Quantile oriented Shapley effect via projected random forest

Ri Wang
Véronique Maume-Deschamps
Universite Claude Bernard Lyon 1, CNRS, Ecole Centrale de Lyon, INSA Lyon, Université Jean Monnet, ICJ UMR5208, 69622 Villeurbanne, France.

Clémentine Prieur
Universite Grenoble Alpes, CNRS, Inria, Grenoble INP, LJK, 38000, Grenoble, France

Global Sensitivity Analysis (GSA) is an important tool to better understand the behavior of black box models. Among the numerous methods for GSA, variance-based approaches have received much attention (Sobol' indices introduced in [1]). Only a few papers focus on Quantile Oriented Sensitivity Analysis (QOSA), which can help in analysing the behavior of the response at different quantile levels [2 , 3, 4, 5]. In [6], we introduced a new estimation procedure of QOSA indices based on the notion of projected random forest [7], with the initial random forest built from a criterion designed for quantiles: the pinball loss also known as quantile loss [8], with theoretical guarantees. Although informative, QOSA indices suffer from the drawback that they do not obey, even in the framework of independent inputs, any analogue to the variance decomposition offered by Sobol' indices through the theorem of Hoeffding [9]. This is the main reason why [5] introduced new indices based on the Shapley values [10]. While [11] introduced Shapley effects as variance-based measure importance, [12] suggested to adapt the value function to reach information on quantiles. In the present work, we propose to estimate the so-called quantile-oriented Shapley effects (QOSE) by combining the projection [7] of random forests built with from the quantile oriented criterion we introduced in [6], as far as arguments from Lundberg et al. Algorithm [13]. We implement our estimation procedure on both analytical examples and real data.

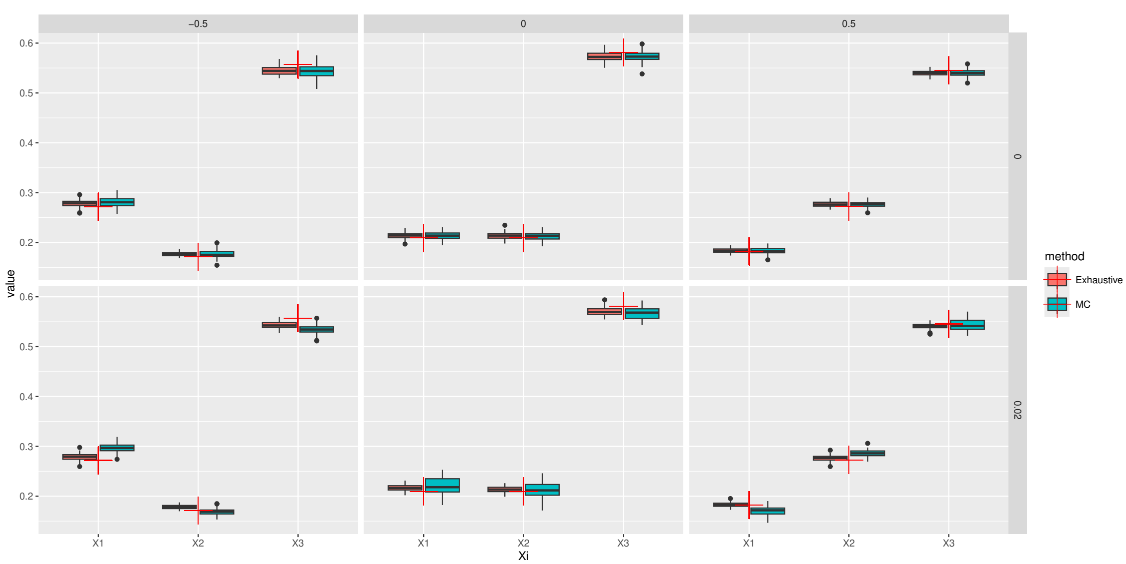

Figure 1: Box-plot on toy example with 100 repetitions. Red cross is ground truth. Red box is estimation from exhaustive search, blue box is our Monte-Carlo approximation result.

Model setting in Fig. 1:

$Y = \boldsymbol{\beta}^\intercal \boldsymbol{X}$, with $\boldsymbol{\beta} = (1,1,1)^\intercal$. $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu} = (0,0,0)^\intercal, \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & \rho\sigma_2\sigma_3 \\ 0 & \rho\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix}, \sigma_1 = \sigma_2 = 1, \sigma_3 = 2, \rho = -0.5, 0, 0.5$. Sample size $n = 2000$.

References:

[1] Sobol' I. M., "Sensitivity estimates for nonlinear mathematical models", *Mathematical Modelling and Computational Experiments*, **1**(4):407–414, 1993.

[2] Thomas Browne, Jean-Claude Fort, Bertrand Iooss, Loïc Le Gratiet, "Estimate of quantile-oriented sensitivity indices", *hal-01450891*, 2017.

[3] Véronique Maume-Deschamps and Ibrahima Niang, "Estimation of quantile oriented sensitivity indices", *Statistics & Probability Letters*, **134**:122–127, 2018.

[4] Zdeněk Kala, "Quantile-oriented global sensitivity analysis of design resistance", *Journal of Civil Engineering and Management*, **25**(4):297–305, 2019.

[5] Kévin Elie-Dit-Cosaque and Véronique Maume-Deschamps, "Random forest based quantile-oriented sensitivity analysis indices estimation", *Computational Statistics* , **39**(4):1747–1777, 2024.

[6] Ri Wang, Véronique Maume-Deschamps, Clémentine Prieur, "Quantile oriented sensitivity analysis with random forest based on pinball loss", *hal-04606380*, 2024.

[7] Clément Bénard, Gérard Biau, Sébastien Da Veiga, Erwan Scornet, "SHAFF: Fast and consistent SHApley eFfect estimates via random Forests", *International Conference on Artificial Intelligence and Statistics*, PMLR, **151**:5563-5582, 2022.

[8] Bhat Harish S., Nitesh Kumar, Garnet J. Vaz. "Towards scalable quantile regression trees", *2015 IEEE International Conference on Big Data (Big Data)*, IEEE, 53–60, 2015.

[9] Hoeffding Wassily, "A Class of Statistics with Asymptotically Normal Distribution." *The Annals of Mathematical Statistics*, **19**(3): 293–325, 1948.

[10] Shapley Lloyd S., "A value for n-person games", *Contribution to the Theory of Games*, **2**:307–317, 1953.

[11] Owen, Art B., "Sobol'indices and Shapley value", *SIAM/ASA Journal on Uncertainty Quantification*, **2**(1):245–251, 2014.

[12] Kévin Elie-Dit-Cosaque and Véronique Maume-Deschamps, "On quantile oriented sensitivity analysis", *arXiv:2102.05895*, 2021

[13] Lundberg Scott M. and Su-In Lee, "A Unified Approach to Interpreting Model Predictions", *Advances in Neural Information Processing Systems*, **30**:4765–4774, 2017.

[Ri Wang; Universite Claude Bernard Lyon 1, ICJ UMR5208]

[`ri.wang@math.univ-lyon1.fr` – `https://github.com/Ri0016`]