IMPS 2018

Columbia University · New York City, New York, USA · July 9-13, 2018

Abstracts

Talks

through cross-validation. To handle both problems, one may recur to penalized regression methods, such as VAR(1) with a lasso penalty. In this paper, we recommend a different solution, called principal component VAR(1) (PC-VAR(1)), in which the variables are first reduced to a few components; next the VAR(1) analysis is applied to the components. Reanalyzing data of a single participant of the COGITO study, we show that PC-VAR(1) has the better predictive performance, and that networks based on PC-VAR(1) give a more informative representation of both the lagged and the contemporaneous relations among the variables.

Symposium 6: 3:20 PM – 4:50 PM

Chair: Willem J. Heiser; Antonio D'Ambrosio

Symposium 6: Individual Differences in Rankings: Aggregation, Representation, Evolution & Prediction

Symposium 6 - Parallel Session: 3.2A: WEIGHTED AGGREGATION OF ORDINAL AND CARDINAL DATA
Josè Luis Garcìa-Lapresta, University of Valladolid; Casilda Lasso De La Vega, Universidad del País Vasco (University of the Basque Country)

In some real problems analyzed in Welfare Economics, Sociology, Psychology, Marketing, Psychological Measurement, and other fields of research, some alternatives are evaluated in several criteria (or dimensions) with the purpose of rank order these alternatives. Alternatives can be countries, regions, products, services, or persons. The criteria on which the alternatives are evaluated can be of different nature. Some criteria may correspond to dichotomous (or binary) variables, as access to electricity, water or Internet, migration status, etc. Other criteria are assessed by linguistic terms of qualitative (or categorical or ordinal) scales, as health status, personal security, environmental quality, educational attainments, etc. In addition, others can be measured through numbers of a quantitative scale, as income, Gross National Product, unemployment rate, life expectancy, literacy rate, etc. Since criteria usually have different importance, frequently they have associated numerical weights. In this contribution some procedures that generate a reciprocal preference relation on the set of alternatives for each criterion are introduced. These kinds of preference relations capture the intensity of preference between each pair of alternatives. Since every weighted average of reciprocal preference relations is also a reciprocal preference relation, overall intensities of preference between pairs of alternatives are obtained, taking into account the weights assigned to the criteria. Then, for each fixed threshold, a ranking on the set of alternatives is attained.

Symposium 6 - Parallel Session: 3.2B: ORDINAL UNFOLDING OF PREFERENCE RANKINGS USING THE KEMENY DISTANCE
Antonio D'Ambrosio, University of Naples Federico II; Willem J. Heiser, Leiden University

Multidimensional unfolding can be seen as a special case of Multidimensional Scaling (MDS) with two sets of points, in which the within sets proximities are missing. Lack of the within-sets information is the major cause of well-known problems of degenerate solutions in analytical procedures, especially for ordinal (or non-metric) unfolding. Over the years, several approaches to avoid degenerate solutions in unfolding have been developed. Our approach belongs to the category of methods that aim to extend the unfolding data with information on the dissimilarities between rankings (Van Deun et al, 2007). By starting from a typical rectangular matrix, in which each row is a preference ranking, we propose a reconstruction strategy of the entire dissimilarity matrix based on the properties of the Kemeny distance (Kemeny and Snell, 1962) and the τ_X extended rank correlation coefficient (Emond and Mason, 2002). We show that our unfolding procedure can be used with any standard MDS program and produces non-degenerate solutions. These solutions are simple to compute, while comparable in quality with the ones returned by the PREFSCAL algorithm (Busing, Groenen and Heiser, 2005), currently the state-of-the-art method for avoiding degeneracies in unfolding.

Symposium 6 - Parallel Session: 3.2C: RANKINGS VARYING IN TIME: A BAYESIAN APPROACH
Elja Aryas, University of Helsinki

Rankings of individual persons, when based on repeated measurement of their performance, will generally vary in time. Moreover, some persons can be absent from the tests that are made either because of delayed entry into the study or early departure, or purely randomly. In this talk a Bayesian version of the well-known Mallows model for rank data is introduced, and then extended to situations in which the rankings can vary in time and be incomplete. The consequent missing data problems are handled by applying, within the considered MCMC, Bayesian data augmentation. The method is illustrated by analyzing some aspects of a data set describing the academic performance, measured by a series of tests, of a class of high school students over a period of four years.

Symposium 6 - Parallel Session: 3.2D: DISTANCE-BASED DECISION TREES FOR RANKING DATA: THE ROLE OF THE WEIGHT SYSTEMS
Mariangela Sciandra, University of Palermo; Antonella Plaia, University of Palermo

In everyday life ranking and classification are basic cognitive skills that people use in order to grade everything that they experience. Grouping and ordering a set of elements is considered easy and communicative; thus, rankings of sport-teams, universities, countries and so on are often observed. A particular case of ranking data is represented by preference data, where individuals show their preferences over a set of items. When individuals specific characteristics are available, an important issue concerns the identification of the profiles of respondents (or judges) giving the same/similar rankings. In order to incorporate respondent-specific covariates distance-based decision tree models (D'Ambrosio 2007, Lee and Yu 2010, Yu et al. 2010, D'Ambrosio and Heiser, 2016, Plaia and Sciandra, 2017) have been recently proposed. Actually, it can happen that one or some of the k items is more important than others, or, similarly, the top of the ordering can deserve more attention than the bottom. In these situations, changing the rank of very important items or changing the top of the ranking require different "weighting". In this contribution we want analyze the role of element and positional information (Kumar and Vassilvitskii 2010) when some distance measures for rankings are evaluated. Several weighting structures will be assumed for both positional and item weights, and we aim at identifying some particular behavior in the distance measures used. Analysis will be carried out both by simulation and by application to real dataset, especially in the framework of tree-based methods for rank data.

Causal Inference and Mediation: 3:20 PM – 4:50 PM

Chair: Joost Van Ginkel

Causal Inference and Mediation

Causal Inference and Mediation - Parallel Session: 3.3A: LOWER-LEVEL MEDIATION IN BINARY SETTINGS
Haeike Josephy, Ghent University; Tom Loeys, Ghent University

In recent literature, researchers have put a lot of time and effort in expanding mediation to multilevel designs. Unfortunately, such extensions are often limited a continuous mediator and outcome, whereas research concerning multilevel mediation with a binary mediator and outcome remains rather sparse. Additionally, in lower-level mediation, the effect of the lower-level mediator on the outcome may oftentimes be confounded by an (un)measured upper-level variable. When such confounding is left unaddressed, the effect of the mediator, as well as the causal mediation effects, will be estimated with bias. In linear settings, bias due to unmeasured additive upper-level confounding is often remedied by separating the effect of the mediator into a within- and between-cluster component, but unfortunately, this solution no longer works when considering binary settings. To assess the severity of this transgression, we aim to tackle binary lower-level mediation from a counterfactual point of view (with a special focus on small clusters), by 1) providing non-parametrical identification assumptions of the direct and indirect effect, 2) parametrically identifying these effects based on multilevel logit-or probit-models, 3) considering estimation models for the mediator and the outcome, and 4) estimating the causal effects through an imputation algorithm that samples counterfactuals. Since steps three and four can be