Bayesian Optimal Change Point Detection in High Dimension

JaeHoon Kim[1], KyoungJae Lee[1], and Lizhen Lin[2]

[1]Department of Statistics, Sungkyunkwan University
[2]Department of Mathematics, The University of Maryland

Presented at KAIST
November 22, 2024

Outline

- Introduction
- Methodologies
  - Change Point Detection in Mean Structure
  - Change Point Detection in Covariance Structure
  - Choice of Hyperparameters
  - Multiscale Approach
- Theoretical Foundations
- Numerical Results

Problem Setting

- Suppose we observe a sequence of random vectors $X_1, \ldots, X_n$ independently from p-dimensional Gaussian distributions:

$$X_i \overset{\text{ind.}}{\sim} N_p(\mu_i, \Sigma_i), \quad i = 1, \ldots, n.$$

- We consider the high-dimensional setting:
  - The number of variables p can increase to infinity as the sample size n gets larger.
  - The number of variables p can even be much larger than n.
- We are interested in:
  - Testing for the existence of a change point.
  - Estimating the number and locations of change points, if they exist.

Problem Setting

- Suppose the changes occur in either the mean vector or the covariance matrix.
- We define K change points $\{i_1, \ldots, i_K\}$ with $1 < i_1 < \cdots < i_K < n$, if:

$$(\mu_{i_k}, \Sigma_{i_k}) = \cdots = (\mu_{i_{k+1}-1}, \Sigma_{i_{k+1}-1}), \quad k = 0, \ldots, K,$$
$$(\mu_{i_k}, \Sigma_{i_k}) \neq (\mu_{i_{k+1}}, \Sigma_{i_{k+1}}), \quad k = 0, \ldots, K,$$

  where $i_0 = 1$ and $i_{K+1} = n + 1$.
- The number of change points and their locations are unknown.

CPD in Mean Structure

- The main idea is to extend the mxPBF for two-sample tests to change point detection by sequentially performing tests on subsets of data within a specified window.
- Assume that the data has a common covariance:

$$X_i \overset{\text{ind.}}{\sim} N_p(\mu_i, \Sigma), \quad i = 1, \ldots, n.$$

- We are interested in testing

$$H_0^\mu : \mu_i = \mu_{i+1} \quad \text{for all } \ 1 \leq i \leq n-1 \quad \text{vs} \quad H_1^\mu : \text{not } H_0^\mu.$$

- The main goals here are:
  - Testing the existence of change points in the mean structure.
  - Estimating the locations of change points in the mean structure.

Testing the existence of CP

- We can decompose the hypothesis test into multiple smaller ones.
- Let $n_w > 1$ be the window size and l be a positive integer.
- For $n_w + 1 \leq l \leq n - n_w + 1$, consider the model:

$$X_{l-n_w}, \ldots, X_{l-1} \mid \mu_{l-1}, \Sigma \overset{\text{i.i.d.}}{\sim} N_p(\mu_{l-1}, \Sigma),$$
$$X_l, \ldots, X_{l+n_w-1} \mid \mu_l, \Sigma \overset{\text{i.i.d.}}{\sim} N_p(\mu_l, \Sigma),$$

where $\mu_{l-1} = (\mu_{l-1,1}, \ldots, \mu_{l-1,p})^T$ and $\mu_l = (\mu_{l,1}, \ldots, \mu_{l,p})^T$.

- Then, the hypothesis becomes:

$$H_0^{\mu,l} : \mu_{l-1} = \mu_l \quad \text{vs} \quad H_1^{\mu,l} : \mu_{l-1} \neq \mu_l$$

in the sense that H_0^{μ} is true if and only if $H_0^{\mu,l}$ is true for all l.

Testing the Existence of CP

- We apply the maximum pairwise Bayes factor (mxPBF) approach (Lee et al., 2021) to construct the Bayes factor for the hypothesis test.
- For $1 \le j \le p$, the marginalized models are given by:

$$X_{l-n_w,j}, \dots, X_{l-1,j} \mid \mu_{l-1,j}, \sigma_{jj} \overset{\text{i.i.d.}}{\sim} N(\mu_{l-1,j}, \sigma_{jj}),$$
$$X_{l,j}, \dots, X_{l+n_w-1,j} \mid \mu_{l,j}, \sigma_{jj} \overset{\text{i.i.d.}}{\sim} N(\mu_{l,j}, \sigma_{jj}),$$

  where σ_{jj} is the (j, j) entry of Σ.
- Finally, the decomposed hypotheses are:

$$H_{0,j}^{\mu,l} : \mu_{l-1,j} = \mu_{l,j} \quad \text{vs} \quad H_{1,j}^{\mu,l} : \mu_{l-1,j} \neq \mu_{l,j}$$

  with $H_0^{\mu,l}$ being true if and only if $H_{0,j}^{\mu,l}$ is true for all j.

