Rank-based multiple change-points detection in multiple time series

Flore Harlé

Sophie Achard, Gipsa-lab
Florent Chatelain, Gipsa-lab
Cédric Gouy-Pailler, CEA-List

UNIVERSITÉ DE GRENOBLE

DE LA RECHERCHE À L'INDUSTRIE

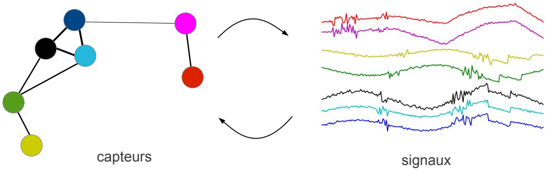
gipsa-lab

Grenoble | images | parole | signal | automatique | laboratoire

Plan of presentation

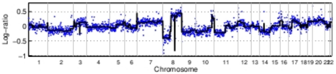
Plan

> Objectives:
- infer the functional links
- detect events



Graph building
Mutual information, correlations, indicators on temporal windows... [XKH11], [CAMGP11], [ASW+06]
→ functional links extracted
→ no temporal information

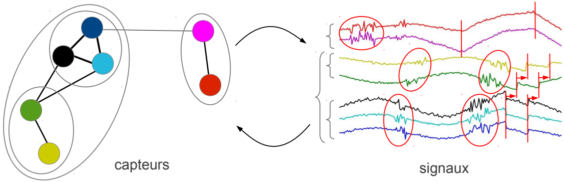
Multivariate analysis
PCA, ICA, dictionnaries, statistical tests... [LYFLLC11]
→ excellent temporal resolution
→ poor spatial information



segmentation of genetic data [BV11]

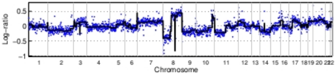
Context

> Objectives:
- infer the functional links
- detect events



Graph building
Mutual information, correlations, indicators on temporal windows... [XKH11], [CAMGP11], [ASW+06]
→ functional links extracted
→ no temporal information

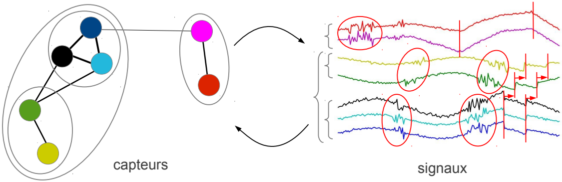
Multivariate analysis
PCA, ICA, dictionnaries, statistical tests... [LYFLLC11]
→ excellent temporal resolution
→ poor spatial information



segmentation of genetic data [BV11]

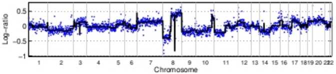
> Objectives:

- infer the functional links
- detect events



Graph building
Mutual information, correlations, indicators on temporal windows...
[XKH11], [CAMGP11], [ASW$^+$06]
→ functional links extracted
→ no temporal information

Multivariate analysis
PCA, ICA, dictionnaries, statistical tests... [LYFLLC11]
→ excellent temporal resolution
→ poor spatial information



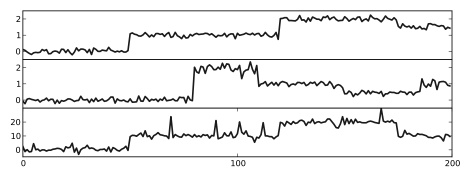
segmentation of genetic data [BV11]

> Our approach :

- detection of events (change-points), using structural priors
- functional relationships inference from a temporal analysis

Problem formulation

Goal: off-line multiple segmentation of multivariate time series



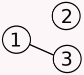
> observations $\boldsymbol{X}$ $(K \times N)$

Approach: statistical test, Bayesian framework, signals dependencies

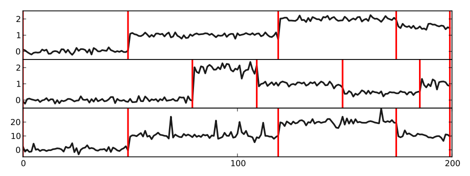
Bayes' theorem:

$$f(\boldsymbol{R}|\boldsymbol{X}) \propto L(\boldsymbol{X}|\boldsymbol{R}) f(\boldsymbol{R})$$

