

4th Workshop on Probabilistic and Statistical Methods

February 1-3, 2016

DEs-UFSCar, São Carlos, SP, Brazil

Invited Speakers

Alexandra Schmidt
Universidade Federal do Rio de Janeiro (UFRJ)

Anders Tolver
University of Copenhagen

Cristian Bayes
Pontificia Universidad Católica del Perú

Florencia Leonardi
Universidade de São Paulo (USP)

Glauco Valle
Universidade Federal do Rio de Janeiro (UFRJ)

Pablo Groisman
Universidad de Buenos Aires

Patricia Giménez
Universidad Nacional de Mar del Plata

Scientific Committee

Luiz Renato Fontes
Instituto de Matemática e Estatística (IME/USP)

Francisco Louzada Neto
Instituto de Ciências Matemáticas e de Computação (ICMC/USP)

Nancy Garcia
Instituto de Matemática, Estatística e
Computação Científica (IMECC/UNICAMP)

Debajyoti Sinha
Florida State University

Organizing Committee

Juliana Cobre
Instituto de Ciências Matemáticas e de Computação (ICMC/USP)

Reiko Aoki
Instituto de Ciências Matemáticas e de Computação (ICMC/USP)

Renato Gava
Universidade Federal de São Carlos (UFSCar)

Sandro Gallo
Universidade Federal de São Carlos (UFSCar)

Minicourse

Dose Finding with Escalation with
Overdose Control in Cancer Clinical Trials
André Rogatko (Samuel Oschin Comprehensive Cancer Institute for Biostatistics)

Further information and registration: <http://estatisticaverao.icmc.usp.br/>

ORGANIZED BY

Programa Interinstitucional de Pós-graduação em Estatística - PIPGEs



Departamento de Matemática
Aplicada e Estatística
Instituto de Ciências Matemáticas e de Computação
Universidade de São Paulo - USP



Departamento de Estatística
Universidade Federal de São Carlos - UFSCar

SPONSORED BY



4rd Workshop on Probabilistic and Statistical Methods

February 1–3, 2016

UFSCar, São Carlos, SP, Brazil

PROGRAM

ICMC/USP and DEs/UFSCar

Organized by



Departamento de Matemática Aplicada e Estatística – SME
Instituto de Ciências Matemáticas e de Computação – ICMC
Universidade de São Paulo – USP



Departamento de Estatística – DEs
Universidade Federal de São Carlos – UFSCar

Sponsored by



Associação Brasileira de Estatística



Fundação Coordenação de Aperfeiçoamento
de Pessoal de Nível Superior



Conselho Nacional de Desenvolvimento
Científico e Tecnológico



Fundação de Amparo à Pesquisa
do Estado de São Paulo

About the 4th Workshop on Probabilistic and Statistical Methods

UFSCar, São Carlos, SP, Brazil

February 1–3, 2016

The Workshop on Probabilistic and Statistical Methods is an initiative of the Programa Interinstitucional de Pós-graduação em Estatística (PIPGEs ICMC/USP and UFSCar), which brings together the Statistics and Probability research groups from ICMC/USP and UFSCar, in São Carlos, SP, Brazil.

This meeting intends to discuss new developments in Statistics, Probability and their applications. Activities include invited plenary sessions, short talks, a poster session and two short courses. The topics of this meeting include Probability and Stochastic Processes, Statistical Inference, Regression Models, Survival Analyses and Complex Stochastic Systems. The aim of the workshop is to provide a unique opportunity for researchers and students to exchange experiences and start collaborations.

Invited Speakers

Anders Tolver, University of Copenhagen, Denmark

Danilo Lourenço Lopes, UFSCar, Brazil

Flávio Bambirra Gonçalves, UFMG, Brazil

Florencia Leonardi, IME-USP, Brazil

Glauco Valle, UFRJ, Brazil

Pablo Groisman, UBA, Argentine

Rafael Bassi Stern, UFSCar, Brazil

Minicourses

André Rogatko, Samuel Oschin Comprehensive Cancer Institute, USA

Scientific Committee

Debajyoti Sinha, FSU, USA

Francisco Louzada Neto, USP, Brazil

Luiz Renato Fontes, IME-USP, Brazil

Nancy Lopes Garcia, UNICAMP, Brazil

Organizing Committee

Juliana Cobre, ICMC/USP (chair)

Reiko Aoki, ICMC/USP

Renato Gava, UFSCar

Sandro Gallo, UFSCar (chair)

Vera Tomazella, UFSCar

4th Workshop on Probabilistic and Statistical Methods

February 1–3, 2016

UFSCar, São Carlos, SP, Brazil

SCHEDULE

ICMC/USP and DEs/UFSCar

4th Workshop on Probabilistic and Statistical Methods

February 1–3, 2016 – UFSCar, São Carlos, SP, Brazil

Time	Monday (2/1)	Tuesday (2/2)	Wednesday (2/3)
08h00 - 09h00	Registration		
09h00 - 10h40	MC: Rogatko	MC: Rogatko	MC: Rogatko
10h40 - 11h00	<i>Coffee Break</i>	<i>Coffee Break</i>	<i>Coffee Break</i>
11h00 - 12h00	Talk: Groisman	Talk: Valle	Talk: Tolver
12h00 - 14h00	—Lunch—	—Lunch—	—Lunch—
14h00 - 15h00	Talk: Stern	Talk: Lopes	
15h00 - 16h00	<i>Coffee Break & Posters</i>	<i>Coffee Break & Posters</i>	
16h00 - 17h00	Short Talks 1 and 2	Short Talks 3 and 4	
17h00 - 18h00	Talk: Gonçalves	Talk: Leonardi	

MC: Mini course.
Short Talks: 20 minutes each.

