

BAYESIAN ESTIMATION OF THE NUMBER OF CHANGE POINTS

Chung-Bow Lee

National Chung-Hsing University

Abstract: The problem of estimating the number of change points in a sequence of independent random variables is considered in a Bayesian framework. We find that, under mild assumptions and with respect to a suitable prior distribution, the posterior mode of the number of change points converges to the true number of change points in the frequentist sense. Furthermore, the posterior mode of the locations of the change points is shown to be within $O_p(\log n)$ of the true locations of the change points where n is the sample size. The prior distribution on the locations of the change points may be taken to be uniform. Finally, some simulated results are given, showing that the method works well in estimating the number of change points.

Key words and phrases: Change points, posterior distribution.

1. Introduction

Consider a sequence of independent random variables $\underline{X} = \underline{X}^{(n)} = (x_1, \dots, x_n)$ with the following distribution :

$$x_i \sim \begin{cases} f(\cdot; \theta_1), & \text{if } 1 \leq i \leq J_1, \\ f(\cdot; \theta_r), & \text{if } J_{r-1} < i \leq J_r, \quad r = 2, \dots, k, \\ f(\cdot; \theta_{k+1}), & \text{if } J_k < i \leq n, \end{cases} \quad (1.1)$$

where $\{f(\cdot; \theta) : \theta \in \Theta\}$ is a family of densities (with respect to Lebesgue measure μ), and $\underline{\theta}^{(k+1)} = (\theta_1, \dots, \theta_{k+1})$, $\underline{J}^{(k)} = (J_1, \dots, J_k)$ and k are unknown parameters. Here k is the number of change points in the sequence $\underline{X}^{(n)}$, $\underline{J}^{(k)}$ consists of the locations of the k change points, and the parameters θ_i satisfy $\theta_i \neq \theta_{i+1}$ for $i = 1, \dots, k$.

We are mainly concerned with estimating the number k of change points as well as their locations $\underline{J}^{(k)}$. Adopting the Bayesian approach, we will place a suitable prior distribution π on k and $\underline{J}^{(k)}$ and estimate them by the posterior mode. The marginal posterior mode \hat{k} of k maximizes the posterior density of k (given $\underline{X}^{(n)}$) which takes the form

$$\pi(k | \underline{X}^{(n)}) \propto \sum_{\underline{J}^{(k)}} \left[\int f(\underline{X}^{(n)}; \underline{\theta}^{(k+1)}, k, \underline{J}^{(k)}) \pi(\underline{\theta}^{(k+1)} | k, \underline{J}^{(k)}) d\underline{\theta}^{(k+1)} \right] \pi(k, \underline{J}^{(k)}), \quad (1.2)$$

where $f(\underline{X}^{(n)}; \underline{\theta}^{(k+1)}, k, \underline{J}^{(k)})$ denotes the joint density of x_1, \dots, x_n given $\underline{\theta}^{(k+1)}$, $k, \underline{J}^{(k)}$. The main result of this paper is that, under regularity assumptions, \hat{k} converges in probability to the true value of k (to be denoted by k_0) in the frequentist sense. Furthermore, let $\hat{\underline{J}}^{(k)} = (\hat{J}_1^{(k)}, \dots, \hat{J}_k^{(k)})$ maximize $\pi(k, \underline{J}^{(k)} | \underline{X}^{(n)})$ over $\underline{J}^{(k)}$ with k fixed. We also show that $\|\hat{\underline{J}}^{(k_0)} - \underline{J}^0\| \equiv [\sum_{i=1}^{k_0} (\hat{J}_i^{(k_0)} - J_i^0)^2]^{1/2} = O_p(\log n)$ where $\underline{J}^0 = (J_1^0, \dots, J_{k_0}^0)$ denotes the true locations of the change points. Note that k_0 and \underline{J}^0 are estimated by \hat{k} and $\hat{\underline{J}}^{(\hat{k})}$, respectively.

Schwarz (1978) considers the problem of model selection, based on decision theory, for the exponential family with parameters of various dimensions. He uses decision theory to find that the Bayes solution is to choose k which maximizes

$$S(\underline{X}^{(n)}, k) \propto \int f(\underline{X}^{(n)}; \underline{\theta}^{(k)}) \pi(\underline{\theta}^{(k)}) d\underline{\theta}^{(k)} \pi(k), \quad (1.3)$$

where $f(\underline{X}^{(n)}; \underline{\theta}^{(k)})$ is the joint density of $\underline{X}^{(n)} = (x_1, \dots, x_n)$ in an exponential family with k -dimensional parameter $\underline{\theta}^{(k)}$ for model k and $\pi(k)$ is a prior probability for model k . For n sufficiently large, he derives an asymptotically optimal solution that is to choose k which maximizes

$$SC(\underline{X}^{(n)}, k) = \log f(\underline{X}^{(n)}; \hat{\underline{\theta}}^{(k)}) - k \log n/2, \quad (1.4)$$

where $\hat{\underline{\theta}}^{(k)}$ is the maximum likelihood estimator (m.l.e.) of $\underline{\theta}^{(k)}$.

Yao (1988) uses Schwarz's criterion for the problem of estimating the number of change points in a sequence of independent normal random variables with common variance σ^2 . He finds that, under mild conditions, the estimator \hat{k} which maximizes

$$SC(\underline{X}^{(n)}, k) = -n \log \hat{\sigma}_k^2/2 - k \log n \quad (1.5)$$

converges to k_0 in probability where $\hat{\sigma}_k^2$ is the m.l.e. of σ^2 given k . Note that given k , the total number of parameters in $\underline{\theta}^{(k+1)}$, $\underline{J}^{(k)}$ and σ^2 equals $2k + 2$.

Our method is to find k which maximizes the integrated likelihood function over $\underline{\theta}^{(k+1)}$ and $\underline{J}^{(k)}$. While this idea in dealing with parameters $\underline{\theta}^{(k+1)}$ and $\underline{J}^{(k)}$ may be considered as an extension of (1.3) in Schwarz's work, our method is somewhat different from that given by Yao (1988), which directly comes from (1.4) of Schwarz (1978) and may be interpreted as finding k that maximizes

$$SC^*(\underline{X}^{(n)}, k) \propto \max_{\underline{\theta}^{(k+1)}, \underline{J}^{(k)}} \log f(\underline{X}^{(n)}; \underline{\theta}^{(k+1)}, k, \underline{J}^{(k)}) - k \log n.$$

Barry and Hartigan (1992) consider the product partition model and give a Bayesian analysis for the problem of multiple change points. They show that,

under suitable choice of prior cohesions and mild assumptions, if there is no change point (i.e. $k_0 = 0$), the posterior probability of $k = 0$ converges to 1 in probability (in the frequentist sense). They also compare, by simulation, the product partition method with that given by Yao (1988) and find that Yao’s method is better at detecting the number of change points, but they suggest that this defect may be overcome by considering different cohesions. Our study attempts to find a large class of suitable prior distributions for which the number of change points can be estimated consistently. Our simulation results show that the precision of estimating the number of change points can be improved greatly by using the uniform prior.

Other references about estimating the number of change points from a Bayesian point of view are given by Chernoff and Zacks (1964), Yao (1984) and Barry and Hartigan (1993) in the context of independent normal random variables.

The present paper is organized as follows. In section 2, assumptions and notations are given. In section 3, the consistency of the posterior mode \hat{k} is proved. Section 4 contains simulation results. The proofs of several lemmas in section 3 are relegated to the appendix.