- > posterior $f(\boldsymbol{R}|\boldsymbol{X})$: estimation of $\boldsymbol{R}$
- > likelihood $L(\boldsymbol{X}|\boldsymbol{R})$: based on a robust statistical test
- > prior $f(\boldsymbol{R})$: introduction of the possible links between signals



Goal: off-line multiple segmentation of multivariate time series



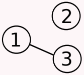
- observations $\boldsymbol{X}$ $(K \times N)$
- indicators $\boldsymbol{R}$ $(K \times N)$

Approach: statistical test, Bayesian framework, signals dependencies

Bayes' theorem:

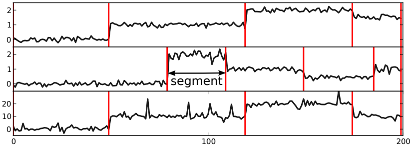
$$f(\boldsymbol{R}|\boldsymbol{X}) \propto L(\boldsymbol{X}|\boldsymbol{R}) f(\boldsymbol{R})$$

- posterior $f(\boldsymbol{R}|\boldsymbol{X})$: estimation of $\boldsymbol{R}$
- likelihood $L(\boldsymbol{X}|\boldsymbol{R})$: based on a robust statistical test
- prior $f(\boldsymbol{R})$: introduction of the possible links between signals



$$f(\boldsymbol{R}|\boldsymbol{X}) \propto L(\boldsymbol{X}|\boldsymbol{R})f(\boldsymbol{R})$$

observations $\boldsymbol{X}$ $(K \times N)$
$x_{j,i}$ mutually independent



indicators $\boldsymbol{R}$ $(K \times N)$

$$r_{j,i} = \begin{cases} 1 & \text{if } x_{j,i} \text{ is a change-point } (H_1), \\ 0 & \text{otherwise } (H_0), \end{cases}$$

for all $1 \leq j \leq K, 1 \leq i \leq N$
by convention $r_1 = r_N = 1$.

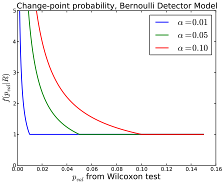
$$\boldsymbol{R} = \begin{pmatrix} 10\ldots010\ldots0\,0\,0\ldots000\ldots01 \\ 10\ldots000\ldots0\,1\,0\ldots010\ldots01 \\ 10\ldots010\ldots0\,0\,0\ldots000\ldots01 \end{pmatrix}$$

$$R_i = \boldsymbol{\epsilon} = (0, 1, 0)^T$$

Plan

Change-point model for $L_*(\boldsymbol{X}|\boldsymbol{R})$

- $\boldsymbol{R}$ defines segments for each signal
- for each $x_{j,i} \in S \rightarrow$ compute p-value $p_{j,i}$, by a statistical test on S
- p-values:
  - under H_0: $x_{j,i}$ not a change-point, $r_{j,i} = 0$, $p_{j,i} \sim \mathcal{U}_{[0,1]}$ [SSC99, SBB01]
  - under H_1: $x_{j,i}$ change-point, $r_{j,i} = 1$, $p_{j,i}$, unknown distribution under H_1: choice of $\mathcal{B}e(\gamma, 1)$ [SBB01] parameter $\gamma \in (0, 1)$:
    - function of an acceptance level α, $f(\alpha|r = 1) = f(\alpha|r = 0)$
    - γ is therefore the unique solution in $(0, 1)$ of $\gamma\alpha^{\gamma-1} = 1, \forall \alpha \in (0, e^{-1})$



- marginal densities of the p-values, taken as random variables:

$$f(p_{j,i}|\boldsymbol{R}) = \begin{cases} \mathbb{1}_{[0,1]}(p_{j,i}) & \text{if } r_{j,i} = 0 \ (H_0,\ x_{j,i} \text{ is not a change-point}), \\ \gamma p_{j,i}^{\gamma-1} \mathbb{1}_{[0,1]}(p_{j,i}) & \text{if } r_{j,i} = 1 \ (H_1,\ x_{j,i} \text{ is a change-point}) \end{cases}$$

- composite marginal likelihood:

$$L_*(\boldsymbol{X}|\boldsymbol{R}) = \prod_{j=1}^{K} \prod_{i=2}^{N-1} \left(\gamma p_{j,i}^{\gamma-1}\right)^{r_{j,i}}$$

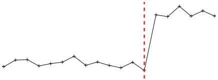
> The Wilcoxon / Mann-Whitney rank-sum test is chosen to compute the p-values [Wil45].