Testing the Existence of CP

- We first calculate the pairwise Bayes factors (PBFs) for each hypothesis and later integrate them.
- Priors under $H_{0,j}^{\mu,l}$:
  - $\mu_j \mid \sigma_{jj} \sim N\left(\overline{\mathbf{X}}_{(l-n_w):(l+n_w-1),j}, \frac{\sigma_{jj}}{2n_w\gamma_{n_w}}\right)$
  - $\pi\left(\sigma_{jj}\right) \propto \sigma_{jj}^{-1}$

  where $\mu_j = \mu_{l-1,j} = \mu_{l,j}$ and $\overline{\mathbf{X}}_{a:b,j} = (b-a+1)^{-1}\sum_{i=a}^{b} X_{ij}$.
- Priors under $H_{1,j}^{\mu,l}$:
  - $\mu_{l-1,j} \mid \sigma_{jj} \sim N\left(\overline{\mathbf{X}}_{(l-n_w):(l-1),j}, \frac{\sigma_{jj}}{n_w\gamma_{n_w}}\right)$
  - $\mu_{l,j} \mid \sigma_{jj} \sim N\left(\overline{\mathbf{X}}_{l:(l+n_w-1),j}, \frac{\sigma_{jj}}{n_w\gamma_{n_w}}\right)$
  - $\pi\left(\sigma_{jj}\right) \propto \sigma_{jj}^{-1}$

  where $\gamma_{n_w} = (n_w \vee p)^{-\alpha}$ and $\alpha > 0$.

Testing the Existence of CP

- The resulting log PBF is:

$$
\begin{aligned}
&\log B_{10}^{\mu}\left(\mathbf{X}_{(l-n_w):(l+n_w-1),j}\right)\\
&= \log \frac{p\left(\mathbf{X}_{(l-n_w):(l+n_w-1),j} \mid H_{1,j}^{\mu,l}\right)}{p\left(\mathbf{X}_{(l-n_w):(l+n_w-1),j} \mid H_{0,j}^{\mu,l}\right)}\\
&= \frac{1}{2}\log\left(\frac{\gamma_{n_w}}{1+\gamma_{n_w}}\right) + n_w \log\left(\frac{2n_w\widehat{\sigma}^2_{n_w,l}}{n_w\widehat{\sigma}^2_{n_w,l1} + n_w\widehat{\sigma}^2_{n_w,l2}}\right),
\end{aligned}
$$

where

$$
\widehat{\sigma}^2_{n_w,l} = \frac{X^T_{(l-n_w):(l+n_w-1),j}(I_{2n_w} - H_{1_{2n_w}})X_{(l-n_w):(l+n_w-1),j}}{2n_w}.
$$

Testing the Existence of CP

- Now, for each l, we can define the mxPBF:

$$B_{max,10}^{\mu,l,n_w}(X_n) := \max_{1\leq j\leq p} B_{10}^{\mu}(X_{(l-n_w):(l+n_w-1),j})$$

- The final mxPBF for testing the existence of change points is:

$$B_{max,10}^{\mu,n_w}(X_n) := \max_{n_w+1\leq l\leq n-n_w+1} B_{max,10}^{\mu,l,n_w}(X_n)$$

- Thus, for a given threshold $C_{cp} > 0$, we can reject H_0^{μ} if

$$B_{max,10}^{\mu,n_w}(X_n) > C_{cp}$$

CPD in Covariance Structure

- The method in the covariance structure is quite similar.
- Assume that the data has a zero mean:

$$X_i \overset{\text{ind.}}{\sim} N_p(0, \Sigma_i), \quad i = 1, \ldots, n.$$

- We consider the following hypothesis test:

$$H_0^{\Sigma} : \Sigma_i = \Sigma_{i+1} \quad \text{for all } \ 1 \leq i \leq n-1 \quad \text{vs} \quad H_1^{\Sigma} : \text{not } H_0^{\Sigma}.$$

- Given n_w and l, this hypothesis becomes:

$$H_0^{\Sigma,l} : \Sigma_{l-1} = \Sigma_l \quad vs \quad H_1^{\Sigma,l} : \Sigma_{l-1} \neq \Sigma_l,$$

meaning that H_0^{Σ} holds if and only if $H_0^{\Sigma,l}$ is true for all l.

Testing the Existence of CP

- We apply the mxPBF idea (Lee et al., 2021) once again to construct the Bayes factor for hypothesis testing.
- For a given pair (i, j) with $1 \leq i \neq j \leq p$, the marginalized models are defined as follows:

$$X_{(l-n_w):(l-1),i} \mid X_{(l-n_w):(l-1),j} \overset{\text{i.i.d.}}{\sim} N_{n_w}\left(a_{l-1,ij} X_{(l-n_w):(l-1),j}, \tau_{l-1,ij} I_{n_w}\right),$$
$$X_{l:(l+n_w-1),i} \mid X_{l:(l+n_w-1),j} \overset{\text{i.i.d.}}{\sim} N_{n_w}\left(a_{l,ij} X_{l:(l+n_w-1),j}, \tau_{l,ij} I_{n_w}\right),$$

where $\Sigma_l = (\sigma_{l,ij})$, $a_{l,ij} = \sigma_{l,ij}/\sigma_{l,jj}$, and $\tau_{l,ij} = \sigma_{l,ii}(1 - \rho^2_{l,ij})$.

