A not so short Introduction to Process Segmentation

Franck Picard[‡,◇]

[‡]UMR 5558 UCB CNRS LBBE, Lyon, France
[◇] Projet BAMBOO, INRIA, F-38330 Montbonnot Saint-Martin, France.
picard@biomserv.univ-lyon1.fr
http://pbil.univ-lyon1.fr/members/fpicard/

RSL - Lyon June 2010

Outline

Segmentation models: definitions and notations

- We observe $\{y_1, \ldots, y_n\}$ a sequence of data modeled by a random process $\mathbf{Y} = \{Y_1, \ldots, Y_n\}$ with

$$Y_t \sim f(\theta_t).$$

- We suppose that there exists $K+1$ change-points $t_0 = 1 < \ldots < t_K = n$ such that θ_t is constant between two changes and different from a change to another.
- $I_k =]t_{k-1}, t_k]$: interval of stationarity, θ_k the parameter between two changes:

$$\forall t \in I_k, \;\; Y_t \sim f(\theta_k)$$

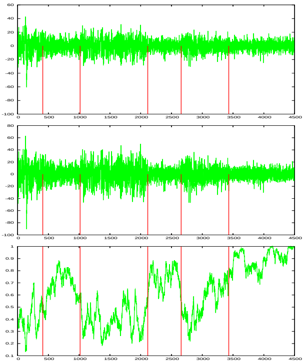
- θ can stand for the mean, variance, spectrum, etc...

Process Segmentation vs. Online Change point detection

- Sequential observations: the detection of a change should be done with past observations only
- Example: quality control, earthquake detection, ...
- Main Reference: Basseville & Nikiforov (93) [3]
- Sequential analysis (mainly based on tests)

Applications of process segmentation

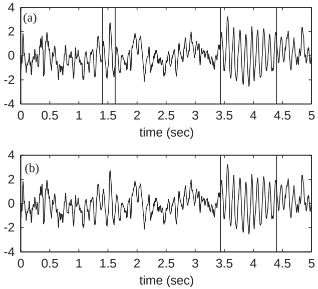
- Econometrics [17, 16]
- Medical Imagery [18]
- Climate series [22]
- Biology (sequence segmentation [6, 5], microarrays [25])



Market prices segmentation [19]

Applications of process segmentation

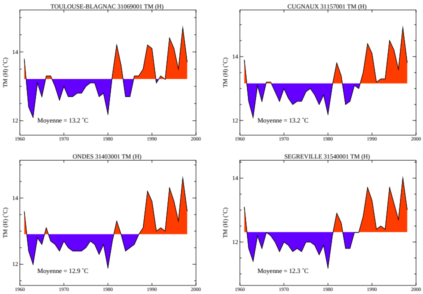
- Econometrics [17, 16]
- Medical Imagery [18]
- Climate series [22]
- Biology (sequence segmentation [6, 5], microarrays [25])



EEG segmentation [18]

Applications of process segmentation

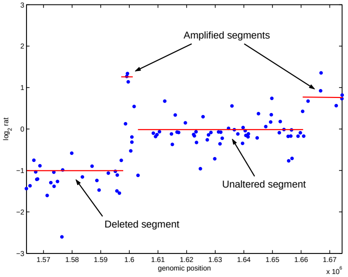
- Econometrics [17, 16]
- Medical Imagery [18]
- Climate series [22]
- Biology (sequence segmentation [6, 5], microarrays [25])



Climate series segmentation [22]

Applications of process segmentation

- Econometrics [17, 16]
- Medical Imagery [18]
- Climate series [22]
- Biology (sequence segmentation [6, 5], microarrays [25])



Array CGH segmentation [25]

Main contributors (non exhaustive) !