2. Assumptions and Notations

2.1. Assumptions

Suppose $\underline{X}^{(n)} = (x_1, \dots, x_n)$ is a sequence of n independent random variables from model (1.1) satisfying the following conditions :

- (A1) The true number k_0 of change points is bounded by a known constant R_0 and the true change point locations $\underline{J}^0 = (J_1^0, \dots, J_{k_0}^0)$ satisfy $0 < J_1^0 < J_2^0 < \dots < J_{k_0}^0 < n$ and $\min_{1 \leq i \leq k_0+1} [(J_i^0 - J_{i-1}^0) / \log n] \rightarrow \infty$, as $n \rightarrow \infty$, where $J_0^0 = 0$ and $J_{k_0+1}^0 = n$.
- (A2) Let $\Theta \subset R$ be an open interval and $\bar{\Theta}$ the closure of Θ . $\forall \theta \in \Theta$ and $\theta' \in \bar{\Theta}$, provided $\theta \neq \theta'$, $\int_{-\infty}^{\infty} |f(x; \theta) - f(x; \theta')| d\mu(x) > 0$.
- (A3) $f(x; \theta)$ is jointly measurable in (x, θ) .
- (A4) $\forall \theta \in \Theta$, the derivatives $\frac{\partial \log f(x; \theta)}{\partial \theta}$, $\frac{\partial^2 \log f(x; \theta)}{\partial \theta^2}$ and $\frac{\partial^3 \log f(x; \theta)}{\partial \theta^3}$ exist, $\forall x$; and are continuous in θ .
- (A5) Let θ_i^0 denote the true parameter value in the interval $(J_{i-1}^0, J_i^0], i = 1, \dots, k_0 + 1$. Then there exist functions $G_1(x), G_2(x)$ and $H(x)$ such that

$$\left| \frac{\partial \log f(x; \theta)}{\partial \theta} \right| \leq G_1(x), \left| \frac{\partial^2 \log f(x; \theta)}{\partial \theta^2} \right| \leq G_2(x), \left| \frac{\partial^3 \log f(x; \theta)}{\partial \theta^3} \right| \leq H(x),$$

for all x and for all θ in a neighborhood of $\theta_i^0, i = 1, \dots, k_0 + 1$; and

$$E_{\theta_i^0}[G_1(x)] < \infty, E_{\theta_i^0}[G_2(x)] < \infty \text{ and } E_{\theta_i^0}[H(x)] < \infty.$$

- (A6) The measures $\prod_{i=1}^n f(x_i; \theta)$ are mutually absolutely continuous for each $n = 1, 2, \dots$. Therefore, a null set will have probability zero for all θ .
- (A7) $\lim_{|\theta| \rightarrow \infty} f(x; \theta) = 0$ a.e. w.r.t. Lebesgue measure.
- (A8) $\forall \theta \in \Theta, E_\theta |\log f(x; \theta)| < \infty$ and $0 < I(\theta) = -E_\theta [\frac{\partial^2}{\partial \theta^2} \log f(x; \theta)]$.
- (A9) $\forall \theta \in \Theta, \rho, r > 0, f(x; \theta, \rho) = \sup_{|\theta' - \theta| \leq \rho} f(x; \theta')$ and $Q(x; r) = \sup_{|\theta| > r} f(x; \theta)$ are measurable functions of x , and for $i = 1, \dots, k_0 + 1$, sufficiently small ρ and sufficiently large r ,

$$E_{\theta_i^0} [\log f(x; \theta, \rho)]^+ < \infty, \quad E_{\theta_i^0} [\log Q(x; r)]^+ < \infty.$$

We now state two assumptions on the prior distribution π which consists of two parts $\pi(k, \underline{J}^{(k)})$ and $\pi(\underline{\theta}^{(k+1)} | k, \underline{J}^{(k)})$. We assume

- (B1) The prior density $\pi(k) > 0$, for $k = 0, 1, \dots, R_0$, and $\pi(\underline{J}^{(k)} | k) = 1/\binom{n-1}{k}$ for each possible location vector $\underline{J}^{(k)}$ (i.e. $\pi(\underline{J}^{(k)} | k)$ is uniform).
- (B2) Given k and $\underline{J}^{(k)}$, the conditional prior of $\underline{\theta}^{(k+1)}$ is such that the $k + 1$ components of $\underline{\theta}^{(k+1)}$ are independent with marginal probability density functions not depending on k and $\underline{J}^{(k)}$, which will be denoted by π' , i.e.

$$\pi(\underline{\theta}^{(k+1)} | k, \underline{J}^{(k)}) = \pi'(\theta_1) \cdots \pi'(\theta_{k+1}).$$

Furthermore, $\pi'(\theta)$ is positive and differentiable in a neighborhood of θ_i^0 , for each $i = 1, \dots, k_0 + 1$.

Remark 1. Conditions (A2)-(A9) are essentially those of Johnson (1970), which are just one set of the many variants (cf. Wald (1949), Wolfowitz (1965), Walker (1969)) to ensure that, (i) when θ_0 is the true value of θ for a random sample $\{x_1, \dots, x_n\}$, $\prod_{i=1}^n \frac{f(x_i; \theta)}{f(x_i; \theta_0)}$ will be sufficiently small for all values of θ outside a neighborhood of the true parameter θ_0 , and (ii) the posterior distribution of $\sqrt{n}(\hat{\theta} - \theta)$ is asymptotically normal, where $\hat{\theta}$ is the m.l.e. of θ . Thus, if there are a large number of x'_i s coming from one distribution and a large number of x''_i s coming from another distribution, say (x_1, \dots, x_j) from $f(\cdot; \theta_1)$, and (x_{j+1}, \dots, x_n) from $f(\cdot; \theta_2)$ with both j and $n - j$ large, it follows from the first result that the value $\prod_{i=1}^n f(x_i; \theta) / [\prod_{i=1}^j f(x_i; \theta_1) \cdot \prod_{i=j+1}^n f(x_i; \theta_2)]$ is small for all values of θ ; and if the random variables (x_1, \dots, x_n) have the same distribution, the second result implies that the integral $\int \prod_{i=1}^n \frac{f(x_i; \theta)}{f(x_i; \theta_0)} \pi(\theta) d\theta$ is of order $O(1/\sqrt{n})$ almost surely. By Taylor's expansion, the condition on the third derivative of the log-likelihood function in (A5) implies that $\hat{\theta}$ is asymptotically normal and $\log \prod_{i=1}^n \frac{f(x_i; \hat{\theta})}{f(x_i; \theta_0)}$ behaves like $1/(2n)$ times the square of the sum of a sequence of independent random variables with mean 0 and variance 1 (cf. lemma 3) so that it is dominated by $O(\log \log n)$ almost surely by the law of the iterated logarithm.

Remark 2. The assumption (B1) on the prior π may be relaxed (cf. Appendix B).

Remark 3. Note that the dependence of \underline{J}^0 on n has been suppressed. Strictly speaking, k_0 and $\underline{\theta}^0$ should depend on n as well. To avoid further assumptions and messy notation, we keep k_0 and $\underline{\theta}^0$ fixed as n increases. For the same reason, the prior π' on θ is kept fixed as n increases.

2.2. Notation

For convenience, we define the following notation:

- (1) $\forall \theta \in \Theta$, for any integers a and b ,

$$\prod_{(a,b] \setminus D} \frac{f(\underline{X}; \theta)}{f(\underline{X}; \underline{\theta}^0)} \equiv \begin{cases} 1, & \text{if set } D \supset (a, b], \\ \prod_{\substack{i=a+1 \\ i \notin D}}^b \frac{f(x_i; \theta)}{f(x_i; \underline{\theta}_{\delta(i)}^0)}, & \text{otherwise,} \end{cases}$$

where $\delta(i) = r$ if $J_{r-1}^0 < i \leq J_r^0$, $r = 1, \dots, k_0 + 1$. Note that $(a, b]$ denotes the set of integers between a and b (including b).

- (2) If set $D = \{(a_i, b_i] \mid i = 1, \dots, d\}$ is a collection of disjoint intervals, then

$$\prod_{(a,b] \setminus D} \frac{f(\underline{X}; \theta)}{f(\underline{X}; \underline{\theta}^0)} \equiv \prod_{(a,b] \setminus \{\cup_{i=1}^d (a_i, b_i]\}} \frac{f(\underline{X}; \theta)}{f(\underline{X}; \underline{\theta}^0)}$$

and

$$\prod_D \frac{f(\underline{X}; \hat{\theta})}{f(\underline{X}; \underline{\theta}^0)} \equiv \prod_{i=1}^d \left[\prod_{(a_i, b_i]} \frac{f(\underline{X}; \hat{\theta}_i)}{f(\underline{X}; \underline{\theta}^0)} \right],$$

where $\hat{\theta}_i$ is the m.l.e. of θ given observations $\{x_t, t \in (a_i, b_i]\}$.

- (3) Let $p(\underline{J}^{(k)}) = \{(0, J_1], (J_1, J_2], \dots, (J_{k-1}, J_k], (J_k, n]\}$ denote the partition of the interval $(0, n]$ induced by $\underline{J}^{(k)} = (J_1, \dots, J_k)$.
- (4) Let $A^* = \cup_{r=1}^{k_0} \{A_r^-, A_r^+\}$ be a collection of $2k_0$ disjoint intervals, where $A_r^- = (J_r^-, J_r^0]$ and $A_r^+ = (J_r^0, J_r^+]$ with integers $J_r^- = J_r^0 - [\alpha \log n]$ and $J_r^+ = J_r^0 + [\alpha \log n]$, $r = 1, \dots, k_0$. Here α is some large constant (cf. lemma 1) and $[\alpha \log n]$ is the largest integer not beyond $\alpha \log n$.
- (5) Let $E_k = \{\underline{J}^{(k)} = (J_1, \dots, J_k) \mid \forall r = 1, \dots, k_0, \exists i \text{ such that } J_i \in A_r^- \cup A_r^+\}$ and E_k^c denotes the complement of E_k .

