Optimal multiple change-point detection and Localization

Nicolas Verzelen

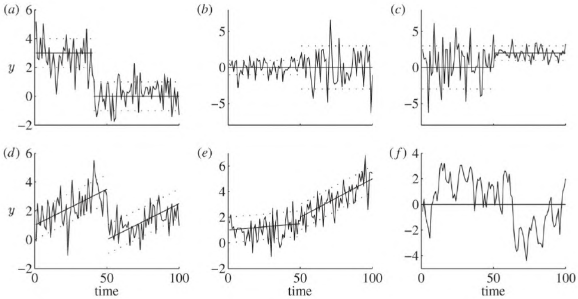
Joint works with A. Carpentier, M. Fromont, M. Lerasle, E. Pilliat, and P. Reynaud-Bouret

https://arxiv.org/abs/2010.11470
https://arxiv.org/abs/2011.07818

MMS Luminy - December 15th

Offline Change-point Analysis

General problem of detecting changes in distribution of a time series



Beaulieu et al.('12)

Old Problem [Wald, 1945] but still vivid.

See [Niu et al., 2016] and [Truong et al., 2020] for recent surveys.

(Sub)-Gaussian univariate mean change-point Model

Data : Time series $\mathbf{Y} \in \mathbb{R}^n$

$$Y_i = \theta_i + \epsilon_i, \qquad \text{where } \epsilon_i \overset{ind.}{\sim} \mathcal{SG}(1) \ ,$$

where we assume that $\boldsymbol{\theta} \in \mathbb{R}^n$ is piece-wise constant.

We leave aside possible time dependencies

Data : Time series $\mathbf{Y} \in \mathbb{R}^n$

$$Y_i = \theta_i + \epsilon_i, \qquad \text{where } \epsilon_i \overset{ind.}{\sim} \mathcal{SG}(1) \ ,$$

where we assume that $\boldsymbol{\theta} \in \mathbb{R}^n$ is piece-wise constant.

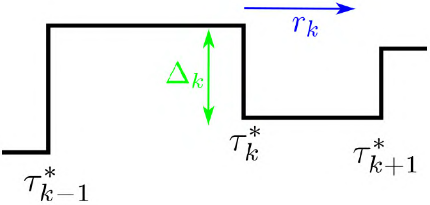
We leave aside possible time dependencies

Notation : change-point vector $\boldsymbol{\tau}^*$
$1 < \tau_1^* < \ldots < \tau_K^* \le n$
s.t. θ is constant over $[\tau_k^*, \tau_{k+1}^*)$.

Height $\Delta_k = \theta_{\tau_k^*} - \theta_{\tau_k^*-1}$

Radius $r_k = \frac{(\tau_{k+1}^* - \tau_k^*)(\tau_k^* - \tau_{k-1}^*)}{\tau_{k+1}^* - \tau_{k-1}^*}$

$$\asymp (\tau_{k+1}^* - \tau_k^*) \wedge (\tau_k^* - \tau_{k-1}^*).$$



(Sub)-Gaussian univariate mean change-point Model

Data : Time series $\mathbf{Y} \in \mathbb{R}^n$

$$Y_i = \theta_i + \epsilon_i, \qquad \text{where } \epsilon_i \overset{ind.}{\sim} \mathcal{SG}(1) \ ,$$

where we assume that $\boldsymbol{\theta} \in \mathbb{R}^n$ is piece-wise constant.

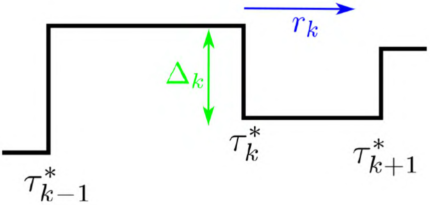
We leave aside possible time dependencies

Notation : change-point vector $\boldsymbol{\tau}^*$
$1 < \tau_1^* < \ldots < \tau_K^* \leq n$
s.t. θ is constant over $[\tau_k^*, \tau_{k+1}^*)$.

Height $\Delta_k = \theta_{\tau_k^*} - \theta_{\tau_k^*-1}$

Radius $r_k = \frac{(\tau_{k+1}^* - \tau_k^*)(\tau_k^* - \tau_{k-1}^*)}{\tau_{k+1}^* - \tau_{k-1}^*}$

$\asymp (\tau_{k+1}^* - \tau_k^*) \wedge (\tau_k^* - \tau_{k-1}^*)$.



Definition of the Energy of τ_k^*

The **Square Energy** of τ_k^* is $\qquad E_k^2 = r_k \Delta_k^2$

l_2 distance between $\boldsymbol{\theta}$ and best approximation by a piece-wise constant vector on $\boldsymbol{\tau}^{(-k)} = (\tau_1^*, \ldots, \tau_{k-1}^*, \tau_{k+1}^*, \ldots)$.

Two mathematical perspectives on change-point Detection

- Denoising/Estimation : Estimating $\boldsymbol{\theta}$ (in l_2 norm).
- Clustering : Recover the change-points $\boldsymbol{\tau}^*$; partition of $[n]$ into segments.

- Denoising/Estimation : Estimating $\boldsymbol{\theta}$ (in l_2 norm).
- Clustering : Recover the change-points $\boldsymbol{\tau}^*$; partition of $[n]$ into segments.

Denoising perspective :
Minimax-Optimal rates (for $K \geq 2$) $$K\left[1 + \log\left(\frac{n}{K}\right)\right]$$

achieved e.g. by penalized least-squares [Birgé and Massart, 2001]

Quadratic computational complexity by dynamic programming.

Change-point detection as a clustering problem

Several lines of literature :

- At Most One Change-point (AMOC) $[K \leq 1]$. **Least-square** estimator detects $\widehat{K} = 1$ if $E_1 \gg \sqrt{\log\log(n)}$ and $|\widehat{\tau}_1 - \tau_1^*| = O(\Delta_1^{-2})$ [Csorgo and Horváth, 1997].

Change-point detection as a clustering problem

Several lines of literature :

- At Most One Change-point (AMOC) $[K \leq 1]$. **Least-square** estimator detects $\widehat{K} = 1$ if $E_1 \gg \sqrt{\log\log(n)}$ and $|\widehat{\tau}_1 - \tau_1^*| = O(\Delta_1^{-2})$ [Csorgo and Horváth, 1997].
- Penalized Least-square Estimator. BIC penalty [Yao and Au, 1989, Wang et al., 2020].

