

2nd Workshop on Probabilistic and Statistical Methods

February 05–07, 2014

UFSCar, São Carlos, SP, Brazil

PROGRAM

ICMC/USP - DEs/UFSCar - INCTMat

Organizers



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



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



Instituto Nacional de Ciência e Tecnologia de Matemática - INCTMat

Support



About the 2WPSM

UFSCar, São Carlos, February 05–07, 2014

The Workshop on Probabilistic and Statistical Methods is an activity of the Programa Interinstitucional de Pós-Graduação em Estatística do ICMC/USP e o DEs/UFSCar which brings together the statistics research groups from ICMC-USP and UFSCar, at São Carlos/SP, Brazil.

This meeting intends to discuss new developments in statistics, probability and their applications. Activities include invited speaker sessions, short talks, a poster session and a short course devoted to graduate students. 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 stimulate the exchange of experiences with scientific investigation in statistics and probability.

Invited Speakers

Afrânio Vieira, UNB

Alina Von Davier, ETS - Educational Testing Service

Carlos Alberto de Bragança Pereira, IME/USP

Ernesto Mordecki, Universidad de la República

Filidor Edilfonso Vilca Labra, IMECC/UNICAMP

Hua-Hua Chang, University of Illinois, Urbana-Champaign

José Antonio Méndez-Bermúdez, Benemérita Universidad Autónoma de Puebla

Juvêncio Santos Nobre, UFC

Sandro Gallo, UFRJ

Minicourse by

Victor Leiva (Universidad de Valparaíso) and Michelli Barros da Silva (UFCEG)

Scientific Committee

Fábio Prates Machado, IME/USP

Francisco Louzada Neto, ICMC/USP

Nancy Lopes Garcia, IMECC/UNICAMP

Mário de Castro Andrade Filho, ICMC/USP

Organizing Committee

Carlos Alberto Ribeiro Diniz (Coordinator), UFSCar

Dorival Leão Pinto Junior, ICMC/USP

Mariana Cúri, ICMC/USP

Pablo Martin Rodriguez (Coordinator), ICMC/USP

2nd Workshop on Probabilistic and Statistical Methods

February 05–07, 2014

UFSCar, São Carlos, SP, Brazil

SCHEDULE

ICMC/USP - DEs/UFSCar - INCTMat

	Wed (05/02)	Thu (06/02)
8h00 a 8h25	Registration	
8h25 a 8h30	Opening	
8h30 a 9h30	Carlos Pereira	Juvêncio Santos
9h30 a 10h00	Coffee Break	Coffee Break / Poster Session 1
10h00 a 11h00	Alina von Davier	Ernesto Mordecki
11h00 a 12h00	Afrânio Vieira	Filidor Labra
12h00 a 14h00	Lunch	Lunch
14h00 a 16h00	Minicourse	Minicourse
16h00 a 16h30	Coffee Break / Poster Session 1	Coffee Break / Poster Session 2
16h30 a 17h00	Salimeh Sekeh	Antonio Bermudez
17h00 a 17h30	Mauricio Castro	
17h30 a 17h50	Jhon Gonzales	Guaraci Requena
17h50 a 18h10	Robson Machado	Achillefs Tzioufas

	Fri (07/02)
8h00 a 10h00	Minicourse
10h00 a 10h30	Coffee Break / Poster Session 2
10h30 a 11h30	Sandro Gallo
11h30 a 12h30	Hua Hua Chang
12h30 a 12h40	Closing

All Conferences/Talks will be held in the Anfiteatro Bento Prado Júnior at UFSCar.

Conferences

Afrânio Vieira (UNB)

Modified Item Response Models for Known and Unknown Heterogeneity Sources.

Alina Von Davier (ETS - Educational Testing Service)

Estimation and Comparability of Test Performance in Multistage Testing.

Carlos Alberto de Bragança Pereira (IME/USP)

Adaptative Significance Levels Using Optimal Decision Rules: Balancing the Error Probabilities.

Ernesto Mordecki (Universidad de la República)

Implied Volatility Smirkness.

Filidor Edilfonso Vilca Labra (IMECC/UNICAMP)

A Bivariate Birnbaum-Saunders Regression Model.

Hua-Hua Chang (University of Illinois at Urbana-Champaign)

A Large Sample Foundation for Computerized Adaptive Testing.

Antonio Méndez Bermúdez (Benemérita Universidad Autónoma de Puebla)

Scattering and Transport Properties of Random Networks.

Juvêncio Santos Nobre (UFC)

U-Tests for Variance Components in Linear Mixed Models.

Sandro Gallo (UFRJ)

Consistent Estimation of the Context Tree of an Ergodic Process.

Minicurso

Victor Leiva (Universidad de Valparaíso) & Michelli Barros da Silva (UFCG)

Modeling Asymmetric Data Using R.

Oral Communications

Achillefs Tzioufas (Universidad de Buenos Aires)

Elementary consequences of the shape theorem for the contact process via coupling techniques.

Guaraci de Lima Requena (UFSCar)

Capital regulatório para risco operacional: um problema de dependência.

Jhon Franky Bernedo Gonzales (UFSCar)

Bayesian Partitioning for long-term survival data.

Mauricio Castro (Universidad de Concepción)

Diagnostics for censored mixed-effects models using the multivariate t-distribution.

Robson José Mariano Machado (UFSCar)

Semiparametric Partially Nonlinear Mixed-Effects Models.

Salimeh Yasaei Sekeh (UFSCar)

Concordance in Additive Models.

Poster Session 1

Alexandre Hiroshi Watanabe (ICMC/USP)

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

Alexandre Pitanguí Calixto (FACET/UFMG e UFSCar)

A Reversible-jump procedure for Image Analysis.

Amanda Morales Eudes (UFSCar/USP)

Métodos comparativos do Tempo Total em Teste (TTT Plot) para modelagem de dados de sobrevivência.

Arthur Necchi Corrêa (UFSCar)

Modelagem da Distribuição de Skellam para Resultados no Futebol.

Bruno Lucian Gonçalves da Costa (UFF)

Estudo de modelos de crescimento de cristais.

Daiane Aparecida Zuanetti (USP/UFSCar)

Estimating the number of unobservable states in hidden Markov models through reversible jump MCMC.

Daniele Cristina Tita Granzotto (USP/UFSCar)

Reference analysis for the complementary exponential geometric model.

Danielle Gonçalves de Oliveira Prado (UFLA)

A Distribuição Gama Pareto na modelagem de períodos de seca.

Dante Reynaldo Baldeon Molleda (USP/UFSCar)

Problema de programación lineal de dos niveles.

Débora Delbem (UFSCar)

Risco Operacional: o cálculo do Capital Regulatório usando Dependência.

Edimar Izidoro Novaes (UNESP/Botucatu)

O Uso de Análise de Sobrevivência para Inferir no Tempo de Vida de Planárias.

Edwin Chaiña Cahui (Universidad Nacional de San Agustín)

Algumas extensões da distribuição Birnbaum Saunders.

Erlandson Ferreira Saraiva (INMA/UFMS e UFSCar)

A Predictive allocation sampler for Analysis of Mixture Model.

Evandro Luiz de Sousa Jackson (USP/UFSCar)

Modelos de Mistura de Betas com Inflação de Zeros.

Eveliny Barroso da Silva (UFMT e UFSCar)

Regressão Beta com Erro de Medida Multiplicativo.

George Lucas Moraes Pezzott (UFSCar)

Estimation of population size under closed capture-recapture model with heterogeneity among individuals.

Poster Session 2

Gladys Dorotea Cacsire Barriga (UNESP)

On estimation and influence diagnostics for zero-inflated Conway-Maxwell-Poisson regression models: Full Bayesian analysis.

Guilherme Barreto Fernandes (USP/UFSCar)

A correlação espacial do risco de inadimplência e o ganho em modelos de credit scoring.

Jurandir Prazeres Filho (UFSCar)

Capacidade preditiva de modelos de *credit scoring* em inferência dos rejeitados.

Katherine Elizabeth Coaguila Zavaleta (UFSCar)

A Bayesian detection Zero-inflated model.

Lia Hanna Martins Morita (UFSCar)

Accelerated Lifetime Modelling with frailty in a non homogeneous Poisson Process for analysis of recurrent events data.

Lorena Yanet Cáceres Tomaya (UFSCar)

Estimação em modelos com erros de medição heteroscedásticos para observações replicadas.

Luiz Alberto Amaral Nardi (UNESP/Botucatu)

Distribuição Espacial do Caramujo Gigante Africano (*Achatina Fulica*) no sítio urbano de Botucatu.

Luiz Roberto Martins Pinto (UESC)

Mathematical modeling of the experimental residue by singular values decomposition.

Mauro Ribeiro de Oliveira Júnior (UFSCar)

Modeling time to default on a personal loan portfolio in presence of disproportionate hazard rates.

Roberto Vila Gabriel (UNB)

Transição de Fase do Modelo de Aglomerados Aleatórios.

Ricardo Rocha (UFSCar)

Inferência Clássica e Bayesiana para o Modelo de Fração de Cura Gompertz Defeituoso.

