A Bayesian Partially Observable Online Change Detection Approach with Thompson Sampling

Jie Guo[1], Hao Yan[2], Chen Zhang[1*]

[1]Tsinghua University

[2]Arizona State University

A: Derivation of Variational Bayesian Inference

In this section, we want to derive the estimation of $\{\mu_{aj}, s_j^2, \alpha_j\}$ using VB inference. The objective of VB inference is to solve the best approximated posterior distribution $\tilde{p}_j(\theta_{aj}, r_j) = \tilde{p}_j(\theta_{aj}|r_j)\tilde{p}_j(r_j)$, indicating to find $\{\mu_{aj}, s_j^2, \alpha_j\}$ that minimize the Kullback-Leibler divergence:

$$KL(\tilde{p}(\boldsymbol{\theta}_a, \mathbf{r})||p(\boldsymbol{\theta}_a, \mathbf{r}|\mathbf{X}_{Z(1)}, \ldots, \mathbf{X}_{Z(n)})) = \int \tilde{p}(\boldsymbol{\theta}_a, \mathbf{r}) \ln \frac{\tilde{p}(\boldsymbol{\theta}_a, \mathbf{r})}{p(\boldsymbol{\theta}_a, \mathbf{r}|\mathbf{X}_{Z(1)}, \ldots, \mathbf{X}_{Z(n)})} d\tilde{p}(\boldsymbol{\theta}_a, \mathbf{r}).$$

This minimization can not be solved directly. To simplify it, we derive the evidence lower bound $J(\tilde{p})$ of $\ln p(\mathbf{X}_{Z(1)}, \ldots, \mathbf{X}_{Z(n)})$ from the Kullback-Leibler divergence:

$$\begin{aligned}
J(\tilde{p}) &= \ln p(\mathbf{X}_{Z(1)}, \ldots, \mathbf{X}_{Z(n)}) - KL(\tilde{p}(\boldsymbol{\theta}_a, \mathbf{r})||p(\boldsymbol{\theta}_a, \mathbf{r}|\mathbf{X}_{Z(1)}, \ldots, \mathbf{X}_{Z(n)})) \\
&= \mathbb{E}_{\tilde{p}(\boldsymbol{\theta}_a, \mathbf{r})}[\ln p\left(\boldsymbol{\theta}_a, \mathbf{r}, \mathbf{X}_{Z(1)}, \ldots, \mathbf{X}_{Z(n)}\right)] - \mathbb{E}_{\tilde{p}(\boldsymbol{\theta}_a, \mathbf{r})}[\ln \tilde{p}(\boldsymbol{\theta}_a, \mathbf{r})].
\end{aligned}$$

To minimize the Kullback-Leibler divergence can be converted to maximizing $J(\tilde{p})$. Given the estimation value of $\tilde{\boldsymbol{\theta}}_t$, we slightly abuse the notation by writing $\tilde{\mathbf{X}}_{1Z(t)} = \mathbf{X}_{Z(t)} - \mathbf{B}_{bZ(t)}\tilde{\boldsymbol{\theta}}_t$.

*Corresponding author

Then the logarithm of the joint posterior distribution $\ln p(\boldsymbol{\theta}_a, \mathbf{r}, \tilde{\mathbf{X}}_{1Z(1)}, \tilde{\mathbf{X}}_{1Z(2)} ... \tilde{\mathbf{X}}_{1Z(n)})$ can be expressed as

$$\ln p(\boldsymbol{\theta}_a, \mathbf{r}, \tilde{\mathbf{X}}_{1Z(1)}, \tilde{\mathbf{X}}_{1Z(2)} ... \tilde{\mathbf{X}}_{1Z(n)}) = \lambda_t^{(n)} \sum_{t=1}^{n} \ln p(\tilde{\mathbf{X}}_{1Z(t)} | \boldsymbol{\theta}_a, \mathbf{r}) + \sum_{j=1}^{k_a} \ln p(\theta_{aj} | r_j) + \sum_{j=1}^{k_a} \ln p(r_j).$$

Further, the first part of $\ln p(\boldsymbol{\theta}_a, \mathbf{r}, \tilde{\mathbf{X}}_{1Z(1)}, \tilde{\mathbf{X}}_{1Z(2)} ... \tilde{\mathbf{X}}_{1Z(n)})$ can be derived as

$$\begin{aligned}\lambda_t^{(n)} \sum_{t=1}^{n} \ln p(\tilde{\mathbf{X}}_{1Z(t)} | \boldsymbol{\theta}_a, \mathbf{r}) = \sum_{t=1}^{n} \lambda_t^{(n)} \Big(- \frac{p \ln(2\pi)}{2} - \frac{1}{2} \ln \sigma_e^2 - \frac{1}{2\sigma_e^2} (\tilde{\mathbf{X}}'_{Z(t)} \tilde{\mathbf{X}}_{Z(t)} - 2 \tilde{\mathbf{X}}'_{Z(t)} \sum_{j=1}^{k_a} \mathbf{B}_{ajZ(t)} \theta_{aj} \\ + (\sum_{j=1}^{k_a} \mathbf{B}_{ajZ(t)} \theta_{aj})' (\sum_{j=1}^{k_a} \mathbf{B}_{ajZ(t)} \theta_{aj})) \Big),\end{aligned}$$

So take the first part's expectation under the distribution of $\tilde{p}(\boldsymbol{\theta}_a, \mathbf{r})$:

$$\begin{aligned}&\mathbb{E}_{\tilde{p}(\boldsymbol{\theta}_a, \mathbf{r})} [\lambda_t^{(n)} \sum_{t=1}^{n} \ln p(\tilde{\mathbf{X}}_{1Z(t)} | \boldsymbol{\theta}_a, \mathbf{r})] \propto - \sum_{t=1}^{n} \frac{\lambda_t^{(n)}}{2\sigma_e^2} \Big(\tilde{\mathbf{X}}'_{1Z(t)} \tilde{\mathbf{X}}_{1Z(t)} - 2 \tilde{\mathbf{X}}'_{1Z(t)} \sum_{j=1}^{k_a} \mathbf{B}_{ajZ(t)} \alpha_j \mu_{aj} \\ &+ 2 \sum_{j=1}^{k_a} \sum_{k=1, k \neq j}^{k_a} \mathbf{B}'_{ajZ(t)} \mathbf{B}_{akZ(t)} \alpha_j \alpha_k \mu_{aj} \mu_{ak} + \sum_{j=1}^{k_a} \mathbf{B}'_{ajZ(t)} \mathbf{B}_{ajZ(t)} \big((\mu_{aj}^2 + s_j^2) \alpha_j \big) \Big).\end{aligned}$$

Also, the second part of $\ln p(\boldsymbol{\theta}_a, \mathbf{r}, \tilde{\mathbf{X}}_{1Z(1)}, \tilde{\mathbf{X}}_{1Z(2)} ... \tilde{\mathbf{X}}_{1Z(n)})$ can be derived as

$$\begin{aligned}\sum_{j=1}^{k_a} \ln p(\theta_{aj} | r_j) &= \sum_{j=1}^{k_a} \ln \Big(\big(\frac{1}{\sqrt{2\pi}\sigma_j} \exp(-\frac{1}{2\sigma_j^2} \theta_{aj}^2) \big)^{r_j} I(\theta_{aj} = 0)^{1 - r_j} \Big) \\ &= \sum_{j=1}^{k_a} r_j (-\frac{\ln(2\pi)}{2} - \frac{1}{2} \ln \sigma_j^2 - \frac{\theta_{aj}^2}{2\sigma_j^2}) + (1 - r_j) \ln I(\theta_{aj} = 0).\end{aligned}$$

Take its expectation under the distribution of $\tilde{p}(\boldsymbol{\theta}_a, \mathbf{r})$:

$$\mathbb{E}_{\tilde{p}(\boldsymbol{\theta}_a, \mathbf{r})} [\sum_{j=1}^{k_a} \ln p(\theta_{aj} | r_j)] = \sum_{j=1}^{k_a} \alpha_j (-\frac{\ln(2\pi)}{2} - \frac{1}{2} \ln \sigma_j^2 - \frac{\mu_{aj}^2 + s_j^2}{2\sigma_j^2}).$$

And the third part can be derived as

$$\sum_{j=1}^{k_a} \ln p(r_j) = \sum_{j=1}^{k_a} \ln \left(w_j^{r_j} (1-w_j)^{1-r_j} \right) = \sum_{j=1}^{k_a} r_j \ln(w_j) + (1-r_j) \ln(1-w_j).$$

Take its expectation under the distribution of $\tilde{p}(\boldsymbol{\theta}_a, \mathbf{r})$:

$$\mathbb{E}_{\tilde{p}(\boldsymbol{\theta}_a,\mathbf{r})}[\sum_{j=1}^{k_a} \ln p(r_j)] = \sum_{j=1}^{k_a} \alpha_j \ln(w_j) + (1-\alpha_j) \ln(1-w_j).$$

To sum up, the expectation of $\ln p(\boldsymbol{\theta}_a, \mathbf{r}, \tilde{\mathbf{X}}_{1Z(1)}, \tilde{\mathbf{X}}_{1Z(2)}...\tilde{\mathbf{X}}_{1Z(n)})$ under $\tilde{p}(\boldsymbol{\theta}_a, \mathbf{r})$ is

$$\mathbb{E}_{\tilde{p}(\boldsymbol{\theta}_a,\mathbf{r})}[\ln p(\boldsymbol{\theta}_a, \mathbf{r}, \tilde{\mathbf{X}}_{1Z(1)}, \tilde{\mathbf{X}}_{1Z(2)}...\tilde{\mathbf{X}}_{1Z(n)})] = -\sum_{t=1}^{n} \frac{\lambda_t^{(n)}}{2\sigma_e^2} \Big(\tilde{\mathbf{X}}'_{1Z(t)} \tilde{\mathbf{X}}_{1Z(t)} - 2\tilde{\mathbf{X}}'_{1Z(t)} \sum_{j=1}^{k_a} \mathbf{B}_{ajZ(t)} \alpha_j \mu_{aj}$$
$$+ 2\sum_{j=1}^{k_a} \sum_{k=1,k\neq j}^{k_a} \mathbf{B}'_{ajZ(t)} \mathbf{B}_{akZ(t)} \alpha_j \alpha_k \mu_{aj} \mu_k + \sum_{j=1}^{k_a} \mathbf{B}'_{ajZ(t)} \mathbf{B}_{ajZ(t)} (\mu_{aj}^2 + s_j^2) \alpha_j \Big) + \sum_{j=1}^{k_a} \alpha_j (-\frac{\ln(2\pi)}{2} - \frac{1}{2} \ln \sigma_j^2$$
$$- \frac{\mu_{aj}^2 + s_j^2}{2\sigma_j^2} + \ln w_j) + (1-\alpha_j) \ln(1-w_j).$$

On the other hand, the approximated posterior distribution $\tilde{p}(\boldsymbol{\theta}_a, \mathbf{r})$ and its logarithm transformation can be expressed as

$$\tilde{p}(\boldsymbol{\theta}_a, \mathbf{r}) = \prod_{j=1}^{k_a} \tilde{p}(\theta_{aj}|r_j) \prod_{j=1}^{k_a} \tilde{p}(r_j), \quad \ln \tilde{p}(\boldsymbol{\theta}_a, \mathbf{r}) = \sum_{j=1}^{k_a} \ln \tilde{p}(\theta_{aj}|r_j) + \sum_{j=1}^{k_a} \ln \tilde{p}(r_j).$$

The first part of $\ln \tilde{p}(\boldsymbol{\theta}_a, \mathbf{r})$ can be derived as

$$\begin{aligned}\sum_{j=1}^{k_a} \ln \tilde{p}(\theta_{aj}|r_j) &= \sum_{j=1}^{k_a} \ln \Big((\frac{1}{\sqrt{2\pi} s_j} \exp(-\frac{(\theta_{aj}-\mu_{aj})^2}{2s_j^2}))^{r_j} I(\theta_{aj}=0)^{1-r_j} \Big) \\ &= \sum_{j=1}^{k_a} r_j (-\frac{\ln(2\pi)}{2} - \frac{1}{2} \ln s_j^2 - \frac{(\theta_{aj}-\mu_{aj})^2}{2s_j^2}) + (1-r_j) \ln I(\theta_{aj}=0).\end{aligned}$$

Take its expectation under the distribution of $\tilde{p}(\boldsymbol{\theta}_a, \mathbf{r})$

$$\mathbb{E}_{\tilde{p}(\boldsymbol{\theta}_a,\mathbf{r})}[\sum_{j=1}^{k_a} \ln \tilde{p}(\theta_a|r_j)] = \sum_{j=1}^{k_a} \alpha_j(-\frac{\ln(2\pi)}{2} - \frac{1}{2}\ln s_j^2 - \frac{s_j^2}{2s_j^2}).$$