Change-point detection as a clustering problem

Several lines of literature :

- At Most One Change-point (AMOC) $[K \leq 1]$. **Least-square** estimator detects $\widehat{K} = 1$ if $E_1 \gg \sqrt{\log\log(n)}$ and $|\widehat{\tau}_1 - \tau_1^*| = O(\Delta_1^{-2})$ [Csorgo and Horváth, 1997].
- Penalized Least-square Estimator. BIC penalty [Yao and Au, 1989, Wang et al., 2020].
- Greedy or Aggregation methods
  Binary segmentation [Scott and Knott, 1974] = iterative bisection.
  Many recent variants [Fryzlewicz, 2014, Fryzlewicz, 2018, Wang and Samworth, 2018, Wang et al., 2020, Kovács et al., 2020]

Change-point detection as a clustering problem

Several lines of literature :

- At Most One Change-point (AMOC) $[K \leq 1]$. **Least-square** estimator detects $\widehat{K} = 1$ if $E_1 \gg \sqrt{\log\log(n)}$ and $|\widehat{\tau}_1 - \tau_1^*| = O(\Delta_1^{-2})$ [Csorgo and Horváth, 1997].
- Penalized Least-square Estimator. BIC penalty [Yao and Au, 1989, Wang et al., 2020].
- Greedy or Aggregation methods
  Binary segmentation [Scott and Knott, 1974] = iterative bisection.
  Many recent variants [Fryzlewicz, 2014, Fryzlewicz, 2018, Wang and Samworth, 2018, Wang et al., 2020, Kovács et al., 2020]

Focus on computational complexity (e.g. $O(n\log(n))$)

Change-point detection as a clustering problem

Several lines of literature :

- At Most One Change-point (AMOC) [$K \leq 1$]. **Least-square** estimator detects $\widehat{K} = 1$ if $E_1 \gg \sqrt{\log\log(n)}$ and $|\widehat{\tau}_1 - \tau_1^*| = O(\Delta_1^{-2})$ [Csorgo and Horváth, 1997].
- Penalized Least-square Estimator. BIC penalty [Yao and Au, 1989, Wang et al., 2020].
- Greedy or Aggregation methods
Binary segmentation [Scott and Knott, 1974] = iterative bisection.
Many recent variants [Fryzlewicz, 2014, Fryzlewicz, 2018, Wang and Samworth, 2018, Wang et al., 2020, Kovács et al., 2020]

Focus on computational complexity (e.g. $O(n\log(n))$)

Theorem (Typical modern result. *sloppy version* ; [Wang et al., 2020, Fryzlewicz, 2018, Kovács et al., 2020]**)**

If $\min_k E_k^2 \gtrsim \log(n)$*, then whp* $\widehat{K} = K$ *and*

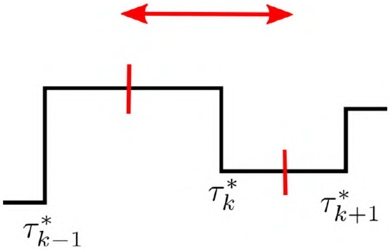
$$d_H(\widehat{\boldsymbol{\tau}}, \boldsymbol{\tau}^*) = \inf_{k=1,\ldots,K} |\widehat{\tau}_k - \tau_k^*| \lesssim \frac{\log(n)}{\min_k \Delta_k^2}$$

Surprisingly, the tightest known results [Frick et al., 2014] are a bit older.

Objectives

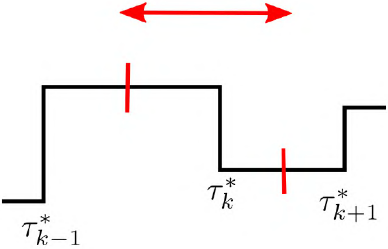
Two sub-problems

Change-Point Detection
= Detecting the existence of the change-point

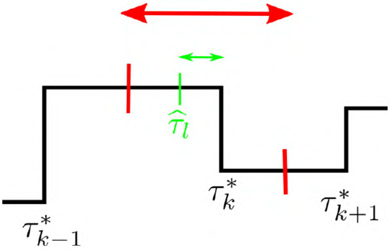


Two sub-problems

Change-Point Detection
= Detecting the existence of the change-point



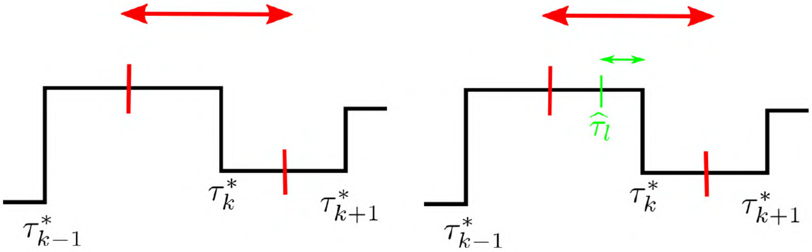
Change-Point Localization
= small estimation error
$d_{H,1}(\widehat{\boldsymbol{\tau}}, \tau^*_k) = \min_l |\widehat{\tau}_l - \tau^*_k|$



Two sub-problems

Change-Point Detection
= Detecting the existence of the change-point

Change-Point Localization
= small estimation error
$d_{H,1}(\widehat{\boldsymbol{\tau}}, \tau_k^*) = \min_l |\widehat{\tau}_l - \tau_k^*|$

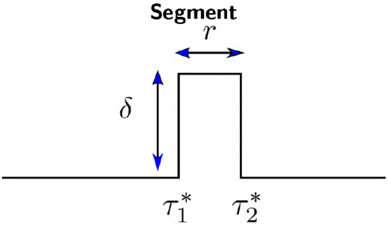


some questions

- What is the energy requirement for detection ?
- How is the transition between detection and localization ?
- Is penalized least-square optimal ? For which penalty ?

Simpler problem : testing $\boldsymbol{\theta} = 0$ versus

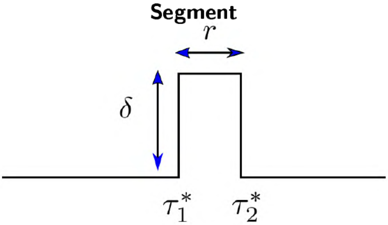
$$\boldsymbol{\theta} \in \Theta[r, \delta] = \left\{ \boldsymbol{\theta} \in \mathbb{R}^n : \exists \tau \in \{n/4, n/4+r, n/4+2r, \ldots, 3n/4\} \text{ such that } \theta_i = \delta \mathbb{1}_{i \in [\tau, \tau+r)} \right\} .$$



Gaussian Change-point Detection

Simpler problem : testing $\boldsymbol{\theta} = 0$ versus

$$\boldsymbol{\theta} \in \Theta[r,\delta] = \left\{\boldsymbol{\theta} \in \mathbb{R}^n : \exists \tau \in \{n/4, n/4+r, n/4+2r, \ldots, 3n/4\} \text{ such that } \theta_i = \delta \mathbb{1}_{i\in[\tau,\tau+r)}\right\} .$$



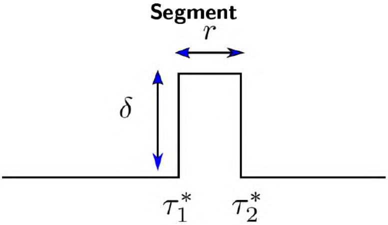
$\lfloor \frac{n}{2r} \rfloor$ possible positions

For each τ, sufficient statistic

$$Z_\tau = r^{-1/2} \sum_{i=\tau}^{\tau+r-1} Y_i \sim \begin{cases} \mathcal{N}(0,1) \\ \mathcal{N}(r^{1/2}\delta, 1) \end{cases}$$

Simpler problem : testing $\boldsymbol{\theta} = 0$ versus

$$\boldsymbol{\theta} \in \Theta[r,\delta] = \left\{\boldsymbol{\theta} \in \mathbb{R}^n : \exists \tau \in \{n/4, n/4+r, n/4+2r, \ldots, 3n/4\} \text{ such that } \theta_i = \delta \mathbb{1}_{i \in [\tau, \tau+r)}\right\} \ .$$



$\lfloor \frac{n}{2r} \rfloor$ possible positions

For each τ, sufficient statistic
$Z_\tau = r^{-1/2} \sum_{i=\tau}^{\tau+r-1} Y_i \sim \begin{cases} \mathcal{N}(0,1) \\ \mathcal{N}(r^{1/2}\delta, 1) \end{cases}$

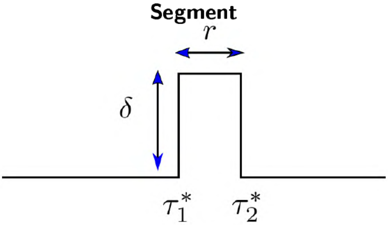
Proposition (Segment Detection ≈ [Arias-Castro et al., 2011])

Fix ξ in $(0,1)$. If $\delta\sqrt{r} \leq \sqrt{2(1-\xi)\log[n/(2r)]}$, then for all tests T
$\mathbb{P}_0[T=1] + \sup_{\boldsymbol{\theta} \in \Theta[\delta,r]} \mathbb{P}_{\boldsymbol{\theta}}[T=0] \geq 1 - c\left(\frac{r}{n}\right)^{c'\xi^2}$.