Valdemiro Piedade Vigas (UFSCar)

A Distribuição Weibull-Poisson de Longa Duração na Presença de Covariáveis.

Vanessa Rufino da Silva (USP/UFSCar)

Um estudo de simulação comparativo para Testes Adaptativo Informatizado (TAI).

Verônica Amparo Quispe Chire (UFSCar)

Estimação dos parâmetros do modelo Poisson bivariado inflacionado de zeros via algoritmo EM.

Vinícius de Castro Nunes de Siqueira (ICMC/USP)

Convergência e simulação do modelo de volatilidade estocástica de Heston.

Wesley Bertoli da Silva (UFSCar)

Modelo de Regressão Poisson Bivariado Para Previsão de Resultados de Futebol.

Willian Luís de Oliveira (UFSCar)

Seleção de variáveis em uma classe de modelos série de potências modificadas inflacionados de zeros.

2nd Workshop on Probabilistic and Statistical Methods

February 05–07, 2014

UFSCar, São Carlos, SP, Brazil

ABSTRACTS

ICMC/USP - DEs/UFSCar - INCTMat

Minicurso

Modeling asymmetric data using R

Víctor Leiva¹ and Michelli Karinne Barros da Silva²

¹Universidad de Valparaíso and ²UFCCG

Abstract: Asymmetric statistical distributions and their modeling have been widely studied and applied during recent decades. Gamma, inverse Gaussian, lognormal and Weibull models are often used for data analyses and, in particular, the Birnbaum-Saunders model is receiving a special attention; see Johnson, Kotz & Balakrishnan (1994). Although almost of these models have their origin in physics and engineering, their applicability has reached also other areas such as earth, financial and medical sciences; see Barros, Paula & Leiva (2008), Paula, Leiva, Barros & Liu (2012) and Leiva, Santos, Cysneiros & Barros (2014). In this course, we present some asymmetric distributions with emphasis in the Birnbaum-Saunders distribution and discuss several diagnostic and goodness-of-fit tools; see Marchant, Leiva, Cavieres & Sanhueza (2013). In addition, we formulate asymmetric regression models with emphasis in the Birnbaum-Saunders distribution; see Leiva, Barros, Paula & Galea (2007). Furthermore, we introduce extensions and generalizations of the Birnbaum-Saunders distribution; see Díaz-García & Leiva (2005). Finally, we sketch some ideas on the more recent methodologies and applications based on Birnbaum-Saunders models; see Villegas, Paula & Leiva (2011). Implementation in the R software and applications in diverse areas of the science are provided; see Barros, Paula & Leiva (2009) and R Development Core Team (2013).

References:

1. Barros M, Paula GA, Leiva V (2008) A new class of survival regression models with heavy-tailed errors: robustness and diagnostics. *Lifetime Data Anal*, 14, 316-332.
2. Barros M, Paula GA, Leiva V (2009) An R implementation for generalized Birnbaum-Saunders distributions. *Comp Stat Data Anal*, 53, 1511-1528.
3. Díaz-García JA, Leiva V (2005) A new family of life distributions based on elliptically contoured distributions. *J Stat Plan Infer*, 128, 445-457.
4. Johnson NL, Kotz S, Balakrishnan N (1994) *Continuous Univariate Distributions*. Volumes 1-2. Wiley, New York.
5. Leiva V, Barros M, Paula GA, Galea M (2007) Influence diagnostics in log-Birnbaum-Saunders regression models with censored data. *Comp Stat Data Anal*, 51, 5694-5707.
6. Leiva V, Santos M, Cysneiros FJA, Barros M (2014) Birnbaum-Saunders statistical modelling: a new approach. *Stat Model*, 14(1)(in press).
7. Marchant C, Leiva V, Cavieres MF, Sanhueza A (2013) Air contaminant statistical distributions with application to PM10 in Santiago, Chile. *Rev Env Contam Tox*, 223, 1-31.
8. Paula GA, Leiva V, Barros M, Liu S (2012) Robust statistical modeling using the Birnbaum-Saunders-t distribution applied to insurance. *Appl Stoch Mod Bus Ind*, 28, 6-34.
9. R Team (2008) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
10. Villegas C, Paula GA, Leiva V (2011) Birnbaum-Saunders mixed models for censored reliability data analysis. *IEEE Trans Rel*, 60, 748-758.

Observações:

1. Este minicurso está direcionado para estudantes de graduação e de pós-graduação.
2. Em todos os capítulos serão dados comandos em R (<http://www.R-project.org>).
3. Slides e materiais serão escritos em inglês porém o minicurso será ministrado em português.

Conferences

Modified Item Response Models for Known and Unknown Heterogeneity Sources

Afrânio Vieira
UNB

Abstract: The Item Response Theory provides the foundations for measurement of latent traits (or skills about a particular content) of individuals who respond to an assessment (test or questionnaire) formed by a number of items (or questions). Besides the measurement of latent traits associated with each individual, parameters such as difficulty and item discrimination, probability of correct answer in the absence of skill, among others, can be estimated for each item of the assessment. This allows, for example, calibration of instruments on educational tests or targeted behavioral research questionnaires.

The item response models assume, among other assumptions, that the information needed to model the probability of a correct answer, in a right/wrong answers type, depends solely on the latent ability of the subjects and the associated item parameters. However, unknown sources of heterogeneity or known heterogeneity factors, at the moment of modeling phase (individuals of a given region, school, learning methods in primary education, etc.) can generate clusters with different within correlations. These correlations disturb the assumption of independence between individuals, creating an effect of overdispersion or underdispersion that compromises inference about the model and the classification of individuals as their ability skills (or attitude).

This seminar we will present the simplest model, though widely used in the Item Response Theory, on which we will make modifications that allow these sources of heterogeneity, known or unknown, can be adjusted and thus generate more consistent and accurate inferences on the characteristics of the items.

Estimation and Comparability of Test Performance in Multistage Testing

Alina A. von Davier
ETS - Educational Testing Service

Abstract: In a Multistage adaptive Test (MST), all test takers take first a set of items (called the routing module), and then, based on their performance there are administered different sets of items (or modules). Parameter estimation, calibration, linking, scoring, and equating are statistical processes required for any practical implementation of an MST design. This presentation discusses the types of parameter estimation and linking necessary at different stages in the life of an operational MST program. Item response theory (IRT) has traditionally provided the framework for these processes in analyzing test data. Justification is provided for application of conventional estimation methods for IRT to an MST. Calibration and linking involve an initial phase for data collection in which conventional test administrations are used to build modules and routing rules to begin use of MST administrations. Once MST administrations begin, data are collected to establish scoring rules, develop new test modules, equate the cut-scores for routing, and ensure comparability of tests over time. Discussion involves both direct use of estimated examinee proficiency in routing and scoring and use of sum scores for these purposes. When many administrations are involved, special procedures are considered for maintenance of stable linkage.

Adaptative Significance Levels Using Optimal Decision Rules: Balancing the Error Probabilities

Carlos Alberto de Bragança Pereira
IME/USP

Abstract: Our purpose is to recommend a change in the paradigm of testing by generalizing a very natural idea exposed by Morris DeGroot (1975), with the aim of developing an approach that is attractive to all schools of statistics, resulting in a procedure better suited to the needs of science. DeGroot's seminal idea is to base testing statistical hypotheses on minimizing a weighted sum of type I and type II error probabilities instead of the prevailing paradigm, which is fixing type I error and minimizing type II error. DeGroot's result is that for simple versus simple hypotheses, the optimal criterion is to reject the null using the likelihood ratio as the evidence (ordering) statistic, with a fixed threshold value instead of a fixed tail probability. By defining expected type I and type II errors, we generalize DeGroot's approach and that the optimal region is defined by the evidence ratio, that is, a ratio of averaged likelihoods (with respect to a prior measure) and a fixed threshold. This approach yields an optimal theory in complete generality, which the classical theory of testing does not. This can be seen as a Bayesian/Non-Bayesian compromise: using a weighted sum of type I and type II error probabilities is Frequentist, but basing the test criterion on a ratio of marginalized likelihoods is Bayesian. We give arguments to push the theory still further, so that the weighting measures (priors) of the likelihoods do not have to be proper and highly informative, but just "well calibrated". That is, priors that give raise to the same evidence (marginal likelihoods) using minimal (smallest) training samples. The theory that emerges, similar to the theories based on objective Bayesian approaches, is a powerful response to criticisms of the prevailing approach of hypothesis testing. See, for example, Ioannidis (2005) and Siegfried (2010), among many others.

Implied Volatility Smirkness

Ernesto Mordecki

Centro de Matemática, Facultad de Ciencias, Universidad de la República

Abstract: An european option is a contract that gives the right to buy a certain asset in the future at a given price K (the “strike price”). A remarkable feature of this contract is that it does not carry the obligation to buy the asset. In this way, this contract can be seen as an insurance against the rise of price of the asset, observing that, if the (unknown) price $S(T)$ of the asset at the moment of execution of the contract T (the “maturity”) is larger than the strike price K , the holder of the contract obtains the difference $S(T) - K$. If the price is lower, he gets nothing. In conclusion, the value at maturity T of the contract is $\max(S(T) - K, 0)$.

