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 $X = X^{(n)} = (x_1, \ldots, x_n)$ with the following distribution:

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

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

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

$$\pi(k \mid X^{(n)}) \propto \sum_{J^{(k)}} \left[ \int f(X^{(n)}; \theta^{(k+1)}, k, J^{(k)}) \pi(\theta^{(k+1)} \mid k, J^{(k)}) d\theta^{(k+1)} \right] \pi(k, J^{(k)}), \quad (1.2)$$
where \( f(\mathbf{X}^{(n)}; \theta^{(k+1)}, k, \mathcal{J}^{(k)}) \) denotes the joint density of \( x_1, \ldots, x_n \) given \( \theta^{(k+1)}, k, \mathcal{J}^{(k)} \). 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 \( \hat{\mathcal{J}}^{(k)} = (\hat{j}_1^{(k)}, \ldots, \hat{j}_k^{(k)}) \) maximize \( \pi(k, \mathcal{J}^{(k)} \mid \mathbf{X}^{(n)}) \) over \( \mathcal{J}^{(k)} \) with \( k \) fixed. We also show that \( \|\hat{\mathcal{J}}^{(k)} - \mathcal{J}^{0}\|^2 = O_p(\log n) \) where \( \mathcal{J}^{0} = (J_1^0, \ldots, J_{k_0}^0) \) denotes the true locations of the change points. Note that \( k_0 \) and \( \mathcal{J}^{0} \) are estimated by \( \hat{k} \) and \( \hat{\mathcal{J}}^{(\hat{k})} \), 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

\[
S(\mathbf{X}^{(n)}; k) \propto \int f(\mathbf{X}^{(n)}; \theta^{(k)})) \pi(\theta^{(k)}) d\theta^{(k)} \pi(k),
\]

where \( f(\mathbf{X}^{(n)}; \theta^{(k)}) \) is the joint density of \( \mathbf{X}^{(n)} = (x_1, \ldots, x_n) \) in an exponential family with \( k \)-dimensional parameter \( \theta^{(k)} \) 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

\[
SC(\mathbf{X}^{(n)}; k) = \log f(\mathbf{X}^{(n)}; \hat{\theta}^{(k)}) - k \log n / 2,
\]

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

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 \( \sigma^2 \). He finds that, under mild conditions, the estimator \( \hat{k} \) which maximizes

\[
SC(\mathbf{X}^{(n)}; k) = -n \log \hat{\sigma}^2 / 2 - k \log n
\]

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

Our method is to find \( k \) which maximizes the integrated likelihood function over \( \theta^{(k+1)} \) and \( \mathcal{J}^{(k)} \). While this idea in dealing with parameters \( \theta^{(k+1)} \) and \( \mathcal{J}^{(k)} \) 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

\[
SC^*(\mathbf{X}^{(n)}; k) \propto \max_{\theta^{(k+1)}, \mathcal{J}^{(k)}} \log f(\mathbf{X}^{(n)}; \theta^{(k+1)}, k, \mathcal{J}^{(k)}) - k \log n.
\]

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 Bay-
esian 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 \( X(n) = (x_1, \ldots, x_n) \) 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 \( J^0 = (J^0_1, \ldots, J^0_{k_0}) \) satisfy \( 0 < J^0_1 < J^0_2 < \cdots < J^0_{k_0} < n \) and \( \min_{1 \leq i \leq k_0+1} |(J^0_i - J^0_{i-1})/\log n| \to \infty \), as \( n \to \infty \),
where \( J^0_0 = 0 \) and \( J^0_{k_0+1} = n \).

(A2) Let \( \Theta \subset R \) be an open interval and \( \Theta \) the closure of \( \Theta \). \( \forall \theta \in \Theta \) and \( \theta' \in \Theta \),
provided \( \theta \neq \theta' \), \( f(x; \theta) \neq f(x; \theta') \) \( \forall \mu(x) > 0 \).

(A3) \( f(x; \theta) \) is jointly measurable in \( (x, \theta) \).

(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 \( \theta \).

(A5) Let \( \theta^0_i \) denote the true parameter value in the interval \( (J^0_{i-1}, J^0_i), i = 1, \ldots, k_0 + 1 \). 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 \( \theta \) in a neighborhood of \( \theta^0_i, i = 1, \ldots, k_0 + 1 \); and
\[
E_{\theta^0_i}[G_1(x)] < \infty, E_{\theta^0_i}[G_2(x)] < \infty \text{ and } E_{\theta^0_i}[H(x)] < \infty.
\]
(A6) The measures $\prod_{i=1}^{n} f(x_i; \theta)$ are mutually absolutely continuous for each $n = 1, 2, \ldots$. Therefore, a null set will have probability zero for all $\theta$.

(A7) $\lim_{|\theta| \to \infty} f(x; \theta) = 0$ a.e. w.r.t. Lebesgue measure.

(A8) $\forall \theta \in \Theta$, $E_\theta | \log f(x; \theta) | < \infty$ and $0 < I(\theta) = -E_\theta \left[ \frac{\partial^2}{\partial \theta^2} \log f(x; \theta) \right]$.