- The resulting hypotheses are:

$$H^{\Sigma,l}_{0,ij} : a_{1,ij} = a_{2,ij} \text{ and } \tau_{1,ij} = \tau_{2,ij} \quad vs \quad H^{\Sigma,l}_{1,ij} : not\ H^{\Sigma,l}_{0,ij}.$$

Testing the Existence of CP

- Priors under $H_{0,j}^{\Sigma,l}$:
  - $a_{ij} \mid \tau_{ij} \sim N\left(\widehat{a}_{l,ij}, \frac{\tau_{ij}}{\gamma_{n_w} \|X_{(l-n_w):(l+n_w-1),j}\|_2^2}\right)$
  - $\tau_{ij} \sim IG\left(a_0, b_{0,ij}\right)$
- Priors under $H_{1,j}^{\Sigma,l}$:
  - $a_{l-1,ij} \mid \tau_{l-1,ij} \sim N\left(\widehat{a}_{1l-1,ij}, \frac{\tau_{l-1,ij}}{\gamma_{n_w} \|X_{(l-n_w):(l-1),j}\|_2^2}\right)$
  - $a_{l,ij} \mid \tau_{l,ij} \sim N\left(\widehat{a}_{2l,ij}, \frac{\tau_{l,ij}}{\gamma_{n_w} \|X_{l:(l+n_w-1),j}\|_2^2}\right)$
  - $\tau_{l-1,ij} \sim IG(a_0, b_{01,ij})$
  - $\tau_{l,ij} \sim IG(a_0, b_{02,ij})$

where

$$a_{ij} = a_{1,ij} = a_{2,ij},\ \tau_{ij} = \tau_{1,ij} = \tau_{2,ij},\ \widehat{a}_{l,ij} = \frac{X_{(l-n_w):(l+n_w-1),i}^T X_{(l-n_w):(l+n_w-1),j}}{\left\|X_{(l-n_w):(l+n_w-1),j}\right\|_2^2}.$$

Testing the Existence of CP

- The resulting log PBF is:

$$
\begin{aligned}
&\log B_{10}^{\Sigma}\left(X_{(l-n_w):(l+n_w-1),i}, X_{(l-n_w):(l+n_w-1),j}\right) \\
&= \frac{1}{2}\log\left(\frac{\gamma_{n_w}}{1+\gamma_{n_w}}\right) + 2\log\Gamma\left(\frac{n_w}{2}+a_0\right) - \log\left\{\Gamma\left(n_w+a_0\right)\Gamma(a_0)\right\} \\
&\quad + a_0\log\left(\frac{b_{01,ij}b_{02,ij}}{b_{0,ij}}\right) - \left(\frac{n_w}{2}+a_0\right)\log\left(b_{01,ij}+\frac{n_w}{2}\widehat{\tau}_{1l-1,ij}\right) \\
&\quad - \left(\frac{n_w}{2}+a_0\right)\log\left(b_{02,ij}+\frac{n_w}{2}\widehat{\tau}_{2l,ij}\right) + \left(n_w+a_0\right)\log\left(b_{0,ij}+n_w\widehat{\tau}_{l,ij}\right),
\end{aligned}
$$

where Γ is the Gamma function and

$$
\widehat{\tau}_{ij} = \frac{X^T_{(l-n_w):(l+n_w-1),i}\left(I_{2n_w} - H_{X_{(l-n_w):(l+n_w-1),j}}\right)X_{(l-n_w):(l+n_w-1),i}}{2n_w}.
$$

Testing the Existence of CP

- For each l, we define the mxPBF as:

$$B_{max,10}^{\Sigma,l,n_w}(X_n) := \max_{1\le i\ne j\le p} B_{10}^{\Sigma}(X_{(l-n_w):(l+n_w-1),i}, X_{(l-n_w):(l+n_w-1),j})$$

- The final mxPBF, used for testing the existence of change points, is:

$$B_{max,10}^{\Sigma,n_w}(X_n) := \max_{n_w+1\le l\le n-n_w+1} B_{10}^{\Sigma,l,n_w}(X_n)$$

- Given a threshold $C_{cp} > 0$, we reject H_0^{Σ} if

$$B_{max,10}^{\Sigma,n_w}(X_n) > C_{cp}$$

- Once the existence of change points is confirmed, we can proceed to estimate their locations.

Estimating the Location of CP

- First, we define an initial estimate of the first change point:

$$\tilde{i}_1 := \min\left\{ n_w + 1 \leq l \leq n - n_w + 1 : B^{l,n_w}_{max,10}(X_n) > C_{cp} \right\}.$$

- Next, a refined estimate $\widehat{i_1}$ can be obtained by maximizing within the interval $[\tilde{i}_1, \tilde{i}_1 + n_w - 1]$:

$$\widehat{i_1} := \underset{\tilde{i}_1 \leq l \leq \tilde{i}_1 + n_w - 1}{\text{argmax}} \log B^{l,n_w}_{max,10}(X_n).$$

- After obtaining $\widehat{i_1}$, the process for detecting subsequent change points follows similarly:

$$\tilde{i}_k := \min\left\{ \widehat{i}_{k-1} + n_w \leq l \leq n - n_w + 1 : B^{l,n_w}_{max,10}(X_n) > C_{cp} \right\}.$$

Choice of Hyperparameters

- The proposed methods require selecting several hyperparameters:
  - α for mean structures
  - $\alpha, a_0, b_{0,ij}, b_{01,ij}$, and $b_{02,ij}$ for covariance structures
- We recommend setting $a_0 = b_{0,ij} = b_{01,ij} = b_{02,ij} = 0.01$ for all $1 \leq i \neq j \leq p$.
- The choice of α has a significant impact on performance.
- However, the theoretical conditions for α tend to be overly conservative in practice.

Choice of Hyperparameters

- We adopt the hyperparameter selection method from Lee et al. (2023), which controls the empirical False Positive Rate (FPR).
  1. Given an observed data matrix X, calculate the sample mean $\hat{\mu}$ and sample covariance matrix $\hat{\Sigma}$.
  2. Generate N simulated datasets $\{X_{\text{sim}}^{(s)} \in \mathbb{R}^{n\times p}\}_{s=1}^{N}$ using $\hat{\mu}$ and $\hat{\Sigma}$.
  3. For each s, calculate the mxPBF $B_{\max,10,\alpha}(X_{\text{sim}}^{(s)})$.
  4. Calculate the empirical FPR as
  $$\widehat{\text{FPR}}_{\alpha} = N^{-1}\sum_{s=1}^{N} I(B_{\max,10,\alpha}(X_{\text{sim}}^{(s)}) > C_{\text{th}}).$$
  5. Choose $\widehat{\alpha}$ that achieves a pre-specified FPR.
- We recommend setting the number of generated samples N between 300 and 500.

