

Workshop on Probabilistic and Statistical Methods

January 28–30, 2013

São Carlos, SP, Brazil

PROGRAM

ICMC/USP - UFSCar - INCTMat

Organizers

- Instituto de Ciências Matemáticas e de Computação - ICMC da Universidade de São Paulo - USP
- Universidade Federal de São Carlos - UFSCar
- Instituto Nacional de Ciência e Tecnologia de Matemática - INCTMat

Support

- ICMC-USP, CAPES, FAPESP

Invited Speakers

- Adriano Suzuki (ICMC - USP)
- Cristian Villegas (ESALQ - USP)
- Francisco Cribari-Neto (UFPE)
- Gisela Tunes (IME - USP)
- Gleici Castro Perdoná (FMRP - USP)
- Hedibert Freitas Lopes (The University of Chicago)
- Javiera Barrera (Universidad Adolfo Ibáñez)
- Luis Ernesto Salazar (UFSCar)
- Marinho Gomes de Andrade Filho (ICMC - USP)
- Nancy Garcia (IMECC - UNICAMP)
- Raydonal Ospina (UFPE)
- Ronaldo Dias (IMECC - UNICAMP)
- Sebastian Grynberg (Universidad de Buenos Aires)
- Viviana Giampaoli (IME - USP)

Scientific Committee

- Carlos Diniz (DEs - UFSCar)
- Ernesto Mordecki (Universidad de la República)
- Fabio Prates Machado (IME - USP)
- Francisco Louzada Neto (ICMC - USP)
- Gilberto A. Paula (IME - USP)
- Mário de Castro Andrade Filho (ICMC - USP)

Organizing Committee

- Cibele Russo (ICMC - USP)
- Danilo Alvares da Silva (ICMC - USP)
- Pablo Martin Rodriguez (ICMC - USP)
- Vera Tomazella (DEs - UFSCar)

Workshop on Probabilistic and Statistical Methods

January 28–30, 2013

São Carlos, SP, Brazil

SCHEDULE

ICMC/USP - UFSCar - INCTMat

Conferences (60 min talks)

1. Francisco Cribari-Neto, UFPE
Testing inference in beta regressions
2. Hedibert Freitas Lopes, The University of Chicago
Modeling of complex stochastic systems via latent factors
3. Nancy Garcia, IMECC - UNICAMP
Functional data analysis of Near-infrared (NIR) spectroscopy data
4. Raydonal Ospina, UFPE
Uma introdução aos modelos GAMLSS
5. Ronaldo Dias, IMECC - UNICAMP
A Hierarchical Model for Aggregated Functional Data
6. Viviana Giampaoli, IME - USP
Proposta de Predição para Modelos Logísticos Multiníveis

Stochastic Processes Session (30 min talks)

1. Javiera Barrera, Universidad Adolfo Ibáñez
Capacity Allocation of Operating Rooms under Waiting Time Constraints: The Chilean AUGÉ program
2. Sebastian Grynberg, Universidad de Buenos Aires
Subcritical Regime in the Discrete Boolean Percolation Model

Regression Models Session (30 min talks)

1. Cristian Villegas, ESALQ - USP
Modelos Lineares Generalizados Simétricos
2. Marinho Gomes de Andrade Filho, ICMC - USP
A Generalized Species-Area Relationship: The Poisson Distribution Case

Survival Analysis Session (30 min talks)

1. Gisela Tunes, IME - USP
Modelagem do tempo de sobrevivência ajustado pela qualidade de vida e aplicação a dados de pacientes internados em UTI
2. Gleici Castro Perdoná, FMRP - USP
A New Weibull Family of Hazard Model for Breast Cancer Data

Young Researcher Session (30 min talks)

1. Adriano Suzuki, ICMC - USP
Modelagem Estatística para Previsão de Resultados de Jogos da Copa do Mundo 2010 de Futebol: Uma Abordagem Bayesiana
2. Luis Ernesto Salasar, UFSCar
Closed-form expressions for the Maximum Integrated Likelihood Estimators for the Closed Population Capture-recapture Model M_t

Oral Communications (20 min talks)

1. Danilo Alvares da Silva, ICMC - USP
Análise e Previsões de Vazões com um Modelo Log-Normal Truncado sob uma Abordagem Bayesiana
2. Eduardo Moreno, Universidad Adolfo Ibañez
Restricted Risk Measures and Robust Optimization
3. Márcio Poletti Laurini, FEA/RP - USP
Approximate Bayesian Computing via Generalized Empirical Likelihood and Method of Simulated Moments
4. Ricardo Z. N. Vêncio, FFCLRP - USP
A Bayesian model and its R package implementation for sugar-cane metabolite probabilistic identification

Minicurso

- Cibele Russo, ICMC - USP
Modelos de regressão para dados correlacionados

Posters

1. Alina Marcondes, ICMC-USP
Modelo de Calibração Ultraestrutural
2. Alexandre Hiroshi Watanabe, ICMC-USP
Estatística de Cramér-Von Mises para dados discretos na presença de censura
3. Carolina C. M. Paraíba, UFSCar
Estimating soil-water characteristic curves using truncated normal nonlinear regression models
4. Cristian Coletti, UFABC
Convergence to the Brownian Web for a generalization of the drainage network model
5. Daniel Francisco Neyra Castañeda, UFS
Percentual de acertos das simulações MA no R.
6. Débora Delbem, UFSCar
A correlação no modelo de distribuição de Perdas Agregadas Operacionais no cálculo do Capital Regulatório
7. Deisy Morselli Gysi, UFPR
Métodos Estatísticos na Análise de Dados Genômicos
8. Eder Angelo Milani, UFSCar
Modelo de risco complementar log-log dependente do tempo
9. Flávia Bolssone do Prado, UFSCar
Modelo de Regressão Bivariado
10. George Lucas Moraes Pezzott, UFSCar
Estimação Bayesiana do Número de Erros de Software a partir de um Modelo de Captura-recaptura com heterogeneidade
11. Guaraci Requena, UFSCar
O Capital Regulatório no Risco Operacional Segundo o Método da Convolução com o Uso de Cópuas
12. João Paulo Taconeli, UFSCar
Modelo de Fragilidade com Fração de Cura na Presença de Covariáveis: Uma Abordagem Paramétrica
13. Katherine Elizabeth Coaguila Zavaleta, UFSCar
Modelo destrutivo com variável terminal em experimentos quimiopreventivos de tumores em animais

