

# 7th Workshop on Probabilistic and Statistical Methods

February 13–15, 2019

ICMC/USP, São Carlos, SP, Brazil

## Poster Presentations

ICMC/USP - DEs/UFSCar

## Poster Session 1

**Alex Rodrigo dos Santos Sousa, Nancy Lopes Garcia, Brani Vidakovic (Universidade Estadual de Campinas)**

Encolhimento bayesiano sob priori Beta de coeficientes de ondaletas em modelos com erros gaussianos e positivos

*Abstract: Considere o problema de regressão não paramétrica de estimação de curva por meio da observação de pontos desta curva. Métodos de encolhimento de coeficientes de ondaletas são aplicados aos dados no domínio das ondaletas para redução de ruído para que a função possa ser estimada por expansão em bases de ondaletas. O presente trabalho propõe uma abordagem bayesiana de encolhimento de coeficientes de ondaletas com a utilização da distribuição beta com suporte em  $(-m, m)$  como distribuição a priori para os coeficientes das ondaletas em modelos com erros aleatórios aditivos gaussianos e positivos. Fórmulas explícitas para casos particulares das regras de encolhimento são obtidas, propriedades estatísticas como viés, risco clássico e bayesiano das regras são analisadas e performances das regras propostas são obtidas em estudos de simulações envolvendo as funções testes de Donoho-Johnstone. Aplicações em conjuntos de dados reais nas áreas de Espectrometria e Spike Sorting são feitas.*

**Amanda Morales Eudes D'Andrea, Vera Lucia Damasceno Tomazella (UFSCar/USP)**

Fragilidade não paramétrica aplicada para dados de múltiplos sistemas reparáveis

*Abstract: Em sistemas reparáveis, é muito importante considerar a escolha do modelo apropriado para modelar o tempo até a ocorrência do evento de interesse. Em muitas situações, o processo de Poisson não-homogêneo modela muito bem os dados. Mas a heterogeneidade entre os sistemas deve ser incluída, isso pode ser feito pelo modelo de fragilidade, isto é, a incorporação de uma variável aleatória não observada multiplicando a taxa de ocorrência de falhas. A proposta deste trabalho é considerar a fragilidade de forma não paramétrica e apresentar uma aplicação com um conjunto de dados real.*

**Ana Paula Jorge do Espirito Santo, Vicente Garibay Cancho (UFSCar/USP)**

A new survival model for lifetime with long-term survivors and unobserved heterogeneity

*Abstract: The Frailty models are often used to modeling unobserved dependence*

and heterogeneity in the individual in survival data, which can be modeled as an unobserved random variable acting multiplicatively on the baseline hazard function. In some situations, it may be appropriate to consider discrete frailty distributions, where individuals with long-term survival times include zero frailty (immune or cured). In this work, we develop a new survival model induced by discrete frailty with Katz distribution which can account for overdispersion, equidispersion and underdispersion. The new model encompasses as special case the mixture cure rate model, promotion cure rate model and cure rate model with dispersion, besides has the proportional hazards structure when covariates are modeled through  $\theta$  that allows a study of the risk ratio independent on time  $t$ . We construct a regression model to evaluate the effects of covariates in the cured fraction, we discussed inference aspects for the proposed model in a classical approach, an EM algorithm is then developed for determining the maximum likelihood estimates of the parameters. Finally, the modeling is fully illustrated on a data set on colorectal cancer in which we observe that the therapy with Levamisole+5-FU increases the lifetime of patients and cured fraction.

### Átila Prates Correia (ICMC-USP)

Kolmogorov extended axioms and negative probabilities

*Abstract:* As an attempt to extend the concept of probability to negative numbers, the content to be exposed in this poster makes use of measure theory in order to achieve this goal. More precisely, given a measurable space  $(\Omega, \Sigma)$ , we associate to it a finite signed measure  $P$  under convenient restrictions (axioms) which turn feasible the concept of negative probability. Although we do not provide an interpretation for the negative probabilities, we propose the structure such concept should fit in.

### Breno Gabriel da Silva , Yana Miranda Borges , Brian A. Ribeiro de Melo (Universidade Estadual Maringá)

Ajuste de extensões da distribuição weibull para dados meteorológicos do estado de Goiás

*Abstract:* Quando o objetivo é descrever situações cujo interesse é analisar o tempo até a ocorrência de um evento, o ponto de partida é escolher uma distribuição de probabilidade adequada ao problema, como nos casos de analisar o tempo até a falha de um componente eletrônico, crescimento de uma determinada bactéria, evolução de alguma doença, entre outros. Se a distribuição candidata não se ajustar bem aos dados, temos então que as estimativas serão imprecisas. Este trabalho tem como objetivo avaliar a capacidade de ganho real no ajuste de onze distribuições de probabilidade derivadas da distribuição Weibull, uma vez que nos últimos tempos foram

sendo criadas inúmeras extensões desta. Em relação aos materiais, utilizaremos um conjunto de dados de precipitação, somente em que esta é maior que zero, do estado de Goiás no período de maio de 2017 a maio de 2018. Em relação aos métodos, utilizaremos o teste não-paramétrico de Kolmogorov-Smirnov (KS), que verifica se determinada distribuição teórica ajusta-se bem à distribuição empírica dos dados e o critério de informação de Akaike (AIC) para a seleção do modelo adequado. Assim, com o modelo selecionado pretendemos em trabalhos futuros verificar a influência de covariáveis no índice de precipitação de uma determinada região.