Gaussian Change-point Detection

Simpler problem : testing $\boldsymbol{\theta} = 0$ versus

$$\boldsymbol{\theta} \in \Theta[r,\delta] = \left\{\boldsymbol{\theta} \in \mathbb{R}^n : \exists \tau \in \{n/4, n/4{+}r, n/4{+}2r, \dots, 3n/4\} \text{ such that } \theta_i = \delta \mathbb{1}_{i \in [\tau, \tau+r)}\right\} .$$



$\lfloor \frac{n}{2r} \rfloor$ possible positions

For each τ, sufficient statistic

$$Z_\tau = r^{-1/2} \sum_{i=\tau}^{\tau+r-1} Y_i \sim \begin{cases} \mathcal{N}(0,1) \\ \mathcal{N}(r^{1/2}\delta, 1) \end{cases}$$

Proposition (Segment Detection ≈ [Arias-Castro et al., 2011])

Fix ξ *in* $(0,1)$. *If* $\delta\sqrt{r} \le \sqrt{2(1-\xi)\log[n/(2r)]}$, *then for all tests* T

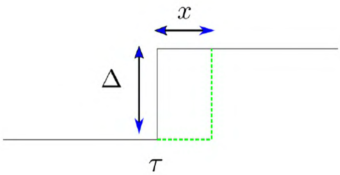
$$\mathbb{P}_0[T=1] + \sup_{\boldsymbol{\theta}\in\Theta[\delta,r]} \mathbb{P}_{\boldsymbol{\theta}}[T=0] \ge 1 - c\left(\frac{r}{n}\right)^{c'\xi^2} .$$

$\kappa > 1$; $q > 0$.

Definition

τ_k^* is a (κ, q)-**high-energy** change-point if $\mathbf{E}_k(\boldsymbol{\theta}) > \kappa\sqrt{2\log\left(\frac{n}{r_k}\right) + q}$.

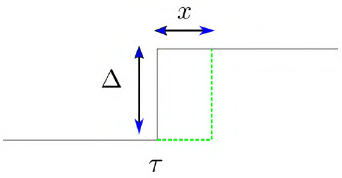
Simplified setting : one change-point ; known means $\boldsymbol{\mu} = (\mu_1, \mu_2)$;
two possible positions for τ^* : τ or $\tau + x$.



Sufficient statistic

$$Z = x^{-1/2} \sum_{i=\tau}^{\tau+x-1} (Y_i - \mu_1) \sim \begin{cases} \mathcal{N}(0,1) \\ \mathcal{N}(x^{1/2}\Delta, 1) \end{cases}$$

Simplified setting : one change-point ; known means $\boldsymbol{\mu} = (\mu_1, \mu_2)$;
two possible positions for τ^* : τ or $\tau + x$.



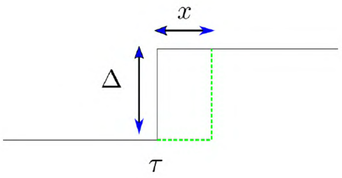
Sufficient statistic
$Z = x^{-1/2} \sum_{i=\tau}^{\tau+x-1} (Y_i - \mu_1) \sim \begin{cases} \mathcal{N}(0,1) \\ \mathcal{N}(x^{1/2}\Delta, 1) \end{cases}$

Lemma (Lower bound for Localization ≈ [Wang and Samworth, 2018])

Write $\Delta = \mu_2 - \mu_1$. *For any* x *in* $[1/2, n/2 - 1 - 2\Delta^{-2})$,

$$\inf_{\widehat{\tau}} \sup_{\tau^* \in \{2,\dots,n\}} \mathbb{P}_{\boldsymbol{\theta}(\tau^*, \boldsymbol{\mu})} \left(|\widehat{\tau} - \tau^*| \geq 2\Delta^{-2} + x \right) \gtrsim e^{-cx\Delta^2} ,$$

Simplified setting : one change-point ; known means $\boldsymbol{\mu} = (\mu_1, \mu_2)$;
two possible positions for τ^* : τ or $\tau + x$.



Sufficient statistic
$Z = x^{-1/2} \sum_{i=\tau}^{\tau+x-1} (Y_i - \mu_1) \sim \begin{cases} \mathcal{N}(0,1) \\ \mathcal{N}(x^{1/2}\Delta, 1) \end{cases}$

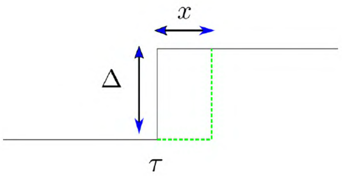
Lemma (Lower bound for Localization ≈ [Wang and Samworth, 2018])

Write $\Delta = \mu_2 - \mu_1$. *For any* x *in* $[1/2, n/2 - 1 - 2\Delta^{-2})$,

$$\inf_{\widehat{\tau}} \sup_{\tau^* \in \{2,\dots,n\}} \mathbb{P}_{\boldsymbol{\theta}(\tau^*, \boldsymbol{\mu})} \left(|\widehat{\tau} - \tau^*| \geq 2\Delta^{-2} + x \right) \gtrsim e^{-cx\Delta^2} ,$$

Small Δ : At best, $|\widehat{\tau} - \tau^*| \asymp \Delta^{-2}$ and has a sub-exponential tail.

Simplified setting : one change-point ; known means $\boldsymbol{\mu} = (\mu_1, \mu_2)$;
two possible positions for τ^* : τ or $\tau + x$.



Sufficient statistic

$Z = x^{-1/2} \sum_{i=\tau}^{\tau+x-1} (Y_i - \mu_1) \sim \begin{cases} \mathcal{N}(0,1) \\ \mathcal{N}(x^{1/2}\Delta, 1) \end{cases}$

Lemma (Lower bound for Localization ≈ [Wang and Samworth, 2018])

Write $\Delta = \mu_2 - \mu_1$. *For any* x *in* $[1/2, n/2 - 1 - 2\Delta^{-2})$,

$$\inf_{\widehat{\tau}} \sup_{\tau^* \in \{2,\ldots,n\}} \mathbb{P}_{\boldsymbol{\theta}(\tau^*, \boldsymbol{\mu})} \left(|\widehat{\tau} - \tau^*| \geq 2\Delta^{-2} + x \right) \gtrsim e^{-cx\Delta^2} ,$$

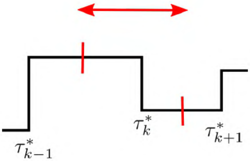
Small Δ : At best, $|\widehat{\tau} - \tau^*| \asymp \Delta^{-2}$ and has a sub-exponential tail.
Large Δ : At best, $\widehat{\tau} = \tau^*$ with proba higher than $1 - c' e^{-c\Delta^2}$.

Desiderata for a suitable change-point procedure

Under an event $\mathcal{A}$ of high (*to be discussed*) probability .