Mini course

- André Rogatko: *Dose Finding with Escalation with Overdose Control in Cancer Clinical Trials.*

Talks

- Flávio Bambirra Gonçalves- *Exact Bayesian Inference in Spatio-temporal Cox Processes Driven by Multivariate Gaussian Processes*
- Anders Tolver - *On the Variation of 3D-acceleration Signals of Trotting Horses*
- Danilo Lourenço Lopes - *Implementation of Censored Regression Models for Survival Analysis*
- Florencia Leonardi - *Change Point Detection for High-dimensional Regression Data with l_1 -regularization*
- Glauco Valle - *Scaling Limit of the Radial Poissonian Web*
- Pablo Groisman - *Branching-selection Particle Systems, Travelling Waves and Quasi-stationary Distributions*
- Rafael Bassi Stern - *A Statistical Contribution to Historical Linguistics*

Short Talks

1. Luca Martino - *Effective Sample Size for Importance Sampling based on the Discrepancy Measures*
2. Mauro R. de Oliveira Jr. - *Zero-inflated Cure Rate Regression Models for Time-to-default with Applications*
3. Willian Luís de Oliveira - *A Class of Bivariate Regression Models for Mixed Discrete and Continuous Responses.*
4. Natalia Oliveira - *The Homogeneity Test for 2×2 Contingency Table.*

Conferences

1. Anders Tolver, University of Copenhagen, Denmark
On the Variation of 3D-acceleration Signals of Trotting Horses
2. Danilo Lourenço Lopes, UFSCar, Brazil
Implementation of Censored Regression Models for Survival Analysis
3. Flávio Bambirra Gonçalves, UFMG, Brazil
Exact Bayesian Inference in Spatio-temporal Cox Processes Driven by Multi-variate Gaussian Processes
4. Florencia Leonardi, IME-USP, Brazil
Change Point Detection for High-dimensional Regression Data with l_1 -regularization
5. Glauco Valle, UFRJ, Brazil
Scaling Limit of the Radial Poissonian Web.
6. Pablo Groisman, UBA, Argentine
Branching-selection Particle Systems, Travelling Waves and Quasi-stationary Distributions.
7. Rafael Bassi Stern, UFSCar, Brazil
A Statistical Contribution to Historical Linguistics [0.5cm]

Minicourse

1. André Rogatko, Samuel Oschin Comprehensive Cancer Institute, USA
Dose Finding with Escalation with Overdose Control in Cancer Clinical Trials.

Oral Communications

1. Luca Martino, Victor Elvira and Francisco Louzada
ICMC/USP, Brazil, University Carlos III de Madrid, Spain, and ICMC/USP, Brazil
Effective Sample Size for Importance Sampling Based on the Discrepancy Measures
2. Mauro R. de Oliveira Jr., Francisco Louzada and Fernando F. Moreira
PIPGEs UFSCar ICMC/USP, ICMC/USP, Brazil and Credit Research Centre, University of Edinburgh Business School, Scotland, UK
Zero-inflated Cure Rate Regression Models for Time-to-default with Applications
3. Natalia Oliveira, Marcio Diniz and Adriano Polpo
PIPGEs UFSCar ICMC/USP, UFSCar and UFSCar, Brazil
The Homogeneity Test for 2x2 Contingency Table

4. Willian Luís de Oliveira and Carlos A. R. Diniz
PIPGEs UFSCar ICMC/USP and UFSCar, Brazil
A Class of Bivariate Regression Models for Mixed Discrete and Continuous Responses.

Poster Session

1. Alexandre C. Maiorano, Francisco Louzada and Jorge L. Bazán
PIPGEs UFSCar ICMC/USP, ICMC/USP and ICMC/USP, Brazil
Eco do Números - Módulo Introdutório em Estatística
2. Aline Campos Reis de Souza and Vicente Garibay Cancho
PIPGEs UFSCar ICMC/USP and ICMC/USP, Brazil
Análise Bayesiana Objetiva para o Modelo de Regressão Linear Heteroscedástico com Erros t-Student
3. Amanda Morales Eudes D'Andrea, Vera Luícia Damasceno Tomazella and Cirdêmia Costa Feitosa
PIPGEs UFSCar ICMC/USP, UFSCar and UFSCar, Brazil
Modelagem de Fragilidade em Sistemas Reparáveis sob a Suposição de Reparo Mínimo
4. Átila Prates Correia, USP, Brazil
Complex Behavior Probability Function
5. Elizbeth Chipa Bedia and Pablo M. Rodriguez
PIPGEs UFSCar ICMC/USP and ICMC/USP, Brazil
Connectivity Threshold for the Erdős-Rényi Random Graph
6. Glauber Márcio Silveira Pereira and Carlos Alberto Ribeiro Diniz
PIPGEs UFSCar ICMC/USP and UFSCar, Brazil
Transmuted Logistic II Distribution
7. Karina Emboaba and Cristian Coletti and Pablo M. Rodriguez.
ICMC/USP, UFABC and ICMC/USP, Brazil
A Stochastic Two-stage Innovation Diffusion Model on a Lattice
8. Lorena Yanet Cáceres Tomaya and Mário de Castro
PIPGEs UFSCar ICMC/USP and ICMC/USP, Brazil
Estimation in a General Skew-t Measurement Error Model
9. Mariele Parteli Florencio and Renato Jacob Gava
PIPGEs UFSCar ICMC/USP and ICMC/USP, Brazil
Calculando o Valor Esperado na Ocorrência de Uma Palavra

