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

Flore Harlé
Sophie Achard, Gipsa-lab
Florent Chatelain, Gipsa-lab
Cédric Gouy-Paillet, CEA-List
Plan of presentation

1. Introduction
   - Context
   - Problem formulation

2. The Bernoulli detector model
   - Change-point model
   - The Wilcoxon rank-sum test
   - Prior on indicators
   - Posterior distribution and implementation

3. Experiments
   - Simulations
   - Applications on real data

4. Conclusion

5. References
Plan

1. Introduction
   - Context
   - Problem formulation

2. The Bernoulli detector model
   - Change-point model
   - The Wilcoxon rank-sum test
   - Prior on indicators
   - Posterior distribution and implementation

3. Experiments
   - Simulations
   - Applications on real data

4. Conclusion

5. References
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, dictionaries, 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, dictionaries, 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

Our approach:
- detection of events (change-points), using structural priors
- functional relationships inference from a temporal analysis
Goal: off-line multiple segmentation of multivariate time series

Approach: statistical test, Bayesian framework, signals dependencies

Bayes’ theorem:
\[ f(R|X) \propto L(X|R) f(R) \]

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

Goal: off-line multiple segmentation of multivariate time series

> observations $X \ (K \times N)$
> indicators $R \ (K \times N)$

Approach: statistical test, Bayesian framework, signals dependencies

Bayes’ theorem:

$$f(R|X) \propto L(X|R)f(R)$$

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

**observations** \( X \ (K \times N) \)

\( x_{j,i} \) mutually independent

**indicators** \( R \ (K \times N) \)

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

for all \( 1 \leq j \leq K, \ 1 \leq i \leq N \)

by convention \( r_1 = r_N = 1 \).

\[
R = \begin{pmatrix}
10\ldots0 & 10\ldots0 & \underbrace{00\ldots000\ldots01} \\
10\ldots000\ldots0 & 10\ldots01 & \underbrace{01\ldots010\ldots01} \\
10\ldots010\ldots0 & \underbrace{00\ldots000\ldots01}
\end{pmatrix}
\]

\( R_i = \epsilon = (0, 1, 0)^T \)
Plan

1. Introduction
   - Context
   - Problem formulation

2. The Bernoulli detector model
   - Change-point model
   - The Wilcoxon rank-sum test
   - Prior on indicators
   - Posterior distribution and implementation

3. Experiments
   - Simulations
   - Applications on real data

4. Conclusion

5. References
Change-point model for $L_*(X \mid R)$

> $R$ 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 $\mathcal{B}(\gamma, 1)$ [SBB01] parameter $\gamma \in (0, 1)$:
    > function of an acceptance level $\alpha$, $f(\alpha \mid r = 1) = f(\alpha \mid r = 0)$
    > $\gamma$ 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:

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

> composite marginal likelihood:

$$L_*(X \mid R) = \prod_{j=1}^{K} \prod_{i=2}^{N-1} \left( \gamma p_{j,i}^{\gamma - 1} \right)^{r_{j,i}}$$
The Wilcoxon rank-sum test

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

> For two segments \( Y \) and \( Z \):

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

\[
Y = (y_1, \ldots, y_M) \\
Z = (z_1, \ldots, z_N)
\]

rank sum \( R_Y \) in sorted vector \((Y, Z)\)

\[
U_Y = MN \frac{M(M + 1)}{2} - R_Y \\
U_Z = MN \frac{N(N + 1)}{2} - R_Z
\]

- 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 \((H_0)\) → the data are not assumed to be normally distributed.
Prior on indicators $f(R)$

> Indicators matrix:

$$R = \begin{pmatrix}
\ldots & 0 & \ldots & 1 & \ldots & 0 & \ldots \\
\ldots & 0 & \ldots & 0 & \ldots & 1 & \ldots \\
\ldots & 1 & \ldots & 0 & \ldots & 0 & \ldots \\
\ldots & 0 & \ldots & 1 & \ldots & 0 & \ldots \\
\end{pmatrix}$$

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

> Dependency: if the signal $k$ depends on the signal $l$, then $R_{k,i} = R_{l,i}$ with a high probability

$P_\epsilon$ is the probability to observe the configuration $\epsilon$ in $R \rightarrow P = (P_\epsilon)_{\epsilon \in \mathcal{E}}$

> $(R_i)_{2 \leq i \leq N-1}$ are assumed to be a priori independent: $f(R) = \prod_{i=2}^{N-1} f(R_i)$

> prior on indicators: $f(R, P) \propto f(R|P)f(P)$, with:

- $f(R|P) = \prod_{\epsilon \in \mathcal{E}} P_\epsilon^{S_\epsilon(R)}$, $S_\epsilon(R)$ is the number of times that the configuration $\epsilon$ appears in the columns of $R$
- vague prior for $P$: $D_L(d)$ [DTD07]

