A Proofs

A.1 Proofs for Section 3

$$(\Sigma)^{-1} = \begin{pmatrix} E^{-1} & -E^{-1}BD^{-1} \\ -D^{-1}CE^{-1} & D^{-1}+D^{-1}CE^{-1}BD^{-1} \end{pmatrix}$$
$$= \begin{pmatrix} A^{-1}+A^{-1}BF^{-1}CA^{-1} & -A^{-1}BF^{-1} \\ -F^{-1}CA^{-1} & F^{-1} \end{pmatrix},$$
$$(\Sigma_t^{'})^{-1} = \begin{pmatrix} A^{-1} & 0 \\ 0 & D^{-1} \end{pmatrix}$$

Theorem A.1 (Hoeffding bound)**.** *Let* $Z_1,Z_2,...,Z_n$ *be independent random variables bounded by the interval* $[a,b]$*. We define the empirical mean of these variables by*

$$\bar{Z} = \frac{1}{n}(Z_1 + Z_2 + ... + Z_n).$$

Then

$$P(\|\bar{Z} - E[\bar{Z}]\| \geq \epsilon) \leq \exp \frac{-2n^2\epsilon^2}{\sum_i (b_i - a_i)^2}.$$

Proof of Lemma **??**. We will first show the bound of quadratic terms, $X^T(\Sigma^{-1} - (\Sigma')^{-1})X$. For a positive semi-definite matrix M, its eigenvectors are orthonormal to each other and all of its eigenvalues are nonnegative. By performing eigenvalue decomposition

$$X^T M X = X^T Q \Lambda Q^T X = \sum_k \lambda_k < X, q_k >^2$$

where q_k is the kth column of Q. Since eigenvectors are orthonormal to each other

$$\sum_k < X, q_k >^2 = X^T Q Q^T x = < X, X > = ||X||^2$$

Therefore,

$$\lambda_{\text{min}}(M) V^2 n \leq X^T M X \leq \lambda_{\text{max}}(M) V^2 n$$

where $\lambda_{\text{min}}(M)$ and $\lambda_{\text{max}}(M)$ represent the smallest and biggest eigenvalue of M respectively.

As Σ is symmetric, its inverse $(\Sigma)^{-1}$ is also symmetric and positive semi-definite. Substituting M by Σ^{-1} gives

$$\frac{1}{\lambda_{\text{max}}(\Sigma)} V^2 n \leq X^T(\Sigma)^{-1}X \leq \frac{1}{\lambda_{\text{min}}(\Sigma)} V^2 n$$

where we use the fact that eigenvalues of inverse matrix is inverse of eigenvalues.

Similarly for $(\Sigma')^{-1}$,

$$\frac{1}{\lambda_{\text{max}}(\Sigma')} V^2 n \leq X^T(\Sigma')^{-1}X \leq \frac{1}{\lambda_{\text{min}}(\Sigma')} V^2 n$$

Then the difference of quadratic terms can be bounded

$$\left(\frac{1}{\lambda_{\text{max}}(\Sigma)} - \frac{1}{\lambda_{\text{min}}(\Sigma')}\right) V^2 n \leq X^T((\Sigma)^{-1} - (\Sigma')^{-1})X \leq \left(\frac{1}{\lambda_{\text{min}}(\Sigma)} - \frac{1}{\lambda_{\text{max}}(\Sigma')}\right) V^2 n$$

For determinant terms, we can use the property of determinant that for any squared matrix N, the determinant of N is the product of all its eigenvalues. Thus determinant of $n \times n$ matrix Σ is bounded as

$$(\lambda_{min}(\Sigma))^n \leq |\Sigma| = \prod \lambda_i \leq (\lambda_{max}(\Sigma))^n$$

and similarly,

$$(\lambda_{min}(\Sigma'))^n \leq |\Sigma'| \leq (\lambda_{max}(\Sigma'))^n.$$

Then the log determinant ratio can be bounded as

$$n(\lambda_{min}(\Sigma) - \lambda_{max}(\Sigma')) \leq \ln\left(\frac{|\Sigma|}{|\Sigma'|}\right) \leq n(\lambda_{max}(\Sigma) - \lambda_{min}(\Sigma'))$$

Combining them, we can conclude

$$(C_1 V^2 + C_3) n \leq \mathfrak{T}_{GLRT} \leq (C_2 V^2 + C_4) n$$

with $C_1 = \left(\frac{1}{\lambda_{\text{max}}(\Sigma)} - \frac{1}{\lambda_{\text{min}}(\Sigma')}\right)$, $C_2 = \left(\frac{1}{\lambda_{\text{min}}(\Sigma)} - \frac{1}{\lambda_{\text{max}}(\Sigma')}\right)$, $C_3 = \lambda_{min}(\Sigma) - \lambda_{max}(\Sigma')$, $C_4 = \lambda_{max}(\Sigma) - \lambda_{min}(\Sigma')$. $\square$

Proof of Lemma C.1. Suppose $Z_1, Z_2, ..., Z_m$ are m samples of $2\mathfrak{L}$ and $\bar{Z} = \frac{1}{m}(Z_1 + Z_2 + ... + Z_m)$. As $X_1, X_2, ..., X_m$ are Gaussian random variables with mean 0, $E[\bar{Z}] = \ln(|I - K_{ab}K_{bb}^{-1}K_{ba}K_{aa}^{-1}|)$ as

$$\begin{aligned} E(2\mathfrak{L}|\mathbb{H}_0) &= E\left[X^T(\Sigma_n)^{-1}X - X^T(\Sigma_n^{'})^{-1}X + \ln\left(\frac{|\Sigma_n|}{|\Sigma_n^{'}|}\right)\right] \\ &= E\left[Tr(XX^T(\Sigma_n)^{-1}) - Tr(XX^T(\Sigma_n^{'})^{-1}) + \ln\left(\frac{|\Sigma_n|}{|\Sigma_n^{'}|}\right)\right] \\ &= Tr(\Sigma_n(\Sigma_n)^{-1}) - Tr(\Sigma_n(\Sigma_n^{'})^{-1}) + \ln\left(\frac{|\Sigma_n|}{|\Sigma_n^{'}|}\right) \\ &= n - n + \ln(|I - K_{ab}K_{bb}^{-1}K_{ba}K_{aa}^{-1}|) \\ &= \ln(|I - K_{ab}K_{bb}^{-1}K_{ba}K_{aa}^{-1}|). \end{aligned}$$

