# Estimation of simultaneous change-point under sparsity conditions

Farida Enikeeva and Zaid Harchaoui

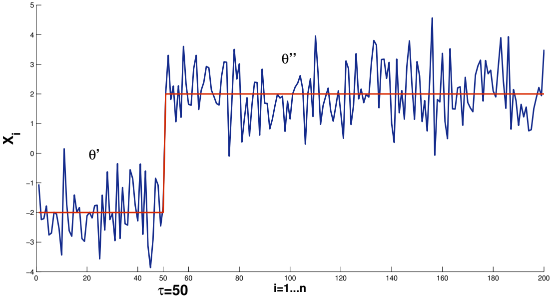
LJK-UJF
INRIA Rhônes-Alpes

7 June 2012

# Change-point problem for Gaussian observations

We observe Gaussian data with the change in mean,

$$X_i = \begin{cases} \theta', & i \le \tau \\ \theta'', & i > \tau \end{cases} + \varepsilon \xi_i, \quad \xi_i \sim \mathcal{N}(0,1), \; i = 1, \dots, n$$




The problem is to estimate the change-point $\tau$

**Historical Overview**

- Chernoff and Zacks (1964) : Bayesian estimate of the mean after the change
- Hinkley (1970) : an MLE of the change-point $\tau$
- Bhattacharya and Brockwell (1976) : limiting behaviour for the likelihood process
- Brodsky and Darkhovksy (1990) : asymptotic distribution of an estimate of $\tau$

Books : Shiryaev (1978), Brodsky and Darkhovsky (1993), Csörgő and Horváth (1997)

# Change-point problem for Gaussian observations

We observe

$$X_i = \begin{cases} \theta', & i \le \tau \\ \theta'', & i > \tau \end{cases} + \varepsilon \xi_i, \quad \xi_i \sim \mathcal{N}(0,1), \ i = 1, \dots, n$$

- Pass to continuous time : $t = i/n, \ t_0 = \tau/n$

$$i = 1, \dots, n \to t \in [a, b], \quad a, b > 0$$

- Size of the jump : $\Delta = |\theta' - \theta''|$

## Asymptotic behavior

$$\widehat{t}_{mle} = \arg \max_{t \in [a,b]} \left\{ \Delta \phi(t) + \frac{\varepsilon}{\sqrt{n}} \frac{B(t)}{\sqrt{t(1-t)}} \right\}$$

where

$$\phi(t) = \sqrt{t(1-t)} \begin{cases} \frac{1-t_0}{1-t}, & t \le t_0 \\ \frac{t_0}{t}, & t > t_0 \end{cases}$$

# Change-point problem for Gaussian observations

**Main results concerning the MLE of $\tau$ :**

- Rate of convergence $n^{-1}$ :

$$\lim_{n\to\infty} \mathbf{E}\left[n(\widehat{t}_{mle} - t_0)\right]^2 = \frac{26\varepsilon^4}{\Delta^4}.$$

- The error of estimation converges to

$$n(\widehat{t}_{mle} - t_0) \xrightarrow{w} \frac{\varepsilon^2}{\Delta^2} \arg\min_{t\in\mathbb{R}} \left\{ -\frac{|t|}{2} + \widetilde{W}(t) \right\}$$

where $\tilde{W}(t) = \begin{cases} W_1(t), & t \geq 0 \\ W_2(-t), & t < 0 \end{cases}$ is a two-sided Wiener process

- Asymptotic distribution of $\widehat{t}_{mle}$ :

$$\mathbf{P}\left\{ n|\widehat{t}_{mle} - t_0| \geq z \right\} \sim \frac{zn\Delta^2}{2\varepsilon^2} e^{-\dfrac{zn\Delta^2}{8\varepsilon^2}}, \quad n \to \infty$$

# Multi-dimensional case

A sequence of Gaussian vectors $X_i \in \mathbb{R}^d$

$$X_i \sim \mathcal{N}(\theta', \varepsilon^2 I_d)\mathbf{1}\{i \le \tau\} + \mathcal{N}(\theta'', \varepsilon^2 I_d)\mathbf{1}\{i > \tau\}$$

with means $\theta', \theta'' \in \mathbb{R}^d$ before and after the change.

