Efficient estimation of Sobol' indices of any order from a single input/output sample.

Da Veiga, Sébastien
Univ Rennes, Ensai, CNRS, CREST - UMR 9194, F-35000 Rennes, France

Gamboa, Fabrice, Klein, Thierry, Lagnoux, Agnès
Institut de Mathématiques de Toulouse and ANITI; UMR5219. Université de Toulouse; CNRS. UT3, F-31062 Toulouse, France

Prieur, Clémentine
Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LJK, Grenoble, France

Despite the numerous advances and insightful proposals in the recent years, estimation of Sobol' indices at any order d (with particular case total indices) is still a challenge. When it comes to theoretical convergence guarantees, two classes of methods are of particular interest. On the one hand, the class of Pick Freeze estimators allows to estimate Sobol' indices at rate $\sqrt{n}$ for any d with minimal assumptions on the computer code, but requires a sample with highly specific structure. On the other hand, local-averaging estimators such as kernel or nearest neighbor estimators can handle any vanilla n-sample of the inputs/output pair (*given-data* context), but the $\sqrt{n}$-parametric rate of convergence was only proved for $d \leq 3$ for nearest neighbors. In addition, such estimators suffer in practice from large bias and variance.

In the present work, we introduce a new class of kernel estimators which enjoys a central limit theorem and asymptotic efficiency for estimating Sobol' indices at rate $\sqrt{n}$ for any d from a vanilla n-sample, unlike all previous works. From a broad perspective, our approach consists of three main ingredients. First, we build upon the explicit expression of the efficient influence function of Sobol' indices which depends on the unknown regression function, and propose a plug-in estimator where the regression function is estimated with a specific kernel estimator, in the same spirit as [1]. Second, for the latter and to ensure $\sqrt{n}$-consistency, we use high-order kernels as classically done in nonparametric regression. Finally, it is crucial to handle boundary effects inherent to kernel estimation procedures: we adapt here recent mirror-type transformations introduced in [2,3]. All in one, we introduce two different estimators that are proved to be asymptotically normal and efficient for Sobol' indices at any order. From a numerical perspective, we conduct extensive comparisons and discuss stability of high-order kernels, showing that one of our estimators performs remarkably well on standard sensitivity analysis examples.

References:

[1] K. Doksum and A. Samarov. Nonparametric estimation of global functionals and a measure of the explanatory power of covariates in regression. *The Annals of Statistics*, 1443–1473, 1995.

[2] K. Bertin, N. Klutchnikoff, J. R. Léon, and C. Prieur. Adaptive density estimation on bounded domains under mixing conditions. *Electronic Journal of Statistics*, **14**(1):2198–2237, 2020.

[3] L. Pujol. Nonparametric estimation of a multivariate density under Kullback-Leibler loss with ISDE. `arXiv preprint arXiv:2205.03199`, 2022.

[Da Veiga Sébastien; Univ Rennes, Ensai, CNRS, CREST - UMR 9194, F-35000 Rennes, France]
[`sebastien.da-veiga@ensai.fr` – `https://sites.google.com/view/sebastien-da-veiga`]

[Prieur Clémentine; Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LJK, Grenoble, France]
[`clementine.prieur@univ-grenoble-alpes.fr` – `https://membres-ljk.imag.fr/Clementine.Prieur/`]