Multiple Changes Offline: Advanced Topics

Arnaud Liehrmann & Guillem Rigaill

Sorbonne Université & INRAE

February 27, 2025

Introduction

- We have seen maximum likelihood approach with a constant penalty serves as a baseline for changepoint detection.
- Three ingredients:
  - A model (we looked at Gaussian errors)
  - A model selection criterion ($2\sigma^2 \log(n)$)
  - An algorithm to explore segmentation space (DP)
- Fairly generic, reasonably fast (for medium n: 10^4-10^5) and good statistical properties
- Good statistical performance in simulations and applications [Fearnhead and Rigaill, 2020, Lai et al., 2005, Picard et al., 2011, Liehrmann et al., 2021, Killick et al., 2010].

Some Limitations...

1. CUSUM algorithm is linear offline but quadratic in an online context.
2. $2\log(n)$ penalty is consistent in infill asymptotics but theoretically unoptimal: misses low energy changepoints.
3. OP algorithm and variants (e.g., Segment Neighborhood) are quadratic and slow for profiles larger than 10^5.
4. DP algorithm is easy to implement for simple log-likelihood functions but complex for robust losses (e.g., ℓ_1, Huber).
5. Most DP algorithm requires that the likelihood is written as a sum over all segments, which is not the case when we have constraints between segments (e.g., piecewise regression with continuity, peaks...).

A very quick overview of some recent solutions

- This chapter highlights recent solutions addressing the first three limitations.
- Aims to guide readers through the evolving changepoint literature.

Plan

Outline

Introduction to Multiscale Penalties

- Constant penalty of order $\log(n)$ in Gaussian case:
  - Detects changes with high energy.
  - Misses low-energy changepoints.

- Suboptimal because the price to pay for a change is the same no matter the position of the change...

Balanced vs. Unbalanced Segmentation

- Penalty is the same for:
  - Balanced segmentation: two equal pieces.
  - Unbalanced segmentation: one segment of length 1.
- Intuition: We should pay less for balanced segmentation.
- Need for penalties that depend on the scale of segmentation.

Mutliscale penalties

- Recent papers show improvement over constant penalty using multi-scale penalties [Pein et al., 2017, Cho and Kirch, 2019, Wang et al., 2020, Verzelen et al., 2023]
- The multiscale penalty proposed in [Verzelen et al., 2023] can be optimized using dynamic programming and in particular the OP algorithm (see exercise)

Multi-Scale Penalty

Multi-Scale Penalty of [Verzelen et al., 2023]

$$\mathrm{pen}_{ms}(\boldsymbol{\tau}) = c_1(\#\boldsymbol{\tau}) + 2c_2 \sum_{k=1}^{\#\boldsymbol{\tau}+1} \log\left(\frac{n}{\tau_k - \tau_{k-1}}\right)$$

where $c_1 > 0$ and $c_2 > 1$.

- Least-square criteria penalized by multi-scale penalty shown to be optimal.

Exercises

Exercises

1. Compare the multiscale penalty with the $2\log(n)$ (and $6\log(n)$) constant penalty for $\#\boldsymbol{\tau} = 1$.
2. Consider some infill asymptotic with fixed changes: for increasing n $\tau_k(n) = \alpha_k n$. How does the penalty vary with n.

Exercises

Exercises

1. Show that the optimal partitioning algorithm presented in the previous chapter applies to the previous multi-scale penalty. Implement it in Python.
2. Calibrate the two constants using some simulations and compare its power with the SIC-like penalty [Yao and Au, 1989]

