Supplement for "Stepwise Signal Extraction via Marginal Likelihood"
by Du, Kao and Kou, Journal of the American Statistical Association

Proofs for the Theoretical Results in Section 3

We use the following notations throughout the proofs. We denote the number of observations within a given interval as $n_{(a,b]} = \#\{i : a < t_i \leq b, 1 \leq i \leq n\}$, the associated likelihood function as $p_{(a,b]}(\theta) = \prod_{t_i \in (a,b]} f(x_i|\theta)$, and the corresponding log-likelihood $l_{(a,b]}(\theta) = \log p_{(a,b]}(\theta)$. The maximum likelihood estimator based on $l_{(a,b]}$ is denoted as $\hat{\theta}_{(a,b]}$. In what follows, we present our proofs for one-dimensional θ, but we want to emphasize that this is only for notational convenience. A general dimensional case can be easily obtained through a straightforward substitution of the one dimensional quantities with their multivariate counterparts. We denote $\hat{\sigma}^2_{(a,b]} = \{-l''_{(a,b]}(\hat{\theta}_{(a,b]})\}^{-1}$, the observed Fisher information, and let $J(\theta_0)$ represent the (expected) Fisher information evaluated at θ_0. We use $\xrightarrow{P}$ and $o_p(1)$ to denote convergence in probability and $O_p(1)$ to denote a sequence bounded in probability.

Next, we list the conditions (A1)-(A5) and (B1)-(B4) discussed in Section 3. Conditions (A1)-(A5) are used to ensure the consistency of the MLE of θ_j. Conditions (B1)-(B4) ensure that the second derivative of log-likelihood is sufficiently smooth for values near θ_j. See Walker (1969).

(A1) Θ is a closed set of points on the real line.

(A2) The set of points $\{x : f(x|\theta) > 0\}$ is independent of θ; we denote this set by $\mathcal{X}$.

(A3) If θ_1, θ_2 are two distinct points of Θ, then the Lebegue measure of $\mu\{x : f(x|\theta_1) \neq f(x|\theta_2)\} > 0$.

(A4) Let $x \in \mathcal{X}$, $\theta' \in \Theta$. For all θ such that $|\theta - \theta'| < \delta$ with δ sufficiently small, we have $|\log f(x|\theta) - \log f(x|\theta')| < H_\delta(x, \theta')$, where $\lim_{\delta \to 0} H_\delta(x, \theta') = 0$, and, for the true value $\theta_0 \in \Theta$, $\lim_{\delta \to 0} \int_{\mathcal{X}} H_\delta(x, \theta') f(x|\theta_0) d\mu = 0$.

(A5) If Θ is not bounded, then for $\theta_0 \in \Theta$ and sufficiently large Δ, we have $\log f(x|\theta) - \log f(x|\theta_0) < K_\Delta(x, \theta_0)$, whenever $|\theta| > \Delta$, where $\lim_{\Delta \to \infty} \int_{\mathcal{X}} K_\Delta(x, \theta_0) f(x|\theta_0) d\mu < 0$.

(B1) $\log f(x|\theta)$ is twice differentiable with respect to θ in some neighborhood of θ_0.

(B2) Let $J(\theta_0) = \int_{\mathcal{X}} f_0 (\frac{\partial \log f_0}{\partial \theta_0})^2 d\mu$, where f_0 denotes $f(x|\theta_0)$. Then $0 < J(\theta_0) < \infty$.

(B3) $\int_{\mathcal{X}} \frac{\partial f_0}{\partial \theta_0} d\mu = \int_{\mathcal{X}} \frac{\partial^2 f_0}{\partial \theta_0^2} d\mu = 0$.

(B4) If $|\theta - \theta_0| < \delta$, where δ is sufficiently small, then $|\frac{\partial^2}{\partial \theta^2} \log f(x|\theta) - \frac{\partial^2}{\partial \theta_0^2} \log f(x|\theta_0)| < M_\delta(x, \delta_0)$, where $\lim_{\delta \to 0} \int_{\mathcal{X}} M_\delta(x, \theta_0) f(x|\theta_0) d\mu = 0$.

