Layman's abstract

Measuring the strength of association between two variables is of essential importance in many research fields. How do we define such measures? What properties should such measures possess? How do we choose an appropriate measure in applications?

The most commonly used measure is the Pearson correlation. This measure is based on the covariance between two variables, which is optimal for the linear relationship between bivariate normal variables. However, the Pearson correlation performs poorly for variables with heavily-tailed or asymmetric distributions, and may be seriously impacted even by a single outlier. As a robust alternative, the Spearman correlation is defined as the covariance between the cumulative distribution functions (or ranks) of two variables, but it may loss the efficiency. Complementing these two measures, the traditional Gini correlations are based on the covariance between one variable and the rank of the other, and hence well balance in efficiency and robustness. The Gini correlations, however, are not symmetric due to different roles of two variables. This asymmetry violates the axioms of correlation measurement.

In this paper, the authors propose the symmetric Gini correlation based on the joint rank function, which takes more dependence information than the marginal rank in the traditional Gini correlations. The properties of the symmetric Gini correlation are fully explored. Theoretical results on efficiency and robustness are obtained. Numerical studies demonstrate that the proposed correlation have satisfactory performance under a variety of situations. The proposed symmetric Gini correlation provides an attractive option for measuring correlation.

Although its formulation is natural, the symmetric Gini loses an intuitive interpretation. It is also more difficult to compute than the Pearson correlation (R code is available at http://olemiss.edu/~xdang/Rcodes/Gcor.R). When applying the proposed measure, one may consider the trade-off among efficiency, robustness, computation and interpretability.