Aust. N. Z. J. Stat. **46**(4), 2004, 555–567

BAYESIAN DETECTION AND ANALYSIS OF CHANGING TRANSITION MATRICES OF STATIONARY MARKOV CHAINS

P.C.N. Groenewald[1]* and A.C. Schoeman[1]

University of the Free State, South Africa

Summary

This paper considers the detection of abrupt changes in the transition matrix of a Markov chain from a Bayesian viewpoint. It derives Bayes factors and posterior probabilities for unknown numbers of change-points, as well as the positions of the change-points, assuming non-informative but proper priors on the parameters and fixed upper bound. The Markov chain Monte Carlo approach proposed by Chib in 1998 for estimating multiple change-points models is adapted for the Markov chain model. It is especially useful when there are many possible change-points. The method can be applied in a wide variety of disciplines and is particularly relevant in the social and behavioural sciences, for analysing the effects of events on the attitudes of people.

Key words: Bayes factors; Gibbs sampling; multinomial; multiple change-points; posterior probability.

1. Introduction

In this paper we consider the detection of change-points in p-state Markov chains of order 1. A change in the transition pattern in a sequence of state variables may have been the result of an abrupt change in the transition probability matrix of the Markov chain. This is a problem that can arise in many disciplines. For example, in geology, the spatial transitions in a sequence of rock types may change (see Gingerich, 1969; Hiscott, 1981); in biochemistry, changes occur in spatial variations in base frequencies of deoxyribonucleic acid (Curnow & Kirkwood, 1989 Section 7). Important applications can also be found in the social sciences. A company may want to determine whether (and when) an advertising campaign or change to a product has any effect on consumer behaviour (Whitaker, 1978). The degree to which certain events affect public attitudes and opinions (Anderson, 1954) can be analysed using the Bayesian change-point model.

There is a very large body of literature on Bayesian change-point analysis dating back to Chernoff & Zacks (1964). The papers deal with a variety of models and we note just a few. Changes in the parameters of some standard parametric models are dealt with by Broemeling (1972, 1974, 1977), Broemeling & Tsurumi (1987), Broemeling & Gregurich (1996) and a large number of earlier references. Other important references are Smith (1975), Smith & Cook (1980), Zacks (1991), Stephens (1994) and Dey & Purkayastha (1997).

Received March 2003; revised November 2003; accepted January 2004.

*Author to whom correspondence should be addressed.

[1] Dept of Mathematical Statistics, University of the Free State, Bloemfontein 9300, South Africa.
e-mail: GroenePC.SCI@mail.uovs.ac.za

© 2004 Australian Statistical Publishing Association Inc. Published by Blackwell Publishing Asia Pty Ltd.

In general, the parametric change-point model assumes a sequence of random variables $x_1, \ldots, x_n$ from the distribution

$$f(x_i) = \begin{cases} f(x_i \mid \boldsymbol{\theta}_1) & i = 1, \ldots, k_1, \\ f(x_i \mid \boldsymbol{\theta}_2) & i = k_1 + 1, \ldots, k_2, \\ \quad \vdots & \\ f(x_i \mid \boldsymbol{\theta}_{r+1}) & i = k_r + 1, \ldots, n. \end{cases}$$

The parameters $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{r+1}$ are usually unknown, but the main interest is in estimating the unknown change-points $\boldsymbol{k} = (k_1, \ldots, k_r)$. Most of the references cited above assume a known number of change-points, usually one. However, if the number of change-points is unknown the problem becomes one of model selection rather than estimation. Authors who deal with an unknown number of change-points include Barry & Hartigan (1992, 1993) who use product partition models, Green (1995), Phillips & Smith (1996) and Chib (1998) who use Markov chain Monte Carlo methods, and Garisch & Groenewald (1999).

The only previous published Bayesian work on the Markov chain change-point problem is by Carlin, Gelfand & Smith (1992). They illustrate the use of the Gibbs sampler in change-point problems in general, with an application to a Markov chain problem. They assume one change-point, thus avoiding the dimensionality problem. In Section 2 we show that the analytical solution is easily obtainable with the no-change model included, as long as the priors are proper. Extension to multiple change-points is given in Section 3. For a moderate number of possible change-points the numerical calculations are not overly extensive. However, for long chains and $r > 3$, the number of numerical calculations becomes prohibitively large. In that case the Gibbs sampling approach of Chib (1998) is very attractive and its application to the changing Markov chain model is described in Section 4. Section 5 contains some simulation results and a few applications.

2. Change-points in Markov chains

Consider a sequence of n observations $\boldsymbol{Y} = (y_1, \ldots, y_n)$ from a process which forms a p-state stationary first-order Markov chain having transition matrix $\boldsymbol{A}$. The elements of $\boldsymbol{A}$ are $a_{ij} = \Pr(Y_{t+1} = j \mid Y_t = i)$ where $a_{ij} \geq 0$, $\sum_j a_{ij} = 1$, $i, j = 1, \ldots, p$. The questions of interest are whether any changes have occurred in the matrix $\boldsymbol{A}$ during the observation period, the possible positions of these change-points and the magnitude of the changes.

Assuming at most one change-point, we consider the Bayes factor and posterior probability in favour of no change. The Bayes factor in favour of the no-change model M_0, when compared with a change just after time k (model M_k) is defined as

$$B_{0k} = \frac{m_0(\boldsymbol{y})}{m_k(\boldsymbol{y})} \,,$$

where $m_0(\boldsymbol{y})$ and $m_k(\boldsymbol{y})$ are the marginal likelihoods under the respective models. Under model M_k we assume a change from $\boldsymbol{A}$ to transition matrix $\boldsymbol{B}$ at time k and take independent Dirichlet priors on the rows of $\boldsymbol{A}$, i.e. $\boldsymbol{a}_i \overset{\mathrm{d}}{=} \mathrm{D}(\boldsymbol{\alpha}_i)$, and similarly for $\boldsymbol{B}$. Then, under M_0,

$$\pi(\boldsymbol{A}) = \prod_{i=1}^{p} f_{\mathrm{D}}(\boldsymbol{a}_i \mid \boldsymbol{\alpha}_i) \,,$$