10. Mauro R. de Oliveira Jr., Francisco Louzada and Fernando F. Moreira
PIPGEs UFSCar ICMC/USP, ICMC/USP, Brazil, and University of Edinburgh
Business School, Scotland
The Zero-inflated Promotion Cure Rate Regression Model Applied to Fraud
Propensity in Bank Loan Applications
11. Michel Bessani, Rodrigo Zempulski Fanucchi, Jorge Alberto Achcar, Carlos
Dias Maciel
EESC/USP, Copel Distribution S/A, FMRP/USP and EESC/USP, Brazil
A Statistical Analysis of Outage Data from a Real Power Distribution System
12. Ricardo Felipe Ferreira and Dorival Leão Pinto Júnior
PIPGEs UFSCar ICMC/USP and ICMC/USP, Brazil
An Euler-Maruyama Approach for Cox-Ingersol-Ross Process
13. Rodrigo Lambert and Miguel Abadi
FAMAT/UFU, IME/USP, Brazil
The Shortest-path Random Variable
14. Taciana K. O. Shimizu, Francisco Louzada and Adriano K. Suzuki
PIPGEs UFSCar ICMC/USP, ICMC/USP, Brazil, and ICMC/USP, Brazil
Compositional Regression Model with Volleyball Data: A Bayesian Approach
15. Taís Roberta Ribeiro and Adriano Kamimura Suzuki
PIPGEs UFSCar ICMC/USP and ICMC/USP, Brazil
The AMH Bivariate Lifetime Copula Model: Modeling, Bayesian Estimation
and Case Influence Diagnostic
16. Wesley Bertoli da Silva and Deisy Morselli Gysi
Universidade Tecnológica Federal do Paraná, Brazil, and University of Leipzig,
Germany
Bayesian Estimation of the Zero-Inflated Quasi Poisson-Lindley Model

4th Workshop on Probabilistic and Statistical Methods

February 1–3, 2016

UFSCar, São Carlos, SP, Brazil

ABSTRACTS

ICMC/USP and DEs/UFSCar

Minicourse

Dose Finding with Escalation with Overdose Control in Cancer Clinical Trials

André Rogatko

Samuel Oschin Comprehensive Cancer Institute, USA

Abstract Escalation With Overdose Control (EWOC) is a Bayesian adaptive dose finding design that produces consistent sequences of doses while controlling the probability that patients are overdosed. EWOC was the first dose-finding procedure to directly incorporate the ethical constraint of minimizing the chance of treating patients at unacceptably high doses. Its defining property is that the expected proportion of patients treated at doses above the maximum tolerated dose (MTD) is equal to a specified value α , the feasibility bound. This value is selected by the clinician and reflects his/her level of concern about overdosing. Among designs with this defining property, EWOC minimizes the average amount by which patients are underdosed. This means that EWOC approaches the MTD as rapidly as possible, while keeping the expected proportion of patients overdosed less than the value α . As a trial progresses, the dose sequence defined by EWOC approaches the MTD (i.e., the sequence of recommended doses converges in probability to the MTD). Eventually, all patients beyond a certain time would be treated at doses sufficiently close to the MTD. Topics discussed in the course include: two-parameter logistic model, correlated priors on ρ_0 and y , varying feasibility bound, cohort size, sample size determination, R115777 trial, use of covariate in prospective clinical trial, PNU trial, use of ordinal toxicity grades, dose finding beyond phase I, designing a trial with EWOC, and new extensions: time to toxicity and drug combinations. The basic text, that will be provided to the participants, is Rogatko and Tighiouart (2015) [Dose Finding with Escalation with Overdose Control (EWOC) in Cancer Clinical Trials, pp 274–305, in *Clinical Trial Biostatistics and Biopharmaceutical Applications*, ed. W. Young and D. Chen, Chapman and Hall, Boca Raton]. We will also learn how to design trials using Web-EWOC, a web-based calculator <https://biostatistics.csmc.edu/ewoc/ewocWeb.php>. The only requirement to use Web-EWOC is access to a standard web browser and all computations are performed on our web server. The tutorial to be followed can be downloaded from the same site: Rogatko and Tighiouart (2013) [Designing a Dose Finding Trial using EWOC].

References

1. Babb J, Rogatko A, Zacks S. 1998. Cancer phase I clinical trials: efficient dose escalation with overdose control. *Statistics in Medicine* 17:1103–1120.
2. Zacks S, Rogatko A, Babb J. 1998. Optimal Bayesian-feasible dose escalation for cancer phase I trials. *Stat. and Prob. Ltrs.* 38:215–220.
3. Babb JS, Rogatko A. 2001. Patient specific dosing in a cancer phase I clinical trial. *Stat. Med.* 20: 2079–2090.
4. Rogatko A, Babb JS, Tighiouart M, Khuri FR, Hudes G. 2005. New paradigm in dose-finding trials: patient-specific dosing and beyond ?phase I?. *Clinical Cancer Research*, 11(15): 5342–6.
5. Tighiouart M, Rogatko A, Babb JS. 2005. Flexible Bayesian methods for cancer phase I clinical trials. Dose escalation with overdose control. *Statistics in Medicine*, 24: 2183–2196.
6. Xu Z, Tighiouart M, Rogatko A. 2007. EWOC 2.0: Interactive Software for Dose Escalation in Cancer Phase I Clinical Trials. *Drug Information Journal*, 41(2): 221–228.
7. Rogatko A, Ghosh P, Vidakovic B, Tighiouart M. Patient-Specific Dose Adjustment in the Cancer Clinical Trial Setting. *Pharmaceutical Medicine*. 2008; 22(6): 345–50
8. Tighiouart M, Rogatko A. 2010. Dose Finding with Escalation with Overdose Control (EWOC) in Cancer Clinical Trials. *Statistical Science* 25(2): 217–226.
9. Tighiouart M, Cook-Wiens G, Rogatko A. 2012. Incorporating a Patient Dichotomous Characteristic in Cancer Phase I Clinical Trials Using Escalation with Overdose Control, *Journal of Probability and Statistics*, vol. 2012, Article ID 567819, 10 pages.
10. Tighiouart M, Rogatko A. 2012. Number of Patients per Cohort and Sample Size Considerations using Dose Escalation with Overdose Control. *Journal of Probability and Statistics*, vol. 2012 , Article ID 692725, 16 pages.
11. Tighiouart M, Cook-Wiens G, Rogatko A. 2012. Escalation with Overdose Control using Ordinal Toxicity Grades for Cancer Phase I Clinical Trials. *Journal of Probability and Statistics*, vol. 2012, Article ID 317634, 18 pages.
12. Tighiouart M, Liu Y, Rogatko A. Escalation with Overdose Control using Time to Toxicity for Cancer Phase I Clinical Trials. *PLoS ONE* 9(3): e93070.
13. Tighiouart M, Piantadosi S, Rogatko A. 2014. Dose-Finding with drug combinations in cancer phase I clinical trials using conditional escalation with overdose control (EWOC). *Statistics in Medicine*, 33(22): 3815–3829.