(NoSp). No spurious change-point is detected :

$$\begin{cases} \left|\{\widetilde{\boldsymbol{\tau}}\} \cap \left(\frac{\tau^*_{k-1}+\tau^*_k}{2}, \frac{\tau^*_k+\tau^*_{k+1}}{2}\right]\right| \leq 1 \ , \text{ for all } k \text{ in } \{2,\ldots,K-1\} \ ; \\ \left|\{\widetilde{\boldsymbol{\tau}}\} \cap \left[2, \frac{\tau^*_1+\tau^*_2}{2}\right]\right| \leq 1 \ ; \ \left|\{\widetilde{\boldsymbol{\tau}}\} \cap \left(\frac{\tau^*_{K-1}+\tau^*_K}{2}, n\right]\right| \leq 1 \ . \end{cases}$$



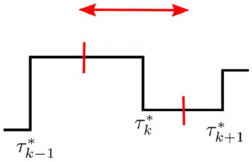
Under an event $\mathcal{A}$ of high (*to be discussed*) probability .

(NoSp). No spurious change-point is detected :

$$\begin{cases} \left|\{\widetilde{\boldsymbol{\tau}}\} \cap \left(\frac{\tau^*_{k-1}+\tau^*_k}{2}, \frac{\tau^*_k+\tau^*_{k+1}}{2}\right]\right| \leq 1 \ \text{, for all } k \text{ in } \{2,\dots,K-1\} \ ; \\ \left|\{\widetilde{\boldsymbol{\tau}}\} \cap \left[2, \frac{\tau^*_1+\tau^*_2}{2}\right]\right| \leq 1 \ ; \ \left|\{\widetilde{\boldsymbol{\tau}}\} \cap \left(\frac{\tau^*_{K-1}+\tau^*_K}{2}, n\right]\right| \leq 1 \ . \end{cases}$$

(Detec[κ, q]). All high-energy change-points are detected.
For all k in $[K]$, if τ^*_k is a (κ, q)-high-energy change-point then

$$d_{H,1}(\widetilde{\boldsymbol{\tau}}, \tau^*_k) \leq \min\left\{\frac{\tau^*_{k+1}-\tau^*_k}{2}, \frac{\tau^*_k-\tau^*_{k-1}}{2}, c\frac{\log\left(1 \vee n\Delta_k^2\right)+q}{\Delta_k^2}\right\} \ .$$



Desiderata for a suitable change-point procedure

Under an event $\mathcal{A}$ of high (*to be discussed*) probability .

(NoSp). No spurious change-point is detected :

$$\begin{cases} \left|\{\widetilde{\boldsymbol{\tau}}\} \cap \left(\frac{\tau^*_{k-1}+\tau^*_k}{2}, \frac{\tau^*_k+\tau^*_{k+1}}{2}\right]\right| \leq 1 \ , \text{ for all } k \text{ in } \{2,\dots,K-1\} \ ; \\ \left|\{\widetilde{\boldsymbol{\tau}}\} \cap \left[2, \frac{\tau^*_1+\tau^*_2}{2}\right]\right| \leq 1 \ ; \ \left|\{\widetilde{\boldsymbol{\tau}}\} \cap \left(\frac{\tau^*_{K-1}+\tau^*_K}{2}, n\right]\right| \leq 1 \ . \end{cases}$$

(Detec[κ, q]). All high-energy change-points are detected.
For all k in $[K]$, if τ^*_k is a (κ, q)-high-energy change-point then

$$d_{H,1}(\widetilde{\boldsymbol{\tau}}, \tau^*_k) \leq \min\left\{\frac{\tau^*_{k+1}-\tau^*_k}{2}, \frac{\tau^*_k-\tau^*_{k-1}}{2}, c\frac{\log\left(1 \vee n\Delta_k^2\right)+q}{\Delta_k^2}\right\} \ .$$

(Loc[κ, q]). High-energy change-points are localized at the optimal rate.
Any (κ, q)-high-energy change-point τ^*_k satisfies

$$\mathbb{P}\left(d_{H,1}(\widetilde{\boldsymbol{\tau}}, \tau^*_k)\mathbb{1}_{\mathcal{A}} \geq cx\Delta_k^{-2}\right) \lesssim e^{-x}, \quad \forall x \geq 1 \ .$$

Penalized least-square estimator

$\boldsymbol{\tau}$= vector of tentative change-points
$\boldsymbol{\Pi}_{\boldsymbol{\tau}}$ = projector onto the space of piece-wise constant vectors with changes at $\boldsymbol{\tau}$

$$\widehat{\tau} = \arg\min_{\tau} \mathrm{Cr}_0(\mathbf{Y}, \boldsymbol{\tau}) = \arg\min_{\boldsymbol{\tau}} \|\mathbf{Y} - \boldsymbol{\Pi}_{\boldsymbol{\tau}}\mathbf{Y}\|^2 + L \ \mathrm{pen}_0(\boldsymbol{\tau}, q) \ ,$$

Multi-scale penalty $\mathrm{pen}_0(\boldsymbol{\tau}, q) = q|\boldsymbol{\tau}| + 2\sum_{k=1}^{|\boldsymbol{\tau}|+1} \log\left(\frac{n}{\tau_k - \tau_{k-1}}\right).$

Penalized least-square estimator

$\boldsymbol{\tau}$= vector of tentative change-points
$\boldsymbol{\Pi}_{\boldsymbol{\tau}}$ = projector onto the space of piece-wise constant vectors with changes at $\boldsymbol{\tau}$

$$\widehat{\tau} = \arg\min_{\tau} \mathrm{Cr}_0(\mathbf{Y}, \boldsymbol{\tau}) = \arg\min_{\boldsymbol{\tau}} \|\mathbf{Y} - \boldsymbol{\Pi}_{\boldsymbol{\tau}}\mathbf{Y}\|^2 + L \ \mathrm{pen}_0(\boldsymbol{\tau}, q) \ ,$$

Multi-scale penalty $\mathrm{pen}_0(\boldsymbol{\tau}, q) = q|\boldsymbol{\tau}| + 2\sum_{k=1}^{|\boldsymbol{\tau}|+1} \log\left(\frac{n}{\tau_k - \tau_{k-1}}\right).$

Remarks :

- Additive Penalty $\rightsquigarrow$ dynamic programming (and its refinements [Killick et al., 2012])
- Over-penalizes **small** segments.
- Highly differs from complexity penalties $\mathrm{pen}_0(\boldsymbol{\tau}, q) = (|\boldsymbol{\tau}| + 1)(1 + \log(n/|\boldsymbol{\tau}|))$.