### **Camila Ozelame, Anderson Ara, Francisco Louzada (UFSCar/USP)**

Redes bayesianas para classificação: diferentes métodos e comparação

*Abstract: Em problemas de classificação, inúmeras técnicas estatísticas e computacionais podem ser aplicadas, uma possibilidade é dada pela metodologia de Redes Bayesianas (RBs) com estruturas específicas, portanto, também conhecidas como classificadores Bayesianos. As RBs unem a teoria de grafos e de probabilidades atuando com maestria em dois dos objetivos que dominam a análise de dados, a descrição e a predição. As RBs conseguem identificar a estrutura, mais provável, das variáveis analisadas, bem como a maneira que interagem umas com as outras. Utilizando a noção gráfico-teórica de  $d$ -separação são capazes de identificar grupos de variáveis que possuem suposição de independência condicional e fornecem a distribuição a posteriori de uma variável resposta dado um conjunto de variáveis explicativas. Diversos classificadores são apresentados na literatura, cada uma com uma abordagem diferente a respeito das probabilidades conjunta e condicional entre as variáveis estudadas. Dentre elas destacam-se: Naïve Bayes (NB); Tree Augmented Naive Bayes (TAN); Bayesian Network Augmented Naive Bayes (BAN); General Bayesian Network (GBN); K-Dependence Bayesian Network (KDB); Averaged One-Dependence Estimator (AODE). Neste trabalho, realizamos uma comparação da capacidade preditiva de tais classificadores utilizando benchmarkings de dados reais e artificiais. Para quantificar a qualidade das estruturas propostas são utilizadas quatro medidas de performance de desempenho: sensibilidade, especificidade, acurácia e o coeficiente de correlação de Matthew.*

### **Carolina Grejo, Pablo Martin Rodriguez (UFSCar/USP)**

A general Maki-Thompson model with directed inter-group interactions

*Abstract: In this work we considered a new extension for the Maki-Thompson model which incorporates inter-group directed contacts. We obtain limit theorems for the remaining proportion of ignorants and discuss some examples and possible applications.*

**Cristel Ecaterin Vera Tapia (UFSCar/USP)**

Estimation of the number of communities in the stochastic block model

*Abstract: The Stochastic Block Model was introduced by Holland et al. (1983) and falls in the general class of random graphs. In this model, the nodes are classied in groups or communities, such that, considering to each node in the graph a associated latent discrete random variable, describing its community label, the probability of a edge connecting two nodes, depends only on the values of the latent variables. In this context, Andressa Cerqueira and Florencia Leonardi (2018) proposed the Krichevsky-Tromov estimator for the number of communities in the stochastic block model. Based on these results, we consider in this work a simple extension, to include the possibility of existence of a nite number of edges between pairs of nodes.*

**Danilo Sarti, Carlos Tadeu dos Santos Dias (USP - LCE)**

Fitting probabilistic models to data and risk exposure

*Abstract: This paper aims to show how a computational methodology can be applied to fitting different probabilistic models to data. Simulations of random variables are used to generate datasets. Cullen and Frey graphical methodology is explored as a guideline for the candidate distributions to model the data. In order to build such graphics higher order moments such as kurtosis and symmetry are used and discussed. Parametric distributions are fitting to the data using likelihood, method of moments as long as other methods focusing on fitting optimization. Several goodness of fitness measures are applied and compared for discussion of their uses and implications. An example of how such method can be used for assessing exposure to risk is made via a R dataset from packages available at CRAN website.*

**George Lucas Moraes Pezzott, José Galvão Leite, Luis Ernesto Bueno Salasar (UFSCar/USP)**

Spatial capture-recapture model for open population

*Abstract: In this work we propose a spatial capture-recapture model to estimate the number of animals in an open population. The statistical model conform to data obtained through individual tag capture-recapture sampling performed in different areas within the habitat, taking into account the rates of births and deaths during the study period, the geographical locations of the catches and the movement of the animals between the sampling regions. Therefore, we was able to estimate the number of the live animals in each occasion and sampling region as well as the size of the area of movement of the species. We propose a Bayesian approach to the inferential process*

