Change point analysis in Low and High dimensions

1. Univariate change-point analysis including multiple changes
2. High-dimensional change-point detection
3. Max (over dimension) type, Sum-of-squares type methods
4. Dimensional reduction/transformation approach
5. Group fused lasso

Univariate change-point problem

- Change-point analysis aims to find and to locate the time of change of the parameter of interest. The simplest setting is to detect a single change in the mean. Consider the model

  $$X_t = \mu_t + \epsilon_t, \quad t = 1, \ldots, n,$$

  and interested whether the mean remains the same. That is

  $$H_0 : \mu_1 = \mu_2 = \ldots = \mu_n$$

  against

  $$H_1 : \mu_1 = \ldots = \mu_k \neq \mu_{k+1} = \ldots = \mu_n$$

  for some k.
- Be aware that k is unknown, hence needs to be estimated.

Change-point estimation based on LSE

- Simple LSE estimator is given by

$$\hat{k}^L = \operatorname*{argmin}_{1\le k\le n-1} \left(\sum_{t=1}^{k} (X_t - \bar{X}_k)^2 + \sum_{t=k+1}^{n} (X_t - \bar{X}_k^*)^2 \right),$$

where

$$\bar{X}_k = \frac{1}{k}\sum_{t=1}^{k} X_t, \quad \bar{X}_k^* = \frac{1}{n-k}\sum_{t=k+1}^{n} X_t.$$

- This is equivalent to

$$\hat{k}^L = \operatorname*{argmax}_{1\le k\le n-1} \sqrt{\frac{k(n-k)}{n}} \left| \frac{1}{k}\sum_{t=1}^{k} X_t - \frac{1}{n-k}\sum_{t=k+1}^{n} X_t \right|$$

Change-point estimation by CUSUM

- A celebrated CUSUM is given by

$$|\chi(k)| = \frac{1}{\sqrt{n}} \left| \sum_{t=1}^{k} X_t - \frac{k}{n} \sum_{t=1}^{n} X_t \right| = \frac{1}{\sqrt{n}} \left| \sum_{t=1}^{k} \left(X_t - \overline{X} \right) \right|,$$

  where $\overline{X} = n^{-1} \sum_{t=1}^{n} X_j$.
- CUSUM change-point estimator is given by

$$\hat{k} = \underset{1 \le k \le n}{\operatorname{argmax}} |\chi(k)|.$$

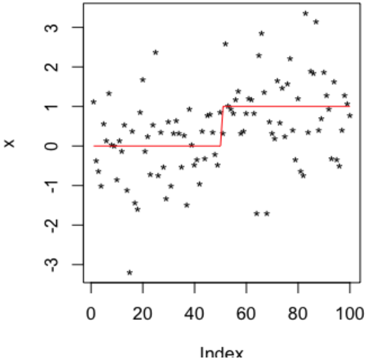
- This can be considered as a variation of LSE estimator since

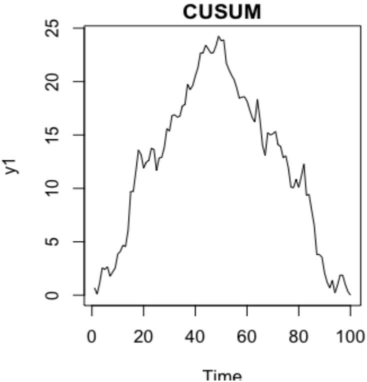
$$\hat{k}^L = \underset{1 \le k \le n-1}{\operatorname{argmax}} |\tilde{\chi}(k)| := \underset{1 \le k \le n-1}{\operatorname{argmax}} \sqrt{\frac{n}{k(n-k)}} |\chi(k)|,$$

  where $|\tilde{\chi}(k)|$ is called adjusted CUSUM in the literature.

CUSUM

- CUSUM became the state-of-the-art due to well-peakedness.





