Change point detection in high dimensions

Sungmin Park

School of Statistics
University of Minnesota

November 4, 2022

Change point detection

We have a sequence of data $\{X_1, \dots, X_T\}$. There may exist a change point in the data s.t.

$$X_t \sim \begin{cases} f_0, & t \leq \tau \\ f_1, & t > \tau \end{cases}$$

The goal is to determine whether a change exists and accurately estimate τ when it exists.

Change point in univariate observation

Recall the univariate case $X_t \sim N(\mu_t, 1)$. We would like to test whether there is a change point in $\{X_t\}_{t=1}^T$. The CUSUM (CUmulative SUM) method determines that a change exists if

$$\max_{1 \le \tau \le n-1} \left\{ \sqrt{\frac{\tau(T-\tau)}{T}} \left| \left(\frac{1}{\tau} \sum_{t=1}^{\tau} X_t \right) - \left(\frac{1}{T-\tau} \sum_{t=\tau+1}^{T} X_t \right) \right| \right\} \ge c.$$

and place a change point at

$$\hat{\tau} = \arg\max_{1 \le \tau \le n-1} \left\{ \sqrt{\frac{\tau(T-\tau)}{T}} \left| \left(\frac{1}{\tau} \sum_{t=1}^{\tau} X_t \right) - \left(\frac{1}{T-\tau} \sum_{t=\tau+1}^{T} X_t \right) \right| \right\}.$$

Change point in multivariate observation

Suppose $X_t \sim N(\mu_t, I)$ where X_t is now a random vector. We still want to perform a test

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_T,$$

vs.

$$H_1 : \exists \tau \text{ s.t. } \mu_1 = \ldots = \mu_\tau \neq \mu_{\tau+1} = \ldots = \mu_T,$$

and correctly locate a change point when it exists. Generalizing the CUSUM method, one can

1. compute the CUSUM statistic $\mathcal{Y}_k$ for each components of $X_t = (X_{t1} \cdots X_{tk} \cdots X_{td})^T$,
2. Aggregate the CUSUM statistic $\mathcal{Y}_k$ to detect a change.

Change point in multivariate observation

Mainly, there are two choices:

1. Maximum: $\mathcal{Y}_\tau^{\text{max}} = \max_{1\le k\le d} \mathcal{Y}_{\tau,k}$,
2. Average: $\mathcal{Y}_\tau^{\text{avg}} = \frac{1}{d}\sum_{k=1}^{d} \mathcal{Y}_{\tau,k}$,

where

$$\mathcal{Y}_{\tau,k} = \sqrt{\frac{\tau(T-\tau)}{T}} \left| \left(\frac{1}{\tau}\sum_{t=1}^{\tau} X_{tk}\right) - \left(\frac{1}{T-\tau}\sum_{t=\tau+1}^{T} X_{tk}\right) \right|.$$

Issues in high dimension

The following examples are adapted from Cho and Fryzlewicz [2015] to illustrate the strengths and weaknesses of maximum / average CUSUM statistic in high dimensions.

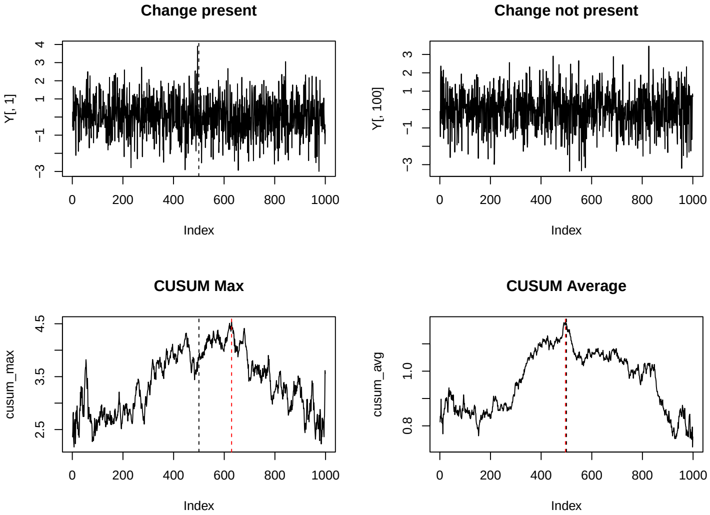
Example 1

$X_t = (X_{t,1} \cdots X_{t,k} \cdots X_{t,100})^T$ is a time series vector with independent $X_{t,k} \sim N(\mu_{t,k}, 1)$ s.t.