- P. Perron (Boston University)
- T.L. Lai (Stanford University)
- D. Siegmund (Stanford University)
- P. Fearnhead (Lancaster University)
- P. Green (Bristol University)
- M. Lavielle (INRIA, Orsay)
- E. Lebarbier (AgroParisTech)

Outline of the presentation

- How to build the model ?
- How to estimate the parameters and the location of the breaks ?
- Properties of the breaks estimators ?
- How many breaks ?
- How to deal with dependent observations ?
- The Bayesian Perspective

Outline

A piece-wise constant regression

- We observe a Gaussian process (iid) $\mathbf{Y} = \{Y_1, \ldots, Y_n\}$ with

$$Y_t \sim \mathcal{N}(\mu_t, \sigma^2).$$

- We suppose that there exists $K+1$ change-points $t_0 < \ldots < t_K$ such that the mean of the signal is constant between two changes and different from a change to another.
- $I_k =]t_{k-1}, t_k]$: interval of stationarity, μ_k the mean of the signal between two changes:

$$\forall t \in I_k, \;\; Y_t = \mu_k + E_t, \;\; E_t \sim \mathcal{N}(0, \sigma^2).$$

Generalization to piece-wise linear regressions

- The parameter subject to changes can be $\mathbb{E}(Y(t))$ and/or $\mathbb{V}(Y(t))$
- The model is extended to piece-wise linear regression
- $I_k =]t_{k-1}, t_k]$: interval of stationarity, $\boldsymbol{\theta}_k$ the set of parameters between two changes:

$$\forall t \in I_k, \ \ Y_t = \sum_{j_1}^{p} \theta_j^k x_j(t) + E_t, \ \ E_t \sim \mathcal{N}(0, \sigma^2).$$

- Difference with splines: no continuity constraint at the breaks

Parameters and estimation strategy

- The parameters: $\mathbf{t} = \{t_0, \dots, t_K\}$, $\boldsymbol{\mu} = \{\mu_1, \dots, \mu_K\}$ and σ^2.
- The estimation is done for a given K which is estimated afterwards.
- The log-likelihood of the model is:

$$\log \mathcal{L}_K(\mathbf{Y}; \mathbf{t}, \boldsymbol{\mu}, \sigma^2) \;=\; \sum_{k=1}^{K} \sum_{t=t_{k-1}+1}^{t_k} f(y_t; \mu_k, \sigma^2).$$

- When K and $\mathbf{t}$ are known, how to estimate $\boldsymbol{\mu}$?
- When K is known, how to estimate $\mathbf{t}$?
- How to choose K ?

Penalized contrast estimators

- Penalized contrast estimators are of the form:

$$(\hat{\mathbf{t}}, \hat{\boldsymbol{\theta}}) = \arg\min_{\mathbf{t},\boldsymbol{\theta}} \{J_K(\mathbf{Y};\mathbf{t},\boldsymbol{\theta}) - \beta \text{pen}(\mathbf{t})\}$$

- $J_K(\mathbf{Y};\mathbf{t},\boldsymbol{\theta})$: to assess the quality of fit of the model
  - locate the changepoints as accurately as possible.
  - Can be broken down into local contrast (log-likelihoods)

$$J_K(\mathbf{Y};\mathbf{t},\boldsymbol{\theta}) = \sum_k \log \ell(Y[t_{k-1} : t_k]; \theta_k)$$

- pen($\mathbf{t}$) only depends on K (increases with K)
- β establishes a trade-off between the contrast and the penalty

Parameter estimation

- When K and $\mathbf{t}$ are known the estimation of $\boldsymbol{\mu}$ is straightforward:

$$\begin{aligned}\widehat{\mu}_k &= \frac{1}{\widehat{t}_k - \widehat{t}_{k-1}} \sum_{t=\widehat{t}_{k-1}+1}^{\widehat{t}_k} y_t, \\ \widehat{\sigma}^2 &= \frac{1}{n} \sum_{k=1}^{K} \sum_{t=\widehat{t}_{k-1}+1}^{\widehat{t}_k} (y_t - \widehat{\mu}_k)^2.\end{aligned}$$

- Find $\widehat{\mathbf{t}}$ such that:

$$\widehat{\mathbf{t}} = \arg\max_{\mathbf{t}} \left\{ \log \mathcal{L}_K(\mathbf{Y}; \mathbf{t}, \boldsymbol{\mu}, \sigma^2) \right\}.$$