The following results from Walker (1969) (Theorem 1 and eq. 24) are needed to prove our results:

Lemma A.1. *Under conditions (A1)-(A5) and (B1)-(B4), if there is no change-point in the interval* $(a, b]$ *and the true value of parameter within this segment is* θ_0*, then as* $n_{(a,b]} \to \infty$*,*

(i) Let $N_0(\delta) = \{\theta : |\theta - \theta_0| < \delta\}$ *be a neighborhood of* θ_0 *contained in* Θ*, the parameter space, there exists a positive number* $k_{\theta_0}(\delta)$*, depending on* θ_0 *and* δ*, such that*

$$\lim_{n_{(a,b]}\to\infty} P[\sup_{\theta\notin N_0(\delta)} n_{(a,b]}^{-1}\{l_{(a,b]}(\theta) - l_{(a,b]}(\theta_0)\} < -k_{\theta_0}(\delta)] = 1;$$

(ii) $(n_{(a,b]}\hat{\sigma}^2_{(a,b]})^{-1} \xrightarrow{P} J(\theta_0)$;

(iii) $l_{(a,b]}(\theta_0) - l_{(a,b]}(\hat{\theta}_{(a,b]}) = O_p(1)$;

(iv) $(p_{(a,b]}(\hat{\theta}_{(a,b]})\,\hat{\sigma}_{(a,b]})^{-1} D(\boldsymbol{x}_{(a,b]}|\alpha) \xrightarrow{P} (2\pi)^{1/2}\pi(\theta_0|\alpha)$.

The following two lemmas are also needed:

Lemma A.2. *Assume regularity conditions 1)- 4). Let* a_n *be a sequence with each element lying between two true change-points* $a_n \in [\tau_j^0, \tau_{j+1}^0]$

(i) If $n_{(\tau_j^0,a_n]} \to \infty$ *and* $n_{(a_n,\tau_{j+1}^0]} \to \infty$*, then*

$$\frac{D(\boldsymbol{x}_{(\tau_j^0,a_n]}|\alpha)D(\boldsymbol{x}_{(a_n,\tau_{j+1}^0]}|\alpha)}{D(\boldsymbol{x}_{(\tau_j^0,\tau_{j+1}^0]}|\alpha)} = O_p\left(\sqrt{\frac{n_{(\tau_j^0,\tau_{j+1}^0]}}{n_{(\tau_j^0,a_n]}n_{(a_n,\tau_{j+1}^0]}}}\right).$$

(ii) If $\limsup n_{(\tau_j^0,a_n]} < \infty$ *and* $n_{(a_n,\tau_{j+1}^0]} \to \infty$*, then*

$$\frac{D(\boldsymbol{x}_{(\tau_j^0,a_n]}|\alpha)D(\boldsymbol{x}_{(a_n,\tau_{j+1}^0]}|\alpha)}{D(\boldsymbol{x}_{(\tau_j^0,\tau_{j+1}^0]}|\alpha)} = O_p(1).$$

(iii) If $n_{(\tau_j^0,a_n]} \to \infty$ *and* $\limsup n_{(a_n,\tau_{j+1}^0]} < \infty$*, then*

$$\frac{D(\boldsymbol{x}_{(\tau_j^0,a_n]}|\alpha)D(\boldsymbol{x}_{(a_n,\tau_{j+1}^0]}|\alpha)}{D(\boldsymbol{x}_{(\tau_j^0,\tau_{j+1}^0]}|\alpha)} = O_p(1).$$

PROOF of Lemma A.2. Let θ_{j+1} denote the true parameter of the segment.

(i) By Lemma A.1(iv), $n_{(\tau_j^0,a_n]} \to \infty$ and $n_{(a_n,\tau_{j+1}^0]} \to \infty$ imply

$$\begin{aligned}
&\frac{D(\boldsymbol{x}_{(\tau_j^0,a_n]}|\alpha)D(\boldsymbol{x}_{(a_n,\tau_{j+1}^0]}|\alpha)}{D(\boldsymbol{x}_{(\tau_j^0,\tau_{j+1}^0]}|\alpha)} \\
= \;& O_p\left(\frac{p_{(\tau_j^0,a_n]}(\hat{\theta}_{(\tau_j^0,a_n]})\,\hat{\sigma}_{(\tau_j^0,a_n]} \times p_{(a_n,\tau_{j+1}^0]}(\hat{\theta}_{(a_n,\tau_{j+1}^0]})\,\hat{\sigma}_{(a_n,\tau_{j+1}^0]}}{p_{(\tau_j^0,\tau_{j+1}^0]}(\hat{\theta}_{(\tau_j^0,\tau_{j+1}^0]})\,\hat{\sigma}_{(\tau_j^0,\tau_{j+1}^0]}}\right).
\end{aligned}$$

Lemma A.1(iii) tells us

$$\begin{aligned}
& \frac{p_{(\tau_j^0,a_n]}(\hat{\theta}_{(\tau_j^0,a_n]})p_{(a_n,\tau_{j+1}^0]}(\hat{\theta}_{(a_n,\tau_{j+1}^0]})}{p_{(\tau_j^0,\tau_{j+1}^0]}(\hat{\theta}_{(\tau_j^0,\tau_{j+1}^0]})} \\
= & \frac{p_{(\tau_j^0,a_n]}(\hat{\theta}_{(\tau_j^0,a_n]})[p_{(\tau_j^0,a_n]}(\theta_{j+1})]^{-1}p_{(a_n,\tau_{j+1}^0]}(\hat{\theta}_{(a_n,\tau_{j+1}^0]})[p_{(a_n,\tau_{j+1}^0]}(\theta_{j+1})]^{-1}}{p_{(\tau_j^0,\tau_{j+1}^0]}(\hat{\theta}_{(\tau_j^0,\tau_{j+1}^0]})[p_{(\tau_j^0,\tau_{j+1}^0]}(\theta_{j+1})]^{-1}} \\
= & O_p(1)
\end{aligned}$$