Conferences

On the variation of 3D-acceleration signals of trotting horses

Anders Tolver

University of Copenhagen, Denmark

Abstract Evaluation of the movement during trot in circles is commonly used in clinical lameness examinations of horses. Lameness as well as the size of the circle is known to affect the symmetry of the movement. Functional observations of accelerations were obtained for a group of sound horses during trot on a straight line and on circles of varying diameter in both directions. In order to describe the different sources of variation in the acceleration signals both the horse-to-horse variation and the possibly non-linear impact of the diameter of the circle must be taken into account. Further, the physical nature of the biomechanical system for quadropedal trot imposes many restrictions on the class of possible idealized acceleration signals that must be build into any reasonable statistical model for the data. Some practical questions of particular interest are: i) do horses have a preferred direction for trot in circles that is reflected in the symmetry of the acceleration signals, ii) at what circle size is the movement comparable to trot on a straight line?

Implementation of Censored Regression Models for Survival Analysis

Danilo Lourenço Lopes

UFSCar, Brazil

Abstract We study the Bayesian censored modeling of several probability distributions that are widely used in reliability and survival analysis. We discuss the implementation a new package for the free software environment R, which provides routines for the Bayesian estimation of several distributions recently introduced in the statistical literature. For some of these distributions, a new parametrization is presented in order to simplify the estimation process. We also perform a simulation study to analyze the quality of the developed procedures. Finally, we discuss on some problems related to the estimation of such complex parametric models.

Exact Bayesian Inference in Spatio-temporal Cox Processes Driven by Multivariate Gaussian Processes

Flávio Bambirra Gonçalves
UFMG, Brazil

Abstract In this paper we present a novel inference methodology to perform Bayesian inference for spatio-temporal Cox processes where the intensity function depends on a multivariate Gaussian process. Dynamic Gaussian processes are introduced to allow for evolution of the intensity function over discrete time. The novelty of the method lies on the fact that no discretisation error is involved despite the non-tractability of the likelihood function and infinite dimensionality of the problem. The method is based on a Markov chain Monte Carlo algorithm that samples from the joint posterior distribution of the parameters and latent variables of the model. A particular choice of the dominating measure to obtain the likelihood function is shown to be crucial to devise a valid MCMC. The models are defined in a general and flexible way but they are amenable to direct sampling from the relevant distributions, due to careful characterisation of its components. The models also allow for the inclusion of regression covariates and/or temporal components to explain the variability of the intensity function. These components may be subject to relevant interaction with space and/or time. Simulated examples illustrate the methodology.

Change Point Detection for High-dimensional Regression Data with l_1 -regularization

Florencia Leonardi
IME-USP, Brazil

Abstract Sequential data is often characterized by some degree of inhomogeneity, translated in sudden changes in the parameters defining the process that generates the data. In this work we consider the problem of detecting such structural changes for high dimensional data in a regression setting. We propose a joint estimator of the locations of the change points and the different parameters for the segments, and prove its consistency even in the case of an unknown number of segments. We evaluate the performance of the proposed estimator on simulated data and we apply the methodology to a real dataset. This is a joint work with Peter Bühlmann (ETHZ).

Scaling Limit of the Radial Poissonian Web

Glauco Valle
UFRJ, Brazil

Abstract We consider a variant of the radial spanning tree introduced by Baccelli and Bordenave. Like the original model, our model is a tree rooted at the origin, built

on the realization of a planar Poisson point process. Unlike it, the paths of our model have independent jumps. We show that locally our diffusively rescaled tree, seen as the collection of the paths connecting its sites to the root, converges in distribution to the *Brownian Bridge Web*, which is roughly speaking a collection of coalescing Brownian bridges starting from all the points of a planar strip perpendicular to the time axis, and ending at the origin.

Branching-selection Particle Systems, Travelling Waves and Quasi-stationary Distributions

Pablo Groisman

Universidad de Buenos Aires, Argentine

Abstract Branching-selection particle systems are Markov processes that are useful to model, from a qualitative point of view, the microscopic behavior of gene evolution in the presence of natural selection (among other things). In these models a front that propagates can be observed both macro and microscopically. In the nineties Brunet, Derrida and coauthors introduced a family of processes in this family to study the shift in the velocity of the front due to the microscopic effects (compared with the macroscopic behavior). We will discuss some of these models. Properties, hydrodynamic limits, selection principles and the relation with conditioned evolution and quasi-stationary distributions.

A Statistical Contribution to Historical Linguistics

Rafael Bassi Stern

UFSCar, Brazil

Abstract Historical Linguistics studies language change over time. If a group of languages derives from changes to a common ancestor language (proto-language) then they are said to be related. Whenever there exists a lack of written records for an ancestor language, a relevant question in Historical Linguistics is to determine whether two languages are related. However, it is hard to find ancient relationships between languages. Given the above challenge, I propose a new (probability) model for the evolution of languages. A relevant innovation of this model is that it captures the regularity of sound changes. I present a new algorithm used to compute the probability of linguistic hypotheses regarding language relationships and the occurrence of regular sound changes. This algorithm is a new variant of Nested Sequential Monte Carlo. I prove the consistency of the algorithm by showing that it is a case of Sequential Importance Resampling. The algorithm is illustrated with tests on Indo-European languages.