*and derive a MCMC algorithm from simple computational implementation, from the use of augmented data techniques. Our proposed methodology is illustrated in a real capture-recapture data set of arachnids in a cave.*

### **Gilson Yuuji Shimizu, Rafael Izbicki (UFSCar/USP)**

#### Conformal Prediction via Densidade Condicional

*Abstract: Os métodos de machine learning tem como principal finalidade a predição de uma variável resposta para um dado conjunto de variáveis explicativas. Geralmente, estamos interessados apenas na esperança condicional desta variável resposta. Contudo, quando a distribuição da variável resposta é complexa, apenas a esperança condicional pode ser insuficiente para descrever a incerteza desta predição. Neste trabalho, nós propomos um método para estimar uma banda de predição, num contexto de regressão, onde nenhuma suposição é feita sobre a distribuição dos dados e qualquer método de machine learning pode ser utilizado. O método conformal prediction é combinado com a estimação de densidade condicional para a construção desta banda. Exemplos com dados simulados são apresentados para demonstrar a eficácia do método. Palavras chave: machine learning, conformal prediction, banda de predição e densidade condicional.*

### **Glauber Márcio Silveira Pereira, Carlos Alberto Ribeiro Diniz (UFSCar)**

#### Distribuição COM-Poisson generalizada parcialmente correlacionada

*Abstract: Desenvolvemos a distribuição COM-Poisson generalizado parcialmente correlacionado (CPGPC), é uma generalização da distribuição Poisson generalizada parcialmente correlacionada desenvolvida por (Luceño, 1995) e COM-Poisson desenvolvida por Shmueli, G Minka & Boatwright (2005). Desenvolvemos a estimação dos parâmetros com uma abordagem clássica. Os estimadores de máxima verossimilhança e estimadores de momentos dos parâmetros são determinados. Intervalos de confiança normal e bootstrap são construídos para os parâmetros da distribuição. Fazemos simulação para os casos citados.*

### **Isis F. Mascarin, Katiane S. Conceição (UFSCar/USP)**

#### Distribuições discretas zero-modificadas para modelar dados de contagem zeros faltantes

*Abstract: Em diversas situações práticas onde o conjunto de dados constitui-se de observações de contagem, é possível notar a ausência de observações zero. A ocor-*

rência deste cenário pode ser devido à probabilidade nula de ocorrência de tal valor ou de algum problema de amostragem. No caso em que a probabilidade de observação zero é positiva, deve-se assumir uma distribuição zero-deflacionada em vez de uma zero-truncada, para obter de forma correta as estimativas dos parâmetros da distribuição que explica adequadamente o comportamento destes dados, bem como a frequência esperada de zeros. Neste trabalho, o interesse é fazer tais estimações por meio do algoritmo EM, utilizando distribuições zero-modificadas adequadas para o estudo. O conteúdo inclui introdução às famílias de distribuições Série de Potência (PS) e Série de Potência Zero-Modificada (ZMPS), incluindo sua versão hurdle, a partir da qual foram desenvolvidos os cálculos para obtenção dos estimadores de máxima verossimilhança dos parâmetros da distribuição ZMPS. Além disso, apresenta-se alguns resultados iniciais obtidos de simulação aplicada à distribuição Poisson, que pertence à família PS.

**Jaime Phasquinel Lopes Cavalcante, Luciana Moura Reinaldo (Universidade Federal do Ceará - UFC)**

Modelos lineares generalizados (MLGs) e sua aplicação em Ciências Atuariais

*Abstract: A utilização de métodos estatísticos na rotina da Ciência Atuarial tem desenvolvido, historicamente, um papel central, tanto nos assuntos teóricos quanto práticos. Nesse sentido, corroborando com estudos outrora publicados, o presente estudo possui como objetivo principal demonstrar a aplicação da metodologia dos Modelos Lineares Generalizados com foco em uma problemática atuarial. A justificativa para o estudo surge do fato de que há uma vasta área de aplicações para o MLG (HABERMAN; RENSHAW, 1996), mas pouco exploradas, especialmente no Brasil. Diante do exposto, considerou-se a base contida em Kaas et al. (2008), que reflete a experiência anual de um portfólio de seguros de automóveis. Com isso, buscou-se relacionar a frequência de sinistros aos fatores de risco: sexo, região, tipo do carro, situação laboral. Portanto, ajustou-se um MLG com relação média-variância do tipo Poisson com função de ligação canônica. Ademais, são apresentados os diagnósticos do modelo que confirmam o bom ajuste da modelagem e a análise dos desvios. Finalmente, foi possível determinar que, em média, um motorista com os piores fatores, em comparação com aquele que tem a melhor combinação, realiza um sinistro a cada 6,6 anos.*

**João Vitor Magri da Silva, Elisangela Aparecida da Silva Lizzi (Universidade Tecnológica Federal do Paraná)**

Modelagem bayesiana espacial aplicada à taxa de homicídios por violência nas unidades federativas brasileiras.