3. Consistency

In this section, we prove the following theorem.

Theorem. Under assumptions (A1)-(A9), as $n \rightarrow \infty$, the posterior mode \hat{k} with respective to a prior distribution π satisfying (B1)-(B2) converges in probability to k_0 . Furthermore, $\|\hat{\underline{J}}^{(k_0)} - \underline{J}^0\| = O_p(\log n)$.

We need several lemmas to establish the theorem. Note that the right hand side of (1.2), divided by $f(\underline{X}; \underline{\theta}^0) = \prod_{i=1}^{k_0+1} [\prod_{r=J_{i-1}^0+1}^{J_i^0} f(x_r; \theta_i^0)]$, equals

$$\begin{aligned} \pi(k \mid \underline{X}^{(n)}) &\propto \sum_{\underline{J}^{(k)} \in E_k} \left[\prod_{(a,b) \in p(\underline{J}^{(k)})} \int \prod_{(a,b)} \frac{f(\underline{X}; \theta)}{f(\underline{X}; \underline{\theta}^0)} \pi'(\theta) d\theta \right] \pi(k, \underline{J}^{(k)}) \\ &+ \sum_{\underline{J}^{(k)} \in E_k^c} \left[\prod_{(a,b) \in p(\underline{J}^{(k)})} \int \prod_{(a,b)} \frac{f(\underline{X}; \theta)}{f(\underline{X}; \underline{\theta}^0)} \pi'(\theta) d\theta \right] \pi(k, \underline{J}^{(k)}). \end{aligned}$$

Lemma 1. $\forall B > 0$, for sufficiently large α ,

$$P \left\{ \int \prod_{(a,b)} \frac{f(\underline{X}; \theta)}{f(\underline{X}; \underline{\theta}^0)} \pi'(\theta) d\theta > n^{-B} \left[\int \prod_{(a,b) \setminus A_r^-} \frac{f(\underline{X}; \theta)}{f(\underline{X}; \underline{\theta}^0)} \pi'(\theta) d\theta + \int \prod_{(a,b) \setminus A_r^+} \frac{f(\underline{X}; \theta)}{f(\underline{X}; \underline{\theta}^0)} \pi'(\theta) d\theta \right], \right. \\ \left. \forall 0 \leq a < b \leq n \text{ with } A_r^- \cup A_r^+ \subset (a, b] \right\}$$

converges to 0 as $n \rightarrow \infty$.

Lemma 2. $\forall \varepsilon > 0$ and any k ,

$$P\{\pi(k, \underline{J}^{(k)} \in E_k^c \mid \underline{X}^{(n)}) > \varepsilon \pi(k_0, \underline{J}^0 \mid \underline{X}^{(n)})\} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

By assumption (A1), for large n , we have $E_k = \phi$ if $k < k_0$. It follows from lemma 2 that, for each $k < k_0$, $P\{\pi(k \mid \underline{X}^{(n)}) < \pi(k_0 \mid \underline{X}^{(n)})\}$ converges to 1 as $n \rightarrow \infty$. Hence $P(\hat{k} \geq k_0)$ converges to 1 as $n \rightarrow \infty$. Also, by lemma 2, as $n \rightarrow \infty$, $P\{\pi(k_0, \underline{J}^{(k_0)} \in E_{k_0}^c \mid \underline{X}^{(n)}) < \pi(k_0, \underline{J}^0 \mid \underline{X}^{(n)})\}$ converges to 1, so that $P(\hat{J}^{(k_0)} \in E_{k_0})$ converges to 1, proving the second part of the theorem. It remains to show that $P(\hat{k} > k_0)$ converges to 0 as $n \rightarrow \infty$, or equivalently $P(\hat{k} = k)$ converges to 0 for each $k_0 < k \leq R_0$.

Fix $k_0 < k \leq R_0$. By lemma 2, it suffices to show that for each $\varepsilon > 0$, $P\{\pi(k, \underline{J}^{(k)} \in E_k \mid \underline{X}^{(n)}) > \varepsilon \pi(k_0, \underline{J}^0 \mid \underline{X}^{(n)})\} \rightarrow 0$ as $n \rightarrow \infty$. For each $\underline{J}^{(k)} \in E_k$, and $i = 1, \dots, k_0$, let c_i be the component of $\underline{J}^{(k)}$ that is closest to J_i^0 . (In case that two components of $\underline{J}^{(k)}$ are both closest to J_i^0 , let c_i be the smaller one.) By the definition of E_k , we have $|c_i - J_i^0| \leq \alpha \log n$. Let $\underline{J}^{*(r_i)} = (J_{i1}^*, \dots, J_{ir_i}^*)$ denote the points of $\underline{J}^{(k)}$ in the interval $(J_{i-1}^0, J_i^0]$ (excluding c_1, \dots, c_{k_0}). Thus, we have $\underline{J}^{(k)} = (\underline{J}^{*(r_1)}, c_1, \underline{J}^{*(r_2)}, c_2, \dots, \underline{J}^{*(r_{k_0})}, c_{k_0}, \underline{J}^{*(r_{k_0+1})})$ and $\sum_{i=1}^{k_0+1} r_i = k - k_0$. Note that $\underline{J}^{*(r_i)} = \phi$ if $r_i = 0$. Denote by $\langle c_i, J_i^0 \rangle$ the interval $(c_i, J_i^0]$ if $c_i \leq J_i^0$, or the interval $(J_i^0, c_i]$ if $c_i > J_i^0$. Let $D(\underline{J}^{(k)}) = \{\langle c_i, J_i^0 \rangle \mid i = 1, \dots, k_0\}$, and let $I(\underline{J}^{*(r_i)})$ be the partition of $(\max(J_{i-1}^0, c_{i-1}), \min(J_i^0, c_i)]$ induced by the points in $\underline{J}^{*(r_i)}$ where $c_0 = 0$ and $c_{k_0+1} = n$. Clearly, $\bigcup_{i=1}^{k_0+1} I(\underline{J}^{*(r_i)}) \cup \{\langle c_i, J_i^0 \rangle \mid i = 1, \dots, k_0\}$ is the partition of $(0, n]$ induced by the points in $\underline{J}^{(k)}$ and \underline{J}^0 .

Now, for each $(a, b] \in p(\underline{J}^{(k)})$, $(a, b] \cap D(\underline{J}^{(k)})$ may be empty or $\langle c_i, J_i^0 \rangle$ or $\langle c_{i-1}, J_{i-1}^0 \rangle \cup \langle c_i, J_i^0 \rangle$ for some i , and $(a, b] \setminus D(\underline{J}^{(k)})$ is an interval (possibly empty). Since $\prod_{(a,b] \cap D(\underline{J}^{(k)})} \frac{f(\underline{X};\theta)}{f(\underline{X};\underline{\theta}^0)} \leq \prod_{(a,b] \cap D(\underline{J}^{(k)})} \frac{f(\underline{X};\hat{\theta})}{f(\underline{X};\underline{\theta}^0)}$, we have

$$\int \prod_{(a,b]} \frac{f(\underline{X};\theta)}{f(\underline{X};\underline{\theta}^0)} \pi'(\theta) d\theta \leq \left[\prod_{(a,b] \cap D(\underline{J}^{(k)})} \frac{f(\underline{X};\hat{\theta})}{f(\underline{X};\underline{\theta}^0)} \right] \left[\int \prod_{(a,b] \setminus D(\underline{J}^{(k)})} \frac{f(\underline{X};\theta)}{f(\underline{X};\underline{\theta}^0)} \pi'(\theta) d\theta \right].$$

