

Group Description	
Title of Research Group:	(RG-MATH-LVT-Lisboa-6-949) Probability, Modelling and Data Analysis
Principal Investigator:	Dinis Duarte Ferreira Pestana
Main Scientific Domain:	Matemática
Group Host Institution:	Faculdade de Ciências - Universidade de Lisboa

Funding, source, dates
Funding, source, dates
ALL THE REFERRED AMOUNTS INCLUDE OVERHEADS TO PAY FFCUL
FCT Base for Research Group Title: (RG-MATH-LVT-Lisboa-6-949) - Probability, Modelling and Data Analysis
Funding:
corresponding to 2008:
date: 29 Abril 2008 / 18957.45 Eur
date: 26 Novembro 2008 / 20379.26 Eur
Corresponding to 2006:
date: 29 Dezembro 2008 / 17162.00 Eur

Objectives & Achievements
Objectives
<p>1. Asymptotic issues: beta randomly extended Panjer's families, and their bearing in stopped sums and stopped maxima, and issues on N-infinite divisibility, N-self-decomposability and N-stability, extensions of Panjer-Sundt-Jewell results on aggregate claims, use of conjugate distributions to investigate large deviations and saddlepoint approximations. [P1], [P3], [MSc4], [ON2], [ON7], [ON8],[ON9], [ON11], [ON13], [ON15], [ON18].</p> <p>2. Statistical modeling: Applications of Zipf-Mandelbrot and discrete lognormal distributions in modeling biological and linguistic data; beta models in population dynamics. Distance sampling and ratio estimators, and error modeling, namely error in covariates issues. Research on parametric and non-parametric methods for longitudinal models, and investigation of link regression in the extended Panjer distributions. Mixtures, namely structure of hyperexponential mixtures. [P6], [P7], [OI2], [OI3], [OI6],[OI7], [OI8], [OI9], [PhD2], [PhD3], [MSc1], [MSc2], [ON10], [ON12], [ON14].</p> <p>3. Statistical inference: bias reduction issues and tail behaviour; extensions of studentization and analysis of scale for comparison of locations in non-gaussian symmetric distributions, with research on robustness properties; and for beta, Pareto and gamma distributions, using the inverse integral transform methods derived by Pestana, Rocha and Brilhante. [P2], [P3], [P4], [P5], [OI1],[OI4], [ON3], [ON6].</p> <p>4. Role of early XXth century Portuguese mathematicians in the History of Probability and Statistics, and teaching of Statistics, namely a discussion on teaching discrete models. [PhD1], [ON1], [ON14], [ON17].</p> <p>5. Applications in Medicine. [P8], [MSc3], [ON4], [ON5], [ON12], [ON14].</p>
Main Achievements
<p>1. Three research members of the group got their Ph. D. degree in 2008, and two others handed their dissertation. The supervisors were senior researchers of the group. Aside from [PhD1], [PhD2], [PhD3], and from the dissertations of Martins and of Felgueiras (degree conferred beginning 2009), this research led to the publication of [P7], [OI2], [OI3], [OI5], [OI6], [OI7], [OI8], [ON2], [ON3], [ON7], [ON11], [ON13], [ON16], [ON17] and [ON18]. In what concerns the supervision of students, we also mention the M.Sc. dissertations [MSc1], [MSc4], [MSc3], [MSc4], and the publications related to the supervision of Ph. D. students research, [P8], [ON4], and [ON5]. The commitment towards the supervision of Ph. D. programs attracted new members to CEAUL: Prof. J. Cadima, Prof. J. L. Rocha, and Dr. C. Ventura (Ph. D.: middle 2009), L. Soares de Almeida (Ph. D.: end of 2009), and Doctor F. Sequeira (Ph. D. June 2009). Hence there has been in 2008-2009 an increase of 8 Ph. D. members in this group, and two more are expected during 2009.</p> <p>2. Research on stability and on infinite divisibility (sums and extremes) led to the publication of [P1], [ON7], [ON8], [ON9], and [ON15], and collateral work [MSc4], [ON2] and [ON11]. S. Mendonça and D. Pestana collaboration with group G1 (Order Statistics, Extremes and Application), and the research on asymptotic results, has been fruitful, see [P3], [P4], [P5], [OI1], [OI4], [OI9],[ON6] and [ON9].</p> <p>3. The research of recently integrated members of this group led to the publication of [P2] (J. Cadima), [OI2], [OI3], [OI6] (J. L. Rocha), [ON10] (T. A. Marques),</p>

