XV Brazilian Meeting of Bayesian Statistics
Beach Hotel Maresias, São Paulo
March 9th - 12th 2020

https://www.ime.usp.br/~isbra/ebeb2020

EBEB Program

Time	Monday	Tuesday	Wednesday
08:30-10:00	Conference	T1: Hahn/T2: Stern	CS3 and CS4
10:00-10:30	registration	coffee break	coffee break
10:30-12:00	& check-in	T1: Hahn/T2: Stern	CS5 and CS6
12:00-14:00	lunch break	lunch break	lunch break
14:00-14:45	IT2: Rosangela Loschi	IT4: Ioanna Manolopoulou	IT6: Richard Hahn
14:45-15:30	IST2: Luis Gutierrez	IST3: Zuanetti	IST4: Rafael Stern
15:30-16:00	coffee break	coffee break	coffee break
16:00-16:45	IT1: Nancy Garcia	IT5: Friel	IT7: Márcia Branco
16:45-17:30	IT3: Miguel de Carvalho	CS1 and CS2	ISBRA Assembly
17:30-18:15	//////////////	CS1 and CS2	//////////////
18:00-20:00	poster & cocktail	//////////////	//////////////
20:00-22:00	dinner	dinner	farewall dinner

T: Tutorial - 3-hour lectures
IT: Invited talks - 40-min-talk + 5-min Q&A
IST: Invited short talks - 30-min talk + 10-min discussant + 5-min reply
CS: Contributed sessions: 3 speakers per session (25-min talk + 5 Q&A)
Poster: Poster session
Cocktail: Get together along poster session
Beach party: Farewell party (reservation required)

Contents

1 Tutorials (T)

1.1 Tutorial 1: Bayesian machine learning for causal inference

Speaker: Richard Hahn
Affiliation: School of Mathematical and Statistical Sciences, Arizona State University
URL: https://math.la.asu.edu/~prhahn

Abstract: Causal inference is a fundamental goal of science and policy evaluation. In situations where randomized experiments are not possible, such as many medical and social science context, one must rely on modeling assumptions to produce valid causal inferences. However, the less credible these modeling assumptions, the less credible the associated inferences. Therefore, flexible regression modeling (supervised learning) is an important tool when performing an applied causal data analysis. It is well known that flexible models require regularization to achieve low generalization error (out-of-sample prediction accuracy). Unfortunately, standard regularization techniques that work well for prediction can strongly bias causal estimators in unforeseeable ways. As such, it is necessary to design regression methods specifically tailored to the causal inference setting. This short course will review the basic formalisms of modern causal inference, using elements of both Rubin?s potential outcomes and also Pearl's causal graphs, and explore how various machine learning approaches can be used to estimate causal effects. We will be interested in estimation of average treatment effects and subgroup average treatment effects using both nonlinear regression adjustment as well as nonlinear instrumental variables models. The models we will discuss are modifications of the Bayesian additive regression tree model of Chipman, George and McCulloch. Specifically the Bayesian causal forest model of Hahn, Murray and Carvalho and Accelerated BART (XBART) will be described and demonstrated. These methods represent the current state-of-the-art in causal machine learning. The approach throughout will be Bayesian. Special attention will be paid to simulation studies comparing well-known frequentist alternatives, in which it is shown that the Bayesian methods have dramatically better frequentist coverage and lower estimation error in finite samples.

1.2 Tutorial 2: Aplicações de inferência Bayesiana à jurimetria

Speaker: Rafael Stern
Affiliation: Department of Statistics, Federal University of São Carlos
URL: https://www.rafaelstern.science

Abstract: A Jurimetria consiste da análise quantitativa de dados Jurídicos. Estas análises podem determinar, por exemplo, impacto regulatório, estratégias para obter Tribunais céleres, estudo de conceitos do Direito envolvendo incerteza, e a criação de ferramentas para auxiliar advogados. Neste tutorial discutirei alguns problemas jurídicos e como a análise quantitativa por meio da Inferência Bayesiana e da Teoria da Decisão podem ser de grande auxílio no seu entendimento e resolução.

2 Invited talks (IT)

2.1 IT1: Nancy Garcia

Affiliation: State University of Campinas
URL: http://www.ime.unicamp.br/~nancy

Title: *Interacting cluster point process model for epidermal nerve fibers*
Abstract: We propose an interacting cluster model for the spatial distribution of epidermal nerve fibers (ENF). The model consists of a spatial process of parent points modeling the base points of nerve fiber bundles. To each base point there is an offspring point process of fibers end points associated. The parent process is, possibly, inhibited by fiber endings that belong to different base bundles. The fibers themselves have random length and spatial orientation. We consider a non-orphan process where we can connect each offspring to a parent. We examine the processes that can be described with our model, how coefficient estimation can be performed under the Bayesian paradigm via Markov chain Monte Carlo methods, and detail approaches for the implementation of efficient MCMC sampling schemes. We study the performance of the estimation procedure through a simulation study. An application to data from skin blister biopsy images of ENFs is presented. Joint work with Peter Guttorp (Norwegian Computing Center) and GuilhermeLudwig (State University of Campinas).

2.2 IT2: Rosangela Loschi

Affiliation: Department of Statistics, Federal University of Minas Gerais
URL: http://www.est.ufmg.br/~loschi

Title: *Handling categorical features with many levels using a product partition model*
Abstract: A common difficulty in data analysis is how to handle categorical predictors with a large number of levels or categories. Few proposals have been developed to tackle this important and frequent problem. We introduce a generative model that simultaneously carries out the model fitting and the aggregation of the categorical levels into larger groups. We represent the categorical predictor by a graph where the nodes are the categories and establish a probability distribution over meaningful partitions of this graph. Conditionally on the observed data, we obtain a posterior distribution for the levels aggregation allowing the inference about the most probable clustering for the categories. Simultaneously, we extract inference about all the other regression model parameters. We compare our and state-of-art methods showing that it has equally good predictive performance and much better interpretable results. Our approach balances out accuracy versus interpretability, a current important concern in machine learning. Joint with Tulio L. Criscuolo, Renato M. Assunção and Wagner Meira Jr. (Department of Computer Science, Federal University of Minas Gerais).