Also the second part of $\ln \tilde{p}(\boldsymbol{\theta}_a, \mathbf{r})$ can be derived as

$$\sum_{j=1}^{k_a} \ln \tilde{p}(r_j) = \sum_{j=1}^{k_a} \ln \left(\alpha_j^{r_j}(1-\alpha_j)^{1-r_j}\right).$$

Take its expectation under the distribution of $\tilde{p}(\boldsymbol{\theta}_a, \mathbf{r})$

$$\mathbb{E}_{\tilde{p}(\boldsymbol{\theta}_a,\mathbf{r})}[\sum_{j=1}^{k_a} \ln \tilde{p}(r_j)] = \sum_{j=1}^{k_a} \alpha_j \ln \alpha_j + (1-\alpha_j)\ln(1-\alpha_j).$$

To sum up, the expectation of $\ln \tilde{p}(\boldsymbol{\theta}_a, \mathbf{r})$ under the distribution of $\tilde{p}(\boldsymbol{\theta}_a, \mathbf{r})$ is

$$\mathbb{E}_{\tilde{p}(\boldsymbol{\theta}_a,\mathbf{r})}[\ln \tilde{p}(\boldsymbol{\theta}_a, \mathbf{r})] = \sum_{j=1}^{k_a} \alpha_j(-\frac{\ln(2\pi)}{2} - \frac{1}{2} - \frac{1}{2}\ln s_j^2 + \ln \alpha_j) + (1-\alpha_j)\ln(1-\alpha_j).$$

To give a summary, the evidence lower bound is

$$\begin{aligned}
J(\tilde{p}) &= \mathbb{E}_{\tilde{p}(\boldsymbol{\theta}_a,\mathbf{r})}[\ln p(\boldsymbol{\theta}_a, \mathbf{r}, \tilde{\mathbf{X}}_{1Z(1)}, \tilde{\mathbf{X}}_{1Z(2)}, ...\tilde{\mathbf{X}}_{1Z(n)})] - \mathbb{E}_{\tilde{p}(\boldsymbol{\theta}_a,\mathbf{r})}[\ln \tilde{p}(\boldsymbol{\theta}_a, \mathbf{r})] \\
&= -\sum_{t=1}^{n} \frac{\lambda_t^{(n)}}{2\sigma_e^2}\Big(\tilde{\mathbf{X}}_{1Z(t)}'\tilde{\mathbf{X}}_{1Z(t)} - 2\tilde{\mathbf{X}}_{1Z(t)}'\sum_{j=1}^{k_a}\mathbf{B}_{ajZ(t)}\alpha_j\mu_{aj} + 2\sum_{j=1}^{k_a}\sum_{k=1,k\neq j}^{k_a}\mathbf{B}_{ajZ(t)}'\mathbf{B}_{akZ(t)}\alpha_j\alpha_k\mu_{aj}\mu_{ak} \\
&+ \sum_{j=1}^{k_a}\mathbf{B}_{ajZ(t)}'\mathbf{B}_{ajZ(t)}(\mu_{aj}^2 + s_j^2)\alpha_j\Big) - \sum_{j=1}^{k_a}\Big((\frac{\mu_{aj}^2 + s_j^2}{2\sigma_j^2} - \frac{1}{2} + \frac{1}{2}\ln\frac{\sigma_j^2}{s_j^2} - \ln w_j + \ln \alpha_j)\alpha_j \\
&+ (1-\alpha_j)\big(\ln(1-\alpha_j) - \ln(1-w_j)\big)\Big).
\end{aligned}$$

Taking the partial derivatives of $J(\tilde{p})$, we obtain the coordinate descent updates for this optimization problem.

$$\frac{\partial J(\tilde{p})}{\partial s_j^2} = \sum_{t=1}^{n} \frac{\lambda_t^{(n)}}{2\sigma_e^2} \mathbf{B}'_{ajZ(t)} \mathbf{B}_{ajZ(t)} \alpha_j + \frac{\alpha_j}{2\sigma_j^2} - \frac{\alpha_j}{2s_j^2} = 0,$$

$$s_j^2 = \frac{1}{\sum_{t=1}^{n} \frac{\lambda_t^{(n)}}{\sigma_e^2} \mathbf{B}'_{ajZ(t)} \mathbf{B}_{ajZ(t)} + \frac{1}{\sigma_j^2}}.$$

$$\frac{\partial J(\tilde{p})}{\partial \mu_{aj}} = \sum_{t=1}^{n} \frac{\lambda_t^{(n)}}{2\sigma_e^2} (-2\tilde{\mathbf{X}}'_{1Z(t)} \mathbf{B}_{ajZ(t)} \alpha_j + 2 \sum_{k=1, k\neq j}^{k_a} \mathbf{B}'_{ajZ(t)} \mathbf{B}_{akZ(t)} \alpha_j \alpha_k \mu_{ak} + 2\mathbf{B}'_{ajZ(t)} \mathbf{B}_{ajZ(t)} \alpha_j \mu_{aj}) + \frac{\alpha_j \mu_{aj}}{\sigma_j^2}$$

$$= 0,$$

$$\mu_{aj} = s_j^2 \sum_{t=1}^{n} \frac{\lambda_t^{(n)}}{\sigma_e^2} (\tilde{\mathbf{X}}'_{1Z(t)} \mathbf{B}_{ajZ(t)} - \sum_{k=1, k\neq j}^{k_a} \mathbf{B}'_{ajZ(t)} \mathbf{B}_{akZ(t)} \alpha_k \mu_{ak}).$$

$$\frac{\partial J(\tilde{p})}{\partial \alpha_j} = \sum_{t=1}^{n} \frac{\lambda_t^{(n)}}{2\sigma_e^2} \Big(-2\tilde{\mathbf{X}}'_{1Z(t)} \mathbf{B}_{ajZ(t)} \mu_{aj} + 2 \sum_{k=1, k\neq j}^{k_a} \mathbf{B}'_{ajZ(t)} \mathbf{B}_{akZ(t)} \alpha_k \mu_{aj} \mu_{ak} + \mathbf{B}'_{ajZ(t)} \mathbf{B}_{ajZ(t)} (\mu_{aj}^2 + s_j^2) \Big)$$

$$-\frac{1}{2} + \frac{\mu_{aj}^2 + s_j^2}{2\sigma_j^2} + \frac{1}{2} \ln \frac{\sigma_j^2}{s_j^2} - \ln w_j + \ln \alpha_j + \ln(1 - w_j) - \ln(1 - \alpha_j) = 0,$$

$$\ln \frac{\alpha_j}{1-\alpha_j} = \ln \frac{w_j}{1-w_j} + \frac{1}{2} - \frac{\mu_{aj}^2 + s_j^2}{2\sigma_j^2} - \frac{1}{2} \ln \frac{\sigma_j^2}{s_j^2} + \frac{\mu_{aj}^2}{s_j^2} - \sum_{t=1}^{n} \frac{\lambda_t^{(n)}}{2\sigma_e^2} \mathbf{B}'_{ajZ(t)} \mathbf{B}_{ajZ(t)} (\mu_{aj}^2 + s_j^2).$$

B: Deviation of Detection Statistic

Some notations are defined as $\tilde{\boldsymbol{\theta}}_n^{[0]}$ equals (12) with $\tilde{\boldsymbol{\mu}}_a = \mathbf{0}$ under H_0, while $\tilde{\boldsymbol{\theta}}_n^{[1]}$ equals (12) with $\tilde{\boldsymbol{\mu}}_a = \tilde{\boldsymbol{\mu}}_a$ under H_1. $\mathbf{A} = \mathbf{B}'_{aZ(n)} \boldsymbol{\Sigma}_{eZ(n)}^{-1} \mathbf{B}_{aZ(n)} + \mathbf{K}_{\mathbf{r}}^{-1}$, $\mathbf{R} = \mathbf{X}'_{Z(n)} \boldsymbol{\Sigma}_{eZ(n)}^{-1} \mathbf{B}_{aZ(n)} + \boldsymbol{\mu}'_{\mathbf{r}} \mathbf{K}_{\mathbf{r}}^{-1}$, $\mathbf{C} = \mathbf{B}'_{bZ(n)} \boldsymbol{\Sigma}_{eZ(n)}^{-1} \mathbf{B}_{aZ(n)}$, $\mathbf{H} = \mathbf{B}'_{bZ(n)} \boldsymbol{\Sigma}_{eZ(n)}^{-1} \mathbf{B}_{bZ(n)} + \tilde{\boldsymbol{\Sigma}}_b^{-1}$,$\mathbf{G}^{[0]} = \mathbf{X}'_{Z(n)} \boldsymbol{\Sigma}_{eZ(n)}^{-1} \mathbf{B}_{bZ(n)} + \tilde{\boldsymbol{\theta}}_n^{[0]\prime} \tilde{\boldsymbol{\Sigma}}_b^{-1}$,$\mathbf{G}^{[1]} = \mathbf{X}'_{Z(n)} \boldsymbol{\Sigma}_{eZ(n)}^{-1} \mathbf{B}_{bZ(n)} + \tilde{\boldsymbol{\theta}}_n^{[1]\prime} \tilde{\boldsymbol{\Sigma}}_b^{-1}$.

The marginal likelihood under H_0, which is the denominator of PBF, can be derived as

$$P(\mathbf{X}_{Z(n)}|H_0) = \int p(\mathbf{X}_{Z(n)}|\boldsymbol{\theta}_n)\tilde{p}(\boldsymbol{\theta}_n|H_0)d\boldsymbol{\theta}_n$$

$$= \sqrt{1/\big((2\pi)^{k_b+m}|\tilde{\boldsymbol{\Sigma}}_b||\boldsymbol{\Sigma}_{eZ(n)}|\big)} \int \exp\Big(-\frac{1}{2}\big((\boldsymbol{\theta}_n - \tilde{\boldsymbol{\theta}}_n^{[0]})'\tilde{\boldsymbol{\Sigma}}_b^{-1}(\boldsymbol{\theta}_n - \tilde{\boldsymbol{\theta}}_n) + (\mathbf{X}_{Z(n)} - \mathbf{B}_{bZ(n)}\boldsymbol{\theta}_n)'\boldsymbol{\Sigma}_{eZ(n)}^{-1}$$

$$(\mathbf{X}_{Z(n)} - \mathbf{B}_{bZ(n)}\boldsymbol{\theta}_n)\big)\Big)d\boldsymbol{\theta}_n$$

$$= \sqrt{1/\big((2\pi)^{k_b+m}|\tilde{\boldsymbol{\Sigma}}_b||\boldsymbol{\Sigma}_{eZ(n)}|\big)} \int \exp\Big(-\frac{1}{2}\big(\boldsymbol{\theta}_n'(\tilde{\boldsymbol{\Sigma}}_b^{-1} + \mathbf{B}_{bZ(n)}'\boldsymbol{\Sigma}_{eZ(n)}^{-1}\mathbf{B}_{bZ(n)})\boldsymbol{\theta}_n - 2(\mathbf{X}_{Z(n)}'\boldsymbol{\Sigma}_{eZ(n)}^{-1}\mathbf{B}_{bZ(n)}$$

$$+ \tilde{\boldsymbol{\theta}}_n^{[0]\prime}\tilde{\boldsymbol{\Sigma}}_b^{-1})\boldsymbol{\theta}_n + \tilde{\boldsymbol{\theta}}_n^{[0]\prime}\tilde{\boldsymbol{\Sigma}}_b^{-1}\tilde{\boldsymbol{\theta}}_n^{[0]} + \mathbf{X}_{Z(n)}'\boldsymbol{\Sigma}_{eZ(n)}^{-1}\mathbf{X}_{Z(n)}\big)\Big)d\boldsymbol{\theta}_n$$

$$= \sqrt{1/\big((2\pi)^{k_b+m}|\tilde{\boldsymbol{\Sigma}}_b||\boldsymbol{\Sigma}_{eZ(n)}|\big)} \exp\big(-\frac{1}{2}(\tilde{\boldsymbol{\theta}}_n^{[0]\prime}\tilde{\boldsymbol{\Sigma}}_b^{-1}\tilde{\boldsymbol{\theta}}_n^{[0]} + \mathbf{X}_{Z(n)}'\boldsymbol{\Sigma}_{eZ(n)}^{-1}\mathbf{X}_{Z(n)})\big) \int \exp\big(-\frac{1}{2}((\boldsymbol{\theta}_n - \mathbf{H}^{-1}\mathbf{G}^{[0]\prime})'$$

$$\mathbf{H}(\boldsymbol{\theta}_n - \mathbf{H}^{-1}\mathbf{G}^{[0]\prime}) - \mathbf{G}^{[0]}\mathbf{H}^{-1}\mathbf{G}^{[0]\prime})\big)d\boldsymbol{\theta}_n$$

$$= \sqrt{1/\big((2\pi)^{m}|\tilde{\boldsymbol{\Sigma}}_b||\boldsymbol{\Sigma}_{eZ(n)}||\mathbf{H}|\big)} \exp\Big(-\frac{1}{2}(\tilde{\boldsymbol{\theta}}_n^{[0]\prime}\tilde{\boldsymbol{\Sigma}}_b^{-1}\tilde{\boldsymbol{\theta}}_n^{[0]} + \mathbf{X}_{Z(n)}'\boldsymbol{\Sigma}_{eZ(n)}^{-1}\mathbf{X}_{Z(n)} - \mathbf{G}^{[0]}\mathbf{H}^{-1}\mathbf{G}^{[0]\prime})\Big).$$

