-Selecione o idioma-   ▼

Associação Brasileira de Estatística
XIV EBEB - Encontro Brasileiro de Estatística Bayesiana - Rio de Janeiro

ABE (http://redeabe.org.br)

Scige

Apresentações em Pôster

Instruções para apresentadores:

- *O pôster devem ser afixado entre 15h20 e 16h40 do respectivo dia da apresentação e retirado ao término da sessão pôster*

- *As dimensões máximas de cada pôster devem ser de 90 cm de largura e 105 cm de altura*

- *O pôster deve estar preferencialmente em inglês*

Pôster 2 (Quarta-feira)

39. **Título:** Exact Bayesian inference for Markov switching Cox processes
Autores: Lívia M. Dutra; Flávio B. Gonçalves; Roger W. C. Silva
Abstract: Statistical modelling of point patterns is an important and common problem in several applications. An important point process, and a generalisation of the Poisson process, is the Cox process, where the intensity function is itself stochastic. We focus on Cox processes in which the intensity function is driven by parametric functional forms that switch among themselves according to a continuous-time Markov chain. We call this as a Markov switching Cox processes (MSCP). We develop a Bayesian methodology to perform exact inference based on MCMC algorithms. Simulated studies are presented in order to investigate the efficiency of the methodology on the estimation of MSCP's intensity function and the parameters indexing its law.
Keywords: *Bayesian inference; Exact posterior distributions; Cox process; Continuous-time Markov chain;*

40. **Título:** Dynamic Forecasting of Wind Speed and Maximization of the Power Function
Autores: Lassance, R.F.L.; Fonseca, T.C.O.; Schmidt, A.M.
Abstract: The objective of this work is to present different classes of models that provide short range forecasts of wind speed, while taking the function of the generated power into account. The statistical analysis for this problem is of primary importance due to the inconstant nature of the wind and the impossibility of storing its energy to tackle

future demands. Since wind power is gradually becoming cheaper and more viable, there is a growing necessity of precise forecasts. Models based on Harvey and Fernandes (1989) are presented, as well as their process of sequential updating. Given that censoring is common on this type of data, an adaptation for the case where a great number of 0's are observed is provided. The power function is explored and a loss function, from Hering and Genton (2010), is presented. Point estimates and model selection methods based on such function are discussed. Lastly, we show an application with real data and how the model adapts to it when an attempt is made to forecast more distant times.

Keywords: *Time Series; Dynamic Models; Power Function; Forecasting;*

41. **Título:** Extending JAGS for spatial data
Autores: Magno Tairone de Freitas Severino; Vinícius Diniz Mayrink; Fábio Nogueira Demarqui

Abstract: Bayesian hierarchical modeling for spatial data is challenging for professionals from other areas than statistics. Setting the model prior distributions and the likelihood are the simplest part of the process. What makes it difficult is the computation of the posterior full conditionals. The BUGS (Bayesian inference Using Gibbs Sampling) family of statistical softwares is very attractive to implement an MCMC for an hierarchical model, since it requires the specification of only the prior distributions and the likelihood. The WinBUGS, released in mid-90s, and OpenBUGS, released in 2005, are a part of this family that can handle spatial data through GeoBUGS. Another interesting alternative similar to the BUGS family is called JAGS (Just Another Gibbs Sampler), released in December 2007. It was built to be extensible, allowing users to write their own functions, distributions and samplers. JAGS currently lacks a module for performing spatial data analyses, but this work intends to fill this gap. In this work we present the GeoJAGS module. It contains implementation of several covariance functions for the point-referenced data case and an implementation for using CAR distribution in the areal data case.

Keywords: *spatial statistics; point-referenced data; areal data; Gibbs sampling;*

42. **Título:** Is there a wage discrimination in IT careers in Santa Catarina?
Autores: Aishameriane Venes Schmidt;Fernando Pozzobon

Abstract: The participation of women in the workforce occurred late in several countries, including Brazil. Even nowadays, there are barriers impeding female participation in the economic active force, such as wage discrimination. Studies point out that women wages are, in average, 30% lower than men. Nevertheless, there are evidences that this differences are smaller in professions where there are a higher demand than surplus for workers, such as STEM (science, technology, engineering and mathematics) careers. In the present work, we investigated the gender effect in wages, controlling age, schooling and time in the job, for 10.919 workers listed as IT analysts in the Brazilian Labor Office for the Santa Catarina State. We combined a methodology based on the economics of discrimination and Oaxaca decomposition and estimated a linear regression model using a conjugated Normal-Gamma prior. As hyperparameters, we used the mean values from Bonini and Pozzobon (2016), who made a broader study comparing IT and industry workers from all three States from Southern Brazil using data for the year of 2011. Our posterior estimates show that the salary for females is approximately 13.34% lower than their males counterparts, with a credibility interval ranging from 11.03% to 15.63%, suggesting that in fact there is a gender wage gap discrepancy between women and men in this sample.

Keywords: *Female wage discrimination; Workforce; IT careers;*

43. **Título:** Joint modeling of longitudinal measurements and event time data: a dynamic generalized hierarchical approach
Autores: Pamela Chiroque Solano Helio S. Migon

Abstract: Our aim is to simultaneously model time-to-event and longitudinal data describing quality of life. A joint dynamic hierarchical multi-state model, based on the Bayesian paradigm, is proposed. The inclusion of time-dependent covariates is appealing when the focus is to compare different medical treatments. Specifically, the longitudinal components are modeled through the introduction of a latent structure associated with the mean trajectory component. The non-progressive health multi-state model is allowed. The mean trajectory function is incorporated as a predictor of the non-proportional hazard function. We assume the data are right censored and we

also include a fragility term to take care of the unobserved heterogeneity. The mean trajectory depends on unobserved Markov switching state variables. An analysis with simulated data is presented to evaluate the predictive power of the model. Finally, we show an application to a real dataset.