2.3 IT3: Miguel de Carvalho

Affiliation: School of Mathematics, University of Edinburgh
URL: https://www.maths.ed.ac.uk/~mdecarv

Title: *Elements of Bayesian geometry*
Abstract: In this talk, I will discuss a geometric interpretation to Bayesian inference that will yield a natural measure of the level of agreement between priors, likelihoods, and posteriors. The starting point for the construction of the proposed geometry is the observation that the marginal likelihood can be regarded as an inner product between the prior and the likelihood. A key concept in our geometry is that of compatibility, a measure which is based on the same construction principles as Pearson correlation, but which can be used to assess how much the prior agrees with the likelihood, to gauge the sensitivity of the posterior to the prior, and to quantify the coherency of the opinions of two experts. Estimators for all the quantities involved in our geometric setup are discussed, which can be directly computed from the posterior simulation output. Some examples are used to illustrate our methods, including data related to on-the-job drug usage, midge wing length, and prostate cancer. Joint work with G. L. Page and with B. J. Barney.

2.4 IT4: Ioanna Manalopoulou

Affiliation: Department of Statistical Science, University College of London
URL: http://www.homepages.ucl.ac.uk/~ucakima

Title: *Computationally efficient methods for Value of Information measures*
Abstract: Value of Information (VoI) measures and their use in public policy decision-making, such as health economics, have seen an increase of research attention in recent years. This is, in part, due to the development of computationally efficient methods for computing VoI measures. We will review three key concepts, namely the Expected Value of Perfect Information (EVPI), the Expected Value of Partial Perfect Information (EVPPI) and the Expected Value of Sample Information (EVSI), highlighting their computational challenges. We will present two sets of methods, one based on computationally efficient Gaussian Process regression based on Integrated Nested Laplace Approximation to compute the EVPPI and the other based on ”moment-matching” to compute the EVSI. Both methods draw on existing methodologies but expand them by providing general-purpose algorithms that can be applied on a wide range of real-life modelling structures. We will present the methods using toy and real-life examples and discuss their advantages and limitations with respect to applicability in practice. Joint work with Gianluca Baio and Anna Heath.

2.5 IT5: Nial Friel

Affiliation: School of Mathematics and Statistics, University College Dublin
URL: https://maths.ucd.ie/~nial

Title: *Informed sub-sampling MCMC: approximate Bayesian inference for large datasets*
Abstract: This talk introduces a framework for speeding up Bayesian inference conducted in the presence of large datasets. We design a Markov chain whose transition kernel uses an unknown fraction of fixed size of the available data that is randomly refreshed throughout the algorithm. Inspired by Approximate Bayesian Computation, the subsampling process is guided by the fidelity to the observed data, as measured by summary statistics. The resulting algorithm, Informed Sub-Sampling MCMC, is a generic and flexible approach which, contrary to existing scalable methodologies, preserves the simplicity of the Metropolis?Hastings algorithm. Even though exactness is lost, i.e the chain distribution approximates the posterior, we study and quantify theoretically this bias and show on a diverse set of examples that it yields excellent performances when the computational budget is limited. If available and cheap to compute, we show that setting the summary statistics as the maximum likelihood estimator is supported by theoretical arguments.

2.6 IT6: Richard Hahn

Affiliation: School of Mathematical and Statistical Sciences, Arizona State University
URL: https://math.la.asu.edu/~prhahn

Title: *Beyond Bayes rule: Simulation experiments for principled data science*
Abstract: Bayesian modeling has many attractive virtues and often leads to estimators that work exceptionally well in practice. But sophisticated Bayesian models have notable drawbacks as well: theoretical intractability, steep computational burden, and poor calibration under model misspecification. Through a detailed case study of my own work on Bayesian causal forests (BCF) and Accelerated Bayesian Additive Regression Trees (XBART) I describe an approach that blends Bayesian modeling with efficient non-Bayesian algorithms and evaluates the resulting estimators via prior predictive simulations. The resulting data analysis approach might be characterized as that of a committed (nonparametric) Bayesian who, in an effort to prevent the perfect from being the enemy of the good, has renounced allegiance to Bayes rule. More concretely, this talk is about the wide practical application of Monte Carlo evaluations of Bayes risk.

2.7 IT7: Marcia Branco

Affiliation: Department of Statistics, University of São Paulo
URL: https://www.ime.usp.br/~mbranco

Title: *Skew-elliptical distributions 20 years later*
Abstract: This talk intends to discuss the state of the art of asymmetric multivariate models related to Skew-Elliptical distributions. In the last 20 years a significant amount of articles have been published on this subject. Many of them developing new classes of multivariate distributions, others more focused on flexible statistical modeling. A historical background of the Skew-Elliptical class will shortly be presented. Then, I focus on more application-oriented works showing severals relevant application in different areas of knowledge. Let us try to understand the advantages and disadvantages of using these models through examples.

2.8 IT8: Marina Vannucci (cancelled)

3 Invited short talks (IST)

3.1 IST1: Isadora Antoniano-Villalobos (cancelled)

3.2 IST2: Luis Gutierrez

Affiliation: Department of Statistics, Pontificia Universidad Católica de Chile
URL: http://statistics-luisgutierrez.com

Title: *Bayesian nonparametric hypothesis testing procedures*
Abstract: Scientific knowledge is firmly based on the use of statistical hypothesis testing procedures. A scientific hypothesis can be established by performing one or many statistical tests based on the evidence provided by the data. Given the importance of hypothesis testing in science, these procedures are an essential part of statistics. The literature of hypothesis testing is vast and covers a wide range of practical problems. However, most of the methods are based on restrictive parametric assumptions. In this talk, we will discuss Bayesian nonparametric approaches to construct hypothesis tests in different contexts. Our proposal resorts to the literature of model selection to define Bayesian tests for multiple samples, paired-samples, and longitudinal data analysis. Applications with real-life datasets and illustrations with simulated data will be discussed.

