Research Statement

Reza Mohammadi

My research interests lie primarily in developing statistical learning methods to analysis high-dimensional data which routinely arise in econometrics and machine learning, neuroscience, and social science. I began my research career by developing statistical methods for mixture distributions with application in queuing systems. My current research is focused on developing Bayesian statistical methods for multivariate statistical analysis to understanding the underlying mechanisms in complex systems. These methods have applications in a wide variety of disciplines, such as capturing causal relationships between brain activities for treatment of psychiatric disorders. My motivation to develop these methods is to encode the relationships between the components in high-dimensional scale. As a continuation of my career, I see myself using my strong statistical background and advanced programming skills to develop statistical methods and softwares to tackle the important scientific questions in Data Science. Following, I describe my scientific works and my future plans.

1 Current Research: Big Data Analysis

Discovering complex relationships among large numbers of variables is one of the main objectives in modern science. For instance, convolutional neural networks play an important role for detecting neuroimaging biomarkers that predict treatment response from fMRI data. in this regards, graphical models provide a powerful framework to uncover complicated patterns and are commonly used in machine learning and Bayesian statistics.

My contribution in this area has been to develop several statistical learning methods for modeling networks by means of graphical models with applications to fMRI research, neuroscience and genetics (Mohammadi, 2015). Besides, to make my statistical methods easily accessible to other researchers, I design two softwares which implements our statistical methods (Mohammadi and Wit, 2019b) and (Mohammadi, 2019b).

In Dobra et al. (2018), we develop a new algorithm for selecting graphical loglinear models that is suitable for analyzing hyper-sparse contingency tables. We show how multi-way contingency tables can be used to represent patterns of human mobility. We analyze a dataset of geolocated tweets from South Africa that comprises 46 million latitude/longitude locations of 476,601 Twitter users that is summarized as a contingency table with 214 variables.

In Dyrba et al. (2018), we propose the novel application of graphical models to study the interregional associations and dependencies between multimodal imaging markers of Alzheimer disease pathology and to compare different hypotheses regarding the spread of the disease. By measuring mean amyloid load, glucose metabolism, and gray matter volume from the six principle nodes of the default mode network- a functional network of brain regions that appears to be preferentially targeted by Alzheimer disease.

In Mohammadi and Wit (2015), we develop statistical method for Gaussian graphical model determination. We construct an efficient search algorithm which explores the graph space to distinguish important links from irrelevant ones and detect the underlying graph structure with high accuracy. In principle, our search algorithm is a trans-dimensional MCMC approach based on a birth-death process with an appropriate stationary distribution. In the paper, we cover the theory and computational details of our method, which is easy to implement and computationally feasible for high-dimensional graphs. To describe how our method can solve the important applied

problems, we apply the method on two real-world applications in genetics. In the first example, we apply our method to the high-dimensional human gene expression dataset from B-lymphocyte cells to discover the complex relationship among genes. In the second example, we focus on graphical models involving time series data to demonstrate how well our proposed method can be extended to other types of graphical models.

The statistical methodology that we propose in Mohammadi and Wit (2015) is not limited only to the Gaussian graphical models. In Mohammadi and Wit (2014) we explain about the extension of our method to deal with outliers, by using robust graphical modeling using Dirichlet **t**-Distributions. In Mohammadi and Kaptein (2016) we explain about the extension of our method for Bayesian regression trees.

The method that we propose in Mohammadi and Wit (2015) is limited only to the data that follows the Gaussianity assumption. In Mohammadi et al. (2017) we present a Bayesian approach for graphical model determination based on a Gaussian copula approach that can deal with continuous, discrete, or mixed data. We embed a graph selection procedure, based on a computationally efficient search algorithm, inside a semiparametric Gaussian copula. We implement our statistical method to discover the effect of potential risk factors of Dupuytren disease and uncover the underlying patterns in our multivariate dataset.

2 Previous Research: Mixture Distributions

Mixture models provide fascinating tools to more closely estimate, predict and infer about complex systems.

In Mohammadi and Salehi-Rad (2012), we introduce a statistical method for the mixture of truncated Normal distributions. We apply our work to approximate the service and re-service time densities in M/G/1 queuing systems with optional second service. Based on our statistical methods, we estimate some performance measures e.g. predictive densities and some performance measures related to this queuing system such as stationary system size and waiting time. We apply the theories in practice by providing an example based on a real data.

In Mohammadi et al. (2013) we develop a Bayesian statistical method on a mixture of Gamma distributions. We implement our method to approximate both the general service and re-service times densities in M/G/1 queuing systems. We estimate the system parameters, predictive densities and some performance measures related to the queuing system such as stationary system size and waiting time.