$$\begin{aligned} P(2\mathfrak{L} \geq \mathfrak{R}_{n,\delta,\mathbb{H}_0}|\mathbb{H}_0) &= P(\bar{Z} \geq nC\sqrt{-1/(2m)\log\delta} + \ln(|I - K_{ab}K_{bb}^{-1}K_{ba}K_{aa}^{-1}|)) \\ &= P(\bar{Z} \geq nC\sqrt{-1/(2m)\log\delta} + E[\bar{Z}]) \\ &= P(\bar{Z} - E[\bar{Z}] \geq nC\sqrt{-1/(2m)\log\delta}) \end{aligned}$$

where $nC = n((C_2 - C_1)V^2 + (C_4 - C_3))$ is the maximal range of $\bar{Z}$ from Lemma **??**. Then from the Theorem A.1 and letting $m = 1$ sample of likelihood ratio,

$$\begin{aligned} P(2\mathfrak{L} - E[2\mathfrak{L}] \geq nC\sqrt{-1/2\log\delta}) &\leq e^{\log\delta}. \\ &= \delta \end{aligned}$$

In conclusion for $\mathfrak{R}_{n,\delta,\mathbb{H}_0} = \ln(|I - K_{ab}K_{bb}^{-1}K_{ba}K_{aa}^{-1}|) + nC\sqrt{-1/2\log\delta})$,

$$P(2\mathfrak{L} \geq \mathfrak{R}_{n,\delta,\mathbb{H}_0}) \leq \delta$$

□

Proof of Lemma C.2. Suppose $Z_1, Z_2, ..., Z_m$ are m samples of $2\mathfrak{L}$ and $\bar{Z} = \frac{1}{m}(Z_1 + Z_2 + ... + Z_m)$. $E[\bar{Z}] = Tr(K_{ab}F^{-1}K_{ba}K_{aa}^{-1}) + Tr(K_{ba}E^{-1}K_{ab}K_{bb}^{-1}) + \ln(|I - K_{ab}K_{bb}^{-1}K_{ba}K_{aa}^{-1}|)$ as

$$\begin{aligned} E(2\mathfrak{L}|\mathbb{H}_1) &= E\left[X^T(\Sigma_n)^{-1}X - X^T(\Sigma_n^{'})^{-1}X + ln\left(\frac{|\Sigma_n|}{|\Sigma_n^{'}|}\right)\right] \\ &= E\left[Tr(XX^T(\Sigma_n)^{-1}) - Tr(XX^T(\Sigma_n^{'})^{-1}) + ln\left(\frac{|\Sigma_n|}{|\Sigma_n^{'}|}\right)\right] \\ &= Tr(\Sigma_n^{'}(\Sigma_n)^{-1}) - Tr(\Sigma_n^{'}(\Sigma_n^{'})^{-1}) + ln\left(\frac{|\Sigma_n|}{|\Sigma_n^{'}|}\right) \\ &= Tr(I + K_{ab}F^{-1}K_{ba}K_{aa}^{-1} + Tr(I + K_{ba}E^{-1}K_{ab}K_{bb}^{-1}) - n + \ln(|I - K_{ab}K_{bb}^{-1}K_{ba}K_{aa}^{-1}|) \\ &= Tr(K_{ab}F^{-1}K_{ba}K_{aa}^{-1}) + Tr(K_{ba}E^{-1}K_{ab}K_{bb}^{-1}) + \ln(|I - K_{ab}K_{bb}^{-1}K_{ba}K_{aa}^{-1}|). \end{aligned}$$

$$\begin{aligned} P(2\mathfrak{L} \leq \mathfrak{R}_{n,\delta,\mathbb{H}_1}|\mathbb{H}_1) &= P(\bar{Z} \leq Tr(K_{ab}F^{-1}K_{ba}K_{aa}^{-1}) + Tr(K_{ba}E^{-1}K_{ab}K_{bb}^{-1}) + \ln(|I - K_{ab}K_{bb}^{-1}K_{ba}K_{aa}^{-1}|) - nC\sqrt{-1/(2m)\log\delta}) \\ &= P(\bar{Z} \leq -nC\sqrt{-1/(2m)\log\delta} + E[\bar{Z}]) \\ &= P(E[\bar{Z}] - \bar{Z} \geq nC\sqrt{-1/(2m)\log\delta}) \end{aligned}$$

where $C = (C_2 - C_1)V^2 + (C_4 - C_3)$ is the maximal range of $\bar{Z}$ from Lemma **??**. Then from the TheoremA.1,

$$\begin{aligned} P(E[\bar{Z}] - \bar{Z} \geq nC\sqrt{-1/(2m)\log\delta}) &\leq e^{\log\delta}. \\ &= \delta \end{aligned}$$

In conclusion,

$$P(2\mathfrak{L} \leq \mathfrak{R}_{n,\delta,\mathbb{H}_1}) \leq \delta$$

□

Proof of Theorem C.1. There are four cases of inequalities that $\mathfrak{R}^0_{n,\delta,H_0}$,$\mathfrak{R}^1_{n,\delta,H_0}$,$\mathfrak{R}^0_{n,\delta,H_1}$,$\mathfrak{R}^1_{n,\delta,H_1}$ can have. Here we will show that in each case the type I error probability is bounded by δ.

