



2012

Società Italiana di Statistica

Proceedings of the XLVI Scientific Meeting

SAPIENZA UNIVERSITY OF ROME

June 20-22, 2012

Plenary sessions
Invited sessions
Contributed sessions



PRESENTATIONS AND AUTHORS

BROWSE PAPERS BY SECTION

PLENARY

Research advances and new challenges in Cluster Analysis

Maurizio Vichi

Handling Measurement Error in Survey Estimation using Accuracy Indicators **PAPER NOT AVAILABLE**

Chris Skinner

Integrating micro and macro data in historical demography **PAPER NOT AVAILABLE**

Marco Breschi

SPECIALIZED

A Bayesian nonparametric model for count functional data

Antonio Canale, David B. Dunson

ROI analysis of pharmafMRI data: an adaptive approach for global testing

Giorgos Minas, John A.D. Aston, Thomas E Nichols, Nigel Stallard

Distance - Based Statistics for Covariance Operators in Functional Data Analysis

Davide Pigoli

Clustering Multivariate Longitudinal Data: Hidden Markov of Factor Analyzers

Antonello Maruotti, Francesca Martella

Model based clustering of multivariate spatio-temporal data: a matrix-variate approach

Cinzia Viroli

Random coefficient based dropout models: a finite mixture approach

Alessandra Spagnoli, Marco Alfò

Bayesian inference for causal effects in randomized experiments with noncompliance: The role of multivariate outcomes

Fan Li, Alessandra Mattei, Fabrizia Mealli

Unconditional and Conditional Quantile Treatment Effect: Identification Strategies and Interpretations

Margherita Fort

Dealing with complex problems of confounding in mediation analysis

Stijn Vansteelandt

A Unified Approach for Defining Optimal Multivariate and Multi-Domains Sampling Designs

Pietro Demetrio Falorsi, Paolo Righi

Forest Inventories: Multi-Phase Sampling Strategies for Estimating Forest and Non-Forest Resources Over Large Areas

Lorenzo Fattorini

Recent Advances in Estimation of Poverty Indicators for Domains and Small Areas

Risto Lehtonen, Ari Veijanen

Weighted likelihood in Bayesian inference

Claudio Agostinelli, Luca Greco

Disclosure risk estimation via nonparametric log-linear models

Cinzia Carota, Maurizio Filippone, Roberto Leombruni, Silvia Poletti

Bayesian Latent Class Models in Veterinary and Human Epidemiology

Luzia Goncalves, Ana Subtil, Nuno Brites, M. Rosario de Oliveira, Ana Margarida Alho, Jose Meireles, L. M. Madeira de Carvalho, Silvana Belo

STAR modeling of pulmonary tuberculosis delay-time in diagnosis

Bruno de Sousa, Dulce Gomes, Patricia A. Filipe, Cristina Areias, Teodoro Briz, Carla Nunes

ROC Curves in medical decision

Ana Cristina Braga, Lino Costa, Pedro Oliveira

On Dividing an Empirical Distribution into Optimal Segments

Jan W. Owsinski

A composite indicator of sustainable well-being: the relative importance of weights in the European Strategy for Sustainable Development

Elena Giachin Ricca, Stefano Tarantola

Non-aggregative assessment of subjective well-being

Marco Fattore

On the Extraction of a Common Persistent Component from Several Volatility Indicators

Fabrizio Cipollini, Giampiero M. Gallo

Estimating jumps in volatility using realized-range measures

Massimiliano Caporin, Eduardo Rossi, Paolo Santucci de Magistris

The Value of Model Sophistication: DJIA Option Pricing

Jeroen V.K. Rombouts, Lars Stentoft, Francesco Violante

Families in Italy: the Quiet Revolution

Silvana Salvini, Gustavo De Santis

- Enterprise in a globalised context and public and private statistical setups
Fabrizio Guelpa, Giovanni Foresti, Stefania Trenti
- Imputation and outlier detection in banking datasets
Andrea Pagano, Domenico Perrotta, Spyros Arsenis
- Something for nothing?
Shirley Coleman
- On consistency of Bayesian variable selection procedures
Elias Moreno, Javier Giron, George Casella, Lina Martínez, F. Vazquez-Polo, Maria Martel
- Dynamic Classification Trees for imprecise data
Massimo Aria, Valentina Cozza
- An innovative procedure for smoothing parameter selection
Gianluca Frasso, Paul Eilers
- The Italian Statistical Institute Macroeconometric Model - MEMo-It
Fabio Bacchini, et al.
- A statistical overview of the economic situation in the euro area
Gian Luigi Mazzi, Filippo Moauro, Rosa Ruggeri Cannata
- M-Estimation of Shape Matrices under Incomplete and Serially Dependent Data **PAPER NOT AVAILABLE**
Gabriel Frahm
- Convergence of Depths and Depth-Trimmed Regions **PAPER NOT AVAILABLE**
Rainer Dyckerhoff
- On Robustifying Some Blind Source Separation Methods for Second Order Nonstationary Data **PAPER NOT AVAILABLE**
Klaus Nordhausen
- The impact of the three crises on health in Italy: evidence and lack of adequate information systems
Giovanni Fattore
- LISA map based on distances for functional data **PAPER NOT AVAILABLE**
Pedro Delicado, Sonia Broner
- Smoothing mortality risks in space and time using flexible models **PAPER NOT AVAILABLE**
Maria Dolores Ugarte, Toma Goicoa, Jaione Etxeberria, Ana F. Militino
- Economic recession and fertility in the developed world: Past evidence and recent trends **PAPER NOT AVAILABLE**
Tomas Sobotka
- Generalized boosted additive models
Sonia Amodio, Jacqueline Meulman
-