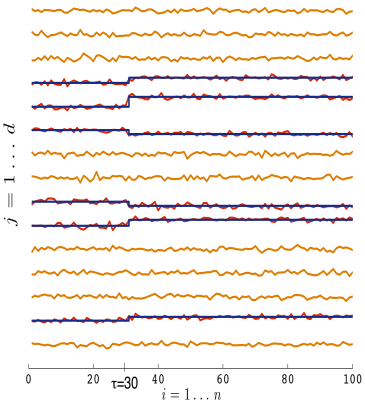
- The norm of the vector of jumps $\Delta\theta = \theta' - \theta'' \in \mathbb{R}^d$ :

$$\|\Delta\theta\|^2 = \sum_{j=1}^{d} (\theta'_j - \theta''_j)^2$$

- The error of estimation converges to

$$n(\widehat{t} - t) \xrightarrow{w} \frac{\varepsilon^2}{\|\Delta\theta\|^2} \arg\min_{t \in \mathbb{R}} \left\{ -\frac{|t|}{2} + \widetilde{W}(t) \right\}, \quad n \to \infty$$

# Multi-channel change-point




**Observations**

$$X_{ij} = \begin{cases} \theta'_j, & i \le \tau \\ \theta''_j, & i > \tau \end{cases} + \xi_{ij}, \; i = 1, \dots, n$$

- $j = 1, \dots, d$ channels
- simultaneous change in the signal mean at some channels
- $\xi_{ij} \sim \mathcal{N}(0, 1)$ i.i.d.
- the number of corrupted channels $J$ is unknown
- $d \to \infty$

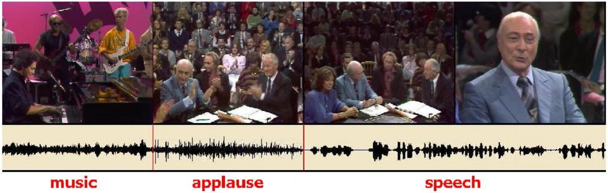
Estimate the common change-point $\tau$

# Motivation

## Applications

- Signal processing :
  Segmentation of audio-visual signals
- Biology :
  analysis of microarrays ; genetic linkage studies
- Cancer research :
  detection of copy-number variation in a gene
  detection of a shared pattern in genomic profiles of patients
- Finance :
  Detection of shifts of volatilities in the stock market

**Segmentation of audio-visual signals**

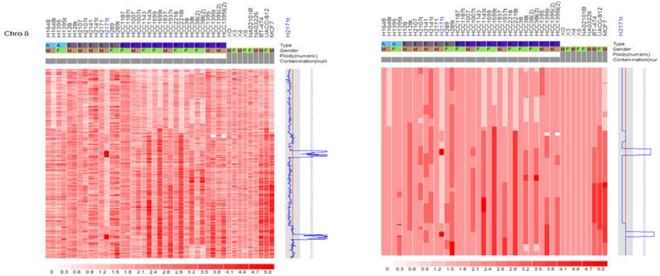



- Take a number of features of the audio-visual signal ($d \to \infty$)
- Some features change simultaneously
  $\Rightarrow$ beginning of a new segment
- The change not necessarily happens for all the features
  $\Rightarrow$ sparsity
- The signal length $n$ is fixed

# Applications

**Detection of copy-number variations (CNV) in a gene**

- CNVs are alterations in a genome resulting in an abnormal number of copies of one or more sections of the DNA :
  A-B-C-D → A-B-C-C-D (a duplication of "C")
  A-B-C-D → A-B-D (a deletion of "C").




- Gene copy number can be elevated in cancer cells
- Identify CNVs that are frequent in population of cancer patients or similar for certain genes.

# Method

## Model

$$X_{ij} = \begin{cases} \theta'_j, & i \le \tau \\ \theta''_j, & i > \tau \end{cases} + \ \xi_{ij}, \ \ i = 1, \dots, n$$

where $\xi_{ij} \sim \mathcal{N}(0,1)$, $j = 1, \dots, d$ and there are $J$ channels with a change .