Multiscale Approach

- Selecting an appropriate window size is critical for performance and is case-specific.
- To reduce sensitivity to window size selection and ensure more robust performance, we propose a multiscale approach inspired by Avanesov and Buzun (2018).
- Assume we have a family of mxPBFs, $\{B^{l,n_w}_{\max,10} : n_w \in \mathcal{N}\}$, where $\mathcal{N}$ is a set of window sizes.

Multiscale Approach

- We integrate the results through a two-step process: grouping and averaging.
- First, we group the change points derived from each mxPBF.
  1. Let $\mathcal{I}_r = \{\widehat{i}_k : k = 1, \ldots, \widehat{K}_r\}$ be the set of estimated change points from the mxPBF with the r^{th} narrowest window $n_w^{(r)}$.
  2. For each $\widehat{i}_k \in \mathcal{I}_1$, construct candidate intervals
  $$[\widehat{i}_k - n_w^{(1)} + 1, \widehat{i}_k + n_w^{(1)} - 1].$$
  3. Identify the interval containing the most change points and group the points within this interval.
  4. Repeat Step 3 until all points are grouped.
  5. If any points remain ungrouped, repeat Steps 2-4 for $\mathcal{I}_r$ with $r \geq 2$.

Multiscale Approach

- Next, we obtain the final estimates through an averaging process.
  1. Remove groups containing fewer than $|\mathcal{N}|/2$ points.
  2. Calculate the average location of points within each remaining group.
- Since the final decision is based on a majority rule, we refer to this multiscale approach as mxPBF_major.
- We found that this approach achieves stable performance across a variety of scenarios, even when $\mathcal{N}$ includes inappropriate window sizes.

Consistency of mxPBF

- To support the proposed change point detection method, we establish the consistency of the mxPBF.
- Consistency of the mxPBF $B^{n_w}_{\text{max, 10}}$ implies that:
  1. $B^{n_w}_{\text{max, 10}}(X_n) \xrightarrow{p} 0$ under H_0
  2. $\{B^{n_w}_{\text{max, 10}}(X_n)\}^{-1} \xrightarrow{p} 0$ under H_1
- We first assume the following condition for the window size n_w:

Condition (A1)

$n_w \leq n/2$ and $\epsilon_{n_w} := \log(n_w \vee p)/n_w = o(1)$ as $n \to \infty$.

Consistency in Mean Structure

Theorem 2.1

Suppose we observe $X_i \overset{\text{ind.}}{\sim} N_p(\mu_{0i}, \Sigma_0)$, $i = 1, \dots, n$, and consider the hypothesis above. Assume a window size n_w satisfies condition (A1). Let $C_{2,\text{low}} = 2 + \epsilon$ and $C_2 = 2 + 2\epsilon$ for some small constant $\epsilon > 0$.

(i) Suppose H_0^{μ} is true.
if $p(n - 2n_w + 1)(n_w \vee p)^{-C_{2,\text{low}}} = o(1)$ and
$\alpha > 2[1 + 2\{C_2 \log(n_w \vee p)\}^{-1/2}]/(1 - 3\sqrt{C_2 \epsilon_{n_w}}) := \alpha_{n_w,p}$,

then $B^{\mu,n_w}_{\text{max},\,10}(X_n) = O_p\{(n_w \vee p)^{-c}\}$, for some $c > 0$.

(ii) Suppose H_1^{μ} is true. If there exist indices k and j such that
$\{i_k - i_{k-1}\} \wedge \{i_{k+1} - i_k\} \geq n_w$ and
$\frac{(\mu_{0i_k,j} - \mu_{0i_{k+1},j})^2}{\sigma_{0,jj}} \geq K \log(n_w \vee p)/n_w := \xi^2_{n_w,p} (\star)$, for some K,

then $\{B^{\mu,n_w}_{\text{max},\,10}(X_n)\}^{-1} = O_p\{(n_w \vee p)^{-c'}\}$, for some $c' > 0$.

Consistency in Mean Structure

- Enikeeva and Harchaoui (2019) proposed a test for change point detection in mean under sparse alternatives.
- Their test can detect a change point if:
  (1) $\|\mu_{0i_1} - \mu_{0i_2}\|_2^2 \geq C\sqrt{p \log(p \log n)}/\{i_1(1 - i_1/n)\}$ in the *moderate sparsity regime*.
  (2) $\|\mu_{0i_1} - \mu_{0i_2}\|_2^2 \geq C' s_0 \log p/\{i_1(1 - i_1/n)\}$ in the *high sparsity regime*, for some positive constants C and C'.
- In contrast, our condition $(\star)$ in Theorem 2.1 can be seen as:

$$\max_{1 \leq k \leq K_0} \|\mu_{0i_k} - \mu_{0i_k}\|_{\max}^2 \geq K^* \log(n_w \vee p)/n_w,$$

which is not directly comparable.