> For two segments Y and Z :

- compute the statistic $U = \min(U_Y, U_Z)$:

$Y = (y_1, \ldots, y_M)$
rank sum R_Y in sorted vector (Y, Z)

$$U_Y = MN\frac{M(M+1)}{2} - R_Y$$



$Z = (z_1, \ldots, z_N)$
rank sum R_Z in sorted vector (Y, Z)

$$U_Z = MN\frac{N(N+1)}{2} - R_Z$$

- tabulated p-values or normal approximation for large samples

$$z = \frac{U - m_U}{\sigma_U}, \quad \text{with} \quad m_U = \frac{MN}{2}, \quad \sigma_U = \sqrt{\frac{MN(M+N+1)}{12}}$$

> High p-values when the differences between the pairs of observations from Y and Z are distributed around 0 (H_0) $\rightarrow$ the data are not assumed to be normally distributed.

- Indicators matrix:

$$\boldsymbol{R} = \begin{pmatrix} \dots & 0\dots1\dots0\dots\boxed{\mathbf{1}}\dots0\dots \\ \dots & 0\dots0\dots0\dots\boxed{\mathbf{1}}\dots0\dots \\ \dots & 1\dots0\dots0\dots\boxed{\mathbf{0}}\dots0\dots \\ \dots & 0\dots1\dots0\dots\boxed{\mathbf{0}}\dots1\dots \end{pmatrix} \qquad R_i = \boldsymbol{\epsilon} = (1,1,0,0)^T$$

- Dependency: if the signal k depends on the signal l, then $R_{k,i} = R_{l,i}$ with a high probability
  P_ϵ is the probability to observe the configuration ϵ in $\boldsymbol{R} \rightarrow \boldsymbol{P} = (P_\epsilon)_{\epsilon\in\mathcal{E}}$
- $(R_i)_{2\leq i\leq N-1}$ are assumed to be *a priori* independent: $f(\boldsymbol{R}) = \prod_{i=2}^{N-1} f(R_i)$
- prior on indicators: $f(\boldsymbol{R},\boldsymbol{P}) \propto f(\boldsymbol{R}|\boldsymbol{P})f(\boldsymbol{P})$, with:
  - $f(\boldsymbol{R}|\boldsymbol{P}) = \prod_{\epsilon\in\mathcal{E}} P_\epsilon^{S_\epsilon(\boldsymbol{R})}$, $S_\epsilon(\boldsymbol{R})$ is the number of times that the configuration ϵ appears in the columns of $\boldsymbol{R}$
  - vague prior for $\boldsymbol{P}$: $\mathcal{D}_L(d)$ [DTD07]
- finally:

$$f(\boldsymbol{R},\boldsymbol{P}) \propto \prod_{\epsilon\in\mathcal{E}} P_\epsilon^{S_\epsilon(\boldsymbol{R})+d_\epsilon-1}$$

Posterior distribution $f(\boldsymbol{R}|\boldsymbol{X})$ and implementation

Posterior distribution

- From the pseudo likelihood and the prior, the posterior expressed as:

$$
\begin{aligned}
f(\boldsymbol{R},\boldsymbol{P}|\boldsymbol{X}) &\propto L_*(\boldsymbol{X}|\boldsymbol{R}) f(\boldsymbol{R}|\boldsymbol{P}) f(\boldsymbol{P}), \\
&\propto \left(\prod_{j=1}^{K}\prod_{i=2}^{N-1}\left(\gamma p_{j,i}^{\gamma-1}\right)^{r_{j,i}}\right)\left(\prod_{\epsilon\in\mathcal{E}} P_\epsilon^{S_\epsilon(\boldsymbol{R})+d_\epsilon-1}\right)
\end{aligned}
$$