Note also that

$$\begin{aligned}
\frac{\hat{\sigma}_{(\tau_j^0,a_n]}\hat{\sigma}_{(a_n,\tau_{j+1}^0]}}{\hat{\sigma}_{(\tau_j^0,\tau_{j+1}^0]}} &= \frac{\hat{\sigma}_{(\tau_j^0,a_n]}\sqrt{n_{(\tau_j^0,a_n]}}\hat{\sigma}_{(a_n,\tau_{j+1}^0]}\sqrt{n_{(a_n,\tau_{j+1}^0]}}}{\hat{\sigma}_{(\tau_j^0,\tau_{j+1}^0]}\sqrt{n_{(\tau_j^0,\tau_{j+1}^0]}}}\sqrt{\frac{n_{(\tau_j^0,\tau_{j+1}^0]}}{n_{(\tau_j^0,a_n]}n_{(a_n,\tau_{j+1}^0]}}} \\
&\xrightarrow{P} \sqrt{J^{-1}(\theta_{j+1})}\sqrt{\frac{n_{(\tau_j^0,\tau_{j+1}^0]}}{n_{(\tau_j^0,a_n]}n_{(a_n,\tau_{j+1}^0]}}}.
\end{aligned}$$

by Lemma A.1(ii). The desired result follows.

(ii) If $\limsup n_{(\tau_j^0,a_n]} < \infty$ and $n_{(a_n,\tau_{j+1}^0]} \to \infty$, then similar argument applies to $D(\boldsymbol{x}_{(a_n,\tau_{j+1}^0]}|\alpha)$ and $D(\boldsymbol{x}_{(\tau_j^0,\tau_{j+1}^0]}|\alpha)$ gives

$$\frac{D(\boldsymbol{x}_{(\tau_j^0,a_n]}|\alpha)D(\boldsymbol{x}_{(a_n,\tau_{j+1}^0]}|\alpha)}{D(\boldsymbol{x}_{(\tau_j^0,\tau_{j+1}^0]}|\alpha)} = O_p\left(\frac{D(\boldsymbol{x}_{(\tau_j^0,a_n]}|\alpha)}{p_{(\tau_j^0,a_n]}(\theta_{j+1})}\sqrt{\frac{n_{(\tau_j^0,\tau_{j+1}^0]}}{n_{(a_n,\tau_{j+1}^0]}}}\right).$$

$\limsup n_{(\tau_j^0,a_n]} < \infty$ implies that $n_{(\tau_j^0,a_n]}$ is bounded, say by $B < \infty$, for all n. This implies $n_{(\tau_j^0,\tau_{j+1}^0]}/n_{(a_n,\tau_{j+1}^0]} \to 1$. Furthermore, note that for all θ, $p_{(\tau_j^0,a_n]}(\theta)$ is a product of up to B i.i.d. random variables and $D(\boldsymbol{x}_{(\tau_j^0,a_n]}|\alpha) = \int_\Theta p_{(\tau_j^0,a_n]}(\theta)\pi(\theta|\alpha)d\theta$. B is finite. $D(\boldsymbol{x}_{(\tau_j^0,a_n]}|\alpha)/p_{(\tau_j^0,a_n]}(\theta_{j+1})$ is, therefore, bounded in probability. The desired result thus follows. □

The proof of (iii) is essentially identical to that of (ii).

Lemma A.3. *Assume regularity conditions 1)- 4). Let $(a_n, b_n]$ be a sequence of intervals that contains one and only one true change-point τ^0.*

(i) If $n_{(a_n,\tau^0]} \to \infty$ and $n_{(\tau^0,b_n]} \to \infty$, then

$$\frac{D(\boldsymbol{x}_{(a_n,b_n]}|\alpha)}{D(\boldsymbol{x}_{(a_n,\tau^0]}|\alpha)D(\boldsymbol{x}_{(\tau^0,b_n]}|\alpha)} \xrightarrow{P} 0. \tag{A.1}$$

(ii) If $\limsup n_{(a_n,\tau^0]} < \infty$ and $n_{(\tau^0,b_n]} \to \infty$, then

$$\frac{D(\boldsymbol{x}_{(a_n,b_n]}|\alpha)}{D(\boldsymbol{x}_{(a_n,\tau^0]}|\alpha)D(\boldsymbol{x}_{(\tau^0,b_n]}|\alpha)} = O_p(1). \tag{A.2}$$