In 1973 Black and Scholes obtained, based on arbitrage considerations, the fair price for this contract at the moment of its signature. This formula, although motivated by a Nobel Prize in Economics in 1997 to Merton and Scholes, gives prices that significantly differ to market prices. Nevertheless it becomes a powerful instrument to translate option prices into volatility values (i.e. the standard deviation in the gaussian model proposed by Black and Scholes). The usual practice to analyze the market is to transform market prices for different strikes (K) and maturities (T) into volatility values, generating a volatility surface, or with fixed T and varying K , volatility curves. Black-Scholes model assumes a constant value for all strikes, but observed surfaces show some characteristic patterns, known as the “volatility smile”, or more recently the “volatility smirk” (mueca, sorrir afetadamente). The problem is now to find a stochastic process to model the asset price, not gaussian now, whose option prices correspond to this observed volatility patterns.

The problem is similar to the substitution of the linear regression by more sophisticated models. Some of them capture the nature of the data at the cost of adding too much parameters. Our proposal, based on Lévy processes, consists first in analyzing a symmetric situation, and afterwards to present a model with essentially two parameters, the volatility and a new parameter, the “smirkness”, that tries to explain the distance from symmetry of the observed model, for fixed values of time.

A bivariate Birnbaum-Saunders regression model

Filidor Edilfonso Vilca Labra
IMECC/UNICAMP

Abstract: Here is proposed a bivariate Birnbaum-Saunders regression model through the use of Sinh-Normal distribution proposed by Rieck (1989). This bivariate regression model can be used to analyze correlated log-time of two units. The proposed regression model has its marginal as the Birnbaum-Saunders regression model of Rieck and Nedelman (1991), which has been extensively discussed by various authors with application in survival and reliability studies. The main aim of this paper is to propose a bivariate Birnbaum-Saunders regression model and discuss some of its properties, in the moment estimation, the maximum likelihood estimation and the observed Fisher information matrix. Hypothesis testing is performed by using the asymptotic normality of the maximum-likelihood estimators.

Finally, the results of a simulation study as well as an application to a real data set are presented.

A Large Sample Foundation for Computerized Adaptive Testing

Hua-Hua Chang

University of Illinois at Urbana-Champaign

Abstract: The paper provides a survey of 18 years' progress that my colleagues and I made in Computerized Adaptive Testing (CAT). We start with a historical review of the establishment of a large sample foundation for CAT. It is worth noting that the asymptotic results were derived under the framework of Martingale Theory, a theoretical perspective of Probability Theory, which may seem unrelated to educational and psychological testing. In addition, we address a number of issues that emerged from large scale implementation and show that how theoretical works can be helpful to solve the problems. Finally we propose that CAT technology can be very useful to support individualized instruction on a mass scale.

Scattering and transport properties of random networks

José Antonio Méndez Bermúdez
Benemérita Universidad Autónoma de Puebla

Abstract: By the use of a Random Matrix Theory approach, we numerically study scattering and transport statistical properties of tight-binding random networks of the Erdős-Renyi type characterized by the number of nodes N and the average connectivity α . In particular, we use the scattering approach to electronic transport and concentrate on the case of a small number of single-channel attached leads. We observe a smooth crossover from insulating to metallic behaviour in the average scattering matrix elements $\langle |S_{mn}|^2 \rangle$, the conductance probability distribution $w(T)$, the average conductance $\langle T \rangle$, the shot noise power P , and the elastic enhancement factor F by varying α from small ($\alpha \rightarrow 0$) to large ($\alpha \rightarrow 1$) values. We also show that all these quantities are invariant for fixed $\xi = \alpha N$. Moreover, we propose a heuristic and universal relation between $\langle |S_{mn}|^2 \rangle$, $\langle T \rangle$, and P and the disorder parameter ξ . We hope our results may shed some light on the transport properties of more realistic models of random networks.

U-Tests for Variance Components in Linear Mixed Models

Juvêncio Santos Nobre
UFC

Abstract: We propose a U -statistics-based test for null variance components in linear mixed models and obtain its asymptotic distribution (for increasing number of units) under mild regularity conditions that include only the existence of the second moment for the random effects and of the fourth moment for the conditional errors. We employ contiguity arguments to derive the distribution of the test under local alternatives assuming additionally the existence of the fourth moment of the random effects. Our proposal is easy to implement and may be applied to a wide class of linear mixed models. We also consider a simulation study to evaluate the behaviour of the U -test in small and moderate samples and compare its performance with that of exact F -tests and of generalized likelihood ratio tests obtained under the assumption of normality. A practical example in which the normality assumption is not reasonable is included as illustration.

Consistent Estimation of the Context Tree of an Ergodic Process

Sandro Gallo
UFRJ

Abstract: We consider the class of discrete time ergodic processes with finite state space. To each of these processes we can associate a context tree, which is informally a set of strings (finite or not) that represents how the process depends on its past. After introducing context trees and explaining their interest, we will speak about (non)existence of a consistent estimators. This is work in progress with Prof. Florencia G. Leonardi.

Oral Communications

Elementary consequences of the shape theorem for the contact process via coupling techniques

Achillefs Tzioufas
Universidad de Buenos Aires

Abstract: A merely using basics argument allows showing that the lower-inclusion part of the shape theorem may take effect from time zero with non-zero probability. Closely related arguments allow to further show the one-dimensional non-nearest neighbours analogue of the endpoints coupling.

Capital regulatório para risco operacional: um problema de dependência

Guaraci de Lima Requena
UFSCar

Abstract: Neste trabalho apresentamos alguns aspectos sobre o risco operacional focando, principalmente, em quatro diferentes métodos para calcular o capital regulatório exigido para tal risco. Um dos métodos, proposto pelos autores, tem como base a teoria de cópulas, usada para modelar a dependência entre as diferentes unidades de risco de uma instituição financeira. Um dos outros três métodos é proposto pelo Comitê de Basileia e é o usualmente empregado nas instituições. Este método superestima o capital regulatório e é apoiado em uma suposição irreal e contraditória: a dependência perfeita positiva entre as unidades de risco.

Um estudo de simulação é desenvolvido com o objetivo de comparar os capitais regulatórios calculados a partir do uso destes métodos, considerando diferentes níveis de dependência entre quaisquer duas unidades de risco.

Bayesian Partitioning for long-term survival data

Jhon Franky Bernedo Gonzales
UFSCar

Abstract: In this paper we propose a Bayesian partition modeling for lifetime data in presence of a cure fraction by considering a local structure generated by a tessellation which depends on covariates. In this modelling we including information of nominal qualitative variables with more than two categories or ordinal qualitative variables. The proposed modeling is based on a promotion time cure model structure but assuming that the number of competing causes follow an exponential discrete power series distribution. It is an alternative modeling strategy to the conventional survival regression modeling generally used for modeling lifetime data in presence of a cure fraction, which models the cure fraction through a (generalized) linear model of the covariates. An advantage of our approach is its ability to capture the effects of covariates in a local structure. The flexibility of having a local structure is crucial to capture local effects and features of the data. The modelling is illustrated with a real data set.

Diagnostics for censored mixed-effects models using the multivariate t -distribution

Mauricio Castro
Universidad de Concepción

Abstract: In biomedical studies on HIV RNA dynamics, the viral loads generate repeated measures that are often subjected to (upper and lower) detection limits, and hence these responses are either left- or right-censored. Linear and non-linear mixed-effects censored (LMEC/NLMEC) models are routinely used to analyze these longitudinal data, with normality assumptions for the random effects and residual errors. However, the derived inference may not be robust when these underlying normality assumptions are questionable, specially presence of outliers and thick-tails. Motivated by this, Matos et al. (2013b) recently proposed an exact EM-type algorithm for LMEC/NLMEC models using a multivariate Student's- t distribution, with closed-form expressions at the E-step. In this paper, we develop influence diagnostics for LMEC/NLMEC models using multivariate Student's- t density, based on the conditional expectation of the complete data log-likelihood which eliminates the complexity associated with the approach of Cook (1977, 1986) for censored mixed-effects models. The new methodology is illustrated through an application to a longitudinal HIV dataset using the NLMEC framework. In addition, a simulation study is presented, which explores the accuracy of the proposed measures in detecting influential observations in heavy-tailed censored data under different perturbation schemes.

Semiparametric Partially Nonlinear Mixed-Effects Models

Robson J. M. Machado

UFSCar

Abstract: Correlated data sets with nonlinear structure are common in many areas such as biostatistics, pharmacokinetics and longitudinal studies. Semiparametric partially nonlinear mixed-effects models are useful alternatives to model those kinds of data (Machado & Russo 2013). This model assumes that the response variable is explained by a nonlinear function, which depends on fixed effects parameters and explanatory variables, by a nonparametric function which provides flexibility about the functional form that underlies the data and by random errors. The random effects are included linearly to the model. It provides advantages to obtain the distribution of the response variable and enables the model to take into account the within-group correlation structure. It is assumed that the random errors and the random effects jointly follow a multivariate normal distribution. The P-spline smoothing technique (Eilers & Marx 1996) is used to approximate the nonparametric function. The basic methodology to obtain the parameters estimates is penalized maximum likelihood method. The random effects may be obtained by using the Empirical Bayes.

