Unsupervised Event Detection in Time Series: Old and New Approaches

Vincent Runge

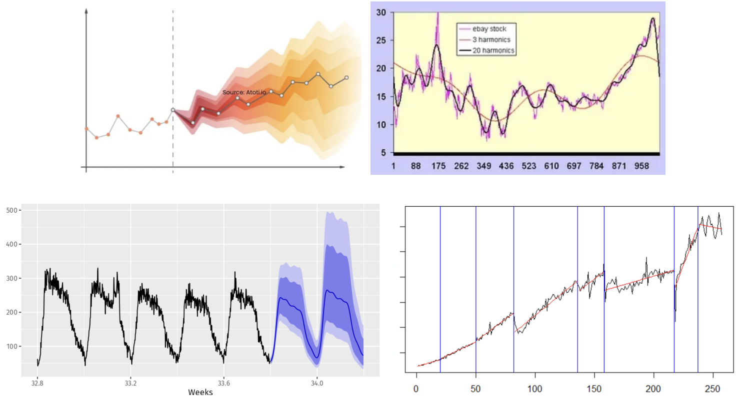
Panorama de la Recherche. ENS Paris-Saclay

November 5th 2025

Outlines

Outlines

Time series: prediction VS analysis

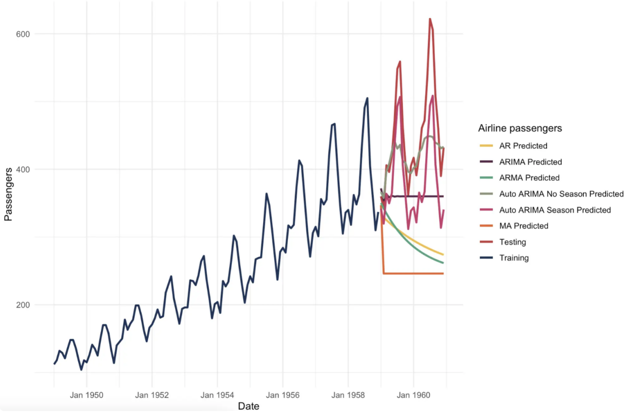


Prediction/forecasting (left) – analysis (right)

Time series: prediction (example)

An **ARMA**(p,q) model (*AutoRegressive Moving Average*) is a stationary time series model combining past values and past random shocks.

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q},$$



example of forcasting model resutls

Time series: Approximation VS Event Detection

A chatGPT explanation!

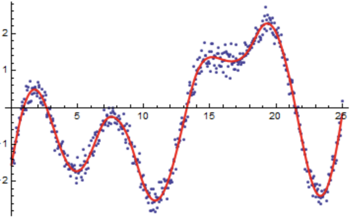
Approximation (Signal Representation)

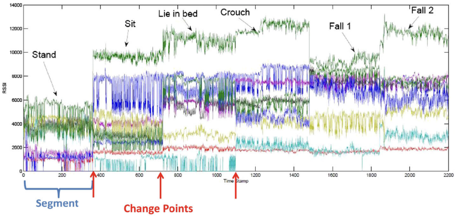
- Focus on smooth or dominant components (trend, cycles, autocorrelation)
- Methods: smoothing filters, splines, kernel regression, spectral analysis
- Goal: reduce noise and reveal the main dynamics

Event Detection (Structural Changes / Anomalies)

- Focus on abrupt changes, peaks, outliers, regime shifts
- Methods: change-point detection, wavelets, residual analysis, thresholding tests
- Goal: locate moments where the process changes

Time series: Approximation VS Event Detection





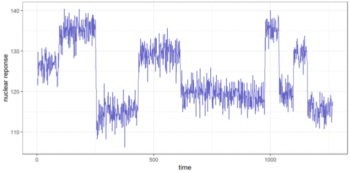
Approximation by smoothing splines (left) – change-point detection analysis (right)

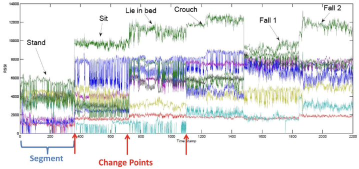
Outlines

The Multiple Change-Point Problem (in pictures!!!)

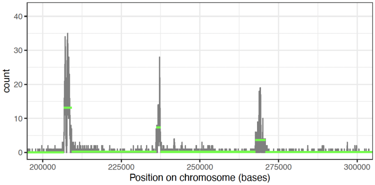
✓ in time-series analysis

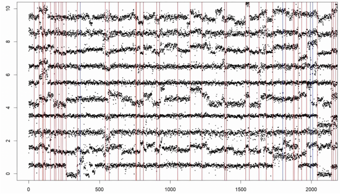
➜ find a structure in segments separated by change points





well-log geological data (left) sensor motion data (Zameni et al. 2019) (right)



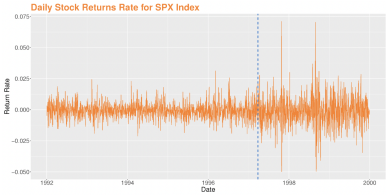


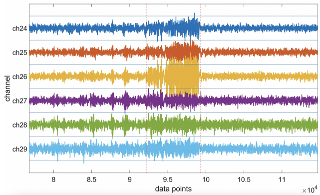
ChIP-seq count data (Hocking et al. 2022) (left) Array Comparative Genomic Hybridization (aCGH) data (Zhang et al. 2024)

The Multiple Change-Point Problem (in pictures!!!)

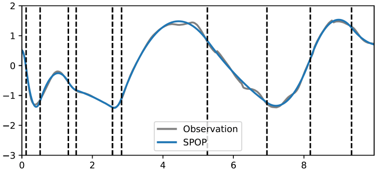
✓ in time-series analysis

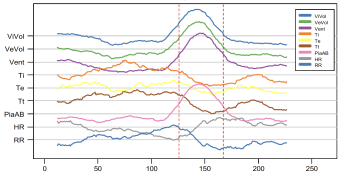
➜ find a structure in segments separated by change points





Financial data (Arrouch et al. 2023) (left) EEG data (Liu et al. 2025) (right)



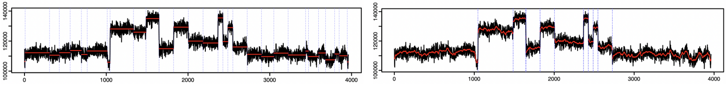


Motion capture data (Cecchi et al. 2025) (left) CO2 inhalation data (cabrieto et al. 2021) (right)

Ingredient No. 1 For Change-Point Detection

We observe a time series $y_1, \dots, y_n$ with $y_t \in \mathbb{R}$ or $\in \mathbb{R}^p$

1. Data model: $Y_t \sim \mathcal{L}(\theta_t)\,, \quad t = 1, \dots, n$



Well-log data. Change-in-mean (left) VERSUS random walk - AR(1) models (right) *Romano et al. (2022)*

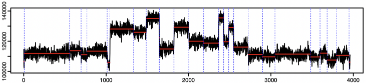
➜ Choice: piece-wise constant $t \mapsto \theta_t \in \mathbb{R}$

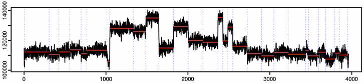
➜ Choice: Gaussian segment cost $\mathcal{C}$

$$\mathcal{C}(y_{ab}) = \min_{\theta \in \mathbb{R}} \mathcal{C}(y_{ab}, \theta) = \min_{\theta \in \mathbb{R}} \left(\sum_{t=a+1}^{b} (y_t - \theta)^2 \right) = \sum_{t=a+1}^{b} (y_t - mean(y_{ab}))^2$$

Ingredient No. 2 For Change-Point Detection

2. Calibration / hyperparameters / penalization / model selection





Well-log data. Change-in-mean model. standard BIC penalty (left) VERSUS inflated BIC penalty (right) *Romano et al. (2022)*

➜ Choice: Linear penalty $\beta > 0$

$$Q_n(\beta) = \min_{\tau \in S_n} \left[\sum_{i=0}^{K-1} \{\mathcal{C}(y_{\tau_i \tau_{i+1}}) + \beta\} \right]$$

$$S_n = \{\tau \in \mathbb{N}^{K+1}, \ 0 = \tau_0 < \tau_1 < \cdots < \tau_{K-1} < \tau_K = n\}$$

Statistic criterion ?

Penalized cost by $\beta > 0$

$$Q_n(\beta) = \min_{\tau \in S_n} \left[\sum_{i=0}^{K-1} \{ \mathcal{C}(y_{\tau_i \tau_{i+1}}) + \beta \} \right]$$

$\Rightarrow$ Choose the right β

$\beta = 2\hat{\sigma}^2 \log(n)$ Yao & Au (1989)

Fixed model size $= K$

$$Q_n^K = \min_{\tau \in S_n} \left[\sum_{i=0}^{K-1} \{ \mathcal{C}(y_{\tau_i \tau_{i+1}}) \} \right]$$

$\Rightarrow$ Choose the right integer K

$$\min_{K=1,\dots,M} \{ Q_n^K + Penalty(K) \}$$

Penalty $=$ slope heuristic, mBIC,...

Statistic criterion ?

