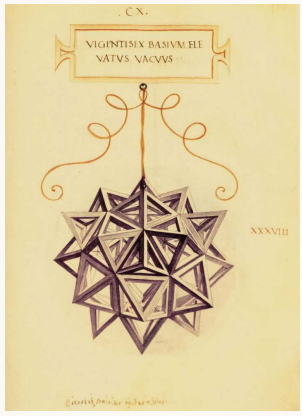
DIPARTIMENTO DI ECONOMIA E GIURISPRUDENZA
UNIVERSITÀ DI CASSINO E DEL LAZIO MERIDIONALE

UNICAS 1979>2019
40 ANNI DALLA FONDAZIONE
UNIVERSITÀ DI CASSINO E DEL LAZIO MERIDIONALE

12-TH SCIENTIFIC MEETING
CLASSIFICATION AND DATA ANALYSIS

CLADAG 2019

11-13 SEPTEMBER 2019
CASSINO



Book of Short Papers

Giovanni C. Porzio
Francesca Greselin
Simona Balzano
Editors

SIS Società Italiana di Statistica

Contents

Keynotes

Invited and contributed sessions

Random projections of variables and units
Laura Anderlucci, Roberta Falcone, Angela Montanari

Sparse linear regression via random projections ensembles
Laura Anderlucci, Matteo Farnè, Giuliano Galimberti, Angela Montanari

High-dimensional model-based clustering via random projections
Laura Anderlucci, Francesca Fortunato, Angela Montanari

Evaluating the school effect: adjusting for pre-test or using gain scores?
Bruno Arpino, Silvia Bacci, Leonardo Grilli, Raffaele Guetto, Carla Rampichini

ACE, AVAS and robust data transformations
Anthony Atkinson

Mixtures of multivariate leptokurtic Normal distributions
Luca Bagnato, Antonio Punzo, Maria Grazia Zoia

Detecting and interpreting the consensus ranking based on the weighted Kemeny distance
Alessio Baldassarre, Claudio Conversano, Antonio D'Ambrosio

Predictive principal components analysis
Simona Balzano, Maja Bozic, Laura Marcis, Renato Salvatore

Flexible model-based trees for count data
Federico Banchelli

Euclidean distance as a measure of conformity to Benford's law in digital analysis for fraud detection
Mateusz Baryła, Józef Pociecha

The evolution of the purchase behavior of sparkling wines in the Italian market
Francesca Bassi, Fulvia Pennoni, Luca Rossetto

Modern likelihood-frequentist inference at work
Ruggero Bellio, Donald A Pierce

Ontology-based classification of multilingual corpuses of documents
Sergey Belov, Salvatore Ingrassia, Zoran Kalinić, Paweł Lula

Modeling heterogeneity in clustered data using recursive partitioning
Moritz Berger, Gerhard Tutz

Mixtures of experts with flexible concomitant covariate effects: a bayesian solution
Marco Berrettini, Giuliano Galimberti, Thomas Brendan Murphy, Saverio Ranciati

Sampling properties of an ordinal measure of interrater absolute agreement
Giuseppe Bove, Pier Luigi Conti, Daniela Marella

Tensor analysis can give better insight
Rasmus Bro

A boxplot for spherical data
Davide Buttarazzi, Giuseppe Pandolfo, Giovanni Camillo Porzio, Christophe Ley

Machine learning models for forecasting stock trends
Giacomo Camba, Claudio Conversano

Tree modeling ordinal responses: CUBREMOT and its applications
Carmela Cappelli, Rosaria Simone, Francesca Di Iorio

Supervised learning in presence of outliers, label noise and unobserved classes
Andrea Cappozzo, Francesca Greselin, Thomas Brendan Murphy

Asymptotics for bandwidth selection in nonparametric clustering
Alessandro Casa, José E Chacón, Giovanna Menardi

Foreign immigration and pull factors in Italy: a spatial approach
Oliviero Casacchia, Luisa Natale, Francesco Giovanni Truglia

Dimensionality reduction via hierarchical factorial structure
Carlo Cavicchia, Maurizio Vichi, Giorgia Zaccaria

Likelihood-type methods for comparing clustering solutions
Luca Coraggio, Pietro Coretto

Labour market analysis through transformations and robust multilevel models
Aldo Corbellini, Marco Magnani, Gianluca Morelli

