Supplementary Materials for "Bayesian Hierarchical Model for Change Point Detection in Multivariate Sequences"

April 26, 2021

A Additional numerical results

A.1 Variance parameter

We assess the impact of using the same ω_s across the sequences in our BHM model. We generate two-dimensional data from model I (homogeneous errors) and model II (heteroscedastic errors). Under each model, the data are generated with the same ($\mathrm{N}_2(\mathbf{0}, \mathbf{I}_2)$) and distinct ($\mathrm{N}_2(\mathbf{0}, \mathbf{\Sigma})$) variance parameters across the sequences, where $\mathbf{I}_2$ is an identity covariance matrix and $\mathbf{\Sigma} = \mathrm{diag}(0.8, 1.2)$. We estimate the change point locations by assuming the two sequences share the same variance or use different variance parameters. The results in Table A.1 show that the two strategies yield similar performances across all scenarios. Hence, we suggest to use the same ω_s in practice which helps to reduce the computational burden and numerical errors as well as facilitating the information borrowing across the sequences.

Table A.1: Comparison results over 500 simulations when using the same or different variance parameters across the sequences with $n = 2$ under model I and model II, respectively. Standard deviations are given in parentheses.

Data-generating Model	Variance Parameter	$\hat{p} - p_0$: ≤ -3	-2	-1	0	1	2	≥ 3	Segmentation Error: $d(\widehat{\mathcal{K}}\|\mathcal{K}_0)$	$d(\mathcal{K}_0\|\widehat{\mathcal{K}})$
I, $(\xi_{1k}, \xi_{2k}) \sim \mathrm{N}_2(\mathbf{0}, \mathbf{I}_2)$	Same	0	0	18	477	5	0	0	2.56 (2.83)	2.59 (3.86)
	Different	0	0	25	473	2	0	0	2.56 (2.83)	2.26 (2.92)
I, $(\xi_{1k}, \xi_{2k}) \sim \mathrm{N}_2(\mathbf{0}, \mathbf{\Sigma})$	Same	0	0	23	474	3	0	0	2.62 (2.91)	2.43 (3.42)
	Different	0	0	25	472	3	0	0	2.62 (2.91)	2.36 (3.21)
II, $(\xi_{1k}, \xi_{2k}) \sim \mathrm{N}_2(\mathbf{0}, \mathbf{I}_2)$	Same	0	0	3	341	135	21	0	1.23 (1.59)	6.32 (7.83)
	Different	0	0	3	345	137	14	0	1.23 (1.59)	6.27 (7.88)
II, $(\xi_{1k}, \xi_{2k}) \sim \mathrm{N}_2(\mathbf{0}, \mathbf{\Sigma})$	Same	0	0	19	479	2	0	0	2.28 (2.74)	2.03 (2.82)
	Different	0	0	19	453	28	0	0	2.28 (2.74)	2.73 (4.07)

A.2 Parameter tuning in the priors of BHM

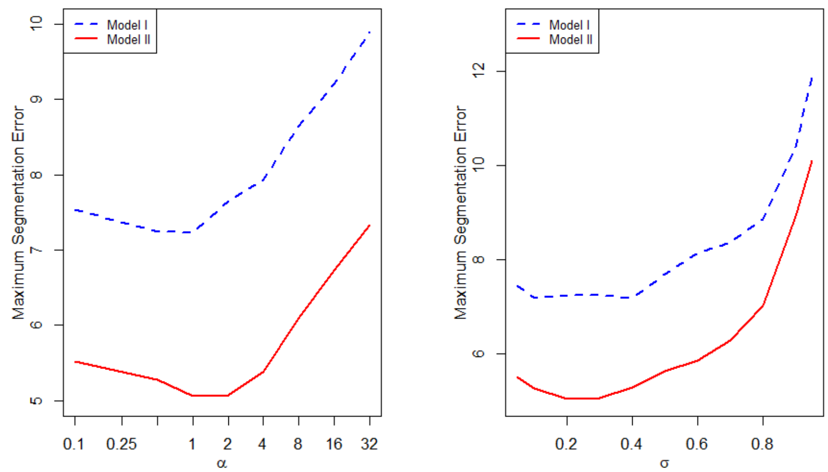

Figure A.1: The maximum segmentation error versus α (left) and σ (right) over 500 simulations with sample size $T = 400$, $n = 2$ under models I and II, respectively.

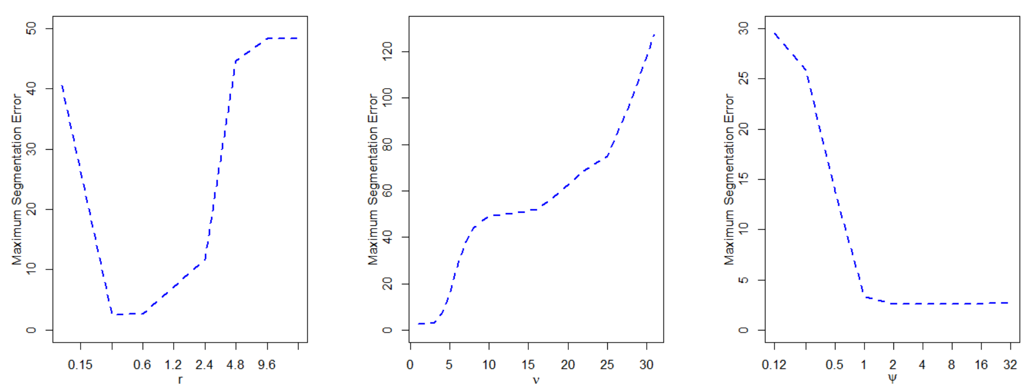

Figure A.2: The maximum segmentation errors versus the corresponding parameters for the inverse moment prior (left), moment prior (middle) and local prior (right) over 500 simulations with sample size $T = 400$, $n = 2$ under model I.

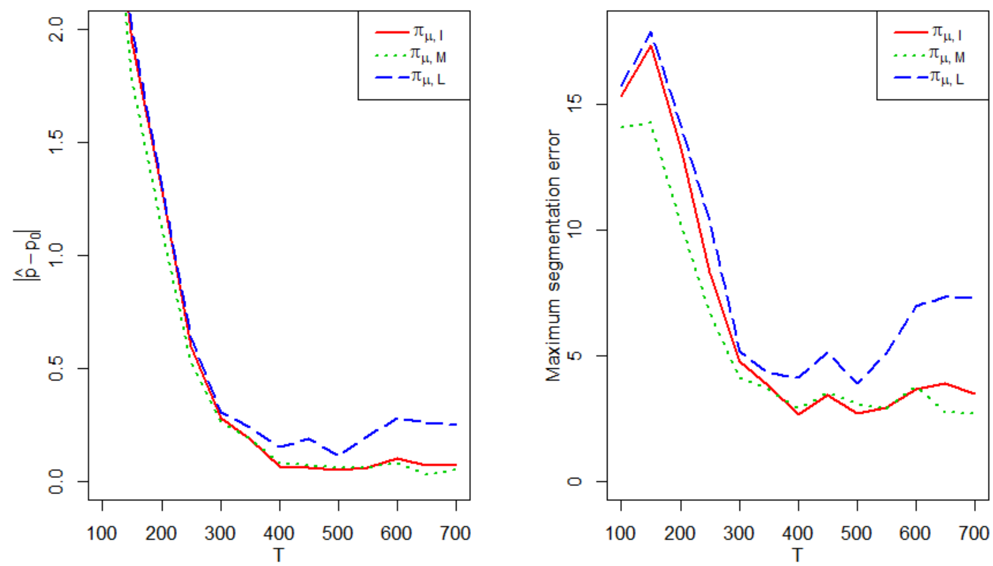

Figure A.3: The absolute difference $|\hat{p} - p_0|$ (left) and the maximum segmentation error (right) versus T over 500 simulations under three priors: the nonlocal inverse moment prior $\pi_{\mu,I}$ with $r = 0.6, \phi = q = 2$, nonlocal moment prior $\pi_{\mu,M}$ with $v = 1$ and local prior $\pi_{\mu,L}$ with $\psi = 2$ under model I with $n = 2$.

A.3 Determination of outliers

In the real data application, we determine the existence of the outliers as follows,

(1) Select a candidate point set $\mathcal{H}(m_I)$ for the dataset with the screening algorithm in the BHM method.

(2) Divide the dataset into segments based on the candidate point set and in this case, we can assume the data points in each segment has homogeneous distribution.

(3) Conduct the generalized extreme studentized deviate (ESD) test (Rosner, 1983) for each segment.

(4) If more than 12% of the segments contain outliers, we adopt the t likelihood; otherwise a normal likelihood is adopted.

In our experiments, this procedure works well as shown in Figure A.4.

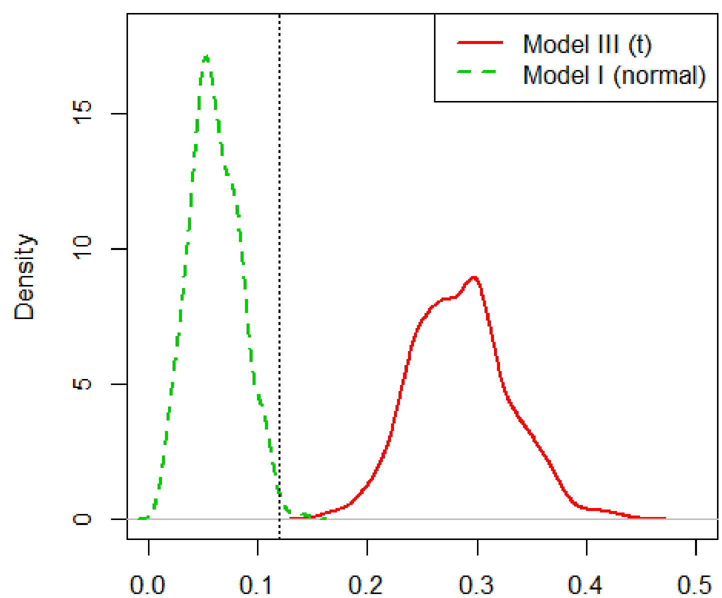

Figure A.4: The densities of proportions of segments containing outliers under model I (normal error) and model III (t error). The black dotted line indicates 12%.

