COMMUN. STATIST.—THEORY METH., 24(5), 1307-1324 (1995)

DETECTING CHANGE POINTS AND MONITORING BIOMEDICAL DATA

Heping Zhang

Yale University School of Medicine
New Haven, Connecticut 06520

Key Words: change point regression; likelihood ratio; linear dynamic model; posterior probability; renal transplant.

ABSTRACT

Bayesian and likelihood approaches to on-line detecting change points in time series are discussed and applied to analyze biomedical data. Using a linear dynamic model, the Bayesian analysis outputs the conditional posterior probability of a change at time $t-1$, given the data up to time t and the status of changes occurred before time $t-1$. The likelihood method is based on a change-point regression model and tests whether there is no change-point.

1. INTRODUCTION

In many situations, a critical decision depends on detecting a sudden change of phenomenon. For example, in screening breast cancer, among several tumor markers CA153 is measured for a period of time, and it is important to know if there is a change in the sequence of the CA153 marker.

Both Bayesian and frequentist methods are used to deal with change-point problems in the literature; see, e.g., Smith (1975) and Carlin, Gelfand, and Smith (1992) for Bayesian analysis, and Siegmund (1986) for the frequentist counterpart.

Motivated by the problem of monitoring renal transplants to see a rejection (or a change) has occurred to a patient who received a kidney transplant, Smith and West (1983) developed a Bayesian approach based on a linear dynamic model. Gordon and Smith (1990) extended the method to the case of unequally spaced data. One of the important problems that these authors addressed is the calculation of the posterior probability that a change occurred at

Copyright © 1995 by Marcel Dekker, Inc

time $t-1$ (or t) given the data up to time t. To evaluate the posterior probability, mixture normal and gamma densities were approximated by single densities. This approximation is not computationally efficient and its adequacy is not clear for the problem. For this consideration, a different but related question is posed here. In clinical applications, it is often necessary to decide on time whether there are changes in the data process. For the application described in Section 5, we need to judge at examination t whether a rejection of the transplanted kidney occurred to a patient since the last examination $t-1$. In making these judgments, we assume that rejections, if any, that have occurred to the patient prior to examination $t-1$, have been determined previously. Statistically, we derive the *conditional* posterior probability that a change occurred at time $t-1$ (or t), assuming that states of the data system are known prior to time t. We will obtain this posterior probability directly using some recursive formulas.

Besides, using a result derived from Hotelling's (1939) Theorem on the volume of tube for regression change-point model (or broken line regression) [see, e.g., Knowles and Siegmund (1989)], a likelihood approach to the on-line detection of change points is proposed. The key advantage of the likelihood method lies in its simplicity; however, its applicability is much more limited. This will be discussed later.

In the next section, a brief discussion of the Bayesian on-line analysis is given. The likelihood method is introduced in Section 3. These methods will be compared by simulations in Section 4 and applications in Section 5. Remarks will be made in Section 6 about the usage of these methods.

2. BAYESIAN ANALYSIS

2.1 The Basic Model

At time t, assume that y_t is observed; $t = 1, 2, \cdots$. Considered a linear dynamic model

$$y_t = H_t\theta_t + \delta y_t^{(j)}, \quad \theta_t = G_t\theta_{t-1} + \delta\theta_t^{(j)}, \tag{1}$$

for $j = 1, \cdots, J$, where θ_t is a vector of system parameters, H_t is a known regression matrix, G_t is a known transition matrix, and

$$\delta y_t^{(j)} \sim N(0, \lambda^{-1}R_y^{(j)}), \quad \delta\theta_t^{(j)} \sim N(0, \lambda^{-1}R_\theta^{(j)}), \tag{2}$$

where $R_y^{(j)}$ and $R_\theta^{(j)}$ are known, but λ is unknown. $\delta y_t^{(j)}$ and $\delta\theta_t^{(k)}$ are assumed to be independent for any t, j, and k.

$J = 4$ is relevant for our present discussion and corresponds to four distinct changes of state for the time series y_t. At time t, states $1, 2, 3$, and 4 stand for no change, a level change, a slope change, and an outlier, respectively. These patterns are exhibited clearly in Figure 4 of Gordon and Smith (1990). The state of y_t is determined by the covariance components $R_y^{(j)}$ and $R_\theta^{(j)}$.

The one-step back posterior probability $O_t^{(j)}$ is the probability that y_{t-1} is at state j given the data $y_1, \cdots, y_t$. Let $M_t^{(j)}$ denote the event that y_t is at state j and $\mathbf{D}_t = (y_t, \cdots, y_1)^T$. Then, $O_t^{(j)} = P(M_{t-1}^{(j)} \mid \mathbf{D}_t)$. We refer to Gordon and Smith (1990) for issues on computing $O_t^{(j)}$. Here, we are concerned with another question, i.e., what is the probability that y_{t-1} is at state j given the data $\mathbf{D}_t$ and the status of the system prior to time $t-1$, i.e., $M_1^{(i_1)}, \cdots, M_{t-2}^{(i_{t-2})}$. We call this probability the conditional one-step back posterior probability. A special case is when the system is stable before time $t-1$, and then this conditional probability detects the first change point.

We now proceed to calculate the conditional one-step back posterior probability, denoted again by $O_t^{(j)}$ for convenience. First, we specify prior information for the system parameters. Assume

$$p(\lambda) \sim \text{gamma}(\frac{1}{2}n_0, \frac{1}{2}r_0)$$

for some known n_0 and r_0. Given λ

$$p(\theta_0 \mid \lambda) \sim N(\mathbf{m}_0, \lambda^{-1}C_0)$$

for some known $\mathbf{m}_0$ and C_0. Then, conditional on data $\mathbf{D}_{t-1}$, $M_{t-1}^{(i_{t-1})}$, and $M_{t-2}^{(i_{t-2})}, \cdots, M_1^{(i_1)}$, beliefs about θ_{t-1} (given λ) and beliefs about λ, respectively, can be expressed in the form

$$p(\theta_{t-1} \mid \mathbf{D}_{t-1}, \lambda, M_{t-1}^{(i_{t-1})}, M_{t-2}^{(i_{t-2})}, \cdots, M_1^{(i_1)}) \sim N(\mathbf{m}_{t-1}^{(i_{t-1})}, \lambda^{-1}C_{t-1}^{(i_{t-1})}) \tag{3}$$

and

$$p(\lambda \mid \mathbf{D}_{t-1}, M_{t-1}^{(i)}, M_{t-2}^{(1)}, \cdots, M_1^{(1)}) \sim \text{gamma}(\frac{1}{2}n_{t-1}, \frac{1}{2}r_{t-1}^{(i_{t-1})}). \tag{4}$$