Outline

Dynamic Programming to optimize the log-likelihood

- Partition n data points into K segments: complexity $\mathcal{O}(n^K)$.
- DP reduces the complexity to $\mathcal{O}(n^2)$ when K is fixed.
- Analogy with the shortest path problem:
  - "subpaths of optimal paths are themselves optimal"
- $\text{RSS}_k(i,j)$ cost of the path connecting i to j in k segments:

$$
\begin{aligned}
\forall 0 \leq i < j \leq n,\ \text{RSS}_1(i,j) &= \sum_{t=i+1}^{j} (y_t - \bar{y}_{ij})^2, \\
\forall 1 \leq k \leq K-1,\ \text{RSS}_{k+1}(1,j) &= \min_{1 \leq h \leq j} \left\{ \text{RSS}_k(1,h) + \text{RSS}_1(h+1,j) \right\}.
\end{aligned}
$$

Dynamic Programming on very large signals ?

- Even if DP reduces the computational burden to $\mathcal{O}(n^2)$ it may be problematic when $n \sim 10^6$
- Constraint the length of segments (lmin, lmax)
- Sequential strategies (Bayesian [13])
- Use the LARS framework [4]
- Find a trick to the trick to decrease the complexity of DP [26]

Outline

Breaks estimator convergence

- Let $\boldsymbol{\tau} = \{0 < \tau_0 < \ldots < \tau_K < 1\}$ and $\boldsymbol{\tau}^\star$ the sequence of (true) normalized change points
- Let $\theta \in \mathbb{R}^d$ and $\theta^\star$ be the (true) parameter subject to changes the true vector of parameters
- Let $\hat{\boldsymbol{\tau}}_n$ and $\hat{\theta}_n$ be minimum contrast estimators
- If K is known, then under very mild conditions [17]

$$(\hat{\boldsymbol{\tau}}_n, \hat{\theta}_n) \underset{\mathbb{P}^\star}{\rightarrow} (\boldsymbol{\tau}^\star, \theta^\star)$$

- Note that the rate of convergence of $\hat{\boldsymbol{\tau}}_n$ is n
- If K is unknown, convergence depends on $\beta_n \text{pen}(\mathbf{t})$ (Ex: $K \log(n)/2$)
- More generally β_n should tend to 0 at an appropriate rate
- Results include strongly mixing and strongly dependent processes [17]

Limit Distribution for the breaks-1

- Central Ingredient to get results ($\delta_k = \mu_{k+1} - \mu_k$)

$$\forall \epsilon > 0, \exists C < \infty, \ \mathbb{P}\left\{ n|\hat{t}_k - t_k^\star| < C/\delta_k^2 \right\} < \epsilon$$

- For one break $t_1, \delta = \mu_2 - \mu_1$, and t_1 lies in a compact set

$$\left\{ |t_1 - t_1^\star| < C\delta^{-2} \right\}$$

- Recall that $\hat{t_1} = \arg\min\{RSS(t_1)\} = \arg\max\{RSS(t_1^\star) - RSS(t_1)\}$

$$RSS(t_1^0) - RSS(t_1) = \begin{cases} -\delta^2(t_1^\star - t_1) + 2\delta \sum_{t_1+1}^{t_1^\star} \epsilon_t + o_p(1) \\ -\delta^2(t_1 - t_1^\star) - 2\delta \sum_{t_1^\star+1}^{t_1} \epsilon_t + o_p(1) \end{cases}$$

- The distribution of these sums depends on t_1

Limit Distribution for the breaks-2

- Let us define W such that $W(0) = 0$ and

$$W(m) = \begin{cases} -\delta^2 m + 2\delta \sum_{t=m+1}^{0} \epsilon_t, & \text{for } m > 0 \\ -\delta^2 m + 2\delta \sum_{t=1}^{m} \epsilon_t, & \text{for } m < 0 \end{cases}$$

- Assuming a stricly stationary distribution for $\{\epsilon_t\}$ then

$$RSS(t_1^\star) - RSS(t_1) = W(t_1 - t_1^\star) + o_p(1)$$

- Using conditions (on ϵ_t) that ensure a unique max for W then

$$\hat{t}_1 - t_1^\star \underset{d}{\rightarrow} \arg\max_m W(m)$$

- More general results can be found in the litterature [29, 23]