- The vector of hyperparameters P_ϵ can be integrated out:

$$
f(\boldsymbol{R}|\boldsymbol{X}) \propto \left(\prod_{j=1}^{K}\prod_{i=2}^{N-1}(\gamma p_{j,i}^{\gamma-1})^{r_{j,i}}\right) \times \frac{\prod_{\epsilon\in\mathcal{E}}\Gamma(S_\epsilon(\boldsymbol{R})+d_\epsilon)}{\Gamma(N+L)}.
$$

Algorithm

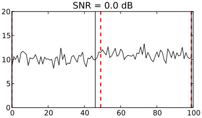
- Estimation of the maximum *a posteriori* of $\boldsymbol{R}$
- Monte Carlo by Markov Chain method
- Gibbs sampling to draw the indicators matrix $\boldsymbol{R}$, column by column

Plan

Choice of the Gibbs sampler

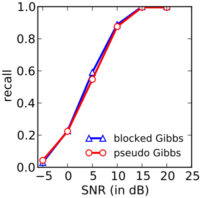
- Single change-point in univariate signal
- Several noise levels:

$$SNR = 10 \log \frac{(\mu_k - \mu_l)^2}{\sigma^2}.$$

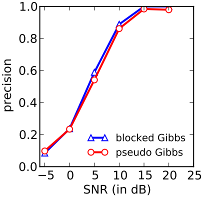


- Detection performances:

$$recall = \frac{TP}{TP + FN} \qquad precision = \frac{TP}{TP + FP}$$



- Gibbs sampler: 2 strategies
  - *blocked Gibbs* sampling
  - conditional probabilities does not form a compatible joint model $\rightarrow$ *pseudo Gibbs* sampling



$$R = (\ldots, 0,\ 1\ , 0, \ldots, 0, \boxed{r_{i-1}, r_i, r_{i+1}}, 0, \ldots, 0,\ 1\ , 0, \ldots)$$

$$P_{val} = (\ldots, \cdot, \boxed{p_{i-}}, \cdot, \ldots, \cdot, \boxed{p_{i-1}, p_i, p_{i+1}}, \cdot, \ldots, \cdot, \boxed{p_{i+}}, \cdot, \ldots)$$

- Single change-point in univariate signal
- Several noise levels:

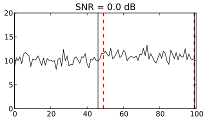
$$SNR = 10 \log \frac{(\mu_k - \mu_l)^2}{\sigma^2}.$$

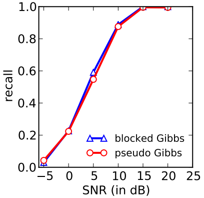
- Detection performances:

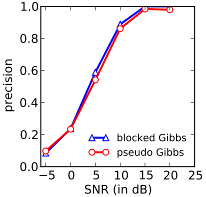
$$recall = \frac{TP}{TP + FN} \qquad precision = \frac{TP}{TP + FP}$$

- Gibbs sampler: 2 strategies
  - *blocked Gibbs* sampling
  - conditional probabilities does not form a compatible joint model $\rightarrow$ *pseudo Gibbs* sampling

$$\begin{aligned} R &= (\dots, 0,\ 1\ , 0, \dots, 0, \boxed{r_i}, 0, \dots, 0,\ 1\ , 0, \dots) \\ P_{val} &= (\dots, \cdot, p_{i-}, \cdot, \dots, \cdot, \boxed{p_i}, \cdot, \dots, \cdot, p_{i+}, \cdot, \dots) \end{aligned}$$