(iii) If $n_{(a_n,\tau^0]} \to \infty$ and $\limsup n_{(\tau^0,b_n]} < \infty$, then

$$\frac{D(\boldsymbol{x}_{(a_n,b_n]}|\alpha)}{D(\boldsymbol{x}_{(a_n,\tau^0]}|\alpha)D(\boldsymbol{x}_{(\tau^0,b_n]}|\alpha)} = O_p(1). \tag{A.3}$$

PROOF of Lemma A.3. Let θ_1 and θ_2 be the two segment-parameters before and after τ^0. By definition, $D(\boldsymbol{x}_{(a_n,b_n]}|\alpha) = \int_\Theta p_{(a_n,\tau^0]}(\theta)p_{(\tau^0,b_n]}(\theta)\pi(\theta|\alpha)d\theta$. Let $N_1(\delta)$ and $N_2(\delta)$ be disjoint neighborhoods of θ_1 and θ_2. We split $D(\boldsymbol{x}_{(a_n,b_n]}|\alpha)$ into three integrals, I_1, I_2 and I_3, taken respectively over the sets $N_1(\delta)$, $N_2(\delta)$ and $\Theta - N_1(\delta) - N_2(\delta)$.

(i) If $n_{(a_n,\tau^0]} \to \infty$ and $n_{(\tau^0,b_n]} \to \infty$, then for the first integral, we can write

$$\begin{aligned} I_1 &= \int_{N_1(\delta)} p_{(a_n,\tau^0]}(\theta)p_{(\tau^0,b_n]}(\theta)\pi(\theta|\alpha)d\theta \\ &= p_{(\tau^0,b_n]}(\hat\theta_{(\tau^0,b_n]})\,\hat\sigma_{(\tau^0,b_n]}\exp[l_{(\tau^0,b_n]}(\theta_2) - l_{(\tau^0,b_n]}(\hat\theta_{(\tau^0,b_n]})] \\ &\times \int_{N_1(\delta)} \hat\sigma^{-1}_{(\tau^0,b_n]}\exp\{l_{(\tau^0,b_n]}(\theta) - l_{(\tau^0,b_n]}(\theta_2)\}p_{(a_n,\tau^0]}(\theta)\pi(\theta|\alpha)d\theta. \end{aligned}$$

According to Lemma A.1(i), the integral on the above right-hand side is less than

$$\begin{aligned} & \hat\sigma^{-1}_{(\tau^0,b_n]}\exp(-n_{(\tau^0,b_n]}k_2(\delta))\int_{N_1(\delta)} p_{(a_n,\tau^0]}(\theta)\pi(\theta|\alpha)d\theta \\ \le\ & \hat\sigma^{-1}_{(\tau^0,b_n]}\exp(-n_{(\tau^0,b_n]}k_2(\delta))\int_{\Theta} p_{(a_n,\tau^0]}(\theta)\pi(\theta|\alpha)d\theta \\ =\ & \{n_{(\tau^0,b_n]}\hat\sigma^2_{(\tau^0,b_n]}\}^{-1/2}n^{1/2}_{(\tau^0,b_n]}\exp(-n_{(\tau^0,b_n]}k_2(\delta))D(\boldsymbol{x}_{(a_n,\tau^0]}|\alpha) \end{aligned}$$

with probability tending to 1. We know from Lemma A.1(ii), (iii) and (iv) that $n_{(\tau^0,b_n]} \to \infty$ implies

$$\begin{aligned} & [n_{(\tau^0,b_n]}\hat\sigma^2_{(\tau^0,b_n]}]^{-1/2} \xrightarrow{P} J^{1/2}(\theta_2) \\ & \exp\{l_{(\tau^0,b_n]}(\theta_2) - l_{(\tau^0,b_n]}(\hat\theta_{(\tau^0,b_n]})\} = O_p(1), \\ & [p_{(\tau^0,b_n]}(\hat\theta_{(\tau^0,b_n]})\,\hat\sigma_{(\tau^0,b_n]}]^{-1}D(\boldsymbol{x}_{(\tau^0,b_n]}|\alpha) \xrightarrow{P} (2\pi)^{1/2}\pi(\theta_2|\alpha). \end{aligned}$$

It follows that

$$\frac{I_1}{D(\boldsymbol{x}_{(a_n,\tau^0]}|\alpha)D(\boldsymbol{x}_{(\tau^0,b_n]}|\alpha)} = O_p(n^{1/2}_{(\tau^0,b_n]}\exp\{-n_{(\tau^0,b_n]}k_2(\delta)\}) \xrightarrow{P} 0. \tag{A.4}$$