Connection between CUSUM and Least-square penalty

Definition (CUSUM Statistic)

For $\mathbf{t} = (t_1, t_2, t_3)$, $\mathbf{C}(\mathbf{Y}, \mathbf{t}) = \left[\overline{\mathbf{Y}}_{[t_2, t_3)} - \overline{\mathbf{Y}}_{[t_1, t_2)}\right] \sqrt{\frac{(t_2 - t_1)(t_3 - t_2)}{t_3 - t_1}}$

Definition (CUSUM Statistic)

For $\mathbf{t} = (t_1, t_2, t_3)$, $\mathbf{C}(\mathbf{Y}, \mathbf{t}) = \left[\overline{\mathbf{Y}}_{[t_2,t_3)} - \overline{\mathbf{Y}}_{[t_1,t_2)}\right]\sqrt{\frac{(t_2-t_1)(t_3-t_2)}{t_3-t_1}}$

Lemma (deletion of a change-point)

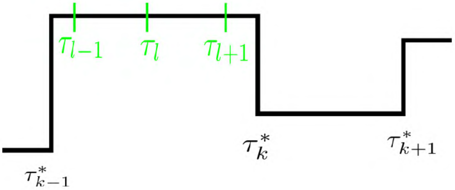
$\boldsymbol{\tau}^{(-l)} = (\tau_1, \ldots, \tau_{l-1}, \tau_{l+1}, \ldots)$

$$\|\mathbf{Y} - \mathbf{\Pi}_{\boldsymbol{\tau}}\mathbf{Y}\|^2 - \|\mathbf{Y} - \mathbf{\Pi}_{\boldsymbol{\tau}^{(-l)}}\mathbf{Y}\|^2 = -\mathbf{C}^2[\mathbf{Y}, (\tau_{l-1}, \tau_l, \tau_{l+1})] \ .$$

$$\begin{aligned}\mathrm{Cr}_0(\mathbf{Y}, \boldsymbol{\tau}) - \mathrm{Cr}_0(\mathbf{Y}, \boldsymbol{\tau}^{(-l)}) \quad = \quad & -\mathbf{C}^2\big(\mathbf{Y}, (\tau_{l-1}, \tau_l, \tau_{l+1})\big) \\ & + L\left[2\log\left(\frac{n(\tau_{l+1} - \tau_{l-1})}{(\tau_{l+1} - \tau_l)(\tau_l - \tau_{l-1})}\right) + q\right] \ .\end{aligned}$$

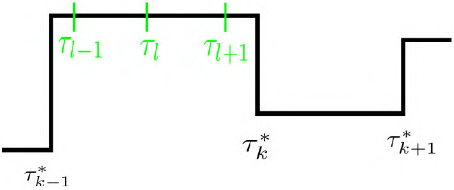
Consider $\boldsymbol{\tau}$ such that
$\boldsymbol{\theta}$ is constant on $[\tau_{l-1}, \tau_{l+1})$

Goal : show that $\boldsymbol{\tau} \neq \widehat{\boldsymbol{\tau}}$?



Consider $\boldsymbol{\tau}$ such that
$\boldsymbol{\theta}$ is constant on $[\tau_{l-1}, \tau_{l+1})$

Goal : show that $\boldsymbol{\tau} \neq \widehat{\boldsymbol{\tau}}$?

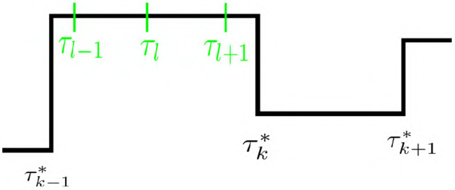


$$\begin{aligned}\mathrm{Cr}_0(\mathbf{Y}, \boldsymbol{\tau}) - \mathrm{Cr}_0(\mathbf{Y}, \boldsymbol{\tau}^{(-l)}) &= -\mathbf{C}^2(\boldsymbol{\epsilon}, (\tau_{l-1}, \tau_l, \tau_{l+1})) \\ &\quad + L\left[2\log\left(\frac{n(\tau_{l+1} - \tau_{l-1})}{(\tau_{l+1} - \tau_l)(\tau_l - \tau_{l-1})}\right) + q\right] .\end{aligned}$$

$\boldsymbol{\tau} \neq \widehat{\boldsymbol{\tau}}$ as long as $\mathbf{C}^2(\boldsymbol{\epsilon}, (\tau_{l-1}, \tau_l, \tau_{l+1}))$ small enough.

Local Optimality and uniform Control of the CUSUM

Consider $\boldsymbol{\tau}$ such that
$\boldsymbol{\theta}$ is constant on $[\tau_{l-1}, \tau_{l+1})$



Goal : show that $\boldsymbol{\tau} \neq \widehat{\boldsymbol{\tau}}$?

$$\mathrm{Cr}_0(\mathbf{Y}, \boldsymbol{\tau}) - \mathrm{Cr}_0(\mathbf{Y}, \boldsymbol{\tau}^{(-l)}) = -\mathbf{C}^2(\boldsymbol{\epsilon}, (\tau_{l-1}, \tau_l, \tau_{l+1})) + L\left[2\log\left(\frac{n(\tau_{l+1} - \tau_{l-1})}{(\tau_{l+1} - \tau_l)(\tau_l - \tau_{l-1})}\right) + q\right] .$$

$\boldsymbol{\tau} \neq \widehat{\boldsymbol{\tau}}$ as long as $\mathbf{C}^2(\boldsymbol{\epsilon}, (\tau_{l-1}, \tau_l, \tau_{l+1}))$ small enough.

Local Optimality $\rightsquigarrow$ Uniform bound for the CUSUM

Lemma (Multi-scale chaining ; in the spirit of [Dumbgen and Spokoiny, 2001])

$$\mathcal{A}_q = \left\{ |\mathbf{C}(\boldsymbol{\epsilon}, \mathbf{t})| \le 2\sqrt{2\log\left(\frac{n(t_3 - t_1)}{(t_3 - t_2)(t_2 - t_1)}\right) + q}, \quad \forall \mathbf{t} = (t_1, t_2, t_3) \right\} .$$

We have $\mathbb{P}[\mathcal{A}_q] \ge 1 - ce^{-c'q}$.

First Analysis of Penalized Least-square

Proposition (V. et al. ('20))

For any L and q large enough, under $\mathcal{A}_q$, the penalized least-square estimator $\widehat{\tau}$ satisfies

(a) **(NoSp)** *No Spurious Jump is detected.*

(b) **(Detec[κ_L, q])** *All (κ_L, q)-high-energy change-points τ_k^* are detected*

$$d_{H,1}(\widehat{\boldsymbol{\tau}}, \tau_k^*) \leq \min\left\{\frac{\tau_{k+1}^* - \tau_k^*}{2}, \frac{\tau_k^* - \tau_{k-1}^*}{2}, \kappa_L \frac{\log\left(n\Delta_k^2\right) + q}{\Delta_k^2}\right\}$$

First Analysis of Penalized Least-square

Proposition (V. et al. ('20))

For any L and q large enough, under $\mathcal{A}_q$, the penalized least-square estimator $\widehat{\tau}$ satisfies

(a) **(NoSp)** *No Spurious Jump is detected.*

(b) **(Detec[κ_L, q])** *All (κ_L, q)-high-energy change-points τ_k^* are detected*

$$d_{H,1}(\widehat{\boldsymbol{\tau}}, \tau_k^*) \leq \min\left\{\frac{\tau_{k+1}^* - \tau_k^*}{2}, \frac{\tau_k^* - \tau_{k-1}^*}{2}, \kappa_L \frac{\log\left(n\Delta_k^2\right) + q}{\Delta_k^2}\right\}$$

[Frick et al., 2014] require $\min_k |\Delta_k|^2 \min_k |\tau_{k+1}^* - \tau_k^*| \gtrsim \log\left(\frac{n}{\min_k |\tau_{k+1}^* - \tau_k^*|}\right)$

Discussion :

- We allow arbitrarily low-energy jumps.
- Local condition for high energy.
- Dependency in q is optimal with respect to the probability $1 - ce^{-c'q}$
- Complexity-based penalties are highly suboptimal.

Localization **(Loc)** by Penalized Least-squares

Proposition (V. et al. ('20))

Fix any L and q large enough. For any (κ_L, q)-high-energy change-point τ_k^*, we have

$$\mathbb{P}\left(d_{H,1}(\widehat{\boldsymbol{\tau}}, \tau_k^*)\mathbb{1}_{\mathcal{A}_q} \geq cx\Delta_k^{-2}\right) \lesssim e^{-x} \quad \forall x \geq 1 \ .$$

Remarks :

- Recovers the optimal subexponential rate of order Δ_k^{-2} for a specific change-point
- **Regional** to **Local** phenomenon :
  Detection= High-Energy **Localization** only depends on Δ_k !