SOLICITED

- Which family model makes couples more happy - dual earner or male breadwinner ?
Anna Baranowska-Rataj, Anna Matysiak
- Socioeconomic determinants of persistence in poor subjective health
Paolo Li Donni, Daria Mendola
- Family structures and subjective wellbeing in Italy
Silvia Montecolle, Francesca Rinesi, Alessandra Tinto
- Identifiability of Discrete Graphical Models with Hidden Variables
Marco Valtorta, Elizabeth S. Allman, John A. Rhodes
- Identifiability of hierarchical loglinear models with one hidden binary variable
Barbara Vantaggi
- Binary models of marginal independence: a comparison of different approaches
Monia Lupparelli, Luca La Rocca
- Small Area Estimation with Uncertain Random Effects
Gauri Sankar Datta, Abhyuday Mandal, Anthony Wanjoya
- Interacting Multiple Try Algorithms
Roberto Casarin, Radu Craiu, Fabrizio Leisen
- Higher-order asymptotics in Bayesian inference
Laura Ventura, Walter Racugno
- Online Detection of Outliers and Structural Breaks
Giovanni Petris
- Estimating the prevalence of cancer patients who have recurred
Angela B. Mariotto, Roberta De Angelis, Lucia Martina
- Multivariate Permutation Test to Compare Survival Curves for Matched Data
Stefania Galimberti
- Patterns of care and related costs of cancer prevalent cases by phase of disease
Silvia Francisci, Anna Gigli
- Estimating the incidence of cancer disease using a Bayesian backcalculation approach
Leonardo Ventura
- Bayesian T-optimal designs by simulation: a case-study on model discrimination
Rossella Berni, Federico M Stefanini
- From Markov moves in contingency tables to linear model estimability
Roberto Fontana, Fabio Rapallo, Maria Piera Rogantin
- Sensitivity Analysis and FANOVA Graphs for Computer Experiments
Jana Fruth, Sonja Kuhnt
- Factorial Graphical Lasso and Slowly Changing Graphical Models for Estimating Dynamic Networks
Antonino Abbruzzo, Ernst Wit
- New Statistics for Estimating the Parameter of the Stochastic Actor-Oriented Model
Viviana Amati
- Graph embedding via dissimilarity mapping for network comparison
Domenico De Stefano

Statistical models for virtual water network analysis
Alessandra Petrucci, Emilia Rocco

Latent Class CUB Models
Leonardo Grilli, Maria Iannario, Domenico Piccolo, Carla Rampichini

Formative and reflective models to determine latent construct
Anna Simonetto

Log-ratios analysis to study the relative information of ordinal variables
Michele Gallo

Ordinal Models for Financial Evaluation
Paola Cerchiello

Bayesian model averaging for financial evaluation
Silvia Figini

Labour market response models for university evaluation
Daniele Checchi, Silvia Salini

A Statistical Framework to Measure Reputation Risk
Tiziano Bellini, Luigi Grossi

The integration of administrative data to analyse the business economic performance: methodological aspects and results of a study
Fulvia Cerroni, Viviana De Giorgi, Marianna Mantuano

The micro economics of trade patterns and firm performances
Giovanni Dosi, Marco Grazzi, Federico Tamagni, Chiara Tomasi

The post-entry effect of exporting on productivity: inference on the counterfactual distribution
Maria Ferrante, Marzia Freo, A. Viviani

Data Integration and Productivity Estimation at a Firm Level
Filippo Oropallo, Stefania Rossetti

A PLS algorithm version working with ordinal variables
Giuseppe Boari, Gabriele Cantaluppi

Bivariate logistic models for the analysis of the Students University "Success"
Marco Enea, Massimo Attanasio

University admission test and students' careers: an analysis through a regression chain graph with a hurdle model for the credits
Leonardo Grilli, Carla Rampichini, Roberta Varriale

Comparing degree programs using unadjusted performance indicators. Assessing the bias from the Potential Confounding Factors
Mariano Porcu, Isabella Sulis

University of Pisa and academic performance: a sample survey on students with no exams in 2011
Lucio Masserini, Monica Pratesi

Assessing the Impact of Financial Aids to Firms: Causal Inference in the presence of Interference
Bruno Arpino, Alessandra Mattei

Inverse probability weighting to estimate causal effects of sequential treatments: a latent class extension to deal with unobserved confounding
Francesco Bartolucci, Leonardo Grilli, Luca Pieroni

A two-part geoaddivitive model for geographical domain estimation
Chiara Bocci, Alessandra Petrucci, Emilia Rocco

Application of Marginal Structural Models in Chronic Kidney Disease (CKD) Epidemiology: practical implementation in the Swedish National CKD Registry
Elena Pasquali, Marie Evans, Juan Jesus Carrero, Rino Bellocco

A Dimension Reduction Method for Approximating Integrals in Latent Variable Models for Binary Data
Silvia Bianconcini, Silvia Cagnone, Dimitris Rizopoulos

Kalman Filter for Maximum Likelihood Estimation of GMRFs
Luigi Ippoliti, Luca Romagnoli

