Master thesis internship in mathematical statistics

Asymptotic efficiency and global sensitivity analysis

Project description

The use of complex computer models for the simulation and analysis of natural systems from physics, engineering, and other fields is by now routine. These models usually depend on many input variables, and it is thus crucial to understand which input parameter or which set of input parameters have an influence on the output. This is the aim of sensitivity analysis which has become an essential tool for system modeling and policy support. Global sensitivity analysis methods consider the input vector as random and propose a measure of influence of each subset of its components on the output of interest. We refer to the seminal book [11] for an overview on global sensitivity analysis or to [4] for a synthesis of recent trends in the field. Among the different measures of global sensitivity analysis, variance-based measures are probably the most commonly used. The definition of the so-called Sobol' indices, introduced in [10] and later revisited in the framework of sensitivity analysis in [12, 13], is based on the Hoeffding decomposition of the variance [6]. More precisely, for the output Y of a computer code $Y = G(V_1, \dots, V_p)$ where the inputs V_i are assumed to be mutually independent, the Sobol' index of Y with respect to a subset of inputs X of dimension d is defined by

$$S^X = \frac{\mathrm{Var}(\mathbb{E}[Y|X])}{\mathrm{Var}(Y)} = \frac{\mathbb{E}[\mathbb{E}[Y|X]^2] - \mathbb{E}[Y]^2}{\mathrm{Var}(Y)}.$$

Since in practice computing explicitly the theoretical value of S^X is out of reach, one of the main tasks in sensitivity analysis is to provide estimators of S^X, with guaranteed asymptotic properties such as consistency, rate of convergence, central limit theorem. In the recent years, a myriad of different estimators has been proposed, see [4, Chapter 4] for a complete review. To compare these different estimators, it is then relevant to define a notion of "optimality" using a concept similar to the Cramér-Rao bound in parametric statistics.

Optimality is assessed via the notion of *asymptotic efficiency* introduced in the seminal works [5, 7] in a parametric setting and further extended to semi and non-parametric models in [2, 9, 1] (see also [3, 14] for an extensive description of the theory of asymptotic efficiency). For Sobol' index inference specifically, the whole difficulty in showing asymptotic efficiency revolves around determining the so-called efficient influence function of the parameter $\psi = \mathbb{E}[\mathbb{E}[Y|X]^2]$. Once this is done, assessing the asymptotic efficiency of a particular estimator boils down to checking if its first order asymptotic Taylor expansion matches the empirical mean of the efficient influence function. Such a study for Sobol' indices has been proposed in [8].

The aim of this project is first to familiarize with the notions of asymptotic efficiency and efficient influence functions. In a second step, the goal is to determine the efficient influence function for several sensitivity analysis indices (e.g., Cramér-von Mises indices, general metric space indices, universal indices) to be able to propose asymptotically efficient estimation procedures.

The PhD project is part of the GATSBII project that is funded by the French national research agency (ANR) in 2024-2029 and hosted at Institut de Mathématiques de Toulouse, with Thierry Klein as principal investigator (PI), Sébastien Gadat, Agnès Lagnoux, and Paul Rochet as co-investigators in Toulouse. The aim of the GATSBII project is to explore the links between Cooperative Game Theory, Global Sensitivity Analysis and Explanability in Atificial Intelligence both from the theoretical and numerical point of view and to apply the results to practical industrial test cases. The GATSBII project gather also other academic partners: Université Paris Dauphine and ENSAI and industrial partners: EDF and Thalès.

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: Thierry Klein, Agnès Lagnoux (Institut de Mathématiques de Toulouse), and Paul Rochet.
- Other collaborators within the GATSBII ANR project: Sébastien Gadat (Toulouse school of echonomics).
- Start: Spring 2025.
- Duration: funded for 3 or 6 months.
- PhD thesis possibility: It is possible that this master thesis will be followed by a PhD thesis. In this case, the student will apply to PhD scholarships, with the help of the master thesis supervisors.
- Location: Institut de Mathématiques de Toulouse (Toulouse, France), with possible visits to EDF and Université Paris Dauphine.

How to apply

Applications will be considered until the position is filled. The candidates should send a CV, application letter and grade transcripts (master level) to Thierry Klein (thierry.klein@math.univ-toulouse.fr).

Keywords

Global sensitivity analysis, asymptotic efficiency, efficient influence function, design of experiments, probability model.

References

[1] J. M. Begun, W. J. Hall, W.-M. Huang, and J. A. Wellner. Information and asymptotic efficiency in parametric-nonparametric models. *The Annals of Statistics*, 11(2):432–452, 1983.

[2] R. Beran. Robust location estimates. *The Annals of Statistics*, pages 431–444, 1977.

[3] P. J. Bickel, C. A. Klaassen, P. J. Bickel, Y. Ritov, J. Klaassen, J. A. Wellner, and Y. Ritov. *Efficient and adaptive estimation for semiparametric models*, volume 4. Springer, 1993.

[4] S. Da Veiga, F. Gamboa, B. Iooss, and C. Prieur. *Basics and Trends in Sensitivity Analysis: Theory and Practice in R*. SIAM, 2021.

[5] J. Hájek. A characterization of limiting distributions of regular estimates. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 14(4):323–330, 1970.

[6] W. Hoeffding. A class of statistics with asymptotically normal distribution. *Ann. Math. Statistics*, 19:293–325, 1948.

[7] N. Inagaki. On the limiting distribution of a sequence of estimators with uniformity property. *Annals of the Institute of Statistical Mathematics*, 22(1):1–13, 1970.

[8] T. Klein and P. Rochet. Efficiency of the averaged rank-based estimator for first order sobol index inference. *Statistics & Probability Letters*, 207:110015, 2024.

[9] B. Y. Levit. On the efficiency of a class of non-parametric estimates. *Theory of Probability & Its Applications*, 20(4):723–740, 1976.

[10] K. Pearson. On the partial correlation ratio. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 91(632):492–498, 1915.

[11] A. Saltelli, K. Chan, and E. Scott. *Sensitivity analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 2000.

[12] I. M. Sobol. Sensitivity estimates for nonlinear mathematical models. *Math. Modeling Comput. Experiment*, 1(4):407–414 (1995), 1993.

[13] I. M. Sobol. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3):271–280, 2001.

[14] A. W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.