The marginal likelihood under H_1, which is the numerator of PBF, can be derived as

$$P(\mathbf{X}_{Z(n)}|H_1) = \sum_{\mathbf{r}} \int\int p(\mathbf{X}_{Z(n)}|\boldsymbol{\theta}_a, \boldsymbol{\theta}_n)\tilde{p}(\boldsymbol{\theta}_n|H_1)\tilde{p}(\mathbf{r}|H_1)\tilde{p}(\boldsymbol{\theta}_a|\mathbf{r}, H_1)d\boldsymbol{\theta}_n d\boldsymbol{\theta}_a$$

$$= \sum_{\mathbf{r}} \tilde{p}(\mathbf{r}|H_1) \int\int p(\mathbf{X}_{Z(n)}|\boldsymbol{\theta}_a, \boldsymbol{\theta}_n)\tilde{p}(\boldsymbol{\theta}_n|H_1)\tilde{p}(\boldsymbol{\theta}_a|\mathbf{r}, H_1)d\boldsymbol{\theta}_n d\boldsymbol{\theta}_a$$

$$= \sum_{\mathbf{r}} \tilde{p}(\mathbf{r}|H_1)\sqrt{1/\big((2\pi)^{k_a+k_b+m}|\mathbf{K}_{\mathbf{r}}||\tilde{\boldsymbol{\Sigma}}_b||\boldsymbol{\Sigma}_{eZ(n)}|\big)} \int\int \exp\Big(-\frac{1}{2}\big((\boldsymbol{\theta}_a - \boldsymbol{\mu}_{\mathbf{r}})'\mathbf{K}_{\mathbf{r}}^{-1}(\boldsymbol{\theta}_a - \boldsymbol{\mu}_{\mathbf{r}}) + (\boldsymbol{\theta}_n - \tilde{\boldsymbol{\theta}}_n)^{[1]\prime}$$

$$\tilde{\boldsymbol{\Sigma}}_b^{-1}(\boldsymbol{\theta}_n - \tilde{\boldsymbol{\theta}}_n^{[1]}) + (\mathbf{X}_{Z(n)} - \mathbf{B}_{bZ(n)}\boldsymbol{\theta}_n - \mathbf{B}_{aZ(n)}\boldsymbol{\theta}_a)'\boldsymbol{\Sigma}_{eZ(n)}^{-1}(\mathbf{X}_{Z(n)} - \mathbf{B}_{bZ(n)}\boldsymbol{\theta}_n - \mathbf{B}_{aZ(n)}\boldsymbol{\theta}_a)\big)\Big)d\boldsymbol{\theta}_a d\boldsymbol{\theta}_n$$

$$= \sum_{\mathbf{r}} \tilde{p}(\mathbf{r}|H_1)\sqrt{1/\big((2\pi)^{k_a+k_b+m}|\mathbf{K}_{\mathbf{r}}||\tilde{\boldsymbol{\Sigma}}_b||\boldsymbol{\Sigma}_{eZ(n)}|\big)} \int\int \exp\Big(-\frac{1}{2}\big(\boldsymbol{\theta}_a'(\mathbf{B}_{aZ(n)}'\boldsymbol{\Sigma}_{eZ(n)}^{-1}\mathbf{B}_{aZ(n)} + \mathbf{K}_{\mathbf{r}}^{-1})\boldsymbol{\theta}_a$$

$$- 2((\mathbf{X}_{Z(n)} - \mathbf{B}_{bZ(n)}\boldsymbol{\theta}_n)'\boldsymbol{\Sigma}_{eZ(n)}^{-1}\mathbf{B}_{aZ(n)} + \boldsymbol{\mu}_{\mathbf{r}}'\mathbf{K}_{\mathbf{r}}^{-1})\boldsymbol{\theta}_a + \boldsymbol{\mu}_{\mathbf{r}}'\mathbf{K}_{\mathbf{r}}^{-1}\boldsymbol{\mu}_{\mathbf{r}} + \boldsymbol{\theta}_n'(\mathbf{B}_{bZ(n)}'\boldsymbol{\Sigma}_{eZ(n)}^{-1}\mathbf{B}_{bZ(n)} + \tilde{\boldsymbol{\Sigma}}_b^{-1})\boldsymbol{\theta}_n$$

$$- 2(\mathbf{X}_{Z(n)}'\boldsymbol{\Sigma}_{eZ(n)}^{-1}\mathbf{B}_{bZ(n)} + \tilde{\boldsymbol{\theta}}_n^{[1]\prime}\tilde{\boldsymbol{\Sigma}}_b^{-1})\boldsymbol{\theta}_n + \mathbf{X}_{Z(n)}'\boldsymbol{\Sigma}_{eZ(n)}^{-1}\mathbf{X}_{Z(n)} + \tilde{\boldsymbol{\theta}}_n^{[1]\prime}\tilde{\boldsymbol{\Sigma}}_b^{-1}\tilde{\boldsymbol{\theta}}_n^{[1]})\Big)d\boldsymbol{\theta}_a d\boldsymbol{\theta}_n$$

$$= \sum_{\mathbf{r}} \tilde{p}(\mathbf{r}|H_1)\sqrt{1/\big((2\pi)^{k_a+k_b+m}|\mathbf{K}_{\mathbf{r}}||\tilde{\boldsymbol{\Sigma}}_b||\boldsymbol{\Sigma}_{eZ(n)}|\big)} \int\int \exp\Big(-\frac{1}{2}\big(\boldsymbol{\theta}_a'\mathbf{A}\boldsymbol{\theta}_a - 2(\mathbf{R} - \boldsymbol{\theta}_n'\mathbf{C})\boldsymbol{\theta}_a + \boldsymbol{\theta}_n'\mathbf{H}\boldsymbol{\theta}_n$$

$$- 2\mathbf{G}^{[1]}\boldsymbol{\theta}_n + \boldsymbol{\mu}_{\mathbf{r}}'\mathbf{K}_{\mathbf{r}}^{-1}\boldsymbol{\mu}_{\mathbf{r}} + \mathbf{X}_{Z(n)}'\boldsymbol{\Sigma}_{eZ(n)}^{-1}\mathbf{X}_{Z(n)} + \tilde{\boldsymbol{\theta}}_n^{[1]\prime}\tilde{\boldsymbol{\Sigma}}_b^{-1}\tilde{\boldsymbol{\theta}}_n^{[1]})\Big)d\boldsymbol{\theta}_a d\boldsymbol{\theta}_n$$

$$= \sum_{\mathbf{r}} \tilde{p}(\mathbf{r}|H_1)\sqrt{1/\big((2\pi)^{k_a+k_b+m}|\mathbf{K}_{\mathbf{r}}||\tilde{\boldsymbol{\Sigma}}_b||\boldsymbol{\Sigma}_{eZ(n)}|\big)} \exp\Big(-\frac{1}{2}(\boldsymbol{\mu}_{\mathbf{r}}'\mathbf{K}_{\mathbf{r}}^{-1}\boldsymbol{\mu}_{\mathbf{r}} + \mathbf{X}_{Z(n)}'\boldsymbol{\Sigma}_{eZ(n)}^{-1}\mathbf{X}_{Z(n)} + \tilde{\boldsymbol{\theta}}_n^{[1]\prime}\tilde{\boldsymbol{\Sigma}}_b^{-1}\tilde{\boldsymbol{\theta}}_n^{[1]})\Big)$$

$$\sqrt{(2\pi)^{k_a}/|\mathbf{A}|} \int \exp\Big(-\frac{1}{2}(\boldsymbol{\theta}_n'(\mathbf{H} - \mathbf{C}\mathbf{A}^{-1}\mathbf{C}')\boldsymbol{\theta}_n - 2(\mathbf{G}^{[1]} - \mathbf{R}\mathbf{A}^{-1}\mathbf{C}')\boldsymbol{\theta}_n - \mathbf{R}\mathbf{A}^{-1}\mathbf{R}')\Big)d\boldsymbol{\theta}_n$$

$$= \sum_{\mathbf{r}} \tilde{p}(\mathbf{r}|H_1)\sqrt{1/\big((2\pi)^{k_b+m}|\mathbf{K}_{\mathbf{r}}||\tilde{\boldsymbol{\Sigma}}_b||\boldsymbol{\Sigma}_{eZ(n)}||\mathbf{A}|\big)} \exp\Big(-\frac{1}{2}(\boldsymbol{\mu}_{\mathbf{r}}'\mathbf{K}_{\mathbf{r}}^{-1}\boldsymbol{\mu}_{\mathbf{r}} + \mathbf{X}_{Z(n)}'\boldsymbol{\Sigma}_{eZ(n)}^{-1}\mathbf{X}_{Z(n)} + \tilde{\boldsymbol{\theta}}_n^{[1]\prime}\tilde{\boldsymbol{\Sigma}}_b^{-1}\tilde{\boldsymbol{\theta}}_n^{[1]})\Big)$$

$$\sqrt{(2\pi)^{k_b}/|\mathbf{H} - \mathbf{C}\mathbf{A}^{-1}\mathbf{C}'|} \exp\Big(-\frac{1}{2}(-\mathbf{R}\mathbf{A}^{-1}\mathbf{R}' - (\mathbf{G}^{[1]} - \mathbf{R}\mathbf{A}^{-1}\mathbf{C}')(\mathbf{H} - \mathbf{C}\mathbf{A}^{-1}\mathbf{C}')^{-1}(\mathbf{G}^{[1]} - \mathbf{R}\mathbf{A}^{-1}\mathbf{C}')')\Big)$$

$$= \sum_{\mathbf{r}} \tilde{p}(\mathbf{r}|H_1)\sqrt{1/\big((2\pi)^{m}|\mathbf{K}_{\mathbf{r}}||\tilde{\boldsymbol{\Sigma}}_b||\boldsymbol{\Sigma}_{eZ(n)}||\mathbf{A}||\mathbf{H} - \mathbf{C}\mathbf{A}^{-1}\mathbf{C}'|\big)} \exp\Big(-\frac{1}{2}(\boldsymbol{\mu}_{\mathbf{r}}'\mathbf{K}_{\mathbf{r}}^{-1}\boldsymbol{\mu}_{\mathbf{r}} + \mathbf{X}_{Z(n)}'\boldsymbol{\Sigma}_{eZ(n)}^{-1}\mathbf{X}_{Z(n)}$$

$$+ \tilde{\boldsymbol{\theta}}_n^{[1]\prime}\tilde{\boldsymbol{\Sigma}}_b^{-1}\tilde{\boldsymbol{\theta}}_n^{[1]} - \mathbf{R}\mathbf{A}^{-1}\mathbf{R}' - (\mathbf{G}^{[1]} - \mathbf{R}\mathbf{A}^{-1}\mathbf{C}')(\mathbf{H} - \mathbf{C}\mathbf{A}^{-1}\mathbf{C}')^{-1}(\mathbf{G}^{[1]} - \mathbf{R}\mathbf{A}^{-1}\mathbf{C}')')\Big).$$

So the posterior Bayes factor is derived as

$$
\begin{aligned}
PBF &= \frac{\sum_{\mathbf{r}} \int\int p(\mathbf{X}_{Z(n)}|\boldsymbol{\theta}_a, \boldsymbol{\theta}_n)\tilde{p}(\boldsymbol{\theta}_n|H_1)\tilde{p}(\mathbf{r}|H_1)\tilde{p}(\boldsymbol{\theta}_a|\mathbf{r}, H_1)d\boldsymbol{\theta}_n d\boldsymbol{\theta}_a}{\int p(\mathbf{X}_{Z(n)}|\boldsymbol{\theta}_n)\tilde{p}(\boldsymbol{\theta}_n|H_0)d\boldsymbol{\theta}_n} \\
&= \sum_{\mathbf{r}} p(\mathbf{r}|H_1)\sqrt{|\mathbf{H}|/(|\mathbf{K}_\mathbf{r}||\mathbf{A}||\mathbf{H}-\mathbf{C}\mathbf{A}^{-1}\mathbf{C}'|)}\exp\Big(-\frac{1}{2}\big(\boldsymbol{\mu}_\mathbf{r}'\mathbf{K}_\mathbf{r}^{-1}\boldsymbol{\mu}_\mathbf{r} + \mathbf{G}^{[0]}\mathbf{H}^{-1}\mathbf{G}^{[0]\prime} - \mathbf{R}\mathbf{A}^{-1}\mathbf{R}' \\
&- (\mathbf{G}^{[1]} - \mathbf{R}\mathbf{A}^{-1}\mathbf{C}')(\mathbf{H}-\mathbf{C}\mathbf{A}^{-1}\mathbf{C}')^{-1}(\mathbf{G}^{[1]} - \mathbf{R}\mathbf{A}^{-1}\mathbf{C}')' + \tilde{\boldsymbol{\theta}}_n^{[1]\prime}\tilde{\boldsymbol{\Sigma}}_b^{-1}\tilde{\boldsymbol{\theta}}_n^{[1]} - \tilde{\boldsymbol{\theta}}_n^{[0]\prime}\tilde{\boldsymbol{\Sigma}}_b^{-1}\tilde{\boldsymbol{\theta}}_n^{[0]}\big)\Big).
\end{aligned}
$$

Note that the form of PBF is too complex to compute easily. We consider further simplifying it by eliminating constants and small values as follows.