- The log-likelihood $L(X, \tau, J, m, \theta', \theta'')$
- Estimate $\theta'$, $\theta'' \to L(X, \tau, J, m)$
- Estimate the number of channels with change :

$$J^*(\tau) = \arg \max_{m, 1 \le J \le d} \Big\{ L(X, \tau, J, m) - \mathrm{Pen}(J) \Big\},$$

- Estimate the change-point

$$\widehat{\tau} = \arg \max_{1 \le \tau \le n} L(X, \tau, J^*(\tau))$$

# Log-likelihood

- $J$ is the number of channels with a change
- $m^* = \{j_1, \ldots, j_J\} \in \mathcal{M}$ is a set of indices of corrupted channels
- $\mathcal{M} = \bigcup_{J=1}^{d} \binom{\{1,\ldots,d\}}{J}$ is a set of all possible combinations of indices

## Log-likelihood

$$L(X; \tau, m) = \sum_{j \in m} Z_j^2(\tau), \quad m \in \mathcal{M}, \quad \tau = 1, \ldots, n$$

where for $k = 1, \ldots, n$ the channels are merged,

$$Z_j^2(k) = \frac{1}{k}\Big(\sum_{i=1}^{k} X_{ij}\Big)^2 + \frac{1}{n-k}\Big(\sum_{i=k+1}^{n} X_{ij}\Big)^2 - \frac{1}{n}\Big(\sum_{i=1}^{n} X_{ij}\Big)^2$$

# Estimator

- Ordered statistics $Z^2_{(1)}(k) > Z^2_{(2)}(k) > \cdots > Z^2_{(d)}(k)$ for each $k = 1, \ldots, n$.
- Penalized likelihood for estimating the number of corrupted channels :

$$J^*(\tau) = \arg \max_{1 \le J \le d} \left\{ \sum_{j=1}^{J} Z^2_{(j)}(\tau) - \mathrm{Pen}(J) \right\},$$

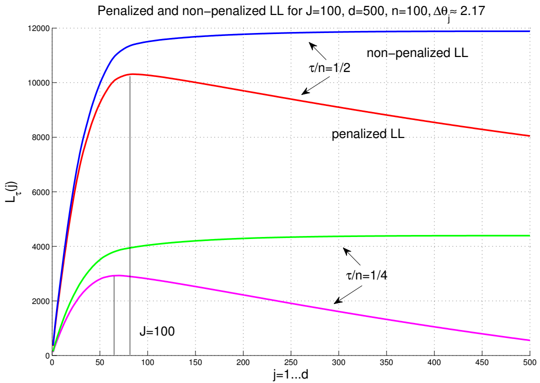
with the penalty chosen according to (Birgé et Massart, 2007)

$$\mathrm{Pen}(J) = (1+\alpha)J + 2J\sqrt{x_J} + 2Jx_J, \quad x_J = \log \frac{de}{J} + 2\frac{\log J}{J}, \quad \alpha > 0.$$

- Estimator of $\tau$ :

$$\widehat{\tau} = \arg \max_{1 \le \tau \le n} \left\{ \sum_{j=1}^{J^*(\tau)} Z^2_{(j)}(\tau) \right\}.$$

# Why penalization ?




- Non-penalized log-likelihood increases
- We subtract some penalty $\mathrm{Pen}(J)$ to penalize the number of "redundant" subsets

# The choice of penalty

$$\mathrm{Pen}(J) = (1+\alpha)J + 2J\sqrt{x_J} + 2Jx_J, \quad \alpha > 0,$$
$$x_J = \log\frac{de}{J} + 2\frac{\log J}{J}$$

- The number of subsets of size $J$ is bounded by

$$\binom{d}{J} \leq \exp\left(J\log\frac{de}{J}\right)$$

$\Rightarrow$ the choice of $x_J$ : $Jx_J = J\log\frac{de}{J} + 2\log J$

- Birgé and Massart condition :

$$\sum_{m\in\mathcal{M}:\ \#m=J} \exp(-Jx_J) < \infty$$

$\Rightarrow$ "putting a prior finite measure on the list of models"
$\Rightarrow$ term $\log d/d$ in $x_J$ to make the sum finite