© 2004 Australian Statistical Publishing Association Inc.

where $\boldsymbol{a}_i = (a_{i1}, \ldots, a_{ip})$, $\boldsymbol{\alpha}_i = (\alpha_{i1}, \ldots, \alpha_{ip})$ and

$$f_{\mathrm{D}}(\boldsymbol{a}_i \mid \boldsymbol{\alpha}_i) = \frac{\Gamma(\alpha_{i\boldsymbol{.}})}{\prod_{j=1}^{p} \Gamma(\alpha_{ij})} \prod_{j=1}^{p} a_{ij}^{\alpha_{ij}-1},$$

with $\alpha_{i\boldsymbol{.}} = \sum_{j=1}^{p} \alpha_{ij}$. Similarly, $\pi(\boldsymbol{A}) = \prod_{j=1}^{p} f_{\mathrm{D}}(\boldsymbol{a}_i \mid \boldsymbol{\lambda}_i)$ and $\pi(\boldsymbol{B}) = \prod_{j=1}^{p} f_{\mathrm{D}}(\boldsymbol{b}_i \mid \boldsymbol{\gamma}_i)$ under M_k.

The likelihood function under M_0 is

$$f(\boldsymbol{y} \mid \boldsymbol{A}) = p(y_1) \prod_{t=1}^{n-1} a_{y_t, y_{t+1}},$$

where $p(y_1)$ denotes the initial state probability, and

$$m_0(\boldsymbol{y}) = \int f(\boldsymbol{y} \mid \boldsymbol{A}) \pi(\boldsymbol{A})\, d\boldsymbol{A} = K_\alpha \int \cdots \int \prod_{t=1}^{n-1} a_{y_t, y_{t+1}} \prod_{i=1}^{p} \prod_{j=1}^{p} a_{ij}^{\alpha_{ij}-1}\, da_{ij}, \qquad (1)$$

where
$$K_\alpha = p(y_1) \prod_{i=1}^{p} \frac{\Gamma(\alpha_{i\boldsymbol{.}})}{\prod_{j=1}^{p} \Gamma(\alpha_{ij})}.$$

Let z_{ij} denote the number of transitions from state i to state j in the $n-1$ steps; then (1) simplifies to

$$m_0(\boldsymbol{y}) = K_\alpha \prod_{i=1}^{p} \int \cdots \int \prod_{j=1}^{p} a_{ij}^{\alpha_{ij}+z_{ij}-1}\, da_{ij} = p(y_1) \prod_{i=1}^{p} \frac{\prod_{j=1}^{p} \Gamma(\alpha_{ij} + z_{ij}) \Gamma(\alpha_{i\boldsymbol{.}})}{\prod_{j=1}^{p} \Gamma(\alpha_{ij}) \Gamma(\alpha_{i\boldsymbol{.}} + z_{i\boldsymbol{.}})}. \qquad (2)$$

Under model M_k $(2 \leq k \leq n-1)$ we have

$$f(\boldsymbol{y} \mid \boldsymbol{A}, \boldsymbol{B}, k) = p(y_1) \prod_{t=1}^{k-1} a_{y_t, y_{t+1}} \prod_{t=k}^{n-1} b_{y_t, y_{t+1}} \quad \text{where } \boldsymbol{B} = [b_{ij}].$$

Let z'_{ij} and z''_{ij} denote the number of transitions from i to j in $Y_1, \ldots, Y_k$ and $Y_k, \ldots, Y_n$ respectively and let $\lambda^*_{ij} = \lambda_{ij} + z'_{ij}$, $\gamma^*_{ij} = \gamma_{ij} + z''_{ij}$, $\alpha^*_{ij} = \alpha_{ij} + z_{ij}$. Then the likelihood is given by

$$m_k(\boldsymbol{y}) = p(y_1) \prod_{i=1}^{p} \Big(\frac{\Gamma(\lambda_{i\boldsymbol{.}}) \Gamma(\gamma_{i\boldsymbol{.}})}{\Gamma(\lambda^*_{i\boldsymbol{.}}) \Gamma(\gamma^*_{i\boldsymbol{.}})} \prod_{j=1}^{p} \frac{\Gamma(\lambda^*_{ij}) \Gamma(\gamma^*_{ij})}{\Gamma(\lambda_{ij}) \Gamma(\gamma_{ij})} \Big), \qquad (3)$$

so that the Bayes factor in favour of M_0 is given by

$$B_{0k} = \prod_{i=1}^{p} \Big(\frac{\Gamma(\alpha_{i\boldsymbol{.}}) \Gamma(\lambda^*_{i\boldsymbol{.}}) \Gamma(\gamma^*_{i\boldsymbol{.}})}{\Gamma(\alpha^*_{i\boldsymbol{.}}) \Gamma(\lambda_{i\boldsymbol{.}}) \Gamma(\gamma_{i\boldsymbol{.}})} \prod_{j=1}^{p} \frac{\Gamma(\alpha^*_{ij}) \Gamma(\lambda_{ij}) \Gamma(\gamma_{ij})}{\Gamma(\alpha_{ij}) \Gamma(\lambda^*_{ij}) \Gamma(\gamma^*_{ij})} \Big). \qquad (4)$$

Since we are conditioning on the data, the initial state probability $p(y_1)$ plays no role and cancels out in the Bayes factor (4).

© 2004 Australian Statistical Publishing Association Inc.

These Bayes factors can be calculated for all k $(2 \le k \le n-1)$, and the posterior probabilities follow as

$$\Pr(M_k \mid \boldsymbol{y}) \propto \frac{\pi_k}{B_{0k}} \qquad (k = 0, 2, 3, \ldots, n-1)\,, \tag{5}$$

where π_k, $k = 0, 2, 3, \ldots, n-1$ are the prior probabilities for the respective models. Also $B_{k0} = B_{0k}^{-1}$ and $B_{ij} = B_{i0}/B_{j0}$.