FCLT for CUSUM

- Also, extensively studied and well understood theoretically. For example, under the null hypothesis of no change-point, functional central limit theorem holds under quite general conditions
$$\max_{1 \le k \le n} |\chi(k)| \xrightarrow{d} \sigma \sup_{0 \le u \le 1} |B^0(u)|,$$
where $B^0(t)$ is a standard Brownian bridge and σ^2 is a long-run variance.
- Similarly for LSE estimator,
$$\max_{1 \le k \le n} |\tilde{\chi}(k)| \xrightarrow{d} \sigma \sup_{0 \le u \le 1} \frac{|B^0(u)|}{\sqrt{u(1-u)}}$$

Multiple change points for means

Four methods are widely used

- Information criteria by Yao (1988). For example,

$$\underset{R}{\mathrm{argmin}} \frac{n}{2} \log \hat{\sigma}_R^2 + R \log n,$$

  where $\hat{\sigma}_R^2$ is the MLE of common variance σ^2 assuming R number of breaks under IID Gaussianity.
- Binary segmentation method by Vostrikova (1981), and bootstrap version called wild binary segmentation by Fryzlewicz (2014).
- MOSUM (moving sum) method by (Antoch et al. (2000), Eichinger and Kirch (2018)).
- fused lasso (Rojas ad Wahlerg (2014)).

Multiple change points: Binary Segmentation

- Most widely used method.
- Top-down method. Graphically,

- However, multiple change points may offset each other. See figure in the next slide.

Multiple change-points; Wild Binary Segmentation

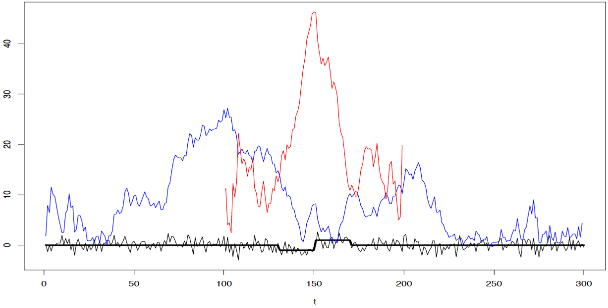


Figure 1: True function f_t, $t = 1, \ldots T = 300$ (thick black), observed X_t (thin black), $|\tilde{X}^b_{1,300}|$ plotted for $b = 1, \ldots 299$ (blue), and $|\tilde{X}^b_{101,200}|$ plotted for $b = 101, \ldots, 199$ (red).

- Localization by random splitting can solve this problem. (Fryzlewicz (2014))

Multiple change points: WBS

function WILDBINSEG(s, e, ζ_T)
    if $e - s < 1$ **then**
        STOP
    else
        $\mathcal{M}_{s,e}$:= set of those indices m for which $[s_m, e_m] \in F_T^M$ is such that $[s_m, e_m] \subseteq [s, e]$
        (Optional: augment $\mathcal{M}_{s,e} := \mathcal{M}_{s,e} \cup \{0\}$, where $[s_0, e_0] = [s, e]$)
        $(m_0, b_0) := \arg\max_{m \in \mathcal{M}_{s,e}, b \in \{s_m, \ldots, e_m - 1\}} |\tilde{X}^b_{s_m, e_m}|$
        if $|\tilde{X}^{b_0}_{s_{m_0}, e_{m_0}}| > \zeta_T$ **then**
            add b_0 to the set of estimated change-points
            WILDBINSEG(s, b_0, ζ_T)
            WILDBINSEG($b_0 + 1$, e, ζ_T)
        else
            STOP
        end if
    end if
end function

Multiple change points: MOSUM

- Localized version of CUSUM statistic with window size G. Eichinger and Kirch (2018) proposed

$$T_n(G) = \max_{G \le k \le n-G} \frac{1}{\sigma_{k,n}\sqrt{2G}} \left| \sum_{t=k+1}^{k+G} X_t - \sum_{t=k-G+1}^{k} X_t \right|$$

- Once G is given, threshold is determined from the extreme value theory (again we are handling max!).

$$a(n/G)T_n(G) - b(n/G) \xrightarrow{d} \Gamma,$$

where Γ is a Gumbel extreme value distribution with
$a(x) = \sqrt{2\log x}$
$b(x) = 2\log x + .5\log\log x + \log(3/2) - .5\log\pi$.

