BOOK OF ABSTRACTS

Polytechnic Institute of Tomar
Tomar - PORTUGAL
May 26–28, 2016

Committees

Executive Committee

Teresa Oliveira (Portugal)
Stanislaw Mejza (Poland)
Christos Kitsos (Greece)
Luís Miguel Grilo (Portugal) – Local Chair

Scientific Committee

Adérito Marcos (Universidade Aberta, Portugal)
Amílcar Oliveira (Universidade Aberta, Portugal)
Ana Maria Porto (UNED, Spain)
António Araújo (Universidade Aberta, Portugal)
António Costa (UNED, Spain)
Arminda Manuela Gonçalves (Universidade do Minho, Portugal)
Augustyn Markiewicz (Poznan University of Life Sciences, Poland)
Barbara Kimmel (University of Texas School of Public Health, USA)
Carlos Agra Coelho (Universidade Nova de Lisboa, Portugal)
Carlos Aldemir Farias da Silva (Universidade Federal do Pará – UFPA, Brazil)
Célia Nunes (Universidade da Beira Interior, Portugal)
Christos Kitsos (Technological Educacional Institution of Athens, Greece)
Christos Skiadas (Technical University of Crete, Greece)
Dário Ferreira (Universidade da Beira Interior, Portugal)
Dinis Pestana (Faculdade de Ciências da Universidade de Lisboa, Portugal)
Edwirde Luiz Silva (Universidade Estadual da Paraíba, Brazil)
Elisa Henning (Universidade do Estado de Santa Catarina – UDESC, Joinville – Brazil)
Elisabete Carolino (Escola Superior de Tecnologias da Saúde de Lisboa, Portugal)
Fernanda Figueiredo (Universidade do Porto, Portugal)
Fernando Caetano (Universidade Aberta, Portugal)
Fernando Carapau (Universidade de Évora, Portugal)
Fernando Costa (Universidade Aberta, Portugal)
Iran Abreu Mendes (Universidade Federal do Rio Grande do Norte – UFRN, Brazil)
Isabel Silva (Universidade do Porto, Portugal)
Iwona Mejza (Poznan University of Life Sciences, Poland)
Jane A. Anderson (Baylor College of Medicine, Houston, Texas, USA)
João Araújo (Universidade Aberta, Portugal)
João Tiago Mexia (FCT, Universidade Nova de Lisboa, Portugal)
José António Pereira (Universidade do Porto, Portugal)
Luciane Mulazani (Univ. do Estado de Santa Catarina – UDESC, Joinville – Brazil)
Luís Gonzaga (Universidade Aberta, Portugal)

Luís Miguel Grilo (Instituto Politécnico de Tomar, Portugal)
M. Ivette Gomes (Faculdade de Ciências da Universidade de Lisboa, Portugal)
Manuela Neves (ISA, Universidade de Lisboa, Portugal)
Marek Kimmel (Rice University, Houston, Texas, USA)
Maria do Rosário Almeida (Universidade Aberta, Portugal)
Maria do Rosário Ramos (Universidade Aberta, Portugal)
Mei-Ling Ting Lee (University of Maryland, USA)
Milan Stehlík (Johannes Kepler University in Linz, Austria)
Pawel Bartoszczuk (Warsaw School of Economics, Poland)
Pedro Serranho (Universidade Aberta, Portugal)
Robert G. Aykroyd (University of Leeds, UK)
Roberval Lima (Universidade Paulista e Universidade Federal do Amazonas, Brazil)
Sandra Ferreira (Universidade de Beira Interior, Portugal)
Sandra Nunes (Instituto Politécnico de Setúbal, Portugal)
Sónia Seixas (Universidade Aberta, Portugal)
Stanislaw Mejza (Poznan University of Life Sciences, Poland)
Teresa Oliveira (Universidade Aberta, Portugal)
Ulisses Azeiteiro (Universidade Aberta, Portugal)
Victor Leiva (Universidad Adolfo Ibáñez, Chile)
Zbigniew Krysiak (Warsaw School of Economics, Poland)
Zuhal Dogan (Nature School EU Project Coordinator, Turkey)

Honour Committee

Rector of Universidade Aberta
President of Instituto Politécnico de Tomar
President of the Polish Biometric Society
President of CRA - International Statistical Institute
President of International Chinese Statistical Association

Organizing Committee

Luís Miguel Grilo (Instituto Politécnico de Tomar, Portugal) – Local Chair
Amílcar Oliveira (Universidade Aberta, Portugal)
Ana Nata (Instituto Politécnico de Tomar, Portugal)
Célia Nunes (Universidade da Beira Interior, Covilhã, Portugal)
Célio Gonçalo Marques (Instituto Politécnico de Tomar, Portugal)
Dário Ferreira (Universidade da Beira Interior, Portugal)
Fernando Carapau (Universidade de Évora, Portugal)
Maria Isabel Pitacas (Instituto Poliécnico de Tomar, Portugal)
Maria João Inácio (Instituto Politécnico de Tomar, Portugal)
Maria do Rosário Ramos (Universidade Aberta, Portugal)
Sandra Ferreira (Universidade da Beira Interior, Portugal)
Teresa Oliveira (Universidade Aberta, Portugal)

Sponsored by

Title

10th Workshop on Statistics, Mathematics and Computation and 5th Portuguese–
Polish Workshop on Biometry – Book of Abstracts

Editor

Authors

Teresa Oliveira, Stanislaw Mejza, Christos Kitsos, Amílcar Oliveira, Ana
Nata and Luís M. Grilo.

Published and printed by:

Preface

Dear participants, colleagues and friends,

WELCOME to the 10th Workshop on Statistics, Mathematics and Computation - WSMC10, and to the 5th Portuguese-Polish Workshop in Biometry - PPWB5. We are delighted to celebrate a decade of strong interaction between researchers from Portugal and abroad, leading to a successful commitment and enthusiasm on promoting research in the broad areas of Statistics, Mathematics and Computation, with special emphasis to Biometry.

It is a great pleasure to receive all our guests and contributors from 26th to 28th May at the Polytechnic Institute of Tomar - IPT, expressing our gratitude to the Institute Coordinators for kindly accepted to embrace this challenge of receiving us again!

WSMC is a platform to exchange knowledge through international contacts involving Statisticians, Mathematicians, Computation professionals and Students, coming from several Universities, Organizations and other Institutions around the World. Our main objective is to assemble researchers and practitioners involved in these fields, and to encourage mutual exchange and interactions among a broad range of Science areas.

PPWB has been promoted every two years, one time in Portugal next in Poland, taking into account the many joint collaborations between Portuguese and Polish researchers, who felt the need of a platform for sharing common interests on the research of Statistical and Biometrical issues.

The format of this joint meeting involves plenary talks, organized sessions, posters and mini-courses. Special attention is given to applications and to new theoretical results having potential of solving real life problems. Discussion will be open in several areas, mainly with regard to challenges in Life Sciences, Medicine, Health Sciences, Education, Management, Insurance and Industry including some topics on Risk Analysis. It is also intended to emphasize the advantages of online learning and teaching, as well as the respective exploitation by using e-textbooks and scientific e-materials, known as the high strand for technology in the current society.

Participants will have several journal opportunities for papers submission. Selected papers will appear in editions of Biometrical Letters, Discussiones Mathematicae and of the Journal of Applied Statistics. A Springer

volume of the Series "Contributions to Statistics" for selected papers is also under preparation.

We are really grateful to all the participants, Invited Speakers, Session organizers and Authors who submitted abstracts, for their valuable contribution and for the enthusiastic way how they assume their participation. We also aknowledge all the sponsors and contributors who made this meeting a reality.

Furthermore, we acknowledge our Honour Committee: Rector of Universidade Aberta, Rector of Universidade da Beira Interior, President of Instituto Politécnico de Tomar, Coordinator of Centro de Estatística e Aplicações da Universidade de Lisboa, President of the Polish Biometric Society, President of the Committee on Risk Analysis of International Statistical Institute and President of Chinese Statistical Association, for their precious support.

Many people and organizations contributed to the planning and execution of this joint meeting, and we are deeply grateful to all for that. However it is not an easy task to enumerate all the people who should be thanked. Nevertheless, there are a few groups and organizations that we would like to single out for special thanks.

We are most grateful to the IPT for hosting the meeting and to all the members of the Organizing Committee and of the Scientific Committee for their crucial help and suggestions. We wish to deeply thank to Luís Grilo, who was responsible for the meetings logistics at IPT and for organizing the social programme.

Last but not least, special thanks are addressed to Amílcar Oliveira, Ana Nata, Maria João Inácio and Luís Grilo for the invaluable contribution on organizing the Webpage, the final programme and the Book of Abstracts details in an incredible short time.

We believe that this joint meeting will be rewarding to all of us and that it will draw us to the next edition -WSMC11, which is planned for 2017.

We wish you a very productive, successful, and enjoyable joint meeting, as well as a very nice stay in Tomar, please try also to find time to enjoy the delights of Portugal!!!

The Executive Committee (EC) - Chairs

Teresa Oliveira (Portugal)	Stanislaw Mejza (Poland)
Christos Kitsos (Greece)	Luís Grilo (Portugal) - Local Chair

Programme

Thursday, MAY 26th

08:30 − 11:00 Registration Desk at the IPT [Room O103]

09:00 − 09:20 Opening Ceremony [Room O106]
09:20 − 09:50 Plenary Session [Room O106]

Plenary Session 1	Room O106
	Chairman: Stanislaw Mejza

09:20	**Choosing the Best Predictive Logit Model**
	Christos Kitsos

09:50 − 10:10 Coffee break [Room O103]

10:10 − 11:30 Plenary Sessions [Rooms O104/O105/O106]

Organized Session 1	Room O106
Computational Mathematics and Applications in Sciences and Engineering	
	Chairman: **Filomena Teodoro**

10:10	**MANOVA for the analyses of the effect of the processing conditions on the dispersion of layered clays in PP polymer**
	Fátima de Almeida
10:30	**Text mining algorithms: a survey on the statistical intuition**
	Anacleto Correia
10:50	**Statistical characterization of noise in different industrial workstations**
	Aldina Correia
11:10	**Evaluation of GPS Static Relative Positioning**
	Filomena Teodoro

Organized Session 2   **Room O105**

Practical applications of multivariate time series

analysis and longitudinal models

Chairman: **Eliana Costa e Silva**

10:10   **Forecasting the number of overnight stays in**

         Algarve

         Isabel Lopes

10:30   **A joint model of breast cancer with spatial efects**

         Ana Borges

10:50   **Joint modelling of longitudinal CEA tumour marker**
         progression and survival data on breast cancer

         Ana Borges

11:10   **Causal influence between times series: a study**
         on imports, exports and economic growth

         Eliana Costa e Silva

Organized Session 3   **Room O104**

Numerical Methods in Fluids and Computation

Chairman: **Fernando Carapau**

10:10   **Some results on the Frobenius coin problem**

         Manuel Branco

10:30   **Animal movement: a simple chaotic model approach**

         Carlos Correia Ramos

10:50   **High Weissenberg number problem: state of art**

         Marília Pires

11:10   **A one-dimensional model for blood flow based on**

         Cosserat Theory

         Paulo Correia

11:30 − 18:30 Cultural Tour

20:00 − 20:30 Cocktail at the "Hotel dos Templários"

         Lobby bar, with piano

20:30 − 22:00 Conference dinner at the "Hotel dos Templários"

Friday, MAY 27th

09:00 − 12:00 Registration Desk at the IPT [Room O103]
09:00 − 10:30 Plenary Sessions [Room O106]

Plenary Sessions 2,3,4	Room O106
	Chairman: **Christos Kitsos**

09:00	**Quasi-stochastic model of the process of recurrent metastasis**
	Marek Kimmel
09:30	**Analyzing data on protein content in two types of hybrid maize cultivars under three models with split units**
	Iwona Mejza
10:00	**Data Science with the R software and applications**
	Víctor Leiva

10:30 − 11:00 Coffee break + Poster Session [Room O103]

11:00 − 12:40 Parallel Sessions [Rooms O104/O105/O106/B260/B257]

Organized Session 4	Room O106
Statistics and Modelling	
	Chairman: **Milan Stehlík**

11:00	**Stochastic bivariate models for bio-medical phenomena with two distinct kinds of dependencies between underlying random quantities**
	Jerzy Filus
11:20	**Effects of violations of prerequisites in sequential designs on sample size and power**
	Karl Moder
11:40	**Impact of Education Support Software on EMM Course Student Achievement**
	Pavel Kolman
12:00	**Testing the difference between two quantiles from independent populations**
	Martina Mittlboeck
12:20	**On robust testing for normality in autoregressive conditional heteroscedasticity models**
	Luboš Střelec

Organized Session 5   **Room O105**

Statistical Quality Control: Theory and Practice

Chairman: **Fernanda Figueiredo**

11:00   **A nonparametric approach for simultaneous monitoring
          of event frequency and magnitude**

          Amitava Mukherjee

11:20   **The effect of using an incorrect model on the
          performance of risk-adjusted control charts**

          Maria João Inácio

11:40   **On the influence of control statistics in the performance
          of joint schemes for location and spread**

          Patrícia Ramos

12:00   **Joint Control Charts applied to monitoring the
          concentration of particles with health effects in
          occupational environment**

          Elisabete Carolino

12:20   **Acceptance sampling plans for sensory and
          chromatography analyses**

          Fernanda Figueiredo

Organized Session 6   **Room O104**

Stochastic processes with applications in System Biology,

Genetics and Epidemiology

Chairman: **Inés M. del Puerto**

11:00   **Stochastic epidemics with heterogeneous contacts in small
          networks**

          António Gómez-Corral

11:20   **Modeling stochastic introgression with branching processes**

          Maria Conceição Serra

11:40   **Mathematical modeling of the population dynamics of
          Black Vulture colonies through Branching Processes**

          Manuel Mota

12:00   **A two-sex branching process as model to explain the fate
          of a recessive allele linked to X chromosome**

          Miguel González

Organized Session 7   **Room B260**

Statistical Distributions and Applications

Chairman: **Filipe Marques**

11:00   **Improving the asymptotically unbiased extreme value index estimation**

Frederico Caeiro

11:20   **Pseudo Maximum Likelihood and Moments Estimators for some Ergodic Diffusions**

Pedro Mota

11:40   **Random polygonal lines iteratively generate on the plane**

Rui Santos

12:00   **Revisiting the choice of block size and the threshold in the extremal index estimation**

Dora Gomes

12:20   **Series representations for the sum of independent Gamma random variables and for the product of independent Beta random variables**

Filipe Marques

Organized Session 8   **Room B257**

Algebra, Mathematics and Computation

Chairman: **João Araújo**

11:00   **Comparing Several Tests of Randomness Based on the Difference of Observations**

Ayana Mateus

11:20   **Nontrivial Ground States for Cooperative Cubic Schrödinger Systems**

F. Oliveira

11:40   **Specht modules and symmetry classes of tensors**

Fátima Rodrigues

12:00   **The polynomial hierarchy of time**

Isabel Oitavem

12:20   **A complete rewriting system for the Plactic monoid using crystals**

António Malheiro

12:30 − 14:00 Lunch [IPT Restaurant]

14:00 − 14:30 Plenary Session [Room O106]

Plenary Session 5	**Room O106**

<div align="right">Chairman: Victor Leiva</div>

14:00	**Big data challenges for statistical approaches to 21st century medical imaging**
	Robert Aykroyd

14:30 − 15:50 Parallel Sessions [Rooms O104/O105/O106/B260]

Invited PPWB5 Session 1	**Room O106**

Applications of quantitative methods in economics and finance

<div align="right">Chairman: Stanislaw Mejza</div>

14:30	**Comparative analysis of classical and positional approaches to the synthetic evaluation of housing conditions in EU countries**
	Joanna Stanisławska
14:50	**The use of biplot in the analysis of poverty determinants in EU countries**
	Andrzej Wołoszyn
15:10	**Identification of spatial effects in the evaluation of socio-economic development of EU regions**
	Romana Głowicka-Wołoszyn
15:30	**Ecoinnovation efficiency**
	Paweł Bartoszczuk

Organized Session 9	**Room O105**

Matrices in Science and Engineering

<div align="right">Chairman: João Cardoso</div>

14:30	**On Some Problems of Analytic Geometry over the Octonions**
	Paulo Saraiva
14:50	**Spatio-temporal modelling of environmental monitoring data from different surveys**
	Luís Margalho
15:10	**Linear Algebra and Image Processing: a new teaching approach**
	Cristina Caridade
15:30	**Finding the Closest Generalized Essential Matrix**
	João Cardoso

Organized Session 10 **Room O104**

Statistical Models in Applied Sciences

Chairman: **Rosário Ramos**

14:30	**t-Student and Mann-Whitney tests applied to small sample data - inevitable but uninformative?**
	Francisco Caramelo
14:50	**Study of the differences in mortality and length of hospital stay in acute myocardial infarction at hospital level**
	Teresa Magalhães
15:10	**A comparative study of the estimators for the demand of Engineering Courses in Portugal**
	Raquel Oliveira
15:30	**At which sample size is Monte Carlo Cross-Validation reliable for dichotomous classifiers?**
	Miguel Patrício

Organized Session 11 **Room B260**

Statistical inference for discrete-valued time series

Chairman: **Isabel Silva**

14:30	**Simple integer-valued bilinear model: inference, prediction and diagnostic**
	Isabel Pereira
14:50	**Surveillance in Discrete Time Series**
	Maria da Conceição Costa
15:10	**On the maxima of integer models based on a new operator**
	Maria da Graça Temido
15:30	**Outliers detection in integer-valued time series**
	Isabel Silva

15:50 − 18:00 Visit to the "Convento de Cristo"

Saturday, MAY 28th

09:00 − 10:00 Plenary Sessions [Room O106]

Plenary Sessions 6,7	Room O106
	Chairman: **Amílcar Oliveira**
09:00	**On optimality of neighbor designs under mixed interference models**
	Augustyn Markiewicz
09:30	**Linear Sufficiency in Models with Orthogonal Block Structure**
	João Mexia

10:00 − 10:30 Special Invited Session [Room O106]

Round Table (Publications)	Room O106
	Chairman: **Adriana Aykroyd**
10:00	**Organized discussion on the aims of the Journal of Applied Statistics and of the Springer book**

10:30 − 10:45 Coffee break + Poster Session [Room O103]

10:45 − 12:45 Parallel Mini-courses[Rooms O106/O105]

Mini-course 1	Room O106
	Chairman: **Teresa Oliveira**
10:45	**Theories and Strategies of Goal Setting and Goal Attainment in Patients' Self-Management of Chronic Diseases: Practical Application of Chronic Disease Management Using Video-Technology**
	Barbara Kimmel, Jane Anderson, Guilherme Oliveira and Cristina Lopes

Mini-courses 2,3	Room O105

<div style="text-align: right">Chairman: **Rosário Ramos**</div>

10:45	**Ruler, Compass and Nail: 360 degree Spherical Perspective made simple**
	António Araújo
11:45	**Creativity and mathematics creation in history**
	Iran Mendes

12:45 − 14:15 Lunch [Lodge Restaurant]

14:15 − 15:45 Plenary Sessions [Room O106]

Plenary Sessions 8,9,10	Room O106

<div style="text-align: right">Chairman: **Luís Grilo**</div>

14:15	**Statistical and probabilistic modeling for cancer risks: mammary cancer and nephroblastoma**
	Milan Stehlík
14:45	**Risk return profile in the hedge fund industry**
	Isabela Pruchnicka-Grabias
15:15	**Tests for equality of mean vectors with structured covariance matrices**
	Carlos A. Coelho

14:45 − 16:10 Coffee break [Room O103]

16:10 − 17:50 Parallel Sessions [Rooms O104/O105/O106/B260]

Invited PPWB5 Session 2	Room O106
Statistical Modelling and Applications	

<div style="text-align: right">Chairman: **Teresa Oliveira**</div>

16:10	**A general method for bivariate stochastic models construction**
	Jerzy Filus
16:30	**Long memory stochastic processes construction**
	Lidia Filus
16:50	**The uniform distribution product as approach to the (Q,r) inventory model using R software**
	J. Seijas-Macias

17:10 **Combination of Statistical and Clinical Methods to Assess the Risk of Tooth Loss**

José Pereira

17:30 **Recent developments on optimal designs and their applications in inference**

Milan Stehlík

Organized Session 12   **Room O105**

Computational Data Analysis and Numerical Methods

Chairman: **Luís Grilo**

16:10 **Multivariate analysis applied to school achievement in mathematics in the 3rd cycle - a case study**

Maria da Luz Soares

16:30 **Shapes and computer generation of the numerical range of complex matrices on a Hilbert space**

Ana Nata

16:50 **Misclassification and asymmetry in binary regression model**

Martin Jacinto

17:10 **Non-parametric Individual Control Chart for Silica in Water**

Luís Grilo

Organized Session 13   **Room O104**

Statistics and Applications in Life Sciences

Chairman: **Manuela Neves**

16:10 **Spatio-temporal structure of ecological assemblages: a comparative study between STATICO and Canonical Correspondence Analysis**

Susana Mendes

16:30 **Applying mixed models in the selection of ancient grapevine varieties**

Elsa Gonçalves

16:50 **Pulse Wave Velocity: statistical characterization of Early Vascular Aging**

Pedro Oliveira

17:10 **Field methods guided by parameter estimation in wind farm monitoring protocols**

Regina Bispo

Organized Session 14   **Room B260**

Applications in Statistics

Chairman: **Manuela Oliveira**

16:10   **Models with orthogonal block structure, commutativity and $B-$Matrices**

Carla Santos

16:30   **Focusing on the Inference**

Cristina Dias

16:50   **Effect of growth medium, temperature and incubation time on *Chlamydomonas reinhardtii***

Vinod Patil

17:10   **Analysis and prediction of heat wave over India during 2015**

Venkata Dodla

17:30   **Determinants behavior analysis of demand for passengers to the Portuguese air travel leisure market: Application with Partial Least Squares Regression (PLS)**

José Vicente

18:00   **Closing Ceremony** [Room O106]

Poster Session [Room O103]

<div align="right">Chairman: **Célio Marques**</div>

Uni- and multivariate approach to evaluate susceptibility to Fusarium head blight of wheat hybrids

Tadeusz Adamski (Poster included in the Invited PPWB5 Session 1)

Principal Component Analysis with Physical Fitness Variables

Domingos Silva (Poster included in the Organized Session 14)

A multilevel analysis of student's achievement in mathematics

Susana Faria (Poster included in the Organized Session 10)

Structural Time Series Models Applied to a Water Quality Monitoring Problem

Manuela Gonçalves (Poster included in the Organized Session 10)

On the Probability of Return Levels for Spatial Extreme Events

Luísa Pereira

Estimating the extremal coefficient: a comparison of methods

Marta Ferreira

A multivariate Birnbaum-Saunders model and its diagnostics and application to biomedicine

Carolina Marchant

Approximate probabilities and critical values for Bartels randomness test

Ayana Mateus

Environmental awareness and education: A higher education-based study

Célio Marques

Contents

Invited Speakers

Invited PPWB5 Sessions

Invited Speakers

Big data challenges for statistical approaches to 21st century medical imaging

Robert G. Aykroyd

University of Leeds, UK

E-mail address: *r.g.aykroyd@leeds.ac.uk*

Abstract

As data collection becomes easier and cheaper, it is clear that ever increasing amounts will be collected. In medical scanning, this might involve multiple modalities, such as SPECT, PET and MR, or it could be high frequency sampling, for example as in electrical tomography. More than two decades ago, substantial advances were made in statistical modelling for imaging problems based on the Bayesian paradigm coupled with iterative and sampling approaches such as the EM and Monte Carlo Markov chain algorithms. However, at the time problems were small-scale compared to those encountered today. Historical background will be given, along with modern practical examples which pose new challenges for statisticians if they want to have further impact. In my view there is scope for new interdisciplinary collaboration which can again lead to new and even greater advances.

Keywords: Bayesian methods, inverse problems, Markov chain Monte Carlo, Markov random fields, SPECT.

—⊙—
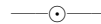
The pioneering work of statisticians Julian Besag and Peter Green, see for example [3–5], on the use of the Bayesian approach to image processing had a widespread influence on a generation of researchers, and it was my great honour to have worked for a short time in the same University department. Their work on spatial modelling for example, set the scene for what would later become the Markov random field models used as prior distributions. This carried over into Green's early work on SPECT imaging. Of course, there were many others in the field with equally far ranging influence on the research direction of many, such as the work of the Geman brothers and their various co-workers. As elsewhere, whenever estimation problems become too complex for routine algebraic approaches, the EM and the Metropolis-Hastings algorithms have become ubiquitous. In addition, the latter permits a full exploration of the posterior distribution creating a more flexible and imaginative output analysis. The key ingredients in any Bayesian model are a likelihood and a prior distribution, which are combined to give a posterior distribution. References for the above review can be found in [1] or [2].

For many medical imaging techniques, such as PET and SPECT, the data are radioactive counts recorded around the patient and the parameters are the concentration of a radioactive tracer drug introduced into the patient. Hence a Poisson likelihood is appropriate which is also a function of attenuation, scattering and geometric factors of the data collections system. A prior distribution, in terms of a Markov random field model, is chosen to regularize by introducing spatial smoothing. Analysis of real SPECT data using this modelling approach appears in [6] and [2]. The estimates use the posterior mean from an MCMC sample obtained with a random walk Metropolis-Hastings algorithm. The same

sample can do used for other posterior summaries such as credible intervals. A similar approach for electrical tomography applied in industrial process tomography can be found in [1]

In contrast to early equipment, modern scanners can collect multi-modal data, such as MR and PET, simultaneously with greater resolution. In particular, using 64 data rings each with 512 detectors. This can produce 64×64 sinograms each of 64×64 values giving a potential data size of over 1 billion values – although fewer than 1 million will be non-zero. Full 3D reconstruction then involves estimation of $64 \times 64 \times 64$ parameters. In the terminology of regression, $n = 1$ million and $p = 0.26$ million with a design matrix of $n \times p = 1$ million $\times 0.26$ million–clearly there is no opportunity of inverting this design matrix. In many complex Bayesian problems estimation is based on posterior samples obtained via a Markov chain Monte Carlo method. It is equally difficult to imagine that such a method could produce a reasonable answer in an acceptable time. To make substantial progress statisticians will need to construct simplified models and approximate numerical procedures, which are fit for purpose but which are easy to implement and fast to execute. If this goal is achieved then equally important impact can be made in the future as was made by the pioneers of the past.

Acknowledgements

References

[1] Aykroyd, R. G. (2015) Statistical image reconstruction, in *Industrial tomography: Systems and applications*, Editor Wang, Woodhead Publishing, 401–427.

[2] Aykroyd, R. G and Zimeras, S. (1999) Inhomogeneous prior models for image reconstruction. *J. Am. Stat. Soc.* **94**: 934–946.

[3] Besag, J. (1986) On the statistical analysis of dirty pictures (with discussion). *J. R. Statist. Soc. B* **48**: 259–302.

[4] Besag, J. and Green, P. J. (1993) Spatial statistics and Bayesian computation. *J. R. Statist. Soc. B* **55**: 25–37.

[5] Besag, J, Green, P. J., Higdon, D. and Mengersen, K. (1995) Bayesian computation and stochastic systems. *Statistical Science* **10**: 1–41.

[6] Green, PJ. (1990) Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE Trans. Medical Imaging* **9**: 84–93.

Tests for equality of mean vectors with structured covariance matrices

Carlos A. Coelho[1,2]

[1]Departamento de Matemática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa
[2]CMA–FCT/UNL Centro de Matemática e Aplicações

E-mail address: *cmac@fct.unl.pt*

Abstract

While the likelihood ratio test for the test of equality of mean vectors, when the covariance matrices are only assumed to be positive-definite, is a common test in Multivariate Analysis, in this presentation we will be dealing with the problem of developing likelihood ratio tests for the equality of mean vectors when the covariance matrices have some common given structure. The structures addressed for the covariance matrices are: (i) the circular or circulant structure, (ii) the compound symmetric structure, and (iii) the spherical structure. Besides obtaining the likelihood ratio statistics, also their exact distributions are obtained, showing that in some cases this distribution may have a very manageable form, while in the other cases very sharp near-exact distributions may be developed.

Keywords: exact distribution, generalized integer gamma distribution, generalized near-integer gamma distribution, likelihood ratio, near-exact distribution.

Acknowledgements

Quasi-stochastic model of the process of recurrent metastasis

Marek Kimmel[1,2] and Monika Kurpas[2]

[1]Rice University
[2]Silesian University of Technology

E-mail addresses: *kimmel@rice.edu; monika .kurpas@polsl.pl*

Abstract

We show how a transport-type partial differential equation with non-local boundary conditions, of the type considered among other by Diekmann or Kimmel and their co-authors, can be used to derive the distribution of the sizes of recurrent metastases from a growing primary tumor. In the simplest cases, we obtain closed-form expressions; in the case in which the growth rates of the metastases are exponentially distributed, we obtain expressions including incomplete Gamma functions.

These purely mathematical results are presented against the background of Darwinian and non-Darwinian theories of development of human solid cancers. We argue that rigorous mathematical models are needed to understand the dynamics of tumor growth. An introduction to the biological set-up of the problem will be provided.

Keywords: modeling, transport equation with non-local boundary conditions, tumor growth, metastases, incomplete gamma function.

Choosing the Best Predictive Logit Model

Christos P. Kitsos

Technological Educational Institute of Athens and JK University Linz

E-mail address: *xkitsos@teiath.gr*

Abstract

There is a number of approaches working with the Logit Model when it is adopted to be applied, estimating the Relative Risk. Although, all these approaches are based, eventually, on estimating the log–odds, there is an underlying different line of thought: some are adopting the optimality theory, so that to obtain the "best" estimates, see [1,2], others to provide an invariant canonical form, see [4], while recently there is an attempt to work through the generalized information criteria, see [3]. The target of this paper is to work within the spirit that we are choosing the model for prediction, see [5,6]. The obtained model might not be the best fitting the data model, but the best for prediction. In such a case, we are working through tolerance regions and not though confidence regions. This is possible to be developed due to the discribed framework in [4]. Therefore we review the existed methods and approaches and we provided the suggested new one.

References

[1]   Bruce, D. (2008) *Optimal Design and Inference for Correlated Bernoulli Variables Using a Simplified Cox Model.* PhD Thesis, Stockholm University.
[2]   Haggstrom, J. (2000) *The MinMax Approach to Optimum Design of Experiments.* PhD Thesis, Dept of Statistics. University of UMEA.
[3]   Kitsos, C. P., Toulias, T. L. (2015) Generalized Information Criteria for the Best Logit Model. In: *Theory and Practice of Risk Assessment*, by Kitsos, C. P., Oliveira, T. A., Rigas, A., Gulati, S. (Eds) Spinger: 3–20.
[4]   Kitsos, C. P. (2011) Invariant Canonical Form for the Multiple Logistic Regression. *Mathematics in Engineering, Science and Aerospace (MESA)* **2(3)**: 267–275.
[5]   Kitsos, C. P. and Zarikas, V. (2013) On the Best Predictive General Linear Model For Data Analysis: A Tolerance Region Algorithm for Prediction. *J. of Appl. Science* **13(4)**: 513–524.
[6]   Zarikas, V. and Kitsos, C. P. (2015) Risk Analysis with Reference Class Forecasting Adopting Tolerance Regions. In: *Theory and Practice of Risk Assessment*, by Kitsos, C. P., Oliveira, T. A., Rigas, A., Gulati, S. (Eds) Spinger: 235–247.

Data Science with the R software and applications

Víctor Leiva

Faculty of Engineering and Science, Universidad Adolfo Ibáñez, Chile

E-mail address: *victorleivasanchez@gmail.com*

Abstract

Data science is an interdisciplinary field of processes and systems to extract knowledge or insights from the data presented in various forms, either structured or unstructured. In particular, data science is a continuation of some of the areas of data analysis, such as statistics, data mining and predictive analytics, similar to KDD (knowledge discovering in databases). The process of data science is sometimes considered as a synonymous of business intelligence [1] and it is wrongly related to the concept of big data [2]. In this talk, we present some tools of data science with the R software and show some real-world applications.

Keywords: big data, business intelligence, databases, data mining, KDD.

References

[1]   Leiva, V. (2016) Business intelligence and its connection with the area of services. *Center of Experiences and Services: Universidad Adolfo Ibáñez*, April: 1–9. Available at `www.cesuai.cl`

[2]   Prajapati, V. (2013) *Big Data Analytics with R and Hadoop.* Birmingham, UK, Packt Publishing.

On optimality of neighbor designs under mixed interference models

Augustyn Markiewicz[1] and Katarzyna Filipiak[2]

[1]Poznan University of Life Sciences, Poland
[2]Poznan University of Technology, Poland

E-mail addresses: *amark@up.poznan.pl; katarzyna.filipiak@put.poznan.pl*

Abstract

The concept of neighbor designs was introduced and defined by Rees [4] who gave also some methods of their construction. Henceforth many methods of construction of neighbor designs as well as of their generalizations are available in the literature. Recently, some results on optimality of specified neighbor designs under various linear models were published (cf., Filipiak and Markiewicz [2,3]). The aim of the talk is to give an overview of study on the problem of optimality of neighbor designs under mixed model. It will include some recent result published by Filipiak and Markiewicz [4] as well as some new results. The study of optimality of designs under mixed model is based on the method of Filipiak and Markiewicz [5].

Keywords: neighbor designs, circular balanced design, universal optimality, mixed model.

References

[1]   Rees, D. H. (1967) Some designs of use in serology. *Biometrics* **23**: 779–791.
[2]   Filipiak, K. and Markiewicz, A. (2012) On universal optimality of circular weakly neighbor balanced designs under an interference model. *Communications in Statistics-Theory and Methods* **41**: 2405–2418.
[3]   Filipiak, K. and Markiewicz, A. (2016) Universally optimal designs under interference models with and without block effects. *Communications in Statistics-Theory and Methods* DOI: 10.1080/03610926.2015.1011786.
[4]   Filipiak, K. and Markiewicz, A. (2014) On the optimality of circular block designs under a mixed interference model. *Communications in Statistics-Theory and Methods* **43**: 4534–4545.
[5]   Filipiak, K. and Markiewicz, A. (2007) Optimal designs for a mixed interference model. *Metrika* **65**: 369–386.

Analyzing data on protein content in two types of hybrid maize cultivars under three models with split units

Iwona Mejza[1], Piotr Szulc[2], Jan Bocianowski[1],
Katarzyna Ambroży-Deręgowska[1] and Kamila Nowosad[3]

[1]Department of Mathematical and Statistical Methods, Poznań University of Life Sciences
[2]Department of Agronomy, Poznań University of Life Sciences
[3]Department of Genetics, Plant Breeding and Seed Production, Wrocław University of Environmental and Life Sciences

E-mail address: *imejza@up.poznan.pl*

Abstract

This paper deals with a way of modeling data from a three-factor field experiment. The models of observations discussed are called randomized-derived models, and are closely connected with the randomization performed for nested and crossed structures of experimental units within blocks.

Consideration was given to three mixed designs, which are combinations of two different systems such as split-plot design, split-block design and randomized complete block design. The following are considered: 1) the split-split-plot (SSP) design, which is a combination of two different split-plot (SP) designs, 2) the split-plot design for three factors as the combination of a randomized complete block (RCB) design and another SP design, and 3) the split-plot × split-block (SPSB) design, which is a combination of a split-block design and a split-plot design. These experimental designs are presented in order to consider statistical consequences in general and particular analysis.

Considerations are based on a one-year study of protein content in the grain of the two types of hybrid maize cultivars under different forms of nitrogen fertilizers and doses of magnesium. The order of the factors and their importance were determined before carrying out the experiment. The reason that these designs were selected is that larger experimental units are required for factor A levels (*forms of fertilizer)* and for factor B levels (*magnesium*) than for factor C levels (*maize cultivars*). Another reason is that less precision in estimation of comparisons is usually required for factors A and B and the interaction A × B, and more for factor C and other interaction effects (except for sampling errors). For each of them an analysis of variance and Tukey's test were performed, particularly for nitrogen fertilizer and maize cultivar interaction effects.

The relative efficiency of the considered designs with respect to estimation of some comparisons among treatment combination effects using stratum estimates of variance errors (except for sampling errors) was examined. Attention was paid to the appropriate relationships between them.

Keywords: protein, relative efficiency, types of nitrogen fertilizer, split-block design, split-plot design.

References

[1] Ambroży K. and Mejza I. (2012) Modeling data from three-factor experiments with split units set up in designs with different block structures (in Polish). *Biul. IHAR* **264**, 23–31.

[2] Ambroży, K., Mejza, I. and Mejza, S. (2014) On the relative efficiency of split-split-plot design to split-plot × split-block design. *Colloquium Biometricum* **44**, 69–78. ISSN 1896-7701.

[3] Federer, W. T., and King, F. (2007) Variations on Split Plot and Split Block Experiment Designs. *Wiley, New Jersey.*

[4] LeClerg E. L., Leonard, W. H. and Clark, A. G. (1962) *Field plot technique.* Burgess, Minneapolis.