For the Bayes factor in (4) to be valid, we must use proper priors because the normalizing constant plays an integral part. In fact, B_{0k} in (4) would be indeterminate if we were to use the non-informative improper prior with all hyperparameters equal to 0. In that case the fractional Bayes factor (FBF) of O'Hagan (1995, 1997) could be used, but only if $z'_{ij}, z''_{ij} > 0$ for all i, j. The FBF uses a fraction of the likelihood to convert an improper prior to a proper one. However, a proper distribution that can be considered non-informative occurs when all probabilities are uniformly distributed over the simplex, i.e. $\alpha_{ij} = \lambda_{ij} = \gamma_{ij} = 1$ for all i, j. Equation (4) then reduces to

$$B_{0k} = \frac{1}{\Gamma^p(p)} \prod_{i=1}^{p} \left(\frac{\Gamma(z'_{i\boldsymbol{\cdot}} + p)\Gamma(z''_{i\boldsymbol{\cdot}} + p)}{\Gamma(z_{i\boldsymbol{\cdot}} + p)} \prod_{j=1}^{p} \frac{\Gamma(z_{ij} + 1)}{\Gamma(z'_{ij} + 1)\Gamma(z''_{ij} + 1)} \right). \tag{6}$$

The Jeffreys prior with $\alpha_{ij} = \lambda_{ij} = \gamma_{ij} = 0.5$ is proper and can also be used. With equal prior weights on the no-change model and the change-point model, and with equal weights for the position of the change-point, given there is one, we have $\pi_0 = \frac{1}{2}$ and $\pi_k = 1/(2(n-2))$.

The posterior distributions of $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$, $i = 1, \ldots, p$, assuming a change-point at k, follow as

$$(\boldsymbol{a}_i \mid \boldsymbol{y}, k) \overset{\mathrm{d}}{=} \mathrm{D}(\boldsymbol{\lambda}_i^*) \quad \text{and} \quad (\boldsymbol{b}_i \mid \boldsymbol{y}, k) \overset{\mathrm{d}}{=} \mathrm{D}(\boldsymbol{\gamma}_i^*)\,,$$

while the marginal posteriors of a_{ij} and b_{ij} are given by

$$(a_{ij} \mid \boldsymbol{y}, k) \overset{\mathrm{d}}{=} \mathrm{Be}(\lambda_{ij}^*, \lambda_{i\boldsymbol{\cdot}}^* - \lambda_{ij}^*) \quad \text{and} \quad (b_{ij} \mid \boldsymbol{y}, k) \overset{\mathrm{d}}{=} \mathrm{Be}(\gamma_{ij}^*, \gamma_{i\boldsymbol{\cdot}}^* - \gamma_{ij}^*)\,. \tag{7}$$

The unconditional posteriors under the change-point model are then

$$\pi(\boldsymbol{a}_i \mid \boldsymbol{y}) = \frac{\pi(\boldsymbol{a}_i \mid y, k)\, \Pr(M_k \mid \boldsymbol{y})}{\Pr(M_0 \mid \boldsymbol{y})}\,, \tag{8}$$

a mixture of Dirichlet distributions.

The multinomial change-point problem occurs as a special case when the y_i are independent and $a_{ij} = a_j$, $b_{ij} = b_j$. The Bayes factor is then

$$B_{0k} = \frac{\Gamma(\alpha_{\boldsymbol{\cdot}})\Gamma(\lambda^*_{\boldsymbol{\cdot}})\Gamma(\gamma^*_{\boldsymbol{\cdot}})}{\Gamma(\alpha^*_{\boldsymbol{\cdot}})\Gamma(\lambda_{\boldsymbol{\cdot}})\Gamma(\gamma_{\boldsymbol{\cdot}})} \prod_{j=1}^{p} \frac{\Gamma(\alpha_j^*)\Gamma(\lambda_j)\Gamma(\gamma_j^*)}{\Gamma(\alpha_j)\Gamma(\lambda_j^*)\Gamma(\gamma_j^*)}\,. \tag{9}$$

Another special case is when the interest is in the possibility of change-points in the transition probabilities of only certain states while the others remain unchanged. Then the transition matrices before and after time k are written as

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{E} \end{bmatrix} \quad \text{and} \quad \boldsymbol{B} = \begin{bmatrix} \boldsymbol{D} \\ \boldsymbol{E} \end{bmatrix}, \tag{10}$$

where $\boldsymbol{C}$ and $\boldsymbol{D}$ are $q \times p$ and $\boldsymbol{E}$ is $(p-q) \times p$.

© 2004 Australian Statistical Publishing Association Inc.

Thus the first q states have a change-point in their transition probabilities at k from $\boldsymbol{C}$ to $\boldsymbol{D}$. The likelihood function under this model (M_{sk}) is

$$f(\boldsymbol{y} \mid \boldsymbol{C}, \boldsymbol{D}, \boldsymbol{E}, k) = p(y_1) \prod_{t=1}^{k-1} c_{y_t, y_{t+1}}^{\delta_t} \prod_{t=k}^{n-1} d_{y_t, y_{t+1}}^{\delta_t} \prod_{t=1}^{n-1} e_{y_t, y_{t+1}}^{1-\delta_t},$$

where $\delta_t = \mathrm{I}(y_t \le q)$. This can be written as

$$f(\boldsymbol{y} \mid \boldsymbol{C}, \boldsymbol{D}, \boldsymbol{E}, k) = p(y_1) \prod_{i=1}^{q} \prod_{\ell=q+1}^{p} \prod_{j=1}^{p} c_{ij}^{z_{1ij}} d_{ij}^{z_{2ij}} e_{\ell j}^{z_{\ell j}},$$

where z_{1ij} and z_{2ij} are the number of transitions from i to j up to and after time k respectively with $i \le q$, and $z_{\ell j}$ is the number of transitions from ℓ to j where $\ell > q$.