Objectives & Achievements

and the successful supervision of Aleixo (J. L. Rocha) and of Miranda and São João (J. Cadima). S. Cabral and H. Gonçalves subproject on the analysis of longitudinal data and medical applications led to the publication of [P6], [ON12] and [ON14]; [ON12] was awarded the “Best Poster Prize” at EQS 2008.

4. [P7] is the first result of a new area of interest (optimal designs). [ON1] 3rd revised and augmented edition is a leading book in teaching of statistics, and the History of Probability and Statistics in Portugal in the early XXth century is reported in [PhD1], [ON16] and [ON17].

Group Productivity

Publications in peer review Journals

[P1] Brillhante, M. F., and Kotz, S. (2008). "Infinite divisibility of the spacings of Kotz-Kozubowski-Podgórski generalized Laplace model", vol 78, Issue 15, 2433-2436.

[P2] Cadima, J., Calheiros, F.L., and Preto, I.P.: "The eigenstructure of block-structured correlation matrices and its implications for Principal Component Analysis." (accepted, Journal of Applied Statistics).

[P3] Gomes, M.I., Canto e Castro, L., Fraga Alves, M.I. & Pestana, D. (2008). Statistics of extremes for iid data and breakthroughs in the estimation of the extreme value index: Laurens de Haan leading contributions. *Extremes* 11:1, 3-34.

[P4] Gomes, M. I., Henriques-Rodrigues, L., Pereira, H., and Pestana, D. (2008). Tail index and second order parameters' semi-parametric estimation based on the log-excesses. *J. Statisti. Comput. and Simul.* (accepted).

[P5] Gomes, M.I., Pestana, D. & Caeiro, F. (2008). A note on the asymptotic variance at optimal levels of a bias-corrected Hill estimator. *Statistics and Probability Letters* (accepted, August 2008).

[P6] Gonçalves, M. Helena and Azzalini, Adelchi. (2008). "Using Markov chains for marginal modelling of binary longitudinal data in an exact likelihood approach." *Metron — International Journal of Statistics*. LXVI,157-181.

[P7] Martins, J.P., Mendonça, S., and Pestana, D.D. (2008). "Optimal and quasi-optimal designs", *REVSTAT* 6, 279-307.

[P8] Soares de Almeida, L, Requena, L., Rütten, A., Kutzner, H., Garbe, C., Pestana, D., and Marques Gomes, M. (2008). "Desmoplastic Malignant Melanoma: a Clinicopathological Analysis of 113 Cases", *American Journal of Dermopathology*, 30, 207-215.

Other publications International

CHAPTERS OF BOOKS

[O1] Gomes, M.I., Pestana, D., Henriques-Rodrigues, L. & Viseu, C. (2008). Tail Behaviour: an Empirical Study. In Arnold, B.C., Balakrishnan, N., Sarabia, J.M. and Míguez, R. (eds.), *Advances in Mathematical and Statistical Modeling*, Chapter 14, 195-207. ISBN 978-0-8176-4625-7, DOI: 10.1007/978-0-8176-4626-4_14, Birkhauser, Boston

PROCEEDINGS

[O12] Aleixo, S., Rocha, J. L., and Pestana, D. "Populational Growth Models Proportional to Beta Densities with Allee Effect", *Proceedings of the BVP2008, Conference on Boundary Value Problems, Mathematical models in Engineering, Biology and Medicine, American Institute of Physics Conference Proceedings Volume 1124*, 3-12.

[O13] Aleixo, S., Rocha, J.L., and Pestana, D. (2008). Populational Growth Models in the Light of Symbolic Dynamics". In Luzar-Stiffler, V., Dobric, V.H. and Bekic, Z. (eds.), *Proceedings of the ITI 2008, 30th International Conference on Information Technology Interfaces*, p. 311-316, Cavtat, Croatia.