Modelling consumers' qualitative perceptions of inflation
Marcella Corduas, Rosaria Simone, Domenico Piccolo

Noise resistant clustering of high-dimensional gene expression data
Pietro Coretto, Angela Serra, Roberto Tagliaferri

A compositional analysis approach assessing the spatial distribution of trees in Guadalajara, Mexico
Marco Antonio Cruz, Maribel Ortego, Elisabet Roca

Joining factorial methods and blockmodeling for the analysis of affiliation networks
Daniela D'Ambrosio, Marco Serino, Giancarlo Ragozini

A latent space model for clustering in multiplex data
Silvia D'Angelo, Michael Fop

Post processing of two dimensional road profiles: variogram scheme application and sectioning procedure
Mauro D'Apuzzo, Rose-Line Spacagna, Azzurra Evangelisti, Daniela Santilli, Vittorio Nicolosi

A new approach to preference mapping through quantile regression
Cristina Davino, Tormod Naes, Rosaria Romano, Domenico Vistocco

Network effect on individual scientific performance: a longitudinal study on an Italian scientific community
Domenico De Stefano, Giuseppe Giordano, Susanna Zaccarin

On the robustness of the cosine distribution depth classifier
Houyem Demni, Amor Messaoud, Giovanni C Porzio

Classify X-ray images using convolutional neural networks
Agostino Di Ciaccio, Federica Crobu

Penalized vs constrained maximum likelihood approaches for clusterwise linear regression modeling
Roberto Di Mari, Stefano Antonio Gattone, Roberto Rocci

Local fitting of angular variables observed with error
Marco Di Marzio, Stefania Fensore, Agnese Panzera, Charles C Taylor

Quantile composite-based path modeling to estimate the conditional quantiles of health indicators
Pasquale Dolce, Cristina Davino, Stefania Taralli, Domenico Vistocco

AUC-based gradient boosting for imbalanced classification
Martina Dossi, Giovanna Menardi

How to measure material deprivation? A latent Markov model based approach
Francesco Dotto

Decomposition of the interval based composite indicators by means of biclustering
Carlo Drago

Consensus clustering via pivotal methods
Leonardo Egidi, Roberta Pappadà, Francesco Pauli, Nicola Torelli

Robust model-based clustering with mild and gross outliers
Alessio Farcomeni, Antonio Punzo

A new proposal for building immigrant integration composite indicator
Mario Fordellone, Venera Tomaselli, Maurizio Vichi

Biodiversity spatial clustering
Francesca Fortuna, Fabrizio Maturo, Tonio Di Battista

A generalization of multivariate depth functions
Giacomo Francisci, Claudio Agostinelli, Alicia Nieto-Reyes

Skewed distributions or transformations? Incorporating skewness in a cluster analysis
Michael Gallaugher, Paul McNicholas, Volodymyr Melnykov, Xuwen Zhu

Robust parsimonious clustering models
Luis Angel Garcia-Escudero, Agustin Mayo-Iscar, Marco Riani

Projection-based uniformity tests for directional data
Eduardo García-Portugués, Paula Navarro-Esteban, Juan Antonio Cuesta-Albertos

Graph-based clustering of visitors' trajectories at exhibitions
Martina Gentilin, Pietro Lovato, Gloria Menegaz, Marco Cristani, Marco Minozzo

Symmetry in graph clustering
Andreas Geyer-Schulz, Fabian Ball

Bayesian networks for the analysis of entrepreneurial microcredit: evidence from Italy
Lorenzo Giammei, Paola Vicard

The PARAFAC model in the maximum likelihood approach
Paolo Giordani, Roberto Rocci, Giuseppe Bove

Structure discovering in nonparametric regression by the GRID procedure
Francesco Giordano, Soumendra Nath Lahiri, Maria Lucia Parrella

A microblog auxiliary part-of-speech tagger based on bayesian networks
Silvia Golia, Paola Zola

Recent advances in model-based clustering of high dimensional data
Claire Gormley

Tree embedded linear mixed models
Anna Gottard, Leonardo Grilli, Carla Rampichini, Giulia Vannucci

Weighted likelihood estimation of mixtures
Luca Greco, Claudio Agostinelli

A canonical representation for multiblock methods
Mohamed Hanafi

An adequacy approach to estimating the number of clusters
Christian Hennig

Classification with weighted compositions
Karel Hron, Julie Rendlova, Peter Filzmoser