Haussdorff and Wasserstein Loss

If $|\widehat{\boldsymbol{\tau}}| = |\boldsymbol{\tau}^*|$,

$$d_W(\widehat{\boldsymbol{\tau}}, \boldsymbol{\tau}^*) = \sum_{k=1}^{K} |\widehat{\tau}_k - \tau_k^*|$$

$$d_H(\widehat{\boldsymbol{\tau}}, \boldsymbol{\tau}^*) = \max_{k=1}^{K} |\widehat{\tau}_k - \tau_k^*|$$

Corollary

Assuming that all change-points have high-energy, we deduce

$$\mathbb{E}\left[d_W\left(\widehat{\boldsymbol{\tau}}, \boldsymbol{\tau}^*\right)\mathbb{1}_{\mathcal{A}_q}\right] \lesssim \sum_{k=1}^{K}\left(e^{-c''\Delta_k^2} \wedge \frac{1}{\Delta_k^2}\right),$$

$$\mathbb{E}\left[d_H\left(\widehat{\boldsymbol{\tau}}, \boldsymbol{\tau}^*\right)\mathbb{1}_{\mathcal{A}_q}\right] \lesssim \max_{k\in\{1,\dots,K\}}\left(Ke^{-c''\Delta_k^2} \wedge \frac{\log K}{\Delta_k^2}\right).$$

Remark : Haussdorff and Wasserstein rates are minimax optimal.

First Step : Detection

CUSUM Statistics $\mathbf{C}[\mathbf{Y}, \mathbf{t}]$ higher than $\sqrt{2\log\left(\frac{n(t_3-t_1)}{(t_3-t_2)(t_2-t_1)}\right)} + \zeta_{1-\alpha}$

$\rightsquigarrow$ Local test of the null $\{$*constant signal over* $[t_1, t_3)\}$

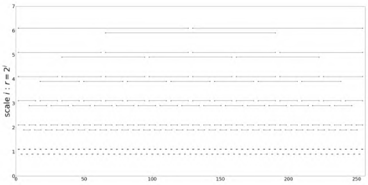
First Step : Detection

CUSUM Statistics $\mathbf{C}[\mathbf{Y}, \mathbf{t}]$ higher than $\sqrt{2\log\left(\frac{n(t_3-t_1)}{(t_3-t_2)(t_2-t_1)}\right)} + \zeta_{1-\alpha}$
$\rightsquigarrow$ Local test of the null $\{$*constant signal over* $[t_1, t_3)\}$

Change-point Detection = Aggregation of Multiple Local Tests

2 Caveats :

- **Too many tests** $n^3/6 \rightsquigarrow$ symmetric intervals + smaller grid $(n \log(n))$
- **Tests** are not always self consistent
  $\rightsquigarrow$ Favoring smaller scales= Bottom-up Approach



CUSUM Statistics $\mathbf{C}[\mathbf{Y},\mathbf{t}]$ higher than $\sqrt{2\log\left(\frac{n(t_3-t_1)}{(t_3-t_2)(t_2-t_1)}\right)}+\zeta_{1-\alpha}$

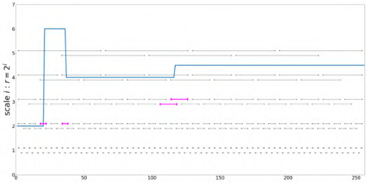
$\rightsquigarrow$ Local test of the null $\{$*constant signal over* $[t_1,t_3)\}$

Change-point Detection = Aggregation of Multiple Local Tests

2 Caveats :

- **Too many tests** $n^3/6$ $\rightsquigarrow$ symmetric intervals + smaller grid $(n\log(n))$
- **Tests** are not always self consistent
  $\rightsquigarrow$ Favoring smaller scales= Bottom-up Approach

Result: $(\hat{\tau}_k)_{k\leq\hat{K}}$
Data: Local test $(T_{l,r})$
$\mathcal{CI}=\varnothing$; $\mathcal{CP}=\varnothing$;
For $r\in$ Scales
For $l\in$ Locations s.t. $T_{l,r}=1$
if $(l-r,l+r)\cap\mathcal{CI}=\varnothing$ **then**
  $\mathcal{CI}\leftarrow\mathcal{CI}\cup(l-r,l+r)$;
  $\mathcal{CP}\leftarrow\mathcal{CP}\cup\{l\}$;
end
return $\mathcal{CP}$



Analysis of the first step

Proposition

With probability higher than $1-\alpha$, $\widehat{\tau}_{ag}$ *satisfies*

(a) **(NoSp)** *No Spurious Jump is detected.*

(b) **(Detec[κ, c_α])** *All* (κ, c_α)*-high-energy change-points* τ_k^* *are detected*

$$d_{H,1}(\widehat{\boldsymbol{\tau}}_{ag}, \tau_k^*) \leq \min\left\{\frac{\tau_{k+1}^* - \tau_k^*}{2}, \frac{\tau_k^* - \tau_{k-1}^*}{2}, \kappa\frac{\log\left(n\Delta_k^2\right) + c_\alpha}{\Delta_k^2}\right\}$$

Similar to the penalized least-square estimator

Analysis of the first step

Proposition

With probability higher than $1-\alpha$, $\widehat{\tau}_{ag}$ *satisfies*

(a) **(NoSp)** *No Spurious Jump is detected.*

(b) **(Detec[κ, c_α])** *All* (κ, c_α)*-high-energy change-points* τ_k^* *are detected*

$$d_{H,1}(\widehat{\boldsymbol{\tau}}_{ag}, \tau_k^*) \leq \min\left\{\frac{\tau_{k+1}^* - \tau_k^*}{2}, \frac{\tau_k^* - \tau_{k-1}^*}{2}, \kappa\frac{\log\left(n\Delta_k^2\right) + c_\alpha}{\Delta_k^2}\right\}$$

Similar to the penalized least-square estimator

... but, $\widehat{\tau}_{ag}$ does not seem to achieve **Loc** at least in worst-case.

Second Step : Localization

For each estimated **change-point** $\widehat{\tau}_k$, we re-estimate the change-point position :

$\rightsquigarrow$ least-square estimator inside a CI $I_{\widehat{\tau}_k}$ of size $\widehat{r}_k$ based on data in a larger interval of size $2\widehat{r}_k - 1$.

$$\widetilde{\tau}_k \in \underset{\tau' \in I_{\widehat{\tau}_k}}{\arg\min} \|\mathbf{Y} - \mathbf{\Pi}_{\tau'} \mathbf{Y}^{(\widehat{\tau}_k, 2\widehat{r}_k - 1)}\|^2 \quad .$$

Second Step : Localization

For each estimated **change-point** $\widehat{\tau}_k$, we re-estimate the change-point position :

⇝ least-square estimator inside a CI $I_{\widehat{\tau}_k}$ of size $\widehat{r}_k$ based on data in a larger interval of size $2\widehat{r}_k - 1$.

$$\widetilde{\tau}_k \in \arg\min_{\tau' \in I_{\widehat{\tau}_k}} \|\mathbf{Y} - \mathbf{\Pi}_{\tau'} \mathbf{Y}^{(\widehat{\tau}_k, 2\widehat{r}_k - 1)}\|^2 \quad .$$

Proposition

The refitted estimator $\widetilde{\boldsymbol{\tau}}$ *satisfies, on an even* $\mathcal{B}_\alpha$, *of probability higher than* $1-\alpha$, **(NoSp)**, **(Detec)**, *and* **(Loc)**.