[O14] Gomes, M.I., Henriques Rodrigues, L., Pereira, H. & Pestana, D. (2008). A semi-parametric estimator of a "scale" second order parameter based upon the log-excesses. In Luzar-Stiffler, V., Dobric, V.H. and Bekic, Z. (eds.) *ITI 2008 Proceedings*, ISBN 978-953-7138-12-7, pp. 329-334, 2008.

[O15] Pestana, D., and Martins, J. P. (2008). Optimal and quasi-optimal designs, 17th International Workshop on Matrices and Statistics, 142-143 (extended abstract).

OTHER PUBLICATIONS

[O16] Aleixo, S., Rocha, J.L., and Pestana, D. (2008). "Populational Growth Model Proportional to Beta Densities", *Notas e Comunicações do CEAUL*, 19/2008.

[O17] Felgueiras, M. (2008). "Gaussian Location Mixtures". *Notas e Comunicações do CEAUL*, 18/2008.

[O18] Felgueiras, M. (2008). "Gaussian Scale Mixtures". *Notas e Comunicações do CEAUL*, 16/2008.

[O19] Gomes, M.I., Henriques-Rodrigues, L., Pereira, H. & Pestana, D. (2008). "Tail Index and Second-Order Parameters Semi-Parametric estimation based on the Log-Excesses". *Notas e Comunicações do CEAUL*, 17/2008.

Other publications National

BOOKS

[ON1] Pestana, D. D. e Velosa, S. (2008). "Introdução à Probabilidade e à Estatística", 3ª ed. revista e aumentada, 1164 pp., Fundação Calouste Gulbenkian, Lisboa.

PROCEEDINGS

[ON2] Felgueiras, M. (2008). Considerações sobre a Distribuição Pareto. In Hill, M.M. et al. (eds.), *Estatística: da Teoria à Prática (Actas XV Congresso Anual da SPE)*, 193-202, 2008.

Group Productivity

[ON3] Martins, J.P. e Mendonça, S. (2008). Transformações Estabilizadoras da Variância. In Hill, M.M. et al. (eds.), Estatística: da Teoria à Prática (Actas XV Congresso Anual da SPE), 343-350, 2008.

[ON4] Soares de Almeida, L, Pestana, D., e Marques Gomes, M. (2008). Evolução da Actividade do Laboratório de Histopatologia Cutânea do Serviço de Dermatologia do Hospital de Santa Maria — Comparação de Três Lustros. Actas da VI Conferência de Estatística e Qualidade na Saúde (in print).

[ON5] Sousa, R. (2008). "O Problema de Dorfman Revisitado — Análises Quantitativas." (in print).

ABSTRACTS AND EXTENDED ABSTRACTS

[ON6] Caeiro, F., Gomes, M.I. & Pestana, D. (2008). Alguns resultados adicionais sobre a variância de um estimador de vies reduzido do índice de cauda. In Oliveira, I. et al. (eds.). XVI Congresso Anual da SPE — Programa e Resumos, 16, SPE editions.

[ON7] Diamantino, M. F. (2008) "O Teorema Limite de Rényi Revisitado". XVI Congresso Anual da SPE — Programa e Resumos, 38, SPE editions.

[ON8] Iglésias Pereira, H. (2008). "As Estáveis no R", XVI Congresso Anual da SPE — Programa e Resumos, 109, SPE editions.

[ON9] Malva, M., Mendonça, S., e Pestana, D. (2008). "Expansões de Edgeworth e aproximações pré-assintóticas". XVI Congresso Anual da SPE — Programa e Resumos, 75, SPE editions.

[ON10] Marques, T. A. & Thomas, L. (2008). "Density estimation for cetaceans from passive acoustic sensors: an example with beaked whales". XVI Congresso Anual da SPE — Programa e Resumos, 75, SPE editions.

[ON11] Santos, R., e Pestana, D. (2008) "Distância entre os extremos de uma linha poligonal 'ao acaso'". XVI Congresso Anual da SPE — Programa e Resumos, 128, SPE editions

[ON12] Trigueirão, E, Gonçalves, H. Cabral, M. S. E. (2008). "Comunicação: Análise estatística de dados longitudinais discretos numa experiência sobre a influência de lípidos no adenocarcinoma mamário". VI Conferência Estatística e Qualidade na Saúde, Lisboa.