A.4 Autocorrelation in the wind turbine data

While we have shown robustness of the BHM method for the moving-average error (refer to model V of Table 1), we also check the autocorrelation function (ACF) of the real data and simulate the datasets with ACF similar to the real data. The ACFs of dataset 1 in the wind turbine data are shown in Figure A.5. From the plots, it is clear that for the wind turbine data, the first three sequences show significant autocorrelation while the other five sequences are not significantly autocorrelated.

Thus, we conduct the simulation studies with a mixed error model where the first three sequences of the simulated dataset are with autoregressive (AR) errors while the other five sequences follows model I in Section 5. Specifically, we adopt the AR(2) model, i.e., $\xi_{ik} = 0.5\xi_{i,k-1} + 0.2\xi_{i,k-2} + a_k$ with $a_k \sim \mathrm{N}(0, 3/5)$. The ACFs of the simulated dataset are shown in Figure A.6 which has similar patterns to Figure A.5. We use an independent normal likelihood in the BHM method and repeat the simulation for 500 times and the results are presented in Table A.2. Based on the results, the BHM-FIX and BHM-MPP methods still yield satisfactory results under the mixed error datasets. However, the nonparameteric ECP method deteriorates dramatically compared with the results in Table 1.

Table A.2: Comparison results over 500 simulations among BHM-FIX, BHM-MPP, ECP and DPMLE when $n = 8$ under the mixed error model. Standard deviations are given in parentheses.

Data-generating Model	Method	$\hat{p} - p_0$: ≤ -3	-2	-1	0	1	2	≥ 3	Segmentation Error: $d(\widehat{\mathcal{K}}\|\mathcal{K}_0)$	$d(\mathcal{K}_0\|\widehat{\mathcal{K}})$
Mixed errors	BHM-FIX	0	0	0	486	14	0	0	0.08 (0.31)	0.54 (2.80)
	BHM-MPP	0	0	0	484	16	0	0	0.03 (0.18)	0.52 (2.77)
	ECP	0	0	0	24	49	101	326	0.22 (0.53)	20.34 (6.59)
	DPMLE	500	0	0	0	0	0	0	59.79 (15.64)	0.12 (1.23)

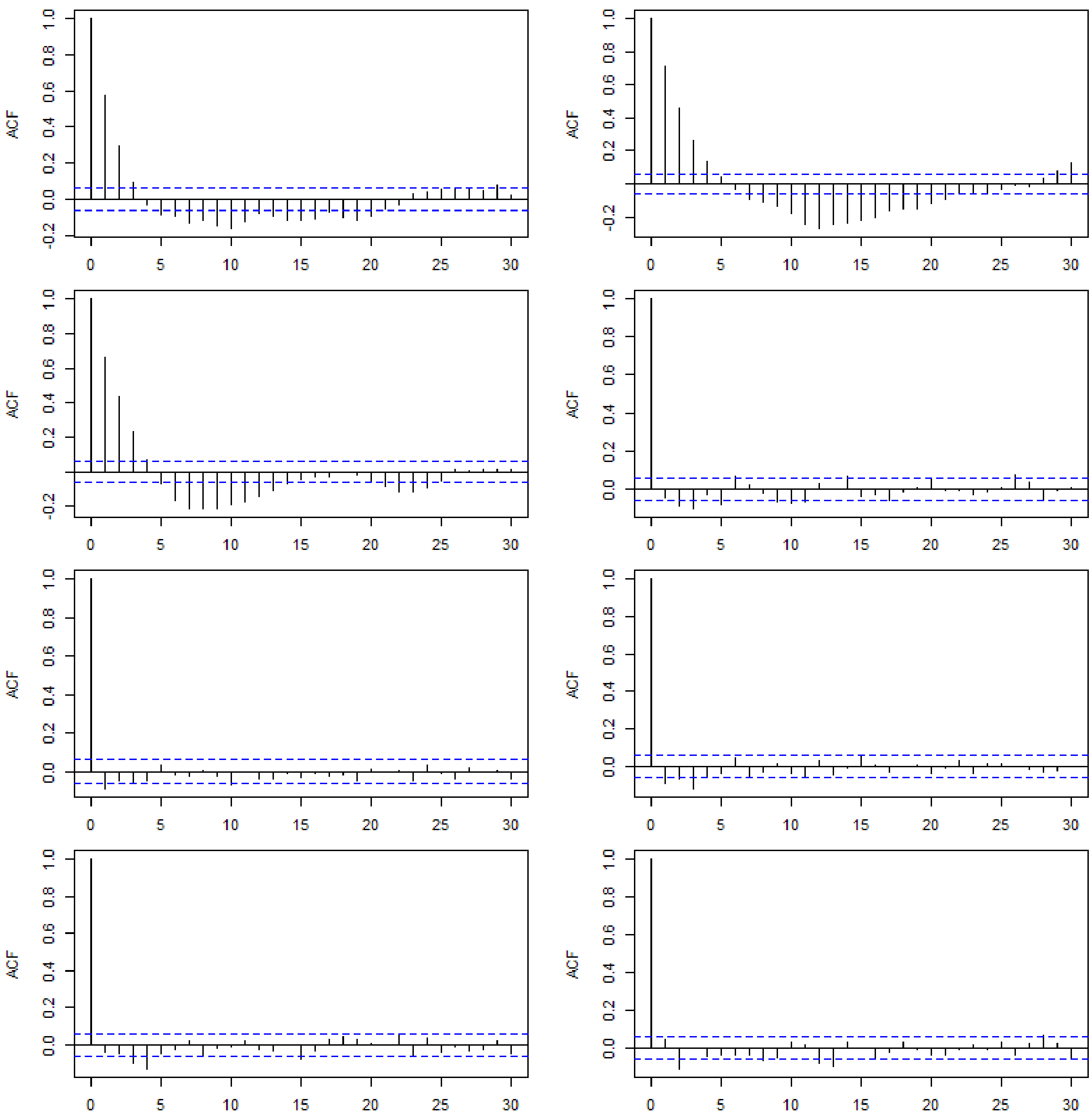

Figure A.5: The autocorrelation functions of dataset 1 in the wind turbine data.

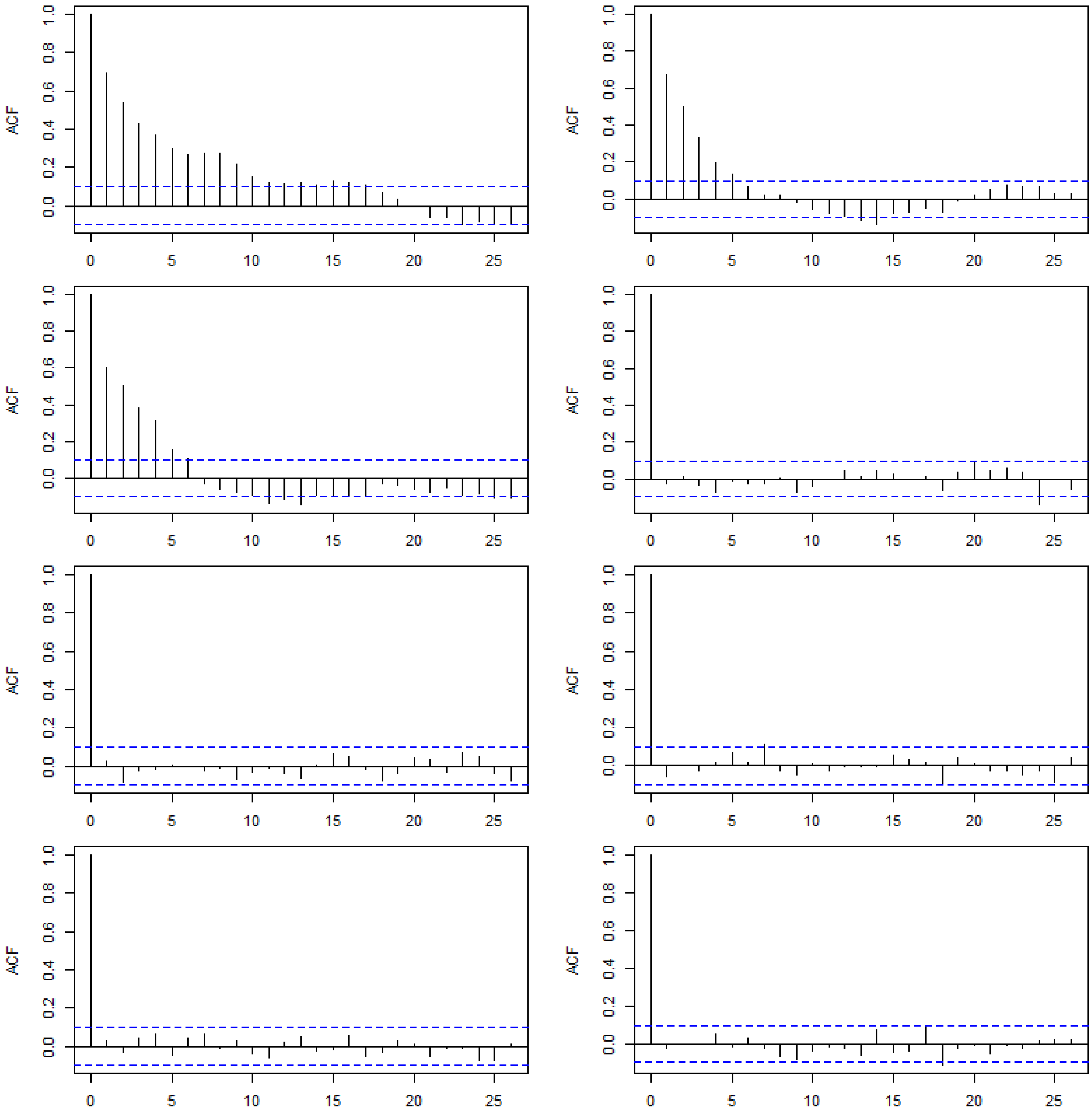

Figure A.6: The autocorrelation functions of the simulated data.