It follows that

$$\begin{aligned} & \prod_{(a,b] \in p(\underline{J}^{(k)})} \int \prod_{(a,b]} \frac{f(\underline{X};\theta)}{f(\underline{X};\underline{\theta}^0)} \pi'(\theta) d\theta \\ & \leq \left[\prod_{D(\underline{J}^{(k)})} \frac{f(\underline{X};\hat{\theta})}{f(\underline{X};\underline{\theta}^0)} \right] \prod_{i=1}^{k_0+1} \left[\prod_{(c,d] \in I(\underline{J}^{*(r_i)})} \int \prod_{(c,d]} \frac{f(\underline{X};\theta)}{f(\underline{X};\underline{\theta}^0)} \pi'(\theta) d\theta \right] \\ & \leq \left[\prod_{i=1}^{k_0} [T_i(\underline{X})] \right] \prod_{i=1}^{k_0+1} \left[\prod_{(c,d] \in I(\underline{J}^{*(r_i)})} \int \prod_{(c,d]} \frac{f(\underline{X};\theta)}{f(\underline{X};\underline{\theta}^0)} \pi'(\theta) d\theta \right], \end{aligned} \tag{3.1}$$

where

$$T_i(\underline{X}) = \max \left\{ \max_{J_i^- \leq m < J_i^0} \sup_{\theta} \left[\prod_{(m, J_i^0]} \frac{f(\underline{X};\theta)}{f(\underline{X};\underline{\theta}^0)} \right], \max_{J_i^0 < m \leq J_i^+} \sup_{\theta} \left[\prod_{(J_i^0, m]} \frac{f(\underline{X};\theta)}{f(\underline{X};\underline{\theta}^0)} \right] \right\}.$$

We now deal with the term inside the second pair of brackets on the right-most side of (3.1). For each $I(\underline{J}^{*(r_i)})$, $i = 1, \dots, k_0 + 1$, consider the first interval (denoted $(a_{i1}, b_{i1}]$), and the final interval (denoted $(a_{i2}, b_{i2}]$). Note that $a_{i1} = \max(J_{i-1}^0, c_{i-1})$, $b_{i1} = J_{ir_i}^*$, $a_{i2} = J_{ir_i}^*$, $b_{i2} = \min(J_i^0, c_i)$ and that $(\sqrt{b_{i1} - a_{i1}})^{-1} \leq O(\sqrt{\log n})/\sqrt{J_{i1}^* - J_{i-1}^0}$ and $(\sqrt{b_{i2} - a_{i2}})^{-1} \leq O(\sqrt{\log n})/\sqrt{J_i^0 - J_{ir_i}^*}$. Thus

$$\int \prod_{(a_{i1}, b_{i1}]} \frac{f(\underline{X};\theta)}{f(\underline{X};\underline{\theta}^0)} \pi'(\theta) d\theta \leq O(\sqrt{\log n}) S_{i1}(\underline{X}) / \sqrt{J_{i1}^* - J_{i-1}^0} \tag{3.2}$$

and

$$\int \prod_{(a_{i2}, b_{i2}]} \frac{f(\underline{X};\theta)}{f(\underline{X};\underline{\theta}^0)} \pi'(\theta) d\theta \leq O(\sqrt{\log n}) S_{i2}(\underline{X}) / \sqrt{J_i^0 - J_{ir_i}^*}, \tag{3.3}$$

where $S_{i1}(\underline{X}) = \max\{1, \max_{J_{i-1}^0 \leq a \leq J_{i-1}^+} \max_{a < b \leq J_i^0} \sqrt{b - a} \int \prod_{(a,b]} \frac{f(\underline{X};\theta)}{f(\underline{X};\underline{\theta}^0)} \pi'(\theta) d\theta\}$ and $S_{i2}(\underline{X}) = \max\{1, \max_{J_i^- \leq b \leq J_i^0} \max_{J_{i-1}^0 \leq a < b} \sqrt{b - a} \int \prod_{(a,b]} \frac{f(\underline{X};\theta)}{f(\underline{X};\underline{\theta}^0)} \pi'(\theta) d\theta\}$. If $r_i = 0$, then $I(\underline{J}^{*(r_i)}) = \{(\max(J_{i-1}^0, c_{i-1}), \min(J_i^0, c_i))\} = \{(a_{i1}, b_{i1}]\} = \{(a_{i2}, b_{i2}]\}$, and $(\sqrt{b_{i1} - a_{i1}})^{-1} \leq O(1)/\sqrt{J_i^0 - J_{i-1}^0}$. Let $L(\underline{J}^{*(r_i)}) = \sqrt{J_i^0 - J_{i-1}^0}/$

$(\sqrt{J_{i1}^* - J_{i-1}^0} \sqrt{J_i^0 - J_{ir_i}^*})$ if $r_i \geq 1$, $= 1$ if $r_i = 0$. By (3.2) and (3.3), for $r_i > 0$,

$$\prod_{j=1}^2 \int \prod_{(a_{ij}, b_{ij}] } \frac{f(\underline{X}; \theta)}{f(\underline{X}; \underline{\theta}^0)} \pi'(\theta) d\theta \leq O(\log n) L(\underline{J}^{*(r_i)}) S_{i1}(\underline{X}) S_{i2}(\underline{X}) / \sqrt{J_i^0 - J_{i-1}^0}$$

and for $r_i = 0$,

$$\int \prod_{(a_{i1}, b_{i1}] } \frac{f(\underline{X}; \theta)}{f(\underline{X}; \underline{\theta}^0)} \pi'(\theta) d\theta \leq O(1) L(\underline{J}^{*(r_i)}) S_{i1}(\underline{X}) S_{i2}(\underline{X}) / \sqrt{J_i^0 - J_{i-1}^0}.$$

So,

$$\prod_{(c,d] \in I(\underline{J}^{*(r_i)})} \int \prod_{(c,d] } \frac{f(\underline{X}; \theta)}{f(\underline{X}; \underline{\theta}^0)} \pi'(\theta) d\theta \leq \left[O(\log n) S_{i1}(\underline{X}) S_{i2}(\underline{X}) / \sqrt{J_i^0 - J_{i-1}^0} \right] \cdot \left[L(\underline{J}^{*(r_i)}) \prod_{(c,d] \in I^0(\underline{J}^{*(r_i)})} \int \prod_{(c,d] } \frac{f(\underline{X}; \theta)}{f(\underline{X}; \underline{\theta}^0)} \pi'(\theta) d\theta \right], \quad (3.4)$$

where $I^0(\underline{J}^{*(r_i)}) = I(\underline{J}^{*(r_i)}) \setminus \{(a_{i1}, b_{i1}], (a_{i2}, b_{i2}]\}$ depends only on $\underline{J}^{*(r_i)}$. By (3.1) and (3.4), we get

$$\begin{aligned} & \sum_{\underline{J}^{(k)} \in E_k} \prod_{(a,b] \in p(\underline{J}^{(k)})} \int \prod_{(a,b] } \frac{f(\underline{X}; \theta)}{f(\underline{X}; \underline{\theta}^0)} \pi'(\theta) d\theta \cdot \pi(k, \underline{J}^{(k)}) \\ & \leq \left[\prod_{i=1}^{k_0} T_i(\underline{X}) \right] \left[\prod_{i=1}^{k_0+1} \frac{O(\log n)}{\sqrt{J_i^0 - J_{i-1}^0}} S_{i1}(\underline{X}) S_{i2}(\underline{X}) \right] \\ & \cdot \left[\sum_{\underline{C}} \sum_{\underline{J}_*^{(k-k_0)}} \left\{ \prod_{i=1}^{k_0+1} \left[L(\underline{J}^{*(r_i)}) \prod_{(c,d] \in I^0(\underline{J}^{*(r_i)})} \int \prod_{(c,d] } \frac{f(\underline{X}; \theta)}{f(\underline{X}; \underline{\theta}^0)} \pi'(\theta) d\theta \right] \right\} \pi(k, \underline{C}, \underline{J}_*^{(k-k_0)}) \right], \end{aligned}$$

where $\underline{C} = (c_1, \dots, c_{k_0})$ and $\underline{J}_*^{(k-k_0)} = (\underline{J}^{*(r_1)}, \dots, \underline{J}^{*(r_{k_0+1})})$. From (A.5) in Appendix A, we can find a constant $\varepsilon_0^* > 0$ such that

$$\prod_{(a,b] \in p(\underline{J}^0)} \int \prod_{(a,b] } \frac{f(\underline{X}; \theta)}{f(\underline{X}; \underline{\theta}^0)} \pi'(\theta) d\theta \cdot \pi(k_0, \underline{J}^0) \geq \varepsilon_0^* \left(\prod_{i=1}^{k_0+1} [J_i^0 - J_{i-1}^0]^{-\frac{1}{2}} \right) \pi(k_0, \underline{J}^0)$$

with probability approaching 1 as $n \rightarrow \infty$. It follows that

$$\begin{aligned} & \frac{\pi(k, \underline{J}^{(k)} \in E_k \mid \underline{X}^{(n)})}{\pi(k_0, \underline{J}^0 \mid \underline{X}^{(n)})} \leq O([\log n]^{k_0+1}) \left[\prod_{i=1}^{k_0} T_i(\underline{X}) \right] \left[\prod_{i=1}^{k_0+1} S_{i1}(\underline{X}) S_{i2}(\underline{X}) \right] \\ & \cdot \left[\sum_{\underline{C}} \sum_{\underline{J}_*^{(k-k_0)}} \left\{ \prod_{i=1}^{k_0+1} \left[L(\underline{J}^{*(r_i)}) \prod_{(c,d] \in I^0(\underline{J}^{*(r_i)})} \int \prod_{(c,d] } \frac{f(\underline{X}; \theta)}{f(\underline{X}; \underline{\theta}^0)} \pi'(\theta) d\theta \right] \right\} \frac{\pi(k, \underline{C}, \underline{J}_*^{(k-k_0)})}{\pi(k_0, \underline{J}^0)} \right]. \end{aligned} \quad (3.5)$$