*Abstract: Introdução: A taxa de homicídios e índices de violência no Brasil são elevados, quando comparados a taxas internacionais, em regiões e estratos específicos da população estas taxas são superiores à países latino-americanos envolvidos em guerrilhas internas e conflitos militares. Objetivo: Modelar a relação entre taxas de homicídios por violência, IDH e o efetivo de policiamento militar distribuídos em todas as 27 unidades federativas do Brasil. Métodos: Os dados relativos à taxa de homicídios por violência, usou o filtro de diferentes causas, como: violência por gênero, mortes violentas por causa indeterminada e óbitos por armas de fogo. Neste cenário, foi proposta uma modelagem bayesiana espacial, usando-se modelo espacial com efeito aleatório BYM, onde modelou-se a dependência espacial com matriz de adjacência de ordem 1 que representa a localização de cada UF e seus respectivos vizinhos de primeira ordem. O número de policiais por 100 mil habitantes e o IDH entrou como preditores. O modelo foi implementado com apoio computacional do software R e utilizou métodos de aproximação de Laplace (INLA). Resultados: Não foi possível estabelecer uma relação entre a taxa de homicídios e o policiamento, porém em relação ao IDH quanto menor o índice maior a taxa de violência naquela região.*

**Jonathan Kevin Jordan Vasquez, Josemar Rodrigues (USP)**

Modelo de Sobrevivência em dois estágios com taxa de cura

*Abstract: Este projeto foi motivado pela análise dos dados de sinusite em pacientes infectados pelo HIV (GIOLO; COLOSIMO, 2006). A estimativa de Kaplan-Meier da função de sobrevivência sugeriu a existência de pacientes imunes a sinusite tornando-se inadequado a utilização dos modelos tradicionais em Análise de Sobrevivência. Para verificar se a infecção pelo HIV aumenta o risco da ocorrência da sinusite, propomos neste projeto o modelo Exponencial e Weibull em dois-estágios (Rodrigues, J. Teoria Unificada de Análise de Sobrevivência, 2009) com taxa de cura.*

**Katy Rocio, Alex Mota, Vera Tomazella, Vicente Calsavara (UFSCar/USP)**

Modelo de fragilidade discreta zero inflacionado poisson

*Abstract: Os modelos de fragilidade são utilizados para a modelagem de heterogeneidade na análise de dados de sobrevivência. Na análise desses dados a distribuição da fragilidade, em geral, é assumida contínua. Entretanto, existem algumas situações nas quais é apropriado considerar a fragilidade distribuída discretamente, por exemplo, quando a heterogeneidade dos tempos de sobrevivência surge por causa da presença de um número aleatório de falhas por unidade ou pela causa da exposição a danos em um número aleatório de ocasiões. Neste trabalho estendemos os modelos*

*de fragilidade de riscos proporcionais permitindo distribuições discretas, em particular a distribuição Zero Inflacionado Poisson (ZIP). Neste contexto podemos observar a possibilidade de indivíduos com fragilidade zero que corresponde a um modelo de falha limitado que contém uma proporção de unidades que nunca falham (sobreviventes de longa duração ou modelo com fração de cura).*

## Poster Session 2

**Lucas Leite Cavalaro, Gustavo Henrique de Araujo Pereira (UFSCar/USP)**  
Comparação de métodos de seleção de variáveis em MLGD com covariáveis correlacionadas

*Abstract: Os modelos lineares generalizados duplos (MLGD), diferentemente dos modelos lineares generalizados (MLG), permitem o ajuste do parâmetro de dispersão da variável resposta em função de variáveis preditoras, aperfeiçoando a forma de modelar fenômenos. Desse modo, os mesmos são uma possível solução quando a suposição de que o parâmetro de dispersão constante não é razoável e a variável resposta tem distribuição que pertence à família exponencial. Considerando nosso interesse em seleção de variáveis nesta classe de modelos, estudamos o esquema de seleção de variáveis em dois passos proposto por Bayer e Cribari-Neto (2015) e, com base neste método, desenvolvemos um procedimento para seleção de variáveis em até " $k$ " passos. Para avaliar o desempenho do nosso procedimento em dados que apresentam comportamento próximo ao observado na prática, realizamos estudos de simulação de Monte Carlo em MLGD considerando covariáveis com diferentes correlações. Os resultados obtidos indicam que o nosso procedimento para seleção de variáveis apresenta, em geral, performance semelhante ou superior à das demais metodologias estudadas sem necessitar de um grande custo computacional.*

**Luís Felipe Barbosa Fernandes, Evandro Marcos Saidel Ribeiro (USP)**  
Modelagem de crédito com a técnica de redes bayesianas aplicadas ao segmento de indústrias de alimentos e bebidas