Case 1: $\mathfrak{R}^0_{n,\delta,H_0} \le \mathfrak{R}^0_{n,\delta,H_1}$ *and* $\mathfrak{R}^1_{n,\delta,H_0} \le \mathfrak{R}^1_{n,\delta,H_1}$

$$\begin{aligned}
\mathbb{P}(\mathfrak{T}^0_{GLRT} = 0 \wedge \mathfrak{T}^1_{GLRT} = 1|\mathbb{H}_0) &= \mathbb{P}((\mathfrak{R}^0_{n,\delta,H_1} \le 2\mathfrak{L} \le \mathfrak{R}^0_{n,\delta,H_0}) \vee (\mathfrak{R}^1_{n,\delta,H_0} \le 2\mathfrak{L} \le \mathfrak{R}^1_{n,\delta,H_1})|\mathbb{H}_0) \\
&= \mathbb{P}(\mathfrak{R}^1_{n,\delta,H_0} \le 2\mathfrak{L} \le \mathfrak{R}^1_{n,\delta,H_1})|\mathbb{H}_0) \\
&\le \mathbb{P}(\mathfrak{R}^1_{n,\delta,H_0} \le 2\mathfrak{L}|\mathbb{H}_0) \\
&\le \delta.
\end{aligned}$$

Case 2: $\mathfrak{R}^0_{n,\delta,H_0} \le \mathfrak{R}^0_{n,\delta,H_1}$ *and* $\mathfrak{R}^1_{n,\delta,H_0} \ge \mathfrak{R}^1_{n,\delta,H_1}$

$$\begin{aligned}
\mathbb{P}(\mathfrak{T}^0_{GLRT} = 0 \wedge \mathfrak{T}^1_{GLRT} = 1|\mathbb{H}_0) &= \mathbb{P}((\mathfrak{R}^0_{n,\delta,H_1} \le 2\mathfrak{L} \le \mathfrak{R}^0_{n,\delta,H_0}) \vee (\mathfrak{R}^1_{n,\delta,H_0} \le 2\mathfrak{L} \le \mathfrak{R}^1_{n,\delta,H_1})|\mathbb{H}_0) \\
&= 0 \\
&\le \delta.
\end{aligned}$$

Case 3: $\mathfrak{R}^0_{n,\delta,H_0} \ge \mathfrak{R}^0_{n,\delta,H_1}$ *and* $\mathfrak{R}^1_{n,\delta,H_0} \le \mathfrak{R}^1_{n,\delta,H_1}$

$$\begin{aligned}
\mathbb{P}(\mathfrak{T}^0_{GLRT} = 0 \wedge \mathfrak{T}^1_{GLRT} = 1|\mathbb{H}_0) &= \mathbb{P}((\mathfrak{R}^0_{n,\delta,H_1} \le 2\mathfrak{L} \le \mathfrak{R}^0_{n,\delta,H_0}) \vee (\mathfrak{R}^1_{n,\delta,H_0} \le 2\mathfrak{L} \le \mathfrak{R}^1_{n,\delta,H_1})|\mathbb{H}_0) \\
&\le \mathbb{P}((2\mathfrak{L} \le \mathfrak{R}^0_{n,\delta,H_0}) \vee (\mathfrak{R}^1_{n,\delta,H_0} \le 2\mathfrak{L})|\mathbb{H}_0) \\
&\le \delta.
\end{aligned}$$

Case 4: $\mathfrak{R}^0_{n,\delta,H_0} \ge \mathfrak{R}^0_{n,\delta,H_1}$ *and* $\mathfrak{R}^1_{n,\delta,H_0} \ge \mathfrak{R}^1_{n,\delta,H_1}$

$$\begin{aligned}
\mathbb{P}(\mathfrak{T}^0_{GLRT} = 0 \wedge \mathfrak{T}^1_{GLRT} = 1|\mathbb{H}_0) &= \mathbb{P}((\mathfrak{R}^0_{n,\delta,H_1} \le 2\mathfrak{L} \le \mathfrak{R}^0_{n,\delta,H_0}) \vee (\mathfrak{R}^1_{n,\delta,H_0} \le 2\mathfrak{L} \le \mathfrak{R}^1_{n,\delta,H_1})|\mathbb{H}_0) \\
&= \mathbb{P}(\mathfrak{R}^0_{n,\delta,H_1} \le 2\mathfrak{L} \le \mathfrak{R}^0_{n,\delta,H_0}|\mathbb{H}_0) \\
&\le \mathbb{P}((2\mathfrak{L} \le \mathfrak{R}^0_{n,\delta,H_0})|\mathbb{H}_0) \\
&\le \delta.
\end{aligned}$$

Marginalizing all the cases, overall $\mathbb{P}(\mathfrak{T}^0_{GLRT}=0 \wedge \mathfrak{T}^1_{GLRT}=1|\mathbb{H}_0)$ is upper bounded by δ.

□

B Detection of Covariance Structure Changes with Unknown Hyperparameters

B.1 Case 1: Variance Change

We are going to write the covariance for the null hypothesis as Σ_n, and the covariance matrix for the alternative hypothesis that has change point at time step t as $\Sigma'_{n,t}$. For the simplest case of covariance structure change, we can think about the case where each data point is independently distributed, i.e., covariance of any two different points is zero. Then the covariance matrix becomes diagonal matrix. We let $\Sigma_n = aI$ for the null hypothesis where there is no change in variance, and $\Sigma'_{n,t} = (I-U)aI + UbI$ for the alternative hypothesis with change at time step t where U is a diagonal matrix with $U_{i,i} = u(i-t)$ for unit step function u. Visualizing the covariance matrices,

$$\Sigma_n = \begin{pmatrix} aI & 0 \\ 0 & aI \end{pmatrix}, \Sigma'_{n,t} = \begin{pmatrix} aI & 0 \\ 0 & bI \end{pmatrix}.$$

In the case of covariance structure change with diagonal covariance matrix, the generalized likelihood ratio can be written as

$$\begin{aligned}
2\mathfrak{L} &= \max_{t \in \mathcal{C}_{n,\alpha}} \left[X^T(\Sigma_n)^{-1}X - X^T(\Sigma'_n)^{-1}X + \ln\left(\frac{|\Sigma_n|}{|\Sigma'_n|}\right) \right] \\
&= \max_{t \in \mathcal{C}_{n,\alpha}} \left[\sum_{i=t+1}^{n} (X_i)^2 \left(\frac{1}{a} - \frac{1}{b}\right) + (n-t)\ln\left(\frac{a}{b}\right) \right].
\end{aligned}$$