$\mathbf{m}_{t-1}^{(i)}, C_{t-1}^{(i)}, n_{t-1}, r_{t-1}^{(i)}$ will be derived recursively in the next subsection. Following the discussion in Section 3.2 of Gordon and Smith (1990), we have

Lemma. Let $\pi_0^{(j)}$ be the initial probability that system is at state j, and

$$\begin{aligned} e_t^{(i)} &= y_t - H_tG_t\mathbf{m}_{t-1}^{(i)}, \\ F_t^{(ij)} &= H_t(G_tC_{t-1}^{(i)}G_t^T + R_\theta^{(j)})H_t^T + R_y^{(j)}, \end{aligned}$$

$$
\begin{aligned}
z_t^{(ij)} &= [F_t^{(ij)}]^{-1/2}[r_{t-1}^{(i)}]^{n_{t-1}/2}\{r_{t-1}^{(i)} + [e_t^{(i)}]^2/F_t^{(ij)}\}^{-n_t/2}, \\
\pi_t^{(ij)} &= z_t^{(ij)}\pi_0^{(j)}\left(\sum_{k=1}^{4}\pi_{t-1}^{(ki)}\right) \Big/ \left[\sum_{i=1}^{4}\sum_{j=1}^{4} z_t^{(ij)}\pi_0^{(j)}\left(\sum_{k=1}^{4}\pi_{t-1}^{(ki)}\right)\right].
\end{aligned}
$$

Then

$$O_t^{(i)} = \sum_{j=1}^{4}\pi_t^{(ij)}.$$

So, this lemma can be applied to find the conditional one-step back posterior probability $O_t^{(i)}$ after the conditional densities in (3) and (4) are determined.

2.2 The Conditional Mean and Covariance

Here we show how to derive $\mathbf{m}_{t-1}^{(i)}, C_{t-1}^{(i)}, n_{t-1}, r_{t-1}^{(i)}$ in (3) and (4) recursively. The following well known fact is very useful [cf. Anderson (1958, Chapter 2)].

Fact 1. Suppose that

$$(\mathbf{y}_1, \mathbf{y}_2) \sim N\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12}^T & \Sigma_{22}\end{pmatrix}\right),$$

then the conditional density of $\mathbf{y}_1$ given $\mathbf{y}_2$ is

$$N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{y}_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T).$$

If the system is at state i_{t-1} at time $t-1$ given the events $M_k^{(i_k)}$ before time $t-1$, it follows from (1) and (2) that, given λ,

$$\begin{pmatrix}\theta_{t-1}\\ \mathbf{D}_{t-1}\end{pmatrix} = \begin{pmatrix}\bar{G}_{t-1}\\ B_{t-1}\end{pmatrix}\theta_0 + \begin{pmatrix}K_{t-1}\\ \Delta_{t-1}\end{pmatrix}\begin{pmatrix}\delta\theta_{t-1}^{(i_{t-1})}\\ \vdots\\ \delta\theta_1^{(i_1)}\end{pmatrix} + \begin{pmatrix}0\\ \delta y_{t-1}^{(i_{t-1})}\\ \vdots\\ \delta y_1^{(i_1)}\end{pmatrix}, \tag{5}$$

where

$$\bar{G}_{t-1} = G_{t-1}\cdots G_1,\ K_{t-1} = (I, \bar{G}_{t-1}\bar{G}_{t-2}^{-1}, \cdots, \bar{G}_{t-1}\bar{G}_1^{-1}),$$

$$B_{t-1} = \begin{pmatrix}H_{t-1}\bar{G}_{t-1}\\ \vdots\\ H_1\bar{G}_1\end{pmatrix}, \text{ and } \Delta_1 = H_1K_1, \cdots, \Delta_{t-1} = \begin{pmatrix}H_{t-1}K_{t-1}\\ 0,\ \ \Delta_{t-2}\end{pmatrix}.$$

Based on (5), we will derive the conditional densities in (3) and (4). Note that θ_{t-1} and $\mathbf{D}_{t-1}$ are jointly normally distributed with mean vector

$$\begin{pmatrix}\bar{G}_{t-1}\\ B_{t-1}\end{pmatrix}\mathbf{m}_0$$

and covariance matrix

$$\frac{1}{\lambda}\left[\begin{pmatrix}\bar{G}_{t-1}\\ B_{t-1}\end{pmatrix}C_0\,(\bar{G}_{t-1}^T, B_{t-1}^T) + \begin{pmatrix}K_{t-1}\\ \Delta_{t-1}\end{pmatrix}\mathbf{R}_\theta^{(t-1)}(K_{t-1}^T, \Delta_{t-1}^T) + \begin{pmatrix}0 & 0\\ 0 & \mathbf{R}_y^{(t-1)}\end{pmatrix}\right],$$

where

$$\mathbf{R}_\theta^{(t-1)} = \begin{pmatrix}R_\theta^{(i_{t-1})} & & \\ & \ddots & \\ & & R_\theta^{(i_1)}\end{pmatrix} \text{ and } \mathbf{R}_y^{(t-1)} = \begin{pmatrix}R_y^{(i_{t-1})} & & \\ & \ddots & \\ & & R_y^{(i_1)}\end{pmatrix}.$$

According to Fact 1,

$$\mathbf{m}_{t-1}^{(i_{t-1})} = \bar{G}_{t-1}\mathbf{m}_0 + (E_{t-1} + \bar{G}_{t-1}F_{t-1})(\mathbf{D}_{t-1} - B_{t-1}\mathbf{m}_0), \tag{6}$$

where

$$E_{t-1} = K_{t-1}\mathbf{R}_\theta^{(t-1)}\Delta_{t-1}^T[B_{t-1}C_0B_{t-1}^T + \mathbf{R}_y^{(t-1)} + \Delta_{t-1}\mathbf{R}_\theta^{(t-1)}\Delta_{t-1}^T]^{-1}, \tag{7}$$

and

$$F_{t-1} = C_0B_{t-1}^T[B_{t-1}C_0B_{t-1}^T + \mathbf{R}_y^{(t-1)} + \Delta_{t-1}\mathbf{R}_\theta^{(t-1)}\Delta_{t-1}^T]^{-1}. \tag{8}$$

