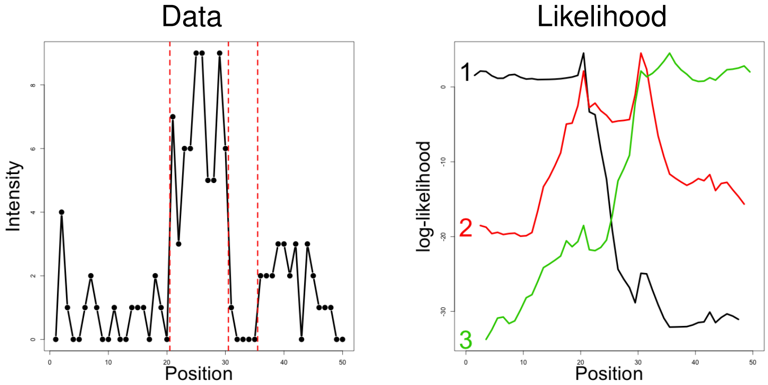
Exact posterior distributions and model selection for multiple change-point detection problems

G. Rigaill, E. Lebarbier, S. Robin

March, 2012

An example

- Dynamic Programming (DP, *Bellman and Dreyfus 1962*)
  - to recover the best segmentation in $K = 1$ to $K = 10$ segments
- Choice of K (model selection)
- Likelihood of the best segmentation having its k-th change at t (*Guédon 2009*)



Change-point model

Notations:

- $K =$ number of segments
- $r =$ region (or segment) $[\![\tau_r, \tau_{r+1}[\![$ $\qquad$ ($n_r =$ length of r)
- $m =$ segmentation: $m = \{r_1, \dots, r_K\}$
- $Y_t =$ signal at position t ($t \in [\![1, n]\!]$)

Model:

- $\{Y_t\}$ independent
- $t \in r$:

$$Y_t \sim p(\cdot|\theta_r)$$

e.g.

$$p(\cdot|\theta_r) = \mathcal{N}(\mu_r, \sigma^2), \qquad \mathcal{N}(\mu_r, \sigma_r^2), \qquad \mathcal{P}(\lambda_r)$$

Inference on the position and number of changes

- Change-points are discrete
- There is a large collection of possible models ($|\mathcal{M}_K| = \binom{n-1}{K-1}$)

Some difficulties:

- Standard model selection criteria (BIC...) are not theoretically justified
- Confidence on change-points, segments: standard MLE properties do not hold

Idea:

- We would like to select a K such that the confidence on the change-points is high/good
- This should ease the interpretation of the result

Outline

Model selection: BIC

The standard Laplace approximation used to derive the BIC criteria

$$\log p(M|\mathbf{Y}) = \log \int p(M, \theta|\mathbf{Y}) d\theta \approx \log p(M|\mathbf{Y}, \widehat{\theta}) - \frac{\log n}{2} \dim(M)$$

is not valid

because the likelihood is not differentiable with respect to the parameters.

Zhang and Siegmund 2007, based on a continuous-version of the segmentation problem derived a modified BIC criteria.

$$\text{pen}(K) = f(|\mathcal{M}_K|) + g\left(\sum_{r \in \widehat{m}(K)} \log n_r\right)$$

Model selection: penalized contrasts

- Best dimension K:

$$\widehat{K} = \arg\min_{K} \ell(\mathbf{Y}, \widehat{m}(K)) + \text{pen}(K)$$

- Best segmentation in $\mathcal{M}_K$:

$$\widehat{m}(K) = \arg \min_{m \in \mathcal{M}_K} \ell(\mathbf{Y}, m)$$

- *Lebarbier 2005*: $\quad \text{pen}(K) = \beta f(|\mathcal{M}_K|)$
- Constant penalty within each dimension $\mathcal{M}_K$.
- Estimation of β.

Outline

Exploring the segmentation space (best segmentation)

For a given dimension K, the optimal segmentation has to be found within

$$\mathcal{M}_K = \{m : |m| = K\}, \qquad |\mathcal{M}_K| = \binom{n-1}{K-1}$$

- An exhaustive search cannot be achieved.