Since $E[\int \prod_{(c,d]} \frac{f(\underline{X};\theta)}{f(\underline{X};\theta^0)} \pi'(\theta) d\theta] = 1$, it follows from Markov's inequality that the term in the last pair of brackets of (3.5) is $O_p(\sum_{\underline{C}} \sum_{\underline{J}^{(k-k_0)}} [\prod_{i=1}^{k_0+1} L(\underline{J}^{*(r_i)})] \frac{\pi(k, \underline{C}, \underline{J}^{(k-k_0)})}{\pi(k_0, \underline{J}^0)})$, which is $O_p(n^{-\varepsilon_0})$ for some $\varepsilon_0 > 0$ by Appendix B. By lemmas 3-5 below, $\prod_{i=1}^{k_0} T_i(\underline{X})$ and $\prod_{i=1}^{k_0+1} S_{i1}(\underline{X})S_{i2}(\underline{X})$ are $O_p([\log n]^{5(k_0+1)})$. Thus we conclude that, $\forall \varepsilon > 0$, the probability of $\{\pi(k, \underline{J}^{(k)} \in E_k \mid \underline{X}^{(n)}) < \varepsilon \pi(k_0, \underline{J}^0 \mid \underline{X}^{(n)})\}$ will converge to 1 as $n \rightarrow \infty$. The proof of the theorem is complete.

Lemma 3. *If $x_i, i = 1, \dots, n$, are i.i.d. random variables having density $f_{\theta_0}(\cdot)$ with respect to Lebesgue measure μ , then,*

$$\lim_{n \rightarrow \infty} \frac{L(\underline{X}; \hat{\theta}) - L(\underline{X}; \theta_0)}{\log \log n} = 1 \quad a.s.$$

where $L(\underline{X}; \theta_0) = \sum_{i=1}^n \log f(x_i; \theta_0)$ and $L(\underline{X}; \hat{\theta}) = \sum_{i=1}^n \log f(x_i; \hat{\theta})$. Here $\hat{\theta}$ is the m.l.e. of θ based on observation $\underline{X} = (x_1, \dots, x_n)$.

Lemma 4. *If $x_i, i = 1, \dots, n$, are i.i.d. random variables having density $f_{\theta_0}(\cdot)$ with respect to Lebesgue measure μ , then, $\forall \varepsilon > 0$*

$$\max_{1 \leq m \leq n} \left[\prod_{[1,m]} \frac{f(\underline{X}; \hat{\theta}_m)}{f(\underline{X}; \theta_0)} \right] = O_p([\log n]^{1+\varepsilon}),$$

where $\hat{\theta}_m$ is the m.l.e. of θ based on observations (x_1, \dots, x_m) .

By lemma 4, we have $\prod_{i=1}^{k_0} [T_i(\underline{X})] = O_p([\log n]^{k_0+\varepsilon})$ for any $\varepsilon > 0$.

Lemma 5. *If $x_i, i = 1, \dots, n$, are i.i.d. random variables having density $f_{\theta_0}(\cdot)$ with respect to Lebesgue measure μ , then*

$$\max_{0 \leq a \leq j^+} \max_{a < m \leq n} \left[\sqrt{m-a} \int \prod_{(a,m]} \frac{f(\underline{X};\theta)}{f(\underline{X};\theta_0)} \pi'(\theta) d\theta \right] = O_p([\log n]^{\frac{5}{2}}),$$

where $j^+ = [\alpha \log n]$.

By lemma 5, we have $\prod_{i=1}^{k_0+1} [S_{i1}(\underline{X}) \cdot S_{i2}(\underline{X})] = O_p([\log n]^{5(k_0+1)})$.

Remark. While we have established the result $\|\hat{\underline{J}}^{(k_0)} - \underline{J}^0\| = O_p(\log n)$, it is of interest to see if this can be pared down to $O_p(1)$.

4. Simulation

In this section, we use Monte Carlo simulation to study the behavior of the posterior mode \hat{k} which maximizes (1.2) and compare its precision with that given by Yao's method in (1.5) for a sequence of independent normal random variables with known common variance $\sigma^2 = 1$. We considered two cases, namely, $k_0 = 1$

(one change point) and $k_0 = 2$ (two change points) with known upper bound $R_0 = 3$. For $k_0 = 1$, we set $J_1^0 = n/2$ and $[n/4]$ with $\theta_1^0 = 0$ and $\theta_2^0 = 0, 1, 2, 3, 5, 7$ where $n = 30, 60, 100$, and for $k_0 = 2$, we set $(J_1^0, J_2^0) = ([n/3], [2n/3])$ with $(\theta_1^0, \theta_2^0, \theta_3^0) = (0, 1, 2), (0, 2, 4)$ and $(0, 3, 6)$, with sample size $n = 60, 100, 150$. We assumed a normal prior with mean μ_0 and variance σ_0^2 for the parameter θ , and selected $\sigma_0^2 = 16$ and $\mu_0 = \bar{X}$, which were used by Barry and Hartigan (1992), but σ^2 is set equal to 1. We also took $\pi(k, \underline{J}^{(k)}) = (1/4) \binom{n-1}{k}^{-1}$. Each case was simulated 1000 times and the results are listed in Table 1 and Table 2. In Table 1, it is evident that our method works very well and is better at detecting the number of change points than Yao's method, especially when the sample size is small ($n = 30$). In Table 2, we observe that Yao's method is better at detecting the two true change points except for the case when the change of means is small like $(0, 1, 2)$.

Table 1. Frequencies of the estimated number of change points

(1) Sample size $n = 30$ with a true changed position = 15 (7)

Mean change	The estimated number of change points							
	0		1		2		3	
0	951	(958)	042	(038)	007	(004)	000	(000)
	782	(761)	095	(112)	079	(079)	044	(048)
1	404	(552)	543	(404)	048	(038)	005	(006)
	239	(349)	512	(427)	140	(130)	109	(094)
2	004	(029)	861	(839)	109	(105)	026	(027)
	001	(014)	719	(716)	173	(150)	107	(120)
3	000	(000)	850	(854)	122	(123)	028	(023)
	000	(000)	730	(741)	160	(170)	109	(089)
5	000	(000)	861	(870)	116	(111)	023	(019)
	000	(000)	718	(737)	173	(153)	109	(110)
7	000	(000)	904	(885)	085	(100)	011	(015)
	000	(000)	728	(733)	159	(170)	113	(097)

(2) Sample size $n = 60$ with a true changed position = 30 (15)

Mean change	The estimated number of change points							
	0		1		2		3	
0	978	(964)	018	(034)	003	(002)	001	(000)
	884	(862)	070	(085)	037	(041)	009	(012)
1	121	(304)	817	(648)	057	(042)	005	(006)
	074	(191)	763	(652)	122	(122)	041	(035)
2	000	(000)	900	(918)	091	(071)	009	(011)
	000	(000)	838	(853)	108	(091)	054	(056)
3	000	(000)	908	(897)	081	(086)	011	(017)
	000	(000)	835	(819)	112	(124)	053	(057)
5	000	(000)	924	(917)	065	(075)	011	(008)
	000	(000)	858	(836)	098	(111)	044	(053)
7	000	(000)	936	(938)	057	(059)	007	(003)
	000	(000)	830	(866)	115	(089)	055	(045)

(3) Sample size $n = 100$ with a true changed position = 50 (25)

Mean change	The estimated number of change points							
	0		1		2		3	
0	977	(979)	021	(020)	002	(001)	000	(000)
	919	(914)	056	(054)	022	(026)	003	(006)
1	016	(083)	929	(868)	050	(043)	005	(006)
	010	(051)	860	(837)	097	(088)	033	(024)
2	000	(000)	928	(927)	068	(069)	004	(004)
	000	(000)	883	(899)	085	(080)	032	(021)
3	000	(000)	932	(925)	057	(066)	011	(009)
	000	(000)	898	(890)	074	(082)	028	(028)
5	000	(000)	931	(935)	066	(059)	003	(006)
	000	(000)	878	(883)	094	(089)	028	(028)
7	000	(000)	937	(946)	056	(051)	007	(003)
	000	(000)	864	(896)	105	(074)	031	(030)

The data are simulated 1000 times with means changing from $\theta_1^0 = 0$ to $\theta_2^0 = 0, 1, 2, 3, 5, 7$. The values in the upper and lower sides of each box are under the posterior mode and Yao's estimator, respectively.