Moreover, we have

$$
\begin{aligned}
& C_{t-1}^{(i_{t-1})} \\
= \ & K_{t-1}\mathbf{R}_\theta^{(t-1)}K_{t-1}^T + \bar{G}_{t-1}C_0\bar{G}_{t-1}^T - (K_{t-1}\mathbf{R}_\theta^{(t-1)}\Delta_{t-1}^T + \bar{G}_{t-1}^TC_0B_{t-1}^T)\cdot \\
& [B_{t-1}C_0B_{t-1}^T + \mathbf{R}_y^{(t-1)} + \Delta_{t-1}\mathbf{R}_\theta^{(t-1)}\Delta_{t-1}^T]^{-1}[\Delta_{t-1}\mathbf{R}_\theta^{(t-1)}K_{t-1}^T + B_{t-1}C_0\bar{G}_{t-1}^T] \\
= \ & K_{t-1}\mathbf{R}_\theta^{(t-1)}K_{t-1}^T + \bar{G}_{t-1}C_0\bar{G}_{t-1}^T \\
- \ & (E_{t-1} + \bar{G}_{t-1}F_{t-1})[\Delta_{t-1}\mathbf{R}_\theta^{(t-1)}K_{t-1}^T + B_{t-1}C_0\bar{G}_{t-1}^T].
\end{aligned} \tag{9}
$$

Apparently, (6) and (9) involve the inversion of a complicated matrix appearing in (7) and (8). We defer this issue to the appendix. To find n_{t-1} and $r_{t-1}^{(i)}$, note that

$$
\begin{aligned}
& p(\lambda \mid \mathbf{D}_{t-1}, M_{t-1}^{(i_{t-1})}, M_{t-2}^{(i_{t-2})}, \cdots, M_1^{(i_1)}) \\
\propto \ & p(\lambda, y_{t-1} \mid \mathbf{D}_{t-2}, M_{t-1}^{(i_{t-1})}, M_{t-2}^{(i_{t-2})}, \cdots, M_1^{(i_1)}) \\
= \ & p(y_{t-1} \mid \lambda, \mathbf{D}_{t-2}, M_{t-1}^{(i_{t-1})}, M_{t-2}^{(i_{t-2})}, \cdots, M_1^{(i_1)})p(\lambda \mid \mathbf{D}_{t-2}, M_{t-2}^{(i_{t-2})}, \cdots, M_1^{(i_1)}),
\end{aligned}
$$

assuming that λ is conditionally independent of $M_{t-1}^{(i_{t-1})}$ given $\mathbf{D}_{t-2}, M_{t-2}^{(i_{t-2})}, \cdots, M_1^{(i_1)}$. We know that

$$p(\lambda \mid \mathbf{D}_{t-2}, M_{t-2}^{(i_{t-2})}, \cdots, M_1^{(i_1)}) \sim \text{gamma}(\frac{1}{2}n_{t-2}, \frac{1}{2}r_{t-2}^{(i_{t-2})})$$

and

$$p(y_{t-1} \mid \lambda, \mathbf{D}_{t-2}, M_{t-1}^{(i_{t-1})}, M_{t-2}^{(i_{t-2})}, \cdots, M_1^{(i_1)}) \sim N(\mu_c, \lambda^{-1}\sigma_c^2),$$

where μ_c and σ_c^2 shall be derived shortly. Then,

$$n_{t-1} = n_{t-2} + 1$$

and

$$r_{t-1}^{(i)} = r_{t-2}^{(i_{t-2})} + (y_{t-1} - \mu_c)^2/\sigma_c^2.$$

It remains to evaluate μ_c and σ_c^2. By (5) and Fact 1, we have

$$\mu_c = H_{t-1}\bar{G}_{t-1}\mathbf{m}_0 + H_{t-1}(G_{t-1}E_{t-2} + \bar{G}_{t-1}F_{t-2})(\mathbf{D}_{t-2} - B_{t-2}\mathbf{m}_0), \tag{10}$$

and

$$\begin{aligned}\sigma_c^2 = {} & R_y^{(i_{t-1})} + H_{t-1}K_{t-1}\mathbf{R}_\theta^{(t-1)}K_{t-1}^T H_{t-1}^T + H_{t-1}\bar{G}_{t-1}C_0\bar{G}_{t-1}^T H_{t-1}^T \\ & - H_{t-1}(G_{t-1}E_{t-2} + \bar{G}_{t-1}F_{t-2})[B_{t-2}C_0\bar{G}_{t-1}^T + \Delta_{t-2}\mathbf{R}_\theta^{(t-2)}K_{t-2}^T G_{t-1}^T]H_{t-1}^T. \end{aligned} \tag{11}$$

We now see that the calculation for $\mathbf{m}_{t-1}^{(i)}, C_{t-1}^{(i)}, n_{t-1}$, and $r_{t-1}^{(i)}$ is straightforward if we are able to update E_{t-1} and F_{t-1} efficiently. An efficient algorithm will be given in the appendix.

2.3 On-line Detection of Change Points

Now, we describe how to apply the method presented above for on-line detection of change points. First, we select a threshold level. At any time t, we find the maximum of $O_t^{(j)}$ over three types of changes. If this maximal one-step probability exceeds the threshold level, a change at $t-1$ will be declared at the state, say i_{t-1}^*, corresponding to the maximum. At the subsequent time $t+1$, we will condition on $M_1^{i_1^*}, \cdots, M_{t-1}^{i_{t-1}^*}$. In fact, the conditional one-step back probability can be viewed as an alternative approximation to the unconditional counterpart evaluated by Gordon and Smith (1990). They used a single normal density to replace a mixture of normal densities by minimizing the Kullback-Leibler distance. If the mixture is replaced with the density corresponding to the states that have the highest "model" probability at each stage, then the conditional and unconditional one-step back probabilities become the same quantity. Therefore, the discussion in Gordon (1986) regarding the choice of threshold may be applied to our study.