Under a summation assumption $(m = \{r_1, \ldots, r_K\})$

$$p(\mathbf{Y}|m, \theta) = \sum_{r \in m} f(Y^r, \theta_r)$$

- Dynamic programming provides the solution $(\widehat{m}, \widehat{\theta})$ with complexity $\mathcal{O}(Kn^2)$.

Dynamic programming algorithm

Cost matrix and cost of a segment $r = [\![i, j [\![$

$$\begin{aligned} \text{if } j > i \quad \mathbf{C}_{ij} &= \quad f(Y^r, \theta_r) = -\log P(Y^r | \widehat{\theta}^r) \\ \text{if } j \leq i \quad \mathbf{C}_{ij} &= \quad +\infty \end{aligned}$$

Optimal cost/likelihood in K of $[\![1, n+1 [\![$:

$$s(K)_{1,n+1} = \min_{m \in \mathcal{M}_K} \sum_k \mathbf{C}_{\tau_k, \tau_{k+1}}$$

Update rule for $K = 2$:

$$s(2)_{1,n+1} = \min_{1<t<n} \{ \mathbf{C}_{1,t+1} + \mathbf{C}_{t+1,n+1} \}$$

Dynamic programming as matrix-vector products

Let's define: $\mathbf{u} * \mathbf{v} = min_i\{u_i + v_i\}$

The update rule for $K = 2$ can be rewritten as:

$$
\begin{aligned}
s(2)_{1,n+1} &= \min_{1<t<n} \{\mathbf{C}_{1,t+1} + \mathbf{C}_{t+1,n+1}\} \\
s(2)_{1,n+1} &= \min_{1\leq t\leq n+1} \{s(1)_{1,t+1} + \mathbf{C}_{t+1,n+1}\} \\
s(2)_{1,n+1} &= \mathbf{s}(1) * \mathbf{C}_{.,n+1}
\end{aligned}
$$

Then the line vector $\mathbf{s(2)}$ is obtained as

$$\mathbf{s(2)} = \mathbf{s(1)} * \mathbf{C}$$

More generally:

$$\mathbf{s(k+1)} = \mathbf{s(k)} * \mathbf{C}$$

Exploring the segmentation space

- Best segmentation in K with its k-th change at t

$$s(k)_{1,t+1} + s(K-k)_{t+1,n+1}$$

- Best segmentation in K with a change at t:

$$min_k\{s(k)_{1,t+1} + s(K-k)_{t+1,n+1}\}$$

- Best segmentation in K with its k-th segment $r = [\![t_1, t_2[\![$

$$s(k-1)_{1,t_1} + \ C_{t_1,t_2} + \ s(K-k-1)_{t_2,n+1}$$

- ...

Outline

Assessing the confidence on those changes

- Discrete nature of breakpoints
- Asymptotic results (*Feder (1975), Bai and Perron (2003); Muggeo (2003)*)
- Bootstrapping (*Husková and Kirch (2008)*)
- Exact exploration of the segmentation space (*Guédon (2009), Fearnhead (2006)*)

Exploring the segmentation space (posterior probabilities)

For a given dimension K,

$$\mathcal{M}_K = \{m : |m| = K\}, \qquad |\mathcal{M}_K| = \binom{n-1}{K-1}$$

- Exhaustive exploration cannot be achieved.

Under a factorisation assumption $(m = \{r_1, \dots, r_K\})$

$$p(\mathbf{Y}|m, \theta) = \prod_{r \in m} f(Y^r, \theta_r)$$

- A DP-like algorithm provides the solution with complexity $\mathcal{O}(Kn^2)$.