Keywords: *Joint hierarchical model; Longitudinal and time-to-event data; Dynamic model;*

44. **Título:** Linear Skew Normal Antedependence Models for Longitudinal Data
Autores: Marta Lucia Corrales Bossio; Edilberto Cepeda Cuervo
Abstract: In recent years, the joint modeling of the mean and the covariance matrix in continuous longitudinal data with multivariate normal errors, by means of the factorization of the precision matrix through antedependence models, has been widely used by authors using Bayesian methods. These models have as advantages the possible use of computers to estimate the parameters and absence of restrictions on them. However, the assumption of multivariate error normality can be questionable in many practical situations: when there are atypical data, when the data exhibit heavy tails or when high asymmetric behavior is evident in the data. When the data have asymmetric distribution, the normal skewed distribution has shown efficiency in treating the skewness. In this work, we propose skew-normal antedependence regression models, where mean, scale, and antedependece parameters follow regression structures.
Keywords: *Antedependence models; longitudinal data; Bayesian method;*

45. **Título:** Maximum entropy distribution on a circular region under mean value constraints
Autores: J.C.S. de Miranda
Abstract: Maximum entropy distributions are a valuable tool in simulation studies where, in some sense, besides the information we already have about a probability structure, we want to assume the least additional information about it. Using variational methods we determine the maximum entropy probability distribution with support on a circular region under mean value constraints. More precisely, we determine the probability density function, $f_{XY},$ of a random vector (X,Y) such that $\mathcal{I}m(X,Y)\subset \mathcal{D},$ where $\mathcal{D}=\{(x,y)\in \mathbb{R}^2: x^2 +y^2\le1\},$ that maximizes the entropy functional $f\rightsquigarrow-\int_\mathcal{D}f\ln f\mathrm{d}\ell$ and satisfies the mean value constraints $\mathbb{E}X=\mu_X$ and $\mathbb{E}Y=\mu_Y,$ where μ_X and $\mu_Y,$ such that $(\mu_X,\mu_Y)\in\mathcal{D},$ are given.
Keywords: *Maximum Entropy Distribution; Variational Calculus; Bessel Functions;*

46. **Título:** Mismeasurement Cure fraction model
Autores: Anna Rafaella da Silva Marinho; Rosangela Helena Loschi
Abstract: The medical advances in cancer treatment and the development of efficient diagnosis techniques in the recent years have contributed to the increase in the fraction of cured patients. Because of this, it is increasing the interest in developing statistical models able to more appropriately deal with lifetime data in the presence of cure fraction. It is well known that some covariates that may influence the patient lifetime can be mismeasured. In this work, a Bayesian cure rate model with mismeasured covariates is developed extending previous models. We consider a structural approach to deal with the explanatory variables with measurement error. One of the main goals is reducing the bias in the estimates of the cure rate. Differently of what has been considered in the literature, the error variance is estimated. Three different prior specifications are proposed to model this parameter behavior. A solution to identifiability problems that arose due to the presence of latent variables in the model is proposed. In all models, the posterior distributions have no closed form expressions. For this reason we used the Gibbs Sampler with Adaptative Metropolis to obtain samples of the posterior distributions. A Monte Carlo simulation study is presented and also an analysis of a melanoma clinical trial that has already been discussed in the literature.
Keywords: *Cure Rate; Structural Model; Mismeasured Covariates; Bayesian Inference;*

47. **Título:** Model selection for log-Gaussian Cox processes using FBST
Autores: Patrícia Viana da Silva; Jony Arrais Pinto Júnior
Abstract: Log-Gaussian Cox processes is a class of models very useful to fit point patterns data. Point patterns are very common in many research areas and their principal goal is to know if there is a spatial pattern governing the

occurrence of the event of interest. In this context, investigate relationships between the point pattern and covariates that are possibly associated with the event of interest is a very interesting issue. Pereira & Stern proposed the Full Bayesian Significance Test (FBST) as a coherent Bayesian significance test for the sharp hypothesis. It is an alternative to frequentists p-value significance tests. Others works mark FBST as an authentic bayesian approach with intuitive interpretation and ease implementation. This paper proposes to investigate the FBST as a model selection criterion for log-Gaussian Cox process and compare their performance with other usual methods as Akaike Information Criterion (AIC) using simulated data implemented in BUGS.
Keywords: *log-Gaussian Cox processes; Full Bayesian Significance Test; model selection;*

48. **Título:** Modelling zero inflated biomass from fisheries in the Lower Amazon River: a Bayesian Approach
Autores: Julio Cesar Pereira; Giovani Loiola da Silva; Victoria J. Isaac
Abstract: In commercial fisheries, catch and fishing effort data are usually the most common data available for stock assessment. This study has been motivated by the difficulty facing researchers in analysing data for catch per unit of effort, from fisheries in the Lower Amazon River due to zero inflation phenomenon. We aimed to develop a statistical model that was able to accommodate the zero inflation on catches allowing a better understanding of variations of catch in weight related to variations in effort and other covariates available. In order to analyse this type of data, we proposed a Bayesian three-stage hierarchical model. At the first stage, we described the number of fishing trips per location (N) according to a Poisson distribution, whereas at the second stage, given N>0, we defined a Bernoulli variable X with probability q of success, where X assumes 1 if catches occurred for a fishing species, and 0 if naught is caught for that species. Eventually at the third stage, we modelled the fishing weight, denoted by Y, which assumes zero when N=0 or X=0 and N>0. So, when X=1 and N>0, we described Y according to a gamma distribution, where the mean was proportional to the number of trips N. This approach provided a useful tool to deal with the variation on catch per unit of effort as function of namely covariates, when the data were inflated by zeros coming from both sources: abstinence of fishing activity and absence of catch in the presence of fishing activity.
Keywords: *Double zero-inflated data; Fisheries; Compound Poisson; Bayesian hierarchical modelling;*