B Dynamic programming

For $l = 0, 1, \ldots, N$, we define

$$H(j|l) \equiv \max_{\mathcal{K} \subseteq \{\tau_0, \tau_1, \ldots, \tau_j, \tau_{j+1}\}, |\mathcal{K}|=l} U(\mathcal{K}|\mathbf{Y}_{(\tau_0, \tau_{j+1}]}).$$

The dynamic programming algorithm is given as Algorithm A.1.

Algorithm A.1 Dynamic programming

Input:

The upper bound of the number of change points M, dataset $\mathbf{Y}$, candidate point set $\mathcal{H}(m_I)$.

Let A be an empty $M \times N$ matrix.
for $i = 0, \ldots, N$ **do**
  $H(i|0) \leftarrow u(\mathbf{Y}_{(\tau_0, \tau_{i+1}]}, s = 0)$
end for
for $l = 1, \ldots, M$ **do**
  for $i = l, \ldots, N$ **do**
    $A_{l,i} \leftarrow \text{argmax}_{l-1 \le k \le i-1} \{H(k|l-1) u(\mathbf{Y}_{(\tau_{k+1}, \tau_{i+1}]}, s = l)\}$
    $H(i|l) \leftarrow \max_{l-1 \le k \le i-1} \{H(k|l-1) u(\mathbf{Y}_{(\tau_{k+1}, \tau_{i+1}]}, s = l)\}$
  end for
end for
$\hat{p} \leftarrow \text{argmax}_{l=0,1,\ldots,M} H(N|l)$
if $\hat{p} = 0$ **then**
  return $\varnothing$
else
  $s \leftarrow \hat{p}$
  $t \leftarrow N$
  $E \leftarrow \varnothing$
  while $s \neq 0$ **do**
    $E \leftarrow E \cup \{A_{s,t} + 1\}$
    $s \leftarrow s - 1$
    $t \leftarrow A_{s,t}$
  end while
  return $\widehat{\mathcal{K}} = \{\tau_i, i \in E\}$
end if

Output:

Estimated change point set $\widehat{\mathcal{K}}$.

C Proofs

We denote $\mathcal{K}_0$ as the true change point set with p_0 change points, $\widehat{\mathcal{K}}$ as the estimated change point set with $\hat{p}$ estimated change points. The m_I-neighbourhood of a time point $\mathbf{Y}_k$ is defined as $\{\mathbf{Y}_l : l \in (k - m_I, k + m_I)\}$. Given an interval $\mathbf{Y}_{(a,b]}$, we denote $p_{(a,b]}(\boldsymbol{\theta}) = \prod_{k\in(a,b]} f(\mathbf{Y}_k|\boldsymbol{\theta})$ as the likelihood function, the corresponding log-likelihood function is $l_{(a,b]}(\boldsymbol{\theta}) = \log p_{(a,b]}(\boldsymbol{\theta})$. We also let $\widehat{\boldsymbol{\theta}}_{(a,b]}$ be the maximum likelihood estimator (MLE) based on $l_{(a,b]}(\boldsymbol{\theta})$, and $\boldsymbol{\theta}_{(a,b]}$ be the true parameters on $(a, b]$. We denote $\hat{\sigma}^2_{(a,b]} = \{-E(\frac{\partial^2 l_{(a,b]}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top})|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_{(a,b]}}\}^{-1}$ and let $J(\boldsymbol{\theta}_0) = -E\frac{\partial^2 \log f(\mathbf{Y}_1|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ be the Fisher information for one observation. We also define d as the dimension of $\boldsymbol{\theta}$. Finally, denote

$$C(\mathbf{Y}_{(\kappa_s,\kappa_{s+1}]}) = \int p_{(\kappa_s,\kappa_{s+1}]}(\boldsymbol{\theta}_s)\pi(\boldsymbol{\theta}_s)d\boldsymbol{\theta}_s.$$

We list the regularity conditions as follows.

(1) The prior for mean difference $\pi_\mu(\mu)$ is continuous with bounded first and second derivatives.

(2) For a segment between two true change points κ_{0j} and $\kappa_{0,j+1}$ $(j = 1, \ldots, p_0)$ with parameters $(\mu_{1,j}, \ldots, \mu_{n,j})$, there exists $\delta_I > 0$ such that for any $i \in \{1, \ldots, n\}$, $|\mu_{i,j}|$ is either greater than δ_I or equal to 0. Further, there is $i \in \{1, \ldots, n\}$ such that $|\mu_{i,j}| > \delta_I$.

(3) The generic prior $\pi(\boldsymbol{\theta})$ is continuous and positive at all $\boldsymbol{\theta}_i$ $(i = 0, \ldots, p_0)$, where $\boldsymbol{\theta}_i$ is the true parameters for interval $(\kappa_{0i}, \kappa_{0,i+1}]$.

(4) The regularity conditions (A1)–(A5) and (B1)–(B4).

Regularity conditions (A1)–(A5) and (B1)–(B4) are listed as follows. All the conditions are multivariate extensions from (Du et al., 2016).

(A1) $\boldsymbol{\Theta}$ is a closed set, and $\boldsymbol{\Theta} \subseteq \mathcal{R}^d$.

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

(A3) If $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$ are two distinct points in $\boldsymbol{\Theta}$, then the Lebesgue measure of $\mu\{\mathbf{x} : f(\mathbf{x}|\boldsymbol{\theta}_1) \neq f(\mathbf{x}|\boldsymbol{\theta}_2)\} > 0$.

(A4) Let $\mathbf{x} \in \mathcal{X}$, $\boldsymbol{\theta}' \in \boldsymbol{\Theta}$. Then for all $\boldsymbol{\theta}$ such that $\|\boldsymbol{\theta} - \boldsymbol{\theta}'\| < \delta$, where $\|\cdot\|$ is the L_2 norm, with δ sufficiently small,

$$|\log f(\mathbf{x}|\boldsymbol{\theta}) - \log f(\mathbf{x}|\boldsymbol{\theta}')| < H_\delta(\mathbf{x}, \boldsymbol{\theta}'),$$

where

$$\lim_{\delta\to 0} H_\delta(\mathbf{x}, \boldsymbol{\theta}') = 0,$$

and for any $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$,

$$\lim_{\delta\to 0} \int_{\mathcal{X}} H_\delta(\mathbf{x}, \boldsymbol{\theta}') f(\mathbf{x}|\boldsymbol{\theta}_0) d\mu = 0.$$

(A5) If $\mathbf{\Theta}$ is not bounded, then for any $\boldsymbol{\theta}_0 \in \mathbf{\Theta}$, and sufficiently large Δ,

$$\log f(\mathbf{x}|\boldsymbol{\theta}) - \log f(\mathbf{x}|\boldsymbol{\theta}_0) < K_\Delta(\mathbf{x}, \boldsymbol{\theta}_0),$$

whenever $\|\boldsymbol{\theta}\| > \Delta$, where

$$\lim_{\Delta \to \infty} \int_{\mathcal{X}} K_\Delta(\mathbf{x}, \boldsymbol{\theta}_0) f(\mathbf{x}|\boldsymbol{\theta}_0) d\mu < 0.$$

(B1) $\log f(\mathbf{x}|\boldsymbol{\theta})$ is twice differentiable with respect to $\boldsymbol{\theta}$ in some neighborhood of $\boldsymbol{\theta}_0$.

(B2) Let

$$J(\boldsymbol{\theta}_0) = \left[\int_{\mathcal{X}} f_0 \frac{\partial \log f_0}{\partial \theta_{0i}} \frac{\partial \log f_0}{\partial \theta_{0j}} d\mu \right]_{i,j=1}^{d},$$

where f_0 denotes $f(\mathbf{x}|\boldsymbol{\theta}_0)$, θ_{0i} is the ith element of $\boldsymbol{\theta}_0$. Then $J(\boldsymbol{\theta}_0)$ is positive definite.

(B3) For any $1 \leq i, j \leq d$,

$$\int_{\mathcal{X}} \frac{\partial f_0}{\partial \theta_{0i}} d\mu = \int_{\mathcal{X}} \frac{\partial^2 f_0}{\partial \theta_{0i} \partial \theta_{0j}} d\mu = 0.$$

(B4) For $\delta > 0$, if $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < \delta$, where δ is small enough, then

$$\left\| \frac{\partial^2 \log f(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} - \frac{\partial^2 \log f(\mathbf{x}|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right\| < M_\delta(\mathbf{x}, \boldsymbol{\theta}_0),$$

where $\lim_{\delta \to 0} \int M_\delta(\mathbf{x}, \boldsymbol{\theta}_0) f(\mathbf{x}|\boldsymbol{\theta}_0) d\mu = 0$.