Identical argument applied to I_2, the integral over $N_2(\delta)$, together with $n_{(a_n,\tau^0]} \to \infty$, gives

$$\frac{I_2}{D(\boldsymbol{x}_{(a_n,\tau^0]}|\alpha)D(\boldsymbol{x}_{(\tau^0,b_n]}|\alpha)} = O_p(n^{1/2}_{(a_n,\tau^0]}\exp\{-n_{(a_n,\tau^0]}k_1(\delta)\}) \xrightarrow{P} 0.$$

For the integral I_3, we apply the same argument, but we note that since the region $\Theta - N_1(\delta) - N_2(\delta)$ contains neither the neighborhood of θ_1 nor the neighborhood of θ_2,

$$\frac{I_3}{D(\boldsymbol{x}_{(a_n,\tau^0]}|\alpha)D(\boldsymbol{x}_{(\tau^0,b_n]}|\alpha)} = O_p((n_{(a_n,\tau^0]}n_{(\tau^0,b_n]})^{1/2}\exp\{-n_{(a_n,\tau^0]}k_1(\delta) - n_{(\tau^0,b_n]}k_2(\delta)\}),$$

which converges to zero even faster. This proves (A.1).

(ii) $n_{(\tau^0,b_n]} \to \infty$ alone gives (A.4) and that

$$\frac{I_3}{D(\boldsymbol{x}_{(a_n,\tau^0]}|\alpha)D(\boldsymbol{x}_{(\tau^0,b_n]}|\alpha)} = O_p(n^{1/2}_{(\tau^0,b_n]}\exp\{-n_{(\tau^0,b_n]}k_2(\delta)\}) \xrightarrow{P} 0.$$

Let us next consider I_2. If $\limsup n_{(a_n,\tau^0]} < \infty$, then we know that $n_{(a_n,\tau^0]}$ is bounded, say by $B < \infty$, for all n.

$$\begin{aligned} I_2 &= \int_{N_2(\delta)} p_{(a_n,\tau^0]}(\theta)p_{(\tau^0,b_n]}(\theta)\pi(\theta|\alpha)d\theta \\ &= p_{(a_n,\tau^0]}(\theta_2)\int_{N_2(\delta)} \exp\{l_{(a_n,\tau^0]}(\theta) - l_{(a_n,\tau^0]}(\theta_2)\}p_{(\tau^0,b_n]}(\theta)\pi(\theta|\alpha)d\theta \end{aligned}$$

Condition (A4) tells us that $\left|l_{(a_n,\tau^0]}(\theta) - l_{(a_n,\tau^0]}(\theta_2)\right| \le \sum H_\delta(x_i,\theta_2)$, where the sum is over $t_i \in (a_n,\tau^0]$, which has up to B terms. It follows that

$$\begin{aligned} I_2 &\le p_{(a_n,\tau^0]}(\theta_2)\exp\{\sum H_\delta(x_i,\theta_2)\}\int_{N_2(\delta)} p_{(\tau^0,b_n]}(\theta)\pi(\theta|\alpha)d\theta \\ &\le p_{(a_n,\tau^0]}(\theta_2)\exp\{\sum H_\delta(x_i,\theta_2)\}D(\boldsymbol{x}_{(\tau^0,b_n]}|\alpha). \end{aligned}$$

Thus

$$\frac{I_2}{D(\boldsymbol{x}_{(a_n,\tau^0]}|\alpha)D(\boldsymbol{x}_{(\tau^0,b_n]}|\alpha)} \le \frac{p_{(a_n,\tau^0]}(\theta_2)\exp\{\sum H_\delta(x_i,\theta_2)\}}{\int_\Theta p_{(a_n,\tau^0]}(\theta)\pi(\theta|\alpha)d\theta} \tag{A.5}$$

Note that $p_{(a_n,\tau^0]}(\theta)$ is a product of up to B i.i.d. random variables and $\sum H_\delta(x_i,\theta_2)$ is a sum of up to B i.i.d. random variables. B is finite. The right hand side of (A.5) is, therefore, bounded in probability. This gives (A.2). $\Box$

The proof of (iii) is essentially identical to that of (ii).