With uniform priors on the rows of $\boldsymbol{C}$, $\boldsymbol{D}$ and $\boldsymbol{E}$, we have $\pi(\boldsymbol{C}) = \pi(\boldsymbol{D}) = \Gamma^q(p)$ and $\pi(\boldsymbol{E}) = \Gamma^{p-q}(p)$, and the marginal likelihood becomes

$$m_{sk}(\boldsymbol{y}) = p(y_1)\Gamma^{p+q}(p) \prod_{i=1}^{q} \frac{\prod_{j=1}^{p} \Gamma(z_{1ij}+1)\Gamma(z_{2ij}+1)}{\Gamma(z_{1i\boldsymbol{.}}+p)\Gamma(z_{2i\boldsymbol{.}}+p)} \prod_{\ell=q+1}^{p} \frac{\prod_{j=1}^{p} \Gamma(z_{\ell j}+1)}{\Gamma(z_{\ell\boldsymbol{.}}+p)}. \qquad (11)$$

This can then be compared with the no-change model in (2) or the completely changing model in (3), and the Bayes factors obtained.

3. Multiple change-points

If the number of possible change-points is unknown but bounded, let $M_{\boldsymbol{k}}^r$ denote the model with r change-points where $\boldsymbol{k} = (k_1, \ldots, k_r)$ denotes the positions of the change-points, and $r = 0, 1, \ldots, R$, where R denotes the maximum possible number of change-points.

Under model $M_{\boldsymbol{k}}^r$ we have transition matrices $\boldsymbol{A}_0, \boldsymbol{A}_1, \ldots, \boldsymbol{A}_r$ with priors

$$\pi(\boldsymbol{A}_\ell) = \prod_{i=1}^{p} f_{\mathrm{D}}(\boldsymbol{a}_{\ell i} \mid \boldsymbol{\lambda}_{\ell i}) \qquad (\ell = 0, 1, \ldots, r),$$

where $\boldsymbol{a}_{\ell i} = (a_{\ell i1}, \ldots, a_{\ell ip})$ and $a_{\ell ij}$ denotes the transition probability from state i to j after the ℓth change-point.

The joint distribution of all quantities is then proportional to

$$p(y_1) \prod_{\ell=0}^{r} \prod_{i=1}^{p} \prod_{j=1}^{p} a_{\ell ij}^{\lambda_{\ell ij}+z_{\ell ij}-1} \pi(r)\pi(\boldsymbol{k} \mid r).$$

Here the distribution of $\boldsymbol{k} \mid r$ can be uniform over all partitions of $\boldsymbol{k}$, while $\pi(r) = 1/(R+1)$ and $z_{\ell ij}$ is the number of transitions from i to j of the observations between k_ℓ and $k_{\ell+1} - 1$, where $k_0 = 1$ and $k_{r+1} = n$.

With $\lambda_{\ell ij} = 1$ for all i, j, the marginal likelihood under $M_{\boldsymbol{k}}^r$ is given by

$$m_{\boldsymbol{k}}^r(\boldsymbol{y}) = p(y_1)\big(\Gamma(p)\big)^{p(r+1)} \prod_{\ell=0}^{r} \prod_{i=1}^{p} \frac{\prod_{j=1}^{p} \Gamma(z_{\ell ij}+1)}{\Gamma(z_{\ell i\boldsymbol{.}}+p)}.$$

© 2004 Australian Statistical Publishing Association Inc.

The Bayes factors in favour of no change-point when compared with models $M_{\boldsymbol{k}}^r$, $B_{0\boldsymbol{k}}^r$, can now be computed as in the previous section. Let

$$B_{\boldsymbol{j}\boldsymbol{k}}^{ir} = \frac{m_{\boldsymbol{j}}^i(\boldsymbol{y})}{m_{\boldsymbol{k}}^r(\boldsymbol{y})}$$

denote the Bayes factor for partition $\boldsymbol{j}$ of i change-points against partition $\boldsymbol{k}$ of r change-points. Then the posterior probabilities with uniform priors on r and $\boldsymbol{k} \mid r$ can be written as

$$\pi(M_{\boldsymbol{k}}^s \mid \boldsymbol{y}) = \pi(s, \boldsymbol{k} \mid \boldsymbol{y}) = \pi(\boldsymbol{k} \mid r = s)\left(\sum_{i=0}^{R} \pi(\boldsymbol{k} \mid r = i) \sum_{\boldsymbol{j}} B_{\boldsymbol{j}\boldsymbol{k}}^{i,s}\right)^{-1},$$

where the summations over $\boldsymbol{j}$ mean over all possible partitionings for the given number of change-points. The posterior probability for the number of change-points is then

$$\pi(r \mid \boldsymbol{y}) = \sum_{\boldsymbol{k}} \pi(r, \boldsymbol{k} \mid \boldsymbol{y}) . \tag{12}$$

4. Markov chain Monte Carlo methods

The direct evaluation of the Bayes factors and posterior probabilities in Section 3 may need extensive numerical calculations if the chain is long and a large number of possible change-points are to be considered. The Gibbs sampling scheme of Carlin *et al.* (1992) works well when one change-point is assumed, but the number of calculations increases exponentially as the number of change-points is increased. Chib (1998) provides a Markov chain Monte Carlo method for estimating multiple change-point models when the number of change-points is known. The value of this approach lies in the fact that the number of change-points assumed has little impact on the computational effort required. Here now follows a brief outline of the model parameterization and sampling scheme as applied to our problem of changing transition matrices in a Markov chain. For more details on the general approach, see Chib (1996, 1998).

Consider the Markov chain model with transition matrices $\boldsymbol{A}_0, \boldsymbol{A}_1, \ldots, \boldsymbol{A}_r$, r change-points and $\boldsymbol{A}_k = [a_{kij}]$. The formulation of Chib (1996) is based on the introduction of a latent variable $s_t \in \{0, 1, \ldots, r\}$, indicating the regime from which the tth observation has been drawn.

The variable s_t is modelled as a Markov chain (not to be confused with the data model) with $(r+1) \times (r+1)$ transition matrix $\boldsymbol{P} = [p_{ij}]$ constrained so that s_{t+1} can only stay in the same state as before or move to one higher, i.e. $p_{ii} + p_{i,i+1} = 1$ for $i = 0, 1, \ldots, r-1$ and $p_{rr} = 1$. The chain starts in state 0 and terminates in state r, indicating the regime the chain is in at each point t. The quantities $s_t, \boldsymbol{P}, \boldsymbol{A}_0, \boldsymbol{A}_1, \ldots, \boldsymbol{A}_r$ are simulated successively from the full conditional distributions using Gibbs sampling. The sample output s_t then determines the distribution of the change-points.