MacroPCA: an all-in-one PCA method allowing for missing values as well as cellwise and rowwise outliers
Mia Hubert, Peter J Rousseeuw, Wannes Van den Bossche

Marginal effects for comparing groups in regression models for ordinal outcome when uncertainty is present
Maria Iannario, Claudia Tarantola

A multi-criteria approach in a financial portfolio selection framework
Carmela Iorio, Giuseppe Pandolfo, Roberta Siciliano

Clustering of trajectories using adaptive distances and warping
Antonio Irpino, Antonio Balzanella

Sampling and learning Mallows and generalized Mallows models under the Cayley distance: short paper
Ekhine Irurozki, Borja Calvo, Jose A Lozano

The gender parity index for the academic students progress
Aglaia Kalamatianou, Adele H. Marshall

Some asymptotic properties of model selection criteria in the latent block model
Christine Keribin

Invariant concept classes for transcriptome classification
Hans Kestler, Robin Szekely, Attila Klimmek, Ludwig Lausser

Clustering of ties defined as symbolic data
Luka Kronegger

Application of data mining in the housing affordability analysis
Viera Labudová, Ľubica Sipková

Cylindrical hidden Markov fields
Francesco Lagona

Comparing tree kernels performances in argumentative evidence classification
Davide Liga

Recent advancement in neural network analysis of biomedical big data
Pietro Liò, Giovanna Maria Dimitri, Chiara Sopegno

Bias reduction for estimating functions and pseudolikelihoods
Nicola Lunardon

Mixture modelling with skew-symmetric component distributions
Geoffrey McLachlan

Large scale social and multilayer networks
Matteo Magnani

Uncertainty in statistical matching by BNs
Daniela Marella, Paola Vicard, Vincenzina Vitale

Evaluating the recruiters' gender bias in graduate competencies
Paolo Mariani, Andrea Marletta

Dynamic clustering of network data: a hybrid maximum likelihood approach
Maria Francesca Marino, Silvia Pandolfi

Stability of joint dimension reduction and clustering
Angelos Markos, Michel Van de Velden, Alfonso Iodice D'Enza

Hidden Markov models for clustering functional data
Andrea Martino, Giuseppina Guatteri, Anna Maria Paganoni

Composite likelihood inference for simultaneous clustering and dimensionality reduction of mixed-type longitudinal data
Antonello Maruotti, Monia Ranalli, Roberto Rocci

Bivariate semi-parametric mixed-effects models for classifying the effects of Italian classes on multiple student achievements
Chiara Masci, Francesca Ieva, Tommaso Agasisti, Anna Maria Paganoni

Multivariate change-point analysis for climate time series
Gianluca Mastrantonio, Giovanna Jona Lasinio, Alessio Pollice, Giulia Capotorti, Lorenzo Teodonio, Carlo Blasi

A dynamic stochastic block model for longitudinal networks
Catherine Matias, Tabea Rebafka, Fanny Villers

Unsupervised fuzzy classification for detecting similar functional objects
Fabrizio Maturo, Francesca Fortuna, Tonio Di Battista

New developments in applications of pairwise overlap

Volodymyr Melnykov, Yana Melnykov, Domenico Perrotta, Marco Riani, Francesca Torti, Yang Wang

Functional approaches for satellite data clustering and fusion
Claire Miller, Marian Scott, Craig Wilkie, Ruth O'Donnell, Mengyi Gong, Anna Sehn

Modelling unobserved heterogeneity of ranking data with the bayesian mixture of extended Plackett-Luce models
Cristina Mollica, Luca Tardella

Issues in nonlinear time series modeling of European import volumes
Gianluca Morelli, Francesca Torti

Gaussian parsimonious clustering models with covariates and a noise component
Keefe Murphy, Brendan Murphy

Illumination in depth analysis
Stanislav Nagy, Jiří Dvořák

Copula-based non-metric unfolding on augmented data matrix
Marta Nai Ruscone, Antonio D'Ambrosio

A statistical model for software releases complexity prediction
Marco Ortu, Giuseppe Destefanis, Roberto Tonelli

Comparison of serious diseases mortality in regions of V4
Viera Pacáková, Lucie Kopecká

Price and product design strategies for manufacturers of electric vehicle batteries: inferences from latent class analysis
Friederike Paetz

