Supplement for “Stepwise Signal Extraction via Marginal Likelihood”

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 \( l_{(a, b]}(\theta) \) is denoted as \( \hat{\theta}_{(a, b]} \). In what follows, we present our proofs for one-dimensional \( \theta \), 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 \( \theta_0 \). We use \( \overset{P}{\to} \) 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 \( \theta_j \). Conditions (B1)-(B4) ensure that the second derivative of log-likelihood is sufficiently smooth for values near \( \theta_j \). See Walker (1969).

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

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

(A3) If \( \theta_1, \theta_2 \) are two distinct points of \( \Theta \), 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 \( \theta \) such that \( |\theta - \theta'| < \delta \) with \( \delta \) 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 \( \theta_0 \in \Theta \), \( \lim_{\delta \to 0} \int_{\mathcal{X}} \delta H_\delta(x, \theta') f(x | \theta_0) d\mu = 0 \).

(A5) If \( \Theta \) is not bounded, then for \( \theta_0 \in \Theta \) and sufficiently large \( \Delta \), 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 \( \theta \) in some neighborhood of \( \theta_0 \).

(B2) Let \( J(\theta_0) = \int_{\mathcal{X}} f_0 \left( \frac{\partial \log f_0}{\partial \theta_0} \right)^2 d\mu \), 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 \( \delta \) is sufficiently small, then \( |\frac{\partial^2}{\partial \theta_0^2} \log f(x | \theta) - \frac{\partial^2}{\partial \theta_0^2} \log f(x | \theta_0)| < M_\delta(x, \theta_0) \), 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 \(\theta_0\), then as \(n_{(a, b]} \to \infty\),

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

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

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

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

(iv) \((p_{(a, b]}(\hat{\theta}_{(a, b]})\hat{\sigma}_{(a, b]})^{-1} D(\theta_{(a, b]}|\alpha) \overset{P}{\to} (2\pi)^{-1/2}(\theta_0|\alpha)\).

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 \(a_n \in [0, \tau_0, \tau_{j+1}]\)

(i) If \(n_{(\tau_{j, a_n}] \to \infty\) and \(n_{(a_n, \tau_{j+1}] \to \infty\), then

\[
\frac{D(\theta_{(\tau_{j, a_n}]}|\alpha)D(\theta_{(a_n, \tau_{j+1}]}|\alpha)}{D(\theta_{(\tau_{j, a_n}]}|\alpha)} = O_p\left(\sqrt{\frac{n_{(\tau_{j, a_n}]}}{n_{(a_n, \tau_{j+1}]}}}\right).
\]

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

\[
\frac{D(\theta_{(\tau_{j, a_n}]}|\alpha)D(\theta_{(a_n, \tau_{j+1}]}|\alpha)}{D(\theta_{(\tau_{j, a_n}]}|\alpha)} = O_p(1).
\]

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

\[
\frac{D(\theta_{(\tau_{j, a_n}]}|\alpha)D(\theta_{(a_n, \tau_{j+1}]}|\alpha)}{D(\theta_{(\tau_{j, a_n}]}|\alpha)} = O_p(1).
\]

**PROOF of Lemma A.2.** Let \(\theta_{j+1}\) denote the true parameter of the segment.

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

\[
\frac{D(\theta_{(\tau_{j, a_n}]}|\alpha)D(\theta_{(a_n, \tau_{j+1}]}|\alpha)}{D(\theta_{(\tau_{j, a_n}]}|\alpha)} = O_p\left(\frac{p_{(\tau_{j, a_n}])(\hat{\sigma}_{(\tau_{j, a_n}]}\times p_{(a_n, \tau_{j+1}]}(\hat{\theta}_{(a_n, \tau_{j+1}]}\hat{\sigma}_{(a_n, \tau_{j+1}]}\overset{P}{\to} p_{(\tau_{j, a_n}])(\hat{\sigma}_{(\tau_{j, a_n}]}\hat{\sigma}_{(a_n, \tau_{j+1}]}\right).
\]

2
Lemma A.1(iii) tells us

\[
\frac{p(\tau_j^0, a_n) | \hat{\theta}(\tau_j^0, a_n)}{p(\tau_j^0, a_n) | \hat{\theta}(\tau_j^0, a_n)} = O_p(1)
\]