(A9) $\forall \theta \in \Theta, \rho, r > 0$, $f(x; \theta, \rho) = \sup_{|\rho' - \theta| \leq \rho} f(x; \theta')$ and $Q(x; r) = \sup_{|\theta| > r} f(x; \theta)$ are measurable functions of $x$, and for $i = 1, \ldots, k_0 + 1$, sufficiently small $\rho$ and sufficiently large $r$,

$$E_{\theta_0} [\log f(x; \theta, \rho)]^+ < \infty, \ E_{\theta_0} [\log Q(x; r)]^+ < \infty.$$  

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

(B1) The prior density $\pi(k) > 0$, for $k = 0, 1, \ldots, R_0$, and $\pi(J^{(k)} | k) = 1/(n-1)$ for each possible location vector $J^{(k)}$ (i.e. $\pi(J^{(k)} | k)$ is uniform).

(B2) Given $k$ and $J^{(k)}$, the conditional prior of $\theta^{(k+1)}$ is such that the $k + 1$ components of $\theta^{(k+1)}$ are independent with marginal probability density functions not depending on $k$ and $J^{(k)}$, which will be denoted by $\pi'$, i.e.

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

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

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 $\theta_0$ is the true value of $\theta$ for a random sample \(\{x_1, \ldots, x_n\}\), $\prod_{i=1}^{n} f(x_i; \theta_0)$ will be sufficiently small for all values of $\theta$ outside a neighborhood of the true parameter $\theta_0$, and (ii) the posterior distribution of $\sqrt{n}(\hat{\theta} - \theta)$ is asymptotically normal, where $\hat{\theta}$ is the m.l.e. of $\theta$. 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 $(x_1, \ldots, x_j)$ from $f(\cdot; \theta_1)$, and $(x_{j+1}, \ldots, x_n)$ 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 $\theta$; and if the random variables $(x_1, \ldots, x_n)$ have the same distribution, the second result implies that the integral $\int f(x; \theta) \pi(\theta) d\theta$ 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 $\log \prod_{i=1}^{n} f(x_i; \theta_0)$ 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 $\pi$ may be relaxed (cf. Appendix B).
Remark 3. Note that the dependence of \( J^0 \) on \( n \) has been suppressed. Strictly speaking, \( k_0 \) and \( \hat{\theta}^0 \) should depend on \( n \) as well. To avoid further assumptions and messy notation, we keep \( k_0 \) and \( \hat{\theta}^0 \) fixed as \( n \) increases. For the same reason, the prior \( \pi' \) on \( \theta \) is kept fixed as \( n \) increases.

### 2.2. Notation

For convenience, we define the following notation:

1. For any integers \( a \) and \( b \),
   \[
   \prod_{(a,b) \notin D} \frac{f(X|\theta)}{f(X|\theta^0)} = \begin{cases} 
   1, & \text{if } D \supset (a,b], \\
   \prod_{i=0}^{b-1} \frac{f(x_i|\theta)}{f(x_i|\theta^0)}, & \text{otherwise},
   \end{cases}
   \]
   where \( \delta(i) = r \) if \( J^0_{r-1} < i \leq J^0_r \), \( r = 1, \ldots, k_0 + 1 \). Note that \( (a,b] \) denotes the set of integers between \( a \) and \( b \) (including \( b \)).

2. If set \( D = \{(a_i, b_i) \mid i = 1, \ldots, d\} \) is a collection of disjoint intervals, then
   \[
   \prod_{(a,b) \notin D} \frac{f(X|\theta)}{f(X|\theta^0)} = \prod_{(a,b) \notin \bigcup_{i=1}^d (a_i, b_i)} \frac{f(X|\theta)}{f(X|\theta^0)}
   \]
   and
   \[
   \prod_{D} \frac{f(X|\theta)}{f(X|\theta^0)} = \prod_{i=1}^d \left[ \prod_{(a_i, b_i)} \frac{f(X|\theta)}{f(X|\theta^0)} \right],
   \]
   where \( \hat{\theta}_i \) is the m.l.e. of \( \theta \) given observations \( \{x_t, t \in (a_i, b_i]\} \).

3. Let \( p(J^{(k)}) = \{0, J_1, (J_1, J_2], \ldots, (J_{k-1}, J_k], (J_k, n]\} \) denote the partition of the interval \( [0, n] \) induced by \( J^{(k)} = (J_1, \ldots, J_k) \).

4. Let \( A^* = \bigcup_{r=1}^{k_0} \{A_r^-, A_r^+\} \) 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 \( J_r^- = J_r^0 - \lceil \alpha \log n \rceil \) and \( J_r^+ = J_r^0 + \lceil \alpha \log n \rceil \), \( r = 1, \ldots, k_0 \). Here \( \alpha \) is some large constant (cf. lemma 1) and \( \lceil \alpha \log n \rceil \) is the largest integer not beyond \( \alpha \log n \).

5. Let \( E_k = \{ J^{(k)} = (J_1, \ldots, J_k) \mid \forall r = 1, \ldots, k_0, \exists i \text{ such that } J_i \in A_r^- \cup A_r^+ \} \)
   and \( E_k \) denotes the complement of \( E_k \).

### 3. Consistency

In this section, we prove the following theorem.