[5] Molga M. (1986) *Agrometeorology* (in Polish). PWRiL, Warsaw.

[6] Szulc, P. and Bocianowski, J. (2012) The effect of soil supplementation with different forms of nitrogen fertiliser on modification of generative yield in two different types of maize (Zea mays L.) hybrids. *Polish Journal of Agronomy*, **11**, 52–64.

Linear Sufficiency in Models with Orthogonal Block Structure

<u>J. T. Mexia</u>[1,2], A. Markiewicz[3,4], F. Carvalho[5,2],
S. Ferreira[6,7], C. Nunes[6,7] and D. Ferreira[6,7]

[1]Department of Mathematics, Sciences and Technology Faculty, New University of Lisbon
[2]CMA-Applied Mathematics Center, Sciences and Technology Faculty, New University of Lisbon
[3]Department of Mathematical and Statistical Methods, Poznań University of Life Sciences
[4]Institute of Plant Genetics, Polish Academy of Science, Poznań
[5]Mathematics and Physics Department, Polytechnic Institute of Tomar
[6]Mathematics Department, Sciences Faculty, University of Beira Interior
[7]Center of Mathematics and Applications, UBI, Covilhã

E-mail addresses: *jtmexia@fct.unl.pt; amark@up.poznan.pl; fpcarvalho@ipt.pt; sandraf@ubi.pt; celian@ubi.pt; dario@ubi.pt*

Abstract

Models with orthogonal block structure, OBS, continue to play an important part in the theory of randomized block designs, see [2] and [3].

In this talk we use linear statistical inference to obtain BLUE estimators for estimable functions and estimable vectors in these models. Namely we will discuss the relation between these estimators and least square estimators, LSE.

Keywords: mixed models sufficiency, variance components.

Acknowledgements

References

[1] Carvalho, F. P., Mexia J. T. and Oliveira M. M. (2009) Estimation in models with commutative orthogonal block structure. *Journal of Statistical Theory and Practice* **3(2)**: 523–533.

[2] Caliński, T. and Kageyama, S. (2000) *Block Designs: A Randomization Approach. Vol. I: Analysis. Lecture Notes in Statistics* **150**, Springer-Verlag, New York.

[3] Caliński, T. and Kageyama, S. (2003) *Block Designs: A Randomization Approach. Vol. II: Design. Lecture Notes in Statistics* **170**, Springer-Verlag, New York.

[4] Fonseca, M. S., Mexia, J. T. and Zmyslony, R. (2008) Inference in normal models with commutative orthogonal block structure, *Acta et Commentationes Universitatis Tartuensis de Mathematica* **12**: 3–16.

[5] Mexia, J. T., Vaquinhas, R., Fonseca, M. and Zmyślony, R. (2010) COBS: Segregation, Matching, Crossing and Nesting. *Latest Trends on Applied Mathematics, Simulation, Modeling* 4th International Conference on Applied Mathematics, Simulation, Modelling (ASM'10), 249–255.

[6] Muller, J. (1987) Sufficiency and Completeness in the linear model, *Journal of Multivariate Analysis* **21(2)**: 312–323.

[7] VanLeeuwen, D. M., Seely, J. F. and Birkes, D. S. (1998) Sufficient conditions for orthogonal designs in mixed linear models, *J. Stat. Plan. Inference* **73**: 373–389.

[8] Zmyślony, R. (1978) A characterization of best linear unbiased estimators in the general linear model, Mathematical Statistics and Probability Theory, *Proc. Sixth Internat. Conf., Wisla, Lecture Notes in Statist.*, Springer, New York-Berlin **2**: 365–373.

Risk return profile in the hedge fund industry

Izabela Pruchnicka-Grabias

Warsaw School of Economics, Institute of Banking, Poland

E-mail address: *ipruch@sgh.waw.pl*

Abstract

Hedge funds are financial institutions which aim at generating absolute rates of return, that is at realizing profits regardless of the market situation. They are the subject of a wide scientific discussion concerning the rates of return generated compared with other forms of investments. However, some scientists claim that although hedge funds really achieve attractive rates of return, at the same time, they generate high risk level. Another problem which has not been solved so far is what is the most adequate measure or group of measures for this type of risk (sometimes named the extreme risk). The most typical risk measure used both by hedge funds and investment funds when they present results of their portfolio management is the standard deviation and the risk–return measure is the Sharpe ratio. Opponents of using it say that it requires the assumption that the rates of return are normally distributed which doesn't have to be true. Assuming that it would be true, one can apply the whole range of alternative risk––return measures. Needless to say that they are more complex, require more advanced knowledge of finance, mathematics and statistics and generate risk of making a mistake during calculations. It is hard to answer the question if the above named disadvantages are worth getting higher accuracy of results. It requires giving an answer to the question if alternative risk measures are really more accurate measures of the risk return profile in the hedge fund world and if the results achieved with the use of them are much different from those based on traditional ratios. If they are different, another question arises, if they are more adequate.

The author presents the results of the research conducted for hedge funds for the period of 1990–2014. They were divided into 10 investment strategies and net asset values calculated for indexes created for them were used. The database used in the research comprises more than 2200 hedge funds, so it can be treated as a sufficient source. The first step was to check if rates of return achieved by hedge funds were normally distributed during this period which turned out not to be true. The second step, if they are not normally distributed, chosen alternative risk-return ratios were calculated and their values were compared with traditional ones for the same period of time. The main conclusion is that these two groups of risk–return measures (traditional and alternative ones) give similar results and lead to extremely similar rankings of hedge funds. This in turn indicates that arguments of opponents of using traditional efficiency ratios in the hedge fund analysis seem to be exaggerated.

Keywords: hedge funds, risk measures, Sharpe ratio, investment strategies, rates of return.

References

[1] Breuer W., Guertler, M. and Schuhmacher, F. (2004) *Portfopiomanagement I–Theoretische Grundlagen und praktische Anwendungen*, Gabler, Wiesbaden.

[2] Brooks C. and Kat, H. M. (2002) The Statistical Properties of Hedge Fund Index Returns and Their Implications for Investors. *The Journal of Alternative Investments* **5(2)**.

[3] Eling M. and Schuhmacher, F. (2007) Does the choice of performance measure influence the evaluation of hedge funds? *Journal of Banking and Finance* **31(9)**: 2632–2647.

[4] Gregoriou, G. N. and Pascalau, R. (ed.) (2011) *Derivatives Pricing, Hedge Funds and Term Structure Models*, Palgrave, Macmillan, Hampshire.

[5] Hedge Fund Research Global Hedge Fund Industry Report–First Quarter 2014 `https://www.hedgefundresearch.com/`.

[6] Hedge Fund Research Global Hedge Fund Industry Report–Fourth Quarter 2013.

[7] Kaplan, P. D. and Knowles, J. A. (2004) *Kappa: A Generalized Downside Risk–Adjusted Performance Measure*, Morningstar Associates and York Hedge Fund Strategies, January.

[8] Kazemi H., Schneeweis, T. and Gupta, B. (2003) Omega as a Performance measure, research paper, Center for International Securities and Derivatives Markets, University of Massachusetts, Amherst, July 2003 `www.cisdm.org.;at:http://www.edge-fund.com/KaSG03.pdf`, 3.11.2015.

[9] Lo, A. W. (2002) The Statistics of Sharpe Ratios. *Financial Analysts Journal*, July–August 2002.

[10] Markowitz H. (1952) Portfolio selection. *Journal of Finance* **7(1)**, 77–91.

[11] The Q1 2015 Preqin Quarterly Update, Hedge Funds `https://www.preqin.com/docs/quarterly/hf/Preqin-Quarterly-Hedge-Fund-Update-Q1-2015.pdf`.

[12] Schneeweis T., Martin, G., Kazemi, H. and Karavas, V. (2005) The Impact of Leverage on Hedge Fund Risk and Return. *The Journal of Alternative Investments* **7(4)**: 10-21.

[13] Titman S. (2010) The leverage of hedge funds. *Finance Research Letters* **7(1)**, 2–7.

Statistical and probabilistic modeling for cancer risks: mammary cancer and nephroblastoma

Milan Stehlík[1,2]

[1]Department of Applied Statistics, Johannes Kepler University in Linz, Austria
[2]Institute of Statistics, Universidad de Valparaíso, Valparaíso, Chile

E-mail address: *Milan.Stehlik@jku.at*

Abstract

In the talk we will discuss discrimination between mammary cancer and mastopathy tissues (see [12,15]). Non invasive techniques generally may produce inverse problems, e.g. estimating a Hausdorff fractal dimension from boundary of examined tissue (see also [7]). We will discuss these issues in the context of our recent results (see e.g. [2,12]). During the talk we will discuss several issues which bring light into both fractal based cancer modelling and more general stochastic geometry models and their comparisons. The algebraic and topologic properties of cancer growth are available via appropriate set structure, e.g. bornology (see [8,9]).

Nephroblastoma (given by Wilms tumour) is the typical tumour of the kidneys appearing in childhood. For introductory work on spatial discrimination see [10,11]. We illustrate on recent pre/post clinical study the effect on chemotherapy to Euclidean volumes of tumors (see [13]). Also we will discuss several direction of potential collaboration.

When we consider fractal based cancer diagnostic, many times a statistical procedure to assess the fractal dimension is needed. We shall look for some analytical tools for discrimination between cancer and healthy ranges of fractal dimensions of tissues. [4] discussed planar tissue preparations in mice which has a remarkably consistent scaling exponents (fractal dimensions) for tumor vasculature even among tumor lines that have quite different vascular densities and growth characteristics. In [5] we provide extensive study of cancer risk assessment on simulated and real data and fractal based cancer. Both non–random and random carpets are modelling the cancer growth.

On the other hand, in previous investigations, it has been shown that the texture of mammary tissue, as seen at low magnification, may be characterized quantitatively in terms of stereology (see [3] and references therein). In [4], the images of the mammary cases were reexamined (20 cases of mastopathy and 20 cases of mammary cancer, each with 10 images). We will construct a statistical test, which is able to distinguish between the two groups and decide for a possibly new image if it belongs to masthopathic group or not (see [6]).

Keywords: fractal cancer hypothesis, mammary cancer, discrimination, statistics, stochastics.

Acknowledgements

References

[1] Baish, J. W. and Jain, R. K. (2000) Fractals and cancer. *Cancer Research*, **60**: 3683–3688.

[2] Filus, J., Filus, L. and Stehlík, M. (2009) Pseudoexponential modelling of cancer diagnostic testing. *Biometrie und Medizinische Informatik, Greifswalder Seminarberichte* **15**: 41–54.

[3] Mattfeldt, T. (2003) Classification of binary spatial textures using stochastic geometry, nonlinear deterministic analysis and artificial neural networks. *Int. J. Pattern Recogn. Artif. Intell.* **17**: 275–300.

[4] Mrkvička, T. and Mattfeldt, T. (2011) Testing histological images of mammary tissues on compatibility with the boolean model of random sets. *Image Analysis and Stereology* **30**: 11–18.

[5] Stehlík, M., Giebel, S. M., Prostakova,J. Schenk, J. P. (2014) Statistical inference on fractals for cancer risk assessment. *Pakistan Journal of Statistic* **30**(4): 439–454.

[6] Stehlík, M., Mrkvička, T., Filus, J. and Filus, L. (2012) Recent development on testing in cancer risk: a fractal and stochastic geometry. *Journal of Reliability and Statistical Studies* **5**: 83–95.

[7] Kiselak, J., Pardasani, K. R., Adlakha, N. Agrawal, M. and Stehlík, M. (2013) On some probabilistic aspects of diffusion models for tissue growth. *The Open Statistics and Probability Journal* **5**: 14–21.

[8] Solovyov, S. A. and Stehlík, M. (2015) Lattice-valued bornological systems. *Fuzzy Sets and Systems* **259(15)**, 68–88.

[9] Solovyov, S. A. and Stehlík, M. On a topological universe of $L-$bornological spaces, submitted for *Soft Computing*, `doi:10.1007/s00500-015-1905-0`.

[10] Giebel, S., Hermann, P., Schenk, J. P. and Stehlík, M. (2015) On analytical methods for cancer research, In Kitsos, C., Oliveira, T., Rigas, A. and Gulati, S. (eds.). Theory and Practice of Risk Assessment. *Springer Proceedings in Mathematics and Statistics* **06/2015**: 37–44.

[11] Hermann, P., Piza, S., Ruderstorfer, S., Spreitzer-Gröbner, S. and Stehlík, M., (2014) Fractal Case Study for Mammary Cancer, *Biometrie und Medizinische Informatik–Greifswalder Seminarberichte*, Teresa Oliveira, Karl-Ernst Biebler, Amilcar Oliveira, Bernd Jäger (Ed.) **23**: 149–166.

[12] Hermann, P., Mrkvicka, T., Mattfeldt, T., Minarova, M., Helisova, K., Nicolis, O., Wartner, F. and Stehlík, M. (2015) Fractal and stochastic geometry inference for breast cancer: a case study with random fractal models and Quermass-interaction process *Statistics in Medicine* **34**: 2636–2661.

[13] Hermann, P., Giebel, S. M., Schenk, J.-P., Stehlík, M. (2015) Dynamic Shape Analysis–before and after chemotherapy. *Proc. International Conference on Risk Analysis (ICRA6)*: 339–346.

[14] Hermann,P. and Stehlík, M. (2015) On some issues on statistical analysis for Wilms tumor, *Proc. 30th International Workshop on Statistical Modelling*, Linz, Austria, July 6–10.

[15] Stehlík M., Hermann, P., and Nicolis, O. (2016) Fractal based cancer modelling. *Revstat* **14(2)**: 139–155.

Invited PPWB5 Sessions

Invited PPWB5 Session 1

Applications of quantitative methods in economics and finance

Organizer: Stanislaw Mejza (Poland)

Ecoinnovation efficiency

Paweł Bartoszczuk

Warsaw School of Economics

E-mail address: *pbarto1@sgh.waw.pl*

Abstract

The goal of this paper is to present eco-innovation implementation in European countries. I assess factors influencing on ecoinnovation in small medium enterprises. Both current and expected energy and material prices are particularly important for pushing firms to reduce pollution. Cost savings are an important main incentive for reducing energy and material use, and for eco-innovation. Customer requirements are very important source for eco-innovations, particularly with regard to products with improved environmental performance and process innovations that increase material efficiency, reduce energy consumption and waste.

Keywords: eco-innovation, environmental impacts, discrete choice models, regulation, cost savings, demand pull, environmental policy.

———⊙———
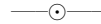
An environmental innovation, according to [5], has been defined as a new or significantly improved product (good or service), process, organizational method or marketing method that creates environmental benefits compared to alternatives. The environmental benefits can be the primary objective of the innovation or the result of other innovation objectives. They can occur during the production of a good or service, or during the after sales use of a good or service by the end user. Concerning environmental product innovations, ZEW econometric results show that present regulations are only effective for reductions of air, water, soil and noise emissions but not suitable for energy consumption and recycling [3]. The companies expect a growing importance of future regulations for all product innovations. In all considered ecoinnovation areas future regulations already trigger eco-innovations. Environmental regulation and environmental policy instruments (soft regulation) are significant driversy for product ecoinnovations). Firms implement environmental product innovation to obey with existing and anticipated legal requirements. Reference [4] demonstrate that a high level of regulatory stringency incentivises companies to implement environmental product innovations which are quite novel to the firm, but this result cannot be corroborated when these innovations are new to the market. Finally, [3] confirm a high importance of expected future regulations for all environmental product innovations. Several attempts to define green innovation have been made in the literature [1]. According to [2], the eco-innovation (or green innovation), is often used to identify those innovations that contribute to a sustainable environment through ecological improvements. In order to identify the factors influencing on whether or not enterprises implemented eco-innovations, a multiple regression model was used, and 27 EU enterprises were analysed. In particular variants of the model, the inclination for ecoinnovations measured as a proportion of enterprises which implemented particular kinds of environmentally-friendly innovations was chosen as a dependent variable. Independent variables used in models were factors influencing the decisions by companies as to whether they should implement eco-innovations.

According to the results of the calculations, the only factor which positively influenced product eco-innovation implementation was an increase in high material prices. With respect to process eco-innovations, the high prices of energy used in production and expected high energy price were only significant variables, and higher energy prices acted contrary to our theory-based expectations as a sedative to the implementation of new or significantly improved eco-innovative production processes or new methods. Considering organizational ecoinnovation, the only significant variable was high material prices. Other factors were not significant, therefore they were not included in the model.We received positive correlation among demand for green product and process organisational ecoinnovation. Concerns related to the development of eco-innovativeness are observed in EU New Member countries. First, politicians do not recognize the significance of eco-innovation and strong lobby supports obsolete industries. On the one hand, there is no pressure on research in the new eco-innovative solutions.

Acknowledgements

References

[1]   Carrillo-Hermosilla, J. and del Rio, P. (2012) Diversity of eco-innovations: reflections from selected case studies. *Journal of Cleaner Production* **18**: 1073–1083.

[2]   Halila, F. and Rundquist, J. (2011) The development and market success of eco-innovations: A comparative study of eco-innovations and "other" innovations in Sweden,*European Journal of Innovation Management* **14(3)**: 278–302.

[3]   Horbach, J., Rennings, K. and Rammer C. (2012) Determinants of eco-innovations by type of environmental impact - The role of regulatory push/pull, technology push and market pull, *Ecological Economics* **6**: 112–122.

[4]   Kammerer, D. (2009) The effects of customer benefit and regulation on environmental product innovation. Empirical evidence from appliance manufacturers in Germany, *Ecological Economics* **68**: 2285–2295.

[5]   Kemp, A. and Pearson, R. (2008) *Measuring eco-innovation* final report of MEI project for DG Research of the European Commission.

Identification of spatial effects in the evaluation of socio-economic development of EU regions

Romana Głowicka-Wołoszyn[1], Andrzej Wołoszyn[2] and
Agnieszka Kozera[1]

[1]Department of Finance and Accounting, Poznan University of Life Sciences, Poland
[2]IT Department, University School of Physical Education in Poznan, Poland

E-mail addresses: *roma@up.poznan.pl; woloszyn@awf.poznan.pl; akozera@up.poznan.pl*

Abstract

Level of social and economic development of EU regions may be influenced not only by their internal demographic and socio-economic potential but also by their geographic location. The location determines natural conditions specific to a region, such as presence of mineral deposits that encourage development of certain industries, or natural assets that favor tourism. Moreover, spatial layout of the regions involves interactions between neighboring areas, as social and economic activities of their denizens transcend the administrative boundaries. The spatial effects are an important factor of local growth and the analysis that takes them into account leads to a better understanding of regional development problems than the traditional methods. This in turn offers better tailored tools for implementing EU regional policies and finding best ways to foster cohesion across the regions. The spatial effects are studied, among many others, in the papers of [3,4,1,2]. In [8] Pietrzak explains how the failure to account for spatial autocorrelation effects leads to an erroneous interpretation of economic phenomena.

The present article aimed to identify spatial effects in the level of social and economic development of NUTS 2 regions in EU-28. Overall clustering of similar level regions was captured by global Moran's I statistic, while assessment of local interactions with neighbors was approximated for each region by Local Indicators of Spatial Association (LISA) [1,9,7]. The study drew on *Regional gross domestic product (million PPS) by NUTS 2 regions* 2014 dataset published by Eurostat for 276 regions. The calculations were performed with *rgeos, rgdal, raster, spdep* and *maptools* R-program packages on *shapefiles* provided by Eurostat's GISCO service.

Keywords: spatial effects, local interactions, economic development, EU regions.

References

[1] Anselin, L. (1995) Local Indicators of Spatial Association-LISA. *Geographical Analysis* **27(2)**: 93–115.

[2] Bivand, R. S., Pebesma, E. J. and Gomez-Rubio, V. (2008) *Applied spatial data analysis with R*, New York, Springer.

[3] Cliff, A. and Ord, J. K. (1981) *Spatial Processes: Models and Applications*, London, Pion.

[4] Getis, A. and Ord, J. K. (1992) The analysis of spatial association by use of distance statistics. *Geographical Analysis* **24(3)**: 189–206.

[5]   Janc, K. (2006) Zjawisko autokorelacji przestrzennej na przykladzie statystyki I
      Morana oraz lokalnych wskaznikow zaleznosci przestrzennej (LISA)–wybrane zagad-
      nienia metodyczne. In T. Komornicki and T. Podgorski (Eds.) Idee i praktyczny
      uniwersalizm geografii. *Dokumentacja geograficzna* **33**: 76–83.

[6]   Kopczewska, K. (2006) *Ekonometria i statystyka przestrzenna z wykorzystaniem pro-
      gramu R*, Warszawa, CeDeWu.

[7]   Kopczewska, K., Kopczewski, T. and Wojcik, P. (2016) *Metody ilosciowe w R. Apli-
      kacje ekonomiczne i finansowe, wydanie II*, Warszawa, CeDeWu.

[8]   Pietrzak, M. B. (2010) Problem identyfikacji struktury danych przestrzennych. *Acta
      Universitatis Nicolai Copernici, Ekonomia XLI, Nauki Humanistyczno-Spoleczne* **397**:
      83–98.

[9]   Suchecka, J. (2014) *Statystyka przestrzenna. Metody analiz struktur przestrzennych*,
      Warszawa, C. H. Beck.

Comparative analysis of classical and positional approaches to the synthetic evaluation of housing conditions in EU countries

<u>Joanna Stanisławska</u>[1], Romana Głowicka-Wołoszyn[1] and
Agnieszka Kozera[1]

[1]Department of Finance and Accounting, Poznan University of Life Sciences, ul. Wojska
Polskiego 28, 60-637 Poznan, Poland

E-mail addresses: *stanislawska@up.poznan.pl; roma@up.poznan.pl; akozera@up.poznan.pl*

Abstract

The process of constructing a synthetic index of development in linear ordering methods is sometimes influenced by atypical observations, outlying or extreme, according to the character of the studied problem. These values may affect analysis results (e.g. in typological classification) and thus require particular attention [4,9,5,7]. In the presence of outliers, adoption of maximum and minimum feature values as ideal objects (in methods such as TOPSIS or Hellwig) would create an excessive distance between the typical values and the extremes, consequently flattening the range of constructed index. The values of the index would then cover only fraction of its variability, potentially invalidating the assessment of development level in the distinguished classes [9]. In the end, even a single outlying observation of a given object may cause it to be unduly assigned an extreme rank in the final classification. Therefore, typological studies of object ranking should place more emphasis on the correct research techniques that take into account outlying or extreme values, including robust methods of linear ordering employing L1-median [2,3,9]. The article aims to compare the results of linear ordering of objects and typological classification using TOPSIS method [1] in classical and positional L1-median approaches, illustrated by an example of housing deficiency problem in EU countries drawing on data from EU-SILC 2014 survey.

Keywords: linear ordering, robust methods, L1-median, housing conditions.

References

[1]   Hwang, C. L. and Yoon, K. (1981) *Multiple attribute decision making: Methods and applications*, Springer, Berlin.
[2]   Rousseeuw, P. J. and Leroy, A. M. (1987) Robust regression and outlier detection. John Wiley and Sons, New York.
[3]   Vardi, Y. and Zhang, C.-H. (2000) The multivariate L1–median and associated data depth. *Proc. National Academy of Science* **97(4)**: 1423–1426.
[4]   Lira, J., Wagner, W. and Wysocki, F. (2002) Mediana w zagadnieniach porzadkowania obiektów wielocechowych. In: Paradysz J. (ed) Statystyka regionalna w służbie samorzadu terytorialnego i biznesu. Akademia Ekonomiczna w Poznaniu: 87–99.
[5]   Maronna, R. A., Martin, R. D. and Yohai, V. J. (2006) *Robust statistics theory and methods*. John Wiley and Sons.

[6]   Wysocki, F. (2010) *Metody taksonomiczne w rozpoznawaniu typów ekonomicznych rolnictwa i obszarów wiejskich*. Wyd. Uniwersytetu Przyrodniczego w Poznaniu.

[7]   Trzesiok M. (2014) Wybrane metody identyfikacji obserwacji oddalonych. Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, Taksonomia 22–Klasyfikacja i analiza danych–teoria i zastosowania **327**: 157–166.

The use of biplot in the analysis of poverty determinants in EU countries

Andrzej Wołoszyn[1], Joanna Stanisławska[2] and F. Wysocki[2]

[1]IT Department, University School of Physical Education in Poznan, Poland
[2]Department of Finance and Accounting, Poznan University of Life Sciences, Poland

E-mail addresses: *woloszyn@awf.poznan.pl; stanislawska@up.poznan.pl; wysocki@up.poznan.pl*

Abstract

Poverty is one of the most important social and economic obstacles in development of modern societies. Recognition and identification of its determinants is necessary for implementation of effective social policies that limit its reach and severity. Betterment of the living conditions of the underprivileged social strata is the aim of contemporary development theories, as well as the primary objective of EU Cohesion Policy. Hence, poverty is the subject of study for scientists from various fields, and for governmental as well as international organizations [1,4,6].

The objective of the present article was multidimensional analysis of poverty in EU countries with identification of interactions between the scope of relative poverty and its selected spatial, demographic, social and economic determinants. An attempt was also made to find links between countries and the determinants thereby characterizing groups of countries with similar poverty circumstances. The study drew on Eurostat datasets and employed PCA Biplot as research tool [3,8,5,2,9], which allowed to examine mutual positions of EU countries as well as poverty determinants with respect to poverty incidence index. The calculations were performed with *BiPlot* script of Statistica program [7].

Keywords: relative poverty, principal component analysis, biplot.

References

[1] Carvalho, S. and White, H. (1997) *Combining the Quantitative and Qualitative Approaches to Poverty Measurement and Analysis. The Practice and Potential*, The World Bank.
[2] Greenacre, M. J. (2010) *Biplots in practice*, Fundacion BBVA.
[3] Jolliffe, I. T (2002) *Principal Component Analysis*, Springer.
[4] Moller, S., Huber, E., Stephens, J. D., Bradley, D. and Nielsen F. (2003) Determinants of Relative Poverty in Advanced Capitalist Democracies. *American Sociological Review* **68(1)**: 22–51.
[5] Papiez, M. and Wanat, S. (2004) Zastosowanie metody biplotu do analizy poziomu zycia ludnosci w Polsce i krajach Unii Europejskiej. *Zeszyty Naukowe Akademii Ekonomicznej w Krakowie* **666**: 53–78.
[6] Ravallion M. (1996) *Issues in Measuring and Modeling Poverty*, The World Bank.
[7] Sagan, A. (2004) Jeden obraz ukazuje wiecej niz 10 liczb, czyli jak budowac mapy zadowolenia klienta z wykorzystaniem programu STATISTICA, StatSoft. Retrieved from `http://www.statsoft.pl/czytelnia/marketing/04obraz.pdf`
[8] Smith, L. I. (2002) *A tutorial on principal components analysis*, Cornell University.
[9] Wysocki, F. (2010) *Metody taksonomiczne w rozpoznawaniu typow ekonomicznych rolnictwa i obszarow wiejskich*, Uniwersytet Przyrodniczy w Poznaniu.

Uni- and multivariate approach to evaluate susceptibility to Fusarium head blight of wheat hybrids

<u>Tadeusz Adamski</u>[1], Maria Surma[1], Halina Wiśniewska[1], Zygmunt Kaczmarek[1], Iwona Mejza[2], Anetta Kuczyńska[1], Karolina Krystkowiak[1], Krzysztof Mikołajaczak[1] and Piotr Ogrodowicz[1]

[1]Institute of Plant Genetics, Polish Academy of Sciences, Poznan, Poland
[2]Poznań University of Life Science, Poznan, Poland

E-mail address: *tada@igr.poznan.pl; msur@igr.poznan.pl*

Abstract

Resistance to diseases is, apart from yield, the main selection criterion in breeding new cereal varieties. In small grain cereals, resistance to Fusarium head blight (FHB) caused by fungi of the genus *Fusarium* is one of the most important characteristic. *F. culmorum* (W.G.Sm.) Sacc, *F. graminearum* Schwabe and *F. avenaceum* (Fries) Saccardo infect spikes of wheat during flowering. The infection reduces grain yield and its quality by contamination of the harvested grain with toxic fungal secondary metabolites. Selection of wheat genotypes with improved resistance to FHB should be started at the beginning of breeding process, when cross combinations are selected for developing new cultivars. The present study was conducted to evaluate variation in the FHB resistance of wheat hybrids derived from crosses between winter wheat cultivars of various origin and different susceptibility to FHB. Plants were inoculated with a mixture of conidial suspension containing *Fusarium culmorum*, *F. graminearum* and *F. avenaceum* isolates. Inoculations were performed individually on each plot at the beginning of anthesis, and repeated about 3 days later at full anthesis. After harvesting, disease symptoms on kernels were observed and expressed as percentage of Fusarium damaged kernels (FDK) in kernel samples. Additionally, thousand kernel weight, number of kernels in spike and kernel weight per spike were evaluated in inoculated and control plants. Univariate (ANOVA) two-way analysis of variance was applied for evaluation of hybrids regarding yield-related traits in control and infected treatments. Additionally, multivariate analysis of variance (MANOVA) and related methods were applied for evaluation of hybrids in terms of all the traits treated simultaneously (Caliński and Kaczmarek 1973; Morrison 1976). Analysis of canonical variables was used to show the mutual position of the lines on a plain, considering all the studied traits as a complex. Besides, hybrids were divided into homogeneous groups in such a way that variability within groups (sum of squares) be minimal while the variability between groups (mixed sum of squares) maximal. To check statistical significance of performed hybrids grouping the proper statistical test was applied.

Keywords: *Triticum aestivum*, Fusarium head blight, wheat hybrids, multivariate analysis, resistance to disease.
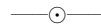
———⊙———

Acknowledgements

References

[1]   Calinski, T. and Kaczmarek, Z. (1973) Statistical methods in the analysis of multi-
      variate experiments. *Third Methodological Colloquium in Agro-Biometrics (in Polish)*:
      258–320, Warsaw, PAN.
[2]   Morrison, D. F. (1976) *Multivariate Statistical methods. Second edition*, Tokyo,
      McGraw-Hill Kogakusha.

Invited PPWB5 Session 2

Statistical Modelling and Applications

Organizer: Teresa Oliveira (Portugal)

A general method for bivariate stochastic models construction

Jerzy K. Filus[1] and Lidia Z. Filus[2]

[1]Department of Mathematics and Computer Science, Oakton Community College,
Des Plaines, Il 60016, USA
[2]Department of Mathematics, Northeastern Illinois University, Chicago, IL 60625, USA

E-mail addresses: *jkfilus98@gmail.com; L-Filus@neiu.edu*

Abstract

We consider arbitrary continuous bivariate probability distribution with positive dependences of a random vector (X, Y), given in advance its marginal distributions. We show that it always has a unique representation of its corresponding joint survival function $P(X \geq x, Y \geq y)$ in the factored form $P(X \geq x)P(Y \geq y)\,\mathrm{e}^{-m(x,y)}$, where the (unique) function $m(x, y)$ satisfies some general conditions, and determines all the stochastic dependencies of the random variables X, Y. This recognition provides a general <u>method</u> for construction of bivariate probability distributions. Fortunately, this method can also be extended for the construction of bivariate stochastic processes $\{(X_t, Y_t)\}_t$ given any two "marginal" stochastic processes $\{X_t\}$ and $\{Y_t\}$ that are completely described. The so obtained bivariate stochastic processes are expected to be applied to model, for example, the time interaction of two econometric quantities such as stock market level versus level of employment, or versus currency inflation. Extension from the bivariate to multivariate models is rather straightforward although can make the underlying statistical analysis complex.

Keywords: modified Cox model and its applications to bivariate survival functions construction, general characterization of the bivariate survival functions with positive dependences.

References

[1]   Aalen, O. O. (1989) A linear regression model for the analysis of life times. *Statistics in Medicine* **8**: 907–925.

[2]   Cox, D. R. (1972) Regression Models and Life Tables (with discussion). *Journal of the Royal Statistical Society, B* **74**: 187–220.

[3]   Filus, J. K. and Filus, L. Z. (2015) FF-Type Multivariate Models in an Enforced Regression Paradigm in *Theory and Practice of Risk Assessment*, Kitsos, C. P., Oliveira, T. A., Rigas, A., Gulati, S. (Eds.), *Springer Proceedings in Mathematics & Statistics* **136**: 399–412.

Long memory stochastic processes construction

Lidia Z. Filus[1] and Jerzy K. Filus[2]

[1]Department of Mathematics, Northeastern Illinois University, Chicago, IL 60625, USA
[2]Department of Mathematics and Computer Science, Oakton Community College,
Des Plaines, Il 60016, USA

E-mail addresses: *jkfilus98@gmail.com; L-Filus@neiu.edu*

Abstract

We will show a method of construction of simple, discrete time, stochastic processes that either are k-Markovian (with $k \geq 2$), or their memories are not limited. As the basic framework for that subject we have chosen the following "reliability with repair" problem. Consider a system which starts to operate as new, and after each failure it is repaired. We do not consider repair times, so only the (random) times T_j ($j = 1, 2, \ldots$) between the consecutive ($j - 1$-th and j-th) failures are of our interest. We will investigate, as the basic stochastic process, the ordered sequence $\{Sn\}$, where $S_n = T_1 + \cdots + T_n$.

In a situation like this the usual assumption is Markovianity of the stochastic process $\{Sn\}$. However, in practical situations this assumption is not always realistic. The probability distribution of a given time to n-th failure S_n may also depend on realizations s_{n-2}, $s_{n-3}, \ldots$ of the random variables S_{n-2}, $S_{n-3}, \ldots$ respectively.

Our goal is then to define (for $n \geq 3$) the corresponding conditional probability distributions $G_n(s_n|s_{n-1}, s_{n-2}, \ldots, s_{n-k})$ for some fixed $k \geq 2$ so that the underlying stochastic process "becomes" k-Markovian. In order to determine the conditional probability distributions $G_n(s_n|s_{n-1}, s_{n-2}, \ldots, s_{n-k})$ we apply the "method of parameter dependence" presented, among others, in our recent paper [2].

Keywords: new stochastic processes construction, k-Markovianity with $k \geq 2$, unlimited memory, reliability and econometric applications.

References

[1] Filus, J. K. and Filus, L. Z. (2008) Construction of new continuous stochastic processes. *Pakistan Journal of Statistics* **24(3)**: 227–251.

[2] Filus, J. K. and Filus, L. Z. (2013) A method for multivariate probability distributions construction via parameter dependence. *Communications in Statistics: Theory and Methods* **42(4)**: 716–721.

[3] Filus, J. K. (2014) An Alternative for Time Series Models. *China-USA Business Review* **13**, No.5 (Serial No. 131): 297–304.

The uniform distribution product as approach to the (Q,r) inventory model using R software

J. Antonio Seijas-Macias[1,2] and Amílcar Oliveira[3,4]

[1]University of Coruña
[2]Universidad Nacional de Educación a Distancia - Madrid
[3]Universidade Aberta - Lisboa
[4]CEAUL (Center of Statistics and Applications of University of Lisbon)

E-mail addresses: *antonio.smacias@udc.es; amilcar.oliveira@uab.pt*

Abstract

This paper analyses the implementation of a procedure in R to calculate the probability density function (PDF) of the product of two Uniform Distributed Random Variables. We use the Rohatgi's theorem for the theoretical form of the product, considering different possibilities for the range of values of the limits of both distributions. The paper considers the application for the product of two uniform variables to manage a (Q,r) model inventory model with the presence of lead-time and uniform demand forecasts. Solution to this model looks up to minimize the total costs through the variables Q (reorder quantity) and r (the reorder point), and there is not always exists an analytical solution of the problem. We have got a procedure for the solution. Implementation in R is straightforward.

Keywords: uniform distribution product, inventory management, (Q,r) model, R software, Rohatgi's theorem.

——⊙——
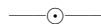
The first part of the paper is dedicated to the implementation in R of the procedure that allow us to calculate the pdf of the product of two uniform distributed variables $x \sim U(a,b)$ and $y \sim U(c,d)$. We consider only the situation where $0 < a < b < \infty$ and $0 < c < d < \infty$. Although another type of intervals could be estimated for the procedure, using a part division of the original interval. At the second part, we consider an application of the product of two uniform variables to the (Q,r) inventory model.

The (Q,r) inventory models attempt to answer two questions: when replenish the inventory and how much to order for replenishment, with the objetive to minimize the expected annual total variable cost. The two variables represent: Q is the order quantity and r is the reorder point. Solution of this problem involves three continuous distributed variables: W is a variable that represents the lead-time occurring between the order set for the product (at time n) and its delivery; and let D be a variable for the product demand. This variables are modeling using several continuous distributions: gamma, normal and uniform. For our proposes we are using the uniform distributions for both of them.