Localization Rate in Mean Structure

- Let $\mathcal{P}_\mu^*(K_0, \delta_n, \psi_{\min}^2, \Sigma_0)$ be the class of distributions X_n, where
  - $X_i \overset{\text{ind.}}{\sim} N_p(\mu_{0i}, \Sigma_0)$, $i = 1, \ldots, n$
  - K_0 change points $\{i_1, \ldots, i_{K_0}\}$ exist
  - $\delta_n = \min_{0 \le k \le K_0} \{i_{k+1} - i_k\}$
  - $\psi_{\min}^2 = \min_{0 \le k \le K_0} \max_{1 \le j \le p} \frac{(\mu_{0i_k,j} - \mu_{0i_{k+1},j})^2}{\sigma_{0,jj}}$

Theorem 2.2

Suppose $X_n \sim P \in \mathcal{P}_\mu^*(K_0, \delta_n, \psi_{\min}^2, \Sigma_0)$.
Assume that $n_w \le \delta_n$, $\psi_{\min}^2 \ge \xi_{n_w,p}^2$. Let $C_{1,\text{low}} = 1+\epsilon$ and $C_{2,\text{low}} = 2+\epsilon'$ for some small constants $\epsilon > 0$ and $\epsilon' > 0$.
If $K_0(n_w \vee p)^{-C_{1,\text{low}}} = o(1)$, $p(n - 2n_w + 1)(n_w \vee p)^{-C_{2,\text{low}}} = o(1)$, and $\alpha > \alpha_{n_w,p}$,
then $\mathbb{P}_0 \left(\widehat{K} = K_0,\ \max_{1 \le k \le K_0} |i_k - \widehat{i}_k| \le n_w \right) \longrightarrow 1$.

Localization Rate in Mean Structure

- Wang and Samworth (2018) proposed a change point estimation method for high-dimensional mean vectors based on a sparse projection.
- To obtain the localization rate for change point detection, they assumed:
$$\min_{1\le k\le K_0} \|\mu_{0i_k} - \mu_{0i_{k+1}}\|_2^2 \ge C\sigma^2 \frac{s_0 \log(np)}{n\tau^2},$$
where $\tau \le n^{-1} \min_{0\le k\le K_0}(i_{k+1} - i_k)$.
- This assumption implies:
$$\min_{1\le k\le K_0} \|\mu_{0i_k} - \mu_{0i_{k+1}}\|_{\max}^2 \ge C\sigma^2 \frac{\log(np)}{\delta_n \tau}.$$
- On the other hand, our condition $(\star)$ in Theorem 2.1 implies:
$$\min_{1\le k\le K_0} \|\mu_{0i_k} - \mu_{0i_{k+1}}\|_{\max}^2 \ge C_\star\sigma^2 \frac{\log(n_w \vee p)}{\delta_n},$$
for some constant $C_\star > 0$, which is a much weaker.

Localization Rate in Mean Structure

- We define the *localization rate* for change point detection as $\epsilon_n^\star$ if

$$\mathbb{P}_0\left(\widehat{K} = K_0,\ \max_{1\le k\le K_0} \frac{|i_k - \widehat{i}_k|}{n} \le \epsilon_n^\star\right) \longrightarrow 1.$$

Theorem 2.3

Let $\mathcal{P}_{\mu,1}^* = \mathcal{P}_{\mu,1}^*(K_0 = 1, \delta_n, \psi_{\min}^2, \Sigma_0)$ be the class of distributions of X_n, where only one change point i_1 exists and $\|\mu_{0i_1} - \mu_{0i_0}\|_0 = 1$. Suppose $1/\delta_n \le \psi_{\min}^2 \le 1$. Then,

$$\inf_{\widehat{i}_1} \sup_{P\in\mathcal{P}_{\mu,1}^*} \frac{1}{n}\mathbb{E}_P|\widehat{i}_1 - i_1| \ge \frac{1}{16n\psi_{\min}^2}.$$

- The lower bound in Theorem 2.3 becomes $\frac{n_w}{n\log(n_w \vee p)}$ if condition $(\star)$ in Theorem 2.1 holds.
- Thus, the rate obtained in Theorem 2.2 is nearly minimax-optimal.

Consistency in Covariance Structure

- We now introduce the following conditions to guarantee consistency of the mxPBF in covariance structure.

Condition (A2)

$\min_{i \neq j} \tau_{0,ij} \gg \{\log(n_w \vee p)\}^{-1}$.

Condition (A3)

There exist an index k and a pair (i, j) with $i \neq j$ satisfying

$$\{\log(n_w \vee p)\}^{-1} \ll \tau_{0i_k,ij} \wedge \tau_{0i_{k+1},ij} \quad \leq \quad \tau_{0i_k,ij} \vee \tau_{0i_{k+1},ij} \ll (n_w \vee p),$$

$$\frac{\tau_{0i_k,ij}}{\tau_{0i_{k+1},ij}} \vee \frac{\tau_{0i_{k+1},ij}}{\tau_{0i_k,ij}} \quad > \quad \frac{1 + C_{\mathrm{bm}}\sqrt{\epsilon_{n_w}}}{1 - 4\sqrt{C_1 \epsilon_{n_w}}}$$

for the constant $C_1 > 1$ and some constant $C_{\mathrm{bm}}^2 > 8(\alpha + 1)$.

Consistency in Covariance Structure

Theorem 3.1

Suppose we observe $X_i \overset{\text{ind.}}{\sim} N_p(0, \Sigma_{0i})$, $i = 1, \ldots, n$, and consider the hypothesis stated above. Assume that a window size n_w satisfies condition (A1). Let $C_{3,\text{low}} = 3 + \epsilon$ and $C_3 = 3 + 2\epsilon$ for some small constant $\epsilon > 0$.