For any square matrix, e.g., $\mathbf{M}$, with spectral radius $\rho(\mathbf{M}) < 1$, according to the Maclaurin series of matrix form, $(\mathbf{I}+\mathbf{M})^{-1} = \sum_{k=0}^{\infty}(-1)^k\mathbf{M}^k$. Here since $\mathbf{K}_\mathbf{r}\frac{\mathbf{B}_{aZ(n)}'\mathbf{B}_{aZ(n)}}{\sigma_e^2}$ is a square matrix and the entries of $\mathbf{K}_\mathbf{r}$ are quite small, the spectral radius $\rho(\mathbf{K}_\mathbf{r}\frac{\mathbf{B}_{aZ(n)}'\mathbf{B}_{aZ(n)}}{\sigma_e^2}) \leq \|\mathbf{K}_\mathbf{r}\frac{\mathbf{B}_{aZ(n)}'\mathbf{B}_{aZ(n)}}{\sigma_e^2}\| < 1$ can be satisfied, where $||\cdot||$ is any norm form. Then we can apply Maclaurin series form as

$$
\begin{aligned}
\mathbf{A}^{-1} &= (\frac{\mathbf{B}_{aZ(n)}'\mathbf{B}_{aZ(n)}}{\sigma_e^2} + \mathbf{K}_\mathbf{r}^{-1})^{-1} = \big(\mathbf{K}_\mathbf{r}^{-1}(\mathbf{K}_\mathbf{r}\frac{\mathbf{B}_{aZ(n)}'\mathbf{B}_{aZ(n)}}{\sigma_e^2} + \mathbf{I})\big)^{-1} = (\mathbf{K}_\mathbf{r}\frac{\mathbf{B}_{aZ(n)}'\mathbf{B}_{aZ(n)}}{\sigma_e^2} + \mathbf{I})^{-1}\mathbf{K}_\mathbf{r} \\
&= \sum_{k=0}^{\infty}(-1)^k(\mathbf{K}_\mathbf{r}\frac{\mathbf{B}_{aZ(n)}'\mathbf{B}_{aZ(n)}}{\sigma_e^2})^k\mathbf{K}_\mathbf{r} = \mathbf{K}_\mathbf{r} - \mathbf{K}_\mathbf{r}\frac{\mathbf{B}_{aZ(n)}'\mathbf{B}_{aZ(n)}}{\sigma_e^2}\mathbf{K}_\mathbf{r} + o(\mathbf{K}_\mathbf{r}^2).
\end{aligned}
$$

Then we can simplify

$$
\begin{aligned}
\mathbf{R}\mathbf{A}^{-1}\mathbf{R}' &= (\frac{\mathbf{X}_{Z(n)}'\mathbf{B}_{aZ(n)}}{\sigma_e^2} + \boldsymbol{\mu}_\mathbf{r}'\mathbf{K}_\mathbf{r}^{-1})(\mathbf{K}_\mathbf{r} - \mathbf{K}_\mathbf{r}\frac{\mathbf{B}_{aZ(n)}'\mathbf{B}_{aZ(n)}}{\sigma_e^2}\mathbf{K}_\mathbf{r})(\frac{\mathbf{X}_{Z(n)}'\mathbf{B}_{aZ(n)}}{\sigma_e^2} + \boldsymbol{\mu}_\mathbf{r}'\mathbf{K}_\mathbf{r}^{-1})' \\
&= \frac{\mathbf{X}_{Z(n)}'\mathbf{B}_{aZ(n)}}{\sigma_e^2}\mathbf{K}_\mathbf{r}\frac{\mathbf{B}_{aZ(n)}'\mathbf{X}_{Z(n)}}{\sigma_e^2} + \boldsymbol{\mu}_\mathbf{r}'\mathbf{K}_\mathbf{r}^{-1}\boldsymbol{\mu}_\mathbf{r} + 2\frac{\mathbf{X}_{Z(n)}'\mathbf{B}_{aZ(n)}}{\sigma_e^2}\boldsymbol{\mu}_\mathbf{r} - \boldsymbol{\mu}_\mathbf{r}'\frac{\mathbf{B}_{aZ(n)}'\mathbf{B}_{aZ(n)}}{\sigma_e^2}\boldsymbol{\mu}_\mathbf{r} \\
&- 2\boldsymbol{\mu}_\mathbf{r}'\frac{\mathbf{B}_{aZ(n)}'\mathbf{B}_{aZ(n)}}{\sigma_e^2}\mathbf{K}_\mathbf{r}\frac{\mathbf{B}_{aZ(n)}'\mathbf{X}_{Z(n)}}{\sigma_e^2} \\
&= \boldsymbol{\mu}_\mathbf{r}'\mathbf{K}_\mathbf{r}^{-1}\boldsymbol{\mu}_\mathbf{r} + 2\frac{\mathbf{X}_{Z(n)}'\mathbf{B}_{aZ(n)}}{\sigma_e^2}\boldsymbol{\mu}_\mathbf{r} - \boldsymbol{\mu}_\mathbf{r}'\frac{\mathbf{B}_{aZ(n)}'\mathbf{B}_{aZ(n)}}{\sigma_e^2}\boldsymbol{\mu}_\mathbf{r} + O(\mathbf{K}_\mathbf{r}).
\end{aligned}
$$

Following the same way, consider $\mathbf{H}^{-1}\mathbf{C}\mathbf{A}^{-1}\mathbf{C}'$ is a square matrix and the entries of $\mathbf{A}^{-1}$ are quite small. The spectral radius $\rho(\mathbf{H}^{-1}\mathbf{C}\mathbf{A}^{-1}\mathbf{C}') \leq \|\mathbf{H}^{-1}\mathbf{C}\mathbf{A}^{-1}\mathbf{C}'\| < 1$ can be satisfied as well. Then,

$$\mathbf{C}\mathbf{A}^{-1}\mathbf{C}' = \frac{\mathbf{B}'_{bZ(n)}\mathbf{B}_{aZ(n)}}{\sigma_e^2}(\mathbf{K_r} - \mathbf{K_r}\frac{\mathbf{B}'_{aZ(n)}\mathbf{B}_{aZ(n)}}{\sigma_e^2}\mathbf{K_r})\frac{\mathbf{B}'_{aZ(n)}\mathbf{B}_{bZ(n)}}{\sigma_e^2} = O(\mathbf{K_r}).$$

$$(\mathbf{H} - \mathbf{C}\mathbf{A}^{-1}\mathbf{C}')^{-1} = \big(\mathbf{H}(\mathbf{I} - \mathbf{H}^{-1}\mathbf{C}\mathbf{A}^{-1}\mathbf{C}')\big)^{-1} = (\mathbf{I} - \mathbf{H}^{-1}\mathbf{C}\mathbf{A}^{-1}\mathbf{C}')^{-1}\mathbf{H}^{-1} = \sum_{k=0}^{\infty}(-1)^k(\mathbf{H}^{-1}\mathbf{C}\mathbf{A}^{-1}\mathbf{C}')^k\mathbf{H}^{-1}$$

$$= \mathbf{H}^{-1} + \mathrm{O}(\mathbf{C}\mathbf{A}^{-1}\mathbf{C}') = \mathbf{H}^{-1} + \mathbf{H}^{-1}\mathbf{C}\mathbf{A}^{-1}\mathbf{C}'\mathbf{H}^{-1} + O(\mathbf{K_r}).$$

Then the following terms can be simplified as

$$(\mathbf{G}^{[1]} - \mathbf{R}\mathbf{A}^{-1}\mathbf{C}')(\mathbf{H} - \mathbf{C}\mathbf{A}^{-1}\mathbf{C}')^{-1}(\mathbf{G}^{[1]} - \mathbf{R}\mathbf{A}^{-1}\mathbf{C}')' = (\mathbf{G}^{[1]} - \mathbf{R}\mathbf{A}^{-1}\mathbf{C}')\mathbf{H}^{-1}(\mathbf{G}^{[1]} - \mathbf{R}\mathbf{A}^{-1}\mathbf{C}')'$$

$$= \mathbf{G}^{[1]}\mathbf{H}^{-1}\mathbf{G}^{[1]\prime} + (\mathbf{R}\mathbf{A}^{-1}\mathbf{C}')\mathbf{H}^{-1}(\mathbf{R}\mathbf{A}^{-1}\mathbf{C}')' - 2\mathbf{R}\mathbf{A}^{-1}\mathbf{C}'\mathbf{H}^{-1}\mathbf{G}^{[1]\prime} + O(\mathbf{K_r}).$$

$$\mathbf{R}\mathbf{A}^{-1}\mathbf{C}' = (\frac{\mathbf{X}'_{Z(n)}\mathbf{B}_{aZ(n)}}{\sigma_e^2} + \boldsymbol{\mu}'_{\mathbf{r}}\mathbf{K}_{\mathbf{r}}^{-1})\mathbf{K_r}\frac{\mathbf{B}'_{aZ(n)}\mathbf{B}_{bZ(n)}}{\sigma_e^2} = \boldsymbol{\mu}'_{\mathbf{r}}\frac{\mathbf{B}'_{aZ(n)}\mathbf{B}_{bZ(n)}}{\sigma_e^2} + O(\mathbf{K_r}).$$

Similarly, consider $\tilde{\boldsymbol{\Sigma}}_b\frac{\mathbf{B}_{bZ(n)}{}'\mathbf{B}_{bZ(n)}}{\sigma_e^2}$ is a square matrix and $\tilde{\boldsymbol{\Sigma}}_b = (\frac{\mathbf{B}'_{bZ(n)}\mathbf{B}_{bZ(n)}}{\sigma_e^2} + \boldsymbol{\Sigma}_b)^{-1}$. Its spectral radius $\rho(\tilde{\boldsymbol{\Sigma}}_b\frac{\mathbf{B}_{bZ(n)}{}'\mathbf{B}_{bZ(n)}}{\sigma_e^2}) \leq \|\tilde{\boldsymbol{\Sigma}}_b\frac{\mathbf{B}_{bZ(n)}{}'\mathbf{B}_{bZ(n)}}{\sigma_e^2}\| < 1$ can be satisfied. Then,

$$\mathbf{H}^{-1} = (\frac{\mathbf{B}'_{bZ(n)}\mathbf{B}_{bZ(n)}}{\sigma_e^2} + \tilde{\boldsymbol{\Sigma}}_b^{-1})^{-1} = \tilde{\boldsymbol{\Sigma}}_b - \tilde{\boldsymbol{\Sigma}}_b\frac{\mathbf{B}_{bZ(n)}{}'\mathbf{B}_{bZ(n)}}{\sigma_e^2}\tilde{\boldsymbol{\Sigma}}_b + o(\tilde{\boldsymbol{\Sigma}}_b^2).$$

$$\mathbf{H}^{-1}\mathbf{G}^{[1]\prime} = (\tilde{\boldsymbol{\Sigma}}_b - \tilde{\boldsymbol{\Sigma}}_b\frac{\mathbf{B}_{bZ(n)}{}'\mathbf{B}_{bZ(n)}}{\sigma_e^2}\tilde{\boldsymbol{\Sigma}}_b)(\tilde{\boldsymbol{\Sigma}}_b^{-1}\tilde{\boldsymbol{\theta}}_n^{[1]} + \frac{\mathbf{B}_{bZ(n)}{}'\mathbf{X}_{Z(n)}}{\sigma_e^2})$$