3.3 IST3: Daiane Zuanetti

Affiliation: Department of Statistics, Federal University of São Carlos
URL: http://www.des.ufscar.br/departamento/docentes-1/daiane-aparecida-zuanetti

Title: *Subset nonparametric Bayesian clustering: An application in genetic data*
Abstract: We propose a nonparametric Bayesian method to cluster big data and apply them to cluster genes by patterns of gene-gene interaction. The approach define model-based clustering with nonparametric Bayesian priors and include an implementation that remains feasible for big data. It is an exact method that divides the data into smaller subsamples and involves local partitions that can be determined in parallel. In a second step, the method requires only the sufficient statistics of each of these local clusters to derive global clusters. Under simulated and benchmark data sets the proposed method compare favorably with other clustering algorithms, including k-means, DP-means, DBSCAN, SUGS, streaming variational Bayes (SVB) and an EM algorithm.

3.4 IST4: Rafael Stern

Affiliation: Department of Statistics, Federal University of São Carlos
URL: https://www.rafaelstern.science

Title: *Predictive model checks for phylogenetical models*
Abstract: Phylogenetical models determine evolutionary relationships between species based on their traits. These relationships are commonly expressed through a tree and established based on a statistical model for the evolution of traits. Since there is a wide variety of evolutionary models, it is desirable to perform model checks. However, it is hard to perform these checks since phylogenetical models are unsupervised. We propose a check based on the accuracy of the models in predicting hidden traits. In a wide variety of simulated data our check recovers the model with the most accurate tree.

4 Contributed sessions (CS)

4.1 CS1: Contributed session 1

- **Bias correction in underreported count data using a clustering strategy**
  Speaker: Guilherme Lopes de Oliveira, CEFET-MG
  Abstract: Data quality from poor and socially deprived regions have given rise to many statistical challenges. One of them is the underreporting of vital events leading to biased estimates for the associated risks. To deal with underreported count data, models based on compound Poisson distributions have been commonly assumed. To be identifiable, such type of models usually require extra and strong information about the probability of reporting the event in all areas of interest, which is not always available. We introduce a novel approach for the compound Poisson model clustering the areas and assuming a structure that relates the reporting probabilities among the clusters. We prove that only prior information about the reporting probability in areas with the best data quality is required for model identification. Several approaches to model the uncertainty about the reporting probabilities are presented, including reference priors. Different features regarding the proposed methodology are studied through simulation. We apply our model to map the early neonatal mortality risks in Minas Gerais, a Brazilian State that presents heterogeneous characteristics and a relevant socio-economical inequality. Joint with Rosangela H. Loschi (UFMG), Renato Assunção (UFMG), Rafaela Argiento (CNR-IMATI), Marcia Branco (USP) and Fabrizio Ruggeri (CNR-IMATI).

- **A Conway-Maxwell-Poisson GARMA Model for Count Time Series Data**
  Speaker: Ricardo Ehlers, University of São Paulo at São Carlos
  Abstract: We propose a flexible model for count time series which has potential uses for both underdispersed and overdispersed data. The model is based on the Conway-Maxwell-Poisson (COM-Poisson) distribution with parameters varying along time to take serial correlation into account. Model estimation is challenging however and require the application of recently proposed methods to deal with the intractable normalising constant as well as efficiently sampling values from the COM-Poisson distribution.

- **Structural equation modeling with time dependence: an application comparing Brazilian energy distributors**
  Speaker: Vinicus Mayrink, Federal University of Minas Gerais
  Abstract: This study proposes a Bayesian structural equation model (SEM) to explore financial and economic sustainability indicators, considered by the Brazilian energy regulator (ANEEL), to evaluate the performance of energy distribution companies. The methodology applies confirmatory factor analysis for dimension reduction of the original multivariate data set into few representative latent variables (factors). In addition, a regression structure is defined to establish the impact of the factors over the response "indebtedness" of the companies; this is a central aspect regularly discussed within ANEEL to identify whether a distributor may have difficulty to manage the concession. Most of the variables in this study are collected for 8 diferent years

(2011-2018), therefore, a time dependence is inserted in the analysis to correlate observations. The SEM approach has several advantages in this context: it avoids using criticisable deterministic formulations to measure non-observable aspects of the distributors, it allows a broad statistical analysis exploring elements that cannot be investigated through the simple descriptive studies currently developed by the regulator and, finally, it provides tools to properly rank and compare distances between companies. The Bayesian view is a powerful option to handle the SEM fit here, since convergence issues, due to sample size and high dimensionality, may be experienced via classical alternatives based on maximization. Joint with Renato Panaro (UFMG) and Marcelo Costa (UFMG)

4.2 CS2: Contributed session 2

- **Logistic modelling of point process: a Bayesian approach for presence-only data**
  Speaker: Guido Moreira, Federal University of Rio de Janeiro
  Abstract: Species distribution models (SDMs) are extremely useful for determining preferences and habitats for different species. Appropriate estimation of species distribution depends on the adequate random sampling scheme which is not always available. Instead, data is frequently composed of geo-referenced locations where the species has been observed, which is commonly referred to as presence-only (PO) data. The statistical modelling of PO type data through Inhomogeneous Poisson Processes (IPP) was proposed by Fithian and Hastie (2013). As has already been noted (Fithian et al, 2015), PO type data presents bias in its sampling pattern, which must be addressed. A natural way to model this bias under IPP is through thinning of the process, which is easily performed using pertinent covariates. A different model for the intensity is proposed using a logistic link function. It maintains the already established flexibility, while adding extra flexibility in the choice of covariates. Therefore, it is possible to have correlated and even the same covariates in the intensity and thinning components of the model. This is shown through simulated results and an application on a real data set. Additionally, it provides computational advantages for handling integrals that appear in the likelihood without resorting to approximations. Finally, all the inference is based on the Bayes paradigm, which provides a way to consider the uncertainty inherent to the unknown components in the analysis. Joint with Dani Gamerman (UFRJ).