14. Leillimar dos Reis Freitas, UFSCar
Qualidade de ajuste de regressão binária pelas funções de ligação logit e probit considerando diferentes tamanhos de amostras
15. Maria da Conceição Farias Freitas Tandel, UNESP
Descrição da composição química e atividade antioxidante da própolis orgânica de *Apis mellifera* produzida no Sul do Brasil
16. Marina M. G. Osio, ICMC-USP
Modelos multiníveis simétricos para dados educacionais
17. Ricardo Ferreira da Rocha, UFSCar
Analysis of the Generalized Mean Function and Voted Classifier as a Combination Technique in Credit Scoring Models with Bagging
18. Robson J. M. Machado, UFSCar
Um modelo semiparamétrico com efeitos aleatórios para dados farmacocinéticos
19. Rosineide F. da Paz, UFSCar
Practical aspects of the estimation of mixture model via Dirichlet Process
20. Silvana Aparecida Meira, UFSCar
Uma Abordagem de EM para Modelo de Mistura
21. Verônica Amparo Quispe Chire, UFSCar
Estimação da distribuição a posteriori do modelo Poisson bivariado a partir de dados ampliados
22. Wesley Bertoli da Silva, UFSCar
Estimação dos Parâmetros do Modelo Poisson Bivariado via Máxima Verossimilhança
23. Willian Luís de Oliveira, UFSCar
Critérios clássicos e bayesianos de seleção de modelos para a classe de modelos de regressão Poisson inflacionado de zero

Workshop on Probabilistic and Statistical Methods

January 28–30, 2013

São Carlos, SP, Brazil

ABSTRACTS

ICMC/USP - UFSCar - INCTMat

Minicurso

Modelos de regressão para dados correlacionados

Cibele Russo

ICMC - Universidade de São Paulo

Abstract: O objetivo principal deste curso é introduzir conceitos de modelos de regressão lineares e não lineares para dados correlacionados, que contemplam modelos para dados longitudinais, medidas repetidas, modelos multiníveis, entre outros. Discutiremos também a forma de incluir efeitos aleatórios em modelos de regressão, a escolha entre o enfoque marginal ou hierárquico, análise de diagnóstico e seleção de modelos. Aplicações serão discutidas com o auxílio do pacote estatístico R.

Obs. Este minicurso está direcionado para estudantes de graduação e mestrado.

Conferences

Testing inference in beta regressions

Francisco Cribari-Neto
Universidade Federal de Pernambuco

Abstract: We consider the issue of performing accurate small-sample testing inference in beta regression models, which are useful for modelling continuous variates that assume values in $(0,1)$, such as rates and proportions. We derive the Bartlett correction to the likelihood ratio test statistic and also consider a bootstrap Bartlett correction. Using Monte Carlo simulations we compare the finite sample performances of the two corrected tests to that of the standard likelihood ratio test and also to its variant that employs Skovgaard's adjustment; the latter is already available in the literature. The numerical evidence favors the corrected tests we propose. We also present an empirical application.

Modeling of complex stochastic systems via latent factors

Hedibert Freitas Lopes
The University of Chicago

Abstract: Factor models, and related statistical tools for dimension reduction, have been widely and routinely used in psychometric, item response theory, geology, econometric and biological, amongst many other fields, since the late 1960's when Karl G. Jöreskog, a Swedish statistician, proposed the first reliable numerical method for maximum likelihood estimation (MLE) in factor analysis (Jöreskog, 1969). Such developments happened, certainly not by chance, around the same time the computer industry was experiencing major advances. From a Bayesian perspective, Martin and McDonald (1975) showed that MLE suffers from several inconsistency issues (for instance, negative idiosyncratic variances). Nonetheless, Bayesian researchers themselves could not produce general algorithms for exact posterior inference for factor models until the early 1990's when the computer industry had another wave of major advances and Markov chain Monte Carlo (MCMC) schemes were almost instantly customized for all fields cited above. In this talk, my goal is to illustrate how such advances, both in factor modeling and statistical computing, have driven my own research in financial econometrics, spatio-temporal modeling and macro and microeconomics, among others. This will be done by linking my own.

Functional data analysis of Near-infrared (NIR) spectroscopy data

Nancy Garcia

IMECC - Universidade Estadual de Campinas

Abstract: Different substances reflect light in a characteristic way. The ultraviolet light has wavelength (in nanometers) in the range 1 – 400 nm, visible light in 400–750 nm and infrared region in 750 to 10^6 nm. The latter is subdivided into near-infrared (NIR) region: 750–2500 nm, mid-infrared (MIR) region: 2500–16,000 nm and far-infrared (FIR): 16,000– 10^6 nm. When materials are submitted to different wavelengths, the overtones and combinations in the NIR band will produce very complex patterns that characterize the constituents of the sample. Usually the samples are submitted to different wavelengths of intervals of less than 10 nm and the spectrum is a curve, although it is only measured at discrete set of points. Hence, the different characteristic curves of the constituents of the sample overlap and give rise to a curve which is a sum of several curves depending on the concentration of each substance. There are several techniques to measure the constituents of a sample. As expected, those very precise are expensive (the so-called analytical techniques), while others are relatively inexpensive at the cost of precision (as in the case of NIR spectroscopy). In this work, we obtain multivariate, simultaneous prediction for NIR spectroscopy data based on a functional interpretation of the Beer-Lambert formula.

Uma introdução aos modelos GAMLSS

Raydonal Ospina

Universidade Federal de Pernambuco

Abstract: Nesta palestra introduziremos uma ampla classe de modelos de regressão semi-paramétricos proposta por Robert Rigby e Mikis Stasinopoulos, a saber: GAMLSS (generalized additive models for location, scale and shape) a qual possui um ambiente computacional bem desenvolvido no R (<http://www.gamlss.org/>) para inúmeras aplicações.