A Mahalanobis-like distance for cylindrical data
Lucio Palazzo, Giovanni Camillo Porzio, Giuseppe Pandolfo

Archetypes, prototypes and other types
Francesco Palumbo, Giancarlo Ragozini

Generalizing the skew-t model using copulas
Antonio Parisi, Brunero Liseo

Contamination and manipulation of trade data: the two faces of customs fraud
Domenico Perrotta, Andrea Cerasa, Lucio Barabesi, Mario Menegatti, Andrea Cerioli

Bayesian clustering using non-negative matrix factorization
Michael Porter, Ketong Wang

Exploring gender gap in international mobility flows through a network analysis approach
Ilaria Primerano, Marialuisa Restaino

On a class of repulsive mixture models
Jose Quinlan, Garritt Page, Fernando Quintana

Clustering two-mode binary network data with overlapping mixture model and covariates information
Saverio Ranciati, Veronica Vinciotti, Ernst Wit, Giuliano Galimberti

A stochastic blockmodel for network interaction lengths over continuous time
Riccardo Rastelli, Michael Fop

Computationally efficient inference for latent position network models
Riccardo Rastelli, Florian Maire, Nial Friel

Clustering of complex data stream based on barycentric coordinates
Parisa Rastin, Basarab Matei, Guénaël Cabanes

An INDSCAL based mixture model to cluster mixed-type of data
Roberto Rocci, Monia Ranalli

Topological stochastic neighbor embedding
Nicoleta Rogovschi, Nistor Grozavu, Basarab Matei, Younès Bennani, Seiichi Ozawa

Functional data analysis for spatial aggregated point patterns in seismic science
Elvira Romano, Jonatan González Monsalve, Francisco Javier Rodríguez Cortés, Jorge Mateu

Multivariate outlier detection in high reliability standards fields using ICS
Anne Ruiz-Gazen, Aurore Archimbaud, Klaus Nordhausen

ROC curves with binary multivariate data
Lidia Sacchetto, Mauro Gasparini

Silhouette-based method for portfolio selection
Marco Scaglione, Carmela Iorio, Antonio D'Ambrosio

The structure, evolution and interaction of multiplex networks of scientific collaboration at a research university
Valerio Leone Sciabolazza, Raffaele Vacca, Till Krenz

Item weighted Kemeny distance for preference data
Mariangela Sciandra, Simona Buscemi, Antonella Plaia

A fast and efficient modal EM algorithm for gaussian mixtures
Luca Scrucca

Probabilistic archetypal analysis
Sohan Seth

Multilinear tests of association between networks
Daniel Sewell

Use of multi-state models to maximise information in pressure ulcer prevention trials
Linda Sharples, Isabelle Smith, Jane Nixon

Partial least squares for compositional canonical correlation
Violetta Simonacci Massimo Guarino, Michele Gallo

Dynamic modelling of price expectations
Rosaria Simone, Domenico Piccolo, Marcella Corduas

Towards axioms for hierarchical clustering of measures
Philipp Thomann, Ingo Steinwart, Nico Schmid

Influence of outliers on cluster correspondence analysis
Michel Van de Velden, Alfonso Iodice D'Enza, Lisa Schut

Earthquake clustering and centrality measures
Elisa Varini, Antonella Peresan, Jiancang Zhuang

Co-clustering high dimensional temporal sequences summarized by histograms
Rosanna Verde, Antonio Irpino, Antonio Balzanella

An algorithmic approach to item selection in the bayesian Mallows model
Valeria Vitelli, Elja Arjas, Arnoldo Frigessi

Statistical analysis of item pre-knowledge in educational tests: latent variable modelling and optimal statistical decision
Chen Yunxiao, Lu Yan, Irini Moustaki

Evaluation of the web usability of the University of Cagliari portal: an eye tracking study
Gianpaolo Zammarchi, Francesco Mola

Application of survival analysis to critical illness insurance data
David Zapletal, Lucie Kopecka

This book collects the short papers presented at CLADAG 2019, the 12th Scientific Meeting of the Classification and Data Analysis Group (CLADAG) of the Italian Statistical Society (SIS).

The meeting has been organized by the Department of Economics and Law of the University of Cassino and Southern Lazio, under the auspices of the SIS and the International Federation of Classification Societies (IFCS). CLADAG is a member of the IFCS, a federation of national, regional, and linguistically-based classification societies. It is a non-profit, non-political scientific organization, whose aims are to further classification research.