**Theorem.** Under assumptions (A1)-(A9), as \( n \to \infty \), the posterior mode \( \hat{k} \) with respective to a prior distribution \( \pi \) satisfying (B1)-(B2) converges in probability to \( k_0 \). Furthermore, \( \| \hat{J}^{(k_0)} - J^0 \| = O_p(\log n) \).
We need several lemmas to establish the theorem. Note that the right hand side of (1.2), divided by \( f(\mathbf{X}^0) = \prod_{i=1}^{k_0+1} \prod_{r=J_{i-1}^0+1}^{J_i^0} f(x_r; \theta_i^0) \), equals

\[
\pi(k \mid \mathbf{X}^{(n)}) \propto \sum_{J(k) \in E_k} \left[ \prod_{(a,b) \in p(J(k))} \int \prod_{(a,b)} \frac{f(\mathbf{X}; \theta)}{f(\mathbf{X}; d\theta)} \pi'(\theta) \pi(k; J(k)) \right] + \sum_{J(k) \in E_k} \left[ \prod_{(a,b) \in p(J(k))} \int \prod_{(a,b)} \frac{f(\mathbf{X}; \theta)}{f(\mathbf{X}; d\theta)} \pi'(\theta) \pi(k; J(k)) \right].
\]

**Lemma 1.** \( \forall B > 0 \), for sufficiently large \( \alpha \),

\[
P\left\{ \int \prod_{(a,b)} f(\mathbf{X}; \theta) \pi'(\theta) d\theta > n^{-B} \left[ \int \prod_{(a,b) \mid A^-} f(\mathbf{X}; \theta) \pi'(\theta) d\theta + \int \prod_{(a,b) \mid A^+} f(\mathbf{X}; \theta) \pi'(\theta) d\theta \right] \right\},
\]

\( \forall 0 \leq a < b \leq n \) with \( A^- \cup A^+ \subset (a, b) \)

converges to 0 as \( n \to \infty \).

**Lemma 2.** \( \forall \epsilon > 0 \) and any \( k \),

\[
P\{ \pi(k; J(k) \in E_k \mid \mathbf{X}^{(n)}) > \epsilon \pi(k_0; J^0 \mid \mathbf{X}^{(n)}) \} \to 0 \quad \text{as} \quad n \to \infty.
\]

By assumption (A1), for large \( n \), we have \( E_k = \emptyset \) if \( k < k_0 \). It follows from lemma 2 that, for each \( k < k_0 \), \( P\{ \pi(k \mid \mathbf{X}^{(n)}) < \pi(k_0 \mid \mathbf{X}^{(n)}) \} \) converges to 0 as \( n \to \infty \). Hence \( P\{ k \leq k_0 \} \) converges to 1 as \( n \to \infty \). Also, by lemma 2, as \( n \to \infty \), \( P\{ \pi(k_0; J(k_0) \in E_k \mid \mathbf{X}^{(n)}) < \pi(k_0; J^0 \mid \mathbf{X}^{(n)}) \} \) converges to 1, so that \( P(J_0) = E_k \) converges to 1, proving the second part of the theorem. It remains to show that \( P(k > k_0) \) converges to 0 as \( n \to \infty \), or equivalently \( P(k = k) \) 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 \( \epsilon > 0 \),

\[
P\{ \pi(k; J(k) \in E_k \mid \mathbf{X}^{(n)}) > \epsilon \pi(k_0; J^0 \mid \mathbf{X}^{(n)}) \} \to 0 \quad \text{as} \quad n \to \infty.
\]

By the definition of \( E_k \), we have \( |c_i - J_i^0| \leq \alpha \log n \). Let \( J^{(i)}(r_i) = (J_{i1}, \ldots, J_{ir_i}) \) denote the points of \( J^{(k)} \) in the interval \( (J_{i-1}^0, J_i^0) \) (excluding \( c_i, c_{k_{i-1}} \)). Thus, we have \( J^{(k)}(c_1, c_2, \ldots, c_{k_{i-1}}, c_{k_{i-1}+1}) \) and \( \sum_{i=1}^{k_0+1} r_i = k - k_0 \). Note that \( J^{(i)}(c_i) = \emptyset \) if \( r_i = 0 \). Denote by \( \langle c_i, J_i^0 \rangle > 0 \) the interval \( (c_i, J_i^0) \) if \( c_i < J_i^0 \), or the interval \( (J_i^0, c_i) \) if \( c_i > J_i^0 \). Let \( D(J(k)) = \{ \langle c_i, J_i^0 \rangle > 0 \mid i = 1, \ldots, k_0 \} \), and let \( I(J^{(i)}(c_i)) \) be the partition of \( (\max(J_{i-1}^0, c_{i-1}), \min(J_i^0, c_i)) \) induced by the points in \( J_{i-1}^0 \) where \( c_0 = 0 \) and \( c_{k_0+1} = n \). Clearly, \( \bigcup_{i=1}^{k_0+1} I(J^{(i)}(c_i)) \cup \{ \langle c_i, J_i^0 \rangle > 0 \mid i = 1, \ldots, k_0 \} \) is the partition of \( (0, n) \) induced by the points in \( J(k) \) and \( J^0 \).
Now, for each \((a, b) \in p(J^{(k)})\), \((a, b) \cap D(J^{(k)})\) may be empty or \(< c_1, J^0_i >\) or \(< c_{i-1}, J^0_i >\) or for some \(i\), and \((a, b) \cap D(J^{(k)})\) is an interval (possibly empty). Since \(\prod_{(a, b) \cap D(J^{(k)})} \frac{f(X^i \theta)}{f(X^{(k)} \theta)} \leq \prod_{(a, b) \cap D(J^{(k)})} \frac{f(X^i \theta)}{f(X^{(k)} \theta)},\) we have

\[
\int \prod_{(a, b) \in p(J^{(k)})} \frac{f(X^i \theta)}{f(X^{(k)} \theta)} \pi'(\theta) d\theta \leq \left[ \prod_{(a, b) \cap D(J^{(k)})} \frac{f(X^i \theta)}{f(X^{(k)} \theta)} \right] \int \prod_{(a, b) \cap D(J^{(k)})} \frac{f(X^i \theta)}{f(X^{(k)} \theta)} \pi'(\theta) d\theta.
\]