We denote by $X = WD$ the sum of the demand forecast for the product made during the lead-time. And this value is the product of the two uniform distributed variables W and D. We have no an explicit expression for de PDF of X, but we have developed a procedure in R for calculating the value of the PDF as a function of X. From this function, we have the cumulative density function (CDF) of X; this is implemented with the procedure CSL, that represent the cycle service level of X that represents the complement to 1 of the probability of stock out. From the PDF of X, we have development another procedure for estimating the value of ESC (expected shortage per cycle) as the difference between

the expected value of X and the value of r times the complement to 1 of the probability of stock out.

The process of minimization of the expected annual total variable cost of the inventory model is solving two equations for optimal values of Q and r. The two equations system cannot be solved analytically except for a special class of distributions.

In this paper, using our implementation in R, we show a graphical solution of the problem, and we can evaluate the approximated optimal solution. Our results are shown using an example and compared with another simulated approximations for the the case were uniform distributions are considered.

Acknowledgements

References

[1]   Glen, A. F., Leemis, L. M. and Drew, J. H. (2004) Computing the distribution of the product of two continuous random variables. *Computational Statistics and Data Analysis* **44**: 451–464.

[2]   Namit, K. and Chen, J. (1999) Solutions to the (Q, r) inventory model for gamma lead-time demand. *International Journal of Physical Distribution & Logistics Management* **29 (2)**: 138–151.

[3]   Wanke, P. F. (2008) The uniform distribution as a first practical approach to new product inventory management. *International Journal of Production Economics* **114**: 811–819.

[4]   R CORE TEAM (2015) *A language and environment for statistical computing*, Vienna, R Foundation for Statistical Computing.

Combination of Statistical and Clinical Methods to Assess the Risk of Tooth Loss

<u>J. A. Pereira</u>[1,2], L. Mendes[1], A. Costa[3] and T. A. Oliveira[4]

[1]Faculdade de Medicina Dentária da Universidade do Porto, Portugal
[2]MBB-Universidade Aberta, Portugal
[3]Universidad Nacional de Educacion a Distancia, Spain
[4]Universidade Aberta and CEAUL, Portugal

E-mail addresses: *lobopereiramail@gmail.com; luziacmmendes@hotmail.com; acosta@mat.uned.es; teresa.oliveira@uab.pt*

Abstract

The estimation of the dental root surface area is an important criterion in therapeutic options. In periodontics the estimation of intra-alveolar root portion is crucial for the prognosis of a periodontally compromised tooth. Traditionally the remaining periodontal attachment has been assessed by direct or proportional linear measurements of intra-alveolar root portion not taking into account the morphology of the tooth root. A method to estimate the area of periodontal attachment would allow a more accurate assessment of tooth loss risk. The aim of this study was to develop a multiple regression model to estimate the intra-alveolar root surface area of the lower seconds premolars from linear measurements obtained noninvasively, and provide an easy method for clinical implementation.

We begin by creating a conceptual model plausible in the context of our research, under the assumption that the root surface area can be modeled as a function of the length and mesiodistal width measured on tooth crown equator.

A sample of five teeth from Portuguese male adults was used to calculate the root surface areas by division planimetry method using X-ray microtomography images, the length and mesiodistal width were measured in 1:2 enlarged photographs. The univariate and multivariate probability distributions were assessed and the correlations matrix and the effect size computed, together with the coefficient representing the remaining percentage of intra-alveolar root surface area per each decile of root length.

From these data the parameters of the test model were estimated and the sample size determined through a Monte Carlo study integrating power analysis and accuracy in parameter estimation. Verified the adequacy of model parameters a simulated sample was generated from the same probability distribution of initial data, and a model with statistically significant parameters was fitted.

The clinical implementations of this method start by calculating the root surface area using the fitted multiple regression model and determining the bone level in root length deciles using a Schei ruler adapted for digital use. Finally, the remaining attachment surface was computed multiplying the coefficient representative of remaining percentage of intra-alveolar root surface area by the root surface area of the tooth.

Keywords: clinical implementation, monte carlo methods, periodontitis, risk, statistical models.

Acknowledgements

References

[1] Cohen, J. (1992) Statistical power analysis. *Current Directions in Psychological Science* **1(3)**: 98–101.

[2] Klock, K. S., Gjerdet, N. R. and Haugejorden, O. (1993) Periodontal attachment loss assessed by linear and area measurements in vitro. *Journal of Clinical Periodontology* **20**: 443–447.

[3] Maxwell, S. E., Kelley, K. and Rausch J. R. (2008) Sample Size Planning for Statistical Power and Accuracy Parameter Estimation. *Annual Review Psychology* **59**: 537–563.

[4] Mowry, J. K., Ching, M. G., Orjansen, M. D., Cobb, C. M., Friesen, L. R., MacNeill, S. R. and Rapley, J. W. (2002) Root Surface Area of the Mandibular Cuspid and Bicuspids. *Journal of Periodontology* **73(10)**: 1095–1100.

[5] Muthén, L. K. and Muthén, B. O. (2002) How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling: A multidisciplinary Journal* **9(4)**: 599–620.

[6] Ramzan, S., Zahid, F. M. and Ramzan, S. (2013) Evaluating Multivariate Normality: A Graphical Approach. *Middle-East Journal of Scientific Research* **13(2)**: 254–263.

[7] Sarkadi K. (1975) The Consistency of the Shapiro-Francia Test. *Biometrika* **62(2)**: 445–450.

[8] Schei, O., Waerhaug, J., Lovdal, A. and Arno, A.(1959) Alveolar bone loss as related to oral hygiene and age. *Journal of Periodontology* **30**: 7–16.

[9] Zheng, B. and Agresti, A. (2000) Summarizing the predictive power of a generalized linear model. *Statistics in Medicine* **19**: 1771–1781.

Recent developments on optimal designs and their applications in inference

Teresa Azinheira Oliveira[1] and <u>Milan Stehlík</u>[2,3]

[1]CEAUL and DCeT, Universidade Aberta, Portugal
[2]Department of Applied Statistics, Johannes Kepler University in Linz, Austria
[2]Institute of Statistics, Universidad de Valparaíso, Valparaíso, Chile

E-mail addresses: *teresa.oliveira@uab.pt; Milan.Stehlik@jku.at*

Abstract

In many chemometrics applications testing and inference of random response models is an important challenge (see e.g. [3]). During the talk we will provide introduction to Bib Designs, factorial designs and Optimum designs. Their importance for robust inference in order to assess degree of polynomial will be explained.

Keywords: Bib designs, optimum design, random response models, algebraic designs.

Acknowledgements

References

[1] Stehlík, M. and Střelec, L. (2012) Invited Discussion on the paper "Optimum design of experiments for statistical inference" by Steven G. Gilmour and Luzia A. Trinca, *Journal of the Royal Statistical Society - Series C* **61(3)**: 392–393.

[2] Oliveira, T. A. (2009) Bib Designs With Repeated Blocks: Review And Perspectives. *Proc. ICCS-X*, Cairo, Egypt December 20-23, 2009, **18**: 1–20.

[3] Stehlík, M., Thulin, M. and Strelec, L. (2014) On robust testing for normality in chemometrics. *Chemometrics and Intelligent Laboratory Systems* **130**: 98–108.

Special Invited Session
Round Table (Publications)

Organized discussion on the aims of the Journal of Applied Statistics and of the Springer book

Adriana Regina Brianezi-Aykroyd

Ass. of Journal of Applied Statistics (Discussion Publications)

Abstract

Chair of publications session. Organized discussion on the aims of the journal and of the Springer book.

Mini-courses

Mini-course 1

Theories and Strategies of Goal Setting and Goal Attainment in Patients' Self-Management of Chronic Diseases: Practical Application of Chronic Disease Management Using Video-Technology

Barbara Kimmel[1], Jane Anderson[2,3,4], Guilherme Oliveira[5] and Cristina Lopes[5]

[1]University of Texas School of Public Health, Houston, TX
[2]Michael E. DeBakey Veterans Medical Center
[3]Houston Health Services Research and Development
[4]Center for Innovations in Quality, Effectiveness & Safety (IQuESt)
[5]USF Rodrigues Miguéis, Lisbon, Portugal

E-mail address: *barbara.kimmel@uth.tmc.edu; jane.anderson@va.gov; guilhermegiloliveira@gmail.com; lopes.cristina@netcabo.pt*

Part 1

Goal Setting/Goal Attainment Theory and Health Behavior Change Strategies for Patients' Self-Management to Reduce Stroke Risk Factors

Barbara Kimmel

The creative challenge of health behavior change for stroke risk factors self-management is to develop theory-based methods and to translate these methods to successful intervention programs aimed to improve stroke risk factors self-management. Goal setting is widely recognized as an integral part of self-management including stroke risk factors control such as hypertension, diabetes, smoking, obesity and stress. Collaborative goal settings with clinicians and patients are used to enhance patients' motivation, adherence and autonomy and improve their satisfaction with intervention programs. In addition, patients' active involvement in goal settings tends to result in a positive correlation between goal-settings process and treatment outcomes. Goal attainment is needed to assess and improve task performance and team work, and to evaluate progress. Several health behavior theories have been used as a theoretical framework of goal settings and goal attainment in the context of the self-management of chronic diseases with the emphasis on stroke. The purpose of this presentation is to present theory-based mechanisms involved in goal settings and goal attainment concepts and summarize the current knowledge of goal attainment measures. We will also present a practical application of the goal settings, action planning and goal attainment framework utilized in the self-management intervention program in the V-Stop study (Patient Self-Management Guide).

Part 2

A Telehealth Exemplar - Video-teleconference Technology for
Self-management TO Prevent - (V-STOP) Stroke

Jane Anderson

Video-teleconference technology among patients experiencing stroke or transient ischemic attack (TIA) will be presented as an exemplar of telehealth solutions to increase access to patient self-management support and follow up care with a specialist. Self-management education has been identified as one key strategy to address the burden of chronic disease to help patients manage their chronic health conditions more effectively. Stroke secondary prevention centers on the management of chronic conditions, like hypertension and diabetes, and lifestyle changes, like smoking and physical inactivity. Specialty care and prevention programs that incorporate frequent provider visits, or involve supportive group therapy and group education are not easily accessible to patients that live in rural setting, far from large medical centers, or are underserved by their current healthcare system.

I V-STOP is a patient self-management support, preventative care program to reduce stroke risk factors among patients who have had a stroke or TIA.

    a. The program is delivered using video-teleconferencing technology (VTEL) from an urban academic medical facilities (hub sites) to local outpatient clinics (spoke sites).

    b. Two types of clinic visits are provided

        1. Group clinic with patient self-management support in a group setting
        2. Individual specialty clinic with a healthcare provider.

II The overall objective of this presentation will be to present results from V-STOP pilot projects within the United States Veterans Affairs system and a new pilot project within a Texas research network-The Loan Star Stroke Consortium.

    a. A formative evaluation of V-STOP outcomes and implementation using a pilot, open trial design will be presented.

    b. Focus groups, survey, pre/posttest and stepped-wedge design methods will be described.

    c. Factors associated with feasibility and effectiveness of telehealth implementation will be presented.

    d. Benefit is to enhance access to telehealth services for rural and underserved patients will be discussed.

    e. Research design to determine comparative effectiveness of telehealth vs. traditional patient care delivery will be presented.

Part 3

Application of a Self-management intervention program to study change in health behavior in a Primary Care Center.

Guilherme Oliveira (co-author Cristina Lopes)

Self-management is crucial for changing health behavior, particularly in chronic diseases. The challenge involving health behavior change is to develop theory-based methods that can translate to successful programs which can improve risk factors self-management. Goal settings is part of self-management programs concerning several health conditions such as hypertension, diabetes, smoking, obesity and stress. For clinicians and patients it should increase the patient's motivation on changing the health behavior. The purpose of this presentation is to discuss a practical application of the goal settings, action planning and goal attainment framework applied in a Primary Health Care Center in Lisbon, Portugal, based in a similar program developed in the Texas Medical Center. The effectiveness of self-management interventions will be considered in this presentation and the mechanism and self-efficacy through which these interventions work will be discussed. In further research the aim is to compare the program effectiveness in both populations.

References

[1]   Anderson, J. A., Goldwin, K. M., Saleem, J., Russell, S., Robinson, J., Netto, A. and Kimmel, B. (2012) Accessibility, Usability and Feasibility of a Web-based Clinical Decision Support Tool to Enhance Provider-Patient Communication around Self-management TO Prevent (STOP) Stroke. *Health Informatics Journal.*

[2]   Anderson, J. A., Godwin, K. M., Petersen, N. J. and Willson, P. C. (2013) A pilot test of videoconferencing to improve access to a stroke risk-reduction programme for Veterans. *Journal of Telemedicine and Telecare* **19(3)**: 135–141.

[3]   Lorig, K. and Holman, H. (2003) Self-management education: History, definition, outcomes, and mechanisms. *Ann Behav. Med* **26**: 1–7. PMID: 12867348.

[4]   Kennedy, A., Reeves, D., Bower P., Lee, V., Middleton E. and Richardson G. (2007) The effectiveness and cost effectiveness of a national lay-led selfcare support programme for patients with long-term conditions: a pragmatic randomized controlled trial. *J. Epidemiol Community Health* **61(3)**: 254–61.

[5]   Brady, T. J. (2013) A meta-analysis of health status, health behaviors, and health care utilization outcomes of the chronic disease self-management program. *Preventing chronic disease* **10**.

Mini-course 2

Ruler, Compass, and Nail: 360 degree Spherical Perspective made simple

<u>António Araújo</u>[1,2,3]

[1]DCeT-Universidade Aberta
[2]CMAF-CIO
[3]CIAC

Abstract

There is a growing interest in curvilinear perspectives for computer graphics, photography, and video applications. We present a method for implementing a full spherical perspective by elementary means

Mini-course 3

Creativity and mathematics creation in history

Iran Abreu Mendes

Federal University of Rio Grande do Norte - Brazil

E-mail address: *iamendes1@gmail.com*

Abstract

In this conference we present some essential aspects of the creative process that characterizes the construction of meanings in mathematics historically produced, since the reorganization of these meanings can contribute to a didactic approach to mathematics to be taught at the basic school and the training of mathematics teachers. This process takes on cognitive exercises connections whose synapses must converge to the understanding and practice of mathematical creation in the classroom, in order to incorporate heuristics as school culture embodied by reinventions of process of the math production. In this sense show how the history of mathematical elaborations defined the outline of the challenges that led to the production of mathematical topics currently discussed at the basic school and higher education.

Organized Sessions

Organized Session 1

Computational Mathematics and Applications in Sciences and Engineering

Organizer: Filomena Teodoro (Portugal)

MANOVA for the analyses of the effect of the processing conditions on the dispersion of layered clays in PP polymer

M. Fátima de Almeida[1], Aldina Correia[1] and Eliana Costa e Silva[1]

[1]CIICESI, ESTGF – Polytechnic of Porto, Portugal

E-mail addresses: *mff@estgf.ipp.pt; aic@estgf.ipp.pt; eos@estgf.ipp.pt*

Abstract

Polymer/clay nanocomposites find application where good mechanical properties, low permeability to gases or flame resistance are important service specifications. Achieving proper intercalation/exfoliation of the clay tactoids is not easy. Several methodologies, such as the use of supercritical fluids, in-situ polymerization and melt compounding, have been proposed. However, from a practical point of view, melt compounding is quite attractive, since twin screw extruders are widely used in industry and the process is straightforward. In fact, melt compounding using twin screw extruders (TSE) allows the required high stress to break-up the tactoid structure of the clays and to promote intercalation/exfoliation. It is generally believed that processing conditions inducing high thermomechanical stress levels will positively impact on dispersion.

In general agreement with experimental observations with high stress applied in the nancomposite materials were found. High screw stress configuration as well as high screw speed, low temperature and feed rate, although the last one with less clearly, promote exfoliation by a mechanism whereby the clay sheets slide transversally over one another. However, despite the various studies focusing on the effect of the operating conditions on the dispersion of layered clays, both the relationship between thermomechanical stresses applied during processing and dispersion levels remain to be fully understood.

In order to investigate the effects of processing conditions, as well as screw profile, upon the formation of PP nanocomposites, a polypropylene-Montmorillonite system was studied. Compounding was performed at fixed composition (80/15/5, w/w/w) of PP/PP-g-MA/C15A. The main goal was to monitor the development of the dispersion of a Cloisite C15A into a matrix consisting of a polypropylene/maleic anhydride grafted polypropylene matrix (PP/PP-g-MA) using a co-rotating intermeshing twin screw extruder (Leistritz LSM 30.34). The screws are 29D long and contain four mixing zones separated by feeding elements. As the material progresses towards the die, it must flow along 12 kneading disks plus a left-handed element, four kneading disks, three kneading disks and a left hand element, respectively. These kneading blocks are staggered positively ($90°$) and negatively ($-30°$). Relatively to the last one it is believed that a high shear screw profile is adequate to ensure the necessary dispersion levels. For this reason, the parameters screw speed, temperature and feed rate, were studied by varying each one independently at 150 rpm and 300 rpm, from $170°C$ to $190°C$ and 2 kg/h to 6 kg/h, respectively.

Samples were collected at die exit and characterized by X-ray diffraction (XRD), transmission electron microscopy, rheology measurements and mechanical properties. Discs were prepared by compression molding for X-ray diffraction and rheology measurements. The X-ray data was obtained using a PANalytical X'Pert PRO spectrometer with a $\lambda = 1.54$ of Cu-K α radiation and 2θ angles ranging from $1°$ to $5°$. Isothermal frequency sweeps from 0.01 to 100 rad/s and was performed in a Paar Physica MCR 300 rotational rheometer

using parallel-plates geometry. To ensure a linear viscoelastic response the strain was kept at 5 percent. Micrographs of material samples were obtained by TEM using an acceleration voltage of 15 kV. For that purpose, ultrathin sections roughly 70 nm thick were cut from samples under cryogenic conditions ($-45°$C), applying a UC6 ultramicrotome (Leica) equipped with a diamond knife. The thickness, length and aspect ratio was measured in the TEM images obtained. Extruded PNC pellets were injection molded using a Ferromatik-Milacron K85 machine into standard tensile specimens. The barrel temperature was of $200°$C, mold temperature was of $40°$C and injection speed was of 40 mm/s. Tensile tests were carried out according to ASTM D638 using an Universal Instron 4505 tester. Crosshead speed was set up at 1 mm/min. An extensometer was used to accurately measure the strain. Young's modulus was calculated by least square formfitting of the experimental data of stress vs. strain in the initial region of strain from 0 to 1%.

Preliminary analysis showed that all the characterization techniques revealed different nanocomposite morphologies. Further more, the results allowed to consider the generalization of some observed relations between the processing conditions and the properties observed in the material. Low temperature, high screws stress with low screw speed allows a highest levels on dispersion of layered clay. However isn't observed significant change by feed rate.

In this work we analyse the effects of the processing conditions on the formation of PP nanocomposites (measured obtained by four experimental techniques), using the multivariate analysis of variance (MANOVA). The goal is to test differences between the means of the dispersion results in the different groups defined according to the processing conditions. The primary purpose of the two-way MANOVA is to understand if there is any interaction between the independent variables (processing conditions, in this work) on the two or more dependent variables (in this case, measures obtained from the various experimental techniques). The sample consisted of 67372 experimental results. Four independent variables (or processing conditions) were considered: screw profile ($-30°$ or $90°$), screw speed (150 rpm or 300 rpm), temperature ($170°$C, $180°$C or $190°$C) and feed rate (2 kg/h, 4 kg/h or 6 kg/h). Nine dependent variables (or measures obtained from the various experimental techniques) were considered: Mechanical properties – stress (σ), and strain (ϵ) in the initial region of strain from 0 to 1; Rheology – G prime (G'), G double prime (G") and complex viscosity (η); TEM – thickness (t) and length (l).

The results obtained with MANOVA allowed to deduce a generalization of the effects observed in experimental studies and prove them by Statistical Inference. The conclusion is that all the processing conditions, except feed rate, have a significant effects on at least one of the dependent variables. That is, the experimental measurements related to several techniques on dispersion of a layered clays in PP polymer vary according to the type of processing conditions. The mechanical properties and TEM depend on all processing conditions. In rheology, the complex viscosity, η, is affected by screw profile and speed, temperature and by the interaction between screw profile and screw speed.

Keywords: dispersion, clay, polymer, processing conditions, material properties, multivariate statistical analysis, MANOVA.

Statistical characterization of noise in different industrial workstations

<u>Aldina Correia</u>[1] and Augusto Miguel Lopes[1]

[1]CIICESI, ESTGF – Polytechnic of Porto, Portugal

E-mail addresses: *aic@estgf.ipp.pt; aml@estgf.ipp.pt*

Abstract

The damage caused by noise in workers health is well known. A summary of main effects of workplace noise is presented in [1]. They define the loss of hearing as the principal effect of noise at work, however, it can also exacerbate stress and increase the risk of accidents. More than one third of employees in the European Union, i.e., approximately 60 million people, are exposed to high levels of noise during a quarter of their working day, according with [3].

In 2003, a Directive of the European Parliament and of the Council on the minimum health and safety defines rules about the exposure of workers to the risks from noise and defines the necessity to transpose it into the national legislation of all Member States. The directive requires that, taking into account the technical progress and the available measures to control the risk at source, the risks arising from exposure to noise shall be eliminated at their source or reduced to a minimum. The directive also sets a new daily exposure limit value of 87 dB(A) [1].

The problem to be addressed in this work is about noise analysis, performed under the PREVENIR program[1]. The data is composed by information about industrial workstation from 280 Portuguese industrial companies distributed by different sectors, including food and beverages (13.9%), rubber and plastic products (18.1%), pottery and glass (14.1 %), jewelery and watchmaking (3.1%), wood and furniture (15.6%), mechanical engineering and metallurgy (14.8%), chemicals (4.9%) and textile and clothing (15.5%). The program was implemented between 2005 and 2011 and was developed by AEP[2] in partnership with ACT[3] and the support of POAT[4].

The descriptive approach of the study is presented in [2]. Our goal in this work is to identify differences of intensity of noise exposure between these industrial sectors in different workstations, using inference techniques; then the existence of significance differences between average levels of Equivalent Sound Level (LAeq,TdB(A)) are verified using ANOVA.

Using this methodology we conclude that workstations with smaller exposure are in the textile and clothing sector (with average 75.065 dB(A)) and in the jewelery and watch-making sector (76.331 dB(A)), followed by the chemicals sector (77.729 dB(A)). These three sectors have not significant differences between them, forming the first subset of sectors. The following subset of sectors includes food and beverages (78.866 dB(A)) and pottery and glass (79.227 dB(A)). Mechanical engineering and metallurgy (84.404 dB(A)) and wood and furniture (85.384 dB(A)) are the sectors with the most affected workstations by noise.

Keywords: noise, industry, industrial jobs, industrial sectors, health.

---

[1] Prevenir – Prevenção como Solução
[2] Associação Empresarial de Portugal
[3] Autoridade para as Condições do Trabalho
[4] Programa Operacional de Assistência Técnica

References

[1] European Agency for Safety and Health at Work (2005) The impact of noise at work. *FACTS* **57**: 176–184.

[2] Lopes, A. M., Pinto, E., Laranjeira, P., Rebelo, M. and Oliveira, P. (2015) Characterization of Workers' Noise Exposure of Portuguese Industry. *SHO 2015 - International Symposium on Occupational Safety and Hygiene. Portugal.*

[3] Pleban, D. (2014) The Use of a Global Index of Acoustic Assessment for Predicting Noise in Industrial Rooms and Optimizing the Location of Machinery and Workstations, *International Journal of Occupational Safety and Ergonomics* **20(4)**: 627–638.

Text mining algorithms: a survey on the statistical process

Anacleto Correia[1] and M. Filomena Teodoro[1,2]

[1]CINAV, Center of Naval Research, Naval Academy, Portuguese Navy
[2]CEMAT, Center for Computational and Stochastic Mathematics, Instituto Superior Técnico, Lisbon University

E-mail addresses: *cortez.correia@marinha.pt; maria.alves.teodoro@marinha.pt*

Abstract

Text mining is an interdisciplinary field which draws on information retrieval, data mining, machine learning, statistics, and computational linguistics [7,4]. Text Mining as a knowledge discovery process aims to deal with unstructured (textual) information, extract meaningful numeric indexes from the text, and, thus, make the information contained in the text accessible to the various data mining algorithms.

Since most currently used information is stored as text, text mining has high commercial value. No wonder that we can currently found applications of text mining on multiple products [2], to mention a few, search engines, email spam filters, product suggestions at check-out, fraud detection, customer relationship management [8], social media analysis, marketing surveys, financial services, and research and development.

With text mining information is extracted to derive summaries for the words contained in the documents [1] or to compute summaries for the documents based on the words contained in them. Hence, we can analyze words, clusters of words used in documents, etc., or we could analyze documents and determine similarities between them or how they are related to other variables of interest in the data mining project [2,6,5].

In this paper we survey currently available text mining algorithms and describe the process workflow of using them - together with supplementary techniques from areas of information retrieval, natural language processing, information extraction and data mining - for deriving patterns and trends from text.

Keywords: association, classification, clustering, data mining, information retrieval, information extraction, text mining.

Acknowledgements

References

[1] Carenini, G., Murray, G. and Ng, R. (2011) *Methods for Mining and Summarizing Text Conversations.* ISBN: 9781608453900, DOI: 10.2200/S00363ED1V01Y201105DTM017, Morgan & Claypool.

[2] Feldman, R. and Sanger, J. (2007) *The Text Mining Handbook – Advanced Approaches in Analyzing Unstructured Data.* ISBN: 13 978-0-521-83657-9, Cambridge University Press.

[3] Hofmann, M. and Chisholm, A. (2015), *Text Mining and Visualization – Case Studies Using Open-Source Tools.* ISBN: 9781482237573, 297 Pages, Data Mining and Knowledge Discovery Series, Chapman and Hall/CRC Press.

[4] Maimon, O. and Rokach, L. (Eds.) (2010) *Data Mining and Knowledge Discovery Handbook.* ISBN 978-0-387-09823-4, DOI 10.1007/978-0-387-09823-4, New York, Springer.

[5] Manning, C. and Schütze, H. (1999) *Foundations of Statistical Natural Language Processing*, ISBN-13: 978-0262133609, 620 pages, Cambridge–MA, MIT Press.

[6] Miner, Elder, Fast, Hill, Nisbet and Delen (2012) *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications*, 1000 pages, 1st Edition, 9780123869791, Academic Press.

[7] Srivastava, A. N. and Sahami, M. (2009) *Text Mining: Classification, Clustering, and Applications.* ISBN 9781420059403, 328 Pages, Chapman and Hall/CRC Press.

[8] Struhl, S. (2015) *Practical Text Analytics – interpreting text and unstructured data for business intelligence.* EAN: 9780749474010, 272 pages, Marketing Science Series, USA, Kogan Page Limited.

Evaluation of GPS Static Relative Positioning

M. Filomena Teodoro[1,2], Fernando M. Boaventura[3] and
Anacleto Correia[1]

[1]CINAV, Center of Naval Research, Naval Academy, Portuguese Navy
[2]CEMAT, Center for Computational and Stochastic Mathematics, Instituto Superior
Técnico, Lisbon University
[3]NGI, Nottigham Geospatial Institute, University of Nottingham, UK

E-mail addresses: *maria.alves.teodoro@marinha.pt; geofactor@vodafone.pt;
cortez.correia@marinha.pt*

Abstract

Networks RTK are common in Europe but this is not the case in emerging economies where huge construction projects are running requiring geodetic support. In such cases, the easiest way to ensure that kind of support still is the static relative positioning using a single reference station. This technique provides surveyors the ability to determine the 3D coordinates of a new point with centimeter-level accuracy relative to a control point located several hundred kilometers away, which in turn can be associated with another GNSS receiver of a continuously operating reference stations (CORS) operated by some institution.

The objective of this work is to evaluate GPS static relative positioning [2–4], regarding accuracy, as the equivalent of a network RTK and to address the practicality of using either a CORS or a passive control point for providing accurate positioning control. The precision of an observed 3D relative position between two GNSS antennas, and how it depends on the distance between these antennas and on the duration of the observing session, was studied.

We analyze the performance of the software for each of the six chosen ranges of length in each of the four scenarios created, considering different intervals of observation time. The relation between observing time and baseline length is established.

In this work are applied different statistical techniques, such as data analysis and elementary/intermediate inference level techniques [5] or multivariate analysis [6,1].

Keywords: data analysis, GPS, long base lines, multivariate statistical methods, statistical modeling.

Acknowledgements

References

[1] Anderson, T. W. (2003) *An Introdution to Multivariate Analysis*, 3rd edition, New York, Jonh Wiley & Sons.

[2] Hofmann-Wellenhof, B., Lichtenegger, H. and Wasle, H. (2008) *GNSS-Global Navigation Satellite Systems GPS, GLONASS, Galileo, and more*, New York, Springer Verlag-Wien.

[3] Kaplan, E. D. and Hegarty, C-J. (2006) *Understanding GPS: principles and applications*, 2nd edition. Norwood, Artech House.

[4] Leick, A. (2004) *GPS Satellite Surveying* 3rd edition. New Jersey, John Wiley & Sons.

[5] Tamhane, Ajit C. and Dunlop, Dorothy D. (2000) *Statistics and Data Analysis: from Elementary to Intermediate*. Upper Saddle River New Jersey 07458, Prentice Hall.

[6] Turkman, M. A. and Silva, G. (2000) *Modelos Lineares Generalizados da teoria a prática*, Lisboa, Sociedade Portuguesa de Estatística.

Organized Session 2

Practical applications of multivariate time series analysis and longitudinal models

Organizer: Eliana Costa e Silva (Portugal)

A joint model of breast cancer with spatial effects

Ana Borges[1], Inês Sousa[2] and Luís Castro[3]

[1]CIICESI, ESTGF, Polytechnic of Porto, Portugal
[2]CMAT, Minho University, Portugal
[3]Braga's Hospital

E-mail addresses: *aib@estgf.ipp.pt; isousa@math.uminho.pt*

Abstract

This work pretends to contribute to the comprehension of the progression and survival of a complex disease such as Breast Cancer, particularly within the Portuguese population. The main purpose of the present analysis is to develop a joint model for longitudinal data, such as repeated measurements over time of the tumour marker Carcinoma Antigen 15-3 (CA15-3), and survival of patients with breast cancer, by incorporating a spatial effect in order to infer how patients' residence location affects the progression of the disease.

Keywords: breast cancer, joint model, longitudinal, spatial effect, survival.

—⊙—
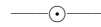
A comparison between previous independent survival [1] and longitudinal analyses [2], with results obtained with a joint model of these two processes, lead us to conclude that independent analysis brings up bias parameter estimates. This points out to the need of acknowledging the expected association between the progression of the tumour marker values with patients' survival. As such, we proceeded with a joint modelling of these two processes to infer on the association between them, and also to understand which factors affect the progression of the tumour marker CA15-3 and the survival process, adopting the methodology of random effects [3,5].

The data analysed gathers information on 540 patients of the Senology Unit of Braga's Hospital, located in the north of Portugal, diagnosed with breast cancer, and englobes 50 variables collected from medical records of the Hospital. To incorporate a spatial effect it was created a variable that groups the patients' residence parish in three categories: (i) parishes that integrate the city (PI), (ii) parishes that are predominantly urban areas (PUA) and (iii) parishes that are moderately urban areas (MUA). The referred categorization was based in information from the Portuguese 2014 typology of urban areas (TIPAU 2014, available at `www.ine.pt`).

Results indicate that the longitudinal progression of CA15-3 is significantly associated with the probability of survival of these patients. Factors such as age at diagnosis, tumour stage and images of vascular invasion affect the longitudinal process. While factors such as triple negative tumour (yes versus no) and tumour degree affect the survival process. The spatial effect is significant on both longitudinal CA15-3 and survival processes. Although no significant differences were register between the survival of patients from PI and PUA areas, the risk of dying is significantly higher for a patient residing in a MUA area compared to a patient who resides in an PI area. However, residing in parishes remote from the center causes a decrease on initial values of the linear mean progression of CA15-3 values.

References

[1]   Borges, A., Sousa, I. and Castro, L. (2014). Breast Cancer Survival at Braga's Hospital - Portugal. In *Theoretical and Applied Issues in Statistics and Demography* C H Skiadas. ISBN 978-618-81257-7-3

[2]   Borges, A., Sousa, I. and Castro, L.(2015). Longitudinal Analysis of Tumor Marker CEA of Breast Cancer Patients from Braga's Hospital. *REVSTAT - Statistical Journal* **13(1)**: 63–78.

[3]   Henderson, R., Diggle, P. J. and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* **1(4)**: 465–480.

[4]   Wulfsohn, M. S. and Tsiatis, A. A. (1997). A Joint model for survival and longitudinal data measured with error. *Biometrics* **53**: 330–339.

Joint modelling of longitudinal CEA tumour marker progression and survival data on breast cancer

Ana Borges[1], Inês Sousa[2] and Luís Castro[3]

[1]CIICESI, ESTGF, Polytechnic of Porto, Portugal
[2]CMAT, Minho University, Portugal
[3]Braga's Hospital

E-mail addresses: *aib@estgf.ipp.pt; isousa@math.uminho.pt*

Abstract

This work proposes the use of Biostatistics methods to study breast cancer in patients of Braga's Hospital Senology Unit, in Portugal. The primary motivation is to contribute to the understanding of the progression of breast cancer, within the Portuguese population, using a more complex statistical model assumptions than the traditional analysis, such as, for example, a possible existence of a serial correlation structure within a same subject observations. The main objective is to develop a joint model for longitudinal data (repeated measurements over time of a tumour marker) and survival (time-to-event of interest) of patients with breast cancer, where death from breast cancer is the event of interest.

Keywords: breast cancer, CEA, joint model, longitudinal, survival.

—⊙—
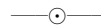
The analysed data gathers information on 540 patients, enclosing 50 variables, collected from medical records of the Hospital. We conducted a previous independent survival analysis from breast cancer related death, to understand what the possible risk factors are, for these patients. Also, an independent longitudinal analysis of tumour marker Carcinoembryonic antigen (CEA) was conducted, to identify risk factors related to the increase in its values.

For survival analysis we made use of the Cox proportional hazards model [1] and the flexible parametric model Royston-Parmar [4]. Generalized linear mixed effect models [2] were applied to study the longitudinal progression of the tumour marker. After the independent survival and longitudinal analysis, we took into account the expected association between the progression of the tumour marker values with patient's survival, and as such, we proceeded with a joint modelling of these two processes to infer on the association between them, adopting the methodology of random effects [3,5].

Results indicate that the longitudinal progression of CEA is significantly associated with the probability of survival of these patients. We also conclude that as the independent analysis returns biased estimates of the parameters, it is necessary to consider the relationship between the two processes when analysing breast cancer data.

Acknowledgements

References

[1]   Cox DR. (1972) Regression models and life-tables. *Journal of the Royal Statistical Society* **34**: 87–220.

[2]   Diggle, P., Heagerty, P., Liang, K. Y. and Zeger, S. (2002) *Analysis of Longitudinal Data*, 2nd edition. Oxford, England: Oxford University Press.

[3]   Henderson, R., Diggle, P. J. and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* **1(4)**: 465–480.

[4]   Royston P. and Parmar M. K. B. (2002) Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine* **21**: 2175–2197.

[5]   Wulfsohn, M. S. and Tsiatis, A. A. (1997) A Joint model for survival and longitudinal data measured with error. *Biometrics* **53**: 330–339.

Forecasting the number of overnight stays in Algarve

Isabel Cristina Lopes[1] and Eliana Costa e Silva[2,3] and
Aldina Correia[2]

[1]LEMA, ESEIG, Polytechnic of Porto, Portugal
[2]CIICESI, ESTGF, Polytechnic of Porto, Portugal
[3]Centre Algoritmi, University of Minho, Portugal

E-mail addresses: *cristinalopes@eu.ipp.pt;eos@estgf.ipp.pt; aic@estgf.ipp.pt*

Abstract

There is a strong dependence in Portuguese economy on tourism. In fact Portugal as a whole and specific regions in particular rely on the performance of tourism for their economic activity [1]. In this paper we present an analysis of the time series of the number of overnight stays in hotel units (hotels, guest houses or hostels, and tourist apartments) in Algarve, to forecast future stays. These values can be used by an hotel manager to predict their occupancy and to determine the best pricing policy.

Keywords: ARIMA models, hotel occupancy, seasonality, time series forecasting.

——⊙——
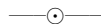
The dataset consisted of three variables concerning the number of overnight stays in: **(i)** hotels, **(ii)** guest houses or hostels, and **(iii)** tourist apartments, all of them located in Algarve, a tourist region in the south of Portugal. Monthly data from January 2006 until September 2015, giving a total of 117 observations, was collected from INE (the Portuguese National Statistic Institute) in November 17th, 2015. This data is based on reports by the hotels. The Algarve region was chosen because it is considered the most appealing location in Portugal for holidays, for its beautiful white sand beaches, golf courses, luxury hotels and resorts, and warm weather all year long.

The three time series showed seasonality and trend, with slight differences, therefore seasonal models were required [3,2].