3 Developing Software

In modern science, developing statistical methods requires to designing statistical software for the applications of the methods in practice. To disseminate the computational tools as an user-friendly software package, I design three softwares as an `R` packages called `BDgraph` (Mohammadi and Wit, 2019b), `ssgraph` (Mohammadi, 2019b), and `bmixture` (Mohammadi, 2019a) which are freely available online.

The `BDgraph` and `ssgraph` packages perform structure learning for graphical models with either continuous or discrete variables. In the packages, I implement recent improvements in the statistical literature, including the methods that I have developed (Mohammadi et al., 2014, Mohammadi and Wit, 2015, Mohammadi et al., 2017, Dobra et al., 2018, Mohammadi et al., 2017).

The core of the `BDgraph` and `ssgraph` packages consists of several search algorithms which implemented in `C++` and interfaced with `R`, to speed up the computations. Besides, the packages contain several functions for simulation and visualization, as well as several multivariate datasets taken from the literature. The packages are well-documented with examples and are regularly maintained. I designed the packages as user-friendly softwares, which include the variety of search algorithms in graphical models. The manuscripts related to these packages are (Mohammadi and Wit, 2019a, Mohammadi and Dobra, 2017).

4 Future Research

My future research line is to develop novel and computationally efficient algorithms based on novel statistical methods to being able to more closely estimate, predict and infer the complex systems in high-dimensional settings. Given my background in statistics and my ability in programing, I very much like to do research in different aspects of statistics with both applied and theoretical contributions. In this regards, I am willing to collaborate with researchers mainly in applied statistics, econometrics and machine learning, which gives me a great opportunity to develop statistical methods that address the important problems in Data Science.

References

Dobra, A., R. Mohammadi, et al. (2018). Loglinear model selection and human mobility. *The Annals of Applied Statistics 12*(2), 815–845.

Dyrba, M., M. J. Grothe, A. Mohammadi, H. Binder, T. Kirste, S. J. Teipel, A. D. N. Initiative, et al. (2018). Comparison of different hypotheses regarding the spread of alzheimer's disease using markov random fields and multimodal imaging. *Journal of Alzheimer's Disease 65*(3), 731–746.

Mohammadi, A. (2015, April). *Bayesian Model Determination in Complex Systems*. Ph. D. thesis, University of Groningen.

Mohammadi, A., F. Abegaz Yazew, E. van den Heuvel, and E. C. Wit (2017). Bayesian modelling of dupuytren disease using gaussian copula graphical models. *Journal of Royal Statistical Society-Series C 66*(3), 629–645.

Mohammadi, A., F. Abegaz Yazew, and E. C. Wit (2014). Bayesian copula gaussian graphical modelling. *(IWSM'14) Proceedings of the 29th International Workshop on Statistical Modelling 1*, 225–230.

Mohammadi, A. and A. Dobra (2017). The R package BDgraph for Bayesian structure learning in graphical models. *ISBA Bulletin 4*, 11–16.

Mohammadi, A. and M. Kaptein (2016). Contributed discussion on article by pratola. *Bayesian Analysis 11*(3), 938–940.

Mohammadi, A., H. Massam, and G. Letac (2017). The ratio of normalizing constants for bayesian graphical gaussian model selection. *arXiv preprint arXiv:1706.04416*.

Mohammadi, A. and M. R. Salehi-Rad (2012). Bayesian inference and prediction in an m/g/1 with optional second service. *Communications in Statistics-Simulation and Computation 41*(3), 419–435.

Mohammadi, A., M. R. Salehi-Rad, and E. C. Wit (2013). Using mixture of gamma distributions for bayesian analysis in an m/g/1 queue with optional second service. *Computational Statistics 28*(2), 683–700.

Mohammadi, A. and E. C. Wit (2014). Contributed discussion on article by finegold and drton. *Bayesian Analysis 9*(3), 577–579.

Mohammadi, A. and E. C. Wit (2015). Bayesian structure learning in sparse gaussian graphical models. *Bayesian Analysis 10*(1), 109–138.

Mohammadi, R. (2019a). *bmixture: Bayesian Estimation for Finite Mixture of Distributions*. R package version 1.3.

Mohammadi, R. (2019b). *ssgraph: Bayesian Graphical Estimation using Spike-and-Slab Priors*. R package version 1.8.

Mohammadi, R. and E. C. Wit (2019a). BDgraph: An R package for Bayesian structure learning in graphical models. *Journal of Statistical Software 89*(3), 1–30.

Mohammadi, R. and E. C. Wit (2019b). *BDgraph: Bayesian Structure Learning in Graphical Models using Birth-Death MCMC*. R package version 2.59.