$$= \tilde{\boldsymbol{\theta}}_n^{[1]} - \tilde{\boldsymbol{\Sigma}}_b\frac{\mathbf{B}'_{bZ(n)}\mathbf{B}_{bZ(n)}}{\sigma_e^2}\tilde{\boldsymbol{\theta}}_n^{[1]} - \tilde{\boldsymbol{\Sigma}}_b\frac{\mathbf{B}_{bZ(n)}{}'\mathbf{B}_{bZ(n)}}{\sigma_e^2}\tilde{\boldsymbol{\Sigma}}_b\frac{\mathbf{B}'_{bZ(n)}\mathbf{X}_{Z(n)}}{\sigma_e^2} + \tilde{\boldsymbol{\Sigma}}_b\frac{\mathbf{B}'_{bZ(n)}\mathbf{X}_{Z(n)}}{\sigma_e^2} + o(\tilde{\boldsymbol{\Sigma}}_b^2) + O(\mathbf{K_r}).$$

$$\mathbf{R}\mathbf{A}^{-1}\mathbf{C}'\mathbf{H}^{-1}\mathbf{G}^{[1]\prime} = \boldsymbol{\mu}_{\mathbf{r}}\frac{\mathbf{B}'_{aZ(n)}\mathbf{B}_{aZ(n)}}{\sigma_e^2}\tilde{\boldsymbol{\theta}}_n^{[1]} - \boldsymbol{\mu}'_{\mathbf{r}}\frac{\mathbf{B}'_{aZ(n)}\mathbf{B}_{bZ(n)}}{\sigma_e^2}\tilde{\boldsymbol{\Sigma}}_b\frac{\mathbf{B}'_{bZ(n)}\mathbf{B}_{bZ(n)}}{\sigma_e^2}\tilde{\boldsymbol{\theta}}_n^{[1]}$$

$$- \boldsymbol{\mu}'_{\mathbf{r}}\frac{\mathbf{B}'_{aZ(n)}\mathbf{B}_{bZ(n)}}{\sigma_e^2}\tilde{\boldsymbol{\Sigma}}_b\frac{\mathbf{B}'_{bZ(n)}\mathbf{B}_{bZ(n)}}{\sigma_e^2}\tilde{\boldsymbol{\Sigma}}_b\frac{\mathbf{B}'_{bZ(n)}\mathbf{X}_{Z(n)}}{\sigma_e^2} + \boldsymbol{\mu}'_{\mathbf{r}}\frac{\mathbf{B}'_{aZ(n)}\mathbf{B}_{bZ(n)}}{\sigma_e^2}\tilde{\boldsymbol{\Sigma}}_b\frac{\mathbf{B}'_{bZ(n)}\mathbf{X}_{Z(n)}}{\sigma_e^2} + o(\tilde{\boldsymbol{\Sigma}}_b^2) + O(\mathbf{K_r}).$$

$$\mathbf{R}\mathbf{A}^{-1}\mathbf{C}'\mathbf{H}^{-1}(\mathbf{R}\mathbf{A}^{-1}\mathbf{C}')' = \boldsymbol{\mu}'_{\mathbf{r}}\frac{\mathbf{B}'_{aZ(n)}\mathbf{B}_{bZ(n)}}{\sigma_e^2}\tilde{\boldsymbol{\Sigma}}_b\frac{\mathbf{B}'_{bZ(n)}\mathbf{B}_{aZ(n)}}{\sigma_e^2}\boldsymbol{\mu}_{\mathbf{r}} - \boldsymbol{\mu}'_{\mathbf{r}}\frac{\mathbf{B}'_{aZ(n)}\mathbf{B}_{bZ(n)}}{\sigma_e^2}\tilde{\boldsymbol{\Sigma}}_b\frac{\mathbf{B}'_{bZ(n)}\mathbf{B}_{bZ(n)}}{\sigma_e^2}$$

$$\tilde{\boldsymbol{\Sigma}}_b\frac{\mathbf{B}'_{bZ(n)}\mathbf{B}_{aZ(n)}}{\sigma_e^2}\boldsymbol{\mu}_{\mathbf{r}} + o(\tilde{\boldsymbol{\Sigma}}_b^2) + O(\mathbf{K_r}).$$

So the term inside the exponential function in PBF is simplified as

$$
\begin{aligned}
&\boldsymbol{\mu}_{\mathbf{r}}'\mathbf{K}_{\mathbf{r}}^{-1}\boldsymbol{\mu}_{\mathbf{r}} + \mathbf{G}^{[0]}\mathbf{H}^{-1}\mathbf{G}^{[0]\prime} - \mathbf{R}\mathbf{A}^{-1}\mathbf{R}' - (\mathbf{G}^{[1]} - \mathbf{R}\mathbf{A}^{-1}\mathbf{C}')(\mathbf{H} - \mathbf{C}\mathbf{A}^{-1}\mathbf{C}')^{-1}(\mathbf{G}^{[1]} - \mathbf{R}\mathbf{A}^{-1}\mathbf{C}')' \\
&+ \tilde{\boldsymbol{\theta}}_n^{[1]\prime}\tilde{\boldsymbol{\Sigma}}_b^{-1}\tilde{\boldsymbol{\theta}}_n^{[1]} - \tilde{\boldsymbol{\theta}}_n^{[0]\prime}\tilde{\boldsymbol{\Sigma}}_b^{-1}\tilde{\boldsymbol{\theta}}_n^{[0]} \\
&= \boldsymbol{\mu}_{\mathbf{r}}'\mathbf{K}_{\mathbf{r}}^{-1}\boldsymbol{\mu}_{\mathbf{r}} + \mathbf{G}^{[0]}\mathbf{H}^{-1}\mathbf{G}^{[0]\prime} - \mathbf{R}\mathbf{A}^{-1}\mathbf{R}' - \mathbf{G}^{[1]}\mathbf{H}^{-1}\mathbf{G}^{[1]\prime} - (\mathbf{R}\mathbf{A}^{-1}\mathbf{C}')\mathbf{H}^{-1}(\mathbf{R}\mathbf{A}^{-1}\mathbf{C}')' \\
&+ 2\mathbf{R}\mathbf{A}^{-1}\mathbf{C}'\mathbf{H}^{-1}\mathbf{G}^{[1]\prime} + \tilde{\boldsymbol{\theta}}_n^{[1]\prime}\tilde{\boldsymbol{\Sigma}}_b^{-1}\tilde{\boldsymbol{\theta}}_n^{[1]} - \tilde{\boldsymbol{\theta}}_n^{[0]\prime}\tilde{\boldsymbol{\Sigma}}_b^{-1}\tilde{\boldsymbol{\theta}}_n^{[0]} \\
&= 2\frac{\mathbf{X}_{Z(n)}'\mathbf{B}_{bZ(n)}}{\sigma_e^2}\tilde{\boldsymbol{\Sigma}}_b\frac{\mathbf{B}_{bZ(n)}'\mathbf{B}_{aZ(n)}}{\sigma_e^2}\tilde{\boldsymbol{\mu}}_a - \tilde{\boldsymbol{\mu}}_a'\frac{\mathbf{B}_{aZ(n)}'\mathbf{B}_{bZ(n)}}{\sigma_e^2}\tilde{\boldsymbol{\Sigma}}_b\frac{\mathbf{B}_{bZ(n)}'\mathbf{B}_{bZ(n)}}{\sigma_e^2}\tilde{\boldsymbol{\Sigma}}_b\frac{\mathbf{B}_{bZ(n)}'\mathbf{B}_{aZ(n)}}{\sigma_e^2}\tilde{\boldsymbol{\mu}}_a \\
&- 2\frac{\mathbf{X}_{Z(n)}'\mathbf{B}_{aZ(n)}}{\sigma_e^2}\boldsymbol{\mu}_{\mathbf{r}} + \boldsymbol{\mu}_{\mathbf{r}}'\frac{\mathbf{B}_{aZ(n)}'\mathbf{B}_{aZ(n)}}{\sigma_e^2}\boldsymbol{\mu}_{\mathbf{r}} - \boldsymbol{\mu}_{\mathbf{r}}'\frac{\mathbf{B}_{aZ(n)}'\mathbf{B}_{bZ(n)}}{\sigma_e^2}\tilde{\boldsymbol{\Sigma}}_b\frac{\mathbf{B}_{bZ(n)}'\mathbf{B}_{aZ(n)}}{\sigma_e^2}\boldsymbol{\mu}_{\mathbf{r}} \\
&+ \boldsymbol{\mu}_{\mathbf{r}}'\frac{\mathbf{B}_{aZ(n)}'\mathbf{B}_{bZ(n)}}{\sigma_e^2}\tilde{\boldsymbol{\Sigma}}_b\frac{\mathbf{B}_{bZ(n)}'\mathbf{B}_{bZ(n)}}{\sigma_e^2}\tilde{\boldsymbol{\Sigma}}_b\frac{\mathbf{B}_{bZ(n)}'\mathbf{B}_{aZ(n)}}{\sigma_e^2}\boldsymbol{\mu}_{\mathbf{r}} + 2\boldsymbol{\mu}_{\mathbf{r}}'\frac{\mathbf{B}_{aZ(n)}'\mathbf{B}_{bZ(n)}}{\sigma_e^2}\tilde{\boldsymbol{\theta}}_n^{[1]} \\
&- 2\boldsymbol{\mu}_{\mathbf{r}}'\frac{\mathbf{B}_{aZ(n)}'\mathbf{B}_{bZ(n)}}{\sigma_e^2}\tilde{\boldsymbol{\Sigma}}_b\frac{\mathbf{B}_{bZ(n)}'\mathbf{B}_{bZ(n)}}{\sigma_e^2}\tilde{\boldsymbol{\theta}}_n^{[1]} + 2\boldsymbol{\mu}_{\mathbf{r}}'\frac{\mathbf{B}_{aZ(n)}'\mathbf{B}_{bZ(n)}}{\sigma_e^2}\tilde{\boldsymbol{\Sigma}}_b\frac{\mathbf{B}_{bZ(n)}'\mathbf{X}_{Z(n)}}{\sigma_e^2} \\
&- 2\boldsymbol{\mu}_{\mathbf{r}}'\frac{\mathbf{B}_{aZ(n)}'\mathbf{B}_{bZ(n)}}{\sigma_e^2}\tilde{\boldsymbol{\Sigma}}_b\frac{\mathbf{B}_{bZ(n)}'\mathbf{B}_{bZ(n)}}{\sigma_e^2}\tilde{\boldsymbol{\Sigma}}_b\frac{\mathbf{B}_{aZ(n)}'\mathbf{X}_{Z(n)}}{\sigma_e^2} \\
&= -2\boldsymbol{\mu}_{\mathbf{r}}\frac{\mathbf{B}_{aZ(n)}'}{\sigma_e^2}(\mathbf{X}_{Z(n)} - \mathbf{B}_{bZ(n)}\tilde{\boldsymbol{\theta}}_n^{[1]}) + \boldsymbol{\mu}_r'\frac{\mathbf{B}_{aZ(n)}'\mathbf{B}_{aZ(n)}}{\sigma_e^2}\boldsymbol{\mu}_{\mathbf{r}} + 2\tilde{\boldsymbol{\mu}}_a'\frac{\mathbf{B}_{aZ(n)}'}{\sigma_e^2}\hat{\mathbf{H}}\mathbf{X}_{Z(n)} - 2\boldsymbol{\mu}_{\mathbf{r}}'\frac{\mathbf{B}_{aZ(n)}'}{\sigma_e^2}\hat{\mathbf{H}}\mathbf{B}_{bZ(n)}\tilde{\boldsymbol{\theta}}_n^{[1]} \\
&- \tilde{\boldsymbol{\mu}}_a'\frac{\mathbf{B}_{aZ(n)}'}{\sigma_e^2}\hat{\mathbf{H}}\mathbf{B}_{aZ(n)}\tilde{\boldsymbol{\mu}}_a,
\end{aligned}
$$

where $\hat{\mathbf{H}} = \mathbf{B}_{bZ(n)}(\mathbf{B}_{bZ(n)}'\mathbf{B}_{bZ(n)})^{-1}\mathbf{B}_{bZ(n)}'$. Since $\mathbf{A}^{-1} \approx \mathbf{K}_{\mathbf{r}}$ and the entries of $\mathbf{C}\mathbf{A}^{-1}\mathbf{C}'$ are quite small, $\sqrt{|\mathbf{H}|/|\mathbf{K}_{\mathbf{r}}||\mathbf{A}||\mathbf{H} - \mathbf{C}\mathbf{A}^{-1}\mathbf{C}'|} \approx 1$, and we drop this constant term. Such that

$$
\begin{aligned}
PBF = \mathbb{E}_{\mathbf{r}}\Bigg[&\exp\Big(\boldsymbol{\mu}_{\mathbf{r}}\frac{\mathbf{B}_{aZ(n)}'}{\sigma_e^2}(\mathbf{X}_{Z(n)} - \mathbf{B}_{bZ(n)}\tilde{\boldsymbol{\theta}}_n^{[1]}) - \frac{1}{2}\boldsymbol{\mu}_r'\frac{\mathbf{B}_{aZ(n)}'\mathbf{B}_{aZ(n)}}{\sigma_e^2}\boldsymbol{\mu}_{\mathbf{r}} - \tilde{\boldsymbol{\mu}}_a'\frac{\mathbf{B}_{aZ(n)}'}{\sigma_e^2}\hat{\mathbf{H}}\mathbf{X}_{Z(n)} \\
&+\boldsymbol{\mu}_{\mathbf{r}}'\frac{\mathbf{B}_{aZ(n)}'}{\sigma_e^2}\hat{\mathbf{H}}\mathbf{B}_{bZ(n)}\tilde{\boldsymbol{\theta}}_n^{[1]} + \frac{1}{2}\tilde{\boldsymbol{\mu}}_a'\frac{\mathbf{B}_{aZ(n)}'}{\sigma_e^2}\hat{\mathbf{H}}\mathbf{B}_{aZ(n)}\tilde{\boldsymbol{\mu}}_a\Big)\Bigg].
\end{aligned}
$$