Outline

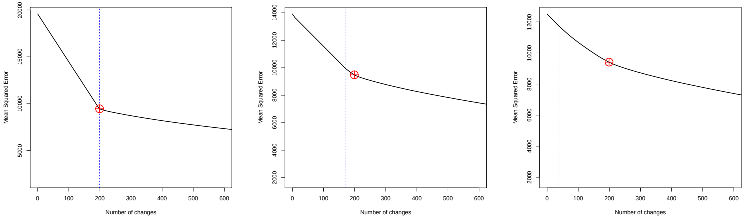
Adaptive Penalty

- Improve upon constant $2\log(n)$ (SIC-penalty) by adaptively choosing the penalty.
- Key idea: Consider minus the log-likelihood as a function of the number of segments.
  - If we are below true number of changes $\#\boldsymbol{\tau}^*$, increasing the number of changepoints should sharply improve the fit to the data.
  - If we are above the true number, increasing the number of changepoints should modestly improve the fit.

The “Elbow Heuristic”

- Plotting minus the log-likelihood vs. number of changes reveals an "elbow."
- Simulation setup:
  - Signal with $n = 10000$, change every 50 points.
  - Change sizes: $\delta = 2$, 1.25, and 1.
  - Total of 199 changepoints.

Simulation Results



- Blue vertical line = number of changes selected using $2\log(n)$.
- Red cross = likelihood for the simulated number of changepoints

Pinpointing the “elbow” by eye

- Without knowing the true number of changepoints, we can hypothesize that the SIC penalty of $2\log(n)$ underestimates the number of changepoints.
- Visually pinpoint the elbow or a range of penalties including it.

Pinpointing the "elbow" by eye

For any changepoint approach

- Penalty or tuning parameters have been calibrated (based on some theory or simulations) under certain distributional assumptions that are likely not always appropriate.
- Consider how the fit varies with the number of changes.
- This can be done using:
  - A grid of penalties.
  - In the case of OP this can be done quickly with CROPS [Haynes et al., 2017].

Adaptive Calibration of Penalty

- Several approaches exploit the "elbow" phenomenon to adaptively calibrate the penalty [Lavielle, 2005, Lebarbier, 2005, Fryzlewicz, 2020, Arlot et al., 2019, Cleynen and Lebarbier, 2017].
- Justification for adaptive calibration:
  - Change the goal from estimating changepoints to estimating the signal.
  - Choosing a multiplicative constant in front of a penalty relates to model selection.
  - This approach extends beyond changepoint detection [Arlot et al., 2019]

Plan

Quadratic Complexity of DP Algorithms

- Optimal partitioning algorithm [Jackson et al., 2005] and Segment Neighborhood algorithm [Auger and Lawrence, 1989] have quadratic complexity.
  - $n = 10^4$: runs in several tens of minutes.
  - $n = 10^5$: runs in one or a few hours.
- For medium-sized datasets, this approach is a good baseline.
- This is a concern for larger datasets

Need for Acceleration

- In certain situations, quadratic runtimes are too slow
- Two exact accelerations:
  - PELT [Killick et al., 2012]:
    - Very generic
    - Expectation linear for a large number of true changepoints.
  - FPOP [Maidstone et al., 2017]
    - Not generic (some univariate and low-dimensional models)
    - Efficient even in the absence of changes, empirically $O(n \log(n))$ (e.g. in the Gaussian case $n = 10^8$ in 1 minute)

DP Algorithm with PELT Pruning

- PELT relies on the assumption:

$$\ell_{i:j} \geq \ell_{i:\tau} + \ell_{\tau+1:j}$$

- This assumption is true for maximum likelihood approaches.
- Leads to the pruning condition
  - If a change τ is bad:

$$\mathcal{L}_{1:\tau} + \ell_{\tau+1:n} > \mathcal{L}_{1:n}$$

  - Then we can prune τ from the candidate changepoints.