Proof of Lemma 1:

We define j as an m_I-flat point if there is no change point in $(j - m_I, j + m_I)$. Let $\mathcal{F}$ be the set of all m_I-flat points. So $|\mathcal{F}| = T - p_0(2m_I - 1)$, where $|\mathcal{F}|$ denotes the cardinality of set $\mathcal{F}$. To prove Lemma 1, it is sufficient to show

$$\Pr\left(\min_{k \in \mathcal{K}_0} R_k > \max_{l \in \mathcal{F}} R_l \right) \to 1,$$

as $T \to \infty$. Note that

$$\Pr\left(\min_{k \in \mathcal{K}_0} R_k > \max_{l \in \mathcal{F}} R_l \right) \geq \Pr\left(\min_{k \in \mathcal{K}_0} R_k > b_T > \max_{l \in \mathcal{F}} R_t \right),$$

where b_T is a positive sequence with respect to T. It follows that

$$\begin{aligned} & \Pr\left(\min_{k \in \mathcal{K}_0} R_k > b_T > \max_{l \in \mathcal{F}} R_l \right) \\ = \; & \Pr\left\{ (\cap_{k \in \mathcal{K}_0} \{R_k > b_T\}) \cap (\cap_{l \in \mathcal{F}} \{R_l < b_T\}) \right\} \\ = \; & 1 - \Pr\left\{ (\cup_{k \in \mathcal{K}_0} \{R_k \leq b_T\}) \cup (\cup_{l \in \mathcal{F}} \{R_l \geq b_T\}) \right\} \\ \geq \; & 1 - \left\{ \Pr\left(\cup_{k \in \mathcal{K}_0} \{R_k \leq b_T\} \right) + \Pr\left(\cup_{l \in \mathcal{F}} \{R_l \geq b_T\} \right) \right\} \\ \geq \; & 1 - \left\{ \sum_{k \in \mathcal{K}_0} \Pr\left(\{R_k \leq b_T\} \right) + \sum_{l \in \mathcal{F}} \Pr\left(\{R_l \geq b_T\} \right) \right\}. \end{aligned}$$

We define

$$R_{ij} = \frac{\int \prod_{l=j+1}^{j+m_I} \exp\{-(Y_{il} - \overline{Y}_{ij} - \mu)^2\}\pi(\mu)d\mu}{\prod_{l=j+1}^{j+m_I} \exp\{-(Y_{il} - \overline{Y}_{ij})^2\}}, \text{for } i = 1, \ldots, n,$$

where $\overline{Y}_{ij} = m_I^{-1} \sum_{l=j-m_I+1}^{j} Y_{il}$. Clearly, $R_j = \prod_{i=1}^{n} R_{ij}$.

For any change point $k \in \mathcal{K}_0$, assume n_x sequences have mean shifts at this change point. By regularity condition (2), we know $n_x \geq 1$ and the absolute change of mean is greater than δ_I.

Without loss of generality, assume the first n_x sequences have mean changes. By Lemma 1 of Jiang, Yin, and Dominici (2018), we have

$$\lim_{T\to\infty} \Pr\{R_{ik} > \exp(Dm_I\delta_I)\} = 1, \tag{1}$$

when there is a mean shift in sequence i at change point k, where $D > 0$ is a constant. Then we set

$$b_T = \exp(D\delta_I m_I/2).$$

For any $l \in \mathcal{F}$, by Lemmas 2, 3, 4 of Jiang, Yin, and Dominici (2018), there exist $c, C > 0$ such that

$$ca_T \leq R_{il} \leq Ca_T, \tag{2}$$

so

$$R_l = \prod_{i=1}^{n} R_{il} = O_p(a_T^n),$$

where $a_T = m_I^{-1/2}$, $m_I^{-v-1/2}$ and $\exp(-m_I^{s/(s+1)})$ correspond to the local prior, moment prior and inverse moment prior. Consequently, we have

$$\begin{aligned} \Pr(R_l \geq b_T) &= O\{a_T^n \exp(-D\delta_I m_I/2)\}, \\ \sum_{l:t_l\in\mathcal{F}} \Pr(R_l \geq b_T) &= O\{Ta_T^n \exp(-D\delta_I m_I/2)\} = o(1), \end{aligned} \tag{3}$$

since $m_I/(\log T)^{1+\epsilon} \to c > 0$.

Next, for $k \in \mathcal{K}_0$, by (1), we know for $i = 1, \ldots, n_x$, we have

$$\lim_{T\to\infty} \Pr\{R_{ik} > \exp(Dm_I\delta_I)\} = 1.$$

As a result,

$$\lim_{T\to\infty} \Pr\left\{\prod_{i=1}^{n_x} R_{ik} > \exp(n_x Dm_I\delta_I)\right\} = 1. \tag{4}$$

Consequently, we obtain

$$
\begin{aligned}
& \Pr(R_k \le b_T) \\
= \; & \Pr\left(\prod_{i=1}^{n_x} R_{ik} \prod_{j=n_x+1}^{n} R_{jk} \le b_T\right) \\
= \; & \Pr\left\{\prod_{i=1}^{n_x} R_{ik} \prod_{j=n_x+1}^{n} R_{jk} \le b_T, \prod_{i=1}^{n_x} R_{ik} > \exp(n_x D \delta_I m_I)\right\} \\
& + \Pr\left\{\prod_{i=1}^{n_x} R_{ik} \prod_{j=n_x+1}^{n} R_{jk} \le b_T, \prod_{i=1}^{n_x} R_{ik} \le \exp(n_x D \delta_I m_I)\right\} \\
\le \; & \Pr\left\{\exp(n_x D \delta_I m_I) \prod_{j=n_x+1}^{n} R_{jk} \le b_T\right\} + \Pr\left\{\prod_{i=1}^{n_x} R_{ik} \le \exp(n_x D \delta_I m_I)\right\}.
\end{aligned}
$$

Combining with (4),

$$
\lim_{T\to\infty} \Pr(R_k \le b_T) \le \lim_{T\to\infty} \Pr\left\{\exp(n_x D m_I \delta_I) \prod_{j=n_x+1}^{n} R_{jk} \le b_T\right\}.
$$

For $j = n_x + 1, \ldots, n$, by (2), $\exists c_1, C_1 > 0$ such that

$$
c_1 a_T^{n-n_x} \le \prod_{j=n_x+1}^{n} R_{jk} \le C_1 a_T^{n-n_x},
$$

and

$$
C_1^{-1} a_T^{n_x-n} \le \left(\prod_{j=n_x+1}^{n} R_{jk}\right)^{-1} \le c_1^{-1} a_T^{n_x-n}.
$$

This implies

$$
\begin{aligned}
& \Pr\left\{\exp(n_x D m_I \delta_I) \prod_{j=n_x+1}^{n} R_{jk} \le b_T\right\} \\
= \; & \Pr\left\{\left(\prod_{j=n_x+1}^{n} R_{jk}\right)^{-1} \ge \exp(n_x D m_I \delta_I) b_T^{-1}\right\} \\
\le \; & \frac{c_1^{-1} a_T^{n_x-n}}{\exp(n_x D m_I \delta_I) b_T^{-1}} \\
= \; & c_1^{-1} a_T^{n_x-n} \exp(-n_x D m_I \delta_I) b_T,
\end{aligned}
$$

where the second to the last inequality holds by the Markov inequality. Therefore

$$
\Pr(R_k \le b_T) = O\{a_T^{n_x-n} \exp(-n_x D m_I \delta_I) b_T\}.
$$

Thus, we obtain

$$\sum_{k\in\mathcal{K}_0} \Pr\left(R_k \le b_T\right) = O\left[p_0 a_T^{n_x-n}\exp\{-(n_x-1/2)Dm_I\delta_I\}\right] = o(1), \tag{5}$$

since $m_I/(\log T)^{1+\epsilon} \to c > 0$. Then using (3) and (5), we achieve:

$$\begin{aligned}
& \Pr\left(\min_{k\in\mathcal{K}_0} R_k > \max_{l\in\mathcal{F}} R_l\right) \\
\ge\ & \Pr\left(\min_{k\in\mathcal{K}_0} R_k > b_T > \max_{l\in\mathcal{F}} R_l\right) \\
=\ & \Pr\left\{(\cap_{k\in\mathcal{K}_0}\{R_k > b_T\}) \cap (\cap_{l\in\mathcal{F}}\{R_l < b_T\})\right\} \\
\ge\ & 1 - \left\{\sum_{k\in\mathcal{K}_0}\Pr\left(\{R_k \le b_T\}\right) + \sum_{l\in\mathcal{F}}\Pr\left(\{R_l \ge b_T\}\right)\right\} \\
=\ & 1 - o(1).
\end{aligned}$$

Finally, we know

$$\Pr\left(\min_{k\in\mathcal{K}_0} R_k > \max_{l\in\mathcal{F}} R_l\right) \longrightarrow 1,$$

as $T \to \infty$. □

Lemma 3. *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 $\boldsymbol{\theta}_{(a,b]}$, then as $(b-a)\to\infty$,*

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

$$\lim_{(b-a)\to\infty} \Pr\left\{\sup_{\boldsymbol{\theta}\notin N_0(\delta)} \frac{l_{(a,b]}(\boldsymbol{\theta}) - l_{(a,b]}(\boldsymbol{\theta}_{(a,b]})}{b-a} < -k_{\boldsymbol{\theta}_{(a,b]}}(\delta)\right\} = 1;$$

2. $l_{(a,b]}(\boldsymbol{\theta}_{(a,b]}) - l_{(a,b]}(\widehat{\boldsymbol{\theta}}_{(a,b]}) = O_p(1)$.

Proof:

The proof of Lemma 3 is a direct multi-dimensional extension from Theorem 1 of Walker (1969). □

The following result is Theorem 3.1 of Fraser and Mcdunnough (1984), and the regularity conditions (A1)–(A5) and (B1)–(B4) imply the three assumptions in Fraser and Mcdunnough (1984).