(i) Suppose H_0^{Σ} is true.
If $p^2(n - 2n_w + 1)(n_w \vee p)^{-C_{3,\text{low}}} = o(1)$ and $\alpha > 6C_3$,
then $B^{\Sigma, n_w}_{\text{max},\,10}(X_n) = O_p\{(n_w \vee p)^{-c}\}$, for some $c > 0$.

(ii) Suppose H_1^{Σ} is true. If there exist k and a pair (i, j) such that $\{i_k - i_{k-1}\} \wedge \{i_{k+1} - i_k\} \geq n_w$ and condition (A3) holds,
then $\{B^{\Sigma, n_w}_{\text{max},\,10}(X_n)\}^{-1} = O_p\{(n_w \vee p)^{-c'}\}$, for some $c' > 0$.

Consistency in Covariance Structure

- Avanesov and Buzun (2018) proposed a testing procedure based on a bootstrap and multiscale approach for change point detection.
- They defined:
  - $\mathcal{N}$ be a set of window sizes.
  - n_- be the minimum window size in $\mathcal{N}$.
  - d be the maximum number of nonzero entries in any column of the true precision matrix.
- To prove the consistency of the proposed test, they assumed

  (1) $d|\mathcal{N}|^3\{\log(np)\}^{10} = o\left((s \wedge n_-^2)^{1/4}\right)$

  (2) $d|\mathcal{N}|\{\log(np)\}^7 = o\left((s \vee n_-)^{1/2}\right)$, where $s \leq n$ is the number of subsamples used to construct bootstrap statistics.

  (3) Sparsity and irrepresentability conditions of the precision matrix.

  which are stronger than those required for mxPBF.

Localization Rate in Covariance Structure

- Let $\mathcal{P}^*_\Sigma(K_0, \delta_n)$ be the class of distributions X_n, where
  - $X_i \overset{\text{ind.}}{\sim} N_p(0, \Sigma_{0i})$, $i = 1, \ldots, n$
  - K_0 change points $\{i_1, \ldots, i_{K_0}\}$ exist
  - $\delta_n = \min_{0 \le k \le K_0}\{i_{k+1} - i_k\}$
  - Condition (A3) holds

Theorem 3.2

Suppose $X_n \sim P \in \mathcal{P}^*_\Sigma(K_0, \delta_n)$.
Assume that $n_w \le \delta_n$ and condition (A1) holds. Let $C_{1,\text{low}} = 1 + \epsilon$, $C_{3,\text{low}} = 3 + \epsilon'$, and $C_3 = 3 + 2\epsilon'$ for some small constants $\epsilon > 0$ and $\epsilon' > 0$.
If $K_0(n_w \vee p)^{-C_{1,\text{low}}} = o(1)$, $p^2(n - 2n_w + 1)(n_w \vee p)^{-C_{3,\text{low}}} = o(1)$, and $\alpha > 6C_3$,
then, $\mathbb{P}_0\left(\widehat{K} = K_0,\ \max_{1 \le k \le K_0} |i_k - \widehat{i}_k| \le n_w\right) \longrightarrow 1.$

Localization Rate in Covariance Structure

- Wang et al. (2021) proposed two procedures to detect change points in covariance structures for sub-Gaussian random vectors.
- Their second procedure, called WBSIP, assumes
  - $p \leq \delta_n / \{8(\log n)^{1+\xi}\}$ (roughly)
  - For any $\xi > 0$, there exists a constant $C > 0$ such that

$$\min_k \|\Sigma_{0i_k} - \Sigma_{0i_{k+1}}\|^2 \geq \frac{Cp(\log n)^{1+\xi}}{n_w}.$$

- On the other hand, condition (A3) is satisfied if

$$\min_k \|\Sigma_{0i_k} - \Sigma_{0i_{k+1}}\|_{\max}^2 \geq \frac{C' \log(n_w \vee p)}{n_w},$$

  for some constant $C' > 0$.
- Their condition always implies ours in their settings, since

$$\|\Sigma_{0i_k} - \Sigma_{0i_{k+1}}\| \leq p \|\Sigma_{0i_k} - \Sigma_{0i_{k+1}}\|_{\max}.$$

Localization Rate in Covariance Structure

- Dette et al. (2022) developed a two-step approach to detect a change point in covariance structures when each component of an observation is a sub-Gaussian random variable.
- They defined:
  - i_1: the true change point
  - λ: a lower bound for the smallest nonzero entry in $|\Sigma_{0i_0} - \Sigma_{0i_1}|$
- To obtain their localization rate Δ/n, they assumed:
  - $p^2 n = o(e^{c'\lambda\Delta})$ for some constant c'.
  - $\lambda \geq C\left(\frac{\tau}{n}\right)^{1/2} \max\left\{\frac{n^2}{(n-i_1)^2}, \frac{n^2}{i_1^2}\right\}$, where τ is a threshold for dimension reduction in their method.
- If we further assume $c'\lambda > 3$, $\tau = C\log(n \vee p)$, and $i_1 \asymp n$, then their localization rate becomes $\log(n \vee p)/n$, which corresponds to ours.
- However, their condition is stronger than ours, as we consider the maximum nonzero entry in $|\Sigma_{0i_0} - \Sigma_{0i_1}|$.