$$\mu_{t,k} = \begin{cases} 0.1, & 1 \leq t \leq 500 \;\&\; 1 \leq k \leq 50, \\ 0, & \text{o.w.} \end{cases}$$

The first 50 streams contain a weak change in mean while the other 50 streams do not.

Issues in high dimension



Issues in high dimension

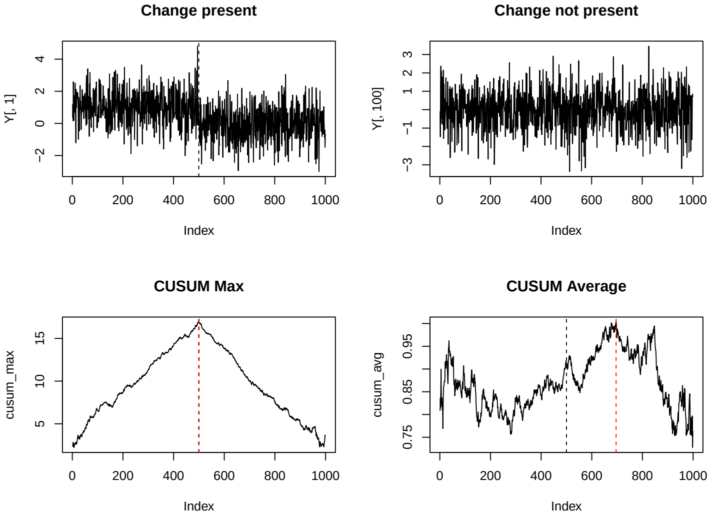
Example 2

$X_t = (X_{t,1} \cdots X_{t,k} \cdots X_{t,100})^T$ is a time series vector with independent $X_{t,k} \sim N(\mu_{t,k}, 1)$ s.t.

$$\mu_{t,k} = \begin{cases} 1, & 1 \leq t \leq 500 \;\&\; k = 1, \\ 0, & \text{o.w.} \end{cases}$$

Change point is only present in the first stream.

Issues in high dimension



CUSUM with threshold

Cho and Fryzlewicz [2015] utilizes thresholding estimator to overcome the limitations in case the signals are sparse:

$$\mathcal{Y}_t^{\text{thr}} = \sum_{k=1}^{d} \mathcal{Y}_{t,k} \mathbf{1}(\mathcal{Y}_{t,k} > \pi_T)$$

where π_T is a threshold on the CUSUM statistic $\mathcal{Y}_{t,k}$. However, the success of a threshold estimator critically depends on the unknown threshold π_T.

CUSUM projection

Instead, Wang and Samworth [2018] presents a data-driven method to detect and estimate change point in high dimensions. Again, suppose

$$X_t \sim \begin{cases} N(\mu_0, I), & t \leq \tau, \\ N(\mu_1, I), & t > \tau. \end{cases}$$

We would like to find a best one-dimensional projection that maximizes the pre- and post-mean difference:

$$a^* = \underset{\|a\|_2=1}{\arg\max} |a^T \mu_1 - a^T \mu_0| = \frac{1}{\|\mu_1 - \mu_0\|_2}(\mu_1 - \mu_0).$$

However, $\mu_1 - \mu_0$ is unknown in most applications.