Note also that

\[
\frac{\hat{\sigma}(\tau_j^0, a_n) \hat{\sigma}(\tau_{j+1}^0, a_n)}{\hat{\sigma}(\tau_j^0, a_n) \hat{\sigma}(\tau_{j+1}^0, a_n)} = \frac{\hat{\sigma}(\tau_j^0, a_n) \sqrt{n(\tau_j^0, a_n)} \hat{\sigma}(\tau_{j+1}^0, a_n) \sqrt{n(\tau_{j+1}^0, a_n)}}{\hat{\sigma}(\tau_j^0, a_n) \sqrt{n(\tau_j^0, a_n)} \hat{\sigma}(\tau_{j+1}^0, a_n) \sqrt{n(\tau_{j+1}^0, a_n)}}
\]

\[
\xrightarrow{P} \sqrt{J-1} \left( \frac{n(\tau_j^0, a_n)}{n(\tau_{j+1}^0, a_n)} \right)
\]

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

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

\[
\frac{D(x(\tau_j^0, a_n) | \alpha) D(x(\tau_{j+1}^0, a_n) | \alpha)}{D(x(\tau_j^0, a_n) | \alpha)} = O_p \left( \frac{D(x(\tau_j^0, a_n) | \alpha)}{p(\tau_j^0, a_n)(\theta_j+1)} \right)
\]

\[
\limsup n(\tau_j^0, a_n) < \infty \text{ implies that } n(\tau_j^0, a_n) \text{ is bounded, say by } B < \infty, \text{ for all } n. \text{ This implies } n(\tau_j^0, a_n) \to 1. \text{ Furthermore, note that for all } \theta, p(\tau_j^0, a_n)(\theta) \text{ is a product of up to } B \text{ i.i.d. random variables and } D(x(\tau_j^0, a_n) | \alpha) = \int \theta p(\tau_j^0, a_n)(\theta) \pi(\theta | \alpha) d\theta. \text{ } B \text{ is finite. } D(x(\tau_j^0, a_n) | \alpha)/p(\tau_j^0, a_n)(\theta_j+1) \text{ is, therefore, bounded in probability. The desired result thus follows. } \]

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 \( \tau^0 \).

(i) If \( n(\tau^0, a_n) \to \infty \) and \( n(\tau^0, b_n) \to \infty \), then

\[
\frac{D(x(\tau^0, a_n) | \alpha)}{D(x(\tau^0, b_n) | \alpha)} \xrightarrow{P} 0.
\]

(ii) If \( \limsup n(\tau^0, a_n) < \infty \) and \( n(\tau^0, b_n) \to \infty \), then

\[
\frac{D(x(\tau^0, a_n) | \alpha)}{D(x(\tau^0, b_n) | \alpha)} = O_p(1).
\]
(iii) If \( n_{(a_n, \tau^o]} \to \infty \) and \( \limsup n_{(\tau^o, b_n]} < \infty \), then

\[
\frac{D(x_{(a_n, \tau^o]}|\alpha)}{D(x_{(a_n, \tau^o]}|\alpha) D(x_{(\tau^o, b_n]}|\alpha)} = O_p(1). 
\]  

(A.3)

PROOF of Lemma A.3. Let \( \theta_1 \) and \( \theta_2 \) be the two segment-parameters before and after \( \tau^o \). By definition, \( D(x_{(a_n, \tau^o]}|\alpha) = \int_{\Theta} p_{(a_n, \tau^o]}(\theta) p_{(\tau^o, b_n]}(\theta) \pi(\theta|\alpha) d\theta \). Let \( N_1(\delta) \) and \( N_2(\delta) \) be disjoint neighborhoods of \( \theta_1 \) and \( \theta_2 \). We split \( D(x_{(a_n, b_n]}|\alpha) \) 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 \( n_{(a_n, \tau^o]} \to \infty \) and \( n_{(\tau^o, b_n]} \to \infty \), then for the first integral, we can write

\[
I_1 = \int_{N_1(\delta)} p_{(a_n, \tau^o]}(\theta) p_{(\tau^o, b_n]}(\theta) \pi(\theta|\alpha) d\theta 
\]

\[
= p_{(\tau^o, b_n]}(\hat{\theta}(\tau^o, b_n)) \hat{\sigma}(\tau^o, b_n) \exp[l_{(\tau^o, b_n]}(\hat{\theta}(\tau^o, b_n))] \times \int_{N_1(\delta)} \hat{\sigma}^{-1}(\tau^o, b_n) \exp[l_{(\tau^o, b_n]}(\theta) - l_{(\tau^o, b_n]}(\hat{\theta}(\tau^o, b_n))] p_{(a_n, \tau^o]}(\theta) \pi(\theta|\alpha) d\theta. 
\]