The proposed model was applied to a pharmacokinetic data set, in which the anti-asthmatic drug theophylline was administered to 12 subjects and serum concentrations were taken at 11 time points over the 25 hours after being administered.

Joint work with C. M. Russo (ICMC-USP).

References:

1. Eilers, P. H. & Marx, B. D. (1996), Flexible smoothing with b-splines and penalties, *Statistical science* **11**, 89-102.
2. Machado, R. J. M. & Russo, C. M. (2013), Semiparametric partially nonlinear mixed-effects models with p-splines, *In: Proceedings of the 28th International Workshop on Statistical Modelling, Muggeo, V. M. R., Capursi, V., Boscaino, G., Lovison, G. (Eds). Palermo: Università Degli Studi di Palermo* **1**, 251-256.

Concordance in Additive Models

Salimeh Yasaei Sekeh
UFSCar

Abstract: Let $Z_1 = g(X_1; Y)$ and $Z_2 = g(X_2; Y)$. Commonly X_1 and X_2 indicate the “individuality” that is associated with Z_1 and Z_2 , whereas Y indicates the factors that give rise to the partial dependence between Z_1 and Z_2 . The influence of Y on the strength of positive dependence between Z_1 and Z_2 is discussed in the talk. In particular, comparisons of dependence based on Kendall’s τ are discussed for the case of additive functions g . Joint work with Franco Pellerey and Moshe Shaked.

Posters

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

Alexandre Hiroshi Watanabe
ICMC/USP

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 [1]. 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).

References:

1. Klein, J., and Moeschberger, M. Survival analysis: techniques for censored and truncated data. Springer Verlag, 2003.

Métodos comparativos do Tempo Total em Teste (TTT Plot) para modelagem de dados de sobrevivência

Amanda Morales Eudes
UFSCar/USP

Abstract: Em algumas aplicações existem informações qualitativas e, muitas vezes, estruturais a respeito do fenômeno em questão, que pode ser utilizada na determinação empírica da forma da função de taxa de falha. Informações estruturais estão diretamente vinculadas ao conhecimento do pesquisador sobre o fenômeno, enquanto que informações qualitativas podem ser extraídas através de análise gráfica. Neste

contexto, um gráfico conhecido como gráfico TTT-plot (tempo total em teste) é de grande utilidade. Este gráfico foi inicialmente proposto por Aarset (1987) e tem sido extensivamente utilizado em várias situações, mas que é calculado utilizando apenas tempos de falha, supondo que os dados são completos. Como em análise de sobrevivência geralmente temos a presença de censuras, Sun & Kececioglu (1999) mostraram uma modificação deste método, incorporando dados censurados no seu cálculo. Neste trabalho, comparação entre os métodos propostos são utilizados através de um estudo de simulação e aplicação com dados de longa duração que tem uma grande proporção de censuras à direita. Trabalho em conjunto com Vera Lúcia Damasceno Tomazella (DEs, UFSCar) e Ricardo Ferreira da Rocha (DEs, UFSCar).

Modelagem da Distribuição de Skellam para Resultados no Futebol

Arthur Necchi Corrêa
UFSCar

Abstract: Nos últimos anos vem sendo observado um crescente interesse por modelagem de previsões relacionadas ao futebol, que partem de questões que vão do entretenimento até o ramo das apostas, assunto que vem ganhando destaque atualmente. No modelo proposto por Lee (1997), temos que o gol de cada equipe é dado por distribuições Poisson independentes. No lugar de modelar o número de gols marcados por cada time, vamos nos concentrar na diferença dos saldos de gol entre as equipes, utilizando a Distribuição de Skellam, obtida a partir da diferença entre duas Distribuições de Poisson independentes.

Estudo de modelos de crescimento de cristais

Bruno Lucian Gonçalves da Costa
UFF

Abstract: Os cristais são utilizados em varias áreas do conhecimento, por isso entender e modelar suas características, tais como forma e velocidade de crescimento, são muito importante. Utilizamos cadeias de Markov em tempo contínuo para modelar crescimento de cristais. Para os modelos, estudos sobre forma foram feitos na literatura de jeito teórico. O presente trabalho busca replicar os estudos por método de simulação. Estudos sobre velocidade de crescimento não foram feitos na literatura de jeito teórico. Estudamos por método de simulação esta característica.

Estimating the number of unobsersable states in hidden Markov models through reversible jump MCMC

Daiane Aparecida Zuanetti

Abstract: Hidden Markov models are a generalization of mixture model and provide a flexible class of models to describe dependent phenomena. In this work, we present the reversible jump MCMC used to estimate the number of components or unobservable states of a hidden Markov model and the parameters associated. We apply this methodology to a simulated and a real dataset.

Reference analysis for the complementary exponential geometric model

Daniele Cristina Tita Granzotto
USP/UFSCar

Abstract: This article applies Bayesian reference analysis, widely considered as the most successful method to produce objective, model-based, posterior distributions, to a problem of estimating parameters of a new two-parameter lifetime distribution with increasing failure rate, the complementary exponential geometric (CEG) model proposed by Louzada et. al. (2011). For the first time, explicit expressions are given for the construction of the priori and posteriori the CEG model. The presented methodology is illustrated by the use of a real data set which presents the study of time until the cure of cervix lesions that are precursors cancer lesions in the cervix. According to INCA (Cancer National Institute), cervical cancer stands as the fourth cause of death among women in Brazil. Together with breast cancer, it is one of the most common malignancy affecting women worldwide. For this reason, patients must be carefully evaluated for metastatic disease. These data were collected in the Woman Clinic which is sited in Maringá city (Paraná State, Brazil). To this end, we presented some of the main properties of this model and its characteristics related to the reliability analysis. A simulation studies is performed to analyse the frequentist properties of credible intervals from the reference posterior distribution beyond the standard error and mean square error (MSE) of estimations. Joint work with Vera L. D. Tomazella (UFSCar) and Francisco Louzada Neto (ICMC-USP).

A Distribuição Gama Pareto na modelagem de períodos de seca

Danielle Gonçalves de Oliveira Prado
UFLA

Abstract: A seca é um fenômeno climático que ocorre periodicamente e pode causar impactos em diferentes seguimentos da atividade humana. Surge então a necessidade da construção de modelos matemáticos e estatísticos que expliquem melhor a ocorrência desse fenômeno. Essa abordagem tem sido tema de muitos artigos, em particular, vários publicados pelo pesquisador Saralees Nadarajah. Neste trabalho são estimados os parâmetros do modelo bivariado: Gama Pareto aplicados a dados de seca e não seca da cidade de Joaquim Távora - PR - Brasil. Trabalho em conjunto com Rodrigo de Andrade Ferrazza, Juscelia Dias Mendonça e Larissa Ribeiro de Andrade.

PROBLEMA DE PROGRAMACION LINEAL DE DOS NIVELES

Dante Reynaldo Baldeon Molleda

Abstract: El problema de programación de dos niveles (BPP) es un problema de optimización jerárquica donde un subconjunto de variables es forzado a tomar una solución de un problema de optimización dado, parametrizado por el resto de variables. La estructura de la optimización jerárquica aparece de forma natural en muchas aplicaciones cuando las acciones del nivel inferior dependen de las decisiones del nivel superior. La formulación original del problema de programación de dos niveles apareció en 1973, en un artículo escrito por J.Bracken y J.McGill. Sin embargo, no fue hasta principios de los ochenta que estos problemas comenzaron a recibir la atención que se merecen. Motivados por la teoría de juegos de H. Strackelberg.

Risco Operacional: o cálculo do Capital Regulatório usando Dependência

Débora Delbem
UFSCar

Abstract: Neste trabalho propomos um novo método para o cálculo do capital regulatório para o risco operacional. O método proposto é utilizado para calcular o capital regulatório para duas classes de risco e é baseado em alguns pressupostos considerados importantes no cálculo deste capital. Entre esses pressupostos se destacam a opinião de especialistas e a captação de dependência entre as variáveis perdas considerando a probabilidade dos eventos de perdas ocorrerem conjuntamente. Essa

probabilidade é captada via cópula. Além disso, apresentamos mais dois métodos, o do somatório, proposto pelo Acordo de Basileia II (2004), e o da correlação não-perfeita, proposto por Frachot et al. (2004). Finalmente, realizamos um estudo de simulação com o objetivo de comparar os capitais regulatórios totais calculados em cada método.

