SIS 2014

47th Scientific Meeting
of the Italian Statistical Society

Cagliari - June 11/13, 2014

PROCEEDINGS

Editors:
S. Cabras, T. Di Battista and W. Racugno

CUEC EDITRICE

Società Italiana di Statistica
SIS
Fondata nel 1939

SIS 2014

47th Scientific Meeting
of the Italian Statistical Society

Cagliari - June 11/13, 2014

PROCEEDINGS

Editors:
S. Cabras, T. Di Battista and W. Racugno

CUEC EDITRICE

9 788884 678744

Program Committee

The Program Committee of the 47th SIS Scientific Meeting is composed by:

Tonio Di Battista *(Chair)*	Università di Pescara
Marco Alfò	Sapienza Università di Roma
Giulio Barcaroli	ISTAT
Gianni Betti	Università di Siena
Daniela Cocchi	Università di Bologna
Stefano Gattone	Università di Roma "Tor Vergata"
Andrea Giommi	Università di Firenze
Salvatore Ingrassia	Università di Catania
Michele La Rocca	Università di Salerno
Letizia Mencarini	Università di Torino
Sonia Petrone	Università Commerciale Bocconi
Alessandra Petrucci	Università di Firenze
Donato Posa	Università di Lecce
Walter Racugno	Università di Cagliari
Giovanna Ranalli	Università di Perugia
Marco Riani	Università di Parma
Roberto Rocci	Università di Roma "Tor Vergata"
Arjuna Tuzzi	Università di Padova

Preface

The edition of this volume gave us the opportunity to perceive that, together with many well-known Italian statisticians belonging to the national and international community, many young researchers are emerging. They presented, at the 47th Scientific Meeting of the Italian Statistical Society, their remarkable contributions, both from the methodological and the applicative point of views.

Although some papers may not look, in their actual form, fully mature from the scientific and communicative point of views, we decided - in agreement with the referees - to publish them since promising and full of ideas. In this respect, the contributions published in this volume provide a comprehensive overview of the current Italian scientific researches in theoretical and applied statistics.

This volume also contains several contributions presented by foreign researchers, highlighting the fact that the Italian Statistical Society has an attractive role in the international scientific community.

Finally, we would like to emphasize that, even from the abstracts of the contributions, the wideness of the collaborations between the statisticians and the experts from other fields emerges. This denotes that, also in Italy, statistical methods are spreading in the different fields of the scientific researches.

Stefano Cabras
Tonio Di Battista
Walter Racugno

Table of Contents

List of papers

Abstracts

You can access to the full paper by clicking on the link.

List of papers

PLENARY SESSIONS

President's Invited Speaker

- Antonio Golini. **Demographic crisis and economic crisis for Italian Mezzogiorno: an iceberg detached from the continent ?**

Plenary Session A

- Adrian Smith FRS. **The Bayesian 21st century. An appreciation of the contributions of Dennis Lindley: 1923-2013**

Plenary Session B

- Chiara Sabatti. **In the mist of the data deluge, how to let the interesting findings surface? Tales from genetics.**

Plenary Session C

- Andrea Cerioli. **How to marry robustness and applied statistics** (Full paper: 3104.pdf)

SPECIALIZED SESSIONS

SP1 - Recent Advances In Biostatistics

- Anne Presanis, Paul Birrell, Daniela De Angelis. **Statistical challenges in epidemic modelling** (Full paper: 3056.pdf)
- Chiara Romualdi. **Recent advances in genomic studies** (Full paper: 3055.pdf)
- Federico Rotolo, Stefan Michelis. **Global interaction tests for predictive gene signatures in randomized clinical trials.** (Full paper: 3054.pdf)

SP2 - Clustering Real Time Data Streams

- Antonio Balzanella, Rosanna Verde, Antonio Irpino. **Monitoring spatially dependent data streams** (Full paper: 3085.pdf)
- Angela De Sanctis, Francesca Fortuna. **Functional classification on convex function spaces** (Full paper: 3141.pdf)
- Donato Malerba, Annalisa Appice. **Discovering trend clusters in sensor data streams** (Full paper: 3140.pdf)

SP3 - Bayesian Nonparametrics: Methods And Applications

- Raffaele Argiento, Ilaria Bianchini, Alessandra Guglielmi. **A Bayesian nonparametric model for density and cluster estimation: the ε-NGG process mixture** (Full paper: 3032.pdf)
- Antonio Lijoi, Bernardo Nipoti, Igor Prunster. **A Bayesian nonparametric model for combining data from different experiments** (Full paper: 3033.pdf)
- Judith Rousseau. **On consistency issues in Bayesian nonparametric testing - a review** (Full paper: 3034.pdf)

SP4 - Recent Advances In Time Series Analysis

- Lara Fontanella, Luigi Ippoliti, Pasquale Valentini. **Regressions in Spatially Dynamic Factor Models** (Full paper: 3082.pdf)

- Liang Peng, Yadong Li, Jinguo Gong, Qiwei Yao. **Estimation of Extreme Quantiles for Functions of Dependent Random Variables** (Full paper: 3080.doc)

- Tommaso Proietti, Alessandra Luati. **Generalised Linear Cepstral Models for the Spectrum of a Time Series** (Full paper: 3081.pdf)

SP5 - New Challenges In Survey Sampling

- Fulvia Mecatti. **Multiple Frame Surveys: a simplified and unified review** (Full paper: 2990.pdf)

- Jean Opsomer, Wade Herndon, F. Jay Breidt. **Testing for Informativeness in Analytic Inference from Complex Surveys** (Full paper: 2998.pdf)

- Vijay Verma. **On sampling elusive populations** (Full paper: 2989.pdf)

SP6 - Directional Data

- Alan Gelfand, Fangpo Wang. **Analyzing spatial and spatio-temporal angular and linear data using Gaussian processes** (Full paper: 2991.pdf)

- Giovanna Jona Lasinio, Gianluca Mastrantonio, Alan Gelfand. **Models for space-time directional data using Wrapped Gaussian processes** (Full paper: 2995.pdf)

- Francesco Lagona. **Regression analysis of correlated circular data** (Full paper: 2870.pdf)

SP7 - Scoring Rules And Pseudo

- A.p. Dawid. **Proper Scoring Rules in Statistical Inference** (Full paper: 3068.pdf)

- Monica Musio, Valentina Mameli, Alexander Dawid. **Comparison of approaches to inference in stationary AR(1) models** (Full paper: 3035.pdf)

- Erlis Ruli, Nicola Sartori, Laura Ventura. **Approximate Bayesian Computation with proper scoring rules** (Full paper: 2973.pdf)

SP8 - Quantile And M

- Giulia Cavrini, Sara Piombo, Alessandra Samoggia. **A Multilevel Quantile Regression Analysis of Chronic Diseases Affects on Physical and Mental Well-being** (Full paper: 3067.docx)

- Marco Geraci, Mc Jones. **Prediction of conditional quantiles on the half line and the unit interval using transformation models** (Full paper: 2825.pdf)

- Maria Francesca Marino, Nikos Tzavidis. **Mixed hidden Markov models for quantiles** (Full paper: 3070.pdf)

SP9 - Methodological Issues For Constructing Composite Indicators

- Filomena Maggino, Marco Fattore. **Evaluating subjective suffering in Italy with partially ordered sets** (Full paper: 3016.pdf)

- Matteo Mazziotta, Adriano Pareto. **A Non-compensatory Composite Index for Comparisons over Time** (Full paper: 3011.pdf)

- Francesco Vidoli, Elisa Fusco, Claudio Mazziotta. **Non-compensability in composite indicators: a robust directional frontier method** (Full paper: 3010.pdf)

SP10 - Parametric And Nonparametric Mixed Effect Models

- Silvia Cagnone, Francesco Bartolucci. **Continuous versus discrete latent structures in dynamic latent variable models** (Full paper: 3065.pdf)

- Paul H.c. Eilers. **Generalized linear mixed models for over-dispersed counts** (Full paper: 3064.pdf)

- Sara Viviani. **Smooth random effect distribution in joint models with an application to cardiomiopathy data.** (Full paper: 3053.pdf)

SP11 - Demography And Environmental Emergency

- Elena Ambrosetti, Enza Petrillo. **Environmental Change, Migration and Displacement. Insights and developments from L'Aquila** (Full paper: 3039.docx)

- Wolfgang Lutz. **Population – Environment Interactions** (Full paper: 3041.pdf)

- Valerio Manno, Giada Minelli, Susanna Conti. **The use of current data to evaluate the health impact of environmental pollution: the "SENTIERI approach" and the case study of Taranto** (Full paper: 3040.docx)

SP12 - Statistical Analysis And Big Data

- Barteld Braaksma. **Big Data and official statistics: local experiences and international initiatives** (Full paper: 3126.pdf)

- Salvatore Rinzivillo, Fosca Giannotti. **Understanding Human Mobility with Big Data** (Full paper: 3027.pdf)

- Simone Vantini, Piercesare Secchi, Paolo Zanini. **The Virtuous Cycle of Big Data and Big Cities: a Case Study from Milan** (Full paper: 3000.pdf)

SOLICITED SESSIONS

SL1 - Bayesian Models For Complex Problems

- Raffaele Argiento, Alessandra Guglielmi. **Bayesian principal curve clustering by species-sampling mixture models** (Full paper: 2877.pdf)

- Laura Azzimonti, Laura Sangalli, Piercesare Secchi. **Modeling prior knowledge on complex phenomena behaviors via partial differential equations** (Full paper: 2923.pdf)

- Michele Guindani, Linlin Zhang, Francesco Versace, Marina Vannucci. **A Bayesian Variable Selection Model for the Clustering of Time Courses in FMRI data** (Full paper: 2879.pdf)

- Andrea Tancredi, Brunero Liseo. **Bayesian Inference with Linked Data** (Full paper: 2878.pdf)

SL2 - Geostatistics And Environmental Applications

- Claudia Cappello, Sandra De Iaco, Donato Posa. **Computing non-separability for space-time covariance functions: a case study on PM10 data** (Full paper: 3138.pdf)

- Sandra De Iaco, Sabrina Maggio, Monica Palma, Donato Posa. **Modeling environmental quality: a case study** (Full paper: 3137.pdf)

- J. Jaime Gòmez-Hernàndez, Teng Teng Xu. **When it is not normal to be Gaussian** (Full paper: 3135.pdf)

- Ana Russo, Amílcar O. Soares. **Hybrid model for urban air pollution forecasting: A stochastic spatio-temporal approach** (Full paper: 3139.pdf)

SL3 - Robust Methods For The Analysis Of Complex Data

- Anthony Atkinson, Aldo Corbellini. **Introducing Prior Information into the Forward Search for Regression** (Full paper: 2829.pdf)

- Andrea Cerasa, Francesca Torti, Domenico Perrotta. **Analysis of complex data in official statistics** (Full paper: 3047.pdf)

- Pietro Coretto, Christian Hennig. **Robust covariance matrix estimation with regularization** (Full paper: 3046.pdf)

- Luis Garcıa-Escudero, Alfonso Gordaliza, Carlos Matran, Agustın Mayo-Iscar. **Robust model-based clustering and mixture** (Full paper: 3058.pdf)

SL4 - Statistics For Environmental Phenomena And Their Interactions

- Paola Berchialla, Marta Blangiardo, Michela Cameletti, Francesco Finazzi, Maria Villoria, Rosaria Ignaccolo. **Spatial modeling for air pollution epidemiology: hospital admission risk for cardio-respiratory diseases in Torino province** (Full paper: 3051.pdf)

- Rosaria Ignaccolo, Maria Franco Villoria. **Kriging for functional data: uncertainty assessment** (Full paper: 2953.pdf)

- Luigi Ippoliti, Lara Fontanella, Pasquale Valentini, Annalina Sarra, Sergio Palermi. **Bayesian structural equation modeling for factors influencing residential radon levels** (Full paper: 3023.pdf)

- Elena Scardovi, Francesca Bruno, Fedele Greco. **Assessment of bayesian models for rainfall field reconstruction** (Full paper: 3013.pdf)

SL5 - Mixture And Latent Variable Models For Causal Inference And Analysis Of Socio

- Silvia Bacci, Francesco Bartolucci. **A multidimensional finite mixture SEM for non-ignorable missing responses to test items** (Full paper: 2831.pdf)

- Paolo Giordani. **Finite mixtures for multivariate mixed data: a Parafac-based approach** (Full paper: 3072.pdf)

- Lucia Modugno, Silvia Cagnone, Simone Giannerini. **A multilevel model for repeated cross-sectional data with stochastic volatility** (Full paper: 3125.pdf)

- Leonardo Grilli, Fulvia Pennoni, Carla Rampichini, Isabella Romeo. **Multivariate multilevel modelling of student achievement data** (Full paper: 2959.pdf)

SL6 - Equity And Sustainability: Theory And Relationships

- Simone Bastianoni, Luca Coscieme, Federico Pulselli. **Categorization of National Economies through Environmental, Social and Economic Indicators.** (Full paper: 2981.pdf)

- Alessandra Ferrara, Angela Ferruzza, Nicoletta Pannuzi. **Environmental resources, landscape and cultural heritage, economic conditions: fundamental components to measure well-being** (Full paper: 3006.doc)

- Tommaso Luzzati, Bruno Cheli. **Measuring the sustainability performances of the Italian regions** (Full paper: 2987.pdf)

- Andrea Regoli, Antonella D'Agostino, Francesca Gagliardi, Laura Neri. **A framework for monitoring country sustainability** (Full paper: 2980.pdf)

SL7 - Advances In Bayesian Statistics

- Stefano Cabras, María Eugenia Castellanos Nueda, Erlis Ruli, Mario Pirastu, Maria Pina Concas. **An ABC/quasi-likelihood approach for linkage/GWAS study of a Sardinian genetic isolate** (Full paper: 3030.pdf)

- Pierpaolo De Blasi, Stephen Walker. **Bayesian nonparametric estimation and asymptotics with misspecified density models** (Full paper: 3087.pdf)

- Edward George, Veronika Ročková. **Determinantal Priors for Variable Selection** (Full paper: 3086.pdf)

- Brunero Liseo. **Approximate Bayesian Inference for Copula Models** (Full paper: 3084.pdf)

SL8 - Statistical Models For The Analysis Of Energy Markets

- Claudia Furlan, Mariangela Guidolin, Renato Guseo. **Are there any effects of Fukushima accident on the diffusion of nuclear energy?** (Full paper: 2826.pdf)

- Luigi Grossi, Fany Nan. **Robust forecasting of electricity prices with nonlinear models and exogenous regressors** (Full paper: 2920.pdf)

- Matteo Pelagatti, Lucia Parisio. **First results on the Italian-Slovenian electricity market coupling** (Full paper: 2958.pdf)

- Alessandro Sapio. **Renewable flows and congested lines in the Italian power grid: Binary time series and vector autoregressions** (Full paper: 2970.pdf)

SL9 - Recent Developments In Sampling Theory

- Pier Luigi Conti. **Weak convergence and empirical processes in survey sampling** (Full paper: 3066.pdf)

- Pier Francesco Perri, Lucio Barabesi, Giancarlo Diana. **On the linearization of inequality indexes in the design-based framework** (Full paper: 2992.pdf)

- Caterina Pisani, Lorenzo Fattorini. **Sampling strategies for diversity indexes estimation in presence of rare species** (Full paper: 3020.pdf)

- Emilia Rocco. **Spatially balanced adaptive web sampling** (Full paper: 3022.pdf)

SL10 - Functional Data Analysis

- Aldo Goia. **Some advances on semi-parametric functional data modelling** (Full paper: 3007.pdf)

- Elvira Romano, Jorge Mateu, Iulian Teodor Vlad. **A Functional Model for Detecting Changes in Evolving Shapes Brain Tumors** (Full paper: 3045.pdf)

- Laura Sangalli. **Functional data analysis in spaces of surfaces** (Full paper: 2986.pdf)

- Simone Vantini, Alessia Pini. **Hypothesis Testing in Functional Data Analysis: a Non-parametric Approach** (Full paper: 3001.pdf)

SL11 - Extremes And Dependent Sequences

- Paola Bortot, Carlo Gaetan. **Latent process models for temporal extremes with an application to rainfall data** (Full paper: 3142.pdf)

- Fabrizio Laurini. **The clustering of extreme values for some asymmetric GARCH-type models** (Full paper: 3143.pdf)

- Giulia Marcon, Simone A. Padoan, Philippe Naveau, Pietro Muliere. **Inference of multivariate dependence structures in extreme value theory** (Full paper: 2996.pdf)

SL12 - Issues In Ecological Statistics

- Linda Altieri, Marian Scott, Janine Illian. **A changepoint analysis on spatio-temporal point processes** (Full paper: 3009.pdf)

- Antonella Bodini, Gianni Gilioli. **Linking metapopulation modelling and Information Theory for area-wide pest management** (Full paper: 2983.pdf)

- Ilaria Rosati, Alessio Pollice, Serena Arima, Giovanna Jona Lasinio, Alberto Basset. **Uncertainty of methodologies assessing ecological status in transitional water systems** (Full paper: 3074.pdf)

- Luca Tardella, Danilo Alunni Fegatelli. **Modelling and inferential issues for behavioral patterns in capture-recapture data** (Full paper: 3059.pdf)

SL13 - Computations With Intractable Likelihood

- Isadora Antoniano-Villalobos, Stephen Walker. **Exact Bayesian inference for discretely observed diffusions** (Full paper: 3005.pdf)

- Luigi Augugliaro, Angelo Mineo. **An efficient algorithm to estimate the sparse group structure of an high-dimensional generalized linear model** (Full paper: 3002.pdf)

- Francesca Greselin, Luis Garcìa-Escudero, Alfonso Gordaliza, Salvatore Ingrassia, Agustìn Mayo-Iscar. **An adaptive method to robustify ML estimation in Cluster Weighted Modeling** (Full paper: 3004.pdf)

- Antonietta Mira, Alberto Caimo. **Delayed rejection algorithm to estimate Bayesian social networks** (Full paper: 3003.pdf)

SL14 - Geographical Information In Sampling And Estimation

- Roberto Benedetti, Federica Piersimoni, Paolo Postiglione. **Spatially balanced samples for land use/land cover surveys** (Full paper: 2997.pdf)

- Chiara Bocci, Emilia Rocco, Patrizia Lattarulo. **Modeling the location decisions of firms** (Full paper: 3136.pdf)

- Elisabetta Carfagna, Simone Maffei, Andrea Carfagna. **Geo-referenced information for agricultural statistics** (Full paper: 3069.pdf)

- Stefano Marchetti, Caterina Giusti, Nicola Salvati. **The use of geographic information under the area-level approach to small area estimation** (Full paper: 3024.pdf)

SL15 - Clinical Designs

- Andrea Ghiglietti, Anna Maria Paganoni. **Statistical properties of urn designs in clinical trials** (Full paper: 2852.pdf)

- Stefania Gubbiotti, Pierpaolo Brutti, Fulvio De Santis. **A predictive look at Bayesian Bandits** (Full paper: 2926.pdf)

- Paola Rebora, Maria Grazia Valsecchi. **Optimal two-phase design and incidence estimation in cohort studies** (Full paper: 2848.pdf)

SL16 - Bayesian Inference For High

- Veronika Ročková, Edward I. George. **Fast EM Inference for Bayesian Factor Analysis with Indian Buffet Process Prior** (Full paper: 3048.pdf)

- Mahlet Tadesse, Alberto Cassese, Michele Guindani, Francesco Falciani, Marina Vannucci. **A Unified Method for CNV Detection and Association with Gene Expression** (Full paper: 3049.pdf)

- Lorenzo Trippa. **Bayesian Adaptive Trials** (Full paper: 3057.pdf)

SL17 - Use Of Big Data For The Production Of Statistical Information

- Stefano Falorsi, Fabio Bacchini, Michele D'Alò, Andrea Fasulo, Carmine Pappalardo. **Does Google index improve the forecast of Italian labour market?** (Full paper: 3019.pdf)

- Dino Pedreschi, Roberta Vivio, Fosca Giannotti, Mirco Nanni, Barbara Furletti, Giuseppe Garofalo, Lorenzo Gabrielli, Letizia Milli,. **Use of mobile phone data to estimate mobility flows. Measuring urban population and inter-city mobility using big data in an integrated approach** (Full paper: 3026.pdf)

- Dino Pedreschi. **Big Data, Social Mining, Diversity, and Wellbeing** (Full paper: 3107.pdf)

- Marco Scarnò, Sergio Salamone, Monica Scannapieco. **Web scraping and web mining: new tools for Official Statistics** (Full paper: 3008.pdf)

SL18 - Measuring The Smart City

- Roberta De Santis, Alessandra Fasano, Nadia Mignolli, Anna Villa. **Smart City: measuring a multidimensional topic** (Full paper: 3029.doc)

- Carlo Giovannella. **Smart Territory Analytics: toward a shared vision** (Full paper: 3031.pdf)

- Mariagrazia Dotoli, Vincenzo Patruno, Luigi Ranieri, Aldo Scarnera. **Staying On The Smart side. Measuring The Smart Communities** (Full paper: 3021.doc)

- Paolo Testa, Nicolò Marchesini. **Measuring the territory to build up the Smart City** (Full paper: 3037.doc)

SL19 - Forecasting Economic And Financial Time Series

- Roberto Baragona, Domenico Cucina. **Outliers in Time Series: an Empirical Likelihood Approach** (Full paper: 3093.pdf)

- Luc Bauwens, Manuela Braione, Giuseppe Storti. **Long term component dynamic models for realized covariance matrices** (Full paper: 3092.pdf)

- Roberto Casarin, Tilmann Gneiting, Francesco Ravazzolo. **Probabilistic Calibration of Predictive Distributions** (Full paper: 3134.pdf)

- Simone Giannerini, Esfandiar Maasoumi, Estela Bee Dagum. **Testing for nonlinear serial dependence in time series with surrogate data and entropy measures** (Full paper: 3123.pdf)

CONTRIBUTED PAPER SESSIONS

CP1 - Demography

- Gustavo De Santis, Mauro Maltagliati, Silvana Salvini. **A measure of the distance between countries based on individual data** (Full paper: 2880.doc)

- Antonino Di Pino, Maria Gabriella Campolo, Marcantonio Caltabiano. **Retirement and Intra-Household Labour Division of Italian Couples: A Simultaneous Equation Approach** (Full paper: 2840.pdf)

- Lucia Leporatti, Paolo Cremonesi, Enrico Di Bella, Luca Persico. **Demographic change and future sustainability of emergency departments: a pilot study for Liguria** (Full paper: 2932.pdf)

- Luigi Marcone, Francesco Borrelli, Stefania Di Domenico. **Error models for weighting estimators in the 15th Italian Population and Household Census** (Full paper: 2917.docx)

- Silvia Meggiolaro, Fausta Ongaro. **Father-child contact after marital dissolution. Evidence from Italy** (Full paper: 2847.pdf)

CP2 - Statistics In Finance

- Alessandra Amendola, Vincenzo Candila. **The use of loss functions in assessing the VaR measures** (Full paper: 2909.pdf)

- Paola Cerchiello, Paolo Giudici. **Systemic risk models** (Full paper: 2977.pdf)

- Giancarlo Ferrara, Francesco Vidoli, Arianna Campagna, Jacopo Canello. **Labour-use efficiency in the Italian machinery industry: a non-parametric stochastic frontier perspective** (Full paper: 2898.doc)

- Andrea Pierini. **Chain Graph for VAR and MARCH parameters reduction. EU index returns case.** (Full paper: 2882.pdf)

- Marianna Succurro, G. Damiana Costanzo. **Predicting financial bankruptcy by a (Robust) Principal Component Analysis based model: an empirical investigation.** (Full paper: 2935.docx)

CP3 - Statistics In Medicine

- Antonio Canale, Dipankar Bandyopadhyay. **Bayesian nonparametric spatial modelling of ordinal periodontal data** (Full paper: 2845.pdf)

- Silvia Liverani, Chris Cantwell. **Elicitation and visualisation of uncertainty in electrograms for activation time maps** (Full paper: 2875.pdf)

- Monia Lupparelli, Alberto Roverato. **Log-mean linear regression models for assessing the effect of HIV-infection on multimorbidity in a case-control study** (Full paper: 2951.pdf)

- Lorenzo Maragoni, Monica Chiogna. **A Quantile-based Test for Detecting Differential Expression in Microarray Data** (Full paper: 2964.pdf)

- Paolo Zanini, Piercesare Secchi, Simone Vantini. **EEG signals decomposition: a multi-resolution analysis** (Full paper: 2843.pdf)

CP4 - Clustering Methods: Theory And Applications

- Marzia Cremona, Pier Giuseppe Pelicci, Laura Riva, Laura Sangalli, Piercesare Secchi, Simone Vantini. **Cluster analysis on shape indices for ChIP-Seq data** (Full paper: 2904.pdf)

- Paolo Giordani, Maria Brigida Ferraro. **fclust: an R package for fuzzy clustering** (Full paper: 2832.pdf)

- Giovanna Menardi, Domenico De Stefano. **Modal Clustering of Social Networks** (Full paper: 2940.pdf)

- Andrea Pastore, Stefano Tonellato. **Identification of multiple clusterings using Gaussian mixtures** (Full paper: 2914.pdf)

CP5 - Functional Data Analysis

- Enea Bongiorno, Aldo Goia. **A clustering method for functional data** (Full paper: 2963.pdf)

- Tommaso Lando, Lucio Bertoli-Barsotti. **A class of bibliometric indices based on a sum of increasing and concave functions** (Full paper: 2902.docx)

- Alessia Pini, Konrad Abramowicz, Charlotte Häger, Kim Hébert-Losier, Lina Schelin, Johan Strandberg, Simone Vantini. **Anterior Cruciate Ligament Rupture: Functional Data Analysis of Knee Motion** (Full paper: 2842.pdf)

- Matteo Ruggiero, Omiros Papaspiliopoulos, Dario Spanò. **Bayesian inference for dynamically evolving distributions** (Full paper: 2820.pdf)

- Nicholas Tarabelloni, Rachele Biasi, Francesca Ieva, Anna Paganoni. **Depth measures for multivariate functional data with data-driven weights** (Full paper: 2834.pdf)

CP6 - Forensic Statistics

- Carlo Cusatelli, Massimiliano Giacalone. **Statistical comparison of European judicial systems according to ICT** (Full paper: 2968.doc)

- Antonia Rosa Gurrieri, Marilene Lorizio. **Efficiency of justice and justice's demand** (Full paper: 2954.doc)

- Laura Antonucci, Francesco D'Ovidio, Domenico Viola. **Measures of Courts Performances and Stochastic Frontier Models** (Full paper: 2974.doc)

- Annamaria Stramaglia, Marilene Lorizio. **The efficiency of justice in an economic perspective: the role of the supply side** (Full paper: 2899.doc)

CP7 - Economic Phenomena

- Alberto Arcagni, Michele Zenga. **Decomposition by sources of the ξ inequality index** (Full paper: 2853.pdf)

- Stefano Falorsi, Michele D'Alò, Silvia Loriga. **LFS quarterly small area estimation of youth unemployment at provincial level** (Full paper: 2945.pdf)

- Francesco Porro, Michele Mario Zenga. **The decomposition by subgroups of the inequality curve Z(p) and the inequality index ξ** (Full paper: 2972.pdf)

- Emanuela Raffinetti, Elena Siletti, Achille Vernizzi. **Inequality measures and the issue of negative incomes** (Full paper: 2894.pdf)

- Daniele Toninelli, Martin Beaulieu, Catalin Dochitoiu. **The Effects of a new Aggregation Structure on Consumer Price Index Estimates** (Full paper: 2892.pdf)

CP8 - Advances In Statistical Modelling

- Alessandra Brazzale, Valentina Mameli. **Small-sample likelihood asymptotics for the equi-correlated bivariate normal model** (Full paper: 2867.pdf)

- Carlo Orsi, Andrea Ongaro. **On Non-central Beta distributions** (Full paper: 2933.pdf)

- Marco Perone Pacifico. **Nonparametric Mode Hunting** (Full paper: 2871.pdf)

- Mario Romanazzi. **Discriminant analysis of von Mises - Fisher distributions** (Full paper: 2895.pdf)

CP9 - Developments In Bayesian Inference

- Pierpaolo Brutti. **Bayesian Inference for the Intrinsic Dimension** (Full paper: 2922.pdf)

- Maria Eugenia Castellanos, Stefano Cabras. **Zellner-Siow Priors for variable selection with censored data** (Full paper: 2886.pdf)

- Francesca Condino, Filippo Domma, Giovanni Latorre. **Likelihood and Bayesian estimation of $P(Y < X)$ using lower record values from a general class of distribution** (Full paper: 2887.pdf)

- Elias Moreno, Francisco Vázquez-Polo, Miguel Negrín. **An objective Bayesian procedure for meta–analysis of binomial data** (Full paper: 2862.pdf)

- Bernardo Nipoti, Stefano Favaro. **Uncertainty quantification for Bayesian nonparametric estimators of rare species variety** (Full paper: 2957.pdf)

CP10 - Educational Statistics

- Matilde Bini, Lucio Masserini. **A finite mixture model approach on the first year university drop-out probability** (Full paper: 2900.doc)

- Stefano Cabras, Juan De Dios Tena Horrillo. **A Bayesian nonparametric modelling to estimate students' response to ICT investment** (Full paper: 2839.pdf)

- Dalit Contini, Elisa Grand. **On the development of school achievement inequalities with cross-sectional data** (Full paper: 2967.docx)

- Franca Crippa, Marcella Mazzoleni, Mariangela Zenga. **The role of the membership function to model university students' flow** (Full paper: 2955.pdf)