Table 2. Frequencies of the estimated number of change points for the two true change points

Sample size	Change position	Mean change	The estimated number of change points			
			0	1	2	3
60	(20, 40)	(0,1,2)	000	526	420	054
			000	485	376	139
		(0,2,4)	000	000	781	219
			000	000	832	168
		(0,3,6)	000	000	759	241
			000	000	816	184
100	(33, 66)	(0,1,2)	000	188	738	074
			000	229	640	131
		(0,2,4)	000	000	819	181
			000	000	865	135
		(0,3,6)	000	000	833	167
			000	000	870	130
150	(50,100)	(0,1,2)	000	044	851	105
			000	071	808	121
		(0,2,4)	000	000	856	144
			000	000	896	104
		(0,3,6)	000	000	881	119
			000	000	919	081

The data are simulated 1000 times and the values in the upper and lower sides of each box are under the posterior mode and Yao's estimator, respectively.

Barry and Hartigan (1992) also compare Yao's (1988) method with the product partition model by using prior probability $g(i) = \frac{4}{i(i+1)(i+2)}$ of the jump variable and find that Yao's method is better at identifying the number of change

points. Especially, when the true model has only “one change point”, the detection will not be precise (cf. Table 1 of Barry and Hartigan (1992)). They also suggest that this defect may be overcome by considering different prior cohesions. For this purpose the uniform prior considered in our method may be used to substantially improve the precision of detecting the number of change points. However, the uniform prior does not meet the assumptions in the work of Barry and Hartigan (1992).

Acknowledgement

The author deeply thanks Professor Yi-Ching Yao for many valuable comments and suggestions which significantly improved the content and presentation of the paper.

Appendix A.

Proof of Lemma 1. Since $\theta_r^0 \neq \theta_{r+1}^0$, there exists a $\delta > 0$ such that the intersection of the neighborhood $N_\delta(\theta_r^0)$ of θ_r^0 and the neighborhood $N_\delta(\theta_{r+1}^0)$ of θ_{r+1}^0 is empty. Hence, for any a, b in $[0, n]$ and $A_r^- \cup A_r^+ \subset (a, b]$, we have, $\forall \underline{X}$ and θ ,

$$\begin{aligned} \prod_{(a,b]} \frac{f(\underline{X};\theta)}{f(\underline{X};\theta^0)} &\leq \left[\sup_{\theta \in N_\delta^c(\theta_r^0)} \prod_{A_r^-} \frac{f(\underline{X};\theta)}{f(\underline{X};\theta_r^0)} \right] \prod_{(a,b] \setminus A_r^-} \frac{f(\underline{X};\theta)}{f(\underline{X};\theta^0)} \\ &+ \left[\sup_{\theta \in N_\delta^c(\theta_{r+1}^0)} \prod_{A_r^+} \frac{f(\underline{X};\theta)}{f(\underline{X};\theta_{r+1}^0)} \right] \prod_{(a,b] \setminus A_r^+} \frac{f(\underline{X};\theta)}{f(\underline{X};\theta^0)}, \end{aligned} \tag{A.1}$$

where $N_\delta^c(\theta_r^0)$ and $N_\delta^c(\theta_{r+1}^0)$ are the complement of $N_\delta(\theta_r^0)$ and $N_\delta(\theta_{r+1}^0)$, respectively. Multiplying by $\pi'(\theta)$ and integrating out θ in (A.1), we obtain

$$\begin{aligned} \int \prod_{(a,b]} \frac{f(\underline{X};\theta)}{f(\underline{X};\theta^0)} \pi'(\theta) d\theta &\leq \left[\sup_{\theta \in N_\delta^c(\theta_r^0)} \prod_{A_r^-} \frac{f(\underline{X};\theta)}{f(\underline{X};\theta_r^0)} \right] \int \prod_{(a,b] \setminus A_r^-} \frac{f(\underline{X};\theta)}{f(\underline{X};\theta^0)} \pi'(\theta) d\theta \\ &+ \left[\sup_{\theta \in N_\delta^c(\theta_{r+1}^0)} \prod_{A_r^+} \frac{f(\underline{X};\theta)}{f(\underline{X};\theta_{r+1}^0)} \right] \int \prod_{(a,b] \setminus A_r^+} \frac{f(\underline{X};\theta)}{f(\underline{X};\theta^0)} \pi'(\theta) d\theta. \end{aligned} \tag{A.2}$$

The terms in the two pairs of brackets of (A.2) do not depend on a or b . By Theorem 1 of Wolfowitz (1949) or the argument given by Wald (1949), we can find a small positive value $h < 1$ such that

$$\sup_{\theta \in N_\delta^c(\theta_r^0)} \prod_{A_r^-} \frac{f(\underline{X};\theta)}{f(\underline{X};\theta_r^0)} \leq n^{\alpha \log h} \text{ and } \sup_{\theta \in N_\delta^c(\theta_{r+1}^0)} \prod_{A_r^+} \frac{f(\underline{X};\theta)}{f(\underline{X};\theta_{r+1}^0)} \leq n^{\alpha \log h} \tag{A.3}$$

with probability approaching 1 as $n \rightarrow \infty$. Hence, let α be sufficiently large to satisfy $\alpha \log h < -B$. Then, from (A.2) and (A.3), the lemma is obtained.

Proof of Lemma 2. Let $E_{r,k}^c$ be the set $\{\underline{J}^{(k)} = (J_1, \dots, J_k) \mid J_i \notin A_r^- \cup A_r^+, \forall i = 1, \dots, k\}$. Since $E_k^c = \bigcup_{r=1}^{k_0} E_{r,k}^c$ and k_0 is finite, it suffices to show that, $\forall \varepsilon > 0$, the probability of the set $\{\pi(k, \underline{J}^{(k)} \in E_{r,k}^c \mid \underline{X}^{(n)}) > \varepsilon \pi(k_0, \underline{J}^0 \mid \underline{X}^{(n)})\}$ will converge to 0 as $n \rightarrow \infty$. For any $\underline{J}^{(k)} = (J_1, \dots, J_k) \in E_{r,k}^c$, there exists i such that $A_r^- \cup A_r^+ \subset (J_{i-1}, J_i]$ and, by lemma 1, for any $B > 0$ and sufficiently large α , the probability of the event

$$\begin{aligned} & \prod_{(a,b) \in p(\underline{J}^{(k)})} \int \prod_{(a,b)} \frac{f(\underline{X}; \theta)}{f(\underline{X}; \underline{\theta}^0)} \pi'(\theta) d\theta \\ = & \left[\prod_{(a,b) \in p(\underline{J}^{(k)}) \setminus (J_{i-1}, J_i]} \int \prod_{(a,b)} \frac{f(\underline{X}; \theta)}{f(\underline{X}; \underline{\theta}^0)} \pi'(\theta) d\theta \right] \left[\int \prod_{(J_{i-1}, J_i]} \frac{f(\underline{X}; \theta)}{f(\underline{X}; \underline{\theta}^0)} \pi'(\theta) d\theta \right] \\ \leq & n^{-B-1} \prod_{(a,b) \in p(\underline{J}^{(k)}) \setminus (J_{i-1}, J_i]} \int \prod_{(a,b)} \frac{f(\underline{X}; \theta)}{f(\underline{X}; \underline{\theta}^0)} \pi'(\theta) d\theta \\ & \cdot \left[\int \prod_{(J_{i-1}, J_i] \setminus A_r^-} \frac{f(\underline{X}; \theta)}{f(\underline{X}; \underline{\theta}^0)} \pi'(\theta) d\theta + \int \prod_{(J_{i-1}, J_i] \setminus A_r^+} \frac{f(\underline{X}; \theta)}{f(\underline{X}; \underline{\theta}^0)} \pi'(\theta) d\theta \right] \end{aligned}$$

for all $\underline{J}^{(k)} \in E_{r,k}^c$ converges to 1 as $n \rightarrow \infty$. Noting that the expectation of the right hand side of the above inequality equals $2n^{-B-1}$, it follows from Markov's inequality that, for any $B > 0$ and sufficiently large α ,

$$\sum_{\underline{J}^{(k)} \in E_{r,k}^c} \left[\prod_{(a,b) \in p(\underline{J}^{(k)})} \int \prod_{(a,b)} \frac{f(\underline{X}; \theta)}{f(\underline{X}; \underline{\theta}^0)} \pi'(\theta) d\theta \right] \pi(k, \underline{J}^{(k)}) = o_p(n^{-B}). \tag{A.4}$$

