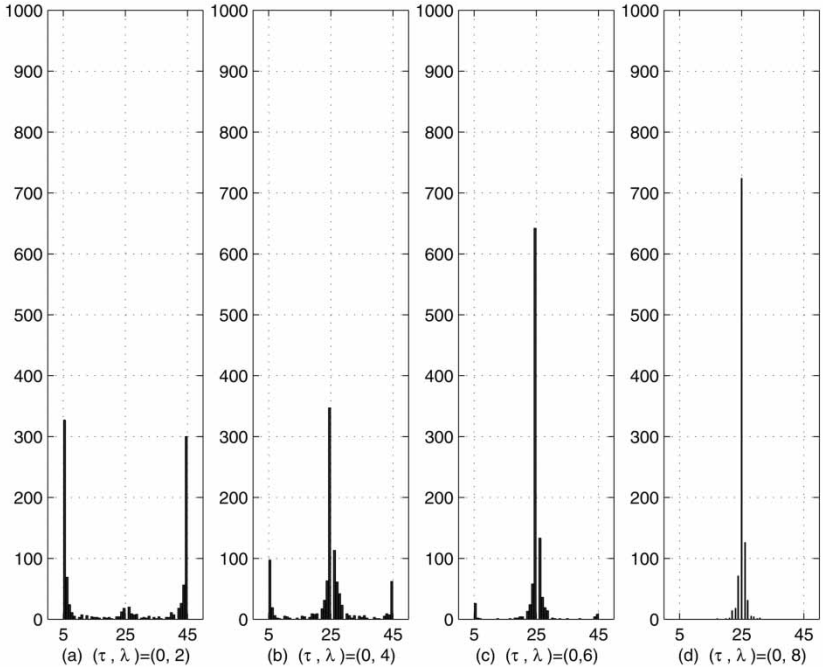

Figure 3. Histograms of proportions on 1000 replications of the mode of change point for a covariance change model.

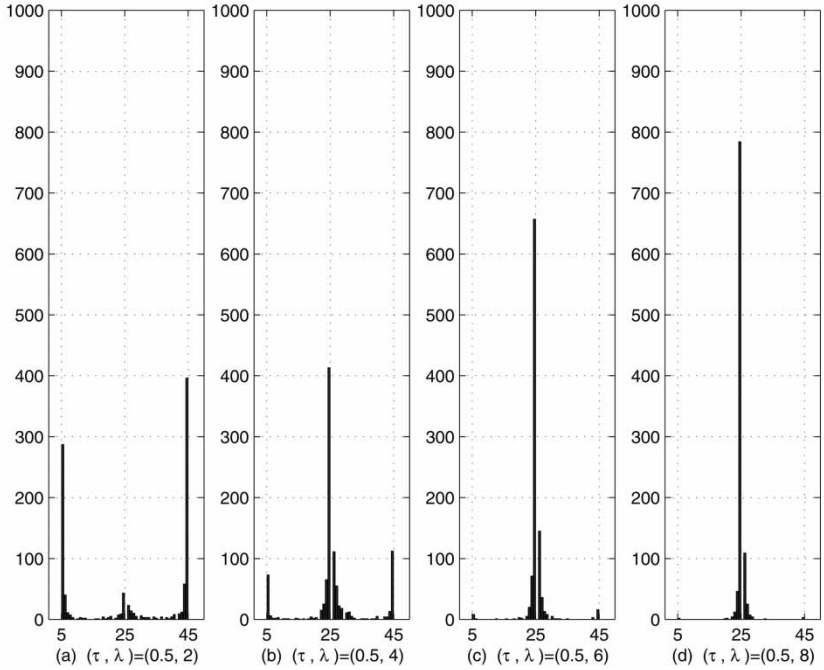

Figure 4. Histograms of proportions on 1000 replications of the mode of change point for a mean and covariance change model with $\tau = 0.5$.