- Francesca Ieva, Tommaso Agasisti, Anna Paganoni. **Multilevel modeling of heterogeneity in math achievements: different class- and school-effects across Italian regions** (Full paper: 2833.pdf)

CP11 - Sanitary Statistics And Epidemiology

- Domenico De Stefano, Stefano Campostrini. **The 2008 Great Recession and Health in Italy. A Study on the Surveillance Data System PASSI** (Full paper: 2962.pdf)

- Marco Geraci, Alessio Farcomeni. **A probabilistic approach to the estimation of principal components with nonignorable missing data: Applications in accelerometer-based physical activity studies** (Full paper: 2881.pdf)

- Stefano Mazzuco, Bruno Scarpa, Lucia Zanotto. **A mortality model based on a mixture distribution function** (Full paper: 2897.pdf)

- Inad Nawajah, Raffaele Argiento, Alessandra Guglielmi, Ettore Lanzarone. **Joint Prediction of Demand and Care Duration in Home Care Patients: a Bayesian Approach** (Full paper: 2893.pdf)

- Emiliano Sironi, Massimo Cannas. **Hospital Differences in Caesarean Deliveries in Sardinia: A Multilevel Analysis** (Full paper: 2948.doc)

CP12 - Survey Methodology

- Alessio Farcomeni. **Heterogeneity for a general class of recapture models based on equality constraints on the conditional capture probabilities** (Full paper: 2822.pdf)

- Pier Luigi Conti, Daniela Marella, Mauro Scanu. **Uncertainty in statistical matching for complex sample surveys** (Full paper: 2859.pdf)

- Flaminia Musella, Daniela Marella, Paola Vicard. **Learning Bayesian networks in complex survey sampling** (Full paper: 2865.pdf)

- Leo Pasquazzi, Lucio De Capitani. **Quantile estimation with auxiliary information** (Full paper: 2857.pdf)

- Silvia Polettini, Serena Arima. **Small Area Estimation with Covariates Perturbed for Disclosure Limitation** (Full paper: 2915.pdf)

CP13 - Statistical Methods For The Analysis Of Fertility And Health

- Bianca Destavola, Lorenzo Richiardi, Daniela Zugna, Rhian Daniel, Rossella Murtas. **Birth Order, Birth Weight and Asthma: how to assess mediation and the presence of Unmeasured Confounding** (Full paper: 2944.pdf)

- Alessio Fornasin, Marco Breschi, Matteo Manfredini, Massimo Esposito. **Reproductive Change in Transitional Italy: Insights from the Italian Fertility Survey of 1961** (Full paper: 2835.pdf)

- Haftu Gebremeskel, Stefano Mazzuco. **Implementing Hierarchical Bayesian Model to Fertility Data: the case of Ethiopia** (Full paper: 2960.pdf)

- Stanislao Mazzoni, Lucia Pozzi, Marco Breschi. **Fertility and Child Mortality in the Sardinian Demographic Transition. Alghero (1866-1935)** (Full paper: 2901.doc)

- Daria Mendola, Annalisa Busetta, Daniele Vignoli. **Persistent Employment Instability and Fertility Intentions** (Full paper: 2907.docx)

CP14 - Advanced In Compositional Data Analysis

- Domenico De Stefano, Maria Rosaria D'Esposito, Giancarlo Ragozini. **Multiple Factor Analysis to Visually Explore Collaboration Structures: the Case of Technological Districts** (Full paper: 2971.pdf)

- Josep Martin-Fernandez, Josep Daunis-i-Estadella, Santiago Thió-Henestrosa. **Information Provided by Absolute, Essential and Structural Zeros in Compositional Data Sets** (Full paper: 2906.docx)

- Alessandra Menafoglio, Alberto Guadagnini, Piercesare Secchi. **Kriging prediction for functional compositional data and application to particle-size curves** (Full paper: 2836.pdf)

- Gianna Monti, Gloria Mateu-Figueras, Vera Pawlowsky-Glahn, Juan José Egozcue. **Scaled-Dirichlet regression for compositional data** (Full paper: 2928.pdf)

- Nickolay Trendafilov, Michele Gallo. **Sparse PCA for compositional data** (Full paper: 2863.pdf)

CP15 - Spatial And Spatio

- Manuela Cattelan, Cristiano Varin. **Composite likelihood estimation in spatial logistic regression** (Full paper: 2924.pdf)

- Enrico Foscolo, Marta Disegna, Fabrizio Durante. **A copula model for tourists' spending behavior** (Full paper: 2949.doc)

- Margherita Gerolimetto, Luisa Bisaglia, Paolo Gorgi. **Estimation and forecasting for binomial and negative binomial INAR(1) time series** (Full paper: 2913.pdf)

- Luca Romagnoli, Luigi Ippoliti, Richard Martin. **Kalman Filter for Estimating Bivariate GMRFs on Regular Lattice** (Full paper: 2919.pdf)

- Grazia Vicario, Giovanni Pistone. **A note on semivariogram** (Full paper: 2925.pdf)

CP16 - Environmental And Poverty Data Analysis

- Lucia Paci, Daniela Cocchi, Alan Gelfand. **Quantifying uncertainty associated with a numerical model output** (Full paper: 2824.pdf)

- Giuliana Passamani, Paola Masotti. **Smoothed Common Trend in Multivariate Time Series Air Pollution Data** (Full paper: 2952.doc)

- Silvia Perra, Stefano Cabras, Alberto Serci, Alessandra Mura, Antonella De Arca, Stefano Renoldi, Antonello Podda. **IDMS: The Sardinian Index of Multiple Deprivation** (Full paper: 2860.doc)

- Alessio Pollice, Vito Muggeo, Federico Torretta, Rocco Bochicchio, Mariana Amato. **Growth curves of sorghum roots via quantile regression with P-splines** (Full paper: 2956.pdf)

- Abel Rodriguez, Fernado Quintana. **On species sampling sequences induced by residual allocation models** (Full paper: 2885.pdf)

CP17 - Topics In Regression Models

- Matilde Bini, Vieri Del Panta, Margherita Velucchi. **Mixtures of Logit Regressions Detection with Forward Search** (Full paper: 2982.pdf)

- Silvia Columbu, Matteo Bottai. **Conditional concordance of the signs of the residuals of quantiles regressions of multivariate outcomes** (Full paper: 2965.pdf)

- Roberto Fontana, Fabio Rapallo, Maria Piera Rogantin. **Indicator functions and saturated fractions for factorial designs: a case study** (Full paper: 2888.pdf)

- Pia Clara Pafundi, Gianmarco Vacca. **Complex Redundancy Analysis models with covariate effect: a simulation study** (Full paper: 2869.doc)

- Mariangela Sciandra. **Variable selection in mixed models: a graphical approach** (Full paper: 2931.pdf)

CP18 - Bayesian Methods And Models

- Julyan Arbel, Kerrie Mengersen, Judith Rousseau. **On diversity under a Bayesian nonparametric dependent model** (Full paper: 2936.pdf)

- Antonio Canale, Bruno Scarpa. **Skew-normal nonparametric mixture models** (Full paper: 2844.pdf)

- Giulia Roli, Meri Raggi. **Bayesian hierarchical models for misaligned data: a simulation study** (Full paper: 2927.pdf)

- Catia Scricciolo, Sophie Donnet, Vincent Rivoirard, Judith Rousseau. **Posterior contraction rates for empirical Bayes procedures with applications** (Full paper: 2918.pdf)

POSTER SESSION

- Antonio Arcos, María Del Mar Rueda, David Molina, Maria Giovanna Ranalli. **Frames2: an R package for estimation in dual frames** (Full paper: 2993.pdf)

- Simona Arcuti, Alessio Pollice, Crescenza Calculli, Nunziata Ribecco, Angelo Tursi. **Spatial smoothing on complex regions: a case study on the median length of deep water rose shrimps in the North-Western Ionian Sea** (Full paper: 2921.pdf)

- Filippa Bono, Marcella Giacomarra. **The effect of support schemes on Photovoltaic installed capacity in Europe: a WDEA-STATIS analysis** (Full paper: 2988.docx)

- Maria Caterina Bramati, Flaminia Musella. **Bayesian Network structural learning in multivariate time series** (Full paper: 2856.pdf)

- Maurizio Brizzi, Alessia Orrù. **Gender Differentiation of Human Longevity in Sardinian Provinces** (Full paper: 2943.doc)

- Simone Del Sarto, M. Giovanna Ranalli, Elena Stanghellini, Davide Cappelletti, Beatrice Moroni, Stefano Crocchianti, Silvia Castellini. **Modelling the effect of vehicular–traffic and meteorology on fine particle concentration using Additive Mixed Models: the case of the town of Perugia** (Full paper: 3148.pdf)

- Rosa Falotico, Paolo Mariani. **Outsourcing in the Italian NHS: a statistical measure of mismatch between private supply and public demand** (Full paper: 2929.doc)

- Francesca Fortuna, Fabrizio Maturo. **Functional analysis of variance for parametric functional data** (Full paper: 3150.pdf)

- Massimiliano Giacalone, Paolo Carmelo Cozzucoli. **A performance comparison of the Lp-norm methods in multicollinearity situations, supposing a generalized normal distribution errors** (Full paper: 2969.doc)

- Clara Grazian, Christian Robert. **Jeffreys Priors for Mixture Models** (Full paper: 2891.pdf)

- Fabrizio Maturo, Francesca Fortuna. **Bell shaped fuzzy numbers associated with the normal curve** (Full paper: 3149.pdf)

- Daria Mendola, Annalisa Busetta, Philippe Van Kerm, Anna Maria Milito. **Material deprivation among foreigners in Italy** (Full paper: 2890.doc)

- Eugenia Nissi, Annalina Sarra. **Local Spatial Analysis of Cardiovascular Diseases in Canadian Health Regions** (Full paper: 3151.pdf)

- Zoe Nivolianitou, Myrto Konstandinidou, Chrys Caroni, Irini Kefalogianni. **Analysis of the causal factors in incidents of the Greek petrochemical industry** (Full paper: 2866.doc)

- Anna Pinto, Giulia Mascarello, Nicoletta Parise, Silvia Bonaldo, Stefania Crovato, Licia Ravarotto. **A classification of Italian consumers based on a proposed measure of their attitudes towards food risks** (Full paper: 2896.pdf)

- Giovanni Pistone. **A version of the geometry of the multivariate Gaussian model, with applications** (Full paper: 2855.pdf)

- Pasquale Recchia, Ernesto Toma. **Family size and educational outcomes: empirical evidence through multilevel approach** (Full paper: 2884.docx)

- Maria Piera Rogantin, Giovanni Pistone. **Fractionalization and Polarization** (Full paper: 2910.pdf)

- Elvira Romano, Maria Dolores Ruiz-Medina, Rosa M. M. Espejo. **A spatial functional approach for curve classification** (Full paper: 3110.pdf)

- Giorgio Russolillo, Laura Trinchera. **An extension of Non-Metric approach to inwards directed PLS Path Models** (Full paper: 3036.pdf)

- Aldo Solari, Jelle Goeman. **Adapting Benjamini-Hochberg by Simes Inequality** (Full paper: 3083.pdf)

- Susanna Zaccarin, Domenico De Stefano, Vittorio Fuccella, Maria Vitale. **Co-authorship Patterns of Italian Statisticians by Combining Different Data Sources** (Full paper: 2941.pdf)

Abstracts

Plenary Sessions

President's Invited Speaker

DEMOGRAPHIC CRISIS AND ECONOMIC CRISIS FOR ITALIAN MEZZOGIORNO: AN ICEBERG DETACHED FROM THE CONTINENT ?

Antonio Golini, (*ISTAT*), pres@istat.it

Chairperson: N. Torelli (Università di Trieste)

Abstract

The paralyzed and confused economic framework, which affects the entire Italian territory, stands out with an exponential intensity and pervasiveness in the South of Italy ("Mezzogiorno"), which both as far as key indicators of the Labour market and of the socio-economic conditions are concerned, represents the "statistics queue" of the Italian breathless society and economy.

Moreover, adding to the *economic and financial crisis* an emerging and abiding *demographic crisis*, the South of Italy is dangerously becoming more a lonely and dying *iceberg* than a marginal contributor to the European multidimensional vitality. In fact, the "Mezzogiorno" is passing through a deep demographic revolution, pivoting on a "*gender revolution*" based on women's emancipation from the "patriarchal culture". Women's empowerment has determined an incredible overturning of the fertility rate: the South of Italy (and the case of Sardinia is illuminating), which was the geographic area with the highest fertility rate in Italy, nowadays is characterized by the lowest fertility rate, also for a lack of births from foreign women. Consequently, the South of Italy is also dramatically involved in the ageing process, but without being ready to face the planetary challenge of ageing: in fact, which kind of continental, national and regional substantial economic and political supports can be reported?

Moreover, the "*Mezzogiorno Iceberg*" is also expected to carry back the other European countries, if considered not more than a national or local question between southern regions: without European shared strategies of social and economic development, starting from the "fall of the Mediterranean Wall" – which is the new fatal barrier between well-being and socio-economic backwardness – an asphyxiated doom would progressively involve all the European countries like an imploding domino. The Union for the Mediterranean and afterwards a Euro-African Union could politically and economically be the saving solution?

Plenary Session A

THE BAYESIAN 21ST CENTURY. AN APPRECIATION OF THE CONTRIBUTIONS OF DENNIS LINDLEY: 1923-2013.

Adrian Smith FRS, (*University of London, UK*), adrian.smith@london.ac.uk

Chairperson: E. George (*The University of Pennsylvania, USA*)

Abstract

The mathematical theory of statistics inference was mainly developed during the 1920s and 1930s by R. A. Fisher and - from a different perspective - by J. Neyman and E. S. Pearson. These differing approaches within the frequentist schools of inference provoked considerable academic controversy and personal animosity and led some individuals to seek to find a more solid axiomatic foundation for the subject, as there were for other branches of the mathematical sciences. Dennis Lindley was one of those who sought a proper mathematical foundation for statistical inference and this talk will trace the development of his thinking and his contribution to the growth of the Bayesian approach to statistics.

Plenary Session B

IN THE MIST OF THE DATA DELUGE, HOW TO LET THE INTERESTING FINDINGS SURFACE? TALES FROM GENETICS.

Chiara Sabatti, (*Stanford University, USA*), sabatti@stanford.edu

Chairperson: S. Petrone (Università Bocconi)

Abstract   A typical genetic study nowadays comprises the evaluation of genetic diversity at millions of loci, in thousands of individuals, with the goal of identifying correlations between this genetic variability and some traits of interest. The number of the traits in question can be very large: the volume, surface, and thickness measurements of many regions of the brain, the expression levels of tens of thousands of genes, or the quantification of thousands of metabolites. Identifying interesting correlations among such large number of variables is computationally challenging. Moreover, assuring the scientific replicability of the findings has proven difficult. In this context, the false discovery rate introduced by Benjamini and Hochberg in 1995 appears an effective criteria to define a measure of global error. I will describe two recently developed approaches that I believe address scientific questions important for genetics and provide us with additional strategies to control FDR, in the context of model selection and structured hypothesis testing.

Plenary Session C

HOW TO MARRY ROBUSTNESS AND APPLIED STATISTICS.

Andrea Cerioli (*Università di Parma*), andrea.cerioli@unipr.it

Chairperson: M. Vichi (Sapienza Università di Roma)

Abstract   A striking feature of most applied statistical analyses is the use of methods that are well known to be sensitive to outliers or to other departures from the postulated model. Since data contamination is often the rule, rather than the exception, we investigate the reasons for this contradictory (and perhaps unintended) choice. We also provide empirical evidence, in a regression setting and in a real- world problem concerning international trade, of the advantages of a new approach to data analysis based on monitoring. Our approach enhances the applicability of robust techniques and the interpretation of their results, thus yielding a positive step towards a reconciliation between robustness and applied statistics.

Specialized Sessions

SP1 - RECENT ADVANCES IN BIOSTATISTICS

Chairperson: M. Chiogna (Università di Padova)

Discussant: F. Ambrogi (Università di Milano)

Statistical challenges in epidemic modelling

Anne Presanis[1], Anne.Presanis@mrc-bsu.cam.ac.uk, Paul Birrell[2], Paul.Birrell@mrc-bsu.cam.ac.uk, Daniela De Angelis[3], daniela.deangelis@mrc-bsu.cam.ac.uk

(1) Cambridge Institute of Public Health (Great Britain, UK), (2) Cambridge Institute of Public Health (Great Britain, UK), (3) Cambridge Institute of Public Health (Great Britain, UK)

Abstract Health-related policy decision-making for epidemic control is increasingly evidence-based, exploiting multiple sources of data. Policy makers rely on models, which are required to be approximating realistically the process of interest and use all relevant information. This requirement poses a number of statically challenging problems. We review some of these challenges in this paper.

Full paper: 3056.pdf

Recent advances in genomic studies

Chiara Romualdi, chiara.romualdi@unipd.it

Università di Padova (Italy)

Abstract Current demand for understanding the behaviour of groups of related genes, combined with the greater availability of data, has led to an increased focus on statistical methods in gene set analysis. In this talk, we will review the most recent advances in the field, focussing in particular on the methodology that exploits the potential of the graphical models theory to incorporate information contained into pathways signalling networks, with the aim to develop statistically sound procedures for gene set analysis.

Full paper: 3055.pdf

Global interaction tests for predictive gene signatures in randomized clinical trials.

Federico Rotolo[1], federico.rotolo@gustaveroussy.fr,
Stefan Michelis[2], stafan.michiels@gustaveroussy.fr

(1) Gustave Roussy (France), (2) Gustave Roussy (France)

Abstract With the advent of the targeted therapy era in oncology, prognostic and predictive gene signatures are becoming increasingly important in clinical research and even in clinical practice. We investigate an approach to develop a gene signature as treatment modifier in a phase III clinical trial. We propose to apply first a permutation procedure in a survival model that controls the family-wise error rate at a pre-specified level. Only if the global test is significant, a classifier can be developed for predicting the treatment effects for future patients. We present the results of a simulation study with 100 candidate biomarkers in a relatively small clinical trial and compare different permutation tests to control the type I error.

Full paper: 3054.pdf

SP2 - CLUSTERING REAL TIME DATA STREAMS

Chairperson: F. Palumbo (Università di Napoli "Federico II")

Discussant: M. Chiodi (Università di Palermo)

Monitoring spatially dependent data streams

Antonio Balzanella[1], antonio.balzanella@gmail.com, Rosanna Verde[2], rosanna.verde@unina2.it, Antonio Irpino[3], antonio.irpino@unina2.it

(1) Università di Napoli "Federico II" (Italy), (2) Università di Napoli "Federico II" (Italy), (3) Università di Napoli "Federico II" (Italy)

Abstract   In many real world problems data streams mining deals with streams recorded by sensor networks located on some geographic area. Often, the placement of the sensors determines the presence of a spatial dependence among the data streams. This paper proposes a new method for monitoring the evolution of the spatial dependence among data streams. It is based on the variogram for histogram data which is used as synopsis for keeping a snapshot of the spatial dependence over time intervals. We propose an efficient way to compute, on-line, the variogram for histogram data and a measure for keeping track of evolutions in the spatial dependence.

Full paper: 3085.pdf

Functional classification on convex function spaces

Angela De Sanctis[1], a.desanctis@unich.it, Francesca Fortuna[2], francesca.fortuna@unich.it

(1) Università di Chieti-Pescara (Italy), (2) Università di Chieti-Pescara (Italy)

Abstract   The problem of clustering functional data is considered. In particular we refer to cases in which the functional form of the observations is known in advance. In this setting, contrary to the classical functional approach, the approximation of the function underlying the data is not required for functional clustering methods. However, clustering functional data is a difficult task because the function space is, generally, of infinite dimension. Thus, the distance among functions may have infinity solutions. For this reason, we restrict the space of the functions to a closed and convex subset in an Hilbert space. In this proposed setting an L^2 metric is applied combined clustering algorithms for finite dimensional data.

Full paper: 3141.pdf

Discovering trend clusters in sensor data streams

Donato Malerba[1], donato.malerba@uniba.it, Annalisa Appice[2], annalisa.appice@uniba.it

(1) Università di Bari (Italy), (2) Università di Bari (Italy)

Abstract   Nowadays sensors are deployed everywhere in order to support real-time data applications. They periodically gather information along a number of attribute dimensions (e.g., temperature and humidity). Applications typically require monitoring these data, fast computing aggregates, predicting unknown data, or issuing alarms. To this aim, this paper introduces a recently defined spatio-temporal pattern, called trend cluster, and its multiple applications to summarize, interpolate and detect outliers in sensor network data. As an example, we illustrate the application of trend cluster discovery to air climate data monitoring.

Full paper: 3140.pdf

SP3 - BAYESIAN NONPARAMETRICS: METHODS AND APPLICATIONS

Chairperson: M. Guindani (The University of Texas, USA)

Discussant: J. Arbel (Collegio Carlo Alberto)

A Bayesian nonparametric model for density and cluster estimation: the ε-NGG process mixture

Raffaele Argiento[1], raffaele@mi.imati.cnr.it, Ilaria Bianchini[2], alessandra.guglielmi@polimi.it, Alessandra Guglielmi[3], alessandra.guglielmi@polimi.it

(1) CNR (Italy), (2) Politecnico di Milano (Italy), (3) Politecnico di Milano (Italy)

Abstract   We define a new class of random probability measures, approximating the well-known normalized generalized gamma (NGG) process. Our new new process is defined from the representation of NGG processes as discrete measures where the weights are obtained by normalization of the jumps of a Poisson process, and the support consists of iid points, however considering only jumps larger than a threshold ε. Therefore, the number of jumps of this new process, called ε-NGG process, is a.s. finite. A prior distribution for ε can be elicited. We will assume the ε-NGG process as the mixing measure in a mixture model for density and cluster estimation. Moreover, a efficient Gibbs sampler scheme to simulate from the posterior is provided. Finally, the performance of our algorithm on the Galaxy dataset will be illustrated.

Full paper: 3032.pdf

A Bayesian nonparametric model for combining data from different experiments

Antonio Lijoi[1], lijoi@unipv.it, Bernardo Nipoti[2], bernardo.nipoti@unito.it, Igor Prunster[3], igor.pruenster@unito.it

(1) Università di Pavia (Italy), (2) Università di Torino (Italy), (3) Università di Torino (Italy)

Abstract   A number of models have been recently proposed in the Bayesian nonparametric literature for dealing with data arising from different related studies. In this paper we consider a modeling approach that relies on canonically correlated Poisson random measures. These lead to define vectors of dependent random probability measures, which are useful in the contexts of density estimation and survival analysis. With reference to the former we point out results useful for devising a Gibbs sampling algorithm. This is then used to emphasize some remarkable features, especially in terms of the clustering behavior and the borrowing information across datasets, of a class of dependent nonparametric priors based on the normalized sigma-stable process.

Full paper: 3033.pdf

On consistency issues in Bayesian nonparametric testing - a review

Judith Rousseau, rousseau@ceremade.dauphine.fr

Université Paris Dauphine and ENSAE-CREST (France)

Abstract   Although there have been a lot of developments in the recent years on estimation in Bayesian nonparametric models, from a theoretical point view as well as from a methodological point of view, little has been done on Bayesian testing in nonparametric frameworks. In this talk I will be interested on asymptotic properties of Bayesian tests when at least one of the hypotheses is nonparametric. I will first give some results on goodness of fit types of tests where one is interested in testing a parametric model against a nonparametric alternative embedding the parametric model. Then I will discuss the more delicate problem where both hypotheses are nonparametric. Such cases involve in particular tests for monotonicity, two-sample tests and estimation of the number of components in nonparametric mixture models. It will be shown that the Bayes factor or equivalently the 0-1 loss function might not be appropriate in such cases and that modifications need to be considered.

Full paper: 3034.pdf

SP4 - RECENT ADVANCES IN TIME SERIES ANALYSIS

Chairperson: F. Battaglia (Sapienza Università di Roma)

Discussant: M. Corduas (Università di Napoli "Federico II")

Regressions in Spatially Dynamic Factor Models

Lara Fontanella[1], lfontan@unich.it, Luigi Ippoliti[2], ippoliti@unich.it, Pasquale Valentini[3], pvalent@unich.it

(1) Università di Chieti-Pescara (Italy), (2) Università di Chieti-Pescara (Italy), (3) Università di Chieti-Pescara (Italy)

Abstract   This paper discusses a number of conceptual issues pertaining to the study of the relationships existing between two groups of variables which are supposed to be spatially and temporally correlated. Since it is assumed that this relationships can be studied in a reduced latent space, we provide an overview of the motivations for including spatial effects in a dynamic factor model, both from a theory-driven as well as from a data-driven perspective. Considerable attention is paid to the inferential framework necessary to carry out estimation and to the different assumptions, constraints and implications embedded in the various model specifications.

Full paper: 3082.pdf

Estimation of Extreme Quantiles for Functions of Dependent Random Variables

Liang Peng[1], peng@math.gatech.edu, Yadong Li[2], yadong.li@barclays.com, Jinguo Gong[3], jinguogong@swufe.edu.cn, Qiwei Yao[4], q.yao@lse.ac.uk

(1) Georgia Institute of Technology (U.S.A.), (2) Quantitative Analytics (U.S.A.), (3) Southwestern University (China), (4) London School of Economics (Great Britain, UK)

Abstract   We propose a new method for estimating the extreme quantiles for a function of several dependent random variables. In contrast to the conventional approach based on extreme value theory, we do not impose the condition that the tail of the underlying distribution admits an approximate parametric form, and, furthermore, our estimation makes use of the full observed data. The proposed method is semiparametric as no parametric forms are assumed on all the marginal distributions. But we select appropriate bivariate vine copulas to model the joint dependence structure. A sample quantile resulted from a large bootstrap sample drawn from the fitted joint distribution is taken as the estimator for the extreme quantile. This estimator is proved to be consistent. The reliable and robust performance of the proposed method is further illustrated by simulation.

Full paper: 3080.doc

Generalised Linear Cepstral Models for the Spectrum of a Time Series

Tommaso Proietti[1], tommaso.proietti@uniroma2.it, Alessandra Luati[2], alessandra.luati@unibo.it

(1) Università di Roma "Tor Vergata" (Italy), (2) Università di Bologna (Italy)

Abstract   We introduce the class of generalised linear cepstral models with Box-Cox link, which is based on the truncated Fourier series expansion of the Box-Cox transformation of the spectral density. The link function depends on a power transformation parameter, and encompasses the exponential model. Other important special cases are the inverse link (which leads to modelling the inverse spectrum), and the identity link. One of the merits of this model class is the possibility of nesting alternative spectral estimation methods (autoregressive, exponential, etc.) under the same likelihood-based framework.

Full paper: 3081.pdf

SP5 - NEW CHALLENGES IN SURVEY SAMPLING

Chairperson: G. Nicolini (Università di Milano)

Discussant: P. Conti (Sapienza Università di Roma)

Multiple Frame Surveys: a simplified and unified review

Fulvia Mecatti, fulvia.mecatti@unimib.it

Università di Milano-Bicocca (Italy)

Abstract   Multiple Frame Surveys are tools for dealing with imperfect sampling frames, difficult to sample populations and other complex sampling situations. However they are challenging with respect to estimation of population parameters. Since they have been introduced in the sixties, several estimators have appeared in the literature, derived under different approaches. This lack of homogeneity as well as their complex structure tricky to implement, with no easy formulae available for variance estimation, can limits the use of a multiple frame design despite its practical appealing. In this paper a review is given at both the selection and the estimation stage of a Multiple Frame Survey, with the purpose of unifying and simplifying the main available methods.

Full paper: 2990.pdf

Testing for Informativeness in Analytic Inference from Complex Surveys

Jean Opsomer[1], jopsomer@stat.colostate.edu, Wade Herndon[2], wade.herndon@colostate.edu, F. Jay Breidt[3], jbreidt@stat.colostate.edu

(1) Colorado State University (U.S.A.), (2) Colorado State University (U.S.A.), (3) Colorado State University (Italy)

Abstract   We discuss tests for informativeness of the design in analytic inference using data from a complex survey. Design informativeness occurs if a model correctly specified for the population does not hold in the sample. We generalize existing methods through a likelihood ratio test that compares the design-based fit of the expanded model to the model-based fit. We derive the asymptotic distribution of the test statistic, which is a linear combination of independent chi- square random variables. The coefficients in the linear combination are eigenvalues of a matrix that can be consistently estimated from the data. We also consider a bootstrap version, and evaluate the tests via simulation and application to real data. Empirical results show that the new test complements existing methodology, providing good power against interesting alternatives.

Full paper: 2998.pdf

On sampling elusive populations

Vijay Verma, vijay.verma46@gmail.com

Università di Siena (Italy)

Abstract   The paper describes some useful techniques for sampling elusive populations: populations for which complete frames are not available, which are small and rare, are mobile, and/or are reclusive to participation in surveys. These include sampling of informal sector establishments, multiplicity and multi-frame sampling, adaptive cluster sampling, capture-recapture and 'time-location' sampling, controlled selection, and snowball and respondent driven sampling.

Full paper: 2989.pdf

SP6 - DIRECTIONAL DATA

Chairperson: A. Fassò (Università di Bergamo)

Discussant: M. Di Marzio (Università di Chieti-Pescara)

Analyzing spatial and spatio-temporal angular and linear data using Gaussian processes

Alan Gelfand[1], alan@stat.duke.edu, Fangpo Wang[2], wangfangpo@gmail.com

(1) Duke University (U.S.A.), (2) Adobe Systems Incorporated (U.S.A.)