Penalized cost by $\beta > 0$

$$Q_n(\beta) = \min_{\tau \in S_n} \left[\sum_{i=0}^{K-1} \{ \mathcal{C}(y_{\tau_i \tau_{i+1}}) + \beta \} \right]$$

$\Rightarrow$ Choose the right β

$\beta = 2\hat{\sigma}^2 \log(n)$ Yao & Au (1989)

Fixed model size $= K$

$$Q_n^K = \min_{\tau \in S_n} \left[\sum_{i=0}^{K-1} \{ \mathcal{C}(y_{\tau_i \tau_{i+1}}) \} \right]$$

$\Rightarrow$ Choose the right integer K

$$\min_{K=1,\dots,M} \{ Q_n^K + Penalty(K) \}$$

Penalty = slope heuristic, mBIC,...

References :

- Guyon and Yao (1999)
- Baraud, Giraud, Huet et al. (2009)
- Lebarbier (2005)
- Zhang and Siegmund (2007)
- Birge & Massart (2007)
- Verzelen (2023)...

Our CHALLENGE in Change-Point Detection

1. Solving the optimization problem in practice!

$$Q_n = \min_{\tau \in S_n} \left[\sum_{i=0}^{K-1} \{ \mathcal{C}(y_{\tau_i \tau_{i+1}}) + \beta \} \right]$$

How to efficiently compute Q_n? (2^{n-1} different segmentations)

There are 2^{n-1} different segmentations
$\Rightarrow$ Combinatorial explosion ! **Question:** How to get the best one ?

Find the global minimum? or accept easily computable approximate solution ?

$\Rightarrow$ battle : $\boxed{\text{Exactness vs. Time complexity}}$

Outlines

Change-point detection

Two main approaches

- Binary segmentation (BS):

Change-point detection

Two main approaches

- Binary segmentation (BS):
  - Status : approximate
  - Time complexity : $O(n \log(n))$

Change-point detection

Two main approaches

- Binary segmentation (BS):
  - Status : approximate
  - Time complexity : $O(n\log(n))$
- Dynamic programming (OP):

Change-point detection

Two main approaches

- Binary segmentation (BS):
  - Status : approximate
  - Time complexity : $O(n \log(n))$
- Dynamic programming (OP):
  - Status : exact
  - Time complexity : $O(n^2)$

Binary Segmentation

Scott and Knott (1974)

Principle

- Iteratively apply a **single change-point method**
- Stop when a **condition** is met

Binary Segmentation

Scott and Knott (1974)

Principle

- Iteratively apply a **single change-point method**
- Stop when a **condition** is met

Example : $y_{1:n}$ and $B > 0$ given

- (1) $\{a, b\} = \{1, ..., n\}$
- (2) find τ^* :

$$\tau^* = \underset{\tau \in \{a, \dots b-1\}}{\mathrm{Argmin}} \left(\mathcal{C}(y_{a:\tau}) + \mathcal{C}(y_{\tau:b})\right)$$

- (3) **if** $\mathcal{C}(y_{a:\tau^*}) + \mathcal{C}(y_{\tau^*:b}) + B < \mathcal{C}(y_{a:b})$
  then repeat (2) with $\{a, b\} = \{a, ..., \tau^*\}$
  and with $\{a, b\} = \{\tau^*, ..., b\}$

Dynamic programming
References

- Bellman and Dreyfus (1962)
- Segment Neighbourhood method. Auger and Lawrence (1989)
- Optimal Partitioning. Yao (1984), Jackson et al. (2005)

Dynamic programming
References

- Bellman and Dreyfus (1962)
- Segment Neighbourhood method. Auger and Lawrence (1989)
- Optimal Partitioning (OP). Yao (1984), Jackson et al. (2005)

Dynamic Programming for Change-Point Detection

Bellman and Dreyfus (1962) Yao (1984), Jackson et al. (2005)

Principle

(1) Solve sub-problems: find the last change point in y_{0t}, $t = 1, \dots, n$.
(2) Backtrack a global solution

$$Q_t = \min_{\tau \in S_t} \left[\sum_{i=0}^{K-1} \{ \mathcal{C}(y_{\tau_i \tau_{i+1}}) + \beta \} \right]$$

Dynamic Programming for Change-Point Detection

Bellman and Dreyfus (1962) Yao (1984), Jackson et al. (2005)

Principle

(1) Solve sub-problems: find the last change point in y_{0t}, $t = 1, \dots, n$.
(2) Backtrack a global solution

$$Q_t = \min_{\tau \in S_t} \left[\sum_{i=0}^{K-1} \{ \mathcal{C}(y_{\tau_i \tau_{i+1}}) + \beta \} \right]$$

$$Q_t = \min_{\tau \in S_t} \left[\sum_{i=0}^{K-2} \{ \mathcal{C}(y_{\tau_i \tau_{i+1}}) + \beta \} + \mathcal{C}(y_{\tau_{K-1} t}) + \beta \right]$$

Dynamic Programming for Change-Point Detection

Bellman and Dreyfus (1962) Yao (1984), Jackson et al. (2005)

Principle

(1) Solve sub-problems: find the last change point in y_{0t}, $t = 1, \dots, n$.
(2) Backtrack a global solution

$$Q_t = \min_{\tau \in S_t} \left[\sum_{i=0}^{K-1} \{ \mathcal{C}(y_{\tau_i \tau_{i+1}}) + \beta \} \right]$$

$$Q_t = \min_{\tau \in S_t} \left[\sum_{i=0}^{K-2} \{ \mathcal{C}(y_{\tau_i \tau_{i+1}}) + \beta \} + \mathcal{C}(y_{\tau_{K-1} t}) + \beta \right]$$

$$Q_t = \min_{s \in \{0,\dots,t-1\}} \Big(Q_s + \mathcal{C}(y_{st}) + \beta \Big)$$

Dynamic Programming for Change-Point Detection

$$Q_n = \min_{\tau \in S_n} \left[\sum_{i=0}^{K-1} \{ \mathcal{C}(y_{\tau_i \tau_{i+1}}) + \beta \} \right] \iff Q_t = \min_{s \in \{0,\dots,t-1\}} \left(Q_s + \mathcal{C}(y_{st}) + \beta \right)$$

Optimal Partitioning (OP)

$Q_0 = 0$
for $t = 1$ to n **do**
    $Q_t = \min_{s \in \{0,\dots,t-1\}} \{ Q_s + \mathcal{C}(y_{st}) + \beta \}$
    $cp(t) = \text{Argmin}_{s \in \{0,\dots,t-1\}} \{ Q_s + \mathcal{C}(y_{st}) + \beta \}$
end for

$$(1,\ cp(2),\ cp(3), \dots\dots\dots,\ \underset{= \cdots}{cp(u)},\ cp(u+1), \dots\dots,\ \underset{= u}{cp(v)},\ cp(v+1), \dots\dots, \underset{= v}{cp(n)})$$

➜ algorithm in $O(n^2)$ time complexity

Time complexity CHALLENGE

4. Efficiently Solving the OP algorithm using pruning rules

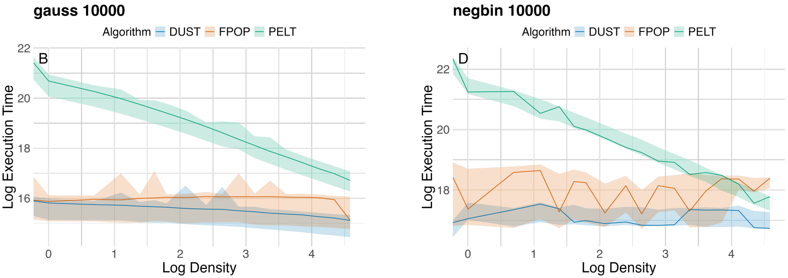
$$Q_n = \min_{\tau \in S_n} \left[\sum_{i=0}^{K-1} \{\mathcal{C}(y_{\tau_i \tau_{i+1}}) + \beta\} \right] \quad \Longleftrightarrow \quad Q_t = \min_{s \in \mathbb{I}_t} \left(Q_s + \mathcal{C}(y_{st}) + \beta \right)$$

➜ OP - PELT - FPOP - DUST algorithms solve the same problem EXACTLY but with different index sets $\mathbb{I}_t$.