Monte Carlo Likelihood Inference in Multivariate Model-Based Geostatistics
Marco Minozzo, Clarissa Ferrari

Statistical Modelling of Spatial Extremes
Anthony Davison, Simone Padoan, Mathieu Ribatet

Nonparametric smoothing of circular data
Agnese Panzera, Charles C. Taylor

Inverse Batschelet Distributions as Models for Circular Data
Arthur Pewsey

Depth Analysis of Directional Data
Claudio Agostinelli, Mario Romanazzi

Simulation of random rotation matrices
John T. Kent, Asaad M. Ganeiber

Dynamically modelling of fuzzy sets for flexible data retrieval
Miroslav Hudec

Factor PD-Co-clustering in Official Statistics
Marina Marino, Germana Scepi, Cristina Tortora

Extracting meta-information by using Network Analysis tools
Agnieszka Stawinoga, Maria Spano, Nicole Triunfo

How the text mining measures complex phenomena in official statistics
Sergio Bolasco, Pasquale Pavone

Assessing assumptions for data fusion procedures
Alfonso Piscitelli, Antonio D'Ambrosio

Filling in long gap sequences by performing EOF and FDA jointly
Antonella Plaia, Francesca Di Salvo, Marianonietta Ruggeri, Gianna Agrò

Missing Data Imputation within the Statistical learning Paradigm
Antonio D'Ambrosio

The use of administrative registers in the 2011 Census in Germany **PAPER NOT AVAILABLE**
Stephanie Hirner

Robust methods for correction and control of Italian Agriculture Census data **PAPER NOT AVAILABLE**
Alessandra Reale, Francesca Torti, Marco Riani

The industry and services continuous census" based on administrative sources: opportunities and problems **PAPER NOT AVAILABLE**

Manlio Calzaroni, Caterina Viviano

Using coarse resolution satellite images for crop area estimation: benchmarking their efficiency

Javier Gallego, Mohamed El-Aydam

How to select sample sites onto a study area?

Lucio Barabesi, S. Franceschi, M. Marcheselli

Do personal characteristics affect the Rasch measures of perceived physical risk? A quantile regression approach.

Fabio Aiello, Giovanni Boscaino, Monica Mandalà

Modeling nonignorable missingness in multidimensional latent class IRT models

Silvia Bacci, Francesco Bartolucci, Bruno Bertaccini

Risk profile versus portfolio selection: a case study

Valeria Caviezel, Sergio Ortobelli Lozza, Lucio Bertoli Barsotti

Cycles Syllogisms and Semantics: Examining the Idea of Spurious Cycles

Stephen Pollock

Spectral filtering for trend estimation

Marco Donatelli, Alessandra Luati, Andrea Martinelli

Robust estimation for multivariate data under the independent contamination model

Claudio Agostinelli, R. A. Maronna, V. J. Yohai

Adaptive robust location-scale estimation

Pietro Coretto

Minimum Volume Peeling: a Multivariate Mode Estimator

Giovanni Porzio, Giancarlo Ragozini, Steffen Liebscher, Thomas Kirschstein

A comparison of robust methods with small sample experimental data

Marco Riani, Andrea Cerioli, Maria Adele Milioli, Gianluca Morelli

Patterns of Mortality Decline and Individual Ageing: An Overview

Elisabetta Barbi

Survival predictive models in centenarians **PAPER NOT AVAILABLE**

Rossella Miglio, Paola Gueresi

Health status in over 85 years old living in Residential Facilities in Italy **PAPER NOT AVAILABLE**

Giulia Cavrini, Claudia Di Priamo, Lorella Sicuro, Alessandra Battisti, Alessandro Solicapa, Giovanni de Girolamo

CONTRIBUTED

Variation in Obstetric Intervention Rates across Hospitals in Sardinia

Massimo Cannas, Emiliano Sironi

On the role of normalized inverse-Gaussian priors in continuous-time models

Matteo Ruggiero

Randomly Reinforced Urn Designs whose Allocation Proportions Converge to Arbitrary Prespecified Values

Giacomo Aletti, Andrea Ghiglietti, Anna Maria Paganoni

Calibration estimation in dual frame surveys

Maria Giovanna Ranalli, Annalisa Teodoro

Comparing model-assisted estimators of structural variables in forest surveys

Ivan Arcangelo Sciascia, Matteo Garbarino, Giorgio Vacchiano, Renzo Motta

A study in panel cointegration and poolability: Long-run money demand equations for Gulf Cooperation Council countries

Stefano Fachin

Uncertainty in statistical matching for discrete categorical variables

Pier Luigi Conti, Daniela Marella, Mauro Scanu

Independent Component Analysis of Milan Mobile Network Data

Paolo Zanini, Piercesare Secchi, Simone Vantini

An Objective Bayesian analysis of dichotomous sensitive data

Maria Maddalena Barbieri, Brunero Liseo

Ensuring comparability over time and between domains by means of complex sample techniques

Tiziana Tuoto, Francesca Inglese

Confidence intervals for the Berger & Boos' procedure in the 2x2 Binomial Trial

Enrico Ripamonti, Piero Quatto

Kalman Filter for Maximum Likelihood Estimation of GMRFs

Luigi Ippoliti, Luca Romagnoli

Bayesian inference for the multivariate skew-normal model: a Population Monte Carlo approach