For each fixed $k = 1, \dots, n$ the merged channels have Gaussian distribution

$$Z_j(k) = \Delta\theta_j \mu(k) + \varepsilon \xi_j(k), \quad \xi_j(k) \sim \mathcal{N}(0,1), \quad j = 1, \dots, d,$$

where

$$\xi_j(k) = \sqrt{\frac{n}{k(n-k)}} \left( \sum_{i=1}^{k} \xi_{ij} - \frac{k}{n} \sum_{i=1}^{n} \xi_{ij} \right)$$

and

$$\xi_j(k) \stackrel{d}{=} \frac{B_j(t)}{\sqrt{t(1-t)}}, \quad t = k/n.$$

# Log-likelihood

The statistic follows non-central $\chi^2$ distribution,

$$\sum_{j\in m} Z_j^2(k) \sim \chi^2(\#m, L(k)), \quad k = 1, \ldots, n$$

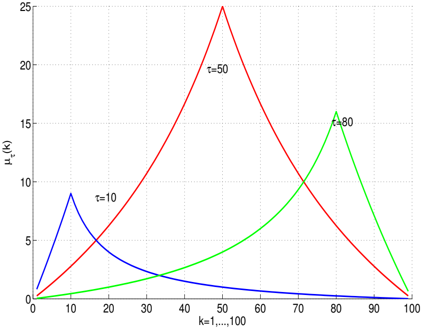
with $\#m$ degrees of freedom (the number of elements in $m$),

- the non-centrality parameter

$$L(k) = \mu^2(k) \sum_{j\in m} (\theta''_j - \theta'_j)^2,$$

- where the function $\mu^2(k)$ attains its maximum at $\tau$,

$$\arg \max_{k=1,\ldots,n} \mu^2(k) = \tau.$$



# Estimation of quadratic functional

- It is impossible to estimate consistently
  - the true subset $m^* = \{j_1, \dots, j_J\}$ of number of corrupted channels
  - the number of channel with a change $J$.
- We estimate the parameter of non-central $\chi^2$ distribution

$$L(\tau) = \mu^2(\tau)\|\Delta\theta\|^2$$

by the estimate (Laurent and Massart, 2000)

$$\widehat{L}(\tau) = \max_{1 \le J \le d} \left\{ \sum_{j=1}^{J} Z^2_{(j)}(\tau) - \mathrm{Pen}(J) \right\}$$

$\Rightarrow$ Consistent estimation of $L(k)$ is possible under certain conditions on $\|\Delta\theta\|^2$ and $J$.

# Consistency of $\widehat{L}(k)$

Estimating $L(k) = \mu^2(k)\|\Delta\theta\|^2$ :

$$Z_j(k) = \Delta\theta_j\mu(k) + \varepsilon\xi_j(k), \quad \xi_j(k) \sim \mathcal{N}(0,1), \quad j = 1, \dots, d,$$

$$\widehat{L}(k) = \max_{1 \le J \le d} \left\{ \sum_{j=1}^{J} Z_{(j)}^2(\tau) - \mathrm{Pen}(J) \right\}$$

**Lemma**

- *For any* $t > 0$

$$\mathbf{P}\left[\widehat{L}(k) - L(k) - 2\mu(k)\langle\Delta\theta, \xi(k)\rangle > t\right] \le Ce^{-t/(2+1/\alpha)}.$$

- *For any* $z > 0$,

$$\mathbf{P}\left[\widehat{L}(k) - L(k) - 2\mu(k)\langle\Delta\theta, \xi(k)\rangle < -z - Q_d(k)\right] \le e^{-z},$$

*where*

$$Q_d(k) = \inf_m \left\{ \frac{3}{4}\mu^2(k)\|\Delta\theta_{\perp m}\|^2 + \mathrm{Pen}(\#m) \right\}.$$

# Corollaries

Assume that for all $k = 1, \dots, n$

$$\lim_{d\to\infty} \frac{Q_d(k)}{\|\Delta\theta\|^2} = 0.$$

Then

- $\frac{\widehat{L}(k)}{\|\Delta\theta\|^2} \xrightarrow{P} \mu^2(k)$, $d \to \infty$
- $\widehat{L}(k)$ is asymptotically normal,

$$\frac{\widehat{L}(k)}{\|\Delta\theta\|^2} - \mu^2(k) \xrightarrow{w} \mathcal{N}\Big(0, 4\mu^2(k)\Big), \quad d \to \infty.$$

- Then $\widehat{\tau}$ is consistent

$$\widehat{\tau} = \arg\max_{k=1,\dots,n} \widehat{L}(k) \xrightarrow{P} \arg\max_{k=1,\dots,n} \mu^2(k) = \tau$$