Time complexity Challenge

$$Q_n = \min_{\tau \in S_n} \left[\sum_{i=0}^{K-1} \{ \mathcal{C}(y_{\tau_i \tau_{i+1}}) + \beta \} \right] \iff Q_t = \min_{s \in \mathbb{I}_t} \left(Q_s + \mathcal{C}(y_{st}) + \beta \right) \quad t = 1, \dots, n$$

➜ **OP - PELT - FPOP - DUST** algorithms solve the same problem EXACTLY but with different index sets $\mathbb{I}_t$.



no change $n = 10^6$	Gauss	Poisson	negbin
FPOP (fpopw, gfpop, gfpop)	0.32s	6.1s	8.2s
DUST	0.20s	0.60s	1.2s

Outlines

Functional Representation

$$Q_t = \min_{s \in \{0,\dots,t-1\}} \left(Q_s + \min_{\theta} \mathcal{C}(y_{st}, \theta) + \beta \right)$$

Functional Representation

$$Q_t = \min_{s \in \{0,\ldots,t-1\}} \left(Q_s + \min_{\theta} \mathcal{C}(y_{st}, \theta) + \beta \right)$$

$$Q_t = \min_{\theta} \min_{s \in \{0,\ldots,t-1\}} \left(Q_s + \mathcal{C}(y_{st}, \theta) + \beta \right)$$

Functional Representation

$$Q_t = \min_{s \in \{0,\dots,t-1\}} \left(Q_s + \min_{\theta} \mathcal{C}(y_{st}, \theta) + \beta \right)$$

$$Q_t = \min_{\theta} \min_{s \in \{0,\dots,t-1\}} \left(Q_s + \mathcal{C}(y_{st}, \theta) + \beta \right)$$

FPOP $\quad Q_t = \min_{\theta} Q_t(\theta)$

Functional Representation

$$Q_t = \min_{s \in \{0,\ldots,t-1\}} \left(Q_s + \min_{\theta} \mathcal{C}(y_{st}, \theta) + \beta \right)$$

$$Q_t = \min_{\theta} \min_{s \in \{0,\ldots,t-1\}} \left(Q_s + \mathcal{C}(y_{st}, \theta) + \beta \right)$$

FPOP $\qquad Q_t = \min_{\theta} Q_t(\theta)$

$$\begin{aligned} Q_t(\theta) &= \min_{s \in \{0,\ldots,t-1\}} \left(Q_s + \mathcal{C}(y_{st}, \theta) + \beta \right) \\ &= \min_{s \in \{0,\ldots,t-1\}} \left(q_t^s(\theta) \right) \end{aligned}$$

Functional Representation: Pruning Principle

$$Q_n = \min_{\tau \in S_n} \left[\sum_{i=0}^{K-1} \{ \mathcal{C}(y_{\tau_i \tau_{i+1}}) + \beta \} \right] \iff Q_t(\theta) = \min_{s \in \mathbb{I}_t} \Big(Q_s + \mathcal{C}(y_{st}, \theta) + \beta \Big) = \min_{s \in \mathbb{I}_t} \Big(q_t^s(\theta) \Big)$$

Functional Representation: Pruning Principle

$$Q_n = \min_{\tau \in S_n} \left[\sum_{i=0}^{K-1} \{ \mathcal{C}(y_{\tau_i \tau_{i+1}}) + \beta \} \right] \iff Q_t(\theta) = \min_{s \in \mathbb{I}_t} \left(Q_s + \mathcal{C}(y_{st}, \theta) + \beta \right) = \min_{s \in \mathbb{I}_t} \left(q_t^s(\theta) \right)$$

$$q_t^s(\theta) = \sum_{u=s+1}^{t} (y_u - \theta)^2 + Q_s + \beta$$

if $Q_t(\theta) < q_t^s(\theta)$ for all $\theta \in \mathbb{R}^p$, we remove s in $\mathbb{I}_t$ for all further iterations

Functional Representation: Pruning Principle

$$Q_n = \min_{\tau \in S_n} \left[\sum_{i=0}^{K-1} \{ \mathcal{C}(y_{\tau_i \tau_{i+1}}) + \beta \} \right] \iff Q_t(\theta) = \min_{s \in \mathbb{I}_t} \left(Q_s + \mathcal{C}(y_{st}, \theta) + \beta \right) = \min_{s \in \mathbb{I}_t} \left(q_t^s(\theta) \right)$$

$$q_t^s(\theta) = \sum_{u=s+1}^{t} (y_u - \theta)^2 + Q_s + \beta$$

if $Q_t(\theta) < q_t^s(\theta)$ for all $\theta \in \mathbb{R}^p$, we remove s in $\mathbb{I}_t$ for all further iterations

$$Q_{t+1}(\theta) \leq Q_t(\theta) + (y_{t+1} - \theta)^2 < q_t^s(\theta) + (y_{t+1} - \theta)^2 = q_{t+1}^s(\theta)$$

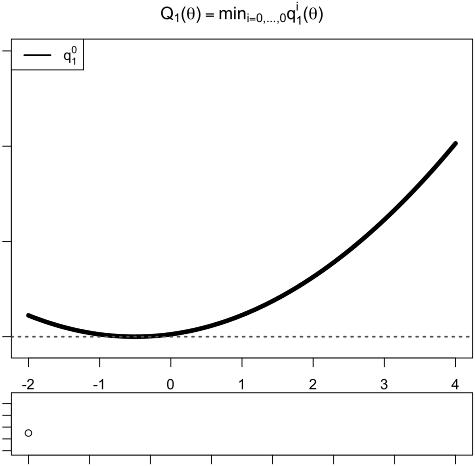
Challenges:

- Testing the indices s in $\mathbb{I}_t$ for covering ($Q_t(\theta) < q_t^s(\theta)$)
- Doing it quickly (very quickly!)

Functional Representation (in 1D)

$$Q_n = \min_{\tau \in S_n} \left[\sum_{i=0}^{K-1} \{ \mathcal{C}(y_{\tau_i \tau_{i+1}}) + \beta \} \right] \iff Q_t(\theta) = \min_{s \in \mathbb{I}_t} \left(Q_s + \mathcal{C}(y_{st}, \theta) + \beta \right) = \min_{s \in \mathbb{I}_t} \left(q_t^s(\theta) \right)$$

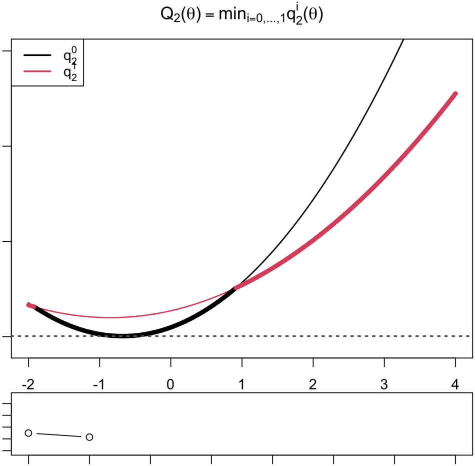
$$\mathbb{I}_1 = \{0\}$$



Functional Representation (in 1D)

$$Q_n = \min_{\tau \in S_n} \left[\sum_{i=0}^{K-1} \{ \mathcal{C}(y_{\tau_i \tau_{i+1}}) + \beta \} \right] \iff Q_t(\theta) = \min_{s \in \mathbb{I}_t} \left(Q_s + \mathcal{C}(y_{st}, \theta) + \beta \right) = \min_{s \in \mathbb{I}_t} \left(q_t^s(\theta) \right)$$

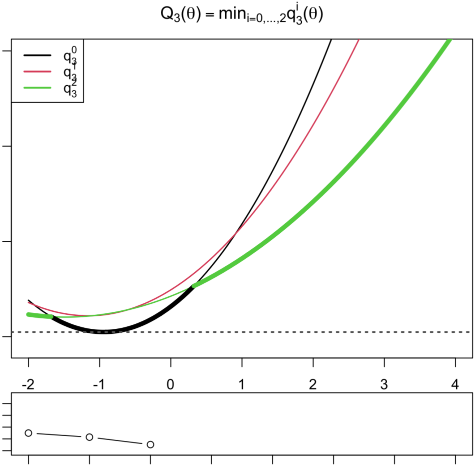
$\mathbb{I}_2 = \{0, 1\}$



Functional Representation (in 1D)

$$Q_n = \min_{\tau \in S_n} \left[\sum_{i=0}^{K-1} \{ \mathcal{C}(y_{\tau_i \tau_{i+1}}) + \beta \} \right] \iff Q_t(\theta) = \min_{s \in \mathbb{I}_t} \Big(Q_s + \mathcal{C}(y_{st}, \theta) + \beta \Big) = \min_{s \in \mathbb{I}_t} \Big(q_t^s(\theta) \Big)$$

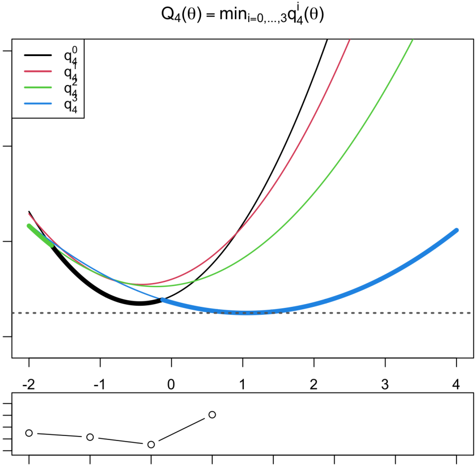
$$\mathbb{I}_3 = \{0, 2\}$$



Functional Representation (in 1D)

$$Q_n = \min_{\tau \in S_n} \left[\sum_{i=0}^{K-1} \{ \mathcal{C}(y_{\tau_i \tau_{i+1}}) + \beta \} \right] \iff Q_t(\theta) = \min_{s \in \mathbb{I}_t} \left(Q_s + \mathcal{C}(y_{st}, \theta) + \beta \right) = \min_{s \in \mathbb{I}_t} \left(q_t^s(\theta) \right)$$