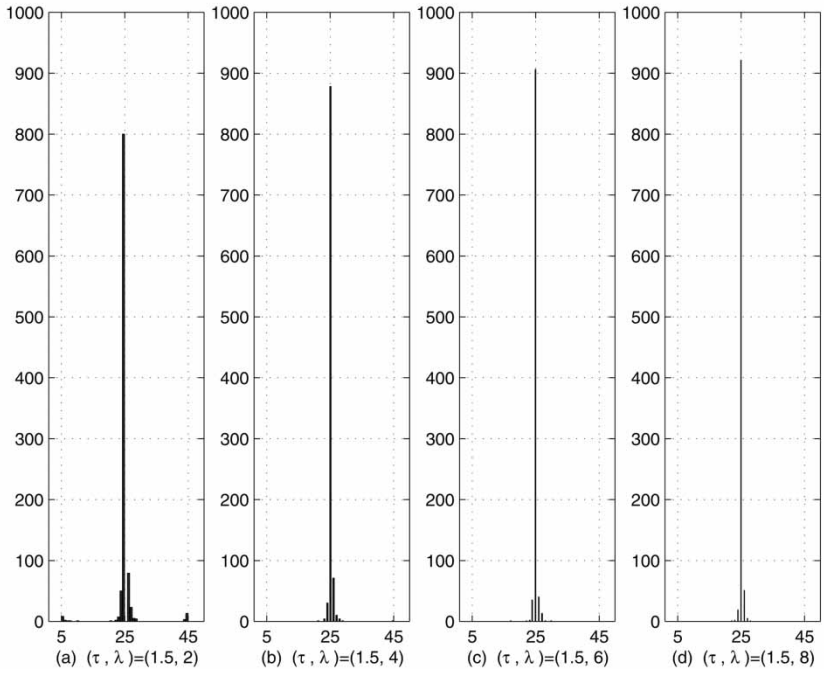

Figure 5. Histograms of proportions on 1000 replications of the mode of change point for a mean and covariance change model with $\tau = 1.5$.

Table 2. Posterior probabilities of change points for three data groups with changes.

Observation no.	Large (%)	Medium (%)	Group 0	Group 2		Group 3	
			$f_1(r_1\|\boldsymbol{x})$	$f_1(r_1\|\boldsymbol{x})$	$f_3(r_3\|\boldsymbol{x})$	$f_1(r_1\|\boldsymbol{x})$	$f_3(r_3\|\boldsymbol{x})$
1	5.4	93.6	·	·	·	·	·
2	3.2	92.6	·	·	·	·	·
3	5.2	91.7	0.0000	·	·	0.0000	0.0000
4	3.5	86.9	0.0000	·	·	0.0000	0.0000
5	2.9	90.4	0.0000	·	·	0.0000	0.0000
6	4.6	92.1	0.0000	·	·	0.0000	0.0000
7	4.4	91.5	0.0000	·	·	0.0000	0.0000
8	5	90.3	0.0000	·	·	0.0000	0.0000
9	8.4	85.1	0.0000	·	·	0.0000	0.0000
10	4.2	89.7	0.0000	·	·	0.0000	0.0000
11	3.8	92.5	0.0000	·	·	0.0000	0.0000
12	4.3	91.8	0.0000	·	·	0.0000	0.0000
13	3.7	91.7	0.0000	·	·	0.0000	0.0000
14	3.8	90.3	0.0000	·	·	0.0000	0.0000
15	2.6	94.5	0.0000	·	·	0.0000	0.0000
16	2.7	94.5	0.0004	·	·	0.0004	0.0000
17	7.9	88.7	0.0002	·	·	0.0005	0.0000
18	6.6	84.6	0.0000	·	·	0.0001	0.0000
19	4	90.7	0.0001	·	·	0.0002	0.0000
20	2.5	90.2	0.0002	·	·	0.0002	0.0000
21	3.8	92.7	0.0012	·	·	0.0013	0.0000
22	2.8	91.5	0.0082	·	·	0.0068	0.0001
23	2.9	91.8	0.0858	·	·	0.0755	0.0028
24	3.3	90.6	0.7407	·	·	0.7359	0.4632
25	7.2	87.3	0.1505	·	·	0.1777	0.5158
26	7.3	79	0.0063	·	·	0.0012	0.0094
27	7	82.6	0.0014	0.0099	0.0301	0.0001	0.0012
28	6	83.5	0.0010	0.0343	0.0044	0.0000	0.0064
29	7.4	83.6	0.0003	0.0452	0.0053	0.0000	0.0009
30	6.8	84.8	0.0002	0.0500	0.0092	0.0000	0.0001
31	6.3	87.1	0.0001	0.0222	0.0049	0.0000	0.0000
32	6.1	87.2	0.0001	0.0138	0.0035	0.0000	0.0000
33	6.6	87.3	0.0001	0.0069	0.0044	0.0000	0.0000
34	6.2	84.8	0.0001	0.0151	0.0155	0.0000	0.0000
35	6.5	87.4	0.0000	0.0086	0.0255	0.0000	0.0000
36	6	86.8	0.0001	0.0101	0.0441	0.0000	0.0000
37	4.8	88.8	0.0001	0.0153	0.0045	0.0000	0.0000
38	4.9	89.8	0.0003	0.0192	0.0027	0.0000	0.0000
39	5.8	86.9	0.0005	0.0401	0.0090	0.0000	0.0000
40	7.2	83.8	0.0004	0.0761	0.0282	0.0000	0.0000
41	5.6	89.2	0.0005	0.0820	0.0525	·	·
42	6.9	84.5	0.0006	0.1918	0.2611	·	·
43	7.4	84.4	0.0005	0.2556	0.4838	·	·
44	8.9	84.3	0.0001	0.0576	0.0083	·	·
45	10.9	82.2	0.0000	0.0122	0.0001	·	·
46	8.2	89.8	0.0000	0.0019	0.0000	·	·
47	6.7	90.4	0.0000	0.0008	0.0000	·	·
48	5.9	90.1	0.0000	0.0006	0.0000	·	·
49	8.7	83.6	0.0000	0.0008	0.0000	·	·
50	6.4	88	0.0000	0.0008	0.0000	·	·
51	8.4	84.7	0.0000	0.0014	0.0000	·	·
52	9.6	80.6	0.0000	0.0213	0.0025	·	·
53	5.1	93	0.0000	0.0065	0.0005	·	·
54	5	91.4	·	·	·	·	·
55	5.9	87.2	·	·	·	·	·
56	5.9	87.2	·	·	·	·	·