- **Geostatistics under preferential sampling in the presence of local repulsion effects**
  Speaker: Gustavo Ferreira, National School of Statistical Sciences
  Abstract: This paper presents an extension of the Geo-statistical model under Preferential Sampling in order to accommodate possible local repulsion effects. This local repulsion can be caused by the researcher in charge of collecting data who, after observing the stochastic process of interest in a specific location, avoids collecting new samples near this place. Proceeding in this way, the resulting sampling design would in practice include a repulsion window centered on each sampling location, even though the researcher was planning the sample preferentially. This perturba-

tion in the Geostatistical model under Preferential Sampling can be modeled through a discrete nonhomogeneous stochastic process over a partition composed of M sub-regions of the study area, where only one sample lies in each sub-region. Simulations and an application to real data are performed under the Bayesian approach and the effects of this perturbation on estimation and prediction are then discussed. The results obtained corroborate the idea that the proposed methodology corrects the distortions caused by this perturbation, thus mitigating the effects on inference and spatial prediction.

- **Model-based Inference for Rare and Clustered Populations from Adaptive Cluster Sampling using covariates**
  Speaker: Izabel Souza, Federal University of Rio de Janeiro
  Abstract: Rare populations, such as endangered animals and plants, drug users and individuals with rare diseases, tend to cluster in regions. Adaptive cluster sampling is generally applied to obtain information from clustered and sparse populations since it increases survey effort in areas where the individuals of interest are observed. This work aims to propose a unit-level model which assumes that counts are related to auxiliary variables, improving the sampling process, assigning different weights to the cells, besides referring them spatially. The proposed model fits rare and grouped populations, disposed over a regular grid, in a Bayesian framework. The proposed model is evaluated using simulated data and a real experiment in which adaptive samples were drawn from an African Buffaloes population in a 24,108km2 area of East Africa. Simulation studies show that the model is efficient under several settings, validating the methodology proposed in this paper for practical situations. Joint with Kelly Gonçalves (UFRJ) and João Pereira (UFRJ).

4.3 CS3: Contributed session 3

- **A new method for sequential learning of states and Parameters for state space models. The particle swarm learning optimization**
  Speaker: Ivan Guzman, Federal University of Espirito Santo
  Abstract: Accuracy of parameter estimation and efficiency of state simulation are common concerns in the implementation of state-space models. Kalman filters combined with Markov Chain Monte Carlo and its variants, presented good results in smoothing, filtering, and predicting the state. Yet, these methods become less accurate and efficient when non-Gaussianity and non-linearity are present in data. Particle Filter technique is among the most successful alternative methods to improve the non linearity and non-Gaussianity. However, it presents several problems regarding the convergence of parameters, besides showing a high level of inefficiency. In this article, we developed a new method combining the structure of particle learning and bare borne particle swarm optimization (BBPSO) to the process of smoothing and filtering the states in the state space models, thus overcoming the efficiency and convergence problem of the particle filter methods. Sampling importance re-sampling is used to estimate the states of the model, then the parameters of the model can be estimated via BBPSO. The newly proposed method is applied to stochastic volatility

models and AR(1) state-space models, which are non-linear models where the assumption of normality is relaxed. Two case studies are illustrated by using the data from Ibovespa and S&P500 index. Empirical results show that the new methodology exhibited better performance when compared to particle filters given a solution the inefficiency and accuracy problems. Joint with P. Corib (USP).

- **Understanding the Economic Policy Uncertainty index using semi-automatic news classification**
  Speaker: Rodrigo Targino, Getulio Vargas Foundation
  Abstract: The Economic Policy Uncertainty (EPU) index from Baker et al. (QJE, 2016) is computed as a proportion of news articles that contain EPU-related words over the total number of articles in the period. Although fully automatic, this approach requires the group of EPU-related words to be chosen in advance, which leaves the index subject to changes in the vocabulary used to describe economic policy uncertainty. In this work we take, instead, a semiautomatic route to text classification: first a sample of the articles is manually read and classified as belonging to the topic (e.g. EPU) or not; then, supervised learning algorithms are used to classify the unread articles and to compute the index. In this paper we present results for both a static and a dynamically sparse regression model. Joint with Angel Arroyo (UFRJ) and Yuri Saporito (FGV).

- **Decoupling Shrinkage and Selection in Gaussian Linear Factor Analysis**
  Speaker: Henrique Bolfarine, University of São Paulo
  Abstract: Sparsity-inducing priors has been a relevant option in variable selection in a variety of statistical models. Despite the interpretability and ease of application, there are still divergences in determining whether a parameter a posteriori is really null. In this context, the decoupling shrinkage and selection (DSS) approach appears as an alternative that preserves the a posteriori information while providing an optimal selection in the set of variables. In this paper, we extend the DSS methodology for the gaussian linear factor analysis model in order to obtain a sparse loadings matrix, reducing to zero the parameters that are not relevant to the model. To perform such selection, we introduce a penalized loss function, a post inference procedure that relies on a penalized predictive version of the expectation-maximization (EM) algorithm, and a graphical summary. The findings are illustrated with simulations and two applications, the first in psychometrics and the latter in denoising of handwritten data. The standard normal and point mass priors were used, resulting in significantly different levels of sparsity in the recovered loading matrix. Joint with Carlos Carvalho (UT Austin), Hedibert Lopes (Insper) and Jared Murray (UT Austin).

4.4 CS4: Contributed session 4 (one cancellation)

- **A zero augmented Birbaum-Saunders mixed model with repeated measures: a Bayesian approach**
  Speaker: Mariana Motta, State University of Campinas
  Abstract: A zero augmented mixed regression model of positive outcome with a Birbaum-Saunders distribution is study in this work. Motivated by a data set where subjects have 20 repeated measures each, one study goal is to investigate model behavior by introducing intra subject random effects, thus being able to estimate intra cluster variability in zero augmented models for positive response. The other study goal is to investigate to the extent to which the amount of repeated measures, along with sample size, an effect estimation of intra subject variation. We adopt a Bayesian hierarchical modeling approach to estimate the latent variables and parameters. using the motivating data we show how the proposed method can be applied to extract valuable information from repeated measures, helping to identify the optimum sample size of repeated measures in the context of zero augmented positive regression models. Joint with Claudia Koda (UNICAMP), Elainy Batista (UNICAMP), Filidor Labra (UNICAMP) and Eliseu Junior (UERJ).