Since $\prod_{(a,b) \in p(\underline{J}^0)} \int \prod_{(a,b)} \frac{f(\underline{X}; \theta)}{f(\underline{X}; \underline{\theta}^0)} \pi'(\theta) d\theta = \prod_{i=1}^{k_0+1} [\int \prod_{(J_{i-1}^0, J_i^0]} \frac{f(\underline{X}; \theta)}{f(\underline{X}; \underline{\theta}_i^0)} \pi'(\theta) d\theta]$ is greater than $\prod_{i=1}^{k_0+1} [\int \prod_{(J_{i-1}^0, J_i^0]} \frac{f(\underline{X}; \theta)}{f(\underline{X}; \hat{\theta}_i)} \pi'(\theta) d\theta]$ for $\hat{\theta}_i$ being the m.l.e. of θ based on the observations $\{x_t, t \in (J_{i-1}^0, J_i^0]\}$, from Johnson (1970) (page 857, (2.21)), we have, for each $i = 1, \dots, k_0 + 1$,

$$\lim_{n \rightarrow \infty} \sqrt{J_i^0 - J_{i-1}^0} \int \prod_{(J_{i-1}^0, J_i^0]} \frac{f(\underline{X}; \theta)}{f(\underline{X}; \hat{\theta}_i)} \pi'(\theta) d\theta = \sqrt{2\pi} \pi'(\theta_i^0) / I^{\frac{1}{2}}(\theta_i^0) \text{ in probability,}$$

where $I(\theta_i^0)$ is the Fisher information of a random variable x with density $f_{\theta_i^0}(\cdot)$. Thus we can find a constant $\varepsilon_0^* > 0$ such that

$$\prod_{(a,b) \in p(\underline{J}^0)} \int \prod_{(a,b)} \frac{f(\underline{X}; \theta)}{f(\underline{X}; \underline{\theta}^0)} \pi'(\theta) d\theta \cdot \pi(k_0, \underline{J}^0) \geq \varepsilon_0^* \left(\prod_{i=1}^{k_0+1} [J_i^0 - J_{i-1}^0]^{-\frac{1}{2}} \right) \pi(k_0, \underline{J}^0) \tag{A.5}$$

with probability approaching 1 as $n \rightarrow \infty$. By assumption (B1), (A.4) and (A.5), we establish the lemma by setting $B > R_0 + (R_0 + 1)/2$.

Proof of Lemma 3. Since $\hat{\theta}$ converges to θ_0 almost surely, for each $\delta > 0$, we can find sufficiently large n so that $\hat{\theta}$ is in the neighborhood $N_\delta(\theta_0)$ of θ_0 . By the Taylor expansion of $L(\underline{X}; \theta)$, and $\frac{\partial^2 L(\underline{X}; \hat{\theta})}{\partial \theta^2}$ at $\hat{\theta}$ and θ_0 , respectively, we obtain

$$\begin{aligned} & L(\underline{X}; \theta_0) - L(\underline{X}; \hat{\theta}) \\ &= \frac{1}{2!} \frac{\partial^2 L(\underline{X}; \hat{\theta})}{\partial \theta^2} (\theta_0 - \hat{\theta})^2 + \frac{1}{3!} \frac{\partial^3 L(\underline{X}; \theta_1^*)}{\partial \theta^3} (\theta_0 - \hat{\theta})^3 \\ &= \frac{1}{2!} \left[\frac{\partial^2 L(\underline{X}; \theta_0)}{\partial \theta^2} (\hat{\theta} - \theta_0)^2 + \frac{\partial^3 L(\underline{X}; \theta_2^*)}{\partial \theta^3} (\hat{\theta} - \theta_0)^3 \right] + \frac{1}{3!} \frac{\partial^3 L(\underline{X}; \theta_1^*)}{\partial \theta^3} (\theta_0 - \hat{\theta})^3 \\ &= \frac{1}{2!} \frac{\partial^2 L(\underline{X}; \theta_0)}{\partial \theta^2} (\hat{\theta} - \theta_0)^2 + \frac{1}{2!} \frac{\partial^3 L(\underline{X}; \theta_2^*)}{\partial \theta^3} (\hat{\theta} - \theta_0)^3 - \frac{1}{3!} \frac{\partial^3 L(\underline{X}; \theta_1^*)}{\partial \theta^3} (\hat{\theta} - \theta_0)^3, \end{aligned}$$

where θ_1^* and θ_2^* are between θ_0 and $\hat{\theta}$. Let

$$\bar{V}_n = -\frac{1}{n} \frac{\partial^2 L(\underline{X}; \theta_0)}{\partial \theta^2} = -\frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \theta^2} \log f(x_i; \theta_0)$$

and

$$\bar{D}_n = \frac{1}{2!} \frac{1}{n} \frac{\partial^3 L(\underline{X}; \theta_2^*)}{\partial \theta^3} (\hat{\theta} - \theta_0) - \frac{1}{3!} \frac{1}{n} \frac{\partial^3 L(\underline{X}; \theta_1^*)}{\partial \theta^3} (\hat{\theta} - \theta_0).$$

Then, as $n \rightarrow \infty$, \bar{V}_n converges to $I(\theta_0)$ (Fisher information of x_1) with probability 1 by the law of large numbers, and \bar{D}_n converges to 0 with probability 1 by the fact (assumption (A5)) that \bar{D}_n is dominated by $\bar{H}_n |\hat{\theta} - \theta_0|$ converging to 0 almost surely as $n \rightarrow \infty$. Here $\bar{H}_n = \frac{1}{n} \sum_{i=1}^n H(x_i)$ converges to some positive value almost surely. Hence, for large n ,

$$L(\underline{X}; \theta_0) - L(\underline{X}; \hat{\theta}) = -\frac{n}{2!} (\hat{\theta} - \theta_0)^2 \bar{V}_n (1 - 2\bar{D}_n / \bar{V}_n). \tag{A.6}$$

Also, by the Taylor expansion of $\frac{\partial L(\underline{X}; \hat{\theta})}{\partial \theta}$ at θ_0 , we obtain

$$\frac{\partial L(\underline{X}; \hat{\theta})}{\partial \theta} - \frac{\partial L(\underline{X}; \theta_0)}{\partial \theta} = \frac{\partial^2 L(\underline{X}; \theta_0)}{\partial \theta^2} (\hat{\theta} - \theta_0) + \frac{1}{2!} \cdot \frac{\partial^3 L(\underline{X}; \theta_3^*)}{\partial \theta^3} (\hat{\theta} - \theta_0)^2,$$

where θ_3^* is between θ_0 and $\hat{\theta}$. Let

$$\bar{S}_n = \frac{1}{n} \frac{\partial L(\underline{X}; \theta_0)}{\partial \theta} = \frac{1}{n} \sum_{i=1}^n \frac{\partial \log f(x_i; \theta_0)}{\partial \theta} \text{ and } \bar{E}_n = \frac{1}{2!n} \frac{\partial^3 L(\underline{X}; \theta_3^*)}{\partial \theta^3} (\hat{\theta} - \theta_0).$$

Then \bar{E}_n is also dominated by $\bar{H}_n |\hat{\theta} - \theta_0|$ and will converge to 0 almost surely as $n \rightarrow \infty$. Thus, for large n ,

$$\bar{S}_n = \bar{V}_n (\hat{\theta} - \theta_0) (1 - \bar{E}_n / \bar{V}_n). \tag{A.7}$$

From (A.6) and (A.7), we have, for large n ,

$$L(\underline{X}; \theta_0) - L(\underline{X}; \hat{\theta}) = -\frac{n \bar{S}_n^2}{2! \bar{V}_n} (1 - 2\bar{D}_n/\bar{V}_n)(1 - \bar{E}_n/\bar{V}_n)^2.$$

This implies that, for large n ,

$$L(\underline{X}; \hat{\theta}) - L(\underline{X}; \theta_0) = \frac{n \bar{S}_n^2}{2! I(\theta_0)} \frac{I(\theta_0)}{\bar{V}_n} \frac{(1 - 2\bar{D}_n/\bar{V}_n)}{(1 - \bar{E}_n/\bar{V}_n)^2}. \tag{A.8}$$