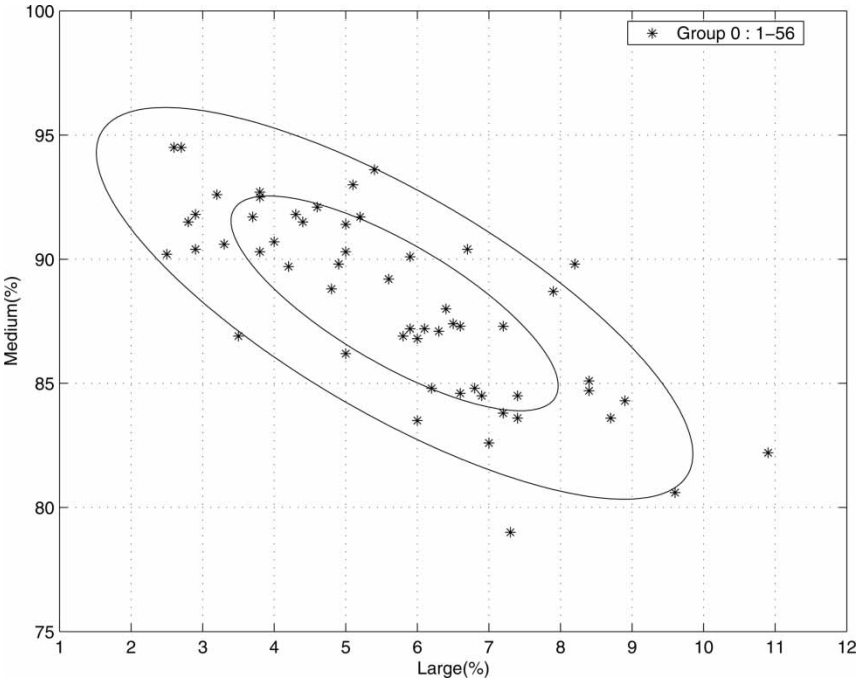

Figure 6. Plot of gravel data with a 50% and a 90% confidence ellipse.

The fourth column of table 2 indicates that the change point in Group 0 is $r_1 = 24$ with the posterior probability 0.7407.

Figure 7 contains several plots of gravel data divided into several groups. First, we assign the first 24 observations to Group 1 and the rest to Group 2. In figure 7a, 50% and 90% confidence regions of the two groups depict visually that there is a mean change but almost no covariance change. All Bayes factors for Group 1 in table 3 select a no-change model M_0. For Group 2, the AIBF selects a mean and covariance change model M_3, whereas the GIBF and the MIBF select a mean change model M_1. Columns five and six in table 2 are posterior probabilities of change points for data of Group 2 under a mean change model and a mean and covariance change model, respectively. It shows that there are change points $r_1 = 43$ and

Table 3. Posterior probabilities of each model for four data groups.