49. **Título:** Modelo Periódico Autorregressivo Aplicado a Modelagem de Vazão
Autores: Marcel de Souza Borges Quintana; Victor Eduardo Leite de Almeida Duca
Abstract: Os modelos autorregressivos são comumente encontrados dentro do contexto de séries hidrológicas, especificamente em séries de vazões e/ou ENA (Energia Natural Afluente). Muitos destes modelos são de ordem 1, possuem parâmetros constantes ou periódicos e necessitam do requisito de normalidade. De acordo com a literatura, séries de vazões anuais podem ser aproximadas para distribuições normais, porém em períodos de tempo curtos como diário, semanal e mensal esta característica não é observada, especialmente pelo problema de assimetria. Devido a isto, uma nova classe de modelo de primeira ordem foi estudada na tentativa de suprir tal problema. O novo modelo mantém estrutura autorregressiva periódica, pode ser aditivo, multiplicativo ou híbrido, porém assumindo distribuição gama. Além disso, os Métodos de Momentos são utilizados para estimação de seus parâmetros. O objetivo deste estudo é propor a estimação dos parâmetros sob abordagem Bayesiana (Monte Carlo via cadeias de Markov-MCMC) do modelo periódico aditivo gama autorregressivo de ordem 1 (PAGAR(1)) em séries do Setor Elétrico Brasileiro. Além disso, o modelo periódico lognormal de ordem 1 também foi proposto no estudo como forma de comparação. Os resultados encontrados evidenciaram problemas durante o processo de validação do modelo PAGAR(1) devido a sua complexidade, mas revelaram resultados satisfatórios ao modelo lognormal de ordem 1. Por fim, foram simuladas séries sintéticas utilizando a distribuição dos parâmetros obtidos em cada cadeia.
Keywords: *Séries Temporais; Modelo Periódico Autorregressivo; Inferência Bayesiana; Monte Carlo via Cadeias de Markov;*

50. **Título:** Modelo de Volatilidade Estocástica via Processo Gaussiano
Autores: Lucas Marques Oliveira; Ralph dos Santos Silva; Fabio Antonio Tavares Ramos.
Abstract: Nesse trabalho será tratado o problema de estimação da volatilidade de um ativo financeiro. A diferença para os modelos convencionais como o GARCH ou outros modelos de volatilidade estocástica é que não foi

definida uma estrutura rígida para a forma funcional de volatilidade. Através disso foi colocado um processo gaussiano como priori no espaço das funções para sua modelagem. Esse modelo é chamado de "Modelo de Volatilidade Estocástica via Processo Gaussiano". Nele além da flexibilização descrita acima, a volatilidade é tratada como uma variável estocástica e não determinística. A partir desse modelo serão utilizadas técnicas de inferência bayesiana, filtro de partícula e aproximações via janela de dados para se conseguir uma estimação precisa de forma eficiente. Apresentamos um estudo de simulação comparando com outros modelos e uma aplicação a dados reais.
Keywords: *Volatilidade Estocástica; Processo Gaussiano; Filtro de Partículas; Kernel; Inferência Bayesiana;*

51. **Título:** Métodos estatísticos de proteção de dados confidenciais sob a condição de differential privacy.
Autores: Augusto Marcolin; Thaís Paiva
Abstract: A quantidade de dados produzidos no mundo digital tem crescido exponencialmente nas últimas décadas. Atentas a este fato, empresas e organizações não medem esforços para analisar toda essa gama de informação. Contudo, há um crescimento na preocupação acerca da privacidade da informação das pessoas. Nesse contexto, surge a àrea de data privacy, cujo objetivo é garantir anonimização das informações em bases de dados. Tendo em vista o problema exposto, este trabalho apresenta métodos para anonimização de variáveis categóricas, através da geração de bases sintéticas sob garantia de differential privacy. Esta condição garante aos indivíduos que, mesmo que um usuário malicioso conheça todos os dados à respeito dos outros sujeitos ele não desvendará seus dados. Devido ao alto nível de distorção imposta em bases de dados sob esta garantia, a inferência tradicional para bases sintéticas possui grande viés. Com isso, apresentamos métodos para fazer inferência para tal tipo de dado, baseado em modelos bayesianos hierárquicos.
Keywords: *Dados Confidencias; Differential Privicy; Dados Sintéticos;*

52. **Título:** Neighborhood definition in the Nearest Neighbor Gaussian Processes
Autores: Mariana Lizarazo O, Dani Gamerman , Thaís C O da Fonseca
Abstract: Spatial process models for analyzing geostatistical data entail computations that become prohibitive as the number of spatial locations become large. The Nearest Neighbor Gaussian Processes (NNGP) provide a scalable alternative by using local information from few nearest neighbors. Scalability is achieved by using the neighbor sets in a conditional specification of the model. The way to select this neighbor depends on: an ordering of the locations, a reference set and the definition of the neighborhood. The original NNGP approach consider a specific form to do this selection. This paper pretends explore alternative ways to select the neighbor set.
Keywords: *Nearest Neighbor Gaussian Processes; large datasets; ordering criteria; neighbor definition;*

53. **Título:** Network reconstruction using classical and Bayesian methods, a comparison
Autores: Henrique Bolfarine
Abstract: Several methods have been proposed for the reconstruction of large-scale networks, which in this case are treated as a Gaussian Graphical Models. In this work we will analyze three different methods; the well-known Graphical Lasso (GLasso), it's Bayesian counter part, the Bayesian Graphical Lasso and a new method called LPC, or Local Partial Correlation. The evaluation will be carried out with high dimensional data, generated from different types of sparse random graphs (Erdos-Renyi, Barabasi-Albert, and Watts-Strogatz). The method used for the evaluation, is the Characteristic Operation Receiver, or ROC curve. We also applied the methods presented in the reconstruction of the gene co-expression network from cervical cancer tumors [4], analyzing later, how much, in terms of quantity (numbers of vertices and nodes) the recovered structures have in common.
Keywords: *Network reconstruction methods; Gaussian Graphical Models; Bayesian Graphical Lasso; Graphical Lasso; LPC;*

54. **Título:** Non-separable space-time SPDE models
Autores: Elias T. Krainski; Finn Lindgren; Daniel Simpson; Håvard Rue
Abstract: The field of space-time statistical models is an ongoing research. There are several valid models in the literature. However, there are some require characteristics that are not fulfilled. This includes smoothness properties

and feasible computational cost. There are approaches in the literature to overcome the computational cost. These includes lowering the model resolution, considering local approximations. In this work we consider the stochastic partial differential equations (SPDE) approach. This is in line of considering the phisical properties of the process. The SPDE's with half-integer linear operators have solutions wich are Markov. This propertie translates into computationally efficient computations. Also, SPDE's solutions have good smoothness properties. The solution of the kind of SPDE's we are working admits a Gaussian Markov representation. Therefore we can compute with sparse precision matrices and there is no need to compute covariance. However, the marginal properties of a model can help understanding the model parameters. This is useful for prior assignement and interpretation of the results. We have computed the marginal variance and correlation in order to provide a map from the marginal properties to the model parameters. We applied the model for global temperature data considering daily temperature data from 12862 stations.