The values of s_t are sampled in reverse order where $s_n = r$, and

$$\begin{aligned}\Pr(s_t = s_{t+1} \mid \boldsymbol{Y}_n, \boldsymbol{S}^{t+1}, \boldsymbol{A}_{s_{t+1}} \boldsymbol{A}_{s_{t+1}-1}, \boldsymbol{P}) \\ = \left(1 + \frac{\Pr(s_t = s_{t+1} - 1 \mid \boldsymbol{Y}_t, \boldsymbol{A}_{s_{t+1}-1}, \boldsymbol{P}) p_{s_{t+1}-1, s_{t+1}}}{\Pr(s_t = s_{t+1} \mid \boldsymbol{Y}_t, \boldsymbol{A}_{s_{t+1}}, \boldsymbol{P}) p_{s_{t+1}, s_{t+1}}}\right)^{-1}\end{aligned} \tag{13}$$

© 2004 Australian Statistical Publishing Association Inc.

for $t = n-1, n-2, \dots, 2,$ and with $s_1 = 1$. Also $\boldsymbol{S}^{t+1} = (s_{t+1}, \dots, s_n)$ and $\boldsymbol{Y}_t = (y_1, \dots, y_t)$. The probabilities in (13) are obtained from

$$\Pr(s_t = k \mid \boldsymbol{Y}_t, \boldsymbol{A}_k, \boldsymbol{P}) \propto \Pr(s_t = k \mid \boldsymbol{Y}_{t-1}, \boldsymbol{A}_k, \boldsymbol{P}) a_{k y_{t-1} y_t} ,$$

where, in turn,

$$\Pr(s_t = k \mid \boldsymbol{Y}_{t-1}, \boldsymbol{A}_k, \boldsymbol{P}) = \sum_{\ell=k-1}^{k} \Pr(s_{t-1} = \ell \mid \boldsymbol{Y}_{t-1}, \boldsymbol{A}_\ell, \boldsymbol{P}) p_{\ell k} .$$

For generating the elements of $\boldsymbol{P}$, assume independent $\mathrm{Be}(a, b)$ priors for the diagonal elements of $\boldsymbol{P}$. The conditional distribution of p_{ii}, $i = 0, \dots, r$, depends only on the set $\boldsymbol{S}_n = (s_1, \dots, s_n)$ and is given by $(p_{ii} \mid s_n) \overset{\mathrm{d}}{=} \mathrm{B}(a+n_{ii}, b+1)$, where n_{ii} denotes the number of time periods spent in state i.

The last step in the cycle is to generate the elements of transition matrices $\boldsymbol{A}_0, \boldsymbol{A}_1, \dots, \boldsymbol{A}_r$. The conditional distribution of the ith row $(i = 1, \dots, p)$ of $\boldsymbol{A}_k$ is $(\boldsymbol{a}_{ki} \mid \boldsymbol{Y}_n, \boldsymbol{S}_n) \overset{\mathrm{d}}{=} \mathrm{D}(\boldsymbol{\lambda}^*_{ki})$, where $\lambda^*_{kij} = \lambda_{kij} + z^{(k)}_{ij}$ and $z^{(k)}_{ij}$ is the number of transitions under regime k. Samples from the Dirichlet distribution can be obtained from its relationship with the Gamma distribution. If $x_i \overset{\mathrm{d}}{=} \mathrm{Ga}(c_i, 1)$, $i = 1, 2, \dots, p$, independently, then $\boldsymbol{y} \overset{\mathrm{d}}{=} \mathrm{D}(\boldsymbol{c})$ where $y_i = x_i / x_{\bullet}$.

Care must be taken with the specification of the hyperparameters a and b. In general we should have $a \gg b$, so that there is some prior resistance to a switch from one regime to the next. The prior mean length of a regime is approximately $(a+b)/b$, so a rule of thumb is to set $(a+b)/b = n/(r+1)$, making the prior expected lengths of all regimes equal, and to choose b small, so as to increase prior variances.

For the comparison of alternative change-point models, the marginal likelihoods and Bayes factors are calculated using the output from the posterior simulations above. Details can be found in Chib (1998 Section 3).

5. Applications

Example 1

In an illustration of the use of Gibbs sampling in change-point analysis Carlin *et al.* (1992) generated 50 observations from a three-state stationary Markov chain with a change after $k = 35$. The transition matrices before and after the change are

$$\boldsymbol{A} = \begin{bmatrix} 0.70 & 0.15 & 0.15 \\ 0.33 & 0.33 & 0.33 \\ 0.33 & 0.33 & 0.33 \end{bmatrix} \quad \text{and} \quad \boldsymbol{B} = \begin{bmatrix} 0.33 & 0.33 & 0.33 \\ 0.15 & 0.70 & 0.15 \\ 0.33 & 0.33 & 0.33 \end{bmatrix} .$$

In Figure 1(a) the results of Carlin *et al.* (1992; CGS) are shown, with uniform priors and the assumption of exactly one change-point. Figure 1(b) shows the posterior probabilities for the position of the change-point using (5) and (6).

As only the transition probabilities from the first two states change in this example, we can apply the theory for the model in (9). Figure 1(c) gives the results for this model from (10) with $q = 2$. The bars at $k = 50$ in Figures 1(b) and 1(c) represent the probabilities for no change-point, which are, respectively, 0.128 and 0.100.

© 2004 Australian Statistical Publishing Association Inc.