Oral Communications

Effective Sample Size for Importance Sampling based on the Discrepancy Measures

Luca Martino
ICMC/USP, Brazil

Abstract The Effective Sample Size (ESS) is an important measure of efficiency of Monte Carlo methods, such as, Markov Chain Monte Carlo (MCMC) and Importance Sampling (IS) techniques. ESS is defined as the equivalent number of independent samples generated directly from the target distribution, which yields the same Monte Carlo efficiency in the estimation. Namely, ESS is theoretically related to the ratio between the variance of the direct Monte Carlo estimator (drawing samples directly from the target) over the variance of the estimator obtained by another Monte Carlo technique, using both with the same number of samples. For IS, in literature, it is common to approximate this theoretical definition with the formula $ESS \approx \frac{1}{\sum_{n=1}^M \bar{w}_n^2}$ where \bar{w}_n are the normalized weights. This expression, obtained after several rough approximations, presents different weaknesses, listed and discussed in this work. However, it can be easily applied and interpreted as a measure of discrepancy between the multinomial probability mass function (pmf) \bar{w}_n , with $n = 1, \dots, N$, with the corresponding discrete uniform pmf, $\frac{1}{N}$, for $n = 1, \dots, N$. In this work, we remark that the expression $ESS \approx \frac{1}{\sum_{n=1}^M \bar{w}_n^2}$ is related to the Euclidean distance between these two pmfs and deduce different other ESS functions based on different discrepancy measures. Furthermore, we define a generic proper ESS function, designing different conditions which should be fulfilled. Several comparisons among the ESS functions are provided, by means of theoretical considerations and numerical results. Joint work with V. Elvira and F. Louzada.

Zero-inflated Cure Rate Regression Models for Time-to-default with Applications

Mauro R. de Oliveira Jr.
PIPGEs UFSCar ICMC/USP, Brazil

Abstract In this talk, we present a methodology based on zero-inflated long-term survival data in order to deal with fraud rate estimation in bank loan portfolios. Our approach enables us to accommodate three different types of loan borrowers,

i.e., fraudsters, those who are susceptible to default and finally, those who are not susceptible to default. Regarding to the survival analysis framework, an advantage of our approach is to accommodate zero-inflated times, which is not possible in the standard cure rate model introduced by Berkson & Gage. To illustrate the proposed method, a real dataset of loan survival times is fitted by the zero-inflated Weibull cure rate model. In this talk, we present a methodology based on zero-inflated long-term survival data in order to deal with fraud rate estimation in bank loan portfolios. Our approach enables us to accommodate three different types of loan borrowers, i.e., fraudsters, those who are susceptible to default and finally, those who are not susceptible to default. Regarding to the survival analysis framework, an advantage of our approach is to accommodate zero-inflated times, which is not possible in the standard cure rate model introduced by Berkson & Gage. To illustrate the proposed method, a real dataset of loan survival times is fitted by the zero-inflated Weibull cure rate model. Joint work with Francisco Louzada and Fernando F. Moreira.

The Homogeneity Test for 2×2 Contingency Table

Natalia Oliveira

PIPGEs UFSCar ICMC/USP, Brazil

Abstract Using the likelihood ratio statistics, we developed a significance index, called P-value, to test the hypothesis of homogeneity in 2×2 contingency tables. The p -value does not depend on asymptotic distributions, and is based on the elimination of the nuisance parameter. That is, we proposed a way to obtain the exact distribution of the likelihood ratio statistics. This procedure is compatible with the likelihood principle. To a better understanding of significance indices to test homogeneity, we performed a simulation study comparing some frequentist indices (likelihood ratio test, chi-square test) and the full Bayesian significance test (FBST). The comparative study showed an interesting relation between all the indices studied, Bayesian and frequentists. Joint work with Marcio Diniz and Adriano Polpo.

A Class of Bivariate Regression Models for Mixed Discrete and Continuous Responses

Willian Luís de Oliveira

PIPGEs UFSCar ICMC/USP, Brazil

Abstract In this work, a wide general class of models for mixed responses is proposed in which joint distributions are constructed by the conditional approach (probability density functions (pdf), as the product of a marginal pdf and a conditional pdf). It is assumed that the distribution of the discrete response and the conditional distribution of the continuous response given the discrete variable belong to one- or two-parameter

exponential family distributions. Furthermore, the marginal means are related to the covariates by link functions using linear and/or nonlinear predictors and a dependency structure between the responses is inserted into the model via the conditional mean. Estimation methods, diagnostic analysis and influence techniques are presented as well as a simulation study considering the Bernoulli-exponential, a particular case of the proposed class. Finally, one of the proposed models is used in a real data set. Joint work with Carlos A. R. Diniz.

Posters

Eco do Números - Módulo Introdutório em Estatística

Alexandre C. Maiorano, Francisco Louzada and Jorge L. Bazán
PIPGEs UFSCar ICMC/USP, ICMC/USP and ICMC/USP, Brazil

Abstract Este trabalho apresenta um módulo online voltado à divulgação da Estatística para professores e alunos do ensino médio. O sistema desenvolvido apresenta conceitos divididos em três diferentes temas: medidas descritivas, gráficos e distribuições. Além da introdução do conceito teórico, são apresentados um exemplo de aplicação e o cálculo dinâmico dessas medidas a partir de um conjunto de observações escolhido pelo próprio usuário. Uma característica importante desse sistema é que foi desenvolvido a partir da linguagem Shiny, um framework em R voltado para aplicações web que possibilita realizações de cálculos instantâneos a partir da relação entre o usuário e o software R instalado no servidor em que o site está hospedado.