However, it is hard to compute the expectation function over nonlinear $\exp(\cdot)$ function. It will further make the objective function of Thompson sampling complicated and require to enumerate all the possible combinations of Z. To simplify PBF and also make Thompson sampling easy to conduct, we would like to remove the nonlinearity by using approximation methods. Consider in normal condition, the term in $\exp(\cdot)$ is quite close to 0, it is reasonable to use the first-order approximation of Taylor expansion to make $\mathbb{E}_{\mathbf{r}}[\exp(\cdot)] \approx \mathbb{E}_{\mathbf{r}}[\cdot]$. On the other hand, under abnormal condition, the term inside the exponential function would deviate from zero, so the first-order approximation of exponential function is more conservative, which may cause the detection power sacrificed a little. However, if we can still estimate $\boldsymbol{\theta}_a$ accurately, that would not be a concern, as shown by the excellent performance of our method in numerical studies. Then we can take expectation function on the linear items and PBF is simplified. So we define the approximated PBF as the detection statistic:

$$\Lambda_n \equiv 2\tilde{\boldsymbol{\mu}}_a'\mathbf{B}_{aZ(n)}'(\mathbf{I}-\hat{\mathbf{H}})(\mathbf{X}_{Z(n)} - \mathbf{B}_{bZ(n)}\tilde{\boldsymbol{\theta}}_n^{[1]}) - \boldsymbol{\mu}_a'(\mathbf{B}_{aZ(n)}'\mathbf{B}_{aZ(n)} \circ \bar{\mathbf{A}})\boldsymbol{\mu}_a + \tilde{\boldsymbol{\mu}}_a'\mathbf{B}_{aZ(n)}'\hat{\mathbf{H}}\mathbf{B}_{aZ(n)}\tilde{\boldsymbol{\mu}}_a,$$

where $\hat{\mathbf{H}} = \mathbf{B}_{bZ(n)}(\mathbf{B}_{bZ(n)}'\mathbf{B}_{bZ(n)})^{-1}\mathbf{B}_{bZ(n)}'$. $\bar{\mathbf{A}}$ has diagonal items $\bar{A}_{ii} = \alpha_i, i = 1, \ldots, k_a$, and other items $\bar{A}_{ij} = \alpha_i\alpha_j$, for all $i, j = 1, \ldots, k_a, i \neq j$.

C: Verification of Subspace Orthogonal Property

Suppose a vector $\mathbf{b} \in \mathcal{R}^p$. Without loss of generality, assume $\|\mathbf{b}\|_2^2 = 1$(since otherwise we can set it as $\frac{\mathbf{b}}{\|\mathbf{b}\|}$). Denote $\mathbf{b}^2 = (b_1^2, b_2^2, ..., b_p^2)'$ and assume that $\|\mathbf{b}^2\|_\infty \leq \frac{c}{p}\|\mathbf{b}\|_2^2$, where $1 \leq c \leq p$. Consider $\mathbf{P}$ is subspace projection matrix from $\mathcal{R}^p \mapsto \mathcal{R}^m$, where m out of p dimensions have $P_{ii} = 1$, and all other entries of $\mathbf{P}$ have values of 0. Without loss of generality, we can assume that $\mathbb{E}[P_{ii}] \geq \mathbb{E}[P_{jj}]$ for any $i \leq j$. Thus, $\mathbb{E}[\mathbf{P}] = \mathrm{diag}\{a_1, a_2, ..., a_p\}$, where $0 \leq a_p \leq a_{p-1} \leq ... \leq a_1 \leq 1$ and $\sum_{i=1}^p a_i = m$. Then we have

$$a_p\|\mathbf{b}\|_2^2 \leq \mathbb{E}[\|\mathbf{P}\mathbf{b}\|_2^2] \leq a_1\|\mathbf{b}\|_2^2.$$

Following Hoeffding (1994); Lim and Durrant (2017), we have one side

$$P(\frac{1}{a_1}\|\mathbf{Pb}\|_2^2 - \|\mathbf{b}\|_2^2 \geq \epsilon) \leq P(\frac{1}{a_1}\|\mathbf{Pb}\|_2^2 - \frac{1}{a_1}\mathbb{E}[\|\mathbf{Pb}\|_2^2] \geq \epsilon) \leq \exp\Big(-\frac{2a_1^2\epsilon^2}{m\|\mathbf{b}^2\|_\infty^2}\Big)$$
$$\leq \exp\Big(-\frac{2a_1^2p^2\epsilon^2}{mc^2}\Big) \leq \exp\Big(-\frac{2(\frac{m}{p})^2p^2\epsilon^2}{mc^2}\Big) = \exp\Big(-\frac{2m\epsilon^2}{c^2}\Big),$$

where $\epsilon \in (0,1]$ is a suitably small value and the other side

$$P(\|\mathbf{b}\|_2^2 - \frac{1}{a_p}\|\mathbf{Pb}\|_2^2 \geq \epsilon) \leq P(\frac{1}{a_p}\mathbb{E}[\|\mathbf{Pb}\|_2^2] - \frac{1}{a_p}\|\mathbf{Pb}\|_2^2 \geq \epsilon) \leq \exp\Big(-\frac{2a_p^2\epsilon^2}{m\|\mathbf{b}^2\|_\infty^2}\Big)$$
$$\leq \exp\Big(-\frac{2a_p^2p^2\epsilon^2}{mc^2}\Big).$$

In our scenario, we instantiate $\mathbf{b}$ as $\frac{\mathbf{b}_{ai}-\mathbf{b}_{bj}}{\|\mathbf{b}_{ai}-\mathbf{b}_{bj}\|}$, for $i = 1,\ldots,k_a$, $j = 1,\ldots,k_b$, where $\mathbf{b}_{ai}$ and $\mathbf{b}_{bj}$ are any column of $\mathbf{B}_a$ and $\mathbf{B}_b$ respectively. And there are k_ak_b such combinations. Thus we get

$$P(\frac{1}{a_1}\|\mathbf{P}(\mathbf{b}_{ai}-\mathbf{b}_{bj})\|_2^2 - \|\mathbf{b}_{ai}-\mathbf{b}_{bj}\|_2^2 \geq \epsilon\|\mathbf{b}_{ai}-\mathbf{b}_{bj}\|_2^2) \leq k_ak_b\exp\Big(-\frac{2m\epsilon^2}{c^2}\Big),$$

and

$$P(\|\mathbf{b}_{ai}-\mathbf{b}_{bj}\|_2^2 - \frac{1}{a_p}\|\mathbf{P}(\mathbf{b}_{ai}-\mathbf{b}_{bj})\|_2^2 \geq \epsilon\|\mathbf{b}_{ai}-\mathbf{b}_{bj}\|_2^2) \leq k_ak_b\exp\Big(-\frac{2a_p^2p^2\epsilon^2}{mc^2}\Big),$$

for any $i = 1,\ldots,k_a$, $j = 1,\ldots,k_b$. Let the right side of these two inequalities equaling a suitably small value $\delta \in (0,1]$. Then we obtain with probability $1-\delta$,

$$\frac{1}{a_1}\|\mathbf{P}(\mathbf{b}_{ai}-\mathbf{b}_{bj})\|_2^2 \leq (1+\epsilon)\|\mathbf{b}_{ai}-\mathbf{b}_{bj}\|_2^2.$$

Also with probability $1-\delta$,

$$\frac{1}{a_p}\|\mathbf{P}(\mathbf{b}_{ai}-\mathbf{b}_{bj})\|_2^2 \geq (1-\epsilon)\|\mathbf{b}_{ai}-\mathbf{b}_{bj}\|_2^2.$$

And assume $\mathbf{B}_a$ and $\mathbf{B}_b$ are two orthogonal spaces, i.e., $\mathbf{B}_a'\mathbf{B}_b = \mathbf{0}$. Following Corollary 1 in Lim and Durrant (2017), for all possible m-dimensional subsets Z in $\mathscr{Z} = \left\{Z_k, k = 1, ..., \binom{p}{m}\right\}$, with probability $1 - 2\delta$, we also have

$$-a_1\epsilon \leq \mathbf{b}_{aiZ}'\mathbf{b}_{bjZ} \leq a_p\epsilon,$$

from the foregoing two-side constraints. Thus we can obtain that when $\frac{c^2}{2\epsilon^2}\ln(k_a k_b/\delta) \leq m \leq \frac{2a_p^2 p^2\epsilon^2}{c^2\ln(k_a k_b/\delta)}$, $-a_1\epsilon \leq \mathbf{b}_{aiZ}'\mathbf{b}_{bjZ} \leq a_p\epsilon$ holds with probability $1 - 2\delta$, where $0 \leq a_p \leq a_1 \leq 1$. So we can verify that the subspaces of $\mathbf{B}_a$ and $\mathbf{B}_b$ are approximately orthogonal when m satisfies the foregoing conditions.

D: Proof of Theorem 1 and Theorem 2

The Thompson sampling procedure is to sample Z by ranking $\Lambda_i = \big(2\hat{\mathbf{X}}_{1i}\mathbf{B}_{ai}\tilde{\boldsymbol{\mu}}_a - \boldsymbol{\mu}_a'(\mathbf{B}_{ai}'\mathbf{B}_{ai} \circ \bar{\mathbf{A}})\boldsymbol{\mu}_a\big), i = 1, ..., p$, from the largest to the smallest and select the top m variables. $\hat{\mathbf{X}}_1$ is generated by sampling $\hat{\boldsymbol{\theta}}_a$ from $\tilde{p}(\boldsymbol{\theta}_a, \mathbf{r})$, sampling $\hat{\mathbf{E}}$ from $N(\mathbf{0}, \boldsymbol{\Sigma}_e)$ and getting $\hat{\mathbf{X}}_1 = \mathbf{B}_a\hat{\boldsymbol{\theta}}_a + \hat{\mathbf{E}}$.

Since the posterior distribution of $\boldsymbol{\theta}_a$ is of spike-slab form, Λ_i follows Gaussian mixture distribution, which means it follows 2^{k_a} Gaussian distributions, each with different probability. Specifically, if denote $S = \{1, 2, ..., k_a\}$, for any subset S_0 of S, we have

$$\begin{aligned}\Lambda_i \sim N\Big(&2\sum_{k\in S\setminus S_0} B_{aik}\mu_{ak}\sum_{j\in S} B_{aij}\mu_{aj}\alpha_j - \sum_{j\in S} B_{aij}^2\mu_{aj}^2\alpha_j - 2\sum_{for\ all\ j_1,j_2\in S, j_1\neq j_2} B_{aij_1}B_{aij_2}\mu_{aj_1}\mu_{aj_2}\alpha_{j_1}\alpha_{j_2}, \\ &4(\sum_{j\in S\setminus S_0} B_{aij}^2 s_j^2 + \sigma_e^2)(\sum_{j\in S} B_{aij}\mu_{aj}\alpha_j)^2\Big),\end{aligned} \tag{1}$$

with probability $\prod_{j\in S\setminus S_0}\alpha_j\prod_{j\in S_0}(1-\alpha_j)$.