It follows that

\[
\prod_{(a, b) \in p(J^{(k)})} \int \prod_{(a, b)} \frac{f(X^i \theta)}{f(X^{(k)} \theta)} \pi'(\theta) d\theta \\
\leq \left[ \prod_{D(J^{(k)})} \frac{f(X^i \theta)}{f(X^{(k)} \theta)} \right]_{k_0}^{k_0 + 1} \left[ \prod_{(c, d) \in I(J^{(r_i)})} \int \prod_{(c, d)} \frac{f(X^i \theta)}{f(X^{(k)} \theta)} \pi'(\theta) d\theta \right] \\
\leq \left[ \prod_{i = 1}^{k_0 - 1} [T_i(X)] \right]_{k_0}^{k_0 + 1} \prod_{(c, d) \in I(J^{(r_i)})} \int \prod_{(c, d)} \frac{f(X^i \theta)}{f(X^{(k)} \theta)} \pi'(\theta) d\theta, \tag{3.1}
\]

where

\[
T_i(X) = \max \left\{ \max_{J_i^0 < m < J_i^*} \sup_{\theta} \left[ \prod_{(m, p_i^0)} \frac{f(X^i \theta)}{f(X^{(k)} \theta)} \right], \max_{J_i^0 < m < J_i^*} \sup_{\theta} \left[ \prod_{(J_i^0, m)} \frac{f(X^i \theta)}{f(X^{(k)} \theta)} \right] \right\}.
\]

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

\[
\int \prod_{(a_{i1}, b_{i1})} \frac{f(X^i \theta)}{f(X^{(k)} \theta)} \pi'(\theta) d\theta \leq O(\sqrt{\log n})S_{11}(X)/\sqrt{J_{i1}^* - J_{i-1}^*} \tag{3.2}
\]

and

\[
\int \prod_{(a_{i2}, b_{i2})} \frac{f(X^i \theta)}{f(X^{(k)} \theta)} \pi'(\theta) d\theta \leq O(\sqrt{\log n})S_{22}(X)/\sqrt{J_{i1}^* - J_{i1}^*}, \tag{3.3}
\]

where \(S_{11}(X) = \max\{1, \max_{J_{i-1}^0 \leq m \leq J_{i1}^*} \max_{\sqrt{b_{i1} - a_{i1}}} \sup_{\theta} \prod_{(a, b)} \frac{f(X^i \theta)}{f(X^{(k)} \theta)} \pi'(\theta) d\theta\}\) and \(S_{22}(X) = \max\{1, \max_{J_{i-1}^0 \leq m \leq J_{i1}^*} \max_{\sqrt{b_{i2} - a_{i2}}} \sup_{\theta} \prod_{(a, b)} \frac{f(X^i \theta)}{f(X^{(k)} \theta)} \pi'(\theta) d\theta\}\).