Taking derivative of likelihood ratio with respect to b and set to zero we can get be as follows,

$$\begin{aligned}
\frac{\partial}{\partial b} 2\mathfrak{L} &= \max_{t \in \mathcal{C}_{n,\alpha}} \left[\sum_{i=t+1}^{n} (X_i)^2 \frac{1}{b^2} - (n-t)\frac{1}{b} \right] = 0 \\
b &= \frac{1}{n-t} \sum_{i=t+1}^{n} (X_i)^2.
\end{aligned}$$

Plugging b in the likelihood, we can get

$$2\mathfrak{L} = \max_{t\in\mathcal{C}_{n,\alpha}} \left[\sum_{i=t+1}^{n} \left(\frac{X_i^2}{a} - 1 \right) + (n-t)\ln\left(\frac{a(n-t)}{\sum_{i=t+1}^{n} X_i^2} \right) \right].$$

The first summation term in the equation can be interpreted as the sum of differences between the data in the second part and the original variance, which increases as t decreases. The second log term indicates the ratio between original variance and the changed variance. This value is big when t is small while the ratio between two variances is big.

B.2 Case 2: Covariance structure change with preserved correlation

For the case where the variance changes while preserving correlation, the inverse of the covariance matrix Σ_t under the null hypothesis, and the inverse of the covariance matrix Σ_t' under the alternative hypothesis can be represented as follows.

$$\Sigma_t^{-1} = \begin{pmatrix} \mathrm{A} & \mathrm{B} \\ \mathrm{B}^\mathrm{T} & \mathrm{C} \end{pmatrix}, (\Sigma_t')^{-1} = \begin{pmatrix} \mathrm{A} & \alpha^{-1}\mathrm{B} \\ \alpha^{-1}\mathrm{B}^\mathrm{T} & \alpha^{-2}\mathrm{C} \end{pmatrix}.$$

To formulate the detection of the change in the covariance matrix, for a fixed $t\in\mathcal{C}_n$, the associated alternative hypothesis with respect to t can be given as,

$$\mathbb{H}_{1,t} : \exists\, \alpha \neq 0, \mathrm{Cov}(X_i, X_j) = \begin{cases} K(X_i,X_j), & i,j<t \\ \alpha^2 K(X_i,X_j), & i,j\geq t \\ \alpha K(X_i,X_j), & \text{otherwise,} \end{cases} \tag{9}$$

where $K(X_i,X_j)$ denotes covariance between X_i and X_j under $\mathbb{H}_0$. Since t is unknown, the alternative hypothesis is given by taking the union of $\mathbb{H}_{1,t}$. Thus, the composite hypothesis testing problem is given by

$$\mathbb{H}_0 : \mathrm{Cov}(X_i, X_j) = K(X_i, X_j), \;\; v.s. \;\; \mathbb{H}_1 = \bigcup_{t\in\mathcal{C}_n} \mathbb{H}_{1,t} \tag{10}$$

We find α to maximize the likelihood ratio.

For convenience we may write Σ_t, the covariance matrices under $\mathbb{H}_0$, and Σ_t', the covariance matrix under $\mathbb{H}_1$ as

$$\Sigma_t = \begin{pmatrix} \mathrm{P} & \mathrm{Q} \\ \mathrm{Q}^\mathrm{T} & \mathrm{R} \end{pmatrix}, \Sigma_t' = \begin{pmatrix} \mathrm{P} & \alpha\mathrm{Q} \\ \alpha\mathrm{Q}^\mathrm{T} & \alpha^2\mathrm{R} \end{pmatrix}.$$

Assuming A and $D-CA^{-1}B$ be nonsingular, we can write inverse of Σ_t and Σ_t' as

$$\Sigma_t^{-1} = \begin{pmatrix} \mathrm{A} & \mathrm{B} \\ \mathrm{B}^\mathrm{T} & \mathrm{C} \end{pmatrix}, (\Sigma_t')^{-1} = \begin{pmatrix} \mathrm{A} & \alpha^{-1}\mathrm{B} \\ \alpha^{-1}\mathrm{B}^\mathrm{T} & \alpha^{-2}\mathrm{C} \end{pmatrix}.$$

where

$$\begin{aligned} A &= P^{-1} + P^{-1}Q(R - Q^\mathrm{T}P^{-1}Q)^{-1}Q^\mathrm{T}P^{-1} \\ B &= -P^{-1}Q(R - Q^\mathrm{T}P^{-1}Q)^{-1} \\ C &= (R - Q^\mathrm{T}P^{-1}Q)^{-1} \end{aligned}$$

with A, P $\in\mathbb{R}^{t\times t}$, B, Q $\in\mathbb{R}^{t\times(n-t)}$, C, R $\in\mathbb{R}^{(n-t)\times(n-t)}$.

Then likelihood ratio can be formulated as

$$\begin{aligned} 2\mathfrak{L} &= \max_{t\in\mathcal{C}_{n,\alpha}} \left[X^T(\Sigma_n)^{-1}X - X^T(\Sigma_n^{'})^{-1}X + \ln\left(\frac{|\Sigma_n|}{|\Sigma_n'|} \right) \right] \\ &= \max_{t\in\mathcal{C}_{n,\alpha}} \left[\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}^T \begin{pmatrix} 0 & (1-\frac{1}{\alpha})\mathrm{B} \\ (1-\frac{1}{\alpha})\mathrm{B}^\mathrm{T} & (1-\frac{1}{\alpha^2})\mathrm{C} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} - 2(n-t)\ln\alpha \right] \\ &= \max_{t\in\mathcal{C}_{n,\alpha}} \left[2(1-\frac{1}{\alpha})X_1^TBX_2 + (1-\frac{1}{\alpha^2})X_2^TCX_2 - 2(n-t)\ln\alpha \right], \end{aligned}$$

where $X_1{=}X_{1:t-1}$, $X_2{=}X_{t:n}$. We find optimal α, which makes the derivative of likelihood ratio with respect to α be zero.