PROOF of Lemma 3.1. First, let us consider the case of $m = 1$.

$$\frac{P(\boldsymbol{x}|\{0,\tau_1,1\})}{P(\boldsymbol{x}|\{0,1\})} = \frac{D(\boldsymbol{x}_{(0,\tau_1]}|\alpha)D(\boldsymbol{x}_{(\tau_1,1]}|\alpha)}{D(\boldsymbol{x}_{(0,1]}|\alpha)}.$$

Lemma A.2(i) tells us that it is $O_p(n^{1/2}_{(0,1]}/(n_{(0,\tau_1]}n_{(\tau_1,1]})^{1/2})$. But $n_{(0,\tau_1]}n_{(\tau_1,1]}/\left\{n^2C_{(0,\tau_1]}C_{(\tau_1,1]}\right\} \xrightarrow{P} 1$ by regularity condition 3, it follows that

$$\frac{P(\boldsymbol{x}|\{0,\tau_1,1\})}{P(\boldsymbol{x}|\{0,1\})} = O_p\left(1/\sqrt{nC_{(0,\tau_1]}C_{(\tau_1,1]}}\right) = O_p(1/\sqrt{n\Delta_\tau}).$$

Next, suppose that the lemma holds for all $m \le M$, $(M > 1)$. Then for $m = M+1$,

$$\frac{P(\boldsymbol{x}|\{0,\tau_1,\cdots,\tau_M,1\})}{P(\boldsymbol{x}|\{0,1\})} = \frac{P(\boldsymbol{x}|\{0,\tau_2,\cdots,\tau_M,1\})}{P(\boldsymbol{x}|\{0,1\})}\frac{P(\boldsymbol{x}|\{0,\tau_1,\cdots,\tau_M,1\})}{P(\boldsymbol{x}|\{0,\tau_2,\cdots,\tau_M,1\})}.$$

By the induction assumption, $P(\boldsymbol{x}|\{0,\tau_2,\cdots,\tau_M,1\})/P(\boldsymbol{x}|\{0,1\}) \xrightarrow{P} 0$. Note that

$$\frac{P(\boldsymbol{x}|\{0,\tau_1,\cdots,\tau_M,1\})}{P(\boldsymbol{x}|\{0,\tau_2,\cdots,\tau_M,1\})} = \frac{D(\boldsymbol{x}_{(0,\tau_1]}|\alpha)D(\boldsymbol{x}_{(\tau_1,\tau_2]}|\alpha)}{D(\boldsymbol{x}_{(0,\tau_2]}|\alpha)}.$$

Lemma A.2(i) again tells us that the above expression converges to 0 in probability. Therefore, the lemma is also true for $m = M+1$: $P(\boldsymbol{x}|\{0,\tau_1,\cdots,\tau_M,1\})/P(\boldsymbol{x}|\{0,1\}) = O_p(1/\sqrt{n\Delta_\tau})$. □

PROOF of Lemma 3.2. We need only to prove this lemma for $m_0 = 2$; the rest can be proved using the same mathematical induction technique as in the proof of Lemma 3.1. We have

$$\frac{P(\boldsymbol{x}|\{0,1\})}{P(\boldsymbol{x}|\{0,\tau_1^0,1\})} = \frac{D(\boldsymbol{x}_{(0,1]}|\alpha)}{D(\boldsymbol{x}_{(0,\tau_1^0]}|\alpha)D(\boldsymbol{x}_{(\tau_1^0,1]}|\alpha)}.$$

Taking $a_n \equiv 0$ and $b_n \equiv 1$ in Lemma A.3(i), we know from its proof that

$$\frac{D(\boldsymbol{x}_{(0,1]}|\alpha)}{D(\boldsymbol{x}_{(0,\tau_1^0]}|\alpha)D(\boldsymbol{x}_{(\tau_1^0,1]}|\alpha)} = O_p(n_{(0,\tau_1^0]}^{1/2}\exp\{-n_{(0,\tau_1^0]}k_1(\delta)\}) + O_p(n_{(\tau_1^0,1]}^{1/2}\exp\{-n_{(\tau_1^0,1]}k_2(\delta)\}).$$

Condition 3 suggests that $O_p(\sqrt{n_{(\tau_1^0,1]}}\exp\{-n_{(\tau_1^0,1]}k_2(\delta)\}) = O_p(\sqrt{n\Delta_0}\exp(-cn\Delta_0))$, for positive constant c, and so does $O_p(\sqrt{n_{(0,\tau_1^0]}}\exp\{-n_{(0,\tau_1^0]}k_1(\delta)\})$. We thus prove the lemma for $m_0 = 2$. □

PROOF of Theorem 3.3. Our proof consists of three steps. Step 1. Let $\mathcal{E}_1$ be the event that there is at least one true change-point τ_j^0 $(0 \le j \le m_0)$ that no estimated change-point is within $\overline{\Delta}/2$ of it, i.e., $\hat{\tau}_i \notin (\tau_j^0 - \overline{\Delta}/2, \tau_j^0 + \overline{\Delta}/2)$ for all i. We will show that the probability of $\mathcal{E}_1$ goes to 0.