- **Bayesian variable selection in Funcional Adaptive Models with application to identification of early responders**
  Speaker: Nancy Lopes, State University of Campinas
  Abstract: We consider a model for which we have a scalar response which belongs to a mixture of normal distributions. The influence of the covariates, which can be scalar or functional, is through the latent variable that determines the mixture. Taking the traditional approach of considering functional covariates as vectors brings the curse of high dimensionality since these covariates can be measured at very small intervals. The goal of this work is to use a Bayesian approach to reduce the dimensionality of the problem. We propose the use of a zero augmented distribution as prior in order to perform variable selection. Joint with Mariana Motta (UNICAMP), Eva Petkova (NYU), Thaddeus Tarpey (Wright State University) and Todd Ogden (Columbia University).

- **Objective Bayesian Analysis for Multiple Repairable Systems (cancelled)**
  Speaker: Vera Tomazella, Federal University of São Carlos

4.5 CS5: Contributed session 5 (one cancelation)

- **Dynamic Degradation Linear model: Application to IRLED data**
  Speaker: Guilherme Augusto, Federal University of Minas Gerais
  Abstract: General path models are usually considered to model degradation data. Its popularity is, in part, due to the simplicity in which different devices features may be accommodated. Particularly, models assuming a linear relationship between time and degradation, in which the degradation rate for each unit under test is not time variant, have received considerable attention. Invariance is a reasonable assumption only if the degradation increment is constant over time. To account for non-regular structures in the degradation increment and, consequently, for piecewise linear behaviors in the degradation paths, we introduce a dynamic linear degradation models. Dynamic is introduced into the model by assuming that both, the baseline degradation and the degradation rate, vary along the time. We run a simulation study to evaluate the proposed model and to compare it to the Weibull linear degradation model. To illustrate its use in real situations, the proposed model is fitted to analyze the IRLED dataset. Results show that the proposed model is a useful approach to model degradation data. Joint with Rosangela Loschi (UFMG)

- Moved to poster number 27
  Robust Bayesian modeling using finite mixtures of Student t distributions
  Speaker: Nivea Bispo, Federal University of Bahia

- **Dependent Mixtures: Modelling Cell Lineages**
  Speaker: Carlos Zanini, University of Texas at Austin
  Abstract: We introduce dependent mixture models for model-based inference in mixtures when the cluster locations are naturally connected by a spanning tree. The motivating application is inference for cell lineage data on the basis of single cell RNA-seq data for cells across different levels of cell differentiation. The terms of the mixture model are interpreted as representing distinct cell types, including a known root cell population and final differentiated cells. We propose two prior models based on shrinkage of the cumulative length of the underlying spanning tree of cluster centers: the hard-MST and the soft-MST, based on minimum spanning trees (MST). Both prior models assume the partition of cells to depend on the lineage structure, which is more biologically reasonable then the usual multistep approach in which partitions are estimated disregarding the underlying tree structure that characterizes cell lineage data. We provide full Bayesian inference via MCMC on the clusters of cells (including number of clusters), on the underlying tree structure and also on pseudo-times. Joint with Giorgio Paulo (UT at Austin) and Peter Müller (UT at Austin).

4.6 CS6: Contributed session 6

- **Student's-t process for geostatistical models with spatial deformation**
  Speaker: Marina Silva Paez, Federal University of Rio de Janeiro
  Abstract: In this paper we propose a geostatistical model for non-stationary processes in space, for data-sets which present atypical observations and/or are generated through heavy tail distributions. This model relaxes the assumption of isotropy and incorporates anisotropy through spatial deformation. Inference is made under a Bayesian point of view, and we propose the use of Markov Chain Monte Carlo methods to sample from the posterior distribution of the model parameters The proposed model is then applied to a hydrology data-set consisting of measurements of water height from the sea level in 93 aquifer locations in Saratoga Valley, Wyoming. The main conclusion of this application is that considering the DIC criterion the proposed t-variate anisotropic model has a better predictive performance than the most usual Gaussian/isotropic models. Joint with Fidel Morales (UFRN), Dimitris Politis (UC San Diego) and Jacek Leskow (NASK Institute).

- **Dynamical non-Gaussian modelling of spatial processes**
  Speaker: Viviana Lobo, Federal University of Rio de Janeiro
  Abstract: Spatio-temporal processes in environmental applications are often assumed to follow a Gaussian model, possibly after some transformation. However, heterogeneity in space and time might have a pattern that will not be accommodated by transforming the data. In this scenario, modelling the variance laws is an appealing alternative. This work adds flexibility to the usual Dynamical Gaussian model by defining the process as a scale mixture between a Gaussian and log-Gaussian process. The scale is represented by a process varying smoothly over space and time. Moreover, this variance process is allowed to depend on covariates. The resultant kurtosis varies with location, allowing the time series at each location to have different distributions with different tail behaviour. Regarding the dynamical model, the specification is based on state-space equations for both response and variance processes resulting in a computationally efficient estimation and prediction algorithm. Analysis of different artificial data show that we can recover the values of the parameters used to generate the data. An application to maximum temperature data illustrates the effects of altitude in the variability of the process and how this dependence may change over time. Joint with Thaís Fonseca (UFRJ) and Alexandra Schmidt (McGill University).