Confidence intervals for break dates

- Results use limit distributions, but may be difficult to handle in practice [30]
- Many techniques use likelihood ratios and sequential analysis [27, 28, 8]
- Resampling strategies are difficult to define in the case of multiple changes [15]
- Bayesian strategies are more suitable for confidence assessment in the case of multiple changes

Outline

Penalized contrasts to estimate the number of segment

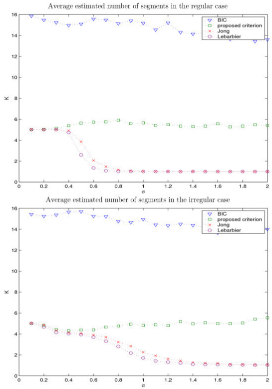
- The number of segments K should be estimated:

$$\widehat{K} = \arg\max_{K} \left\{ \log \mathcal{L}_K(\mathbf{Y}; \widehat{\mathbf{t}}, \widehat{\boldsymbol{\mu}}, \widehat{\sigma}^2) - \beta \text{pen}(K) \right\}.$$

- Main difficulty: breakpoints are discrete parameters
  - the likelihood is not differentiable wrt $\mathbf{t}$
  - C_{n-1}^{K-1} possible segmentations for a model with K segments.
  - how to define the dimension of the model ?
- How to define pen(K) ?
- How to define β ?

Comparison of segmentation results

Criterion	pen(K)	β
AIC	$2K$	1
BIC	$2K$	$\log(n)/2$
Lavielle	$2K$	adaptive
mBIC	$f(K, \sum_k \log n_k)$	$\log(n)/2$
Lebarbier	$c_1 + c_1 \log(n/K)$	adaptive



Simulations [25] and $\hat{K} = f(\sigma)$ with $K^\star = 5$

Construction of a Bayesian Criterion mBIC [31]

- Define $\mathcal{M}_K$ the segmentation model with K segments, then

$$\text{mBIC}(K) = \frac{\log \Pr\{\mathbf{Y}|\mathcal{M}_K\}}{\Pr\{\mathbf{Y}|\mathcal{M}_0\}}$$

- Derive an asymptotic approximation of the Bayes factor
- Use asymptotic results $\lim_{n\to\infty} t_k/n = \tau_k$ and a prior for $\boldsymbol{\tau}$ of the form:

$$\pi(\boldsymbol{\tau}) = g(\boldsymbol{\tau})/n^K, \;\; C_1 < \max g(\boldsymbol{\tau}) < C_2$$

- In the case σ^2 known we get:

$$\text{mBIC}(K) = SSB(\hat{\mathbf{t}}) - \sum_k \log(\hat{t}_{k+1} - \hat{t}_k) + (0.5 - K)\log(n) + O_p(1)$$

Penalty function in a non asymptotic framework-1

- Consider the regression: $Y(t) = s(t) + \epsilon(t)$
- Define $\mathcal{S}_m$ the set of piece-wise constant functions on partition $m = \{I_k\}_{k=1,K_m}$:

$$\mathcal{S}_m = \left\{ u = \sum_{k=1}^{K_m} u_k \mathbb{I}\{I_k\},\ (u_k)_k \in \mathbb{R}^{K_m} \right\}$$

- The approach of Birge-Massart is to consider that $s \notin \mathcal{S}_m$ but that $\mathcal{S}_m$ is just an approximation set

Penalty function in a non asymptotic framework-2

- Define $\bar{s}_m$ the projection of s on $\mathcal{S}_m$: it is an approximation of s but is unknown
- Define $\hat{s}_m$ the estimator of $\bar{s}_m$ in $\mathcal{S}_m$ whose quadratic risk is $\mathbb{E}\|s - \hat{s}_m\|^2$
- This risk can be broken down such that (bias/variance trade-off)

$$\mathbb{E}\|s - \hat{s}_m\|^2 = \mathbb{E}\|s - \bar{s}_m\|^2 + \mathbb{E}\|\bar{s}_m - \hat{s}_m\|^2$$

- Bias term: $\mathbb{E}\|s - \bar{s}_m\|^2$ measures the distance of the unknown s to its approximator $\bar{s}_m$ in $\mathcal{S}_m$
- Variance term: $\mathbb{E}\|\bar{s}_m - \hat{s}_m\|^2$ measures the quality of estimation
- The **ideal** estimator will achieve the **best** Bias/Variance trade-off