*Abstract: O trabalho consiste na elaboração de um modelo Bayesiano capaz de prever o risco de inadimplência de uma empresa com base em um conjunto de oito indicadores financeiros. Para isso, foi usado um banco de dados da Serasa Experian formado por um conjunto de 368 empresas, com balanços patrimoniais para três períodos consecutivos. O trabalho divide-se em duas etapas: modelagem e construção do aplicativo. A modelagem foi realizada através das técnicas de padronização dos dados, categorização e seleção do modelo. Após essa etapa, a Rede Bayesiana resultante foi elaborada através de um aplicativo. Esse aplicativo, desenvolvido através do pacote Shiny, do RStudio, permite que um usuário obtenha o risco de uma empresa, a partir da atribuição de valores para os oito indicadores financeiros selecionados. O usuário deve digitar os valores dos oito indicadores financeiros da empresa em questão. Após digitar os valores, o aplicativo exibe os valores digitados padronizados e categorizados em uma tabela visível ao usuário. Os valores categorizados são submetidos à rede, que calcula a probabilidade da empresa pertencer a cada uma das 22 classes de risco. A classe com maior probabilidade é a classe de risco da empresa.*

**Luiz Carlos Medeiros Damasceno, Luís Aparecido Milan (IFNMG/UFSCar/USP)**  
Modelos de mistura aplicados a dados com dependência espacial

*Abstract: O objetivo desse poster é apresentar instrumentos para inferência para modelos de misturas com dependência espacial entre as variáveis usando metodologia de M.V. e/ou bayesiana. Entre os aspectos a serem tratados esta a seleção de modelos, estimação via MCMC, bem como verificação de performance. Serão realizados estudos de simulação para verificar a performance dos métodos propostos e comparação com métodos propostos na literatura. Os métodos propostos também serão aplicados a dados reais.*

**Luiz Gabriel Fernandes Cotrim, Daiane Aparecida Zuanetti (UFSCar/USP)**  
Inferência Bayesiana para Modelos de Mistura de Regressão: uma aplicação em dados educacionais brasileiros

*Abstract: É comum que a variável resposta não se relacione com as covariáveis de forma homogênea para toda a população. Uma maneira interessante de modelar esses dados é considerar que a heterogeneidade observada surgiu devido ao fato de que a população é composta por  $K$  subpopulações e que as variáveis a serem analisadas se relacionam de maneira diferente em cada subpopulação. Se as subpopulações são desconhecidas, podemos assumir que a distribuição da variável resposta condicionada as covariáveis é um modelo de mistura. Denominamos esta classe de modelos de modelos de mistura de regressão. Apresentamos métodos de estimação e seleção Bayesianos para o caso em que temos uma mistura de regressões normais, comparamos os resultados ao algoritmo EM e aplicamos a metodologia em dados educacionais brasileiros, considerando o índice de desenvolvimento da educação básica (IDEB) como nossa variável de interesse.*

**Marcelo da Silva, Jorge Bazán (UFSCar/USP)**

Validação da matriz  $Q$  em modelos da teoria da resposta ao item multidimensionais

*Abstract: A matriz  $Q$  é um componente bastante simples e intuitivo utilizado originalmente por uma nova classe de modelos de variáveis latentes multidimensionais, conhecida como modelos de diagnóstico cognitivo (MDC) com o objetivo de especificar a relação item-traço em um instrumento de medição. Recentemente, a matriz  $Q$  foi incorporada nos modelos da teoria da resposta ao item multidimensionais (TRIM). A construção da matriz  $Q$  é tipicamente feita por especialistas no tema dos itens, ou seja, é um processo subjetivo que pode implicar em equívocos e, conseqüentemente, resultar em importantes implicações práticas. Assim, a verificação da exatidão da*

*matriz  $Q$  faz-se necessário tanto em MDC como nos modelos da TRIM. Baseando-se no método para validar teoricamente as especificações da matriz  $Q$  apresentado por Jimmy de la Torre em MDC, propomos um método de validação da matriz  $Q$  em modelos da TRIM. Para isso, definimos e comparamos o comportamento de alguns critérios para buscar uma matriz  $Q$  adequada aos dados, adaptamos um algoritmo de busca de matriz chamado Algoritmo de Troca por Ponto e realizamos um estudo de simulação para avaliar o desempenho do método proposto.*

**Naiara Caroline Aparecido dos Santos, Breno Gabriel da Silva, Yana Miranda Borges, Brian A. Ribeiro de Melo (Universidade Estadual de Maringá)**

Modelo de regressão quasi-poisson para dados de tentativas de suicídio por intoxicação exógena do estado de São Paulo