[ON13] Aleixo, S. M., Diamantino, M. F. and Pestana, D. D. (2008). "Família parametrizada GLE". Notas e Comunicações do CEAUL, 15/2008.

[ON14] Cabral, M. S. E. (2008). "O modelo marginal na análise de contagens ao longo de tempo: uma aplicação ao estudo da influência de lípidos no adenocarcinoma mamário". Notas e Comunicações do CEAUL, 23/2008.

[ON15] Iglésias Pereira, H. (2008). "As Estáveis no R", (submitted).

[ON16] Santos, R. (2008). "Rodolpho Guimarães (1904)", Episódios na História da Estatística do Boletim da Sociedade Portuguesa de Estatística, p. 11.

[ON17] Santos, R. (2008). "José de Sousa Pinto (1913)", Episódios na História da Estatística do Boletim da Sociedade Portuguesa de Estatística, p. 12.

[ON18] Santos, R., e Pestana, D. (2008) "Distância entre os extremos de uma linha poligonal 'ao acaso'". (accepted, Actas do XVI Congresso Anual da SPE)

Master and Ph.D. thesis completed

Ph. D. Thesis:

[PhD1] Rui Sousa Santos: "Probabilidade Circa 1914 e a Construção de Pacheco d'Amorim".

Local: Universidade de Lisboa

Supervisor: Dinis Pestana.

[PhD2] Sandra Aleixo: "Métodos Analíticos em Probabilidade e Métodos Probabilísticos em Análise: Fractalidade Associada aos Modelos Beta(P,q), Evolução de Populações e Dimensões de Hausdorff".

Local: Universidade Nova de Lisboa

Supervisors: José Leonel Rocha and Dinis Pestana.

[PhD3] Fernanda Diamantino: "Contribuição ao Estudo de Dados em Falta".

Local: Universidade Nova de Lisboa

Supervisors: Dinis Pestana.

M. Sc. Thesis:

[MSC1] Regina Bispo: "Loglinear models for ordinal tables with zeros".

Supervisor: Dinis Pestana.

[MSC2] Ana Catarina Sequeira Nunes Coutinho de Miranda: "Modelação da Distribuição de Aves Limícolas durante a Maré-baixa no estuário do Tejo".

Supervisor: Jorge Cadima.

[MSC3] Ricardo Miguel Vieira de São João: "Modelação da Insuficiência Cardíaca Congestiva por Modelos Lineares Generalizados com variável resposta binária".

Supervisor: Jorge Cadima.

[MSC4] Ivanilda Cabral: "Sobre Estatísticas Ordinais e a Teoria dos Valores Extremos".

Supervisor: Sandra Mendonça.

Organization of conferences

Group Productivity

D. Pestana was a member of the scientific committee of EQS 2008 — VI Conferência de Estatística e Qualidade na Saúde, <http://eqs.estesl.ipl.pt>.

Internationalization

S. Cabral and H. Gonçalves workteam on longitudinal data analysis continued their long-standing collaboration with Prof. Azzalini.

T. A. Marques maintained his strong bonds with S. Buckland's St. Andrews team working in distance sampling.

D. Pestana is a Portuguese member of CENE, a joint committee contributing to the multilingual ISI dictionary of Statistics (CENE provided Portuguese translation for more than 95% of the entries) and standardization of Portuguese-Brazilian glossary of Statistics.

J. Cadima, S. Mendonça and D. Pestana act as reviewers for JABES, Computational Statistics and Data Analysis, J. Applied Statistics, J. Korean Statistical Society, Mathematical Reviews, Revstat, and Acquaculture. J. Cadima is associate editor of JABES, D. Pestana is associate editor of Revstat.

D. Pestana has been an invited speaker at the 17th International Workshop on Matrices and Statistics ("Optimal and Quasi-Optimal Designs", with J. P. Martins). T. A. Marques has been an instructor at the workshop "Introduction to distance sampling", September 2008 – Marshal, California.