The number of overnight stays in hotels showed a linear trend and seasonal effects depending on the level. Therefore an Holt-Winters exponential smoothing model was applied, with level parameter $\alpha = 0.114$, trend parameter $\gamma = 0.299$, and seasonality parameter $\delta = 0.999$. The stationary R-squared of this model was 0.287, the MAPE (Mean Absolute Percent Error) was 6.542% and the normalized BIC (Bayesian Information Criterion) was 20.962.

For modelling the number of overnight stays in guest houses or hostels, the ARIMA(1,0,0)(0,1,0) model was applied. This model includes an autoregressive component with lag 1 and seasonal differentiation of order 1. The Ljung Box Test resulted in a $p-value = 0.452 > 0.05$, thus the null hypothesis that the residuals are white noise is not rejected, therefore we may conclude that the model is adequate for the data, with a significant level of 5%. The PACF and ACF correlograms stayed between the critical levels for 5% significance, showing that the residuals are independent and identically distributed.

For modelling the number of stays in tourist apartments, a simple seasonal model without trend was used, with level parameter $\alpha = 0.5$, and seasonality parameter $\delta = 0.00002$, which reflected the fact that the seasonal component was constant along the

time. The stationary R-squared of this model was 0.624, the MAPE was 7.109% and the normalized BIC was 20.124.

Having validated the models, they were used to forecast the hotel occupancy for the months between October 2015 to December 2016. In addition to the calculation of the point forecasts, we also calculated the prediction intervals for these months.

References

[1]   Andraz, J., Gouveia, P. and Rodrigues, P. (2009) Modelling and forecasting the UK tourism growth cycle in Algarve. *Tourism Economics* **15(2)**: 323–338.

[2]   Brockwell, P. J. and Davis, R. A. (1996) *Introduction to Time Series and Forecasting*, New York, Springer-Verlag.

[3]   Gonçalves, E. and Mendes Lopes, N. (2008) *Séries temporais: Modelações lineares e não lineares*, 2nd edition, Lisboa, Sociedade Portuguesa de Estatística.

Causal influence between times series: a study on imports, exports and economic growth

Eliana Costa e Silva[1,2] and A. Manuela Gonçalves[3] and Marco Costa[4]

[1]CIICESI, ESTGF, Polytechnic of Porto, Portugal
[2]Centre Algoritmi, University of Minho, Portugal
[3]CMAT, DMA, University of Minho, Portugal
[4]CIDMA, ESTGA, University of Aveiro, Portugal

E-mail addresses: *eos@estgf.ipp.pt; mneves@math.uminho.pt; marco@ua.pt*

Abstract

Studying the stochastic process of several time series in simultaneous may improve forecast when compared to single time series analysis. For example, the history of a variable may help to forecast future values of other variables. Also, it may help to expect future developments of one or more variables if changes occur in some other variable (or variables). As a study case, in this work the causal direction of the effects of trade openness on economic growth in Portugal is studied.

Keywords: causality, international trade, multivariate time series, VAR.

———⊙———
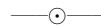
There are several phenomena which can be described by multivariate time series that exhibit causal relationships, from economics to neuroscience and neuroinformatics. For example, [4] presents a study on the relationship between imports and economic growth in Turkey. While [3], in neuroscience, study the causal interactions between brain regions based on multivariate autoregressive models. See also [1] for examples of neuroinformatics applications. Giving the large amount of applications and areas of applications, the study of causal interactions between multivariate time series is of great interest in applied statistics. Granger causality is a useful statistical test to search for the causal influence of one time series to another.

Multivariate vector autoregressive (VAR) models are used since they have proven to be especially useful for describing and forecasting the behavior of econometric time series. In economics the Export-Led Growth hypothesis (ELG), the Growth-Led Export hypothesis (GLE), the Import-Led Growth hypothesis (ILG)have been proposed and tested in different economies and periods (see e.g. [4]).

The data used in this work was obtained in *PORDATA, Base de Dados de Portugal Contemporâneo* and *INE, Instituto Nacional de Estatística*, and concerns the period from 1974 to 2012. For this period three variables are considered: Growth Domestic Product (GDP); volume importation and exportation. The mean value (and standard deviation) of GDP, import and export volumes were of approximately, 132.40(9.52), 189.80(214.89) and 122.90(140.7) million euros, respectively.

First, for testing the stationarity of each time series, the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) was used. For GDP, import and export the KPSS statistics were 0.1485, 0.1903, 0.2388, respectively, therefore, below the critical value for 5% of 0.1146. GDP, import and export are non-stationary and contain unit roots $I(1)$. Using also the KPSS test, the null hypothesis of stationarity is not rejected at 5% for the first order difference of the logaritmized time series. In order to determine the best order lag length of the VAR model

the Akaike information criterion (AIC), the Bayesian information criterion (BIC) and the Hannan-Quinn Information criterion (HQ) were used. For BIC and HQ the selected lag is 2 (BIC= -18.184, HQ= -18.684), and for AIC the lag is 3 however the results is quite close to lag 2 (AIC= -18.960 and AIC= -18.961, respectively). The multivariate Portmanteau test indicates that there are no strong serial or cross-correlations in the residuals of the simplified VAR(2) model. Furthermore, the covariance matrix of of residuals R_0 is not diagonal, therefore GDP, import and export are not instantaneous correlated. This means that GDP, import and export have instantaneous Granger causality.

In summary, we found that, for the analyzed period: **(i)** GDP does not depend on the lagged export, i.e. the ELG hypothesis is not satisfied **(ii)** export does not depend on GDP (and Import) for lag 1, but depends for lag 2, i.e., GLE hypothesis is satisfied (for lag 2) **(iii)** import depends on the GDP, i.e., ILG hypothesis is satisfied **(iv)** import depends on export lag 1 but not lag 2. **(v)** import depends on GDP and GDP depends on import, i.e., import and GDP are dynamically related.

References

[1] Youssofzadeh, V., Prasad, G., Naeem, M. and Wong-Lin, K. (2016) Temporal Information of Directed Causal Connectivity in Multi-Trial ERP Data using Partial Granger Causality. *Neuroinformatics* **14(1)**: 99–120.

[2] Hamilton, J. D. (1994) *Time series analysis*, vol. **2**. Princeton university press Princeton.

[3] Michalareas, G., Schoffelen, J.M., Paterson, G. and Gross, J. (2013). Investigating causality between interacting brain areas with multivariate autoregressive models of MEG sensor data. *Hum Brain Mapp* **34**: 890–913.

[4] Ugur, A. (2008) Import and Economic Growth in Turkey: Evidence from Multivariate VAR Analysis. *Journal of Economics and Business*, vol. **XI**: 54–75.

[5] Tsay, R. S. (2014) *Multivariate Time Series Analysis: with R and Financial Applications*. Wiley.

Organized Session 3

Numerical Methods in Fluids and Computation

Organizer: Fernando Carapau (Portugal)

Some results on the Frobenius coin problem

<u>Manuel B. Branco</u>[1,2] and J. C. Rosales[3]

[1]Departamento de Matemática, ECT, Universidade de Évora
[2]Centro de Investigação em Matemática e Aplicações (CIMA), IIFA-UE
[3]Departamento de Algebra, Universidad de Granada

E-mail addresses: *mbb@uevora.pt; jrosales@ugr.es*

Abstract

Let $\mathbb{N}$ denote the set of nonnegative integers. A numerical semigroup is a subset S of $\mathbb{N}$ closed under addition, it contains the zero element and has finite complement in $\mathbb{N}$. Given a nonempty subset A of $\mathbb{N}$ we will denote by $\langle A \rangle$ the submonoid of $(\mathbb{N}, +)$ generated by A, that is,

$$\langle A \rangle = \{\lambda_1 a_1 + \cdots + \lambda_n a_n \mid n \in \mathbb{N}\backslash\{0\}, a_i \in A, \lambda_i \in \mathbb{N} \text{ for all } i \in \{1, \ldots, n\}\}.$$

It is well known that $\langle A \rangle$ is a numerical semigroup if and only if $\gcd(A) = 1$. If S is a numerical semigroup and $S = \langle A \rangle$ then we say that A is a system of generators of S. Moreover, if $S \neq \langle X \rangle$ for all $X \subsetneq A$, then we say that A is a minimal system of generators of S. It is well known that every numerical semigroup admits a unique minimal system of generators, which in addition is finite. The cardinality of its minimal system of generators is called the embedding dimension of S, denoted by e(S). Following a classic line, two invariants have special relevance to a numerical semigroups: the greatest integer that does not belong to S, called the Frobenius number of S denoted by F(S), and the cardinality of $\mathbb{N}\backslash S$, called the gender of S denoted by g(S). The Frobenius coin problem (often called the linear Diophantine problem of Frobenius) consists in finding a formula, in terms of the elements in a minimal system of generators of S, for computing F(S) and g(S). This problem was solved by Sylvester for numerical semigroups with embedding dimension two. Sylvester demonstrated that if $\{n_1, n_2\}$ is a minimal system of generators of S, then F(S) $= n_1 n_2 - n_1 - n_2$ and g(S) $= \frac{1}{2}(n_1 - 1)(n_2 - 1)$. The Frobenius coin problem remains open for numerical semigroups with embedding dimension greater than or equal to three. In this talk we will present some classes of numerical semigroups for which this problem is solved (see References).

Keywords: numerical semigroup, Frobenius number, embedding dimension, Genus, Mersenne numbers.

Acknowledgements

References

[1]   Ramirez Alfonsín, J. L. (2005) *The Diophantine Forbenius Problem.* London: Oxford University Press.

[2]   Rosales, J. C. and Branco, M. B. (2011) The Frobenius problem for numerical semi-groups with multiplicity four. *Semigroup Forum* **83(3)**: 468–478.

[3]   Rosales, J. C. and Branco, M. B. (2011) The Frobenius problem for numerical semi-groups. *J. Number Theory* **131(12)**: 2310–2319.

[4]   Rosales, J. C. and Branco, M. B. (2003) Irreducible numerical semigroups. *Pacific J. Math.* **209(1)**: 131–143.

[5]   Rosales, J. C. and García-Sánchez, P. A. (2009) Numerical semigroups. *Developments in Mathematics* **20**, New York: Springer.

[6]   Rosales, J. C., Branco, M. B. and Torrão, D. (2015) The Frobenius problem for Thabit numerical semigroups, to appear in *J. Number Theory.*

[7]   Rosales, J. C., Branco, M. B. and Torrão, D. (2014) The Frobenius problem for Repunit numerical semigroups, submitted.

[8]   Rosales, J. C., Branco, M. B. and Torrão, D. (2015) The Frobenius problem for Mersernne numerical semigroups, submitted.

A one-dimensional model for blood flow based on Cosserat Theory

Fernando Carapau[1,2] and <u>Paulo Correia</u>[1,2]

[1]Departamento de Matemática, ECT, Universidade de Évora
[2]Centro de Investigação em Matemática e Aplicações (CIMA), IIFA-UE

E-mail addresses: *flc@uevora.pt; pcorreia@uevora.pt*

Abstract

In this talk, we study the unsteady motion of a generalized viscoelastic fluid of third-grade where specific normal stress coefficient depends on the shear rate by using a power-law model. For that, we use the Cosserat theory approach which reduces the exact three-dimensional equations to a system depending only on time and on a single spatial variable. This one-dimensional system is obtained by integrating the linear momentum equation over the cross-section of the tube, taking a velocity field approximation provided by the Cosserat theory. The velocity field approximation satisfies exactly both the incompressibility condition and the kinematic boundary condition. From this reduced system, we obtain unsteady equations for the wall shear stress and mean pressure gradient depending on the volume flow rate, Womersley number, viscoelastic coefficients and flow index over a finite section of the tube geometry with constant circular cross-section. Attention is focused on some numerical simulations.

Keywords: one-dimensional model, blood flow model, shear-thinning, volume flow rate, Cosserat theory.

—⊙—
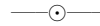
This work is motivated by the study of mathematical models for blood flow in the vascular system. The mathematical modelling of blood flow is a complex and challenging problem. Some issues related to that difficulty are the following: the human circulatory system is a closed network of vessels carrying blood; blood is a suspension of particles in a fluid called plasma; the heart is the pump that moves blood through the system by muscular contraction and ventricular volume reduction; the vessels can be regarded as hollow tubes with different geometries and scales; the study of specific diseases such as artherosclerosis, aneurysms, hypertension and high level of blood viscosity also contribute to the complexity of the mathematical modelling; the structure of the vessels consists in many layers, with different mechanical characteristics and interacts mechanically with the blood flow field (fluid–structure interaction), for example, the wall vessels can be permeable and elastic; in large vessels blood may be considered as an homogeneous fluid with the standard behavior of a Newtonian fluid and the wall may be considered elastic or mildly viscoelastic; in small vessels the wall is characterized by a strong branching and may in general be considered rigid; in small vessels phenomena like aggregation and deformability of red blood cells have great influence on the rheological behavior of blood specially on its viscosity at low shear rate. Blood exhibits non-Newtonian phenomenon due to shear-thinning viscosity and viscoelasticity effects in small vessels. Considering our physical problem, we propose a new three-dimensional model to study blood flow in small vessels, with which it is possible to capture the shear-thinning viscosity and viscoelasticity effects of the blood. Our new model is a generalization of the thermodynamically compatible fluid of third-grade (see [1] and [2]). A possible simplification related with the new three-dimensional model is to consider the evolution of average flow quantities

using simpler one-dimensional models. Usually, in the case of flow in a tube, the classical one-dimensional models are obtained by imposing additional assumptions and integrating both the equations of conservation of linear momentum and mass over the cross-section of the tube. Here, we introduce a one-dimensional model based on the hierarchical approach developed by Caulk and Naghdi [3]. From this reduce system, we obtain (using the Runge-Kutta method to solve the ODE) the unsteady equations for the mean pressure gradient and wall shear stress depending on the volume flow rate, Womersley number, viscoelastic coefficients and flow index over a finite section of the tube geometry. Finally, we compare our results with the resulting obtained for the same physical problem with other different proposed models, see for example [4] and [5].

Acknowledgements

References

[1] Truesdell, C. and Noll, W. (1992) The non-linear field theories of mechanics, 2nd edition , Springer, New-York.

[2] Fosdick, R.L. and Rajagopal, K.R. (1980) Thermodynamics and stability of fluids of third grade, *Proc. R. Soc. Lond. A.*, **339**, 351–377.

[3] Caulk, D.A. and Naghdi, P.M. (1987) Axisymmetric motion of a viscous fluid inside a slender surface of revolution, *Journal of Applied Mechanics*, **54** (1), 190–196.

[4] Robertson, A.M. and Sequeira, A. (2005) A director theory approach for modeling blood flow in the arterial system: an alternative to classical 1d models, Math. Models Methods Appl. Sci., **15** (6), 871–906.

[5] Carapau, F. and Sequeira, A. (2006) 1D Models for blood flow in small vessels using the Cosserat theory, *WSEAS Transactions on Mathematics*, **5** (1), 54–62.

High Weissenberg number problem: state of art

Marília Pires[1,2,3]

[1]Departamento de Matemática, ECT, Universidade de Évora
[2]Centro de Investigação em Matemática e Aplicações (CIMA), IIFA-UE
[3]Centro de Matemática Computacional e Estocástica (CEMAT, IST, UTL)

E-mail address: *marilia@uevora.pt*

Abstract

A outstanding and crucial problem in computational rheology of non-Newtonian fluids, in particular viscoelastic fluids Oldroyd type, is the non-convergence of the numerical algorithms used in the simulation flow of such fluids with high Weissenberg values. Despite some progress in solving high Weissenberg number problem (HWNP), the solution has not been found. The purpose of this contribution is to focus on the issues that arise in the numerical simulation of fluid flow with high Weissenberg numbers and summarize the latest developments to solve this problem, including the work undertaken and continued by Marília Pires, Tomas Bodnar and J. Janela [1].

Keywords: Oldroyd, Weissenberg number, numerical instabilities, convergence.

Acknowledgements

References

[1]   Bodnár, T, Pires, M. and Janela, J. (2014) Blood Flow Simulation Using Traceless Variant of Johnson–Segalman Viscoelastic. *Model, Math. Model. Nat. Phenom.* **9(6)**: 117–141.

[2]   Fattal, R. and Kupferman, R. (2004) Constitutive laws for the matrix–logarithm of the conformation tensor. *J. Non-Newton. Fluid Mech.* **124**: 281–285.

Animal movement: a simple chaotic model approach

C. Correia Ramos[1,2]

[1]Departamento de Matemática, ECT, Universidade de Évora
[2]Centro de Investigação em Matemática e Aplicações (CIMA), IIFA-UE

E-mail address: *ccr@uevora.pt*

Abstract

The study of animal movement has attracted much attention since Mandelbrot observed that it shows scale invariance and fractal properties. The properties, of statistical nature, manifest as a power law-like tail in the distribution of movement lengths. Most of the approaches in the literature follows these aspects, and have been applied to several different animal movements, from bird flight, insect movement and also human movement. We propose a discrete dynamical system, deterministic, which serves as a model to produce and study a variety of types of movements, either two dimensional or three dimensional. The complexity description and the characterization of the movements are based on the topological classification of the discrete dynamical system. Several experiments are proposed to understand how this model can be useful in practice.

Keywords: animal movement, deterministic, discrete dynamical system, statistical properties, topological classification.

References

[1] Mandelbrot, B. (1977) *The Fractal Geometry of Nature*, New York, Freeman.
[2] Viswanathan G. M., et al. (1996) Levy flight search patterns of wandering albatrosses. *Nature* **381**: 413–415.
[3] Edwards, A. M., et al. (2007) Revisiting Lévy flight search patterns of wandering albatrosses, bumblebees and deer. *Nature* **449**: 1044–1048.
[4] Mashanova, A., Oliver, T. H. and Jansen, V. A. A. (2010) Evidence for intermittency and a truncated power law from highly resolved aphid movement data. *J. R. Soc. Interface* **7**: 199–208.
[5] Brockmann, D., Hufnagel, L. and Geisel, T. (2006) The scaling laws of human travel. *Nature* **439**: 462–465.
[6] Turchin, P. (1996) Fractal analyses of animal movement: A critique. *Ecology* **77**: 2086–2090.
[7] Sergei, P., Alla, M. and Vincent, A. A. J. (2011) PNAS **108(21)**: 8704–8707.

Organized Session 4

Statistics and Modelling

Organizer: Milan Stehlík (Austria)

Stochastic bivariate models for bio-medical phenomena with two distinct kinds of dependencies between underlying random quantities

Jerzy K. Filus[1] and Lidia Z. Filus[2]

[1]Department of Mathematics and Computer Science, Oakton Community College, Des Plaines, Il 60016, USA
[2]Department of Mathematics, Northeastern Illinois University, Chicago, IL 60625, USA

E-mail addresses: *jkfilus98@gmail.com; L-Filus@neiu.edu*

Abstract

Bivariate probability densities are constructed and applied as stochastic models for some bio-medical phenomena. We consider two classes of models; each of them describing different physical and corresponding distinct stochastic mechanism of dependence between the same random quantities. The considered random quantities also have some bio-medical meanings that may vary from case to case. One of the two "dependence classes" contains (exclusively) the first type of Gumbel bivariate exponential pdfs. This probabilistic model and some of its extensions describe first kind of the stochastic dependencies as an "ingredient" of the overall model we construct. The second ingredient of any single constructed bivariate model is any member of other class of pdfs that contains the bivariate pseudo-exponential generalization of the exponential distributions that were constructed in our previous works. After some elaboration of the models we construct the fusion of the considered two classes, such that the resulting composite "Gumbel-pseudoexponential" densities become, at least potentially, stochastic models for some aspects of bio-objects subjected to two distinct physical processes. Each of the processes, separately, may add its part to the stochastic behavior of the two interacting random bio-quantity. As it will be pointed out, the fusion of the two models into the one has a nice analytical reflection by a nontrivial factorization of the composite mathematical description.

Keywords: stochastic models in biomedical investigations fusion of two stochastic mechanisms into one model, Gumbel-pseudoexponential bivariate probability densities, factorization.

References

[1] Filus, J. K. and Filus, L. Z. (2006) On Some New Classes of Multivariate Probability Distributions. *Pakistan Journal of Statistics* **1(22)**: 21–42.
[2] Gumbel, E. J. (1960) Bivariate exponential distributions. *J. Amer. Statist. Assoc.* **55**: 698–707.

Impact of Education Support Software on EMM Course Student Achievement

Pavel Kolman[1] and Josef Holoubek[2]

[1]Mendel University in Brno, Department of Statistics and Operation Analysis (FBE),
Zemědělská 1, 613 00 Brno, Czech Republic
[2]Mendel University in Brno, Department of Statistics and Operation Analysis (FBE),
Zemědělská 1, 613 00 Brno, Czech Republic

E-mail addresses: *xkolman@mendelu.cz; holoubek@mendelu.cz*

Abstract

The authors of the paper have long been engaged in the development of educational software for optimization courses. Currently, under the so-called "Education Software Support in the Economic-Mathematical Methods Course of Faculty of Business and Economics" grant and, based on their gained knowledge they have partly created and in part innovated educational optimization software for the above mentioned course. The paper summarizes the created optimization software and compares the differences in achievements of students who study using this software and in achievements of students who don't. The comparative process of both groups will be carried out following various criteria. The primary objective is, of course, the demonstrability of the differences between both groups of monitored students. Other criteria include, for example, determining the differences between full-time and part-time students as well as the differences in results (achievements) based on individual use of various types of software (programs), trends in academic progress, etc.

The paper aims to analyze the impact of e-learning on the Economic-Mathematical Methods Course student achievement and to use, respectively implement the results in grant implementation (execution) as well as in the implementation and further development of educational software.

Testing the difference between two quantiles from independent populations

<u>Martina Mittlboeck</u>[1] and Harald Heinzl[1]

[1]Center for Medical Statistics, Informatics, and Intelligent Systems
Medical University of Vienna, Spitalgasse 23, A-1090 Vienna, Austria

E-mail addresses: *martina.mittlboeck@meduniwien.ac.at; harald.heinzl@meduniwien.ac.at*

Abstract

An empirical distribution function estimator for the difference of two order statistics has been used to develop a test of the difference between two quantiles from independent populations [1]. Properties of this test, in particular, effects of sample size, location of investigated quantiles, and underlying distributions have been studied. Thereby, common and also rather extreme situations have been considered [2].

Small to moderate sample sizes, tail quantiles, and quantiles which do not coincide with the expectation of an order statistic can lead to both, conservative but also extreme liberal behavior of the test. Consequently, this statistical test should be used with care, in particular, in those areas of application where liberal statistical tests are usually not acceptable, like marketing authorization of medicines or medical devices.

References

[1] Hutson, A. (2009) A distribution function estimator for the difference of order statistics from two independent sampls. *Stat Papers* **50**: 203–208.
[2] Heinzl, H. and Mittlboeck, M. (2016) Assessing a Hypothesis Test for the Difference between Two Quantiles from Independent Populations. *Communications in Statistics - Simulation and Computation* (Epub ahead of print; DOI: `10.1080/03610918.2015.1096379`).

Effects of violations of prerequisites in sequential designs on sample size and power

Karl Moder

University of Natural Resources and Life Sciences, Vienna, Austria

E-mail address: *karl.moder@boku.ac.at*

Abstract

In sequential designs for means (one-and two-sample problems) the test statistic is compared with the $K-$percentile of the standard normal distribution, Therefore the test is approximate and the formulae for the expected sample size too. By simulation for several situations empirical sample sizes for the O'Brien Fleming test and for triangular sequential designs as described in [4] were calculated . The tests and situations considered are: The group sequential test O'Brien-Fleming Method (1979), the triangular method of Kittelson and Emerson (1999) and the method of Whiteheads and Stratton for continuous monitoring. Simulations were carried out for several situations of assumed population variances and deviations from these assumptions for the sample distributions. Various situations of non normal distributions in regard to skewness and kurtosis (based on [1]) and their influence on alpha and power were examined. All simulations very carried out using SAS 9.2 (2008) by means of proc seqdesign and proc seqtest.

Keywords: group sequential test, continuous monitoring, sample size.

References

[1] Fleishman, A. I. (1978) A method for simulating non-normal distributions. *Psychometrika* **43(4)**: 521–532.
[2] Kittelson, J. M. and Emerson, S. S. (1999) A Unifying Family of Group Sequential Test Designs. *Biometrics* **55**: 874882.
[3] O'Brien, P. C. and Fleming, T. R. (1979) A Multiple Testing Procedure for Clinical Trials. *Biometrics* **35**: 549556.
[4] Rasch, D., Pilz, J., Gebhardt, A. and Verdooren, R. L. (2011) *Optimal Experimental Design with R*, Boca Raton, Chapman and Hall.
[5] SAS Institute Inc. (2008) *SAS/STAT*® *9.2 Users Guide*. Cary, NC: SAS Institute Inc.
[6] Whitehead, J. and Stratton, I. (1983) Group Sequential Clinical Trials with Triangular Continuation Regions. *Biometrics* **39**: 227236.

On robust testing for normality in autoregressive conditional heteroscedasticity models

Luboš Střelec[1,2] and Milan Stehlík[3,4]

[1]Mendel University in Brno
[2]Department of Statistics and Operation Analysis
[3]Johannes Kepler University in Linz
[4]Institute of Applied Statistics

E-mail addresses: *lubos.strelec@mendelu.cz; milan.stehlik@jku.at*

Abstract

Classic parametric statistical significance tests, such as analysis of variance and $t-$test in least squares regression, are widely used by researchers in many disciplines, e.g. psychology, economics and finance, among others. For classic parametric tests to produce accurate results, the basic assumptions underlying them, e.g. normality and homoscedasticity, must be satisfied. However, these assumptions are rarely met when analysing real data sets. The use of classic parametric methods with violated assumptions can result in the inaccurate computation of $p-$values, effect sizes as well as confidence intervals, which may lead to substantive errors in the interpretation of analysed data sets. In other words, failure to asses non-normality may lead to incorrect results, because significant deviations from normality can substantially affect the performance of usual statistical inference techniques.

In this contribution, we focus on simulations applied to explore power of several classical and robust tests of normality in models with dependence, especially models with presence of autoregressive conditional heteroscedasticity. Within this work we also introduce the RT class of normality tests based on robust modification of the classical Jarque-Bera test statistic.

Keywords: autoregressive conditional heteroscedasticity, error terms, robustness, RT tests of normality, testing for normality.

Acknowledgements

References

[1] Bartolucci, F. and Scaccia, L. (2005) The use of mixtures for dealing with non-normal regression errors. *Computational Statistics & Data Analysis* **48(4)**: 821–834.

[2] Erceg-Hurn, D. and Mirosevich, V. (2008) Modern robust statistical methods: an easy way to maximize the accuracy and power of your research. *American Psychologist* **63**: 591–601.

[3] Önder, A. Ö. and Zaman, A. (2003) A test for normality based on robust regression residuals. *Dutter, R., Filzmoser, P., Gather, U., Rousseeuw, P.J. (eds.), Developments in Robust Statistics. Physica-Verlag, Heidelberg*: 296–306.

[4] Önder, A. Ö. and Zaman, A. (2005) Robust tests for normality of errors in regression models. *Economics Letters* **86(1)**: 63–68.

[5] Stehlík, M., Fabián, Z., and Střelec, L. (2012) Small sample robust testing for normality against Pareto tails. *Communications in Statistics – Simulation and Computation* **41(7)**: 1167–1194.

[6] Stehlík, M., Střelec, L., and Thulin, M. (2014) On robust testing for normality in chemometrics. *Chemometrics and Intelligent Laboratory Systems* **130**: 98–108.

[7] Thode, H. C. (2002) *Testing for normality*, New York, Marcel Dekker.

Statistical Quality Control: Theory and Practice

Organizer: Fernanda Figueiredo (Portugal)

Joint Control Charts applied to monitoring the concentration of particles with health effects in occupational environment

Elisabete Carolino[1,3], Rosário Ramos[2,4], Carla Viegas[1,3]
and Susana Viegas[1,3]

[1]Instituto Politécnico de Lisboa, Escola Superior de Tecnologias da Saúde
(ESTeSL-IPL), Portugal
[2]Universidade Aberta, Lisboa, Portugal
[3]Research Group Environment and Health
[4]CMAF-CIO Center for Mathematics, Fundamental Applications and Operations
Research, Universidade de Lisboa

E-mail addresses: *etcarolino@estesl.ipl.pt; marosram@uab.pt; carla.viegas@estesl.ipl.pt; susana.viegas@estesl.ipl.pt*

Abstract

The health effects associated with occupational exposure to suspended particulates, because of their importance, have been studied by several authors. In this context, it is important to ensure a monitoring system of concentration of particles with adverse health effects, in occupational environments, in support of the implementation of corrective and preventive measures.

We present here an application of statistical quality control charts methodology to a selection of variables that are among the particles with adverse health effects. Classical control systems using the Shewhart charts are often difficult to manage in real time. Moreover, the use of independent control charts when measurements relate to simultaneous observations and possibly correlated, distorts multivariate vector monitoring through type I error and the probability of being within the control limits to be not adjusted to real value. This distortion of the control procedure increases with the number of variables to be controlled.

We experimented with an alternative method that uses the graphs of Shewhart charts in order to use a simplified method to a plan for monitoring the particle concentration values measured in different locations on the same occupational environment. Multivariate charts are used based on Hotelling's T^2 supplemented with graphics Shewhart and Bonferroni limits.

The study was performed in animal feed production industry. The device that performed the measurements provides information on the concentration of five different particle diameter (PM0.5, PM1.0, PM2.5, and PM5.0 PM10.0) and which correspond to different degrees of severity, as respect to effects on health.

Keywords: control charts, multivariate process, T^2 Hotelling charts, Bonferroni limits, occupational exposure, particle concentration.

References

[1] Montgomery, Douglas C. (2009) *Introduction to Statistical Quality Control.* 6th Ed. Wiley & Sons.

[2] Tracy, N. D., Young, J. C. and Mason, R. L. (1995) Multivariate Control Charts for individual observations. *Journal of Quality Technology* **24(2)**: 88–95.

Acceptance sampling plans for sensory and chromatography analyses

Fernanda Otília Figueiredo[1,3], Adelaide Maria Figueiredo[1,3,4]
and Maria Ivette Gomes[2,3]

[1]Universidade do Porto, Faculdade de Economia, Portugal
[2]Universidade de Lisboa, Faculdade de Ciências, Portugal
[3]CEAUL, Universidade de Lisboa, Portugal
[4]LIAAD–INESC TEC Porto, Universidade do Porto, Portugal

E-mail addresses: *otília@fep.up.pt; adelaide@fep.up.pt; ivette.gomes@fc.ul.pt*

Abstract

The performance of some acceptance sampling plans for use in sensory and chromatography analysis is evaluated in terms of the probability of acceptance of the lots and its average outgoing quality level.

Keywords: acceptance sampling plans, chromatography analysis, sensory analysis, statistical quality control.

———⊙———
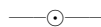
Haloanisoles cause musty off-aromas (off-odors) in many consumer products, for instance in the food industry, but also affects the quality of many other manufacturing goods. Sensory analysis together with analytical methods based on chromatography measurements are usually performed for identification and quantification of these compounds. Apart from being necessary to fulfill the multiple regulations imposed to the industrial and manufacturing industries, these analysis are also useful for marketing decisions, such as product positioning with respect to competitors, customer relationship management and price policies.

Acceptance sampling is one of the most important off-line techniques in Statistical Quality Control. Traditionally it is used before and after the process production to filter the raw-material from suppliers and to screen the finished products for customers. But in some industries it is also performed on-line to filter the products along the different phases of its production. The decision of accepting or rejecting a lot of products is based on appropriate predefined sampling plans. The most commonly used are the attributes and the variables sampling plans. For further details see, for instance, [4], [3] and [1].

Based on a consulting problem we evaluate the performance of some acceptance sampling plans of interest for the company, for use in sensory and chromatography analysis. These sampling plans are developed to inspect batches of raw material or final products from an industrial process, and are evaluated in terms of the probability of acceptance of the lots and its average outgoing quality level. In a previous study Figueiredo *et al.* [2] used the bootstrap methodology combined with Monte Carlo simulations to evaluate the performance of complex acceptance sampling plans in the detection of chemical substances in lots of raw-material. In this study we compare single and double sampling plans that have being used in the company for sensory analysis, and we suggest some improvements on these plans with the expense of increasing the sample size. In what concerns the analysis based on the chromatography measurements, first we model such type of data with an

inflated Pareto distribution, and then, we develop and compare some variables sampling plans based on this distribution.

References

[1] Carolino, E. and Barão, I. (2013) Robust methods in acceptance sampling. *Revstat* **11(1)**: 67–82.

[2] Figueiredo, F., Figueiredo, A. and Gomes, M. I. (2014) Comparison of Sampling Plans by Variables using the Bootstrap and Monte Carlo Simulations. In *Proceedings of the International Conference of Computational Methods in Sciences and Engineering 2014 (ICCMSE 2014)*, 535–538.

[3] Gomes, M. I. (2011) Acceptance sampling. In *International Encyclopedia of Statistical Science, Part 1*, Lovric, M. (ed.), 5–7. ISBN: 978-3-642-04897-5, Springer.

[4] Montgomery, D. C. (2009) *Introduction to Statistical Quality Control: a Modern Introduction, 6th edition*, John Wiley & Sons.

The effect of using an incorrect model on the performance of risk-adjusted control charts

Maria João Inácio[1,4], Paulo Infante[2,4] and
Fernanda Otília Figueiredo[3,5]

[1]Instituto Politécnico de Tomar
[2]Universidade de Évora, Escola de Ciências e Tecnologia
[3]Universidade do Porto, Faculdade de Economia
[4]CIMA/IIFA, Universidade de Évora
[5]CEAUL, Universidade de Lisboa

E-mail addresses: *mjantunes@ipt.pt; pinfante@uevora.pt; otilia@fep.up.pt*

Abstract

Over the past few years risk-adjusted control charts have become popular for monitoring processes in health context. In these control charts, a model is adjusted to predict, for each patient, the risk of an adverse event. The suitability of the model affect the performance of the control charts. In this work, considering a real data set, we studied the effect of the estimation error, prediction error and mortality rate changes on the performance of the risk-adjusted control charts. The effect was measured by the change on the average run length.

Keywords: attributes control charts, estimation error, prediction error, risk-adjusted, statistical quality control.

—⊙—
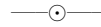
Control charts have become popular in the health context. However, unlike the most industrial processes where the subjects are relatively homogeneous, concerning the health context the patients often vary widely in terms of the base-line risk of the event that we are interested to monitoring, and in order to avoid undesired false alarms, this variability needs to be considered when designing the control charts. This can be achieved by estimating and taking into account, for each patient, the probability of occurrence of the event that is to be monitored. Control charts that implement this procedure are often referred as Risk-Adjusted Control Charts.

Over the past few years several control charts for attributes adapted to risk adjusted have been proposed. The most popular are the RAP-charts (Risk-Adjusted P-charts), the VLAD (Variable Life-Adjust Display) or CRAM (Cumulative Risk-Adjusted Mortality) charts, the RASPRT (Risk- Adjusted Sequential Probability Ratio Tests) and RARSPRT (Risk- Adjusted Resetting Sequential Probability Ratio Tests) charts, the RAEWMA (Risk-Adjusted Weight Exponentially Moving Average) and RACUSUM (Risk-Adjusted Cumulative SUM) control charts. For details, see, for example, Cook et al. [1] and Grigg and Farewell [2].

The adjustment of the risk model is an important issue in the design of these charts. However, it is not possible to know if the initial set model is suitable for future observations due to the changes that may occur during the monitoring process. Although there is already a variety of control charts adapted to risk adjusted, until now, according to our research, very little has been done to assess the impact of using an incorrect model. With

this propose, two different approaches for the RACUSUM chart were discussed by Jones and Steiner [3] and Tian et al. [5].

The aim of this paper is to study the sensitivity of the charts according to the use of an incorrect model. The ARL (Average Run Length) was the measure chosen to evaluate the performance of the charts.

We used a real data set, studied initially by Steiner et al. [4], consisting of 6994 operations from a single UK cardiac surgical central over the period 1992-1998, and our event of interest is the 30 day post-operative mortality rate. We assess the effect of the estimation error, prediction error and mortality rate changes on the performance of risk adjusted control charts in the estimation of the model parameters. Our results show that the control charts are sensitive to model changes. It can be seen that changes that underestimate the risk of death have larger impact on the ARL of the charts than changes that overestimate it. In addition, the impact of the underestimation is greater as larger is the magnitude of the error. When we compared the performance of the charts designed to detect different changes in the odds ratio, we realized that the impact of overestimation on the ARL is smaller on the charts designed to detect large changes.

References

[1]   Cook, D., Duke, G., Hart, G., Pilcher, D. and Mullany, D. (2008) Review of the application of risk-adjusted charts to analyse mortality outcomes in critical care. *Critical Care and Resuscitation* **10**: 239–251.

[2]   Grigg O. and Farewell, V. (2004) An overview of risk adjusted charts. *Journal of the Royal Statistical Society* **167**: 523–539.