Multiple change points: MOSUM

- Then the local maxima over a threshold will give the number of change points and locations. This does not suffer from size problem as binary segmentation, but bandwidth selection remains an issue.

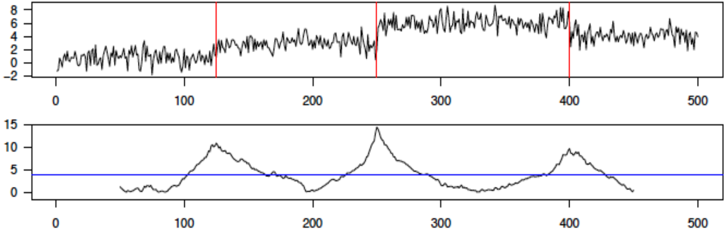


Figure 3.1: Time series with i.i.d. errors and three change points marked by the vertical lines (upper panel) as well as $\hat{\sigma}_{k,n}^{-1}|T_{k,n}(G)|$ (lower panel), where the horizontal line gives the asymptotic critical value at the 5% level.

Multiple change points: fused lasso

- Multiple change-point estimation can be formulated as

$$\underset{\mu_1,\ldots,\mu_n}{\operatorname{argmin}} \frac{1}{n} \sum_{t=1}^{n} (X_t - \mu_t)^2 \quad \text{subject to} \quad \sum_{t=1}^{n-1} |\mu_{t+1} - \mu_t| \le s.$$

- Reformulate this in a regression setting

$$\underset{\gamma,\beta}{\operatorname{argmin}} \|\mathbf{X} - 1_n \mu_1 - \mathbf{W}\beta\|^2 \quad \text{subject to } \|\beta\|_1 \le s,$$

where

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}, \quad \mathbf{W}_{n\times(n-1)} = \begin{pmatrix} 0 & \ldots & 0 \\ 1 & \ldots & 0 \\ \vdots & \vdots & \vdots \\ 1 & \ldots & 1 \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_2 \\ \vdots \\ \beta_n \end{pmatrix},$$

where $\beta_t = \mu_t - \mu_{t-1}$ is the jump size.

Multiple change points: fused lasso

- Since the minimum in γ is $\mathbf{1}'(\mathbf{X} - \mathbf{W}\beta)/n$, this is further reduced to

$$\underset{\beta}{\operatorname{argmin}} \|\overline{\mathbf{X}} - \overline{\mathbf{W}}\beta\|^2 \quad \text{subject to } \|\beta\|_1 \leq s,$$

  where $\overline{\mathbf{X}}$ and $\overline{\mathbf{W}}$ are obtained by centering each column of $\mathbf{X}$ and $\mathbf{W}$.
- However, this formulation is known to fail to recover the so-called stair-case scenario (Rojas ad Wahlerg (2014)). This could be fixed with proper refinement (Son and Lim (2019), Qian and Jia (2016)).
- Optimal tuning parameter may be selected from information criteria.

Change point (for mean) in HDTS

- We are now interested in mean change in HDTS setting. Let $X_1, \ldots, X_n$ be a sequence of p-dimensional vectors. Interested in whether mean vector remains the same.
- As in HD mean testing, two types of statistics can be considered by utilizing univariate CUSUM statistic in each coordinate.
- Jirak (2015) proposed max-type of test statistic

$$T = \max_{1 \le i \le p} \max_{1 \le k \le n} \frac{|\chi_i(k)|}{\widehat{\sigma}_i},$$

for suitable long-run variance estimator $\widehat{\sigma}_i$.

Max-type: Jirak (2015)

- Then, under both temporal and spatial (over the dimension) assumptions for $\{X_t\}$, T converges to Gumbel distribution

$$\lim_{n\to\infty} P(T \le u_p(e^{-x})) = \exp(-e^{-x}), \quad p \sim n^{\xi}, \xi > 0,$$

where $u_p(e^{-x}) = x/e_p + f_p$ with $e_p = 2\sqrt{2\log(2p)}$, $f_p = e_p/4$.