Every two years, CLADAG organizes a scientific meeting, devoted to the presentation of theoretical and applied papers on classification and related methods of data analysis in the broad sense. This includes advanced methodological research in multivariate statistics, mathematical and statistical investigations, survey papers on the state of the art, real case studies, papers on numerical and algorithmic aspects, applications in special fields of interest, and the interface between classification and data science. The conference aims at encouraging the interchange of ideas in the above-mentioned fields of research, as well as the dissemination of new findings.

CLADAG conferences, initiated in 1997 in Pescara (Italy), were soon considered as an attractive information exchange market and became a most important meeting point for people interested in classification and data analysis. One reason was certainly the fact that a selection of the presented papers is regularly published in (post-conference) proceedings, typically by Springer Verlag.

The Scientific Committee of CLADAG2019 conceived the Plenary and Invited Sessions to provide a fresh perspective on the state of the art of knowledge and research in the field. The scientific program of CLADAG 2019 is particularly rich. All in all, it comprises 5 Keynote Lectures, 32 Invited Sessions promoted by the members of the Scientific Program Committee, 16 Contributed Sessions, a Round Table and a Data Competition. We thank all the session organizers for inviting renowned speakers, coming from 28 countries. We are greatly indebted to the referees, for the time spent in a careful review.

The editors would like to express their gratitude to the Rector of the University of Cassino and Southern Lazio and the Director of the Department of Economics and Law for having hosted the meeting. Special thanks are finally due to the members of the Local Organizing Committee and all the people who with their abnegation and enthusiasm have worked for CLADAG 2019.

Special thanks go to Alfiero Klain and Livia Iannucci for the editorial and administrative support.

Last but not least, we thank all the authors and participants, without whom the conference would not have been possible.

Cassino, September 11, 2019

Giovanni C. Porzio

Francesca Greselin

Simona Balzano

COPULA-BASED NON-METRIC UNFOLDING ON AUGMENTED DATA MATRIX

Marta Nai Ruscone[1] and Antonio D'Ambrosio[2]

[1] School of Economics and Management, LIUC Università Cattaneo, (e-mail: mnairuscone@liuc.it)

[2] Department of Economics and Statistics, University of Naples Federico II, (e-mail: antdambr@unina.it)

ABSTRACT: A multidimensional unfolding technique that is not prone to degenerate solutions and is based on multidimensional scaling of a complete data matrix is proposed. We adopt the strategy of augmenting the data matrix, trying to build a complete dissimilarity matrix, by using Copulas-based association measures among rankings (the individuals), and between rankings and objects (namely, a rank-order representation of the objects through tied rankings). The proposed technique leads to acceptable recovery of given preference structures.

KEYWORDS: copulas, unfolding, multidimensional scaling.

1 The copulas function

Copulas are functions that join multivariate distribution functions to their marginal distribution functions (Nelsen, 2013). They describe the dependence structure existing across pairwise marginal random variables. In this way we can consider bivariate distributions with dependency structures different from the linear one that characterizes the multivariate normal distribution.

A bivariate copula $C : I^2 \to I$, with $I^2 = [0,1] \times [0,1]$ and $I = [0,1]$, is the cumulative bivariate distribution function of a random variable (U_1, U_2) with uniform marginal random variables in [0,1]

$$C(u_1, u_2; \theta) = P(U_1 \leq u_1, U_2 \leq u_2; \theta), \quad 0 \leq u_1 \leq 1 \quad 0 \leq u_2 \leq 1 \qquad (1)$$

where θ is a parameter measuring the dependence between U_1 and U_2.

The following theorem by Sklar (Nelsen, 2013) explains the use of the copula in the characterization of a joint distribution. Let (Y_1, Y_2) be a bivariate random variable with marginal cdfs $F_{Y_1}(y_1)$ and $F_{Y_2}(y_2)$ and joint cdf $F_{Y_1,Y_2}(y_1, y_2; \theta)$, then there always exists a copula function $C(\cdot, \cdot; \theta)$ with $C : I^2 \to I$ such that

$$F_{Y_1,Y_2}(y_1, y_2; \theta) = C\big(F_{Y_1}(y_1), F_{Y_2}(y_2); \theta\big), \quad y_1, y_2 \in \mathbb{R}. \qquad (2)$$