Penalty function in a non asymptotic framework-3

- In the case of process segmentation, this framework leads to a penalty of the form[21]

$$\beta \times \mathrm{pen}(K) = \frac{K}{n}\sigma^2 \left(c_1 + c_2 \log \frac{n}{K}\right)$$

- (c_1, c_2) to be calibrated and σ^2 to be estimated
- Emipirical behavior: minimization of the risk can lead to a lack of power in detection

Using the Slope heuristic-1

- General heuristic that is very effective and easy to implement in practice
- Idea: construct the sequence of β_i using $\{(\text{pen}(K_i), J_{K_i})\}$ the convex hull of the set $\{(\text{pen}(K), J_K)\}$

$$\beta_i = \frac{J_{K_i} - J_{K_{i+1}}}{\text{pen}(K_{i+1}) - \text{pen}(K_i)}$$

- Look at the length ℓ_i of intervals $[\beta_i, \beta_{i+1}]$ and retain the value(s) of K_i such that $\ell_j >> \ell_i$: find the "biggest jump" of dimension
- Strategy close to L-curve strategies [18]

Using the Slope heuristic-2

- Normalize J_K s.t. (average slope=-1)

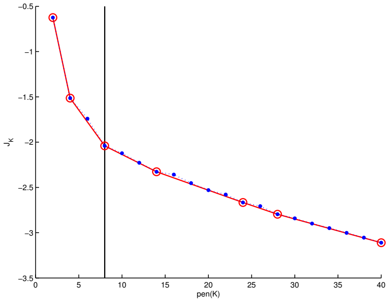
$$\widetilde{J}_K = \frac{J_{\text{Kmax}} - J_K}{J_{\text{Kmax}} - J_1}(\text{Kmax}-1)+1$$

- Use the empirical second derivative

$$D_K^2 = \widetilde{J}_{K-1} - \widetilde{J}_K + \widetilde{J}_{K+1}$$

- Choose the best $K(S)$ s.t.

$$\hat{K}(S) = \arg\max\{D_K^2 > S\}$$



blue line: contrast, red line: convex hull

Outline

Principle of the Bayesian view of process segmentation

- Model determination using hierarchical modelling:

$$\Pr\{\mathbf{Y}, \boldsymbol{\mu}, K\} = \Pr\{K\} \times \Pr\{\boldsymbol{\mu}|K\} \times \Pr\{\mathbf{Y}|\boldsymbol{\mu}, K\}$$

- Change-points $\{t_k\}_k$ are considered as random variables with distribution $\pi(\mathbf{t}; \lambda, K)$
- The objective : recover the *posterior* distribution: $\pi(\mathbf{t}|\mathbf{Y}, \boldsymbol{\mu}, \sigma^2, K)$
- Considering random variables makes some issues easier to assess: confidence intervals, dependent data, uncertainty about model choice
- Computationnaly intensive: MCMC, Hasting Metropolis, Reversible Jump, Forward-Backward Recursions

The multiple change point problem and the Reversible Jump algorithm-1 [14]

- Suppose that the number of segments is $K \sim \mathcal{P}(\lambda)$
- With K given, breaks positions are uniformely distributed on $[0; n]$:

$$t_1 < \ldots < t_K | K \sim \mathcal{U}[0; n]$$

- Then the mean of each segment $\{\mu_k\}_k$ are *iid* s.t.:

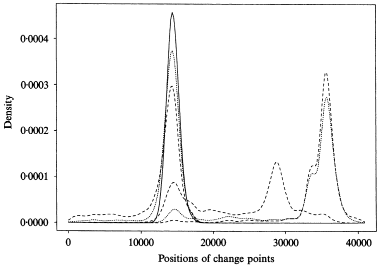
$$\boldsymbol{\mu} | \mathbf{t}, K \sim \Gamma(\alpha, \beta)$$

