Changepoint Inference for latent variable models, links to geometry

To apply please contact guillem.rigaill@inrae.fr, mahendra.mariadassou@inrae.fr and julien.chiquet@inrae.fr with a CV.

Context

Changepoint inference and geometry. The increasing volume of data streams poses computational challenges for detecting changepoints. Exact likelihood-based methods are effective, but often slow due to high computational costs[1,2]. We recently established **a geometric connection between the detection of a changepoint in p-dimensional data streams and the construction of a convex hull of points** in the (p+1)-dimensional Euclidean space[3,4]. This link makes it possible to compute exactly the maximum likelihood (ML) estimate by considering only an expected number $O_p(\log^p(t))$ of changepoints at time step t rather than O(t). Empirically, this leads to an overall $O(n\log^p(n))$ algorithm, practical for p up to 6. For larger p, this becomes impractical but the link to geometry remains promising to improve existing approximations[4]. Further work is needed to gain a better understanding of this geometric link, particularly its real-time updates in the online setting. **This link between geometry and changepoint detection is surprisingly generic** as it applies to any loss that can be algebraically written as the likelihood of a natural exponential family, even with priors or constraints on the parameter space. This includes classical distributions (Gaussian, Poisson), more exotic ones, such as the Pareto type I distribution, and even non-parametric extensions such as e-detectors[5].

Latent variable models. Latent variable models are very popular in statistics as the *de facto* framework to capture complex dependencies. Exact inference in those models requires integrating the usually complete likelihood over all possible values of the latent variable to compute the observed likelihood before maximizing it. This integration typically becomes impossible for complex models. A popular method to alleviate this difficulty is the EM algorithm and its variants (MCMC-based approaches, variational approximations, Laplace method, etc.) which use iterative procedures to approximate the observed likelihood, maximize it, and repeat until convergence[6].

The goal of this post-doctoral fellowship would be to explore and push forward the connection between geometric shapes and changepoint detection inference for latent variable models. We will start with simple models with both latent variables and emission distributions in the exponential family, such as Generalized Mixed Effect Models, where the changepoint concerns (i) the fixed effect, and/or (ii) a low dimensional parametrization of the variance parameters of the mixed effect before moving on to more complex models. The aim would be to apply the geometrical connection within the EM to improve the pruning of the changepoints compared to well-studied inequality-based pruning[1]. Motivating examples concern the segmentation of longitudinal multivariate data, as are now increasingly available in omics data (e.g. multi-conditions expression level data along the genome, multi-population recombination rates along the genome, etc).

Skills

We are looking for a Ph.D in applied mathematics, statistics, or computational statistics with an interest in optimization and algorithmics. Knowledge of a programming language would be a plus.

Supervision

The position will be based at the MIA-PS AgroParisTech/INRAE The supervision will be provided by Guillem Rigaill (LaMME, Evry), Mahendra Marriadassou (MaIAGE, INRAE, Jouy en Josas), and Julien Chiquet (MIA-PS).

Bibliography

1) Killick et al. Optimal detection of changepoints with a linear computational cost. Journal of the American Statistical Association (2012)
2) Fearnhead and Rigaill. Relating and comparing methods for detecting changes in mean. Stat (2020)
3) Romano et al. Fast online changepoint detection via functional pruning CUSUM statistics. Journal of Machine Learning Research (2023)
4) Pishchagina et al. Online Multivariate Changepoint Detection: Leveraging Links With Computational Geometry. arXiv:2311.01174 (2023).
5) Jaehyeok et al. E-detectors: a nonparametric framework for sequential change detection. arXiv:2203.03532 (2022).
6) Bishop, Christopher M. Pattern recognition and machine learning. Springer New York (2007)