O Uso de Análise de Sobrevivência para Inferir no Tempo de Vida de Planárias

Edimar Izidoro Novaes
UNESP/Botucatu

Abstract: O trabalho que segue é referente a uma abordagem de Análise de Sobrevivência em dados de observações feito com planárias presentes nas águas de diferentes pontos e diferentes rios do polo petroquímico (REPLAN), onde o objeto da abordagem foi fazer uma abordagem com Análise de Sobrevivência e ajustar um modelo tendo como resposta o tempo de vida das planárias observadas nas observações. Usando o software SAS foi aplicado o estimador de Kaplan Meier para verificar a existência ou não de risco proporcional. Depois de verificado que os dados não apresentaram riscos proporcionais, foi selecionado pelos valores de Wilcoxon e Log-Rank, as variáveis significativas e com tais variáveis, foi verificado dentre os modelos paramétricos Weibull, Gama, Exponencial e Lognormal, qual modelo melhor se ajustou aos dados. Analisando os valores de AIC, AICC e BIC no ajuste dos modelos paramétricos, verificou-se que o modelo Gama foi o que melhor se ajustou, pois foi o ajuste que apresentou menor valor para AIC AICC e BIC. Depois usando o modelo Gama, foi observado o p-valor na relação das variáveis significativas com o tempo a censura e o modelo nulo, onde foi selecionado a relação que proporcionou o menor p-valor e então foi realizado uma análise de resíduo para concluir se o ajuste foi realmente satisfatório. Por fim foi feito a interpretação dos resultados de maneira que para isso, o cálculo foi através da exponencial dos coeficientes angulares obtidos na estimação do modelo. Trabalho em conjunto com Miriam Harumi Tsunemi e Dilza Aparecida Nalin (Depto de Bioestatística UNESP-Botucatu).

Algumas extensões da distribuição Birnbaum Saunders

Edwin Chaiña Cahui
Universidad Nacional de San Agustín

Abstract: A distribuição Birnbaum-Saunders (BS) está baseada em um argumento físico de dano cumulativo que produz a fadiga de materiais. Esta fadiga foi identificada como uma importante causa de falhas em estruturas de engenharia. Um

dos principais problemas para escolher uma distribuição estatística, é que frequentemente vários modelos ajustam os dados bem na parte central, porém, no entanto, os extremos da distribuição colocam em dúvida a decisão para selecionar alguns dos modelos propostos. (...) Observamos que os modelos log-BSSMN se ajustam melhor aos dados que o modelo log BS. Entre os modelos *logBSSMN* podemos destacar que o modelo *logBS-t* se ajusta melhor aos dados na presença de observações influentes.

A Predictive allocation sampler for Analysis of Mixture Model

Erlandson Ferreira Saraiva
INMA/UFMS e UFSCar

Abstract: Mixture models are widely used for modeling data where each observation is assumed to have arisen from one of the k subpopulations, where k can be known or unknown, and each subpopulation may be suitably modelled by a density from some parametric family.

In this WORK, we develop an MCMC procedure to jointly estimate the number of components and allocation probabilities of a mixture model. For this, we follow [1] and [2], integrating the parameters out of the model and working in terms of number of components and latent allocation variables. In order to makes draws from the joint posterior distribution of the number of components and the latent allocation variables we propose the PAS.

The PAS is a Metropolis-Hastings within Gibbs sampling, where the Gibbs sampling is performed to update the latent allocation variables and the Metropolis-Hastings is performed to change the number of components using the pair of ejection/absorption moves that change k in the neighborhood $k \pm 1$. These moves are developed using allocation probabilities which are calculated using the predictive distribution. As advantages, the ejection/absorption moves allow a major change in configuration of the latent variables in a single iteration of the algorithm, allowing a greater exploration of clusters; the moves can be easily proposed and tested and new clusters are created based on information from clusters of observations.

We apply the method to simulated data sets to verify its performance and to a real data set to illustrate its use. Simulation results show a good performance of the method to estimate the number of components and allocation probabilities. The first real data set used consists of *Escherichia coli* bacterium gene expression data, described in [?]. The second real data set is a shotgun proteomics experiment extracted from site <http://cybert.ics.uci.edu/controlexp> Baldi.

References:

1. CASELLA, G., ROBERT, C. P. and WELLS, M. T. (2004). Mixture models, latent variables and partitioned importance sampling. *Statistical Methodology* **1** 1–18.
2. NOBILE, A. (2004). On the posterior distribution of the number of components in a finite mixture. *The Annals of Statistics* **32** 2044–2073.

Modelos de Mistura de Betas com Inflação de Zeros

Evandro Luiz de Sousa Jackson
USP/UFSCar

Abstract: Diversos problemas de modelagem possuem variável resposta como sendo uma proporção, e é comum utilizar modelos de regressão beta quando se deseja modelar taxas, ou razões. No entanto em alguns conjuntos de dados o excesso de valores zeros não permite que a distribuição beta convencional os descreva satisfatoriamente. Uma solução apresentada na literatura para esse problema foi a de adicionar uma massa de probabilidade ao ponto zero, chegando-se assim a uma nova distribuição denominada por distribuição beta inflacionada. Analisando histogramas das variáveis (resposta) de interesse para alguns desses dados de proporções e taxas levantados, foi possível observar também que em alguns casos estes apresentavam uma forma bimodal devido a sua característica intrínseca do problema em questão, sugerindo que uma mistura de distribuições teóricas pudessem descrever mais satisfatoriamente sua distribuição empírica. Sendo assim este trabalho propõe uma extensão da distribuição beta inflacionada, para uma distribuição de mistura de betas inflacionadas.

Regressão Beta com Erro de Medida Multiplicativo

Eveliny Barroso da Silva
UFMT e UFSCar

Abstract: Neste trabalho serão apresentados a construção de alguns modelos estatísticos motivados por problemas reais na área financeira. Instituições financeiras têm a necessidade de determinar a proporção de gasto do limite a ser oferecido no LIS (cheque especial) para futuros clientes. O conhecimento desta proporção permite à instituição alocar um montante de capital para fazer frente a este risco, além de ter a possibilidade de calcular o lucro futuro oriundo dos juros que serão pagos. Um dado extremamente importante para a determinação desta proporção é a renda do cliente. Por se tratar de clientes ainda não correntistas, informações sobre a renda

real podem não estar disponíveis no cadastro. Por essa razão, a instituição obtém no mercado uma renda presumida do cliente, ou seja, uma renda medida com erro. Apesar da motivação financeira estes modelos são facilmente adaptados para serem usados em qualquer outra área do conhecimento.

Estimation of population size under closed capture-recapture model with heterogeneity among individuals

George Lucas Moraes Pezzott
UFSCar

Abstract: In this work, we discuss maximum likelihood estimation for a closed capture-recapture model which allows heterogeneity among individuals by means of a mixture model. Suppose there are N individuals in the population which is sampled in t different occasions, $t > 1$. We shall assume that each individual i is captured in the j occasion, independently from the other individuals and other occasions, with probability p_i , which means, the capture probabilities varies with individuals but not with occasions. Furthermore, we suppose that the capture probabilities p_1, \dots, p_N are i.i.d random variables with discrete distribution function F which takes values $\theta_1, \dots, \theta_k$ with probabilities $\omega_1, \dots, \omega_k$, respectively, $k > 1$. For identifiability purposes, we shall adopt the restriction $\theta_1 < \dots < \theta_k$. Thus, we interpret the parameters $\theta_1, \dots, \theta_k$ as different levels of catchability for the individuals and $\omega_1, \dots, \omega_k$ as proportions of individuals in the population pertaining to each level. For maximum likelihood estimation of N , we propose an iterative method based on the EM algorithm which overcomes the difficulty of the imposed parametric restriction and the discreteness of the parameter N . The proposed method is applied to two real data sets. The first one is related to an animal trapping study performed in 18 consecutive days and the second concerns a software review designed to identify the number of errors in a specific feature of AT&T 5 ESS switch. The choice of the number of catchability levels k is done by means of AIC and BIC model selection criteria. Joint work with Luis E. B. Salazar and José G. Leite (UFSCar).

On estimation and influence diagnostics for zero-inflated Conway-Maxwell-Poisson regression models: Full Bayesian analysis

Gladys Dorotea Cacsire Barriga
UNESP

Abstract: We consider the use of Markov chain Monte Carlo methods to develop a Bayesian analysis for parameter estimation of zero-inflated regression models based on

Conway-Maxwell-Poisson (COM-Poisson) distribution. Moreover, some discussions on the model selection to compare the fitted models are given, as well as, case deletion influence diagnostics are developed for the joint posterior distribution based on the ϕ -divergence, which has several divergence measures as particular cases, such as the Kullback-Leibler (K-L), J-distance, L1 norm and χ^2 -square divergence measures. Simulation studies are performed and experimental results are illustrated based on a real data.

A correlação espacial do risco de inadimplência e o ganho em modelos de credit scoring

Guilherme Barreto Fernandes
USP/UFSCar

Abstract: O modelo de Vasicek (2002) propõe que a probabilidade de inadimplência de uma empresa depende de suas características idiossincráticas e de um fator macroeconômico comum a todas as empresas. Entretanto, empresas de pequeno e médio porte devem ser mais afetadas por um fator econômico regional. Apesar de não ser possível observar esse fator, seu efeito pode ser percebido através da análise espacial da inadimplência. O resultado dessa análise é o fator de risco espacial. No presente trabalho, propomos a hipótese que a estrutura de correlação espacial existe e não é a mesma para todas as regiões do estado de São Paulo. A estimativa de um fator de risco espacial é obtida através da krigagem ordinária (Matheron, 1963) da inadimplência das empresas. A importância do fator de risco espacial é verificada através de sua inclusão em modelos de credit scoring. A base de dados de um banco de médio porte foi utilizada na verificação dessa segunda hipótese e os métodos de regressão logística (Hosmer e Lemeshow, 2000) e regressão logística com erro de medida (Cook e Stefanski, 1994) foram aplicados. O primeiro resultado encontrado é a estrutura de correlação espacial de três regiões paulistas: 1) grande São Paulo, 2) interior paulista e 3) Vale do Paraíba e litoral. Outro resultado importante é a confirmação da relevância do fator de risco espacial nos modelos de credit scoring, já que a inserção desse como covariável representa um aumento de cerca de 8 p.p. de KS para o caso apresentado. Por fim, a utilização do modelo com erro de medida não apresentou diferença relevante em termos de desempenho ou efeito estimado. Trabalho em conjunto com Rinaldo Artes (Departamento de Economia, Insper-SP).

