High-dimensional change-point regression with structured information

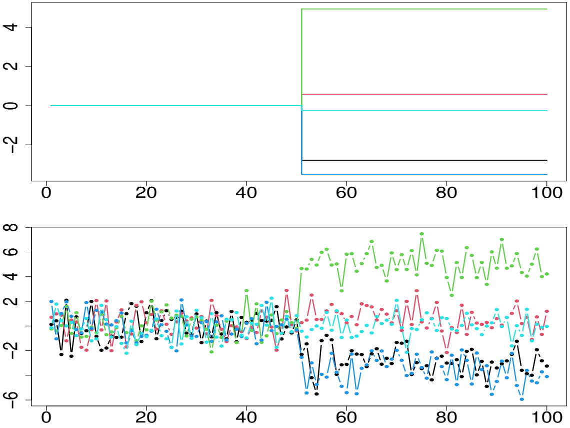
Florian Pein
joint work with Hyeyoung Maeng, Paul Fearnhead and Idris Eckley

Lancaster University

Friday 16th December
StatScale ECR Meeting
Brighton, UK

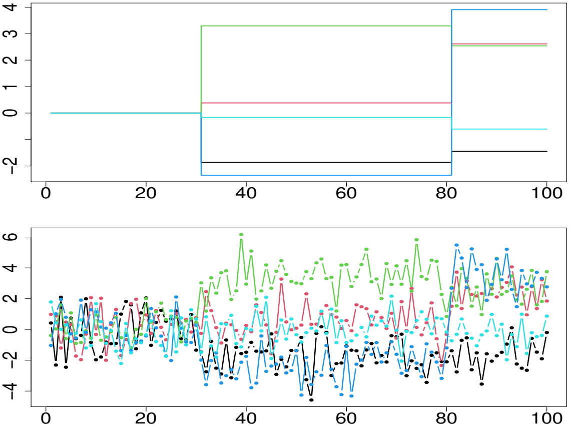
Multivariate change-point regression

Single changepoint at the same time in all coordinates.



p = 5 coordinates, n = 100 observation in each of them.

Multiple changepoints



$K = 2$ changes

Model

- p time series each consisting of n observations

$$x_{j,t},\ j = 1, \ldots, p,\ t = 1, \ldots, n$$

- For simplicity we assume Gaussian mean model with

$$x_{j,t} \sim \mathcal{N}(\mu_{j,t}, 1) \quad \text{independently of each other}$$

- K change-points at $\tau_0 = 0 < \tau_1 < \cdots < \tau_K < n = \tau_{K+1}$
- $\mu_{j,t} \neq \mu_{j,t+1} \iff t = \tau_k, k = 1 \ldots, K,\ j = 1, \ldots, p$

Methods

- LRT for a single change-point at a given location b:

$$T_b = \sum_{j=1}^{p} C_{j,b}^2,$$

where $C_{j,b}$ is the CUSUM statistic

$$C_{j,b} := \sqrt{\frac{b(n-b)}{n}} \left(\frac{1}{b} \sum_{t=1}^{b} x_{j,t} - \frac{1}{n-b} \sum_{b+1}^{n} x_{j,t} \right)$$

- Unknown location: take maximum

$$T = \max_{1 \le b < n} T_b$$

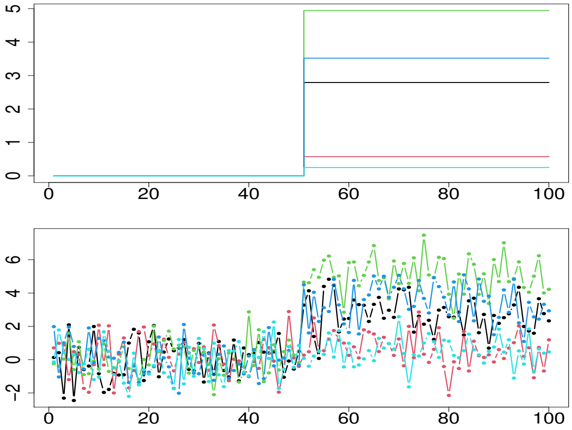
- Compare with threshold q,
  e.g. obtained by Monte-Carlo simulations
- Multiple change-points: BS, MOSUM, seeded BS, WBS