[3]   Jones, M. and Steiner, S. (2012) Assessing the effect of estimation error on risk-adjusted CUSUM chart performance. *International Journal for Quality in Health Care* **24(2)**: 176–178.

[4]   Steiner, S., Cook, R., Farewell, V. and Treasure, T. (2000) Monitoring Surgical Performance Using Risk Adjusted Cumulative Sum Charts. *Biostatistics* **1**: 441–452.

[5]   Tian, W., Sun, H, Zang, X. and Woodall, W. (2014) The impact of varying patient populations on the in-control performance of the risk-adjusted CUSUM chart. *International Journal for Quality in Health Care* **27(1)**: 31–36.

A nonparametric approach for simultaneous monitoring of event frequency and magnitude

Amitava Mukherjee[1], Yuan Cheng[2] and Min Xie[2]

[1]XLRI - Xavier School of Management
Production, Operation and Decision Sciences Area
Jamshedpur-831001, India
[2]Department of Systems Engineering and Engineering Management
College of Science and Engineering,
City University of Hong Kong,
83 Tat Chee Ave, Kowloon Tong, Hong Kong

E-mail address: *amitmukh2@yahoo.co.in*

Abstract

Simultaneous monitoring of the event frequency and magnitude, such as, time between the two successive contract labour strikes and the duration of the strike; call centre response times and its service quality; machine downtime and the consequential financial loss or time between the two major earthquakes and the magnitude in Richter scale etc., is an important research topic in various areas ranging from production, operations and manufacturing to event environmental science. Most of the traditional approaches for simultaneous monitoring of event frequency and magnitude either assume that the event frequency and magnitude are two independent variables or consider a bivariate parametric model, such as bivariate gamma or bivariate normal. In reality, it is often difficult to justify such assumptions and therefore, in the present work, we introduce a nonparametric Shewhart type procedure taking into account the dependence between the event frequency and its magnitude. We assume that the underlying process distribution is unknown but random samples from retrospective phase are obtained as preliminary information. The concept of ranks and empirical copula are utilized in constructing the proposed nonparametric monitoring procedure. Our proposed procedure can also be applied in monitoring shifts in any general bivariate processes and using the proposed follow-up scheme, practitioners may identify the variable where shift has actually occurred. Design and implementation procedure of the proposed scheme are described in details. We illustrate our proposed scheme with a real example and show the various construction steps. Some properties of the proposed procedure are studied through Monte-Carlo simulation.

Keywords: bivariate process, Cramér-von Mises criterion, empirical copula, Monte-Carlo, nonparametric, phase–II Shewhart chart.

On the influence of control statistics in the performance of joint schemes for location and spread

<u>Patrícia Ferreira Ramos</u>[1,2], Manuel Cabral Morais[3,2],
António Pacheco[3,2] and Wolfgang Schmid[4]

[1]Universidade Autónoma de Lisboa, Portugal
[2]CEMAT, Instituto Superior Técnico
[3]Universidade de Lisboa, Portugal
[4]European University Viadrina, Frankfurt(Oder), Germany

E-mail addresses: *patriciaferreira@tecnico.ulisboa.pt; maj@math.ist.utl.pt;*
apacheco@math.tecnico.ulisboa.pt; schmid@euv-frankfurt-o.de

Abstract

The performance of a product frequently relies on more than one quality characteristic. In such a setting, joint schemes are used to determine whether or not we are in the presence of unfavourable disruptions in the location and spread of a vector $\mathbf{X}$ of p quality characteristics.

A common joint scheme comprises two individual control charts: one for the mean vector ($\boldsymbol{\mu}$) based on the control statistic $T^{(1)}$, a weighted Mahalanobis distance between the vector of sample means ($\bar{\mathbf{X}}$) and the target mean vector ($\boldsymbol{\mu}_0$); another one for the covariance matrix ($\boldsymbol{\Sigma}$) depending on the control statistic $U^{(1)}$, the ratio between the determinants of the sample covariance matrix ($\mathbf{S}$) and the target covariance matrix ($\boldsymbol{\Sigma}_0$).

As we are well aware that there are plenty of quality control practitioners who are still reluctant to use sophisticated control statistics, this paper tackles two other Shewhart-type joint schemes for the location and spread: the first alternative is based on the multivariate analogue of the univariate t statistic, $T^{(2)} = n\,(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)^\top \mathbf{S}^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu}_0)$; the second alternative joint scheme depends on the maximum likelihood estimator of $\boldsymbol{\Sigma}$ when $\boldsymbol{\mu}$ is at its target value $\boldsymbol{\mu}_0$, $\mathbf{S}^* = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{X}_i - \boldsymbol{\mu}_0)(\mathbf{X}_i - \boldsymbol{\mu}_0)^\top$.

To sum up, we shall consider the following pairs of control statistics, which can be thought as the $p-$variate extensions of the ones used by [1]:

	Rule to trigger a signal while using the chart for	
Scheme	$\boldsymbol{\mu}$	$\boldsymbol{\Sigma}$
1	$T^{(1)} = n\,(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_0^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu}_0) > UCL_{\boldsymbol{\mu}}^{(1)}$	$U^{(1)} = \frac{det((n-1)\mathbf{S})}{det(\boldsymbol{\Sigma}_0)} > UCL_{\boldsymbol{\Sigma}}^{(1)}$
2	$T^{(2)} = n\,(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)^\top \mathbf{S}^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu}_0) > UCL_{\boldsymbol{\mu}}^{(2)}$	$U^{(2)} \equiv U^{(1)} > UCL_{\boldsymbol{\Sigma}}^{(2)}$
3	$T^{(3)} = n\,(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)^\top (\mathbf{S}^*)^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu}_0) > UCL_{\boldsymbol{\mu}}^{(3)}$	$U^{(3)} = \frac{det(n\mathbf{S}^*)}{det(\boldsymbol{\Sigma}_0)} > UCL_{\boldsymbol{\Sigma}}^{(3)}$

Keywords: multivariate normal output, run length, statistical process control.

———⊙———
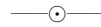
The quicker the joint scheme triggers a signal in the presence of a shift, in the process mean or covariance matrix, the better; nonetheless, the joint scheme should also provide a good protection against false alarms. Unsurprisingly, the number of samples collected until

a signal is triggered by a joint scheme — the run length (RL) — is a vital performance measure.

In view of the fact that the samples are assumed to be independent, the RL of any of the joint schemes has a geometric distribution and its average run length (ARL) is equal to the reciprocal of the associated power function, given that the mean vector and the covariance matrix are equal to $\boldsymbol{\mu} = \boldsymbol{\mu}_0 + \boldsymbol{\Sigma}_0^{1/2} \boldsymbol{\delta}/\sqrt{n}$ and $\boldsymbol{\Sigma} = [\sigma_{ij}]_{i,j=1,\dots,p}$, as in [3].

The control statistics $T^{(1)}$ and $U^{(1)}$ have very well known properties, one of them is independence and the associated average run length, $ARL^{(1)}(\boldsymbol{\delta}, \boldsymbol{\Sigma})$, follows quite trivially. As for the ARL of the remaining joint schemes, the lengthy and technical derivations of $ARL^{(i)}(\mathbf{0}, \boldsymbol{\Sigma})$, for $i = 2, 3$, can be found in [2]; in the presence of assignable causes responsible for a shift in the mean vector, we have to resort to Monte Carlo simulation to obtain estimates of the corresponding ARL, $ARL^{(i)}(\boldsymbol{\delta}, \boldsymbol{\Sigma})$, for $i = 2, 3$ and $\boldsymbol{\delta} \neq \mathbf{0}$.

To compare the performance of these three joint schemes, we take: the set of different out-of-control scenarios considered in [3]; $p = 2, 3, 4$; $n = p + 2$. Furthermore, the upper control limits of the individual charts of each joint scheme for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are chosen so that both individual charts have the same in-control ARL and the joint scheme yields a certain specified in-control ARL, say $ARL^{\star} = 500$ samples.

Mathematica programs were written to obtain $ARL^{(1)}(\boldsymbol{\delta}, \boldsymbol{\Sigma})$ and $ARL^{(i)}(\mathbf{0}, \boldsymbol{\Sigma})$, $i = 2, 3$. We had to resort to R to obtain the Monte Carlo estimates of $ARL^{(i)}(\boldsymbol{\delta}, \boldsymbol{\Sigma})$, for $i = 2, 3$ and $\boldsymbol{\delta} \neq \mathbf{0}$.

The preliminary results suggest, for instance, that when $\boldsymbol{\delta} = \mathbf{0}$: joint schemes 1 and 2 have a very similar ARL behavior; the use of $(T^{(3)}, U^{(3)})$ leads to an expected and systematic reduction of the ARL.

References

[1] Morais, M. C., Ramos, P. F. and Pacheco, A. (2015) Strategies to reduce the probability of a misleading signal. In *Frontiers in Statistical Quality Control (Vol. 11)*, 183–199. Knoth, S. and Schmid, W. (eds.), Springer International Publishing Switzerland.

[2] Morais, M. C., Pacheco, A., Ramos, P. F. and Schmid, W. (2016) On the influence of control statistics in the ARL of joint schemes for multivariate normal i.i.d. output. To be submitted.

[3] Ramos, P. F., Morais, M. C., Pacheco, A. and Schmid, W. (2016) On the misleading signals in simultaneous schemes for the mean vector and covariance matrix of multivariate i.i.d. output. *Statistical Papers* **57(2)**: 471–498.

Organized Session 6

Stochastic processes with applications in System Biology, Genetics and Epidemiology

Organizers: Miguel González Velasco (Spain), Manuel Mota Medina (Spain) and Inés M. del Puerto Garcia (Spain)

Stochastic epidemics with heterogeneous contacts in small networks

A. Gómez-Corral[1] and M. López-García[2]

[1]Instituto de Ciencias Matemáticas CSIC-UAM-UC3M-UCM, Spain
[2]University of Leeds, United Kingdom

E-mail addresses: *antonio.gomez@icmat.es; m.lopezgarcia@leeds.ac.uk*

Abstract

In this talk, we focus on stochastic models for the spread of SIS and SIR epidemics among a population consisting of N individuals, each having heterogeneous infectiousness and/or susceptibility. Under the practically relevant situation when N is small, the SIS and SIR models under consideration are respectively formulated in [1] and [2] as finite time-homogeneous continuous-time Markov chains $\mathcal{X}$, which amount to directed graphs $\mathcal{G}$ with members of the population represented by nodes, and with connections between individuals represented by edges. Based on an appropriate labeling of states, we first construct their infinitesimal rate matrices by using an iterative argument, and we then present algorithmic procedures for computing steady-state measures, such as the number of infected individuals, the length of an outbreak, the maximum number of infectives, and the number of infections suffered by a marked individual during an outbreak. The time till the epidemic extinction is characterized as a phase-type random variable when there is no external source of infection, and its Laplace-Stieljtes transform and moments are derived in terms of a forward elimination backward substitution solution. The inverse iteration method is applied to the quasi-stationary distribution of $\mathcal{X}$, which provides a good approximation of the process $\mathcal{X}$ at a certain time, conditional on non-extinction, after a suitable waiting time. The basic reproduction number $\mathcal{R}_0$ is defined here as a random variable, rather than an expected value. In our numerical work, the interest is in the spread of the syndrome *Acute coryza* within a family, and infections caused by nosocomial pathogens in an intensive care unit.

Keywords: basic reproduction number, Markov chain model, maximum number of infectives, quasi-stationary regime, stochastic SIS and SIR epidemics.

Acknowledgements

References

[1] Economou, A., Gómez-Corral, A. and López-García, M. (2015) A stochastic SIS epidemic model with heterogeneous contacts. *Physica A* **421**: 78–97.
[2] López-García, M. (2016) Stochastic descriptors in an SIR epidemic model for heterogeneous individuals in small networks. *Mathematical Biosciences* **271**: 42–61.

A two-sex branching process as model to explain the fate of a recessive allele linked to X chromosome

Miguel González[1], Cristina Gutiérrez[1], Rodrigo Martínez[1] and Manuel Mota[1]

[1]Departamento de Matemáticas and Instituto de Computación Científica Avanzada de la Universidad de Extremadura, Spain

E-mail addresses: *mvelasco@unex.es; cgutierrez@unex.es; rmartinez@unex.es; mota@unex.es*

Abstract

Some defective alleles of certain genes can cause severe diseases or serious disorders in the organisms that carry them. Some of these genes, as could be the responsible for hemophilia, are linked to X chromosome. If the alleles causing the disorder are dominant, all the carriers are affected and most of them do not reach breeding age so they are rarely detected in a population. However, recessive pernicious alleles can survive since they only affect to carrier males and homozygous carrier females (the last ones must be daughters of a carrier male, so they rarely exist). Hence, heterozygous carrier females are not affected but can pass the allele onto offspring. Recently, in [1], a multitype two-sex branching process has been introduced for describing the evolution of the number of individuals carrying the alleles, R and r, of a gene linked to X chromosome. The R allele is considered dominant and the r allele is supposed to be recessive and defective, responsible of a disorder. For this model we study under which conditions the recessive allele eventually disappears from the population. We also investigate conditions for the fixation of the dominant allele and for the extinction of the population. Finally, conditions for the coexistence of both alleles are conjectured, and shown through a simulation-based study.

Keywords: two-sex branching processes, extinction, unlimited growth, X-linked genes.

Acknowledgements

References

[1] González, M., Gutiérrez, C., Martínez, R. and Mota, M. (2016) Extinction probability of some recessive alleles of X-linked genes in the context of two-sex branching processes. In Branching Processes and their Applications. *Lecture Notes in Statistics–Proceedings* (del Puerto, I.M. et al., Eds). In press. DOI: 10.1007/978-3-319-31641-3

Mathematical modeling of the population dynamics of Black Vulture colonies through Branching Processes

Manuel Mota[1,2] Casimiro Corbacho[1,3] and Manuel Molina[1,2]

[1]University of Extremadura
[2]Department of Mathematics
[3]Department of Anatomy, Celular Biology and Zoology

E-mail addresses: *mota@unex.es; ccamado@unex.es; mmolina@unex.es*

Abstract

The Black vulture (Aegypius monachus) is a big Eurasian raptor species with a highly fragmented distribution; the western limits of its breeding range, but also some of the main populations, are located in the Iberian peninsula (Spain and Portugal). Worldwide, the species has declined over most of its range in the last 200 years, and it is currently listed as Near Threatened in the IUCN Red List. The decline was the greatest in the western half of the range, with extinction in many European and North African countries. Nowadays, the European population shows an increasing trend, but the much larger Asian population appears to be in decline (see [1]).

In this context, an appropriate mathematical model could help us to understand the population dynamics of this species and to reveal some parameters responsible for its decline. In this work we propose a discrete-time branching process that describes the evolution in the number of females in a black vulture colony, taking into account the special features distinctive of this species.

In particular, the model proposed presents the novelty of being indexed by the time instead of the generation, as usual in discrete time branching procesess. Such kind of processes have been investigated in several recent works (see for example [2–4]). On the basis of real data, we use this model to make inference on some population parameters that can be of interest to forecast future trends of the population of black vultures.

Keywords: branching processes, population dynamics.

Acknowledgements

References

[1]   BirdLife International (2013) *Aegypius monachus*, The IUCN Red List of Threatened Species 2013: e.T22695231A48120733. `http://dx.doi.org/10.2305/IUCN.UK.2013-2.RLTS.T22695231A48120733.en`.

[2]   Jagers, P. and Sagitov, S. (2008) General branching processes in discrete time as random trees. *Bernoulli* **14(4)**: 949–962.

[3]   Pénisson, S. and Jacob, C. (2012) Stochastic methodology for the study of an epidemic decay phase, based on a branching model. *Int. J. Stoch. Anal.* Art. ID 598701: 32 pp

[4]   Pénisson, S. (2014) Estimation of the infection parameter of an epidemic modeled by a branching process. *Electron. J. Stat.* **8(2)**: 2158–2187.

Modeling stochastic introgression with branching processes

Maria Conceição Serra[1] and Patsy Haccou[2]

[1]Centre of Mathematics, Minho University
[2]Leiden University College the Hague

E-mail address: *mcserra@math.uminho.pt*

Abstract

Introgression is the permanent incorporation of genes from one population into another through hybridization and backcrossing. It is currently of particular concern as a possible mechanism for the spread of modified crop genes to wild populations. In this work, multitype branching processes are used to model the evolution of populations that are exposed to introgression events.

It is important to be able to quantify the risk of an introgression event. We use the hazard rate, that is the probability, per time unit, that such an escape event takes place, given that it has not happened before. It is a quantitative measure of the introgression risk that takes the stochastic elements inherent in introgression processes into account. We present a methodology to calculate the hazard rate for situations with: i) constant; ii) time-varying; iii) random; gene flow from a crop to a large recipient wild population. We analyze the effect that conditions i) - iii) have on the behaviour of the hazard rate, namely on the long term behaviour. Finally, we discuss the risk of introgression in meta-populations.

Keywords: branching processes, multitype, introgression, hazard rates.

References

[1]   Serra, M. C., Haccou, P. (2007) Dynamics of escape mutants. *Theoretical Population Biology* **72(1)**: 167–178.
[2]   Ghosh, A., Serra, M. C. and Haccou, P. (2012) Quantifying time-inhomogeneous stochastic introgression processes with hazard rates. *Theoretical Population Biology* **81(4)**: 253–263.
[3]   Ghosh, A., Serra, M. C. and Haccou, P. (2015) Quantifying stochastic introgression processes in random environments with hazard rates. *Theoretical Population Biology* **100**: 1–5.

Organized Session 7

Statistical Distributions and Applications

Organizer: Filipe Marques (Portugal)

Improving the asymptotically unbiased extreme value index estimation

<u>Frederico Caeiro</u>[1,2], Ivanilda Cabral[3] and M. Ivette Gomes[4,5]

[1]Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal
[2]Centro de Matemática e Aplicações (CMA), Universidade Nova de Lisboa
[3]Departamento de Ciência e Tecnologia, Universidade de Cabo Verde, Cabo Verde
[4]DEIO, Faculdade de Ciências, Universidade de Lisboa, Portugal
[5]Centro de Estatística e Aplicações da Universidade de Lisboa (CEAUL)

E-mail addresses: *fac@fct.unl.pt; ivanilda.cabral@docente.unicv.edu.cv; ivette.gomes@fc.ul.pt*

Abstract

The extreme value index characterizes the tail behavior of a distribution, and indicates the size and frequency of certain extreme events under a given probability model.

In this work, we are interested in improvements attained through the reduction of bias of the extreme value index estimators introduced in [1,2] and further studied in [3,5,6]. A comparison with other reduced bias estimators, such as the corrected-Hill estimator [4] is also performed.

Keywords: Bias reduction, extreme value index, heavy tails, semi-parametric estimation.

Acknowledgements

References

[1] Caeiro, F. and Gomes, M. I. (2002a) A class of asymptotically unbiased semi-parametric estimators of the tail index. *Test* **11(2)**: 345–364.

[2] Caeiro, F. and Gomes, M. I. (2002b) Bias reduction in the estimation of parameters of rare events. *Theory of Stochastic Processes* **8(24)**: 67–76.

[3] Caeiro, F. and Gomes, M. I. (2014) Comparison of asymptotically unbiased extreme value index estimators: a Monte Carlo simulation study. *AIP Conference Proceedings* **1618**: 551–554.

[4] Caeiro, F., Gomes, M. I. and Pestana, D. D. (2005) Direct reduction of bias of the classical Hill estimator. *Revstat* **3(2)**: 111–136.

[5] Gomes, M. I., Caeiro, F. and Figueiredo, F. (2004) Bias Reduction of a Tail Index Estimator Through an External Estimation of the second-order parameter. *Statistics* **38(6)**: 497–510.

[6] Penalva, H., Caeiro, F., Gomes, M. I. and Neves, M. M. (2016) *An efficient naive generalisation of the Hill estimator—discrepancy between asymptotic and finite sample behaviour.* Notas e Comunicações CEAUL 02/2016.

Revisiting the choice of block size and the threshold in the extremal index estimation

<u>Dora Prata Gomes</u>[1,2] and M. Manuela Neves[3,4]

[1]Universidade Nova de Lisboa, FCT
[2]CMA/FCT, Portugal,
[3]Universidade de Lisboa, ISA
[4]CEAUL/ULisboa

E-mail addresses: *dsrp@fct.unl.pt; manela@isa.ulisboa.pt*

Abstract

The extremal index, θ, a key parameter in extreme value theory in a dependent set-up, is a measure of clustering of extreme events characterizing the degree of local dependence in the extremes of a stationary time series. Its estimation has been addressed by several authors but some difficulties still remain. Resampling computer intensive methodologies, like the bootstrap and the jackknife have been recently considered in a reliable estimation of parameters of rare events, among which the extremal index. Because of the serial dependence that arises in many applications, the classical bootstrap cannot be applied and the block bootstrap procedure needs to be considered. However, the block size for resampling strongly affects the estimates and needs to be properly estimated. In this work procedures for the choice of the block size for resampling as well as an heuristic procedure for the adaptive choice of the threshold will be considered. Results from a simulation study will illustrate the performance of the aforementioned procedures.

Keywords: block bootstrap, block size, clusters of extremes, extremal index.

——⊙——
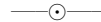
In Extreme Value Theory (EVT) the extremal index (EI), θ, is a key parameter that measures the propensity of the large observations in a dataset to cluster, Leadbetter *et al.* [5]. Like other semi-parametric estimators, EI estimators show nice asymptotic properties, but a high variance for small values of k, the number of upper order statistics used in the estimation, and a high bias for large values of k. This brings a real need for the choice of k, one of the problems here addressed.

After a brief reference to some estimators of the EI and their asymptotic properties, the goal of this work is to improve the performance of those estimators through computational procedures based on resampling blocks of observations, that need to be adequately chosen.

Resampling methodologies have provided very fruitful results in the field of statistics of extremes. Classical bootstrap procedures, derived for independence, have to be adapted to the dependent context. One of the procedures proposed for the bootstrap resampling in this situation, see for example Lahiri [3], consists of defining blocks for resampling, instead of resampling the individual observations. But the performance of the bootstrap estimator crucially depends on the block size that must be supplied by the user.

Several authors such as Hall *et al.*[2], Bühlman and Künsch [1], Politis and White [6] and Lahiri *et al.* [4] proposed ways of estimating the optimal block size.

Here we follow Lahiri *et al.* [4], who proposed a nonparametric plug-in (NPPI) method for the empirical choice of the optimal block size for the block bootstrap estimation of

characteristics of an estimator. Two of these characteristics are the bias and the variance of the estimator.

Some simulation results and applications have been presented in Prata Gomes and Neves [7,8], dealing with those characteristics.

Here some improvements on the previous works will be considered and more results from a more complete simulation study that has been performed will be discussed.

Acknowledgements

References

[1] Bühlmann, P. and Künsch, H. (1999). Block length selection in the bootstrap for time series. *J. Comput. Statist. and Data Analysis* **31**, 295–310.

[2] Hall, P., Horowitz, J. L. and Jing, B.-Y. (1995) On blocking rules for the bootstrap with dependent data. *Biometrika* **50**: 561–574.

[3] Lahiri, S. (2002) On the jackknife after bootstrap method for dependent data and its consistency properties. *Econometric Theory* **18**: 79–98.

[4] Lahiri, S., Furukawa, K. and Lee, Y-D. (2007). Nonparametric plug-in method for selecting the optimal block lengths. *Statistical Methodology* **4**: 292–321.

[5] Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983) *Extremes and related properties of random sequences and series.* Springer-Verlag, New York.

[6] Politis, D.R. and White, H. (2004). Automatic Block-Length Selection for the Dependent Bootstrap. *Econometric Reviews* **23(1)**: 53–70.

[7] Prata Gomes, D. and Neves, M. M. (2015a) Bootstrap and other resampling methodologies in statistics of extremes. *Communications in Statistics–Simulation and Computation* **44(10)**: 2592–2607.

[8] Prata Gomes, D. and Neves, M. Manuela (2015b) Adaptive choice and resampling techniques in extremal index estimation. In Kitsos, C., Oliveira, T., Rigas, A. and Gulati, S. (eds.), *Theory and Practice of Risk Assessment*, Springer Proceedings in Mathematics and Statistics, 321–332.

Series representations for the sum of independent Gamma random variables and for the product of independent Beta random variables

Filipe J. Marques[1,2]

[1]Universidade NOVA de Lisboa, Almada, Portugal
[2] Centro de Matemática e Aplicações (CMA), FCT, UNL, Portugal

E-mail address: *fjm@fct.unl.pt*

Abstract

The authors show that using well known series expansions it is possible to represent a single Gamma distribution, and also the logarithm of a single Beta distribution, as an infinite mixture of Gamma distributions. Then, using these representations, it is possible to derive simple gamma-series representations for the sum of independent Gamma random variables and for the product of independent Beta random variables. It is shown that the representation obtained for the sum of independent Gamma random variables is somehow related to the one obtained in [2] which was already used in [3] to derive very accurate near-exact distributions, introduced in [1], for the linear combination of independent Gamma random variables. The numerical studies developed support the applicability of the representations obtained.

Keywords: binomial expansion, exponential series expansion, characteristic functions, mixtures, near-exact approximations.

Acknowledgements

References

[1]　Coelho, C. A. (2004) The generalized near-integer Gamma distribution: a basis for "near-exact" approximations to the distribution of statistics which are the product of an odd number of independent Beta random variables. *Journal of Multivariate Analysis* **89**: 191–218.
[2]　Moschopoulos, P. G. (1985) The distribution of the sum of independent Gamma random variables. *Ann. Inst. Statist. Math.* **37**: 541-Ű544.
[3]　Marques, F. J. and Coelho, C. A. (2014) A note on the distribution of the linear combination of independent Gamma random variables. *American Institute of Physics Conference Proceedings* **1618**: 527–530.

Pseudo Maximum Likelihood and Moments Estimators for some Ergodic Diffusions

Pedro Mota[1,2] and Manuel L. Esquível[1,2]

[1]Faculdade de Ciências e tecnologia, Universidade NOVA de Lisboa
[2]Centro de Matemática e Aplicações (CMA)

E-mail addresses: *pjpm@fct.unl.pt; mle@fct.unl.pt*

Abstract

When $(X_t)_{t \geq 0}$ is an ergodic process, the density function of X_t converges to some invariant density, when $t \to \infty$. We will compute and study some asymptotic properties of pseudo maximum likelihood and moments estimators obtained from this invariant density, for a specific class of ergodic processes. In this class of processes we can find the Cox-Ingersoll & Ross ([2]) or Dixit & Pindyck processes ([3]), among others. A comparative study of the proposed estimators with the usual estimators obtained from discrete approximations of the likelihood function will be carried out.

Keywords: Consistency; Ergodic; Maximum Likelihood and Moments Estimators.

———⊙———

Ergodic diffusion processes are frequently used in many practical applications, but most of the times no explicit likelihood function is known and because of that, the usual estimation procedure uses approximations of the likelihood function based in some kind of discretization or uses martingale estimating functions, see [1] or [4] and the references therein.

We will study the processes satisfying the stochastic differential equation (SDE),

$$dX_t = b(a - X_t)X_t^\gamma dt + \sigma\sqrt{X_t^{\gamma+1}}dB_t, a, b, \sigma > 0, \gamma \geq 0, \tag{1}$$

and for the combination of parameters that makes this processes ergodic, we will present pseudo maximum likelihood (see [5]) and moments estimators.

The processes satisfying the SDE (1) are ergodic when $2ab > \sigma^2(\gamma + 1)$ and their invariant density is then given by,

$$f_\theta(x) = \frac{x^{\alpha-1}e^{-\beta x}\beta^\alpha}{\Gamma(\alpha)} \sim Gamma(\alpha, \beta), \text{ with, } \alpha = \frac{2ab}{\sigma^2} - \gamma, \beta = \frac{2b}{\sigma^2} \tag{2}$$

Suppose that we have the observations $X_1, ..., X_n$ of the process and that γ and σ are known parameters. If we deal with the observations like if they were independent and identically distributed, we are able to compute pseudo maximum likelihood and pseudo moments estimators.

The pseudo maximum likelihood estimators are $(\hat{a}_n, \hat{b}_n)$, with $\hat{b}_n$ the solution of

$$\frac{1}{n}\sum_{i=1}^n \log(X_i) + \log\left(\frac{2b}{\sigma^2}\right) - \psi\left(\frac{2\bar{X}_n b}{\sigma^2}\right) = 0 \tag{3}$$

where $\psi(.)$ is the digamma function, $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ and with

$$\hat{a}_n = \bar{X}_n + \frac{\sigma^2\gamma}{2\hat{b}_n}. \tag{4}$$

The pseudo moments estimators are $(\tilde{a}_n, \tilde{b}_n)$, using the (non-central) sample variance $S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$,

$$\tilde{a}_n = \bar{X}_n + \frac{S_n^2}{\bar{X}_n}\gamma \quad \wedge \quad \tilde{b}_n = \frac{\sigma^2 \bar{X}_n}{2S_n^2}. \tag{5}$$

Our main result is that, both pseudo maximum likelihood and moments estimators are consistent estimators.

Theorem 1. *If $2ab > \sigma(\gamma + 1)$, the pseudo maximum likelihood estimators $\hat{a}_n$ and $\hat{b}_n$ are a.s. consistent estimators for a and b.*

Theorem 2. *If $2ab > \sigma(\gamma + 1)$, the pseudo moments estimators $\tilde{a}_n$ and $\tilde{b}_n$ are a.s. consistent estimators for a and b.*

Proofs of both theorems are provided and a comparative study, through si-mulation, will be carried out for different values of γ.

Acknowledgements

References

[1] Bibby, B. M. and Sørensen, M. (1995) Martingale estimation functions for discretely observed diffusion processes, *Bernoulli* **1(1/2)**: 17–39.

[2] Cox, J., Ingersoll, J. and Ross, S. (1985) A theory of the term structure of interest rates, *Econometrica* **53(2)**: 385–408.

[3] Dixit, A. K. and Pindyck, R. S. (1994) *Investment Under Uncertainty*, Princeton University Press, New Jersey.

[4] Kessler, M. (1997) Estimation of an ergodic diffusion from discrete observations, *Scandinavian Journal of Statistics* **24(2)**: 211–229.

[5] Mota, P. and Esquível, M. L. (2016) Estimation using the Invariant Density for some Ergodic Processes. *Submitted*

Random polygonal lines iteratively generate on the plane

[Rui Santos]<u></u>[1], Miguel Felgueiras[1,2] and João Paulo Martins[1]

[1]School of Technology and Management, Polytechnic Institute of Leiria, CEAUL–Center of Statistics and Applications of University of Lisbon
[2]CIGS–Centre for Research in Management for Sustainability of Polytechnic Institute of Leiria

E-mail addresses: *rui.santos@ipleiria.pt; mfelg@ipleiria.pt; jpmartins@ipleiria.pt*

Abstract

Starting from a straight line of limited length, Pacheco d'Amorim defines an iterative procedure of breaking each side at the middle point and randomly folding the two resulting segments in the plane, considering that at the limit it is obtained a random curve. With this procedure he concludes that the distance between the two endpoints of a random curve generated in the plane is almost surely zero.

In this work, we define a random polygonal line as the result of an infinite iterative procedure where in each iteration each polygonal segment is randomly (in what concerns the angle they form) folded up at its middle point, and re-scaled by a factor 2^H, where $0 \leq H \leq 1$. We aim to find a scale factor such that the square of the distance between the two endpoints of a random polygonal line generated in the plane is a non-degenerate random variable with finite mean, and then to investigate its distribution. We also analyse more general random polygonal lines and the corresponding limit curves, folded at random points of the segments at each iterative step of its construction, instead of the middle point. Thus, our main goal is to investigate the limiting behaviour of these random polygonal lines, mainly the distribution of the square of the distance between its two endpoints.

Keywords: non-degenerate asymptotic laws, renormalization, Hausdorff dimension, simulation, plane random curves.

References

[1]   Pacheco d'Amorim, D. (1914) *Elementos de Cálculo das Probabilidades*, Tese de doutoramento, Imprensa da Universidade de Coimbra, Coimbra.

[2]   Santos, R. (2011) Simulation Analysis of Statistical Properties of Random Polygonal Lines Iteratively Generate on the Plane, *Proceedings of the ITI 2011, 33rd International Conference on Information Technology Interfaces*: 557–562.

Organized Session 8

Algebra, Mathematics and Computation

Organizer: João Araújo (Portugal)

A complete rewriting system for the Plactic monoid using crystals

<u>António Malheiro</u>[1,2], A. Cain[3] and R. Gray[4]

[1]Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa
[2]Centro de Matemática e Aplicações, UNL
[3]CMA, UNL, Portugal
[4]East Anglia University, UK

E-mail address: *ajm@fct.unl.pt*

Abstract

In this talk I will present the main definitions and concepts that relate crystal graphs, Young tableaux, and the Plactic monoid. It will then be possible to show that Plactic monoids admit a finite complete rewriting system.

Keywords: crystal graphs, monoid presentation, Plactic monoid, rewriting systems, Young tableaux.

Acknowledgements

References

[1]   Cain, A.J., Gray, R. and Malheiro, A. (2015). Finite Gröbner–Shirshov bases for Plactic algebras and biautomatic structures for Plactic monoids. *Journal of Algebra* **423**: 37–53.

Comparing Several Tests of Randomness Based on the Difference of Observations

Ayana Mateus[1,2] and Frederico Caeiro[1,2]

[1]Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal
[2]Centro de Matemática e Aplicações (CMA)

E-mail addresses: *amf@fct.unl.pt; fac@fct.unl.pt*

Abstract

Randomness is a common assumption in many statistical methods. When such assumption is not fulfilled, researchers may draw wrong conclusions. In many applications a simple graphical analysis is enough to check such assumption. But there are cases where a test of hypothesis is required. There is a wide range of nonparametric statistical tests that can be used to test the null hypothesis that the sequence is random, i.e. the variables are independent and occur in a random order and therefore contains no recognizable patterns or regularities. In this research we consider three nonparametric randomness tests based on the sign of the difference of different observations from the sample, namely the turning point test, the difference-sign test and the rank test and compare their performance through a Monte Carlo simulation study. Our choice on these tests has been based on the popularity that they have in the context of nonparametric tests for serial independence. They are often used in time series analysis for testing the independence of the residuals in fitted deterministic trend or cyclic models [1,2]. Although aware that parametric tests are usually more powerful, nonparametric tests have the advantage of being more robust since fewer assumptions have to be made about the data.

Keywords: Monte Carlo comparison, nonparametric test, randomness.

Acknowledgements

References

[1] Cammarota, C. (2011) The difference-sign runs lenght distribution in testing for serial independence. *Journal of Applied Statistics* **38**: 1033–1043.
[2] Brockwell, P. J. and Davis, R. A. (2002) *Introduction to Time Series and Forecasting*, Springer.
[3] Hofmann, G. and Balakrishnan, N. (2006) A nonparametric test for trend based on initial ranks. *Journal of Statistical Computation and Simulation* **76(9)**: 829–837.
[4] Mann, H. B. (1945) Nonparametric test against trend. *Econometrica* **13**: 245–259.
[5] Moore, G. H. and Wallis, W. A. (1943) Time Series Significance Tests Based on Signs of Differences. *Journal of the American Statistical Association* **38**: 153–154.

The polynomial hierarchy of time

[Isabel Oitavem][1,2]

[1]DM, FCT, Universidade Nova de Lisboa
[2]CMA-Centro de Matemática e Aplicações

E-mail address: *oitavem@fct.unl.pt*

Abstract

We characterize the polynomial hierarchy of time (FPH) and its levels in a recursion-theoretic manner, in the vein of Cobham's characterization of FPtime. This is based on [1] and it combines monotonicity constraints with a known characterization of FPspace [2].

Ptime (or just P) is the class of polynomial time problems and Pspace is the class of polynomial space problems. However, if one allows write-once memory, instead of read/write memory, we impose a monotone change of the memory content. Intuitively, to restrict the time corresponds to restrict the number of re-writes. So, it does not surprise that P corresponds also to the class of polynomial write-once space problems. The challenge is then to consider other descriptions of FPspace, not machine based, and mimic in those contexts the monotonicity corresponding to write-once memory.

In this talk we consider two recursion-theoretic approaches to FPspace, and we impose to them the same monotonicity constraint. In one case, we get FPtime (the first level of the polynomial hierachy of time), and in the other case we characterize the full hierarchy FPH.

Keywords: computational complexity, polynomial hierarchy, recursion, write-once memory.

Acknowledgements

References

[1]   Ben-Amram, A., Loff, B. and Oitavem, I. (2012) Monotonicity constraints in characterisations of PSPACE. *Journal of Logic and Computation* **22**: 179–195.
[2]   Oitavem, I. (1997) New recursive characterizations of the elementary functions and the functions computable in polynomial space. *Revista Matemática de la Universidad Complutense de Madrid* **10(1)**: 109–125.