Functional Pruning Overview

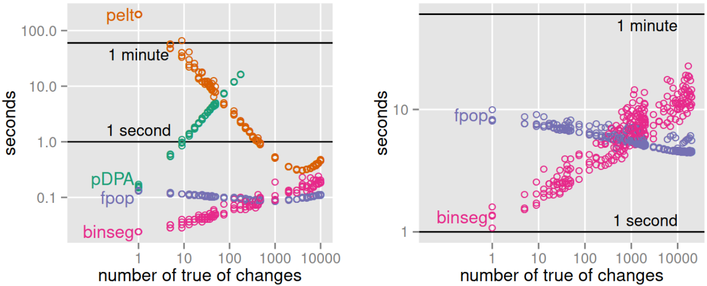
- Another way to prune candidate changepoints is through functional pruning [Rigaill, 2015, Maidstone et al., 2017].
- Assumes log-likelihood can be expressed as:

$$\ell_{i:j}(\theta) = -\sum_{t=i}^{j} \log(f_{Y_t}(\theta)).$$

- Rewrite likelihood as a function of the last segment parameters μ.
- The goal is to recover the optimal log-likelihood as a function of μ:

Compared to binseg

- Runtimes for $n = 200000$ and $n = 10^7$
- Varying number of changepoints



(adapted from [Maidstone et al., 2017])

Plan

Model selection

Multiscale Penalties

Adaptive Penalty and Elbow-like Heuristics

Computational Speed-Up

Isolation, Local search or optimization

Local Search and Isolation

- When DP is too slow or is not possible

First idea: Approximate the likelihood using Binary Segmentation.

- Split 1: $1:n$ τ_1
- Split 2:
  - Find the best split of $1:\tau_1$ and $\tau_1+1:n$: τ_2 and τ_3
  - Pick the best of τ_2 and τ_3 (say τ_2)

  Split 3:
  - Find the best split of $1:\tau_2$ and $\tau_2+1:\tau_1$: τ_4 and τ_5
  - Pick the best of τ_3, τ_4, τ_5

- Empirically $O(n\log(n))$
- Statistical limitations
  [Fryzlewicz, 2014, Venkatraman and Olshen, 2007].

Improving Initial Solutions

Fairly Recent Idea: Apply AMOC LRT/Cusum-like strategy not to the whole data but to diverse chunks of data. Intuition

- Each true changepoint will be detected in at least one chunk.
- Use a well-thought-out penalty and aggregation procedure.

References: [Fryzlewicz, 2014, Baranowski et al., 2019, Eichinger and Kirch, 2018, Cho and Kirch, 2019, Anastasiou and Fryzlewicz, 2022, Kovács et al., 2023, Verzelen et al., 2023]

Some packages

Among many others

- Packages with many models/approaches and penalties
  - https://centre-borelli.github.io/ruptures-docs/
  - https://cran.r-project.org/web/packages/changepoints/index.html
  - https://cran.r-project.org/web/packages/breakfast/index.html
- Some specialised fast DP packages in R in the Gaussian case
  - i.i.d weighted : https://cran.r-project.org/web/packages/fpopw/index.html
  - i.i.d + multiscale penalty: https://github.com/aLiehrmann/MsFPOP
  - AR(1): https://cran.r-project.org/web/packages/DeCAFS/index.html

Anastasiou, A. and Fryzlewicz, P. (2022).
Detecting multiple generalized change-points by isolating single ones.
Metrika, 85(2):141–174.

Arlot, S., Celisse, A., and Harchaoui, Z. (2019).
A kernel multiple change-point algorithm via model selection.
Journal of machine learning research, 20(162):1–56.

Auger, I. E. and Lawrence, C. E. (1989).
Algorithms for the optimal identification of segment neighborhoods.
Bulletin of mathematical biology, 51(1):39–54.

Baranowski, R., Chen, Y., and Fryzlewicz, P. (2019).
Narrowest-over-threshold detection of multiple change points and change-point-like features.
Journal of the Royal Statistical Society Series B: Statistical Methodology, 81(3):649–672.

Cho, H. and Kirch, C. (2019).

Localised pruning for data segmentation based on multiscale change point procedures.
arXiv preprint arXiv:1910.12486.