Localization Rate in Covariance Structure

Theorem 3.3

For given constant $C_d > 0$ and $0 < \tilde{\sigma}_{\mathrm{diff}} < 3/4$, define a class of joint distributions of $(X_1, \ldots, X_n)$,

$$\begin{aligned} \mathcal{G}(\tilde{\sigma}_{\mathrm{diff}}, C_d) \quad := \quad & \Big\{ P : X_i \overset{ind.}{\sim} N_p(0, \Sigma_{0i}),\ i = 1, \ldots, n, \\ & \text{there is only } i_1 \text{ satisfying } \frac{C_d \log p}{\tilde{\sigma}^2_{\mathrm{diff}}} \le \delta_n \le \frac{n}{3}, \\ & \max_{1 \le i \le j \le p} \frac{(\sigma_{0i_0,ij} - \sigma_{0i_1,ij})^2}{\sigma_{0i_0,ii}\sigma_{0i_0,jj} + \sigma_{0i_1,ii}\sigma_{0i_1,jj}} \ge \tilde{\sigma}^2_{\mathrm{diff}} \Big\}. \end{aligned}$$

If $0 < C_d < 1/2$, then we have

$$\inf_{\widehat{i}} \sup_{P \in \mathcal{G}(\tilde{\sigma}_{\mathrm{diff}}, C_d)} \frac{1}{n} \mathbb{E}_P |\widehat{i} - i_1| \quad \ge \quad c$$

for some constant $c > 0$ and all sufficiently large p.

Localization Rate in Covariance Structure

- Theorem 3.3 implies that if $\tilde{\sigma}^2_{\text{diff}} < \frac{\log p}{2\delta_n}$, the localization rate is larger than some constant $c > 0$, at least in the worst case.

Theorem 3.4

If $C_d > 2$, then

$$\inf_{\widehat{i}} \sup_{P \in \mathcal{G}(\tilde{\sigma}_{\text{diff}}, C_d)} \mathbb{E}_P |\widehat{i} - i_1| \geq c\, \tilde{\sigma}^{-2}_{\text{diff}}$$

for some constant $c > 0$ and for all sufficiently large p.

- Under condition (A3), we can derive that $\tilde{\sigma}^2_{\text{diff}} \asymp \frac{\log(n_w \vee p)}{n_w}$.
- Thus, the localization rate in Theorem 3.2 is nearly minimax-optimal, since $\tilde{\sigma}^{-2}_{\text{diff}} n^{-1}$ has the same rate as $\frac{n_w}{n \log(n_w \vee p)}$.

Simulation Study

- We compare our methods with those from:
  - Dette (Dette et al., 2022)
  - Geomcp (Grundy et al., 2020)
  - Inspect (Wang and Samworth, 2018)
  - Edivisive (Matteson and James, 2014)
- Settings for mxPBF:
  - Set of window sizes $\mathcal{N} = \{25, 60, 100\}$
  - Number of generated samples $N = 300$
  - Prespecified FPR $= 0.05$
  - Threshold $C_{\text{th}} = 10$
- 50 replications for each scenario.
- Evaluation metrics: F1 score, Hausdorff distance, and covering metric.

Performance with Data without Change Points

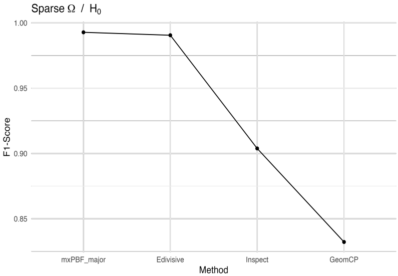
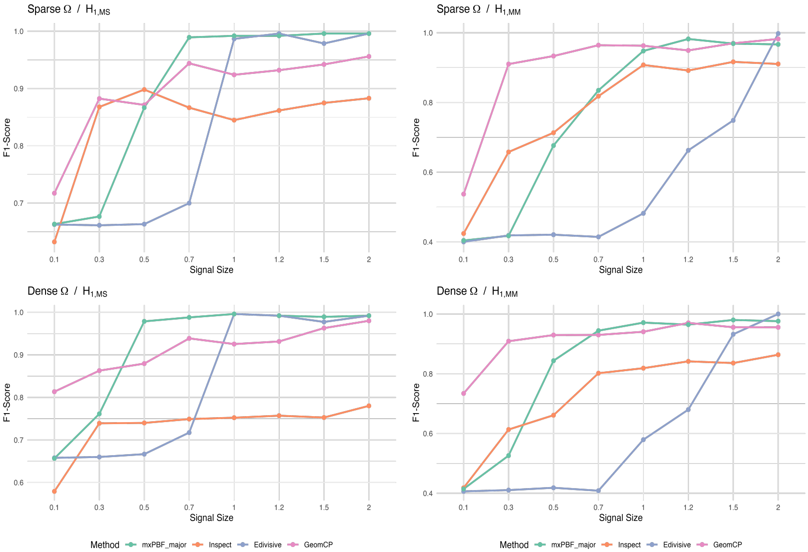
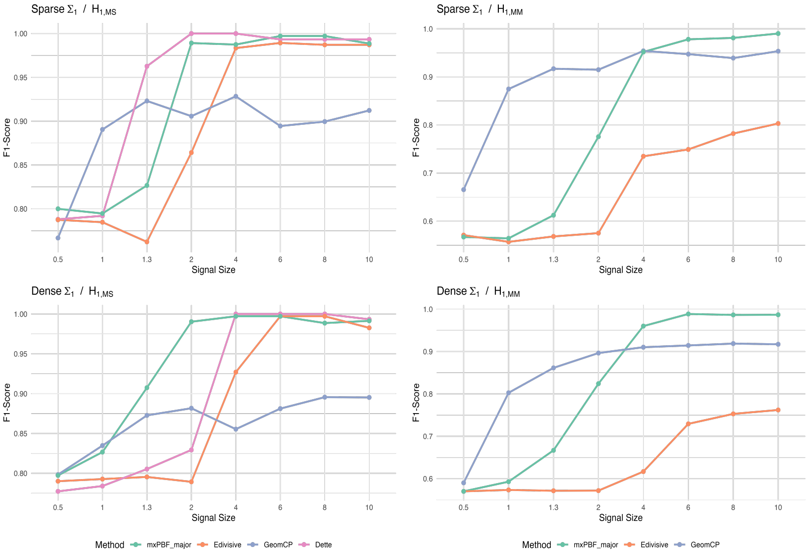