*Abstract: Segundo a Organização Mundial da Saúde (OMS), a intoxicação exógena é um dos três principais motivos de tentativas de suicídios no mundo. Assim, objetiva-se verificar qual grupo (sexo e faixa etária) possui a maior taxa de tentativas de suicídio por intoxicação exógena. Desta forma, foi realizado um estudo dos registros de casos do Estado de São Paulo, Brasil. Os dados foram extraídos do Sistema de Informações sobre Mortalidade (DATASUS) e características populacionais do Instituto Brasileiro de Geografia e Estatística (IBGE) no ano de 2017. As análises estatísticas foram realizadas no software R versão 3.5.0, em que utilizou-se o modelo de Regressão Poisson, também conhecido como Modelo Log-Linear de Poisson, o qual pertence a família de Modelos Lineares Generalizados. Inicialmente ajustou-se um modelo log-linear de Poisson, que apresentou fortes evidências de superdispersão, indicando que o modelo não era o mais apropriado. Sendo assim, ajustou-se o modelo quasi-Poisson, de modo a acomodar a variabilidade presente, mostrando-se adequado. Verificou-se que homens cometem suicídio 0.376 vez mais que as mulheres, e que a maior taxa de tentativa de suicídio por intoxicação ocorre entre 20 e 39 anos. Por outro lado, a menor taxa ocorre em indivíduos acima de 80 anos.*

**Oilson Alberto Gonzatto Junior, Marcos Jardel Henriques, Camila Sgarioni Ozelame, Anderson Ara, Mariana Curi, Francisco Louzada Neto (UFSCar/USP)**

Um Modelo de regressão com respostas Beta Inflacionada e componentes principais

*Abstract: A distribuição Beta é comumente utilizada para descrever o comportamento de uma variável aleatória com suporte em um intervalo aberto e, em particular, quando a problemática envolve uma proporção, o intervalo  $(0,1)$ . Essa distribuição*

*pode ser estendida com a adição apropriada de pontos de interesse ao seu suporte, o que caracteriza a distribuição Beta Inflacionada. Se o objetivo é examinar a existência e quantificar a influência de fatores que atuam sobre o comportamento de uma proporção de interesse, o modelo de Regressão Beta Inflacionado é uma alternativa interessante. Se houver uma alta quantidade de covariáveis possivelmente correlacionadas, pode-se associar o modelo de regressão com a técnica das componentes principais. Para ilustrar a aplicação de tal metodologia, um conjunto de dados reais foi utilizado.*

**Oscar Holguín Villamil, Jonas Rafael Dos Santos (Universidade Estadual Paulista "Júlio de Mesquita Filho")**

IRAMUTEQ y ATLAS.TI en la descomposición genética como herramientas para el análisis de contenido de la práctica docente y de políticas públicas: aplicación del análisis de correspondencia.

*Abstract: El trabajo describe el proceso de categorización, clasificación y la implementación de la técnica de estadística multivariada de análisis de correspondencia en el desarrollo del análisis de primer orden que arrojan los datos obtenidos en una investigación de tipo bibliográfico cuyo objetivo por una parte es la caracterización de la práctica educativa de docentes universitarios en la experiencia de la especialización en educación en tecnología de la UFScar y los componentes del concepto EaD apropiados por un experto investigador del campo. Y por otra la identificación de patrones de comparación entre programas de política educativa de primera infancia implementadas en Brasil y en Colombia. Dada la naturaleza de rastreo documental y sistematización de experiencias; se desarrolla un planteamiento inferencial de descomposición genética del fenómeno lingüístico y comunicativo que soporta los dos estudios y desde los cuales es posible el trabajo analítico de la estadística; para ello se implementan los programas ATLAS.ti e IRAMUTEQ, los cuales permiten almacenar todos los datos, la codificación, categorización y el análisis inferencial de los resultados obtenidos a partir de la descomposición genética como opción de análisis de contenido y modelación en investigación cualitativa.*

**Patty Mercedes Arce Flores (Pontificia Universidad Católica del Perú)**

Estimativas para la probabilidad de eventos sorpresivos en una cadena de Markov.

*Abstract: En este trabajo estudiaremos el siguiente aspecto sobre cadenas de Markov con espacios de estados finitos  $S$ . Denotaremos por  $\tau(y)$  el tiempo que le toma a la cadena en llegar por primera vez a un estado  $y$ . Fijado un tiempo  $t$  y un estado inicial  $x$ , buscamos estimar la probabilidad de que la cadena llegue por primera*