# Consistency of $\widehat{\tau}$

- Let the number of channels with change be $J = d^\beta$
- Define $\rho = \min(\tau/n, 1 - \tau/n)$
- Assume that

$$\lim_{d\to\infty} \frac{\min\left(J \log \frac{de}{J}, \sqrt{d}\right)}{\|\Delta\theta\|^2 \rho} = 0,$$

**Consistency of $\widehat{\tau}$**

If for some $K > 0$

$$\delta_{n,d} = K \frac{\min\left(J \log \frac{de}{J}, \sqrt{d}\right)}{n \|\Delta\theta\|^2 \rho},$$

the estimator $\widehat{\tau}$ is consistent

$$\mathbf{P}\left[ |\widehat{\tau} - \tau| \geq n\delta_n \right] \to 0, \quad d \to \infty.$$

# Rate

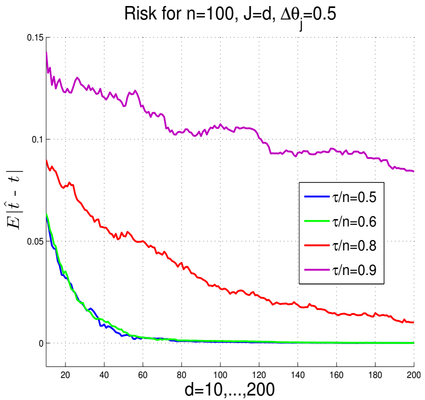
The rate

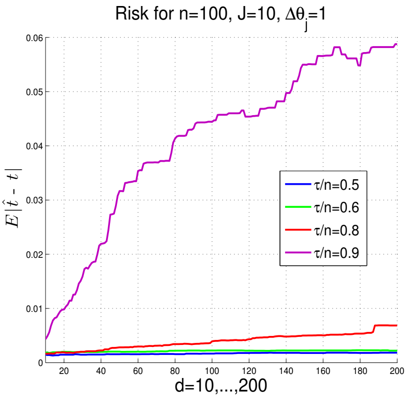
$$\delta_{n,d} \sim \frac{\min\left(J \log \frac{de}{J}, \sqrt{d}\right)}{n\|\Delta\theta\|^2 \rho}$$

depends on

- the norm of jump sizes $\|\Delta\theta\|^2 = \sum_{j=1}^{d} (\theta'_j - \theta''_j)^2$
- the location of the change-point $\tau/n \Rightarrow$ $\rho = \min(\tau/n, 1 - \tau/n)$
- the number of channels $d$ and the number of corrupted channels $J$

# Rate




$$\|\Delta\theta\|^2 = d/4,\ J = d$$
$$\delta_{n,d} \sim \frac{d^{-1/2}}{n\rho}$$

$$\|\Delta\theta\|^2 = J,\ J = d^\beta,\ \beta < 1/2$$
$$\delta_{n,d} \sim \frac{1+(1-\beta)\log d}{n\rho}$$

# Future work

- Why $Q_d(k) \sim \min\left(J \log \frac{de}{J}, \sqrt{d}\right)$ ? $\Rightarrow$ Construct an estimator with different penalties whether $J < \sqrt{d}$ or $J > \sqrt{d}$.
- Asymptotic distribution of $\widehat{\tau}$ :

$$\widehat{t}_{mle} = \arg \max_{t \in [a,b]} \left\{ \|\Delta\theta\| \phi(t) \left( \|\Delta\theta\| \phi(t) + \frac{1}{\sqrt{n}} \frac{B(t)}{\sqrt{t(1-t)}} \right) \right\}$$

where

$$\phi(t) = \sqrt{t(1-t)} \begin{cases} \frac{1-t_0}{1-t}, & t \leq t_0 \\ \frac{t_0}{t}, & t > t_0 \end{cases}$$

- Optimality $\Rightarrow$ lower bounds

$$\lim_{d \to \infty} \inf_{\widehat{t}} \sup_{t \in [a,b]} \sup_{\Delta\theta \in \Theta} \delta_{n,d}^{-1} \mathbf{E}_{\Delta\theta} |\widehat{t} - t_0| \geq C_0.$$