Remark :

- Similar to penalized least-square estimator.
- Computational complexity $O(n \log(n))$

Summary

Wrap-up :

- Regional to Local phenomenon.
- Low-energy change-points are (almost) unharmful.
- Localization errors behave almost independently.

Summary

Wrap-up :

- Regional to Local phenomenon.
- Low-energy change-points are (almost) unharmful.
- Localization errors behave almost independently.

One versus Multiple change-points.
When $K = 1$, $\log$ conditions are replaced by $\log\log$ conditions.

Summary

Wrap-up :

- Regional to Local phenomenon.
- Low-energy change-points are (almost) unharmful.
- Localization errors behave almost independently.

One versus Multiple change-points.
When $K = 1$, $\log$ conditions are replaced by $\log\log$ conditions.

Possible Extensions/ Open Questions :

- Heavier **tail** distribution, **time dependencies** :
  $\rightsquigarrow$ uniform control of the CUSUM (e.g.[Cho and Kirch, 2019])
- **Exact constant** for detection ?

High-Dimensional Setting

Gaussian Multivariate Change-point Model

$$Y_i = \boldsymbol{\theta}_i + \epsilon_i, \qquad \text{where } \boldsymbol{\theta}_i \in \mathbb{R}^p \text{ and } \epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \mathbf{I}_p).$$

Objective : Detecting **times** $\tau_1^*, \ldots, \tau_K^*$ such that $\boldsymbol{\theta}_{\tau_k^*} \neq \boldsymbol{\theta}_{\tau_k^*-1}$ (with the side information that the difference $\boldsymbol{\theta}_{\tau_k^*} - \boldsymbol{\theta}_{\tau_k^*-1}$ is possibly sparse)

[Wang and Samworth, 2018, Enikeeva and Harchaoui, 2019, Liu et al., 2019]

General analysis of the bottom-up algorithm

We are given :

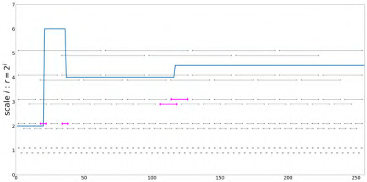
- A grid $\mathcal{G}$ of (l, r) (positions, radius) corresponding to $(l - r, l + r)$.
- A collection of local homogeneity tests $\mathcal{T} = \{T_{l,r}\}$

General analysis of the bottom-up algorithm

We are given :

- A grid $\mathcal{G}$ of (l, r) (positions, radius) corresponding to $(l - r, l + r)$.
- A collection of local homogeneity tests $\mathcal{T} = \{T_{l,r}\}$

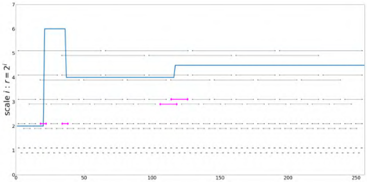
Result: $(\hat{\tau}_k)_{k \leq \hat{K}}$
Data: Local test $(T_{l,r})$
$\mathcal{CI} = \varnothing$; $\mathcal{CP} = \varnothing$;
For $r \in$ Scales
For $l \in$ Locations s.t. $T_{l,r} = 1$
if $(l - r, l + r) \cap \mathcal{CI} = \varnothing$ **then**
  $\mathcal{CI} \leftarrow \mathcal{CI} \cup (l - r, l + r)$;
  $\mathcal{CP} \leftarrow \mathcal{CP} \cup \{l\}$;
end
return $\mathcal{CP}$



We are given :

- A grid $\mathcal{G}$ of (l, r) (positions, radius) corresponding to $(l-r, l+r)$.
- A collection of local homogeneity tests $\mathcal{T} = \{T_{l,r}\}$

Result: $(\hat{\tau}_k)_{k \le \hat{K}}$
Data: Local test $(T_{l,r})$
$\mathcal{CI} = \emptyset$; $\mathcal{CP} = \emptyset$;
For $r \in$ Scales
For $l \in$ Locations s.t. $T_{l,r} = 1$
if $(l-r, l+r) \cap \mathcal{CI} = \emptyset$ **then**
  $\mathcal{CI} \leftarrow \mathcal{CI} \cup (l-r, l+r)$;
  $\mathcal{CP} \leftarrow \mathcal{CP} \cup \{l\}$;
end
return $\mathcal{CP}$



Proposition

If $FWER(\mathcal{T}) \le \alpha$, then $\widehat{\tau}_{ag}$ satisfies $(\mathbf{NoSp})$ with probability higher than $1-\alpha$. All change points $\boldsymbol{\tau}_k^*$ detected by a local test (up to some margin), are detected by $\widehat{\boldsymbol{\tau}}_{ag}$.

Generic Schemes :

- Introducing a sensible notion of energy
- Optimal testing with respect to that energy.

Energy of a Change-Point

$$E_k^2 = r_k \frac{\|\boldsymbol{\theta}_{\tau_k^*} - \boldsymbol{\theta}_{\tau_k^*-1}\|^2}{\sigma^2}$$

Local Homogeneity Tests on $[l-r, l+r)$
1st Simplification : two-sample tests over data in $[l-r, l)$ versus $[l, l+r)$.

2nd Simplification : (possibly-sparse) signal detection test with multivariate CUSUM statistics

$$\mathbf{C}_{l,r} = \left[\overline{\mathbf{Y}}_{[l,l+r)} - \overline{\mathbf{Y}}_{[l-r,l)}\right] \frac{\sqrt{2r}}{\sigma} \sim \mathcal{N}\left[\left(\overline{\boldsymbol{\theta}}_{[l,l+r)} - \overline{\boldsymbol{\theta}}_{[l-r,l)}\right) \frac{\sqrt{2r}}{\sigma}, \mathbf{I}_p\right]$$

Energy and Optimal Tests for sparse high-dimensional data

Energy of a Change-Point

$$E_k^2 = r_k \frac{\|\boldsymbol{\theta}_{\tau_k^*} - \boldsymbol{\theta}_{\tau_k^*-1}\|^2}{\sigma^2}$$

Local Homogeneity Tests on $[l-r, l+r)$
1st Simplification : two-sample tests over data in $[l-r, l)$ versus $[l, l+r)$.

2nd Simplification : (possibly-sparse) signal detection test with multivariate CUSUM statistics

$$\mathbf{C}_{l,r} = \left[\overline{\mathbf{Y}}_{[l,l+r)} - \overline{\mathbf{Y}}_{[l-r,l)}\right] \frac{\sqrt{2r}}{\sigma} \sim \mathcal{N}\left[\left(\overline{\boldsymbol{\theta}}_{[l,l+r)} - \overline{\boldsymbol{\theta}}_{[l-r,l)}\right) \frac{\sqrt{2r}}{\sigma}, \mathbf{I}_p\right]$$

Old toy detection Problem : [Donoho and Jin, 2004, Collier et al., 2015]
$\rightsquigarrow$ Higher-Criticism + χ^2 type statistics (minimax optimal wrt sparsity s and p)

Energy and Optimal Tests for sparse high-dimensional data

Energy of a Change-Point

$$E_k^2 = r_k \frac{\|\boldsymbol{\theta}_{\tau_k^*} - \boldsymbol{\theta}_{\tau_k^*-1}\|^2}{\sigma^2}$$

Local Homogeneity Tests on $[l-r, l+r)$
1st Simplification : two-sample tests over data in $[l-r, l)$ versus $[l, l+r)$.