Group	Observation no.	Bayes factor	$P(M_0\|\boldsymbol{x})$	$P(M_1\|\boldsymbol{x})$	$P(M_2\|\boldsymbol{x})$	$P(M_3\|\boldsymbol{x})$
0	1–56	AIBF	0.0001	0.9963	0.0000	0.0035
		GIBF	0.0005	0.9988	0.0000	0.0007
		MIBF	0.0003	0.9988	0.0000	0.0009
1	1–24	AIBF	0.3667	0.1963	0.2122	0.2248
		GIBF	0.7054	0.1704	0.0904	0.0338
		MIBF	0.6278	0.1937	0.1340	0.0445
2	25–56	AIBF	0.0560	0.3209	0.2099	0.4131
		GIBF	0.2082	0.4876	0.1672	0.1370
		MIBF	0.1581	0.4629	0.2164	0.1627
3	1–43	AIBF	0.0001	0.3446	0.0008	0.6545
		GIBF	0.0012	0.7468	0.0015	0.2505
		MIBF	0.0010	0.7377	0.0019	0.2594

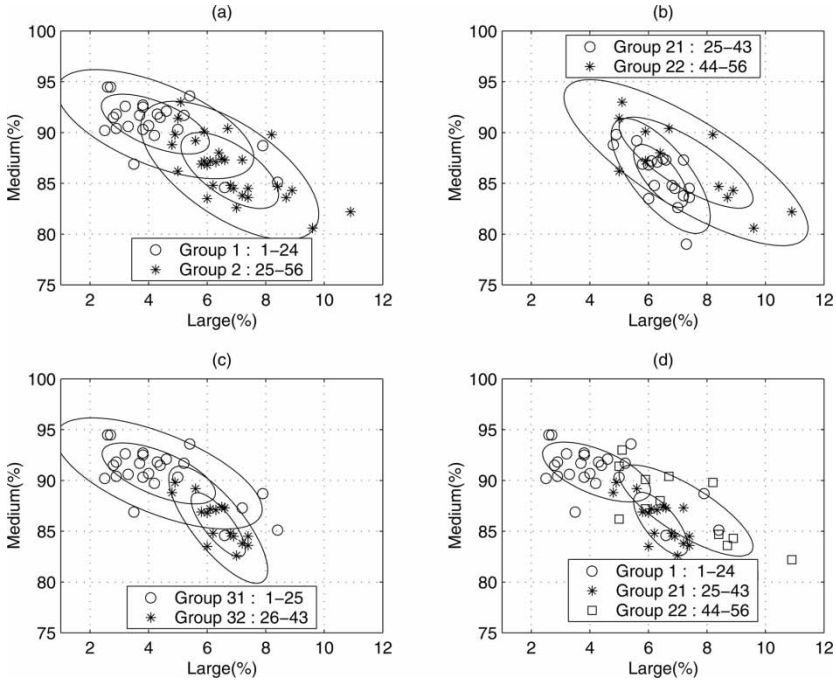

Figure 7. Plots of gravel data divided by several groups.

$r_3 = 43$. Figure 7b is the plot of 50% and 90% confidence regions for the two groups, Group 21 (observations 25–43) and Group 22 (observations 44–56). It visually indicates that there are changes not only in the mean but in the covariance.

Suppose that we consider a portion of the data with observations 1–43 and denote this by Group 3. It turned out that a mean and covariance change model is selected by the AIBF, whereas a mean change model is selected by the GIBF and the MIBF. From the seventh and the eighth columns of table 2, the change point for Group 3 is $r_1 = 24$ on the basis of the result of the GIBF and the MIBF, and $r_3 = 25$ and $r_3 = 24$ with a little smaller posterior probability under the selection of the AIBF. It is due to the fact that eliminating observations 44–56 makes the change point to be shifted from $r_1 = 24$ for Group 0 (the whole data) to $r_3 = 25$ for Group 3. Plots of Group 31 (observations 1–25) and Group 32 (observations 26–43) in figure 7c indicate the existence of change in both the mean and the covariance. Finally, three elliptical regions in figure 7d are 50% confidence regions of three datasets, Group 1 (observations 1-24), Group 21 (observations 25–43), and Group 22 (observations 44–56). On the basis of the results shown in table 3, for Groups 2 and 3, which seem to be a mixture of shifts in the mean vector and the covariance matrix, the GIBF and the MIBF are insensitive to detecting covariance change in the perspective of a combined shift. We can see that the results of Bayesian change point detection by the AIBF coincide with our visual output and also with those of Sullivan and Woodall [20] by the control chart.

5. Concluding remarks

In this article, we have identified a no-change model and three change models with a single change point using the AIBF, the GIBF, and the MIBF. Once a change point model is selected, the change point is estimated from the posterior probability density of the change point. The

fractional Bayes factor (FBF) of O'Hagan [22] can also be considered as a default or an automatic Bayes factor. Generally, the FBF is somewhat simpler than the IBF in the sense that the FBF does not need training sample computation. However, the FBF approach has some shortcomings in change point problems discussed in this article. That is, the possible range of the change point is very restrictive as the sample size gets larger. When computing the FBF, the range of change points must be adjusted to ensure that the arguments of the gamma functions included in the correction term of the FBF are positive. After analysing gravel data, we can see that default Bayesian procedure for detecting the change with only one change point is recursively repeated and can be used in detecting multiple change points.