Nontrivial Ground States for Cooperative Cubic Schrödinger Systems

F. Oliveira[1,2], S. Correia[3] and H. Tavares[4]

[1]Departamento de Matemática, FCT, Universidade Nova de Lisboa
[2]Centro de Matemática e Aplicações CMA-FCT-UNL
[3]CMAF-CIO and FCUL
[4]CAMGSD, Mathematics Department Instituto Superior Técnico, Universidade de Lisboa

E-mail address: *fso@fct.unl.pt; sfcorreia@fc.ul.pt; htavares@math.ist.utl.pt*

Abstract

In this talk we will present some new results concerning the existence of semitrivial and fully nontrivial ground states for cubic Schrödinger Systems of cooperative type. In particular, we will focus on a recently found structure arizing in the case of systems with at least 3 equations.

Keywords: cooperative systems, ground states, Schrödinger systems, semitrivial, fully nontrivial solutions.

——⊙——
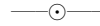
In this talk we will consider the weakly coupled Schrödinger cubic system

$$-\Delta u_i + \lambda_i u_i = \mu_i u_i^3 + u_i \sum_{j \neq i} b_{ij} u_j^2$$

with $u_i \in H^1(R^N; R)$, $i = 1, \ldots, d$, and where $1 \le N \le 3$, $\lambda_i, \mu_i > 0$ and $b_{ij} = b_{ji} > 0$ for $i \neq j$.

This system admits semitrivial solutions, that is, solutions $u = (u_1, \ldots, u_d)$ with null components. We provide optimal qualitative conditions on the parameters λ_i, μ_i and b_{ij} under which the *ground state solutions* have all components nontrivial, or, conversely, are semitrivial.

This question had been clarified only in the $d = 2$ equations case. For $d \ge 3$ equations, prior to the present paper, only very restrictive results were known, namely when the above system was a small perturbation of the super-symmetrical case $\lambda_i \equiv \lambda$ and $b_{ij} \equiv b$. We treat the general case, uncovering in particular a much more complex and richer structure with respect to the $d = 2$ case.

Acknowledgements

References

[1]   S. Correia, F. Oliveira and H. Tavares, Semitrivial vs. fully nontrivial ground states in cooperative cubic Schrödinger systems with $d \ge 3$ equations, arXiv:1508.01783, 2016.

Specht modules and symmetry classes of tensors

Fátima Rodrigues[1,2], J. Dias da Silva[3], Pedro C. Silva[5,6] and
Maria M. Torres[3,4]

[1]Departamento de Matemática, FCT, UNL
[2]Centro de Matemática e Aplicações (CMA), FCT, UNL
[3]Departamento de Matemática, FC, UL
[4]Centro de Análise Funcional Estruturas Lineares e Aplicações, UL
[5]Departamento de Ciências e Engenharia de Biossistemas, ISA, UL
[6]Centro de Estudos Florestais, ISA, UL

E-mail addresses: *mfsr@fct.unl.pt; japsilva@fc.ul.pt; pcsilva@isa.ulisboa.pt;
mmtorres@fc.ul.pt*

Abstract

Specht modules are irreducible representations of the symmetric group and they have
a basis of polytabloids indexed by Young tableaux. We present bases to the critical orbital
subspaces of the symmetry classes of tensors and a canonical isomorphism between Specht
modules and this subspaces.

Keywords: decomposable tensor, orbital spaces, Specht modules, symmetry class of tensors.

Acknowledgements

References

[1] Dias da Silva, J. A. (1981) Irreducible Representations of the Full Linear Group.
Linear Algebra and its Applications **40**: 61–82.
[2] Fulton, W. (1997) *Young tableaux, wiyh applications to representation theory and
geometry*, London Math. Soc. Student Texts, vol. 35, Cambridge University Press,
Cambridge.
[3] Harmon, H. G. and Holmes, R. R. (2014) Geometry of standard symmetrized tensors.
Linear Algebra and its Applications **460**: 136–150.
[4] Torres, M. and Silva, P. (2012) Tensors, Matchings and Codes. *Linear Algebra and
its Applications* **436**: 1606–1622.

Organized Session 9

Matrices in Science and Engineering

Organizer: João Cardoso (Portugal)

Finding the Closest Generalized Essential Matrix

João R. Cardoso[1,2] and Pedro Miraldo[3,4]

[1]Polytechnic Institute of Coimbra/ISEC, Portugal
[2]Institute of Systems and Robotics, University of Coimbra, Portugal
[3]University of Lisbon, Instituto Superior Técnico
[4]Institute for Systems and Robotics (LARSyS), University of Lisbon, Portugal

E-mail addresses: *jocar@isec.pt; pmiraldo@isr.tecnico.ulisboa.pt*

Abstract

One of the most common strategies to do the scene reconstruction using computer vision techniques is to use a pair of stereo cameras (two cameras, in different positions, looking at the same scene). Usually, due to their simplicity, authors use perspective cameras to build up this stereo pair [3]. However, these cameras have several disadvantages. For example, they have a limited field of view, which results in a small 3D reconstruction area. To overcome this drawback, authors created a new camera system, with a special emphasis on omnidirectional cameras, which have a wide field of view.

Here we consider a stereo camera system, formed by two general camera models [6]. In the perspective stereo problem, authors defined a 3×3 matrix that encodes the incident relation between pixels in both cameras, which can be used later to compute the triangulation, giving the 3D scene reconstruction [3]. However, in the case of these new camera systems, this cannot be used, because the required constraints are not verified. Instead, it can be proved that there is a 6×6 matrix that expresses this relationship, for any type of camera system (non-constrained camera systems–pixels are mapped onto generic 3D straight lines in the world) [2,7]. Such a matrix is called *generalized essential* and has a special block structure involving rotation and skew–symmetric matrices of order 3. In general, computer vision problems where this matrix has to be determined are affected with noise and the result is a matrix of order 6 that fails to fit the structure of an generalized essential matrix. In these cases, one needs to find the closest generalized essential matrix from a generic 6×6 matrix. From a mathematical point of view, this is a nonlinear constrained matrix optimization problem. It is nonconvex which raises many difficulties to finding a global minimum.

In this talk we describe our current investigations. First, we show that the original problem can be converted into an optimization problem with orthogonal constraints and then present a preliminar algorithm for finding a global minima. Some open questions will be pointed out.

We recall that methods for problems with orthogonal constraints are available in the literature [1,5,8,4], but its difficulty depends significantly on the objective function. In our problem, the objective function is not easily to handle which raises many challenging issues. Techniques from matrix theory and optimization on manifolds are used throughout.

Keywords: computer vision, generalized essential matrix, nonlinear optimization, orthogonality.

References

[1] Edelman, A., Arias, T. and Smith, S. (1998) The geometry of algorithms with orthogonal constraints. *SIAM Journal on Matrix Analysis and Applications* **20**: 303–353.

[2]   Grossberg, M. and Nayar, S. (2001) A General Imaging Model and a Method for Finding its Parameters. *Proceedings of 8th IEEE International Conference on Computer Vision* **2**: 108–115.

[3]   Hartley, R. and Zisserman, A. (2004) *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press.

[4]   Jiang B. and Dai, Y.-H. (2015) A framework of constraint preserving update schemes for optimization on Stiefel manifold. *Mathematical Programming, Series A* **153**: 535–575.

[5]   Manton, J. (2002) Optimization Algorithms Exploiting Unitary Constraints. *IEEE Transactions on Signal Processing* **50**: 635–650.

[6]   Nayar, S. (1997) Catadioptric Omnidirectional Camera. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*: 482–488.

[7]   Sturm, P. (2005) Multi-View Geometry for General Camera Models. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **1**: 206–212.

[8]   Wen, Z. and Yin, W. (2013) A feasible method for optimization with orthogonality constraints. *Mathematical Programming, Series A* **142**: 397–434.

Linear Algebra and Image Processing: a new teaching approach

C. M. R. Caridade[1,2]

[1]Polytechnic Institute of Coimbra/ISEC, Portugal
[2]Geo-Space Sciences Research Centre, University of Porto

E-mail address: *caridade@isec.pt*

Abstract

Linear Algebra is one of the main mathematical fields, which is used in many technical subjects. Inability to realize the importance of Linear Algebra and its applications in technical subjects is one of the reasons of poor academic performance of students. The typical teaching approach separates the Linear Algebra from technical subject. This does not necessarily promote students awareness in both subjects. This paper presents a new teaching approach which combines Linear Algebra and Image Processing contents. The methodology presented is based in teaching Linear Algebra contents such as matrices and matrix transformation using Image Processing applications. This could lead to a whole new interesting learning environment, new teaching approaches that are more stimulating for students and more productive and dynamic for teachers.

Keywords: mathematics, matrices, image processing, teaching.

———⊙———

The Digital Image Processing develops computer algorithms to perform image processing on digital images. It is a highly relevant area, extremely rich in mathematical ideas, that allows an interesting pedagogical tool to explore Linear Algebra concepts. The combination of Digital Image Processing with Linear Algebra is a natural link that provides a visual component for Linear Algebra concepts [1].

Some experiences have been made, for example the work presented by Silverman [2] that uses Digital Image Processing and Linear Algebra to teach concepts and terminologies of Linear Algebra while developing self confidence students. The Digital Image Processing helps to develop visual and intuitive understanding of the concepts, which are usually abstract and entirely new for students [3,4]. So, the Digital Image Processing can improve the teaching of Linear Algebra in a completely different way by increasing the motivation and participation.

The purpose of this paper is to present some experiences using Image Processing to teach Linear Algebra to students. Linear Algebra concepts like linear equations, matrix operations, matrix algebra, matrix transformations, and inverse transformation can be explored using applications of digital image processing (image representation, image transformation, image enhancement, image restoration, image segmentation or image compression) [5]. Image filtering techniques are usually applied to increase the quality of magnetic resonance images, segmentation of skin cancer in dermoscopy images, edge detection techniques for fingerprint images, etc. All this Image Processing techniques involve Linear Algebra subjects.

The concepts to be learned by students must be meaningful for them, so that they can assimilate it. Using matrices operations to eliminate the image background or identify parts of the image can enrich teaching practices and improve student learning. The use of the

2D-geometric transformation can bring one image into alignment with another increasing the motivation and improvement of students skills in this concepts. Linear Algebra and Image Processing can be taught simultaneously. With this combined lessons, it is hoped that students can overcome the difficulties in learning Linear Algebra through the real world applications. Furthermore, teachers are also able to teach the contents of Linear Algebra to students in a more stimulating and motivating environment.

References

[1] Gonzalez, R. C. and Woods, M. P. (2002) *Digital image processing*, Prentice Hall, 2nd edition.

[2] Silverman, J. and Rosen, G. (2010) Supporting students interest in mathematics throught applications from digital image processing. *Journal of the Research Center for Education Technology* **6(2)**: 63–77.

[3] Berriochoa, E. et al. (2009) An approach for teaching the linear algebra for students of engineering. *Proc. of International Conference of Education, Research and Innovation* 3756–3763.

[4] Caridade, C. M. R. (2011) Applying image processing techniques to motivate students in linear algebra classes. *Proc. of the* 1st *World Engineering Flash Week, Lisbon* 114–121.

[5] Zin, Z. M., Sallehb T. S. and Bakrib, N. (2015) Transforming Teaching and Learning Approach of Mathematics and Image Processing. *Journal of Science and Engineering Technology JSET* **2(2)**.

Spatio-temporal modelling of environmental monitoring data from different surveys

L. Margalho[1,3], R. Menezes[2,3] and I. Sousa[2,3]

[1]Polytechnic Institute of Coimbra/ISEC, Portugal
[2]University of Minho, Portugal
[3]CMAT, University of Minho

E-mail addresses: *lmelo@isec.pt; rmenezes@math.uminho.pt; isousa@math.uminho.pt*

Abstract

Environmental monitoring networks are providing large amounts of spatio-temporal data. Air pollution data, as other environmental data, exhibit a spatial and a temporal correlated nature. To improve the accuracy of predictions at unmonitored locations, there is a growing need for models capturing those spatio-temporal correlations.

With this work, we propose a spatio-temporal model for gaussian data collected in a few number of surveys. We assume the spatial correlation structure to be the same in all surveys. In an application of this model to real data, concerning heavy metal concentrations in mosses collected from three surveys occurring between 1992 and 2002 in mainland Portugal, the data set is dense in the spatial dimension but sparse in the temporal one, thus our model-based approach corresponds to a saturated correlation model in the time dimension. A novel interpretation for the space-time covariance function is introduced. A simulation study, aiming to validate the model, provided better results in terms of accuracy with the novel covariance function.

Prediction maps of the observed variable for the most recent survey, and of the interpolation error as a measure of accuracy, are presented.

Keywords: environmental pollution monitoring, space-time modelling, sparse time dimension, separable covariance matrix structure.

—⊙—
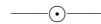
Nowadays, due to technology developments and worldwide policies, environmental monitoring networks are providing large amounts of data exhibiting a spatial and a temporal correlated nature, and as a consequence a large number of models and techniques to analyze this sort of data has emerged.

Although in environmental sciences data are typically collected through monitoring stations, it may also be collected through biomonitoring surveys covering extensive areas. Some examples of studies involving moss samples as biomonitors of atmospheric heavy metal deposition are included in [1,3] with data from Galicia, northern Spain, in [6,7] concerning Norway data, and in [8] with data from Austria.

It is common to have studies, such as the ones mentioned before, involving environmental spatio-temporal data containing a dense time dimension but only a sparse spatial one, as a result of the easiness of gathering data enabled by modern technologies. That is not the case of the biomonitoring data being used in this work, which are related to measurements of heavy metal concentrations made at 146 spatial locations in only 3 surveys.

Our aim is to propose a naive spatio-temporal framework which incorporates into the model both time and space correlations, capable to fit spatio-temporal data containing a reduced number of time observations. Due to this particular characteristic of having few

temporal records, and under the hypothesis of separability of the correlation structure, it may be the case that the number of parameters to estimate in the temporal correlation function equals the number of temporal observations, which corresponds to have a saturated correlation model in the time dimension, *i.e.*, a model perfectly reproducing the data.

References

[1] Aboal, J., Real, C., Fernández, J. and Carballeira, A. (2006) Mapping the results of extensive surveys: The case of atmospheric biomonitoring and terrestrial mosses. *Science of The Total Environment* **356**: 256–274.

[2] Cressie, N. and Wikle, C. (2011) *Statistics for spatio-temporal data*. John Wiley & Sons Inc, New Jersey.

[3] Diggle, P., Menezes, R. and Su, T. (2010) Geostatistical inference under preferential sampling. *Applied Statistics* **59(2)**: 191–232.

[4] Margalho, L., Menezes, R. and Sousa, I. (2014) Assessing interpolation error for space-time monitoring data. *Journal of Stochastic Environmental Research and Risk Assessment* **28(5)**: 1307–1321.

[5] Rodriguez-Iturbe, I. and Mejia, J. (1974) The design of networks in time and space. *Water Resources Research* **10(4)**: 713–728.

[6] Steinnes, E., Berg, T. and Sjøbakk, T. (2003) Temporal and spatial trends in Hg deposition monitored by moss analysis. *The Science of the Total Environment* **304**: 215–219.

[7] Steinnes, E., Berg, T. and Uggerud, H. (2011) Three decades of atmospheric metal deposition in Norway as evident from analysis of moss samples. *The Science of the Total Environment* **412–413**: 351–358.

[8] Zechmeister, H., Hohenwallner, D., Hanus–Illnar, A., Hagendorfer, H., Roder, I. and Riss, A. (2008) Temporal patterns of metal deposition at various scales during the last two decades. *Atmospheric Environment* **42**: 1301–1309.

On Some Problems of Analytic Geometry over the Octonions

P. Saraiva[1,2,3]

[1]University of Coimbra, Faculty of Economics, Portugal
[2]CMUC–Centre for Mathematics, University of Coimbra
[3]CeBER–Center for Business and Economics Research, University of Coimbra

E-mail address: *psaraiva@fe.uc.pt*

Abstract

The existence of a vector cross product in the real linear space $\mathbb{R}^7$ having similar properties to those of the usual 3-dimensional version is closely related with the notion of a real composition algebra. This is an algebra A over $\mathbb{R}$ with multiplication $*$ where we can define a nondegenerate quadratic form $N : A \to \mathbb{R}$ which is multiplicative, *i.e.*, for any $x, y \in A$, $N(x * y) = N(x)N(y)$. The so-called *Hurwitz problem*, which consists on finding multiplicative quadratic forms on algebras over a field $\mathbb{K}$ ($char(\mathbb{K}) \neq 2$) lead to the *Hurwitz theorem*, which asserts that (assuming anisotropy of N) the only finite dimensional (unital) real composition algebras are, up to isomorphism, the real division algebras: real numbers, $\mathbb{R}$; complex numbers, $\mathbb{C}$; quaternions (or Hamilton numbers), $\mathbb{H}$; octonions (or Cayley numbers), $\mathbb{O}$. As proved in [2], if it is possible to define a vector cross product in the real Euclidean space $\mathbb{R}^n$, satisfying the same requirements as the usual one considered in $\mathbb{R}^3$, then $n = 1$ (trivial case), $n = 3$ or $n = 7$.

One can formulate the definition of the vector cross product $\times$ in $\mathbb{R}^7$ in terms of the product of octonions, recalling that $\mathbb{R}^7$ can be regarded as the real linear space Im $\mathbb{O}$ of the pure octonions. Hypercomplex numbers - *i.e.*, quaternions and oconions - play an important role both in Mathematics and in Physics. We will mention *en passant* some of these (*e.g.*, in [1], connections with Clifford algebras, spinors, Bott periodicity, projective and Lorentzian geometry, Jordan algebras, exceptional Lie groups, quantum logic, special relativity and supersymmetry are described).

This talk is devoted to two Analytic Geometry problems. Concerning the first one, we start obtaining the orthogonal projection of the origin of the coordinates onto an octonion equation of a straight line. Further, taking the norm of the pure octonion belonging to the given line which is closest to the origin, we find the distance from the mentioned point to this line. The cited problem is then extended to the case of the orthogonal projection of an external point onto a line over the octonions. In our approach, we give a geometric construction where properties of the octonions and the matrix representation of the vector cross product are applied. Finally, the best pair of two skew lines over the octonions is obtained, that is, we present the points defining the shortest segment connecting two skew lines over the octonions, allowing us to compute the distance between those lines. As our ground space is an inner product one, by [3], we have guaranteed the existence and uniqueness of the best approximation points we are looking for. The above results and approach appear in [9].

The second problem is focused on the generalized Euler's formula for rotations in $\mathbb{R}^7$ that preserve an axis of rotation (observe that, contrarily to what happens in $\mathbb{R}^3$, it is not necessarily true that every rotation matrix has a rotation axis). The key issue for the alternative proof here presented is the construction of the rotation matrix in $\mathbb{R}^7$.

Keywords: octonion equation of a line, orthogonal projection, parallel and perpendicular octonions, matrix representation of the vector cross product in $\mathbb{R}^7$, best approximation pair, Euler's rotation formula in $\mathbb{R}^7$ (see [6]).

—⊙—
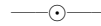
   Though not mentioned in the Abstract, the oral communication includes two tables (the octonion multiplication table and the vector cross product multiplication table in $\mathbb{R}^7$) and two figures (the Fano plane and one example of the orthogonal projection).

References

[1]   Baez, J. C. (2002) The Octonions. *Bulletin of the American Mathematical Society* **39**: 145–205.

[2]   Brown, R. B. and Gray, A. (1967) Vector Cross Products. *Commentarii Mathematici Helvetici* **42**: 222–236.

[3]   Laurent, P.-J. (1972) *Approximation et Optimisation*, Hermann, Paris.

[4]   Leite, F. S. (1993) The Geometry of Hypercomplex Matrices. *Linear and Multilinear Algebra* **34**: 123–132.

[5]   Leite, F. S. and Crouch, P. (1997) The Triple Cross Product in $\mathbb{R}^7$. *Reports on Mathematical Physics* **39**: 407–424.

[6]   Leite, F. S. and Saraiva, P. (1997) An alternative proof for the generalized Euler's formula. *Proc. Meeting on Matrix Analysis and Applications*: 307–312, Seville, Spain.

[7]   Luenberger, D. G. (1969) *Optimization by Vector Space Methods*, John Wiley and Sons, New York .

[8]   Okubo, S. (2005) *Introduction to Octonion and other Non-Associative Algebras in Physics*, Cambridge University Press, New York.

[9]   Saraiva, P., Beites, P. D., Fernandes, J., Costa, C. and Vitória, J. (2015) Best pair of two skew lines over the octonions. *Advances in Applied Clifford algebras* **25(3)**: 657–672.

[10]   Ward, J. P. (1997) *Quaternions and Cayley Numbers*, Kluwer, Dordrecht.

Organized Session 10

Statistical Models in Applied Sciences

Organizer: Rosário Ramos (Portugal)

t-Student and Mann-Whitney tests applied to small sample data - inevitable but uninformative?

F. Caramelo[1,2], B. Oliveiros[1,2] and M. Patrício[1,2]

[1]University of Coimbra, Faculty of Medicine
[2]Laboratory of Biostatistics and Medical Informatics and IBILI

E-mail addresses: *fcaramelo@fmed.uc.pt; boliveiroso@fmed.uc.pt; mjpd@uc.pt*

Abstract

Health Science research has strong boundaries due to ethical issues, which frequently impact in different aspects, notably the sample size. Sample size tends to be very small. In this work we focused in the comparison between the *t*-Student and the Mann-Whitney test applied to small independent samples.

Keywords: Mann-Whitney, power analysis, small sample size, *t*-Student.

——⊙——

There is the general notion [1,2] that samples sized less than 6 should be tested resorting to nonparametric procedures since the normal distribution assumption cannot be tested properly. Even so, many studies have been performed to assess the feasibility of using the *t*-test despite the violation of the assumptions. The published studies can be grouped into two main categories: those that evaluate the *t*-test under different conditions [3,4] and those that perform a direct comparison between the *t*-test and the Mann-Whitney test [5–7]. Despite the knowledge acquired about the two tests, some points regarding small data application remain unclear.

In this simulation study, in order to assess which test, *t*-Student or Mann-Whitney, should be preferred for analysing small sample data the power attained by each test was computed considering different situations. The power function of the *t*-Student and the Mann-Whitney tests was computed using a Monte-Carlo simulation for different population distributions (Normal, $Beta(2, 2)$, $Beta(6, 2)$, Uniform, $Weibull(1, 1)$, $Chi-square(4)$, $Gamma(2, 1)$ and *t*-Student(5-30)), different effect sizes (0.5, 0.8, 1.2 and 1.8) and samples sizes (from 3 to 30). Two sets of values, following one of the aforementioned distributions and having equal variance but different mean values according to one of the effect size values, were created. Samples were then taken from each distribution and the two tests were applied to compare the two samples. The process was repeated 500 times and the power was computed as the frequency of rejection of the null hypothesis.

As expected, the mean power of the two tests increases both with the sample size and the effect size. For extremely small sample sizes ($n < 5$) the *t*-test shows in average (without statistical meaning) more power than the Mann-Whitney test, however the power attained is very small. For small effect sizes both tests present very small power, for example for effect size 0.5 and sample size $n = 5$ the confidence intervals for the power obtained, respectively for t-Student and Mann-Whitney, with a population following a normal distribution were IC95%[7.6%; 11.6%] and IC95%[5.0%; 9.0%].

The Mann-Whitney test performed better than the t-test in samples taken from skewed distributions (e.g., $Beta(2, 2)$, $Weibull(1, 1)$, $Chi-square(4)$ and $Beta(6, 2)$) whereas

the t-test showed better results in the other studied distributions. Regardless of the distribution for larger effect sizes (>1.8), the two tests showed equivalent results for samples sized larger than 20.

Two main conclusions result of this work: firstly, the two tests may present different results if samples are taken from skewed distributed populations. Secondly, small sample sizes should only be considered if the expected effect size is high and, even in those situations, the power of tests can be very small. The resulting p-value should be interpreted with caution.

Acknowledgements

References

[1] Siegel, S. (1957) *Nonparametric statistics.* The American Statistician **11**: 13–19.

[2] Elliott, A. C. and Woodward, W. A. (2007) *Comparing one or two means using the t-test.* Statistical Analysis Quick Reference Guidebook. SAGE.

[3] Harry O. P. (1979) The robustness of the one-sample t-test over the pearson system. *Journal of Statistical Computation and Simulation* **9(2)**: 133–149.

[4] deWinter, J. F. (2013) Using the Student's t-test with extremely small sample sizes. *Practical Assessment, Research & Evaluation* **18(10)**: 1–12.

[5] Harry, O. P. (1982) Two-sample wilcoxon power over the pearson system and comparison with the t-test. *Journal of Statistical Computation and Simulation* **16(1)**: 1–18.

[6] Zimmerman, D. W. (1987) Comparative Power of Student T Test and Mann-Whitney U Test for Unequal Sample Sizes and Variances. *The Journal of Experimental Education* **55(3)**: 171–174.

[7] Gibbons, J. D. and Chakraborti, S. (1991) Comparisons of the Mann-Whitney, Student's t, and Alternate t Tests for Means of Normal Distributions. *The Journal of Experimental Education* **59(3)**: 258–267.

Study of the differences in mortality and length of hospital stay in acute myocardial infarction at hospital level

Teresa Magalhães[1], Sílvia Lopes[1] and João Ferreira Gomes[2]

[1]Public Health Research Center, Department of Health Systems Policy and Management, National School of Public Health, Universidade Nova de Lisboa, Lisbon, Portugal
[2]Centro de Matemática e Aplicações Fundamentais, Faculdade de Ciências da Universidade de Lisboa, Lisbon, Portugal

E-mail addresses: *teresa.magalhaes@ensp.unl.pt; silvia.lopes@ensp.unl.pt; jjgomes@campus.ul.pt*

Abstract

Context: The appropriate use of hospital resources is important for controlling spending and improving health care outcomes. Understand the factors that explain the results in length of hospital stay (LOS) and mortality, through the unjustified variation of associated waste is the objective of this investigation. The Acute Myocardial Infarction (AMI) was chosen for the high morbidity and mortality, relevant length of hospital stay and mortality variability and high-related treatment costs.

Methodology: To study the differences in mortality and length of hospital stay in acute myocardial infarction at hospital level, the patients discharged from NHS hospitals in Portugal mainland from 2011-13 with anonymized administrative data were selected. The relevant variables to explain the differences in mortality and length of hospital stay were grouped in a hierarchical data structure: patient characteristics (demographics, AMI type, comorbidities and procedures), other characteristics (time to hospital) and hospital characteristics (Hospital with "coronary green way", volume and faculty hospital). In data analysis multilevel models were used (1st level-patient; 2nd level-hospital). Generalized linear mixed models and Gaussian mixture models were applied:to counting and continuous models for the hospital length of stay and binary regression for mortality. We quantified waste in length of hospital stay as the excess in number of days and avoidable deaths in mortality, through the hospital random effect variance associated with the considered covariates unexplained variability at hospital level and at patient level.

Results: 28.681 cases were considered and differences were found in mortality and in length of hospital stay in the group of hospitals analysed. We observed that the variable groups selected for the study explained 38% of the differences in length of hospital stay and 29,8% of the differences in mortality. Considering groups, the main patient characteristics that explain the differences between hospitals are: (1) Comorbidities in both indicators (length of hospital stay 24,1%; mortality 23,2%), (2) the procedures in length of hospital stay (7,8%) and (3) the demographic variables in mortality (7,8%). The hospital characteristics explained few of these differences (length of hospital stay 8%; mortality - not significant). We observed a protective factor from the primary angioplasty, since all other procedures showed a comparatively increased risk of death and length of hospital stay, with the angioplasty without fibrinolysis as an exception in mortality. The existence of one hospital with "coronary green way" was also an effective protector for the length of hospital stay. The variability found at hospital level corresponded to a total waste of 2392 treatment days and an average day variation by patient between -2,4 (better hospital) and 3 (hospital with higher waste). In mortality, the variability found corresponded to a

total 137 avoidable deaths with a variation between -81 deaths (better hospital) and +133 deaths (worst hospital). The associated waste to those days was evaluated between 1.5M€ and 2.3M€.

Conclusion: this study has shown that there are differences between hospitals in the length of hospital stay and mortality in acute myocardial infarction, indicating variability of practices. The waste found in length of hospital stay and in avoidable deaths shows values that demand root cause investigation, mainly in the hospitals with worse performance. As evaluated, the hospital characteristics have little explanatory power of the detected differences and therefore the inefficiencies can be attributed to the internal organization of hospitals and how the process of stay of the patient is managed. These results raise relevant concerns at system and hospitals levels, in particular the compliance with treatment guidelines in AMI and the integrated care level on the monitoring of the patient in the prevention and post-discharge.

Keywords: acute myocardial infarction, Gaussian mixture models, generalized linear mixed models, hospital waste, length of hospital stay, mortality.

References

[1]   Häkkinen, U., Rosenqvist, G., Peltola, M., Kapiainen, S., Rättö, H., Cots, F., et al. (2014) Quality, cost, and their trade-off in treating AMI and stroke patients in European hospitals. *Health Policy* **117(1)**: 15–27. Doi: `10.1016/j.healthpol.2014.05.001`.

[2]   Stargardt, T., Schreyogg, J. and Kondofersky, I. (2014) Measuring the relationship between costs and outcomes: the example of acute myocardial infarction in german hospitals. *Health Economics* **23**: 653–669.

[3]   Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A. and Smith, G. M. (2009) *Mixed effects models and extensions in ecology with R*, New York: Springer-Verlag. (Statistics for Biology and Health). Doi: `10.1016/B978-0-12-387667-6.00013-0`.

A comparative study of the estimators for the demand of Engineering Courses in Portugal

Raquel Oliveira[1,2], A. Manuela Gonçalves[1,3] and
Rosa M. Vasconcelos[1,4]

[1]University of Minho, Portugal
[2]CMAT - Centre of Mathematics, Portugal, IPCA-EST
[3]CMAT - Centre of Mathematics, DMA - Department of Mathematics and Applications
[4]2C2T - Centre for Textile Science and Technology, DET - Department of Textile
Engineering

E-mail addresses: *rmro_17@hotmail.com; mneves@math.uminho.pt;
rosa@det.uminho.pt*

Abstract

With the purpose of modelling the demand in Engineering Courses in Portugal we analyzed the possible regression models for panel count data models, establishing a comparison between the estimators encountered, finding the most appropriate for our data set. A precise quantification of the demand for each academic program is facilitated by the rules of access to higher education, in a national contest, where candidates must list up to six preferences of institution and program. The data used in this paper covers the results of the national contest concerning the years from 1997 to 2015, provided by the Portuguese Ministry of Education. Multivariate methodologies were performed in order to allow a better understanding of the student's allocation behavior before and after the implementation of the Bologna process. Results seems to indicate that the negative binomial estimates fits better the data set analyzed.

Keywords: higher education, students access, panel count data, regression models.

Acknowledgements

References

[1] Cameron, A. C. and Trivedi, P. K. (1986) *Econometric models based on count data:comparisons and apllications of some estimators and tests*, Cambridge: Cambridge University Press.

[2] Cameron, A. C. and Trivedi, P. K. (1998) *Regression analysis of count data*, Cambridge: Cambridge University Press.

[3] Cardoso, A. R., Portela, M., Sá, C. and Alexandre, F. (2008) Demand for higher education programs: the impact of the Bologna Process. *CESifo Economic Studies* **54(2)**: 229–247.

[4] Oliveira, R., Gonçlves, A. M. and Vasconcelos, R. M. (2013) negative binomial model
     for student allocation to higher education in Portugal during the pre-Bologna period.
     AIP Conf. Proc. **1558**: 1881–1884.

[5] Portela, M., Sá, C., Alexandre, F. and Cardoso, A. R. (2009) Perceptions of the
     Bologna process: what do students' choices reveal? *High Educ* **58**: 46–474.

[6] Winkelman, R. (2003) *Econometric analysis of count data*, Springer-Verlag, Heidel-
     berg.

At which sample size is Monte Carlo Cross-Validation reliable for dichotomous classifiers?

M. Patrício[1,2] and F. Caramelo[1,2]

[1]University of Coimbra, Faculty of Medicine
[2]Laboratory of Biostatistics and Medical Informatics and IBILI

E-mail addresses: *mjpd@uc.pt; fcaramelo@fmed.uc.pt*

Abstract

The present work shows that large sample sizes have to be considered for cross-validation estimates to be reliable. However, the combination of Support Vector Machine classifiers, based on quantitative predictors, with Monte Carlo Cross-Validation ensures internal validity with more modest sample sizes. An algorithm to compute the adequate sample size is proposed.

Keywords: Monte Carlo, validation, classifier, simulation.

———⊙———
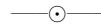
It is common practice that the design of clinical studies makes use of formulas to compute the minimum sample size that ensures that a clinically significant result will border on being statistically significant, [1]. However, it is less clear how to compute the sample size required for building up valid dichotomous classifiers, which are known to be data-hungry, [2, 3]. A distinction must be made between assessing performance of the models or checking for their validity, [4, 5, 6]. Focusing solely on internal validity, a recent study found that a stable area under the Receiver Operating Characteristic curve (AUC) could be reached with logistic regression modelling, following a split-sample approach, by taking a sample size 20 to 50 times the number of variables, [7]. The goal of the present study is to assess the reliability of the Monte Carlo Cross-Validation [8], in relation to the sample size, as a method to perform internal validation.

Throughout this simulation study, results were obtained for both Logistic Regression (LR) and Support Vector Machines (SVM). The data was generated artificially by replicating 20 times the data set of the study on breast cancer reported in [9] and deriving the outcomes from the models considered, likewise to the procedure described in [7]. Separately for LR and SVM, the corresponding data set was then randomly split in equal parts between two sets, henceforth called development set and validation set and: 1- For each number n in the set 30, 40, 50, ..., 250, a random sample (without repetition) with n elements was drawn from the development set. A Monte Carlo Cross-Validation procedure, with 500 random splits of the data, was then undertaken. Each of the 500 models generated was also applied to the validation set. The AUC was computed for each model both over the corresponding test set and the validation set, thus allowing the computation of the percentage of models for which the two AUC values were close, which was defined as occurring when the absolute difference between them was less than 0.1. The entire procedure was repeated 100 times for each value of n. This allowed computing the average percentage of times the cross-validation procedure yielded an AUC value close to that obtained over the validation set. 2- The procedure described in 1 was repeated. However, this time we resorted to the secant method to find a sample size for which the percentage

of times the cross-validation procedure yielded an AUC value close to that obtained over the validation set ranged between 78 and 82 percent.

The sample size was found to have to be nearly 15 times greater than the number of variables for the Monte Carlo Cross-Validation procedure to attain reliable results for the logistic regression. For the Support Vector Machine procedures, the sample size was found to have to be 9 times greater.

Acknowledgements

References

[1] McCrum-Gardner, E. (2010) Sample size and power calculations made simple. *International Journal of Therapy and Rehabilitation* **17(1)**.

[2] Foster, K. R., Koprowski, R. and Skufca, J. D. (2014) Machine learning, medical diagnosis, and biomedical engineering research - commentary. *Biomed Eng Online* **13**: p. 94.

[3] Selker, H. P., et al. (1995) A comparison of performance of mathematical predictive methods for medical diagnosis: identifying acute cardiac ischemia among emergency department patients. *J Investig Med.* **43(5)**: 468–76.

[4] Lasko, T. A., et al. (2005) The use of receiver operating characteristic curves in biomedical informatics. *J Biomed Inform.* **38(5)**: 404–15.

[5] Steyerberg, E. W. and Vergouwe, Y. (2014) Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J.* **35(29)**: 1925–31.

[6] Steyerberg, E. W., et al. (2001) Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* **54(8)**: 774–81.

[7] van der Ploeg, T., Austin, P. C. and Steyerberg, E. W. (2014) Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. *BMC Med Res Methodol* **14**: p. 137.

[8] Xu, Q. S. and Liang, Y. Z. (2001) Monte Carlo cross validation. *Chemometrics and Intelligent Laboratory Systems* **56(1)**: 1–11.

[9] Crisóstomo, J., et al. (2016) *Hyperresistinemia and metabolic dysregulation: the close crosstalk in obese breast cancer.* Endocrine.

Organized Session 11

Statistical inference for discrete-valued time series

Organizer: Isabel Silva (Portugal)

Surveillance in Discrete Time Series

Maria da Conceição Costa[1,2], Isabel Pereira[1,2] and
Manuel G. Scotto[3,4]

[1]University of Aveiro
[2]Center for Research and Development in Mathematics and Applications, CIDMA
[3]Instituto Superior Técnico, University of Lisbon
[4]Center for Computational and Stochastic Mathematics, CEMAT

E-mail addresses: *lopescosta@ua.pt; isabel.pereira@ua.pt;
manuel.scotto@tecnico.ulisboa.pt*

Abstract

Time series of small non-negative counts have become available in a wide variety of contexts over the last years and there are many applications for which a critical situation related to a level-crossing event can be defined. In this work, we focus our attention in a particular class of models for non-linear time series of counts and explore the implementation of an optimal alarm system for this class. The class of models is presented, the methodologies involved are illustrated and the prediction capability of the alarm system is demonstrated in practice.