Abstract   It is valuable to have a better understanding of factors that influence sea motion in order to provide more accurate forecasts. Our application here is to data on wave heights and outgoing wave directions over a region in the Adriatic sea during the time of a storm, with the overarching goal of understanding the association between wave directions and wave heights and enabling improved prediction of wave behavior. Our contribution is to develop a fully model-based approach to capture joint structured spatial and temporal dependence between a linear and an angular variable. Model fitting is carried out using a suitable data augmented Markov chain Monte Carlo (MCMC) algorithm. Spatial interpolation and temporal forecasting is straightforward. We illustrate with data outputs from a deterministic wave model for region in the Adriatic Sea.

Full paper: 2991.pdf

Models for space-time directional data using Wrapped Gaussian processes

Giovanna Jona Lasinio[1], giovanna.jonalasinio@uniroma1.it, Gianluca Mastrantonio[2], gianluca.mastrantonio@uniroma3.it, Alan Gelfand[3], alan@stat.duke.edu

(1) Sapienza Università di Roma (Italy), (2) Università di Roma Tre (Italy), (3) Duke University (U.S.A.)

Abstract   In the present work we review modeling strategies based on wrapped Gaussian processes defined to model directional spatio-temporal data. We first illustrate the model-based approach to handle spatial periodic data. The wrapped Gaussian spatial process is here induced by a customary linear Gaussian process. We formulate the model as a Bayesian hierarchical one and we show that the fitting of the model is possible using standard Markov chain Monte Carlo methods. Then we move to some spatio-temporal generalizations of the spatial model. In the spatio-temporal setting we present a simulation study of our proposal aiming at understanding its computational and statistical properties. We highlight the pros and cons of this model and the difficulties arising in the implementation of the MCMCs. Eventually we provide some general advice in the use of spatio-temporal wrapped Gaussian process and we provide a real data example.

Full paper: 2995.pdf

Regression analysis of correlated circular data

Francesco Lagona, francesco.lagona@uniroma3.it

Università di Roma Tre (Italy)

Abstract   A regression model for correlated circular data is proposed by assuming that samples of angular measurements are drawn from a multivariate von Mises distribution with mean and concentration parameters that depend on covariates through suitable link functions. The model accommodates for heteroscedasticity, unstructured correlation, and specific autoregressive correlation structures. Inference is based on a Monte Carlo approximation of the log-likelihood, due to the intractability of the normalizing constant. The model is illustrated on two case studies: a longitudinal study of animal orientation and a study on the spatial distribution of sea current directions.

Full paper: 2870.pdf

SP7 - SCORING RULES AND PSEUDO

Chairperson: M. Musio (Università di Cagliari)

Discussant: F. Corradi (Università di Firenze)

Proper Scoring Rules in Statistical Inference

Philip Dawid, apd@statslab.cam.ac.uk

University of Cambridge (Great Britain, UK)

Abstract   Proper scoring rules can be use to motivate You to assess Your true uncertainty honestly, as well as to measure the quality of Your past probability forecasts in the light of the actual outcomes. They also have many other statistical applications. In this overview paper I discuss characterisations, properties and specialisations of proper scoring rules, and describe some of their uses, including robust estimation and Bayesian model selection.

Full paper: 3068.pdf

Comparison of approaches to inference in stationary AR(1) models

Monica Musio[1], mmusio@unica.it, Valentina Mameli[2], vmameli@unica.it, Alexander Dawid[3], A.P.Dawid@statslab.cam.ac.uk

(1) Università di Cagliari (Italy), (2) Università di Cagliari (Italy), (3) University of Cambridge (Great Britain, UK)

Abstract   In this paper, we consider an inference approach based on Hyvarinen's local homogeneous scoring rule within the stationary first order autoregressive framework for which both full and pairwise likelihood-based inference are available. Simulation studies were conducted to compare the estimator found by resorting to the Hyvarinen score to the full and pairwise maximum likelihood estimators.

Full paper: 3035.pdf

Approximate Bayesian Computation with proper scoring rules

Erlis Ruli[1], erlis.ruli@gmail.com, Nicola Sartori[2], sartori@stat.unipd.it, Laura Ventura[3], ventura@stat.unipd.it

(1) Università di Padova (Italy), (2) Università di Padova (Italy), (3) Università di Padova (Italy)

Abstract   We illustrate a novel approach for developing robust posterior distributions using Approximate Bayesian Computation (ABC) methods with proper scoring rules. This is formally motivated by the use of unbiased estimating functions as automatic informative summary statistics in ABC. Examples with the Tsallis score are illustrated, and comparisons with robust M-estimating functions are considered.

Full paper: 2973.pdf

SP8 – QUANTILE AND M

Chairperson: R. Miglioi (Università di Bologna)

Discussant: A. Farcomeni (Sapienza Università di Roma)

A Multilevel Quantile Regression Analysis of Chronic Diseases Affects on Physical and Mental Well-being

Giulia Cavrini[1], gcavrini@unibz.it, Sara Piombo[2], s.piombo@unibo.it, Alessandra Samoggia[3], alessandra.samoggia@unibo.it

(1) Università di Bolzano/Bozen (Italy), (2) ISTAT (Italy), (3) Università di Bologna (Italy)

Abstract   The main objective of the analysis is to evaluate how much chronic diseases affect the perceived physical and mental well-being. The effects are evaluated on Physical and Mental Component Summary indexes scores. Data are draw from the Italian Health Survey carried out by Italian National Institute of Statistics in 2005. Our analysis take into account the hierarchical data structure and consider both individual characteristics than information related to socio-economic conditions and household context. Due to the pronounced asymmetry of the response variables and not normally distributed residuals a more robust estimation methods like Linear Quantile Mixed Models.

Full paper: 3067.doc

Prediction of conditional quantiles on the half line and the unit interval using transformation models

Marco Geraci[1], m.geraci@ucl.ac.uk, Mc Jones[2], chris.jones@open.ac.uk

(1) University College London (Great Britain, UK), (2) Open University (Great Britain, UK)

Abstract   In statistical applications, transforming data may serve a number of purposes. In parametric regression, transformations are often applied to the response variable in hopes of meeting one or more assumptions of a simplified model that are otherwise unsupported by the raw (untransformed) data. A number of transformation families have been developed to address the violation of the standard assumptions of linear mean regression, and recently some of these have been extended to quantile regression to model nonlinear predictors. In this talk, we will discuss recent advances in the application of transformations to predict conditional quantiles of continuous variables with support $\mathbb{R}^+$ or the unit interval.

Full paper: 2825.pdf

Mixed hidden Markov models for quantiles

Maria Francesca Marino[1], mariafrancesca.marino@uniroma1.it, Nikos Tzavidis[2], n.tzavidis@soton.a.uk

(1) Sapienza Università di Roma (Italy), (2) Southampton University (Great Britain, UK)

Abstract   We propose a mixed hidden Markov model for continuous longitudinal data, to the quantile regression perspective. Time-constant and time-varying random parameters are added in the quantile regression model to account for time-invariant and dynamic unobserved factors affecting the variable of interest. A non-parametric maximum likelihood approach is applied to solve the numerical integration problem typically arising in the mixed model framework. Parameter estimates are then obtained by means of an EM algorithm, easily derived by exploiting the forward and backward variables defined in the so called Baum-Welsh recursion.

Full paper: 3070.pdf

Chairperson: F. Aiello (Università Kore di Enna)

Discussant: S. Iezzi (Università di Roma "Tor Vergata")

Evaluating subjective suffering in Italy with partially ordered sets

Filomena Maggino[1], maggino@unifi.it, Marco Fattore[2], marco.fattore@unimib.it

(1) Università di Firenze (Italy), (2) Università di Milano-Bicocca (Italy)

Abstract   The goal of the chapter is to investigate the effects of the global economic crisis on subjective well-being, and on its connections to objective life conditions. The study focuses primarily on suffering people that is on people who are suffering from health, economic, cultural or relational aspects, at objective or subjective level. What are the patterns of "suffering"? How are they changing over time and during the economic crisis? Is a polarization process going on in the Italian society, worsening the conditions of suffering people and sharpening the distance between them and other social groups? The chapter deals with these questions, analyzing data from the "Multipurpose Survey on Families" held by the Italian National Institute of Statistics (ISTAT), over years from 2005 to 2010. The analysis will be pursued both at national and regional level, over different years, to give a comprehensive and dynamical picture of "suffering" in Italy. To deal with such complex datasets, comprising many variables of an ordinal kind, new statistical tools will be used, based on the theory of partial orders. These tools allow us to address the construction of synthetic indicators avoiding aggregative procedures, which are unfeasible with qualitative data. To present and justify the use of such new statistical tools, a part of the chapter will be devoted to the methodological issues to face when using subjective data and, more generally, large sets of weakly interdependent ordinal variables.

Full paper: 3016.pdf

A Non-compensatory Composite Index for Comparisons over Time

Matteo Mazziotta[1], mazziott@istat.it, Adriano Pareto[2], pareto@istat.it

(1) ISTAT (Italy), (2) ISTAT (Italy)

Abstract   Most of the socio-economic phenomena such as development, quality of life and well-being have a multidimensional nature and require the definition of a set of individual indicators in order to be properly assessed. Often, individual indicators are summarized and a composite index is created. One of the main problems in constructing composite indices is the choice of a method that allows to assess changes over time. In this paper, we propose a variant of the Mazziotta-Pareto Index (MPI) which allows time comparisons across units to be made. An application to a set of individual indicators of well-being in the Italian regions is presented.

Full paper: 3011.pdf

Non-compensability in composite indicators: a robust directional frontier method

Francesco Vidoli[1], fvidoli@gmail.com, Elisa Fusco[2], elisa.fusco@libero.it, Claudio Mazziotta[3], claudio.mazziotta@uniroma3.it

(1) SOSE S.p.A. (Italy), (2) SOSE S.p.A. (Italy), (3) Università di Roma Tre (Italy)

Abstract   This paper follows the research mainstream aimed to link the efficiency frontier approaches [13] and the composite indicators (CI) methods [28]. More in detail, the main drawbacks of the CI methods based on Benefit of the Doubt (BoD) approach are the sensitivity to the outliers, the non-compensability and the lack of consideration about the marginal rate of substitution between simple indicators. Following [32] results, we propose a weighting method that bypassing all previous shortcomings suggests a comprehensive approach to construct robust and non compensatory composite indicators. This approach is based on the integration of BoD by a directional distance function.

Full paper: 3010.pdf

SP10 – PARAMETRIC AND NONPARAMETRIC MIXED EFFECT MODELS

Chairperson: P. Giordani (Sapienza Università di Roma)

Discussant: P. Brutti (Sapienza Università di Roma)

Continuous versus discrete latent structures in dynamic latent variable models

Silvia Cagnone[1], marco.alfo@uniroma1.it, Francesco Bartolucci[2], bart@stat.unipg.it

(1) Università di Bologna (Italy), (2) Università di Perugia (Italy)

Abstract   The paper proposes a comparison between dynamic models for panel data with continuous and discrete latent variables. We consider Limited Dependent Variable models (LDV) in the first case, and Latent Markov (LM) models in the second case. In both instances we use the maximum likelihood estimation method through the EM algorithm. Since the likelihood of LDV models is not tractable analytically, we implement the Gauss Hermite and the Adaptive Gauss Hermite quadrature methods for approximating the integrals involved. The comparison between the two classes of models is carried out by means of a simulation study.

Full paper: 3065.pdf

Generalized linear mixed models for over-dispersed counts

Paul H.c. Eilers, p.eilers@erasmusmc.nl

Erasmus University Medical Centre (Netherlands)

Abstract   Over-dispersion can be a nuisance when modeling count data with a generalized linear model, when it is assumed that the Poisson distribution holds. A solution may be to use the negative binomial distribution, but there are limitations to the dispersion it can handle. This paper presents a first look at modeling over-dispersed counts using a semi-parametric smooth mixing distribution. The basis is the penalized composite link model. Several applications are presented.

Full paper: 3064.pdf

Smooth random effect distribution in joint models with an application to cardiomiopathy data.

Sara Viviani, sara.viviani@fao.org

FAO (Italy)

Abstract   Longitudinal studies often entail non Gaussian primary responses. If individuals prematurely leave the study, potential non-ignorability of the missingness process may occur. A joint model for the primary response and a time-to-event may represent an appealing tool to account for dependence between the two processes. As an extension to the generalized linear mixed joint models, recently proposed, and based on Gaussian latent effects, we assume that the random effects follow a smooth, P-spline based density. To estimate model parameters, we adopt a two-step conditional Newton-Raphson algorithm. Since the maximization of the penalized log-likelihood requires numerical integration over the random effect, which is often cumbersome, we opt for a Pseudo-Adaptive Gaussian quadrature rule to approximate the model likelihood. We discuss the proposed model by analyzing an original dataset on dilated cardiomyopathies.

Full paper: 3053.pdf

SP11 - DEMOGRAPHY AND ENVIRONMENTAL EMERGENCY

Chairperson: A. De Rose (Sapienza Università di Roma)

Discussant: D. Cocchi (Università di Bologna)

Environmental Change, Migration and Displacement. Insights and developments from L'Aquila

Elena Ambrosetti[1], elena.ambrosetti@uniroma1.it, Enza Petrillo[2], enzaroberta.petrillo@uniroma1.it

(1) Sapienza Università di Roma (Italy), (2) Sapienza Università di Roma (Italy)

Abstract   This paper analyzes the environmentally-induced migration and displacement resulting from disasters and natural hazards, looking at the case study of L'Aquila's earthquake of 2009. After a general critical overview of the social science literature on this topic, the essay analyzes roots and trajectories of the forced human displacement that followed L'Aquila's earthquake, reflecting on the challenges related to post-earthquake demographic movements and post-disaster resettlement.

Full paper: 3039.pdf

Population – Environment Interactions

Wolfgang Lutz, wolfgang.lutz@oeaw.ac.at

Austrian Academy of Sciences (Austria)

Abstract   With a time horizon of up to 2050, the possible effects of environmental change on the population outlook are typically seen by experts as having only marginal impacts. This may be due to the fact that the international population experts included in the survey were almost exclusively social scientists who rarely deal with environmental issues and consider social and economic factors as the key drivers in their analyses. The slow progression of global climate change, at least on the scale of human measurement, is also a factor in the experts' judgments. Another dampening factor on the rating of environmental change is that even the more dramatic climate change scenarios do not predict significant changes over the next few decades. This paper will address the environmental change issue directly by analyzing the global population and human capital outlook for the rest of this century in a broader context of sustainable development and global environmental change.

Full paper: 3041.pdf

The use of current data to evaluate the health impact of environmental pollution: the "SENTIERI approach" and the case study of Taranto

Valerio Manno[1], valerio.manno@iss.it, Giada Minelli[2], giada.minelli@iss.it, Susanna Conti[3], susanna.conti@iss.it

(1) Istituto Superiore di Sanità (Italy), (2) Istituto Superiore di Sanità (Italy), (3) Istituto Superiore di Sanità (Italy)

Abstract   The study of the health impact of pollution is a relevant issue in Public Health and for this purpose is very useful to analyze current data (regarding mortality, hospitalization, demographic characteristics of the populations) that are of good quality, exhaustive for the whole Country at municipality level and available for a long time period. In this framework, the Study SENTIERI has been carried out, developing an original approach, aimed at assessing the health status of people resident in the Italian National Polluted Sites. This approach has been considered valid by WHO and is currently adopted by the Istituto Superiore di Sanità to evaluate the health status of residents in contaminated sites. This paper regards the main characteristics of such an approach, and presents its application to the site of Taranto, the situation of which has been object of scientific studies, heated debate on the media and also of official inquiries

Full paper: 3040.pdf

SP12 – STATISTICAL ANALYSIS AND BIG DATA

Chairperson: M. Pratesi (Università di Pisa)

Discussant: B. Scarpa (Università di Padova)

Big Data and official statistics: local experiences and international initiatives

Barteld Braaksma, b.braaksma@cbs.nl

Statistics Netherlands (Netherlands)

Abstract   Big Data is an interesting new source for official statistics, which promises abundant opportunities but also implies non-trivial challenges. The paper illustrates these with concrete examples. A particular characteristic of many Big Data sources is that their nature makes them very well suited for cross-border and multidisciplinary approaches.

Full paper: 3126.pdf

Understanding Human Mobility with Big Data

Salvatore Rinzivillo[1], Salvatore.Rinzivillo@isti.cnr.it, Fosca Giannotti[2], fosca.giannotti@gmail.com

(1) Università di Pisa (Italy), (2) Università di Pisa (Italy)

Abstract   The paper illustrates the basic methods of mobility data mining, designed to extract from the big mobility data the patterns of collective movement behavior, i.e., discover the subgroups of travelers characterized by a common purpose, and the profiles of individual movement activity, i.e., characterize the routine mobility of each traveler. We illustrate a number of concrete case studies where mobility data mining is put at work to create powerful analytical services for policy makers, businesses, public administrations, and individual citizens.

Full paper: 3027.pdf

The Virtuous Cycle of Big Data and Big Cities: a Case Study from Milan

Simone Vantini[1], simone.vantini@polimi.it, Piercesare Secchi[2], piercesare.secchi@polimi.it, Paolo Zanini[3], paolo.zanini@polimi.it

(1) Politecnico di Milano (Italy), (2) Politecnico di Milano (Italy), (3) Politecnico di Milano (Italy)

Abstract   We here present the analysis of mobile network data from the city of Milan. In particular we aim at identifying spatiotemporal patterns characterizing specific locations and/or specific periods possibly associated to different human activities taking place within the city. The analysis is carried out by means of a new dimensional reduction technique, named Hierarchical Independent Component Analysis, providing a low-dimension and sparse representation of the phenomenon.

Full paper: 3000.pdf

Solicited Sessions

SL1 - BAYESIAN MODELS FOR COMPLEX PROBLEMS

Chairperson: B. Scarpa (Università di Padova)

Bayesian principal curve clustering by species-sampling mixture models

Raffaele Argiento[1], raffaele@mi.imati.cnr.it, Alessandra Guglielmi[2], alessandra.guglielmi@polimi.it

(1) CNR (Italy), (2) Politecnico di Milano (Italy)

Abstract   In this work we are interested in clustering data whose support is "curved". For this purpose, we will follow a Bayesian nonparametric approach by considering a species sampling mixture model. Our first goal is to define a general/flexible class of distributions, such that they can model data from clusters with non standard shape. To this end, we extend the definition of principal curve given in [8] (Tibshirani 1992) into a Bayesian framework. We propose a new hierarchical model, where the data in each cluster are parametrically distributed around the Bayesian principal curve, and the prior cluster assignment is given on the latent variables at the second level of hierarchy according to a species sampling model. As an application we will consider the detection of seismic faults using data coming from Italian earthquake catalogues.

Full paper: 2877.pdf

Modeling prior knowledge on complex phenomena behaviors via partial differential equations

Laura Azzimonti[1], laura.azzimonti@polimi.it, Laura Sangalli[2], laura.sangalli@polimi.it, Piercesare Secchi[3], piercesare.secchi@polimi.it

(1) Politecnico di Milano (Italy), (2) Politecnico di Milano (Italy), (3) Politecnico di Milano (Italy)

Abstract   In this talk we shall describe an innovative method for the analysis of spatially distributed data, when a prior knowledge on the phenomenon under study is available and can be formalized in terms of a partial differential model. The prior knowledge may for instance derive from physics, physiology, morphology, chemistry or mechanics of the problem at hand, and may as well concern conditions characterizing the phenomenon at the boundary of the problem domain. The proposed models exploit advanced numerical techniques and specifically make use of the Finite Element method.

Full paper: 2923.pdf

A Bayesian Variable Selection Model for the Clustering of Time Courses in FMRI data

Michele Guindani[1], mguindani@mdanderson.org, Linlin Zhang[2], lz17@rice.edu, Francesco Versace[3], fversace@mdanderson.org, Marina Vannucci[4], marina@rice.edu

(1) University of Texas (U.S.A.), (2) Rice University (U.S.A.), (3) University of Texas (U.S.A.), (4) Rice University (U.S.A.)

Abstract   We describe a wavelet-based Bayesian nonparametric regression model for the analysis of functional magnetic resonance imaging (fMRI) data. Our framework detects regions of the brain exhibiting neuronal activity in response to a stimulus and, simultaneously, infers the association, or clustering, of spatially remote voxels exhibiting similar temporal characteristics. We use spike and slab, Markov Random Field and Dirichlet Process priors to detect brain activations, account for the complex spatial correlation structure of the brain as well as cluster correlated time courses. The model performance is illustrated on a set of fMRI synthetic data.

Full paper: 2879.pdf

Bayesian Inference with Linked Data

Andrea Tancredi[1], andrea.tancredi@uniroma1.it, Brunero Liseo[2], brunero.liseo@uniroma1.it

(1) Sapienza Università di Roma (Italy), (2) Sapienza Università di Roma (Italy)

Abstract   In this paper we will describe some advances on a Bayesian methodology for performing record linkage and making inference using the resulting matched units. In particular, we frame the record linkage process into a formal statistical model which comprises both the matching variables and the other variables included at the inferential stage. This way, the researcher is able to account for the matching process uncertainty in inferential procedures based on probabilistically linked data, and at the same time, he/she is also able to generate a feed-back propagation of the information between the working statistical model and the record linkage stage.

Full paper: 2878.pdf

SL2 – GEOSTATISTICS AND ENVIRONMENTAL APPLICATIONS

Chairperson: S. De Iaco (Università del Salento)

Computing non-separability for space-time covariance functions: a case study on PM10 data

Claudia Cappello[1], claudia.cappello@unisalento.it, Sandra De Iaco[2], sandra.deiaco@unisalento.it, Donato Posa[3], donato.posa@unisalento.it

(1) Università del Salento (Italy), (2) Università del Salento (Italy), (3) Università del Salento (Italy)

Abstract   Different forms of non-separability for space-time covariance functions have been recently defined in the literature, then various well-known non-separable space-time stationary covariance models are analyzed and classified according to the notion of non-separability. These results can be helpful to generate as well as to select appropriate covariance models for describing space-time data. Box-plots of sample non-separability ratios, classified for spatial lags and temporal lags, have been proposed as an appropriate diagnostic tool to detect different forms of nonseparability and to support the choice of the type of non-separability which should characterize a suitable covariance model for the data. The analysis of the nonseparability index for some covariance models and the space-time sample covariance function for PM10 data are presented.

Full paper: 3138.pdf

Modeling environmental quality: a case study

Sandra De Iaco[1], sandra.deiaco@unisalento.it, Sabrina Maggio[2], sabrina.maggio@unisalento.it, Monica Palma[3], monica.palma@unisalento.it, Donato Posa[4], donato.posa@unisalento.it

(1) Università del Salento (Italy), (2) Università del Salento (Italy), (3) Università del Salento (Italy), (4) Università del Salento (Italy)

Abstract   Assessing environmental quality usually requires the observation of two or more correlated variables, which are measured in several points of the study area. Sometimes, the characteristic of interest is sparsely sampled over the area, then it is convenient to incorporate some auxiliary variables, correlated with the variable of interest, into the estimation procedure. Indeed they carry relevant information for the variable being estimated, especially if they are more densely available over the domain. In this paper three different spatial interpolation approaches have been used in order to obtain spatial predictions of the variable of interest characterized by a severe lack of data.

Full paper: 3137.pdf

When it is not normal to be Gaussian

J. Jaime Gòmez-hernàndez[1], jgomez@upv.es, Teng Teng Xu[2], tenxu@posgrado.upv.es

(1) Universitat Politècnica de València (Spain), (2) Università di Pavia (Spain)

Abstract   The parameters characterizing geological formations display spatial patterns that are very often not amenable to be modeled using a multiGaussian random function. Multimodal marginal distributions, spatially connected low or high values and curvilinear features are some of the observed characteristics that a multiGaussian random function will fail to reproduce. Hydraulic conductivity is one such parameter, of great importance for the prediction of the movement of groundwater. Many efforts have been carried out in the last decades to create algorithms capable to generate hydraulic conductivity realizations according to non-Gaussian random functions. And, in the last years, these efforts have shifted to creating algorithms for the stochastic inverse modeling of hydraulic conductivity in a non-Gaussian context. The ensemble Kalman filter is one of the methods which have been favored in stochastic inverse modeling; however, it fails when hydraulic conductivity is best modeled as non-Gaussian. This paper discusses why it fails and shows an alternative algorithm based on a simple modification of the standard formulation of the ensemble Kalman filter.

Full paper: 3135.pdf

Hybrid model for urban air pollution forecasting: A stochastic spatio-temporal approach

Ana Russo[1], acrusso@fc.ul.pt, Amílcar O. Soares[2], asoares@ist.utl.pt

(1) Universidade de Lisboa (Portugal), (2) Universidade Técnica de Lisboa (Portugal)

Abstract   Air pollution is usually driven by a complex combination of factors in which meteorology, physical obstacles, and interactions between pollutants play significant roles. Considering the characteristics of urban atmospheric pollution and its consequent impacts on human health and quality of life, forecasting models emerge as an effective tool to identify and forecast AP episodes. The overall objective of the present work was to produce forecasts of pollutant concentrations with high spatio-temporal resolution and to quantify the uncertainty in those forecasts. Therefore, a new approach was developed based on a two-step methodology. Firstly, neural network models were used to generate short-term temporal forecasts based on air pollution and meteorology data. The accuracy of those forecasts was then evaluated against an independent set of historical data. Secondly, local conditional distributions of the observed values with respect to the predicted values were used to perform spatial stochastic simulations with local distributions for the entire geographic area of interest.

Full paper: 3139.pdf

SL3 – ROBUST METHODS FOR THE ANALYSIS OF COMPLEX DATA

Chairperson: M. Riani (Università di Parma)

Introducing Prior Information into the Forward Search for Regression

Anthony Atkinson[1], A.C.Atkinson@lse.ac.uk, Aldo Corbellini[2], aldo.corbellini@unipr.it

(1) London School of Economics (Great Britain, UK), (2) Università di Parma (Italy)

Abstract   The forward search provides a flexible and informative form of robust regression. We describe two ways of introducing prior information into the regression model used in the search, either through fictitious observations or through prior distributions of the parameters. The relationship between the two methods is established.The extension to the forward search is not entirely straightforward, requiring weighted regression. Forward plots are used to exhibit the effect of prior information on inferences.

Full paper: 2829.pdf

Analysis of complex data in official statistics

Andrea Cerasa[1], andrea.cerasa@jrc.ec.europa.eu, Francesca Torti[2], francesca.torti@jrc.ec.europa.eu, Domenico Perrotta[3], domenico.perrotta@ec.europa.eu

(1) JRC (Italy), (2) JRC (Italy), (3) JRC (Italy)

Abstract   International trade data are often affected by multiple linear populations and heteroscedasticity. An immediate consequence is the false declaration of outliers. We propose the monitoring of the White test statistic through the Forward Search as a new robust tool to test the presence of heteroscedasticity. We show that, if the data are considered on a monthly basis, the heteroscedastic problem can be often bypassed.

Full paper: 3047.pdf

Robust covariance matrix estimation with regularization

Pietro Coretto[1], pcoretto@unisa.it, Christian Hennig[2], c.hennig@ucl.ac.uk

(1) Università di Salerno (Italy), (2) University College London (Great Britain, UK)

Abstract   The sample covariance matrix performs poorly in situations with outliers and high dimensionality. Most high-breakdown point estimators of multivariate location and scatter cannot cope with datasets with small n/p ratio, where n is the sample size and p the dimension. In this paper we introduce and discuss an estimator of location and scatter defined as the solution of a system of nonlinear (implicit) equations. These equations are based on a pseudo-model representation of the underlying probability measure. The latter consists in an improper density function called "Black Hole Density". The estimator smoothly trims data points far away from the center. The robustness of the procedure is based on tunings that have a straightforward interpretation. For datasets with small n/p ratio, regularization is provided by a constraint that bounds the condition number of the covariance matrix. Empirical evidence suggests that the proposed estimator compare well with other robust location-scatter estimators.

Full paper: 3046.pdf

Robust model-based clustering and mixture

Luis Garcıa-Escudero[1], lagarcia@eio.uva.es, Alfonso Gordaliza[2], alfonsog@eio.uva.es, Carlos Matran[3], matran@eio.uva.es, Agustın Mayo-iscar[4], agustinm@eio.uva.es

(1) Universidad de Valladolid (Spain), (2) Universidad de Valladolid (Spain), (3) Universidad de Valladolid (Spain), (4) Universidad de Valladolid (Spain)

Abstract   The use of trimming procedures is common in several statistical frameworks to achieve robustness against anomalous observations. Trimming can be also applied in model-based clustering and mixture modeling and this work reviews some trimming proposals aimed at robustifying these two widely used techniques. The consideration of computationally feasible constraints on the scatters of the clusters and mixture components also plays an important role in these proposals. This work also presents some recent extensions of these methods to related problems. Some new tools helping the user to fix the different tuning parameters are shown.

Full paper: 3058.pdf

SL4 – STATISTICS FOR ENVIRONMENTAL PHENOMENA AND THEIR INTERACTIONS

Chairperson: F. Bruno (Università di Bologna)

Spatial modeling for air pollution epidemiology: hospital admission risk for cardio-respiratory diseases in Torino province

Paola Berchialla[1], paola.berchialla@unito.it, Marta Blangiardo[2], m.blangiardo@imperial.ac.uk, Michela Cameletti[3], michela.cameletti@unibg.it, Francesco Finazzi[4], francesco.finazzi@unibg.it, Maria Villoria[5], maria.francovilloria@unito.it, Rosaria Ignaccolo[6], rosaria.ignaccolo@unito.it

(1) Università di Torino (Italy), (2) Imperial College London (Great Britain, UK), (3) Università di Bergamo (Italy), (4) Università di Bergamo (Italy), (5) Università di Torino (Italy), (6) Università di Torino (Italy)

Abstract   We analyse the association between atmospheric pollution and hospital admissions for respiratory and cardiovascular causes in the province of Torino in 2004. The proposed model, which is fitted using INLA, includes fixed effects at individual and municipality level and spatially structured random effects for pollutants. Preliminary results suggests higher risk of hospitalization for females aged less than 15 and over 64 years and on days with higher temperatures.