Lemma 4. *Suppose conditions (A1)–(A5) and (B1)–(B4) hold. For i.i.d samples $\{\mathbf{Y}_1,\ldots,\mathbf{Y}_T\}$ from $f(\mathbf{Y}|\boldsymbol{\theta}_0)$, let $\hat{\sigma}^2 = \{-E(\frac{\partial^2 log p(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top})\big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}\}^{-1}$ and $p(\boldsymbol{\theta}) = \prod_{k:t_k\in(0,1]} f(\mathbf{Y}_k|\boldsymbol{\theta})$, where $\widehat{\boldsymbol{\theta}}$ is* *the MLE of* $\boldsymbol{\theta}_0$. *If* $w(\boldsymbol{\theta}) \geq 0$ *for all* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ *and satisfies* $\int w(\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta} < \infty$ *and is continuous and nonzero at the true* $\boldsymbol{\theta}_0$, *then*

$$\frac{\det(\hat{\sigma})w(\widehat{\boldsymbol{\theta}})p(\widehat{\boldsymbol{\theta}})}{\int w(\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}} \xrightarrow{a.s.} (2\pi)^{-d/2}.$$

Lemma 5. *Assume conditions in Theorem 1 hold. Suppose that there are* r *change points in* (a,b), *say* $\{\kappa_1,\ldots,\kappa_r\}$, *with* $\kappa_1 < \ldots < \kappa_r$. *Further assume* $(\kappa_{i+1}-\kappa_i) \to \infty, i = 0,\ldots,r$ *(let* $\kappa_0 = a, \kappa_{r+1} = b$*) as* $(b-a) \to \infty$. *Then let* $\underline{\kappa} = \min_{i=0,\ldots,r}(\kappa_{i+1}-\kappa_i)$, $\exists c_2 > 0$ *such that*

$$\frac{C(\mathbf{Y}_{(a,b]})}{C(\mathbf{Y}_{(a,\kappa_1]})\cdots C(\mathbf{Y}_{(\kappa_r,b]})} = O_p\left\{(b-a)^{rd/2}\exp(-\underline{\kappa}c_2)\right\}.$$

Proof:

The r change points separate the sequences into $r+1$ segments. We first assume all the $r+1$ segments have different parameters denoted as $\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_{r+1}$. Then we can find a δ and define $N_i(\delta) = \{\boldsymbol{\theta} : \|\boldsymbol{\theta}-\boldsymbol{\theta}_i\| < \delta\}, i = 1,\ldots,r+1$ such that $N_i(\delta)\cap N_j(\delta) = \varnothing$ for $i \neq j$. We write

$$C(\mathbf{Y}_{(a,b]}) = \sum_{i=0}^{r+1} I_i$$

where

$$\begin{aligned} I_i &= \int_{N_i(\delta)} p_{(a,b]}(\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}, \quad \text{for} \quad i = 1,\ldots,r+1, \\ I_0 &= \int_{\boldsymbol{\Theta}-\cup_{i=1}^{r+1}N_i(\delta)} p_{(a,b]}(\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}. \end{aligned}$$

By Lemma 4, we have

$$\begin{aligned} C(\mathbf{Y}_{(\kappa_{i-1},\kappa_i]}) &= p_{(\kappa_{i-1},\kappa_i]}(\hat{\boldsymbol{\theta}}_{(\kappa_{i-1},\kappa_i]})\pi(\widehat{\boldsymbol{\theta}}_{(\kappa_{i-1},\kappa_i]})\det(\hat{\sigma}_{(\kappa_{i-1},\kappa_i]})O_p(1) \\ &\neq p_{(\kappa_{i-1},\kappa_i]}(\hat{\boldsymbol{\theta}}_{(\kappa_{i-1},\kappa_i]})\pi(\widehat{\boldsymbol{\theta}}_{(\kappa_{i-1},\kappa_i]})\det(\hat{\sigma}_{(\kappa_{i-1},\kappa_i]})o_p(1). \end{aligned} \tag{6}$$

where $i = 1,\ldots,r+1$. Note that by definition of $\hat{\sigma}_{(\kappa_{i-1},\kappa_i]}$,

$$\begin{aligned} \det(\hat{\sigma}_{(\kappa_{i-1},\kappa_i]}) &= O_p\{(\kappa_i-\kappa_{i-1})^{-d/2}\}, \\ \det(\hat{\sigma}_{(\kappa_{i-1},\kappa_i]}) &\neq o_p\{(\kappa_i-\kappa_{i-1})^{-d/2}\}. \end{aligned} \tag{7}$$

By (6), for $j \neq 0$, we obtain

$$
\begin{aligned}
& \frac{I_j}{C(\mathbf{Y}_{(a,\kappa_1]}) \cdots C(\mathbf{Y}_{(\kappa_r,b]})} \\
= & \frac{\int_{N_j(\delta)} \pi(\boldsymbol{\theta}) p_{(\kappa_0,\kappa_1]}(\boldsymbol{\theta}) \cdots p_{(\kappa_r,\kappa_{r+1}]}(\boldsymbol{\theta}) d\boldsymbol{\theta}}{C(\mathbf{Y}_{(a,\kappa_1]}) \cdots C(\mathbf{Y}_{(\kappa_r,b]})} \\
= & \frac{O_p(1) \int_{N_j(\delta)} \pi(\boldsymbol{\theta}) p_{(\kappa_0,\kappa_1]}(\boldsymbol{\theta}) \cdots p_{(\kappa_r,\kappa_{r+1}]}(\boldsymbol{\theta}) d\boldsymbol{\theta}}{C(\mathbf{Y}_{(\kappa_{j-1},\kappa_j]}) \prod_{i\neq j} p_{(\kappa_{i-1},\kappa_i]}(\hat{\boldsymbol{\theta}}_{(\kappa_{i-1},\kappa_i]}) \pi(\widehat{\boldsymbol{\theta}}_{(\kappa_{i-1},\kappa_i]}) \det(\hat{\sigma}_{(\kappa_{i-1},\kappa_i]})} \\
= & \frac{O_p(1) \int_{N_j(\delta)} \pi(\boldsymbol{\theta}) p_{(\kappa_0,\kappa_1]}(\boldsymbol{\theta}) \cdots p_{(\kappa_r,\kappa_{r+1}]}(\boldsymbol{\theta}) d\boldsymbol{\theta}}{C(\mathbf{Y}_{(\kappa_{j-1},\kappa_j]}) \prod_{i\neq j} p_{(\kappa_{i-1},\kappa_i]}(\boldsymbol{\theta}_i) \pi(\widehat{\boldsymbol{\theta}}_{(\kappa_{i-1},\kappa_i]}) \det(\hat{\sigma}_{(\kappa_{i-1},\kappa_i]})} \\
= & \frac{\int_{N_j(\delta)} \pi(\boldsymbol{\theta}) p_{(\kappa_0,\kappa_1]}(\boldsymbol{\theta}) \cdots p_{(\kappa_r,\kappa_{r+1}]}(\boldsymbol{\theta}) d\boldsymbol{\theta}}{C(\mathbf{Y}_{(\kappa_{j-1},\kappa_j]}) \prod_{i\neq j} p_{(\kappa_{i-1},\kappa_i]}(\boldsymbol{\theta}_i) \pi(\widehat{\boldsymbol{\theta}}_{(\kappa_{i-1},\kappa_i]}) O_p\{(\kappa_i - \kappa_{i-1})^{-d/2}\}} \\
= & \frac{\int_{N_j(\delta)} \pi(\boldsymbol{\theta}) p_{(\kappa_0,\kappa_1]}(\boldsymbol{\theta}) \cdots p_{(\kappa_r,\kappa_{r+1}]}(\boldsymbol{\theta}) d\boldsymbol{\theta}}{C(\mathbf{Y}_{(\kappa_{j-1},\kappa_j]}) \prod_{i\neq j} p_{(\kappa_{i-1},\kappa_i]}(\boldsymbol{\theta}_i) \pi(\boldsymbol{\theta}_i) O_p\{(\kappa_i - \kappa_{i-1})^{-d/2}\}}, \qquad (8)
\end{aligned}
$$

where the third equality in the above equation is due to the second result of Lemma 3, the fourth is by (7), and the last one is by the continuous mapping theorem.
Using the first result of Lemma 3, $\exists k(\delta) > 0$ such that

$$
\begin{aligned}
& \frac{\int_{N_j(\delta)} \pi(\boldsymbol{\theta}) p_{(\kappa_0,\kappa_1]}(\boldsymbol{\theta}) \cdots p_{(\kappa_r,\kappa_{r+1}]}(\boldsymbol{\theta}) d\boldsymbol{\theta}}{C(\mathbf{Y}_{(\kappa_{j-1},\kappa_j]}) \prod_{i\neq j} p_{(\kappa_{i-1},\kappa_i]}(\boldsymbol{\theta}_i) \pi(\boldsymbol{\theta}_i)} \\
= & \frac{1}{C(\mathbf{Y}_{(\kappa_{j-1},\kappa_j]})} \int_{N_j(\delta)} \pi(\boldsymbol{\theta}) p_{(\kappa_{j-1},\kappa_j]}(\boldsymbol{\theta}) \prod_{i\neq j} \exp\{l_{(\kappa_{i-1},\kappa_i]}(\boldsymbol{\theta}) - l_{(\kappa_{i-1},\kappa_i]}(\boldsymbol{\theta}_i)\} d\boldsymbol{\theta} \\
< & \frac{1}{C(\mathbf{Y}_{(\kappa_{j-1},\kappa_j]})} \prod_{i\neq j} \exp\{-(\kappa_i - \kappa_{i-1}) k(\delta)\} \int_{N_j(\delta)} p_{(\kappa_{j-1},\kappa_j]}(\boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
\leq & \frac{1}{C(\mathbf{Y}_{(\kappa_{j-1},\kappa_j]})} \prod_{i\neq j} \exp\{-(\kappa_i - \kappa_{i-1}) k(\delta)\} \int_{\boldsymbol{\Theta}} p_{(\kappa_{j-1},\kappa_j]}(\boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
= & \prod_{i\neq j} \exp\{-(\kappa_i - \kappa_{i-1}) k(\delta)\} \qquad (9)
\end{aligned}
$$

with probability tending to unit as $(b-a) \to \infty$. Combining (8) and (9), we achieve

$$
\begin{aligned}
& \frac{I_j}{C(\mathbf{Y}_{(a,\kappa_1]}) \cdots C(\mathbf{Y}_{(\kappa_r,b]})} \\
= & O_p\left[\prod_{i\neq j} (\kappa_i - \kappa_{i-1})^{d/2} \exp\{-(\kappa_i - \kappa_{i-1}) k(\delta)\}\right] \\
= & O_p\left\{(b-a)^{rd/2} \exp(-\underline{\kappa} k(\delta))\right\}.
\end{aligned}
$$

For I_0, we apply the same argument, but note that the region $\boldsymbol{\Theta} - \cup_{i=1}^{r+1} N_i(\delta)$ does not contain the neighborhood of any $\boldsymbol{\theta}_i$, so

$$\frac{I_0}{C(\mathbf{Y}_{(a,\kappa_1]})\cdots C(\mathbf{Y}_{(\kappa_r,b]})}$$

would have a faster convergence rate compared with

$$\frac{I_j}{C(\mathbf{Y}_{(a,\kappa_1]})\cdots C(\mathbf{Y}_{(\kappa_r,b]})}.$$

Thus we achieve

$$\frac{C(\mathbf{Y}_{(a,b]})}{C(\mathbf{Y}_{(a,\kappa_1]})\cdots C(\mathbf{Y}_{(\kappa_r,b]})} = \frac{\sum_{i=0}^{r+1} I_i}{C(\mathbf{Y}_{(a,\kappa_1]})\cdots C(\mathbf{Y}_{(\kappa_r,b]})} = O_p\left\{(b-a)^{rd/2}\exp(-\underline{\kappa}k(\delta))\right\}.$$

If some segments share the same parameters, without loss of generality, we assume only $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_3$ then $N_1(\delta) = N_3(\delta)$. The argument is analogous when more than two segments share the same parameters.