- Long-run variance estimator $\widehat{\sigma}_i^2$ is the weighted average of before/after (possible) change point

$$\widehat{\sigma}_i^2 = \hat{\tau}(\widehat{\sigma}_i^L)^2 + (1-\hat{\tau})(\widehat{\sigma}_i^R)^2,$$

where

$$\widehat{\tau} = \underset{\delta \le k \le n-\delta}{\operatorname{argmax}} |\chi_i(k)|$$

and $(\widehat{\sigma}_i^L)^2$, $(\widehat{\sigma}_i^R)^2$ are Bartlett long-run variance estimator using only data before/after change point $\widehat{\tau}$, respectively.

Max-type: Jirak (2015)

- Again, block multiplier bootstrap can improve finite sample performance.
- That is, calculate

$$T^* = \max_{1 \le i \le p} \max_{1 \le k \le n} \frac{|\chi_i^*(k)|}{\widehat{\sigma}_i^*}$$

by using bootstrap sample $\{\epsilon_j \widehat{X}_t\}$, where $\widehat{X}_t$ is mean-adjusted series and $\{\epsilon_j\}$ is IID $\mathcal{N}(0,1)$ multiplier for t-th observation falling into j-th block.

- Then, under mild conditions and conditional on data X,

$$\sup_{x \in \mathbb{R}} |P_{|X}(T^* \le x) - P(T \le x)| = O_P(n^{-C}), \quad C > 0.$$

SS-type: Horváth and Hušková (2012)

- Based on quasi-likelihood argument, and studied in the panel setting, that is, independent across dimensions.
- The test statistic is given by

$$V = \max_{1\leq k\leq n} \frac{1}{\sqrt{p}} \sum_{i=1}^{p} \left| \frac{|\chi_i(k)|^2}{\widehat{\sigma}_i^2} - \frac{k(n-k)}{n^2} \right|$$

- If $p/n^2 \to 0$, under H_0,

$$V \to \sup_{0\leq x\leq 1} |\Gamma(x)|,$$

where $\Gamma(x)$ is a Gaussian process with zero mean and $\mathbb{E}\Gamma(x)\Gamma(y) = 2x^2(1-y^2), 0 \leq x \leq y \leq 1$ and under H_1,

$$V \xrightarrow{p} +\infty.$$

SS-type: Enikeeva and Harchaoui (2014)

- Interested in detecting changes only in a small subset of components but under HD with i.i.d. errors setting.
- Enikeeva and Harchaoui (2014) proposed the combination of linear and scan statistics given by

$$T_{lin} = \max_{1 \le k < n} \frac{1}{\sqrt{2p}} \sum_{i=1}^{p} \left\{ |\tilde{\chi}_i(k)|^2 - 1 \right\},$$

$$T_{scan} = \max_{1 \le k < n} \max_{1 \le m \le p} \frac{1}{\sqrt{2m} T_m} \sum_{i=1}^{m} \left\{ |\tilde{\chi}_{(i)}(k)|^2 - 1 \right\},$$

where $|\tilde{\chi}_i(k)|$ is the adjusted CUSUM test statistic and its order statistics $|\tilde{\chi}_{(1)}(k)| \ge |\tilde{\chi}_{(2)}(k)| \ge \ldots \ge |\tilde{\chi}_{(m)}(k)|$.

SS-type; Enikeeva and Harchaoui (2014)

- Hence, reject the null hypothesis of no change if either

$$T_{lin} > H \quad \text{or} \quad T_{scan} > 1.$$

- Roughly speaking $|\tilde{\chi}_i(k)|^2$ follows Chi-square distribution with df=1.
- H and T_m are approximate quantiles of Chi-square distribution for faster computation in high-dimension.
- Scan statistic is to improve detection for sparse changes (change over a small number of dimensions).