Acknowledgements

This work was supported by Grant No. R02-2000-00024 from the Basic Research Program of the Korea Science Engineering Foundation.

References

[1] Smith, A.F.M., 1975, A Bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika*, **62**, 407–416.
[2] Lee, A.F.S. and Heghinian, S.M., 1977, A shift of the mean level in a sequence of independent normal random variables – a Bayesian approach. *Technometrics*, **19**, 503–506.
[3] Menzefricke, U., 1981, A Bayesian analysis of a change in the precision of a sequence of independent normal random variables at an unknown time point. *Applied Statistics,* **30**, 141–146.
[4] Perreault, L., Haché, M., Slivitzky, M. and Bobée, B., 1999, Detection of changes in precipitation and runoff over eastern Canada and U.S. using a Bayesian approach. *Stochastic Environmental Research and Risk Assessment*, **13**, 201–216.
[5] Perreault, L., Bernier, J., Bobée, B. and Parent, E., 2000, Bayesian change-point analysis in hydrometeorological time series. Part 1. The normal model revisited. *Journal of Hydrology*, **235**, 221–241.
[6] Perreault, L., Bernier, J., Bobée, B. and Parent, E., 2000, Bayesian change-point analysis in hydrometeorological time series. Part 2. Comparison of change-point models and forecasting. *Journal of Hydrology*, **235**, 242–263.
[7] Booth, N.B. and Smith, A.F.M., 1982, A Bayesian approach to retrospective identification of change-points. *Journal of Economics,* **19**, 7–22.
[8] Perreault, L., Parent, E., Bernier, J., Bobée, B. and Slivitzky, M., 2000, Retrospective multivariate Bayesian change-point analysis: a simultaneous single change in the mean of several hydrological sequences. *Stochastic Environmental Research and Risk Assessment*, **14**, 243–261.
[9] Berger, J.O. and Pericchi, L.R., 1996, The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association,* **91**, 109–122.
[10] Varshavsky, J.A., 1996, Intrinsic Bayes factors for model selection with autoregressive data. In: J.M. Bernardo *et al.* (Eds) *Bayesian Analysis* V. (Oxford: Oxford University Press), pp. 757–763.
[11] Lingham, R.T. and Sivaganesan, S., 1997, Testing hypothesis about the power law process under failure truncation using intrinsic Bayes factors. *Annals of the Institute of Statistical Mathematics*, **49**, 693–710.
[12] Kim, S.W., 2000, Intrinsic priors for testing exponential means. *Statistics and Probability Letters*, **46**, 195–201.
[13] Kim, S.W. and Ibrahim, J.G., 2000, Default Bayes factors for generalized linear models. *Journal of Statistical Planning and Inference*, **87**, 301–315.
[14] Berger, J.O. and Pericchi, L.R., 1998, Accurate and stable Bayesian model selection: the median intrinsic Bayes factor. Sankhya Series B, **60**, 1–18.
[15] Berger, J.O. and Mortera, J., 1999, Default Bayes factors for nonnested hypothesis testing. *Journal of the American Statistical Association,* **94**, 542–554.
[16] Spiegelhalter, D.J. and Smith, A.F.M., 1982, Bayes factors for linear and log-linear models with vague prior information. *Journal of the Royal Statistical Society: Series B*, **44**, 377–387.
[17] Johnson, R.A. and Wichern, D.W., 1998, *Applied Multivariate Statistical Analysis* (4th edn) (Prentice-Hall, Inc).
[18] Kim, H.J. and Kim, S.W., 2001, Two-sample Bayesian test using intrinsic Bayes factors for multivariate normal observations. *Communications in Statistics: Series B*, **30**, 423–436.
[19] Marks, S. and Dunn, O.J., 1974, Discriminant functions when covariance matrices are unequal. *Journal of the American Statistical Association,* **69**, 555–559.
[20] Sullivan, J.H. and Woodall, W.H., 2000, Change-point detection of mean vector or covariance matrix shifts using multivariate individual observations. *IEEE Transactions*, **32**, 537–549.
[21] Holmes, D.S. and Mergen, A.E., 1993, Improving the performance of the T^2 control chart. *Quality Engineering*, **5**, 619–625.
[22] O'Hagan, A., 1995, Fractional Bayes factors for model comparison. *Journal of the Royal Statistical Society: Series B*, **57**, 99–138.

Copyright of Statistics is the property of Taylor & Francis Ltd. The copyright in an individual article may be maintained by the author in certain cases. Content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.