Conversely, if $C(\cdot,\cdot;\theta)$ is a copula function and $F_{Y_1}(y_1)$ and $F_{Y_2}(y_2)$ are marginal cdfs, then $F_{Y_1,Y_2}(y_1,y_2;\theta)$ is a joint cdf.
If $F_{Y_1}(y_1)$ and $F_{Y_2}(y_2)$ are continuous functions then the copula $C(\cdot,\cdot;\theta)$ is unique. Moreover, if $F_{Y_1}(y_1)$ and $F_{Y_2}(y_2)$ are continuous the copula can be found by the inverse of (2):

$$C(u_1,u_2) = F_{Y_1,Y_2}(F_{Y_1}^{-1}(u_1),F_{Y_2}^{-1}(u_2)) \tag{3}$$

with $u_1 = F_{Y_1}(y_1)$ and $u_2 = F_{Y_2}(y_2)$. This theorem states that each joint distribution can be expressed in term of two separate but related issues, the marginal distributions and the dependence structures between them. The dependence structure is explained by the copula function $C(\cdot,\cdot;\theta)$. Moreover the (2) provides a general mechanism to construct new multivariate models in a straightforward manner. By changing the copula function we can construct new bivariate distributions with different dependence structures, with the association parameter indicating the strength of the dependence, also different from the linear one that characterizes the multivariate normal distribution.

Each copula is related to the most important measures of dependency: the Pearson correlation coefficient and the Spearman grade correlation coefficient. The Spearman grade correlation coefficient (see Nelsen, 2013 pp. 169-170 for the definition of the grade correlation coefficient for continuous random variables) measure the association between two variables and can be expressed as a function of the copula. More precisely, if two random variables are continuous and have copula C with parameter θ, then the Spearman grade correlation is

$$\rho_s(C) = 12\int_{I^2} C_\theta(u_1,u_2)du_1du_2 - 3. \tag{4}$$

For continuous random variables it is invariant with respect to the two marginal distributions, i.e. it can be expressed as a function of its copula. This property is also known as 'scale invariance'. Note that not all measures of association satisfy this property, e.g. Pearson's linear correlation coefficient (Embrechts *et al.* , 2002).

In the following, we focus on observations Y_{ik} of the latent continuous random variable Y_{ik}^*, describing the preference of the consumer i ($i \in N = \{1,...,n\}$) for the object k. Let $y_i = (y_{i1},...,y_{ik})$ be the vector of ranks of consumer i for the k objects, where y_{ik} is the rank of object k for the subject $i-th$. Be $U = F(Y_{ik}^*)$ and $V = F(Y_{jk}^*)$ the marginal cumulative distributions (cds). We assume that (Y_{ik},Y_{jk}) correspond to the bivariate discrete random variable obtained by a discretization of the continuous latent variable $(U = F(Y_{ik}^*), V = F(Y_{jk}^*))$ with support $[0,1]\times[0,1]$ and cdf given by $C_\theta(\cdot,\cdot)$.

Let $A_{r,s} = [u_{r-1}, u_r] \times [v_{s-1}, v_s]$ $r,s = 1,...,k$ be rectangles defining the discretization. Let $p_{k,k}$ be the joint probabilities corresponding to the rectangle $A_{r,s}$ for $r,s = 1,...,k$ with value $1/k$ if the pair (y_{ik}, y_{jk}) is observed and 0 otherwise. Let $V_{C_\theta}(A_{11}), \ldots, V_{C_\theta}(A_{kk})$ be the volumes of the rectangles under the copula C_θ, then there exists a unique element in the family of copula for which the following relationship holds true:

$$(V_{C_\theta}(A_{11}), \ldots, V_{C_\theta}(A_{kk})) = (p_{11}, \ldots, p_{kk}). \tag{5}$$

Given the ranking of two subjects i and j, a $k \times k$ contingency table K_{rs} $(r,s = 1,...,k)$ is defined. A cell in this table takes value $1/k$ if (y_{ik}, y_{jk}) is observed and 0 otherwise. This contingency table provides the basis for our estimation procedure.