Keywords: *Non-separable; Stochastic Partial Differential Equations; Space-time;*

55. **Título:** On a family of autoregressive conditional duration models based on the log-symmetric distributions
Autores: Helton Saulo; Rafael Paixão; Jeremias Leão; Ming-Hui Chen
Abstract: This paper adapts the Hamiltonian Monte Carlo method for application in log-symmetric autoregressive conditional duration models. These recent models are based on a class of log-symmetric distributions. In such class, it is possible to model both median and skewness of the data, which is specially useful in the case of high frequency financial data. In this context, we use the Bayesian approach to estimate the parameters of some log-symmetric autoregressive conditional duration models, as well as evaluate its performance using a Monte Carlo simulation study. Furthermore, the usefulness of the estimation methodology is demonstrated by analyzing a real-world high frequency financial data set, related to April of 2016, from the german company BASF-SE.
Keywords: *ACD models; Bayesian inference; Log-symmetric distributions; High frequency financial data;*

56. **Título:** Particle Filters and Adaptive Metropolis-Hastings Sampling Applied to Volatility Models
Autores: Iago Cunha; Ralph Silva
Abstract: Markov Chain Monte Carlo (MCMC) simulation methods are widely applied in statistical inference to sample from a probability density. However, it is common to come across problems when calculating the likelihood, commonly for non-linear non-Gaussian models, or in situations where the choice of powerful proposals is not particularly easy. When the likelihood does not have a closed form we can approximate it using the particle filter algorithm. Two particle filtering methods are the standard particle filter (SIR) developed by Gordon et al. (1993) and the auxiliary particle filter (ASIR) proposed by Pitt \& Shephard (1999). Moreover, Andrieu et al. (2010) proved that MCMC methods still converge to the correct posterior density function even if the simulated likelihood via SIR or ASIR is used. To work around problems when choosing effective proposal distributions to use on MCMC, we can apply some adaptive sampling techniques. In such methods, the parameters of the proposal distribution are tuned by using previous draws and the difference between these successive parameters of the proposal converges to zero (diminishing adaptation). Important theoretical and practical contributions to diminishing adaptation sampling were made by Haario et al. (1999), Haario et al. (2001), Roberts \& Rosenthal (2007), Roberts \& Rosenthal (2009) and Giordani \& Kohn (2010). We will work with particle filters and adaptive sampling techniques to estimate volatility models such as the generalized autoregressive conditionally heteroscedastic model with noise and several stochastic volatility models. The results will be compared using marginal likelihoods and some likelihood based information criteria.
Keywords: *MCMC; Particle Filters; Adaptive Sampling;*

57. **Título:** Portfolios: Modeling and Optimization
Autores: Pedro Helal Chafir; Helio dos Santos Migon; Ralph dos Santos Silva
Abstract: Dynamic Linear Models are widely used to obtain predictive distributions of financial series. In this work we aimed to model the series of the price log-returns of a big number of stocks using regression dynamic linear models with certain variables that helps to explain the stock price. Each serie was modeled separetly and with bayesian inference techniques predictives distributions of each serie was obtained (which gives us the expected value and tha variance one step ahead of all series). The covariance matrix of the multiveriate distribution can also

be obtained. The portfolio value is the scalar product of a vector of weights with the vector of the stocks prices. Markowitz (1952) has introduced techniques to optimize the capital division among the stocks based on the expected value and the covariance matrix of the multivariate predictive distribution of the stocks log-return. One of those techniques consists in fix a lower bound to the expected value of the porfolio and find the vector of weights that minimize the covariance matrix. With the values of the moments of the predictive distribution obtained from the model, the optimization technique of Markowitz is applied to determine the portfolio weights. It can be updated daily or every k days, for some specified time interval, and then results can be compared to other portfolios, market indexes, bonds, etc.

Keywords: *portfolio; optimization; dynamic;*

58. **Título:** Prediction of credit card defaults through logistic regression models
Autores: Catarina Dall Agnol Zidde

Abstract: This project aims to model the default of credit card payment clients through a Bayesian logistic regression model, performing inference via Hamiltonian Monte Carlo (HMC) methods. The chosen data set contains information on demographic factors, billing statements and previous payments of 30000 clients from Taiwan that were collected through 8 months in 2008. When modeling this kind of data, some questions arise. First, there are too many variables, so the HMC posterior sampling becomes computationally intensive, and convergence is hard to achieve. Second, since the data set is imbalanced, having a naturally low default probability, scoring models according to their predictive ability is not trivial (e.g. a model that assumes no one will default will have high accuracy). This project, then, discusses the fit and comparison of models that use different subsets of variables as covariates, along with principal components, aiming to fit a model that can be estimated more rapidly, losing the least possible amount of predictive power.

Keywords: *logistic regression; credit card; principal component analysis;*

59. **Título:** Prior specification in change point problems
Autores: Fernando Corrêa Victor Fossaluza

Abstract: Change point models looks for independent partitions of a random sequence. Barry and Hartigan proposed a well behaved and computationaly cheap bayesian model for this task. The good behavior of the model required an appropriate selection of the priors. In this paper we investigate why the care is necessary. Why some priors are better than others? Which priors should not be chosen in change point models? This question is partially answered on the binomial change point model. Our main result states that too much prior weight on "there's a change point" leads to an inference that always detect a change point, even if there's none. We backup our findings on other cenarios through empirical evalutations. The results suggests that prior choice in change point models should always be a point of attention, even in low dimensions.