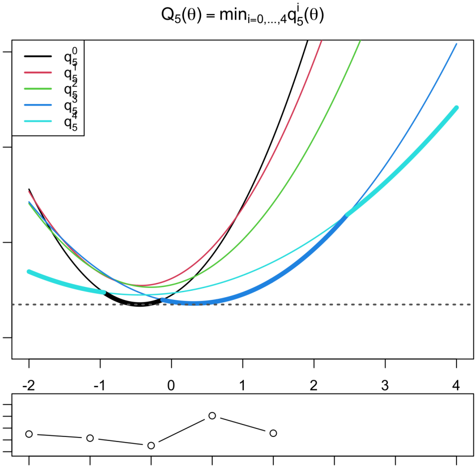
$$\mathbb{I}_4 = \{0, 2, 3\}$$



Functional Representation (in 1D)

$$Q_n = \min_{\tau \in S_n} \left[\sum_{i=0}^{K-1} \{ \mathcal{C}(y_{\tau_i \tau_{i+1}}) + \beta \} \right] \iff Q_t(\theta) = \min_{s \in \mathbb{I}_t} \left(Q_s + \mathcal{C}(y_{st}, \theta) + \beta \right) = \min_{s \in \mathbb{I}_t} \left(q_t^s(\theta) \right)$$

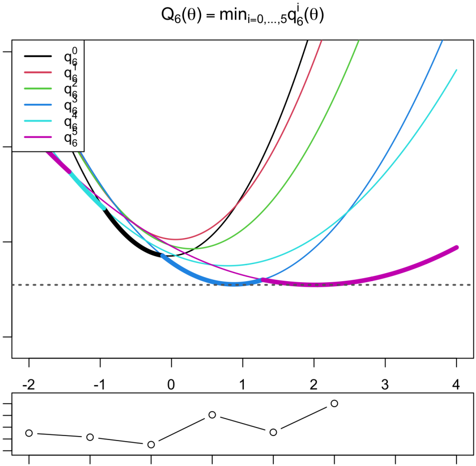
$\mathbb{I}_5 = \{0, 3, 4\}$



Functional Representation (in 1D)

$$Q_n = \min_{\tau \in S_n} \left[\sum_{i=0}^{K-1} \{ \mathcal{C}(y_{\tau_i \tau_{i+1}}) + \beta \} \right] \iff Q_t(\theta) = \min_{s \in \mathbb{I}_t} \left(Q_s + \mathcal{C}(y_{st}, \theta) + \beta \right) = \min_{s \in \mathbb{I}_t} \left(q_t^s(\theta) \right)$$

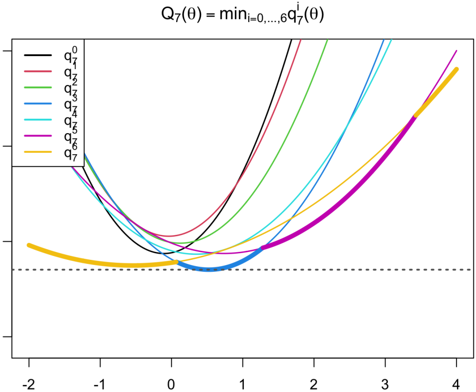
$\mathbb{I}_6 = \{0, 3, 4, 5\}$

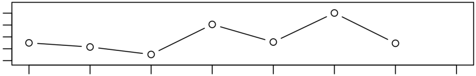


Functional Representation (in 1D)

$$Q_n = \min_{\tau \in S_n} \left[\sum_{i=0}^{K-1} \{ \mathcal{C}(y_{\tau_i \tau_{i+1}}) + \beta \} \right] \iff Q_t(\theta) = \min_{s \in \mathbb{I}_t} \left(Q_s + \mathcal{C}(y_{st}, \theta) + \beta \right) = \min_{s \in \mathbb{I}_t} \left(q_t^s(\theta) \right)$$

$\mathbb{I}_7 = \{3, 5, 6\}$

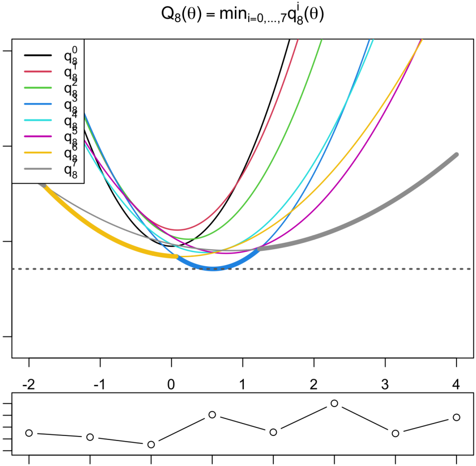




Functional Representation (in 1D)

$$Q_n = \min_{\tau \in S_n} \left[\sum_{i=0}^{K-1} \{ \mathcal{C}(y_{\tau_i \tau_{i+1}}) + \beta \} \right] \iff Q_t(\theta) = \min_{s \in \mathbb{I}_t} \left(Q_s + \mathcal{C}(y_{st}, \theta) + \beta \right) = \min_{s \in \mathbb{I}_t} \left(q_t^s(\theta) \right)$$

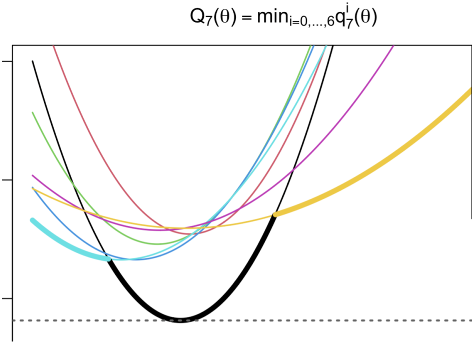
$\mathbb{I}_8 = \{3, 6, 7\}$



FPOP pruning VS. DUST pruning

FPOP

- Update couples (Intervals, q_t^i)
- From left to right
- Adapted to 1D Q_t functions



DUST:

- Solves an optimization problem on the q_t^i functions
- Adapted to multivariate Q_t functions

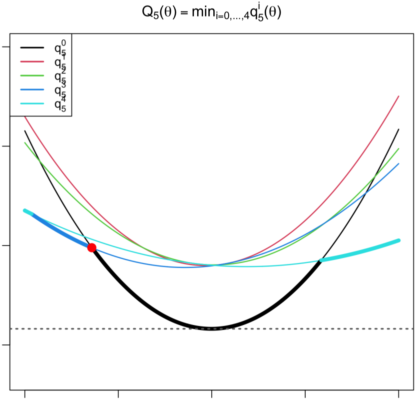
Outlines

DUST Pruning

$$Q_n = \min_{\tau \in S_n} \left[\sum_{i=0}^{K-1} \{ \mathcal{C}(y_{\tau_i \tau_{i+1}}) + \beta \} \right] \iff Q_t(\theta) = \min_{s \in \mathbb{I}_t} \left(Q_s + \mathcal{C}(y_{st}, \theta) + \beta \right) = \min_{s \in \mathbb{I}_t} \left(q_t^s(\theta) \right)$$

We solve for $r < s$:

$$\min_{\theta \in \mathbb{R}^p} q_t^s(\theta) \quad \text{under constraint} \quad q_t^s(\theta) - q_t^r(\theta) \leq 0$$



Example:

$$\min_{\theta \in \mathbb{R}^p} q_5^3(\theta)$$

under constraint:

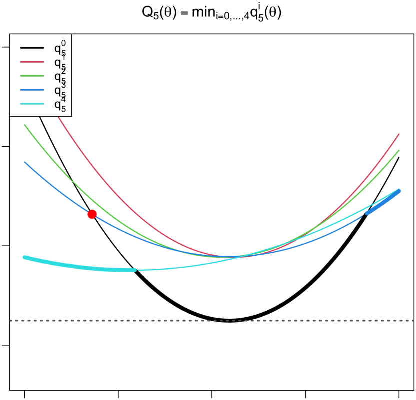
$$q_5^3(\theta) - q_5^0(\theta) \leq 0$$

DUST Pruning

$$Q_n = \min_{\tau \in S_n} \left[\sum_{i=0}^{K-1} \{ \mathcal{C}(y_{\tau_i \tau_{i+1}}) + \beta \} \right] \iff Q_t(\theta) = \min_{s \in \mathbb{I}_t} \Big(Q_s + \mathcal{C}(y_{st}, \theta) + \beta \Big) = \min_{s \in \mathbb{I}_t} \Big(q_t^s(\theta) \Big)$$

We solve for $r < s$:

$$\min_{\theta \in \mathbb{R}^p} q_t^s(\theta) \quad \text{under constraint} \quad q_t^s(\theta) - q_t^r(\theta) \leq 0$$



Example:

$$\min_{\theta \in \mathbb{R}^p} q_5^3(\theta)$$

under constraint:

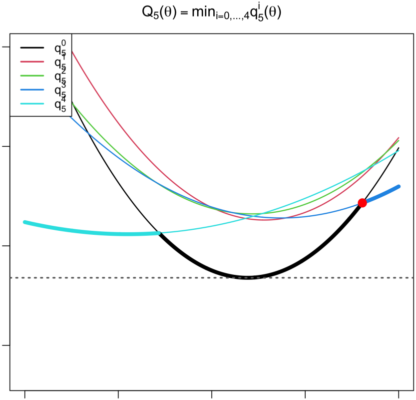
$$q_5^3(\theta) - q_5^0(\theta) \leq 0$$