For $j \neq 1$ or 3, the argument is identical to the above. When $j = 1$ (and there is no I_3, since $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_3$), following similar discussions for (8) and (9),

$$\begin{aligned}
&\frac{I_1}{C(\mathbf{Y}_{(a,\kappa_1]})\cdots C(\mathbf{Y}_{(\kappa_r,b]})} \\
= \;& \frac{\int_{N_j(\delta)} \pi(\boldsymbol{\theta}) p_{(\kappa_0,\kappa_1]}(\boldsymbol{\theta})\cdots p_{(\kappa_r,\kappa_{r+1}]}(\boldsymbol{\theta}) d\boldsymbol{\theta}}{C(\mathbf{Y}_{(\kappa_0,\kappa_1]})C(\mathbf{Y}_{(\kappa_2,\kappa_3]})\prod_{i\neq 1,3} p_{(\kappa_{i-1},\kappa_i]}(\boldsymbol{\theta}_i)\pi(\boldsymbol{\theta}_i) O_p\{(\kappa_i-\kappa_{i-1})^{-d/2}\}} \\
\leq \;& \frac{\int_{\boldsymbol{\Theta}} p_{(\kappa_0,\kappa_1]}(\boldsymbol{\theta}) p_{(\kappa_2,\kappa_3]}(\boldsymbol{\theta})\pi(\boldsymbol{\theta}) d\boldsymbol{\theta}}{C(\mathbf{Y}_{(\kappa_0,\kappa_1]})C(\mathbf{Y}_{(\kappa_2,\kappa_3]})} \prod_{i\neq 1,3} \exp\{-(\kappa_i-\kappa_{i-1})k(\delta)\} O_p\{(\kappa_i-\kappa_{i-1})^{d/2}\}. \quad (10)
\end{aligned}$$

Using similar discussion for (15),

$$\frac{\int_{\boldsymbol{\Theta}} p_{(\kappa_0,\kappa_1]}(\boldsymbol{\theta}) p_{(\kappa_2,\kappa_3]}(\boldsymbol{\theta})\pi(\boldsymbol{\theta}) d\boldsymbol{\theta}}{C(\mathbf{Y}_{(\kappa_0,\kappa_1]})C(\mathbf{Y}_{(\kappa_2,\kappa_3]})} = O_p\left\{\frac{(\kappa_3-\kappa_2)(\kappa_1-\kappa_0)}{(\kappa_3-\kappa_2)+(\kappa_1-\kappa_0)}\right\}^{d/2}. \quad (11)$$

Combining (10) and (11), we achieve

$$\begin{aligned}
&\frac{I_1}{C(\mathbf{Y}_{(a,\kappa_1]})\cdots C(\mathbf{Y}_{(\kappa_r,b]})} \\
= \;& O_p\left\{\frac{\prod_{i=1}^{r+1}(\kappa_i-\kappa_{i-1})}{(\kappa_3-\kappa_2)+(\kappa_1-\kappa_0)}\right\}^{d/2} O_p\left[\prod_{i\neq 1,3}\exp\{-(\kappa_i-\kappa_{i-1})k(\delta)\}\right] \\
= \;& O_p\left\{(b-a)^{rd/2}\exp(-\underline{\kappa}k(\delta))\right\}.
\end{aligned}$$

Then we obtain

$$\frac{C(\mathbf{Y}_{(a,b]})}{C(\mathbf{Y}_{(a,\kappa_1]})\cdots C(\mathbf{Y}_{(\kappa_r,b]})} = O_p\left\{(b-a)^{rd/2}\exp(-\underline{\kappa}k(\delta))\right\}.$$

□

Lemma 6. *Assume conditions in Theorem 1 hold, and $\mathcal{K}_0$ is a subset of $\mathcal{H}(m_I)$. Let $\widehat{\mathcal{K}}$ be the estimated change point set determined by our algorithm. Suppose that there exists a true change point $\kappa_{0j} \notin \widehat{\mathcal{K}}$. Let $\hat{\kappa}_i$ and $\hat{\kappa}_{i+1}$ be the estimated change point which sandwich κ_{0j}, and $\hat{\kappa}_i < \kappa_{0,j-l} < \ldots < \kappa_{0j} < \ldots < \kappa_{0,l+r} < \hat{\kappa}_{i+1}$, where $l, r \geq 0$. Considering a new estimated change point set*

$$\widetilde{\mathcal{K}} = \{\hat{\kappa}_1, \ldots, \hat{\kappa}_i, \kappa_{0,j-l}, \ldots, \kappa_{0,j+r}, \hat{\kappa}_{i+1}, \ldots, \hat{\kappa}_{\hat{p}}\},$$

then

$$\frac{\Pr(\widehat{\mathcal{K}}|\mathbf{Y})}{\Pr(\widetilde{\mathcal{K}}|\mathbf{Y})} = o_p(1).$$

Proof: Let $T_0 = \hat{\kappa}_{i+1} - \hat{\kappa}_i$, $t_1 = \kappa_{0,j-l} - \hat{\kappa}_i, \ldots, t_{l+r+2} = \hat{\kappa}_{i+1} - \kappa_{0,j+r}$. By the Stirling formula, we have

$$\begin{aligned} & \frac{\Pr(\widehat{\mathcal{K}}|\mathbf{Y})}{\Pr(\widetilde{\mathcal{K}}|\mathbf{Y})} \\ = & \frac{\prod_{j=1}^{T_0-1}(j-\sigma)\prod_{h=1}^{l+r+2} t_h!}{T_0!\prod_{h=1}^{l+r+2}\prod_{j=1}^{t_h-1}(j-\sigma)} \prod_{s=\hat{p}+1}^{\hat{p}+l+r+1}\left(\frac{s+1}{\alpha+s\sigma}\right) \frac{C(\mathbf{Y}_{(\hat{\kappa}_i,\hat{\kappa}_{i+1}]})}{C(\mathbf{Y}_{(\hat{\kappa}_i,\kappa_{0,j-l}]})\cdots C(\mathbf{Y}_{(\kappa_{0,j+r},\hat{\kappa}_{i+1}]})} \\ = & \frac{C(\mathbf{Y}_{(\hat{\kappa}_i,\hat{\kappa}_{i+1}]})}{C(\mathbf{Y}_{(\hat{\kappa}_i,\kappa_{0,j-l}]})\cdots C(\mathbf{Y}_{(\kappa_{0,j+r},\hat{\kappa}_{i+1}]})} O\left(\frac{\prod_{j=1}^{l+r+2} t_j}{T_0}\right)^{1+\sigma}. \end{aligned}$$

By Lemma 5, let $\underline{\kappa} = \min_{i=1,\ldots,l+r+2} t_i$, there exists $c_2 > 0$ such that

$$\frac{C(\mathbf{Y}_{(\hat{\kappa}_i,\hat{\kappa}_{i+1}]})}{C(\mathbf{Y}_{(\hat{\kappa}_i,\kappa_{0,j-l}]})\cdots C(\mathbf{Y}_{(\kappa_{0,j+r},\hat{\kappa}_{i+1}]})} = O_p\left\{T_0^{(l+r+1)d/2}\exp(-\underline{\kappa}c_2)\right\}.$$

Thus we obtain

$$\begin{aligned} & \frac{\Pr(\widehat{\mathcal{K}}|\mathbf{Y})}{\Pr(\widetilde{\mathcal{K}}|\mathbf{Y})} \\ = & O\left(\frac{\prod_{j=1}^{l+r+2} t_j}{T_0}\right)^{1+\sigma} \frac{C(\mathbf{Y}_{(\hat{\kappa}_i,\hat{\kappa}_{i+1}]})}{C(\mathbf{Y}_{(\hat{\kappa}_i,\kappa_{0,j-l}]})\cdots C(\mathbf{Y}_{(\kappa_{0,j+r},\hat{\kappa}_{i+1}]})} \\ = & O\left(\frac{\prod_{j=1}^{l+r+2} t_j}{T_0}\right)^{1+\sigma} O_p\left\{T_0^{(l+r+1)d/2}\exp(-\underline{\kappa}c_2)\right\} \\ = & O_p\left\{T_0^{(l+r+1)(d/2+1+\sigma)}\exp(-\underline{\kappa}c_2)\right\}. \end{aligned} \tag{12}$$