CUSUM projection

Let $\mathcal{T} : \mathbb{R}^{T\times p} \to \mathbb{R}^{(T-1)\times p}$ be the CUSUM transformation:

$$[\mathcal{T}(X)]_{t,k} = \sqrt{\frac{t(T-t)}{T}} \left(\frac{1}{T-t} \sum_{i=t+1}^{T} X_{ij} - \frac{1}{t} \sum_{i=1}^{t} X_{ij} \right).$$

Note that

$$\mathbf{E}\left[\mathcal{T}(X)\right]_{t,\cdot} = \begin{cases} \sqrt{\frac{t(T-t)}{T}} \left(\frac{T-\tau}{T-t}\mu_1 + \frac{\tau-t}{T-t}\mu_0 - \mu_0 \right)^T, & t < \tau \\ \sqrt{\frac{\tau(T-\tau)}{T}}(\mu_1 - \mu_0)^T, & t = \tau \\ \sqrt{\frac{t(T-t)}{T}}(\mu_1 - \frac{\tau}{t}\mu_0 - \frac{t-\tau}{t}\mu_1)^T, & t > \tau \end{cases}$$

$$= \begin{cases} \sqrt{\frac{t}{T(T-t)}}(T-\tau)(\mu_1 - \mu_0)^T, & t < \tau \\ \sqrt{\frac{\tau(T-\tau)}{T}}(\mu_1 - \mu_0)^T, & t = \tau \\ \sqrt{\frac{T-t}{Tt}}\tau(\mu_1 - \mu_0)^T, & t > \tau \end{cases}$$

CUSUM projection

which shows that $\mathbf{E}\mathcal{T}(X) = u(\mu_1 - \mu_0)^T$ is a rank-1 matrix with left singular vector

$$u = \left(\sqrt{\frac{1}{T(T-1)}}(T-\tau) \cdots \sqrt{\frac{\tau(T-\tau)}{T}} \cdots \sqrt{\frac{1}{T(T-1)}}\tau \right)^T,$$

and right singular vector $(\mu_1 - \mu_0)$. The idea is to estimate $(\mu_1 - \mu_0)$ with the right singular vector $\hat{v}$ of $\mathcal{T}(X)$, and perform a one-dimensional CUSUM test on $X\hat{v}$.

Simulations

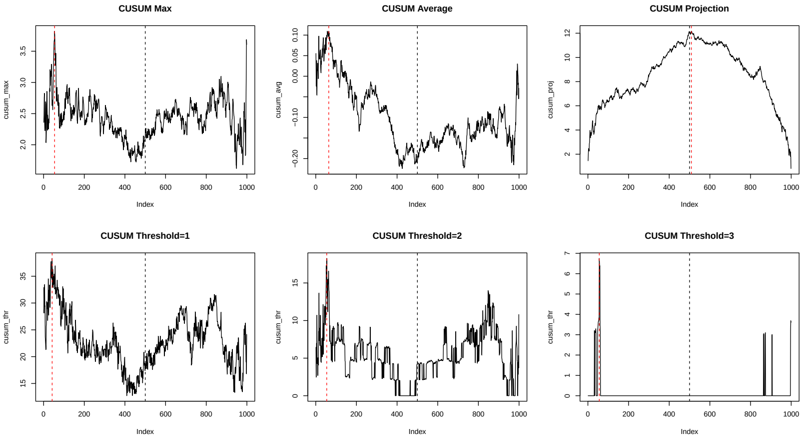
Example 3

$X_t = (X_{t,1} \cdots X_{t,k} \cdots X_{t,100})^T$ is a time series vector with independent $X_{t,k} \sim N(\mu_{t,k}, 1)$ s.t.

$$\mu_{t,k} = \begin{cases} 0.2, & 1 \leq t \leq 500 \text{ \& } 1 \leq k \leq 5, \\ 0, & \text{o.w.} \end{cases}$$

The signal is weak and sparse at the same time.

Simulations



Reference

Haeran Cho and Piotr Fryzlewicz. Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(2):475–507, 2015.

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