Antonio Parisi, Brunero Liseo

Reconstructing a multinormal covariance matrix from its spherically truncated projection

Filippo Palombi, Simona Toti, Romina Filippini

Clustering of financial time series in extreme scenarios

Roberta Pappadà, Fabrizio Durante

Investments in Renewable energies: evidence from a panel of countries

Giuseppe Scandurra

A Topological Definition of Phase and Amplitude Variability of Functional Data

Simone Vantini

Nonparametric saddlepoint test and pairwise likelihood inference

Nicola Lunardon, Elvezio Ronchetti

On the stationarity of the Threshold Autoregressive process: the two regimes case

Marcella Niglio, Francesco Giordano, Cosimo Damiano Vitale

Filling in long gap sequences by performing EOF and FDA jointly

Francesca Di Salvo, Mariantonietta Ruggieri, Gianna Agro'

Parallel Adaptive Markov chain Monte Carlo with applications

Mauro Bernardi, Lea Petrella
Modern Bayesian Inference in Zero-Inflated Poisson Models
Erlis Ruli, Laura Ventura
A comparison of different procedures for combining high-dimensional multivariate volatility forecasts
Giuseppe Storti, Alessandra Amendola
Estimation of wind speed prediction intervals by multi-objective genetic algorithms and neural networks
Valeria Vitelli
On a predictive measure of discrepancy between classical and Bayesian estimators
Stefania Gubbiotti
A prediction error for a linear regression model with fuzzy random elements
Maria Brigida Ferraro
Some further results for the two-parameter Poisson-Dirichlet partition model
Annalisa Cerquetti
Matching immigrant and native workers: evidence from the recent downturn in Italy
Adriano Paggiaro
The analysis of firm demography: an approach based on micro-geographic data
Diego Giuliani, Simonetta Cozzi, Giuseppe Espa
Regression estimators for capture-recapture frequency data
Irene Rocchetti
Towards an integrated surveillance system of road accidents
Tiziana Tuoto, Silvia Bruzzone, Luca Valentino, Giordana Baldassarre, Nicoletta Cibella, Marilena Pappagallo
On the Design Based Inference for Continuous Spatial Populations
Giorgio Eduardo Montanari, Giuseppe Cicchitelli
A Critical Look at Compositional Analysis for Assessing Habitat Selection
Caterina Pisani
Small area estimation for panel data
Annamaria Bianchi
Interpreting Deviations from Long-run Parity in an I(2) Model
Giuliana Passamani
Contributions from income components to Zenga's point and synthetic inequality measures: an application to EU countries
Michele Zenga, Leo Pasquazzi
An income mobility measure based on Zenga's inequality index
Mauro Mussini
Equivalence scales, inflation, and PPP: a unique (and simple) approach to estimation
Gustavo De Santis
Sensitivity analysis on a Cellular Automata model for the diffusion of Pleural Mesothelioma
Claudia Furlan, Cinzia Mortarino
Reproducibility Probability Estimation and Testing for some common nonparametric tests
Daniele De Martini, Lucio De Capitani
Multilevel algorithmic models to measure item importance on latent variables' indicators
Marica Manisera, Marika Vezzoli
A multiple imputation procedure of censored values in family-based genetic association studies
Fabiola Del Greco M., Cristian Pattaro, Cosetta Minelli, Peter P. Pramstaller, John R Thompson
Ridge analysis through profile likelihoods
Valeria Sambucini, Ludovico Piccinato
Migrant students classroom allocation policy in Italian schools
Andrea Scagni
Testing Phase and Amplitude Variability in Functional Data Analysis: a Hierarchical Permutation Test Approach
Alessia Pini, Simone Vantini
A hierarchical bayesian model for modelling benthic macroinvertebrates densities in lagoons
Serena Arima, Alberto Basset, Giovanna Jona Lasinio, Alessio Pollice, Ilaria Rosati
Life-Course Transitions, Market Work and Domestic Work of Italian Couples
Antonino Di Pino
Adjusting Time Series of Possible Unequal Lengths
Ilaria Lucrezia Amerise, Agostino Tarsitano
Variable selection in competing risks model
Marialuisa Restaino, Alessandra Amendola
A Bayesian Semiparametric Fay-Herriot-type model for Small Area Estimation
Silvia Poletini
Assessing Multivariate Measurement Systems in Multisite Testing
Michele Scagliarini, Stefania Evangelisti
Predicting EQ-5D responses from SF-12: should we take into account dependence and ordering?
Caterina Conigliani, Andrea Tancredi, Andrea Manca
Large sample properties of Gibbs-type priors
Pierpaolo De Blasi, Antonio Lijoi, Igor Pruenster
Bayesian Unit Root Tests: a Monte Carlo Study
Margherita Gerolimetto, Isabella Procidano
Data fusion in pharmaceutical marketing: new perspective from administrative data.
Paolo Mariani
How internal and international migrations shape the age structure of the Italian regions, 1955-2008
Paola Di Giulio, Cecilia Reynaud, Luca Vergaglia
Bayesian modeling of presence-only data
Fabio Divino, Giovanna Jona Lasinio, Natalia Golini
An Evaluation of the Student Satisfaction based on CUB Models
Barbara Cafarelli
Limited Information Estimation Methods for Paired Comparison Data

Manuela Cattelan
 Closed Likelihood-Ratio Testing Procedures to Assess Similarity of Covariance Matrices
Francesca Greselin, Antonio Punzo