A hierarchical model for aggregated functional data

Ronaldo Dias

IMECC - Universidade Estadual de Campinas

Abstract: In many areas of science one aims to estimate latent sub-population mean curves based only on observations of aggregated population curves. By aggregated curves we mean linear combination of functional data that cannot be observed individually. We assume that several aggregated curves with linear independent coefficients

are available. More specifically, we assume each aggregated curve is an independent partial realization of a Gaussian process with mean modeled through a weighted linear combination of the disaggregated curves. We model the mean of the Gaussian processes as a smooth function approximated by a function belonging to a finite dimensional space \mathcal{H}_K which is spanned by K B-splines basis functions. We explore two different specifications of the covariance function of the Gaussian process: one that assumes a constant variance across the domain of the process, and a more general variance structure which is itself modelled as a smooth function, providing a nonstationary covariance function. Inference procedure is performed following the Bayesian paradigm allowing experts' opinion to be considered when estimating the disaggregated curves. Moreover, it naturally provides the uncertainty associated with the parameters estimates and fitted values. Our model is suitable for a wide range of applications. We concentrate on two different real examples: calibration problem for NIR spectroscopy data and an analysis of distribution of energy among different type of consumers.

Proposta de Predição para Modelos Logísticos Multiníveis

Viviana Giampaoli

IME - Universidade de São Paulo

Abstract: Os modelos multiníveis são apropriados quando unidades pertencem a um grupo específico, isto é, os dados apresentam uma estrutura de agrupamento hierárquica. Tratamos do problema de prever a variável resposta para novos grupos quando esta foi modelada por meio de modelos logísticos mistos. Apresentaremos a comparação entre novos métodos de predição para o modelo misto com dois efeitos aleatórios, considerando métodos de estimação numérica apropriados. A comparação é realizada através de estudos de simulação e utilizando uma base de dados real.

Stochastic Processes Session

Capacity allocation of operating rooms under waiting time constraints: the Chilean AUGE program

Javiera Barrera
Universidad Adolfo Ibáñez

Abstract: We study the problem of allocating fixed Operating Room (OR) capacity among a set of surgeries that share personnel and infrastructure.

In the Chilean context, this problem must take into consideration the legal bounds imposed by the Health Reform (Plan AUGE) in terms of the maximum time a patient can wait for treatment, and also in terms of the costs of failing to provide treatment in time.

We will study the problem faced by any surgical specialty chief, who must decide how to allocate fixed OR capacity among surgeries from her department. For some examples, we will determine the optimal allocation of fixed OR capacity and the total expected cost of such decision. We will study the case when the decision should be made once at the beginning of the year and remain fixed for every following month. As well as the case where decisions can be made on a monthly basis, and quantifying the improvement of the new optimal solution. Joint work with Susana Mondschein and Paula Montes.

Subcritical regime in the discrete boolean percolation model

Sebastian Grynberg
Universidad de Buenos Aires

Abstract: We consider the Bernoulli Boolean model of discrete percolation. We show that there is a subcritical phase if the density of points on the d -dimensional lattice is small and if the random radius has d -moment finite.

Regression Models Session

Modelos lineares generalizados simétricos

Cristian Villegas

ESALQ - Universidade de São Paulo

Abstract: O objetivo deste trabalho é introduzir os modelos lineares generalizados simétricos no sentido dos modelos lineares generalizados, isto é, uma função de ligação é definida para estabelecer a relação entre os valores médios das distribuições simétricas e os preditores lineares. A classe de distribuições simétricas possui distribuições com caudas mais leves e mais pesadas do que a distribuição normal e oferece maior flexibilidade para analisar dados com simetria. Um algoritmo de mínimos quadrados ponderados é usado para estimar os parâmetros do modelo. Ilustramos o uso de esses modelos com um conjunto de dados reais. Trabalho em conjunto com Gilberto A. Paula, Francisco José A. Cysneiros e Manuel Galea.

A Generalized Species-Area Relationship: The Poisson Distribution Case

Marinho Gomes de Andrade Filho

ICMC - Universidade de São Paulo

Abstract: Species richness is related to various factors, such as mutation, interaction, competition and the amount of available resources for survival, among others. Additionally, species richness is related to the size of the habitat area (species-area relationship). A major problem in ecology is to identify the best function that models this relationship. The main aim of the paper is to propose a generalized fitting formula for the species area relationship. The advantage of our model is that it leads to a unique formulation for species-area relationships, which takes into account both the effects of the minimal area and asymptotic behavior of the growth curve for large areas. This approach provides a unique algorithm for fitting different data sets, and choosing the best model in light of the data. The applicability of our approach was tested via a simulation study conducted to determine if two usual selection criteria are suitable to choose the best formulation to be considered to describe species-area relationships. A real data set involving a species richness of fish in 70 lakes was analysed using the proposed methodology.

Survival Analysis Session

Modelagem do tempo de sobrevivência ajustado pela qualidade de vida e aplicação a dados de pacientes internados em UTI

Gisela Tunes

IME - Universidade de São Paulo

Abstract: O índice QAST (quality-adjusted survival time) é uma medida que quantifica simultaneamente a duração e a qualidade de vida. O ICESP - Instituto do Câncer do Estado de São Paulo - desenvolve um projeto em que procura avaliar, por meio do tempo de sobrevivência ajustado pela qualidade de vida, como se dará a qualidade de vida, além do tempo de sobrevida, de pacientes com câncer internados em sua Unidade de Terapia Intensiva. Neste trabalho, será discutida a modelagem do tempo de sobrevivência ajustado pela qualidade de vida utilizando métodos desenvolvidos recentemente. A aplicabilidade das técnicas consideradas é avaliada na base de dados do ICESP. Trabalho em conjunto com Antonio Carlos Pedroso de Lima.

A new Weibull family of hazard model for breast cancer data

Gleici Castro Perdoná

FMRP - Universidade de São Paulo

Abstract: In this work, we present a family Weibull Modified of hazard model to breast cancer problematic. The breast cancer is addressed here by the high incidence and lack of knowledge in survival among women worldwide. The model is very flexible, and accommodate several particular cases. Inference procedure is based on maximum likelihood. A simulation study is performed. A real example on breast cancer is addressed.

Young Researcher Session

Modelagem estatística para previsão de resultados de jogos da Copa do Mundo 2010 de Futebol: uma abordagem bayesiana

Adriano Suzuki

ICMC - Universidade de São Paulo

Abstract: Neste trabalho propomos uma metodologia estatística bayesiana com enfoque subjetivista para previsão dos resultados dos jogos da Copa do Mundo de 2010. A informação a priori considerada pelo modelo é a opinião de especialistas e os escores FIFA divulgados previamente ao início da competição. As probabilidades de vitória, empate e derrota em cada jogo foram obtidas de maneira exata e, a partir de uma simulação estocástica, estimamos as probabilidades de classificação na fase de grupo, de se chegar a cada uma das rodadas da fase eliminatória e de vencer o torneio.