Análise Bayesiana Objetiva para o Modelo de Regressão Linear Heteroscedástico com Erros t-Student

Aline Campos Reis de Souza and Vicente Garibay Cancho
PIPGEs UFSCar ICMC/USP and ICMC/USP, Brazil

Abstract Neste trabalho, apresentamos uma extensão da análise bayesiana objetiva feita em Fonseca *et al.* (2008), baseada nas distribuições *a priori* de Jeffreys para o modelo de regressão linear com erros t-Student, considerando a suposição de heteroscedasticidade. Mostramos que a distribuição *a posteriori* gerada pela distribuição *a priori* que propomos é própria. Através de estudo de simulação, analisamos as propriedades frequentistas dos estimadores bayesianos obtidos e testamos a robustez do modelo através de perturbações na variável resposta, comparando seu desempenho com o de outros modelos encontrados na literatura. Finalmente, um conjunto de dados reais é utilizado para o ajuste do modelo proposto. Detectamos os possíveis pontos influentes através da medida de divergência de Kullback-Leibler e utilizamos os critérios de seleção DIC e LPML para comparação de modelos.

Modelagem de Fragilidade em Sistemas Reparáveis sob a Suposição de Reparo Mínimo

Amanda Morales Eudes D’Andrea, Vera Lúcia Damasceno Tomazella and Cirdêmia Costa Feitosa
PIPGEs UFSCar ICMC/USP, UFSCar and UFSCar, Brazil

Abstract Em sistemas reparáveis, um aspecto fundamental a ser considerado é prever a confiabilidade dos sistemas em estudo. Porém, segundo Vaupel et al. (1979), os métodos padrões em análise de dados de sistemas reparáveis ignoram a heterogeneidade não observada. Assim, neste trabalho são explorados os modelos de fragilidade, caracterizados pela utilização de um efeito aleatório, isto é, uma variável aleatória não observável que representa as informações que não puderam ou não foram observadas. O método inferencial para estimação dos parâmetros será apresentado para os modelos com reparo mínimo. Por fim, foi feita uma aplicação a dados reais (Toledo, 2014).

Complex Behavior Probability Function

Átila Prates Correia
USP, Brazil

Abstract To begin with, let us explain what is meant by the title. The probability function’s predicate “complex behavior” refers to the fact we are interested in describing the exchange probability between states which are represented by complex numbers. Particularly, in the present poster, it has been chosen the political environment in order to contextualize the theory to be introduced. The main idea basically reduces to sampling an odd number of times (different populations) and rearranging the results into (equally odd) “partisan vectors”, so to speak, where each of them represents one political party. Thereafter their coordinates are conveniently mapped into complex numbers of $S^1 = \{z \in \mathbb{C} | z\bar{z} \leq 1\}$, from which we build the probability function itself. At last, it is worth mentioning these results come from previous studies done at [1] and [2].

Connectivity Threshold for the Erdős-Rényi Random Graph

Elizabeth Chipa Bedia and Pablo M. Rodriguez
PIPGEs UFSCar ICMC/USP and ICMC/USP, Brazil

Abstract We say that a random graph is connected if there exists a path of open edges between any two pairs of vertices. In this poster, we study a connectivity threshold on the edge probability (probability to join two vertices by an edge) for the Erdős-Rényi random graph. A classical reference concerning random graphs, and containing the results that are presented here is the book of Bela Bollobas [2]. The proofs are taken from the book of Remco Van Der Hofstad [1].

Transmuted Logistic II Distribution

Glauber Márcio Silveira Pereira and Carlos Alberto Ribeiro Diniz
PIPGEs UFSCar ICMC/USP and UFSCar, Brazil

Abstract A generalization of the generalized inverse Logistic II distribution so-called transmuted generalized inverse Logistic II distribution is proposed and studied. The studies we have done: calculate the first and second time, set the survival function, risk function and a study of statistical order.

A Stochastic Two-stage Innovation Diffusion Model on a Lattice

Karina Emboaba and Cristian Coletti and Pablo M. Rodriguez
ICMC/USP, UFABC and ICMC/USP, Brazil

Abstract We propose a stochastic model describing a process of awareness, evaluation and decision-making by agents on the d -dimensional integer lattice. Each agent may be in any of the three states belonging to the set $\{0, 1, 2\}$. In this model 0 stands for ignorants, 1 for aware and 2 for adopters. Aware and adopters inform its nearest ignorant neighbors about a new product innovation at rate λ . At rate α an agent in aware state becomes an adopter due to the influence of adopters neighbors. Finally, aware and adopters forget the information about the new product, thus becoming ignorant, at rate one. Our purpose is to analyze the influence of the parameters on the qualitative behavior of the process. We obtain sufficient conditions under which the innovation diffusion (and adoption) either becomes extinct or propagates through the population with positive probability. This is a joint work with Cristian Coletti (UFABC) and Pablo Rodríguez (ICMC-USP).

Estimation in a General Skew-t Measurement Error Model

Lorena Yanet Cáceres Tomaya and Mário de Castro
PIPGEs UFSCar ICMC/USP and ICMC/USP, Brazil

Abstract The usual regression model fits data under the assumption that the explanatory variable is measured without error. However, in many situations the explanatory variable is observed with measurement errors. These will seriously affect the statistical inferences if measurement errors are ignored. We study an extension of measurement error model by assuming a skew-t distribution for the true explanatory variable and a t-distribution with centered parameterization for the error terms. It simultaneously captures skewness and heavy tailedness in data, while allowing the

true covariate and error distributions to have different degrees of freedom. Parameter estimation was obtained by the maximum likelihood methods. The behavior of the estimators was assessed in a simulation study.