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

\[
\hat{\sigma}^{-1}(\tau^o, b_n) \exp(-n_{(\tau^o, b_n]} k_2(\delta)) \int_{N_1(\delta)} p_{(a_n, \tau^o]}(\theta) \pi(\theta|\alpha) d\theta 
\]

\[
\leq \hat{\sigma}^{-1}(\tau^o, b_n) \exp(-n_{(\tau^o, b_n]} k_2(\delta)) \int_{\Theta} p_{(a_n, \tau^o]}(\theta) \pi(\theta|\alpha) d\theta 
\]

\[
= \{n_{(\tau^o, b_n]} \hat{\sigma}^2(\tau^o, b_n)\}^{-1/2} n_{(\tau^o, b_n]}^{1/2} \exp(-n_{(\tau^o, b_n]} k_2(\delta)) D(x_{(a_n, \tau^o]}|\alpha) 
\]

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

\[
[n_{(\tau^o, b_n]} \hat{\sigma}^2(\tau^o, b_n)]^{-1/2} \xrightarrow{p} \mathcal{J}(\theta_2) 
\]

\[
\exp[l_{(\tau^o, b_n]}(\theta_2) - l_{(\tau^o, b_n]}(\hat{\theta}(\tau^o, b_n))] = O_p(1), 
\]

\[
[p_{(a_n, \tau^o]}(\hat{\theta}(\tau^o, b_n)) \hat{\sigma}(\tau^o, b_n)]^{-1} D(x_{(\tau^o, b_n]}|\alpha) \xrightarrow{p} (2\pi)^{1/2} \mathcal{P}(\theta_2|\alpha). 
\]

It follows that

\[
\frac{I_1}{D(x_{(a_n, \tau^o]}|\alpha) D(x_{(\tau^o, b_n]}|\alpha)} = O_p(n_{(\tau^o, b_n]}^{1/2} \exp\{-n_{(\tau^o, b_n]} k_2(\delta)\}) \xrightarrow{p} 0. 
\]  

(A.4)

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

\[
\frac{I_2}{D(x_{(a_n, \tau^o]}|\alpha) D(x_{(\tau^o, b_n]}|\alpha)} = O_p(n_{(a_n, \tau^o]}^{1/2} \exp\{-n_{(a_n, \tau^o]} k_1(\delta)\}) \xrightarrow{p} 0. 
\]

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 \( \theta_1 \) nor the neighborhood of \( \theta_2 \),

\[
\frac{I_3}{D(x_{(a_n, \tau^o]}|\alpha) D(x_{(\tau^o, b_n]}|\alpha)} = O_p((n_{(a_n, \tau^o]} n_{(\tau^o, b_n]})^{1/2} \exp\{-n_{(a_n, \tau^o]} k_1(\delta) - n_{(\tau^o, b_n]} k_2(\delta)\}),
\]
which converges to zero even faster. This proves (A.1).

(ii) \( n^{(a,\tau_0)} \to \infty \) alone gives (A.4) and that

\[
I_3 = \frac{D(x^{(a,\tau_0)}_0|\alpha)D(x^{(a,\tau_0)}_0|\alpha)}{D(x^{(a,\tau_0)}_0|\alpha)} = O_p\left( n^{1/2} \exp\left\{ -n^{(a,\tau_0)}_0k_2(\delta) \right\} \right) \to 0.
\]

Let us next consider \( I_2 \). If \( \lim \sup n^{(a,\tau_0)} \leq M \), then we know that \( n^{(a,\tau_0)} \) is bounded, say by \( B < \infty \), for all \( n \).