If \(r_i = 0\), then \(I(J^{(r_i)}) = \{(\max(J_{i-1}^0, c_{i-1}), \min(J_i^0, c_i)\} \equiv \{(a_{i1}, b_{i1})\} = \{(a_{i2}, b_{i2})\}\) and \((\sqrt{b_{i1} - a_{i1}})^{-1} \leq O(1)/\sqrt{J_i^0 - J_{i-1}^0}\). Let \(L(J^{(r_i)}) = \sqrt{J_i^0 - J_{i-1}^0}/\sqrt{\log n}\).
\[\left(\sqrt{J_{1i}^* - J_{i-1}^0} - \sqrt{J_i^0 - J_{ri}}\right)\] if \( r_i \geq 1 \), = 1 if \( r_i = 0 \). By (3.2) and (3.3), for \( r_i > 0 
abla\]
\[
\prod_{j=1}^{2} \prod_{(a_i, b_j)} f(X^{(j)}) \pi'(\theta) \log n^2 S_{i1}(X) S_{i2}(X) / \sqrt{J_i^0 - J_i}.
\]
and for \( r_i = 0 \),
\[
\prod_{(a_i, b_1)} f(X^{(j)}) \pi'(\theta) \log n^2 S_{i1}(X) S_{i2}(X) / \sqrt{J_i^0 - J_i}.
\]
So,
\[
\prod_{(c, d) \in I^{(r_i)}} \prod_{(c, d) \in I^{(r_i)}} f(X^{(j)}) \pi'(\theta) \log n^2 S_{i1}(X) S_{i2}(X) / \sqrt{J_i^0 - J_i}.
\]
and for \( r_i = 0 \),
\[
\prod_{(c, d) \in I^{(r_i)}} f(X^{(j)}) \pi'(\theta) \log n^2 S_{i1}(X) S_{i2}(X) / \sqrt{J_i^0 - J_i}.
\]
where \( I^{(r_i)} = I^{(r_i)} \setminus \{(a_1, b_1), (a_2, b_2)\} \) depends only on \( J^{(r_i)} \). By (3.1) and (3.4), we get
\[
\sum_{\sum_{i=1}^{k_0} T_i(X)} \prod_{(a, b) \in \pi(\theta)} f(X^{(j)}) \pi'(\theta) \log n^2 S_{i1}(X) S_{i2}(X)
\]
\[
\sum_{C} \sum_{J^{(k-k_0)}} \left[ \prod_{i=1}^{k_0+1} \prod_{(c, d) \in I^{(r_i)}} f(X^{(j)}) \pi'(\theta) \log n^2 S_{i1}(X) S_{i2}(X) \right]
\]
where \( C = (c_1, \ldots, c_{k_0}) \) and \( J^{(k-k_0)} = (J^{(r_1)}, \ldots, J^{(r_{k_0+1})}) \). From (A.5) in Appendix A, we can find a constant \( \varepsilon_0 > 0 \) such that
\[
\prod_{(a, b) \in \pi(\theta)} f(X^{(j)}) \pi'(\theta) \log n^2 S_{i1}(X) S_{i2}(X)
\]
with probability approaching 1 as \( n \to \infty \). It follows that
\[
\frac{\pi(k, J^{(k)} \in E_k \mid X^{(n)})}{\pi(k_0, J_0^0 \mid X^{(n)})} \leq O\left(\log n \right) \prod_{i=1}^{k_0+1} T_i(X) \prod_{i=1}^{k_0+1} S_{i1}(X) S_{i2}(X)
\]
\[
\sum_{C} \sum_{J^{(k-k_0)}} \left[ \prod_{i=1}^{k_0+1} \prod_{(c, d) \in I^{(r_i)}} f(X^{(j)}) \pi'(\theta) \log n^2 S_{i1}(X) S_{i2}(X) \right]
\]
(3.5)
Since \( E[\int_{c,d} f(X|\theta) \pi'(\theta) d\theta] = 1 \), it follows from Markov’s inequality that the term in the last pair of brackets of (3.5) is \( O_p(\sum \sum \pi^{(k-k_0)} [\prod i=1^{k_0+1} L_j^{(r_i)}] \pi^{(k;C)\pi^{(k-k_0)}}) \), which is \( O_p(n^{-\varepsilon_0}) \) for some \( \varepsilon_0 > 0 \) by Appendix B. By lemmas 3-5 below, \( \prod i=1^{k_0} T_i(\hat{X}) \) and \( \prod i=1^{k_0+1} S_1(\hat{X})S_2(\hat{X}) \) are \( O_p([\log n]^{5(k_0+1)}) \). Thus we conclude that, \( \forall \varepsilon > 0 \), the probability of \( \{\pi(k, J^{(k)}) \in E_k | X^{(n)} < \varepsilon \pi(k_0, J^{(0)}) \} \) will converge to 1 as \( n \to \infty \). The proof of the theorem is complete.

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

\[
\lim_{n \to \infty} \frac{L(X; \hat{\theta}) - L(X; \theta_0)}{\log \log n} = 1 \quad a.s.
\]

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

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

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

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

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

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

\[
\max_{0 \leq a \leq j^+} \max_{a < m \leq n} \left[ \sqrt{m-a} \int_{(a,m)} \frac{f(X|\theta_0)}{f(X|\hat{\theta}_0)} \pi' (\theta) d\theta \right] = O_p([\log n]^\frac{5}{2}),
\]

where \( j^+ = [a \log n] \).

By lemma 5, we have \( \prod i=1^{k_0+1}[S_1(\hat{X}) \cdot S_2(\hat{X})] = O_p([\log n]^{5(k_0+1)}) \).

**Remark.** While we have established the result \( ||\hat{X}^{(k_0)} - X^0|| = O_p(\log n) \), 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 \( k \) 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 \( \theta_2^0 = 1, 2, 3, 5, 7 \) 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 \( \mu_0 \) and variance \( \sigma_0^2 \) for the parameter \( \theta \), and selected \( \sigma_0^2 = 16 \) and \( \mu_0 = \bar{X} \), which were used by Barry and Hartigan (1992), but \( \sigma^2 \) is set equal to 1. We also took \( \pi(k, J^{(k)}) = (1/4)^{(n-1)} \). 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