Capacidade preditiva de modelos de *credit scoring* em inferência dos rejeitados

Jurandir Prazeres Filho
UFSCar

Abstract: A concessão de crédito é uma decisão a ser tomada num contexto de incertezas. No momento em que o credor decide conceder um empréstimo, realizar um financiamento ou venda a prazo sempre existe a possibilidade de perda, e, se for atribuída uma probabilidade a esta perda ou utilizado algum método de modelagem de dados, a decisão de conceder ou não crédito será mais confiável. Com o objetivo de auxiliar a tomada de decisão de aceitar ou não o pedido de crédito de clientes são utilizados os modelos *credit scoring*, que estimam a probabilidade de perda associada à concessão de crédito. Porém, um dos problemas envolvendo estes modelos é que somente informações a respeito dos proponentes aceitos são utilizadas, o que causa um viés amostral, pois, são descartados os solicitantes recusados. Com intuito de solucionar este problema tem-se a inferência dos rejeitados, em que são considerados os indivíduos que tiveram pedido de crédito rejeitado. No entanto, somente considerar a inferência dos rejeitados e o uso de apenas um método de modelagem de dados, muitas vezes, não é suficiente para que se tenha um desempenho satisfatório. Desta forma, foram utilizados resultados combinados de três metodologias, regressão logística, *probit* e árvore de decisão/classificação. O objetivo dessa combinação é verificar se ocorre aumento da capacidade preditiva e as métricas utilizadas são sensibilidade, especificidade, valor preditivo positivo, valor preditivo negativo e acurácia. Trabalho em conjunto com Prof. Dr. Francisco Louzada (ICMC/USP) Prof. Dr. Gilberto Pereira Araújo (DES/ UFTM).

A Bayesian detection Zero-inflated model

Katherine Elizabeth Coaguila Zavaleta
UFSCar

Abstract: We studied the destructive model formulated by Rodrigues (2010) to observable variables. The modeling was applied in the analysis of biological tests on animals in order to identify and assess chemical agents that can prevent cancer. The goal of these biological experiments is to evaluate the effect of a particular treatment on incidence rate of tumours observed in animals. The observation of these tumors is possible until death of the animal, we shall call the observable terminal variable. Through an example, we observe that the results can be sensitive to parametric shapes of distributions assumed, in particular for the distribution of the number of tumors per animal. The data have presence of zeros, so will be proposed for the number of tumors detected a zero-inflated distributions with use of Poisson and negative binomial models. The Gamma model is assumed for the time of detection of tumors.

The study of this work will be addressed from the Bayesian point of view, where the estimation of the parameters are obtained through the MCMC method for doing inference.

Accelerated Lifetime Modelling with frailty in a non homogeneous Poisson Process for analysis of recurrent events data

Lia Hanna Martins Morita
UFSCar

Abstract: In survival analysis, the hazard function for each individual may be influenced by risk variables, but commonly we have some variables that can not be observed nor measurable. Besides, when lifetime data presents more than one event observed for each individual, frailty is a common factor among such recurrence times. In this paper we present a natural extension of the conventional accelerated failure time model for recurrent events with frailty, in order to take into account possible correlations and heterogeneity among failure times. We include the effects of covariates on the intensity function of non homogeneous Poisson process for recurrent events. The proposed model retains the direct physical interpretation of the original accelerated failure time model in that the role of the covariates is to accelerate or decelerate the time to each recurrence. These include parametric approaches to model fitting, we consider estimate the vector of regression parameters under this model and the baseline hazard functions are considered specified up to a parameter vector “theta”. This methodology is illustrated with a simulation study and also with a known real data set. Joint work with Vera Tomazella (UFSCar) and Francisco Louzada-Neto (ICMC-USP).

Estimação em modelos com erros de medição heteroscedásticos para observações replicadas

Lorena Yanet Cáceres Tomaya
UFSCar

Abstract: Neste trabalho estudamos um modelo com erros de medição heteroscedásticos para observações replicadas. A estimação dos parâmetros do modelo proposto foi efetuada pelos métodos de máxima verossimilhança de máxima pseudoverossimilhança. O comportamento dos estimadores de alguns parâmetros foi analisado por meio de simulações para diferentes números de réplicas. Resultados sobre as matrizes de covariâncias dos estimadores assim como a probabilidade de cobertura dos intervalos de confiança assintóticos foram apresentados. Trabalho em conjunto com Mário de Castro (ICMC-USP).

Distribuição Espacial do Caramujo Gigante Africano (*Achatina Fulica*) no sítio urbano de Botucatu

Luiz Alberto Amaral Nardi
UNESP/Botucatu

Abstract: O Caramujo Gigante Africano (*Achatina Fulica*) é uma das cem maiores espécies exóticas invasoras da atualidade. Este molusco encontra-se atualmente em todo o Brasil e pode ser considerado sob dois aspectos: uma praga agrícola e um problema de zoonose já que atua como hospedeiro intermediário de um nematódeo causador da angiostrongilíase meningoencefálica e angiostrongilíase abdominal. Utilizando-se dados obtidos junto à secretaria municipal de saúde da cidade de Botucatu sobre os locais de ocorrência do *A. fulica* nos anos de 2010, 2011 e 2012, aplicou-se uma técnica da Estatística Espacial, o estimador de intensidade kernel e determinou-se as áreas com maiores intensidades de aparecimento destes moluscos. Notou-se, ao final da análise, que houve um deslocamento da área de maior intensidade de ocorrências no transcorrer dos anos pesquisados. Trabalho em conjunto com Paulo Milton Barbosa Landim (Departamento de Geologia Aplicada, Instituto de Geociências e Ciências Exatas, UNESP, Rio Claro/SP) e Cassiano Victoria (Departamento de Higiene Veterinária e Saúde Pública, Faculdade de Medicina Veterinária e Zootecnia, Unesp, Botucatu/SP).

Modeling time to default on a personal loan portfolio in presence of disproportionate hazard rates

Mauro Ribeiro de Oliveira Júnior
UFSCar

Abstract: We consider a parametric Weibull mixture cure model for modeling time to default on a personal loan portfolio in presence of disproportionate hazard rate. A straight comparison with the well known Cox proportional hazard mixture cure model provides evidence that risk measurements derived from a Cox proportional hazard mixture cure model can be greatly affected if required proportional conditions are not satisfied. In fact, taking into account presence of covariates, if the covariates levels do not have proportional hazards rate over time, the adjustment with models that assume proportional hazard rates will not be appropriate, and then, erroneous measurements may be derived, i.e., under or overestimate expected losses of a portfolio can be observed. Our approach can be seen as a complement to the modeling framework presented in Tong et al. (2012), which require a proportional hazard rate structure. A credit data from a Brazilian commercial bank illustrates the procedure. Joint work with Francisco Louzada (ICMC-USP), Vicente G. Cancho (ICMC-USP) and Bao Yiqi (DEs-UFSCar).

Transição de Fase do Modelo de Aglomerados Aleatórios

Roberto Vila Gabriel

UNB

Abstract: O Modelo de Aglomerados Aleatórios ou Modelo FK (Devido aos matemáticos Fortuin-Kesten) pode-se ver como uma generalização do conhecido Modelo de Bernoulli no sentido que: a escolha dos elos (de ser aberto o fechado) de algum espaço lattice finito (fixado a priori) vai depender de certas condições de fronteira que o modelo vai sofrer. É por isso que nossa situação agora é muito mais complicada no sentido que as ferramentas que são comumente usadas na área de percolação como: Desigualdade FKG e algumas noções de independência, não vão ser válidas neste novo modelo. Porém novas ferramentas geométricas precisam ser introduzidas para poder obter alguns resultados análogos ao modelo de Bernoulli, como por exemplo: unicidade do aglomerado aleatório, unicidade da medida FK, valor exato do ponto crítico onde o Modelo de Aglomerados Aleatórios sofre transição de fase.

Portanto nosso problema principal neste seminário é definir o Modelo de Aglomerados Aleatórios sobre a Rede Bidimensional e obter o valor do ponto crítico deste modelo. O Modelo de Aglomerados Aleatórios é um modelo que depende de dois parâmetros $p \in [0, 1]$ e $q \geq 1$; onde o parâmetro p descreve a probabilidade de um elo estar aberto ou fechado e q é o parâmetro de dependência que de certa maneira influencia na contagem dos aglomerados no Modelo de Aglomerados Aleatórios. Na realidade o parâmetro q vai tornar nosso modelo a ser dependente.