Full paper: 3051.pdf

Kriging for functional data: uncertainty assessment

Rosaria Ignaccolo[1], rosaria.ignaccolo@unito.it, Maria Franco Villoria[2], maria.francovilloria@unito.it

(1) Università di Torino (Italy), (2) Università di Torino (Italy)

Abstract   We predict a curve at an unmonitored site taking into account exogenous variables using a functional kriging model with external drift and, alternatively, an additive model with a spatio-temporal smooth term. To evaluate uncertainty of the predicted curves, a semi-parametric bootstrap approach is used for the first, while standard inference is used for the second. The performance of both approaches is illustrated on pollutant functional data.

Full paper: 2953.pdf

Bayesian structural equation modeling for factors influencing residential radon levels

Luigi Ippoliti[1], ippoliti@unich.it, Lara Fontanella[2], lfontan@unich.it, Pasquale Valentini[3], pvalent@unich.it, Annalina Sarra[4], asarra@dmqte.unich.it, Sergio Palermi[5], s.palermi@artaabruzzo.it

(1) Università di Chieti-Pescara (Italy), (2) Università di Chieti-Pescara (Italy), (3) Università di Chieti-Pescara (Italy), (4) Università di Chieti-Pescara (Italy), (5) Agency of Environmental Protection of Abruzzo (Italy)

Abstract   Motivated by the problem of modelling high-dimensional multivariate referenced data, arising in many areas of research, this articles proposes a Generalised Latent Spatial Quantile Regression (GLSQR) model as a reliable solution in studying the effects of some covariates across the quantiles of the response distribution. In addition, the discussed model warrants consideration when the matrix of the explanatory variables is defined through a set of spatial common latent factors. The latent factors and quantile regression components are estimated through a hierarchical Bayesian procedure and MCMC algorithms are used to provide full probabilistic inference. We illustrate the use of our GLSQR model with application to radon data,as motivating example coming from the environmental protection research.

Full paper: 3023.pdf

Assessment of bayesian models for rainfall field reconstruction

Elena Scardovi[1], elena.scardovi2@unibo.it, Francesca Bruno[2], francesca.bruno@unibo.it, Fedele Greco[3], fedele.greco@unibo.it

(1) Università di Bologna (Italy), (2) Università di Bologna (Italy), (3) Università di Bologna (Italy)

Abstract   We develop a Bayesian hierarchical model for predicting rain occurrence and amount on a fine-pixel grid, by modelling the relationship between rain gauge and radar data. An enrichment consists in including neighborhood information as a covariate for dealing with spatial misalignment. A focus of this work is on evaluation of the predictions, which can consist in point or probabilistic forecasts. In the first case, competing models can be assessed and ranked on the basis of consistent scoring functions. On the other hand, probabilistic forecasts consist in full predictive distributions and ought to be calibrated and sharp. Suitable adaptation to the zero-inflated model of standard tools for evaluation are applied to the prediction of hourly rainfall in Emilia-Romagna region.

Full paper: 3013.pdf

SL5 – MIXTURE AND LATENT VARIABLE MODELS FOR CAUSAL INFERENCE AND ANALYSIS OF SOCIO

Chairperson: C. Rampichini (Università di Firenze)

A multidimensional finite mixture SEM for non-ignorable missing responses to test items

Silvia Bacci[1], silvia.bacci@stat.unipg.it, Francesco Bartolucci[2], bart@stat.unipg.it

(1) Università di Perugia (Italy), (2) Università di Perugia (Italy)

Abstract   A structural equation model is proposed to deal with dichotomously-scored items in the presence of missing not at random responses. Two types of latent variables are introduced: one refers to the abilities measured by the questionnaire and the other describes the propensity to answer. A semi-parametric approach is adopted, where both types of latent variable are discretely distributed. In such a way, the proposed structural equation model reduces to a multidimensional latent class item response theory model. Individual covariates may also be included in order to explain the probabilities of belonging to each latent class. For this aim, a multinomial logistic parametrization is introduced. A simulation study is then performed to evaluate the finite-sample properties of the parameter estimates.

Full paper: 2831.pdf

Finite mixtures for multivariate mixed data: a Parafac-based approach

Paolo Giordani, paolo.giordani@uniroma1.it

Sapienza Università di Roma (Italy)

Abstract   We propose a flexible regression model for multivariate mixed responses. Discrete, outcome-specific, latent effects are used to account for potential dependence between outcomes. For this purpose, we define a multidimensional approach, where possibly different numbers of locations are used for each margin, and joined by a full association structure. This structure may be further simplified by using a Parafac-based approach.

Full paper: 3072.pdf

A multilevel model for repeated cross-sectional data with stochastic volatility

Lucia Modugno[1], lucia.modugno@unibo.it, Silvia Cagnone[2], silvia.cagnone@unibo.it, Simone Giannerini[3], simone.giannerini@unibo.it

(1) Università di Bologna (Italy), (2) Università di Bologna (Italy), (3) Università di Bologna (Italy)

Abstract   In this paper we propose a multilevel approach for the analysis of repeated cross-sectional data that exhibit volatility effects. We treat individuals as clustered within time-points so that the dynamics over time is modelled at the second level. Items sold in auction present a structure like that of repeated cross-sectional surveys since different goods are sold at different time-points. For prices of artworks, as well as for other assets (financial, insurance, etc.) the hypothesis of constant volatility appears unreasonable. In this work we combine a multilevel model with autoregressive random effects and a stochastic volatility model in order to account for the kurtosis and the volatility pattern of prices. We apply the model to Tribal art auction prices and show improvement over existing proposals both in terms of fit and forecasting.

Full paper: 3125.pdf

Multivariate multilevel modelling of student achievement data

Leonardo Grilli[1], grilli@disia.unifi.it, Fulvia Pennoni[2], fulvia.pennoni@unimib.it, Carla Rampichini[1], rampichini@disia.unifi.it, Isabella Romeo[2], isabella.romeo@unimib.it

(1) Università di Firenze (Italy), (2) Università di Milano-Bicocca (Italy)

Abstract   The paper analyses the Italian sample of the TIMSS and PIRLS 2011 Combined International Database, containing the results of the assessment on Reading, Mathematics and Science for a sample of 4th grade pupils, alongside with several variables about the pupils, their teachers and their schools. This is the first time that TIMSS and PIRLS surveys are conducted on the same sample, thus allowing to jointly analyse the achievement for the considered subjects. We propose a multivariate multilevel model, with pupils nested within classes, to explore the determinants of achievement in the three subjects.

Full paper: 2959.pdf

SL6 – EQUITY AND SUSTAINABILITY: THEORY AND RELATIONSHIPS

Chairperson: A. Lemmi (Università di Siena)

Categorization of National Economies through Environmental, Social and Economic Indicators.

Simone Bastianoni[1], bastianoni@unisi.it, Luca Coscieme[2], coscieme2@unisi.it, Federico Pulselli[3], fpulselli@unisi.it

(1) Università di Siena (Italy), (2) Università di Siena (Italy), (3) Università di Siena (Italy)

Abstract The definition of a sustainable economic model has to deal with environmental, social and economic aspects. National Economies can be described by a combination of measures in an input-state-output scheme. In this framework, a set of three indicators is used to identify different expressions of world economy. Energy per capita is used as an environmental measure of input; employed to population ratio, as a social measure of the internal organization of the system; per capita GDP, as an economic measure of output. In this framework of analysis, different national economies can be compared, clusters are observed, and dynamics can be studied by taking into account natural, social and economic changes.

Full paper: 2981.pdf

Environmental resources, landscape and cultural heritage, economic conditions: fundamental components to measure well-being

Alessandra Ferrara[1], ferrara@istat.it, Angela Ferruzza[2], ferruzza@istat.it, Nicoletta Pannuzi[3], pannuzi@istat.it

(1) ISTAT (Italy), (2) ISTAT (Italy), (3) ISTAT (Italy)

Abstract The present analysis aims to outline the socio–economic profiles of different territories (the Italian regions), based on a set of indicators drawn from those selected for the dimensions Environment, Landscape and cultural heritage and Economic well-being within the Istat - Cnel Bes Project for the measurement of equitable and sustainable well-being. The indicators describe both the state of common goods such as environmental, heritage and landscape values, and the economic conditions of the citizens; for each territory, an assessment of the economic conditions is compared with some measures of erosion or impact on environmental and landscape resources, seeking to highlight how different economic condition are associated with more or less efficient maintenance or enhancement of environmental and landscape assets.

Full paper: 3006.doc

Measuring the sustainability performances of the Italian regions

Tommaso Luzzati[1], tluzzati@ec.unipi.it, Bruno Cheli[2], bcheli@ec.unipi.it

(1) Università di Pisa (Italy), (2) Università di Pisa (Italy)

Abstract The aim of this paper is twofold, methodological and empirical. From the methodological point of view it aims at contributing to the debate about composite indicators. From the empirical one it assesses the relative sustainability of the Italian regions. Instead of building a single composite indicator (score) for each region, we calculate many composite indices by combining different weighting systems and rules of normalization and aggregation. In this way, we get a distribution function of the ranks (and a plausible rank range) for each country. Such an approach represents a good compromise between the need of synthesising the information provided by many variables and the need to avoid the loss of relevant information that occurs when several indicators are aggregated into a single composite index.

Full paper: 2987.pdf

A framework for monitoring country sustainability

Andrea Regoli[1], andrea.regoli@uniparthenope.it, Antonella D'Agostino[2], dagostino@uniparthenope.it, Francesca Gagliardi[3], gagliardi10@unisi.it, Laura Neri[4], laura.neri@unisi.it

(1) Università di Napoli "Parthenope" (Italy), (2) Università di Napoli "Parthenope" (Italy), (3) Università di Siena (Italy), (4) Università di Siena (Italy)

Abstract   The development of systemic monitoring of country sustainability requires to tackle the complexity of human–environmental systems. To this end starting with a simple input/state/output scheme that shows the links among environment, community's economy/society and finally "real economy", we discuss on some more specific aspects of such scheme and we develop a classification of economies according to the proposed framework.

Full paper: 2980.doc

SL7 - ADVANCES IN BAYESIAN STATISTICS

Chairperson: L. Tardella (Sapienza Università di Roma)

An ABC/quasi-likelihood approach for linkage/GWAS study of a Sardinian genetic isolate

Stefano Cabras[1], s.cabras@unica.it, María Eugenia Castellanos Nueda[2], maria.castellanos@urjc.es, Erlis Ruli[3], erlis@stat.unipd.it, Mario Pirastu[4], m.pirastu@igp.cnr.it, Maria Pina Concas[5], m.p.concas@igp.cnr.it

(1) Università di Cagliari and Universidad Carlos III de Madrid (Italy), (2) Universidad Rey Juan Carlos (Spain), (3) Università di Padova (Italy), (4) CNR (Italy), (5) CNR (Italy)

Abstract   In this work we discuss an application of recent developments of Approximate Bayesian Computation (ABC) methods coupled with the quasi-likelihood theory, to the problem of estimating the relation between genotype and phenotype in genetic isolates with known genealogy between subjects from the common founders. The combination of quasi-likelihoods and ABC allows automatic ABC inference when the likelihood function is intractable. Indeed, the quasi-likelihood delivers an approximation of the intractable likelihood that can enter either into an ABC-MCMC algorithm as a proposal density or can be used as a surrogate of the intractable likelihood function directly in the Bayes theorem. The proposed method is applied to a Genome Wide Association Study of the reduced Mean Cell Volume (MCV<72) on a Sardinian genetic isolate for which the genealogy relating all the observed subjects is known.

Full paper: 3030.pdf

Bayesian nonparametric estimation and asymptotics with misspecified density models

Pierpaolo De Blasi[1], pierpaolo.deblasi@unito.it, Stephen Walker[2], s.g.walker@math.utexas.edu

(1) Università di Torino (Italy), (2) University of Texas (U.S.A.)

Abstract   In this paper we summarize some recent findings that can be found in [1, 2] and concern Bayesian misspecified density models. We first discuss a prior summability condition for the posterior to accumulate around the densities in the model closest in the Kullback–Leibler sense to the data generating density. This condition is shown to be satisfied by popular nonparametric priors such as infinite mixtures of normal densities and Gaussian process priors. In smooth parametric models, the posterior shrinks at a $\sqrt{n}$-rate of convergence around the parameter value minimizing the Kullback–Leibler divergence. In this setting we show how Gaussian process priors can be used to consistently estimate the discrepancy of the parametric model from the data generating density. A novel Monte Carlo Markov Chain methods is devised for dealing with intractable normalizing constants.

Full paper: 3087.pdf

Determinantal Priors for Variable Selection

Edward George[1], edgeorge@wharton.upenn.edu, Veronika Ročková[2], vrockova@wharton.upenn.edu

(1) University of Pennsylvania (U.S.A.), (2) University of Pennsylvania (U.S.A.)

Abstract   Determinantal point processes (DPPs) provide a probabilistic formalism for modeling repulsive distributions over subsets. Such priors encourage diversity between selected items through the introduction of a kernel matrix that determines which items are similar and therefore less likely to appear together. We investigate the usefulness of such priors in the context of spike-and-slab variable selection, where penalizing predictor collinearity may reveal more interesting models.

Full paper: 3086.pdf

Approximate Bayesian Inference for Copula Models

Brunero Liseo, brunero.liseo@uniroma1.it

Sapienza Università di Roma (Italy)

Abstract   We propose a method for making inference on a functional of a multivariate distribution. The method is based on a copula representation and it is based on the properties of an Approximate Bayesian algorithm based on empirical likelihood.

Full paper: 3084.pdf

SL8 – STATISTICAL MODELS FOR THE ANALYSIS OF ENERGY MARKETS

Chairperson: L. Grossi (Università di Verona)

Are there any effects of Fukushima accident on the diffusion of nuclear energy?

Claudia Furlan[1], furlan@stat.unipd.it, Mariangela Guidolin[2], guidolin@stat.unipd.it, Renato Guseo[3], renato.guseo@unipd.it

(1) Università di Padova (Italy), (2) Università di Padova (Italy), (3) Università di Padova (Italy)

Abstract   The year 2012 registered a decline of nuclear power consumption in several countries such as the US, France, Germany, and other OECD members. Was it a post-Fukushima outcome? We focus on the nuclear consumption leaders - the US, France and Germany - and explore, through diffusion models, whether and to what extent Fukushima had an effect on their consumption dynamics. In particular, we compare the evolutionary behavior estimated with the entire time series and that obtained by excluding the last two observations (2011 and 2012): how would the forecasts have been before Fukushima? The results show that nuclear energy policy of these countries does not seem to be affected by the accident: nuclear technology is being dismissed with different speeds and timing due to rising costs in risk management and possible shortages in Uranium 235 supply.

Full paper: 2826.pdf

Robust forecasting of electricity prices with nonlinear models and exogenous regressors

Luigi Grossi[1], luigi.grossi@univr.it, Fany Nan[2], fany.nan@univr.it

(1) Università di Verona (Italy), (2) Università di Verona (Italy)

Abstract   It is well known that GM estimators for linear models are consistent and lead to a small loss of efficiency with respect to LS estimator. When they are extended to threshold models, which are piecewise linear models, the consistency of GM estimators is guaranteed only under certain choices of the objective function. In this paper we suggest the use of robust SETAR (Self Exciting Threshold AutoRegressive) processes to model and forecast electricity prices observed on deregulated markets. The main advantages of estimating robust SETAR models is the possibility to capture two very well-known stylized facts of electricity prices: nonlinearity produced by changes of regimes and the presence of sudden spikes due to inelasticity of demand.

Full paper: 2920.pdf

First results on the Italian-Slovenian electricity market coupling

Matteo Pelagatti[1], matteo.pelagatti@unimib.it, Lucia Parisio[2], lucia.parisio@unimib.it

(1) Università di Milano-Bicocca (Italy), (2) Università di Milano-Bicocca (Italy)

Abstract   Since January 1st, 2011 the electricity exchanges of Italy and Slovenia are working under a mechanism of market coupling for their respective day-ahead markets. Similar mechanisms are being implemented in many European countries to foster the integration of electricity markets that eventually will merge into one large European power market. This short paper is one of the first works in which, by analysing market results, we try to assess the degree of integration of the Italian and Slovenian electricity markets due to the market coupling policy.

Full paper: 2958.pdf

Renewable flows and congested lines in the Italian power grid: Binary time series and vector autoregressions

Alessandro Sapio, alessandro.sapio@uniparthenope.it

Università di Napoli "Parthenope" (Italy)

Abstract   This paper estimates the impact of renewables on congestion in the Sicily-Southern Italy transmission line, using data for 2012. Two empirical approaches are followed: a binary autoregression of congestion probabilities, and a vector autoregressive (VAR) model that allows to explore the interrelationships between RE supply, market power exercise, and congestion patterns.

Full paper: 2970.pdf

SL9 – RECENT DEVELOPMENTS IN SAMPLING THEORY

Chairperson: T. Di Battista (Università di Chieti-Pescara)

Weak convergence and empirical processes in survey sampling

Pier Luigi Conti, pierluigi.conti@uniroma1.it

Sapienza Università di Roma (Italy)

Abstract   In this paper an "equivalent version" of the empirical process is introduced and studied in sampling finite populations under fairly general sampling designs. This allow to establish the asymptotic normality of estimators commonly used in practice. As a by-product, estimates of the variance are also obtained.

Full paper: 3066.pdf

On the linearization of inequality indexes in the design-based framework

Pier Francesco Perri[1], pierfrancesco.perri@unical.it, Lucio Barabesi[2], lucio.barabesi@unisi.it, Giancarlo Diana[3], diana@stat.unipd.it

(1) Università della Calabria (Italy), (2) Università di Siena (Italy), (3) Università di Padova (Italy)

Abstract   Linearization methods are customarily adopted in sampling surveys to obtain approximated variance formulae for estimators of nonlinear functions of finite-population totals which can be usually rephrased in terms of statistical functionals. In the present paper, by considering Deville's (1999) approach stemming on the concept of design-based influence curve, we provide a general result for linearizing large families of inequality indexes. As an example, the achievement is applied to the Gini and the Amato indexes. We also discuss the case when income data are supposed to be collected by means of the randomized response technique.

Full paper: 2992.pdf

Sampling strategies for diversity indexes estimation in presence of rare species

Caterina Pisani[1], caterina.pisani@unisi.it, Lorenzo Fattorini[2], lorenzo.fattorini@unisi.it

(1) Università di Siena (Italy), (2) Università di Siena (Italy)

Abstract   The estimation of diversity indexes is considered when species abundance is estimated by means of plots thrown onto the study area in accordance with probabilistic schemes. In presence of rare species the sample diversity index estimators heavily underestimate their population counterparts. Under the uniform random sampling variance estimation and reduction of finite-sample negative bias are performed using jackknife. Despite its theoretical simplicity, uniform random sampling may lead to uneven coverage of the study area. To avoid the shortcoming, the use of stratified sampling is adopted. Large sample properties of diversity indexes estimators are considered under this alternative scheme as well as the use of jackknife to deal with negative bias.

Full paper: 3020.pdf

Spatially balanced adaptive web sampling

Emilia Rocco, rocco@disia.unifi.it

Università di Firenze (Italy)

Abstract   The paper deals with sampling from a finite population that is distributed over space and has an highly uneven spatial distribution. It suggests a sampling design that allocates a portion of the sample units well-spread over the population and selects sequentially the remaining units in sub-areas that appear of more interest according to the study variable values observed during the survey. In order to estimate the population mean while using this sampling design, a computational intense estimator, obtained via the Rao-Blackwell approach, is proposed and a resampling method that makes the inference computationally feasible is used. The whole sampling strategy is evaluated through several Monte Carlo experiments.

Full paper: 3022.pdf

SL10 - FUNCTIONAL DATA ANALYSIS

Chairperson: R. Verde (Seconda Università di Napoli)

Some advances on semi-parametric functional data modelling

Aldo Goia, aldo.goia@eco.unipmn.it

Università del Piemonte Orientale (Italy)

Abstract   The aim of this work is to present some recent advances in single index modelling when the covariate is a functional variable and the response a scalar. We pay for special attention to the situation of possible structural changes which produce unsmooth relevant direction. An estimation procedure combining spline functions and the well-known Nadaraya-Watson approach is illustrated. From an example of interest in the spectrometry, it emerges that the method provides a nice exploratory tool both for analyzing structural changes in the spectrum and for visualizing the most informative directions, still keeping good predictive power.

Full paper: 3007.pdf

A Functional Model for Detecting Changes in Evolving Shapes Brain Tumors

Elvira Romano[1], elvira.romano@unina2.it, Jorge Mateu[2], mateu@mat.uji.es, Iulian Teodor Vlad[3], vlad@mat.uji.es

(1) Università di Napoli "Federico II" (Italy), (2) Universitat Jaume I (Spain), (3) Universitat Jaume I (Spain)

Abstract   In this work we propose a dynamic functional model for inspecting the evolution and predicting the deformation of a brain tumor shape. We illustrate the usefulness of the proposed model on real and simulated data.

Full paper: 3045.pdf

Functional data analysis in spaces of surfaces

Laura Sangalli, laura.sangalli@polimi.it

Politecnico di Milano (Italy)

Abstract   The talk presents a novel functional data analysis technique for surface estimation and spatial smoothing, at the interface between statistics and numerical analysis.

Full paper: 2986.pdf

Hypothesis Testing in Functional Data Analysis: a Non-parametric Approach

Simone Vantini[1], simone.vantini@polimi.it, Alessia Pini[2], alessia.pini@polimi.it

(1) Politecnico di Milano (Italy), (2) Politecnico di Milano (Italy)

Abstract   We present the Interval Testing Procedure (ITP), a non parametric procedure based on permutation tests that enables inference for functional data. The procedure is based on an interval-wise control of the Family Wise Error Rate (FWER), which allows selecting the basis components which present a statistically significant test result. We present some theoretical results on the power and error control of the ITP. Then, by means of a simulation study, we empirically compare the power and the FWER of the ITP with the ones of the Benjamini-Hochberg procedure.

Full paper: 3001.pdf

SL11 - EXTREMES AND DEPENDENT SEQUENCES

Chairperson: S. Padoan (Università Bocconi)

Latent process models for temporal extremes with an application to rainfall data

Paola Bortot[1], paola.bortot@unibo.it, Carlo Gaetan[2], gaetan@unive.it

(1) Università di Bologna (Italy), (2) Università di Venezia (Italy)

Abstract   The paper reviews a hierarchical modeling procedure for extreme values of stationary series proposed by Bortot and Gaetan (2013). A modification of the model to allow for different types of marginal tail behavior is presented. The two models are applied to the analysis of extreme hourly and daily rainfalls and their performances compared.

Full paper: 3142.pdf

The clustering of extreme values for some asymmetric GARCH-type models

Fabrizio Laurini, fabrizio.laurini@unipr.it

Università di Parma (Italy)

Abstract   Several models with conditional heteroschedasticity have been studied in financial econometrics, with the simple GARCH(1,1) with Gaussian innovation representing the standard benchmark. There is evidence of asymmetry in some daily data and more flexible models, which take such an asymmetry into account, have become recently popular. Understanding the extremal behavior of asymmetric processes becomes very important to build proper inference about extremal events. For processes satisfying mild mixing conditions the clustering of extreme values is characterized by a single key-parameter, known as the extremal index, which represents the average clusters size of values which exceed a high-level threshold. An approach extending results for the GARCH(1, 1) is presented, with skew-t innovation.

Full paper: 3143.pdf

Inference of multivariate dependence structures in extreme value theory

Giulia Marcon[1], giulia.marcon@phd.unibocconi.it, Simone A. Padoan[2], simone.padoan@unibocconi.it, Philippe Naveau[3], Philippe.Naveau@lsce.ipsl.fr, Pietro Muliere[4], pietro.muliere@unibocconi.it

(1) Università Bocconi (Italy), (2) Università Bocconi (Italy), (3) Laboratoire des Sciences du Climat et l'Environnement (France), (4) Università Bocconi (Italy)

Abstract   Modelling dependence structures in high-dimensional problems of extreme events is of interest in several application areas. Current dependence models for multivariate extremes are based upon max-stable distributions and one approach is to investigate the Pickands dependence function through nonparametric estimators. In the bivariate setting, there exist several estimators while more problematic are the assumptions that must be satisfied in multivariate extremes. The aim is to briefly review an existing nonparametric inference method for estimating the Pickands function, which assume known marginal distributions, in the multivariate framework.

Full paper: 2996.pdf

SL12 – ISSUES IN ECOLOGICAL STATISTICS

Chairperson: A. Pollice (Università di Bari)

A changepoint analysis on spatio-temporal point processes

Linda Altieri[1], linda.altieri@unibo.it, Marian Scott[2], marian.scott@glasgow.ac.uk, Janine Illian[3], janine@mcs.st-and.ac.uk

(1) Università di Bologna (Italy), (2) University of Glasgow (Great Britain, UK), (3) University of St Andrews (Great Britain, UK)

Abstract This work introduces a likelihood ratio based test statistic for detecting change points over time in the inhomogeneous intensity of a spatio-temporal point process. We propose a new method for detecting changes by fitting a spatio-temporal log-Gaussian Cox process model using INLA. A simulation study assessing the validity and properties of the test is presented.

Full paper: 3009.pdf

Linking metapopulation modelling and Information Theory for area-wide pest management

Antonella Bodini[1], antonella.bodini@mi.imati.cnr.it, Gianni Gilioli[2], gianni.gilioli@med.unibs.it

(1) National Reserch Council (Italy), (2) Università di Brescia (Italy)

Abstract In [3], management strategies for threatened species conservation are evaluated by measuring the distance of a vector of presence/absence predicted by a metapopulation model from the vector representing some specified target state, like an extinction state, by the Kullback-Leibler divergence. In [4], it has been shown how this method supports the evaluation of short-term strategies for a pest, the Pine processionary moth metapopulation in Aspromonte (Italy). Here, we show by simulations that the method can deal with the management of pest organized in finite networks of discrete habitat patches with given areas and spatial locations.

Full paper: 2983.pdf

Uncertainty of methodologies assessing ecological status in transitional water systems

Ilaria Rosati[1], ilaria.rosati@unisalento.it, Alessio Pollice[2], alessio.pollice@uniba.it, Serena Arima[3], serena.arima@uniroma1.it, Giovanna Jona Lasinio[4], giovanna.jonalasinio@uniroma1.it, Alberto Basset[5], alberto.basset@unisalento.it

(1) Università del Salento (Italy), (2) Università di Bari (Italy), (3) Sapienza Università di Roma (Italy), (4) Sapienza Università di Roma (Italy), (5) Universitat Politècnica de València (Italy)

Abstract Assessments of ecological status according to the principles devised by the Water Framework Directive (WFD) are always associated with some degree of uncertainty. This uncertainty stems from the inevitable imperfection of the assessment criteria (i.e. development of indicators of ecological status, refinement of reference conditions and class boundaries, and routines for integrated assessment) and from the uncertainty of measurements. Here, we address the ecological status classification issue in transitional water ecosystems, using benthic macroinvertebrates and four proposed multimetric indices (BAT, BITS, M-AMBI and ISS), likely to respond differently to different sources of stress and natural variability adding uncertainty to resulting classifications. In order to investigate the possible contrasting behavior of the four multimetric indices, we propose two multivariate Bayesian hierarchical models in which the multimetric indices are jointly modeled as function of abiotic covariates and indicators of anthropogenic pressures.

Full paper: 3074.pdf

Modelling and inferential issues for behavioral patterns in capture-recapture data

Luca Tardella[1], luca.tardella@uniroma1.it,
Danilo Alunni Fegatelli[2], danilo.alunnifegatelli@uniroma1.it

(1) Sapienza Università di Roma (Italy), (2) Sapienza Università di Roma (Italy)

Abstract   A comprehensive understanding of the factors that affect capture probability in a capture-recapture study plays a fundamental role in determining the accuracy and precision of population parameter estimates. In this paper we provide an updated account of recent advances in the statistical modelling of capture recapture experiments for closed populations which highlight general frameworks to explore and fit alternative behavioural patterns and point out some key issues which are relevant for developing possibly efficient and flexible inferential tools.

Full paper: 3059.pdf

SL13 – COMPUTATIONS WITH INTRACTABLE LIKELIHOOD

Chairperson: E. Mineo (Università di Palermo)

Exact Bayesian inference for discretely observed diffusions

Isadora Antoniano-Villalobos[1], isadora.antoniano@unibocconi.it,
Stephen Walker[2], s.g.walker@math.utexas.edu

(1) Università Bocconi (Italy), (2) University of Texas (U.S.A.)

Abstract   Inference for discretely observed real-valued diffusion processes is commonly based on approximations. The only methods available for exact inference profit from a retrospective rejection sampler for exact simulation of diffusion paths which provides Maximum Likelihood estimates of the model parameters. In the context of Bayesian analysis, however, the current approach resorts to MCMC estimation for which the complexity of the algorithm grows linearly with the size of the data set. We propose a reinterpretation of the exact simulation algorithm, in terms of a set of latent variables which transform the Bayesian parametric diffusion model into a Bayesian non or semi parametric model. We then propose an estimation method, involving trans-dimensional MCMC methods, which allows exact inference for a specific family of diffusions.