Closed-form expressions for the maximum integrated likelihood estimators for the closed population capture-recapture model M_t

Luis Ernesto Salazar

DEs - Universidade Federal de São Carlos

Abstract: In this work, we consider the closed population capture-recapture model M_t , where the unknown number of individuals is the parameter of interest and the capture probabilities vary with the capture occasions. In order to eliminate the nuisance parameters (capture probabilities), the likelihood function is integrated with respect to a weight function (uniform and Jeffreys) of the nuisance parameters resulting in an integrated likelihood function depending only on the interest parameter, the population size. For these integrated likelihood functions, an analytical expression for the maximum likelihood estimate is obtained and it is proved that it is always finite and unique.

Oral Communications

Análise e Previsões de Vazões com um Modelo Log-Normal Truncado sob uma Abordagem Bayesiana

Danilo Alvares da Silva
ICMC - Universidade de São Paulo

Abstract: Este trabalho apresenta a construção de modelos Log-Normais Truncados para uma série de vazões afluentes, desenvolvidos a partir de modelos periódicos autorregressivos de ordem 1 (PAR(1)). Além disso, sob uma perspectiva Bayesiana, são adotadas distribuições a priori informativas para os ajustes dos modelos e obtenção das estimativas dos parâmetros. Com isso, são feitas previsões das vazões médias mensais para um horizonte de 36 meses. Assim, esses resultados podem ser utilizados no planejamento energético da operação de sistemas hidroelétricos. Trabalho em conjunto com Marinho Gomes de Andrade Filho.

Restricted Risk Measures and Robust Optimization

Eduardo Moreno
Universidad Adolfo Ibañez

Abstract: We study the restriction of coherent and distortion risk measures to the spaces of linear and affine linear random variables and the robust uncertainty sets associated to these risk measures. We show that for discrete distributions the uncertainty sets associated to these risk measures are scalings of previously considered uncertainty sets. We present computational result that suggest that these measures could help mitigate estimation errors.

Approximate Bayesian Computing via Generalized Empirical Likelihood and Method of Simulated Moments

Márcio Poletti Laurini
FEA/RP - Universidade de São Paulo

Abstract: An important recent literature discusses the use of Bayesian methods using numerical approximations which are likelihood free (Approximate Bayesian Computation - ABC). These methods allow inferences on issues hitherto untractable due to the complexity of the evaluation of the likelihood function. We discuss how to formulate ABC methods using Generalized Empirical Likelihood, extending the work of Mergensen et al (2012). These methods allow a formulation computationally more efficient than the usual ABC methods, and can be used even when it is not possible to calculate the moment conditions analytically through the use of simulated moment conditions. We apply these methods in the estimation of Stochastic Volatility Models.

A Bayesian model and its R package implementation for sugar-cane metabolite probabilistic identification

Ricardo Z. N. Vêncio
FFCLRP - Universidade de São Paulo

Abstract: Metabolomics is an emerging field of study in post-genomics which aims at comprehensive analysis of small organic molecules in biological systems. Experimental techniques involving mass spectrometry coupled to liquid chromatography (LC-MS) stand out as one of the dominant methods in metabolomics. In spite of high accuracy LC-MS machines available, identity inference of all metabolites is still subject to strong uncertainty. Rogers et al., (Bioinformatics 2010) proposed a Bayesian methodology to assign probabilities to this metabolite identification problem that takes into account important a priori information. This model uses a Gibbs sampler scheme to reach the distribution of a binary matrix which represents connections among metabolites and measured signals. In the present work we extended this probabilistic model suggesting (i) an improved likelihood function and (ii) introducing biologically oriented a priori information. We applied the developed methodology to an original sugar-cane metabolome dataset generated by the FAPESP Bioenergy Program in order to probabilistically identify important signaling molecule candidates. Finally, the theoretical model improvements were computationally implemented in R codes that also organizes available spectra pre-processing tools in a single coherent R package (*probmetab*) to be submitted to CRAN for the benefit of all statistical bioinformatics community. Joint work with Ricardo R. Silva.

Posters

Estatística de Cramér-Von Mises para dados discretos na presença de censura

Alexandre Hiroshi Watanabe
ICMC - Universidade de São Paulo

Abstract: Um dos principais problemas em testes de hipóteses para a homogeneidade de curvas de sobrevivência ocorre quando as taxas de falha (ou funções intensidade) não são proporcionais. O ponto inicial para análise consiste em assumir que os dados são contínuos e neste caso processos Gaussianos apropriados podem ser utilizados para testar a hipótese de homogeneidade. Aqui, citamos o testes de Renyi e Cramér von-Mises, ver [3]. Apesar destes testes não paramétricos apresentarem bons resultados para dados contínuos, estes podem ter problemas para dados discretos ou arredondados. Neste poster, propomos um novo teste não paramétrico baseado na estatística de Cramér von-Mises que nos permite detectar taxas de falha não proporcionais sujeitas a censuras arbitrárias para dados discretos ou arredondados. Ao aplicarmos nossa estatística para dados discretos ou arredondados, o teste desenvolvido neste trabalho apresenta uma função poder melhor do que os testes clássicos. Trabalho em conjunto com Dorival Leão Pinto Júnior (ICMC-USP).

Modelo de Calibração Ultraestrutural

Alina Marcondes
ICMC - Universidade de São Paulo

Abstract: A sociedade utiliza os programas de Ensaio de proficiência (EP) para avaliar a competência e a confiabilidade dos laboratórios ao executar medições específicas. Atualmente, diversos grupos de EP foram estabelecidos pelo INMETRO, entre estes, o grupo de testes de motores. Cada grupo é formado por diversos laboratórios que medem o mesmo artefato e suas medições são comparadas através de métodos estatísticos. A GM Powertrain, gentilmente, cedeu um motor gasolina 1.0, que foi escolhido como artefato pelo grupo de motores. A potência do artefato foi medida em 10 pontos de rotação por 6 laboratórios. Neste trabalho, motivados por este conjunto de dados, estendemos o modelo de calibração comparativa de Barnett (1969) para avaliar a compatibilidade dos laboratórios considerando a distribuição t de Student. Para avaliar tal compatibilidade, usamos o algoritmo EM e testes de hipóteses assintóticos. Trabalho em conjunto com Reiko Aoki (ICMC-USP).