\[
I_2 = \int_{N_2(\delta)} p^{(a,\tau_0)}(\theta) \pi(\theta|\alpha) \exp\left\{ I^{(a,\tau_0)}_0(\theta) - I^{(a,\tau_0)}_1(\theta) \right\} d\theta
\]

Condition (A4) tells us that \( |I^{(a,\tau_0)}_0(\theta) - I^{(a,\tau_0)}_1(\theta)| \leq \sum H_\delta(x_0, \theta_2) \), where the sum is over \( t_i \in (a, \tau_0 \rangle \), which has up to \( B \) terms. It follows that

\[
I_2 \leq p^{(a,\tau_0)}(\theta_2) \exp\left\{ \sum H_\delta(x_0, \theta_2) \right\} \int_{N_2(\delta)} p^{(a,\tau_0)}(\theta) \pi(\theta|\alpha) d\theta
\]

Thus

\[
\frac{I_2}{D(x^{(a,\tau_0)}_0|\alpha)D(x^{(a,\tau_0)}_0|\alpha)} \leq \frac{p^{(a,\tau_0)}(\theta_2) \exp\left\{ \sum H_\delta(x_0, \theta_2) \right\}}{\int_{\Theta} p^{(a,\tau_0)}(\theta) \pi(\theta|\alpha) d\theta} \tag{A.5}
\]

Note that \( p^{(a,\tau_0)}(\theta) \) is a product of up to \( B \) i.i.d. random variables and \( \sum H_\delta(x_0, \theta_2) \) 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). \( \square \)

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

**Proof of Lemma 3.1.** First, let us consider the case of \( m = 1 \).

\[
\frac{P(x|\{0, \tau_1, 1\})}{P(x|\{0, 1\})} = \frac{D(x_{(0,\tau_1)}|\alpha)D(x_{(\tau_1, 1)}|\alpha)}{D(x_{(0,1)}|\alpha)}.
\]

Lemma 2(i) tells us that it is \( O_p\left( n^{1/2}_{(0,1)}/n^{(\tau_1,1)}_{(\tau_1,1)} \right) \). But \( n^{(\tau_1,1)}_{(\tau_1,1)}/n^{2}_{(0,\tau_1)}C(\tau_1,1) \to 1 \) by regularity condition 3, it follows that

\[
\frac{P(x|\{0, \tau_1, 1\})}{P(x|\{0, 1\})} = O_p\left( 1/ \sqrt{nC(\tau_1,1)} \right) = O_p(1/ \sqrt{n\Delta}).
\]

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

\[
\frac{P(x|\{0, \tau_1, \ldots, \tau_M, 1\})}{P(x|\{0, 1\})} = \frac{P(x|\{0, \tau_2, \ldots, \tau_M, 1\})P(x|\{0, \tau_1, \ldots, \tau_M, 1\})}{P(x|\{0, 1\})P(x|\{0, \tau_2, \ldots, \tau_M, 1\})}.
\]
By the induction assumption, \( P(x|\{0, \tau_2, \cdots, \tau_M, 1\})/P(x|\{0, 1\}) \xrightarrow{P} 0. \) Note that
\[
P(x|\{0, \tau_1, \cdots, \tau_M, 1\}) = \frac{D(x_{(0,\tau_1)}|\alpha)D(x_{(\tau_1,\tau_2)}|\alpha)}{D(x_{(0,\tau_2)}|\alpha)}.
\]
 Lemma A.2(i) again tells us that the above expression converges to 0 in probability. Therefore, the lemma is also true for \( m = M + 1: P(x|\{0, \tau_1, \cdots, \tau_M, 1\})/P(x|\{0, 1\}) = O_p(1/\sqrt{n\Delta}) \). □

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

\[
P(x|\{0, 1\}) = \frac{D(x_{(0,\tau_1)}|\alpha)}{D(x_{(0,\tau_1^0)}|\alpha)D(x_{(\tau_1^0,1)}|\alpha)}.
\]

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

\[
\frac{D(x_{(0,\tau_1)}|\alpha)}{D(x_{(0,\tau_1^0)}|\alpha)D(x_{(\tau_1^0,1)}|\alpha)} = O_p(n^{1/2}\exp\{-n(\tau_1^0)k_1(\delta)\}) = O_p(n^{1/2}\exp\{-n(\tau_1^0)k_2(\delta)\}).
\]

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

PROOF of Theorem 3.3. Our proof consists of three steps. Step 1. Let \( E_1 \) be the event that there is at least one true change-point \( \tau_j^0 \) \((0 \leq j \leq m_0)\) that no estimated change-point is within \( \Delta/2 \) of it, i.e., \( \hat{\tau}_j \not\in (\tau_j^0 - \Delta/2, \tau_j^0 + \Delta/2) \) for all \( i \). We will show that the probability of \( E_1 \) goes to 0.