vez al estado "y" en "t" pasos, es decir buscamos estimar el valor de  $Px(\tau(y) = t)$ . En este trabajo enunciaremos cotas superiores para esta probabilidad, que dependen del tamaño del espacio de estado y el tiempo "t" y no tanto de las probabilidades de transición que tiene la cadena. Estas estimativas son Obtenidas por J. Norris et y nuestro trabajo de tesis se basa en ese artículo. Las estimativas que presentaremos en este trabajo son las siguientes: (a) Para una cadena de Markov con n estados,  $Px(\tau(y) = t) \leq n * t$ . (b) En una cadena de Markov reversible con n estados,  $Px(\tau(y) = t) \leq (2n * t)^{(1/2)}$  para  $t \geq 4n + 4$  (para un tiempo "t" suficientemente grande). (c) Para un camino aleatorio sobre un grafo simple con  $n \geq 2$  vertices,  $Px(\tau(y) = t) \leq 4e(\log n) * t$ . En dicho artículo se construyen ejemplos que muestran que estas cotas están muy cerca de ser óptimas.

### **Roberta de Souza, Carlos A. R. Diniz (UFSCar/USP)**

Modelo de regressão geométrico de ordem k: uma aplicação em operações de crédito.

*Abstract: Atrasos em operações de crédito, como no pagamento de consecutivas parcelas de empréstimo, faz com que o cliente se torne inadimplente e possa ser transferido do sistema bancário para uma empresa de cobranças. Algumas políticas internas bancárias consideram 90 dias em atraso o tempo para um cliente se tornar inadimplente, o que corresponde a três parcelas consecutivas não pagas. A sequência de parcelas pagas em dia e em atraso, até que três parcelas consecutivas em atraso ocorram, pode ser observada numa sequência de respostas binárias e a quantidade de respostas desta sequência representa uma variável com distribuição geométrica de ordem k (Philippou & Muwafi, 1980). Neste trabalho, um modelo de regressão geométrico de ordem k foi proposto para a avaliação de covariáveis na probabilidade de atraso da parcela de crédito, como também no tempo médio até a inadimplência, em clientes inadimplentes de um banco brasileiro (neste caso,  $k=3$ ). Os parâmetros foram estimados por método Bayesiano, neste caso as distribuições a posteriori dos parâmetros foram obtidas por método de simulação estocástica de MCMC utilizando o algoritmo de Metropolis-Hastings (Gamerman & Lopes, 2006). O diagnóstico bayesiano para ajuste do modelo e verificação de observações influentes foi analisado por resíduos quantílicos aleatorizados (Jiang et al., 2013) e medidas de divergência (Peng & Dey, 1995).*

### **Solange Ferreira Silvino, Yana Miranda Borges, Breno Gabriel da Silva, Naiara Caroline Aparecido dos Santos, Brian Alvarez Ribeiro de Melo. (Instituto Federal de Roraima - IFRR)**

Estudo de modelos de previsão para demanda de processos distribuídos em uma unidade judiciária do Estado de Roraima

*Abstract: A previsão de demanda em unidades judiciárias é um importante fator de impacto no desempenho de tribunais, onde flutuações de demanda não previstas causam altos impactos nas unidades judiciárias, gerando sobrecarga de trabalho aos servidores e, conseqüentemente, demora em tramitação de processos. Neste contexto, este artigo tem como objetivo analisar o comportamento da série de processos distribuídos no período de 2000 a 2015 na unidade judiciária 2ª Vara Cível do Tribunal de Justiça do Estado de Roraima, sob o olhar da previsão de demanda baseada em métodos de séries temporais comparando modelos de previsão. Neste trabalho, aplicar-se-á vários testes estatísticos que direcionam para os melhores modelos de previsão a serem utilizados.*

**Thiago Gottardi, Rosana Teresinha Vaccare Braga (ICMC-USP)**

Providing past and future awareness to adaptive systems using stochastic processes

*Abstract: In the context of self-adaptive systems, software behavior may vary at real-time. These systems depend on self-awareness in order to self-improve and self-adapt. According to several road-maps for these systems, many authors have pointed that this unpredictability requires specific measures to monitor the system status. In this work, we discuss how these systems can be modelled as stochastic processes. We present an analytic proven metric based on Markov chain to assess behavior by comparing their execution to past successes and failures. It is expected that the metric and associated models can be used to improve the management of those systems, as well as providing a basis for risk analysis of their execution. After presenting a case study, we also discuss how our proposal is feasible to improve the monitoring of these systems. In conclusion, this study indicates how stochastic methods could eventually become essential for monitoring self-adaptive systems by enabling past and future awareness. We also discuss how this monitoring can be effective for dev-ops development techniques, which are composed by distributed teams generating frequent releases. These teams require intensive usage of monitoring tools to cope with unpredictable risks that may arise throughout development.*

**Vanessa Helena Pereira, Theodore Gyle Lewis, Leonardo Tomazeli Duarte (UNICAMP)**

Criteria importance through intercriteria correlation method: a mathematical implementation for NBA players efficiency analysis

*Abstract: Usually, solving problems involves determining and choosing criteria,*