Full paper: 3005.pdf

An efficient algorithm to estimate the sparse group structure of an high-dimensional generalized linear model

Luigi Augugliaro[1], luigi.augugliaro@unipa.it, Angelo Mineo[2], angelo.mineo@unipa.it

(1) Università di Palermo (Italy), (2) Università di Palermo (Italy)

Abstract   Massive regression is one of the new frontiers of computational statistics. In this paper we propose a generalization of the group least angle regression method based on the differential geometrical structure of a generalized linear model specified by a fixed and known group structure of the predictors. An efficient algorithm is also proposed to compute the proposed solution curve.

Full paper: 3002.pdf

An adaptive method to robustify ML estimation in Cluster Weighted Modeling

Francesca Greselin[1], francesca.greselin@unimib.it, Luis Garcìa-Escudero[2], lagarcia@eio.uva.es, Alfonso Gordaliza[3], alfonsog@eio.uva.es, Salvatore Ingrassia[4], s.ingrassia@unict.it, Agustìn Mayo-Iscar[5], agustinm@eio.uva.es

(1) Università di Milano-Bicocca (Italy), (2) Universidad de Valladolid (Spain), (3) Universidad de Valladolid (Spain), (4) Università di Catania (Italy), (5) Universidad de Valladolid (Spain)

Abstract   Cluster-Weighted Models are a wide family of mixture distributions for modeling the joint probability of data coming from a heterogeneous population, and includes mixtures of distributions and mixtures of regressions as special cases. Unfortunately, they suffer from non-regular maximum likelihood issues, due to possible spikes and unboundedness in the target function. We propose an improved version of the Gaussian Cluster-Weighted estimation methodology, by trimming a portion α of the data and imposing constraints to the estimated variances. Trimming provides robustness properties to the estimators and constraints move the maximization problem to a well-posed setting and allow to avoid spurious solutions, i.e. fitting a small localized random pattern in the data rather than a proper underlying cluster structure. Theoretical results are illustrated using a few empirical studies.

Full paper: 3004.pdf

Delayed rejection algorithm to estimate Bayesian social networks

Antonietta Mira[1], antonietta.mira@usi.ch, Alberto Caimo[2], alberto.caimo@usi.ch

(1) University of Lugano (Switzerland), (2) University of Lugano (Switzerland)

Abstract   The Delayed rejection (DR) strategy is a modification of the Metropolis-Hastings (MH) algorithms that reduces the variance of the resulting Markov chain Monte Carlo estimators and allows partial adaptation of the proposal distribution. This strategy is exploited to estimate dyadic independence network models leading to an average 40% variance reduction relative to the competing MH algorithm, confirming that DR dominates, in terms of Peskun ordering, the MH algorithm.

Full paper: 3003.pdf

SL14 – GEOGRAPHICAL INFORMATION IN SAMPLING AND ESTIMATION

Chairperson: A. Petrucci (Università di Firenze)

Spatially balanced samples for land use/land cover surveys

Roberto Benedetti[1], benedett@unich.it, Federica Piersimoni[2], piersimo@istat.it, Paolo Postiglione[3], postigli@unich.it

(1) Università di Chieti-Pescara (Italy), (2) ISTAT (Italy), (3) Università di Chieti-Pescara (Italy)

Abstract   The particular characteristics of geographically distributed data should be taken into account in designing a land use/land cover survey. This paper discusses several methods for sampling spatial units that have been recently introduced in literature. The methodological framework followed is of design-based typology. The techniques outlined are: the GRTS, the Cube, the SPCS, and the LPMs. These methods will be verified on data deriving from LUCAS 2012.

Full paper: 2997.pdf

Modeling the location decisions of firms

Chiara Bocci[1], chiara.bocci@irpet.it, Emilia Rocco[2], rocco@disia.unifi.it, Patrizia Lattarulo[3], patrizia.lattarulo@irpet.it

(1) IRPET (Italy), (2) Università di Firenze (Italy), (3) IRPET (Italy)

Abstract   The paper analizes the birth process of small and medium manufacturing firms in Tuscany, an Italian region. In particular it explores, through a spatial microeconometric approach, the possible determinants of the location decisions of the new firms. The geographical distribution of the manufacturing firms born in Tuscany between the 2005 and the 2008 is defined in terms of a inhomogeneous marked point process in the continuous space and we evaluate the effect of space-varyingfactors, both exogenous and endogenous, on the location decisions of new firms by parametrically modeling the intensity of the process. Results show that the choice is influenced on the one hand by the availability of infrastructures and the level of accessibility, and on the other by the presence and characteristics of existing firms.

Full paper: 3136.pdf

Geo-referenced information for agricultural statistics

Elisabetta Carfagna[1], elisabetta.carfagna@unibo.it, Simone Maffei[2], simone.maffei@gmail.com, Andrea Carfagna[3], andreacarfagna@virgilio.it

(1) Università di Bologna (Italy), (2) TERRA NOVA GIS (Italy), (3) Independent consultant (Italy)

Abstract   In this paper, methods for evaluating the impact of interpolation, aggregation and disaggregation of georeferenced information on models for producing agricultural statistics are discussed. The focus is on small area estimators and yield forecasting models.

Full paper: 3069.pdf

The use of geographic information under the area-level approach to small area estimation

Stefano Marchetti[1], stefano.marchetti@ec.unipi.it, Caterina Giusti[2], caterina.giusti@ec.unipi.it, Nicola Salvati[3], salvati@ec.unipi.it

(1) Università di Pisa (Italy), (2) Università di Pisa (Italy), (3) Università di Pisa (Italy)

Abstract   The Fay and Herriot (1979) area-level model, widely used in statistics to obtain reliable small area estimates, assumes that the area-level direct estimates are spatially uncorrelated, an hypothesis that can be unrealistic with agricultural, environmental, economic and epidemiological data. The aim of this paper is to compare, using some simulation studies, the classical Fay-Herriot model with three different extensions of it, that are able to take into account the spatial proximity effects between the small areas: the Semiparametric Fay-Herriot model proposed by Giusti et al (2012), the Spatial Fay-Herriot model proposed by Salvati (2004), Singh et al (2005) and Petrucci and Salvati (2006), and, finally, the Spatial Nonstationary Fay-Herriot model recently proposed by Chandra et al (2014).

Full paper: 3024.pdf

SL15 – CLINICAL DESIGNS

Chairperson: V. Sambucini (Sapienza Università di Roma)

Statistical properties of urn designs in clinical trials

Andrea Ghiglietti[1], andrea.ghiglietti@mail.polimi.it, Anna Maria Paganoni[2], anna.paganoni@polimi.it

(1) Politecnico di Milano (Italy), (2) Politecnico di Milano (Italy)

Abstract   We propose a response-adaptive design, described in terms of urn model, whose allocation proportion converge to prespecified values.The main asymptotic results concerning the urn designs are presented and discussed. We adopt the urn model to implement the random allocation procedure of an experiment that aims at testing the mean effect of two treatments. We conduct a statistical analysis on the inferential performance of different tests and we show that, given a non adaptive test T_0, the response adaptive model construct a test T that is better than T_0, in terms of (a) higher power and (b) fewer subjects assigned to the inferior treatment. A retrospective real case study is presented.

Full paper: 2852.pdf

A predictive look at Bayesian Bandits

Stefania Gubbiotti[1], stefania.gubbiotti@uniroma1.it, Pierpaolo Brutti[2], pierpaolo.brutti@uniroma1.it, Fulvio De Santis[3], fulvio.desantis@uniroma1.it

(1) Sapienza Università di Roma (Italy), (2) Sapienza Università di Roma (Italy), (3) Sapienza Università di Roma (Italy)

Abstract   A multi-armed bandit problem models an agent that simultaneously attempts to acquire new information (exploration) and optimizes the decisions based on existing knowledge (exploitation). In clinical trials, this framework applies to Bayesian multi-armed randomized adaptive designs. The allocation rule of experimental units involves the posterior probability of each treatment being the best. The trade-off between gain in information and selection of the most promising treatment is modulated by a quantity γ, typically prefixed or linearly increasing with accumulating sample size. We propose a predictive criterion for selecting γ that also allows its progressive reassessment based on interim analyses data.

Full paper: 2926.pdf

Optimal two-phase design and incidence estimation in cohort studies

Paola Rebora[1], paola.rebora@unimib.it, Maria Grazia Valsecchi[2], grazia.valsecchi@unimib.it

(1) Università di Milano-Bicocca (Italy), (2) Università di Milano-Bicocca (Italy)

Abstract   Two-phase studies are attractive for their economy and efficiency in research settings where large cohorts are available for investigating the prognostic and predictive role of novel genetic and biological factors. In this type of study, information on novel factors is collected only in a convenient subcohort (phase II) drawn from the cohort (phase I) according to a given (optimal) sampling strategy. In order to estimate event incidence in the subcohort, a Kaplan-Meier method accounting for the design has been derived with a proper variance estimator. The proposed method is applied in the context of a two-phase study on childhood acut lymphoblastic leukaemia.

Full paper: 2848.pdf

SL16 - BAYESIAN INFERENCE FOR HIGH

Chairperson: M. Guindani (The University of Texas, USA)

Fast EM Inference for Bayesian Factor Analysis with Indian Buffet Process Prior

Veronika Ročková[1], vrockova@wharton.upenn.edu, Edward I. George[2], edgeorge@wharton.upenn.edu

(1) University of Pennsylvania (U.S.A.), (2) University of Pennsylvania (U.S.A.)

Abstract   We consider a nonparametric Bayesian approach to parsimonious latent factor modeling of high-dimensional covariance matrices. Central to our approach is the Indian Buffet Process (IBP) prior, which allows the number of factors to be unbounded while remaining finite with probability one. The combinatorial complexity of the IBP makes posterior simulation very challenging. As an alternative, we present a fast EM algorithm for MAP learning in Bayesian factor analysis with spike-and-slab priors, exploiting the stick-breaking representation of IBP.

Full paper: 3048.pdf

A Unified Method for CNV Detection and Association with Gene Expression

Mahlet Tadesse[1], mgt26@georgetown.edu, Alberto Cassese[2], alberto.cassese@rice.edu, Michele Guindani[3], mguindani@mdanderson.org, Francesco Falciani[4], F.Falciani@liverpool.ac.uk, Marina Vannucci[5], marina@rice.edu

(1) Georgetown University (U.S.A.), (2) Rice University (U.S.A.), (3) University of Texas (U.S.A.), (4) University of Liverpool (Great Britain, UK), (5) Rice University (U.S.A.)

Abstract   Copy number variants (CNV) are chromosomal aberrations resulting in DNA segments having an abnormal number of copies. The problem of detecting CNVs has received a lot of attention and several methods have been developed to infer CNVs from high-throughput array-based technologies. There is also a strong interest in identifying associations between CNVs and biological functions. These analyses are commonly done in two stages, by first inferring the CNV calls and using them in the analysis as if they were the true copy numbers. Another commonly used procedure performs the association analysis using the normalized raw intensity measurements. These approaches have several limitations. Here, we propose a hierarchical Bayesian model that handles both the CNV detection and association analysis in a unified manner, by integrating array CGH and gene expression data collected on the same set of subjects. We specify a measurement error model that relates the gene expression levels to the latent copy number states, which in turn are related to the observed surrogate fluorescence intensity measurements via a hidden Markov model. Latent selection indicators that exploit the dependencies between copy number states at adjacent chromosomal locations are incorporated into the model. Model fitting and posterior inference are accomplished via MCMC stochastic search techniques. We demonstrate the performance of the method on simulated data and illustrate its application on a genomic study.

Full paper: 3049.pdf

Bayesian Adaptive Trials

Lorenzo Trippa, ltrippa@jimmy.harvard.edu

Dana-Farber Institute (U.S.A.)

Abstract   Adaptive randomization schemes are designed for obtaining a more desirable assignment of competing treatments to the overall group of patients enrolled in the study compared to balanced designs. I discuss major tools in this area of research, including (i) the development of testing procedures and decision rules for adaptive phase II trials, (ii) the comparison of designs with different levels of complexity to provide recommendations about their relative advantages and disadvantages and (iii) subgroup-based adaptive designs that simultaneously search for patients subgroups and adaptively allocate patients to the best subgroup-specific treatments during the course of the trial.

Full paper: 3057.pdf

SL17 – USE OF BIG DATA FOR THE PRODUCTION OF STATISTICAL INFORMATION

Chairperson: P. Righi (ISTAT)

Does Google index improve the forecast of Italian labour market?

Stefano Falorsi[1], stfalors@istat.it, Fabio Bacchini[2], bacchini@istat.it, Michele D'Alò[3], dalo@istat.it, Andrea Fasulo[4], Fasulo@istat.it, Carmine Pappalardo[5], cpappalardo@istat.it

(1) ISTAT (Italy), (2) ISTAT (Italy), (3) ISTAT (Italy), (4) ISTAT (Italy), (5) ISTAT (Italy)

Abstract It is possible to exploit Google Trend Index (GTI) for different statistical purposes. At national level GTI allows to exploit the time series of query shares to improve the quality of estimates of short-term (monthly or quarterly) socio-economic indicators. In particular GTI may be used for improving the forecasting or nowcasting of such short-term indicators. In this paper the first results of an empirical study aimed to evaluate the use of GTI for the forecasting (one month ahead) of the number of people seeking for job from Labour Force Survey.

Full paper: 3019.pdf

Use of mobile phone data to estimate mobility flows. Measuring urban population and inter-city mobility using big data in an integrated approach

Dino Pedreschi[1], pedre@di.unipi.it, Roberta Vivio[2], vivio@istat.it, Fosca Giannotti[3], fosca.giannotti@isti.cnr.it, Mirco Nanni[4], mirco.nanni@isti.cnr.it, Barbara Furletti[5], barbara.furletti@isti.cnr.it, Giuseppe Garofalo[6], garofalo@istat.it, Lorenzo Gabrielli[7], lorenzo.gabrielli@isti.cnr.it, Letizia Milli[8], letizia.milli@isti.cnr.it,

(1) Università di Pisa (Italy), (2) ISTAT (Italy), (3) CNR (Italy), (4) CNR (Italy), (5) CNR (Italy), (6) ISTAT (Italy), (7) CNR (Italy), (8) Università di Pisa (Italy),

Abstract   The Big Data, originating from the digital breadcrumbs of human activities,sensed as a by-product of the technologies that we use for our daily activities, let us to observe the individual and collective behavior of people at an unprecedented detail. Many dimensions of our social life have big data "proxies", as the mobile calls data for mobility. In this paper we investigate to what extent such "big data", in integration with administrative ones, could be a support in producing reliable and timely estimates of inter-city mobility. The study has been jointly developed by Istat, CNR, University of Pisa in the range of interest of the "Commissione di studio avente il compito di orientare le scelte dell'Istat sul tema dei Big Data". In an ongoing project at ISTAT, called "Persons and Places" – based on an integration of administrative data sources, it has been produced a first release of Origin Destination matrix – at municipality level – assuming that the places of residence and that of work (or study) be the terminal points of usual individual mobility for work or study. The coincidence between the city of residence and that of work (or study) – is considered as a proxy of the absence of intercity mobility for a person (we define him a static resident). The opposite case is considered as a proxy of presence of mobility (the person is a dynamic resident: commuter or embedded). As administrative data do not contain information on frequency of the mobility, the idea is to specify an estimate method, using calling data as support, to define for each municipality the stock of standing residents, embedded city users and daily city users (commuters).

Full paper: 3026.pdf

Big Data, Social Mining, Diversity, and Wellbeing

Dino Pedreschi, pedre@di.unipi.it

Università di Pisa (Italy)

Abstract   Big data are the digital traces that we humans leave behind us continuously, as an aware or unaware side effect of the use of information and communication technologies for all our activities. By means of social mining, we have the chance of discovering knowledge hidden in Big Data, about social phenomena that hard to observe and measure. We discuss here how elusive and important social properties, such as diversity, collective intelligence and wellbeing, can be to some extent quantified and related to each other, thanks to novel analytics made possible by Big Data.

Full paper: 3107.pdf

Web scraping and web mining: new tools for Official Statistics

Marco Scarnò[1], mscarno@caspur.it, Sergio Salamone[2], sesalamo@istat.it, Monica Scannapieco[3], scannapi@istat.it

(1) CINECA (Italy), (2) ISTAT (Italy), (3) ISTAT (Italy)

Abstract   The amount of information that can be accessed through the net constitutes an enormous resource, useful to describe many aspects of the actual society. The possibility to extract this information has many merits; between these a "constant measurement" of a phenomenon, or the replacement of some surveys. However, such process presents new challenges that derive mainly from the nature of the information (huge, sparse and not structured). In this paper we will illustrated a study made jointly by ISTAT and CINECA about the possibility to use web scraping techniques, associated with text and data mining algorithms, in order to substitute traditional instruments of data collection and estimation, and to combine them in an integrated approach. In particular we referred to the ISTAT sampling survey on "ICT in enterprises" and to its section in which it is treated the use of the net by Italian enterprises for various purposes (e-commerce, e-recruitment, advertisement, e-tendering, e-procurement, e-government). The 8,600 websites have been "scraped" and acquired texts processed in order to try the reproduction of the same information collected via the questionnaire. Preliminary results are quite encouraging, stating an almost satisfactory predictive capability of fitted models.

Full paper: 3008.pdf

SL18 – MEASURING THE SMART CITY

Chairperson: N. Mignolli (ISTAT)

Smart City: measuring a multidimensional topic

Roberta De Santis[1], rdesantis@istat.it, Alessandra Fasano[2], fasano@istat.it, Nadia Mignolli[3], mignolli@istat.it, Anna Villa[4], avilla@istat.it

(1) ISTAT (Italy), (2) ISTAT (Italy), (3) ISTAT (Italy), (4) ISTAT (Italy)

Abstract   Transforming a city into a Smart City is a complex and multidimensional process which changes over time since all the involved stakeholders work to achieve more and better results. "To be smart" affects many aspects of a city including economics, government, people, living, mobility, environment, energy and services.This paper aims at critically analysing the main features related to smart cities such as terminological issues, the heterogeneous theoretical background and the methodological limits of the few existing measurement experiences.

Full paper: 3029.doc

Smart Territory Analytics: toward a shared vision

Carlo Giovannella, mifav@roma2.infn.it

Università di Roma "Tor Vergata" (Italy)

Abstract Smart Territory Analytics is a very young domain that by a great number of people is expected to support benchmarking of cities' and territories' smartness. Analytics, however, is first of all, a route toward awareness but the question is: about what? In other words what does smartness mean? Here, it is argued that smartness is the ability to: 1a) adopt a systemic, holistic and co-evolutive thinking, focusing also on person/citizen and environmental relevant dimensions; 1b) promote a merging of bottom-up and top-down strategies; 2) maintain and measure the city/territorial state of flow; 3a) monitor the traces produced by the crowd and transform them in smart data; 3b) communicate the evolution of the territorial ecosystem adopting engaging strategies, to foster awareness and participation. As an example of 3a), text analytics of open answers given to a questionnaire on Smart City is presented and discussed. It emerges that the achievement of awareness is a process that implies a deeper understanding of the territory and of individual perception and, as well, a greater attention to the well being of citizens and of the environment they populate, being the former at the same time the main source and the end-target of the territorial state of flow.

Full paper: 3031.pdf

Staying On The Smart side. Measuring The Smart Communities

Mariagrazia Dotoli[1], dotoli@poliba.it, Vincenzo Patruno[2], vincenzo.patruno@istat.it, Aldo Scarnera[2], scarnera@istat.it, Luigi Ranieri[3], luigi.ranieri@unisalento.it

(1) Politecnico di Bari (Italy), (2) ISTAT (Italy), (3) Università del Salento (Italy)

Abstract Due to the continuous increase of the world population living in cities, it is crucial to identify strategic plans and perform associated actions to make cities smarter, i.e., more operationally efficient, socially friendly, and environmentally sustainable, in a cost effective manner. To achieve these goals, emerging smart cities need to be optimally and intelligently monitored. We propose the development of a framework for classifying performance indicators of a smart city. It is based on two dimensions: the degree of objectivity of observed variables and the level of technological advancement for data collection. The paper shows an application of the presented framework to the case of the Bari municipality (Italy).

Full paper: 3021.doc

Measuring the territory to build up the Smart City

Paolo Testa[1], testa@cittalia.it, Nicolò Marchesini[2], marchesini@cittalia.it

(1) Cittalia - Fondazione Anci Ricerche (Italy), (2) Cittalia - Fondazione Anci Ricerche (Italy)

Abstract The functional and operational structure of Metropolitan City is still a present topic in the debate concerning the Chapter V of Italian Constitution. It is anyway certain that the large cities in Italy are and will be involved in a radical transformation shaping the territory, not just in terms of geography but of governance and policy too. The Smart City paradigm allows to overcome the sector approach of governance and management in favour of a smart and integrated approach. This paper aims to compare the main characteristics of the areas affected by the reform to outline the future Metropolitan City profile. This comparison is made with innovative tools, comparing the core and the fringe of the Metropolitan City, where the latter equals to the actual Province territory.

Full paper: 3037.doc

SL19 – FORECASTING ECONOMIC AND FINANCIAL TIME SERIES

Chairperson: M.M. Barbieri (Università di Roma Tre)

Outliers in Time Series: an Empirical Likelihood Approach

Roberto Baragona[1], r.baragona@lumsa.it, Domenico Cucina[2], domenico.cucina@uniroma1.it

(1) Università Lumsa (Italy), (2) Sapienza Università di Roma (Italy)

Abstract   The empirical likelihood method is known to be a flexible and effective approach for testing hypotheses and constructing confidence regions in a nonparametric setting. This framework is adopted here for dealing with the outlier problem in time series where conventional distributional assumptions may be inappropriate in most cases. The procedure is illustrated by a simulation experiment.

Full paper: 3093.pdf

Long term component dynamic models for realized covariance matrices

Luc Bauwens[1], luc.bauwens@uclouvain.be, Manuela Braione[2], manuela.braione@uclouvain.be, Giuseppe Storti[3], storti@unisa.it

(1) Université Catholique de Louvain (Belgium), (2) Université Catholique de Louvain (Belgium), (3) Università di Salerno (Italy)

Abstract   Dynamic models for realized covariance matrices are proposed, which include a secular component that captures the changing levels of realized variances and correlations. They generalize the realized DCC models of Bauwens et al. (2012), where the long term level is assumed to be constant. The long term component is specified either as a nonparametric function or as a MIDAS term. Estimation can be done in steps for large dimensional matrices.

Full paper: 3092.pdf

Probabilistic Calibration of Predictive Distributions

Roberto Casarin[1], r.casarin@unive.it, Tilmann Gneiting[2], Tilmann.Gneiting@h-its.org, Francesco Ravazzolo[3], Francesco.Ravazzolo@norges-bank.no

(1) Università di Venezia (Italy), (2) Heidelberg Institute for Theoretical Studies and Karlsruhe Institute of Technology (Germany), (3) Norges Bank and BI Norwegian Business School (Norway)

Abstract   This contribution studies the calibration approach recently proposed in [2] for predictive densities.We consider a set of simulated experiments in order to study the effectiveness of the method.

Full paper: 3134.pdf

Testing for nonlinear serial dependence in time series with surrogate data and entropy measures

Simone Giannerini[1], simone.giannerini@unibo.it, Esfandiar Maasoumi[2], esfandiar.maasoumi@emory.edu, Estela Bee Dagum[3], estela.beedagum@unibo.it

(1) Università di Bologna (Italy), (2) Emory University (U.S.A.), (3) Università di Bologna (Italy)

Abstract   In this work we propose a nonparametric test for the identification of nonlinear dependence in time series. The approach is based on a combination of a test statistic based on an entropy dependence metric together with a suitable extension of surrogate data methods, a class of Monte Carlo tests introduced in the field of nonlinear dynamics. We focus on the null hypothesis of linear Gaussian processes and we derive the asymptotic theory for the test statistics. Since the asymptotic approximations depend on unknown quantities and require long series to be feasible we advocate the use of surrogate methods. We prove the asymptotic validity of the inference derived from the test and show the finite sample performance through a small simulation study.

Full paper: 3123.pdf

Contributed Paper Sessions

CP1 - DEMOGRAPHY

Chairperson: S. Bertino (Sapienza Università di Roma)

A measure of the distance between countries based on individual data

Gustavo De Santis[1], gustavo.desantis@unifi.it, Mauro Maltagliati[2], mmaltag@disia.unifi.it, Silvana Salvini[3], salvini@disia.unifi.it

(1) Università di Firenze (Italy), (2) Università di Firenze (Italy), (3) Università di Firenze (Italy)

Abstract   We present a new method for evaluating the relative distance between any two countries, among several, on the basis of individual data. First, clusters of respondents are formed and the proportions of each country's respondents who belong to the various clusters are calculated. If respondents in the same cluster are similar to one another, and if two countries are close to each other when their nationals distribute similarly among clusters, the Euclidean distance between the observed distributions provides a measure of the distance between countries. The method is applied to the 21 European countries of the WVS (World Value Survey) for the years 1994-2007, first by "domain" (opinions and attitudes on, e.g., religion, politics, and family), and then globally.

Full paper: 2880.pdf

Retirement and Intra-Household Labour Division of Italian Couples: A Simultaneous Equation Approach

Antonino Di Pino[1], dipino@unime.it, Maria Gabriella Campolo[2], mgcampolo@unime.it, Marcantonio Caltabiano[3], caltabiano@unime.it

(1) Università di Messina (Italy), (2) Università di Messina (Italy), (3) Università di Messina (Italy)

Abstract   A Difference-in-Differences estimation procedure of market and domestic work of both partners is adopted to study the changes in the allocation of market and domestic work among Italian married or cohabiting adults in their 50s and 60s after the retirement of man. Our results show that men increase their commitment in domestic activities as an effect of the transition to retirement, but the influence of gender ideology significantly hampers this propensity. This result has been obtained providing proper methodological instruments to specify the gender attitudes' influence.

Full paper: 2840.pdf

Demographic change and future sustainability of emergency departments: a pilot study for Liguria

Lucia Leporatti[1], lucia.leporatti@unige.it, Paolo Cremonesi[2], paolo.cremonesi@galliera.it, Enrico Di Bella[3], edibella@economia.unige.it, Luca Persico[4], persico@economia.unige.it

(1) Università di Genova (Italy), (2) E.O. Galliera (Italy), (3) Università di Genova (Italy), (4) Università di Genova (Italy)

Abstract   The progressive ageing of the population, together with increasing migration flows from non western countries are significantly changing the shape of the population pyramid in Italy. In this paper we estimate how these phenomena will impact on the future sustainability of Accident and Emergency Departments (AEDs) in Liguria, providing a forecasting of the total AED expenditure for the next decades (2012-2065). Results show that the increase in the number of foreign residents will be the most relevant determinant of the raise in AED accesses and expenditure, boosting expenditure and increasing the problems of overcrowding, waiting times and inappropriate use of AEDs.

Full paper: 2932.pdf

Error models for weighting estimators in the 15th Italian Population and Household Census

Luigi Marcone[1], lmarcone@istat.it, Francesco Borrelli[2], borrelli@istat.it, Stefania Di Domenico[3], sdidomenico@istat.it

(1) ISTAT (Italy), (2) ISTAT (Italy), (3) ISTAT (Italy)

Abstract   The 15th Italian Population and Household Census adopted an innovative frame in order to comply with the requests of better timeliness in data dissemination. The load of traditional Censuses operations usually increases with the population size of the municipalities. Therefore the Italian National Institute of Statistics (ISTAT) decided to plan a mixed Census, based on both exhaustive and sampling surveyed subsets of variables. The use of municipal population registers (henceforth Lac) enabled the mailing out of short/long questionnaire models to each family according to the sampling design. The estimates of long-form variables and their intersections with exhaustive variables require the use of calibration weighting estimators as well as the analysis of the variability in geographic domains.

Full paper: 2917.docx

Father-child contact after marital dissolution. Evidence from Italy

Silvia Meggiolaro[1], meg@stat.unipd.it, Fausta Ongaro[2], ongaro@stat.unipd.it

(1) Università di Padova (Italy), (2) Università di Padova (Italy)

Abstract   With the diffusion of marital instability also in Italy, the number of children who spend some of their childhood without one of their parent, usually the father, is not negligible even in this country. For these children, examining contact with their father after separation is worthwhile, since a good father-child interaction contributes to their well-being. In this paper we consider the frequency of child-father contact after separation for children aged 6-17 living with their mothers, with data from two rounds (2003 and 2009) of the Italian survey "Family and Social Subjects". Multivariate analyses show a negative effect of mother's repartnering in the frequency of father-child contact in 2009, and an unexpected weak positive effect in 2003. A negative effect of paternal repartnering is observed in 2003 and no significant effect in 2009.

Full paper: 2847.pdf

CP2 - STATISTICS IN FINANCE

Chairperson: E. Mineo (Università di Palermo)

The use of loss functions in assessing the VaR measures

Alessandra Amendola[1], alamendola@unisa.it, Vincenzo Candila[2], vcandila@unisa.it

(1) Università di Salerno (Italy), (2) Università di Salerno (Italy)

Abstract   The identification of the best model in terms of volatility forecast accuracy is a troublesome task and many evaluation methods have been proposed on the basis of a statistical or economic approach. The aim of this work is to investigate the opportunity to use a statistical approach in a VaR framework, i.e. evaluating the VaR measures by means of a loss function. By using high-frequency data it is possible to achieve a consistent estimate of the VaR bootstrapping the intraday increments of an asset. Hence, the performances of the volatility models are compared with that employing the VaR consistent estimate. In particular, the 'true' VaR is used to find a threshold discriminating low from high loss function values for each volatility model. The proposed procedure is assessed by means of a Monte Carlo simulation.

