Statistics | CIDMA CENTRO DE I&D EM MATEMÁTICA E APLICAÇÕES / CENTER FOR R&D IN MATHEMATICS AND APPLICATIONS | FCT FUNDAÇÃO PARA A CIÊNCIA E A TECNOLOGIA — universidade de aveiro

Probability and Statistics Group

Members: Adelaide Freitas, Ana Helena Tavares, Andreia Hall, Cláudia Santos, Helena Alvélos, Isabel Pereira (coordinator), M. Eduarda Silva, Magda Monteiro, Marco Costa, M. Conceição Costa, Pedro Macedo, Pedro Sá Couto, Vera Afreixo

Collaborators: Anabela Rocha, Isabel Magalhães, Cristina Miranda, Manuela Souto, Nélia Silva, Sónia Gouveia, Andreia Monteiro

PhD Students: Alberto Silva, Diana Rocha, Rodney Sousa

Group Characterization: The work developed by PSG is characterized by methodological developments, substantive applications in genetics, environmental, management and industry, biomedical areas and education and promotion of Maths and Stats

Research Topics:

. Time Series Analysis and its Applications

. Regularization and Maximum Entropy

. Genomics and Multivariate Statistics

. Data Analysis and General Applications

. Robustness

Exterior Visibility:

. Collaborative projects with industry, medicine and clinical research groups.

. Supervision of MSc thesis within industrial projects

. Members are involved in Academic Societies, particularly in Boards of Portuguese Statistics Society and of the Federation of the European National Statistics Societies

. Promotion of Maths and Statistics in elementary, middle and high schools

. Annual workshop of PSG – Interdisciplinarity and Applications

. Members are regularly giving invited talks in specialized sessions

. Annual Workshop on Maximum Entropy and its Applications

International Collaborations:

. Short-term visits of well known researchers such as A. Golan (Info-Metrics Institute, American Univ.), M. Vichi (Sapienza Univ. Rome), B. McCabe (Univ. Liverpool), J. Raymackers (KU Leuven), E. Ciavolino (Univ Salento), J. Husler (Univ. Bern), B. Mizinskias (Univ. Wroclaw), resulting in collaborative work.

. Close collaborations have been established with J.R. Pardinas (Univ Vigo), C. Weiss (Univ. Hamburg), P. Galindo (Univ. Salamanca), A. Alonso (Univ Carlos III).

. Visits to research centers and Universities of Hamburg, Weimar, Liverpool, Vigo, Salamanca, Carlos III, Bern, Carnegie Mellon University (CMU)-Pittsburgh, among others.

Current Research Topics:

. Dynamic factor state space models for multivariate discrete valued time series

. Analysis of truncated and censored time series

. Development of a Kalman type algorithm to reduce the bias of state space model estimates

. Variable selection and large scale data analysis with normalized entropy

. Multivariate analysis and genomics: data visualization, analysis of compositional data; data dimensionality reduction

Methodological with statistical development / Statistical development driven by applications

BAYESIAN OUTLIER DETECTION IN NON-GAUSSIAN AUTOREGRESSIVE TIME SERIES

MARIA EDUARDA SILVA,a,b* ISABEL PEREIRAb,c AND BRENDAN McCABEd
a Faculdade de Economia, Universidade do Porto, Porto, Portugal
b CIDMA, Aveiro, Portugal
c Departamento de Matemática, Universidade de Aveiro, Aveiro, Portugal
d Management School, The University of Liverpool, Liverpool, UK

Impact Factor: 0.826

This work investigates outlier detection and modelling in non Gaussian autoregressive time series models with margins in the class of a convolution closed parametric family. This framework allows for a wide variety of models for count and positive data types. The paper investigates additive outliers which do not enter the dynamics of the process but whose presence may adversely influence statistical inference based on the data. The Bayesian approach proposed here allows one to estimate, at each time point, the probability of an outlier occurrence and its corresponding size thus identifying the observations that require further investigation. The methodology is illustrated using simulated and observed data sets.

Comparing Reverse Complementary Genomic Words Based on Their Distance Distributions and Frequencies

Ana Helena Tavares[1] · Jakob Raymaekers[2] · Peter J. Rousseeuw[2] · Raquel M. Silva[3] · Carlos A. C. Bastos[4] · Armando Pinho[4] · Paula Brito[5] · Vera Afreixo[6]

Impact Factor: 0.64

Abstract

In this work, we study reverse complementary genomic word pairs in the human DNA, by comparing both the distance distribution and the frequency of a word to those of its reverse complement. Several measures of dissimilarity between distance distributions are considered, and it is found that the peak dissimilarity works best in this setting. We report the existence of reverse complementary word pairs with very dissimilar distance distributions, as well as word pairs with very similar distance distributions even when both distributions are irregular and contain strong peaks. The association between distribution dissimilarity and frequency discrepancy is also explored, and it is speculated that symmetric pairs combining low and high values of each measure may uncover features of interest. Taken together, our results suggest that some asymmetries in the human genome go far beyond Chargaff's rules. This study uses both the complete human genome and its repeat-masked version.