DUST Pruning

$$Q_n = \min_{\tau \in S_n} \left[\sum_{i=0}^{K-1} \{ \mathcal{C}(y_{\tau_i \tau_{i+1}}) + \beta \} \right] \iff Q_t(\theta) = \min_{s \in \mathbb{I}_t} \Big(Q_s + \mathcal{C}(y_{st}, \theta) + \beta \Big) = \min_{s \in \mathbb{I}_t} \Big(q_t^s(\theta) \Big)$$

We solve for $r < s$:

$$\min_{\theta \in \mathbb{R}^p} q_t^s(\theta) \quad \text{under constraint} \quad q_t^s(\theta) - q_t^r(\theta) \leq 0$$



Example:

$$\min_{\theta \in \mathbb{R}^p} q_5^3(\theta)$$

under constraint:

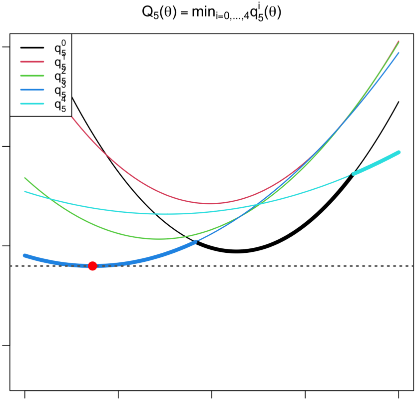
$$q_5^3(\theta) - q_5^0(\theta) \leq 0$$

DUST Pruning

$$Q_n = \min_{\tau \in S_n} \left[\sum_{i=0}^{K-1} \{\mathcal{C}(y_{\tau_i \tau_{i+1}}) + \beta\} \right] \iff Q_t(\theta) = \min_{s \in \mathbb{I}_t} \Big(Q_s + \mathcal{C}(y_{st}, \theta) + \beta \Big) = \min_{s \in \mathbb{I}_t} \Big(q_t^s(\theta) \Big)$$

We solve for $r < s$:

$$\min_{\theta \in \mathbb{R}^p} q_t^s(\theta) \quad \text{under constraint} \quad q_t^s(\theta) - q_t^r(\theta) \leq 0$$



Example:

$$\min_{\theta \in \mathbb{R}^p} q_5^3(\theta)$$

under constraint:

$$q_5^3(\theta) - q_5^0(\theta) \leq 0$$

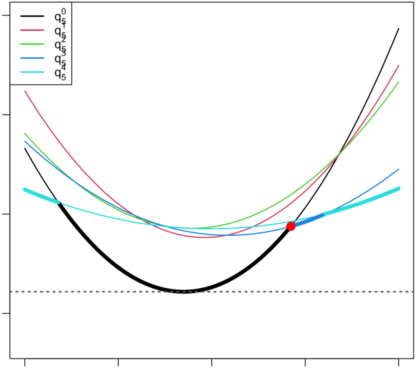
DUST Pruning

$$Q_n = \min_{\tau \in S_n} \left[\sum_{i=0}^{K-1} \{\mathcal{C}(y_{\tau_i \tau_{i+1}}) + \beta\} \right] \iff Q_t(\theta) = \min_{s \in \mathbb{I}_t} \Big(Q_s + \mathcal{C}(y_{st}, \theta) + \beta \Big) = \min_{s \in \mathbb{I}_t} \Big(q_t^s(\theta) \Big)$$

We solve for $r < s$:

$$\min_{\theta \in \mathbb{R}^p} q_t^s(\theta) \quad \text{under constraint} \quad q_t^s(\theta) - q_t^r(\theta) \le 0$$

$Q_5(\theta) = \min_{i=0,\ldots,4} q_5^i(\theta)$



Example:

$$\min_{\theta \in \mathbb{R}^p} q_5^3(\theta)$$

under constraint:

$$q_5^3(\theta) - q_5^0(\theta) \le 0$$

DUST Pruning

We solve the non-convex optimization problem ($r < s$):

$$\min_{\theta \in \mathbb{R}^p} q_t^s(\theta) \quad \text{under constraint} \quad q_t^s(\theta) - q_t^r(\theta) \leq 0$$

DUST Pruning

We solve the non-convex optimization problem ($r < s$):

$$\min_{\theta \in \mathbb{R}^p} q_t^s(\theta) \quad \text{under constraint} \quad q_t^s(\theta) - q_t^r(\theta) \leq 0$$

Primal problem

$$p_t^{rs}(\theta) = q_t^s(\theta) + \mu\Big(q_t^s(\theta) - q_t^r(\theta)\Big)$$

DUST Pruning

We solve the non-convex optimization problem ($r < s$):

$$\min_{\theta \in \mathbb{R}^p} q_t^s(\theta) \quad \text{under constraint} \quad q_t^s(\theta) - q_t^r(\theta) \leq 0$$

Primal problem

$$p_t^{rs}(\theta) = q_t^s(\theta) + \mu\Big(q_t^s(\theta) - q_t^r(\theta)\Big)$$

Dual problem

$$d_t^{rs}(\mu) = q_t^s(\theta^*(\mu)) + \mu\Big(q_t^s(\theta^*(\mu)) - q_t^r(\theta^*(\mu))\Big)$$

DUST Pruning

We solve the non-convex optimization problem ($r < s$):

$$\min_{\theta \in \mathbb{R}^p} q_t^s(\theta) \quad \text{under constraint} \quad q_t^s(\theta) - q_t^r(\theta) \leq 0$$

Primal problem

$$p_t^{rs}(\theta) = q_t^s(\theta) + \mu\Big(q_t^s(\theta) - q_t^r(\theta)\Big)$$

Dual problem

$$d_t^{rs}(\mu) = q_t^s(\theta^*(\mu)) + \mu\Big(q_t^s(\theta^*(\mu)) - q_t^r(\theta^*(\mu))\Big)$$

Result

$$d_t^{rs}(\mu) \leq \text{solution} \quad \text{for all} \quad \mu \in [0, \mu_{max})$$

Evaluation of the dual: always less than the optimum under constraint

DUST Pruning: theory

- DUST Pruning Remove k in $\mathbb{I}_t$ if

$$d_t^{rs}(\mu_0) \geq Q_t + \beta$$

Proposition

No duality gap: $\max_{\mu \in [0,\mu_{max})} d_t^{rs}(\mu) =$ solution $=$ red point 

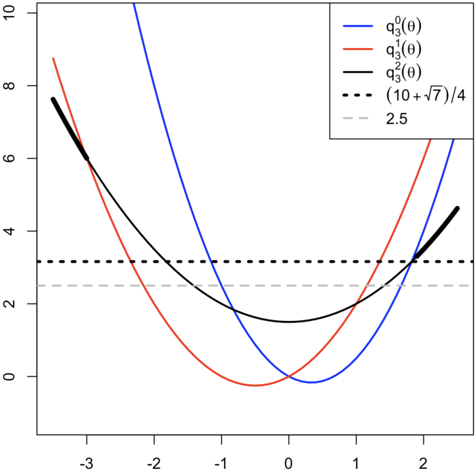
- With a cost from the exponential family (function A strictly convex):

$$\mathcal{C}(y_{st}, \theta) = (t-s)A(\theta) - \theta \cdot \mathbf{S}_{s+1\dots t}$$

➜ True even for $\theta \in \mathbb{R}^p$

Also true for $d \leq p$ constraints, in that case the dual is $d_t^{(j_1,\dots,j_d)s}(\mu \in \mathbb{R}^d)$

DUST Pruning: theory



DUST Pruning: dual shape

For simplicity, we present the one-constraint, one-parameter case first!

DUST Pruning: dual shape

$$\mathcal{C}(y_{st}, \theta) = (t-s)\Big(A(\theta) - \theta \cdot \overline{y_{st}}\Big)$$

We consider the test:

$$(\star) = \frac{\mathcal{D}(\mu) - (Q_t + \beta)}{t-s} > 0$$

Proposition

At time t, the index s is pruned by index r $(r < s < t)$ when we find $\mu \in [0, \mu_{max}]$ (with $\mu_{max} \leq 1$) such that:

$$\boxed{(\star) = -(1-\mu)\mathcal{D}^*\left(\frac{\overline{y_{st}} - \mu \overline{y_{rs}}}{1-\mu}\right) + \mu \frac{Q_s - Q_r}{s-r} - \frac{Q_t - Q_s}{t-s} > 0}$$

with strictly convex function $\mathcal{D}^*(x) = x \cdot (\nabla A)^{-1}(x) - A((\nabla A)^{-1}(x))$

DUST Pruning: dual shape - Examples

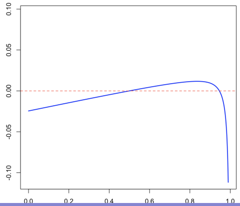
Gauss:

$$\mathcal{C}(y_{st}, \theta) = (t-s)\left(A(\theta) - \theta \cdot \overline{y_{st}}\right) = (t-s)\left(\frac{\theta^2}{2} - \theta \cdot \overline{y_{st}}\right)$$