DP-like algorithm

Probability matrix and probability of a segment $r = [\![i, j[\![$

$$\begin{aligned} \text{if } j > i \quad \mathbf{A}_{ij} &= \quad f(Y^r, \theta_r) = \int p(Y^r|\theta_r)p(\theta_r)\mathrm{d}\theta_r \\ \text{if } j \leq i \quad \mathbf{A}_{ij} &= \quad 0 \end{aligned}$$

Posterior probability of K for $[\![1, n+1[\![$:

$$p(K)_{1,n+1} = \sum_{m \in \mathcal{M}_K} \prod_k \mathbf{A}_{\tau_k, \tau_{k+1}}$$

Update rule for $K = 2$:

$$p(2)_{1,n+1} = \sum_{1<t<n} \mathbf{A}_{1,t+1} \mathbf{A}_{t+1,n+1}$$

Matrix-vector products

As for the optimization problem this can be seen as a matrix-vector product:

$$\mathbf{uv} = min_i\{u_i v_i\}$$

The line vector $\mathbf{p(2)}$ is obtained as

$$\mathbf{p(2)} = \mathbf{p(1)}\ \mathbf{A}$$

More generally:

$$\mathbf{p(k+1)} = \mathbf{p(k)}\ \mathbf{A}$$

and

$$\mathbf{p(k+1)} = \mathbf{p(1)}\ \mathbf{A}^k$$

Exploring the segmentation space

- Localisation of the k-th change

$$\Pr\{\tau_k = t | K\} = p(k)_{1,t+1} p(K-k)_{t+1,n+1}$$

- The probability that there is a breakpoint at position t:

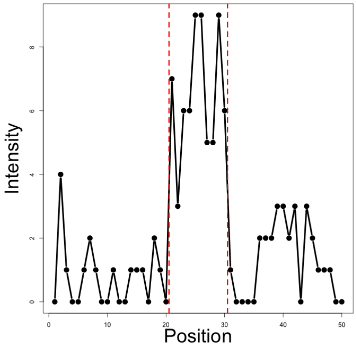
$$\Pr\{\exists k : \tau_k = t | \mathbf{Y}, K\} = \sum_{k=1}^{K} \Pr\{\tau_k = t | K\}$$

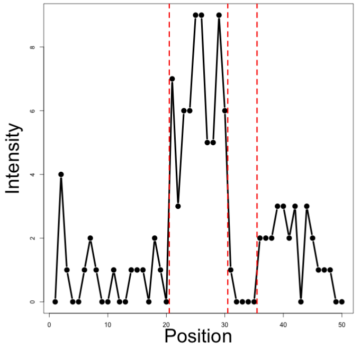
- The probability of segment $r = [\![t_1, t_2[\![$ for a given K
- The posterior entropy of m within a dimension:

$$\mathcal{H}(K) = -\sum_{m \in \mathcal{M}_K} p(m|\mathbf{Y}, K) \log p(m|\mathbf{Y}, K)$$

An example, K=3 and K=4

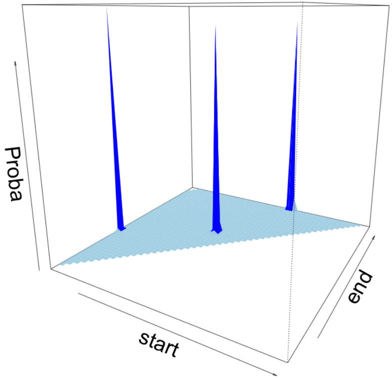
Best segmentation

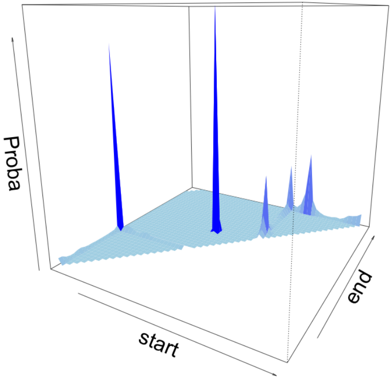




An example, K=3 and K=4

Segment probability





Outline

Back to model selection

- Posterior probability of a segmentation

$$P(m|Y) \qquad \text{, or "exact" BIC} \qquad -\log(P(m|Y))$$

- "Exact" Deviance Information Criteria (*Spiegelhalter et al. (2002)*)
  - $f(Y)$ is the likelihood of the saturated model.
  - Deviance: $D(\Theta) = -2\log P(Y|\Theta) + 2\log f(Y)$

$$DIC(K) = -D(\mathbb{E}[\Theta|Y,K]) + 2\mathbb{E}[D(\Theta)|Y,K]$$

- "Exact" Integrated Completed Likelihood (*Biernacki et al. (2000)*)