<table>
<thead>
<tr>
<th>Mean change</th>
<th>( 0 )</th>
<th>( 1 )</th>
<th>( 2 )</th>
<th>( 3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>951 (958)</td>
<td>042 (038)</td>
<td>007 (004)</td>
<td>000 (000)</td>
</tr>
<tr>
<td></td>
<td>782 (761)</td>
<td>095 (112)</td>
<td>079 (079)</td>
<td>044 (048)</td>
</tr>
<tr>
<td>1</td>
<td>404 (552)</td>
<td>543 (404)</td>
<td>048 (038)</td>
<td>005 (006)</td>
</tr>
<tr>
<td></td>
<td>239 (349)</td>
<td>512 (427)</td>
<td>140 (130)</td>
<td>109 (094)</td>
</tr>
<tr>
<td>2</td>
<td>004 (029)</td>
<td>861 (839)</td>
<td>109 (105)</td>
<td>026 (027)</td>
</tr>
<tr>
<td></td>
<td>001 (014)</td>
<td>719 (716)</td>
<td>173 (150)</td>
<td>107 (120)</td>
</tr>
<tr>
<td>3</td>
<td>000 (000)</td>
<td>850 (854)</td>
<td>122 (123)</td>
<td>028 (025)</td>
</tr>
<tr>
<td></td>
<td>000 (000)</td>
<td>730 (741)</td>
<td>160 (170)</td>
<td>109 (089)</td>
</tr>
<tr>
<td>5</td>
<td>000 (000)</td>
<td>861 (870)</td>
<td>116 (111)</td>
<td>023 (019)</td>
</tr>
<tr>
<td></td>
<td>000 (000)</td>
<td>718 (737)</td>
<td>173 (154)</td>
<td>109 (110)</td>
</tr>
<tr>
<td>7</td>
<td>000 (000)</td>
<td>904 (885)</td>
<td>085 (100)</td>
<td>011 (015)</td>
</tr>
<tr>
<td></td>
<td>000 (000)</td>
<td>728 (733)</td>
<td>159 (170)</td>
<td>113 (097)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mean change</th>
<th>( 0 )</th>
<th>( 1 )</th>
<th>( 2 )</th>
<th>( 3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>978 (964)</td>
<td>018 (034)</td>
<td>003 (002)</td>
<td>001 (000)</td>
</tr>
<tr>
<td></td>
<td>884 (862)</td>
<td>070 (085)</td>
<td>037 (041)</td>
<td>009 (012)</td>
</tr>
<tr>
<td>1</td>
<td>121 (304)</td>
<td>817 (648)</td>
<td>057 (042)</td>
<td>005 (006)</td>
</tr>
<tr>
<td></td>
<td>074 (191)</td>
<td>763 (652)</td>
<td>122 (122)</td>
<td>041 (035)</td>
</tr>
<tr>
<td>2</td>
<td>000 (000)</td>
<td>900 (918)</td>
<td>091 (071)</td>
<td>009 (011)</td>
</tr>
<tr>
<td></td>
<td>000 (000)</td>
<td>838 (853)</td>
<td>108 (091)</td>
<td>054 (056)</td>
</tr>
<tr>
<td>3</td>
<td>000 (000)</td>
<td>908 (897)</td>
<td>081 (086)</td>
<td>011 (017)</td>
</tr>
<tr>
<td></td>
<td>000 (000)</td>
<td>835 (819)</td>
<td>112 (124)</td>
<td>053 (057)</td>
</tr>
<tr>
<td>5</td>
<td>000 (000)</td>
<td>924 (917)</td>
<td>065 (075)</td>
<td>011 (008)</td>
</tr>
<tr>
<td></td>
<td>000 (000)</td>
<td>858 (836)</td>
<td>098 (111)</td>
<td>044 (055)</td>
</tr>
<tr>
<td>7</td>
<td>000 (000)</td>
<td>936 (938)</td>
<td>057 (059)</td>
<td>007 (003)</td>
</tr>
<tr>
<td></td>
<td>000 (000)</td>
<td>830 (866)</td>
<td>115 (089)</td>
<td>055 (045)</td>
</tr>
</tbody>
</table>
Barry and Hartigan (1992) also compare Yao’s (1988) method with the product partition model by using prior probability \( g(i) = \frac{4}{(i+1)(i+2)} \) of the jump variable and find that Yao’s method is better at identifying the number of change points.
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 $\theta_0^0 \neq \theta_0^{0+1}$, there exists a $\delta > 0$ such that the intersection of the neighborhood $N_\delta(\theta_0^0)$ of $\theta_0^0$ and the neighborhood $N_\delta(\theta_0^{0+1})$ of $\theta_0^{0+1}$ is empty. Hence, for any $a, b$ in $[0, n]$ and $A_\neg \subset (a, b]$, we have, \( \forall X \) and $\theta$,

\[
\prod_{a,b} \frac{f(x; \theta)}{f(x; \theta')} \leq \left[ \sup_{\theta \in N_\delta(\theta_0^0)} \prod_{A_\neg} \frac{f(x; \theta)}{f(x; \theta')} \right] \prod_{a,b \setminus A_\neg} \frac{f(x; \theta)}{f(x; \theta')} + \left[ \sup_{\theta \in N_\delta(\theta_0^{0+1})} \prod_{A_\neg} \frac{f(x; \theta)}{f(x; \theta')} \right] \prod_{a,b \setminus A_\neg} \frac{f(x; \theta)}{f(x; \theta')}, \tag{A.1}
\]

where $N_\delta(\theta_0^0)$ and $N_\delta(\theta_0^{0+1})$ are the complement of $N_\delta(\theta_0^0)$ and $N_\delta(\theta_0^{0+1})$, respectively. Multiplying by $\pi'(\theta)$ and integrating out $\theta$ in (A.1), we obtain

\[
\int \prod_{a,b} \frac{f(x; \theta)}{f(x; \theta')} \pi'(\theta) \, d\theta \leq \left[ \sup_{\theta \in N_\delta(\theta_0^0)} \prod_{A_\neg} \frac{f(x; \theta)}{f(x; \theta')} \right] \int \prod_{a,b \setminus A_\neg} \frac{f(x; \theta)}{f(x; \theta')} \pi'(\theta) \, d\theta + \left[ \sup_{\theta \in N_\delta(\theta_0^{0+1})} \prod_{A_\neg} \frac{f(x; \theta)}{f(x; \theta')} \right] \int \prod_{a,b \setminus A_\neg} \frac{f(x; \theta)}{f(x; \theta')} \pi'(\theta) \, d\theta. \tag{A.2}
\]