Suppose $\hat{\boldsymbol{\tau}}$ is such an estimate. Let $\hat{\tau}_i$ and $\hat{\tau}_{i+1}$ be the estimated change-points that sandwich τ_j^0: $\hat{\tau}_i < \tau_j^0 < \hat{\tau}_{i+1}$. Let $\tau_{j-l}^0 < \cdots < \tau_{j+r}^0$ be the sequence of true change-points $(l, r \ge 0)$ that are between $\hat{\tau}_i$ and $\hat{\tau}_{i+1}$

$$\tau_{j-l-1}^0 < \hat{\tau}_i < \tau_{j-l}^0 < \ldots < \tau_{j+r}^0 < \hat{\tau}_{i+1} < \tau_{j+r+1}^0.$$

Consider an alternative choice of change-points

$$\tilde{\boldsymbol{\tau}} = \{\hat{\tau}_0, \hat{\tau}_1, \cdots, \hat{\tau}_i, \tau_{j-l}^0, \ldots, \tau_{j+k}^0, \hat{\tau}_{i+1}, \cdots, \hat{\tau}_{\hat{m}}\},$$

which is formed by inserting $\tau_{j-l}^0 < \cdots < \tau_{j+r}^0$ into $\hat{\boldsymbol{\tau}}$. It is clear that

$$\frac{P(\boldsymbol{x}|\hat{\boldsymbol{\tau}})}{P(\boldsymbol{x}|\tilde{\boldsymbol{\tau}})} = \frac{D(\boldsymbol{x}_{(\hat{\tau}_i,\hat{\tau}_{i+1}]}|\alpha)}{D(\boldsymbol{x}_{(\hat{\tau}_i,\tau_{j-l}^0]}|\alpha) \times D(\boldsymbol{x}_{(\tau_{j-l}^0,\tau_{j-l+1}^0]}|\alpha) \times \cdots \times D(\boldsymbol{x}_{(\tau_{j+k-1}^0,\tau_{j+k}^0]}|\alpha) \times D(\boldsymbol{x}_{(\tau_{j+k}^0,\hat{\tau}_{i+1}]}|\alpha)}.$$

Since $n_{(\hat{\tau}_i,\tau_j^0]} \to \infty$, $n_{(\tau_j^0,\hat{\tau}_{i+1}]} \to \infty$ and $n_{(\tau_k^0,\tau_{k+1}^0]} \to \infty$ (for any k) by condition 5, it follows from Lemma A.3 that the ratio $P(\boldsymbol{x}|\hat{\boldsymbol{\tau}})/P(\boldsymbol{x}|\tilde{\boldsymbol{\tau}})$ would go to zero in probability. Another way to look at it is to think of $\tilde{\boldsymbol{\tau}}$ as being created by inserting the true change-points one at a time from the left. The first and last insertions would have probability contribution of $O_p(1)$ by Lemma A.3, while the middle ones would have probability contribution $o_p(1)$ by Lemma 3.2. Therefore, the probability of having such an estimate $\hat{\boldsymbol{\tau}}$ goes to zero, i.e., the probability of $\mathcal{E}_1$ goes to zero. This in fact proves equation (3.1), since $\overline{\Delta}/2 \to 0$ by condition 5.

Step 2. The previous step tells us that, with probability going to one, for each true change-point τ_j^0, there would be at least one estimated change-point $\hat{\tau}_i$ such that $|\hat{\tau}_i - \tau_j^0| < \overline{\Delta}/2$. On the other hand, since $\hat{\tau}_{i+1} - \hat{\tau}_i \geq \overline{\Delta}$ by the definition, we know that there cannot be two estimated change-point within $(\tau_j^0 - \overline{\Delta}/2, \tau_j^0 + \overline{\Delta}/2)$. Hence, with probability going to one, for each true change-point τ_j^0, there would be one and only one estimated change-point $\hat{\tau}_i$ such that $|\hat{\tau}_i - \tau_j^0| < \overline{\Delta}/2$.

Step 3. In order to establish $\hat{m} \xrightarrow{P} m_0$, it remains to show that, with probability going to one, the union of $\bigcup_j (\tau_j^0 - \overline{\Delta}/2, \tau_j^0 + \overline{\Delta}/2)$ contains all the estimated change-points. Suppose $\hat{\tau}_i$ is outside the union. Let τ_j^0 and τ_{j+1}^0 be the adjacent true change-points that sandwich $\hat{\tau}_i$: $\tau_j^0 < \hat{\tau}_i < \tau_{j+1}^0$. We must have $\hat{\tau}_i - \tau_j^0 \geq \overline{\Delta}/2$ and $\tau_{j+1}^0 - \hat{\tau}_i \geq \overline{\Delta}/2$. From Steps 1 and 2, we know that with probability going to one, there are two estimated change-points of which one is within $\overline{\Delta}/2$ of τ_j^0 and the other is within $\overline{\Delta}/2$ of τ_{j+1}^0. Let $\hat{\tau}_{i-l} < \cdots < \hat{\tau}_{i+r}$ be the sequence of estimated change-points $(l, r \geq 0)$ that are between τ_j^0 and τ_{j+1}^0:

$$\tau_j^0 < \hat{\tau}_{i-l} < \cdots < \hat{\tau}_{i+r} < \tau_{j+1}^0$$

Consider the following alternative change-points:

$$\tilde{\boldsymbol{\tau}} := \hat{\boldsymbol{\tau}} - \{\hat{\tau}_{i-l}, \ldots, \hat{\tau}_{i+r}\} = \{\tau_0, \tau_1, \cdots, \tau_{i-l-1}, \tau_{i+r+1}, \ldots, \tau_m\}.$$