Proposition

At time t, the index s is pruned by index r ($r < s < t$) when we find $\mu \in [0, \mu_{max}]$ (with $\mu_{max} \leq 1$) such that:

$$(\star) = -\frac{1-\mu}{2}\left(\frac{\overline{y_{st}} - \mu \overline{y_{rs}}}{1-\mu}\right)^2 + \mu \frac{Q_s - Q_r}{s-r} - \frac{Q_t - Q_s}{t-s} > 0$$



DUST Pruning: dual shape - Examples

Poisson:

$$\mathcal{C}(y_{st}, \theta) = (t-s)\Big(A(\theta) - \theta \cdot \overline{y_{st}}\Big) = (t-s)\Big(\exp\theta - \theta \cdot \overline{y_{st}}\Big)$$

Proposition

At time t, the index s is pruned by index r ($r < s < t$) when we find $\mu \in [0, \mu_{max}]$ (with $\mu_{max} \leq 1$) such that:

$$(\star) =$$
$$-(1-\mu)\left(\frac{\overline{y_{st}} - \mu\overline{y_{rs}}}{1-\mu}\right)\left(\log\left(\frac{\overline{y_{st}} - \mu\overline{y_{rs}}}{1-\mu}\right) - 1\right) + \mu\frac{Q_s - Q_r}{s-r} - \frac{Q_t - Q_s}{t-s} > 0$$

Dual Function To Decision Function

We normalise again the dual test $(\star)$:

$$(\star) = \frac{\mathcal{D}(\mu) - (Q_t + \beta)}{t - s} > 0 \iff \frac{\mathcal{D}(\mu) - (Q_t + \beta)}{(t-s)(1-\mu)} > 0$$

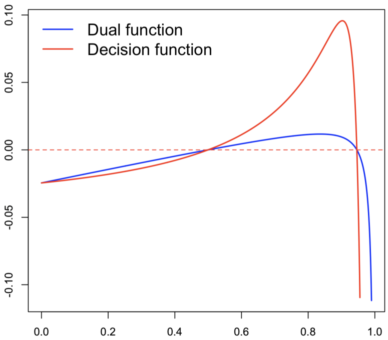
and $\mu = \frac{x}{1+x}$ and we get a simpler test:

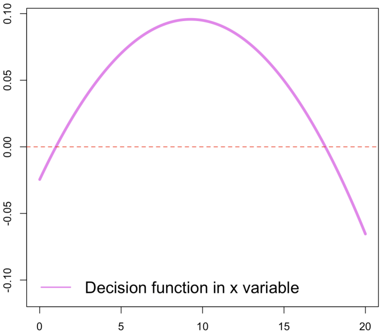
Theorem

At time t, the index s is pruned by index r ($r < s < t$) when we find $x \in [0, x_{max}]$ such that:

$$-\mathcal{D}^*\left(\overline{y_{st}} + x(\overline{y_{st}} - \overline{y_{rs}})\right) - \left(\frac{Q_t - Q_s}{t - s} + x\left(\frac{Q_t - Q_s}{t - s} - \frac{Q_s - Q_r}{s - r}\right)\right) > 0$$

Dual Function To Decision Function





$$\frac{\mathcal{D}(\mu) - (Q_t + \beta)}{t - s}$$

$$\frac{\mathcal{D}(\mu) - (Q_t + \beta)}{(t - s)(1 - \mu)}$$

$$\frac{\mathcal{D}(\frac{x}{1+x}) - (Q_t + \beta)}{(t - s)(1 - \frac{x}{1+x})}$$

Dual Function To Decision Function

Proposition

At time t, the index s is pruned by index r ($r < s < t$) with the following inequality test if the maximum is inside its domain (otherwise we evaluate at the boundary value):

$$d_t^{rs}(\mu^*) \geq Q_t + \beta \iff q_t^s\left(-\frac{\frac{Q_t - Q_s}{t-s} - \frac{Q_s - Q_r}{s-r}}{\overline{y_{st}} - \overline{y_{rs}}}\right) > Q_t + \beta$$

Dual Function To Decision Function

Proposition

At time t, the index s is pruned by index r ($r < s < t$) with the following inequality test if the maximum is inside its domain (otherwise we evaluate at the boundary value):

$$d_t^{rs}(\mu^*) \geq Q_t + \beta \iff q_t^s\left(-\frac{\frac{Q_t - Q_s}{t-s} - \frac{Q_s - Q_r}{s-r}}{\overline{y_{st}} - \overline{y_{rs}}}\right) > Q_t + \beta$$

Proof: We differentiate the decision function:

$$-\mathcal{D}^*\left(\overline{y_{st}} + x(\overline{y_{st}} - \overline{y_{rs}})\right) - \left(\frac{Q_t - Q_s}{t - s} + x\left(\frac{Q_t - Q_s}{t - s} - \frac{Q_s - Q_r}{s - r}\right)\right)$$

Dual Function To Decision Function

Proposition

At time t, the index s is pruned by index r ($r < s < t$) with the following inequality test if the maximum is inside its domain (otherwise we evaluate at the boundary value):

$$d_t^{rs}(\mu^*) \geq Q_t + \beta \iff q_t^s\left(-\frac{\frac{Q_t - Q_s}{t-s} - \frac{Q_s - Q_r}{s-r}}{\overline{y_{st}} - \overline{y_{rs}}}\right) > Q_t + \beta$$

Proof: We differentiate the decision function:

$$-\mathcal{D}^*\Big(\overline{y_{st}} + x(\overline{y_{st}} - \overline{y_{rs}})\Big) - \left(\frac{Q_t - Q_s}{t-s} + x\left(\frac{Q_t - Q_s}{t-s} - \frac{Q_s - Q_r}{s-r}\right)\right)$$

Comment: the PELT test is

$$q_t^s\Big((\nabla A)^{-1}(\overline{y_{st}})\Big) > Q_t + \beta$$

Outlines

DUST algorithm

For $t = 1$ to n:

- Optimal Partitioning EXPLORE s in $\mathbb{I}_t$

$$Q_s + \mathcal{C}(y_{st}) + \beta = d_t^{rs}(0) = q_t^s(\theta^*(0)) + 0\Big(q_t^s(\theta^*(0)) - q_t^r(\theta^*(0))\Big)$$

We find $Q_t = \min_{s \in \mathbb{I}_t} \left\{ d_t^{rs}(0) \right\}$ (dynamic programming)

DUST algorithm

For $t = 1$ to n:

- Optimal Partitioning EXPLORE s in $\mathbb{I}_t$

$$Q_s + \mathcal{C}(y_{st}) + \beta = d_t^{rs}(0) = q_t^s(\theta^*(0)) + 0\Big(q_t^s(\theta^*(0)) - q_t^r(\theta^*(0))\Big)$$

We find $Q_t = \min_{s \in \mathbb{I}_t} \left\{ d_t^{rs}(0) \right\}$ (dynamic programming)

- DUST Pruning REMOVE k in $\mathbb{I}_t$ if

$$d_t^{rs}(\mu^*) \geq Q_t + \beta$$

r is uniformly sampled in $\mathbb{I}_t$ with $r < s$

DUST algorithm

For $t = 1$ to n:

- Optimal Partitioning EXPLORE s in $\mathbb{I}_t$

$$Q_s + \mathcal{C}(y_{st}) + \beta = d_t^{rs}(0) = q_t^s(\theta^*(0)) + 0\Big(q_t^s(\theta^*(0)) - q_t^r(\theta^*(0))\Big)$$

We find $Q_t = \min_{s \in \mathbb{I}_t} \left\{ d_t^{rs}(0) \right\}$ (dynamic programming)

- DUST Pruning REMOVE k in $\mathbb{I}_t$ if

$$d_t^{rs}(\mu^*) \geq Q_t + \beta$$

r is uniformly sampled in $\mathbb{I}_t$ with $r < s$

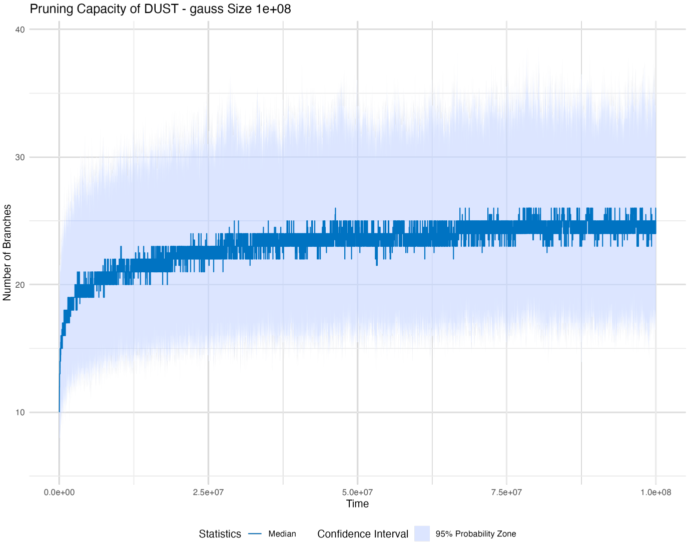
Exact max evaluation:

If the maximum is within the domain $(0, \mu_{max})$:

$$d_t^{rs}(\mu^*) \geq Q_t + \beta \iff q_t^s\left(- \frac{\frac{Q_t - Q_s}{t-s} - \frac{Q_s - Q_r}{s-r}}{\overline{y_{st}} - \overline{y_{rs}}} \right) > Q_t + \beta$$

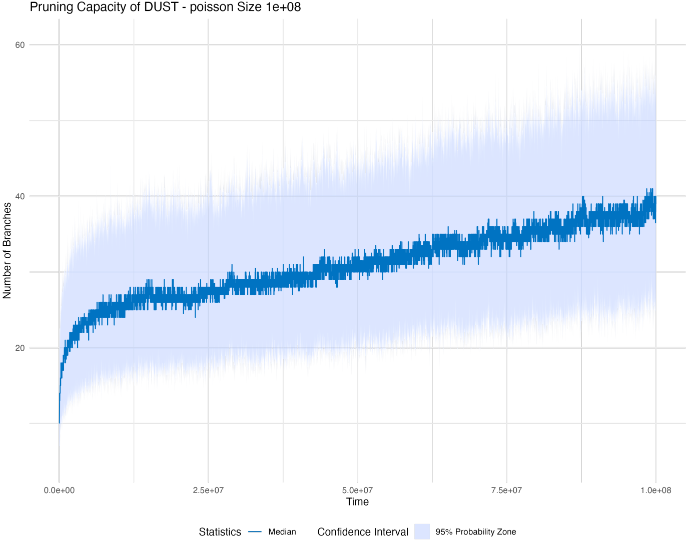
Simulations

Size of $\mathbb{I}_t$ (Gaussian model. No change)



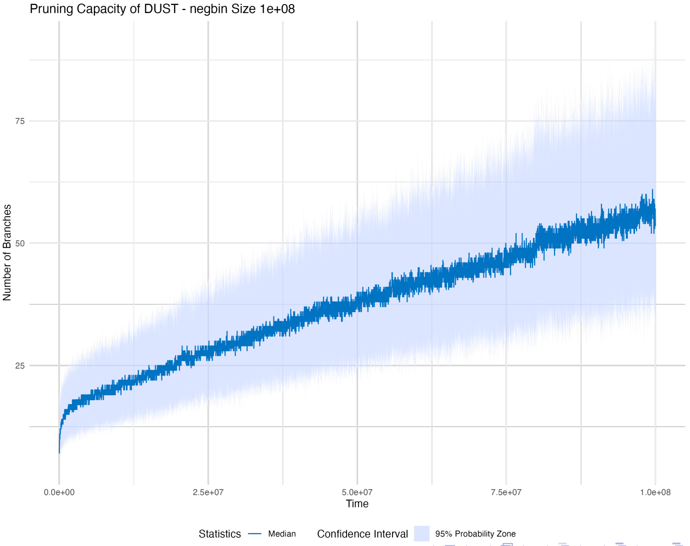
Simulations

Size of $\mathbb{I}_t$ (Poisson model. No change)

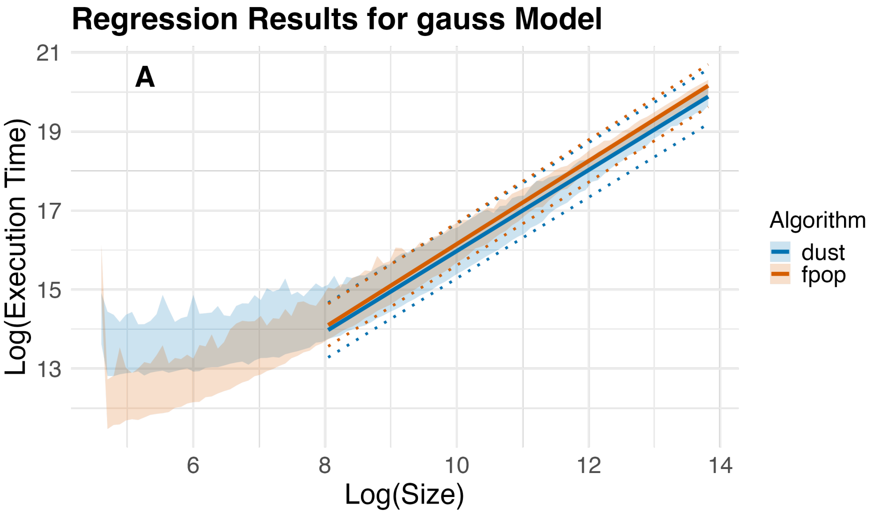


Simulations

Size of $\mathbb{I}_t$ (Negative Binomial model. No change)