$$\frac{\partial 2\mathfrak{L}}{\partial\alpha} = \frac{2}{\alpha^2}X_1^TBX_2 + \frac{2}{\alpha^3}X_2^TCX_2 - \frac{2(n-t)}{\alpha} = 0$$

Guaranteed that $\alpha>0$, multiplying α^3 to both side,

$$(n-t)\alpha^2 - X_1^T B X_2 \alpha - X_2^T C X_2 = 0.$$

Solving the equation we can get

$$\alpha = \frac{b+\sqrt{b^2+4ac}}{2a},$$

where $a=n-t$,$b=X_1^T BX_2$, $c=X_2^T CX_2$. To calculate α dynamically as n increases, we need to keep track of b and c

$$X_1^T B X_2 = X_1^T B X_2 + X_{n+1}\sum_{k=1}^{t} X_k K(n+1,k)$$

$$X_2^T C X_2 = X_2 C^T X_2 + 2X_{n+1}\sum_{k=n-t+1}^{n} K(n+1,k) + X_{n+1}^2 K(n+1,n+1).$$

Computational complexity of the likelihood ratio is determined by computational complexity of matrix inversion, matrix multiplication, and calculating matrix determinant which are known to be $O(n^{2.373})$. This will be repeated for maximum of n times, thus the total computational complexity of the likelihood ratio is $O(n^{3.373})$.

C A Structural Break with a Kernel

Here, we consider two identical GPs. Formally, the hypotheses can be written as $\mathbb{H}_0$:$\mathrm{Cov}(X_i,X_j)=K(i,j)$ and $\mathbb{H}_1=\bigcup_{t\in\mathcal{C}_n}\mathbb{H}_{1,t}$, where the specific alternative hypothesis with change point t, $\mathbb{H}_{1,t}$, is defined as

$$\mathbb{H}_{1,t} : \mathrm{Cov}(X_i, X_j) = \begin{cases} K(i,j), & i,j<t \\ K(i,j), & i,j\geq t \\ 0, & \text{otherwise.} \end{cases}$$

The corresponding covariance matrices are be written as

$$\Sigma = \begin{pmatrix} K_{aa} & K_{ab} \\ K_{ba} & K_{bb} \end{pmatrix}, \Sigma_t' = \begin{pmatrix} K_{aa} & 0 \\ 0 & K_{bb} \end{pmatrix}.$$

Here, K_{ij} indicates the covariance matrix between $X_a{:=}X_{1:t}$ and $X_b{:=}X_{t+1:n}$ with kernel K.

The likelihood ratio in Equation 4 is written as $\max_{t\in\mathcal{C}_n,\alpha}\left[X^T\bar{\Sigma}X+\ln(|I+E-K_{aa}|)\right]$, where

$$\bar{\Sigma} = \begin{pmatrix} K_{aa}^{-1}K_{ab}F^{-1}K_{ba}K_{aa}^{-1} & -E^{-1}K_{ab}K_{bb}^{-1} \\ -K_{bb}^{-1}K_{ba}E^{-1} & K_{bb}^{-1}K_{ba}E^{-1}K_{ab}K_{bb}^{-1} \end{pmatrix},$$

$E=K_{aa}-K_{ab}K_{bb}^{-1}K_{ba}$, and $F=K_{bb}-K_{ba}K_{aa}^{-1}K_{ab}$, and is conditioned such that K_{aa}, K_{dd}, E, and F are nonsingular.[4]

We now define a likelihood ratio test as $\mathfrak{T}_{GLRT}=\mathbb{I}(2\mathfrak{L}\leq\mathfrak{R}_\delta)$. We reject $\mathbb{H}_0$ if $\mathfrak{T}_{GLRT}=1$ and accept $\mathbb{H}_0$ if $\mathfrak{T}_{GLRT}=0$.

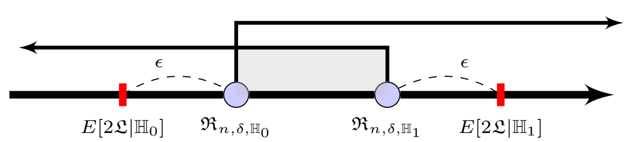

Figure 3: Range of thresholds that guarantee bounded type I (arrow pointing right) and type II (arrow pointing left) errors.

Lemma C.1. *The probability that the GLRT $\mathfrak{T}_{GLRT}$ is correct under the null hypothesis (absence of CP) is at least $1-\delta$,*

$$P(2\mathfrak{L}\geq\mathfrak{R}_{n,\delta,\mathbb{H}_0}|\mathbb{H}_0)\leq\frac{\delta}{2},$$

with $\mathfrak{R}_{n,\delta,\mathbb{H}_0}=\max_t\left(n-Tr(\Sigma(\Sigma_t')^{-1})+\ln\left(\frac{|\Sigma|}{|\Sigma_t'|}\right)\right)+(C_{0,2}-C_{0,1})\sqrt{0.5\ln(2n/\delta)}$*,*

$C_{0,1}=\left(\frac{1}{\lambda_{max}(\Sigma)}-\frac{1}{\lambda_{min}(\Sigma)\wedge\lambda_{min}(\Sigma_t')}\right)$ *and*

$C_{0,2}=\left(\frac{1}{\lambda_{min}(\Sigma)}-\frac{1}{\lambda_{max}(\Sigma)\vee\lambda_{max}(\Sigma_t')}\right)$.

[4]Similarly, using the property of block matrices, $\frac{|\Sigma|}{|\Sigma_t'|}=\frac{|K_{aa}|\cdot|F|}{|K_{aa}|\cdot|K_{bb}|}=|I+E-K_{aa}|$.

Lemma C.2. *The probability that the GLRT $\mathfrak{T}_{GLRT}$ is correct under the alternative hypothesis (existence of CP) is at least $1-\delta$,*

$$P(2\mathfrak{L} \leq \mathfrak{R}_{n,\delta,\mathbb{H}_1} | \mathbb{H}_1) \leq \frac{\delta}{2},$$

with $\mathfrak{R}_{n,\delta,\mathbb{H}_0} = \min_t \left(Tr(\Sigma_t'(\Sigma)^{-1}) - n + \ln\left(\frac{|\Sigma|}{|\Sigma_t'|}\right)\right) - (C_{0,2} - C_{0,1})\sqrt{0.5\ln(2n/\delta)}$.