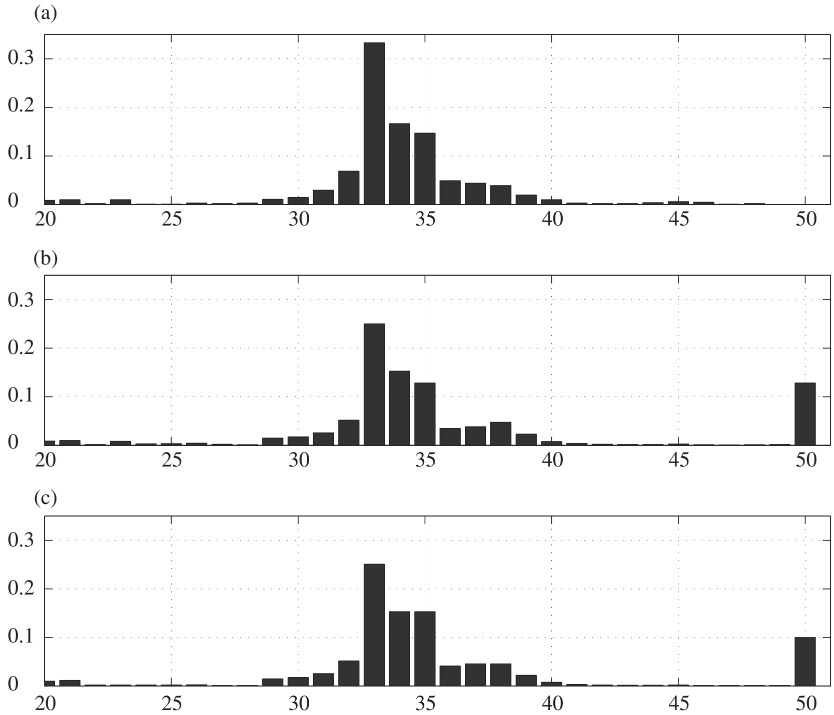

Figure 1. Posterior probabilities for change-point k for the CGS example: (a) CGS results, assuming a change-point; (b) results from (5) and (6) with uniform prior; (c) results for model with partial change, from (11). Bar at $k = 50$ represents probability for no change.

When comparing the model with changes in the first two states, M_{sk}, with the model with changes in all states, M_k, the posterior probability for M_{sk} is 0.548, a very slight indication that this is the more appropriate model.

Simulation study

To get an idea of the sensitivity and frequentist properties of the Bayesian procedure, a limited simulation study was done, based on the form of Example 1. Fifty observations were generated from a three-state Markov chain with a possible change after $k = 35$. The transition matrices are

$$\boldsymbol{A} = \begin{bmatrix} 0.3 & 0.1 & 0.6 \\ 0.6 & 0.3 & 0.1 \\ 0.1 & 0.6 & 0.3 \end{bmatrix} \quad \text{and} \quad \boldsymbol{B} = \begin{bmatrix} 0.3+0.5d & 0.1+0.5d & 0.6-d \\ 0.6-d & 0.3+0.5d & 0.1+0.5d \\ 0.1+0.5d & 0.6-d & 0.3+0.5d \end{bmatrix},$$

so that the differences in the transition probabilities increase with d. Three thousand chains were generated for each value of $d = 0(0.1)0.6$, and for the uniform prior as well as the Jeffreys prior. Table 1 gives the results. The first two rows give the average posterior probability of no change for each value of d. The last two rows show the proportion of correct decisions, i.e. the proportion of times the posterior probability is larger than 0.5 when $d = 0$, or smaller than 0.5 when $d > 0$.

With the uniform prior, the average posterior probability for no change over the 3000 samples is 0.776 when the true model is M_0. Small or moderate changes in the transition

© 2004 Australian Statistical Publishing Association Inc.

TABLE 1

Simulation results

d	0	0.1	0.2	0.3	0.4	0.5	0.6
	Average posterior probability of no change						
Uniform prior	0.776	0.664	0.525	0.333	0.173	0.059	0.008
Jeffreys prior	0.787	0.699	0.574	0.380	0.203	0.066	0.005
	Proportion correct decisions						
Uniform prior	0.884	0.245	0.440	0.695	0.886	0.980	0.9997
Jeffreys prior	0.870	0.238	0.388	0.628	0.835	0.965	0.9997

matrix ($d \leq 0.2$) are not easily detected, as reflected in the low percentage of correct decisions in those categories. The uniform prior seems to perform better than the Jeffreys prior, at least for this particular configuration, and a small posterior probability, say less than 0.3, indicates a substantial change in the transition matrix.

Example 2

Colwell, Jones & Gillett (1990) show that the outcomes of the Ashes cricket tests between Australia and England can be modelled very well by a three-state first-order Markov chain with an Australian win, an English win, and a draw as the three states. We use the data from the last 125 tests (from 1959 to 2001) to examine the chain for possible change-points in the transition probabilities.

The data show strong evidence of exactly one change-point. Assuming at most two change-points, and using (12), the posterior probabilities for 0, 1 or 2 change-points are, respectively 0.061, 0.894 and 0.055 with uniform priors. The results, assuming at most one change-point, are shown in Figure 2.

Figure 2(a) gives the exact posterior probabilities for the position of the change-point according to (6) with the bar at $k = 0$ the probability of no change, which is now 0.086. Figure 2(b) shows the result from 10 000 Gibbs samples drawn according to the approach of Chib (1998) given in Section 4. Here exactly one change-point is assumed, and hyperparameters $a = 30$ and $b = 0.5$ are used. From Figure 2 it can be seen that a change is likely to have occurred somewhere between the 83rd and 116th observations (between 1986 and 1997) with posterior probability of 0.80 for that interval. The maximum posterior probabilities, assuming two change-points, are at observations 27 (1968) and 86 (1989).

Assuming one change, Figure 3 shows the unconditional marginal posterior densities of a_{11} and b_{11}, the probability of an Australian win given an Australian win in the previous test match, before and after the change-point, according to (7) and (8).

The unconditional means of the transition matrices before ($\boldsymbol{A}$) and after the change-point ($\boldsymbol{B}$) are (with an Australian win, an English win, and a draw, respectively)

$$\boldsymbol{A} = \begin{bmatrix} 0.320 & 0.160 & 0.520 \\ 0.400 & 0.358 & 0.242 \\ 0.277 & 0.302 & 0.421 \end{bmatrix} \quad \text{and} \quad \boldsymbol{B} = \begin{bmatrix} 0.528 & 0.294 & 0.178 \\ 0.525 & 0.141 & 0.334 \\ 0.571 & 0.245 & 0.184 \end{bmatrix},$$

showing the increasing dominance of the Australians during the recent past. Interestingly, the probability of an English win after an Australian win has increased significantly even though the overall probability of a win has decreased.