Fiducial Distributions for Real Exponential Families
Piero Veronese, Eugenio Melilli

Hedonic Indexes and GDP estimate in the USA
Gabriele Serafini

Some results on stochastic comparisons of ROC curves
Silvia Figini, Chiara Gigliarano, Pietro Muliere

Efficiency in the use of natural non-renewable resources from mining and quarrying in Italy. Time series analysis and Economy-wide Material Flows Accounts
Donatella Vignani

Marital Disruption and Subjective Well-being: Evidence from an Italian Panel Survey
Giulia Rivellini, Alessandro Rosina, Emiliano Sironi

The Decision Making Process of Leaving Home: A Longitudinal Analysis of Italian Women
Giulia Ferrari, Alessandro Rosina, Emiliano Sironi

Alternative Bayesian analysis of capture recapture data with behavioral effect modelling
Daniilo Alumni Fegatelli

PDE penalization for spatial fields smoothing
Laura Azzimonti, Maurizio Domanin, Laura Maria Sangalli, Piercesare Secchi

Multivariate Nonlinear Least Squares: Direct and Beauchamp and Cornell Methodologies
Renato Guseo, Cinzia Mortarino

Handling weak dependence structures with copulas
Enrico Foscolo, Fabrizio Durante

Bayesian nonparametric predictions for count time series
Luisa Bisaglia, Antonio Canale

How to integrate macro and micro perspectives: an example on Human Development and Multidimensional Poverty.
Silvia Terzi

Composite Indicator of Social Inclusion in the European Union
Erasmo Vassallo, Francesca Giambona

Depth measures for the study of real and simulated ECG signals
Francesca Ieva

Experimental design for the estimation of Rician-distributed intensity fields in MRI
Stefano Baraldo, Francesca Ieva, Luca Mainardi, Anna Maria Paganoni

A von Mises Markov random field model for the analysis of spatial circular data
Francesco Lagona

A Multivariate VEC-BEKK Model for Portfolio Selection
Andrea Federico Pierini, Alessia Naccarato

Combining the complete-data and nonresponse models for drawing imputations under MAR
Shahab Jolani, Stef van Buuren, Laurence E. Frank

A Well-being Index Based on the Weighted Product Method
Matteo Mazziotto, Adriano Pareto

A comparison of semiparametric density estimation methods for multivariate risk management
Marco Bee

Modelling poverty transitions in Luxembourg: true state dependence or heterogeneity?
Alessio Fusco, Nizamul Islam

Causal analysis of education and birth inequalities through a latent class SEM
Silvia Bacci, Francesco Bartolucci, Luca Pieroni

Prediction of nonstationary functional data: Universal Kriging in a Hilbert Space
Alessandra Menafoglio, Matilde Dalla Rosa, Piercesare Secchi

School tracking and equality of opportunity in a multilevel perspective
Isabella Romeo, Emanuela Raffinetti

Indexing the Worthiness of Social Agents
Giulio D'Epifanio

Asymptotic estimation of right and left kurtosis measures, with applications to finance
Anna Maria Fiori, Davide Beltrami

Ordinal Lorenz Regression with application in Customer Satisfaction Surveys
Emanuela Raffinetti

A computational method to estimate sparse multiple Gaussian graphical models
Rossella Onorati, Luigi Augugliaro, Angelo Marcello Mineo

Deterministic or stochastic seasonality in daily electricity prices?
Paolo Chirico

Social capital and its impact on poverty reduction: measurement issues in logitudinal and cross-country comparisons. The case of UE.
Isabella Santini, Anna De Pascale

The diffusion of nuclear energy in the developing countries
Alessandra Dalla Valle, Claudia Furlan

A model for the joint distribution of income and wealth
Markus Jantti, Eva Sierminska, Philippe Van Kerm

Mothers with children aged 0-2 years: work/family reconciliation and support networks
Cinzia Castagnaro, Alessandra Fasano, Antonella Guarneri

A novel method for spatial smoothing
Laura M. Sangalli, James O. Ramsay

Poverty transitions in Italy
Lucia Coppola, Davide Di Laurea, Daniela Lo Castro, Mattia Spaziani

Spatial smoothing over non-planar domains
Bree Ettinger, Simona Perotto, Laura M. Sangalli

Lattice Models for the analysis of Urban Crime

Enrico di Bella, Luca Persico, Lucia Leporatti
Family resources and cognitive decline among elderly in Italy
Stefano Mazzuco

The median of a set of histogram data
Lidia Rivoli, Rosanna Verde, Antonio Irpino

Estimating the Homeless Population through Indirect Sampling and Weight Sharing Method
Claudia De Vitiis

Rates for Bayesian estimation of location-scale mixtures of super-smooth densities
Catia Luisa Scricciolo

Data gathering for elusive population. The case of foreigners during the XV Italian Census. A focus on Prato
Linda Porciani

Immigrant entrepreneurship through the economic crisis in Italy
Benedetta Cassani, Cristina Giudici, Roberta Rizzi

International Mobility of University Students: the Italian case
Domenica Fioredistella Iezzi, Mario Mastrangelo, Scipione Sarlo

Chronological analysis of textual data and curve clustering: preliminary results based on wavelets
Matilde Trevisani, Arjuna Tuzzi

