Are all distributions distorted?

M. Capaldo[a], J.M. Arevalillo[b] and J. Navarro [c]

[a] *Institute of Statistics, RWTH Aachen University, 52056 Aachen, Germany*
[b] *Department of Statistics and Operational Research, University Nacional Educación a Distancia (UNED), Madrid, Spain*
[c] *Departamento de Estadística e Investigación Operativa, Facultad de Matemáticas, Universidad de Murcia, Murcia, Spain*

Distortion functions were introduced by Yaari [5] within the theory of choice under risk, aiming to perform a "distortion" on the initial risk distribution. Specifically, an increasing continuous function $q : [0, 1] \to [0, 1]$ such that $q(0) = 0$ and $q(1) = 1$ is named distortion function. Then, given a cumulative distribution function (CDF) F, the distorted CDF from F through q is $F_q = q(F)$. Similarly for the survival function (SF) $\overline{F} = 1 - F$ by considering the dual distortion function with respect to q, namely $\tilde{q}(u) = 1 - q(1 - u)$, for $u \in [0, 1]$. Indeed, $\overline{F}_q = \tilde{q}(\overline{F})$ is the distorted SF of F_q. Nowadays distorted distributions are employed in several fields, such as Order Statistics, Reliability Theory and Survival Analysis, to name a few (see Navarro [2] and references therein). In this talk, we define and study new distortion functions based on the well-known Receiver Operating Characteristic (ROC) curve. They can be used to provide the distortion function which connects two CDFs (or two SFs) when distortion-based model assumptions are unknown. We also consider new distortions which are related with the concept of partial area under the ROC curve through recent univariate skewed models (see Navarro and Arevalillo [3]). Note that, any distortion function is a CDF with support included in $[0, 1]$. In this sense, various stochastic orders and aging classes (cf. Shaked and Shanthikumar [4]) are characterized by using the interpretation of ROC-based distortions as CDFs of suitable relative random variables. Some connections with equilibrium distribution, Lorenz curve and Gini's index are provided too. A real data application of the distortion-based results allows us to evaluate the performance of some Machine Learning classifiers by using semiparametric estimations of the ROC. The talk is based on Capaldo et al. [1].

References:

1. Capaldo M., Arevalillo J M., & Navarro J. (2025). Distorted distributions and ROC curves. *Scandinavian Journal of Statistics* **52**, 1786–1815.

2. Navarro J. (2022). *Introduction to System Reliability Theory*. Springer.

3. Navarro J., Arevalillo J.M. (2023). On connections between skewed, weighted and distorted distributions: applications to model extreme value distributions. *Test* **32**, 1307–1335.

4. Shaked M., Shanthikumar J.G. (2007). *Stochastic Orders*. Springer, New York.

5. Yaari M.E. (1987). The dual theory of choice under risk. *Econometrica: Journal of the Econometric Society* **55**, 95–115.