Estimating soil-water characteristic curves using truncated normal nonlinear regression models

Carolina C. M. Paraíba
DEs - Universidade Federal de São Carlos

Abstract: Soil-water characteristic curves are useful graphical tools important to study the relationship between soil and water, a physical phenomenon that affects soil use in many different purposes. These curves relate water volume content (response variable) with matric potential (regressor variable), being usually estimated by nonlinear regression models fitted using nonlinear least squares procedures to data sets obtained from laboratory experiments. However, given the nature of the data, it is known that the observed water content at a matric potential will be such that it is not less than the residual soil-water content, and no more than the saturated soil-water content, a phenomenon known in statistics as truncation. In this scenario, the well established usual least squares procedures can be highly biased, inefficient, and inconsistent, which can seriously affect the estimated curve and prediction based on it. Therefore, it is important to account for truncation in soil-water retention data analysis. In the present paper, we propose an alternative approach for estimating soil-water characteristic curves based on truncated normal nonlinear regression models. Maximum likelihood estimator of the curve parameters are obtained by direct maximization of the likelihood function. Simulation studies are provided to assess the quality of estimates for the proposed regression model and diagnostic analysis are considered to check for model adequacy. We also provide a comparison study, based on simulation results, between the proposed methodology and the usual nonlinear least squares procedure. A real data set is analyzed using the proposed methodology. Joint work with Carlos A. R. Diniz (UFSCar), Aline H. N. Maia (Embrapa - Meio Ambiente) and Lineu N. Rodrigues (Embrapa - Cerrados).

Convergence to the Brownian Web for a generalization of the drainage network model

Cristian Coletti
CMCC - Universidade Federal do ABC

Abstract: We introduce a system of one-dimensional coalescing nonsimple random walks with long range jumps allowing crossing paths and exhibiting dependence before coalescence. We show that under diffusive scaling this system converges in distribution to the Brownian Web. Joint work with Glauco Valle da Silva Coelho (UFRJ).

Percentual de acertos das simulações MA no R

Daniel Francisco Neyra Castañeda
Universidade Federal de Sergipe

Abstract: One of the goals of the series over time is to identify the process generating this series. In practical situations the process of de generating a series can be identified by a particular case of models such as the ARIMA models (autoregressive integrated moving-average), following the methodology of Box and Jenkins, however unaware of the tasks allocated in these models in R. To identify this performance (which was classified) were generated simulations for various parameters of AR an MA models and compared with their respective theoretical series. This paper tries to clarify some gaps as to the performance of functions: `arima.sim()` and `auto.arima()` in R.

A correlação no modelo de distribuição de Perdas Agregadas Operacionais no cálculo do Capital Regulatório

Débora Delbem
DEs - Universidade Federal de São Carlos

Abstract: O termo Risco Operacional foi mencionado pela primeira vez na década de 1990 com a falência do Banco Barings, em Londres. A partir deste acontecimento, os bancos passaram a ter um olhar mais crítico a este tipo de risco, que até então era desconsiderado.

Em setembro de 2001, o Comitê da Basileia definiu o Risco Operacional como sendo: “O risco de perdas resultantes de processos internos falhos ou inadequados, pessoas e sistemas, ou de eventos externos”. Tal Comitê definiu algumas modalidades para a mensuração do Risco Operacional. Dentre elas, se destaca a Metodologia Avançada. Uma importante modalidade desta metodologia é o Modelo de Distribuição de Perdas (Loss Distribution Approach - LDA).

Os tipos de correlação presentes neste modelo, encontrados na literatura, são correlação perfeita, nula e não perfeita. Esta última foi proposta por Frachot et al (2004). O objetivo principal deste trabalho é analisar o cálculo do Capital Regulatório (capital exigido para que instituições financeiras possam se proteger de perdas que envolvam o Risco Operacional) considerando cada uma dessas correlações. A forma de cálculo deste capital também foi definida pelo Comitê da Basileia.

Atualmente, o banco utiliza a correlação perfeita, apesar de estar fora da realidade. Por meio dessa correlação, usamos o método do somatório que superestima o Capital Regulatório. Assim, o total de perdas L do banco é a soma das perdas agregadas para cada classe: linha de negócio \times eventos de perda. No contexto da Basileia, existem 7 linhas de eventos de perdas e 8 linhas de negócios.

Trabalho em conjunto com Guaraci Requena e Carlos Diniz (UFSCar).

Métodos Estatísticos na Análise de Dados Genômicos

Deisy Morselli Gysi
Universidade Federal do Paraná

Abstract: This work is a case-control study of Rheumatoid Arthritis which has available a large amount of genetic Single Nucleotide Polymorphism (SNP) markers. The proposed approach is to investigate which of these SNPs are associated with Rheumatoid Arthritis enabling the identification of SNPs with potential connection with this disease. For this purpose, we consider an adapted logistic regression model with covariates: the amount of risk alleles presents in DRB1, the SNP under investigation, the first six associated principal components and the interaction between the SNP and DRB1, without considering interactions between SNPs. On chromosome 6 were found a number of 166 SNPs associated with Rheumatoid Arthritis (p -value $\cdot 10^{-4}$). Such SNPs are located in HLA (Human Leukocyte Antigen) region, which was expected since Rheumatoid Arthritis is an autoimmune multi-factorial. This is a joint work with Wesley Bertoli da Silva (UFSCar).

Modelo de risco complementar log-log dependente do tempo

Eder Angelo Milani
DEs - Universidade Federal de São Carlos

Abstract: Modelos que não sustentam a suposição de riscos proporcionais são alternativas para ajustar conjuntos de dados que não apresentam a suposição de riscos proporcionais. Neste trabalho propomos uma nova família de modelos de sobrevivência de riscos não proporcionais, a partir da função de risco complementar log-log. Estendemos esta família apresentando os modelos com fragilidade gama e gaussiana inversa; para encontrar a função de sobrevivência não condicional a metodologia utiliza a transformada de Laplace. Um estudo de simulação para verificar as propriedades frequentista dos modelos em questão é exposto. Os modelos desenvolvidos são ajustados ao conjunto de dados de câncer de pulmão e os resultados são comparados com os obtidos nos ajustes dos modelos TDL e TDL com fragilidade gama. Trabalho em conjunto com Carlos Alberto Ribeiro Diniz e Vera L. D. Tomazella (UFSCar).