Full paper: 2909.pdf

Systemic risk models

Paola Cerchiello[1], paola.cerchiello@unipv.it, Paolo Giudici[2], giudici@unipv.it

(1) Università di Pavia (Italy), (2) Università della Basilicata (Italy)

Abstract   The latest financial crisis has stressed the need of understanding financial systems, as network of interconnected institutions, where financial linkages play a fundamental role in the spread of systemic risks. The statistical estimation of financial networks is a rather challenging data analysis task, due to the presence of big data, in two main dimensions: a) the number of institutions may be quite large and, therefore, the number of possible networks from which to select the most representative one becomes a huge number; b) the number of data variables on financial institutions is very large, and may come from different sources: for example, balance sheet data, actual and expected; market prices of shares, bonds and options; macroeconomic data; financial analysts and rating agency evaluations; news and media data. In this paper we propose a modelling strategy aimed at estimating financial networks, within a structured statistical framework, that of graphical Gaussian models and suggest ways to reduce data complexity by means of conditional independence arguments. We use financial market share prices as the variables on which to base the estimation of linkages between institutions. We then show how conditioning on a country effect can reduce cross border links between banks, and apply the procedure to Eurozone banks. Second we show, for a country that maintains many linkages between banks, how such linkages can be reduced conditioning on the available idiosyncratic (quarterly) balance sheet information of each bank.

Full paper: 2977.pdf

Labour-use efficiency in the Italian machinery industry: a non-parametric stochastic frontier perspective

Giancarlo Ferrara[1], gferrara@sose.it, Francesco Vidoli[2], fvidoli@sose.it, Arianna Campagna[3], acampagna@sose.it, Jacopo Canello[4], jcanello@sose.it

(1) SOSE S.p.A. (Italy), (2) SOSE S.p.A. (Italy), (3) SOSE S.p.A. (Italy), (4) SOSE S.p.A. (Italy)

Abstract   Firms' efficiency is a mainstream in the study of economic growth. Within this broad research area, the present work analyses the labor-use efficiency in the Italian machinery industry through the application of a non-parametric stochastic frontier model with the aim of suggesting new insights to interpret the recent dynamics of the Italian manufacturing system. An extended panel data of manufacturing Small and Medium Enterprises (SMEs) operating in the mechanical industry for the period 2002-2012 has been extracted from the Italian Ministry of Economy and Finance annual survey (Studi di Settore) and used for the implementation of the proposed method.

Full paper: 2898.doc

Chain Graph for VAR and MARCH parameters reduction. EU index returns case.

Andrea Pierini, andrea.pierini@uniroma3.it

Università di Roma Tre (Italy)

Abstract   In this paper a use of Chain Graph models is applied to a multivariate time series of EU index returns, in order to construct graphs with minimum BIC among a particular class of graphs called decomposable, which have the desirable property of a closed form estimation. Then based on the previous identified relationships among present and past values of the multivariate time series, thanks to the chain graph modelling, a VAR(1) - MArch(1) model is constructed by restricting to zero the parameters which are not indicated by the graphs. In this way a great reduction of parameters is put in place, using the opportune multidimensional modelling only if it is necessary.

Full paper: 2882.pdf

Predicting financial bankruptcy by a (Robust) Principal Component Analysis based model: an empirical investigation.

Marianna Succurro[1], marianna.succurro@unical.it, G. Damiana Costanzo[2], dm.costanzo@unical.it

(1) Università della Calabria (Italy), (2) Università della Calabria (Italy)

Abstract   Starting by a series of financial ratios data we introduce in this paper a new composite index to identify over-indebted firms. First, we build up an over-indebtedness index which takes account of the firm's debt level structure and at same time its sustainability, secondly, we investigate to what extent such over-indebtedness index predicts financial distressed companies. For the construction of such composite index a Robust Principal Component Analysis for skewed data is used. The results obtained in the empirical analysis are then compared with those of the popular Altman Z-Score which is instead based on Discriminant Analysis.

Full paper: 2935.docx

CP3 - STATISTICS IN MEDICINE

Chairperson: A. Decarli (Università Milano – Bicocca)

Bayesian nonparametric spatial modelling of ordinal periodontal data

Antonio Canale[1], antonio.canale@unito.it, Dipankar Bandyopadhyay[2], dbandyop@umn.edu

(1) Università di Torino and Collegio Carlo Alberto (Italy), (2) University of Minnesota (U.S.A.)

Abstract   Clinical attachment loss (CAL) is a measure often used to assess periodontal disease (PD) status at a tooth site. While being ideally continuous, CAL measures are usually rounded and recorded as ordered categorical data. In addition, these CAL measures are hypothesized to be spatially-referenced. Traditional analysis model this integer-valued CAL via a linear mixed model with appropriate spatial random effects. In this paper, we propose a flexible nonparametric Bayesian approach to model the ordinal categories using an ordered probit model, simultaneously accounting for the within mouth spatial-referencing, yet preserving computational simplicity. An application to a real dataset on PD is presented.

Full paper: 2845.pdf

Elicitation and visualisation of uncertainty in electrograms for activation time maps

Silvia Liverani[1], s.liverani@imperial.ac.uk, Chris Cantwell[2], c.cantwell@imperial.ac.uk

(1) Imperial College London (Great Britain, UK), (2) Imperial College London (Great Britain, UK)

Abstract   When treating atrial arrhythmias using catheter ablation it is standard practice to create local activation maps of a patient's atrium to guide treatment. Generating these maps involves recording hundreds of electrograms from the inside surface of the chamber and then using automated algorithms to identify deflections in the signal and choose the activation time for each recording point. The activation times are then visually represented on a 3D model of the chamber which is used to guide treatment. However, uncertainty in the selection of the activation time is not considered by the algorithm and may lead to poor quality data collection. To the best of our knowledge, this is the first attempt to quantify and communicate the uncertainty of each activation time.

Full paper: 2875.pdf

Log-mean linear regression models for assessing the effect of HIV-infection on multimorbidity in a case-control study

Monia Lupparelli[1], monia.lupparelli@unibo.it, Alberto Roverato[2], alberto.roverato@unibo.it

(1) Università di Bologna (Italy), (2) Università di Bologna (Italy)

Abstract   We propose a multivariate regression approach for binary variables basedon the log-mean linear link function for the response variables. This approach has shown to provide useful insights for assessing the effect of HIV-infection on multimorbidity, defined as the occurrence of co-existing noninfectious diseases for a sample of patients deriving from a case-control study. The coefficients of these regression models are log-linear combinations of relative risks and we show that submodels identified by zero regression coefficients encode relevant hypotheses for the considered application.

Full paper: 2951.pdf

A Quantile-based Test for Detecting Differential Expression in Microarray Data

Lorenzo Maragoni[1], lorenzo.maragoni@gmail.com, Monica Chiogna[2], monica.chiogna@unipd.it

(1) Università di Padova (Italy), (2) Università di Padova (Italy)

Abstract   In the last decade the development of statistical methods for microarray data analysis has been object of active research. A crucial problem in this area is the search for genes that show a differential expression between two or more groups like, for example, healthy and sick tissues. A huge number of statistical tools has been proposed for this task, most of which focus on testing equality of means. The aim of the present work is to introduce a simple quantile-based statistic that can test differences at different levels of the distribution, and has the desirable property of invariance with respect to monotone data transformations.

Full paper: 2964.pdf

EEG signals decomposition: a multi-resolution analysis

Paolo Zanini[1], paolo.zanini@polimi.it, Piercesare Secchi[2], piercesare.secchi@polimi.it, Simone Vantini[3], simone.vantini@polimi.it

(1) Politecnico di Milano (Italy), (2) Politecnico di Milano (Italy), (3) Politecnico di Milano (Italy)

Abstract   In this work we analyze a Blind Source Separation (BSS) problem for the EEG signals of patients affected by alcoholism. We solve the BSS problem through an innovative algorithm, named Hierarchical Independent Component Analysis (HICA), able to provide a multi-resolution non-orthogonal data-driven basis. Moreover we show the improvements obtained by HICA in terms of the phenomenological interpretation of the components with respect to other popular BSS techniques.

Full paper: 2843.pdf

CP4 – CLUSTERING METHODS: THEORY AND APPLICATIONS

Chairperson: G. Pistone (Università di Torino)

Cluster analysis on shape indices for ChIP-Seq data

Marzia Cremona[1], marziaangela.cremona@polimi.it, Pier Giuseppe Pelicci[2], piergiuseppe.pelicci@ieo.eu, Laura Riva[3], laura.riva@iit.it, Laura Sangalli[4], laura.sangalli@polimi.it, Piercesare Secchi[5], piercesare.secchi@polimi.it, Simone Vantini[6], simone.vantini@polimi.it

(1) Politecnico di Milano (Italy), (2) European Institute of Oncology (Italy), (3) Istituto Italiano di Tecnologia (Italy), (4) Politecnico di Milano (Italy), (5) Politecnico di Milano (Italy), (6) Politecnico di Milano (Italy)

Abstract   In recent years many techniques have been developed to study genetic and epigenetic processes. Here we focus on a particular Next Generation Sequencing method called ChIP-Seq (Chromatin ImmunoPrecipitation Sequencing), that permits to investigate protein-DNA interactions. At present, in the relevant literature, the analysis of ChIP-Seq data is mainly restricted to the detection of enriched regions (peaks) in the genome, considering only signal intensity. The innovative approach that we propose takes into consideration also the shape of such peaks. We introduce some indices to summarize the shape and we use multivariate clustering techniques in order to detect statistically significant differences in peak shape, with the idea that it can be associated with a functional role and a biological meaning.

Full paper: 2904.pdf

fclust: an R package for fuzzy clustering

Paolo Giordani[1], paolo.giordani@uniroma1.it, Maria Brigida Ferraro[2], mariabrigida.ferraro@uniroma1.it

(1) Sapienza Università di Roma (Italy), (2) Sapienza Università di Roma (Italy)

Abstract   In this work the R package fclust is introduced. It contains functions to perform fuzzy clustering analysis. The well-known Fuzzy k-Means algorithm, along with some variants, is available in the package. The package also contains some functions for computing fuzzy cluster validity indices.

Full paper: 2832.pdf

Modal Clustering of Social Networks

Giovanna Menardi[1], menardi@stat.unipd.it, Domenico De Stefano[2], domenico.destefano@unive.it

(1) Università di Padova (Italy), (2) Università di Venezia (Italy)

Abstract   The idea underlying modal clustering is to associate groups with the regions around the modes of the probability density function underlying the data. This correspondence between clusters and dense regions in the sample space is here exploited to discuss a possible extension of modal clustering to the analysis of social networks. Such extension, albeit non trivial, seems particularly appealing: conceptually, the notion of high-density cluster fits well the one of cluster in a network, where groups are usually regarded as collections of individuals with dense local ties in their neighborhood. Additionally, modal clustering often resorts to graph theory for the operational detection of clusters, which is another condition that seems particularly appropriate to deal with relational data.

Full paper: 2940.pdf

Identification of multiple clusterings using Gaussian mixtures

Andrea Pastore[1], pastore@unive.it, Stefano Tonellato[2], s.tonellato@unive.it

(1) Università di Venezia (Italy), (2) Università di Venezia (Italy)

Abstract In this paper we present a generalisation of the algorithm by Raftery and Dean, which is tailored for the selection of variables in Gaussian mixture models. Our proposal allows for the detection of multiple independent clusterings. A simulation study is provided.

Full paper: 2914.pdf

CP5 - FUNCTIONAL DATA ANALYSIS

Chairperson: L. Ippoliti (Università di Chieti-Pescara)

A clustering method for functional data

Enea Bongiorno[1], enea.bongiorno@gmail.com, Aldo Goia[2], aldo.goia@unipmn.it

(1) Università del Piemonte Orientale (Italy), (2) Università del Piemonte Orientale (Italy)

Abstract This work proposes an unsupervised classification algorithm for curves. It extends the density based multivariate cluster approach to the functional framework. In particular, the modes of the small-ball probability are used as starting points to build the clusters. A simulation study is proposed.

Full paper: 2963.pdf

A class of bibliometric indices based on a sum of increasing and concave functions

Tommaso Lando[1], tommaso.lando@unicatt.it, Lucio Bertoli-Barsotti[2], lucio.bertoli-barsotti@unibg.it

(1) Technical University Ostrava (Czech Republic), (2) Università di Bergamo (Italy)

Abstract The h-index is used by major citation database to evaluate the academic performance of a researcher. We propose a new class of bibliometric indicators to measure the quality of the publications within the h-core (the group of the h most highly cited papers): the aim is to reduce the effect of extremely cited publications (which affects some well-known alternatives to the h-index such as the E- and A-indices) basing on a sum of increasing and concave functions.

Full paper: 2902.docx

Anterior Cruciate Ligament Rupture: Functional Data Analysis of Knee Motion

Alessia Pini[1], alessia.pini@polimi.it, Konrad Abramowicz[2], konrad.abramowicz@math.umu.se, Charlotte Häger[3], charlotte.hager@umu.se, Kim Hébert-Losier[4], khebertlosier@huskers.com, Lina Schelin[5], lina.schelin@math.umu.se, Johan Strandberg[6], johan.strandberg@math.umu.se, Simone Vantini[7], simone.vantini@polimi.it

(1) Politecnico di Milano (Italy), (2) Umeå University (Sweden), (3) Umeå University (Sweden), (4) Mid Sweden University (Sweden), (5) Umeå University (Sweden), (6) Umeå University (Sweden), (7) Politecnico di Milano (Italy)

Abstract   Human movement data are traditionally summarized into discrete parameters prior to analysis, only considering certain points in time. Nevertheless, the quantities of interest are smooth functions of time, making it possible to analyse motion data using Functional Data Analysis techniques. We present the analysis of a functional data set, that involved comparing the knee movement patterns during a one-leg hop in individuals suffering from an Anterior Cruciate Ligament injury treated with surgery and physiotherapy (ACL_R), with physiotherapy alone (ACL_{PT}), as well as uninjured controls (ACL_C). We compare our results with the ones of a classical technique focusing on the maximum knee flexion data only, and present the advantages of considering a functional approach.

Full paper: 2842.pdf

Bayesian inference for dynamically evolving distributions

Matteo Ruggiero[1], matteo.ruggiero@unito.it, Omiros Papaspiliopoulos[2], omiros.papaspiliopoulos@upf.edu, Dario Spanò[3], D.Spano@warwick.ac.uk

(1) Università di Torino and Collegio Carlo Alberto (Italy), (2) ICREA and Universitat Pompeu Fabra (Spain), (3) University of Warwick (Great Britain, UK)

Abstract   Filtering for hidden Markov models amounts to sequential Bayesian inference on some dynamic parameter given observations collected at discrete time points. The filtering is exact if the sequence of posterior distributions are evaluated in closed form rather than approximated. Relying on the notion of duality for Markov processes, we provide general sufficient conditions for constructing a filter which sequentially evaluates the posterior distribution of finite-dimensional parameters on uncountable state spaces. We then exploit the application of such results to some specific models for extending the filtering strategies to some important classes of measure-valued diffusions, which describe dynamically evolving nonparametric distributions.

Full paper: 2820.pdf

Depth measures for multivariate functional data with data-driven weights

Nicholas Tarabelloni[1], nicholas.tarabelloni@gmail.com, Rachele Biasi[2], rachele.biasi@mail.polimi.it, Francesca Ieva[3], francesca.ieva@unimi.it, Anna Paganoni[4], anna.paganoni@polimi.it

(1) Politecnico di Milano (Italy), (2) Politecnico di Milano (Italy), (3) Università di Milano (Italy), (4) Politecnico di Milano (Italy)

Abstract   The notion of statistical depth have recently been extended to the case of multivariate functional data. Its definition involves the choice of proper weights, averaging the univariate functional depths of each component. The choice of weights is crucial and must be carefully done according to the problem at hand. We describe a procedure that, starting from data, allows to compute a set of weights which are suitable for classification based on depths. These weights incorporate information on distances between covariance operators of the sub-populations. We show the validity of our strategy through a case study in which we perform supervised classification on ECG traces referring to both physiological and pathological subjects.

Full paper: 2834.pdf

CP6 - FORENSIC STATISTICS

Chairperson: P. Conti (Sapienza Università di Roma)

Statistical comparison of European judicial systems according to ICT

Carlo Cusatelli[1], carlo.cusatelli@uniba.it, Massimiliano Giacalone[2], massimilia.giacalone@unibo.it

(1) Università di Bari (Italy), (2) Università di Bologna (Italy)

Abstract   The need for strong regulatory and valid instruments is clear in contemporary society to resolve disputes in a timely manner and penalize wrongdoing. The judicial function must be measured against this complexity, in order to ensure the safety of citizens, social cohesion, and economic competitiveness at an international level. The Information and Communication Technology (ICT), that facilitates knowledge and exchange of data and information through an analytical approach to problem solving, is of paramount help in tackling this complexity.

Full paper: 2968.doc

Efficiency of justice and justice's demand

Antonia Rosa Gurrieri[1], antoniarosa.gurrieri@unifg.it, Marilene Lorizio[2], marilene.lorizio@unifg.it

(1) Università di Foggia (Italy), (2) Università di Foggia (Italy)

Abstract   A high level of quality and efficiency of an economic system is very relevant for the growth of a Country. This mainstream is particularly true in relation with the justice. The aim of this work is to estimate the extent to which the identified variables affecting the length of civil proceedings, particularly with regard to the role of lawyers.

Full paper: 2954.doc

Measures of Courts Performances and Stochastic Frontier Models

Laura Antonucci[1], laura.antonucci@unifg.it, Francesco D'Ovidio[2], francescodomenico.dovidio@uniba.it, Domenico Viola[3], domenico.viola@uniba.it

(1) Università di Foggia (Italy), (2) Università di Bari (Italy), (3) Università di Bari (Italy)

Abstract   Italy is frequently reprimanded by the European Court of Human Rights (ECtHR) over the amount of time it takes Italian courts to reach verdicts. As stated by President Giorgio Napolitano, European Court decisions made urgent an intervention in order to save time and costs in Italian judicial system. Efficiency and effectiveness are key goals for managing justice in Italy, but are not easy to achieve. In this paper, using a Stochastic Frontier Model (SFM) we compare the Italian courts efficiency to identify strong and weak points.

Full paper: 2974.doc

The efficiency of Justice in an economic perspective: the role of the supply side

Annamaria Stramaglia[1], annamaria.stramaglia@unifg.it, Marilene Lorizio[2], marilene.lorizio@unifg.it

(1) Università di Foggia (Italy), (2) Università di Foggia (Italy)

Abstract   The work contributes to the multi-disciplinary debate on the efficiency of the justice service in an economic perspective. The aim is to analyze some of the several factors which characterize the system on the supply side. More specifically, we take into account the economic theories which can explain the contribution of: (i) the economic actors at an individual level; (ii) the factors which determine the supply level at system level.

Full paper: 2899.doc

CP7 - ECONOMIC PHENOMENA

Chairperson: P. Veronese (Università Bocconi)

Decomposition by sources of the ξ inequality index

Alberto Arcagni[1], alberto.arcagni@unimib.it, Michele Zenga[2], michele.zenga@unimib.it

(1) Università di Milano-Bicocca (Italy), (2) Università di Milano-Bicocca (Italy)

Abstract   In this work we present the decomposition by sources of the inequality index ξ proposed by Zenga in 1984. The decomposition approach is inspired by the general one presented by Zenga et al. in 2012 and by Zenga in 2013 that allows to compare different point and synthetic inequality measures, such as the Gini and the Bonferroni indexes, and the Zenga index I.

Full paper: 2853.pdf

LFS quarterly small area estimation of youth unemployment at provincial level

Stefano Falorsi[1], stfalors@istat.it, Michele D'Alò[2], dalo@istat.it, Silvia Loriga[3], siloriga@istat.it

(1) ISTAT (Italy), (2) ISTAT (Italy), (3) ISTAT (Italy)

Abstract
Small area estimation methods are applied to compute quarterly youth unemployment on Italian provinces. Results are compared with calibration estimates and with Census results. The small area methods perform well in terms of efficiency and respect to the Census data, especially the estimator exploiting the spatial and temporal correlation. Nevertheless further evaluation are being conducted to ensure the coherence with the regular official estimates produced by Istat at higher level of aggregation.

Full paper: 2945.pdf

The decomposition by subgroups of the inequality curve Z(p) and the inequality index ξ

Francesco Porro[1], francesco.porro1@unimib.it, Michele Mario Zenga[2], michele.zenga@unimib.it

(1) Università di Milano-Bicocca (Italy), (2) Università di Milano-Bicocca (Italy)

Abstract
This paper provides a procedure to decompose by subgroups the inequality index ξ, proposed by Zenga. Through this decomposition, such index can be seen as the weighted average of two terms, which represent the "within" and the "between" components. This decomposition arises from the decomposition by subgroups of the inequality curve $Z(p)$, that originates the index itself.

Full paper: 2972.pdf

Inequality measures and the issue of negative incomes

Emanuela Raffinetti[1], emanuela.raffinetti@unimi.it, Elena Siletti[2], elena.siletti@unimi.it, Achille Vernizzi[3], achille.vernizzi@unimi.it

(1) Università di Milano (Italy), (2) Università di Milano (Italy), (3) Università di Milano (Italy)

Abstract
Income distribution studies have a long history in economic and statistical literature. Many results in such research area are provided by the standard inequality measure Gini coefficient, traditionally defined for non-negative incomes. In this paper the issue of negative incomes is faced and a specific reformulation of the Gini coefficient is introduced. More precisely, a new Gini coefficient normalization, held by the Pigou-Dalton transfers principle fulfillment, is presented.

Full paper: 2894.pdf

The Effects of a new Aggregation Structure on Consumer Price Index Estimates

Daniele Toninelli[1], daniele.toninelli@unibg.it, Martin Beaulieu[2], MartinJ.Beaulieu@statcan.gc.ca, Catalin Dochitoiu[3], Catalin.Dochitoiu@statcan.gc.ca

(1) Università di Bergamo (Italy), (2) Statistics Canada (Canada), (3) Statistics Canada (Canada)

Abstract
Recent studies were developed in order to improve the representativeness of price indices. This paper shows the effect on the Consumer Price Index (CPI) estimates of an alternative aggregation structure. This new proposed methodology includes a stratification of the sampled units (stores) by type. Starting from an elementary aggregate of collected prices, market shares are used to reflect the contribution of each stratum. The Jevons index is used at a lower level of aggregation; then the market shares are used in the following levels to compile the final index. The behaviour of the estimates is also evaluated in the mid-to-long term through the computation of chained indices. The results underlined a potential improvement in the quality of the estimates, in comparison to the traditional method.

Full paper: 2892.pdf

CP8 - ADVANCES IN STATISTICAL MODELLING

Chairperson: P. Muliere (Università Bocconi)

Small-sample likelihood asymptotics for the equi-correlated bivariate normal model

Alessandra Brazzale[1], brazzale@stat.unipd.it, Valentina Mameli[2], vmameli@unica.it

(1) Università di Padova (Italy), (2) Università di Cagliari (Italy)

Abstract   We discuss two likelihood-based small-sample confidence intervals for the skewness parameter of the distribution of the maximum (or minimum) of the equi-correlated bivariate normal model. These are compared numerically to their large-sample counterpart, and to an approximate confidence interval whose construction derives from theoretical findings on the intraclass correlation coefficient and Fisher's transformation. The performance of the confidence intervals is analyzed in terms of actual coverage, symmetry of errors, and expected length. A simulation revealed that the considered small-sample procedures perform well even when the sample size is limited. A real-life application to a mono-zygotic twin study is also given.

Full paper: 2867.pdf

On Non-central Beta distributions

Carlo Orsi[1], c.orsi@campus.unimib.it, Andrea Ongaro[2], andrea.ongaro@unimib.it

(1) Università di Milano-Bicocca (Italy), (2) Università di Milano-Bicocca (Italy)

Abstract   A new Non-central Beta distribution is defined. Several properties are derived (including various representations and moments expressions) both for the new and the standard Non-central Beta distribution, showing in many respects a greater tractability of the former.

Full paper: 2933.pdf

Nonparametric Mode Hunting

Marco Perone Pacifico, marco.peronepacifico@uniroma1.it

Sapienza Università di Roma (Italy)

Abstract   We propose a procedure for detecting the modes of a density estimate and test their significance. We use a data-splitting approach: potential modes are identified using the first half of the data and their significance is tested with the second half of the data. The mode test is based on nonparametric confidence intervals for the eigenvalues of the Hessian. In order to get valid bootstrap confidence sets even in presence of multiplicity of the eigenvalues, we use a bootstrap based on an elementary-symmetric-polynomial transformation.

Full paper: 2871.pdf

Discriminant analysis of von Mises - Fisher distributions

Mario Romanazzi, romanaz@unive.it

Università di Venezia (Italy)

Abstract   The geometry of von Mises - Fisher distributions offers useful clues to discriminant analysis. The discriminant functions turn out to depend on the cosine of a non random vector with a von Mises - Fisher random vector and their probability density functions are explicitly derived. This considerably improves our understanding of discriminant analysis with hyper - spherical data in arbitrary dimension, especially in the two - group case. Specific applications include misclassification probabilities and ROC curve.

Full paper: 2895.pdf

CP9 - DEVELOPMENTS IN BAYESIAN INFERENCE

Chairperson: L. Piccinato (Sapienza Università di Roma)

Bayesian Inference for the Intrinsic Dimension

Pierpaolo Brutti, pierpaolo.brutti@uniroma1.it

Sapienza Università di Roma (Italy)

Abstract   In this work we propose a new Bayesian method for inference on the intrinsic dimension of point–cloud data sampled from a low–dimensional structure embedded in a high–dimensional ambient space. The basic ingredient of our Bayesian recipe is a composite marginal likelihood built under independence assumptions, that was suggested by MacKay and Ghahramani (2005) to improve on an earlier proposal based on local Poisson process approximations (Levina and Bickel, 2005). In order to get a posterior with approximately correct asymptotic behavior and curvature, we calibrate this pseudolikelihood as in Pauli et al. (2011). In simulated and real examples we compare a standard MCMC method against a variation of the default Bayesian technique described in Ventura et al. (2013).

Full paper: 2922.pdf

Zellner-Siow Priors for variable selection with censored data

Maria Eugenia Castellanos[1], maria.castellanos@urjc.es, Stefano Cabras[2], s.cabras@unica.it

(1) Universidad Rey Juan Carlos (Spain), (2) Università di Cagliari and Universidad Carlos III de Madrid (Italy)

Abstract   In this paper we analyze the problem of variable selection in regression analysis when the response variable, Y, follows a lognormal distribution and observations are right censored. The use of conventional priors, namely g-priors and Zellner-Siow priors are analyzed in the context of censored data. Of particular interest is the understanding of the covariance matrix that must be used for specifying such priors in the presence of censored data. A comparison among different options is made through a simulation study.

Full paper: 2886.pdf

Likelihood and Bayesian estimation of P(Y < X) using lower record values from a general class of distribution

Francesca Condino[1], francesca.condino@unical.it, Filippo Domma[2], f.domma@unical.it, Giovanni Latorre[3], g.latorre@unical.it

(1) Università della Calabria (Italy), (2) Università della Calabria (Italy), (3) Università della Calabria (Italy)

Abstract   In this paper, we will study different inferential procedures for the stress-strength reliability based on lower record data, where the stress and the strength variables are modeled by two independent but not identically distributed random variables from a general class of distributions. Likelihood and Bayesian estimators are derived and the corresponding confidence intervals and credible sets are obtained. Finally, some numerical results are reported in order to show the performance of the proposed procedures.

Full paper: 2887.pdf

An objective Bayesian procedure for meta–analysis of binomial data

Elias Moreno[1], emoreno@ugr.es, Francisco Vázquez-Polo[2], fjvpolo@dmc.ulpgc.es, Miguel Negrín[3], mnegrin@dmc.ulpgc.es

(1) University of Granada (Spain), (2) Univesidad de Las Palmas (Spain), (3) Univesidad de Las Palmas (Spain)

Abstract   This paper presents a Bayesian model for meta–analysis of sparse discretebinomial data which are out of the scope of the usual hierarchical normal random effect models. To assure coherency between the marginal and conditional prior distributions utilized in the analysis, the crucial linking distribution between the experimental effectiveness and the unconditional effectiveness (meta–effectiveness) is constructed from specific bivariate classes of distributions with given marginals. Applications to real multicenter data are given and compared with previous meta–analysis.

Full paper: 2862.pdf

Uncertainty quantification for Bayesian nonparametric estimators of rare species variety

Bernardo Nipoti[1], bernardo.nipoti@carloalberto.org, Stefano Favaro[2], stefavano.favaro@unito.it

(1) Università di Torino (Italy), (2) Università di Torino (Italy)

Abstract   Species sampling problems have regained popularity in recent years due to their frequent appearance in challenging applications arising from ecology, genetics, linguistic, etc. Interest often lies in estimating the number of rare species that appear in a sample, that is species with a frequency smaller than a specific abundance threshold. The Bayesian nonparametric approach has proved successful by providing closed form estimators for rare species variety. In this paper we present a novel methodology for endowing such estimators with asymptotic credible intervals. We illustrate it through the analysis of some genomic datasets.