According to Theorem 5 of Wang and Blei (2019), the variational Bayesian posterior converges to the point mass of the true parameter value in distribution. Note that the posterior in (5) is not exactly the true posterior, but the exponentially weighted posterior. Thus we need $\lambda \to 0$, i.e., representing all samples receive almost equivalent weights, to guarantee as $n \to \infty$, all the samples could be used for estimating $\boldsymbol{\theta}_a$. This assumption of $\lambda \to 0$ is common in theoretical analysis of other online process monitoring schemes with exponential weights (Zhou et al. 2012; Zou et al. 2012). In our case, to guarantee the algorithm can detect the anomaly efficiently and select the most anomalous variables, we need $\boldsymbol{\theta}_a$ is accurately estimated, which can be guaranteed only if $n \to \infty$. However, the estimation of $\boldsymbol{\theta}_n$ and $\boldsymbol{\theta}_a$ are coupled. To remove the influence of estimation error of $\boldsymbol{\theta}_n$ on $\boldsymbol{\theta}_a$, first consider a special case that there is no background signal, i.e., $\mathbf{B}_b\boldsymbol{\theta}_n = \mathbf{0}$. In normal condition, the true value of θ_{aj} equals 0, for $j = 1, \ldots, k_a$. The posterior distribution that we obtain through variational Bayesian method is of spike-slab form. For example, $\tilde{p}_j(\theta_{aj}) \sim N(\mu_{aj}, s_j^2)$ with probability α_j and $\theta_{aj} = \delta_0$ with probability $1 - \alpha_j$. Then suggested by Theorem 5 of Wang and Blei (2019), as $\lambda \to 0$, $n \to \infty$,

$$\tilde{p}_j(\theta_{aj}) \xrightarrow{d} \delta_0, \quad j = 1, \ldots, k_a, \tag{2}$$

where δ_0 is a point mass at 0. That suggests $\mu_{aj} \to 0$ and $s_j^2 \to 0$, for $j = 1, ..., k_a$. So, from (1), in normal condition, $\mathbb{E}[\Lambda_i] \to 0$ and $\mathrm{Var}[\Lambda_i] \to 0$, for $i = 1, ..., p$, which means under the limit conditions, we sample the variables $Z(n+1)$ randomly.

Following a similar way, in the abnormal condition, assume the anomaly relates to certain bases $\mathcal{A} \subset S$. For $l \in \mathcal{A}$, assume the anomaly relates to the l^{th} base has change magnitude ϕ_l. Then suggested by Theorem 5 of Wang and Blei (2019), as $\lambda \to 0$, $n \to \infty$,

$$\tilde{p}_l(\theta_{al}) \xrightarrow{d} \delta_{\phi_l}, \quad l \in \mathcal{A}, \tag{3}$$

$$\tilde{p}_j(\theta_{aj}) \xrightarrow{d} \delta_0, \quad j \in S - \mathcal{A}, \tag{4}$$

where δ_{ϕ_l} is a point mass at ϕ_l. That suggests $\mu_{al} \to \phi_l$, $\alpha_l \to 1$ and $s_l^2 \to 0$, for $l \in \mathcal{A}$.

The same as normal condition, $\mu_{aj} \to 0$ and $s_j^2 \to 0$, for $j \neq l$. So, from (1), in abnormal condition, $\mathbb{E}[\Lambda_i] \to (\sum_{l\in\mathcal{A}} B_{ail}\phi_l)^2$ and $\text{Var}[\Lambda_i] \to 4\sigma_e^2(\sum_{l\in\mathcal{A}} B_{ail}\phi_l)^2$, for $i = 1, \ldots, p$. Thus in abnormal condition, we prefer to choose the variables most influenced by the abnormal patterns.

For general cases with $\mathbf{B}_b\boldsymbol{\theta}_n$, when $p \to \infty$, $m \to \infty$, and $0 < m/p \leq 1$, according to the asymptotic normality and consistency of Bayesian theory, $\tilde{\boldsymbol{\theta}}_n$ will converge to Gaussian distribution with mean equal to $\boldsymbol{\theta}_n$ and variance matrix equal to $(mJ(\boldsymbol{\theta}_n))^{-1}$, where $J(\boldsymbol{\theta}_n)$ is the Fisher information matrix evaluated at the true value $\boldsymbol{\theta}_n$. Hence, $\tilde{\boldsymbol{\theta}}_n$ will converge to its true value $\boldsymbol{\theta}_n$ as $m \to \infty$. Then the properties of $\boldsymbol{\theta}_a$ in (2) and (3) still hold. Consequently, Theorem 1 and Theorem 2 hold.

E: Simulation Results for 1D and 2D Cases

Table 1: Average Detection Delays/ADDs(Standard Deviation of Detection Delays/STDs) for 1D data with $p = 100$, $m = 10$.

ϕ	TRAS	CMAB(s)	NAS	R-SADA	SASAM	TSSRP	TS-BSSCD	TS-BSSCD(I)	ORACLE
0.0	200(148)	200(192)	200(310)	200(192)	200(191)	200(191)	200(207)	200(215)	200(456)
0.1	188(136)	98.2(75.5)	198(337)	198(194)	192(169)	200(76.7)	41.6(40.1)	**40.3**(36.8)	1.33(0.88)
0.2	168(124)	38.3(16.7)	152(241)	186(183)	183(168)	188(73.4)	**11.1**(8.27)	13.3(11.9)	1.07(0.27)
0.3	162(126)	28.1(11.6)	91.1(132)	183(180)	151(133)	150(59.7)	**7.23**(5.60)	10.0(10.0)	1.09(0.09)
0.4	150(124)	23.7(10.6)	58.2(77.0)	164(156)	114(94.3)	98.5(39.6)	**5.82**(4.33)	8.98(14.0)	1.00(0.00)
0.5	149(122)	21.4(10.1)	36.4(41.2)	152(149)	74.8(61.4)	63.6(23.9)	**5.03**(4.33)	9.09(9.64)	1.00(0.00)
0.6	143(122)	20.5(10.4)	26.3(25.0)	126(118)	54.8(47.1)	45.9(14.8)	**4.77**(4.18)	9.06(14.0)	1.00(0.00)
0.7	135(121)	19.3(10.6)	21.4(19.4)	105(111)	42.3(32.8)	35.9(10.8)	**4.49**(4.06)	8.97(9.67)	1.00(0.00)
0.8	139(122)	18.4(10.7)	18.5(14.3)	92.0(91.6)	33.9(25.2)	29.7(7.11)	**3.84**(3.34)	9.33(10.3)	1.00(0.00)
0.9	136(122)	17.9(10.6)	16.2(11.2)	83.4(96.6)	28.5(20.2)	26.1(5.93)	**3.98**(3.44)	9.22(14.6)	1.00(0.00)
1.0	131(123)	18.2(10.9)	14.1(9.13)	71.4(79.6)	23.0(15.0)	23.0(5.06)	**3.85**(3.36)	9.21(10.0)	1.00(0.00)

Table 2: Average Detection Delays/ADDs(Standard Deviation of Detection Delays/STDs) for 1D data with $p = 100$, $m = 20$.

ϕ	TRAS	CMAB(s)	NAS	R-SADA	SASAM	TSSRP	TS-BSSCD	TS-BSSCD(I)	ORACLE
0.0	200(137)	200(169)	200(348)	200(194)	200(175)	200(139)	200(255)	200(215)	200(456)
0.1	181(120)	58.3(38.7)	161(282)	135(141)	193(196)	185(111)	**14.3**(13.8)	22.1(18.9)	1.33(0.88)
0.2	148(113)	25.2(11.1)	76.2(133)	131(132)	175(166)	164(98.8)	**5.14**(3.82)	7.39(7.60)	1.07(0.27)
0.3	126(101)	19.2(9.67)	34.3(56.9)	116(112)	147(134)	112(57.4)	3.**17**(2.32)	4.97(5.49)	1.09(0.09)
0.4	118(101)	16.5(8.66)	20.0(25.0)	109(109)	100(88.3)	69.1(28.5)	**2.66**(2.01)	4.53(5.02)	1.000(0.00)
0.5	105(93.9)	15.1(8.67)	15.3(16.5)	97.2(98.8)	66.9(49.4)	46.1(16.2)	**2.40**(1.97)	4.62(5.60)	1.000(0.00)
0.6	94.5(91.2)	13.8(8.61)	11.8(12.3)	76.4(83.6)	48.4(35.1)	35.0(10.5)	**2.11**(1.59)	4.97(8.96)	1.000(0.00)
0.7	102(102)	13.1(8.72)	10.3(8.85)	52.6(56.8)	36.4(28.1)	27.8(7.65)	**2.18**(1.68)	5.07(5.68)	1.000(0.00)
0.8	100(100)	12.4(8.64)	8.60(7.04)	41.0(42.1)	29.2(20.9)	23.4(5.31)	**1.91**(1.48)	4.88(5.46)	1.000(0.00)
0.9	87.6(91.9)	11.9(8.87)	7.92(5.92)	30.7(28,7)	24.2(16.0)	20.4(4.25)	**1.93**(1.42)	4.79(5.55)	1.000(0.00)
1.0	86.5(92.6)	12.2(8.98)	7.24(5.08)	26.7(26.9)	19.9(12.1)	18.3(3.34)	**1.85**(1.33)	4.59(4.44)	1.000(0.00)

Table 3: Average Detection Delays/ADDs(Standard Deviation of Detection Delays/STDs) for 1D data with $p = 100$, $m = 30$.

ϕ	TRAS	CMAB(s)	NAS	R-SADA	SASAM	TSSRP	TS-BSSCD	TS-BSSCD(I)	ORACLE
0.0	200(147)	200(169)	200(313)	200(209)	200(172)	200(134)	200(335)	200(256)	200(456)
0.1	169(122)	43.4(28.6)	175(263)	189(189)	200(192)	176(122)	**8.69**(8.96)	15.4(15.4)	1.33(0.88)
0.2	132(92.5)	18.9(9.35)	77.5(119)	185(197)	176(163)	148(93.3)	**3.08**(2.41)	4.94(4.97)	1.07(0.27)
0.3	105(80.9)	14.4(8.08)	35.1(51.9)	186(197)	142(137)	91.6(50.8)	**2.11**(1.41)	3.33(3.19)	1.09(0.09)
0.4	94.4(78.5)	12.0(7.45)	18.9(23.8)	163(171)	95.7(85.0)	57.5(23.6)	**1.81**(1.16)	3.29(3.24)	1.00(0.00)
0.5	85.1(75.7)	11.2(7.33)	13.4(15.0)	151(176)	63.2(52.4)	38.7(13.5)	**1.69**(1.12)	3.11(3.02)	1.00(0.00)
0.6	74.9(72.8)	10.0(7.12)	10.9(11.9)	102(123)	43.6(28.5)	29.7(8.51)	**1.54**(0.97)	3.27(4.21)	1.00(0.00)
0.7	70.1(72.9)	9.70(7.34)	9.12(8.20)	60.3(73.9)	32.6(20.6)	24.1(5.91)	**1.47**(0.90)	3.17(3.08)	1.00(0.00)
0.8	64.3(69.1)	8.55(7.00)	7.54(6.57)	37.4(46.5)	26.5(16.3)	20.9(4.89)	**1.42**(0.77)	3.17(3.25)	1.00(0.00)
0.9	65.7(75.6)	8.44(7.20)	6.80(4.98)	26.6(28.3)	21.3(11.2)	18.1(3.66)	**1.36**(0.71)	3.24(3.07)	1.00(0.00)
1.0	59.1(70.1)	8.33(7.37)	6.37(4.52)	18.4(16.9)	18.0(9.15)	16.0(3.11)	**1.45**(0.87)	3.36(3.46)	1.00(0.00)

F: Sensitivity of Parameters

Sensitivity of λ

Actually, we borrow the idea of exponentially weight moving average (EWMA) in (5). It is used to accumulate historical online samples for better small shift detection. The decayed weight guarantees more recent time points have larger weights, while more previous time points have smaller ones since they may be before the true change point and contain normal samples. Generally λ is set to be a small value. The popular choices of λ are $0.05, 0.1, 0.2$. For small changes, the smaller λ is, the better detection will be (Montgomery 2020). Here we set $\lambda = 0.05, 0.1, 0.2$ respectively and run one case of 2D experiment with $m = 40$ to

Table 4: Average Detection Delays/ADDs(Standard Deviation of Detection Delays/STDs) for 2D data with $p = 400$, $m = 20$.