Using Lemma 3.2 and the concentration inequality, Lemmas C.1 and C.2 show that we can control the type I and type II errors to be below $\delta/2$. Details of the proof are provided in Appendix D.

Theorem C.1. *For*

$$\mathfrak{R}_{n,\delta,H_0} = E[2\mathfrak{L}|\mathbb{H}_0] + (C_{0,2} - C_{0,1})\sqrt{0.5\ln(2n/\delta)},$$
$$\mathfrak{R}_{n,\delta,H_1} = E[2\mathfrak{L}|\mathbb{H}_1] - (C_{0,2} - C_{0,1})\sqrt{0.5\ln(2n/\delta)},$$

if $\mathfrak{R}_{n,\delta,H_0} \leq \mathfrak{R}_{n,\delta,H_1}$ and $\mathfrak{R}_\delta = \mathfrak{R}_{n,\delta,H_1}$, the conditional detection error probability is bounded as follows:

$$\varphi_n(\mathfrak{T}) = \mathbb{P}(2\mathfrak{L} \geq \mathfrak{R}_\delta | \mathbb{H}_0) + \max_{t\in\mathcal{C}_n} \mathbb{P}(2\mathfrak{L} \leq \mathfrak{R}_\delta | \mathbb{H}_{1,t}) \leq \delta.$$

Proof. By Lemma C.2, the probablity of type I error is bounded as $\mathbb{P}(\mathfrak{T}_{GLRT}=1|\mathbb{H}_0) \leq \frac{\delta}{2}$ and the probablity of type II error is bounded as $\mathbb{P}(\mathfrak{T}_{GLRT}=0|\mathbb{H}_1) \leq \frac{\delta}{2}$. Summing the two, we obtain the desired result. □

From Theorem C.1, we guarantee that, the likelihood ratio test is statistically correct for any error bound δ under some conditions.

Figure 3 illustrates the thresholds that guarantee a type I error or type II error. If we set the threshold to be greater than or equal to the upper epsilon bound of the null distribution, $\mathfrak{R}_{n,\delta,\mathbb{H}_0}$, we can guarantee a bounded type I error. If we set the threshold to be less than or equal to the lower epsilon bound of the alternative distribution, $\mathfrak{R}_{n,\delta,\mathbb{H}_1}$, we can guarantee a bounded type II error.

The inequalities for $\mathfrak{R}_{n,\delta,\mathbb{H}_0}$ and $\mathfrak{R}_{n,\delta,\mathbb{H}_1}$ can take three possible forms. If $\mathfrak{R}_{n,\delta,\mathbb{H}_0} > \mathfrak{R}_{n,\delta,\mathbb{H}_1}$, there is no threshold guaranteeing both type I and type II errors. If $\mathfrak{R}_{n,\delta,\mathbb{H}_0} = \mathfrak{R}_{n,\delta,\mathbb{H}_1}$, there is only one threshold that can guarantee both type I and type II errors. If $\mathfrak{R}_{n,\delta,\mathbb{H}_0} < \mathfrak{R}_{n,\delta,\mathbb{H}_1}$, the thresholds that can guarantee both type I and type II errors are indicated by the shaded area in Figure 3.

D Proofs for Section 3

Let $\Sigma, \Sigma' \in \mathbb{R}^{n\times n}$ be covariance matrices which are defined element wise as $\Sigma_{i,j} = K(X_i, X_j)$ and $\Sigma'_{i,j} = K'(X_i, X_j)$. Let Σ'_t be as

$$\Sigma'_{t,(i,j)} = \begin{cases} \Sigma_{i,j}, & i,j<t \\ \Sigma'_{i,j}, & i,j\geq t \\ 0, & \text{otherwise} \end{cases}.$$

Definition 2 (Subgaussianity)**.** *A random variable X is σ-subgaussian if for all $\lambda\in\mathbb{R}$ it holds that* $E[\exp(\lambda X)] \leq \exp(\lambda^2\sigma^2/2)$

Lemma D.1. *If X is a bounded and centered random variable, with $X \in [a,b]$, then X is $\frac{b-a}{2}$-subgaussian.*

Theorem D.1 (Chernoff Bound)**.** *If X is σ-subgaussian, then for any $\epsilon \geq 0$,*

$$P(X \geq \epsilon) \leq \exp(-\frac{\epsilon^2}{2\sigma^2}), \quad \textit{and} \quad P(X \leq -\epsilon) \leq \exp(-\frac{\epsilon^2}{2\sigma^2})$$

Theorem D.2 (Cauchy Interlace Theorem)**.** *Let A be a Hermitian matrix of order n, and let B be a principal submatrix of A of order $n-1$, and denote the kth largest eigenvalues of a general $n\times n$ matrix by $\lambda_k(\Sigma)$ so that*

$$\lambda_1(\Sigma) \geq \lambda_2(\Sigma) \geq \cdots \geq \lambda_n(\Sigma), \ \forall k \in [1,n].$$

If $\lambda_n(A) \leq \lambda_{n-1}(A) \leq \cdots \leq \lambda_2(A) \leq \lambda_1(A)$ and $\lambda_{n-1}(B) \leq \lambda_{n-2}(B) \leq \cdots \leq \lambda_2(B) \leq \lambda_1(B)$, then $\lambda_n(A) \leq \lambda_{n-1}(B) \leq \lambda_{n-1}(A) \leq \lambda_{n-2}(B) \leq \cdots \leq \lambda_2(A) \leq \lambda_1(B) \leq \lambda_1(A)$.

Corollary D.1. *Let $A, B, C_t \in \mathbb{R}^{n\times n}$ be positive semi-definite matrix defined element wise as $A_{i,j} = K(X_i, X_j)$, $B_{i,j} = K'(X_i, X_j)$ and C_t is defined as*

$$C_{t,(i,j)} = \begin{cases} K(X_i,X_j), & i,j<t \\ K'(X_i,X_j), & i,j\geq t \\ 0, & \textit{otherwise} \end{cases}$$

for all $t \in [1,n]$. Then, $\forall 1 \leq k \leq n, (\lambda_n(A) \wedge \lambda_n(B)) \leq \lambda_k(C_t) \leq (\lambda_1(A) \vee \lambda_1(B))$, where $\wedge$ and $\vee$ stand for minimum and maximum operators.