2nd Simplification : (possibly-sparse) signal detection test with multivariate CUSUM statistics

$$\mathbf{C}_{l,r} = \left[\overline{\mathbf{Y}}_{[l,l+r)} - \overline{\mathbf{Y}}_{[l-r,l)}\right] \frac{\sqrt{2r}}{\sigma} \sim \mathcal{N}\left[\left(\overline{\boldsymbol{\theta}}_{[l,l+r)} - \overline{\boldsymbol{\theta}}_{[l-r,l)}\right) \frac{\sqrt{2r}}{\sigma}, \mathbf{I}_p\right]$$

Old toy detection Problem : [Donoho and Jin, 2004, Collier et al., 2015]
$\rightsquigarrow$ Higher-Criticism + χ^2 type statistics (minimax optimal wrt sparsity s and p)

No Sufficient : $\Omega[n\log(n)]$ tests are considered
$\rightsquigarrow$ one also needs optimal dependencies wrt Types I and II error probabilities :
e.g. variants of HC [Liu et al., 2019] or Pilliat et al.('20).

Optimal Detection

$\delta \in (0,1)$; s_k sparsity of change-point τ_k^*.

High-energy change-point

τ_k^* is a high-energy change-point if $E_k^2 \geq c\psi_{s,p,s_k,\delta}$ where

$$\psi_{s,p,s_k,\delta} = s_k \log\left(1 + \frac{\sqrt{p}}{s_k}\sqrt{\log\left(\frac{n}{r_k\delta}\right)}\right) + \log\left(\frac{n}{r_k\delta}\right) \quad .$$

Optimal Detection

$\delta \in (0,1)$; s_k sparsity of change-point τ_k^*.

High-energy change-point

τ_k^* is a high-energy change-point if $E_k^2 \geq c\psi_{s,p,s_k,\delta}$ where

$$\psi_{s,p,s_k,\delta} = s_k \log\left(1 + \frac{\sqrt{p}}{s_k}\sqrt{\log\left(\frac{n}{r_k\delta}\right)}\right) + \log\left(\frac{n}{r_k\delta}\right) \quad .$$

Theorem (Pilliat et al.('20))

With probability higher than $1-\delta$, $\widehat{\boldsymbol{\tau}}_{ag}$ *achieves* **(NoSp)** *and* **(Detects)** *all* **high-energy** *change-points* τ_k^* *with* $E_k^2 \geq c_+\psi_{s,p,s_k,\delta}$.

$\delta \in (0,1)$; s_k sparsity of change-point τ_k^*.

High-energy change-point

τ_k^* is a high-energy change-point if $E_k^2 \geq c\psi_{s,p,s_k,\delta}$ where

$$\psi_{s,p,s_k,\delta} = s_k \log\left(1 + \frac{\sqrt{p}}{s_k}\sqrt{\log\left(\frac{n}{r_k\delta}\right)}\right) + \log\left(\frac{n}{r_k\delta}\right) .$$

Theorem (Pilliat et al.('20))

With probability higher than $1-\delta$, $\widehat{\boldsymbol{\tau}}_{ag}$ *achieves* **(NoSp)** *and* **(Detects)** *all* **high-energy** *change-points* τ_k^* *with* $E_k^2 \geq c_+\psi_{s,p,s_k,\delta}$.

Conversely, no procedure achieving **(NoSp)** *is able to* **(Detect)** *high-energy change-points (up to a constant)* τ_k^* *with* $E_k^2 \geq c_-\psi_{s,p,s_k,\delta}$

Remark : For $K \leq 1$, see [Liu et al., 2019].

Main Message

Sloppy Conjecture

For general change-points models, optimal detection (almost) amounts to optimal multiple homogeneity testing

Main Message

Sloppy Conjecture

For general change-points models, optimal detection (almost) amounts to optimal multiple homogeneity testing

Open Questions

Localization rates require model-specific techniques.

For (Sparse) High-dimensional change-points, there seem to exist several phase transitions from regional to local (*work in progress*)

References I

Arias-Castro, E., Candès, E. J., and Durand, A. (2011).
Detection of an anomalous cluster in a network.
Ann. Statist., 39(1) :278–304.

Birgé, L. and Massart, P. (2001).
Gaussian model selection.
J. Eur. Math. Soc. (JEMS), 3(3) :203–268.

Cho, H. and Kirch, C. (2019).
Localised pruning for data segmentation based on multiscale change point procedures.
arXiv preprint arXiv :1910.12486.

Collier, O., Comminges, L., and Tsybakov, A. B. (2015).
Minimax estimation of linear and quadratic functionals on sparsity classes.
arXiv preprint arXiv :1502.00665.

Csorgo, M. and Horváth, L. (1997).
Limit theorems in change-point analysis.
John Wiley & Sons Chichester.

Donoho, D. and Jin, J. (2004).
Higher criticism for detecting sparse heterogeneous mixtures.
Ann. Statist., 32(3) :962–994.

References II

Dumbgen, L. and Spokoiny, V. G. (2001).
Multiscale testing of qualitative hypotheses.
Annals of Statistics, pages 124–152.

Enikeeva, F. and Harchaoui, Z. (2019).
High-dimensional change-point detection under sparse alternatives.
Ann. Statist., 47(4) :2051–2079.

Frick, K., Munk, A., and Sieling, H. (2014).
Multiscale change point inference.
Journal of the Royal Statistical Society : Series B (Statistical Methodology), 76(3) :495–580.

Fryzlewicz, P. (2014).
Wild binary segmentation for multiple change-point detection.
The Annals of Statistics, 42(6) :2243–2281.

Fryzlewicz, P. (2018).
Tail-greedy bottom-up data decompositions and fast multiple change-point detection.
The Annals of Statistics, 46(6B) :3390–3421.

Killick, R., Fearnhead, P., and Eckley, I. A. (2012).
Optimal detection of changepoints with a linear computational cost.
Journal of the American Statistical Association, 107(500) :1590–1598.

References III

Kovács, S., Li, H., Bühlmann, P., and Munk, A. (2020).
Seeded binary segmentation : A general methodology for fast and optimal change point detection.
arXiv preprint arXiv :2002.06633.

Liu, H., Gao, C., and Samworth, R. J. (2019).
Minimax rates in sparse, high-dimensional changepoint detection.
arXiv preprint arXiv :1907.10012.

Niu, Y. S., Hao, N., and Zhang, H. (2016).
Multiple change-point detection : A selective overview.
Statistical Science, 31(4) :611–623.

Scott, A. J. and Knott, M. (1974).
A cluster analysis method for grouping means in the analysis of variance.
Biometrics, pages 507–512.

Truong, C., Oudre, L., and Vayatis, N. (2020).
Selective review of offline change point detection methods.
Signal Processing, 167 :107299.

Wald, A. (1945).
Sequential tests of statistical hypotheses.
The annals of mathematical statistics, 16(2) :117–186.

Wang, D., Yu, Y., and Rinaldo, A. (2020).
Univariate mean change point detection : Penalization, CUSUM and optimality.
Electron. J. Stat., 14(1) :1917–1961.

Wang, T. and Samworth, R. J. (2018).
High dimensional change point estimation via sparse projection.
J. R. Stat. Soc. Ser. B. Stat. Methodol., 80(1) :57–83.

Yao, Y.-C. and Au, S. T. (1989).
Least-squares estimation of a step function.
Sankhyā Ser. A, 51(3) :370–381.