Knockoff'ed Total Indices

Elmar Plischke
Helmholtz-Zentrum Dresden-Rossendorf, Dresden, Germany

Emanulele Borgonovo
Uni Bocconi, Milano, Italy

Clémentine Prieur
Université Grenoble Alpes, Grenoble, France

Total indices are global sensitivity indicators subject of intensive investigation in the statistical, machine learning as well as simulation literature. They aim to capture the strength of dependence between a quantity of interest/target and covariates/features. In general, a total index is the fraction of the variance of Y that is left unexplained when all features are fixed but X_i.

Applying total indices in a dependent-input setting generally requires conditionally independent realizations. In [6], non-Cartesian input domains are studied using a rejection technique, in [2] the pick-and-freeze methodology is also applied in the dependent input case, with the introduction of an additional density quotient to adjust for the disparity between the product of marginal densities (where the pick-and-freeze sampling is formed) and the joint density (where the conditionally independent distributions are found).

In the machine learning context, it was noted by [5] that model-X knockoffs introduced by [3] may be used for assessing feature importance. We apply this reasoning to the sensitivity analysis context, and arrive at the surprising result that pick-and-freeze algorithms can be applied unmodified also in the dependent case when the alternative sample block is generated as knockoff.

This approach is extended to moment-invariant measures using kernel-based dissimilarity measures [4] and optimal-transport-based measures [1] which can be applied in case of stochastic output.
References:

[1] E. Borgonovo, A. Figalli, E. Plischke, and G. Savarè. Global sensitivity analysis via optimal transport. *Management Science*, 2024. Accepted.

[2] E. Borgonovo, E. Plischke, and C. Prieur. Total effects with constrained features. *Statistics and Computation*, 34(87):1–25, 2024.

[3] E. Candès, Y. Fan, L. Janson, and J. Lv. Panning for gold: 'model-X' knockoffs for high dimensional controlled variable selection. *Journal of the Royal Statistical Society, Series B*, 3(80):551–577, 2018.

[4] S. Da Veiga. Kernel-based ANOVA decomposition and Shapley effects – Application to global sensitivity analysis. preprint, Hyper articles en ligne, 2021. `https://hal.archives-ouvertes.fr/hal-03108628`.

[5] G. Hooker, L. Mentch, and S. Zhou. Unrestricted permutation forces extrapolation: variable importance requires at least one more model, or there is no free variable importance. *Statistics and Computing*, 31(6):82:1–16, 2021.

[6] S. Kucherenko, O. V. Klymenko, and N. Shah. Sobol´ indices for problems defined in non-rectangular domains. *Reliability Engineering&System Safety*, 167:218–231, 2017.

[Elmar Plischke; Helmholtz-Zentrum Dresden-Rossendorf, Bautzner Landstr. 400, 01328 Dresden]
[`e.plischke@hzdr.de` – `https://artefakte.rz-housing.tu-clausthal.de/epl/`]