3. LIKELIHOOD METHOD

In this section, we are concerned with only the slope (or amplitude) change ($j = 3$). Suppose that we independently observed y_i associated with a (time dependent and non-

random) covariate x_i ($i = 1, \cdots, t$), and that $x_1 < x_2 < \cdots < x_t$. In fact, x_i is often taken as i. It is of interest to know if and where a change, say τ, occurred before x_i given the data $\mathbf{D}_t$. This is a different, perhaps less ambitious issue than the one discussed in the previous section, and shall be elaborated later.

Assume that

$$y_i \sim N(\alpha_1 + \beta_1 x_t, \sigma^2) \text{ for } x_i \le \tau$$

and otherwise

$$y_i \sim N(\alpha_2 + \beta_2 x_t, \sigma^2).$$

Moreover, it is required that

$$\alpha_1 + \beta_1\tau = \alpha_2 + \beta_2\tau,$$

which corresponds to the slope change. This is the so-called two-phase regression [see Hinkley (1971)], change-point regression problem [Carlin, Gelfand, and Smith (1992)], or the broken line regression [see Siegmund and Zhang (1990)]. It is also a basis for Friedman's MARS (1991).

Finding the maximum likelihood estimate $\hat{\tau}$ for the change point τ has been illustrated by a number of authors; see, e.g., Hinkly (1971). Here, we construct the likelihood ratio test for whether the change point is real, i.e., the hypothesis $H_0 : \tau \in [x_2, x_{t-1})$.

Let $(x-\tau)^+$ stand for an t-vector whose ith component is zero if $x_i \le \tau$ and $x_i - \tau$ otherwise. Let r_x and r_y denote the residual vectors from regressions (with intercepts) $(x-\tau)^+$ vs x and y vs x, respectively. Let $\gamma(\tau) = r_x/||r_x||$ and $U = r_y/||r_y||$; here $||\cdot||$ denotes the norm of a vector.

According to Knowles and Siegmund (1990) [also see Naiman (1986) and (1987), Johnstone and Siegmund (1989), Johansen and Johnstone (1990)], the p-value under H_0 can be computed by the following formula,

$$p_t = t_{t-3}\left(\sqrt{\frac{t-3}{1-w^2}}w\right) + |\gamma|(1-w^2)^{\frac{(t-5)}{2}}/2\pi, \tag{12}$$

where t_k is the Student's t-distribution with degrees of freedom k, $|\gamma|$ is the arc length of the curve $\gamma(\tau)$ when τ traverses from x_2 to x_{t-1}, and

$$w = |\langle\gamma(\hat{\tau}), U\rangle|, \tag{13}$$

where $\langle\cdot\rangle$ denotes inner-product of vectors.

Note that the curvature of $\gamma(\tau)$ is zero when $\tau \neq x_i$ for some i. This implies that γ is along great circles of a unit sphere in R^t and changes its directions when $\tau = x_i$ for some i. The arc length $|\gamma|$ is simply the summation of the arc length of each segment between $\gamma(x_i)$ and $\gamma(x_{i+1})$ $(i = 2, \cdots, t-2)$, and the angle between $\gamma(x_i)$ and $\gamma(x_{i+1})$ determines the arc length between them because the segment is one piece of a great circle. More precisely, we have

$$|\gamma| = \sum_{2}^{t-2} \cos^{-1}\{\langle \gamma(x_i), \gamma(x_{i+1})\rangle\}.$$

Next, we use (12) to identify the first change point when data are collected sequentially, based on the belief that the system generating the data will usually be adjusted when a change occurs. We proceed as follows. Start from (x_4, y_4). At time $t(\geq 4)$, we use data from (x_1, y_1) to (x_t, y_t) and compute the p value p_t in (12) for the possible change point τ before x_t. If $p_t \leq \alpha$ (pre-specified by users, e.g., 0.05), then a change point will be declared at the position $\hat{\tau}$, and a new process will start after $\hat{\tau}$.

It is worthwhile to point out the implications of the p value p_t and the (conditional or unconditional) one-step-back posterior probability $O_t^{(j)}$. When a new data y_t is observed, $O_t^{(j)}$ is calculated with the current data available and will not be updated afterwards. Because of this, we may not be able to detect a change at a right time. Conceptually one can compute many-step back posterior probabilities; however, doing so increases the complexity of the calculation and may be redundant since it is not always needed to look back so many steps.

In contrast, p_t is the p value for the hypothesis that a change occurred before x_t, not necessarily at x_{t-1}. Of course, it would be ideal if we can detect a change right after it has occurred. This is sometimes possible but not always; see Section 4. As a by-product, the information about how many steps back are needed to identify a change is also gathered, and the steps vary along time. So, p_t takes the advantage of all the data available and is a implicitly multi-step-back procedure. As a reminder, the likelihood approach is designed for the slope change although it can also be applied to detect other states of changes by renewing the data process and repeating the technique. For example, multiple (two or more) change points, which are usually consecutive, may be associated with the transient change (see examples below), i.e., an outlier.

4. SIMULATIONS

Our simulations are designed to demonstrate the sensitivity and specificity of the methods introduced earlier.

The Bayesian framework in Section 2 can be applied to a variety of models; see, e.g., Gordon and Smith (1990). To compare it with the likelihood method, we concentrate on its applications to the linear growth model,

$$\begin{aligned} y_t &= \mu_t + \delta y_t^{(j)} \\ \mu_t &= \mu_{t-1} + \beta_t + \delta y_t \\ \beta_t &= \beta_{t-1} + \delta \beta_t^{(j)}. \end{aligned}$$

namely,

$$H_t = (1, 0), \quad \text{and} \quad G_t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Remark. It is important in practical sense to point out a minor variation between the slope change-point T in the linear growth model and the change-point τ in the regression model introduced in Section 3. Because the measurements are taken at discrete times, $\tau \leq T$. This relationship may give the likelihood approach some advantage to detect change-points more promptly than the Bayesian one as illustrated in Section 5.

To use the Bayesian method, priors should be specified. Here, we bellow the specifications of Smith and West (1983). At different states assume that the variance components to be:

	$R_y^{(j)}$	$R_\mu^{(j)}$	$R_\beta^{(j)}$
$j=1$	1	0	0
$j=2$	1	90	0
$j=3$	1	0	60
$j=4$	100	0	0.