Take a step back

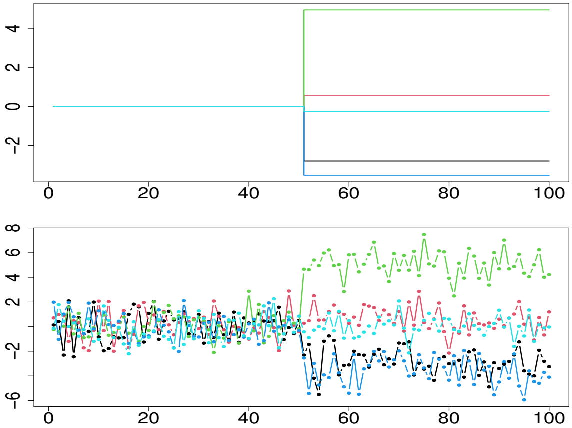
- Why do we assume that changes are at the same location?
- One event causes changes in related processes (same process at different locations, different individuals etc.)
- Often also reasonable that all changes have the same sign, i.e. all changes go upwards or all changes go downwards

Example

Single changepoint with (positive) sign information



How does our data look like?



Motivating Applications

- Copy Number Variation
- Environmental data, e.g. soil moisture
- Economic data e.g. stock prices in the same sectors
- House price rising or dropping together
- ...

Model with sign information

- p time series each consisting of n observations

$$x_{j,t},\ j = 1, \ldots, p,\ t = 1, \ldots, n$$

- For simplicity we assume Gaussian mean model with

$$x_{j,t} \sim \mathcal{N}(\mu_{j,t}, 1) \quad \text{independently of each other}$$

- K change-points at $\tau_0 = 0 < \tau_1 < \cdots < \tau_K < n = \tau_{K+1}$
- $\mu_{j,t} \neq \mu_{j,t+1} \iff t = \tau_k, k = 1 \ldots, K$

- Sign information
  $\text{sign}(\mu_{1,\tau_1+1} - \mu_{1,\tau_1}) = \cdots = \text{sign}(\mu_{p,\tau_K+1} - \mu_{p,\tau_K}) = \pm 1$

Let's use it

Likelihood ratio test:

- If positive sign

$$T_{\text{pos},b} = \sum_{j=1}^{p} \max(C_{j,b}, 0)^2$$

- If negative sign

$$T_{\text{neg},b} = \sum_{j=1}^{p} \min(C_{j,b}, 0)^2$$

- If (unknown) sign

$$T_{\text{sign},b} = \max(T_{\text{pos}}, T_{\text{neg}})$$

- e.g. $T_{\text{pos}} = \max_{1 \leq b < n} T_{\text{pos},b}$

More powerful test

3 Views:

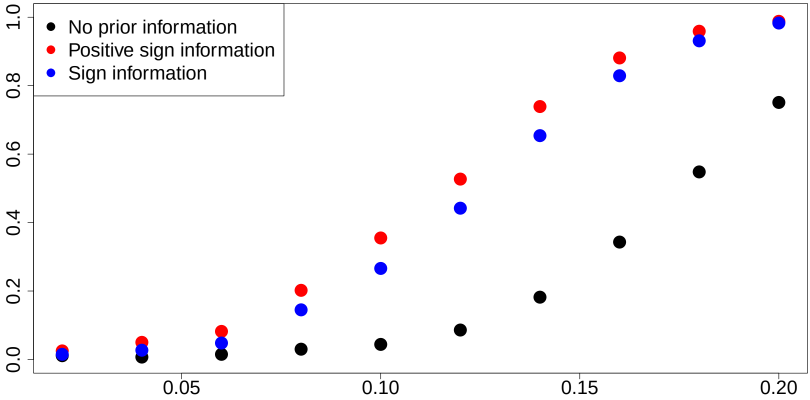
1. Intuition
2. Simulations
3. (Minimax) Theory