© 2004 Australian Statistical Publishing Association Inc.

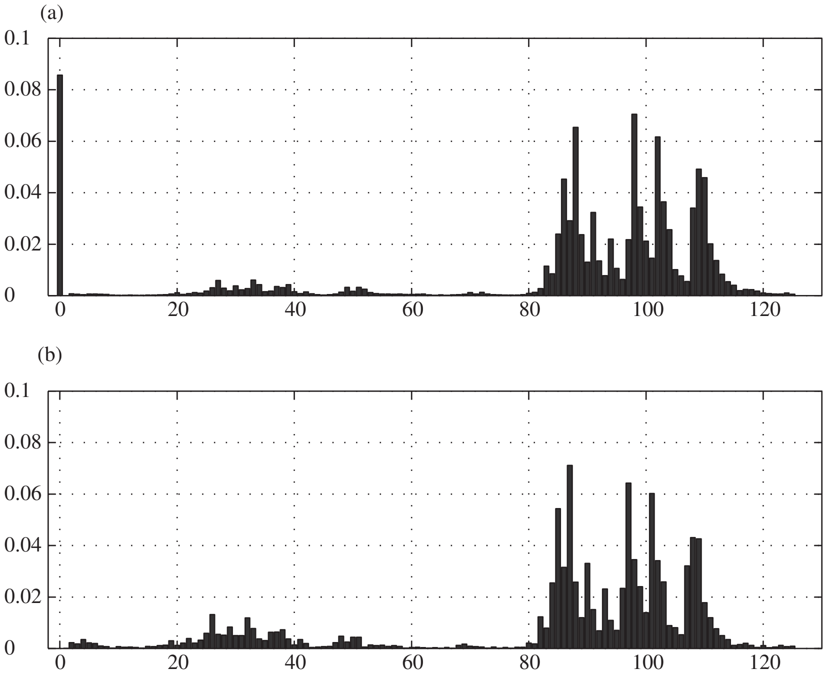

Figure 2. Posterior probabilities for a single change-point k for the cricket test-matches data: (a) exact results with uniform prior, assuming no or one change-point; (b) results from Gibbs sampling, assuming one change-point

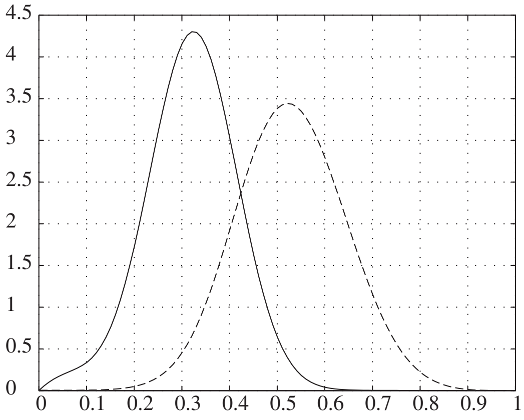

Figure 3. Posterior densities of a_{11} (before the change-point —) and b_{11} (after the change-point - - -)

To present a frequentist evaluation of the results in the above example, 2000 sequences of the same length were generated with no change-point, using the maximum likelihood estimator of the single transition matrix, and the posterior probability of no change calculated in each case. The average of these probabilities is 0.614 with standard deviation 0.183. The proportion of samples yielding a posterior probability of less than the observed one of 0.086 is 0.012. This can be interpreted as a P-value.

© 2004 Australian Statistical Publishing Association Inc.

TABLE 2

Transition counts

	2nd interview				3rd interview				4th interview		
1st	R	D	N	2nd	R	D	N	3rd	R	D	N
R	125	5	16	R	124	3	16	R	146	2	4
D	7	106	15	D	6	109	14	D	6	111	4
N	11	18	142	N	22	9	142	N	40	36	96

Example 3

The following example appears in Anderson (1954) and was also used by Bhat (1984 Chapter 5). It consists of data collected by the Bureau of Applied Social Research on voter attitudes in Erie County, Ohio, during 1940. A group of people was interviewed about their voting preferences (D = Democrat, R = Republican, N = Do not know, or Other candidates). The group was interviewed four times, twice before the national conventions, once between the Republican and Democratic conventions, and once after both conventions. Table 2 gives the transition counts for the pairs of three successive interviews.

The pertinent question is whether these results reflect the same behaviour on the part of the voters over the four-month period, or if one or both of the conventions had a significant effect on voters' preferences. In this situation, with multiple chains, the change-point model reduces to four models; no change (M_0), a change only after the Republican convention (M_1), a change only after the Democratic convention (M_2), or two changes, one after each convention (M_{12}). The Bayes factor gives overwhelming evidence in favour of model M_2. With equal prior weights for the four models and uniform priors on the transition probabilities, the posterior probabilities are 0, 0, 0.972 and 0.028 respectively. so a change occurred after both conventions. It can be seen that this change is mainly due to people from preference N finally deciding which way to vote; it is not really due to a swing towards a particular party.

What do the classical likelihood ratio tests tell us? When testing the null hypothesis of stationarity (model M_0) against each of the other three models, the results are all highly significant with $P < 0.005$ for model M_1 and $P < 0.001$ for models M_2 and M_{12}. When testing the null hypothesis of one change (M_2) against the alternative of two changes (M_{12}), the result is not significant, as expected, with $P > 0.25$. The classical tests, however, cannot convey the overall strength of evidence in favour of a particular model in the same way as posterior probabilities do.

6. Discussion

In this paper we discuss the use of Bayes factors and posterior probabilities for the detection of change-points in the transition matrix of a stationary Markov chain. When the number of change-points is fixed and known, even the use of improper priors yields reasonable answers. That is because the parameters under the models being compared have essentially the same interpretation; that is, the priors are exchangeable, causing the implied normalizing constants to cancel out. So most papers on Bayesian change-point analysis assume the number of change-points known. However, when there is uncertainty about the existence of a change-point and one of the possible models is of no change, the dimensions of the parameter space are different under different models, and improper priors are no longer applicable.

© 2004 Australian Statistical Publishing Association Inc.