- RJ-MCMC is used to compute $\pi(K, \mathbf{t}, \boldsymbol{\mu} | \mathbf{Y})$

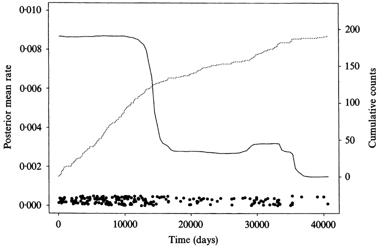
The multiple change point problem and the Reversible Jump algorithm-2 [14]

- The target distribution is $\pi(K, \mathbf{t}, \boldsymbol{\mu} | \mathbf{Y})$
- Dimension of the model change according to K: how to design appropriate moves ?
    - a change to the mean of a randomly chosen segment
    - b change of a position of a randomly chosen break
    - c 'birth' of a new segment at a randomly chosen location on $[0, n]$
    - d 'death' of a randomly chosen segment

Posterior inference of change-points location



$\pi(\mathbf{t}|\mathbf{Y},\boldsymbol{\mu},K)$



$\pi(\boldsymbol{\mu}|\mathbf{Y},\mathbf{t},K)$

Limitations of the RJ-MCMC algorithm

- Jumps in dimension lead to very demanding algorithms
- The posterior of the mean is very smooth: not in accordance with the "abrupt-changes" model
- Reparametrization of the model with $\mathbf{r} = \{r_t\}$, a sequence of length n s.t.:

$$\{r_t = 1\} \text{ if } t = t_k$$

- Use a temperature parameter during the Hastings-Metropolis algorithm to discriminate the local and global maxima of the posterior

A new formulation of the Bayesian change-point problem [20]

- $r_t \sim \mathcal{B}(\lambda)$, and $K = \sum_t r_t \sim \mathcal{B}(n-1, \lambda)$
- The sequence $\mathbf{r}$ is of fixed length: no need of jumps in dimension to assess $\pi(\mathbf{r}|\mathbf{Y})$
- The model on $\mathbf{Y}$ is unchanged:

$$\forall t \in I_k \;\; Y_t \sim \mathcal{N}(\mu_k, \sigma^2)$$

- The sequence of means μ_k is modelled s.t.

$$\boldsymbol{\mu} \sim \mathcal{N}(\mathbf{m}; s^2)$$

Back to penalized constrast estimators

- For any configuration of change-points the posterior distribution of $\mathbf{r}$ is

$$\pi(\mathbf{r}|\mathbf{Y};\boldsymbol{\theta}) \propto \exp\left\{-\phi RSS(\mathbf{r}, K_r) - \gamma K_r\right\}$$

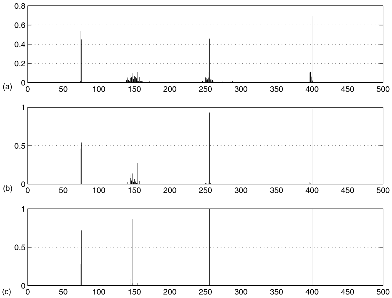
- This is the *joint* distribution of a vector of size $n-1$
- The MAP estimator of $\mathbf{r}$ is a penalized contrast estimator !
- This posterior distribution can be computed with a Hastings-Metropolis algorithm
- Use SAEM [9] to estimate $\boldsymbol{\theta}$ the set of hyperparameters

Running the H-M algorithm at low temperature

- Strategy inspired from Simulated Annealing algorithms
- Introduce a temperature parameter T s.t. $\pi_T(\mathbf{r}|\mathbf{Y};\boldsymbol{\theta})$ changes to

$$\exp\left\{-\frac{\phi}{T}RSS(\mathbf{r},K_r)-\frac{\gamma}{T}K_r\right\}$$

- When $T\to 0$, $\pi_T(\mathbf{r}|\mathbf{Y};\boldsymbol{\theta})$ CV to the uniform distribution of the seg of global maxima of $\pi(\mathbf{r}|\mathbf{Y};\boldsymbol{\theta})$



$\pi_T(\mathbf{r}|\mathbf{Y};\boldsymbol{\theta})$ with decreasing T

Running the H-M algorithm at low temperature

- Difficulties in using MCMC: how to design moves (between different models) which enable the MCMC algorithm to mix well, and being able to detect convergence of the chain.
- Idea: use recursions inspired from the Forward-Backward algorithm [12]
- Objective : perform direct simulation from the posterior distribution of $\mathbf{t}$ and K