Exponential Random Graph Model for multivariate networks: an application in knowledge network analysis
Domenico De Stefano, Susanna Zaccarin

The Role of Social Capital in Preventing Irregular Work in Italian Regions
Maria Felice Arezzo

Bayesian model averaging for financial credit risk measurement
Silvia Figini

Considerations about the Quotient of two Correlated Normals
Angiola Pollastri, Vanda Tulli

Estimating Business Statistics from administrative data: a study on small and medium enterprises
Orietta Luzi, Giovanni Seri, Viviana De Giorgi, Giampiero Sesto

The Coverage Survey of the 6th Agricultural Census
Matteo Mazziotta, Antonella Bernardini, Loredana De Gaetano, Lorenzo Soriani

A Local Price Observatory – Price minimarket: innovations and additional knowledge about prices - The experience of Umbria
Cristina Carbonari, Angiona Sabrina, Paradisi Francesca

Frailty Multi-State Models based on Maximum Penalized Partial Likelihood
Federico Rotolo, Catherine Legrand

Reconciliation of Time Series according to a Growth Rates Preservation Principle
Tommaso Di Fonzo

A decision support system for duopolies with incomplete information
Paola Vicard, Julia Mortera

False discovery rate control and the dependence structure of test statistics
Claudio Lupi

Neural Network Approach Applied for Classification in Business and Trade Statistics
Jana Juriová

The analysis of the material deprivation of foreigners in Italy
Anna Maria Milioto, Annalisa Busetta, Antonino Mario Oliveri

Effective Facebook population: the Italian case
Cristiano Tessitore, Ester Macri

A Clustream strategy for Functional Boxplots on multiple streaming time series
Antonio Balzanella, Elvira Romano

New approach to the identification of the Inverse Weibull model
Biagio Palumbo, Giuliana Pallotta

Stochastic Frontiers Approach: an Empirical Analysis of Italian Environmental Spending
Sabrina Auci, Annalisa Castelli, Donatella Vignani

Estimating student learning value-added models from repeated cross-sections
Dalit Contini

Intergenerational Mobility and Gender Gap: Evidence from Mediterranean Countries
Rosalina Castellano, Gennaro Punzo, Antonella Rocca

The role of Istat territorial offices for data quality control in the 15th Population and Housing Census. The case of Tuscany
Alessandro Valentini, Sabina Giampaolo

Longitudinal patterns of financial product ownership: a latent growth mixture approach
Francesca Bassi

Measuring job quality: a composite indicator
Giovanna Boccuzzo, Martina Gianecchini

On Measuring Inequity in Taxation Between Groups of Tax Payers
Achille Vernizzi, Simone Pellegrino, Maria Giovanna Monti

Capital income inequality: evidences from ECHP data
Francesca Greselin, Leo Pasquazzi, Ricardas Zitkis

Note on a new generalization of the skew-normal distribution
Valentina Mameli, Monica Musio

Estimates of Foreign Trade Using Genetic Programming
Miroslav Klucik

Machine learning techniques for Propensity score matching with clustered data. A simulation study.
Massimo Cannas, Bruno Arpino, Francesco Billari

Impact of Audio Tools in Web Surveys
Daniele Toninelli, Silvia Biffignandi

Early-life circumstances and late-life income
Omar Paccagnella, Christelle Garrouste

A Simple Risk-Adjusted CUSUM chart for monitoring binary health data

- Marco Marchi*
From theory to practice: a methodological proposal for operationalising and summarizing the concept of quality of work
Marco Centra, Maurizio Curtarelli, Valentina Gualtieri
- Partners' income and decision making
Lucia Coppola, Domenica Quartuccio
- Dealing With a Potential Bias in Estimating the Share of Discriminated Women
Rosa Giaimo, Giovanni Luca Lo Magno
- Border surveys and Time Location Sampling (TLS): an application on incoming tourism in Sicily
Stefano De Cantis, Mauro Ferrante
- The Use of Administrative Data for Short Term Business Statistics: Lessons from a Cross-Country Experience
Ciro Baldi, Francesca Ceccato, Silvia Pacini, Donatella Tuzi
- Price transmission and market power in the food market
Maria Caterina Bramati
- Data imputation processes based on statistical analysis: the case of Kosovo census data
Marco Scarnò, Bekim Canolli, Servete Muriqi, Hisni Ferizi
- Dimension reduction for measuring the multidimensional demographic convergence
Maria Rita Sebastiani
- Is financial fragility a matter of illiquidity? An appraisal for Italian households
Marianna Brunetti
- Burnout, learning and self-esteem at school: an empirical study.
Cristiana Ceccatelli
- Autocorrelated non-normal data in control charts
Claudio Giovanni Borroni, Manuela Cazzaro, Paola Maddalena Chiodini
- Social welfare orderings of the Generalized-Lorenz Type: applications of an extended equivalence theorem
Alessandra Giovagnoli
- Timely Indices for Residential Construction Sector
Attilio Gardini, Enrico Foscolo
- Dimensions of well-being and their statistical measurements
Carla Ferrara, Francesca Martella, Maurizio Vichi
- The diagnostics of the mean squared error of the Eblup in small area estimation models
Renato Salvatore, Maria Chiara Pagliarella

On a predictive measure of discrepancy between classical and Bayesian estimators

Pierpaolo Brutti, Fulvio De Santis and Stefania Gubbiotti

Abstract In the presence of prior information on an unknown parameter of a statistical model, Bayesian and frequentist estimates based on the same observed data do not coincide. However it is well known that, in many standard parametric problems, their discrepancy tends to be reduced as the sample size increases. In this paper we consider a measure of discrepancy, D_n , between a frequentist and a Bayesian point estimator and we study its predictive distribution. In some specific examples we analyze the main characteristics of this predictive distribution for increasing sample sizes. We also consider the use of the predictive density of D_n for the assessment of a prior distribution informativeness. Some explicit results are given for the normal model.

