Contributed talk
"Asymptotic behaviour of U-statistics on row-column exchangeable matrices"

Tâm Le Minh*

*Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France

This is the abstract of a contributed talk given at the Bernoulli-IMS 11th World Congress in Probability and Statistics in Bochum, in 2024. It is based on the works [2, 1].

We suggest a methodology to analyze bipartite network data. This methodology is based on U-statistics of row-column exchangeable matrices. Row-column exchangeable matrices are random matrices, the joint probability distribution of which is invariant by simultaneous permutations of rows and columns. U-statistics correspond to the class of statistics defined as the empirical mean of a function of a subset, over all subsets of observations. U-statistics of matrices are the average of a submatrix function over the entire matrices. In network analysis, row-column exchangeable matrices are the adjacency matrices of bipartite node-exchangeable networks and U-statistics can be used as estimators of quantities of interest.

Our work focuses on the asymptotic behavior of the U-statistics of row-column exchangeable matrices. We derive a decomposition for them, based on orthogonal projections on probability spaces generated by sets of Aldous-Hoover-Kallenberg variables [2, 1]. The specificity of these sets is that they are indexed by bipartite graphs, which allows for the use of concepts from graph theory to describe this decomposition. The decomposition is used to identify the asymptotic behavior of U-statistics of row-column exchangeable matrices, including in degenerate cases. In particular, it depends only on a few terms of the decomposition, corresponding to the non-zero elements that are indexed by the smallest graphs, named principal support graphs. Hence, we show that the asymptotic behavior of a U-statistic and its degeneracy are characterized by the properties of their principal support graphs. Indeed, their number of nodes gives the convergence rate of a U-statistic to its limit distribution. Specifically, the latter is degenerate if and only if this number is strictly greater than 1. Also, when the principal support graphs are connected, we find that the limit distribution is Gaussian, even in degenerate cases.

These statistical developments are applied to the analysis of bipartite networks, more specifically mutualistic ecological networks. Many ecological questions are interested in the general structure of networks rather than the collection of present species. This makes exchangeable random network models, the adjacency matrices of which are row-column exchangeable, well-suited to analyze these networks. U-statistics are used as estimators of quantities of interest such as the degree heterogeneity, motif densities or graphon metrics. It possible to obtain statistical guarantees on these estimators, for example in the form of confidence intervals, owing to the theoretical results and the methodology developed in this work.

References

[1] T. Le Minh. Characterization of the asymptotic behavior of U-statistics on row-column exchangeable matrices. *arXiv preprint arXiv:2401.07876*, 2024.

[2] T. Le Minh, S. Donnet, F. Massol, and S. Robin. Hoeffding-type decomposition for U-statistics on bipartite networks. *arXiv preprint arXiv:2308.14518*, 2023.