Using point process to describe the sequence of changes

- Introduce some dependency among change-points $\pi(t_k|t_{k-1})$
- Introduce a point-process on integers with $g(t) > 0$ the time between two successive points (product-partition model)
- $G(t) = \sum_{s=1}^{t} g(s)$ is the distribution function of the distance bewteend two successive points ($g_0(t)$ the mass function of the first point after 0) then

$$\pi_K(\mathbf{t}) = g_0(t_1)\left(\prod_{k=2}^{K} g(t_k - t_{k-1})\right)(1 - G(t_{k+1} - t_k))$$

- Suppose a Negative Binomial distribution for $g(t)$ (discrete version of Gamma distributions, $k = 1$ leads to Markov distrib.)

$$g(t) = \mathcal{C}_{t-k}^{k-1} p^k (1-p)^{t-k}$$

Basic Recursions

- Define $\forall s \geq t \;\; P(t,s) = \Pr\{Y_{t:s} | t, s \text{ in the same segment}\}$
- Define $Q(t) = \Pr\{Y_{t:n} | \text{changepoint at } t-1\}$

$$Q(t) = \sum_{s=t}^{n-1} P(t,s) Q(s+1) g(s+1-t) + P(t,n)(1 - G(n-t))$$

- Then the posterior distribution of the change points is:

$$\Pr\{t_k | t_{k-1}, Y_{1:n}\} = P(t_{k-1}+1, t_k) Q(t_k+1) g(t_k - t_{k-1}) / Q(t_{k-1}+1)$$

Model selection using posterior distributions

- Model selection can be performed using

$$\pi(K|Y_{1:n}) \propto \pi(K)\pi(Y_{1:n}|K)$$

- Redefine the recursion conditioning by K
- Define $Q_j^K(t) = \Pr\{Y_{t:n}|t_j = t-1, K\}$

$$Q_j^K(t) = \sum_{s=t}^{n-K+j} P(t,s)Q_{j+1}^K(s+1)\pi_K(t_j = t-1|t_{j+1} = s)$$

- Finally

$$\Pr\{Y_{1:n}|K\} = \sum_{s=1}^{n-K} P(1,s)Q_1^K(s+1)$$

Outline

The piece-wise AutoRegressive Model [16]

- Piece-wise constant volatility and regression parameters $\theta_t = (\mu_t, \alpha_{\bullet,t})^T$:

$$Y_t = \mu_t + \sum_{s=1}^{k} \alpha_{s,t} Y_{t-s} + \sigma_t \epsilon_t \quad t > k$$

- The jump process is modeled with r_t s.t. $r_t \sim \mathcal{B}(p)$
- Denoting by (Z_t^T, γ_t) the set of new parameters:

$$(\theta_t^T, \sigma_t) = (1 - r_t) \times (\theta_{t-1}^T, \sigma_{t-1}) + r_t \times (Z_t^T, \gamma_t),$$

- Forward-Backward recursions are used to calculate:

$$\mathbb{E}(\boldsymbol{\theta}_n^T, \sigma_n | Y_{1:n})$$

Conclusions & perspectives

- Very old/wide subject !!!
- Sequential analysis are taking some new importance due to the increase in the size of the datasets
- Other projects involve the segmentation of many series [10, 24, 1, 7]
- Towards semi parametric models and links with functional data [24, 2, 11]
- Slides @ `http://pbil.univ-lyon1.fr/members/fpicard/`

F. Picard an E. Lebarbier, E. Budinska, and S. Robin.
Joint segmentation of multivariate gaussian processes using mixed linear models.
Technical Report 5, SSB, 2007.

A. Aue, R. Gabrys, L. Horvath, and P. Kokoszka.
Estimation of a change-point in the mean function of functional data.
JMVA, 100:2254–2269, 2009.

N. Basseville and I. Nikiforov.
Detection of abrupt changes. Theory and application.
Prentice Hall Information and system sciences series, 1993.

K. Bleakley and J.P. Vert.
Joint segmentation of many acgh profiles using fast group lars.
Technical report, HAL-00422430, 2009.