Key words: Clinical trials; Conflict between estimators; Predictive approach.

1 Introduction

Bayesian methods offer the theoretical framework for combining experimental data and pre-experimental information on an unknown parameter, that is formalized by a prior probability distribution. In the presence of prior information, frequentist and Bayesian procedures, such as point or interval estimates based on the same observed sample, do not coincide. However, in many standard parametric problems, the discrepancy between frequentist and Bayesian procedures tends to disappear as the sample size increases. This is typically shown in most of introductory books on Bayesian inference (see, among others, [1]).

Let us consider the estimation problem for the expected value θ of a normal distribution. Given n observations from i.i.d. normal random variables, the standard

Pierpaolo Brutti e-mail: pierpaolo.brutti@uniroma1.it · Fulvio De Santis e-mail: fulvio.desantis@uniroma1.it · Stefania Gubbiotti e-mail: stefania.gubbiotti@uniroma1.it
Sapienza Università di Roma, Dipartimento di Scienze Statistiche

Bayesian estimate of θ is a linear combination of the sampling mean, \bar{x}_n , and of a prior guess on θ , μ_A , i.e. $\omega_n \bar{x}_n + (1 - \omega_n) \mu_A$, where ω_n tends to one as n diverges. Therefore, for a sufficiently large sample size, the sample mean provides a good approximation of the Bayesian estimate.

In this paper, we are interested in analysing a measure of discrepancy between two competing estimators. This measure is random before observing the data. In Section 2, we introduce a specific measure of discrepancy, D_n , between a frequentist and a Bayesian estimator and in Section 2.1 we derive its explicit expression, its predictive cumulative distribution function (cdf) and its expected value for the normal model with conjugate priors. After briefly discussing the asymptotic behaviour of D_n , in Section 2.2 we focus on a fixed sample size, optimal with respect to a given criterion, and we assess the informativeness of the prior by evaluating D_n .

2 A discrepancy between estimators

Let $\mathbf{X}_n = (X_1, X_2, \dots, X_n)$ be a random sample from a probability distribution $f_n(\cdot|\theta)$, where θ is an unknown real-valued parameter that belongs to the parameter space, $\Theta \subseteq \mathbb{R}$. Let $\mathbf{x}_n = (x_1, x_2, \dots, x_n)$ be an observed sample, $\pi_A(\cdot)$ the prior density function of θ , $f_n(\mathbf{x}_n|\theta)$ the likelihood function and $\pi_A(\theta|\mathbf{x}_n)$ the posterior distribution. We will refer to π_A as to the *analysis-prior*. It models pre-experimental knowledge/uncertainty on θ taken into account in posterior analysis.

We denote a Bayesian estimator of θ as $\hat{\theta}_B(\mathbf{X}_n)$, whereas $\hat{\theta}_F(\mathbf{X}_n)$ is a generic classical estimator. Let $D_n(\mathbf{X}_n)$ be a measure of discrepancy between $\hat{\theta}_B$ and $\hat{\theta}_F$. Specifically, we consider the standard squared difference between estimators, i.e.

$$D_n(\mathbf{X}_n) = [\hat{\theta}_B(\mathbf{X}_n) - \hat{\theta}_F(\mathbf{X}_n)]^2.$$

Before observing the data, $\hat{\theta}_B$, $\hat{\theta}_F$ and D_n are random variables (functions of \mathbf{X}_n). For instance, in the following we consider the posterior expectation of the parameter θ , $E(\theta|\mathbf{X}_n) = \int_{\Theta} \theta \pi(\theta|\mathbf{x}_n) d\theta$, as $\hat{\theta}_B$, and the maximum likelihood estimator (MLE) as $\hat{\theta}_F$, although the methodology could be extended to other estimators. We want to evaluate the probability of observing a small/large discrepancy between $\hat{\theta}_B$ and $\hat{\theta}_F$. As for any other pre-posterior Bayesian analysis, to this purpose two alternative distributions for the data can be used. The *conditional* approach prescribes the use of the sampling distribution $f_n(\cdot|\theta)$, with $\theta = \mu_D$, a “design value” for the unknown parameter, whereas the *predictive* approach implies the use of the predictive distribution $m_D(\mathbf{x}_n) = \int_{\Theta} f_n(\mathbf{x}_n|\theta) \pi_D(\theta) d\theta$, where π_D (*design-prior*) is a density function that accounts for uncertainty on the design value of θ . Note that the conditional approach is a special case of the predictive one, when π_D is a point-mass prior on μ_D ; for this reason in the following we adopt the most general approach. See [4] and [2] for a detailed discussion on this point.