> finally:

$$f(R, P) \propto \prod_{\epsilon \in \mathcal{E}} P_\epsilon^{S_\epsilon(R) + d_\epsilon - 1}$$
Posterior distribution $f(R|X)$ and implementation

**Posterior distribution**

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

$$f(R, P|X) \propto L_*(X|R)f(R|P)f(P),$$

> The vector of hyperparameters $P_\epsilon$ can be integrated out:

$$f(R|X) \propto \prod_{j=1}^{K} \prod_{i=2}^{N-1} (\gamma P_{j,i}^{\gamma-1})^{r_{j,i}} \prod_{\epsilon \in \mathcal{E}} P_{\epsilon}^{S_\epsilon(R)+d_\epsilon - 1}.$$

**Algorithm**

> Estimation of the maximum \textit{a posteriori} of $R$
> Monte Carlo by Markov Chain method
> Gibbs sampling to draw the indicators matrix $R$, column by column
1 Introduction
   - Context
   - Problem formulation

2 The Bernoulli detector model
   - Change-point model
   - The Wilcoxon rank-sum test
   - Prior on indicators
   - Posterior distribution and implementation

3 Experiments
   - Simulations
   - Applications on real data

4 Conclusion

5 References
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:

\[
recall = \frac{TP}{TP + FN}\quad precision = \frac{TP}{TP + FP}
\]

> Gibbs sampler: 2 strategies

- *blocked Gibbs* sampling
- conditional probabilities does not form a compatible joint model \(\rightarrow\) *pseudo Gibbs* sampling

\[
R = (...) , 0 , 1 , 0 , ... , 0 , \overline{r_{i-1}, r_i, r_{i+1}} , 0 , ... , 0 , 1 , 0 , ...
\]

\[
P_{val} = (...) , \overline{p_{i-1}, p_i, p_{i+1}} , ... , \overline{p_{i-1}, p_i, p_{i+1}} , ... , \overline{p_{i-1}, p_i, p_{i+1}} , ...
\]
Choice of the Gibbs sampler

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

\[ SNR = 10 \log \left( \frac{(\mu_k - \mu_l)^2}{\sigma^2} \right). \]

> Detection performances:

\[
\text{recall} = \frac{TP}{TP + FN} \quad \text{precision} = \frac{TP}{TP + FP}
\]

> Gibbs sampler: 2 strategies
- \textit{blocked Gibbs} sampling
- conditional probabilities does not form a compatible joint model \(\rightarrow\) \textit{pseudo Gibbs} sampling

\[
R = (..., 0, 1, 0, ..., 0, \underbrace{r_i}_{\text{r_i}}, 0, ..., 0, 1, 0, ...)
\]

\[
P_{\text{val}} = (..., \cdot, p_{i-}, \cdot, ..., \cdot, \underbrace{p_i}_{\text{p_i}}, \cdot, ..., \cdot, p_{i+}, \cdot, ...)
\]
Impact of data distribution

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

Observations on segment \(k\):

\[ N(\mu_k, \sigma) \]

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

Recall and Precision plots for different SNR values.
Control of the false discovery rate (FDR)

Definition:
> multiple hypothesis testing
> maximizing the probability of detecting the true positive by controlling the false positives

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

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 \( \alpha \)

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

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

Figure: \( FDR = f(\alpha) \), 320 points, 15 change-points, SNR = 5 dB
Household electrical power consumption

> 4 time series
> Dependencies known → noninformative or informative prior on $P$

$$R = \begin{pmatrix} \ldots & 1 & \ldots & 0 & \ldots & 1 & \ldots & \ldots \\
\ldots & 0 & \ldots & 1 & \ldots & 0 & \ldots & \ldots \\
\ldots & 1 & \ldots & 0 & \ldots & 0 & \ldots & \ldots \\
\ldots & 0 & \ldots & 1 & \ldots & 0 & \ldots & \ldots \end{pmatrix}$$

Noninformative prior
Household electrical power consumption

> 4 time series
> Dependencies known → noninformative or informative prior on $P$

$$R = \begin{pmatrix} \ldots & 1 & \ldots & 0 & \ldots & 1 & \ldots & 1 & \ldots \\ \ldots & 0 & \ldots & 0 & \ldots & 1 & \ldots \end{pmatrix}$$

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
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]

Fused lasso [Tib11], $\lambda = 3.0$, each patient 53 individually
1 Introduction
   - Context
   - Problem formulation

2 The Bernoulli detector model
   - Change-point model
   - The Wilcoxon rank-sum test
   - Prior on indicators
   - Posterior distribution and implementation

3 Experiments
   - Simulations
   - Applications on real data

4 Conclusion

5 References
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 $\alpha$ (empirically)

Drawbacks, limitations

> high complexity (linear with the number of configurations $\epsilon$), 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 $\hat{P}$
  • graphical representation
Thank you for your attention!
References I