Sparse projection: Wang and Samworth (2018)

- Dimension reduction: want to find univariate sequence that attenuate changes in mean by linear combination of data.
- The main idea is to apply CUSUM first due to well peakedness of CUSUM, then consider linear combination of them. Denote CUSUM transformed data by $T_{p\times n} = \mathcal{T}(\mathbf{X})$ for $p \times n$ data vector $\mathbf{X}$.
- Now, do dimension reduction by considering linear transformation of X. That is, consider dimension reduction using PCA (singular value decomposition). Interested in finding $\nu \in \mathbb{R}^p$ such that

$$\underset{\nu}{\operatorname{argmax}} \|\nu' T\|_2, \quad \nu'\nu = 1.$$

- Since there are at most m changes, ν has some sparse constraints. For constant mean changes, the optimal vector ν is in fact a function of $\mu_t - \mu_{t-1}$. Under this sparsity constraint, it is known that the solution is the leading left singular vector of T, but non-convex so computationally NP-hard.

Sparse projection: Wang and Samworth (2018)

- Wang and Samworth (2018) proposed convex relaxations,

$$\hat{M} \in \underset{M \in S_1 (\text{or} S_2)}{\text{argmax}} \; (\langle M, T \rangle - \lambda \|M\|_1),$$

  where $S_1 = \{M \in \mathbb{R}^{p \times (n-1)} : \|M\|_* \leq 1\}$ and $S_2 = \{M \in \mathbb{R}^{p \times (n-1)} : \|M\|_2 \leq 1\}$. Then the optimal projection is given by

$$\hat{\nu} = \text{ leading left singular vector of } \hat{M}$$

- Now we find change-point from reduced vector $\bar{T} = \hat{\nu}' T$.

$$\hat{k} = \underset{t}{\text{argmax}} |\bar{T}_t|$$

- For known σ and $\lambda = 2\sigma\sqrt{\log(p \log n)}$,

$$P\left(\frac{1}{n}|\hat{k} - k| \leq C \frac{\sigma^2 \log \log n}{n}\right) \geq 1 - \frac{7}{\sqrt{\log(n/2)}}$$

Sparse projection: Wang and Samworth (2018)

- For multiple change-points, they suggested to use WBS scheme by Fryzlewicz (2014).

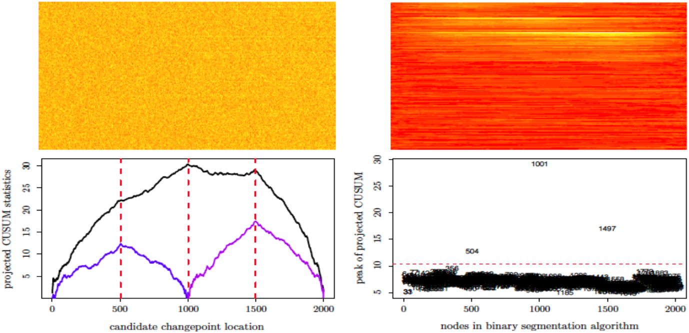


Figure 1: Example of `inspect` algorithm in action. Top-left: visualisation of the data matrix. Top-right: its CUSUM transformation. Bottom-left: overlay of the projected CUSUM statistics for the three changepoints detected. Bottom-right: visualisation of thresholding; the three detected changepoints are above the threshold (dotted red line) whereas the remaining numbers are the test statistics obtained if we run the wild binary segmentation to completion without applying a termination criterion.

Group fused lasso for HDTS

- Can naturally be extended to multidimensional case using group fused lasso (Bleakley and Vert (2011)).
- Consider

$$\min_{\mu_1,\ldots,\mu_n} \frac{1}{2}\sum_{t=1}^{n} \|X_t - \mu_t\|_2^2 + \lambda \sum_{t=1}^{n-1} \frac{\|\mu_{t+1} - \mu_t\|_2}{d_t},$$

where the weight

$$d_t = \sqrt{\frac{n}{t(n-t)}}$$

controls boundary effect and stair-case issue.