Fixed the copula C_θ and defined the Spearman grade correlation coefficients $\rho_s(C_\theta)$ (Nelsen, 2013) to each pair (Y_{ik}, Y_{jk}), $i \neq j$ with $i, j \in N$, we define the dissimilarity coefficient d_{ij}:

$$d_{ij} = \sqrt{1 - \frac{\rho_s + 1}{2}} \tag{6}$$

where ρ_s performs well in measuring the agreement between two rankings Y_{ik} Y_{jk}.

Notice that other ways of findings a correlation-type distance matrix have been provided in the literature (Kaufman & Rousseeuw, 2009). For instance, one may consider $d_{ij} = 1 - \rho$ or $d_{ij} = 1 - |\rho|$.

The parameter θ can be estimated via maximum likelihood. Estimating the value of the copula dependence parameter θ we obtain the grade of association between two rankings.

2 Unfolding as a special case of multidimensional scaling on Copulas based association between rankings

Unfolding applies multidimensional scaling (Cox & Cox, 2000) to an off-diagonal $n \times m$ matrix, usually representing the scores (or the rank) assigned to a set of m items by n individuals or judges (Borg & Groenen, 1997). The goal is to obtain two configuration of points representing the position of the judges (X) and the items (Y) in a reduced geometrical space. Each point representing the individuals is considered as an ideal point so that its distances to the object points correspond to the preference scores (Coombs, 1964). Unfolding can be seen as a special case of multidimensional scaling because the off-diagonal

matrix is considered as a block of an ideal distance matrix in which both the within judges and the within items dissimilarities are missing. The presence of blocks of missing data causes the phenomenon of the so-called degenerate solutions, i.e., solutions that return excellent badness of fit measures but not graphically interpretable at all. To tackle the problem of degenerate solutions, several proposals have been presented in the literature (Borg & Groenen, 1997). By following the approach introduced by Van Deun *et al.* , 2007, we adopt the strategy of augmenting the data matrix, trying to build a complete dissimilarity matrix, and then applying any MDS algorithms. In order to augment the data matrix, we use Copulas-based association measures among rankings (the individuals), and between rankings and objects (namely, a rank-order representation of the objects through tied rankings). Both experimental evaluations and applications to well-known real data sets show that the proposed strategy produces non-degenerate non-metric unfolding solutions.

References

BORG, I., & GROENEN, P.J. 1997. *Modern multidimensional scaling*. Springer Series in Statistics. Springer-Verlag, New York. Theory and applications.

CHERUBINI, U., LUCIANO, E., & VECCHIATO, W. 2004. *Copula methods in finance*. John Wiley & Sons.

COOMBS, C.H. 1964. *A theory of data.* Wiley.

COX, T.F., & COX, M.A.A. 2000. *Multidimensional scaling*. Chapman and hall/CRC.

EMBRECHTS, P., MCNEIL, A., & STRAUMANN, D. 2002. Correlation and dependence in risk management: properties and pitfalls. *Risk management: value at risk and beyond*, **1**, 176–223.

JOE, H. 1997. *Multivariate models and multivariate dependence concepts*. CRC Press.

KAUFMAN, L., & ROUSSEEUW, P. J. 2009. *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons.

NELSEN, R.B. 2013. *An introduction to copulas*. Springer Science & Business Media.

VAN DEUN, K., HEISER, W.J., & DELBEKE, L. 2007. Multidimensional unfolding by nonmetric multidimensional scaling of Spearman distances in the extended permutation polytope. *Multivariate Behavioral Research*, **42**(1), 103–132.

CLADAG 2019
Cassino (ITALY)
11-13 September, 2019

Società Italiana di Statistica

UNICAS 1979>2019
40 YEARS FROM FOUNDATION
UNIVERSITÀ DI CASSINO E DEL LAZIO MERIDIONALE

The CLAssification and Data Analysis Group of the Italian Statistical Society (SIS) promotes advanced methodological research in multivariate statistics with a special vocation in Data Analysis and Classification.

CLADAG supports the interchange of ideas in these fields of research, including the dissemination of concepts, numerical methods, algorithms, computational and applied results.

CLADAG is a member of the International Federation of Classification Societies (IFCS).

Among its activities, CLADAG organizes a biennial international scientific meeting, schools related to classification and data analysis, publishes a newsletter, and cooperates with other member societies of the IFCS to the organization of their conferences.

Founded in 1985, the IFCS is a federation of national, regional, and linguistically-based classification societies. It is a non-profit, nonpolitical scientific organization, whose aims are to further classification research.