**The Institute for**
**Information Transmission Problems**

# Multiscale approach for change point detection

Alexandra Suvorikova, Nazar Buzun, Vladimir Spokoiny

## Algorithm

“ Materia – this is what is changing ”

B. Gita

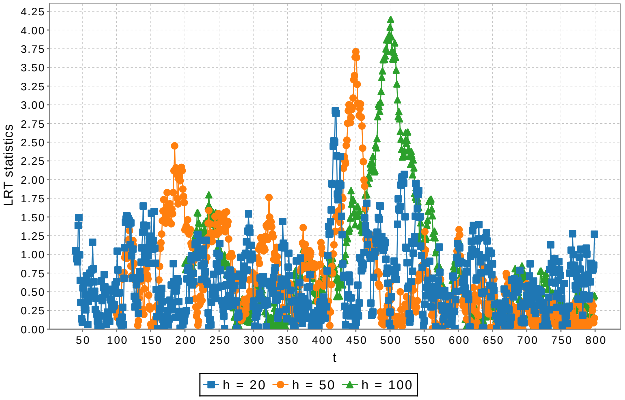
$$YYYYYYYY\,\overbrace{\underbrace{YYYYYYYY}_{L_1(\theta)}\,\underbrace{YYYYYYYY}_{L_2(\theta)}}^{L(\theta)}\,YYYYYYYY$$

Statistic foreach window position $t$:

$$T^2(t) = L_1(t)(\widehat{\theta}_1) + L_2(t)(\widehat{\theta}_2) - L(t)(\widehat{\theta})$$

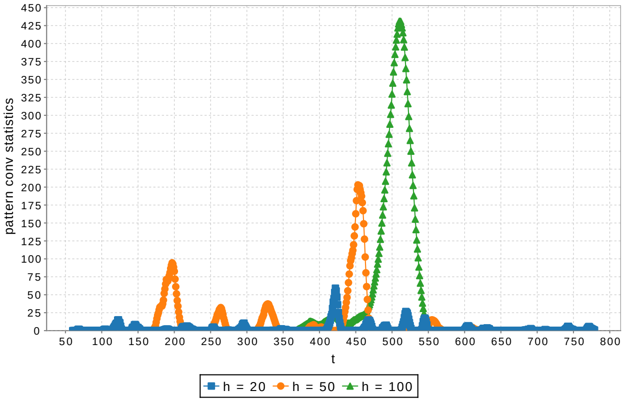
Different window sizes: $h = 20, 50, 100$

$$T(t) \approx \|\xi_{12} + \theta_2 - \theta_1\|$$



## Smoothing with fit linear pattern $P(t)$

$$\mathbb{T}_h(\tau) = \sum_{t=\tau}^{\tau+h} (T(t) - \widehat{b}) * P(t), \quad P(t) = \widehat{a}t$$



$$L(\theta) \approx L(\theta^*) + \xi^T D(\theta - \theta^*) + \frac{1}{2} \|D(\theta - \theta^*)\|^2$$

$$T(t) \approx \|\xi_{12} + D(\theta_2^* - \theta_1^*)\|$$

Critical bound

$$\max_\tau \sum_{t=\tau}^{\tau+h} P(t) \|\xi_{12}(t)\|$$

Independent data $\{Y_i\}_{i=1}^n$

$$\xi_{12}(t) = \sum_{i=t}^{t+h} \xi_i$$

Multiplier bootstrap

$$\xi_{12}^\flat(t) = \sum_{i=t}^{t+h} \xi_i \varepsilon_i^\flat, \quad \varepsilon_i^\flat \sim \mathcal{N}(0,1)$$

## Bootstrap approximation for $\xi_{12}(t)$

1. $\max_\tau \sum_t P_\tau(t) \|\xi_{12}(t)\| \approx \max_\tau \sum_t P_\tau(t) \left\|\widetilde{\xi_{12}}(t)\right\|$
2. $\max_\tau \sum_t P_\tau(t) \left\|\widetilde{\xi_{12}}(t)\right\| \approx \max_\tau \sum_t P_\tau(t) \left\|\xi_{12}^\flat(t)\right\|$

$$\max_{\tau} \sum_{t} P_{\tau}(t) \left\|\xi_{12}(t)\right\| = \max_{\tau} \max_{\gamma} \sum_{t} P_{\tau}(t) \gamma_t^T \xi_{12}(t) = \max_{\eta=(\gamma,\tau)} X_{\eta}$$

$$X_{\eta} = \sum_{t=\tau}^{\tau+h} P(t) \gamma_t^T \sum_{i=1}^{h} \xi_{t+i} = \sum_{i=1}^{2h} X_{\eta}(\xi_{\tau+i})$$