1. Intuition

- Consider the null hypothesis of having no change-point
- Distribution of CUSUM statistic is symmetric around 0
- $\Rightarrow$ So $\max(C_{j,b}, 0)$ discards half of the noise contribution
- $\Rightarrow$ Lower critical value
- Positive jump size gets picked up with probability $> 50\%$
- Increase of detection power

2. Simulations

$n = 100, \ p = 50$
$\tilde{\mu} \sim \mathcal{N}(0, 0.2)$
$\mu = C\tilde{\mu}/\sqrt{\|\tilde{\mu}\|p}$
Change-point in the middle



3. Minimax Theory

Let $\theta = \frac{\tau(n-\tau)}{n}\left(\mu_{\cdot,\tau+1} - \mu_{\cdot,\tau+1}\right) \in \mathbb{R}^p$ be the effective jump size.

Theorem 1

Suppose our data comes from the model with positive sign information. Allow p to depend on n and assume $\log\log(n)/p \to 0$*. If* θ_n *such that*

$$\frac{1}{2}\|\theta_n \mathbb{1}_{\theta_n \leq 1}\|_2^2 + \sqrt{\frac{2}{\pi}}\|\theta_n \mathbb{1}_{\theta_n \leq 1}\|_1 + \|\theta_n \mathbb{1}_{\theta_n > 1}\|_2^2 > \sqrt{\frac{5}{2}}\sqrt{p \log\log n},$$

then there exists a sequence of critical values $q_{n,p}$ *such that*

$$\lim_{n\to\infty} \mathbb{P}_0\left(T_{\text{pos}} \leq q_{n,p}\right) = 1,$$

and

$$\lim_{n\to\infty} \mathbb{P}_{\theta_n}\left(T_{\text{pos}} > q_{n,p}\right) = 1.$$

Proof sketch

- Show that
$$X^2 \mathbb{1}_{X>0} - \mathbb{E}\left[X^2 \mathbb{1}_{X>0}\right]$$
is sub-Gamma distributed with variance factor Var $\left[X^2 \mathbb{1}_{X>0}\right]$ and scale parameter 2, where X is Gaussian distributed
- Similar to (Liu et al., 2021, Theorem 5)
- Tests on a dyadic grid
- Union bound
- Use concentration bounds for sub-Gamma r.v.
- $q_{n,p} = \frac{p}{2} + \sqrt{\frac{5}{2}}\sqrt{(1+\delta_1)p \log\log n}$ for a specific constant δ_1
- Require expectation under H_1 to be larger than $q_{n,p}$ + variance term (also of order $q_{n,p}$)

Interpretation

$$\frac{1}{2}\|\theta_n \mathbb{1}_{\theta_n \leq 1}\|_2^2 + \sqrt{\frac{2}{\pi}}\|\theta_n \mathbb{1}_{\theta_n \leq 1}\|_1 + \|\theta_n \mathbb{1}_{\theta_n > 1}\|_2^2 > \sqrt{\frac{5}{2}}\sqrt{p \log\log n}$$

Motivation for the two terms if $\theta_n \leq 1$:

- $\mathbb{E}[\chi^2_{k,\lambda}] = k + \lambda$
- $\frac{1}{2}\|\theta_n \mathbb{1}_{\theta_n \leq 1}\|_2^2$: in more than 50% of the cases we have the contribution of a non-central chi-square distribution
- $\sqrt{\frac{2}{\pi}}\|\theta_n \mathbb{1}_{\theta_n \leq 1}\|_1$: coordinate "selected" with probability

$$\mathbb{P}_\theta(C_{j,b} > 0) = \frac{1}{2} + \frac{\theta}{\sqrt{2\pi}} + \mathcal{O}(\theta^2)$$

Interpretation

$$\frac{1}{2}\|\theta_n \mathbb{1}_{\theta_n \leq 1}\|_2^2 + \sqrt{\frac{2}{\pi}}\|\theta_n \mathbb{1}_{\theta_n \leq 1}\|_1 + \|\theta_n \mathbb{1}_{\theta_n > 1}\|_2^2 > \sqrt{\frac{5}{2}}\sqrt{p \log \log n} \tag{1}$$