Full paper: 2957.pdf

CP10 - EDUCATIONAL STATISTICS

Chairperson: S. Zaccarin (Università di Trieste)

A finite mixture model approach on the first year university drop-out probability

Matilde Bini[1], mbini.unier@gmail.com, Lucio Masserini[2], l.masserini@adm.unipi.it

(1) Università Europea di Roma (Italy), (2) Università di Pisa (Italy)

Abstract   University drop-out is one of the most important problems occurring in degree courses. The aim of this study is to analyze the first year university drop-out at the University of Pisa (Italy). In particular, we are focused on identifying the covariates affecting the response variable (if the student dropped out or not), and on detecting unobserved subgroups of students, if they exist, having different probabilities of dropping out. To perform this analysis we propose the use of a finite mixture logit model that allow us to consider a methodological framework where the population is made up by an unknown but finite number of subpopulations (latent classes). A dataset formed by administrative data of the University of Pisa, collected at enrollment time for the academic year 2009-2010, is used for this purpose. The analysis is limited to students of the first cycle degree courses. The characteristics detected of subgroups and the influential covariates, should represent useful information for the implementation of academic policy changes that could affect the drop-out rate.

Full paper: 2900.doc

A Bayesian nonparametric modelling to estimate students' response to ICT investment

Stefano Cabras[1], s.cabras@unica.it, Juan De Dios Tena Horrillo[2], juande@uniss.it

(1) Università di Cagliari and Universidad Carlos III de Madrid (Italy), (2) Università di Sassari and Universidad Carlos III de Madrid (Spain)

Abstract This paper estimates the causal impact of Information Communication Technology (ICT) investment on student performances as measured in the Program for International Student Assessment (PISA) 2012 for Spain. We employ a new methodology in this context known as Bayesian Additive Regression (BART) that has important advantages over more standard parametric specifications, namely, (1) it is far simpler and more flexible than many recent competitors as it does not require ad-hoc assumptions about model fitting, number of covariates and how they interact, (2) it easily deals with missing data, and (3) in order to treat endogeneity it does not require the estimation of two equations one for ICT investment and one for the score. Our preliminary results suggest that, contrary to most previous studies, ICT have a moderate effect on math scores.

Full paper: 2839.pdf

On the development of school achievement inequalities with cross-sectional data

Dalit Contini[1], dalit.contini@unito.it, Elisa Grand[2], elisa.grand@unito.it

(1) Università di Torino (Italy), (2) Università di Torino (Italy)

Abstract Despite the spread of standardized assessments of student learning, longitudinal data are still lacking in Italy. This article raises the following question: can we exploit repeated cross-sectional assessments held at different schooling stages to evaluate how achievement inequalities related to individual ascribed characteristics develop over time? Consistently with a simple learning accumulation model, we propose an imputed regression strategy delivering consistent estimates of the parameters of interest. We then apply the model to Italian achievement data and investigate how inequalities develop between primary and secondary school.

Full paper: 2967.docx

The role of the membership function to model university students' flow

Franca Crippa[1], franca.crippa@unimib.it, Marcella Mazzoleni[2], m.mazzoleni8@campus.unimib.it, Mariangela Zenga[3], mariangela.zenga@unimib.it

(1) Università di Milano-Bicocca (Italy), (2) Università di Milano-Bicocca (Italy), (3) Università di Milano-Bicocca (Italy)

Abstract Students' pathways have long been the focus of educational research. Recent efforts have been directed to model levels of a progression embodied in a programme or in a theory, providing evidence about an individual level on that progression(Choi, 2012). In order to meet different schooling systems, the recognition of educational pathways as Markov processes (Trivellato, 1980, Ghellini et al., 2009) has unfolded into extensions such as fuzzy Markov chains (Symeonaki and Kalamatianou, 2011; Crippa, Mazzoleni and Zenga, 2013). The latter models, together with fuzzy set theory contributions to learning evaluation (Hameed and Sorensen, 2010), have enhanced the purpose of the membership function, whose role is investigated in this paper with respect to the issue of Italian university students' retention.

Full paper: 2955.pdf

Multilevel modeling of heterogeneity in math achievements: different class- and school-effects across Italian regions

Francesca Ieva[1], francesca.ieva@unimi.it, Tommaso Agasisti[2], tommaso.agasisti@polimi.it, Anna Paganoni[3], ann.paganoni@polimi.it

(1) Università di Milano (Italy), (2) Politecnico di Milano (Italy), (3) Politecnico di Milano (Italy)

Abstract   Catching the differences in educational attainments between groups of students and across schools is becoming increasingly interesting. With the aim of assessing the extent of these differences in the context of Italian educational system, the paper applies multilevel modeling to a dataset containing detailed information of students' math attainments at grade 6 of primary school in the year 2011/12, provided by the Italian Institute for the Evaluation of Educational System.

Full paper: 2833.pdf

CP11 – SANITARY STATISTICS AND EPIDEMIOLOGY

Chairperson: F. De Santis (Sapienza Università di Roma)

The 2008 Great Recession and Health in Italy. A Study on the Surveillance Data System PASSI

Domenico De Stefano[1], domenico.destefano@unive.it, Stefano Campostrini[2], stefano.campostrini@unive.it

(1) Università di Venezia (Italy), (2) Università di Venezia (Italy)

Abstract   In the present work we study possible local/regional effects of the 2008Great Recession utilizing data from the national Italian surveillance system PASSI (Progressi delle Aziende Sanitarie per la Salute in Italia). PASSI, adopting the behavioral risk factor surveillance (BRFS) model to the Italian context, tracks health conditions and risk behaviors across Italian regions. Here we are interested in time and context effects of some BRFS indicators. We specify a multilevel logistic model following the general form of the generalized linear mixed model (GLMM) in order to disentangle the differences of both health perception and risk behavior considering three sources of variation: individual, local and regional context effects. Results show a noticeable context effect on the main BRFS indicators.

Full paper: 2962.pdf

A probabilistic approach to the estimation of principal components with nonignorable missing data: Applications in accelerometer-based physical activity studies

Marco Geraci[1], m.geraci@ucl.ac.uk, Alessio Farcomeni[2], alessio.farcomeni@uniroma1.it

(1) University College London (Great Britain, UK), (2) Sapienza Università di Roma (Italy)

Abstract   Our paper is motivated by an accelerometer-based study of physical activity (PA) behaviours in a large cohort of UK school-aged children. Given the multiplicity of dimensions associated with PA, a comprehensive individual PA profile can include more than forty different outcomes which clearly poses challenges for the analysis. We explore the application of principal component analysis (PCA) to accelerometer measurements which are aggregated daily over several days of the week and are affected by nonresponse. The probabilistic approach to PCA with latent scores (Tipping and Bishop, 1999, JRSS-B) is extended to include nonignorable missing data. The extended likelihood is maximized through a Monte Carlo EM algorithm via adaptive rejection Metropolis sampling.

Full paper: 2881.pdf

A mortality model based on a mixture distribution function

Stefano Mazzuco[1], mazzuco@stat.unipd.it, Bruno Scarpa[2], scarpa@stat.unipd.it, Lucia Zanotto[3], zanotto@stat.unipd.it

(1) Università di Padova (Italy), (2) Università di Padova (Italy), (3) Università di Padova (Italy)

Abstract   A new mortality model based on a mixture distribution function is proposed. We mix a half-normal distribution with a generalization of the skew-normal distribution. As a result we get a six-paramters distribution function having a good fit with a wide variety of mortality patterns. This model is fitted to several mortality data and compared with the Siler model (five parameters). The main feature of the proposed model is that it has a relatively good fit even in mortality pattern with a high accident hump (this happens, for instance, in countries with high HIV incidence), while the Siler model cannot catch this hump.

Full paper: 2897.pdf

Joint Prediction of Demand and Care Duration in Home Care Patients: a Bayesian Approach

Inad Nawajah[1], inad.nawajah@mail.polimi.it, Raffaele Argiento[2], raffaele@mi.imati.cnr.it, Alessandra Guglielmi[3], alessandra.guglielmi@polimi.it, Ettore Lanzarone[4], ettore@mi.imati.cnr.it

(1) Politecnico di Milano (Italy), (2) CNR (Italy), (3) Politecnico di Milano (Italy), (4) CNR (Italy)

Abstract   Home Care (HC) service is an alternative to conventional hospitalizationand consists of delivering medical, paramedical and social services to patients at their homes rather than in hospitals or nursing homes. Human resource planning in HC is a difficult task and, for a good quality of planning, knowledge of future patients' demands is required. The aim of this paper is to propose a Bayesian model for predicting the number of visits required by HC patients, which is fundamental for planning human and material resources, and at the same time describing the natural history of Care Profiles. We model patients' holding times, i.e., the duration of Care Profiles, and the number of nurses' visits at each future time slot. The model has been applied to the real data of one of the largest public HC providers in Italy. We computed the estimates of all model parameters and the predictions for both new patients and patients already in the charge. Preliminary results show the applicability of the approach in the practice and good quality of predictions.

Full paper: 2893.pdf

Hospital Differences in Caesarean Deliveries in Sardinia: A Multilevel Analysis

Emiliano Sironi[1], emiliano.sironi@unicatt.it, Massimo Cannas[2], massimo.cannas@unica.it

(1) Università di Milano (Italy), (2) Università di Cagliari (Italy)

Abstract   Using data from official hospital abstracts on deliveries occurred in Sardinia during 2010/2011, we implement a multilevel logistic regression in order to assess the effect of the determinants of caesarean deliveries. Results underline as the between hospital variation is 0.39 in a model with the only intercept. After including individual characteristics of delivered mothers the variation does not reduce and it is estimated to be 0.38.

Full paper: 2948.doc

CP12 – SURVEY METHODOLOGY

Chairperson: E. Carfagna (Università di Bologna)

Heterogeneity for a general class of recapture models based on equality constraints on the conditional capture probabilities

Alessio Farcomeni, alessio.farcomeni@uniroma1.it

Sapienza Università di Roma (Italy)

Abstract   We define a class of discrete-time capture-recapture models which takes into account time (t), behavioural (b), observed heterogeneity (o) and unobserved heterogeneity (h) effects. The resulting M_{hotb} model is completely general in the tb part. We specify a model in the class through equality constraints on the conditional capture probabilities. A general EM algorithm is described to fit the chosen model.

Full paper: 2822.pdf

Uncertainty in statistical matching for complex sample surveys

Pier Luigi Conti[1], pierluigi.conti@uniroma1.it, Daniela Marella[2], daniela.marella@uniroma3.it, Mauro Scanu[3], scanu@istat.it

(1) Università di Roma Tre (Italy), (2) Sapienza Università di Roma (Italy), (3) ISTAT (Italy)

Abstract   An important feature of statistical matching is the estimation of the underlying joint distribution of variables separately available from independent sample surveys. Unless special assumptions are made, the absence of joint information on the variables of interest leads to uncertainty about the data generating model. The aim of this paper is to analyze the uncertainty in statistical matching for complex survey data.

Full paper: 2859.pdf

Learning Bayesian networks in complex survey sampling

Flaminia Musella[1], flaminia.musella@uniroma3.it, Daniela Marella[2], daniela.marella@uniroma3.it, Paola Vicard[3], paola.vicard@uniroma3.it

(1) Università di Roma Tre (Italy), (2) Università di Roma Tre (Italy), (3) Università di Roma Tre (Italy)

Abstract   Bayesian networks are multivariate statistical models satisfying sets of conditional independence statements. The association structure can be learnt from data by a sequence of independence and conditional independence tests using the PC algorithm. The learning process is based on assumption of independent and identically distributed observations. This assumption is almost never valid for sample surveys data since most of the commonly used survey designs employ stratification and/or cluster sampling and/or unequal selection probabilities. Here, a PC algorithm correction is proposed for taking into account the sampling design complexity.

Full paper: 2865.pdf

Quantile estimation with auxiliary information

Leo Pasquazzi[1], leo.pasquazzi@unimib.it, Lucio De Capitani[2], lucio.decapitani1@unimib.it

(1) Università di Milano-Bicocca (Italy), (2) Università di Milano-Bicocca (Italy)

Abstract   In this work we compare finite population quantile estimators in a simulation study. We consider settings where complete auxiliary information is available, and quantile estimators that are obtained from inversion of several well-known estimators for the population cdf. The simulation results show that estimators based on separate estimates of the regression function and the error distributions are usually the most efficient ones.

Full paper: 2857.pdf

Small Area Estimation with Covariates Perturbed for Disclosure Limitation

Silvia Polettini[1], silvia.polettini@uniroma1.it, Serena Arima[2], serena.arima@uniroma1.it

(1) Sapienza Università di Roma (Italy), (2) Sapienza Università di Roma (Italy)

Abstract   We exploit the connections between measurement error and data perturbation for disclosure limitations in the context of small area estimation. Our starting point is the model in Ybarra and Lohr (2008), where some of the covariates (all continuous) are measured with error. Using a fully Bayesian approach, we extend the aforementioned model including continuous and categorical auxiliary variables, both perturbed by disclosure limitation methods, with masking distributions fixed according to the assumed protection mechanism. In order to investigate the feasibility of the proposed method, we conduct an extensive simulation study exploring the effect of different protection scenarios on the small area mean predictions. We also perform a comparative analysis of the proposed estimator.

Full paper: 2915.pdf

CP13 – STATISTICAL METHODS FOR THE ANALYSIS OF FERTILITY AND HEALTH

Chairperson: E. Moreno (Universidad de Granada, Spain)

Birth Order, Birth Weight and Asthma: how to assess mediation and the presence of Unmeasured Confounding

Bianca Destavola[1], bianca.destavola@lshtm.ac.uk, Lorenzo Richiardi[2], lorenzo.richiardi@unito.it, Daniela Zugna[3], daniela.zugna@unito.it, Rhian Daniel[4], rhian.daniel@lshtm.ac.uk, Rossella Murtas[5], rmurtas@unica.it

(1) London school of hygiene and tropical medicine (Great Britain, UK), (2) Università di Torino (Italy), (3) Università di Torino (Italy), (4) London school of hygiene and tropical medicine (Great Britain, UK), (5) Università di Cagliari (Italy)

Abstract   If the goal is to measure the causal relation between two variables when a third is involved and plays the role of mediator, it is essential to explicitly define the relational assumptions among these and others relevant variables. However if any of this assumptions are not met, estimates of mediated effects may be affected by bias. One example where this occurs is the widely discussed situation known as the Birth Weight paradox which arises when birth weight is the mediator of interest and unmeasured confounders affect the mediator-outcome relation. In this paper we will focus on a setting where such paradox might arise where birth order act as the exposure for childhood asthma (by age 1.5). After estimating the direct and indirect effects, we give a plausible graphical explanation of the empirical results and explore the magnitude of the causal relations using sensitivity analysis.

Full paper: 2944.pdf

Reproductive Change in Transitional Italy: Insights from the Italian Fertility Survey of 1961

Alessio Fornasin[1], fornasin@uniud.it, Marco Breschi[2], breschi@uniss.it, Matteo Manfredini[3], matteo.manfredini@unipr.it, Massimo Esposito[4], mesposito@uniss.it

(1) Università di Udine (Italy), (2) Università di Sassari (Italy), (3) Università di Parma (Italy), (4) Università di Sassari (Italy)

Abstract   In 1961, a second complete fertility survey was re-proposed some thirty years later the survey of 1931, introducing more details in the information concerning ever-married women. The Central Institute of Statistics processed this mass of data only in part and moreover with considerable delay. It is therefore not surprising that the results of the census of 1961 have received less attention despite some of its innovative characteristics. Our starting point is the original "Family sheets" of the Census of 1961 regarding (as of today, March 2013) four populations located in four different regions. We are currently working to expand our analysis to other communities. This approach would allow us to focus our attention (at the micro-level) on specific groups of women who have lived as protagonists the dramatic socio-economic transformation of Italy. Moreover, the most part of such women lived their childhood, marriage and reproductive life during the Fascist era. We have therefore the opportunity to get also some insight in the effects of the pro-natalistic policy of the regime on the fertility of the various socio-economic strata of the population.

Full paper: 2835.pdf

Implementing Hierarchical Bayesian Model to Fertility Data: the case of Ethiopia

Haftu Gebremeskel[1], gebrehiwot@stat.unipd.it, Stefano Mazzuco[2], mazzuco@stat.unipd.it

(1) Università di Padova (Italy), (2) Università di Padova (Italy)

Abstract   The modeling of fertility patterns is an essential method researchers use to understand world-wide population patterns. Various types of fertility studies have been carried out particularly in developed countries and developing countries outside of Africa to model and look at fertility variations across countries. Nevertheless, much less attention has been given to local or regional fertility curves, where we expect a wider variety of patterns than for country level. Given this variety of possible fertility patterns, in this article, we will propose a model that best captures the different age-specific fertility patterns of Ethiopia at country and regional levels. However, considering the low sample size at local level, a hierarchical Bayesian approach will be considered, in order to have less imprecise estimates.

Full paper: 2960.pdf

Fertility and Child Mortality in the Sardinian Demographic Transition. Alghero (1866-1935)

Stanislao Mazzoni[1], smazzoni@uniss.it, Lucia Pozzi[2], lpozzi@uniss.it, Marco Breschi[3], breschi@uniss.it

(1) Università di Sassari (Italy), (2) Università di Sassari (Italy), (3) Università di Sassari (Italy)

Abstract   The reduction in infant and child mortality rates has been considered one of the key factors prompting fertility decline. The relationship between early life mortality and fertility is complex since involves multiple dimensions. Furthermore up to now research on this topic has not permitted to measure adequately the interaction mechanisms involved, since it has mostly relied on aggregate data. This paper aims to explore this complex relationship during the Sardinian demographic transition on the basis of an individual level data set reconstructed for the community of Alghero for the years 1866-1935. Sardinian experience is particularly interesting because the island, where infant mortality rates were amongst the lowest in the country, has known the most delayed fertility decline process.

Full paper: 2901.doc

Persistent Employment Instability and Fertility Intentions

Daria Mendola[1], daria.mendola@unipa.it, Annalisa Busetta[2], annalisa.busetta@unipa.it, Daniele Vignoli[3], vignoli@disia.unifi.it

(1) Università di Palermo (Italy), (2) Università di Palermo (Italy), (3) Università di Firenze (Italy)

Abstract   Our paper adds to the growing literature on the measurement of employment instability and on that on its consequences on fertility dynamics. Many of these studies disregard that the persistence (duration) in an unstable condition, more than the status itself may have the most severe consequences on subsequent family choices. In this paper, we propose an index of persistence in employment instability (EINI) that synthesizes all the information inside the individual sequence of employment statuses in a single number accounting simultaneously for the duration, sequencing, intensity, and labour market circumstances. Then, we test its impact on short-term childbearing intentions of Italian couples and separately for different parities.

Full paper: 2907.docx

CP14 – ADVANCED IN COMPOSITIONAL DATA ANALYSIS

Chairperson: F. Corradi (Università di Firenze)

Multiple Factor Analysis to Visually Explore Collaboration Structures: the Case of Technological Districts

Domenico De Stefano[1], domenico.destefano@unive.it, Maria Rosaria D'Esposito[2], mdesposi@unisa.it, Giancarlo Ragozini[3], giragoz@unina.it

(1) Università di Venezia (Italy), (2) Università di Salerno (Italy), (3) Università di Napoli "Federico II" (Italy)

Abstract   In the present work we propose to analyze the relational structure of collaboration networks emerging in Technological Districts (TD) observed across time occasions through the use of Multiple Factorial Analysis (MFA). This methods allows to deal with complex relational structures that lead to highly heterogeneous networks. We show how the solutions provided by the MFA can be interpreted in a natural and suitable way in the relational setting and can furnish valuable information about the strength and typology of the collaboration structure and its evolution. We focus as a case study on the analysis of the IMAST TD.

Full paper: 2971.pdf

Information Provided by Absolute, Essential and Structural Zeros in Compositional Data Sets

Josep   Martin-Fernandez[1],   josepantoni.martin@udg.edu,   Josep   Daunis-i-Estadella[2], pepus.daunis@udg.edu, Santiago Thió-Henestrosa[3], santiago.thio@udg.edu

(1) Universitat de Girona (Spain), (2) Universitat de Girona (Spain), (3) Universitat de Girona (Spain)

Abstract   There is general agreement among researchers that log-ratio methodology is the most adequate in compositional data analysis. Zero entries in compositional data matrix require special treatment. When zero values are structural, absolute and essential then the analyst should select a treatment or provide a data structure through compositions and its parts. The Use of Time Survey in Catalonia is used for illustrative purposes.

Full paper: 2906.docx

Kriging prediction for functional compositional data and application to particle-size curves

Alessandra Menafoglio[1], alessandra.menafoglio@polimi.it, Alberto Guadagnini[2], alberto.guadagnini@polimi.it, Piercesare Secchi[3], piercesare.secchi@polimi.it

(1) Politecnico di Milano (Italy), (2) The University of Arizona and Politecnico di Milano (Italy), (3) Politecnico di Milano (Italy)

Abstract   We present a new geometric approach to krige functional compositional data which embraces the viewpoints of both Functional and Compositional Data Analysis. Our theoretical framework enables one to characterize and predict random fields valued in the Hilbert space of functional compositions endowed with the Aitchison geometry. We show the application of the methodology to a field case scenario dealing with particle-size data collected within a heterogeneous aquifer near Tubingen, Germany. We consider particle-size densities, interpreted as functional compositional data, and perform kriging of these curves to obtain a complete characterization of the soil textural properties within the aquifer.

Full paper: 2836.pdf

Scaled-Dirichlet regression for compositional data

Gianna Monti[1], gianna.monti@unimib.it, Gloria Mateu-Figueras[2], gloria.mateu@udg.edu, Vera Pawlowsky-glahn[3], vera.pawlowsky@udg.edu, Juan José Egozcue[4], juan.jose.egozcue@upc.edu

(1) Università di Milano-Bicocca (Italy), (2) University of Girona (Spain), (3) University of Girona (Spain), (4) Technical University of Catalonia (Spain)

Abstract   The Dirichlet distribution, the historical model for compositional data, provides a handy tool for modelling data restricted to the unit simplex. It has been widely used in several disciplines, such as geology, biology, and chemistry. The drawbacks of the Dirichlet distribution are described extensively in Aitchison (1986). They refer to its almost completely negative correlation and its very strong independence structure. For this reason, a lot of generalizations were proposed to compensate for its lack of flexibility. Among others, in Monti et al. (2011) a scaled-Dirichlet distribution was proposed as a translation in the simplex of a Dirichlet random composition. In particular, the scaled Dirichlet distribution can be obtained starting from a perturbed random composition with a Dirichlet density. In this contribution, following the approach suggested by Campbel and Mosimann (1987), and also studied by Hijazi and Jernigan (2009), we propose a new family of models based on scaled-Dirichlet regression, where, given a vector of covariates one considers the response vector with a conditional scaled-Dirichlet distribution. The scaled-Dirichlet Covariate Model is characterized by the fact that its parameters are a linear model on the covariates, i.e. they change with the covariates. Parameter estimation and prediction issues will be explored and an example with real data will complete the presentation.

Full paper: 2928.pdf

Sparse PCA for compositional data

Nickolay Trendafilov[1], nickolay.trendafilov@open.ac.uk, Michele Gallo[2], mgallo@unior.it

(1) Open University (Great Britain, UK), (2) Università di Napoli "L'Orientale" (Italy)

Abstract   A great number of procedures for sparse principal component analysis (PCA) were proposed in the last decade. However, they cannot be applied directly for PCA of compositional data (CoDa). We introduce a new procedure for sparse PCA which takes into account the additional constraints specific for CoDa. The proposed method is very effective to find log contrasts in data, which is illustrated on a real example.

Full paper: 2863.pdf

CP15 - SPATIAL AND SPATIO

Chairperson: A. Mira (Università della Svizzera Italiana, Swetzerland)

Composite likelihood estimation in spatial logistic regression

Manuela Cattelan[1], manuela.cattelan@unipd.it, Cristiano Varin[2], cristiano.varin@unive.it

(1) Università di Padova (Italy), (2) Università di Venezia (Italy)

Abstract   A regression model for binary spatial data with dependence expressed in terms of cross-ratios is discussed. In order to avoid difficulties associated with the specification of multivariate binary distributions, a conditional pairwise likelihood method is developed for inferential purposes. The methodology is illustrated through the analysis of data about the effect of acid rains on trout populations in Norwegian lakes.

Full paper: 2924.pdf

A copula model for tourists' spending behavior

Enrico Foscolo[1], enrico.foscolo@unibz.it, Marta Disegna[2], marta.disegna@unibz.it, Fabrizio Durante[3], fabrizio.durante@unibz.it

(1) Università di Bolzano/Bozen (Italy), (2) Università di Bolzano/Bozen (Italy), (3) Università di Bolzano/Bozen (Italy)

Abstract   Analysing spending behaviour at micro-level is an important issue in order to understand how the consumer is choosing among different possibilities. In this paper, the Copula-based Logit model, a novel nonlinear approach, is presented in order to analyze the propensity to spend on different tourist expenditure categories simultaneously. A real case in tourism is exploited to assess the relevance of the proposed approach; managerial and marketing implications are derived from results.

Full paper: 2949.doc

Estimation and forecasting for binomial and negative binomial INAR(1) time series

Margherita Gerolimetto[1], margherita.gerolimetto@unive.it, Luisa Bisaglia[2], bisaglia@stat.unipd.it, Paolo Gorgi[3], paolo.gorgi@unipd.it

(1) Università di Venezia (Italy), (2) Università di Padova (Italy), (3) Università di Padova (Italy)

Abstract   In spite of the central role of the Box-Jenkins ARMA modeling, there is no such a leading technique for count time series. A proposal is represented by the INteger-values AutoRegressive process (INAR) introduced by Al-Osh and Azaid (1987) and McKenzie (1988). Typically, the Poisson distribution provides the main instrument for modeling count time series data: however, other distributional assumptions may be used instead. In this work we discuss the issue of estimating and forecasting in case of INAR(1) time series with over and under-dispersion. To take into account these features, for the distribution of the innovation term we resort to the binomial and negative binomial. We calculate the maximum likelihood function for the considered cases via a Monte Carlo experiment, we show that the resulting estimators have a good performance. Moreover, we also concentrate on the problem of producing coherent predictions based on estimates of the p-step ahead predictive mass function assuming binomial and negative binomial distributions of the error term. Finally, we compare the forecast accuracy of binomial and negative binomial INAR with that of Poisson INAR and ARMA models with a Monte Carlo experiment.

Full paper: 2913.pdf

Kalman Filter for Estimating Bivariate GMRFs on Regular Lattice

Luca Romagnoli[1], luca.romagnoli@unimol.it, Luigi Ippoliti[2], ippoliti@unich.it, Richard Martin[3], r.j.martin26@gmail.com

(1) Università del Molise (Italy), (2) Università di Chieti-Pescara (Italy), (3) Università di Chieti-Pescara (Italy)

Abstract   Bivariate conditional autoregressive CAR models are widely used classes of multivariate spatial models. In this paper, we consider their maximum likelihood estimation which, under general boundary conditions, can be problematic. As in time series analysis, it will be shown that the Kalman filter provides an alternative and a computationally efficient solution for computing the likelihood.

Full paper: 2919.pdf

A note on semivariogram

Grazia Vicario[1], grazia.vicario@polito.it, Giovanni Pistone[2], giovanni.pistone@carloalberto.org

(1) Politecnico di Torino (Italy), (2) Collegio Carlo Alberto (Italy)

Abstract   Variograms are usually discussed in the framework of stationary or intrinsically stationary processes. We retell here this piece of theory in the setting of generic Gaussian vectors.

Full paper: 2925.pdf

CP16 - ENVIRONMENTAL AND POVERTY DATA ANALYSIS

Chairperson: B. Liseo (Sapienza Università di Roma)

Quantifying uncertainty associated with a numerical model output

Lucia Paci[1], lucia.paci2@unibo.it, Daniela Cocchi[2], daniela.cocchi@unibo.it, Alan Gelfand[3], alan@stat.duke.edu

(1) Università di Bologna (Italy), (2) Università di Bologna (Italy), (3) Duke University (U.S.A.)

Abstract   Environmental numerical models are deterministic tools widely used to simulate and predict complex systems. However, they are unsatifying since they do not provide information about the uncertainty associated with their predictions. Conversely, uncertainty assessment of model outputs can be useful to guide environmental agencies in improving computer models. We propose a Bayesian hierarchical model to obtain spatially varying uncertainty associated with a numerical model output. We show how we can learn about such uncertainty through suitable stochastic data fusion modeling using some external validation data. The model is illustrated by providing the uncertainty map associated with a temperature output over the northeastern United States.

Full paper: 2824.pdf

Smoothed Common Trend in Multivariate Time Series Air Pollution Data

Giuliana Passamani[1], giuliana.passamani@unitn.it, Paola Masotti[2], paola.masotti@unitn.it

(1) Università di Trento (Italy), (2) Università di Trento (Italy)

Abstract   Air quality improvement is a fundamental issue, given the possible effects of air pollution on human health. The paper focuses on detecting the trend component of an unobserved dynamic factor that we consider as a pollution indicator. This factor emerges from the multivariate monthly time series analysis of the levels of four air pollutants measured at four different monitoring sites in the Alpine Province of Trento over the last ten years. In particular, we suggest a procedure that can be used in order to assess whether any improvement in the pollution level has been observed during the period of observation. The suggested procedure is based on the double exponential smoothing which is simpler than state space time series analysis and gives comparable results.