Keywords: *change point; structural change; prior choice; product partition model;*

60. **Título:** Quantifying the long-term impact of the Tucurui dam on riverine hydrology under non-stationary conditions and presence of substantial data gaps using a sparse infinite factor model.
Autores: Denis Valle; David Kaplan

Abstract: Environmental impact assessments require a comparison between observed post-impact outcomes with what "would have happened" in the absence of the impact (i.e., the counterfactual). Standard impact assessment approaches rely on very strong assumptions. Furthermore, a key challenge in assessing impact is the presence of substantial data gaps, particularly in long time-series. We propose a novel statistical method that enables us to predict the counterfactual in the presence of these data gaps. A key feature of this method is that it automatically determines the number of factors that is best supported by the data, leading to improved predictions due the sparse representation of the covariance matrix. In this study, we apply this model to the Tocantins River (Brazil), one of the most dam-altered rivers in the Amazon. We find that our statistical model had good out-of-sample predictive skill and captured uncertainty well for the negative control period between 1979 and 1984. Despite the high predictive skill of our model for this period, there were substantial discrepancies between predictions and observations from 1984 onwards, which can be attributed to the dam impact. Most of these discrepancies were associated with a higher than expected water level during the dry season but there were also substantial changes in seasonality for

the closest upstream gauge. Dam impacts were evident even in gauges 176 km away from the dam. Appropriately assessing the impact of large-scale infrastructure projects such as dams is particularly important given the current spree of dam construction in the Amazon region.

Keywords: *Sparse infinite factor model; Counterfactual; Impact assessment; Hydroelectric dam; Missing data;*

61. **Título:** SPATIO-TEMPORAL DISTRIBUTION OF VEHICLE THEFT IN THE MUNICIPALITY OF RIO DE JANEIRO FROM JANUARY 2012 TO JULY 2016

Autores: Adriel Costa Maia Larissa de Carvalho Alves Gustavo da Silva Ferreira

Abstract: This study proposes an analysis of the occurrences of vehicle theft in the municipality of Rio de Janeiro, with emphasis on its spatio-temporal evolution. In this work, it was proposed a dynamic generalized linear model utilizing, in the prior distribution of the parameters, the concept of the intrinsic conditional autoregressive gaussian models. The proposal allowed the temporal and spatial effects of occurrences to be analyzed separately. In this way, the evolution of the phenomenon was described in the global level, in the municipality as a whole, and at the local level, detailing the behavior of each neighborhood individually. The analyzed data was obtained from the occurrence bulletins registered in the municipality between January 2012 and July 2016.

Keywords: *spatio-temporal distribution; dynamic generalized linear model; conditional autoregressivel gaussian models (CAR); relative risk; public security;*

62. **Título:** Semiparametric Bayesian modeling via mixtures

Autores: Nívea B. da Silva; Marcos O. Prates; Flávio B. Gonçalves

Abstract: Recently, because its flexibility, finite mixtures of distributions have been used to model the errors distribution in univariate and multivariate linear regression models. This work introduces a novel methodology based on finite mixtures of Student-t distributions to model the errors distribution in linear regression models. The proposed approach contemplates a hierarchical structure for the mixture distribution in which the first level models the number of modes, responsible to accommodate multimodality and skewness features, and the second level models tail behavior. This hierarchical structure allow for modeling the tail structure of the model without estimating degree of freedom which is known to be a difficult problem in the literature. Inference is performed via Markov chain Monte Carlo and simulation studies are conducted to evaluate the performance of the proposed methodology. Results from the analysis of two real data are also presented.

Keywords: *Finite mixtures; hierarchical modelling ; heavy tail distributions ; MCMC;*

63. **Título:** Sensibility Analysis of the Propensity Score in Bayesian Nonparametric Models for Causal Inference

Autores: Pedro Henrique Filipini dos Santos; Hedibert Freitas Lopes

Abstract: An efficient method to estimate treatment effects from observational data is through Bayesian nonparametric modeling, specifically Bayesian regression tree models, like seen in Hill (Journal of Computational and Graphical Statistics, 20 (1), 217-240, 2011). Two of these models are under the scope of this study: the Bayesian Additive Regression Trees (BART), which is, according to many researchers, the current gold standard to causal effect estimation, and the Bayesian Causal Forests (BCF), which is an extension of the former model proposed by Hahn, Murray and Carvalho (Submitted for publication, 2017), and is focused on estimating the treatment effects separately from the prognostic effect. In the presence of confounding, the regularization of these models may cause biased estimations of the treatment effects. A way to ease this problem in those models is through the inclusion of the propensity score as a covariate, as this can show a reduction of bias in estimates of the Average Treatment Effects (ATE) under strong ignorability. Since the true propensity score is unknown, this study performed a sensibility analysis by estimating the propensity score by different methods and measured how its misspecification affected the ATE estimation in each model. To ensure ignorability and aiming to preserve the applicability of the analysis, only the outcomes and the true propensity scores were simulated, keeping the other covariates as real data.

Keywords: *Causal inference; Bayesian regression trees; Propensity score;*

64. **Título:** Skew-Normal Regression Models. Bayesian Joint Modeling of Location, Scale and Shape Parameters

Autores: Edilberto Cepeda-Cuervo; Martha Corrales

Abstract: In normal linear regression models, the assumption of normality of the errors may be questionable in cases where the data present outliers, heavy tails or asymmetric behavior. A possible way to overcome these weaknesses without transformation of the data is to assume heavy-tailed distributions, skewed distributions or other distributions. This work proposes Bayesian skew normal regression models where location, scale and shape parameters follow (linear or nonlinear) regression structures, and where the variable of interest follows an Azzalini or a Sahu skew normal distribution. A Bayesian method is developed to fit the proposed models, using working variables to build the kernel transition functions. To illustrate the performance of the proposed Bayesian method and application of the model in statistical analysis, we present results of simulated studies and of application to the forced displacement data in Colombia.
Keywords: *Bayesian analysis; Skew-normal distribution; Regression models;*

65. **Título:** Sparse Bayesian model of binary response with asymmetric link function for text categorization
Autores: Hugo M. Agurto; Márcia D. Branco