Proof of Theorem 3.1. The likelihood ratio can be written as

$$2\mathfrak{L} = \max_{t\in\mathcal{C}_{n,\alpha}} \left[X^T(\Sigma)^{-1}X - X^T(\Sigma_t')^{-1}X + \ln\left(\frac{|\Sigma|}{|\Sigma_t'|}\right)\right]$$
$$= X^T(\Sigma^{-1} - \Sigma_t'^{-1})X + \ln\left(\frac{|\Sigma|}{|\Sigma_t'|}\right).$$

Then, $2\mathfrak{L} - ln\left(\frac{|\Sigma|}{|\Sigma_t'|}\right) = X^T((\Sigma)^{-1} - (\Sigma_t')^{-1})X$. As $(\Sigma)^{-1} - (\Sigma_t')^{-1}$ is a symmetric matrix, Lemma C.2.2 in [Ruben, 1961; Isupova, 2017] gives the generalized chi-squared distribution.

□

Proof of Lemma 3.2. We will first show the bound of quadratic terms, $X^T(\Sigma^{-1} - (\Sigma_t')^{-1})X$. For a positive semi-definite matrix M, its eigenvectors are orthonormal to each other and all of its eigenvalues are nonnegative. By performing eigenvalue decomposition

$$X^TMX = X^TQ\Lambda Q^TX = \sum_k \lambda_k < X, q_k >^2,$$

where q_k is the kth column of Q. Since eigenvectors are orthonormal to each other

$$\sum_k < X, q_k >^2 = X^TQQ^Tx =< X, X >= ||X||^2.$$

By the restriction that X_t is bounded with $X_t \in [-V, V]$ for all t,

$$\lambda_{\text{min}}(M)V^2n \le X^TMX \le \lambda_{\text{max}}(M)V^2n,$$

where $\lambda_{\text{min}}(M)$ and $\lambda_{\text{max}}(M)$ represent the smallest and biggest eigenvalue of M respectively.

As Σ is symmetric, its inverse $(\Sigma)^{-1}$ is also symmetric and positive semi-definite. Substituting M by Σ^{-1} gives

$$\frac{1}{\lambda_{\text{max}}(\Sigma)}V^2n \le X^T(\Sigma)^{-1}X \le \frac{1}{\lambda_{\text{min}}(\Sigma)}V^2n,$$

where we use the fact that eigenvalues of inverse matrix is inverse of eigenvalues.

Similarly for $(\Sigma_t')^{-1}$,

$$\frac{1}{\lambda_{\text{max}}(\Sigma_t')}V^2n \le X^T(\Sigma_t')^{-1}X \le \frac{1}{\lambda_{\text{min}}(\Sigma_t')}V^2n.$$

Then the difference of quadratic terms can be bounded

$$\left(\frac{1}{\lambda_{\text{max}}(\Sigma)} - \frac{1}{\lambda_{\text{min}}(\Sigma_t')}\right)V^2n \le X^T((\Sigma)^{-1} - (\Sigma_t')^{-1})X \le \left(\frac{1}{\lambda_{\text{min}}(\Sigma)} - \frac{1}{\lambda_{\text{max}}(\Sigma_t')}\right)V^2n.$$

Therefore,

$$C_{t,1}V^2n \le X^T((\Sigma)^{-1} - (\Sigma_t')^{-1})X \le C_{t,2}V^2n,$$

with $C_{t,1} = \left(\frac{1}{\lambda_{\text{max}}(\Sigma)} - \frac{1}{\lambda_{\text{min}}(\Sigma_t')}\right)$, $C_{t,2} = \left(\frac{1}{\lambda_{\text{min}}(\Sigma)} - \frac{1}{\lambda_{\text{max}}(\Sigma_t')}\right)$. Thus, we can conclude proof by Definition 2 and Lemma D.1. □

Proof of Lemma C.1. For brevity, let $Z_t = X^T(\Sigma^{-1} - \Sigma_t'^{-1})X + \ln\left(\frac{|\Sigma|}{|\Sigma_t'|}\right)$, $t = 1, \ldots, n$. Lemma 3.2 implies that $Z_t - E[Z_t]$ is $\frac{C_{t,2} - C_{t,1}}{2}$-subgaussian. Under the null hypothesis the expectation of Z_t is defined as

$$E(Z_t|\mathbb{H}_0) = E\left[X^T(\Sigma)^{-1}X - X^T(\Sigma_t^{'})^{-1}X + \ln\left(\frac{|\Sigma|}{|\Sigma_t'|}\right)\middle|\mathbb{H}_0\right]$$
$$= E\left[Tr(XX^T(\Sigma)^{-1}) - Tr(XX^T(\Sigma_t^{'})^{-1}) + \ln\left(\frac{|\Sigma|}{|\Sigma_t'|}\right)\middle|\mathbb{H}_0\right]$$
$$= Tr(\Sigma(\Sigma)^{-1}) - Tr(\Sigma(\Sigma_t^{'})^{-1}) + \ln\left(\frac{|\Sigma|}{|\Sigma_t'|}\right)$$
$$= n - Tr(\Sigma(\Sigma_t^{'})^{-1}) + \ln\left(\frac{|\Sigma|}{|\Sigma_t'|}\right).$$

Theorem D.1 implies that $P(Z_t \geq n - Tr(\Sigma(\Sigma_t^{'})^{-1}) + \ln\left(\frac{|\Sigma|}{|\Sigma_t^{'}|}\right) + (C_{t,2} - C_{t,1})\sqrt{0.5\ln(2n/\delta)}) \leq \frac{\delta}{2n}$. By Corollary D.1, there exist constants $C_{0,1}$ and $C_{0,2}$ such that $C_{t,2} - C_{t,1} \leq C_{0,2} - C_{0,1}$ for all $t \in [1, n]$. Thus,

$$P\left[Z_t \geq \max_t \left(n - Tr(\Sigma(\Sigma_t^{'})^{-1}) + \ln\left(\frac{|\Sigma|}{|\Sigma_t^{'}|}\right)\right) + (C_{0,2} - C_{0,1})\sqrt{0.5\ln(2n/\delta)}\right] \leq \frac{\delta}{2n}.$$