Calculando o Valor Esperado na Ocorrência de Uma Palavra

Mariele Parteli Florencio and Renato Jacob Gava
PIPGEs UFSCar ICMC/USP and ICMC/USP, Brazil

Abstract Considere uma sequência de ensaios independentes em que cada um gera uma letra. Ao considerarmos uma palavra, um conjunto finito de letras, nosso interesse é conhecer o tempo de espera até que ela ocorra pela primeira vez nos ensaios. Por exemplo, considere uma sequência de lançamentos de uma moeda honesta, e denotamos o resultado por H, se sair cara ou T se sair coroa. Com essas letras podemos formar a palavra HTHT, na realização TTTHTHT o tempo de espera até ocorrer a palavra foi 8. Introduzamos o problema formalmente, considere a variável aleatória Z que assume valores em um alfabeto finito Σ , e seja $\{Z, Z_k\}_{k \geq 0}$ uma sequência independente e identicamente distribuídas de variáveis aleatórias. Assuma que nos é dada uma palavra A pertencente a Σ . Denotamos por τ o tempo de espera até A ocorrer pela primeira vez ao executar a sequência Z_1, Z_2, \dots . Encontraremos o valor esperado de τ . Considere agora que temos um conjunto $C = A_1, \dots, A_k \in \Sigma$ de k palavras concorrendo pela primeira ocorrência na série $\{Z_n\}_{n \geq 1}$, nesse caso tomamos $\tau = \min\{\tau_{A_1}, \dots, \tau_{A_k}\}$ e encontremos o seu valor esperado.

The Zero-inflated Promotion Cure Rate Regression Model Applied to Fraud Propensity in Bank Loan Applications

Mauro R. de Oliveira Jr., Francisco Louzada and Fernando F. Moreira
PIPGEs UFSCar ICMC/USP, ICMC/USP, Brazil, and University of Edinburgh
Business School, Scotland

Abstract In this paper we extend the promotion cure rate model proposed by Chen et al. [3], by incorporating excess of zeros in the modelling. Despite allowing to relate the covariates to the fraction of cure, the current approach does not enable to relate the covariates to the fraction of zeros. The presence of zeros in survival data, unusual in medical studies, can frequently occur in banking loan portfolios, as presented in Louzada et al. [4], where they deal with propensity to fraud in lending loans in a major Brazilian bank. To illustrate the new cure rate survival method, the same real dataset analyzed in Louzada et al. [4] is fitted here, and the results are compared.

A Statistical Analysis of Outage Data from a Real Power Distribution System

Michel Bessani, Rodrigo Zempulski Fanucchi, Jorge Alberto Achcar, Carlos Dias
Maciel
EESC/USP, Copel Distribution S/A, FMRP/USP and EESC/USP, Brazil

Abstract Power Distribution Systems are responsible for distributing electricity to the final customers, e.g., homes, hospitals and factories. Such structures are classified as critical infrastructures, once its failure directly affects the functioning of essential systems for our society. The adoption of the Smart Grid philosophy leads to changes in such systems, and one is the capability of monitoring systems? outages. In this study, we analyzed the outage data from a single substation of a Brazilian Power Distribution System using both, Bayesian and frequentist statistical analysis. We investigated the outages duration, or repair time, related to the following different causes of outages: atmospheric, environmental, urban, operational needs, and defects in equipment. We performed a one-way ANOVA assuming different variance among groups and fitted a Weibull reliability function using the Maximum Likelihood methodology and a Bayesian approach (MCMC). The results of the three methodologies indicate that the causes directly affects the outages duration. Being the outages caused by atmospheric causes the one with the longest duration. In the other hand, the outages caused by operational needs have the shortest duration. A better knowledge of the effects of different causes in the duration of electricity interruption can be useful from an operational point of view once they can assist in the targeting of financial and human resources to reduce the impact of long-term outages, thus improving the power distribution quality indicators.

An Euler-Maruyama Approach for Cox-Ingersoll-Ross Process

Ricardo Felipe Ferreira and Dorival Leão Pinto Júnior
PIPGEs UFSCar ICMC/USP and ICMC/USP, Brazil

Abstract The Cox-Ingersoll-Ross (CIR) process was originally proposed by John C. Cox, Jonathan E. Ingersoll Jr. and Stephen A. Ross in 1985. Nowadays, this process is widely used in financial modelling, e.g. as a model for short-time interest rates or a volatility process in the Heston model. The stochastic differential equation (SDE) which defines this model does not have closed form solution, so we need to approximate the process by a numerical method. In the literature, there are several numerical approximations which are based in interval discretization. In this work, we approximate the CIR process by the Euler-Maruyama-type method based in random discretization proposed by Leão e Ohashi (2013) under the Feller condition. In this context, we obtain an exponential convergence order for this approximation and we use Monte Carlo techniques to compare the numerical results with theoretical values.

The shortest-path random variable

Rodrigo Lambert and Miguel Abadi
FAMAT/UFU, IME/USP, Brazil

Abstract We consider a discrete time stationary stochastic process over a finite or countable alphabet. For each pair of n -cylinders, we define the Shortest path function as the minimum numbers of steps the process can take to go from one cylinder to another. For two independent ergodic processes with positive entropy we prove a concentration of mass for this random variable. Under some mixing conditions, we prove a large deviation principle and a convergence in law for the shortest-path function.

Compositional Regression Model with Volleyball Data: A Bayesian Approach

Taciana K. O. Shimizu, Francisco Louzada and Adriano K. Suzuki
PIPGEs UFSCar ICMC/USP, ICMC/USP, Brazil, and ICMC/USP, Brazil

Abstract Compositional data consist of known compositions vectors whose components are positive and defined in the interval $(0,1)$ representing proportions or fractions of a “whole?”. We introduce a Bayesian analysis for compositional regression applying additive log-ratio (ALR) transformation and assuming uncorrelated and correlated errors. The Bayesian inference procedure based on Markov Chain Monte Carlo Methods (MCMC). A simulation study was conducted for the evaluation of the estimation procedure in the compositional regression model. The methodology is illustrated on a real data set of volleyball.