Cleynen, A. and Lebarbier, É. (2017).
Model selection for the segmentation of multiparameter exponential family distributions.
Electron. J. Statist.

Eichinger, B. and Kirch, C. (2018).
A mosum procedure for the estimation of multiple random change points.

Fearnhead, P. and Rigaill, G. (2020).
Relating and comparing methods for detecting changes in mean.
Stat, 9(1):e291.

Fryzlewicz, P. (2014).
Wild binary segmentation for multiple change-point detection.

Fryzlewicz, P. (2020).

Detecting possibly frequent change-points: Wild binary segmentation 2 and steepest-drop model selection.
Journal of the Korean Statistical Society, 49(4):1027–1070.

Haynes, K., Eckley, I. A., and Fearnhead, P. (2017).
Computationally efficient changepoint detection for a range of penalties.
Journal of Computational and Graphical Statistics, 26(1):134–143.

Jackson, B., Scargle, J. D., Barnes, D., Arabhi, S., Alt, A., Gioumousis, P., Gwin, E., Sangtrakulcharoen, P., Tan, L., and Tsai, T. T. (2005).
An algorithm for optimal partitioning of data on an interval.
IEEE Signal Processing Letters, 12(2):105–108.

Killick, R., Eckley, I. A., Ewans, K., and Jonathan, P. (2010).
Detection of changes in variance of oceanographic time-series using changepoint analysis.
Ocean Engineering, 37(13):1120–1126.

Killick, R., Fearnhead, P., and Eckley, I. A. (2012).

Optimal detection of changepoints with a linear computational cost.
Journal of the American Statistical Association, 107(500):1590–1598.

Kovács, S., Bühlmann, P., Li, H., and Munk, A. (2023).
Seeded binary segmentation: a general methodology for fast and optimal changepoint detection.
Biometrika, 110(1):249–256.

Lai, W. R., Johnson, M. D., Kucherlapati, R., and Park, P. J. (2005).
Comparative analysis of algorithms for identifying amplifications and deletions in array cgh data.
Bioinformatics, 21(19):3763–3770.

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

Lebarbier, É. (2005).

Detecting multiple change-points in the mean of gaussian process by model selection.
Signal processing, 85(4):717–736.

Liehrmann, A., Rigaill, G., and Hocking, T. D. (2021).
Increased peak detection accuracy in over-dispersed chip-seq data with supervised segmentation models.
BMC bioinformatics, 22:1–18.

Maidstone, R., Hocking, T., Rigaill, G., and Fearnhead, P. (2017).
On optimal multiple changepoint algorithms for large data.
Statistics and computing, 27:519–533.

Pein, F., Sieling, H., and Munk, A. (2017).
Heterogeneous change point inference.
Journal of the Royal Statistical Society Series B: Statistical Methodology, 79(4):1207–1227.

Picard, F., Lebarbier, E., Hoebeke, M., Rigaill, G., Thiam, B., and Robin, S. (2011).

Joint segmentation, calling, and normalization of multiple cgh profiles.
Biostatistics, 12(3):413–428.

Rigaill, G. (2015).
A pruned dynamic programming algorithm to recover the best segmentations with 1 to $k\_\{max\}$ change-points.
Journal de la Société Française de Statistique, 156(4):180–205.

Venkatraman, E. and Olshen, A. B. (2007).
A faster circular binary segmentation algorithm for the analysis of array cgh data.
Bioinformatics, 23(6):657–663.

Verzelen, N., Fromont, M., Lerasle, M., and Reynaud-Bouret, P. (2023).
Optimal change-point detection and localization.
The Annals of Statistics, 51(4):1586–1610.

Wang, D., Yu, Y., and Rinaldo, A. (2020).
Univariate mean change point detection: Penalization, cusum and optimality.

Yao, Y.-C. and Au, S.-T. (1989).
Least-squares estimation of a step function.
Sankhyā: The Indian Journal of Statistics, Series A, pages 370–381.