Let (.85, .06, .07, .02) be the initial probabilities π_0 that the process $\{y_t\}$ starts at states $(1, 2, 3, 4)$, respectively. We will specify priors for λ and θ_0 in accordance with particular data.

The first experiment proceeds as follows. Ten iid N(0,1) noisy data are generated. We take $x_i = i$ and y_i is the corresponding noise. The purpose is to see how often the methods set false alarms when there is no change in the data sequence. One set of sample data is given in Table 1. For these data, assume that

$$\lambda \sim \Gamma(5/2, 3/2),$$

$$\theta_0 \,\Big|\, \lambda \sim N(0, \lambda^{-1} I_2).$$

The results can be sensitive to the choices of priors and general principles of choosing priors are suggested in Smith and West (1983) and Gordon and Smith (1990). For example, n_0

TABLE I
Results for simulated N(0,1) data

t	uncond. $O_t^{(2)}$	cond. $O_t^{(2)}$	uncond. $O_t^{(3)}$	cond. $O_t^{(3)}$	uncond. $O_t^{(4)}$	cond. $O_t^{(4)}$	p_t	τ	y_t
1	0.060	0.060	0.070	0.070	0.020	0.020			0.415
2	0.013	0.013	0.013	0.013	0.003	0.003			0.114
3	0.013	0.013	0.013	0.013	0.003	0.003			1.002
4	0.013	0.013	0.013	0.013	0.007	0.007	0.342	3	-0.389
5	0.031	0.033	0.064	0.066	0.002	0.003	0.172	3	-1.374
6	0.041	0.048	0.047	0.055	0.003	0.003	0.052	3	-2.196
7	0.012	0.011	0.016	0.012	0.017	0.017	0.114	6	0.098
8	0.056	0.029	0.016	0.012	0.012	0.011	0.400	6	-1.339
9	0.015	0.012	0.023	0.030	0.009	0.004	0.075	6	0.901
10	0.022	0.020	0.011	0.010	0.043	0.042	0.349	6	-1.689

and r_0 are usually so chosen that $E(\lambda^{-1}) = r_0/(n_0-2)$ corresponds to the noise level. Thus, our choice of the prior for λ reflects the noise level $\sigma^2 = 3/(5-2) = 1$, though this has to be estimated for real data. In Table 1, both conditional and unconditional one-step back probabilities are listed and they are numerically indistinguishable despite of their distinct implications.

Following the discussion in Section 2.3, we use the same threshold, 0.2, suggested by Gordon (1986) for the posterior probability. When 0.2 and 0.05 are selected as thresholds for $O_t^{(j)}$ and p_t, respectively, above which signals a change, then neither Bayesian nor likelihood approach sets off false alarm. Many such experiments have been repeated and have led to similar conclusions.

To get a little more insight into the likelihood method, the average number of observations before a false alarm went off is computed as follows. At any time t a normal $N(0,\sigma^2)$ noise is generated. Since σ is a nuisance parameter and does not affect our results at all, it is taken as 1. Let α is the significance level. Let T denote the number of observations before a false alarm. The average of T is obtained from 10,000 repetitions, and, as shown by Figure 1, is strikingly linearly correlated with the level α^{-1}. Their relation is that $\bar{T} = 5 + 0.64/\alpha$. For example, when α is chosen as 0.01, in the average case the likelihood approach would set off a false alarm during 69 units of time.

Our next simulation is designed to see the sensitivity of these methods. A series of sample data are generated from the linear growth model to test the procedures. Here, we

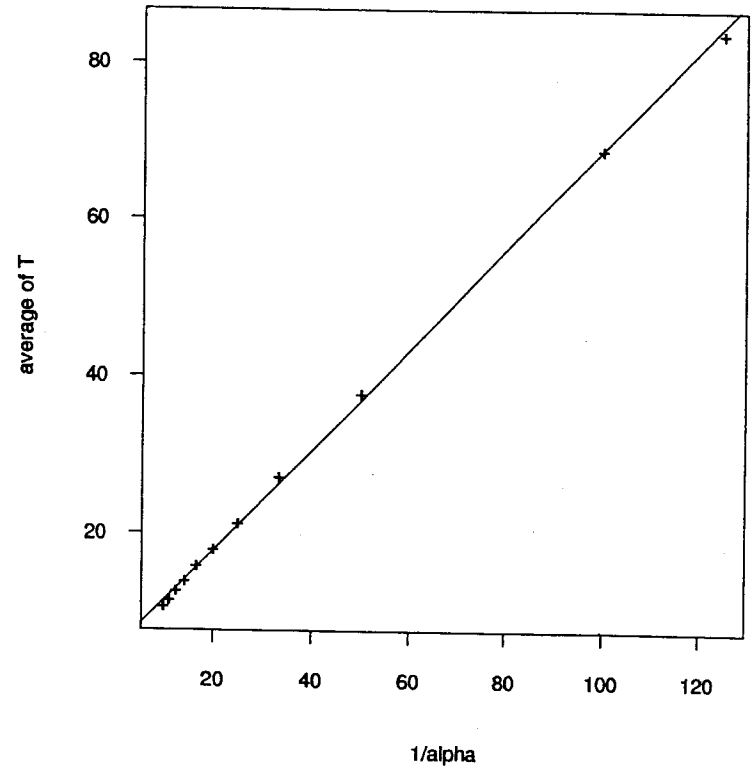

FIG I. Average Number of Observations before a False Alarm.

merely report the results for the simulated data of Gordon (1986). Figure 2 presents the actual data, which consists of 100 samples generated from the linear growth model with the error $\delta y_t^{(1)} \sim N(0, 15)$; also see Figure 3.31 of Gordon (1986). Four change points are put in the data. They are at 25 (the slope was reversed in sign), 35 (a transient change or outlier), 50 (a level change or a shift of means), and 80 (a second outlier). The prior distributions for λ and θ_0 are assumed to be

$$\lambda \sim \Gamma(5/2, 45/2) \text{ and } \theta_0 \mid \lambda \sim N(\mathbf{m}_0, \lambda^{-1}C_0),$$