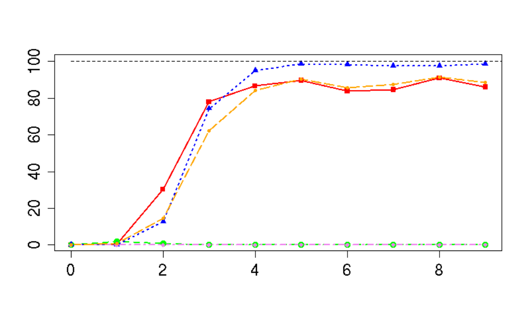
$$\text{ICL}(K) = -\log P(K|Y) + \mathcal{H}(K)$$

  - It favors a K where the best segmentation is by far the most probable

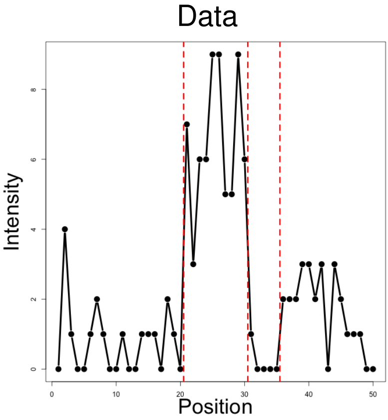
Selection of the number of breakpoints

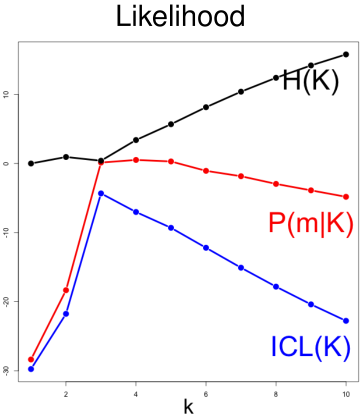
Simulations

- Comparison of P(m|Y), DIC(K) and ICL(K)
- 150 observations with 6 breakpoints
- Increasing signal to noise ratio



Back to the example





Conclusion and perspectives

- DP as a matrix-vector product ($\mathcal{O}(Kn^2)$ runtime)
  - Best segmentation
  - Best segmentation with a change at t
  - Posterior probability of a change at t
  - Posterior probability of a segment
  - Posterior entropy
- Model selection
  - "Exact" BIC for segmentation
  - "Exact" DIC for segmentation
  - "Exact" ICL for segmentation (using the entropy)
  - Priors
- More details in our paper
- Runtime for large n? (see The Minh Luong and Alice Cleynen's presentations)

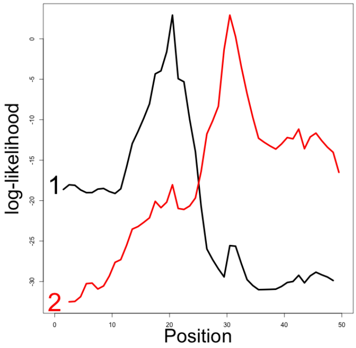
Acknowledgements

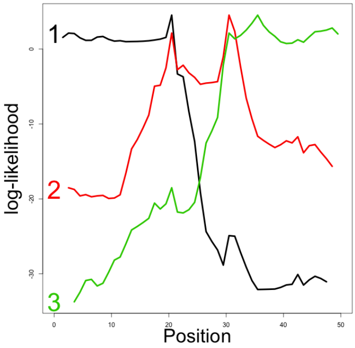
- Emilie Lebarbier, Stéphane Robin
- Alice Cleynen, Michel Koskas
- Lodewyk Wessels

Thank you

An example, K=3 and K=4

Best segmentation in K with its k-th change at t





An example, K=3 and K=4

Change-point probability