2.1 Results for the normal model

Let \mathbf{X}_n be a random sample from a $N(\theta, \sigma^2)$ distribution. The MLE of θ is $\hat{\theta}_F = \bar{x}_n$. Assume for θ a normal prior density $\pi_A(\theta) = N(\theta|\mu_A, \sigma^2/n_A)$, where n_A is given the standard interpretation of *prior sample size*. From standard results on conjugate analysis [1], the posterior distribution of θ is a normal density of mean $(n+n_A)^{-1}(n\bar{x}_n + n_A\mu_A)$ and variance $(n+n_A)^{-1}\sigma^2$. Hence $\hat{\theta}_B = \omega_n\bar{x}_n + (1-\omega_n)\mu_A$, with $\omega_n = n/(n+n_A)^{-1}$.

Letting $\pi_D(\theta) = N(\theta|\mu_D, \sigma^2/n_D)$, the predictive density function of \bar{x}_n is $m_D(\bar{x}_n) = N(\bar{x}_n|\mu_D, \psi_n^2)$, where $\psi_n^2 = b_n\sigma^2$ and $b_n = (n+n_D)(nn_D)^{-1}$. Given the above assumptions, letting $a_n = 1 - \omega_n$ the explicit expression of D_n is as follows

$$D_n = a_n^2(\bar{X}_n - \mu_A)^2.$$

It is straightforward to check that the predictive expected value of D_n is

$$e_n = a_n^2[b_n\sigma^2 + \delta^2],$$

where $\delta = \mu_A - \mu_D$, while the cdf of D_n is

$$p_n(d) = \Phi\left[b_n^{-1/2}(\delta + a_n^{-1}d^{1/2})\sigma^{-1}\right] - \Phi\left[b_n^{-1/2}(\delta - a_n^{-1}d^{1/2})\sigma^{-1}\right],$$

where $\Phi(\cdot)$ is the standard Normal cdf.

Noting that $b_n = O(1)$ and $a_n = o(n^{-1})$, it follows that $e_n = o(n^{-2})$ and that, as n diverges, D_n converges in probability to zero as fast as n^{-2} . Both e_n and p_n depend on the prior means only through the absolute difference $|\delta|$.

2.2 Quantifying the informativeness of the prior

Let us consider the set up of an efficacy clinical trial, with positive values of θ indicating an effective treatment. Let us suppose for instance to select the minimum sample size n^* such that the frequentist conditional power reaches a desired level (see among others [3]). For a normal model, under the assumptions of the previous section the power is $\beta = \Phi\left(\frac{\theta\sqrt{n}}{\sigma} + z_{\alpha/2}\right)$, where z_α denotes the quantile of a standard normal at level α . If we set for instance a design value for θ equal to 0.5, when $\sigma = 2$ and $\alpha = 0.05$, the optimal sample size required to reach a 0.80 power is $n^* = 126$. Let us assume a design prior of parameters $\mu_D = 0.5, n_D = 20$. Based on n^* , we can compute e_{n^*} for a given analysis prior. The predictive expected discrepancy thus provides a measure of the conflict between the two alternative estimators and, at the same time, it represents the level of informativeness of the prior. In fact, the larger e_{n^*} , the stronger the impact of the prior in $\hat{\theta}_B$. In this way it is also possible to compare different choices for the analysis prior in terms of their informativeness level. In Table 1a we report the values of e_{n^*} for several choices of

the analysis prior parameters. For illustrative purposes we may consider a threshold value such as $d = 0.2$ (values exceeding 0.2 are bolded in the table): when a e_{n^*} is below this threshold the corresponding analysis prior can be considered relatively non informative, otherwise its impact on the Bayes estimator is remarkable. As expected, for increasing values of n_A , the analysis prior becomes more informative and, consequently, the expected discrepancy is larger. The increments of the conflict measure appear to be smaller when the analysis prior mean coincides with θ_D . Similar considerations can be drawn by computing $p_{n^*}(d)$, but in this case a threshold on a probability scale can be set: for instance in Table 1b the values of $p_{n^*}(d)$ below a given level, say 0.5, identify the analysis prior parameters with stronger impact on the Bayesian estimator.

Table 1 e_{n^*} and p_{n^*} for several choices of the analysis prior, with $n^* = 126$.

μ_A	n_A					
	1	10	20	50	100	200
(a)						
-2	0	0.035	0.122	0.523	1.269	2.440
-1	0	0.013	0.047	0.200	0.486	0.934
0	0	0.003	0.009	0.039	0.094	0.181
0.5	0	0.001	0.004	0.019	0.045	0.087
(b)						
-2	1	1	0.944	0.027	0.001	0.000
-1	1	1	1	0.561	0.155	0.055
0	1	1	1	0.987	0.855	0.677
0.5	1	1	1	0.999	0.964	0.870

• $\mu_D = 0.5, n_D = 20$

In summary, in this work we have introduced a measure of conflict between classical and Bayesian point estimators. A similar methodology could be considered extending the idea of discrepancy both to different objects to be compared (such as interval estimators, Bayes factors, etc.) and to more complex models.

References

1. Bernardo, J.M. and Smith, A.F.M. (1994). Bayesian Theory. Wiley.
2. De Santis, F. (2006). Sample size determination for robust Bayesian analysis. Journal of the American Statistical Association, 101, n. 473, 278-291.
3. Spiegelhalter, D.J, Abrams, K.R. and Myles, J.P. (2004). Bayesian approaches to clinical trials and health-care evaluation. Wiley.
4. Wang, F., and Gelfand, A.E. (2002). A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models. Statistical Science, 17, n. 2, 193-208.