Figure: F-scores for change point detection methods computed using 50 simulated data sets for H_0 with $p = 200$.

Performance with Data Including CPs in Mean



Performance with Data Including CPs in Covariance

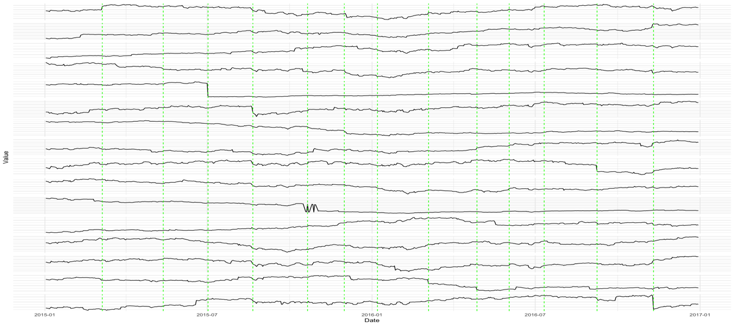


Real Data Analysis

- We analyze two datasets scaled by the median absolute deviation (MAD):
  - The closing stock prices for companies in the S&P 500:
    - Daily log returns from January 2015 to December 2016
    - 504×470 matrix
    - scaled median absolute deviation
  - The comparative genomic hybridization microarray dataset:
    - Fluorescence intensity levels of DNA fragments
    - 2215×43 matrix
- Utilized the same settings for mxPBF in the simulation study.

S&P 500 Stock Prices Data

- Observed change points correspond to events such as:
  - Major declines in Chinese stock markets in August 2015 and January 2016
  - Impact of Brexit in July 2016
  - U.S. presidential election in November 2016
  - Toyota's recall of 5 million cars on May 13, 2015
  - $190 million penalty imposed on Wells Fargo on September 8, 2016



Comparative Genomic Hybridization Data

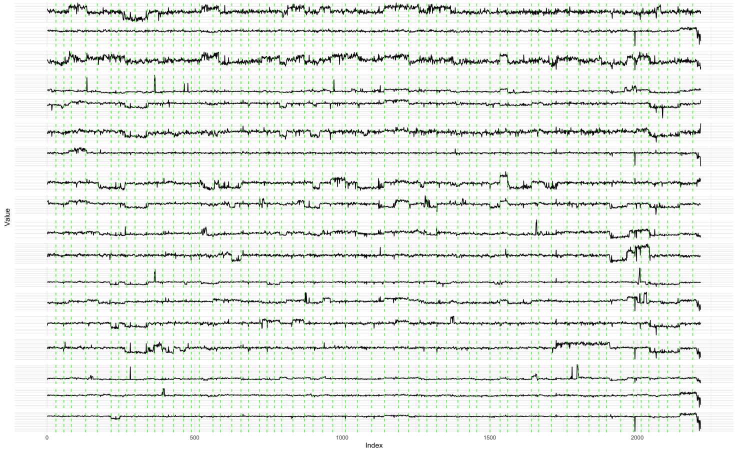


Figure: Log-intensity ratio measurements from microarray data of 18 out of 43 individuals, with vertical lines marking the identified change points.

References I

Avanesov, V. and Buzun, N. (2018). Change-point detection in high-dimensional covariance structure. *Electronic Journal of Statistics*, 12(2):3254–3294.

Dette, H., Pan, G., and Yang, Q. (2022). Estimating a change point in a sequence of very high-dimensional covariance matrices. *Journal of the American Statistical Association*, 117(537):444–454.

Enikeeva, F. and Harchaoui, Z. (2019). High-dimensional change-point detection under sparse alternatives. *The Annals of Statistics*, 47(4):2051–2079.

Grundy, T., Killick, R., and Mihaylov, G. (2020). High-dimensional changepoint detection via a geometrically inspired mapping. *Statistics and Computing*, 30(4):1155–1166.

Lee, K., Lin, L., and Dunson, D. (2021). Maximum pairwise Bayes factors for covariance structure testing. *Electronic Journal of Statistics*, 15(2):4384 – 4419.

Lee, K., You, K., and Lin, L. (2023). Bayesian optimal two-sample tests for high-dimensional gaussian populations. *Bayesian Analysis*, 1(1):1–25.

References II

Matteson, D. S. and James, N. A. (2014). A nonparametric approach for multiple change point analysis of multivariate data. *Journal of the American Statistical Association*, 109(505):334–345.

Wang, D., Yu, Y., and Rinaldo, A. (2021). Optimal covariance change point localization in high dimensions. *Bernoulli*, 27(1):554–575.

Wang, T. and Samworth, R. J. (2018). High dimensional change point estimation via sparse projection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(1):57–83.