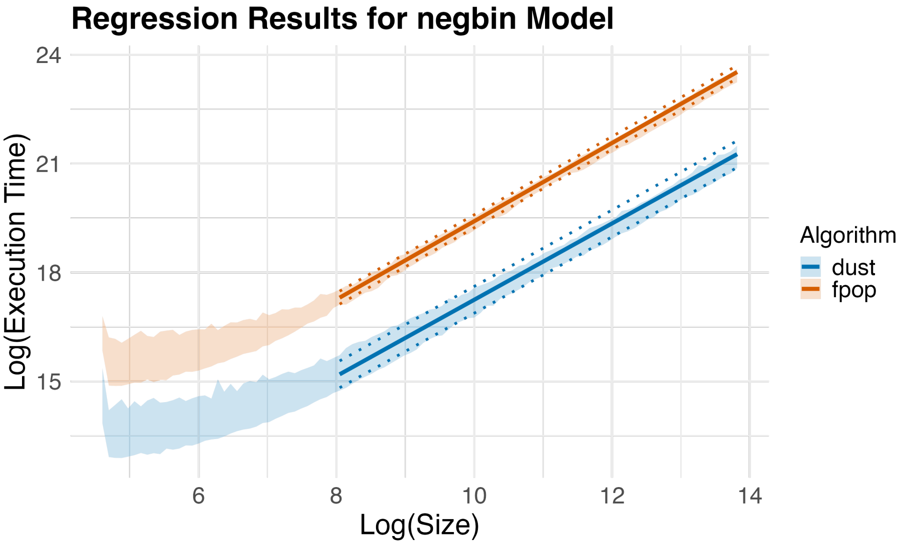


Simulations

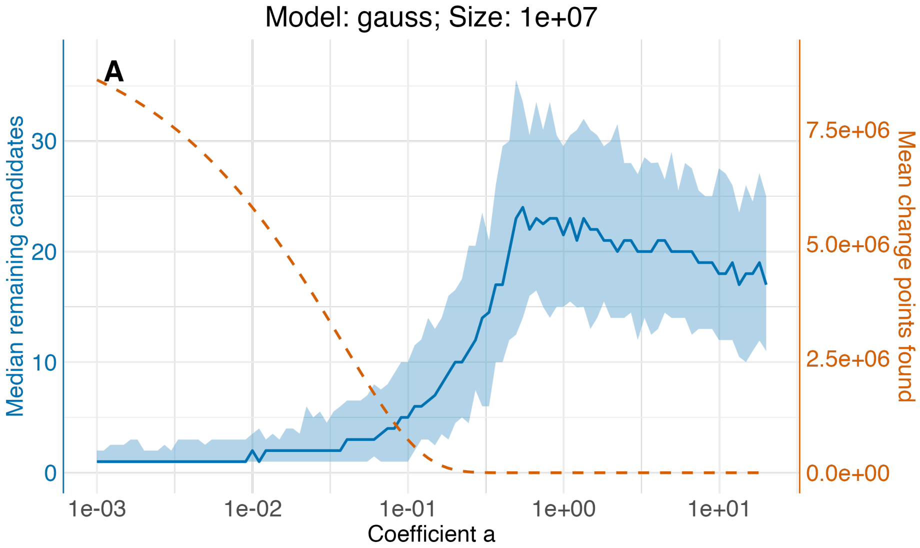


Simulations

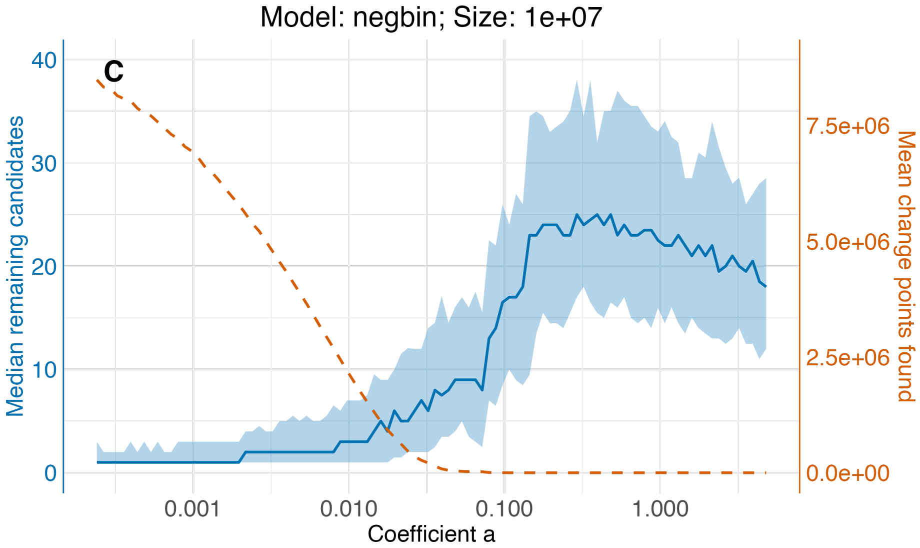
Simulations



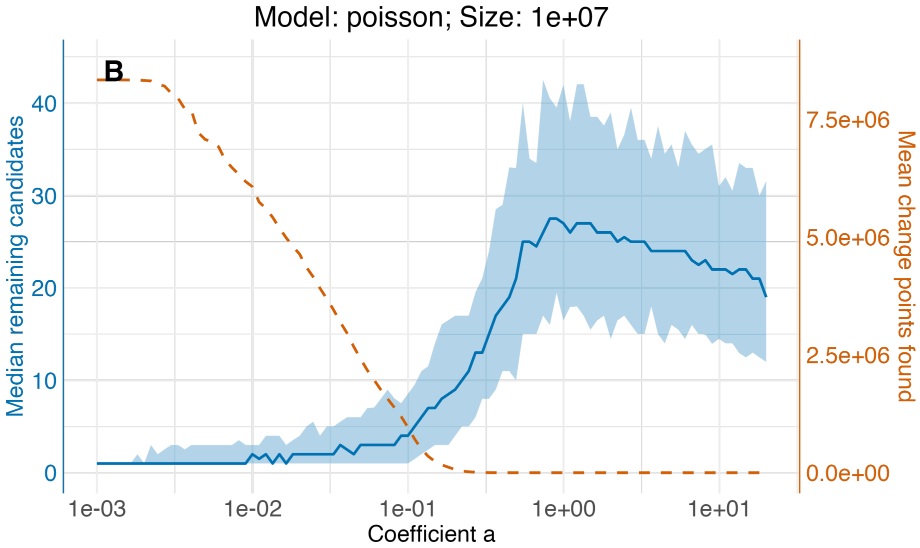
Simulations



Simulations



Simulations



Outlines

In The Exponential Family, dimension d

The index s at time $t > s$ is pruned by the DUST rule when we find x such that
From this

$$-\mathcal{D}^*\left(\overline{y_{st}} + {\color{blue}x}(\overline{y_{st}} - \overline{y_{rs}})\right) - \left(\frac{Q_t - Q_s}{t-s} + {\color{blue}x}\left(\frac{Q_t - Q_s}{t-s} - \frac{Q_s - Q_r}{s-r}\right)\right) > 0?$$

to:

$$\boxed{-\mathcal{D}^*\left(\overline{\mathbf{S}}_{st} + \sum_{r \neq s} {\color{blue}x_r} \Delta \overline{\mathbf{S}}_{rst}\right) - \left(\overline{Q}_{st} + \sum_{r \neq s} {\color{blue}x_r} \Delta \overline{Q}_{rst}\right) > 0?}$$

with $\mathcal{D}^*(x) = x \cdot (\nabla A)^{-1}(x) - A((\nabla A)^{-1}(x))$

Example: Change in Mean and Variance

Pruning Decision Rule

The decision function with $A(\theta_1, \theta_2) = -\frac{\theta_1^2}{4\theta_2} + \frac{1}{2}\log(-\frac{1}{2\theta_2})$ and a single constraint is: fo

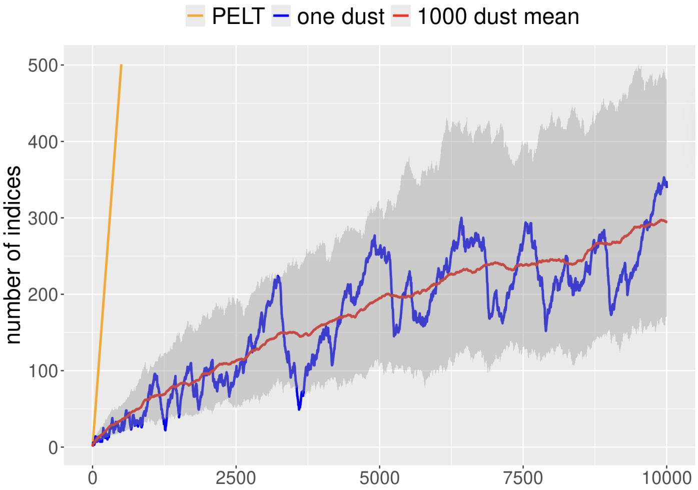
$$\mathbb{D}(x) = \frac{1}{2}\left[1+\log\left(\overline{y_{st}^2}+x\Delta\overline{y_{rst}^2}-(\overline{y}_{st}+x\Delta\overline{y}_{rst})^2\right)\right]-\left(\overline{Q}_{st}+x\Delta\overline{Q}_{rst}\right),$$

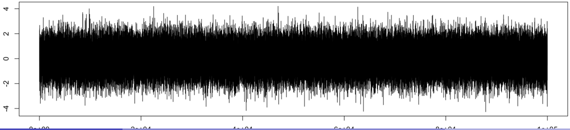
$$x_0 = \frac{1}{2}\Big(\frac{V(y_{st}) - V(y_{rs})}{(\Delta\overline{y}_{rst})^2} - 1\Big) \quad x_1 = x_0^2 + \frac{V(y_{st})}{(\Delta\overline{y}_{rst})^2}$$

and its maximum, with notation $x_2 = \Delta\overline{Q}_{rst}$, is evaluated in:

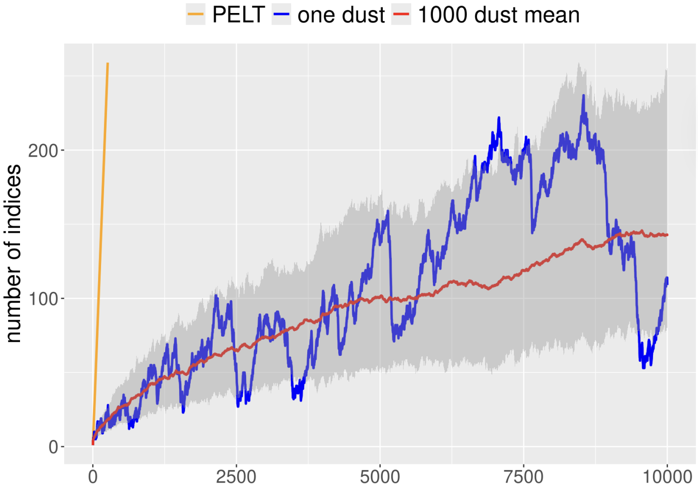
$$x^\star = \max\left\{0, x_0 + (2x_2)^{-1} - \text{sign}(x_2)\sqrt{x_1 + (2x_2)^{-2}}\right\}$$

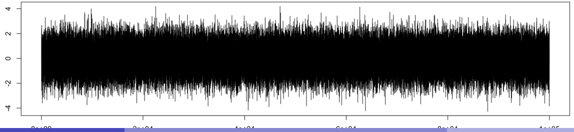
Example: Change in Mean and Variance (1 constraint)



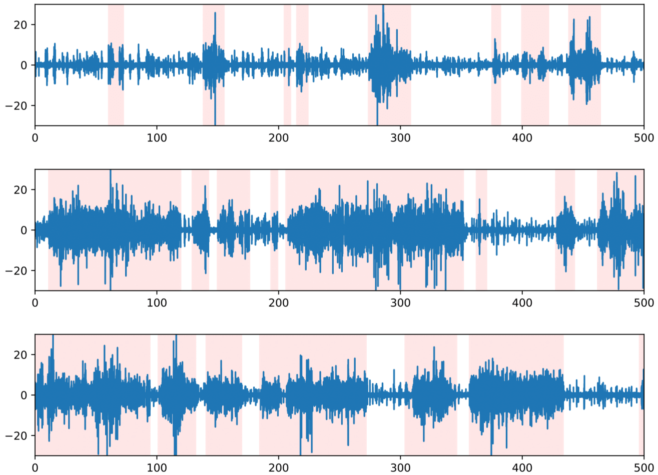


Example: Change in Mean and Variance (2 constraints)





Example: Change in Variance (Mouse activity measured by a force platform)



About 400 000 data points (12h of record) analysed in a few seconds.

Conclusion

DUST pruning

- As simple as PELT (one comparison) (DUality Simple Test)
- Efficient in all change-point regimes
- Often quasi-linear time complexity (univariate data)
- Efficient for multi-parametric/multivariate time series
- Rcpp package **dust** available in github

What next?

- ☞ Efficient maximum dual evaluation with multivariate dual function (better than Quasi-Newton algorithm)
- ☞ Exact maximum evaluation as in 1D?

Merci !

References

- (arXiv) DUST: Duality-Based Pruning Methods For Exact Multiple Change Point Detection.
  Vincent Runge, Charles Truong, Simon Querné, 2025
- An Efficient Algorithm for Exact Segmentation of Large Compositional and Categorical Time Series.
  stats. Charles Truong, Vincent Runge, 2024

Rcpp package: `https://github.com/vrunge/dust`