J. V. Braun, R.K. Braun, and H.G Muller.
Multiple change-point fitting via quasilikelihood, with application to DNA sequence segmentation.
Biometrika, 87:301–314, 2000.

J. V. Braun and H.G Muller.
Statistical methods for DNA sequence segmentation.
Statistical Science, 13(2):142–162, 1998.

H. Caussinus and O. Mestre.
Detection and correction of artificial shifts in climate series.
JRSS-C, 53(3):405–425, 2004.

G.W. Cobb.
The problem of the Nile. Conditional solution to a change-point problem.
Biometrika, 65(2):243–251, 1978.

B. Delyon, M. Lavielle, and E. Moulines.
Convergence properties of a stochastic approximation verison of the em algorithm.

Ann. Stat., 27(1):94–128, 1999.

N. Dobigeon, J.-Y. Tourneret, and J. Scargle.
Joint segmentation of multivariate astronomical time series: Bayesian sampling with a hierarchical model.
IEEE Trans. Signal Processing, 55(1), 2007.

Y. Fan, J.and Wang.
Multi-scale jump and volatility analysis for high-frequency financial data.
JASA, 102:1349–1362, 2007.

P. Fearnhead.
Exact abd efficient bayesian inference for multiple changepoint problems.
Stat. and Comp., 16:203–213, 2006.

P. Fearnhead and Z. Liu.
Online inference for multiple changepoint problems.
JRSS-B, 69:589–605, 2007.

P.J. Green.
Reversible jump Markov chain Monte Carlo computation and Bayesian model determination.
Biometrika, 82(4):711–732, 1995.

M. Huskova and Kirch C.
Boostraping confidence intervals fort the change point of time series.
JTSA, 29(6):947–971, 2008.

T.L. Lai, H. Liu, and H. Xing.
Autoregressive models with piecewise constant volatility and regression parameters.
Statistica Sinica, 15:279–301, 2005.

M. Lavielle.
Detection of multiple changes in a sequence of dependent variables.
Stoch. Proc. and Appl., 83:79–102, 1999.

M. Lavielle.

Using penalized contrasts for the change-point problem.
Signal Processing, 85(8):1501–1510, 2005.

M. Lavielle and Teyssière G.
Detection of multiple change-points in multiple time-series.
Lithuanian Mathematical Journal, 46(4):351–376, 2006.

M. Lavielle and E. Lebarbier.
An application of MCMC methods for the multiple change-points problem.
Signal Processing, 81:39–53, 2001.

E. Lebarbier.
Detecting multiple change-points in the mean of Gaussian process by model selection.
Signal Processing, 85:717–736, 2005.

O. Mestre.
Methodes statistiques pour l'homogeneisation de longues series climatiques.
PhD thesis, Université Paul Sabatier, Toulouse, 2000.

Pierre Perron.
Dealing with structural break.
Technical report, Boston University, 2005.

F. Picard, E. Lebarbier, M. Hoebeke, B. Thiam, and S. Robin.
Joint segmentation, calling and normalization of multiple cgh profiles.
Technical Report 29, SSB, 2010.

F. Picard, S. Robin, M. Lavielle, C. Vaisse, and J-J. Daudin.
A statistical approach for CGH microarray data analysis.
BMC Bioinformatics, 6:27, 2005.

Guillem Rigaill.
Pruned dynamic programming for optimal multiple change-point detection.
Technical report, arXiv:1004.0887v1, 2010.

D. Siegmund.
Confidence sets in change-point problems.
International Statistical Review, 56(1):31–48, 1988.

K.J. Worsley.
Confidence regions and tests for a change-point in a sequence of exponential family random variables.
Biometrika, 73(1):91–104, 1986.

Y.C. Yao and S.T. Au.
Least square estimation of a step function.
Sankhya, 3:370–381, 1989.

Achim Zeileis, Christian Kleiber, Walter Krämer, and Kurt Hornik.
Testing and dating of structural changes in practice.
Computational Statistics & Data Analysis, 44(1–2):109–123, 2003.

NR. Zhang and DO. Siegmund.
A modified bayes information criterion with applications to the analysis of comparative genomic hybridization data.
Biometrics, 63(1):22–32, 2007.