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

\[
\sup_{\theta \in N_\delta(\theta_0^0)} \prod_{A_\neg} \frac{f(x; \theta)}{f(x; \theta')} \leq n^{\alpha \log h} \quad \text{and} \quad \sup_{\theta \in N_\delta(\theta_0^{0+1})} \prod_{A_\neg} \frac{f(x; \theta)}{f(x; \theta')} \leq n^{\alpha \log h} \tag{A.3}
\]

with probability approaching 1 as $n \to \infty$. Hence, let $\alpha$ 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 \{\(J^{(k)} = (J_1, \ldots, J_k) \mid J_i \notin A^-_r \cup A^+_r, \forall i = 1, \ldots, k\)\}. Since $E_{k_0}^c = \bigcup_{r=1}^{k_0} E_{r,k}^c$ and $k_0$ is finite, it suffices to show that, for any $\varepsilon > 0$, the probability of the set \(\{\pi(k, J^{(k)} \in E_{r,k}^c \mid X^{(n)} > \varepsilon \pi(k_0, J^{(0)} \mid X^{(n)})\}\) will converge to 0 as $n \to \infty$. For any $J^{(k)} = (J_1, \ldots, J_k) \in E_{r,k}^c$, 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 $\alpha$, the probability of the event

\[
\prod_{(a,b) \in p(J^{(k)})} \int \prod_{(a,b)} \frac{f(x,d)}{f(x,d')} \pi'(\theta)d\theta
\]

\[
= \left[ \prod_{(a,b) \in p(J^{(k)}) \setminus (J_{i-1}, J_i]} \int \prod_{(a,b)} \frac{f(x,d)}{f(x,d')} \pi'(\theta)d\theta \right] \left[ \prod_{(J_{i-1}, J_i]} \frac{f(x,d)}{f(x,d')} \pi'(\theta)d\theta \right]
\]

\[
\leq n^{-B-1} \prod_{(a,b) \in p(J^{(k)}) \setminus (J_{i-1}, J_i]} \int \prod_{(a,b)} \frac{f(x,d)}{f(x,d')} \pi'(\theta)d\theta
\]

\[
\cdot \left[ \int \prod_{(J_{i-1}, J_i]} \frac{f(x,d)}{f(x,d')} \pi'(\theta)d\theta + \int \prod_{(J_{i-1}, J_i]} \frac{f(x,d)}{f(x,d')} \pi'(\theta)d\theta \right]
\]

for all $J^{(k)} \in E_{r,k}^c$ converges to 1 as $n \to \infty$. 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 $\alpha$,\n
\[
\sum_{J^{(k)} \in E_{r,k}^c} \left[ \prod_{(a,b) \in p(J^{(k)})} \int \prod_{(a,b)} \frac{f(x,d)}{f(x,d')} \pi'(\theta)d\theta \right] \pi(k, J^{(k)}) = o_p(n^{-B}). \quad (A.4)
\]

Since $\Pi_{(a,b) \in p(J^{(k)})} \int \prod_{(a,b)} \frac{f(x,d)}{f(x,d')} \pi'(\theta)d\theta = \Pi_{i=1}^{k_0+1} \int \prod_{(J^0_{i-1}, J^0_i)} \frac{f(x,d)}{f(x,d')} \pi'(\theta)d\theta = \Pi_{i=1}^{k_0+1} \int \prod_{(J^0_{i-1}, J^0_i)} \frac{f(x,d)}{f(x,d')} \pi'(\theta)d\theta$ is greater than $\Pi_{i=1}^{k_0+1} \int \prod_{(J^0_{i-1}, J^0_i)} \frac{f(x,d)}{f(x,d')} \pi'(\theta)d\theta$ for $\hat{\theta}_i$ being the m.i.e. of $\theta$ based on the observations \(\{x_i, t \in (J^0_{i-1}, J^0_i)\}\), from Johnson (1970) (page 857, (2.21)), we have, for each $i = 1, \ldots, k_0 + 1$,

\[
\lim_{n \to \infty} \sqrt{n} \int_{(J^0_{i-1}, J^0_i)} \prod_{(J^0_{i-1}, J^0_i)} \frac{f(x,d)}{f(x,d')} \pi'(\theta)d\theta = \sqrt{2\pi} \pi'(\theta^0_i)/I(\theta^0_i)
\]