Group fused lasso for HDTS

- This is reformulated as the group lasso problem:

$$\underset{\beta}{\mathrm{argmin}} \frac{1}{2} \|\overline{X} - \overline{W}\beta\|_2^2 + \lambda \sum_{t=1}^{n-1} \|\beta_t\|_2,$$

where $\overline{X}$ and $\overline{W}$ are the centered matrix of

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}, \quad W = \begin{pmatrix} 0 & \dots & 0 \\ d_1 & \dots & 0 \\ \vdots & \vdots & \vdots \\ d_1 & \dots & d_{n-1} \end{pmatrix}_{n \times (n-1)} \otimes I_p$$

and

$$\beta_{(n-1)p \times 1} = \begin{pmatrix} \beta_2 \\ \vdots \\ \beta_n \end{pmatrix}, \quad \beta_t = \frac{\mu_{t+1} - \mu_t}{d_t}$$

Group fused lasso for HDTS

- Some theoretical results for single change point are provided in Bleakley and Vert (2011).
- For the selection of λ, they proposed two-stage strategy. Try to select λ to over-estimate the number of change-points, say $k_1, \ldots, k_m$. Then, for each $n \leq m$, find a set of change points $\tilde{k}_1, \ldots, \tilde{k}_n$ that minimizes SSE. Then the final estimate is the one with the smallest information criteria or from a scree plot of SSE.
- In practice, works quite well even for high dimension. Open questions include theoretical analysis for HDTS, justification of weights d_t, etc.

(Auto)covariance change in HDTS

- Growing interest recently. Real life applications include functional connectivity on the brain, association on genes, dynamic networks etc.
- Aue et al. (2009): Based on CUSUM statistic for vectorized $X_t X_t'$.
- Factor model based method: Try to detect changes in loading matrix or (low dimensional) factors. Barigozzi et al. (2018).
- Spectral domain test: Preuss et al. (2015).

References

- Aue, A., Hormann, S., Horvath, L and Reimherr, M. (2009). Break detection in the covariance structure of multivariate time series models. The Annals of Statistics, 37(6B), 4046–4087.
- Antoch, J., Hušková, M., and Jaruškovà, D. (2000). Change point detection. 5th ERS IASC Summer School.
- Barigozzi, M, Cho, H. and Fryzlewicz, P. (2018). Simultaneous multiple change-point and factor analysis for high-dimensional time series, Preprint.
- Bleakley, K. and Vert, J. P. (2011). The group fused Lasso for multiple change-point detection. Technical Report. Technical report HAL-00602121.
- Eichinger, B. and Kirch, C. (2018). A MOSUM procedure for the estimation of multiple random change points. Bernoulli, 526–564.
- Fryzlewicz, P. (2014). Wild Binary Segmentation for multiple change-point detection. Annals of Statistics 42 2243–2281.
- Horváth and Hušková (2012). Change-point detection in panel data. Journal of Time Series Analysis 33 631–648.
- Jirak, M. (2015). Uniform change point tests in high dimension. Annals of Statistics 43 2451–2483.
- Qian, J., Jia, J. (2016). On stepwise pattern recovery of the Fused Lasso. Comput. Statist. Data Anal. 94, 221–237.
- Preuss P., Puchstein, R. and Dette, H. (2015). Detection of multiple structural breaks in multivariate time series, JASA, 654-668.
- Rojas, C.R., Wahlberg, B. (2014). On change point detection using the Fused Lasso method arXiv preprint arXiv:1401.5408.
- Son, W. and Lim, J. (2019). Modified path algorithm of fused Lasso signal approximator for consistent recovery of change points. Journal of Statistical Planning and Inference, 223–238.
- Vostrikova, L. J. (1981). Detecting disorder in multidimensional random processes. Soviet Doklady Mathematics 24 55–59.
- Wang, T. and Samworth, R. J. (2018). High-dimensional changepoint estimation via sparse projection. JRSS-B, 57–83.
- Yao, Y.-C. (1988). Estimating the number of change-points via Schwarz criterion. Statist. Probab. Lett. 6 (3), 181–189.