By definition of $\mathcal{H}(m_I)$,

$$\underline{\kappa} \geq m_I \geq c\{\log(T)\}^{1+\epsilon}$$

for some $c > 0$ and $\epsilon > 0$ when T is large enough. Clearly, $T \geq T_0$. Thus as $T \to \infty$,

$$\begin{aligned} & T_0^{(l+r+1)(d/2+1+\sigma)} \exp(-\underline{\kappa} c_2) \\ \leq \; & T^{(l+r+1)(d/2+1+\sigma)} \exp\left[-\{\log(T)\}^{1+\epsilon} c c_2\right] \\ = \; & \exp\left[\log(T) c_0 - \{\log(T)\}^{1+\epsilon} c c_2\right] \\ = \; & \exp\left(\log(T)[c_0 - \{\log(T)\}^{\epsilon} c c_2]\right) \longrightarrow 0 \end{aligned} \tag{13}$$

where $c_0 = (l+r+1)(d/2+1+\sigma)$. With (12) and (13), we achieve

$$\frac{\Pr(\widehat{\mathcal{K}}|\mathbf{Y})}{\Pr(\widetilde{\mathcal{K}}|\mathbf{Y})} = o_p(1).$$

□

Lemma 7. *Assume the conditions in Theorem 1 hold, and $\mathcal{K}_0$ is a subset of $\mathcal{H}(m_I)$. Let $\widehat{\mathcal{K}}$ be the estimated change point set determined by our algorithm. Suppose that there exists an estimated change point $\hat{\kappa}_i$, such that no true change point is within its m_I-neighbourhood, i.e., $\kappa_{0j} \notin (\hat{\kappa}_i - m_I, \hat{\kappa}_i + m_I)$ for all j. Considering a newly estimated change point set*

$$\widetilde{\mathcal{K}} = \{\hat{\kappa}_1, \dots, \hat{\kappa}_{i-1}, \hat{\kappa}_{i+1}, \dots, \hat{\kappa}_{\hat{p}}\},$$

then

$$\frac{\Pr(\widehat{\mathcal{K}}|\mathbf{Y})}{\Pr(\widetilde{\mathcal{K}}|\mathbf{Y})} = o_p(1).$$

Proof: By the Stirling formula, we have

$$\begin{aligned} & \frac{\Pr(\widehat{\mathcal{K}}|\mathbf{Y})}{\Pr(\widetilde{\mathcal{K}}|\mathbf{Y})} \\ = \; & \frac{\alpha + (\hat{p}+1)\sigma}{\hat{p}+2} \frac{C(\mathbf{Y}_{(\hat{\kappa}_{i-1}, \hat{\kappa}_i]}) C(\mathbf{Y}_{(\hat{\kappa}_i, \hat{\kappa}_{i+1}]})}{C(\mathbf{Y}_{(\hat{\kappa}_{i-1}, \hat{\kappa}_{i+1}]})} \\ & \times \frac{\prod_{j=1}^{\hat{\kappa}_{i+1} - \hat{\kappa}_i - 1} (j-\sigma)/(\hat{\kappa}_{i+1} - \hat{\kappa}_i)!}{\prod_{j=1}^{\hat{\kappa}_{i+1} - \hat{\kappa}_{i-1} - 1} (j-\sigma)/(\hat{\kappa}_{i+1} - \hat{\kappa}_{i-1})!} \prod_{j=1}^{\hat{\kappa}_i - \hat{\kappa}_{i-1} - 1} (j-\sigma)/(\hat{\kappa}_i - \hat{\kappa}_{i-1})! \\ = \; & \frac{\alpha + (\hat{p}+1)\sigma}{\hat{p}+2} \frac{C(\mathbf{Y}_{(\hat{\kappa}_{i-1}, \hat{\kappa}_i]}) C(\mathbf{Y}_{(\hat{\kappa}_i, \hat{\kappa}_{i+1}]})}{C(\mathbf{Y}_{(\hat{\kappa}_{i-1}, \hat{\kappa}_{i+1}]})} \\ & \times \frac{\Gamma(\hat{\kappa}_i - \hat{\kappa}_{i-1} - \sigma)\Gamma(\hat{\kappa}_{i+1} - \hat{\kappa}_i - \sigma)(\hat{\kappa}_{i+1} - \hat{\kappa}_{i-1})!}{\Gamma(1-\sigma)\Gamma(\hat{\kappa}_{i+1} - \hat{\kappa}_{i-1} - \sigma)(\hat{\kappa}_i - \hat{\kappa}_{i-1})!(\hat{\kappa}_{i+1} - \hat{\kappa}_i)!} \\ = \; & O\left\{ \frac{(\hat{\kappa}_{i+1} - \hat{\kappa}_{i-1})}{(\hat{\kappa}_{i+1} - \hat{\kappa}_i)(\hat{\kappa}_i - \hat{\kappa}_{i-1})} \right\}^{1+\sigma} \frac{C(\mathbf{Y}_{(\hat{\kappa}_{i-1}, \hat{\kappa}_i]}) C(\mathbf{Y}_{(\hat{\kappa}_i, \hat{\kappa}_{i+1}]})}{C(\mathbf{Y}_{(\hat{\kappa}_{i-1}, \hat{\kappa}_{i+1}]})}. \end{aligned} \tag{14}$$

By Lemma 6, every true change point κ_{0j} is in $\widehat{\mathcal{K}}$. Thus, there is no true change point between $\hat{\kappa}_{i-1}$ and $\hat{\kappa}_{i+1}$. Then using Lemma 4, we obtain

$$\begin{aligned} C(\mathbf{Y}_{(\hat{\kappa}_{i-1}, \hat{\kappa}_i]}) &= p_{(\hat{\kappa}_{i-1}, \hat{\kappa}_i]}(\hat{\boldsymbol{\theta}}_{(\hat{\kappa}_{i-1}, \hat{\kappa}_i]}) \pi(\widehat{\boldsymbol{\theta}}_{(\hat{\kappa}_{i-1}, \hat{\kappa}_i]}) \det(\hat{\sigma}_{(\hat{\kappa}_{i-1}, \hat{\kappa}_i]}) O_p(1), \\ C(\mathbf{Y}_{(\hat{\kappa}_i, \hat{\kappa}_{i+1}]}) &= p_{(\hat{\kappa}_i, \hat{\kappa}_{i+1}]}(\hat{\boldsymbol{\theta}}_{(\hat{\kappa}_i, \hat{\kappa}_{i+1}]}) \pi(\widehat{\boldsymbol{\theta}}_{(\hat{\kappa}_i, \hat{\kappa}_{i+1}]}) \det(\hat{\sigma}_{(\hat{\kappa}_i, \hat{\kappa}_{i+1}]}) O_p(1), \\ C(\mathbf{Y}_{(\hat{\kappa}_{i-1}, \hat{\kappa}_{i+1}]}) &= p_{(\hat{\kappa}_{i-1}, \hat{\kappa}_{i+1}]}(\hat{\boldsymbol{\theta}}_{(\hat{\kappa}_{i-1}, \hat{\kappa}_{i+1}]}) \pi(\widehat{\boldsymbol{\theta}}_{(\hat{\kappa}_{i-1}, \hat{\kappa}_{i+1}]}) \det(\hat{\sigma}_{(\hat{\kappa}_{i-1}, \hat{\kappa}_{i+1}]}) O_p(1). \end{aligned}$$

It follows that

$$
\begin{aligned}
& \frac{C(\mathbf{Y}_{(\hat\kappa_{i-1},\hat\kappa_i]})C(\mathbf{Y}_{(\hat\kappa_i,\hat\kappa_{i+1}]})}{C(\mathbf{Y}_{(\hat\kappa_{i-1},\hat\kappa_{i+1}]})} \\
= & \frac{p_{(\hat\kappa_{i-1},\hat\kappa_i]}(\hat{\boldsymbol{\theta}}_{(\hat\kappa_{i-1},\hat\kappa_i]})\det(\hat\sigma_{(\hat\kappa_{i-1},\hat\kappa_i]})}{p_{(\hat\kappa_{i-1},\hat\kappa_{i+1}]}(\hat{\boldsymbol{\theta}}_{(\hat\kappa_{i-1},\hat\kappa_{i+1}]})} \times \frac{p_{(\hat\kappa_i,\hat\kappa_{i+1}]}(\hat{\boldsymbol{\theta}}_{(\hat\kappa_i,\hat\kappa_{i+1}]})\det(\hat\sigma_{(\hat\kappa_i,\hat\kappa_{i+1}]})}{\pi(\widehat{\boldsymbol{\theta}}_{(\hat\kappa_{i-1},\hat\kappa_{i+1}]})} \\
& \times \frac{\pi(\widehat{\boldsymbol{\theta}}_{(\hat\kappa_{i-1},\hat\kappa_i]})\pi(\widehat{\boldsymbol{\theta}}_{(\hat\kappa_i,\hat\kappa_{i+1}]})}{\det(\hat\sigma_{(\hat\kappa_{i-1},\hat\kappa_{i+1}]})} O_p(1).
\end{aligned}
$$

As $\widehat{\boldsymbol{\theta}}_{(\hat\kappa_{i-1},\hat\kappa_i]} \xrightarrow{\mathcal{P}} \boldsymbol{\theta}_0$, $\widehat{\boldsymbol{\theta}}_{(\hat\kappa_i,\hat\kappa_{i+1}]} \xrightarrow{\mathcal{P}} \boldsymbol{\theta}_0$ and $\widehat{\boldsymbol{\theta}}_{(\hat\kappa_{i-1},\hat\kappa_{i+1}]} \xrightarrow{\mathcal{P}} \boldsymbol{\theta}_0$ where $\boldsymbol{\theta}_0$ is the true parameter, and $\pi(\boldsymbol{\theta})$ is continuous, by the continuous mapping theorem, we know

$$
\begin{aligned}
& \frac{C(\mathbf{Y}_{(\hat\kappa_{i-1},\hat\kappa_i]})C(\mathbf{Y}_{(\hat\kappa_i,\hat\kappa_{i+1}]})}{C(\mathbf{Y}_{(\hat\kappa_{i-1},\hat\kappa_{i+1}]})} \\
= & \frac{p_{(\hat\kappa_{i-1},\hat\kappa_i]}(\hat{\boldsymbol{\theta}}_{(\hat\kappa_{i-1},\hat\kappa_i]})\det(\hat\sigma_{(\hat\kappa_{i-1},\hat\kappa_i]})}{p_{(\hat\kappa_{i-1},\hat\kappa_{i+1}]}(\hat{\boldsymbol{\theta}}_{(\hat\kappa_{i-1},\hat\kappa_{i+1}]})} \times \frac{p_{(\hat\kappa_i,\hat\kappa_{i+1}]}(\hat{\boldsymbol{\theta}}_{(\hat\kappa_i,\hat\kappa_{i+1}]})\det(\hat\sigma_{(\hat\kappa_i,\hat\kappa_{i+1}]})}{\det(\hat\sigma_{(\hat\kappa_{i-1},\hat\kappa_{i+1}]})} O_p(1)
\end{aligned}
$$