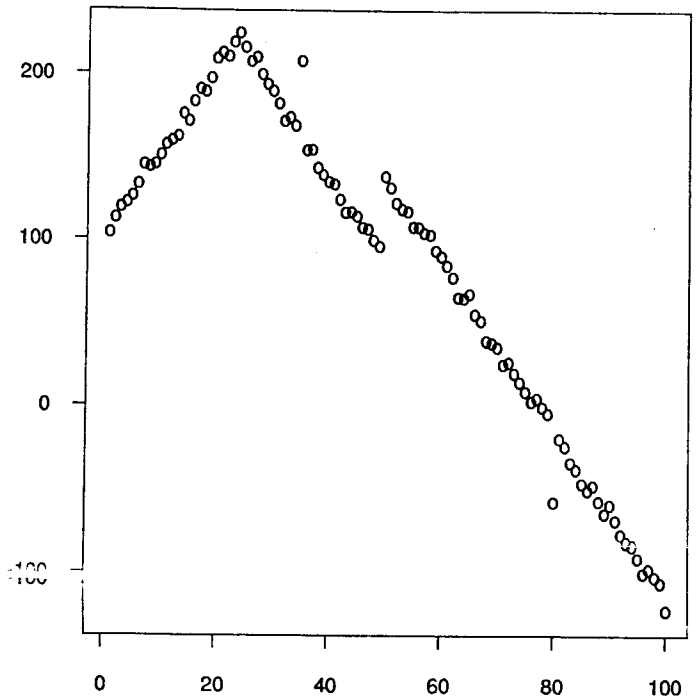

FIG II. Simulated Data from Linear Growth Model.

where

$$\mathbf{m}_0 = \begin{pmatrix} 100 \\ 5 \end{pmatrix} \text{ and } C_0 = \begin{pmatrix} 10 & 0 \\ 0 & 0.5 \end{pmatrix}.$$

At five incidences, the unconditional one-step posterior probability exceeds 0.2: $O_{26}^{(2)} = 0.79, O_{36}^{(4)} = 0.97, O_{51}^{(2)} = 0.99, O_{81}^{(4)} = 0.95$, and $O_{88}^{(2)} = 0.25$. In contrast, the conditional one-step posterior probability picked up the four change points precisely, with the probabilities $O_{26}^{(2)} = 0.93, O_{36}^{(4)} = 0.98, O_{51}^{(2)} = 0.99, O_{81}^{(4)} = 0.97$. It is interesting to note here that the conditional probability is larger than the corresponding unconditional one at any of the four change points, while it is smaller at 88 where the unconditional Bayesian procedure sets off a false alarm. The generality of this conclusion needs further investigation.

Now, we present the results from the likelihood procedure. At the significance level 0.01, all four change points are detected without any delay. This also explains why the one-step

probability can locate the change points well in this data set. For the two transient changes, their neighbors are also declared as change points because the likelihood procedure considers the slope change only and one transient change may be converted into three consecutive slope changes.

These simulations represent a variety of similar experiments done by Gordon (1986) and suggest that the three procedures discussed in this paper may be reliable and consistent to some extent. Further remarks will be made later.

5. THE CREATININE SERIES

In this section, we re-analyze the creatinine series from Gordon and Smith (1990). This application is for monitoring the functioning of an individual renal transplant for signs of rejection. Serum creatinine concentration is measured at time intervals; see e.g., Knapp, Blamey, Cove-Smith, and Heath (1977) and Smith and West (1983) for more information. In Figure 3(a), weight-corrected reciprocal serum creatinine measurements are plotted daily. This is a duplication of Figure 1 in Gordon and Smith (1990). As pointed out by Gordon and Smith (1990), the primary interest of this problem is to detect the slope change.

Since the measurements are taken at unequal time intervals, one-step back refers to the previous measuring time not the previous day. Some modification is necessary for the Bayesian analysis. Gordon and Smith (1990) discussed the issue in detail and their derivation is also applicable to our conditional analysis. Eight change points, on days 7, 11, 17, 42, 82, 105, 112, and 119, are detected by the likelihood analysis. In Table 2, accompanying the day of change the extra number of measurements needed to take after that day and the corresponding p value are given. On the average, three additional measurements are needed to make the decision, so we would expect that the current application is challenging for the one-step Bayesian schemes.

To apply the Bayesian theory, our initial attempt was to choose the same prior specifications used by Gordon and Smith (1990). To be certain, we first tried to reproduce their numerical results following the procedure of the authors. We did not succeed because their report contains numerical errors. Our numerical results were different from theirs for the creatinine series even though we were able to get the same results as Gordon (1986) for the simulation data described in Section 4. Then, we used the information from the likelihood analysis and took the following priors:

$$\lambda \sim \Gamma(5/2, 10^{-6}/2) \text{ and } \theta_0 \,\Big|\, \lambda \sim N(\mathbf{m}_0, \lambda^{-1} I_2),$$

