Funded PhD position in mathematical statistics
Post-selection inference for latent variable models
With optional funded internship before
At *Institut de Mathématiques de Toulouse*

Project description

Classical inference tools, in particular hypothesis tests and confidence intervals, can dramatically fail when applied to data-driven statistical models. Post-selection inference refers to a set of recent research works that design and analyze statistical methods tailored to these data-driven models. In particular, [3] addresses Gaussian linear models and [2] provides extensions to non-linear non-Gaussian settings, based on asymptotic arguments.

The goal of the PhD project is to extend post-selection inference to latent variables models. These models have become the method of choice in a wide range of applications [4, 6, 8] and are the object of many recent contributions [1, 5]. Nevertheless, post-selection inference guarantees are currently missing for them, while model selection often takes place in practice [7, 9].

This extension, relying on [2], will necessitate to obtain uniform joint central limit theorems for parameter estimators with latent variables. Also, from a computational point of view, parameter estimation will be performed thanks to the Expectation Maximization (EM) algorithms and their extensions. This will also necessitate mathematical developments to account for the post-selection inference context.

Candidate profile

We are seeking for candidates with a degree in mathematics, with a specialization in probability, statistics, machine learning or applied mathematics. Solid theoretical skills are expected.

Details

- **Supervisors:** François Bachoc and Juliette Chevallier (Institut de Mathématiques de Toulouse).

- **Start:** Fall 2025, with possible master internship during Spring and Summer 2025.

- **Duration:** PhD funded for 3 years.

- **Location:** Institut de Mathématiques de Toulouse (Toulouse, France).

- **Funding source:** The PhD is funded by the QUTHY project involving industrial actors. The selected PhD student will have the option (non-mandatory) to attend workshops with these industrial actors and to address real data sets from the QUTHY project.

- **Further information:** The current scientific context in Toulouse, in the field of statistics and related fields, is very rich. There are over 20 PhD students, with a PhD student mathematical seminar and another group seminar on statistics and optimization. Many events are also organized by the ANITI AI cluster.

How to apply

Applications will be considered starting from November 2024 and until the position is filled. The candidates should send a CV and grade transcripts (bachelor and master level) to François Bachoc (francois.bachoc@math.univ-toulouse.fr) and Juliette Chevallier (juliette.chevallier@insa-toulouse.fr).

References

[1] P. Abry, J. Chevallier, G. Fort, and B. Pascal. Pandemic intensity estimation from stochastic approximation-based algorithms. In *2023 IEEE 9th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 356–360. IEEE, 2023.

[2] F. Bachoc, D. Preinerstorfer, and L. Steinberger. Uniformly valid confidence intervals post-model-selection. *The Annals of Statistics*, 48(1):440–463, 2020.

[3] R. Berk, L. Brown, A. Buja, K. Zhang, and L. Zhao. Valid post-selection inference. *The Annals of Statistics*, pages 802–837, 2013.

[4] D. M. Blei. Build, compute, critique, repeat: Data analysis with latent variable models. *Annual Review of Statistics and Its Application*, 1(1):203–232, 2014.

[5] J. Chevallier, V. Debavelaere, and S. Allassonniere. A coherent framework for learning spatiotemporal piecewise-geodesic trajectories from longitudinal manifold-valued data. *SIAM Journal on Imaging Sciences*, 14(1):349–388, 2021.

[6] B. Everett. *An introduction to latent variable models*. Springer Science & Business Media, 2013.

[7] S. Lotfi, P. Izmailov, G. Benton, M. Goldblum, and A. G. Wilson. Bayesian model selection, the marginal likelihood, and generalization. In *International Conference on Machine Learning*, pages 14223–14247. PMLR, 2022.

[8] B. O. Muthén. Beyond SEM: General latent variable modeling. *Behaviormetrika*, 29(1):81–117, 2002.

[9] Y.-Q. Zhang, G.-L. Tian, and N.-S. Tang. Latent variable selection in structural equation models. *Journal of Multivariate Analysis*, 152:190–205, 2016.