Further,

$$
\begin{aligned}
& \frac{C(\mathbf{Y}_{(\hat\kappa_{i-1},\hat\kappa_i]})C(\mathbf{Y}_{(\hat\kappa_i,\hat\kappa_{i+1}]})}{C(\mathbf{Y}_{(\hat\kappa_{i-1},\hat\kappa_{i+1}]})} \\
= & \frac{p_{(\hat\kappa_{i-1},\hat\kappa_i]}(\hat{\boldsymbol{\theta}}_{(\hat\kappa_{i-1},\hat\kappa_i]})\det(\hat\sigma_{(\hat\kappa_{i-1},\hat\kappa_i]})}{p_{(\hat\kappa_{i-1},\hat\kappa_{i+1}]}(\hat{\boldsymbol{\theta}}_{(\hat\kappa_{i-1},\hat\kappa_{i+1}]})} \times \frac{p_{(\hat\kappa_i,\hat\kappa_{i+1}]}(\hat{\boldsymbol{\theta}}_{(\hat\kappa_i,\hat\kappa_{i+1}]})\det(\hat\sigma_{(\hat\kappa_i,\hat\kappa_{i+1}]})}{\det(\hat\sigma_{(\hat\kappa_{i-1},\hat\kappa_{i+1}]})} O_p(1) \\
= & \frac{p_{(\hat\kappa_{i-1},\hat\kappa_i]}(\hat{\boldsymbol{\theta}}_{(\hat\kappa_{i-1},\hat\kappa_i]})p_{(\hat\kappa_i,\hat\kappa_{i+1}]}(\hat{\boldsymbol{\theta}}_{(\hat\kappa_i,\hat\kappa_{i+1}]})}{p_{(\hat\kappa_{i-1},\hat\kappa_{i+1}]}(\hat{\boldsymbol{\theta}}_{(\hat\kappa_{i-1},\hat\kappa_{i+1}]})} O_p\left\{\frac{\hat\kappa_{i+1}-\hat\kappa_{i-1}}{(\hat\kappa_{i+1}-\hat\kappa_i)(\hat\kappa_i-\hat\kappa_{i-1})}\right\}^{d/2}. \qquad (15)
\end{aligned}
$$

Note that by the second result of Lemma 3,

$$
\begin{aligned}
& \frac{p_{(\hat\kappa_{i-1},\hat\kappa_i]}(\hat{\boldsymbol{\theta}}_{(\hat\kappa_{i-1},\hat\kappa_i]})p_{(\hat\kappa_i,\hat\kappa_{i+1}]}(\hat{\boldsymbol{\theta}}_{(\hat\kappa_i,\hat\kappa_{i+1}]})}{p_{(\hat\kappa_{i-1},\hat\kappa_{i+1}]}(\hat{\boldsymbol{\theta}}_{(\hat\kappa_{i-1},\hat\kappa_{i+1}]})} \\
= & \frac{p_{(\hat\kappa_i,\hat\kappa_{i+1}]}(\hat{\boldsymbol{\theta}}_{(\hat\kappa_i,\hat\kappa_{i+1}]})p_{(\hat\kappa_{i-1},\hat\kappa_i]}(\hat{\boldsymbol{\theta}}_{(\hat\kappa_{i-1},\hat\kappa_i]})p_{(\hat\kappa_{i-1},\hat\kappa_i]}(\boldsymbol{\theta}_0)^{-1}}{p_{(\hat\kappa_i,\hat\kappa_{i+1}]}(\boldsymbol{\theta}_0)p_{(\hat\kappa_{i-1},\hat\kappa_{i+1}]}(\hat{\boldsymbol{\theta}}_{(\hat\kappa_{i-1},\hat\kappa_{i+1}]})p_{(\hat\kappa_{i-1},\hat\kappa_{i+1}]}(\boldsymbol{\theta}_0)^{-1}} \\
= & O_p(1). \qquad (16)
\end{aligned}
$$

Thus, combining (14), (15) and (16), we obtain

$$
\begin{aligned}
\frac{\Pr(\widehat{\mathcal{K}}|\mathbf{Y})}{\Pr(\widetilde{\mathcal{K}}|\mathbf{Y})} &= O\left\{\frac{(\hat\kappa_{i+1}-\hat\kappa_{i-1})}{(\hat\kappa_{i+1}-\hat\kappa_i)(\hat\kappa_i-\hat\kappa_{i-1})}\right\}^{1+\sigma} \frac{C(\mathbf{Y}_{(\hat\kappa_{i-1},\hat\kappa_i]})C(\mathbf{Y}_{(\hat\kappa_i,\hat\kappa_{i+1}]})}{C(\mathbf{Y}_{(\hat\kappa_{i-1},\hat\kappa_{i+1}]})} \\
&= O\left\{\frac{(\hat\kappa_{i+1}-\hat\kappa_{i-1})}{(\hat\kappa_{i+1}-\hat\kappa_i)(\hat\kappa_i-\hat\kappa_{i-1})}\right\}^{1+\sigma} O_p\left\{\frac{\hat\kappa_{i+1}-\hat\kappa_{i-1}}{(\hat\kappa_{i+1}-\hat\kappa_i)(\hat\kappa_i-\hat\kappa_{i-1})}\right\}^{d/2} \\
&= O_p\left\{\frac{\hat\kappa_{i+1}-\hat\kappa_{i-1}}{(\hat\kappa_{i+1}-\hat\kappa_i)(\hat\kappa_i-\hat\kappa_{i-1})}\right\}^{d/2+1+\sigma}.
\end{aligned}
$$

Note that $(\hat{\kappa}_{i+1} - \hat{\kappa}_i) + (\hat{\kappa}_i - \hat{\kappa}_{i-1}) = (\hat{\kappa}_{i+1} - \hat{\kappa}_{i-1})$, thus either $(\hat{\kappa}_{i+1} - \hat{\kappa}_{i-1})/(\hat{\kappa}_{i+1} - \hat{\kappa}_i)$ or $(\hat{\kappa}_{i+1} - \hat{\kappa}_{i-1})/(\hat{\kappa}_i - \hat{\kappa}_{i-1})$ will go to a constant $c > 0$ as $T \to \infty$. Therefore,

$$\frac{\hat{\kappa}_{i+1} - \hat{\kappa}_{i-1}}{(\hat{\kappa}_{i+1} - \hat{\kappa}_i)(\hat{\kappa}_i - \hat{\kappa}_{i-1})} \to 0,$$

since both $(\hat{\kappa}_i - \hat{\kappa}_{i-1})$ and $(\hat{\kappa}_{i+1} - \hat{\kappa}_i)$ go to infinity as $T \to \infty$. It follows that

$$\frac{\Pr(\widehat{\mathcal{K}}|\mathbf{Y})}{\Pr(\widetilde{\mathcal{K}}|\mathbf{Y})} = O_p \left\{ \frac{\hat{\kappa}_{i+1} - \hat{\kappa}_{i-1}}{(\hat{\kappa}_{i+1} - \hat{\kappa}_i)(\hat{\kappa}_i - \hat{\kappa}_{i-1})} \right\}^{d/2+1+\sigma} = o_p(1).$$

□

Proof of Theorem 1:

By Lemma 6, we know all the true change points will fall into $\widehat{\mathcal{K}}$ with probability one as $T \to \infty$. Lemme 7 implies that all the estimated change points out of m_I-neighbourhood of true change points can be removed in probability as $T \to \infty$. By the definition of set $\mathcal{H}(m_I)$, for any two points τ_i and τ_j with $\tau_i < \tau_j$ in $\mathcal{H}(m_I)$, $(\tau_j - \tau_i) > m_I$. We obtain $\widehat{\mathcal{K}}$ by optimizing over $\mathcal{H}(m_I)$. Thus for any true change point, there is one and only one point in $\widehat{\mathcal{K}}$ within its m_I-neighbourhood, i.e., the true change point itself. Therefore, with $T \to \infty$,

$$\hat{p} \xrightarrow{\mathcal{P}} p_0 \qquad \text{and} \qquad \sup_{b \in \mathcal{K}_0} \inf_{a \in \widehat{\mathcal{K}}} |a - b| = O_p(1).$$

□

References

Du, C., Kao, C. L. M., and Kou, S. C. (2016), "Stepwise Signal Extraction via Marginal Likelihood," *Journal of the American Statistical Association*, 111, 314–330. 9

Fraser, D. A. S. and Mcdunnough, P. (1984), "Further Remarks on Asymptotic Normality of Likelihood and Conditional Analyses," *The Canadian Journal of Statistics*, 12, 183–190. 13

Jiang, F., Yin, G., and Dominici, F. (2018), "Bayesian model selection approach to boundary detection with non-local priors," in *Advances in Neural Information Processing Systems*. 11

Rosner, B. (1983), "Percentage points for a generalized ESD many-outlier procedure," *Technometrics*, 25, 165–172. 4

Walker, A. M. (1969), "On the Asymptotic Behaviour of Posterior Distributions," *Journal of the Royal Statistical Society: Series B*, 31, 80–88. 13