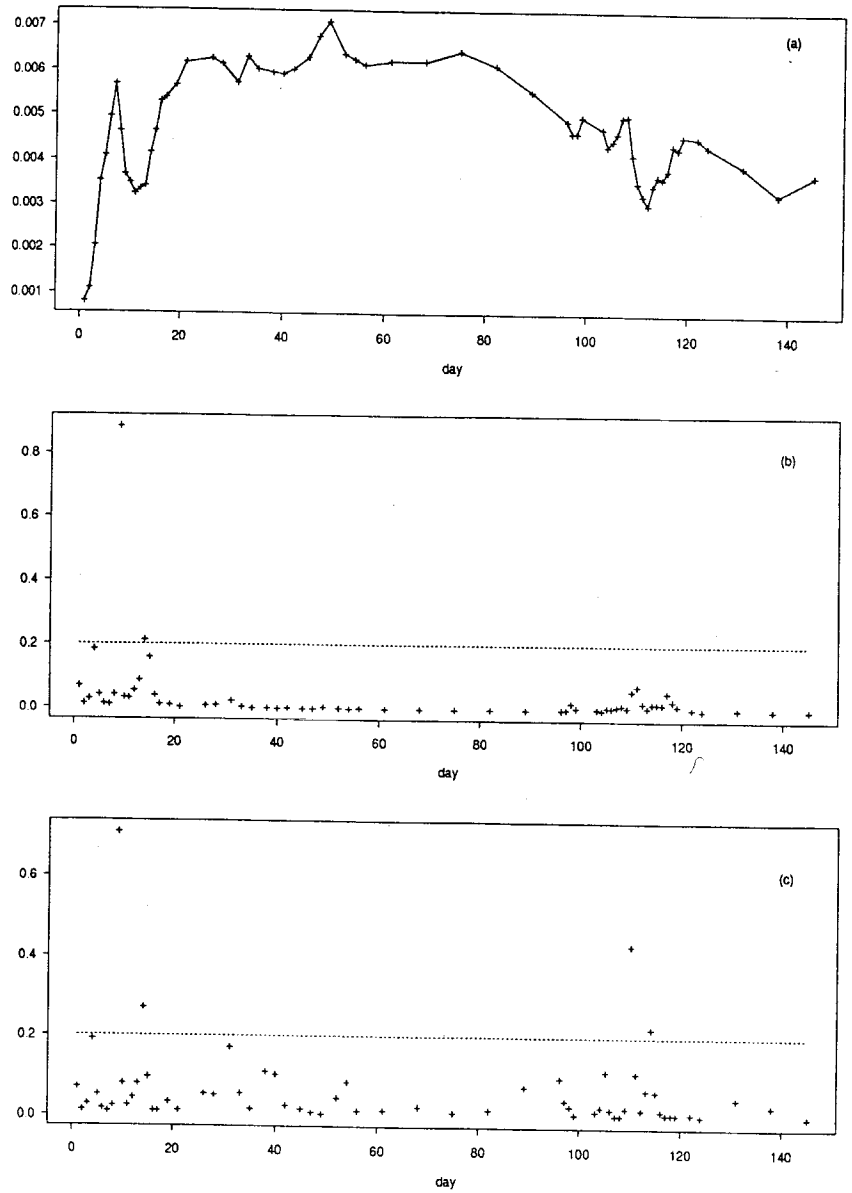

FIG III. The Comparison of the Bayesian Analyses. (a) Creatinine Series; (b) Unconditional Probability of Slope Change; (c) Conditional Probability of Slope Change.

TABLE II
Likelihood Analysis of the Creatinine Series.

t	τ	p_t	steps delayed
8	7	0.006	1
15	11	0.006	4
26	17	0.002	3
49	42	0.005	3
97	82	0.008	3
108	105	0.005	3
116	112	0.006	3
124	119	0.010	1

where $\mathbf{m}_0 = (0.001, 0)^T$.

As can be seen from Figure 3(b), two changes are reported on days 8 and 13 by the unconditional Bayesian procedure. In fact, $O_9^{(2)} = 0.89$ and $O_{14}^{(2)} = 0.21$. The big value of $O_9^{(2)}$ and the marginal level of $O_{14}^{(2)}$ confirm our anticipation based on Table 2. The conditional Bayesian analysis shows four changes on days 8, 13, 109, and 113; see Figure 3(c). Note that change-points are detected earlier by the likelihood method in four occasions where the conditional Bayesian approach also signals the changes. Rejection therapy was initiated on days 9 and 112, so changes before these days were presumably diagnosed by clinicians. It is also expected that changes may occur after the therapy. Therefore, for the current application, the conditional Bayesian would have been more appropriate than the unconditional Bayesian one suggested in the literature. Since information is not available, we cannot determine the status of the other changes declared by the likelihood method.

6. DISCUSSIONS

In this paper, we discussed Bayesian and likelihood approaches to the on-line detection of change points for time series. As demonstrated by our examples, both approaches are competitive and led to consistent results. The likelihood method is much easier to use but limited to the linear growth model and the slope change. The Bayesian one has greater capacity to deal with general problems, but may require that one has substantial knowledge of the data. These methods can be more useful when used as complementary tools in applications.

APPENDIX

Following the discussion in Section 2.2, we now show how to obtain E_{t-1} and F_{t-1} from E_{t-2} and F_{t-2} recursively, which avoids the direct inversion of the matrix in (7) and (8). We will use another well known fact.

Fact 2.

$$\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} & -A^{-1}\Sigma_{12}\Sigma_{22}^{-1} \\ -\Sigma_{22}^{-1}\Sigma_{12}^T A^{-1} & \Sigma_{22}^{-1} + \Sigma_{22}^{-1}\Sigma_{12}^T A^{-1}\Sigma_{12}\Sigma_{22}^{-1} \end{pmatrix},$$

where

$$A = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T.$$

The proof of this result is straightforward.

Let

$$\begin{aligned} \Sigma_{11} &= R_y^{(i_{t-1})} + H_{t-1}(\tilde{G}_{t-1}C_0\tilde{G}_{t-1}^T + K_{t-1}\mathbf{R}_\theta^{(t-1)}K_{t-1}^T)H_{t-1}^T, \\ \Sigma_{12} &= H_{t-1}(\tilde{G}_{t-1}C_0B_{t-2}^T + G_{t-1}K_{t-2}\mathbf{R}_\theta^{(t-2)}\Delta_{t-2}^T], \\ \Sigma_{22} &= B_{t-2}C_0B_{t-2}^T + \mathbf{R}_y^{(t-2)} + \Delta_{t-2}\mathbf{R}_\theta^{(t-2)}\Delta_{t-2}^T. \end{aligned}$$

Then,

$$B_{t-1}C_0B_{t-1}^T + \mathbf{R}_y^{(t-1)} + \Delta_{t-1}\mathbf{R}_\theta^{(t-1)}\Delta_{t-1}^T = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{pmatrix}.$$