Modelo de Regressão Bivariado

Flávia Bolssone do Prado
DEs - Universidade Federal de São Carlos

Abstract: Neste trabalho desenvolvemos um modelo de regressão para respostas bivariadas, discreta e contínua, com a variável discreta seguindo distribuição Bernoulli e a variável contínua, condicionada na discreta, seguindo distribuição exponencial. Um procedimento de ajuste, via abordagem Bayesiana, e análise de resíduos Bayesianos são também apresentados. A metodologia é ilustrada através de um estudo de simulação.

Estimação Bayesiana do Número de Erros de Software a partir de um Modelo de Captura-recaptura com heterogeneidade

George Lucas Moraes Pezzott
UFSCar

Abstract: Na área de confiabilidade de software existe o interesse de se conhecer o número N de erros existentes em um software. Neste trabalho, vamos discutir como esta quantidade de interesse pode ser estimada a partir do processo de amostragem conhecido como captura-recaptura. O processo de amostragem considerado consiste em examinar um software, em paralelo, por certo número de revisores, sendo que os erros podem ser identificados por mais de um revisor. O modelo probabilístico considerado é utilizado sob a hipótese de independência e homogeneidade entre os revisores e independência e heterogeneidade entre os erros, na qual esta heterogeneidade se divide em 2 diferentes níveis de dificuldade de detecção: erro fácil ou difícil de se detectar. A partir da criação de uma variável artificial de tal sorte a ajudar na criação do modelo probabilístico, as estimativas foram obtidas via metodologia Bayesiana para média e mediana dos parâmetros. Para fins ilustrativos, comparou-se os resultados com estimativas Bayesianas encontradas por Basu e Ebrahimi (2001) no qual prioris diferentes para N foram adotados. A comparação foi realizada através de dados reais, com dados provenientes na estimação de erros na revisão de interruptores e estimação populacional de coelhos contontail. As estimativas pelos dois modelos foram próximas, e assim podemos considerar que ambos se adequaram bem aos dois tipos de dados. Trabalho em conjunto com José Galvão Leite (USP) e Luis Ernesto Bueno Salasar (UFSCar).

O Capital Regulatório no Risco Operacional Segundo o Método da Convolução com o Uso de Cópulas

Guaraci Requena
DEs - Universidade Federal de São Carlos

Abstract: O Comitê de Supervisão Bancária de Basileia tem como principal objetivo assegurar um nível adequado de capital, ao qual comumente chamamos de capital regulatório, que tem como fim proteger e reforçar a segurança e solidez do sistema financeiro internacional e também minimizar desigualdades competitivas entre as instituições financeiras internacionalmente ativas.

A exposição aos riscos nas instituições financeiras é tão antigo quanto as próprias instituições. Porém, somente em 2004, com o Acordo de Basileia II, se formalizou a preocupação com o Risco Operacional. Esse tipo de risco pode ser visto como a exposição à incerteza ao se realizar um trabalho. Pelo Basileia II, Risco Operacional é “a possibilidade de ocorrência de perdas resultantes de falha, deficiência ou inadequação de processos internos, pessoas e sistemas, ou de eventos externos”.

São exemplos de eventos em Risco Operacional fraudes internas e externas; segurança deficiente no local de trabalho; prática inadequada com clientes ou transações; falhas em sistemas entre outros.

No entanto, a proteção de qualquer risco exige sua mensuração e, por esse motivo, nesse trabalho são estudados dois métodos de cálculo (mensuração) do capital regulatório: o método usual (do somatório) e um método alternativo (da convolução). Ambos os métodos propõem um capital para que as instituições financeiras possam se proteger das perdas operacionais inesperadas. O método do somatório superestima o capital regulatório por considerar as perdas perfeitamente correlacionadas, ou seja, fazendo um paralelo com teoria de cópulas, é como se o método utilizasse implicitamente a cópula limite superior de Fréchet para modelar a dependência entre as variáveis perdas-operacionais. Porém, considerar a correlação perfeita entre as perdas não reflete a realidade e isso motiva o estudo de métodos alternativos. Nesse trabalho estudamos o método da convolução e este capta as relações de dependência entre as variáveis perdas através da integral de convolução a partir da distribuição conjunta construída com base na Teoria de Cópulas, isto é, nesse método há a liberdade de escolher a cópula para captar de maneira mais realística a dependência existente.

Nesse trabalho é utilizada a cópula Gaussiana, com diversos coeficientes de correlação linear, em um estudo de simulação para comparar os resultados numéricos obtidos a partir das duas metodologias mostrando suas implicações, vantagens e desvantagens. Trabalho em conjunto com Débora Delbem e Carlos Diniz (UFSCar).

Modelo de Fragilidade com Fração de Cura na Presença de Covariáveis: Uma Abordagem Paramétrica

João Paulo Taconeli

DEs - Universidade Federal de São Carlos

Abstract: Neste trabalho consideramos modelo de mistura padrão, originalmente proposto por Boag (1949) e Berkson & Cage (1952). Entretanto adicionamos um componente de fragilidade e incorporamos covariáveis tanto na proporção de curados quanto na função de sobrevivência dos não curados. Além disso utilizamos um método de estimação paramétrico, onde o tempo de vida dos indivíduos em risco segue uma distribuição Weibull e a variável de fragilidade uma distribuição gama.

Além disso, a proporção de pacientes não curados foi estimada através de modelos de regressão com funções de ligação logístico, probito e complementar log-log, a fim de dectarmos possíveis diferenças entre estas funções. Através de aplicações em dados reais verificamos as estimativas dos parâmetros dos modelos propostos.

Trabalho em conjunto com Vera L. D. Tomazella e Jhon F. B. Gonzales (UFSCar).

Modelo destrutivo com variável terminal em experimentos quimiopreventivos de tumores em animais

Katherine Elizabeth Coaguila Zavaleta
DEs - Universidade Federal de São Carlos

Abstract: A indução química de substâncias cancerígenas em experimentos quimiopreventivos em animais são cada vez mais frequentes em pesquisas biológicas. O objetivo destes experimentos biológicos é avaliar o efeito de um determinado tratamento na taxa de incidência de tumores em animais. Neste trabalho o número de tumores promovidos por animal será modelado parametricamente seguindo as sugestões dadas em [2],[3] e [5]. O estudo desses experimentos quimiopreventivos será apresentado no contexto do modelo destrutivo proposto por Rodrigues et al.(2010) (vide [4]) com variável terminal que condiciona ou censura o experimento no instante de morte do animal. Os dados analisados possuem uma grande quantidade de zeros, será proposto para o número de tumores promovidos as distribuições: binomial negativa, a distribuição de Poisson com zeros in acionados e a distribuição binomial negativa com zeros in acionados. A seleção destes modelos será feita através do teste da razão de verossimilhança e os critérios AIC, BIC. As estimativas dos respectivos parâmetros serão obtidas utilizando o método de máxima verossimilhança. Trabalho em conjunto com Josemar Rodrigues (DEs-UFSCar) e Vicente Cancho Garibay (ICMC-USP).