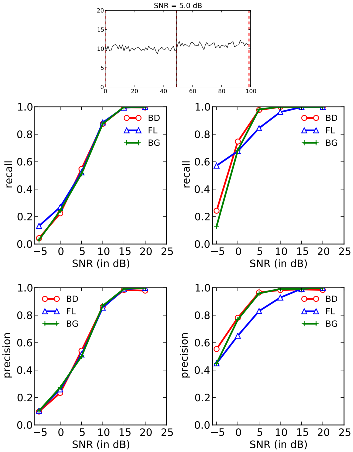


Impact of data distribution

Comparison with the fused lasso [Tib11] ($\lambda = 22.3$) and the Bernoulli Gaussian model

Observations on segment k:

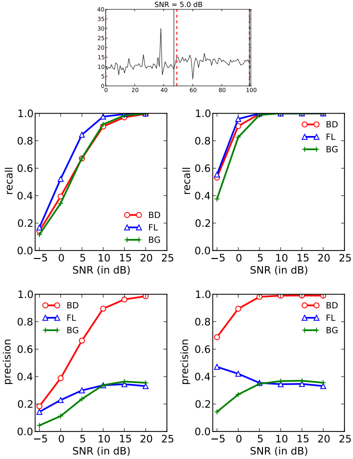
$$\mathcal{N}(\mu_k, \sigma)$$



exact position

position ± 5 points

$$t(\nu, \mu_k, \sqrt{\frac{\nu}{\nu-2}})$$



exact position

position ± 5 points

Control of the false discovery rate (FDR)

Definition:

- multiple hypothesis testing
- maximizing the probability of detecting the true positive by controlling the false positives

$$FDR = E\left[\frac{V}{R \vee 1}\right], V = \text{number of false positives}, R = \text{number of positives}$$

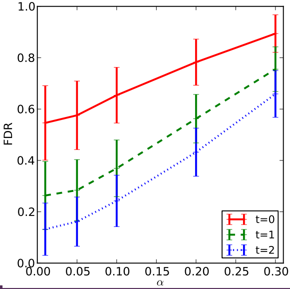
Control of the FDR:

- m tests independent $\rightarrow$ Benjamini-Hochberg procedure [BH95]
- our model:
  - p-values computed by the statistical test highly dependent
  - control by acceptance level α

acceptance level α:
$f(\alpha|r=1) = f(\alpha|r=0)$
$\gamma\alpha^{\gamma-1} = 1$

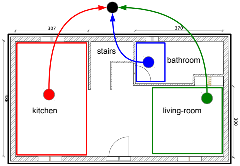
$$L_*(X|R) = \prod_{2=1}^{N-1} \left(\gamma p_i^{\gamma-1}\right)^{r_i}$$

Figure: $FDR = f(\alpha)$, 320 points, 15 change-points, SNR = 5 dB

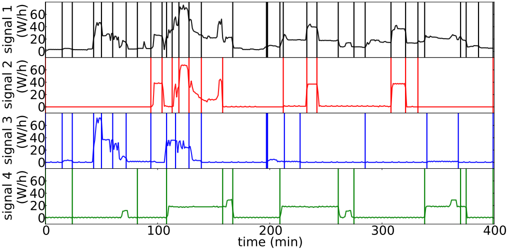


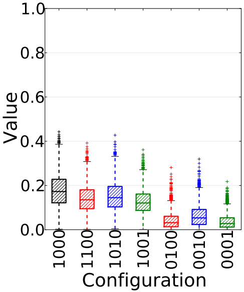
- 4 time series
- Dependencies known $\rightarrow$ noninformative or informative *prior* on $\boldsymbol{P}$

$$\boldsymbol{R} = \begin{pmatrix} \dots & 1\dots1\dots0\dots1\dots1\dots \\ \dots & 0\dots0\dots1\dots1\dots0\dots \\ \dots & 1\dots0\dots0\dots0\dots0\dots \\ \dots & 0\dots1\dots0\dots0\dots1\dots \end{pmatrix}$$