Full paper: 2952.doc

IDMS: The Sardinian Index of Multiple Deprivation

Silvia Perra[1], silviaperra@gmail.com, Stefano Cabras[2], s.cabras@unica.it, Alberto Serci[3], alberto.serci@gmail.com, Alessandra Mura[4], alessandra.mura@gmail.com, Antonella De Arca[5], adearca@gmail.com, Stefano Renoldi[6], stefano.renoldi@gmail.com, Antonello Podda[7], antonellopodda@gmail.com

(1) Università di Cagliari (Italy), (2) Università di Cagliari and Universidad Carlos III de Madrid (Italy), (3) Regione Autonoma della Sardegna (Italy), (4) Regione Autonoma della Sardegna (Italy), (5) Regione Autonoma della Sardegna (Italy), (6) Università di Cagliari (Italy), (7) Università di Cagliari (Italy)

Abstract   The aim of this paper is to briefly illustrate the construction methodology of an index of deprivation. Deprivation is a concept wider than poverty and can be defined as a general lack of resources and possibilities. That's why we investigate different aspects -or domains- of interest to the citizenship, and that's why we define the index as an Index of Multiple Deprivation for Sardinia (IDMS). In particular, we studied the following seven domains: income, availability of services, education, health, environment, criminality and employment. The estimation of IDMS poses interesting statistical problems as well as of data collection because the index is estimated using administrative registers, which have their own peculiarities and need suitable estimation methods. For each domain we briefly illustrate the problems faced with data and the statistical methods used. Results for each domain and for the global index are summarized trough cartography.

Full paper: 2860.doc

Growth curves of sorghum roots via quantile regression with P-splines

Alessio Pollice[1], alessio.pollice@uniba.it, Vito Muggeo[2], vito.muggeo@unipa.it, Federico Torretta[3], federico.torretta@unipa.it, Rocco Bochicchio[4], bochicchiorocco@gmail.com, Mariana Amato[5], mariana.amato@unibas.it

(1) Università di Bari (Italy), (2) Università di Palermo (Italy), (3) Università di Palermo (Italy), (4) Università del Salento (Italy), (5) Università del Salento (Italy)

Abstract   Plant roots are a major pool of total carbon in the planet and their dynamics are directly relevant to greenhouse gas balance. Composted wastes are increasingly used in agriculture for environmental and economic reasons and their role as a substitute for traditional fertilizers needs to be tested on all plant components. Here we propose a regression quantile approach based on P-splines to assess, quantify and compare the root growth patterns in two treatment groups respectively undergoing compost and traditional fertilization.

Full paper: 2956.pdf

On species sampling sequences induced by residual allocation models

Abel Rodriguez[1], abel@soe.ucsc.edu, Fernado Quintana[2], quintana@mat.puc.cl

(1) University of California (U.S.A.), (2) Pontificia Universidad Catolica de Chile (Chile)

Abstract   We discuss fully Bayesian inference in a class of species sampling models that are induced by residual allocation (sometimes called stick-breaking) priors on almost surely discrete random measures. This class provides a generalization of the well-known Ewens sampling formula that allows for additional flexibility while retaining computational tractability. In particular, the procedure is used to derive the exchangeable predictive probability functions associated with the generalized Dirichlet process of Hjort, (2000). The procedure is illustrated with an application to genetics.

Full paper: 2885.pdf

CP17 – TOPICS IN REGRESSION MODELS

Chairperson: M.E. Castellanos (Universidad Rey Juan Carlos, Spain)

Mixtures of Logit Regressions Detection with Forward Search

Matilde Bini[1], mbini.unier@gmail.com, Vieri Del Panta[2], vieri.delpanta@virgilio.it, Margherita Velucchi[3], margherita.velucchi@unier.it

(1) Università Europea di Roma (Italy), (2) Università Europea di Roma (Italy), (3) Università Europea di Roma (Italy)

Abstract   The forward search is a general method to detect multiple outliers and to determine their effect on inference about models fitted to data. From the monitoring of a series of statistics based on subsets of data of increasing size we obtain multiple views of any hidden structure. Sometimes, some features emerge unexpectedly during the progression of the forward search only when a specific combination of forward plots is inspected at the same time. These features have to be harmonized and linked together in order to give an exhaustive description of a complex problem. In this paper, we use a set of new robust graphical tools on a mixture of logit regressions. We use simulated data and we show the dynamic interaction with different "robust plots" to highlight the presence of groups of outliers and regression mixtures in the context of logit regression and highlight the effect that these hidden groups provide on the fitted model.

Full paper: 2982.pdf

Conditional concordance of the signs of the residuals of quantiles regressions of multivariate outcomes

Silvia Columbu[1], silvia.columbu@unica.it, Matteo Bottai[2], matteo.bottai@ki.se

(1) Università di Cagliari (Italy), (2) Karolinska Institutet (Sweden)

Abstract   The extension of the concept of quantile of a univariate response variable to a multivariate response is not straightforward because of the absence of a natural definition of multivariate quantile. This paper describes an approach to studying the dependence of a multivariate response variable that is based on the comparison of signs of the residuals of univariate quantile regressions. The method permits to capture important aspects of the association structure of dependent variables given a set of covariates.

Full paper: 2965.pdf

Indicator functions and saturated fractions for factorial designs: a case study

Roberto Fontana[1], roberto.fontana@polito.it, Fabio Rapallo[2], fabio.rapallo@unipmn.it, Maria Piera Rogantin[3], rogantin@dima.unige.it

(1) Politecnico di Torino (Italy), (2) Università del Piemonte Orientale (Italy), (3) Università di Genova (Italy)

Abstract   We consider a recently developed algebraic criterion to check whether a fraction is saturated or not for a given model. Such criterion is based on combinatorial algebraic objects, namely the circuit basis of the design matrix of the model. We show on a case study how to use indicator functions of the circuits to classify saturated fractions. The connection with two optimality criteria is also analyzed.

Full paper: 2888.pdf

Complex Redundancy Analysis models with covariate effect: a simulation study

Pia Clara Pafundi[1], p.pafundi@campus.unimib.it, Gianmarco Vacca[2], g.vacca@campus.unimib.it

(1) Università di Milano-Bicocca (Italy), (2) Università di Milano-Bicocca (Italy)

Abstract   In the field of Structural Equation Models (SEM), the focus of the present work will be on the Redundancy Analysis (RA) framework. The extension of RA to more than two sets of variables will be the main topic of interest, with the recently developed Extended Redundancy Analysis (ERA) as the major outline. Drawbacks of the model in presence of concomitant indicators will be highlighted, subsequently introducing a new methodological development, called Generalized Redundancy Analysis (GRA). Motivations for this enhancement will be given, along with a simulation study aimed to assess the performance of the GRA model, in three structural models with increasing complexity.

Full paper: 2869.doc

Variable selection in mixed models: a graphical approach

Mariangela Sciandra, mariangela.sciandra@unipa.it

Università di Palermo (Italy)

Abstract   Model selection can be defined as the task of estimating the performance of different models in order to choose the (approximate) best one. The purpose of this article is to introduce an extension of the graphical representation of deviance proposed in the framework of classical and generalized linear models to the wider class of mixed models. The proposed plot is useful in determining which are the important explanatory variables conditioning on the random effects part. The applicability and the easy interpretation of the graph are illustrated with a real data examples.

Full paper: 2931.pdf

CP18 – BAYESIAN METHODS AND MODELS

Chairperson: I. Pruenster (Collegio Carlo Alberto)

On diversity under a Bayesian nonparametric dependent model

Julyan Arbel[1], julyan.arbel@carloalberto.org, Kerrie Mengersen[2], k.mengersen@qut.edu.au, Judith Rousseau[3], rousseau@ceremade.dauphine.fr

(1) Collegio Carlo Alberto (Italy), (2) Queensland University of Technology (Australia), (3) Université Paris-Dauphine (France)

Abstract   We present a dependent Bayesian nonparametric model for the probabilistic modelling of species-by-site data, i.e. population data where observations at different sites are classified into distinct species. We use a dependent version of the Griffiths-Engen-McCloskey distribution, the distribution of the weights of the Dirichlet process, in the same lines as the Dependent Dirichlet process is defined. The prior is thus defined via the stick-breaking construction. Some distributional properties of this model are presented.

Full paper: 2936.pdf

Skew-normal nonparametric mixture models

Antonio Canale[1], antonio.canale@unito.it, Bruno Scarpa[2], scarpa@stat.unipd.it

(1) Università di Torino and Collegio Carlo Alberto (Italy), (2) Università di Padova (Italy)

Abstract   In this paper we investigate nonparametric mixtures of skew-normal kernels under a Bayesian perspective. The choice of this kernel function allows us to formulate nonparametric location-scale-shape mixture prior with large support and good mathematical and computational properties. The performance of the methods are tested through simulations and applications to galaxy velocity data

Full paper: 2844.pdf

Bayesian hierarchical models for misaligned data: a simulation study

Giulia Roli[1], g.roli@unibo.it, Meri Raggi[2], meri.raggi@unibo.it

(1) Università di Bologna (Italy), (2) Università di Bologna (Italy)

Abstract   In this paper, the problem of combining information from different sources of data is considered. We focus our attention on spatially misaligned data, where available information (typically counts or rates from administrative sources) refers to spatial units that are different from the ones of interest. A hierarchical Bayesian perspective is considered, as firstly proposed by Mugglin *et al.* in 2000, to provide a fully model-based approach in an inferential, and not only descriptive, sense. In particular, explanatory covariates are arranged to be modeled according to spatial correlation through a conditionally autoregressive prior structure. In order to assess model performance and its robustness we generate artificial data basing on a real study and a simulation exercise is then carried out.

Full paper: 2927.pdf

Posterior contraction rates for empirical Bayes procedures with applications

Catia Scricciolo[1], catia.scricciolo@unibocconi.it, Sophie Donnet[2], donnet@ceremade.dauphine.fr,
Vincent Rivoirard[3], rivoirard@ceremade.dauphine.fr,
Judith Rousseau[4], rousseau@ceremade.dauphine.fr

(1) Università Bocconi (Italy), (2) Université Paris Dauphine (France), (3) Université Paris Dauphine (France), (4) Université Paris Dauphine (Italy)

Abstract   We consider the asymptotic behavior of empirical Bayes posterior distributions for infiniite-dimensional statistical models. We give general results on contraction rates for empirical Bayes posterior distributions which are applied to several examples including priors on monotone non-increasing intensities of Aalen multiplicative intensity processes and Dirichlet process Gaussian mixtures.

Full paper: 2918.pdf

Poster Session

Frames2: an R package for estimation in dual frames

Antonio Arcos[1], arcos@ugr.es, María Del Mar Rueda[2], mrueda@ugr.es, David Molina[3], dmolinam@ugr.es, Maria Giovanna Ranalli[4], giovanna@stat.unipg.it

(1) Universidad de Granada (Spain), (2) Universidad de Granada (Spain), (3) Universidad de Granada (Spain), (4) Università di Perugia (Italy)

Abstract Dual frame surveys are very useful when it is not possible to guarantee a complete coverage of the target population and may result in considerable cost savings over a single frame design with comparable precision. In fact, dual frame methodology has experienced a substantial development along last years and nowadays there are more and more statistical agencies and private organizations using dual frame surveys to take advantage of their widely proved benefits. However, most statistical commercial software does not incorporate estimation procedures for data coming from dual frame surveys. In this work, we present main features of Frames2, a new R package incorporating functions for point and confidence estimation for most of the estimators in dual frame proposed so far: from the one suggested by Hartley (1962) to the recently arisen calibration estimators (Ranalli et al. (2013)).

Full paper: 2993.pdf

Spatial smoothing on complex regions: a case study on the median length of deep water rose shrimps in the North-Western Ionian Sea

Simona Arcuti[1], simona.arcuti@gmail.com, Alessio Pollice[2], alessio.pollice@uniba.it, Crescenza Calculli[3], calculli.enza@gmail.com, Nunziata Ribecco[4], nunziata.ribecco@uniba.it, Angelo Tursi[5], angelo.tursi@uniba.it

(1) Università di Bari (Italy), (2) Università di Bari (Italy), (3) Università di Bari (Italy), (4) Università di Bari (Italy), (5) Università di Bari (Italy)

Abstract Conventional smoothing may produce errors across the boundary when used over complicated coastal and island regions, due to the Euclidean distances used to measure similarity between data points. In this work we want to compare the performance, in terms of prediction, of a conventional smoother based on Thin Plate Regression Splines (Wood, 2006) with the one of an alternative smoother, called Soap Film smoother (Wood, 2008). Both smoothers are considered inside a generalized additive model estimated to evaluate the changes in the median length (mm) of an aquatic population, over spatio-temporal scales and in response to anthropogenic and environmental factors.

Full paper: 2921.pdf

The effect of support schemes on Photovoltaic installed capacity in Europe: a WDEA-STATIS analysis

Filippa Bono[1], filippa.bono@unipa.it, Marcella Giacomarra[2], marcella.giacomarra@unipa.it

(1) Università di Palermo (Italy), (2) Università di Palermo (Italy)

Abstract Since 1985 the European Union (EU) has been committed in the Renewable Energy Sources (RES) promotion, focusing the attention on the financial coordination needed to support the sector, and on the possibility of harmonizing support schemes systems. Starting from this assumption, aim of this work is to provide, by means of STATIS together with windows DEA analysis, an answer about which kind of support schemes have worked best in a sample of 11 EU Member States in the last 15 years in terms of overall installed capacity, focusing the attention on the Photovoltaic sector, the most expensive RES technology.

Full paper: 2988.docx

Bayesian Network structural learning in multivariate time series

Maria Caterina Bramati[1], mariacaterina.bramati@uniroma1.it,
Flaminia Musella[2], flaminia.musella@uniroma3.it

(1) Sapienza Università di Roma (Italy), (2) Università di Roma Tre (Italy)

Abstract   Structural learning of Bayesian network is usually fulfilled by the expert knowledge, whenever available, or by some efficient algorithmic procedures. Despite the vast literature on structural learning, still little has been done specifically aimed at the multivariate time series modeling. We suggest a testing procedure able to learn the DAG structure whose vertex set only consists of the components of a stationary vector autoregressive VAR(p) model. The proposal procedure follows a constraint-based approach by using a test between blocks of variables. The class of tests proposed is based on multivariate ranks of distances and it is asymptotically distribution-free under very mild assumptions on the noise.

Full paper: 2856.pdf

Gender Differentiation of Human Longevity in Sardinian Provinces

Maurizio Brizzi[1], maurizio.brizzi@unibo.it, Alessia Orrù[2], ceylon2005@yahoo.it

(1) Università di Bologna (Italy), (2) Università di Cagliari (Italy)

Abstract   One of the most interesting features of Sardinian population is the tendency to longevity, and within this phenomenon there is a neat gender differentiation: the sex ratio (Female/Male) at final classes of age, which is usually very favorable to females, in Sardinia is sensibly more reduced than in the whole Italy. In this paper some geographic differences are pointed out, in order to find a local kernel of specificity, which could be located in the provinces of Nuoro and Olbia-Tempio. This specificity seems to be also stable in time, since many results are relatively similar in 1991 and 2011 Italian Census.

Full paper: 2943.doc

Modelling the effect of vehicular–traffic and meteorology on fine particle concentration using Additive Mixed Models: the case of the town of Perugia

Simone Del Sarto, delsarto@stat.unipg.it, M. Giovanna Ranalli, giovanna.ranalli@stat.unipg.it, Elena Stanghellini, elena.stanghellini@stat.unipg.it, Davide Cappelletti david.cappelletti@unipg.it, Beatrice Moroni, moroni@impact.dyn.unipg.it, Stefano Crocchianti, croc@impact.dyn.unipg.it, Silvia Castellini, castellini@impact.dyn.unipg.it

Università di Perugia (Italy)

Abstract   The purpose of this work consists in understanding how fine particles concentration in the town of Perugia is affected by vehicular traffic and meteorological variables. In particular we analyse data from the PMetro project (http://www.pmetro.it), which studies urban pollution dynamics in the town of Perugia. The particles concentration data are collected by an instrument located on a cabin of Minimetro, a public conveyance that moves on a monorail throughout the town. We merge these data with vehicular traffic and several meteorological variables (temperature, wind speed, etc.). We analyse this dataset using Additive Mixed Models, which use a smooth function (such as penalized splines) to model the effect of each covariate, instead of a polynomial pre–specified function.

Full paper: 3148.pdf

Outsourcing in the Italian NHS: a statistical measure of mismatch between private supply and public demand

Rosa Falotico[1], rosafalotico@hotmail.it, Paolo Mariani[2], paolo.mariani@unimib.it

(1) Università di Milano-Bicocca (Italy), (2) Università di Milano-Bicocca (Italy)

Abstract   Our study deals with the relationship between ASLs (Aziende Sanitarie Locali) and private operators outsourcing in Italy. We hypothesize that the level of matching between public demand and private supply of outsourcing can affect the outsourcing development in Italian National Health Service. Under this hypothesis we present an analysis that aims at quantifying the level of mismatch between pharmaceutical companies offer and ASL needs in order to understand the future developments of outsourcing. We studied 44 pharmaceutical companies (with a yearly turnover of about € 6 billion) and 34 ASLs (with about 12 million users). The results suggest that the low public and private level of outsourcing mismatch are increasing in the future.

Full paper: 2929.doc

Functional analysis of variance for parametric functional data

Francesca Fortuna[1], francesca.fortuna@unich.it, Fabrizio Maturo[2], fbmatu@gmail.com

(1) Università di Chieti-Pescara (Italy), (2) Università di Chieti-Pescara (Italy)

Abstract   A parametric functional analysis of variance (FANOVA) is proposed to measure the functional variability explained by other variables. This technique presents the advantage of using all the information in the curves, instead of some specific values on them. In particular, we refer to cases in which the functional observations belong to a parametric family of functions, whose functional form is known in advance. In this framework, the approximation of the function underlying the data is not required as in the classical functional approach. Therefore, we avoid the impact that smoothing techniques might have on FANOVA. The proposed method is applied to a real data set concerning lichen biodiversity in Liguria region, in northwestern Italy.

Full paper: 3150.pdf

A performance comparison of the Lp-norm methods in multicollinearity situations, supposing a generalized normal distribution errors

Massimiliano Giacalone[1], maxgiacalone@yahoo.it, Paolo Carmelo Cozzucoli[2], cozzucoli@unical.it

(1) Università Bocconi (Italy), (2) Università della Calabria (Italy)

Abstract   The normal distributions of order p, known also as exponential power distributions, are used in the description of non normal random errors in the general situation of L_p-norm estimation. L_1-norm estimators (least absolute deviation);L_2-norm estimators (least squares) and other L_p norm proposal estimators are reviewed and applied to estimate the parameters of a particular linear regression model. In the paper are also reported few results concerning three alternative p-estimation methods to evaluate their performance for a linear model.

Full paper: 2969.doc

Jeffreys Priors for Mixture Models

Clara Grazian[1], clara.grazian@ceremade.dauphine.fr,
Christian Robert[2], Christian.Robert@ceremade.dauphine.fr

(1) Sapienza Università di Roma (Italy), (2) Université Paris Dauphine (France)

Abstract   Mixture models may be a useful and flexible tool to describe data with a complicated structure, for instance characterized by multimodality or asymmetry.In a Bayesian setting, it is a well established fact that one need to be careful in using improper prior distributions, since the posterior distribution may not be proper. This feature leads to problems in carry out an objective Bayesian approach. In this work an analysis of Jeffreys priors in the setting of finite mixture models will be presented.

Full paper: 2891.pdf

Bell shaped fuzzy numbers associated with the normal curve

Fabrizio Maturo[1], fabmatu@gmail.com, Francesca Fortuna[2], francesca.fortuna@unich.it

(1) Università di Chieti-Pescara (Italy), (2) Università di Chieti-Pescara (Italy)

Abstract   Fuzzy regression models often consider triangular or trapezoidal member- ship functions. Of course this kind of fuzzy numbers has the advantage that they are much easy to use; nevertheless several doubts arise about the appropriateness of the triangular shape. As known, the fuzzy sets are very useful when there are imprecise data and in particular when, within an interval, the degree of truth of the values is not always the same. Fuzzy numbers are also useful to translate terms of human language into numbers. Triangular fuzzy numbers offer a good solution for dealing with this kind of data. However in this paper it is proposed an alternative member- ship function which appears to be much more appropriate; it has been called "fuzzy number associated with the normal curve". In this framework it will be justified the introduction of this membership function as a valid substitute for triangular fuzzy numbers.

Full paper: 3149.pdf

Material deprivation among foreigners in Italy

Daria Mendola[1], daria.mendola@unipa.it, Annalisa Busetta[2], annalisa.busetta@unipa.it, Philippe Van Kerm[3], philippe.vankerm@ceps.lu, Anna Maria Milito[4], annamaria.milito@unipa.it

(1) Università di Palermo (Italy), (2) Università di Palermo (Italy), (3) CEPS/INSTEAD (Luxembourg), (4) Università di Palermo (Italy)

Abstract   In all European countries, migrant populations tend to have worse living conditions than native; this is particularly true for those born outside the EU. This paper proposes a new way to look at the relative living conditions of foreigners by looking at non-monetary indicators of material deprivation in Italy. In particular, we discuss differences in material deprivation among groups of foreigners once we control for the demographic and socioeconomic characteristics of each group using a flexible standardization methodology. Our results show that, in Italy, foreigners from African and Mediterranean countries and, to a lesser extent, from South Asia are most deprived and that the construction of the counterfactual distributions only marginally explain the gap between different foreigner groups.

Full paper: 2890.pdf

Local Spatial Analysis of Cardiovascular Diseases in Canadian Health Regions

Eugenia Nissi[1], nissi@unich.it, Annalina Sarra[2], asarra@unich.it

(1) Università di Chieti-Pescara (Italy), (2) Università di Chieti-Pescara (Italy)

Abstract   Cardiovascular diseases are one of the major worldwide health concerns, responsible for premature death and disability all over the world. The present analysis has as its primary intention in assessing the extent to which the association between cardiovascular disease prevalence in the Canadian Health Regions and some well-established risk factors vary spatially. We adopt the GWR to explore if the effects of some cardiovascular disease predictors are heterogeneous across space and hence vary from place to place. Since GWR amplifies the negative effects of correlation in the weighted explanatory variables, in this study we also employ a GWR model where a ridge regression parameter has been incorporated to reduce model complications arising from collinearity.

Full paper: 3151.pdf

Analysis of the causal factors in incidents of the Greek petrochemical industry

Zoe Nivolianitou[1], zoe@ipta.demokritos.gr, Myrto Konstandinidou[2], myrto@ipta.demokritos.gr, Chrys Caroni[3], ccar@math.ntua.gr, Irini Kefalogianni[4], eirhnh.kefalogiannh@gmail.com

(1) National Center for Scientific Research of Greece (Greece), (2) National Center for Scientific Research of Greece (Greece), (3) National Technical University of Athens (Greece), (4) National Technical University of Athens (Greece)

Abstract   This paper presents a statistical analysis of all reported incidents in the Greek petrochemical industry from 1997 to 2003. A comprehensive database has been developed to include industrial accidents (fires, explosions and substance releases), occupational accidents, incidents without significant consequences and near misses. The study concentrates on identifying and analysing the causal factors related to different consequences of incidents, in particular, injury, absence from work and material damage. Methods of analysis include logistic regression with one of these consequences as dependent variable. The causal factors that are considered cover four major categories related to organizational issues, equipment malfunctions, human errors (of commission or omission) and external causes. Further analyses aim to confirm the value of recording near misses by comparing their causal factors with those of more serious incidents. The statistical analysis highlights the connection between the human factor and the underlying causes of accidents or incidents.

Full paper: 2866.doc

A classification of Italian consumers based on a proposed measure of their attitudes towards food risks

Anna Pinto[1], APinto@izsvenezie.it, Giulia Mascarello[2], GMascarello@izsvenezie.it, Nicoletta Parise[3], nicoletta.parise@gmail.com, Silvia Bonaldo[4], SBonaldo@izsvenezie.it, Stefania Crovato[5], SCrovato@izsvenezie.it, Licia Ravarotto[6], LRavarotto@izsvenezie.it

(1) Istituto Zooprofilattico Sperimentale delle Venezie (Italy), (2) Istituto Zooprofilattico Sperimentale delle Venezie (Italy), (3) Università di Padova (Italy), (4) Istituto Zooprofilattico Sperimentale delle Venezie (Italy), (5) Istituto Zooprofilattico Sperimentale delle Venezie (Italy), (6) Istituto Zooprofilattico Sperimentale delle Venezie (Italy)

Abstract   This paper aims to propose a measure of Italian consumers' attitudes towards food risks in order to distinguish self-protective consumers from non-self-protective consumers towards food risks. Through the survey method CATI (Computer Assisted Telephone Interviewing) a sample of 1,000 consumers, representative of the Italian population, was interviewed on the issues of risk perception and eating habits in general. Factor analysis combined into a single factor three variables: the consumers' propensity to read the label when buying a food product for the first time, the propensity to seek food-related information and the perceived level of exposure to food-borne diseases. This technique has allowed us to define the attitudes of Italian consumers towards food risks. The identification of characteristics that distinguish self-protective consumers from non-self-protective consumers represent important information for the authorities concerned to reduce exposure of consumers to food risks.

Full paper: 2896.pdf

A version of the geometry of the multivariate Gaussian model, with applications

Giovanni Pistone, giovanni.pistone@carloalberto.org

Collegio Carlo Alberto (Italy)

Abstract   We present a version of the classical geometry of the Gaussian multivariate model that has some advantage in the treatment of operations on the tangent bundle on the model. Applications and generalisations are briefly discussed.

Full paper: 2855.pdf

Family size and educational outcomes: empirical evidence through multilevel approach

Pasquale Recchia[1], pasqualerecchia86@gmail.com, Ernesto Toma[2], ernesto.toma@uniba.it

(1) Università di Bari (Italy), (2) Università di Bari (Italy)

Abstract   The aim of this work is to test if family size is correlated with educational outcomes of pupils, analyzing data from the survey INVALSI 2011 on the 5th grade of Italian primary school. Using a multilevel regression model we verified that children with only one sibling get math scores slightly higher than only children, even with the same other observable characteristics, while children of large families (three or more siblings) obtain significantly worse average results. Furthermore, we also found the presence of a contextual effect that leads to worse scores for students attending classes with many children from large families.

Full paper: 2884.docx

Fractionalization and Polarization

Maria Piera Rogantin[1], rogantin@dima.unige.it, Giovanni Pistone[2], giovanni.pistone@gmail.com

(1) Università di Genova (Italy), (2) Collegio Carlo Alberto (Italy)

Abstract   Fractionalization is a measure of concentration of a qualitative distribution, while Polarization is a measure of its bimodality. We study the steepest ascent lines of these functions.

Full paper: 2910.pdf

A spatial functional approach for curve classification

Elvira Romano[1], elvira.romano@unina2.it, Maria Dolores Ruiz-Medina[2], mruiz@ugr.es, Rosa M. M. Espejo[3], rmespejo@ugr.es

(1) Dipartimento di Studi Europei e mediterranei (Italy), (2) Universidad de Granada (Spain), (3) Universidad de Granada (Spain)

Abstract   This paper proposes a spatial functional formulation of the normal mixed effect model for the statistical classification of spatially dependent Gaussian curves. Fixed effect parameters are represented in terms of a functional multiple regression model whose regression operators can change in space. Local spatial homogeneity of these operators is measured in terms of their Hilbert-Schmidt distances, leading to the classification of fixed effect curves in different groups. A real data example in the financial context is analyzed as an illustration.

Full paper: 3110.pdf

An extension of Non-Metric approach to inwards directed PLS Path Models

Giorgio Russolillo[1], giorgio.russolillo@cnam.fr, Laura Trinchera[2], laura.trinchera@neoma-bs.fr

(1) CNAM (France), (2) NEOMA Business School (France)

Abstract   Partial Least Squares Path Modeling (PLS-PM) aims at modeling multivariate relations among several blocks of variables observed on the same units. Since its first presentation in 1975 by Herman Wold, several authors have developed specific techniques related to the PLS-PM and investigated its mathematical, numerical and statistical properties. In recent years researchers mainly focused on the optimization criteria behind the PLS-PM algorithm and on modifications of such criteria in order to handle variables independently of the measurement scale on which they have been observed. In this paper we extend the Non-Metric approach for the analysis of variables measured on different scales presented by Russolillo in 2012 to inwards directed PLS Path Models.

Full paper: 3036.pdf

Adapting Benjamini-Hochberg by Simes Inequality

Aldo Solari[1], aldo.solari@unimib.it, Jelle Goeman[2], j.j.goeman@lumc.nl

(1) Università di Milano-Bicocca (Italy), (2) Leiden University Medical Center (Netherlands)

Abstract
Benjamini and Hochberg introduced, in 1995, a simple but very successful procedure which bounds the False Discovery Rate at $\alpha\pi_0$, where π_0 is the unknown proportion of true null hypotheses [1]. Since that time, various Authors have tried to achieve adaptive control of the FDR by incorporating an estimate of π_0 into the Benjamini and Hochberg procedure. We propose a two-stage procedure in which Simes inequality [2] is used to derive an upper bound for π_0 in the first stage, providing an adjusted level for use of the Benjamini and Hochberg procedure in the second stage. The advantage of this approach lies in the link between Simes inequality and Benjamini-Hochberg procedure, thereby making the first stage a very natural way to estimate π_0.

Full paper: 3083.pdf

Co-authorship Patterns of Italian Statisticians by Combining Different Data Sources

Susanna Zaccarin[1], susannaz@deams.units.it, Domenico De Stefano[2], domenico.destefano@unive.it, Vittorio Fuccella[3], vfuccella@unisa.it, Maria Vitale[4], mvitale@unisa.it

(1) Università di Trieste (Italy), (2) Università di Venezia (Italy), (3) Università di Salerno (Italy), (4) Università di Salerno (Italy)

Abstract
The present paper aims at describing the scientific collaboration patterns of the Italian academic Statisticians by exploiting the information retrieved through different bibliographic archives: ISI-WoS, Current Index to Statistics, and the bibliographic information of the database of nationally funded research projects. A data merging procedure is adopted to obtain a unified dataset containing both top-international as well as nationally oriented production. The unified dataset will be used as new basis for network analysis.

Full paper: 2941.pdf

CAGLIARI, JUNE 11, 2014