It follows from Fact 1 that

$$\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{pmatrix}^{-1} = \sigma_c^{-2}\begin{pmatrix} 1 \\ -\Sigma_{22}^{-1}\Sigma_{12}^T \end{pmatrix}(1, -\Sigma_{12}\Sigma_{22}^{-1}) + \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_{22}^{-1} \end{pmatrix},$$

where σ_c is defined in (11). By the definitions, we know

$$\begin{aligned} E_{t-1} &= K_{t-1}\mathbf{R}_\theta^{(t-1)}\Delta_{t-1}^T \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{pmatrix}^{-1} \\ &= \sigma_c^{-2}K_{t-1}\mathbf{R}_\theta^{(t-1)}\Delta_{t-1}^T \begin{pmatrix} 1 \\ -\Sigma_{22}^{-1}\Sigma_{12}^T \end{pmatrix}(1, -\Sigma_{12}\Sigma_{22}^{-1}) + (0, G_{t-1}E_{t-2}), \quad (14) \end{aligned}$$

and

$$\begin{aligned} F_{t-1} &= C_0B_{t-1}^T \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{pmatrix}^{-1} \\ &= \sigma_c^{-2}C_0B_{t-1}^T \begin{pmatrix} 1 \\ -\Sigma_{22}^{-1}\Sigma_{12}^T \end{pmatrix}(1, -\Sigma_{12}\Sigma_{22}^{-1}) + (0, F_{t-2}). \quad (15) \end{aligned}$$

Note that $\Sigma_{12}\Sigma_{22}^{-1} = H_{t-1}(G_{t-1}E_{t-2} + \tilde{G}_{t-1}F_{t-2})$. So, E_{t-1} and F_{t-1} can be derived from E_{t-2} and F_{t-2} without matrix inversions. Furthermore, the following two recurrences are also useful in calculations of E_{t-1} and $C_{t-1}^{(i_{t-1})}$. Let $Y_{t-1} = K_{t-1}\mathbf{R}_\theta^{(t-1)}K_{t-1}^T$ and $Z_{t-1} = K_{t-1}\mathbf{R}_\theta^{(t-1)}\Delta_{t-1}^T$. Then

$$\begin{aligned} Y_{t-1} &= R_\theta^{(i_{t-1})} + G_{t-1}Y_{t-2}G_{t-1}^T, \\ Z_{t-1} &= (Y_{t-1}H_{t-1}^T, G_{t-1}Z_{t-2}). \end{aligned}$$

To summarize, we start with $E_1 = R_\theta^{(i_1)}H_1^T/(R_y^{(1)} + B_1C_0B_1^T + H_1R_\theta^{(i_1)}H_1^T)$ and $F_1 = C_0B_1^T/(R_y^{(i_1)} + B_1C_0B_1^T + H_1R_\theta^{(i_1)}H_1^T)$. Then, we can update E_t and F_t via (14) and (15), respectively. Next, substituting E_t and F_t into (6) and (9), we can derive $\mathbf{m}_t^{(i)}$ and $C_t^{(i)}$; substituting E_t and F_t into (10) and (11), we obtain μ_c and σ_c, and hence $r_t^{(i)}$.

ACKNOWLEDGMENT

I thank Professor Robert Makuch for bringing Gordon and Smith's work to my attention, and for his helpful comments. I am also grateful to Professors John Hartigan and David O. Siegmund for their invaluable suggestions. This research is supported in part by an NIH grant HD30712.

BIBLIOGRAPHY

Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*, New York: Wiley.

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

Friedman, J. H. (1991). "Multivariate adaptive regression splines" (with discussions) *Ann. Statist.*, 19, 1–141.

Gordon, K. (1986). Modeling and monitoring of medical time series. PhD Thesis, University of Nottingham.

Gordon, K. and Smith, A. F. M. (1990). "Modeling and monitoring biomedical time series." *J. Amer. Statist. Assoc.*, 85, 328–337.

Hinkley, D. (1971). "Inference in two-phase regression." *J. Amer. Statist. Assoc.*, 66, 736–743.

Hotelling, H. (1939). "Tubes and spheres in n–space and a class of statistical problems." *Amer. J. Math.*, 61, 440–460.

Johansen, S. and Johnstone, I. (1990). "Hotelling's theorem on the volume of tubes: Some illustrations in simultaneous inference and data analysis." *Ann. Statist.*, 18, 652–684.

Johnstone, I. and Siegmund, D. (1989). "On Hotelling's formula for the volume of tubes and Naiman's inequality." *Ann. Statist.*, 17, 184–194.

Knapp, M. S., Blamey, R. W., Cove-Smith, J. R., and Heath, M. (1977). "Monitoring the function of renal transplants." *Lancet*, 2, 1183.

Knowles, M. and Siegmund, D. O. (1989). "On Hotelling's approach to testing for nonlinear parameter in regression." *Internat. Statist. Review*, 57, 205–220.

Naiman, D. Q. (1986). "Conservative confidence bands in curvilinear regression." *Ann. Statist.*, 14, 896–1906.

Naiman, D. Q. (1986). "Simultaneous confidence bands in multiple regression using predictor variable constraints." *J. Amer. Statist. Assoc.*, 82, 214–219.

Siegmund, D. (1986). "Boundary crossing probabilities and statistical applications." *Ann. Statist.*, 14, 361–404.

Siegmund, D. and Zhang, H. (1990). Confidence regions in broken line regression. Technical Report #12, Department of Statistics, Stanford University.

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

Smith, A. F. M. and Cook, D. G. (1980). "Straight lines with a change point: A Bayesian analysis of some renal transplant data." *Applied Statistics*, 29, 180–189.

Smith, A. F. M. and West, M. (1983). "Monitoring renal transplants: An application of the multiprocess Kalman filter." *Biometrics*, 39, 867–878.

Zhang, H. (1991). Confidence Regions in Nonlinear Regression and Geometry. PhD Thesis, Stanford University.

Received July, 1994; Revised December, 1994.

COMMUN. STATIST.—THEORY METH., 24(5), 1325-1341 (1995)

CHOICE OF NONINFORMATIVE PRIORS FOR THE VARIANCE COMPONENTS OF AN UNBALANCED ONE-WAY RANDOM MODEL

Éric Belzile and Jean-François Angers

Dép. de mathématiques et de statistique and CRM, Université de Montréal
C.P. 6128 Succ. Centre-ville, Montréal, Qc H3C 3J7, Canada

Key Words and Phrases: noninformative prior; unbalanced random model; variance components.

ABSTRACT

Several noninformative priors for the variance components of an unbalanced one-way model are proposed. A simulation study is performed to compare the posterior means of the variance components based on the different priors. These estimators are also compared to the restricted maximum likelihood estimator (REML). Under the squared error loss, this simulation study indicates that the invariant priors lead to the optimal estimators of the variance components.

Copyright © 1995 by Marcel Dekker, Inc.