Abstract: A typical problem when dealing datasets with a large amount of covariates compared to small sample sizes is to satisfactorily estimate the parameters associated with each covariate. When the number of covariates greatly exceeds the sample size, the parameter estimation becomes very difficult. In various areas of application such as text categorization, it is necessary the task of selecting important covariates and avoiding the overfitting of the model. In this work, we developed a Sparse Bayesian binary regression model with asymmetric link function for text categorization. In addition, we assign a sparse prior distribution (double exponential) for regression parameters to favor sparsity and to reduce the number of covariates in the model. The performance of the proposed model is demonstrated with real data set, the Reuters R8 corpus. The dataset contains the eight most frequent classes from the Reuters-21578 collection of newswire articles. The eight classes consist of a minimum of 51 up to 3923 documents and sum up to a total of 7674 texts. Parameter estimation is performed considering Hamiltonian Monte Carlo estimation method on No-U-Turn Sampler (NUTS) extension, using the Stan software in the R package.
Keywords: *Bayesian lasso; Skew link; Sparsity; Text categorization;*

66. **Título:** Spatial pattern analysis of prison locations in Brazil
Autores: Rebecca de Oliveira Souza; Marina Silva Paez; Vinícius Pinheiro Israel

Abstract: The imprisonment has been studied in the international scene and it has worrying outcomes with a rapid growth in the recent decades. According to data from the Institute for Criminal Policy Research in Birkbeck University in London, the amount of the Brazilian prison populations increased twenty-fold over 1973 through the present days. Therefore, there is the interest in investigating the distribution of prison units in Brazil and studying the association of covariates in order to understand the disposition of these units. This work aims to model the location of this units by Cox Process. Firstly, the intensity function was defined by the combination of normal distributions which goal is to describe the formation of points clusters. The covariates in this model refer to the prison units, and they are included in the dimensions of the points, that leads to the formation of clusters made of prison units geographically close and with possible similarities. The second part of this work uses a model proposed by Diggle et al.(1997) that allows distances between prision unit locations and influence sources previously defined, taken for granted in intensity function, besides spatial covariates. Such sources may be defined as the centers of the clusters estimated in the first part of this work. In both parts, the inference is done under the Bayesian approach.
Keywords: *Spatial analysis; Bayesian inference; Cox Process; Imprisonment in Brazil;*

67. **Título:** Spatio-Temporal analysis of cases of death from cerebrovascular diseases in the municipalities of Rio de Janeiro from 2010 to 2015
Autores: Isabele Martins Araujo (ENCE/IBGE); Ludmila Freitas Simões Souza (ENCE/IBGE); Gustavo da Silva Ferreira (ENCE/IBGE)

Abstract: In this study we analyzed the cases of death due to cerebrovascular disease in the municipalities of the state of Rio de Janeiro using a Space-Temporal approach. In the spatial analysis we used Census Data in order to verify the influence of the average schooling and the Human Development Index of each municipality in the risks of

death from cerebrovascular disease in 2010. On the other hand, in the spatio-temporal analysis we studied the evolution of the risk of death due to cerebrovascular disease in the municipalities from 2010 to 2015, aggregating the information for each municipality and year. Initially, an exploratory analysis was performed in two stages: data visualization and analysis of the spatial autocorrelation. The first one allowed us to verify how data were distributed and clarify their evolution over time. Secondly, we computed spatial autocorrelation measurements to verify the existence of spatial dependence between municipalities. Based on the exploratory analysis, traditional spatial models were adjusted in order to capture global spatial effects (more specifically, the Spatial AutoRegressive Model - SAR and the Spatial Error Model - SEM). Finally, it was also adjusted a Bayesian Spatial Dynamic Generalized Linear Model using MCMC methods in order to identify and quantify local spatial effects that could be thought as causes of increases (or decreases) in the risk of death by this disease in each municipality. The results showed differences in the risks over time and between areas of the Rio de Janeiro state.
Keywords: *cerebrovascular disease; spatial analysis; Bayesian Spatial Dynamic Generalized Linear Model ;*

68. **Título:** Spatio-temporal Poisson models applied to ecological populations
Autores: Izabel Nolau de Souza; João Batista de Morais Pereira
Abstract: In Ecology, researchers are continuously challenged to understand complex ecological processes and ecosystems. In particular, in the study organisms population like plants and animals, one is often interested in understanding how individuals of a particular species behave in their environment as well as how their population evolve over time. In the modeling of processes characterized by a temporal structure, dynamic models are constantly explored. On the other hand, process that present spatial variation can be modeled under the spatial statistics approach. In this work, Poisson spatio-temporal models are proposed to investigate the spatio-temporal behavior of the bird species known as horned lark in California, United States, during the years 1968-2015. We assume that counts of bird specimens in a particular year and in a particular location are governed by a Poisson distribution, whose intensity is composed of temporal and spatial components, with the first being assumed to evolve dynamically over time and the latter being modeled as a Gaussian process. Applications for artificial and real data are made. Proposed models are discussed and compared for the particular applications by means of comparison criteria as DIC and RPS. The inference procedure is done under the Bayesian approach and Monte Carlo Markov chain methods (MCMC) are used to obtain samples of the quantities of interest.
Keywords: *counting data; spatial statistics; dynamic models; spatio-temporal models; Bayesian inference;*

69. **Título:** Spatio-temporal modeling for Transformed Gaussian Markov Random Fields
Autores: Douglas R. Mesquita Azevedo; Marcos Oliveira Prates
Abstract: Models that are capable of capturing the spatial and temporal characteristics of the data are applicable in many science fields. Non-separable spatio-temporal models were introduced in the literature to capture these features, however, these models are commonly complicated in its interpretation and construction. In this work we introduce a class of nonseparable Transformed Gaussian Markov Random Fields (TGMRF) where the dependence structure is not only flexible but also provides simple interpretation to the spatial, temporal and spatio-temporal parameters in the random effects. Another advantage is that the TGMRF settings allow for specialists to define any desired margins. Therefore, the construction of spatio-temporal models using the TGMRF framework leads to a new class of general models such as spatio-temporal gamma random fields, that can be directly used to model Poisson intensity for space-time data. The proposed models were applied to the abundance data of Nenia Tridens to pick out important environmental variables that affect their abundance and also study possible spatial and temporal trends.
Keywords: *Spatio-temporal; Gaussian copula; GLMM; Spatial confounding;*