- **Time series and multilevel modeling for longitudinal item response theory data**
  Speaker: Caio Lucidius, State University of Campinas
  Abstract: Longitudinal Item Response Theory (IRT) data occurs when experimental units are submitted to measurement instruments (e.g., cognitive test, psychiatric questionnaires, biological essays among others) along different assessment conditions, as different time points. Very often, in this kind of study, we are interested in the so-called latent variables (or latent traits) and their behavior along these conditions, including the modeling of their inter-dependency structure. In this work we use some stationary and non-stationary time series and multilevel models to repre-

sent longitudinal IRT data. More specifically, we consider first order autoregressive (AR(1)), first order moving average (MA(1)), first order autoregressive moving average (ARMA(1,1)), antedependence (AD) time series models as well as the Uniform and Hankel dependency structures, induced by appropriate multilevel models. These structures are studied under a time-homocedastic and time-heteroscedastic fashions. We developed a Bayesian inference framework, which includes parameter estimation, model fit assessment and model comparison, through MCMC algorithms. Simulation studies are conducted in order to measure the parameter recovery and model comparison tools. A real data analysis, concerning a longitudinal cognitive study of Mathematics achievement, conducted by the Federal Brazilian government, is performed. All computational implementations are made through the WinBUGS program, using the R2WinBUGS package, from R program. Joint with Andrade (UFSC) and Fox (University of Twente).

5 Poster session (PS)

1. Afonso Cesar Borges, Federal University of Rio de Janeiro

   Title: *Bayesian quantile factor models*

   Abstract: The technological advances and the improvements of computational tools have enabled the availability of large amounts of data, making the dimensionality reduction of great importance in data analysis. However, the main dimensionality reduction technique, the factor analysis, do this at mean association between the components of the original variable. Recent studies have shown a particular interest in evaluating the linear predictors effects on distributions tail parts, not only on a center measure. To answer these questions, the quantile regression is a technique that works well with this kind of analysis. The main goal of the Quantile Factor Model is to incorporate the advantages of factor analysis and quantile regression, which allow the description of the original dependence of a set of observed correlated variables by a smaller number of not observable variables, called latent factors by the quantile dependence structure. The proposed method in this work overcomes two factor analysis restrictive features: (i) hidden factors that may shift characteristics (moments or quantiles) of the distribution of original variable rather than its mean are captured in the method; (ii) the factor loadings are not fixed at only one original distributional characteristic, but it is possible to make it more flexible from another characteristics. Joint with Kelly Gonçalves (UFRJ).

2. Aldo Medina Garay, Federal University of Pernambuco

   Title: *Bayesian analysis of the p-order integer valued AR process with zero-inflated Poisson innovations*

   Abstract: In recent years, there has been considerable interest to study count time series with a dependence structure and appearance of excess of zeros values. Such series are commonly encountered in diverse disciplines, such as economics, nancial research, environmental science, public health, among others. In this paper, we propose a stationary p-order integer-valued autoregressive process with zero inflated Poisson innovations, called the ZINAR(p) times series model. We study some of its theoretical properties and develop a MCMC algorithm for inferring parameters from Bayesian perspectives. Finally, we demonstrate the utility of proposed ZINAR(p) model through simulated and real data examples. Joint with Francyelle Medina (UFPE), Celso Cabral (UFAM) and Tsung-I Lin (National Chung Hsing University).

3. Alex Sousa, University of São Paulo

   Title: *Bayesian wavelet shrinkage with beta priors*

   Abstract: We present a Bayesian approach for wavelet shrinkage in the context of non-parametric curve estimation with the use of the beta distribution with symmetric support around zero as the prior distribution for the location parameter in the wavelet domain in mod- els with additive Gaussian errors. Explicit formulas of shrinkage rules for particular cases are obtained, statistical properties such as bias, classical and Bayesian risk of the rules are analyzed and performance of the proposed rules is assessed in simulations studies involving standard test functions. Application to Spike Sorting real data set is performed. Joint with Nancy Garcia (UNICAMP) and Brani Vidakovic (Georgia Tech).

4. André Martins, University of São Paulo

   Title: *Going fully Bayesian: Infinities in probabilistic inference and the consequences for scientific practice*

   Abstract: Using Bayesian methods to estimate the probability of theories involve dealing with several impossible infinities. We must be able to fully characterise each theory plus all possible sets of auxiliary hypothesis. In principle, we need to be able to calculate likelihoods from all those variations and that basically implies a logical omniscience we can not achieve. And, even if those two steps were feasible, we would need to sum and integrate over a huge, potentially in finite set of possibilities in order to obtain posteriors. The way we usually work around those problems is to limit our inferences to a few reasonable cases and ignore all other possibilities, hoping they would not change our conclusions in any major way. Indeed, that might be the best we can do. However, discussing the full problem is worth it and it can help us improve the ways we make science. In this talk, I will review why it is fundamental we try to use Bayesian methods for theory estimation as close to the impossible ideal as possible. And I will explain how accepting the normative force of Bayesian methods can lead to new insights on methodological problems. Examples from quantum mechanics and string theory will show both how even physicists need to improve their inferential tools and how, in some cases, their reasoning might be seen as a valid approximation to our limited abilities. As a second example, the problems with hypothesis testing will also be discussed.

5. Cristian Villegas, Escola Superior de Agricultura Luiz de Queiroz

   Title: *Retrospective estimation of growth of Chilean common hake based on the otolith measurements*

   Abstract: The age-at-capture and length-at-capture of common hake (*Merluccius gayi gayi*) was collected in San Antonio and Talcahuano areas in Chile between 1967 and 2010. In addition, the number of annual rings in sagittal otolith, a common estimate of the age of the fish, is determined. This estimate is, however, interval-censored. Research questions such as the comparison of the age-at-capture over the years and the relationship of age- with length-at-capture require appropriate modeling techniques. It was also of interest to know whether there are different growth patterns inferred from the otolith radii in the hake and whether there is a trend in growth over the many years data were collected. This involves mixture modeling of longitudinal data. Because of the complexity of the data, a Bayesian approach was used to address the research questions. Joint with Emmanuel Lesafre (KU Leuven) and Víctor Espejo (Subsecretaría de Pesca y Acuicultura, Valparaío, Chile).