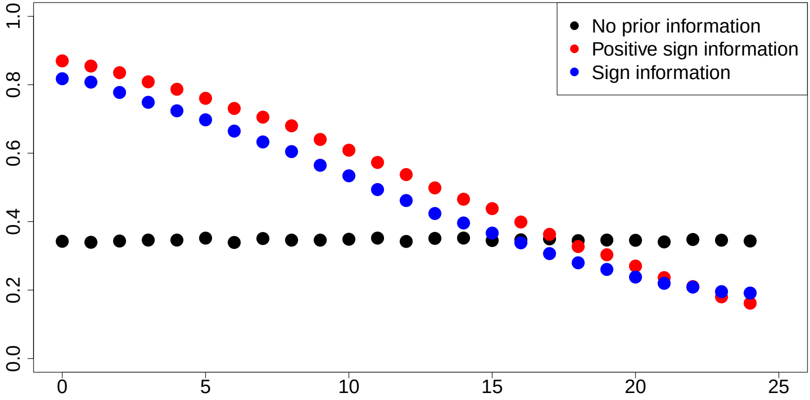
- (Liu et al., 2021, Theorem 5) shows a general minimax statement for

$$\|\theta_n\|_2^2 > 2\sqrt{p \log \log n}.$$

- LHS in (1) $\geq \|\theta_n\|_2^2$, so we win at least on the constant
- Assume that all changes are of the same size θ_1
- Without prior information: $\theta_1 \geq C_1 \left(\frac{\log \log n}{p}\right)^{1/4}$
- Positive sign information: $\theta_1 \geq C_2 \left(\frac{\log \log n}{p}\right)^{1/2}$
- Significant rate gain

Robustness

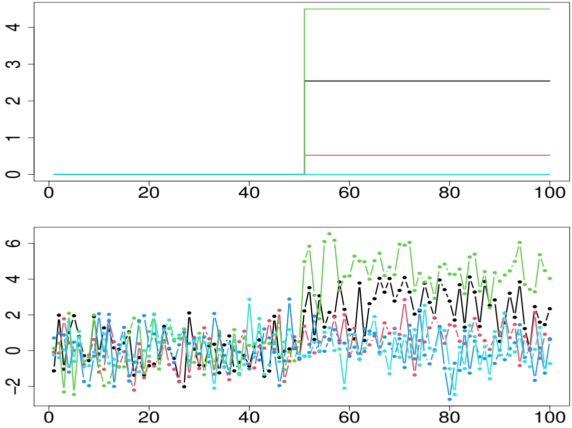
x-Axis: number of coordinates with negative sign



Highly robust: Better until 16/50 coordinates have negative sign

High-dimensional: sparse changes

Only $s = 3$ out of 5 coordinates change; with positive sign.



Model

- p time series each consisting of n observations.

$$x_{j,t},\ j = 1, \ldots, p,\ \ t = 1, \ldots, n.$$

- For simplicity we assume Gaussian mean model with

$$x_{j,t} \sim \mathcal{N}(\mu_{j,t}, 1) \quad \text{independently of each other.}$$

- K change-points at $\tau_0 = 0 < \tau_1 < \cdots < \tau_K < n = \tau_{K+1}$.
- $\mu_{j,t} \neq \mu_{j,t+1} \Longleftrightarrow t = \tau_k, k = 1 \ldots, K$.

- Sparse changes: $\mu_{j,\tau_1+1} - \mu_{j,\tau_1} = 0$ for $p - s$ coordinates
- Sign information
  $\text{sign}(\mu_{i_1,\tau_1+1} - \mu_{i_1,\tau_1}) = \cdots = \text{sign}(\mu_{i_s,\tau_K+1} - \mu_{i_s,\tau_K}) = \pm 1.$