Noninformative prior

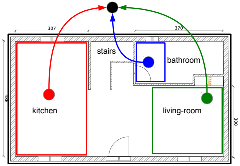




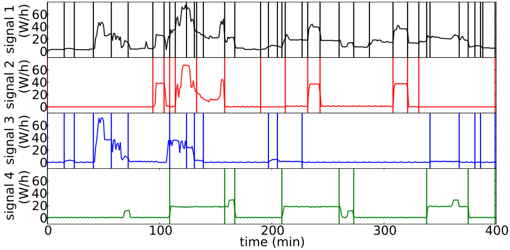
Household electrical power consumption

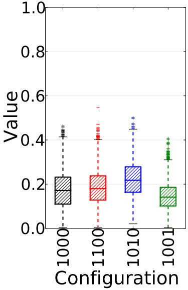
- 4 time series
- Dependencies known → noninformative or informative *prior* on $\boldsymbol{P}$

$$\boldsymbol{R} = \begin{pmatrix} \dots & 1\dots1\dots0\dots1\dots1\dots \\ \dots & 0\dots0\dots0\dots1\dots0\dots \\ \dots & 1\dots0\dots0\dots0\dots0\dots \\ \dots & 0\dots1\dots0\dots0\dots1\dots \end{pmatrix}$$



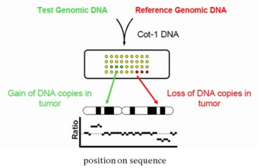
Informative prior



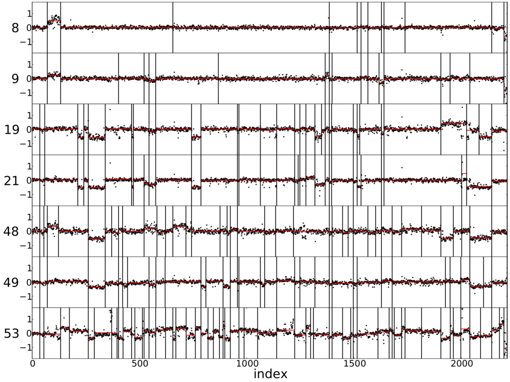


Array Comparative Genomic Hybridization

- Tumorous cells: deregulations in DNA copy number
- Samples: transcription of the chromosomes of patients, labelled with red fluorescent molecules
- Hybridization with reference gene copies, labelled with green fluorescent molecules
- Measure of the $\log_2$-ratio
- Objective: to localize the DNA portions over or under-expressed [AGH$^+$02, BV11]

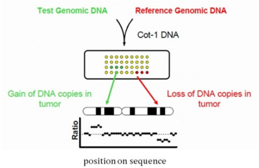


Bernoulli detector model, all patients jointly

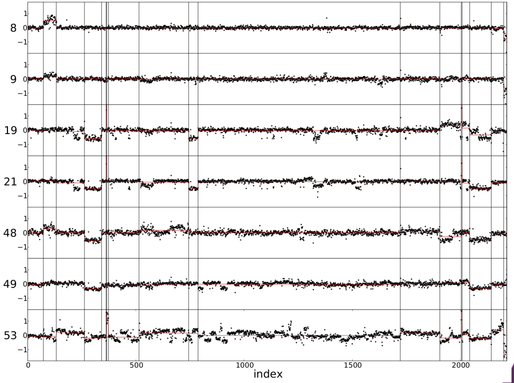


Array Comparative Genomic Hybridization

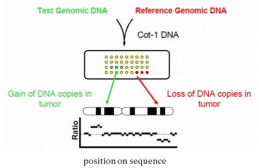
- Tumorous cells: deregulations in DNA copy number
- Samples: transcription of the chromosomes of patients, labelled with red fluorescent molecules
- Hybridization with reference gene copies, labelled with green fluorescent molecules
- Measure of the $\log_2$-ratio
- Objective: to localize the DNA portions over or under-expressed [AGH$^+$02, BV11]



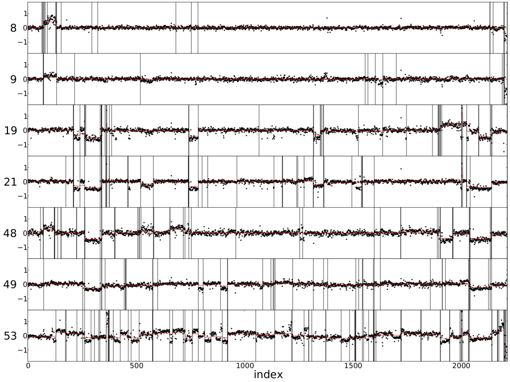
Group fused lasso [BV11], all patients jointly