Keywords: ergodicity, maximum likelihood, non linear time series, observation-driven models, optimal alarm systems.

———⊙———

The analysis of low integer-valued time series is an area of growing interest as, in the last three decades, time series of counts arising from many different areas have become available. Statistical quality control, computer science, economics and finance, medicine and epidemiology and environmental sciences are just some of the fields that we can mention to point out the wide variety of contexts from which discrete time series have emerged.

In many of these areas is not just the analysis and modelling of count data that matters. For instance, in environmental sciences or epidemiology, surveillance is critical and timely intervention is mandatory in order to ensure safety and public health. Actually, a major issue in the analysis of a large variety of random phenomena relates to the ability to detect and warn the occurrence of a catastrophe or some other event connected with an alarm system [1,3].

In this work, the principles for the construction of optimal alarm systems are discussed and their implementation is described. As there is no unifying approach to the modelling of all integer-valued time series, we will focus our attention in the class of observation-driven models and the implementation of the optimal alarm system will be described in detail for a particular non linear model in this class, the INteger-valued Asymmetric Power ARCH, or, in short, INAPARCH(p,q). The INAPARCH model is the integer-valued counterpart for the APARCH representation for the volatility introduced by Ding et al. [2] and is able to accommodate asymmetric responses relative to the mean of the process. The probabilistic properties and asymptotic theory related to maximum likelihood estimation of this model are also covered.

Acknowledgements

References

[1] de Maré, J. (1980) Optimal prediction of catastrophes with applications to Gaussian processes. *The Annals of Probability* **8(4)**: 841–850.

[2] Ding, Z., Granger, C. W. and Engle, R. F. (1993) A long memory property of stock market returns and a new model. *Journal of Empirical Finance* **1(1)**: 83–106.

[3] Lindgren, G. (1985) Optimal prediction of level crossings in Gaussian processes and sequences. *The Annals of Probability* **13(3)**: 804–824.

Simple integer-valued bilinear model: inference, prediction and diagnostic

Isabel Pereira[1,2] and Nélia Silva[1,2]

[1]University of Aveiro
[2]Center for Research and Development in Mathematics and Applications, CIDMA

E-mail addresses: *isabel.pereira@ua.pt; neliasilva@ua.pt*

Abstract

In this work, the integer-valued bilinear INBL (1,0,1,1) model is discussed. For the estimation of the parameters, the maximum likelihood estimator and Bayesian methodology are considered and a comparison of results is presented using a simulation study. Furthermore, the problem of predicting future observations from the classical and Bayesian approaches is analyzed. Since the evaluation of predictive performance and suitability of the model are important issues, the probabilistic forecast is compared with the true data-generating distribution. This comparison is made using the Probability Integral Transform (PIT) and applied to real data sets.

Keywords: Bayesian methodology, integer-valued bilinear process, probability integral transformation.

——⊙——

Several models that take the discreteness of the data explicitly into account have been developed in the literature. In particular, using the concept of binomial thinning, also conventional bilinear models can be adapted to the integer case leading to the class of integer-valued bilinear (INBL) models. This class of models is particularly suitable for modeling processes which assume low values with high probability, but exhibiting, at the same time, sudden bursts of large values. Doukhan et al. [1] analyzed the special INBL(1, 0, 1, 1) model and obtained conditions to ensure that this process is second--order stationary. Considering binomial thinning operators as well, Drost et al. [2] introduced the super diagonal INBL process, which is a particular class of the more general INBL(p, q, m, n) and derived a condition guaranteeing the existence of a unique strictly stationary process.

Considering the binomial thinning operator, the present work intends to continue the research in the particular INBL(1,0,1,1) model. This first-order integer-valued bilinear model is discussed and some theoretical properties are reviewed. Classical and Bayesian methodologies are considered and compared, namely to obtain estimates of model parameters and to calculate point and intervalar predictions. Probability Integral Transform(PIT) values were used to evaluate predictive performance of the process. Finally, the suggested approaches are illustrated with its application to real datasets of E.coli infections and meningitis cases.

Acknowledgements

References

[1] Doukhan, P., Latour, A. and Oraichi, D. (2006) A simple integer-valued bilinear time series models. *Advances in Applied Probability* **38**: 559–578.

[2] Drost, F. C., van den Akker, R. and Werker, B. J. M. (2008) Note on integer-valued bilinear time series models. *Statistics and Probability Letters* **78**: 992–996.

Outliers detection in integer-valued time series

Isabel Silva[1,3] and Maria Eduarda Silva[2,3]

[1]Faculdade de Engenharia, Universidade do Porto
[2]Faculdade de Economia, Universidade do Porto
[3]CIDMA

E-mail addresses: *ims@fe.up.pt; mesilva@fep.up.pt*

Abstract

The presence of outliers or discrepant observations has a negative impact in time series modelling. In this paper, we consider first-order integer-valued autoregressive (INAR(1)) models contaminated with additive and innovational outliers. We propose two methods based on wavelets in order to address the problem of identifying the times of outlier occurrence. The effectiveness of the proposed methods is illustrated and compared through a simulation study. A real dataset application is also presented.

Keywords: discrete wavelet transform, integer-valued autoregressive model, outlier detection, parametric resamplig, threshold.

———⊙———

A problem of interest in time series modelling is detecting outliers to assess data quality, to study the robustness of the statistical inference in the presence of inconsistent observations and to find eventual important phenomena underlying the data. It is well known that the presence of outliers in time series can lead to biased parameter estimation, incorrect inference and inadequate predictions. Several methodologies for detecting and estimating outliers have been established for ARMA models. The emphasis is on iterative procedures and likelihood based statistics [8,2,3]. However, the problem of detection and estimation of outliers in time series of counts has received less attention in the literature. Recently, [7] suggested a Bayesian approach in order to detect additive outliers in Poisson first-order INteger-valued AutoRegressive, INAR(1), models. The class of INAR models for time series of counts, first proposed by [1], has been extensively studied in the literature and applied to many real-world problems because of its easiness of interpretation. These models are apparently autoregressive models in which the usual multiplication has been replaced by a random operation, called thinning operation (for details see [6]).

In this work we propose to identify the times of outlier occurrence in INAR(1) time series using wavelets. Wavelets are a family of basis functions used to localize a given function in both space and scaling [5]. In particular, we propose to apply discrete wavelet transform (DWT), which is a powerful tool for a time-scale multi-resolution analysis. DWT can be considered as filters of different cut-off frequencies used to analyse the signal at different scales. The high-scale (low-frequency) components of the signal are contained in the approximation coefficients. On the other hand, the low-scale (high-frequency) components are represented by the detail coefficients.

In a first approach, similar to that of [4], the detail coefficients derived from DWT, using the Haar wavelet, are compared with a threshold. In a second approach, the parametric resampling method of [9] is used in order to obtain the empirical distribution of these detail coefficients. The proposed procedures are illustrated and compared with synthetic data. Furthermore, the methods are also applied on an observed dataset.

Acknowledgements

References

[1] Al-Osh, M. A. and Alzaid, A. A. (1987) First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis* **8**: 261–275.

[2] Chang, I., Tiao, G. C. and Chen, C. (1988) Estimation of time series parameters in the presence of outliers. *Technometrics* **30**: 193–204.

[3] Chen, C. and Liu, L. M. (1993) Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association* **88**: 284–297.

[4] Grané, A. and Veiga, H. (2010) Wavelet-based detection of outliers in financial time series. *Computational Statistics & Data Analysis* **54**: 2580–2593.

[5] Percival, D. and Walden, A. (2006) *Wavelet methods for time series analysis*, Cambridge Series in Statistical and Probabilistic Mathematics, New York, Cambridge University Press.

[6] Scotto, M. G., Weiß, C. H. and Gouveia, S. (2015) Thinning-based models in the analysis of integer-valued time series: a review. *Statistical Modelling* **15**: 590–618.

[7] Silva, M. E. and Pereira, I. (2015) Detection of additive outliers in Poisson INAR(1) time series. In Bourguignon, J. P. *et al.* (Eds.) *CIM Series in Mathematical Sciences - Mathematics of Energy and Climate Change*, Springer: 377–388.

[8] Tsay, R. S. (1986) Time series model specification in the presence of outliers. *Journal of the American Statistical Association* **81**: 132–141.

[9] Tsay, R. S. (1992) Model checking via parametric bootstraps in time series analysis. *Journal of the Royal Statistical Society: Series C* **41**: 1–15.

On the maxima of integer models based on a new thinning operator

Maria da Graça Temido[1,2] and Sandra Dias[3,4]

[1]University of Coimbra
[2]Dept. of Mathematics, Faculty of Sciences and Technology/CMUC
[3]University of Trás-os-Montes e Alto Douro
[4]Dept. of Mathematics, School of Sciences and Technology/CMAT-UTAD and CEMAT

E-mail addresses: *mgtm@mat.uc.pt; sdias@utad.pt*

Abstract

In this work we introduce and study an integer-valued process referred as $\Psi-$INARMA(1,1). We prove its strictly stationarity and specify its asymptotic independence and local dependence behaviour. As a consequence, we conclude that the sequence of maxima converges in distribution to a Gumbel discrete distribution, when the residual sequence belongs to Anderson's class.

Keywords: Anderson's class, extreme value theory, integer-valued stationary models, thinning operators.

———⊙———
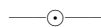
van Harn *et al.* [3] introduced a new thinning operator, denoted by $\odot_\Psi$, an extension of the well known binomial thinning operator proposed by Steutel and van Harn [4]. For a given integer valued random variable (r.v.), X, and $\eta \in (0,1)$, they define

$$\eta \odot_\Psi X = \sum_{i=1}^{X} Y_i,$$

where $\{Y_i\}$ is a sequence of i.i.d. r.v.'s independent of X, with common probability generating function $\Psi_{-\ln \eta}(s)$, that belongs to a special semi-group of functions. In Aly and Bouzar [1] this thinning operator was used instead of the binomial thinning operator to develop a family of integer processes. We propose and study a generalization of the geometric ARMA(1,1) process, proposed by McKenzie [5],

$$X_n = \beta \star Z_n + V_n W_{n-1}, \ \text{with} \ \ W_n = \eta \star W_{n-1} + U_n Z_n,$$

where $\{Z_n\}$, $\{U_n\}$ e $\{V_n\}$ are i.i.d. r.v.'s, $\{U_n\}$ and $\{V_n\}$ have a Bernoulli distribution (with parameters $1-\eta$ e $1-\beta$, respectively) and W_0 is independent of all the other r.v.'s. The generalized process, based on the operator $\odot_\Psi$, here called Ψ-INARMA(1,1) process, is defined by

$$X_n = \beta \odot_\Psi Z_n + V_n W_{n-1}, \ \text{where} \ W_n = \eta \odot_\Psi W_{n-1} + U_n Z_n,$$

with $\{Z_n\}$, $\{U_n\}$ and $\{V_n\}$ satisfying the same conditions. After proving the strict stationarity of the process, we hold true the asymptotic independence and the local dependence behaviour induced by conditions $D_{k_n}(u_n)$ and $D'_{k_n}(u_n)$, introduced in Temido and Canto e Castro [7], where $\{k_n\}$ is a non-decreasing integer sequence such that $k_{n+1}/k_n \to r > 1$, $n \to +\infty$. In this work we assume that the marginal distribution $\{Z_n\}$ belongs to

the so-called Anderson's class [2], here denoted by $\mathcal{C}_{\mathcal{A}}(r_Z)$, which includes all distribution functions F_Z satisfying

$$1 - F_Z(z) \sim A[z]^{\xi} r_Z^{-[z]}, \ z \to +\infty,$$

where $\xi \in \mathbb{R}$, $A > 0$ and $r_Z > 1$. We also take into account the main result of Temido [6], where it is proved that if $\{Z_n\}$ belongs to $\mathcal{C}_{\mathcal{A}}(r_Z)$, then $F_Z^{k_n}(x + b_n)$, $b_n \in \mathbb{R}^+$, converges in distribution to a discretized Gumbel distribution, $G(x) = \exp(-r_Z^{-[x]})$, $x \in \mathbb{R}$. Establishing that $\{W_n\}$ and $\{X_n\}$ also belong to $\mathcal{C}_{\mathcal{A}}(r_Z)$, we conclude that the sequence of maxima of the process $\Psi-\text{INARMA}(1,1)$ is attracted in distribution to a discretized Gumbel distribution.

Acknowledgements

References

[1]   Aly, E. A. and Bouzar, N. (2005) On a class of $\mathbb{Z}_+-$valued autoregressive moving average (ARMA) processes. *REVSTAT- Statistical Journal* **6**: 101–121.

[2]   Anderson, C. W. (1970) Extreme value theory for a class of discrete distribution with applications to some stochastic processes. *Journal of Applied Probability* **7**: 99–113.

[3]   van Harn, K., Steutel, F. W. and Vervaat, W. (1982) Self-decomposable discrete distributions and branching processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **61**: 97–118.

[4]   Steutel, F. W. and van Harn, K. (1979) Discrete analogues of self-decomposability and stability. *Annals of Probability* **7**: 893–899.

[5]   McKenzie, E. (1986) Auto regressive-moving-average processes with negative binomial and geometric marginal distribution. *Advances in Applied Probability* **18**: 679–705.

[6]   Temido, M. G. (2002) Domínios de atracção de funções de distribuição discretas. In Carvalho, L. *et al.* (eds): *Novos Rumos em Estatística*, 415–426, Edições SPE.

[7]   Temido, M. G. and Canto e Castro, L. (2003) Max-semistable laws in extremes of stationary random sequences. *Theory of Probability and its Applications* **47**: 365–374.

Organized Session 12

Computational Data Analysis and Numerical Methods

Organizer: Luís Grilo (Portugal)

Non-parametric Individual Control Chart for Silica in Water

Luís M. Grilo[1,2,4], Mário A. Santos[3] and
Helena L. Grilo[4]

[1]Departamento de Matemática e Física, Instituto Politécnico de Tomar
[2]Centro de Matemática e Aplicações - CMA/FCT-UNL, Portugal
[3]Departamento de Engenharia, Instituto Politécnico de Tomar, Portugal
[4]Centro de Sondagens e Estudos Estatísticos, Instituto Politécnico de Tomar, Portugal

E-mail addresses: *lgrilo@ipt.pt; magsantos91@gmail.com;*
helenagrilo56@gmail.com

Abstract

The soluble silica content in the demineralized water is a continuous variable measured and controlled in the Chemical Laboratory of a Portuguese thermoelectric central, in order to keep the equipment operating in the best conditions, allowing, in particular, to extend its useful life. In this case study, this variable could be considered approximately normal distributed, and because we just have one measure for each group of the sample, an individual control chart to monitor the silica content is obtained based on average moving range. Once the sample size available is small, robust control limits using a non-parametric method based on empirical quantiles (that performs also well under the normality of the observations) are also estimated with the bootstrap procedure. Since the silica content should be as small as possible, it is very important for technicians to compare the control limits obtained with different approaches. Although the process seems stable, does not stay within the engineering specification limits.

Keywords: Bootstrap, demineralized water, empirical quantiles, process capability, small sample.

——⊙——

Demineralized water is indispensable for the energy production process in a Portuguese thermoelectric central, since the formation of high pressure water steam is obtained by heat transfer between the boiler and demineralized water, which circulates in adjacent pipes. This water steam provides the movement of the turbine blades producing mechanical energy, which is then converted into electricity in a generator. The variable silica (in μg SiO2/L), soluble in water, has to be removed because this chemical compound has a tendency to form deposits on the walls of the equipment and piping. Furthermore, when water boils in the boiler, this silica has a high abrasive effect on its metal walls and being entrained in the steam duct, will cause wear on turbine blades. To monitor eventual changes in the process, Shewhart individual control chart (X) for the variable silica is obtained with individual observations stem from a process which is statistically in control (as in [3,4,2,3]), computing the control limits based on the average moving range (AMR). Because the sample size is small, we also estimate robust control limits using a non-parametric method based on empirical quantiles (EQ), to turn the X control chart into a more sensitive one to persistent assignable causes [2]. This alternative control chart is a special case of the bootstrapping control chart and is not only quite robust

against deviations from normality but also performs reasonably well under normality of the observations ([4,5]).

The individual control chart and the approaches considered (AMR and EQ) to obtain different control limits are very important for the members of the Chemical Laboratory of the thermoelectric central, once they allow to evaluate the silica variable and their comparison shows how important is the sample size, no matter the involved costs. The process capability developed also gives a considerable information about how much the process should be improved, since here it is necessary to continually try to minimize the variation and strive to centre the process.

Acknowledgements

References

[1]   Wheeler, D. J. (1995) *Advanced Topics in Statistical Process Control*, SPC Press, Knoxville, Tennessee.

[2]   Grilo, L. M. and Grilo, H. L. (2015) Individual and moving range control charts in the production of olive oil. *ICNAAM 2014, AIP Conf. Proc.* **1648**: 840011, 1–4.

[3]   Grilo, L. M. and Grilo, H. L. (2016) Comparison of individual charts to monitor peroxide index of olive oil. *Proceedings in Advances in Mathematics and Computer Science and their Applications (Math. and Comp. in Sc. and Eng. Series* **57**: 272–275.

[4]   Grilo, L. M., Silva, D. S., Nogueira, I. M., Grilo, H. L. and Oliveira, T. A. (2016) Individual Control Charts in Paperboard Industry. *ICCMSE 2016, AIP Conf. Proc.* (accepted for publication).

[5]   Vermaat, M. B., Does, R. J. M. M. and Klaassen, C. A. J. (2003) A comparison of Shewhart individuals control charts based on normal, non-parametric, and extreme-value theory. *Quality and Reliability Engineering International*, **19(4)**: 337–353.

[6]   Willemain, T. R. and Runger, G. C. (1996) Designing control charts using an empirical reference distribution. *Journal of Quality Technology*, **28(1)**: 31–38.

Misclassification and asymmetry in binary regression model

<u>Jacinto Martín</u>[1], Lizbeth Naranjo[2] and Carlos J. Pérez[3]

[1]Facultad de Ciencias and Instituto de Computación Científica Avanzada (ICCAEx), Universidad de Extremadura, Badajoz, Spain
[2]Facultad de Ciencias, Universidad Nacional Autónoma de México, México D.F., Mexico
[3]Facultad de Veterinaria, Universidad de Extremadura, Cáceres, Spain

E-mail addresses: *jrmartin@unex.es; lizbethna@ciencias.unam.mx; carper@unex.es*

Abstract

Sometimes, data-generating processes are not error free when data are collected in the real world. For categorical data this is named misclassification. In this context, misclassification parameters are necessary to correct the bias yielded by the use of the non error-free data. If the error in a data generating process is not properly modeled, the information may be perceived as being more accurate than it actually is, leading, in many cases, to a non optimal decision making. Therefore, statistical models should address misclassification (see [2]).

On the other hand, sometimes a critical issue when modeling binary response data is the choice of the link function. The most popular models for misclassified binary response data are the logistic and probit ones. However, in some applications the overall fit can be significantly improved by the use of asymmetric links. In order to describe a link, [1] considered the rates at which the probabilities of a given binary response approaches 1 or 0. Under this notion, a link is symmetric if the rates are the same, otherwise the link is skewed or asymmetric. A skewed link can be characterized as positively skewed if the rate approaching 1 is faster than the rate approaching 0, otherwise it is negatively skewed. [3] studied the asymmetric exponential power distribution from a Bayesian viewpoint. Its cumulative distribution function could be used as an asymmetric link.

An approach for binary regression model in the presence of misclassified data under the Bayesian methodology is proposed. An asymmetric link based on the asymmetric exponential power distribution is used. The computational difficulties have been avoided by using data augmentation. The idea of using a data augmentation framework with two types of latent variables is exploited to derive efficient Markov chain Monte Carlo algorithms. Although the augmented models increase the dimensionality, the generation process becomes easier. The potential applicability of this approach to many fields of knowledge makes this proposal interesting.

Keywords: asymmetric exponential power distribution, Bayesian methods, binary regression, link functions, misclassification.

—⊙—
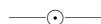
A simulation example illustrates the model performance when comparing with standard methods that do not consider skewness and/or that do not consider misclassification. It has been proved that when data are generated from skewed link models and errors are introduced in the responses, the skewed models considering misclassification perform better than the symmetric link model (considering misclassification) and than the standard skewed link models (error free). This suggests that the proposed models can be used as an

alternative for data where the data generation process is not known and it can be possibly related to skewed link-based regressions and data are misclassified.

References

[1] Chen, M.-H., Dey, D. K. and Shao, Q.-M. (1999) A new skewed link model for dichotomous quantal response data. *Journal of the American Statistical Association* **94(448)**: 1172–1186.

[2] Naranjo, L., Martín, J., Pérez, C. J. and Rufo, M. J. (2014) Addressing misclassification for binary data: probit and t-link regressions. *Journal of Statistical Computation and Simulation* **84(10)**: 2187–2213.

[3] Naranjo, L., Pérez, C. J. and Martín, J. (2015) Bayesian analysis of some models that use the asymmetric exponential power distribution. *Statistics and Computing* **25(3)**: 497–514.

Shapes and computer generation of the numerical range of complex matrices on a Hilbert space

<u>Ana Nata</u>[1,2], Natália Bebiano[3,2] and J. da Providência[4]

[1]Polytechnic Institute of Tomar, Portugal
[2]CMUC - Centro de Matemática da Universidade de Coimbra
[3]Department of Mathematics, University of Coimbra, Portugal
[4]Department of Physics, University of Coimbra, Portugal

E-mail addresses: *anata@ipt.pt; bebiano@mat.uc.pt; providencia@teor.fis.uc.pt*

Abstract

In this talk we develop, implement and test an alternative algorithm to compute the numerical range of a complex matrix, constructing 2 by 2 matrix compressions and their easily determined elliptical numerical ranges. The numerical results reported here indicate that this method is very efficient, since it is faster and more accurate than either of the existing algorithms. As a consequence of this characterization, a Matlab program that accurately exhibits the numerical range is presented and its performance will be illustrated by several examples.

Keywords: numerical range, Hilbert space, matrix compression, numerical algorithm.

——⊙——

Let M_n be the algebra of $n \times n$ complex matrices, and $\mathbb{C}^n$ endowed with the Euclidean inner product $\langle x, y \rangle = y^* x$, where y^* denote the conjugate transpose of y. The numerical range of a matrix $A \in M_n$ is the subset of $\mathbb{C}$ denoted and defined by

$$W(A) := \{ x^* A x : x \in \mathbb{C}^n, \ x^* x = 1 \}.$$

Despite the conceptual simplicity of the definition of $W(A)$, this is a remarkable set for many reasons. For example, it succinctly captures information about A as a transformation, particularly about the eigenvalues and, perhaps more importantly, about the eigenspaces of A. Also there are many results concerning the interplay between the algebraic properties of a matrix and the geometrical properties of its numerical range.

In this talk we develop, implement and test an alternative algorithm to compute the numerical range of a complex matrix, constructing 2 by 2 matrix compressions. Our algorithm revisit an earlier idea of Marcus and Pesce from 1987 for generating Hilbert space numerical ranges of matrices of size n. According to this result, for any $A \in M_n$,

$$W(A) = \bigcup_{\tilde{u}, \tilde{v}} W(A_{\tilde{u}\tilde{v}}),$$

where $A_{\tilde{u}\tilde{v}}$ is the 2-dimensional real orthogonal compression of A, and $\tilde{u}$ and $\tilde{v}$ run over all pairs of real orthonormal vectors. Our algorithm improve the Marcus-Pesce algorithm especially for large dimensions complex matrices. Instead of using randomly generated vectors u and v, we make use of suitably chosen vectors u and v which generate boundary points of the numerical range.

References

[1]   Bebiano, N., Providência, J. da, Nata, A. and Providência J. P. da (2015) Computing the numerical range of Krein space operators. *Open Mathematics* **13(1)**: 146–156.

[2]   Marcus, M. and Pesce, C. (1987) Computer generated numerical ranges and some resulting theorems. *Linear and Multilinear Algebra* **20**: 121–157.

[3]   Johnson, C. R. (1978) Numerical determination of the field of values of a general complex matrix. *SIAM J. Numer. Anal.* **15(3)**: 595–602.

[4]   Uhlig, F. (2014) Faster and more accurate computation of the field of values boundary for n by n matrices. *Linear and Multilinear Algebra* **62(5)**: 554–567.

Multivariate analysis applied to school achievement in mathematics in the 3rd cycle - a case study

Maria Soares[1] and Catarina Nunes[1]

[1]Universidade Aberta, Departamento de Ciências e Tecnologia, Portugal

E-mail addresses: *1300709@estudante.uab.pt; CatarinaS.Nunes@uab.pt*

Abstract

The objective of the study is to identify factors that may caracterize the success or failure of students in mathematics.It is also aims to know their opinion about causes of failure and strategies to increase success in this discipline. The study carried out in the academic year 2014/2015, in the 3rd cycle of basic education in Primary and Secondary School Joaquim Inácio da Cruz Sobral in Sobral. The school failure is understood as "the repetition or retention for one or more years during the school career of students" (CNE, 2013: 40). In Portugal, although general school failure as measured by retention rates decreased significantly, still have high values. Currently, there are about 150,000 students who are retained in the same grade. Withholdings begin in very early educational levels and will be accentuated as it advances schooling. Mathematics is one of the subjects that most contributes to the educational failure of many students (Bridge 1988: 10). The research used a questionnaire administered to 200 students covering issues related to the student, the family and the discipline of mathematics. A univariate logistic regression model is tried and all the analysis was performed with IBM SPSS Statistics software, version 22.0, in order to build a model to predict the result of the success / failure of a student in mathematics. The dependent variable of the problem is "reproved ever since the 5th year" reflected in the model as a binary response (yes/no) and the independent variables chosen as candidates to predict failure were sixteen.

Keywords: school failure, logistic regression, mathematics.

References

[1] Ponte, João Pedro (1988) Matemática, Insucesso e Mudança: Problema Possível, Impossível ou Indeterminado? *Aprender* **6**: 10–19.
[2] Conselho Nacional de Educação (2014) *Estado da educação 2013*, Lisboa: Conselho Nacional de Educação (CNE).
[3] Maroco, J. (2011) Análise Estatística com utilização do SPSS. ReportNumber, 5th ed., 801–911.

Organized Session 13

Statistics and Applications in Life Sciences

Organizer: Manuela Neves (Portugal)

Field methods guided by parameter estimation in wind farm monitoring protocols

Regina Bispo[1,2], Joana Bernardino[3,4] and
Tiago A. Marques[5,6]

[1]ISPA - Instituto Universitário
[2]MARE Marine and Environmental Sciences Centre
[3]Bioinsight Environmental Services Lda
[4]CIBIO-InBio Research Center in Biodiversity and Genetic Resources
[5]Centre for Research into Ecological and Environmental Modeling, The Observatory,
University of St Andrews
[6]Centro de Estatística e Aplicações da Universidade de Lisboa, Faculdade de Ciências da
Universidade de Lisboa

E-mail addresses: *rbispo@ispa.pt; joanavbernardino@hotmail.com;
tam2@st-andrews.ac.uk*

Abstract

Wind farm monitoring protocols aiming to estimate avian fatality include a field method known as removal trials. Typically daily visits are used to verify the time until removal of a (frequently small) sample of bird carcasses up to a pre-determined time. Until now these methods have been almost solely guided by empirical information and financial criteria. Information about how to plan field methods aiming to minimize estimation error is almost nonexistent. Hence, here we have studied the error associated with the estimation of carcass persistence time as a function of field methods.

We have analysed data collected from camera trapping trials to compare the influence of using continuous *vs.* censored data on estimation. Using a simulation framework the accuracy and precision of the parameter estimates was studied as a function of sample size, inspection protocol and the risk of carcass removal. The time of carcass persistence was modelled using parametric survival methods.

The results show that conventional carefully planned removal trials, in terms of total length and visits distribution over time, allow to obtain estimates as accurate as the obtained from camera trapping trials. The simulation results indicate that the precision of the estimates may be compromised by (1) a small risk of removal, (2) using daily checks for short follow-up time periods and/or (3) using small sample sizes. The precision increases using longer inspection protocols or increasing the sample size.

We concluded that camera-trapping trials do not necessarily represent an advantage as the estimates obtained from continuous *vs.* censored data are practically the same for well planned conventional trials. From the simulation exercise, we conclude that small length inspection protocols should be replaced by protocols composed by daily visits in the first days after carcass placement, followed by visits spaced in time. Under a high carcass removal risk scenario, we recommend the use a minimum sample size of 30 carcasses; under a lower removal risk scenario (or mix/unknown removal scenarios) our advice is to use the greatest practicable sample size.

Keywords: parameter estimation, fatality estimation, removal trials, field methods, wind farms.

Acknowledgements

Applying mixed models in the selection of ancient grapevine varieties

Elsa Gonçalves[1,2] and Antero Martins[1,2]

[1]Universidade de Lisboa, Instituto Superior de Agronomia, Portugal
[2]LEAF/ISA and Portuguese Association for Grapevine Diversity, PORVID

E-mail addresses: *elsagoncalves@isa.ulisboa.pt; anteromart@isa.ulisboa.pt*

Abstract

The genotypes within an ancient grapevine variety (*Vitis vinifera L.*) are markedly different with respect to the most important cultural and oenological traits. They are precisely these differences within the ancient variety that justify the selection to respond to new demands of growers and consumers and to adapt the variety to new environmental contexts. In Portugal, the grapevine selection methodology is strongly supported by quantitative genetics and all the issues related to the selection process are answered through the application of the theory of mixed models, being a crucial point the proper choice of the structure of the covariance matrices (Gonçalves *et al.*, 2007, Gonçalves *et al.*, 2016).

To understand the application of mixed models in this biologic context, the initial stage of grapevine selection can be analyzed as an example. The initial field trial is established with a random sample of genotypes of the target ancient variety. The size of this sample is usually in the hundreds, therefore, the field trial contains a large numbers of genotypes and ideally is established according to an experimental design belonging to the class of incomplete blocks (alpha and row-column designs) as alternative to randomized complete block design (Gonçalves *et al.*, 2010). A mixed model is fitted to the data of this trial, with genotypic effects as random effects and experimental design effects as fixed and/or random, depending on the experimental design. The objective is to quantify the intra-varietal genetic variability of the most important traits because it is the raw material for selection (the greater the intra-varietal genetic variability, the greater the gains of selection). On the other hand, knowledge on variability and its geographical distribution is an essential basis for all approaches to ancient varieties: genetic conservation, understanding of their origin and expansion, etc.. Additionally, the estimates of genetic variance components, as well as all other covariance parameters estimates, are essential for selection. In fact, the empirical best linear unbiased predictors (EBLUPs) of the genotypic effects for the several traits under study, on which selection is performed, directly depend on the structure of the covariance matrices.

In this work is exemplified the application of linear mixed models with adequate covariance structures to be used in the first stage of grapevine selection. The study is focused in the quantification of intra-varietal genetic variability of economically important traits in the several growing regions of an ancient variety and in the predicted genetic gains obtained from the selection of a superior group of genotypes.

Keywords: mixed models, variance components, genetic variability, grapevine selection.

References

[1]   Gonçalves, E., St. Aubyn, A. and Martins, A. (2007) Mixed spatial models for data analysis of yield on large grapevine selection field trials.*Theoretical and Applied Genetics* **115**: 653–663.

[2]   Gonçalves, E., St. Aubyn, A. and Martins, A. (2010) Experimental designs for eva-
      luation of genetic variability and selection of ancient grapevine varieties: a simulation
      study. *Heredity* **104**: 552–562.

[3]   Gonçalves, E., Carrasquinho, I., Almeida, R., Pedroso, V. and Martins, A. (2016)
      Genetic correlations in grapevine and their effects on selection. *Australian Journal of
      Grape and Wine Research* **22**: 52–63.

Spatio-temporal structure of ecological assemblages: a comparative study between STATICO and Canonical Correspondence Analysis

<u>Susana Mendes</u>[1,2], M. José Fernández-Gómez[3], Ulisses Miranda Azeiteiro[4] and M. Purificación Galindo-Villardón[3]

[1]Instituto Politécnico de Leiria, ESTM, Portugal
[2]MARE–Marine and Environmental Sciences Centre
[3]University of Salamanca, Department of Statistics, Spain
[4]Department of Sciences and Technology, University Aberta and University of Coimbra, CEF-Department of Life Sciences, Portugal

E-mail addresses: *susana.mendes@ipleiria.pt; mjfg@usal.es; ulisses@uab.pt; pgalindo@usal.es*

Abstract

Environmental and biotic data sets can be multidimensional and have a complex structure. Usually, they are collected as sets (tables) of objects and variables obtained under different experimental circumstances or for various sampling periods, etc. Putting all tables together results in data with three-way structure. An example for such data is when in samples collected at different sampling sites, the species abundances/densities and the values of several physical-chemical components are measured during a period of time (sites x species/parameters x time). There are many tools helping to explore and interpret multitable structure of the data. The present work deals with the data exploration of three-way species and environmental data with the use of two multitable methods: (1) Canonical Correspondence Analysis (CCA, ter Braak, 1986); and (2) STATICO, an application of STATIS to co-inertia operators (Simier et al, 1999; Thioulouse et al, 2004). While CCA is the most popular one as also is the classic technique to understand which variables are influencing the composition of species, STATICO appeared as a very attractive three-way exploratory tool due to their robust properties and ability to deal with sets of two-way tables (matrices), that do not have the same dimension for columns or rows. Here, the performances of them are compared and the features are demonstrated on real data sets. More than that, the differences and the similarities between STATICO and CCA are emphasized as well as their advantages and/or disadvantages. The example data set is from January 2002 to June 2003 in Canal de Mira, Ria de Aveiro, Western Portugal and was extracted from the study published by Paula Resende, Ulisses Azeiteiro and Mário Jorge Pereira (2005). The species data consist in the abundance of 231 diatoms in three sites, measured at two tide conditions, and the environmental data consist of 7 variables measured in the same sites and at the same dates. So, in this case the sites are the same at all dates. The relationship between diatom assemblage and the environmental parameters governing their composition and structure was focused. The parameters and diatom species that stood out in the CCA biplot as well as the results of the interstructure, compromise, and trajectories steps of the STATICO method were analyzed. Analyzing data it is important to choose an appropriate structural basis and appropriate constraints. A poor choice of either structure or constraints can grossly impact the results of the analysis. A main issue here is the problem of choosing structure and constraints based on a priori knowledge, exploratory analysis of the data, and the goal of the analysis. Besides this, it is interesting to develop models and algorithms according to the nature of the

data, instead of trying to adjust the data to the nature of the model. Finally, this or other comparisons between various methods can be produce, but all should be based on two types of considerations: the mathematical properties of the methods, and the biological aims of the study.

Keywords: CCA, estuaries, multitable analysis, STATICO, STATIS.

References

[1]   Resende, P., Azeiteiro, U. M. and Pereira, M. J (2005) Diatom ecological preferences in a shallow temperate estuary (Ria de Aveiro, Western Portugal. *Hydrobiologia* **544**: 77–88.

[2]   Simier, M., Blanc, L., Pellegrin, F. and Nandris, D. (1999) Approche simultanée de K douples de tableaux: application à l'étude des relations pathologie végétale–environnement. *Rev. Statist. Appl.* **47**: 31–46.

[3]   ter Braak, C. J. F. (1986) Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis, *Ecology* **67**: 1167–1179.

[4]   Thioulouse, J., Simier, M. and Chessel, D. (2004) Simultaneous analysis of a sequence of saired ecological tables. *Ecology* **85**: 272–283.

Pulse Wave Velocity: statistical characterization of Early Vascular Aging

P. Oliveira[1,2] and P. G. Cunha[3,4]

[1]University of Porto
[2]ICBAS/EPIUnit
[3]Centro Hospitalar do Alto Ave, University of Minho
[4]IICVS

E-mail addresses: *pnoliveira@icbas.up.pt; pedrocunha@ecsaude.uminho.pt*

Abstract

Portuguese population in Northen Portugal presents high prevalence of hypertension and stroke incidence. This work presents a cohort study to identify and characterize arterial damage and premature arterial aging, based on Pulse Wave Velocity measurements, in the population of two adjacent cities, Guimarães and Vizela, in Northern Portugal.

Keywords: multiple regression, logistic regression, pulse wave velocity.

——⊙——

Portuguese population exhibits a high incidence of stroke that might be associated with a high salt consumption, hypertension and other cardiovascular risks. Early vascular aging (EVA) corresponds to an accelerated aging of the blood vessels, with arteries displaying characteristics typically observed at older (chronological) ages. Pulse wave velocity (PWV) can be very useful in understanding the effects of accumulated cardiovascular risks in the process of vascular aging. In order to identify and charcaterize the distribution of PWV and EVA, a randomly sample dwellers in two cities, aged 18–96 years from Northern Portugal, was selected with 2542 individuals completing the evaluation.

Subjects were classified as EVA if the corresponding PWV values were above the 97.5 percentile, adjusted for age according to the European reference values. Individuals exhibiting PWV values above 10m/s were claasified as having arterial stiffness.