Since \bar{V}_n converges to $I(\theta_0)$ almost surely and \bar{D}_n and \bar{E}_n both converge to 0 almost surely as $n \rightarrow \infty$,

$$\lim_{n \rightarrow \infty} \frac{(1 - 2\bar{D}_n/\bar{V}_n)}{(1 - \bar{E}_n/\bar{V}_n)^2} = 1 \quad \text{a.s.}$$

Also, \bar{S}_n is the sample mean of random variables $\frac{\partial \log f(x_i; \theta_0)}{\partial \theta}$, $i = 1, \dots, n$, with mean 0 and variance $I(\theta_0)/n$. By the law of the iterated logarithm, we obtain,

$$\overline{\lim}_{n \rightarrow \infty} \frac{|n \bar{S}_n|}{\sqrt{2n(\log \log n)I(\theta_0)}} = 1 \quad \text{a.s.}$$

Therefore, it follows from (A.8) that

$$\overline{\lim}_{n \rightarrow \infty} \frac{L(\underline{X}; \hat{\theta}) - L(\underline{X}; \theta_0)}{\log \log n} = 1 \quad \text{a.s.}$$

Proof of Lemma 4. $\forall N$ and ε , the event $\{\max_{N \leq m \leq n} [\prod_{[1,m]} \frac{f(\underline{X}; \hat{\theta}_m)}{f(\underline{X}; \theta_0)}] > (\log n)^{1+\varepsilon}\}$ is contained in $\{\max_{N \leq m \leq n} \{[\log \prod_{[1,m]} \frac{f(\underline{X}; \hat{\theta}_m)}{f(\underline{X}; \theta_0)}] / [\log \log m]\} > 1 + \varepsilon\}$. By lemma 3, we establish the lemma.

Proof of Lemma 5. Consider the two cases $m \leq 2j^+$ and $m > 2j^+$. If $m \leq 2j^+$, by Markov's inequality and observing that $E[\int \prod_{(a,m]} \frac{f(\underline{X}; \theta)}{f(\underline{X}; \theta_0)} \pi'(\theta) d\theta] = 1$, we have $\max_{0 \leq a \leq j^+} \max_{a < m \leq 2j^+} \sqrt{m-a} \int \prod_{(a,m]} \frac{f(\underline{X}; \theta)}{f(\underline{X}; \theta_0)} \pi'(\theta) d\theta = O_p([\log n]^{\frac{5}{2}})$. If $m > 2j^+$, by $m - a \leq 2(m - j^+)$, we have

$$\begin{aligned} & \max_{0 \leq a \leq j^+} \max_{2j^+ < m \leq n} \sqrt{m-a} \int \prod_{(a,m]} \frac{f(\underline{X}; \theta)}{f(\underline{X}; \theta_0)} \pi'(\theta) d\theta \leq O(1) \left[\max_{0 \leq a \leq j^+} \prod_{(a,j^+]} \frac{f(\underline{X}; \hat{\theta}_a)}{f(\underline{X}; \theta_0)} \right] \\ & \cdot \left[\max_{2j^+ < m \leq n} \prod_{(j^+,m]} \frac{f(\underline{X}; \hat{\theta}_m)}{f(\underline{X}; \theta_0)} \right] \left[\max_{2j^+ < m \leq n} \sqrt{m-j^+} \int \prod_{(j^+,m]} \frac{f(\underline{X}; \theta)}{f(\underline{X}; \theta_m)} \pi'(\theta) d\theta \right], \end{aligned}$$

where $\hat{\theta}_a$ and $\hat{\theta}_m$ are the m.l.e.'s of θ given observations $\{x_t, t \in (a, j^+]\}$ and $\{x_t, t \in (j^+, m]\}$, respectively. By lemma 4, the terms in the first and second

pairs of brackets are both $O_p([\log n]^{1+\varepsilon}) \forall \varepsilon > 0$. By Johnson (1970) (page 857, (2.21)), we obtain the term in the last pair of brackets is $O_p(1)$. Thus we establish the lemma.

Appendix B.

1. For the uniform prior $\pi(k, \underline{J}^{(k)}) = O(n^{-k})$, using the fact that $\sum_{1 \leq J_1 < \dots < J_r < n} \sqrt{\frac{n \cdot n}{J_1(n-J_r)}} \frac{1}{n^r}$ converges as $n \rightarrow \infty$, we obtain, for each fixed r_i ,

$$\begin{aligned} \sum_{J_{i-1}^0 < J_{i1}^* < \dots < J_{ir_i}^* < J_i^0} L(\underline{J}^{*(r_i)}) &\leq O\left([J_i^0 - J_{i-1}^0]^{r_i - \frac{1}{2}}\right) = O\left(n^{r_i - \frac{1}{2}}\right) \quad \text{if } r_i \geq 1 \\ &= O(1) \quad \text{if } r_i = 0 \end{aligned}$$

so that

$$\begin{aligned} &\sum_{\underline{C}} \sum_{\underline{J}_*^{(k-k_0)}} \left[\prod_{i=1}^{k_0+1} L(\underline{J}^{*(r_i)}) \right] \frac{\pi(k, \underline{C}, \underline{J}_*^{(k-k_0)})}{\pi(k_0, \underline{J}^0)} \\ &\leq \left\{ \sum_{r_1 + \dots + r_{k_0+1} = k - k_0} \prod_{i=1}^{k_0+1} \left[\sum_{\underline{J}^{*(r_i)}} L(\underline{J}^{*(r_i)}) \right] \right\} O([\log n]^{k_0} / n^{k-k_0}) \\ &= O\left(\left\{ \prod_{i=1}^{k_0+1} [n]^{\max(r_i - \frac{1}{2}, 0)} \right\} \cdot [\log n]^{k_0} / n^{k-k_0}\right) \\ &= O\left(n^{k-k_0 - \frac{1}{2}} \cdot [\log n]^{k_0} / n^{k-k_0}\right) = O(n^{-\varepsilon_0}) \quad \text{for some } \varepsilon_0 > 0. \end{aligned}$$

2. From (3.5), the theorem still holds if we replace the assumption (B1) by the assumption (B1)' that the prior information $\pi(k, \underline{J}^{(k)})$ satisfies the condition

$$\sum_{\underline{C}} \sum_{\underline{J}_*^{(k-k_0)}} \left[\prod_{i=1}^{k_0+1} L(\underline{J}^{*(r_i)}) \right] \frac{\pi(k, \underline{C}, \underline{J}_*^{(k-k_0)})}{\pi(k_0, \underline{J}^0)} = O(n^{-\varepsilon_0}) \quad \text{for some } \varepsilon_0 > 0.$$

References

Barry, D. and Hartigan, J. A. (1992). Product partition models for change point problems. *Ann. Statist.* **20**, 260-279.
 Barry, D. and Hartigan, J. A. (1993). A Bayesian analysis for change point problems. *J. Amer. Statist. Assoc.* **88**, 309-319.
 Chernoff, H. and Zacks, S. (1964). Estimating the current mean of a normal distribution which is subject to changes in time. *Ann. Math. Statist.* **35**, 999-1018.
 Hartigan, J. A. (1983). *Bayes Theory*. Springer, New York.
 Hartigan, J. A. (1990). Partition models. *Comm. Statist. Theory Methods* **19**, 2745-2756.
 Johnson, R. A. (1970). Asymptotic expansions associated with posterior distributions. *Ann. Math. Statist.* **41**, 851-864.

- Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6**, 461-464.
- Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **20**, 595-601.
- Walker, A. M. (1969). On the asymptotic behaviour of posterior distributions. *J. Roy. Statist. Soc. Ser. B* **31**, 80-88.
- Wolfowitz, J. (1949). On Wald's proof of the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **20**, 601-602.
- Wolfowitz, J. (1965). Asymptotic efficiency of the maximum likelihood estimators. *Theory Probab. Appl.* **10**, 247-260.
- Yao, Y.-C. (1984). Estimation of a noisy Discrete-Time step function : Bayes and empirical Bayes approaches. *Ann. Statist.* **12**, 1434-1447.
- Yao, Y.-C. (1988). Estimating the number of change-points via Schwarz' criterion. *Statist. Probab. Lett.* **6**, 181-189.

Department of Applied Mathematics, National Chung-Hsing University, 40227 Taichung, Taiwan.

E-mail: cblee@mail.nchu.edu.tw

(Received June 1996; accepted August 1997)