We can conclude the proof by a union bound argument as

$$P\left[\max_t Z_t \geq \max_t \left(n - Tr(\Sigma(\Sigma_t^{'})^{-1}) + \ln\left(\frac{|\Sigma|}{|\Sigma_t^{'}|}\right)\right) + (C_{0,2} - C_{0,1})\sqrt{0.5\ln(2n/\delta)}\right] \leq \frac{\delta}{2}.$$

□

Proof of Lemma C.2. For brevity, let $Z_t = X^T(\Sigma^{-1} - \Sigma_t^{'-1})X + \ln\left(\frac{|\Sigma|}{|\Sigma_t^{'}|}\right)$, $t = 1, \ldots, n$. Lemma 3.2 implies that $Z_t - E[Z_t]$ is $\frac{C_{t,2} - C_{t,1}}{2}$-subgaussian. Under the alternative hypothesis the expectation of Z_t is defined as

$$\begin{aligned}
E(Z_t|\mathbb{H}_{1,t}) &= E\left[X^T(\Sigma)^{-1}X - X^T(\Sigma_t^{'})^{-1}X + \ln\left(\frac{|\Sigma|}{|\Sigma_t^{'}|}\right)\middle|\mathbb{H}_{1,t}\right] \\
&= E\left[Tr(XX^T(\Sigma)^{-1}) - Tr(XX^T(\Sigma_t^{'})^{-1}) + \ln\left(\frac{|\Sigma|}{|\Sigma_t^{'}|}\right)\middle|\mathbb{H}_{1,t}\right] \\
&= Tr(\Sigma_t^{\prime}(\Sigma)^{-1}) - Tr(\Sigma_t^{\prime}(\Sigma_t^{'})^{-1}) + \ln\left(\frac{|\Sigma|}{|\Sigma_t^{'}|}\right) \\
&= Tr(\Sigma_t^{\prime}(\Sigma)^{-1}) - n + \ln\left(\frac{|\Sigma|}{|\Sigma_t^{'}|}\right).
\end{aligned}$$

Theorem D.1 implies that $P(Z_t \leq Tr(\Sigma^{\prime}(\Sigma)^{-1}) - n + \ln\left(\frac{|\Sigma|}{|\Sigma_t^{'}|}\right) - (C_{t,2} - C_{t,1})\sqrt{0.5\ln(2n/\delta)}) \leq \frac{\delta}{2n}$. By Corollary D.1, there exist constants $C_{0,1}$ and $C_{0,2}$ such that $C_{t,2} - C_{t,1} \leq C_{0,2} - C_{0,1}$ for all $t \in [1, n]$. Thus,

$$P\left[Z_t \leq \min_t \left(Tr(\Sigma^{\prime}(\Sigma)^{-1}) - n + \ln\left(\frac{|\Sigma|}{|\Sigma_t^{'}|}\right)\right) - (C_{0,2} - C_{0,1})\sqrt{0.5\ln(2n/\delta)}\right] \leq \frac{\delta}{2n}.$$

We can conclude the proof by a union bound argument as

$$P\left[\max_t Z_t \leq \min_t \left(Tr(\Sigma^{\prime}(\Sigma)^{-1}) - n + \ln\left(\frac{|\Sigma|}{|\Sigma_t^{'}|}\right)\right) - (C_{0,2} - C_{0,1})\sqrt{0.5\ln(2n/\delta)}\right] \leq \frac{\delta}{2}.$$

□

Proof of Theorem 4.2. Let's define the gain of CBOCPD over BOCPD as

$$|E[x_t|\emptyset] - E_B[x_t|x_{1:t-1}]| - |E[x_t|\emptyset] - E_C[x_t|x_{1:t-1}]|.$$

In case $\mathfrak{T}^*_{GLRT} = 1$, the gain is written as

$$\begin{aligned}
&|E[x_t|\emptyset] - E_B[x_t|x_{1:t-1}]| - |E[x_t|\emptyset] - E[x_t|\emptyset]| \\
&= |E[x_t|\emptyset] - E_B[x_t|x_{1:t-1}]| \geq \epsilon_L \cdot \gamma
\end{aligned}$$

where $\gamma = 1 - P(r_t = 0|x_{1:t-1})$. In case $\mathfrak{T}^*_{GLRT} = 0$, the loss of CBOCPD is written as

$$\begin{aligned}
&|E[x_t|\emptyset] - \sum_{r_t=1}^{t-1} E[x_t|x^{(r_t)}] \cdot P(r_t|x_{1:t-1})| \\
&\leq \epsilon_U.
\end{aligned}$$

Then the gain is bounded from below as

$$\begin{aligned}
&|E[x_t|\emptyset] - E_B[x_t|x_{1:t-1}]| - |E[x_t|\emptyset] - E_C[x_t|x_{1:t-1}]| \\
&\geq \epsilon_L \cdot \gamma - \epsilon_U.
\end{aligned}$$

As $P(\mathfrak{T}^*_{GLRT} = 1) = (1 - \delta_1)(1 - \delta_2)$ and $P(\mathfrak{T}^*_{GLRT} = 0) = \delta_1\delta_2$ in non-stationary case, the expected gain is bounded from below as

$$\begin{aligned}
&E(|E[x_t|\emptyset] - E_B[x_t|x_{1:t-1}]| - |E[x_t|\emptyset] - E_C[x_t|x_{1:t-1}]|) \\
&\geq \epsilon_L\gamma(1 - \delta_1)(1 - \delta_2) + (\epsilon_L\gamma - \epsilon_U)\delta_1\delta_2 \geq 0
\end{aligned}$$

where the last inequality follows the assumption. Thus we can conclude that the expected gain is positive.

□