Linear regression models were studied with PWV as the dependent variable and with independent variables that included age, age2, sex, Systolic Blood Pressure (SBP) (mean of four measurements), heart rate (HR), BMI, years of education, tobacco use, family history of premature Cardio Vascular Disease (CVD), antihypertensive medication use, fasting glucose, lipid profile, mean estimated glomerular filtration rate, C-reactive protein, antidiabetic treatment, antilipidic treatment, and known CVD. Similarly, logistic regression models were constructed to study contributing variables to both the development of large artery damage (PWV above 10 m/s) and EVA.

Mean PWV was 8.4 m/s (men: 8.6 m/s, women: 8.2 m/s). PWV prevalence was 12.5% (26.1% with less than 30 years; 18.7% of the population exhibited PWV values above 10 m/s). Differences were observed in the odds ratio regrading sex, with women reaching PWV values above 10 m/s approximately ten years later when compared to men.

The distribution of PWV by sex and age is, for the first time in Portugal, realized in a population based study. PWV values are higher than expected according to the European reference values, particularly amongst younguer individuals of the male sex. In accordance to other studies, regression models have shown the association of age and Systolic Blood Pressure with PWV. These results have great relevance for the prevention of caridovascular

disease, namely regarding the cardiovascuklar risk of the population and, therefore, the need for urgent clinic and public health interventions.

References

[1]   Cunha, P. G., Cotter, J., Oliveira, P., Vila, I. and Sousa, N. (2014) The Rationale/Design of the Guimarães/Vizela Study: a multimodal population-based cohort study to determine global cardiovascular risk and disease. *Journal of Investigative Medicine* **62**: 812–8290.

[2]   Cunha, P. G., Cotter, J., Oliveira, P., Vila, I., Boutouyrie, P., Laurent, L., Nilsson, P. N., Scuteri, A. and Sousa, N. (2014) Pulse wave velocity distribution in a cohort study: from arterial stiffness to early vascular aging. *Journal of Hypertension* **33(7)**: 1438–45.

[3]   The Reference Values for Arterial Stiffness' Collaboration (2010) Determinants of pulse wave velocity in healthy people and in the presence of cardiovascular risk factors: "establishing normal and reference values". *European Heart Journal* **31**: 2338–2350.

Organized Session 14

Applications in Statistics

Organizer: Manuela Oliveira (Portugal)

Focusing on the Inference

<u>Cristina Dias</u>[1,2], Carla Santos[3,2], Manuela Oliveira[5,6] and
João T. Mexia[7,2]

[1]Escola Superior de Tecnologia e Gestão, Instituto Politécnico de Portalegre
[2]Centro de Matemática e Aplicações (CMA)
[3]Escola Superior de Tecnologia, Instituto Politécnico de Beja
[5]Escola de Ciências e Tecnologia, Universidade de Évora
[6]CIMA-Centro de Matemática e Aplicações, Universidade de Évora
[7]Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa

E-mail addresses: *cpsilvadias@gmail.com; carla.santos@ipbeja.pt; mmo@uevora.pt;
jtm@fct.unl.pt*

Abstract

We developed two aspects of inference on families structured symmetric stochastic matrices ([4]). Initially we simplify the models that are apply to these matrices and in the second phase we estimate the relative relevances of the various hypotheses that can be tested.

Keywords: stochastic matrices, inference, relative relevancies.

———⊙———
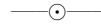
The models are developed using the spectral analysis of the matrices $\boldsymbol{\mu}$ ([3]). The adjustment and validation of the model requires the usage of the vector $\tilde{\boldsymbol{\beta}}_i = \theta_i \boldsymbol{\gamma}_i$, $i = 1, \ldots, k$, which an estimator of the structure vector $\boldsymbol{\beta}_i = \lambda_i \boldsymbol{\alpha}_i$, $i = 1, \ldots, k$ of $\mathbf{M}$ ([4]). With $\lambda_1, \ldots, \lambda_k$ the non-null eingenvalues and $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_k$ are the eigenvectors, of the matrix $\mathbf{M}$. As the degree (k) of the model is given by the mean matrix characteristic, and thus the model study involves all k eigenvalues. For the models with a degree $k > 1$, we still consider the possibility of truncating the model, when there are eigenvalues, $\theta_1, \ldots, \theta_k$ much greater than the other myth. In this case we can considering the interests on their relative relevance, which can be measured by

$$\nu_j = \frac{\delta_j}{\sum\limits_{l=1}^{h} \delta_l} = \frac{\boldsymbol{\beta}_j^t \mathbf{W}_j^{-1} \boldsymbol{\beta}_j}{\sum\limits_{l=1}^{h} \boldsymbol{\beta}_l^t \mathbf{W}_j^{-1} \boldsymbol{\beta}_l} \quad , j = 1, \ldots, h$$

Our main result is that, whenever $\theta_1 > \cdots > \theta_h >> \theta_h + 1 > \cdots > \theta_k$ that is, when θ_h is much greater than $\theta_h + 1$, obtains a truncated model with structure vector set given

$$\tilde{\boldsymbol{\beta}}^0 = [\tilde{\boldsymbol{\beta}}_1^t, \ldots, \tilde{\boldsymbol{\beta}}_h^t]^t$$

which simplifies the calculations without great loss of accuracy of the results.

Acknowledgements

References

[1]   Areia, A. M., Oliveira, M. and Mexia, J. T. (2008) Models for a series of studies based on geometrical representation. *Statistical Methodology* **5(3)**: 277–88.

[2]   Dias, C. (2013) *Modelos e Famílias de Modelos para Matrizes Estocásticas Simétricas*, Tese de Doutoramento, Universidade de Evora.

[3]   Silvey, SD. (1975) *Statistical inference*, New York, Chapman and Hall.

Analysis and prediction of heat wave over India during 2015

<u>Venkata B. Dodla</u>[1], C. S. Gubbala[1], S. Desamsetti[1] and
José Carlos Tiago Oliveira[2]

[1]K L University, India
[2]University of Évora

E-mail address: *dvbrao@kluniversity.in*

Abstract

Heat waves are recognized as a severe natural disaster globally due to its societal impact not only causing loss of human lives but also effecting crop failures, accentuating wild fires and impacting industrial and social activities. The term "Heat wave" refer to abnormal heat spells, and is generally defined relative to normal temperatures (Meehl and Tebaldi, 2004) such as "persistent temperature anomalies". Human losses due to heat waves were about 70,000 in Europe during 2003 (Robine et al.(2008)); 11,000 in Russia during 2010 (http://www.wwfblogs.org/climate/content/great-russian-heat-wave-2010-caused-11000-deaths-moscow-alone); 1,718 in Japan during 2010 (Matsuyama and Sato(2011)); about 500 in US (Chicago) in 1995 (Dematte, Jane E. et al. (1998)) which indicate the severity of heat waves as a disastrous event. Heat wave of 2015 over India caused about 2500 human deaths evoking concern among the public and the government. This heat wave has been studied to understand the characteristics, associated atmospheric circulation and evaluate its predictability. Although temperatures are highest in May over India, occurrence of heat wave conditions over southeast coastal parts of India was unanticipated. Surface temperature data from Global analysis fields and station observations have been analyzed to assess the characteristics of the heat wave. It has been inferred that isolated region of Andhra Pradesh (AP) had experienced the heat wave conditions during 23-27 May, 2015, with temperatures escalating by 7-10 C within a short span of 2-3 days leading to human deaths. Short range weather predictions with lead time of 72-hours have been made using ARW (Advanced Research Weather Research and Forecasting) model at 3-km resolution. Predictions up to 72-hours lead time have been accurate with statistical metrics of mean absolute error, root mean square error to be very less and with high index of agreement confirming the predictability of the heat wave evolution. Atmospheric patterns over Eurasia have indicated that regional changes of atmospheric pressure within the Eurasia region leading to increased pressure gradient between Middle East and India have been responsible for increased northwest wind flow over to northwest India and to southeast India which have advected higher temperatures. Estimates of warm air advection have shown heat accumulation over AP region, due to sea breeze effect. The study brought out the reasons are to be the changing pressure gradients between Middle East and India, enhancement of northwest wind flow with warm air advection and sea breeze effect along southeast coast blocking the free flow all leading to the occurrence of heat wave conditions over coastal Andhra Pradesh.

Keywords: heat wave, ARW model, temperature, atmospheric circulations.

References

[1]   Meehl, G. A. and Tebaldi, C. (2004) More intense, more frequent and longer lasting heat waves in the 21-st century, *Science* **305**: 994–997.

[2]   Robine, J. M., Cheung, S. L. K., Le Roy, S., Van Oyen, H., Griffiths, C., Michel, J. P. and Herrmann, F. R. (2008) Death toll exceeded 70,000 in Europe during the summer of 2003. *Comptes Rendus Biologies* **331(2)**: 171–178.

[3]   Matsuyama, K. and Sato, S. (2011) Heatstroke Deaths Quadruple as Japan Shuns Air Conditioners to Save Power. Bloomberg L. P. Retrieved 1 June 2012.

[4]   Dematte, Jane E., et al. (1998) Near-Fatal Heat Stroke during the 1995 Heat Wave in Chicago. *Annals of Internal Medicine* **129(3)**: 173–181. doi: 10.1059/0003-4819-129-3-199808010-00001. PMID 9696724.

[5]   Sherwood, S. C. and Matthew Huber (2010) An adaptability limit to climate change due to heat stress. *PNAS* **107(21)**: 9552–9555.

Effect of growth medium, temperature and incubation time on *Chlamydomonas reinhardtii*

Vinod Kumar Patil[1,2], Manuela Oliveira[1,3,4] and
Birgit Arnholdt Schmitt[1,2]

[1]University of Évora
[2]Marie Curie Chair, ICAAM-Institute of Agricultural Science and Environmental
Mediterranean
[3]Department of Mathematics
[4]CIMA-Centro de Matemática e Aplicações, Universidade de Évora

E-mail addresses: *eu_chair@uevora.pt; mmo@uevora.pt*

Abstract

Synchronized cultures of Chlamydomonas reinhardtii were grown phototrophically under a wide a range of environmental conditions including temperature (18, 24 and 30° C) and nitrogen (0.5 and 1.0 %) and check the differences of their cell cycle, biomass growth and lipid content at different time periods (5, 10, 15 and 20 days). Under standard growth conditions, algae biomass is composed primarily of proteins, cell wall carbohydrates and membrane lipids. Since specific growth rate (SGR) is the single most criteria to decide the biomass yield of microalgae, for biomass growth and lipid production which is accumulated as one of the reserve compound under directed stress conditions such as N_2 deprivation. Improvement of algae strain performances can be achieved for increasing biomass production or synthesis of other specific compounds.

——⊙——
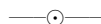
Under stress condition, microalgae undergoes variety of changes, eventually alter the lipid biosynthesis pathway and accumulate more amount of intracellular lipid pool, accumulated lipid pool during stress constitutes triacylglycerides (TAGs) as dominant form of stress responsive energy reserve. Variety of nutritional and environmental causes has been shown to influence the lipid content and fatty acid profile in algae. Among nutritional stress, nitrogen starvation is a prominent one to enhance the lipid level in majority of algae. The duration of the life cycle varies considerably depending on temperature, nutrients and luminance. Growth of microalgae is likely to be affected by culture conditions such as temperature, salinity, illumination and/or nutrients. Temperature is one of the important stress which affects growth through respiration and calorespirometry allows measuring metabolic heat rates and CO_2 emission of respiring cells/tissues, across different temperatures. Thus, identification of optimal temperatures for biomass growth and lipid synthesis becomes predictable. Temperature is an important factor for the growth of algae, it strongly influences cellular chemical composition, uptake of nutrients, CO_2 fixation and growth rate. Up to 40-fold reduction in photosynthetic proteins level were observed in green alga Chlamydomonas reinhardtii, thus it is important to study in detail about the growth level under different stress levels. Ten percent of optimally grown culture was inoculated in 300 ml of Tris Acetate Phsophate (TAP) Medium and kept under $25^o \pm 2^oC$ with a photoperiod of 14:10 h light-dark cycle for 20 days. At every 5 days intervals, samples were withdrawn aseptically and the growth was measured spectrophotometrically at 750 nm. In order to determine the dry biomass yield, 50 ml samples were

filtered on to pre-dried and weighed GF/C fiber filters every 5th day of experimental period. The filtered biomasses were oven dried overnight at $105^{o}C$ along with filter paper, and reweighed using an analytical balance. The dry biomass of Chlamydomonas under experimental condition was expressed g/L. The specific growth rate was measured number of generation (the number of doublings) that occurs per unit of time in an exponentially growing culture. The phase of growth was carefully determined and specific growth rate was obtained using the following equation: $l = ln(N_t/N_0)/T_t t_0$; $N_t =$ no of cells at the end of the log phase; $N_0 =$ no of cells at the start of log phase; $T_t =$ final day of log phase; $T_0 =$ starting day of log phase. If T expressed in days from the growth rate (l) can be converted to division or doubling per day (K) by dividing (l) by the natural log of 2 (0.6931). The cell count by haemocytometer, dry weight and lipid content by soxhlet method were measured at different times. All the data was read in triplicates and statistically analyzed by SPSS. This will help to select the strain by their growth conditions, biomass and compound production rate under stress conditions.

Acknowledgements

Models with orthogonal block structure, commutativity and B-matrices

<u>Carla Santos</u>[1,5], Célia Nunes[2] , Cristina Dias[3,5] and
João Tiago Mexia[4,5]

[1]Polytechnical Institute of Beja
[2]Department of Mathematics and Center of Mathematics University of Beira Interior
[3]Polytechnical Institute of Portalegre
[4]Department of Mathematics, Faculty of Science and Technology, New University of
Lisbon
[5]Center of Mathematics and its Applications, New University of Lisbon

E-mail addresses: *carla.santos@ipbeja.pt; celian@ubi.pt; cpsilvadias@gmail.com;
jtm@fct.unl.pt*

Abstract

A model with orthogonal block structure, OBS, is a linear mixed model whose variance-covariance matrix, V, is a linear combination of known pairwise orthogonal orthogonal projection matrices, $\mathbf{Q}_i$, $i = 1, \ldots, w$, that add up to the identity matrix.

Models with commutative orthogonal block structure, COBS, are a special case of OBS in which the orthogonal projection matrix on the space spanned by the mean vector, $\mathbf{T}$, commutes with the $\mathbf{Q}_1, \ldots, \mathbf{Q}_w$. This condition ensures that for estimable vectors the least squares estimators, LSE, are the Best linear unbiased estimator, BLUE, whatever the variance components. Resorting to B-matrices (see [8]) we present a general condition, for $\mathbf{T}$ to commute with $\mathbf{V}$, which is an equivalent condition to the first commutativity condition we here refer to.

Keywords: models with commutative orthogonal block structure, Jordan algebra, estimable vectors, B-matrices.

Acknowledgements

References

[1]   Carvalho, F., Mexia, J. T., Santos, C. and Nunes, C. (2014) Inference for types and structured families of commutative orthogonal block structures. *Metrika* **78**: 337–372, doi: `10.1007/s00184-014-0506-8`.

[2]   Houtman, A. M. and Speed, T. P. (1983). Balance in designed experiments with orthogonal block structure. *Ann. Statist* **11(4)**: 1069-1983.

[3]   Jordan, P., von Neumann, J. and Wigner, E. (1934). On the algebraic generalization of the quantum mechanical formalism. *Annals of Mathematics* **36**: 26–64.

[4]   Nelder, J. A. (1965a) The analysis of randomized experiments with orthogonal block structure I. Block structure and the null analysis of variance. *Proceedings of the Royal Society, Series A*, **283**: 147162.

[5] Nelder, J. A. (1965b) The analysis of randomized experiments with orthogonal block structure II. Treatment structure and the general analysis of variance. *Proceedings of the Royal Society*, Series A, **283**: 163178.

[6] Nunes, C., Santos, C. and Mexia, J. T. (2008) Relevant statistics for models with commutative orthogonal block structure and unbiased estimator for variance components. *Journal of Interdisciplinary Mathematics* **11(4)**: 553–564, ISSN 0972–0502, doi: `10.1080/09720502.2008.10700581`.

[7] Santos, C., Nunes, C. and Mexia, J. T. (2007) OBS, COBS and Mixed Models associated to commutative Jordan Algebra. *In proceedings of 56th session of the International Statistical Institute.* Lisboa, Portugal, August 22–29.

[8] Santos, C. (2012) Error orthogonal models. Structure, Operations and Inference. PhD Thesis, Universidade da Beira Interior.

[9] Seely, J. (1971) Quadratic subspaces and completeness. *Ann. Math. Stat.* **42**: 710–721.

[10] Seely, J. e Zyskind, G. (1971). Linear spaces and minimum variance estimators. *The Annals of Mathematical Statistics* **42(2)**.

Determinants behavior analysis of demand for passengers to the Portuguese air travel leisure market: Application with Partial Least Squares Regression (PLS)

<u>José Manuel Vicente</u>[1,2], Andreia Dionísio[1,2] and Manuela Oliveira[1,3]

[1]University of Évora
[2]DREAMS-Universidade Lusófona de Humanidades e Tecnologias
[3]CIMA-Centro de Matemática e Aplicações, Universidade de Évora

E-mail addresses: *jose.manuel.vicente81@gmail.com; andreia@uevora.pt; mmo@uevora.pt*

Abstract

This article aims to determining the behavior analysis of the demand for passengers to the Portuguese air travel leisure market (between Lisbon and Oporto and the Azores and Madeira), between 2010 and 2014, using an application of Partial Least Squares regression (PLSR). The PLS regression analysis used in the essay was developed by Wold [2] to allow the modeling of complex data sets in terms of matrices chains [4]. The PLS regression takes the two description of goals and forecasting, combining principal component analysis and multiple linear regression analysis. In the first section we present a contextualization of the PLSR and applicability to air transport. In a second section we present the data and the model. In a third section we present the discussion of results. Finally, we present the final considerations. The results of our empirical analysis suggest that the demand for air transport of passengers between Portugal and the Azores and Madeira is more affected by factors of destination attraction's (leisure and events, sun and beach), by aviation industry factors (airfare and presence of LCC's on routes) and geographical variable (distance). The results were promising and provide important information to improve business decision making, with regard to the operation of business level policies, but also the airport business development policies.

Keywords: partial least squares regression, data panel, modelling, air transport, Portugal.

Acknowledgements

References

[1]   Tenenhaus, M., Esposito Vinzi, V., Chatelin, Y. and Lauro, C. (2005) PLS path modeling. *Computational Statistical and Data Analysis* **48(1)**: 159–205.

[2]   Wold, H. (1985) *Partial Least Squares. In Samuel Kotz and Norman L. Johnson.* Encyclopedia of Statistical Sciences, **6**, New York: Wiley, 581–591.

Posters included in an Organized Session

Principal Component Analysis with Physical Fitness Variables

Poster included in the Organized Session 14

Domingos J. Lopes da Silva[1,2] and Manuela Oliveira[3,4]

[1]Institute of Higher Studies of Fafe
[2]Fafe Education School
[3]Department of Mathematics, University of Évora
[4]CIMA-Centro de Matemática e Aplicações, Universidade de Évora

E-mail addresses: *domingosjlsilva@gmail.com; mmo@uevora.pt*

Abstract

Principal Component Analysis (PCA) is a useful tool to reduce data dimensionality. Using physical fitness tests (sit-and-reach, lower-limbs explosive power, sit-up's, and cardiorespiratory fitness), a PCA was conducted in 11th grade students. It was performed a descriptive and correlation analysis between variables. The adequacy of PCA was measured by KMO and chi-square tests. It was found that the proportion of variance of each variable, explained by principal components (communalities), was greater than 80%, which means that it is appropriate to describe the latent correlation structure between physical fitness tests. The eigenvalues and the proportion of variance explained by each component showed that the PCA followed by Varimax rotation gave the retention of two components; explaining about 87% of the total variability. Regarding component matrix, both without and with rotation of the components, it was recorded high and balanced factorial loads ($>88\%$). The factorial loads were analyzed by component transformation matrix. The solution was presented after Varimax rotation, leading to a better definition of the components. With Kaiser normalization criterion we observe the scores of each variable in the components. All variables saturate in only one component, meaning that these tests can be explained in a single component. The component one had a close relationship with strength-resistance; the component two were more associated with flexibility.

Keywords: principal component analysis, physical fitness, secondary education.

Acknowledgements

A multilevel analysis of student's achievement in mathematics

<u>Susana Faria</u>[1,2] and João Silva[1]

Poster included in the Organized Session 10

[1]University of Minho
[2]Centre of Mathematics, University of Minho

E-mail addresses: *sfaria@math.uminho.pt; joaosilva17@sapo.pt*

Abstract

In this work, we pretend to analyse the students academic achievement in several countries, with the purpose of understand which are the factors that can influence the students achievement in Mathematic literacy. For our analysis we use data from the Programme for International Student Assessment (PISA) 2012, collected by the Organisation for Economic Co-operation and Development (OECD) in a large number of countries. Given the hierarchical structure of data, the models adopted for statistical analysis were multilevel models, which can take into account data variability within and among the hierarchical levels.

Keywords: multilevel regression models, hierarchical structure, programme for International Student Assessment (PISA), students achievement in Math.

———⊙———
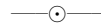
Multilevel regression models are regression models which usually apply in situations where data are hierarchically structured. For that reason, these models are very important in the analysis of studies related with education because the population of these type of study are found structured in a hierarchical way. An example of the hierarchical structure in educational data is the students grouped in classes, the classes grouped in schools, the schools grouped in regions, etc.

This work presents the rationale for multilevel regression models, by applying them to a case study in education area. In particular, this models were applied to data collected under the Programme for International Student Assessment (PISA) 2012 in Mathematics literacy in several countries. With this application, we pretend to analyse the students academic achievement in several countries, with the purpose of understand which are the factors that can influence the students achievement in Mathematic literacy.

Applying a multilevel regression model with three levels concluded that the fact of the student were male, the index of economic, social and cultural, the proportion of computers connected to the Internet, the proportion of girls in school, the student behavior, the fact of the students attend private schools independent from government or the school autonomy in the allocation of school resources positively the students achievement in Math. On the other hand, the fact that the students are immigrants or have already repeated a school year, the ratio between the number of students and the number of math teachers and the number of computers for educational purposes per student has a negative influence on student achievement in Math.

Acknowledgements

References

[1]   Fuchs, T. and Woessmann, L. (2007) What accounts for international differences in student performance? A reexamination using PISA data. *Empirical Economics* **32**: 433–464.
[2]   Goldstein, H. (2011) *Multilevel Statistical Models*, John Wiley and Sons.
[3]   OECD (2015) *Students, Computers and Learning: Making the Connection*, PISA, OECD Publishing.
[4]   Raudenbush, S. W. and Bryk, A. S. (2002) *Hierarchical Linear Models* (Second Edition), Thousand Oaks, Sage Publications.

Structural Time Series Models Applied to a Water Quality Monitoring Problem

A. Manuela Gonçalves[1], Olexandr Baturin[2] and Marco Costa[3]

Poster included in the Organized Session 10

[1]CMAT-Centre of Mathematics, DMA-Department of Mathematics and Applications[1],
University of Minho, Portugal
[2]DMA-Department of Mathematics and Applications, University of Minho, Portugal
[3]CIDMA-Centre for Research and Development in Mathematics and Applications,
University of Aveiro, Portugal

E-mail addresses: *mneves@math.uminho.pt; olexandr.baturin@gmail.com; marco@ua.pt*

Abstract

State space models provide a very flexible tool for analyzing dynamic phenomena and evolving systems, and have significantly contributed to extending the classical domains of application of statistical time series analysis to non-stationary, irregular processes, etc. In the context of a water quality monitoring problem we applied structural time series models, the state space models associated to the Kalman filter, in the modeling process of monthly measurements of physico-chemical and microbiological variables in a network of water quality monitoring sites in a river basin. The data concerns the River Ave hydrological basin located in the Northwest of Portugal, where monitoring has become a priority in water quality planning and management because its water has been in a state of obvious environmental degradation for many years. As a result, the watershed is now monitored by eight sampling stations distributed along the River Ave and its main streams. For the modeling process we consider data between January 1999 to January 2014. The framework of the state space models shows versatility to incorporate unobserved components, such as trends, cycles and seasonals, that have a natural interpretation and represent the salient features of the series under investigation: environmental time series. From the environmental point of view, the approach proposed allows to obtain pertinent findings concerning water surface quality interpretation and of change point of view, thus highlighting the potential value of this type of analysis, and it is relevant in order to identify unanticipated changes that are important in the management process and for the assessment of water quality.

Keywords: water quality monitoring, time series, structural components, state space modeling, Kalman filter.

References

[1] Gonçalves, A. M. and Costa, M. (2011) Application of Change-Point Detection to a Structural Component of Water Quality Variables. In Simos, T.E., Psihoyios, G., Tsitouras, Ch. (eds.): *AIP Conference Proceedings* **1389**: 1565–1568, AIP American Institute of Physics Publishing.

[2] Gonçalves, A. M. and Costa, M. (2013) Predicting seasonal and hydro-meteorological impact in environmental variables modelling via Kalman filtering. *Stochastic Environmental Research and Risk assessment* **127(5)**: 1021–1038.

[3] Harvey, A. C. (1996) *Forecasting, structural time series models and Kalman filter*, Cambridge: Cambridge University Press.

[4] Petris, G., Petrone, S. and Campagnoli, P. (2009) *Dynamic Linear Models with R*, Springer, New York.

[5] R Team, D. (2015) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing.

Contributed Poster

Estimating the extremal coefficient: a comparison of methods

Marta Ferreira[1,2]

[1]CMAT - University of Minho
[2]CEMAT, CEAUL - University of Lisbon

E-mail address: *msferreira@math.uminho.pt*

Abstract

Tail dependence is an important issue to evaluate risk. The multivariate extreme values theory is the most suitable to deal with the extremal dependence. The extremal coefficient measures the degree of dependence between the marginals of max-stable distributions, a natural class of models in this framework. The estimation of the extremal coefficient is addressed and a new estimator is compared through simulation with existing methods. An illustration with real data is presented.

Keywords: multivariate extreme values, tail dependence, max-stable models.

Acknowledgements

A multivariate Birnbaum-Saunders model and its diagnostics and application to biomedicine

<u>Carolina Marchant</u>[1], Víctor Leiva[2], Francisco J. A. Cysneiros[1] and Juan F. Vivanco[2]

[1]Universidade Federal de Pernambuco, Recife, Brazil
[2]Universidad Adolfo Ibáñez, Viña del Mar, Chile

E-mail addresses: *cimf1@de.ufpe.br; victorleivasanchez@gmail.com; cysneiros@de.ufpe.br; juan.vivanco@uai.cl*

Abstract

Birnbaum-Saunders (BS) models are receiving considerable attention in the literature. Multivariate regression models are a useful tool of the multivariate analysis. Diagnostic analysis is an important aspect to be considered in the statistical modeling. In this work, we formulate multivariate generalized BS regression models and carry out a diagnostic analysis for these models. We consider the Mahalanobis distance (MD) as a global influence measure to detect multivariate outliers and use it for evaluating the adequacy of the distributional assumption. We also consider the local influence approach and study how a perturbation may impact on the estimation of model parameters. We illustrate the obtained results with the R software and real-world data to show their potential applications.

Keywords: Birnbaum-Saunders distributions, global and local influence, goodness-of-fit, multivariate data analysis, R software.

——⊙——

The BS distribution originates from the cumulative damage law related to fatigue and strength of materials. The parameter estimates of BS distribution, obtained from the ML method, are sensitive to atypical observations. In order to attenuate this sensitivity, one can obtain a BS distribution based on the Student-t distribution. BS and BS-t distributions are members of a wider family of distributions generated from elliptically contoured (EC) distributions, known as generalized BS (GBS) distributions. Marchant et al (2016), [1], developed multivariate logarithmic GBS (log-GBS) distributions, which are generated from EC kernels. Similarly to the univariate case, the ML estimation procedure based on the multivariate log-BS-t distribution is non-sensitive to atypical observations because it attribute smaller weights to these observations. Multivariate regressions are a useful tool of the multivariate analysis. The main advantage of a multivariate regression model over marginal regressions is that it takes into account the correlation between the response variables. In our case, the multivariate BS log-linear regression model is formulated as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{E}, \tag{1}$$

where $\boldsymbol{Y} = (Y_{ij}) \in \mathbb{R}^{n \times m}$ is the log-response matrix, and $\boldsymbol{X} = (x_{is}) \in \mathbb{R}^{n \times p}$ the design matrix of rank p containing the values of p covariates. Note that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are connected by the regression coefficient matrix $\boldsymbol{\beta} = (\beta_{sj}) = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m) \in \mathbb{R}^{p \times m}$ to be estimated, and $\boldsymbol{E} = (\varepsilon_{ij}) \in \mathbb{R}^{n \times m}$ is the matrix of random errors. In the model given in (1), let $\boldsymbol{Y}_i^\top$, $\boldsymbol{x}_i^\top$ and $\boldsymbol{\varepsilon}_i^\top$ be the ith rows of $\boldsymbol{Y}$, $\boldsymbol{X}$ and $\boldsymbol{E}$, respectively. Then, we can write $\boldsymbol{Y}_i = \boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_i = \boldsymbol{\beta}^\top \boldsymbol{x}_i + \boldsymbol{\varepsilon}_i$, $i = 1, \ldots, n$, where $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n$ are independently and identically

log-GBS$_m(\alpha\mathbf{1}_{m\times1}, \mathbf{0}_{m\times1}, \boldsymbol{\Psi}, g^{(m)})$ distributed, with $\boldsymbol{\Psi} = (\Psi_{rs}) \in \mathbb{R}^{m\times m}$ the correlation matrix, and $g^{(m)}$ the multivariate EC kernel.

The main objective of this work is to carry out diagnostics in multivariate GBS log-linear regression models by (i) GOF techniques; (ii) global and local influence methods; and (iii) applying them to real-world data. To meet these objectives, we estimate the parameters of multivariate BS and BS-t models with the ML method. Then, we derive fitting and diagnostic tools to evaluate the adequacy and stability of these multivariate models. First, we use the MD to test goodness of fit in the distributional assumption. Second, we consider the MD as a global influence method to identify multivariate outliers. Third, we derive the total local influence method assuming perturbations of (i) case-weight, (ii) correlation matrix, (iii) response, and (iv) a continuous explanatory variable. We implement the obtained results in the R statistical software. We carry out an illustration with real-world multivariate data collected by the authors. Specifically, we consider multivariate data useful for regression models based on computed tomography (CT) to study the bone quality. Particularly, we investigate four types of densities: (i) bulk density, which considers the mass of the intact core, including fat and water; (ii) dry density, which excludes fat and water; (iii) mineral density, which is related to the mineral content; and (iv) clinical CT, which is obtained from a calibration equation that is derived from known mineral content phantoms. Note that the BS distribution is appropriate for modeling physical properties of bone, such as its densities extracted from CT data, which are related to mechanical properties, as strength, which plays a role as human bone ages affecting its fatigue properties that can be described by the BS distribution.

Acknowledgements

References

[1] Marchant, C., Leiva, V. and Cysneiros, F. (2016) A multivariate log-linear model for Birnbaum-Saunders distributions. *IEEE Transactions on Reliability*, in press.

Environmental awareness and education: A higher education-based study

Célio Gonçalo Marques[1,2,3] and Vasco Gestosa da Silva[1]

[1]Unidade Departamental de Tecnologias de Informação e Comunicação, Instituto Politécnico de Tomar
[2]Centro de Administração e Políticas Públicas, Universidade de Lisboa
[3]Laboratório de Tecnologia Educativa, Universidade de Coimbra

E-mail addresses: *celiomarques@ipt.pt; vasco.silva@ipt.pt*

Abstract

The use of electrical and electronic devices has grown exponentially in the last decade thanks to technological advances and its wide dissemination. On the other hand, the number of obsolete equipment has also increased significantly, becoming a major problem for society. This communication deals with electronic waste and explains the concept and risks for the particular case of Portugal. A survey is presented which was carried out in two Portuguese Higher Education institutions with a view to determining students' awareness, attitudes and opinions on this matter.

Keywords: e-Waste, higher education, legislation, environment, sustainability.

References

[1] Official Journal of the European Union: Directive 2012/19/EU of the European Parliament and of the Council of 4 July 2012 on waste electrical and electronic equipment (WEEE). *Official Journal of the European Union (2012) L 197/38 - L 197/71 de 24/7/2012.*

[2] Wilkinson, S., Duffy, N. and Crowe, M. (2001) Waste from electrical and electronic equipment in Ireland: a status report. *Ireland: Environmental Protection Agency.*

[3] Eurostat: Waste Electrical and Electronic Equipment (WEEE) (2014). Available from: `http://ec.europa.eu/eurostat/web/products-datasets/-/env\_waselee(22/01/2014).`

Approximate probabilities and critical values for Bartels randomness test

Ayana Mateus[1,2] and Frederico Caeiro[1,2]

[1]Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal
[2]Centro de Matemática e Aplicações (CMA)

E-mail addresses: *amf@fct.unl.pt; fac@fct.unl.pt*

Abstract

Testing randomness is an important problem in statistics, since most inferential procedures concerning population characteristics require a random sample. In this work we consider and study the statistical properties of the Bartels randomness test [1] which is based on the von Neumann's ratio statistic. The exact null distribution can be obtained from the full set of permutations of the first n positive integers. However, this approach becomes impractical for moderate samples size. For larger sample sizes, the test statistic is asymptotically normally distributed with mean 2 and variance approximately equal to $20/(5n + 7)$, but the rate of convergence to the normal distribution is slow. Consequently, it is useful to have accurate approximations for the exact null distribution of the test statistic. Bartels obtained critical values for samples of size $10 \leq n \leq 100$, based on the beta approximation. In this work we study the precision of such critical values and present new approximations based on the Edgeworth expansion.

Keywords: Bartels test, Edgeworth expansion, nonparametric test, randomness, rank tests.

Acknowledgements

References

[1] Bartels, R. (1982) The Rank Version of von Neumann's Ratio Test for Randomness, *Journal of the American Statistical Association* **77(377)**: 40–46.
[2] Gibbons, J. D. and Chakraborti, S. (2011) *Nonparametric Statistical Inference*, CRC Press.
[3] Neumann, John von (1941) Distribution of the Ratio of the Mean Square Successive Difference to the Variance, *Annals of Mathematical Statistics* **12(4)**: 367–395.

On the Probability of Return Levels for Spatial Extreme Events

Luísa Pereira

Centro de Matemática e Aplicações da Universidade da Beira Interior (CMA-UBI), Covilhã, Portugal

E-mail address: *lpereira@ubi.pt*

Abstract

The impact of environmental extreme events, ranging from disturbances in ecosystems to economic impacts on society as well as loss of life, motivated the study of extremes of random fields.

In this paper the main question of interest is about risk: if occurs one exceedance of a high level in a given location, $x \in \mathbb{R}^2$, and the maximum over a neighborhood of x does not exceed the level then, what will be the probability that an exceedance occurs in another location? This probability is independent of the univariate marginal distribution of the random field and can be related with extremal coefficients, which will provide immediate estimators. We illustrate the results with an application to annual maxima temperatures over Texas.

Keywords: max-stable random fields, extremal coefficients, return level, extreme values.

Acknowledgements

References

[1] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004) *Statistics of Extremes: Theory and Applications,* John Wiley.

[2] Cooley, D., Cisewski, R. J., Erhardt, R. J., Jeon, S., Mannshardt, E., Omolo, B. O. and Sun, Y. (2012) A survey of spatial extremes: Measuring spatial dependence and modeling spatial effects. *Revstat*, **10**, 135–165.

[3] Ferreira, H. and Ferreira, M. (2012) On extremal dependence of block vectors. *Kybernetika* **48**: 988–1006.

[4] Ferreira, M. and Ferreira, H. (2012). On extremal dependence: some contributions. *Test* **21**: 566–583.

[5] Ferreira, H. and Pereira, L. (2015). Dependence maxima in space. *Journal of Physics: Conference Series (JPCS)* **574**: 12021–12027.

[6] Fonseca, C., Pereira, L., Ferreira, H. and Martins, A. P. (2015) Generalized madogram and pairwise dependence of maxima over two regions of a random field. *Kybernetika* **51(2)**: 193–211.

[7] Fonseca, C., Ferreira, H., Pereira, L. and Martins, A. P. (2014) Stability and contagion measures for spatial extreme value analyses. *Kybernetika* **50(6)**: 914–928.

[8]   Naveau, P., Guillou, A., Cooley, D. and Diebolt, J. (2009). Modelling pairwise depen-
      dence of maxima in space. *Biometrika* **96(1)**: 1–17.

[9]   Schlather, M. and Tawn, J. (2003). A dependence measure for multivariate and spatial
      extreme values: Properties and inference. *Biometrika* **90**: 139–156.

[10] Smith, R. L. (1990). Max-stable processes and spatial extremes. Preprint, Univ. North
      Carolina, USA.

Index of authors