70. **Título:** Spatiotemporal diffusion of influenza A (H1N1): starting point, risk factors, and proposed vaccination targets
Autores: Ana Carolina Carioca da Costa; Aline Araújo Nobre; Claudia Torres Codeço; Elias Teixeira Krainski; Marcelo Ferreira da Costa Gomes
Abstract: Influenza constitutes a major challenge to world health authorities due to high transmissibility and the capacity to generate large epidemics. This study aimed to characterize the diffusion process of influenza A (H1N1)

by identifying the starting point of the epidemic as well as climatic and sociodemographic factors associated with the ocurrence and intensity of transmission of the disease. The study was carried out in the Brazilian state of Paraná, where H1N1 caused the largest impact. Under the Bayesian paradigm, parametric inference was performed through a two-part spatiotemporal model and the integrated nested Laplace approximation (INLA) algorithm. We identified the most likely starting points through the effective distance measure based on mobility networks. The proposed estimation methodology allowed for rapid and efficient implementation of the spatiotemporal model, and provided evidence of different patterns for chance of occurrence and risk of influenza throughout the epidemiological weeks. The results indicate the capital city of Curitiba as the probable starting point, and showed that the interventions that focus on municipalities with greater migration and density of people, especially those with higher Human Development Indexes (HDIs) and the presence of municipal air and road transport, could play an important role in mitigation of effects of future influenza pandemics on public health. These results provide important information on the process of introduction and spread of influenza, and could contribute to the identification of priority areas for surveillance as well as establishment of strategic measures for disease prevention and control. The proposed model also allows identification of epidemiological weeks with high chance of influenza occurrence, which can be used as a reference criterion for creating an immunization campaign schedule.

Keywords: *Influenza A (H1N1); spatiotemporal modeling; Hurdle models; INLA;*

71. **Título:** State space mixed models for binary response with asymmetric Laplace distribution link
Autores: Vanessa S. Santos; Carlos A. Abanto-Valle
Abstract: A state space mixed model with asymmetric Laplace distribution link is proposed. The threshold latent approach to represent the binary system as a linear state space model is considered. Under a Bayesian perspective, Markov chain Monte Carlo (MCMC) methods are use for parameter estimation.
Keywords: *Quantile regression ; binary regression; state space models; asymmetric laplace distribution;*

72. **Título:** Statistical pairs that preserve statistical relations
Autores: Jaime Lincovil ; Alexandre Patriota
Abstract: Commons interpretations of the principles and theorem in Birnbaum(1962) are related with estimators or evidence measures (Berger,1988). Evans(2013) proposed to formulate of these principles and Birnbaum's theorem by using set theory. Following to Evans(2013), we denote by $\mathcal{I}$ the class of all pairs $I=(E,x)= \big((\mathcal{X}_E,\mathcal{F}_E),x \big)$, such that $\mathcal{X}_E$ and the parameter space of the model E (Θ_E) are finite subsets (or finite unions) of Euclidean spaces and $x\in\mathcal{X}_E$ } We define the class $\mathcal{B}$ of all pairs $((E,x),\Theta_0)$, such that $(E,x)\in \mathcal{I}$ and $\Theta_0 \subseteq \Theta_E$. Let $R: \mathcal{B} \rightarrow \mathcal{C}$ be a function, where $\mathcal{C}$ is doted by an equivalence relation $\stackrel{\bullet}{=}$. A \textit{statistical pair} is defined as a pair $(R,\stackrel{\bullet}{=})$. Let $L \subset \mathcal{I}\times \mathcal{I}$ be the \textit{statistical relation} which relates all pairs (E,x) with the same parameter space and proportional likelihood function. The \textit{likelihood principle} could be formulated as ``choose $(R,\stackrel{\bullet}{=})$ such that $R((E_1,x_1),\Theta_0) \stackrel{\bullet}{=} R((E_2,x_2),\Theta_0) \ \forall \ ((E_1,x_1),(E_2,x_2)) \in L$ and $\forall \ \Theta_0 \subseteq \Theta_{E_1}$". Using this framework, we define the conditional and sufficiency principle and provide a proof for the Birnbaum's theorem.
Keywords: *Statistical pair; Statistical relation; Statistical principle; Birnbaum's theorem;*

73. **Título:** Term structure of interest rates: robust methods
Autores: William Lima Leão; Carlos A. Abanto-Valle
Abstract: The Term Structure of Interest Rates plays a key role in the economic scenario so that the shape of the interest rate curve gives an idea of current economic activity and provides relevant information to predict possible changes in future rates. Understanding the mechanism of the curves is fundamental for several areas, such as: securities pricing, portfolio management, asset allocation, etc. In this paper we introduce an extension to the basic model of Nelson and Siegel with the adoption of robust structures that improve both fit and forecasts, analyzing the main stylized facts present in the rates and identifying the relation between the term structure and the conjuncture current macroeconomic scenario. An empirical application based on US government bond interest rates through an estimation process, under the Bayesian paradigm, is performed with the construction of an efficient algorithm based

on Monte Carlo simulations via Markov Chains (MCMC). The results reveal that both in-sample fit and out-of-sample forecast improve significantly in relation to the Nelson and Siegel's yield curve.

Keywords: *Term Structure; Bayesian Inference; MCMC; Stochastic Volatility; Markov Switching;*

74. **Título:** ZIP model regression for censored data: An application to university dropout

Autores: Diana Milena Galvis; Victor Hugo Lachos; Mauricio Castro

Abstract: Counts data arise in different areas. This type of data can be analyzed using models such as Poison or Negative Binomial. However, it is also possible to observe censored data as well as a quantity of zeros that exceeds the expectation for these models. In this case the previously mentioned models cannot be used, thus in this work, we propose the zero inflated Poisson (ZIP) censored regression models and specifically we applied it to modeling the problem of university dropout. Our motivating data comes from a cohort study performed by the University of Quindio - Colombia where every student, who was enrolled in the first semester of 2012, was followed during 8 semesters. In this study the recorded variables include age, sex, results of academic tests in different areas, number of subjects registered who approve the disciplines, among others. The initial analysis allow us to conclude that around 20% of these students left your course at the end of the first semester and approximately 40% finished all the eighth semesters. Based on these results, we propose to use the ZIP censored regression model to identify students with high probability of leave their early studies because this would reduce the Students' desertion and increase the graduation rate.