Cabe mencionar que o Modelo de Aglomerados Aleatórios é bastante interessante, pois mediante este modelo pode-se estudar o Modelo de Ising e o Modelo de Potts de q estados de uma maneira mais simples. Para $q = 2$ o Modelo FK está relacionado com o Modelo de Ising. O cálculo exato do ponto crítico do Modelo de Ising foi feito primeiramente pelo Físico-Químico Lars Onsager e por este feito Onsager recebeu em 1968 o prêmio nobel de Química. Alguns anos mais tarde em um célebre artigo, Smirnov [Ann. of Math. 172 (2010) 1435-1467] definiu um observável para o Modelo de Aglomerados Aleatórios no ponto auto-dual com $q = 2$ e o usou para obter a invariância conforme no limite de escalas do modelo de Ising. Isto possibilitou que Smirnov reobtivesse o resultado de Onsager, mas introduzindo uma nova maneira de fazer o cálculo explícito do ponto crítico do modelo de Ising. Por este importante resultado Smirnov em 2010 acabou recebendo a Medalha Fields que é um dos prêmios de maior prestígio na comunidade matemática.

Inferência Clássica e Bayesiana para o Modelo de Fração de Cura Gompertz Defeituoso

Ricardo Rocha
UFSCar

Abstract: Na teoria de análise de sobrevivência, estuda-se o caso em que há imunes numa certa população, gerando assim uma fração de cura. Esses casos são usualmente tratados utilizando um modelo de mistura padrão, no entanto, aqui utilizamos uma metodologia alternativa, baseada em modelos defeituosos. Modelos defeituosos são aqueles que permitem mudar o domínio que está definido um certo parâmetro de uma distribuição, gerando uma distribuição imprópria que se comporta naturalmente como um modelo de fração de cura. Na literatura encontram-se duas distribuições com essa propriedade, a inversa Gaussiana e a Gompertz. Propomos uma abordagem bayesiana para o modelo Gompertz Defeituoso, juntamente com uma comparação ao estimador de máxima verossimilhança. Os resultados indicam que o modelo bayesiano se aproxima do clássico, mas com algumas pequenas vantagens adicionais em termos da confiança associada. Trabalho em conjunto com Vera Tomazella (Departamento de Estatística, UFSCar) e Francisco Louzada-Neto (ICMC-USP).

A Distribuição Weibull-Poisson de Longa Duração na Presença de Covariáveis.

Valdemiro Piedade Vigas
UFSCar

Abstract: Em diversas aplicações da área de estatística a variável resposta é o tempo até a ocorrência de um evento, este tempo é denominado por tempo de falha ou tempo de sobrevivência. Também é comum que nem sempre é possível observar o valor exato desta variável e sim o valor mínimo ou máximo dos tempos. Interpretando essa situação como o tempo de vida de um sistema (em série ou em paralelo). Pode-se observar isto como uma competição entre fatores para verificar qual deles é o responsável pela falha. Especificamente, considera-se o caso de que não há informações sobre quais fatores foram responsáveis pela falha do sistema, ou seja, tanto o número como a identificação do(s) fator(es) que provocou (provocaram) a falha não são observáveis, mas sim o valor do tempo de vida mínimo ou máximo desses fatores, neste caso observável. Outra característica importante presente em dados de sobrevivência é a presença de uma fração de itens ou indivíduos que não estão sujeitos ao evento de interesse, considerados como imunes ou curados, nesta situação, são recomendados os modelos de sobrevivência de longa duração. Outra particularidade que se pode notar é a presença de covariáveis, sendo que essas podem influenciar na proporção de indivíduos curados em estudo. Uma forma eficiente de observar como se comporta o efeito das covariáveis é utilizar um modelo com a presença destas. De

acordo com as considerações citadas é proposto neste trabalho o modelo de longa duração Weibull-Poisson (LWP) na presença de covariáveis. Este modelo provém da distribuição Weibull-Poisson (Louzada, Bereta e Franco, 2011) o qual generaliza a distribuição exponencial-Poisson (EP) proposta por Kuss (2007) e acomoda dados numa estrutura de longa duração. Trabalho em conjunto com Francisco Louzada (ICMC/USP) Josmar Mazucheli (DES/UEM).

Um estudo de simulação comparativo para Testes Adaptativo Informatizado (TAI)

Vanessa Rufino da Silva
USP/UFSCar

Abstract: Nessa apresentação é mostrado um estudo de simulação de Testes Adaptativos Informatizados (TAI) para o modelo logístico de três parâmetros (ML3P) da Teoria de Resposta ao Item (TRI). Os parâmetros dos itens (dificuldade, discriminação e acerto ao acaso) foram fixados em valores conhecidos e as respostas dos indivíduos a cada item foram simuladas de acordo com o ML3P da TRI. As estimativas dos traços latentes foram obtidas por Máxima Verossimilhança (MV) e pelo método bayesiano Esperança a Posteriori (EAP) utilizando como critério de seleção de itens no TAI a Informação Total de Fisher. Os resultados mostram que os dois métodos geram boas estimativas da proficiência do indivíduo. No entanto, quando se tem um número pequeno de itens no banco, o EAP se mostrou mais eficaz com menor Erro Quadrático Médio, para quase todas as estimativas quando comparado com as de MV. Trabalho em conjunto com Mariana Cúri (ICMC-USP).

Estimação dos parâmetros do modelo Poisson bivariado inflacionado de zeros via algoritmo EM

Verônica Amparo Quispe Chire
UFSCar

Abstract: Na análise de dados de contagem, a ocorrência de muitos zeros é usual em diversas áreas, tais como epidemiologia, medicina, agricultura, indústria entre outras. Em tais situações, a distribuição Poisson inflacionada de zeros tem sido largamente aplicada. No caso bivariado, o excesso de zeros é também muito comum. Neste trabalho, consideramos uma versão inflacionada de zeros do modelo Poisson bivariado proposto por Holgate (1964). Para ilustrar a aplicação do modelo, será utilizado um conjunto de dados reais em que a estimação será determinada através do método máxima verossimilhança via algoritmo EM. Este algoritmo de estimação utiliza variáveis latentes e simplifica a inferência sobre os parâmetros. Trabalho em

conjunto com Luis Ernesto Bueno Salazar (DEs - UFSCar) e Vera Lucia Damasceno Tomazella (DEs - UFSCar).

Convergência e simulação do modelo de volatilidade estocástica de Heston

Vinícius de Castro Nunes de Siqueira
ICMC/USP

Abstract: Neste trabalho, utilizamos uma abordagem do tipo Euler-Maruyama para implementação de um método numérico de aproximação do modelo de volatilidade estocástica de Heston (veja, por exemplo Hobson (2004) [1] dado por

$$dS_t = S_t \left(\mu V_t dt + \sqrt{V_t} dW_t^1 \right) \quad (1)$$

$$dV_t = 2\lambda (\theta - V_t) dt + 2\sigma \sqrt{V_t} dW_t^2. \quad (2)$$

em que V_t corresponde ao termo de volatilidade e que é incorporado na equação do preço do ativo (??) e W_t^1 e W_t^2 são dois movimentos Brownianos possivelmente correlacionados com coeficiente de correlação constante ρ . Assumimos que s_0 e v_0 , as condições iniciais de S e V , são constantes positivas. Para um inteiro positivo k e para $j = 1, 2$ definimos os seguintes tempos de parada que podem ser encontrados em Leão e Ohashi (2013) [?] dados por $T_0^{k,1} : 0$ q.c e

$$T_n^{k,j} := \inf \left\{ T_{n-1}^{k,j} < t < \infty; \left| W_t^j - W_{T_{n-1}^{k,j}}^j \right| = 2^{-k} \right\}, \quad n \geq 1, \quad (3)$$

em que $W := (W^1, W^2)$ é um movimento Browniano bidimensional. O método de Euler-Maruyama para o sistema dado pelas equações (??) e (??) calcula as aproximações numéricas V^k para V e S^k para S através das equações

$$V_{T_n^k}^k = V_{T_{n-1}^k}^k + 2\lambda(\theta - V_{T_{n-1}^k}^k) \Delta T_n^k + 2\sigma \sqrt{|V_{T_{n-1}^k}^k|} \Delta W_{T_n^k}^2 \quad (4)$$

$$S_{T_n^k}^k = S_{T_{n-1}^k}^k + \mu S_{T_{n-1}^k}^k V_{T_{n-1}^k}^k \Delta T_n^k + S_{T_{n-1}^k}^k \sqrt{|V_{T_{n-1}^k}^k|} \Delta W_{T_n^k}^1 \quad (5)$$

em que $T_n^k \in \mathcal{T}$, e \mathcal{T} é a partição mais fina do intervalo $[0, T]$ gerada pelos tempos de parada definidos em (??).

Demonstramos a convergência forte do método numérico proposto e utilizamos técnicas de simulação Monte Carlo para precificar alguns tipos de opções de mercado. Trabalho em conjunto com Dorival Leão (ICMC-USP).

References:

1. D. Hobson. Stochastic volatility models, correlation, and the q-optimal measure. *Mathematical Finance*, 14(4):537-556, 2004.
2. D. Le ao and A. Ohashi. Weak approximations for wiener functionals. *The Annals of Applied Probability*, 23(4):1660-1691, 2013.