Members of this research group presented their work at the International Conference BVP2008, Mathematical Models in Engineering, Biology and Medicine, Conference on Boundary Value Problems, Santiago de Compostela, Spain (S. Aleixo and D. Pestana), at International Conference DYN2008, Dynamics & Applications, Braga, Portugal (S. Aleixo and D. Pestana), 30th International Conference on Information Technology and Interfaces, Cavtat, Croatia (S. Aleixo and D. Pestana), at 17th International Workshop on Matrices and Statistics, 2008 (J. Cadima, D. Pestana and J. P. Martins), at the 7th World Congress in Probability and Statistics, Singapore (M. Felgueira, D. Pestana and J. P. Martins), at the Conference on Asymptotic Statistics, Barcelona (J. Paulo Martins), at

1st International Statistical Ecology Conference – St Andrews, UK (T. A. Marques), and at the II Iberian Mathematical Meeting, Badajoz (D. Pestana and F. Sequeira).

D. Pestana lectured at the international Ph. D. program on Computational Biology sponsored by Calouste Gulbenkian Foundation. He has also been invited to make a conference on "Meta-Analysis and other Statistical Tools for Knowledge Building" for the international researchers at IGC.

Future Research

Objectives

- Meta analysis and applications in Medicine, namely in studies of desmoplastic melanoma. Uniformity issues related to combined p-values and random p-values. Research team: F. Sequeira, F. Brilhante, S. Mendonça, D. Pestana, L. Soares de Almeida, R. Vasconcelos, M. Malva, J. P. Martins.
- Structure of population matrices and Principal Components Analysis. Research team: J. L. Cadima and co-workers from ISA.
- History of Science — Research in Probability and Statistics in Portugal during the first Republic. Research team: R. Santos, S. Velosa, E. Graça Martins, D. Pestana, S. Maciel Pires, E. Mendes de Oliveira.
- Fractal issues in population models — analysis of the Gompertz model. Research team: J. L. Rocha, S. Aleixo, D. Pestana.
- Scale analysis using the independence of scale and location estimators. Research team: D. Pestana, M. Marques dos Santos, F. Brilhante, S. Mendonça, M. Malva, M. Reis, H. Iglésias Pereira.
- Analysis of binary longitudinal data, in view of medical applications. Research team: S. Cabral, H. Gonçalves.
- Further work on errors in distance sampling models, and modeling bird death rates in eolic plants. Research team: T. A. Marques, jointly with R. Bispo Carita (ISPA) and his co-workers at St. Andrews.
- Teaching of Statistics: Documentation for teachers, and collaboration with INE's Alea Project. Research team: E. Graça Martins, E. Mendes de Oliveira.
- Extensions of Dorfman's theory on analysis of combined blood samples: specificity and sensibility issues in quantitative tests; mixture models. Research team: D. Pestana, M. Felgueiras, R. Sousa, H. Iglésias Pereira, R. Vasconcelos.
- Analysis of messy data, and development of models for contingency tables. Research team: J. Paulo Martins, R. Vasconcelos, P. Semblano.
- Simulation studies of the Rényi's rarefaction, and applications to sampling theory. Research team: S. Aleixo, F. Diamantino, S. Mendonça, D. Pestana, H. Iglésias Pereira.
- Scientific translation in the area of Probability and Statistics, and development of an e-learning platform for professional translators for TRADULÍNGUAS. Research team: C. Ventura, S. Velosa, D. Pestana.

Funding, source, dates

ALL THE REFERRED AMOUNTS INCLUDE OVERHEADS TO PAY FFCUL

1. FCT Base for Research Group (RG-MATH-LVT-Lisboa-6-949) Probability, Modelling and Data Analysis:

A. Funding left to 2009 from 2008:

Due to the delay and late transferring dates from FCT (at the end of the year), the Group was not able to schedule a convenient execution of a percentage from the budget of previous year, namely: around 47651.61 Eur.

B. Funding from 2009:

Future Research

Already transferred the first semester amount from FCT on the 2nd April 2009: 16839,38 Eur

Expected Second semester amount from FCT : similar to first semester