$$X = \sum_{i=1}^{2h} X(i)$$

$$I\!\!P(\max_{\eta} X_{\eta} > z) = I\!\!E I[\max_{\eta} X_{\eta} > z] \approx I\!\!E f_{\triangle}(X)$$

$$|I\!\!E f(X) - I\!\!E f(\widetilde{X})| \le \|\nabla^3 f\|_1 \sum_{i} I\!\!E \left\|X(i)\right\|_{\infty}^3 + I\!\!E \left\|\widetilde{X}(i)\right\|_{\infty}^3 \le \frac{c}{\triangle^3} \frac{1}{\sqrt{h}}$$

$$\left| I\!\!P(\max_{\eta} X_{\eta} > z) - I\!\!P(\max_{\eta} \widetilde{X}_{\eta} > z \pm 2\triangle) \right| \le \frac{c}{\triangle^3} \frac{1}{\sqrt{h}}$$

## Gaussian approximation for $\xi_{12}(t)$

$$\left| I\!P(\max_\eta X_\eta > z) - I\!P(\max_\eta \widetilde{X}_\eta > z \pm 2\triangle) \right| \leq \frac{c}{\triangle^3} \frac{1}{\sqrt{h}}$$

$$\left| I\!P(\max_\eta \widetilde{X}_\eta > z) - I\!P(\max_\eta \widetilde{X}_\eta > z \pm 2\triangle) \right| \leq c_1 \triangle$$

$$\Downarrow$$

$$\left| I\!P(\max_\eta X_\eta > z) - I\!P(\max_\eta \widetilde{X}_\eta > z) \right| \leq \left( \frac{c_2}{h} \right)^{1/8}$$

## Bootstrap approximation for $\xi_{12}(t)$

1. $\max_\tau \sum_t P_\tau(t) \|\xi_{12}(t)\| \approx \max_\tau \sum_t P_\tau(t) \left\|\widetilde{\xi_{12}}(t)\right\|$
2. $\max_\tau \sum_t P_\tau(t) \left\|\widetilde{\xi_{12}}(t)\right\| \approx \max_\tau \sum_t P_\tau(t) \left\|\xi^\flat_{12}(t)\right\|$

## Gaussian comparison $\widetilde{X} \sim \mathcal{N}(0, S)$, $X^\flat \sim \mathcal{N}(0, S^\flat)$

$$|\mathbb{E} f(\widetilde{X}) - \mathbb{E} f(X^\flat)| \leq \|\nabla^2 f\|_1 \left\| S - S^\flat \right\|_{\mathrm{op}} \leq \frac{1}{\triangle^2} \left\| S - S^\flat \right\|_{\mathrm{op}}$$

$$S^{-1/2} S^\flat S^{-1/2} - I$$

$$\widetilde{X}_\eta = \sum_{i=\tau}^{\tau+2h} V_i \widetilde{\xi}_i, \quad S = V \Sigma V^T$$

$$X^\flat_\eta = \sum_{i=\tau}^{\tau+2h} V_i \xi_i \varepsilon^\flat_i, \quad S^\flat = V \,\mathrm{blockDiag}(\xi \xi^T) V^T$$

$$S^{-1/2} S^\flat S^{-1/2} - I = \mathcal{U} \,\mathrm{blockDiag}(\Sigma^{-1/2} \xi \xi^T \Sigma^{-1/2} - I) \mathcal{U}^T = \sum_i \mathcal{U}_i \mathsf{block}_i \mathcal{U}_i^T$$

# Bernstein matrix inequality [1]

$$\forall i: \; \mathbb{E}\psi^2\left(\frac{\|S_i\|}{M}\right) \le 1,$$

$$R = M\psi^{-1}\left(\frac{2}{\delta}\frac{nM^2}{\mathrm{v}^2}\right), \quad \delta \in (0, 2/\psi(1)).$$

$$\mathrm{v}^2 = \left\|\sum_{i=1}^{n} \mathbb{E}S_i^2\right\|_{\mathrm{op}} \le \max_i \left\|\mathcal{U}_i^T \mathcal{U}_i\right\|_{\mathrm{op}} \|\mathsf{block}_i\|_{\mathrm{op}}^2$$

Then for $zR \le (e-1)(1+\delta)\mathrm{v}^2$ and $Z = S_1 + \ldots + S_n$

$$\mathbb{P}\{\|Z\|_{\mathrm{op}} \ge z\} \le 2p\exp\left\{-\frac{z^2}{2(1+\delta)\mathrm{v}^2 + 2Rz/3}\right\}$$

[1] V. Koltchinskii. A remark on low rank matrix recovery and noncommutative Bernstein type inequalities.