ϕ	TRAS	CMAB(s)	NAS	R-SADA	SASAM	TSSRP	TS-BSSCD	TS-BSSCD(I)	ORACLE
0.0	200(128)	200(171)	200(137)	200(195)	200(181)	200(6.75)	200(194)	200(204)	200(479)
0.1	188(128)	160(123)	205(146)	187(191)	136(108)	194(6.62)	**97.5**(102)	165(167)	1.71(1.57)
0.2	166(101)	85.8(59.9)	188(152)	186(186)	128(107)	187(10.1)	**21.9**(19.2)	83.3(99.1)	1.09(0.34)
0.3	143(80.1)	54.4(27.9)	167(160)	167(174)	102(80.8)	171(14.7)	**12.4**(7.42)	34.9(46.2)	1.02(0.15)
0.4	121(67.4)	39.2(17.7)	147(169)	164(161)	78.2(58.9)	155(16.8)	**9.48**(5.74)	17.6(19.5)	0.00(0.08)
0.5	108(62.9)	31.7(12.8)	94.4(116)	162(168)	59.5(47.4)	141(17.9)	**7.87**(5.32)	11.4(10.7)	1.00(0.00)
0.6	102(59.7)	27.0(11.2)	56.2(58.2)	140(150)	48.8(40.9)	128(16.7)	**6.87**(4.60)	7.96(7.17)	1.00(0.00)
0.7	92.0(57.1)	24.1(10.3)	39.0(33.0)	128(140)	37.9(27.6)	117(16.2)	**6.34**(4.26)	6.60(6.03)	1.00(0.00)
0.8	87.1(57.0)	21.8(9.84)	33.4(22.5)	118(135)	32.4(23.3)	109(15.1)	**5.71**(3.93)	5.82(5.45)	1.00(0.00)
0.9	82.1(56.8)	20.8(9.91)	27.7(17.0)	111(128)	29.7(21.5)	102(13.9)	5.22(3.90)	**4.64**(4.41)	1.00(0.00)
1.0	77.1(52.0)	19.6(9.90)	26.4(14.7)	105(128)	24.3(17.8)	98.2(14.5)	5.17(3.97)	**4.28**(4.24)	1.00(0.00)

Table 5: Average Detection Delays/ADDs(Standard Deviation of Detection Delays/STDs) for 2D data with $p = 400$, $m = 40$.

ϕ	TRAS	CMAB(s)	NAS	R-SADA	SASAM	TSSRP	TS-BSSCD	TS-BSSCD(I)	ORACLE
0.0	200(149)	200(171)	200(208)	200(192)	200(166)	200(9.51)	200(219)	200(206)	200(479)
0.1	195(128)	114(93.2)	190(218)	193(185)	126(109)	187(10.0)	**35.1**(48.0)	144(159)	1.71(1.57)
0.2	160(91.2)	46.7(27.2)	178(213)	189(183)	117(100)	178(12.5)	**8.96**(6.87)	46.8(60.2)	1.09(0.34)
0.3	127(68.5)	29.7(10.7)	174(259)	171(172)	98.4(76.7)	160(15.8)	**5.95**(3.97)	16.5(18.0)	1.02(0.15)
0.4	108(56.4)	23.6(7.59)	136(233)	178(183)	78.5(62.8)	143(16.4)	**4.45**(3.04)	8.40(8.15)	0.00(0.08)
0.5	95.9(51.3)	20.3(7.29)	74.1(120)	157(183)	55.7(47.7)	128(16.5)	**3.75**(2.68)	5.94(5.30)	1.00(0.00)
0.6	82.6(46.7)	18.4(7.24)	43.3(56.1)	145(179)	42.5(31.7)	116(15.1)	**3.29**(2.44)	4.31(3.81)	1.00(0.00)
0.7	76.4(44.9)	16.3(7.14)	25.7(28.3)	140(207)	37.1(27.6)	106(15.1)	**2.96**(2.10)	3.23(2.55)	1.00(0.00)
0.8	68.1(38.2)	15.6(7.28)	20.1(19.3)	130(203)	30.4(21,2)	98.3(13.8)	**2.63**(1.99)	2.77(2.31)	1.00(0.00)
0.9	64.8(41.1)	15.2(7.29)	16.1(13.5)	94.0(142)	27.9(22.1)	91.3(12.8)	2.54(1.84)	**2.44**(1.97)	1.00(0.00)
1.0	60.0(38.6)	14.0(7.35)	13.8(9.70)	82.3(124)	22.8(14.6)	85.1(11.5)	2.47(1.79)	**2.31**(1.94)	1.00(0.00)

justify the influence of choosing different λ. The results are in Table 7. As we expect, a small λ can achieve better performance for small change magnitude, while a large λ is better for large changes. Since $\lambda = 0.1$ can balance the whole range of changes well, we set $\lambda = 0.1$ in our manuscript.

Sensitivity of estimation of σ_e and σ_0

In practice, we can estimate σ_e and $\mathbf{\Sigma}_0 = \sigma_0^2\mathbf{I}$ from historical reference samples in Phase I. In particular, assume we have N historical reference samples. We can estimate their $\boldsymbol{\theta}_t$ using maximum likelihood estimation (MLE). Define $\mathbf{X}_t$ is the t^{th} sample. Its fitted coefficients are $\hat{\boldsymbol{\theta}}_t = (\mathbf{B}_b'\mathbf{B}_b)^{-1}\mathbf{B}_b'\mathbf{X}_t$ and the fitted residuals are $\hat{\mathbf{E}}_t = (\mathbf{I} - \mathbf{B}_b(\mathbf{B}_b'\mathbf{B}_b)^{-1}\mathbf{B}_b')\mathbf{X}_t$,

Table 6: Average Detection Delays/ADDs(Standard Deviation of Detection Delays/STDs) for 2D data with $p = 400$, $m = 60$.

ϕ	TRAS	CMAB(s)	NAS	R-SADA	SASAM	TSSRP	TS-BSSCD	TS-BSSCD(I)	ORACLE
0.0	200(136)	200(180)	200(196)	200(203)	193(165)	200(11.5)	200(258)	200(210)	200(479)
0.1	180(119)	91.2(71.7)	199(222)	196(198)	111(90.4)	193(11.6)	**18.9**(24.7)	136(161)	1.71(1.57)
0.2	145(77.5)	35.5(17.2)	180(205)	185(170)	102(88.2)	181(14.6)	**5.82**(4.46)	36.5(53.9)	1.09(0.34)
0.3	115(57.6)	23.6(7.53)	169(249)	185(190)	83.3(65.5)	161(16.9)	**3.77**(2.63)	12.0(12.8)	1.02(0.15)
0.4	92.3(46.3)	19.2(6.54)	161(260)	178(188)	70.2(56.1)	141(17.3)	**2.96**(1.97)	6.15(5.65)	0.00(0.08)
0.5	80.7(41.3)	16.8(6.12)	93.4(166)	168(208)	53.5(39.1)	126(16.4)	**2.47**(1.74)	4.23(3.41)	1.00(0.00)
0.6	74.6(38.3)	14.8(6.11)	46.5(67.6)	161(208)	41.7(30.7)	114(15.5)	**2.23**(1.44)	3.05(2.31)	1.00(0.00)
0.7	63.4(35.3)	13.7(6.14)	24.9(28.7)	175(296)	37.0(26.9)	104(14.3)	**1.98**(1.25)	2.36(1.84)	1.00(0.00)
0.8	59.4(34.7)	12.7(6.18)	18.2(15.2)	146(276)	31.0(21.7)	95.7(12.8)	**1.91**(1.31)	1.93(1.37)	1.00(0.00)
0.9	54.0(33.0)	12.0(6.21)	14.5(11.0)	104(203)	25.8(17.8)	88.6(12.3)	**1.75**(1.07)	1.84(1.28)	1.00(0.00)
1.0	50.6(31.1)	11.85(6.12)	12.8(9.88)	97.9(201)	22.2(15.7)	83.1(11.9)	**1.70**(1.15)	1.73(1.29)	1.00(0.00)

Table 7: ADDs(STDs) for different settings of λ with $p = 400$, $m = 40$ for 2D experiments.

ϕ	$\lambda = 0.05$	$\lambda = 0.1$	$\lambda = 0.2$
0	200(204)	200(209)	200(209)
0.1	28.4(24.5)	35.5(36.8)	68.3(79.0)
0.2	10.9(7.12)	9.64(6.47)	11.9(13.4)
0.3	6.89(4.65)	5.94(3.77)	5.60(3.95)
0.4	5.44(3.90)	4.46(2.98)	4.04(2.36)
0.5	4.46(3.30)	3.70(2.60)	3.19(2.08)
0.6	4.05(3.10)	3.34(2.38)	2.80(1.76)
0.7	3.58(2.85)	3.05(2.07)	2.66(1.77)
0.8	3.34(2.56)	2.71(1.89)	2.47(1.60)
0.9	2.99(2.36)	2.58(1.90)	1.26(1.52)
1.0	2.88(2.40)	2.61(2.02)	2.11(1.37)

$t = 1, \dots, N$. Therefore, σ_e is estimated as $\hat{\sigma}_e^2 = \frac{1}{Np}\sum_{t=1}^N ||\hat{\mathbf{E}}_t - \bar{\mathbf{E}}||_2^2$, where $\bar{\mathbf{E}}$ is the mean of $\hat{\mathbf{E}}_t$. And $\hat{\sigma}_0^2 = \frac{1}{Nk_b}\sum_{t=1}^N ||\hat{\boldsymbol{\theta}}_t - \bar{\boldsymbol{\theta}}_b||_2^2$, where $\bar{\boldsymbol{\theta}}_b$ is the mean of $\hat{\boldsymbol{\theta}}_t$. In our simulation study, there are two phases. During Phase I, we first generate N historical reference samples, and use them to estimate σ_0 and σ_e. During Phase II, we generate abnormal samples and conduct the monitoring scheme based on the estimated $\hat{\sigma}_e$ and $\hat{\sigma}_0$. We repeat the experiment for $n_{rep} = 1000$ times to calculate the average detection delay (ADD). Also with the n_{rep} replications, we can compute the average estimation bias as $bias_{\sigma_e} = \frac{1}{n_{rep}}\sum_{i=1}^{n_{rep}} |\hat{\sigma}_e^i - \sigma_e|/\sigma_e$ and $bias_{\sigma_0} = \frac{1}{n_{rep}}\sum_{i=1}^{n_{rep}} |\hat{\sigma}_0^i - \sigma_0|/\sigma_0$. We set N equal to three different levels, i.e., $N = p$, $N = p/5$ and $N = p/10$ to see how sensitive our estimation accuracy is to the number of historical reference data. The ADDs are shown in Table 8 for 1D experiment with $m = 10$. Meanwhile, the estimation performance is $bias_{\sigma_e} = 0.020$ and $bias_{\sigma_0} = 0.065$ for $N = p$, $bias_{\sigma_e} = 0.040$ and $bias_{\sigma_0} = 0.140$ for $N = p/5$, and $bias_{\sigma_e} = 0.065$ and $bias_{\sigma_0} = 0.210$ for $N = p/10$, respectively. As we see, with N decreasing, the estimation accuracy decreases. On the one hand, it would not be a concern that the estimation accuracy of σ_0 is not high under small N, since $\hat{\sigma}_0$ is only used as prior value, i.e., $\mathbf{\Sigma}_b = \hat{\sigma}_0^2\mathbf{I}$, in estimating $\boldsymbol{\theta}_n$. On the other hand, the estimation accuracy of σ_e is high even under small N. Therefore, the detection power of our algorithm is not affected by the estimation of σ_e and σ_0 very much, as shown in Table 8.

Table 8: ADDs(STDs) using different estimation values of σ_e and σ_0 with $p = 100$, $m = 10$ for 1D experiment.

ϕ	N=p	N=p/5	N=p/10
0	200(201)	200(241)	200(222)
0.1	40.2(40.4)	38.6(37.0)	39.0(36.0)
0.2	11.3(8.29)	10.8(8.06)	10.8(7.83)
0.3	7.06(5.36)	7.30(5.49)	7.40(5.72)
0.4	5.77(4.73)	5.73(4.46)	5.68(4.78)
0.5	4.80(4.23)	5.13(4.26)	4.84(4.32)
0.6	4.55(3.92)	4.89(4.33)	4.60(3.97)
0.7	4.28(3.79)	4.22(3.81)	4.19(3.58)
0.8	4.17(3.73)	4.40(3.80)	4.04(3.60)
0.9	3.85(3.50)	4.16(3.66)	4.01(3.72)
1.0	3.88(3.50)	3.88(3.43)	3.71(3.48)

References

Hoeffding, W. (1994), "Probability inequalities for sums of bounded random variables," in *The Collected Works of Wassily Hoeffding*, Springer, pp. 409–426.

Lim, N. and Durrant, R. J. (2017), "Linear dimensionality reduction in linear time: Johnson-lindenstrauss-type guarantees for random subspace," *arXiv preprint arXiv:1705.06408*.

Montgomery, D. C. (2020), *Introduction to statistical quality control*, John Wiley & Sons.

Wang, Y. and Blei, D. M. (2019), "Frequentist consistency of variational Bayes," *Journal of the American Statistical Association*, 114, 1147–1161.

Zhou, Q., Zou, C., Wang, Z., and Jiang, W. (2012), "Likelihood-based EWMA charts for monitoring Poisson count data with time-varying sample sizes," *Journal of the American Statistical Association*, 107, 1049–1062.

Zou, C., Wang, Z., and Tsung, F. (2012), "A spatial rank-based multivariate EWMA control chart," *Naval Research Logistics (NRL)*, 59, 91–110.