In the case of the Markov chain model all parameters are bounded, so non-informative but proper priors are available, for which the exact analytical results are given. However, the improper prior can be used in conjunction with the fractional Bayes factor of O'Hagan (1995, 1997) if the data contain transitions from every state to every other state in every partition of the time space, which is often the case with multiple chains as in Example 3.

A note on the computational aspects: the programs are written in MATLAB and run on a 1 GHz PC. When no or one change-point are considered the calculation is very quick, even for very long chains. For Example 2 with $n = 125$ and with no, one or two possible change-points, the program runs for about 7 minutes. This is about the same length of time it takes to run 10 000 simulations using the Gibbs sampling scheme of Section 4 for two change-points. When the number of possible change-points is increased to three, the exact evaluation takes 8 hours, while the time for the Gibbs sampling only increases to 15 minutes. However, in the latter case additional simulations are required to find the Bayes factors for model comparisons, which follows directly in the case of the exact evaluation.

Gupta & Chen (1996) use the Schwartz information criterion and a step-wise procedure to detect multiple change-points. First they test for a single change-point, and if one is found the sequence is split at that point and the two subsequences are each tested for an additional change-point. The process is repeated until no further change-points are found. However, the flaw in this procedure is that the estimated set of change-points under a particular model is not necessarily a subset of the set of change-points under a larger model.

The Bayesian analysis of change-points in a Markov chain is a useful tool, especially in the social and behavioural sciences, when the effect of an event on the attitudes of people is to be evaluated. A possible extension to the models discussed is one in which a more gradual change in attitude occurs.

References

ANDERSON, T.W. (1954). Probability models for analyzing time changes in attitudes. In *Mathematical Thinking in the Social Sciences*, ed. P.F. Lazarsfeld, pp . 17–66. Glencoe, Ill: The Free Press.

BARRY, D. & HARTIGAN, J.A. (1992). Product partition models for change-point problems. *Ann. Stat.* **20**, 260–279.

BARRY, D. & HARTIGAN, J.A. (1993). A Bayesian analysis for change-point problems. *J. Amer. Statist. Assoc.* **88**, 309–319.

BHAT, U.N. (1984). *Elements of Applied Stochastic Processes.* New York: Wiley.

BROEMELING, L.D. (1972). Bayesian procedures for detecting a change in a sequence of random variables. *Metron.* **30**, 1–14.

BROEMELING, L.D. (1974). Bayesian inferences about a changing sequence of random variables. *Comm. Statist.* **3**, 243–255.

BROEMELING, L.D. (1977). Forecasting future values of a changing sequence. *Comm. Statist.* **A6**, 87–102.

BROEMELING, L.D. & GREGURICH, M.A. (1996). On a Bayesian approach for the shift point problem. *Comm. Statist. Theory Methods* **25**, 2267–2279.

BROEMELING, L.D. & TSURUMI, H. (1987). *Econometrics and Structural Change.* New York: Marcel Dekker Inc.

CARLIN, B.P., GELFAND, A.E. & SMITH, A.F.M. (1992). Hierarchical Bayesian analysis of change-point problems. *Appl. Statist.* **41**, 389–405.

CHERNOFF, H. & ZACKS, S. (1964). Estimating the current mean of a normal distribution which is subject to changes in time. *Ann. Math. Statist.* **35**, 999–1018.

CHIB, S. (1996). Calculating posterior distributions and model estimates in Markov mixture models. *J. Econometrics* **75**, 79–98.

CHIB, S. (1998). Estimation and comparison of multiple change-point models. *J. Econometrics* **86**, 221–241.

© 2004 Australian Statistical Publishing Association Inc.

COLWELL, D., JONES, B. & GILLETT, J. (1991). A Markov Chain in Cricket (MCC!). *The Mathematical Gazette*, June 1991, 183–185.

CURNOW, R.N. & KIRKWOOD, T.B.L. (1989). Statistical analysis of deoxyribonucleic acid sequence data — a review. *J. Roy. Statist. Soc. Ser. A* **152**, 199–220.

DEY, D.K. & PURKAYASTHA, S. (1997). Bayesian approach to change-point problems. *Comm. Statist. Theory Methods* **26**, 2035–2047.

GARISCH, I. & GROENEWALD, P.C.N. (1999). The Nile revisited: Change-point analysis with autocorrelation. In *Bayesian Statistics 6*, eds J.M. Bernardo, J.O. Berger, A.P. Dawid & A.F.M. Smith, pp. 753–760. Oxford: Oxford University Press.

GINGERICH, P.D. (1969). Markov analysis of cyclic alluvial sediments. *J. Sedimentary Petrology* **39**, 330–332.

GREEN, P.J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711–732.

GUPTA, A.K. & CHEN, J. (1996). Detecting changes of mean in multidimensional normal sequences with application to literature and geology. *Comput. Statist.* **11**, 211–221.

HISCOTT, R.N. (1981). Chi-square tests for Markov chain analysis. *Math. Geol.* **13**, 69–80.

O'HAGAN, A. (1995). Fractional Bayes factors for model comparison. *J. Roy. Statist. Soc. Ser. B* **57**, 99–148.

O'HAGAN, A. (1997). Properties of intrinsic and fractional Bayes factors. *Test* **6**, 101–118.

PHILLIPS, D.B. & SMITH, A.F.M. (1996). Bayesian model comparison via jump diffusions. In *Markov Chain Monte Carlo in Practice*, eds W.R. Gilks, S. Richardson & D.J. Spiegelhalter, pp. 215–239. London: Chapman and Hall.

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. & COOK, D.G. (1980). Straight lines with a change-point: a Bayesian analysis of some renal transplant data. *Appl. Statist.* **29**, 180–189.

STEPHENS, D.A. (1994). Bayesian retrospective multiple change-point identification. *Appl. Statist.* **43**, 159–178.

WHITAKER, D. (1978). The derivation of a measure of brand loyalty using a Markov brand switching model. *J. Opl. Res. Soc.* **29**, 959–970.

ZACKS, S. (1991). Detection and change-point problems. In *Handbook of Sequential Analysis*, eds B.K. Ghosh & P.K. Sen, pp. 531–562. New York: Dekker.

© 2004 Australian Statistical Publishing Association Inc.