6. Dimas Soares Lima, Federal University of Rio de Janeiro

   Title: *Flexible collaborative filtering: a Bayesian approach*

   Abstract: Recommendation systems seek to predict the “rating” or “preference” that a user would give to an item. We propose a new method for collaborative filtering that allows flexible recommendations to users through Markov Chain Monte Carlo algorithms. With this approach, one can draw from the predictive posterior distribution and use it to produce point estimators. After convergence, each sampled value can be used as a recommendation. Our approach allows fast results to be produced since it doesn?t require waiting of the simulation of a full chain before making predictions. This is not only welcome by the users themselves but also help with the learning mechanism of the algorithm. Also, one of the biggest concerns of this study was to create an algorithm that is scalable. To do so, we suggest optimization steps to bypass matrix inversion. That way, even under high dimensional scenarios, the computational time of our algorithm is satisfactory. Finally, an application to the Movie Lens data set Harper and Konstan (2016) is presented as an illustration. Joint with Marina Paez (UFRJ) and Hugo Carvalho (UFRJ).

7. Douglas Azevedo, Federal University of Minas Gerais

   Title: *Projection-based approach to alleviate spatial confounding in spatial frailty models*

   Abstract: Spatial Confounding is the name given to the confounding between fixed and random effects in Generalized Linear Mixed Models. It has been widely studied and it gained attention in the past years in spatial literature, as it may generates unexpected results in modeling. The projection-based approach (also known as restricted models) appears as a good way to work around the Spatial Confounding in such kind of models. However, when the support of fixed effects is different of the spatial effect one, this approach can no longer be applied directly. Spatial Frailty models are able to incorporate spatially structured effects and it is common to observe more than one sample unit per area what means that the support of fixed and spatial effects differ. In this work we introduce a projection-based approach for Spatial Frailty models where the support of fixed and spatial effects (areal data) do not match. To provide a fast inference for the parameters we used the Integrated Nested Laplace Approximation (INLA) methodology. The Restricted Spatial Frailty model is applied to modelling cases of Lung and Bronchus cancer in California state and the results prove the methodology efficiency. Joint with Marcos Prates (UFMG) and D. Bandyopadhyay (Virginia Commonwealth University).

8. Eriton dos Santos, Instituto de Psiquiatria do Hospital das Clínicas de São Paulo

   Title: *Using reversible jump for graphical modeling*

   Abstract: The random graph model is used to model the relationship between random variables using graph theory. The inference of parameters of this type of modeling is commonly based on maximum likelihood estimation. However, this type of methodology requires the adjustment of all possible models to verify which model best represents the relationship between the variables. In the case of any model, among all possible, presents estimation problem, the result may not represents the true relationship between the variables. We propose a procedure based on Reversible Jump and Metropolis-Hastings for model selection and model fitting of a graphical model. The results found are satisfactory, but they can be improved. Joint with Luis Milan (UFSCar).

9. Felipe Polo, University of São Paulo

   Title: *Skills to not fall behind in school*

   Abstract: Many recent studies emphasize how important the role of cognitive and social-emotional skills can be in determining people's quality of life. Although skills are of great importance in many aspects, in this paper we will focus our efforts to better understand the relationship between several types of skills with academic progress delay. Our dataset contains the same students in 2012 and 2017, and we consider that there was an academic progress delay for a specific student if he or she progressed less than expected in school grades. Our methodology primarily includes the use of a Bayesian logistic regression model and our results suggest that both cognitive and social-emotional skills may impact the conditional probability of falling behind in school, and the magnitude of the impact between the two types of skills can be comparable.

10. Gabriela Massoni, Federal University of São Carlos

    Title: *The impact of servant unification on the effectiveness of Brazilian judicial management*

    Abstract: In order to obtain legal celerity, it is crucial to have effective Court administration. This article compares the traditional model of legal servant management and the unification model, where legal servants are subject to all judges in the Court and managed by a central administrator. In order to compare the models of legal staff administration it was necessary to estimate how much time was used so far by the legal staff in each lawsuit. For this, we calculate the posterior probability that the current procedure is being performed by a legal servant using hidden Markov models. Since the total servant time per suit is censored for active cases, we use a proportional hazards model to compare the types of legal staff administration. We use the suit's subject, the type of legal staff administration and process distribution date covariates and estimate the parameters using LASSO. We have come to the conclusion that the staff unification contributes to the decreasing of approximately 3 hours in the resolution of each lawsuit, the modernization contributes to decreasing of approximately 3 hours/year in the resolution of each process, and some processes have spent a long time without being effectively worked. Joint with Rafael Stern (UFSCar) and Julio Trecenti (Associação Brasileira de Jurimetria).

11. José Ordoñeza, State University of Campinas
Title: *Objective Bayesian analysis for the spatial Student-t regression model*

Abstract: The choice of the prior distribution is a key aspect for Bayesian analysis. In the spatial regression setting a subjective prior choice for the parameters may not be trivial, from this perspective, using the objective Bayesian analysis framework a reference and two Jeffreys priors are introduced for the Spatial Student-t regression model with unknown degrees of freedom. The spatial Student-t regression model poses two main challenges when eliciting priors: one for the spatial dependence parameter and the other one for the degrees of freedom. It is well-known that the propriety of the posterior distribution over objective priors is not always guaranteed, whereas the use of proper prior distributions may dominate and bias the posterior analysis. In this paper, we show the conditions under which our proposed reference prior, and the two introduced Jeffreys priors yield to a proper posterior distribution. Simulation studies are used in order evaluate the performance of the reference prior over the two Jeffereys priors and a non-informative proper prior. Joint with Victor Lachos (University of Connecticut) and Marcos Prates (UFMG).

12. José Sartori Jr., University of São Paulo (cancelled)
Title: *Identifiability analysis of dynamic models using data cloning*

Abstract: Unidentifiable statistical models may make impossible unique inferential conclusions. Therefore, finding parametric constraints that ensure identifiability is of utmost importance in Statistics. In this work, we investigate the use of Data Cloning, a modern technique for classical inference in latent variable models with roots in Bayesian computational methods, as a tool for assessing identifiability of nonlinear Gaussian dynamic models. Joint with Márcia Branco (USP).