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

\[
\tau_j^0 < \cdots < \tau_j^{r_j} < \cdots < \tau_j^{r_j}.
\]

Consider an alternative choice of change-points

\[
\bar{\tau} = \{\hat{\tau}_0, \hat{\tau}_1, \cdots, \hat{\tau}_i, \tau_j^0, \hat{\tau}_i, \cdots, \hat{\tau}_n\},
\]

which is formed by inserting \( \tau_j^0 < \cdots < \tau_j^{r_j} \) into \( \hat{\tau} \). It is clear that
\[
P(x|\bar{\tau}) = \frac{D(x(\hat{\tau}_i,\hat{\tau}_{i+1})|\alpha)}{D(x(\hat{\tau}_i,\tau_j^0)|\alpha)}D(x(\tau_j^0,\tau_j^0)|\alpha)D(x(\tau_j^0,\tau_j^{r_j})|\alpha)D(x(\tau_j^{r_j},\tau_j^{r_j})|\alpha).
\]

Since \( n(\hat{\tau}_i,\tau_j^0) \to \infty \), \( n(\tau_j^0,\hat{\tau}_{i+1}) \to \infty \) and \( n(\tau_j^{r_j},\tau_j^{r_j}) \to \infty \) (for any \( k \)) by condition 5, it follows from Lemma A.3 that the ratio \( P(x|\bar{\tau})/P(x|\bar{\tau}) \) would go to zero in probability. Another way to look at it is to think of \( \bar{\tau} \) 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 $\hat{\tau}$ goes to zero, i.e., the probability of $\mathcal{E}_1$ goes to zero. This in fact proves equation (3.1), since $\Delta/2 \to 0$ by condition 5.

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

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 $\bigcup_j (\tau^0_j - \Delta/2, \tau^0_j + \Delta/2)$ contains all the estimated change-points. Suppose $\hat{\tau}_i$ is outside the union. Let $\tau^0_j$ and $\tau^0_{j+1}$ be the adjacent true change-points that sandwich $\hat{\tau}_i$: $\tau^0_j < \hat{\tau}_i < \tau^0_{j+1}$. We must have $\hat{\tau}_i - \tau^0_j \geq \Delta/2$ and $\tau^0_{j+1} - \hat{\tau}_i \geq \Delta/2$. From Steps 1 and 2, we know that with probability going to one, there are two estimated change-points of which one is within $\Delta/2$ of $\tau^0_j$ and the other is within $\Delta/2$ of $\tau^0_{j+1}$. Let $\hat{\tau}_{i-l} < \cdots < \hat{\tau}_{i+r}$ be the sequence of estimated change-points $(l, r \geq 0)$ that are between $\tau^0_j$ and $\tau^0_{j+1}$:

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

Consider the following alternative change-points:

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

We can think of $\tilde{\tau}$ as being created by deleting from $\hat{\tau}$ 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 $\hat{\tau}_{k+1} - \hat{\tau}_k \geq \Delta$ by the definition and $n\Delta \to \infty$. It follows that the ratio $P(\mathbf{x}|\hat{\tau})/P(\mathbf{x}|\tilde{\tau})$ would go to zero in probability. Therefore, the probability of having a $\hat{\tau}_i$ outside the union of $\bigcup_j (\tau^0_j - \Delta/2, \tau^0_j + \Delta/2)$ 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

$$D(x_{(a,b)}|\hat{\alpha}_n) \xrightarrow{P} (2\pi)^{1/2}\pi(\theta_0|\alpha^*).$$  \hspace{1cm} (A.6)$$

To do so, let $N(\delta)$ be a neighborhood of $\theta_1$. 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 $\alpha^*$, 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 $\alpha^*$ 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.  \( \square \)

References