Qualidade de ajuste de regressão binária pelas funções de ligação logit e probit considerando diferentes tamanhos de amostras

Leillimar dos Reis Freitas
DEs - Universidade Federal de São Carlos

Abstract: Este trabalho abordou a análise de regressão binária, utilizando das funções de ligação logit e probit, com o objetivo de verificar a robustez das funções de ligação diante da variação do tamanho da amostra. As funções de ligação logit e probit abordadas pela regressão binária utilizam, respectivamente as distribuições logística e normal, cuja principal diferença são os valores de probabilidades nos extremos da variável x . Dentro desse contexto, foram realizadas simulações com 500 repetições utilizando de amostras de 10 diferentes tamanhos, desde 10 a 91, com uma diferença entre as sucessivas amostras de 9 unidades. Com a utilização de medidas de desempenho que foram: percentual de convergência, estimativas do erro quadrático médio da probabilidade geral, estimativas do erro quadrático médio da probabilidade específica, teste Wald para avaliação dos betas, foi possível estabelecer uma recomendação para o uso das duas diferentes funções de ligação quando os dados foram gerados com o uso do logit e probit e analisados por ambas as funções de ligação em cada tamanho de amostra. Concluiu-se que o objetivo desse trabalho foi atingido ao estabelecer uma recomendação para o uso da função de ligação logit para tamanhos inferiores a 20 e logit ou probit para maiores tamanhos de amostras.

Descrição da composição química e atividade antioxidante da própolis orgânica de *Apis mellifera* produzida no Sul do Brasil

Maria da Conceição Farias Freitas Tandel
UNESP

Abstract: A própolis é uma substância resinosa coletada pelas abelhas de diversas partes das plantas, como brotos, botões florais e exudados resinosos, conhecida por várias atividades biológicas como antimicrobiana, anti-inflamatória, antiproliferativa e antioxidante. Sua composição química depende de vários fatores, como a localização geográfica, vegetação e clima. A própolis analisada neste trabalho é proveniente da região Sul do Brasil e possui certificação de produto orgânico, por ser produzida em florestas nativas e áreas de reflorestamento, locais livres de contaminação por insumos agrícolas, metais pesados e poluição industrial. O objetivo deste trabalho é avaliar a composição química e atividade antioxidante desta própolis, considerando a variação sazonal nas estações verão, outono e primavera. Não há coleta no inverno, uma vez que o frio intenso e constante na região desfavorece drasticamente a propolização e a coleta deve ser evitada para que a colméia não seja prejudicada. No período de Fevereiro de 2011 a Janeiro de 2012, foram coletadas 78 amostras provenientes de 14 apiários de 11 apicultores, localizados em 14 municípios: 10 apiários da região sul do Paraná e 4 do norte de Santa Catarina. Foram realizadas análises físico-químicas do extrato etanólico da própolis orgânica. Descreve-se no presente trabalho somente alguns dos vários resultados obtidos pelos diversos métodos empregados: teores de compostos Fenólicos e Flavonoides obtidos por métodos colorimétricos e atividade antioxidante obtida pelos métodos de seqüestro dos radicais livres DPPH_o e ABTS. Pode-se concluir que: a própolis orgânica apresenta, em sua composição química, baixos teores de flavonóides, cuja variabilidade é maior no verão (Levene, $p = 6\%$); os teores de compostos fenólicos apresentam maior variabilidade no outono e maior homogeneidade na primavera (Levene, $p = 0,04\%$); a atividade antioxidante foi detectada segundo os critérios do MAPA: Ministerio da Agricultura, Pecuária e de Abastecimento, sendo que na primavera apresenta atividade antioxidante mais elevada, uma vez que os teores de ABTS são maiores (Kruskal Wallis, $p = 0,3\%$) e é mais homogênea quando medida pelos teores de DPPH* (Levene, $p = 7\%$); os teores de Flavonoides e atividade antioxidante não apresentaram correlação, isso indica que a atividade antioxidante da própolis orgânica está relacionada, provavelmente, com outros compostos fenólicos que não simplesmente aos da classe dos flavonóides; a correlação entre os teores de Fenólicos totais e a atividade antioxidante é de 0,63 se utilizarmos o método do seqüestro de ABTS, apresentando um salto para 0,79 se medida através do seqüestro de DDPH*. O conhecimento da composição química da própolis orgânica e de sua atividade antioxidante contribui para a fixação dos apicultores nos locais de mata nativa, corroborando assim, com a preservação ambiental de um ecossistema que está diretamente envolvido na produção dessa própolis.

Modelos multiníveis simétricos para dados educacionais

Marina M. G. Osio
ICMC - Universidade de São Paulo

Abstract: Os modelos multiníveis podem ser uma alternativa bastante apropriada para analisar dados educacionais que possuem uma estrutura hierárquica, onde podemos considerar que as informações das características relacionadas aos alunos e escola estão, respectivamente, nos primeiro e segundo níveis de hierarquia. Neste trabalho utilizamos um modelo com dois níveis, que pode ser visto como um modelo com efeitos mistos e aplicamos a dados educacionais a fim de investigar a relação do desempenho com alguns dos fatores, disponíveis no conjunto de dados, por exemplo, época de ingresso na vida escolar (nível 1) e infraestrutura da escola (nível 2). Com a finalidade de encontrar o melhor modelo para explicar o rendimento escolar, consideramos além da distribuição normal, as distribuições simétricas t-Student e exponencial potência, que pertencem à classe elíptica e, por atribuir pesos para as observações aberrantes, podem evitar a influência desproporcional das mesmas no cálculo das estimativas dos parâmetros. Trabalho em conjunto com Cibele M. Russo (ICMC-USP) e Bruno Feres de Souza (ICMC-USP).