- Tumorous cells: deregulations in DNA copy number
- Samples: transcription of the chromosomes of patients, labelled with red fluorescent molecules
- Hybridization with reference gene copies, labelled with green fluorescent molecules
- Measure of the $\log_2$-ratio
- Objective: to localize the DNA portions over or under-expressed [AGH$^+$02, BV11]



Fused lasso [Tib11], $\lambda = 3.0$,
each patient 53 individually



Plan

Conclusion

Advantages

- non parametric, robust statistical test, high power for the change-point model we choose
- weak assumptions on the data distribution
- flexible dependency structure learning, or used to improve the segmentation
- FDR controlled by α (empirically)

Drawbacks, limitations

- high complexity (linear with the number of configurations ϵ), MCMC method $\rightarrow$ slow, can't handle large number of time series
- composite marginal likelihood, dependency between the p-values
- approximation by the pseudo Gibbs sampler
- control of the FDR not formalized

Future work

- higher dimensions
- likelihood:
  - other statistical tests (Student's t-test, Welch's t-test...)
  - semi-parametric approach with the empirical likelihood [Owe10]
- dependency structure:
  - estimation of the causality from $\widehat{\boldsymbol{P}}$
  - graphical representation

Thank you for your attention !

Plan

References I

[AGH+02] David B Allison, Gary L Gadbury, Moonseong Heo, José R Fernández, Cheol-Koo Lee, Tomas A Prolla, and Richard Weindruch, *A mixture model approach for the analysis of microarray gene expression data*, Computational Statistics & Data Analysis **39** (2002), no. 1, 1–20.

[ASW+06] Sophie Achard, Raymond Salvador, Brandon Whitcher, John Suckling, and Ed Bullmore, *A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs*, The Journal of Neuroscience **26** (2006), no. 1, 63–72.

[BH95] Yoav Benjamini and Yosef Hochberg, *Controlling the false discovery rate: a practical and powerful approach to multiple testing*, Journal of the Royal Statistical Society. Series B (Methodological) (1995), 289–300.

[BV11] K. Bleakley and J.-P. Vert, *The group fused lasso for multiple change-point detection*, ArXiv e-prints (2011).

[CAMGP11] F. Chatelain, S. Achard, O. Michel, and C. Gouy-Pailler, *Multivariate approach for brain decomposable connectivity networks*, Statistical Signal Processing Workshop (SSP), 2011 IEEE, june 2011, pp. 817 –820.

[DTD07] N. Dobigeon, J. Y Tourneret, and Manuel Davy, *Joint segmentation of piecewise constant autoregressive processes by using a hierarchical model and a bayesian sampling approach*, Signal Processing, IEEE Transactions on **55** (2007), no. 4, 1251–1263.

[LYFLLC11] Alexandre Lung-Yut-Fong, Céline Lévy-Leduc, and Olivier Cappé, *Homogeneity and change-point detection tests for multivariate data using rank statistics*, 2011.

[Owe10] A.B. Owen, *Empirical likelihood*, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Taylor & Francis, 2010.

[SBB01] Thomas Sellke, M. J. Bayarri, and James O. Berger, *Calibration of p Values for Testing Precise Null Hypotheses*, The American Statistician **55** (2001), no. 1, 62–71.

[SSC99] Harold Sackrowitz and Ester Samuel-Cahn, *P values as random variables-expected p values*, The American Statistician **53** (1999), no. 4, 326–331.

[Tib11] Ryan Joseph Tibshirani, *The solution path of the generalized lasso*, Stanford University, 2011.

[Wil45] Frank Wilcoxon, *Individual comparisons by ranking methods*, Biometrics **1** (1945), no. 6, 80–83.

[XKH11] K.S. Xu, M. Kliger, and A.O. Hero, *A shrinkage approach to tracking dynamic networks*, Statistical Signal Processing Workshop (SSP), 2011 IEEE, june 2011, pp. 517 –520.