We can think of $\tilde{\boldsymbol{\tau}}$ as being created by deleting from $\hat{\boldsymbol{\tau}}$ the estimated change-points one at a time starting from $\hat{\tau}_{i-l}$. According to Lemma A.2, deleting $\hat{\tau}_{i-l}$ and $\hat{\tau}_{i+r}$ would have probability contribution of either $O_p(1)$ or $o_p(1)$, while deleting the middle ones would have probability contribution of $o_p(1)$, since $\hat{\tau}_{k+1} - \hat{\tau}_k \geq \overline{\Delta}$ by the definition and $n\overline{\Delta} \to \infty$. It follows that the ratio $P(\boldsymbol{x}|\hat{\boldsymbol{\tau}})/P(\boldsymbol{x}|\tilde{\boldsymbol{\tau}})$ would go to zero in probability. Therefore, the probability of having a $\hat{\tau}_i$ outside the union of $\bigcup_j (\tau_j^0 - \overline{\Delta}/2, \tau_j^0 + \overline{\Delta}/2)$ goes to zero. This concludes our proof. □

PROOF of Corollary 3.4. Since in the proof of Theorem 3.3 the only place that $\pi(\theta|\alpha)$ appears is in Lemma A.1 (iv), for the proof we only need to show that

$$(p_{(a,b]}(\hat{\theta}_{(a,b]})\hat{\sigma}_{(a,b]})^{-1} D(x_{(a,b]}|\hat{\alpha}_n) \xrightarrow{P} (2\pi)^{1/2}\pi(\theta_0|\alpha^*). \tag{A.6}$$

To do so, let $N(\delta)$ be a neighborhood of θ_1. Then, we have

$$D(x_{(a,b]}|\hat{\alpha}_n) = \int_{N(\delta)} p_{(a,b]}(\theta)\pi(\theta|\hat{\alpha}_n)d\theta + \int_{\Theta - N(\delta)} p_{(a,b]}(\theta)\pi(\theta|\hat{\alpha}_n)d\theta.$$

For the first term, since $\pi(\theta|\alpha)$ is continuous at α^*, and $\hat{\alpha}_n \xrightarrow{P} \alpha^*$, we have,

$$\int_{N(\delta)} p_{(a,b]}(\theta)\pi(\theta|\hat{\alpha}_n)d\theta = \int_{N(\delta)} p_{(a,b]}(\theta)\frac{\pi(\theta|\hat{\alpha}_n)}{\pi(\theta|\alpha^*)}\pi(\theta|\alpha^*)d\theta = (1-o_p(1))\int_{N(\delta)} p_{(a,b]}(\theta)\pi(\theta|\alpha^*)d\theta.$$

For the second term, by Lemma A.1 (i) and a similar analogue to Lemma A.3, one could show that

$$(p_{(a,b]}(\hat{\theta}_{(a,b]})\hat{\sigma}_{(a,b]})^{-1}\int_{\Theta-N(\delta)} p_{(a,b]}(\theta)\pi(\theta|\hat{\alpha}_n)d\theta = o_p(1).$$

Similarly, we have

$$(p_{(a,b]}(\hat{\theta}_{(a,b]})\hat{\sigma}_{(a,b]})^{-1}\int_{\Theta-N(\delta)} p_{(a,b]}(\theta)\pi(\theta|\alpha^*)d\theta = o_p(1).$$

Combining them, we know that replacing $\hat{\alpha}_n$ by α^* does not change the asymptotics of the left hand side of (A.6), that is,

$$(p_{(a,b]}(\hat{\theta}_{(a,b]})\hat{\sigma}_{(a,b]})^{-1}D(x_{(a,b]}|\hat{\alpha}_n) = (p_{(a,b]}(\hat{\theta}_{(a,b]})\hat{\sigma}_{(a,b]})^{-1}D(x_{(a,b]}|\alpha^*) + o_p(1) = (2\pi)^{1/2}\pi(\theta_0|\alpha^*) + o_p(1)$$

This completes the proof. □

References

[1] Walker, A. M. (1969). On the asymptotic behaviour of posterior distributions. *J. Roy. Statist. Soc.*, B, 31, 80-88.