Modelo de Regressão Poisson Bivariado Para Previsão de Resultados de Futebol

Wesley Bertoli da Silva
UFSCar

Abstract: Diversas são as situações na qual o modelo de regressão Poisson bivariado encontra aplicações. De forma geral, podemos citar a modelagem conjunta da compra de dois produtos, o cross merchandising, além de estudos relacionados à combinação de índices econômicos. Em outros segmentos, podemos destacar a modelagem de dados epidemiológicos, bem como a análise de incidência de doenças em um mesmo indivíduo. No contexto esportivo, o modelo de regressão Poisson bivariado pode ser utilizado para modelar o número de gols marcados por duas equipes oponentes, em esportes como o futebol, o polo aquático e assim por diante. Em particular, a modelagem de dados de futebol via regressão é abordada por Lee (1997), que considera o ajuste de um modelo Poisson duplo; por Dixon e Coles (1997) que consideram uma abordagem voltada para o mercado de apostas, baseada em um modelo Poisson duplo modificado; por Karlis e Ntzoufras (2003) que propõem abordagens baseadas no modelo Poisson bivariado; dentre outros. Em virtude disso, neste trabalho propomos uma aplicação do modelo Poisson bivariado aos dados da Série A do Campeonato Brasileiro de 2012. Apresentaremos brevemente a estrutura deste modelo, bem como suas principais características. A estimação via máxima verossimilhança dos coeficientes de regressão é considerada por meio do método numérico de otimização BFGS. Além disso, a qualidade preditiva do modelo ajustado poderá ser acessada por meio da medida de DeFinetti. Trabalho em conjunto com Luis Ernesto Bueno Salasar e Adriano Kamimura Suzuki.

Seleção de variáveis em uma classe de modelos série de potências modificadas inflacionados de zeros

Willian Luís de Oliveira
UFSCar

Abstract: Neste trabalho, consideramos uma classe de modelos de regressão série de potências modificadas inflacionados de zeros, que tem sido bastante explorada nos

dias atuais. Nosso objetivo é desenvolver um procedimento de seleção de variáveis para esta classe de modelos utilizando ambas as abordagens clássica e Bayesiana. Alguns métodos são considerados para esta classe de modelos. Nós adaptamos ao modelo série de potências modificadas inflacionados de zeros os métodos clássicos AIC e BIC e os métodos Bayesianos CPO, DIC, EAIC, EBIC e medida L. Usando as linguagens de programação OX e R, alguns conjuntos de dados simulados são utilizados para ilustrar os métodos e comparar suas performances. Trabalho em conjunto com Carlos Alberto Ribeiro Diniz (UFSCar) e Marinho G. Andrade (ICMC-USP).

Participants

Adriano Kamimura Suzuki, ICMC/USP
 Afrânio Vieira, UNB
 Adriano Polpo, UFSCar
 Alan Carvalho Oliveira, USP
 Álaze Gabriel do Breviário, UFSCar
 Alexandre Hiroshi Watanabe, USP/UFSCar
 Alexandre Pitanguí Calixto, UFSCar
 Alina Von Davier, ETS
 Allan Gustavo Brigola, UFSCar
 Amanda Buosi Gazon, USP/UFSCar
 Amanda Cristina Estevam, UFSCar
 Amanda Morales Eudes, UFSCar/USP
 Amélia M. Correia Fernandes, ICMC/USP
 Ana Carolina Ottaviani, UFSCar
 André Ricardo Tavares, ICMC/USP
 Arthur Necchi Corrêa, UFSCar
 Bao Yiqi, UFSCar
 Bruno Lucian Gonçalves da Costa, UFF
 Carlos A. de Bragança Pereira, IME/USP
 Carlos Alberto Ribeiro Diniz, UFSCar
 Cherlynn Daniela da Silva Arce, UFSCar
 Cristel Ecaterin Vera Tapia, USP/UFSCar
 Daiane Aparecida Zuanetti, USP/UFSCar
 Daniele C. Tita Granzotto, USP/UFSCar
 Danielle G. de Oliveira Prado, UFLA
 Dante R. Baldeon Molleda, UFSCar/USP
 Debora Delbem, UFSCar
 Dorival Leão Pinto Junior, ICMC/USP
 Eder Angelo Milani, UFSCar
 Edimar Izidoro Novaes, UNESP/Botucatu
 Eduardo Sartoretto, USP
 Edwin Chaiña Cahui, UFSCar
 Emanuel de Cássio dos Santos, UFSCar
 Erlandson Ferreira Saraiva, UFSCar
 Ernesto Mordecki, Universidad de la República
 Evandro Luiz de Sousa Jackson, USP/UFSCar
 Eveliny Barroso da Silva, UFSCar
 Fábio Prativiera, UFSCar
 Felipe Cesar Marangoni Santos, UFSCar/USP
 Filidor E. Vilca Labra, IMECC/UNICAMP
 Francisco Louzada Neto, ICMC/USP
 George Lucas Moraes Pezzott, UFSCar
 Giovani Zamboim de Vasconcellos, UFSCar
 Gladys Dorotea Cacsire Barriga, UNESP
 Guaraci de Lima Requena, UFSCar
 Guilherme Barreto Fernandes, USP/UFSCar
 Helena Martins Oliveira, IME/USP
 Hiron Pereira Farias, UFG
 Hua-Hua Chang, University of Illinois
 Jean Mimar Santa Cruz Yabarrena, USP
 Jhon Franky Bernedo Gonzales, UFSCar
 João Victor de Souza Cunha, IFSC/USP
 Jociene Santana Pimentel, UTFPR
 Jorge Luis Bazan, USP
 Jorge Oishi, Statsol
 José Galvão Leite, UFSCar
 Antonio M. Bermúdez, Benemérita Universidad Autónoma de Puebla
 José Augusto Fioruci, UFSCar
 Juan Pablo Mamani Bustamante, USP/UFSCar
 Juliana Ottaviani, UFSCar
 Jurandir Prazeres Filho, UFSCar
 Juvêncio Santos Nobre, UFC
 Karina Bindandi Emboaba de Oliveira, USP
 Katherine E. Coaguila Zavaleta, UFSCar
 Lia Hanna Martins Morita, UFSCar
 Lorena Yanet Cáceres Tomaya
 Lucas de Freitas Rosa, UFSCar
 Luciana Leite de Mattos Alcantara, UFF
 Luis Aparecido Milan, UFSCar
 Luis Ernesto Bueno Salazar, UFSCar
 Luis Felipe Dos Santos Esteves, UFES
 Luiz Alberto Amaral Nardi, UNESP/Botucatu
 Luiz Clementino Campos, CAIXA
 Luiz Roberto Martins Pinto, UESC
 Manuel Saturnino, FMRP/USP
 Marcelo Hartmann, USP/UFSCar
 Márcia Aparecida Centanin Macera, UFSCar
 Marcio Diniz, UFSCar
 Marco Antonio H. Carnaz, UFSCar
 Marco Aurélio de Menezes Franco, UFSCar
 Mariana Cúri, ICMC/USP
 Mariana Ottaviani, UFSCar
 Marinho Gomes de Andrade Filho, ICMC/USP
 Mário de Castro Andrade Filho, USP
 Mauricio Castro, Universidad de Concepción
 Mauro Ribeiro de Oliveira Júnior, UFSCar
 Michelli Barros da Silva, UFCG

Orlando da Silva Junior, UFABC
Pablo Martin Rodriguez, ICMC/USP
Patrícia Viana da Silva, UFSCar
Rafael Grisi, CMCC/UFABC
Reiko Aoki, ICMC/USP
Renato Alexandre Silva Junior, USP
Renato Frazzato Viana, USP/UFSCar
Renato Gava, UFSCar
Ricardo Ehlers, ICMC/USP
Ricardo Felipe Ferreira, UFSCar/USP
Ricardo Rocha, UFSCar
Roberto Vila Gabriel, UNB
Robson José Mariano Machado, UFSCar
Rodrigo de Andrade Ferrazza, UNESP/Botucatu
Salimeh Yasaei Sekeh, UFSCar
Sandro Gallo, UFRJ
Suely Aparecida Kfour, FSP/USP
Taiuani Marquine Raymundo, USP
Teresa Cristina Martins Dias, UFSCar
Thales Akira Matsumoto Ricarte, ICMC/USP
Tzioufas Achillefs, Universidad de Buenos Aires
Valdemiro Piedade Vigas, UFSCar
Vanessa Masitéli, UFSCar/USP
Vanessa Rufino da Silva, USP/UFSCar
Vera Lucia Damasceno Tomazella, UFSCar
Verônica Amparo Quispe Chire, UFSCar
Vicente Garibay Cancho, ICMC/USP
Victor Leiva, Universidad de Valparaíso
Victor Sae Hon Sung, UFSCar/USP
Vinícius de Castro N. de Siqueira, ICMC/USP
Vinicius Raniero Angelo, UFMT
Vitor Alex Alves de Marchi, UFSCar
Wesley Bertoli da Silva, UFSCar
Willian Luís de Oliveira, UFSCar

