MASCOT-NUM 2022
Clermont Ferrand, France
7-9 June 2022

Robustness assessment using quantile-constrained Wasserstein projections

Marouane Il Idrissi
EDF R&D, Institut de Mathématiques de Toulouse (IMT), SINCLAIR AI Lab

Supervisor(s): J.-M. Loubes (IMT), F. Gamboa (IMT), N. Bousquet (EDF R&D, SINCLAIR) and B. Iooss (EDF R&D, SINCLAIR)

PhD expected duration: Mar. 2021 - Mar. 2024

Address: EDF Lab Chatou, 6 Quai Watier, 78401 Chatou

E-mail: marouane.il-idrissi@edf.fr

Abstract:

The robust analysis of predictive models, whether they are phenomenological or learned from data, requires the ability to propose interpretable perturbations of the inputs (or features) and outputs (or labels) of these models. These are random variables with known or empirically manipulable distributions. An important issue is therefore to propose an interpretable and minimally subjective approach to perturbing these distributions, adapted to the partially available or assumed information about them (e.g., support, shape, expert's assessment on some of the inputs statistics).

In the sensitivity analysis (SA) literature, several perturbation mechanisms have been proposed, based on the Kullback-Leibler (KL) divergence [4], or based on the Fisher metric [2]. These perturbation schemes lead to the class of "perturbed-law indices" (PLI), aiming at assessing input importance through the study of sensitivity of the model output with respect to the amplitude of the perturbation. In the machine learning (ML) community, KL divergence based perturbation have also been implemented in order to assess feature importance [1]. These methods rely on three main ingredients:

- Knowledge of the distribution, or the observation of an i.i.d. sample of the inputs/features, denoted P ;
- The choice of a metric (or quasi-metric) $\mathcal{D}$ on a particular space of probability measures, denoted $\mathcal{P}$;
- The definition of a set of probability measures respecting the desired perturbations, denoted $\mathcal{C}$.

Whenever only an i.i.d. sample of the inputs is observed, P is defined as the empirical probability measure associated with the sample. The distributional perturbation problem can then be formalized as the projection Q of P onto $\mathcal{C}$ with respect to $\mathcal{D}$, in other words:

$$Q = \underset{G \in \mathcal{P}}{\operatorname{argmin}}\ \mathcal{D}(P, G)$$
$$\text{s.t. } G \in \mathcal{C}$$

This work explores the particular choice of the 2-Wasserstein distance as a metric between probability measures, and more specifically on the space $\mathcal{P}_2(\mathbb{R})$ of real-valued probability measures with finite 2-nd order moments, defined for $P, G \in \mathcal{P}_2(\mathbb{R})$, as [6]:

$$W_2 = \int_0^1 \left(F_P^{\rightarrow}(x) - F_Q^{\rightarrow}(x)\right)^2 dx$$

where $F_G^{\rightarrow}$, for a probability measure $G \in \mathcal{P}_2(\mathbb{R})$, denotes the right-continuous generalized inverse of the probability distribution F_G of G, i.e.,

$$F_G^{\rightarrow}(y) := \sup\{x \in \overline{\mathbb{R}} \mid F_G(x) \leq y\} = \inf\{x \in \overline{\mathbb{R}} \mid F_G(x) > y\}.$$

Additionally, $\mathcal{C}$ is characterized as constraints over the quantile values of the desired optimally perturbed projection Q.

It can be shown that the resulting optimization problem can be equivalently written as an $L^2([0,1])$ projection of $F_P^{\rightarrow}$, with respect to the usual L^2 norm, with monotonicity and interpolation constraints. In its simplest form, it admits an analytical solution: atoms with a specific mass are added to P at the desired quantile values, in order to verify the quantile constraint.

In order to further take advantage of the equivalent representation of the distributional perturbation problem, and to better control the shape of the solution, the choice of restricting the solution to piece-wise continuous monotone polynomials is explored. This restriction can be written as a convex constrained quadratic program, through the representation of sum-of-squares polynomials using semi-definite matrices [3, 5]. Figure 1 illustrates this particular projection of $F_P^{\rightarrow}$, and the subsequent simulated sample obtained from the solution, on the `airquality` dataset.

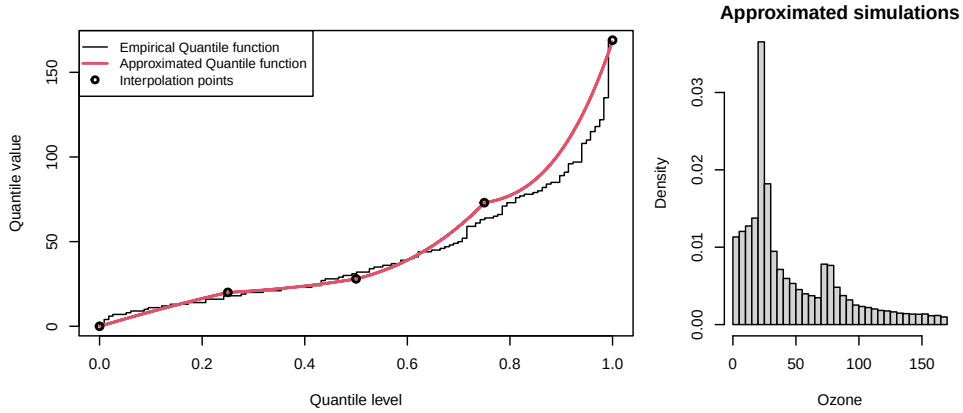

Figure 1: Piece-wise continuous monotone approximation of the Ozone variable under quantile constraints, and resulting simulations.

References

[1] F. Bachoc, F. Gamboa, M. Halford, J-M. Loubes, and L. Risser. Explaining Machine Learning Models using Entropic Variable Projection. *arXiv:1810.07924 [cs, stat]*, December 2020. arXiv: 1810.07924.

[2] C. Gauchy, J. Stenger, R. Sueur, and B. Iooss. An information geometry approach to robustness analysis for the uncertainty quantification of computer codes. *Technometrics*, 64(1):80–91, 2022.

[3] J-B. Lasserre. *An Introduction to Polynomial and Semi-Algebraic Optimization*. Cambridge Texts in Applied Mathematics. Cambridge University Press, Cambridge, 2015.

[4] P. Lemaître, E. Sergienko, A. Arnaud, N. Bousquet, F. Gamboa, and B. Iooss. Density modification-based reliability sensitivity analysis. *Journal of Statistical Computation and Simulation*, 85(6):1200–1223, 2015.

[5] P. A. Parrilo. *Semidefinite Optimization and Convex Algebraic Geometry*, chapter 3, pages 47–157. 2012.

[6] C. Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, March 2003. ISSN: 1065-7339.

Short biography – After graduating from ENSAI and Rennes 1 University in 2020, I started a CIFRE PhD track in 2021 at EDF R&D and Institut de Mathématiques de Toulouse, working on the development of interpretability methods for ML models. My research interests are at the crossroads between sensitivity analysis and explainable artificial intelligence methods, and more specifically applied cooperative game theory and probability measure perturbations.