Existing methods

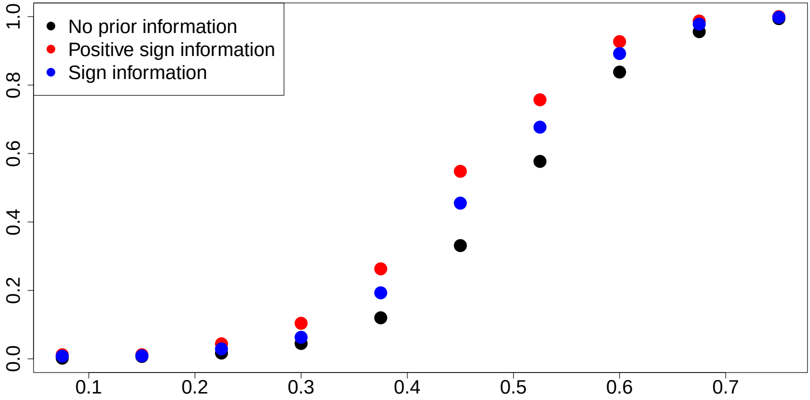
Without prior information:

- Inspect (Wang and Samworth, 2018):
  cusum, then convex relaxation of sparse leading left singular vector to project it to a one-dimensional time series
- Double CUSUM (Cho, 2016):
  cusum, order them, then apply cusum again to extended vector
- Hard thresholding (Liu et al., 2021):
  cusum, apply hard thresholding and take the sum

$\Longrightarrow$ We can use sign information in all of them

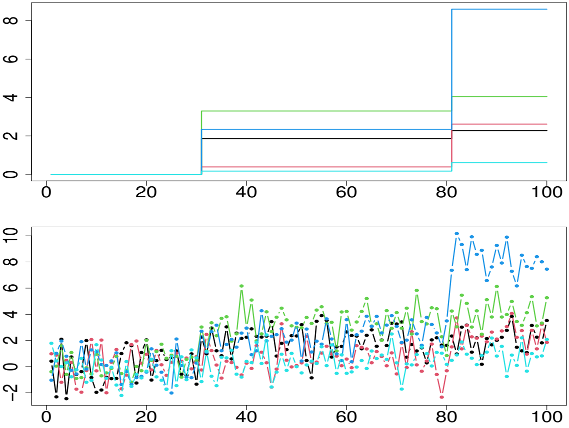
Simulations

- $n = 100$, $p = 50$, only $s = 5$ coordinates have a change
- positive sign
- hard thresholding
- adaptive: multiple thresholds, detect change if one rejects, Monte-Carlo simulations under the null



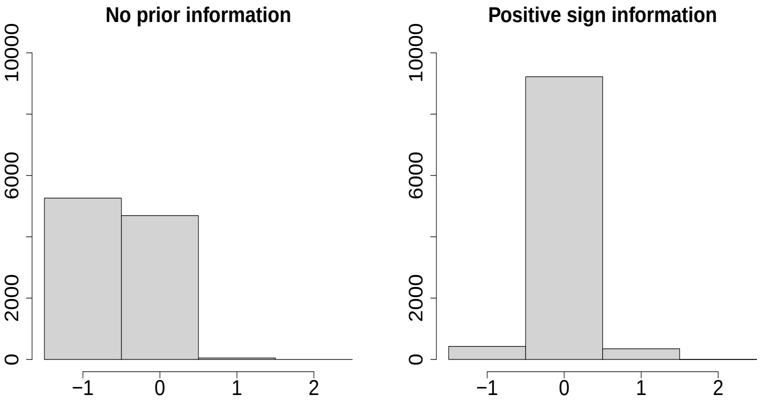
Multiple changes

$K = 2$ changepoints with (positive) sign information (for both of them).



Simulations

- $K = 2$, equidistant, $n = 100$, $p = s = 50$
- Binary segmentation
- $\hat{K} - K$



Conclusion

- We have used what everyone assumes anyway!
- Assuming sign information increases detection power
- Existing extensions to sparse and multiple changes usable
- R-package and paper in 2023

- Extension to non-parametric changes?

Thank you!

Cho, H. (2016). Change-point detection in panel data via double CUSUM statistic. *Electronic Journal of Statistics*, 10(2):2000–2038.

Liu, H., Gao, C., and Samworth, R. J. (2021). Minimax rates in sparse, high-dimensional change point detection. *The Annals of Statistics*, 49(2):1081–1112.

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