The AMH Bivariate Lifetime Copula Model: Modeling, Bayesian Estimation and Case Influence Diagnostic

Taís Roberta Ribeiro and Adriano Kamimura Suzuki
PIPGEs UFSCar ICMC/USP and ICMC/USP, Brazil

Abstract In this work we propose models based on the Ali-Mikhail-Haq archimedean copula to model the dependence of bivariate survival data in the presence of covariates and censored observations. For inferential purposes a Bayesian approach via Monte Carlo Markov Chain (MCMC) methods is considered. Some discussions on the model selection criteria are given. In order to examine outlying and influential observations in the analyzed data, we present a Bayesian case deletion influence diagnostics based on the τ -divergence. The newly developed procedures are illustrated on artificial and real data.

Bayesian Estimation of the Zero-Inflated quasi Poisson-Lindley Model

Wesley Bertoli da Silva and Deisy Morselli Gysi

Universidade Tecnológica Federal do Paraná, Brazil, and University of Leipzig, Germany

Abstract Currently, most applications involving the analysis of count data are performed using the Poisson and Negative Binomial distributions. The latter is a well known 2-parameter Poisson compound model that arises as alternative to fit under (over)-dispersed data once Poisson models are not applicable in this case. The literature concerning discrete models that accommodate different levels of dispersion is very wide and provides several composed distributions as Poisson-Lindley, Negative Binomial-Lindley among others. On the other hand, a relevant drawback of such models is the fact that they do not fit well when a large amount of zeros is observed. To overcome this issue, several zero-inflated and hurdle approaches for standard Poisson and Negative Binomial models were proposed. In addition, zero-inflated models may be useful to deal with overdispersion, for example. Once a brand new Poisson compound model using the more flexible quasi-Lindley distribution is to be released in literature, we will anticipate some works applying the zero-inflated version of the quasi Poisson-Lindley distribution (the actual model has three parameters) to fit a real data set with many zeros and variance much greater than the mean. For this purpose will be considered a reparameterization of the probability function which allow model likelihood to be separable in parameters. It will be shown that this trick also leads to standard Bayesian procedures once marginal posterior functions are available for fixed priors, one being known. Gibbs-Sampling and Metropolis-Hastings algorithms will be applied to estimate the parameters. Model comparison respect to Poisson and Negative Binomial zero-inflated distributions will be performed using DIC in order to determine the best fit to the proposed data set.

Registered Participants

Last update: February 5, 2016

Adriano Kamimura Suzuki
Alan Henrique de Jesus
Alexandre Cristovão Maiorano
Alexsandro G G Gallo
Aline Campos Reis de Souza
Amanda Buosi Gazon
Amanda Morales Eudes
Anders Tolver
André Rogatko
Andrey Luan Gomes Contel
Átila Prates Correia
Bruno Pinheiro de Andrade
Bruno Venâncio Barbosa
Carlos Alberto Ribeiro Diniz
Caroline Tenorio Mendes de Aquino
Cleber Martins Xavier
Daiane de Souza Santos
Danila Maria Almeida de Abreu Silva
Danilo Lourenço Lopes
Demerson Poli
Denis Bertazzo Watashi
Diego Mattozo Bernardes da Silva
Eduardo Schneider Bueno de Oliveira
Elizabeth Chipa Bedia
Eveliny Barroso da Silva
Fatemeh Gholi zadeh kharrat
Flávio Bambirra Gonçalves
Florenzia Graciela Leonardi
Francisco Louzada Neto
Gabriela Cintra Raquel
George Lucas Moraes Pezzott
Glauber Márcio Silveira Pereira
Glauco Valle da Silva Coelho
Gretta Rossi Ferreira
Gustavo Pereira
Jeremias da Silva Leão
Joceline Cristina Sartore
Jorge Bazán
Jose Augusto Fiorucci
José Clelto Barros Gomes
Juliana Cecília da Silva Teixeira
Juliana Cobre
Karina Bindandi Emboaba de Oliveira
Karine Zanuto Mendes
Katherine Elizabeth Coaguila Zavaleta
Lorena Yanet Cáceres Tomaya
Luan Pablo Mamani
Luca Martino

Lucas Pereira lopes
Luis Carlos Otte Junior
Luis Ernesto Bueno Salasar
Marcelo Andrade da Silva
Marco Pollo Almeida
Marcos Antonio Alves Pereira
Máriele Parteli Florencio
Marinho de Andrade Gomes Filho
Mário de Castro Andrade Filho
Mauro Ribeiro de Oliveira Júnior
Michel Bessani
Mônica Thiemi Alves Tanaka
Murilo Cantoni
Natália Lombardi de Oliveira
Nicholas Wagner Eugenio
Pablo Groisman
Patricia Akemi da Silva
Pedro Luiz Ramos
Rafael Catoia Pulgrossi
Rafael Izbicki
Rafael Soares Paixão
Rafael Stern
Raul Caram de Assis
Reiko Aoki
Renata Soares da Costa
Renato Jacob Gava
Ricardo Felipe Ferreira
Roberta de Souza
Rodrigo Lambert
Susan Alicia Chumbimune Anyosa
Taciana Kasaki Oliveira Shimizu
Taís Roberta Ribeiro
Tatyana Zabanova
Themis da Costa Abensur
Vanessa Rufino da Silva
Wesley Bertoli da Silva
Wesley Henrique Silva Pereira
Willian Luís de Oliveira
Yigi Bao
Yury Rojas Benites