Efficiency in the European agricultural sector: environment and resources

Victor Moutinho[1] · Mara Madaleno[1] · Pedro Macedo[2] · Margarita Robaina[1] · Carlos Marques[1]

Impact Factor: 2.8

This article intends to compute agriculture technical efficiency scores of 27 European countries during the period 2005–2012, using both data envelopment analysis (DEA) and stochastic frontier analysis (SFA) with a generalized cross-entropy (GCE) approach, for comparison purposes. Afterwards, by using the scores as dependent variable, we apply quantile regressions using a set of possible influencing variables within the agricultural sector able to explain technical efficiency scores. Results allow us to conclude that although DEA and SFA are quite distinguishable methodologies, and despite attained results are different in terms of technical efficiency scores, both are able to identify analogously the worst and better countries. They also suggest that it is important to include resources productivity and subsidies in determining technical efficiency due to its positive and significant exerted influence.

Discrimination of water quality monitoring sites in River Vouga using a mixed-effect state space model

Marco Costa[1] · Magda Monteiro[1]

Impact Factor: 2.668

Abstract The surface water quality monitoring is an important concern of public organizations due to its relevance to the public health. Statistical methods are taken as consistent and essential tools in the monitoring procedures in order to prevent and identify environmental problems. This work presents the study case of the hydrological basin of the river Vouga, in Portugal. The main goal is to discriminate the water monitoring sites using the monthly dissolved oxygen concentration dataset between January 2002 and May 2013. This is achieved through the extraction of trend and seasonal components in a linear mixed-effect state space model. The parameters estimation is performed with both maximum likelihood method and distribution-free estimators in a two-step procedure. The application of the Kalman smoother algorithm allows to obtain predictions of the structural components as trend and seasonality. The water monitoring sites are discriminated through the structural components by a hierarchical agglomerative clustering procedure. This procedure identified different homogenous groups relatively to the trend and seasonality components and some characteristics of the hydrological basin are presented in order to support the results.

Applications in health sciences, economics, clinical research, industry and social sciences

Statistical Methods in Medical Research

2.284 Impact Factor | 5-Year Impact Factor 3.199 | *Journal Indexing & Metrics »*

Permutations of functional magnetic resonance imaging classification may not be normally distributed

Mohammed S Al-Rawi, Adelaide Freitas, João V Duarte, more... Show all authors

Impact Factor: 2.583

Research Article

Health-Related Quality of Life in Pulmonary Hypertension and Its Clinical Correlates: A Cross-Sectional Study

Abílio Reis,[1] Mário Santos,[1,2,3] Margarida Vicente,[4] Inês Furtado,[5] Célia Cruz,[5] Alzira Melo,[1] Luísa Carvalho,[1,5] Fabienne Gonçalves,[1,5] Pedro Sa-Couto,[6] and Luís Almeida[7]

Impact Factor: 5.47

SCIENTIFIC REPORTS

OPEN **DNA word analysis based on the distribution of the distances between symmetric words**

Ana H. M. P. Tavares[1,2], Armando J. Pinho[3,5], Raquel M. Silva[3,4], João M. O. S. Rodrigues[3,5], Carlos A. C. Bastos[3,5], Paulo J. S. G. Ferreira[3,4] & Vera Afreixo[1,2,4]

Promotion of Maths, Statistics, Education

. Summer Schools for children (ages 9 - 10) at the university of Aveiro

. Mathematical Circus project

. Professional development courses for mathematics teachers of all levels.

. Workshops for teachers and students in other institutions.

. Experiences using wikis and educational videos in higher education.

Exploring the Usage of MOOCs in Higher Education Institutions: Characterization of the Most Used Platforms

Carolina Costa (DEGEIT - University of Aveiro, Aveiro, Portugal), Leonor Teixeira (DEGEIT / IEETA - University of Aveiro, Aveiro, Portugal) and Helena Alvelos (DEGEIT / CIDMA - University of Aveiro, Portugal)

Interlacing Mathematics and Culture: Symmetry in Traditional Pavements and Crafts
Andreia Hall – CIDMA and Ricardo Cunha Teixeira - NICA