Combinatorial regression model in abstract simplicial complexes

Andrej Srakar, Institute for Economic Research (IER) and School of Economics and Business, University of Ljubljana, Slovenia, srakara@ier.si

Abstract

In regression analysis of the market share data four main parametric type models are prevalent: multinomial logistic regression, attraction models of various types, Dirichlet covariance models, and compositional regression. Nonparametric regressions in simplex spaces include local polynomial (Di Marzio et al, 2015), simplicial spline (Machalová, Hron and Talská, 2019) and simplicial wavelet (Srakar and Fry, 2019) approaches. We extend this arsenal of possibilities with a new type, a completely novel regression perspective, labelled combinatorial regression, based on combining n-tuplets of sampling units into groups and treating them in abstract simplical complex spaces (Lee, 2011; Korte, Lovász and Schrader, 1991). The novel perspective, estimated using Bradley-Terry based maximum likelihood approach and in combination with Multivariate Distance Matrix Regression approach (Anderson, 2001; McArdle and Anderson, 2001), allows extensive number of perspectives in the analysis of, for example, triplets, quadruplets or quintuplets (or any n-tuplet) and using as measure of disparity between the units (to construct regressors) different divergence measures. It also allows applications to very small datasets as the number of units in the new model can be expressed in terms of factorial products of units of original sample. We provide the analysis of new approach for different n-tuple combinations using Jensen-Shannon and generalized Jensen-Shannon divergence measures and provide the Gaussian asymptotic limits of the approach with exploring its properties also in a Monte Carlo simulation study. In a short application we analyze sessile hard-substrate marine organisms image data from Italian coast areas which allows to explore the new approach in relative abundance data setting.

Joint work with Prof. Miroslav Verbič, University of Ljubljana, Slovenia.
Presented previously on CFE-CMStatistics Conference 2020 and ICAS2020 – The 14th International Conference on Applied Statistics.

Basic keywords: regression models, abstract simplicial complexes, symplectic data, algebraic statistics, algebraic topology, Gaussian processes, Monte Carlo, divergence measures