Analysis of the Generalized Mean Function and Voted Classifier as a Combination Technique in Credit Scoring Models with Bagging

Ricardo Ferreira da Rocha
DEs - Universidade Federal de São Carlos

Abstract: In order to increase the predictive capabilities in credit scoring models, we propose an approach using simulation to analyze the strategies of model combinations. The models are generated using bagging and combined using the generalized mean function and by vote counting in classification models. The analysis will be made in relation to the parameters of the methods, so that you can manipulate in order to get the best results for the measures of major interest. We will show that not always the best performance is given in the standard cases of techniques, i.e., in the combination by ordinary average and by majority vote.

Um modelo semiparamétrico com efeitos aleatórios para dados farmacocinéticos

Robson J. M. Machado

DEs - Universidade Federal de São Carlos

Abstract: Modelos lineares mistos desempenham papel importante para a análise de dados com respostas (contínuas) correlacionadas, tais como, dados longitudinais ou dados agrupados. No entanto, esses modelos apresentam limitações para modelar dados que não apresentam relação linear entre a variável resposta e as covariáveis. Uma alternativa é considerar modelos semiparamétricos e não paramétricos, em que a relação entre a variável resposta e as covariáveis não é totalmente modelada de forma paramétrica. Seguimos Wand (2003), para designar o termo suavização como sendo qualquer modelo de regressão que contenha ao menos uma função modelada de forma não paramétrica. A grande vantagem desse método de modelagem é que o mesmo pode ser visto como uma generalização dos modelos lineares mistos e todos os procedimentos de ajustes de modelos e inferências, podem ser realizados nos softwares clássicos de estatística como o R e o SAS. Consideramos nesse trabalho um dos métodos existentes para construir esses modelos no contexto de modelos de regressão com efeitos mistos, sobre os pressupostos de normalidade dos erros envolvidos e dos efeitos aleatórios, apresentando métodos de estimação dos parâmetros e predição dos efeitos aleatórios. Para ilustrar a utilização desses modelos, apresentamos a aplicação realizada para estudar a cinética do anti-asmático Theophylline, em que diferentes doses foram administradas a dose pacientes e 11 concentrações séricas desse medicamento foram coletadas nas 25 horas seguintes. Trabalho em conjunto com Cibele M. Russo (ICMC-USP).

Practical aspects of the estimation of mixture model via Dirichlet Process

Rosineide F. da Paz

DEs - Universidade Federal de São Carlos

Abstract: We review the Dirichlet process mixture model and investigate its performance as a classification method. The first aspect considered is its sensibility to the choice of location parameter of base distribution. The second aspect consider the performance of the model regarding the departure of the parameters of the component distributions. Simulation results with mixture of normal distributions indicate sensibility to location parameters choices and good performance even when component normal distributions differ only in variances. Finally, we apply the method to two data sets. Joint work with Luís A. Milan (UFSCar).

Uma Abordagem de EM para Modelo de Mistura

Silvana Aparecida Meira
DEs - Universidade Federal de São Carlos

Abstract: Os modelos que utilizam misturas de distribuições têm sido empregados em várias áreas do conhecimento, tais como na astronomia, biologia, genética, medicina, economia, engenharia, entre outras áreas da ciência. Neste trabalho consideramos que variável aleatória observada $Y = (Y_1, \dots, Y_T)$ é dependente da variável aleatória não observável $S = (S_1, \dots, S_T)$ na composição do modelo de mistura em que S associa Y a uma componente de mistura. Para a estimação dos parâmetros do modelo de mistura utilizamos o algoritmo EM com uma modificação no procedimento de estimação e utilizamos os critérios de informação AIC e BIC para determinar o número de componentes de mistura do modelo. Para aplicação do modelo de mistura utilizamos conjuntos de dados reais. Trabalho em conjunto com Luis Aparecido Milan (UFSCar).

Estimação da distribuição a posteriori do modelo Poisson bivariado a partir de dados ampliados

Verônica Amparo Quispe Chire
DEs - Universidade Federal de São Carlos

Abstract: Neste trabalho, o método iterativo para obtenção da distribuição a posteriori proposto por Tanner e Wong será aplicado ao modelo Poisson bivariado de Holgate. O algoritmo baseia-se na ampliação dos dados observados a partir de dados latentes de tal forma que os dados ampliados são mais simples de se analisar. O algoritmo será ilustrado para um conjunto de dados reais.

Estimação dos Parâmetros do Modelo Poisson Bivariado via Máxima Verossimilhança

Wesley Bertoli da Silva
DEs - Universidade Federal de São Carlos

Abstract: A modelagem de dados provenientes de contagens pareadas é um tópico com abordagem frequente na literatura estatística, sendo a distribuição Poisson bivariada uma das alternativas de modelagem mais utilizadas. Neste trabalho, abordamos a estimação dos parâmetros do modelo Poisson bivariado utilizando o método de máxima verossimilhança. Uma vez que a obtenção dos estimadores não é trivial através dos métodos usuais, o uso de relações de recorrência para o modelo Poisson bivariado torna essa tarefa algo factível. Nesse contexto, consideramos um estudo de simulação para validar o método de estimação, e verificou-se que, conforme o tamanho amostral cresce, o viés e o erro quadrático médio das estimativas de máxima verossimilhança se aproximam de zero. Além disso, consideramos uma aplicação a um levantamento socioeconômico alemão realizado entre os anos de 1984 e 1994. Os parâmetros foram estimados via máxima verossimilhança e, através do método Bootstrap, foram obtidos os intervalos de confiança estimados para cada parâmetro. No mais, verificou-se que o modelo Poisson bivariado mostrou-se o mais indicado para a modelagem dos dados considerados. Trabalho em conjunto com Luis Ernesto Bueno Salazar (UFSCar).

Critérios clássicos e bayesianos de seleção de modelos para a classe de modelos de regressão Poisson inflacionado de zero

Willian Luís de Oliveira
DEs - Universidade Federal de São Carlos

Abstract: Neste trabalho, consideramos o modelo de regressão Poisson inflacionado de zero, utilizado com bastante frequência em dados de contagem com excesso de zero. A estimação dos parâmetros de interesse foi estudada por meio de uma abordagem clássica e bayesiana. Além disso, critérios clássicos e bayesianos de seleção de modelos foram aplicados para o modelo em questão. Por fim, um estudo de simulação foi realizado com o objetivo de analisar o comportamento dos critérios apresentados, para algumas situações predeterminadas. Toda parte computacional foi desenvolvida em linguagem R.