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

\[
\prod_{(a,b) \in p(J^{(0)})} \int \prod_{(a,b)} \frac{f(x,d)}{f(x,d')} \pi'(\theta)d\theta \cdot \pi(k_0, J^{(0)}) \geq \varepsilon_0 \prod_{i=1}^{k_i \in 1} [J^0_i - J^0_{i-1}] \cdot \frac{1}{2} \pi(k_0, J^{(0)})
\]

with probability approaching 1 as $n \to \infty$. 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 \( \theta_0 \) 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 \( \theta_0 \). By the Taylor expansion of \( L(X; \theta) \), and \( \frac{\partial^2 L(X; \hat{\theta})}{\partial \theta^2} \) at \( \hat{\theta} \) and \( \theta_0 \), respectively, we obtain

\[
L(X; \theta_0) - L(X; \hat{\theta}) = \frac{1}{2!} \frac{\partial^2 L(X; \hat{\theta})}{\partial \theta^2} (\theta_0 - \hat{\theta})^2 + \frac{1}{3!} \frac{\partial^3 L(X; \hat{\theta})}{\partial \theta^3} (\theta_0 - \hat{\theta})^3
\]

which converges to 0 almost surely for each \( \delta > 0 \) almost surely, for each \( \delta > 0 \). Hence, for large \( n \),

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

Also, by the Taylor expansion of \( \frac{\partial L(X; \hat{\theta})}{\partial \theta} \) at \( \theta_0 \), we obtain

\[
\frac{\partial L(X; \hat{\theta})}{\partial \theta} - \frac{\partial L(X; \theta_0)}{\partial \theta} = \frac{\partial^2 L(X; \theta_0)}{\partial \theta^2} (\theta_0 - \hat{\theta}) + \frac{1}{2!} \frac{\partial^3 L(X; \theta_0)}{\partial \theta^3} (\hat{\theta} - \theta_0)^2,
\]

where \( \theta^*_3 \) is between \( \theta_0 \) and \( \hat{\theta} \). Let

\[
\tilde{S}_n = \frac{1}{n} \frac{\partial L(X; \theta_0)}{\partial \theta} = \frac{1}{n} \sum_{i=1}^n \frac{\partial \log f(x_i; \theta_0)}{\partial \theta} \quad \text{and} \quad \tilde{E}_n = \frac{1}{2!n} \frac{\partial^2 L(X; \theta_0)}{\partial \theta^2} (\hat{\theta} - \theta_0).
\]

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

\[
\tilde{S}_n = \bar{V}_n (\hat{\theta} - \theta_0) (1 - \tilde{E}_n/\bar{V}_n). \tag{A.7}
\]
From (A.6) and (A.7), we have, for large $n$,

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

This implies that, for large $n$,

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

(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 $n \to \infty$,

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

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

$$\lim_{n \to \infty} \frac{|n S_n|}{\sqrt{2n(\log \log n) I(\theta_0)}} = 1 \quad \text{a.s.}$$

Therefore, it follows from (A.8) that

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

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

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

$$\max_{0 \leq a \leq j^+} \max_{2j^+ < m \leq n} \sqrt{m - a} \int \Pi_{[a,m]} \frac{f(X_\theta; \theta)}{f(X_\theta; \theta_0)} \pi'(\theta) d\theta \leq O(1) \left[ \max_{0 \leq a \leq j^+} \prod_{(a,m)} \frac{f(X_\theta; \theta)}{f(X_\theta; \theta_0)} \right] \left[ \max_{2j^+ < m \leq n} \prod_{(j^+,m)} \frac{f(X_\theta; \theta)}{f(X_\theta; \theta_0)} \right] \left[ \sqrt{m - j^+} \int \Pi_{(j^+,m)} \frac{f(X_\theta; \theta)}{f(X_\theta; \theta_0)} \pi'(\theta) d\theta \right],$$

where $\hat{\theta}_a$ and $\hat{\theta}_m$ are the m.l.e.’s of $\theta$ 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 $\pi(k, J^{(k)}) = O(n^{-k})$, using the fact that $\sum_{1 \leq J_1 < \cdots < J_r < n} \frac{n-n}{J_r(n-J_r) n^2}$ converges as $n \to \infty$, we obtain, for each fixed $r_i$,

$$
\sum_{J_{r_i}^0 < J_{r_i}^1 < \cdots < J_{r_i}^{k_0} < J_i^0} L(J^{(r_i)}) \leq O\left(\left[J_i^0 - J_{i-1}^0\right]^{r_i - \frac{1}{2}}\right) = O\left(n^{r_i - \frac{1}{2}}\right) \quad \text{if } r_i \geq 1
$$

$$
= O(1) \quad \text{if } r_i = 0
$$

so that

$$
\sum_{C} \sum_{J^{(k-k_0)}} \left[ \prod_{i=1}^{k_0+1} L(J^{(r_i)}) \right] \frac{\pi(k, C, J^{(k-k_0)})}{\pi(k_0, J^0)}
$$

$$
\leq \left\{ \sum_{r_1 + \cdots + r_{k_0+1} = k-k_0} \prod_{i=1}^{k_0+1} \left[ \sum_{J^{(r_i)}} L(J^{(r_i)}) \right] \right\} O(\max\{\log n\}^{k_0} / n^{k-k_0})
$$

$$
= O\left(\left\{ \prod_{i=1}^{k_0+1} \left[ n^{\max\{r_i - \frac{1}{2}, 0\}} \right] \cdot \log n \right\}^{k_0} / n^{k-k_0} \right)
$$

$$
= O\left(n^{k-k_0 - \frac{1}{2}} \cdot \log n \right)^{k_0} / n^{k-k_0} = O(n^{-\varepsilon_0}) \quad \text{for some } \varepsilon_0 > 0.
$$

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

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

References


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

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

(Received June 1996; accepted August 1997)