*the consequences of which are critical in a decision-making process. The determination of subjective preferences is always difficult, especially when it comes to involving multiple criteria decisions. An interesting mathematical method, the CRITIC (Critical Importance Through Intercriteria Correlation) involves important tools of Descriptive Statistics and Principal Component Analysis (PCA). The objective of this work is to implement the CRITIC method in open computing environment for scientific applications, the SciLab, in a classification study of players of NBA (National Basketball League). This type of method allows to classify mathematically and by multiple criterias the players according to their performances. The data are selected from the General List of Official Leading Players in the Regular Season 2017-2018. Our approach points to new types of statistical methods of analyzing athlete in addition to standard methods for efficiency analysis.*

**Walkiria M. Oliveira Macerau, Luis A. Milan (UFSCar/USP)**

Inferência bayesiana de modelos de misturas de distribuições assimétricas aplicados à dados de criptomoedas.

*Abstract: Resumo: Criptomoedas são moedas virtuais descentralizadas que tem recebido muita atenção recentemente, tendo sido apontada por alguns analistas como sendo o futuro das transações financeiras. Uma característica que esse mercado tem apresentado é alta volatilidade. Bitcoin é a pioneira dentre as criptomoedas, mas desde seu lançamento em 2009 um grande número de criptomoedas tem sido lançadas e esse número já ultrapassa 2100 [1]. O volume de recursos aplicados em criptomoedas também aumenta a cada dia o faz desse fenômeno algo a ser estudado e melhor entendido. Nesse trabalho testamos algumas alternativas de modelos que expliquem o comportamento de algumas características, em particular o retorno, de algumas criptomoedas. Utilizaremos as distribuições normal e t de Student assimétricas e também misturas dessas distribuições. Utilizaremos a abordagem bayesiana para ajustar os modelos combinada com os métodos MCMC (Monte Carlo Markov Chain). Métodos de seleção de modelos serão empregados para identificar quais modelos explicam melhor o comportamento dessas criptomoedas, com ênfase para os critérios EAIC, EBIC e DIC. A implementação dos métodos será feita utilizando o software R.*

*Palavras-chave: Modelos de mistura, distribuições assimétricas, criptomoedas, distribuição normal, distribuição t de Student, abordagem bayesiana.*

*Referência: [1] Site: <https://coinmarketcap.com/all/views/all/#BRL>*

**Yana Miranda Borges, Breno Gabriel da Silva, Naiara Caroline Aparecido dos Santos, Brian Alvarez Ribeiro de Melo (Universidade Estadual de Maringá)**

Uma análise de modelos de regressão Poisson para dados de doenças isquêmicas do coração no Brasil em 2016

*Abstract: Ao estudarmos situações cuja variável dependente descreve dados de contagem, uma abordagem possível é estimar um Modelo Linear Generalizado com distribuição de Poisson. No entanto, devido às especificidades da distribuição Poisson, em que média e variância são iguais, há casos em que a ausência de equidispersão pode subestimar ou superestimar as estimativas, havendo necessidade de ajustar o modelo ou de procurar modelos alternativos mais adequados aos dados. O presente estudo avaliou ajustes pelo modelo Poisson, quasi-Poisson e binomial negativa para analisar a taxa de doenças isquêmicas do coração, considerando sexo, faixa etária e região. Dos modelos verificados neste estudo, o modelo quasi-Poisson apresentou um melhor ajuste quando comparado com o modelo log-linear Poisson e binomial negativa. Como resultado, verificou-se que entre os homens a taxa de ocorrência de morte por doenças isquêmicas do coração é 1.83 maior que entre as mulheres. Quanto à faixa etária, a taxa apresenta-se crescente, tomando-se como referência o grupo de menores que vinte anos. As regiões não apresentam grandes diferenças, no entanto, a região Norte é a que apresenta a menor taxa de mortalidade, quando comparada à região Centro-Oeste.*

**Yury Rojas Benites, prof. Vicente Garibay Cancho (UFSCar/USP)**

Um novo modelo de regressão para taxas e proporções

*Abstract: Neste trabalho é apresentado um novo modelo estatístico para modelar dados no intervalo contínuo  $(0,1)$ . O modelo proposto é baseado na distribuição SB de Johnson, onde consideramos como transformação o quantil da distribuição valor extremo generalizado ao invés do quantil da distribuição logística. A nova família é estendida para modelos de regressão, onde fazendo uma reparametrização do modelo na mediana da variável resposta modelamos a mediana e o parâmetro de dispersão conjuntamente. A estimação dos parâmetros é baseada no método de máxima verossimilhança. Estudos de simulação são feitos para avaliar o desempenho das estimativas de máxima verossimilhança. Uma aplicação com dados reais de câncer colorretal é apresentado para ver a utilidade do modelo.*