75. **Título:** Predicting Churn with Statistical Learning

Autores: Nathália Demetrio V. Moura; Victor Fossaluza

Abstract: The churn rate refers to the proportion of customers who leave a service during a given time period. It is a possible indicator of dissatisfaction, better offers from the competitors, or even reasons having to do with the customer's life cycle. However there are contexts where the churn does not occur officially, the case of account holders who start to use another bank as their main option for example , where there is no churn, in the definition point of view, but the financial loss is equivalent. Such cases have as a challenge not only to identify which behavior change can be treated as churn, but also to do so early enough for the company to take actions to prevent the exit. In this work we will consider a simulated data set based on a bank institution which wishes to detect the changes in the costumer behavior that will be detrimental to its business, using the costumer historical itself, what includes deposits, types of spending, and bill payments. In addition we will generate possible external information, like fidelity programs, Customer Relationship Management (CRM), and even social networks use, since the possibility of including potentially relevant but unstructured information is a increasingly present reality in the analyzes nowadays. Then, comparing diferents statistical learning methods, we will indicate which one will best predict the churn proposed, considering also the discussion of the advantagens and limitations that each methodology presents.

Patrocínio
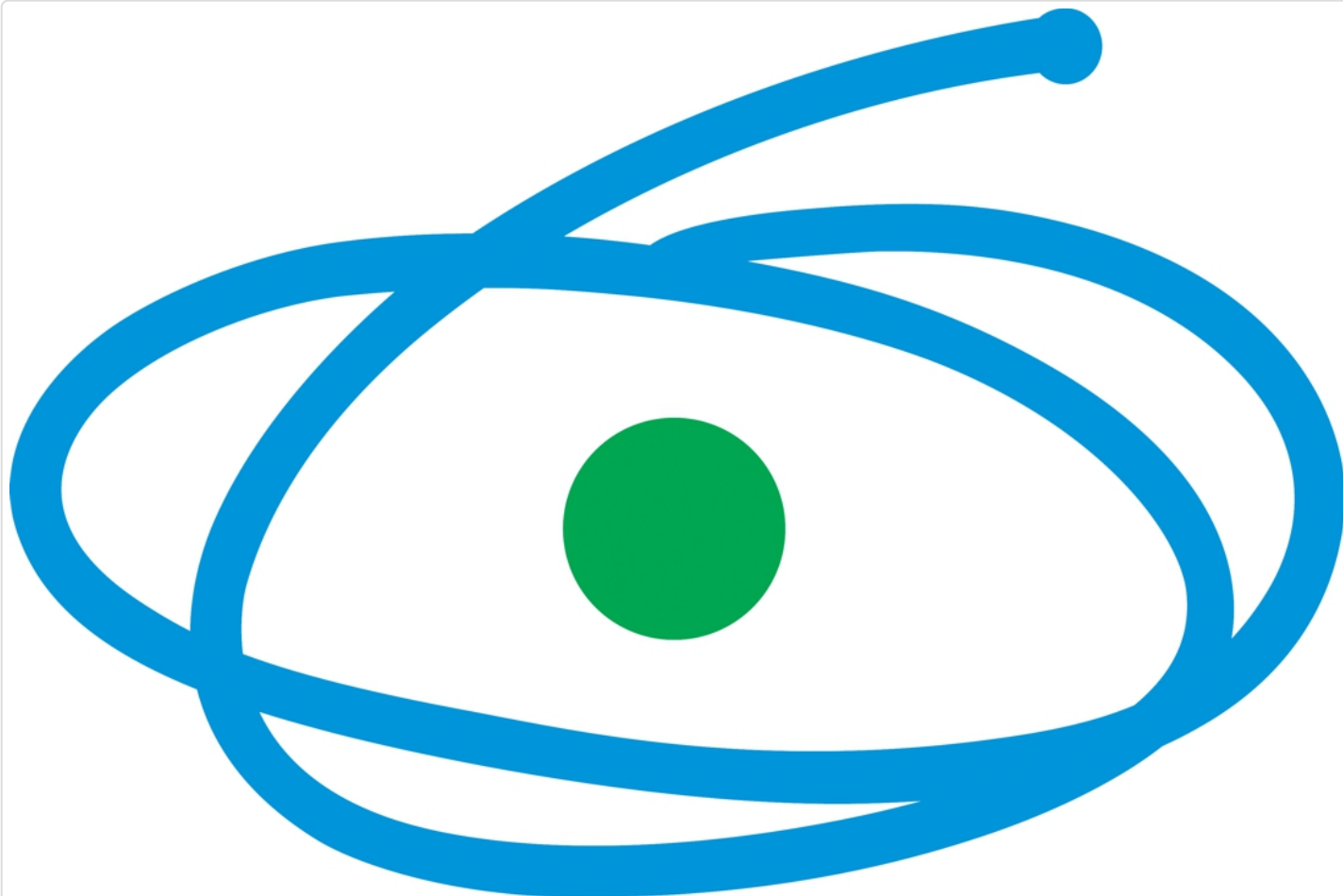
Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (http://www.capes.gov.br/)

Instituto Brasileiro de Geografia e Estatística (http://www.ibge.gov.br/)

Organização

ISBrA - Seção Brasileira da International Society for Bayesian Analysis (http://www.ime.usp.br/~isbra/)

Apoio

International Society for Bayesian Analysis (http://bayesian.org/)

Instituto Nacional de Matemática Pura e Aplicada (http://impa.br/)

Associação Brasileira de Estatística (http://www.redeabe.org.br/site/)

Instituto de Matemática - UFRJ (http://www.im.ufrj.br/)