13. Juliana Silva, Federal University of Minas Gerais
Title: *Flexible joint model for longitudinal and survival data using Bernstein polynomials*

Abstract: The main topic of the present work is to approximate the hazard function, which is a part of the survival sub-model within the joint modelling approach, will be done through the Bernstein Polynomials (BP) (Bernstein, 1912). The BP were developed by Sergei Natanovich Bernstein. Using polynomials to approximate a function encompasses analytic advantages, once they are easily written in the form of summation. Thus, as highlighted by Osman and Ghosh (2012), calculations such as derivatives and gradients are much easier to obtain trough this structure. Regarding the BP, particularly, they stand out by presenting the best shape preserving considering all polynomial approximation, as stated by Carnicer and Peña (1993) and Osman and Ghosh (2012). The goals of this work focus on jointly model longitudinal and survival data using Bernstein Polynomial to approximate the unknown baseline hazard function. Joint with Vinícius Diniz Mayrink (UFMG), Fábio Demarqui (UFMG) and Sujit Ghosh (NC State University).

14. Leon Zatti, Federal University of Minas Gerais

    Title: *Factor analysis and structural equation modeling: a comparison study of different methods*

    Abstract: The study developed here was part of a research project supported by the energy distribution company CEMIG with activities in the state of Minas Gerais - Brazil. Factor analysis (FA) and structural equation modeling (SEM) are the main methods being applied. The work is basically focused on a simulation study comparing different approaches (including the Bayesian point of view) to fit these models. The software R1 is used for all analyses. The Bayesian SEM takes advantage of the software STAN, which applies the Hamiltonian Monte Carlo MCMC method to sample from unknown and complex joint posterior distributions. Some interesting results were obtained in this study. The idea behind this work is to use FA in order to create latent factors (new variables) which represent a general concept using the observed variables, for more on FA read Johnson and Wichern (2007). Then, the SEM comes doing linear regressions among these created latent variables, in order to understand the direct and indirect effects between the factors, for more on linear regression see Montgomery, Peck and Vining (2012). Lastly, the basis of the Bayesian methods utilized can be found in Gelman et al (2013) and Gamerman and Lopes (2006). Joint with Vinícius Mayrink (UFMG).

15. Mariana Sá, Federal University of Rio de Janeiro

    Title: *Hierarchical Dirichlet regression model to benthic cover Abrolhos Bank*

    Abstract: In this work, we developed an approach to modeling the benthic coral-reef dynamics. The Abrolhos Bank is a 46,000 km2 extension of the continental shelf off Bahia and Espírito Santo states, eastern Brazil. Reef structures are built by corals, macroalgae, turf, bryozoans, sponge, fire coral, CCA, cyano, and others. Thus we have compositional data with nine categories (corals, macroalgae, turf, bryozoans, sponge, fire coral, CCA, cyano, and others), summing to 1 for any given observation. Based on this, we propose the Dirichlet regression model including hierarchical effects by sites for benthic cover. In addition, we use our model to estimate whether environmental conditions impact on each component. Joint with Jéssica Areias (UFRJ), Larissa Martins (UFRJ) and Pamela Solano (UFRJ).

16. Milton Ferreira, Federal University of Minas Gerais

    Title: *Spatio-temporal factor model with non-linear interactions for cluster analysis*

    Abstract: Neste estudo, desenvolvemos dois modelos lineares generalizados com efeito aleatório estruturado pela modelagem fatorial para explorar dados coletados no espaço e no tempo. O objetivo principal é incorporar ao modelo fatorial uma estrutura de interações não lineares proposta em Mayrink e Lucas (2013). Inspirado nos trabalhos de Lopes et al. (2011), a dependência espacial entre regiões e estabelecida através das colunas da matriz de cargas, para a qual assumimos a configuração do modelo autoregressivo condicional (CAR, Besag, 1974). A dependência temporal é considerada associação das colunas da matriz de escores dos fatores, também inserida por meio do modelo CAR. A presença de interações não lineares visa melhorar a explicação das complexas interrelações entre as regi oes do espaçp e, também, permitir a detecção de conglomerados (clusters). Novos tipos de clusters podem surgir como uma combinação dos efeitos principais dos fatores e do efeito da interação. Em termos de aplicação real, este estudo é motivado pela análise do desfecho (morte ou não) de pacientes de municípios do estado de Minas Gerais que sofreram infarto agudo do miocárdio (IAM). Joint with Vinícius Mayrink (UFMG) and Antônio Ribeiro (UFMG).

17. Montcho Djidenou, Federal University of São Carlos

    Title: *Using data driven reversible jump for modeling schizophrenia data*

    Abstract: Symptom based diagnosis are known to be limited specially concerning complex disorders such as schizophrenia. Modern attempts in providing predictive risk for such disease, to assist existing diagnosis tools, integrate genetic and brain information. In this paper, our goal is both inferential and predictive. Given functional Magnetic Resonance Imaging(fMRI) and genetic information on Single Nucleotide Polymorphisms (SNPs) from healthy and people diagnosed with schizophrenia, we aim at selecting the most discriminatory ROIs and SNPs in a joint framework. For this purpose, a data driven reversible jump is proposed to efficiently handle model selection and parameters estimates notably the additive and dominant effect of selected SNPs. We also compare our model to existing methodologies. Joint with Daiane Zuanetti (UFSCar) and Luis Milan (UFSCar).

18. Nathalie Deziderio, Federal University of Rio de Janeiro

    Title: *A variational inference approach to music emotion recognition*

    Abstract: How can we identify which emotions are evoked by a piece of music? Within the Music Information Retrieval (MIR) community this problem is called "Music Emotion Recognition", a very challenging and interesting one. Several MIR researchers are trying to obtain advances in this area, but the current approaches fail to give an interesting and informative statistical contextualization to the employed methods. Here, we attempt to stretch that boundary by using Statistical Learning and modern Bayesian techniques to retrieve more from data and create a model that shows better results when compared with similar techniques. More specifcally, we will present two models, using Bayesian Lasso and Automatic Relevance Determination (ARD) respectively, to introduce sparsity and estimate the coefficients of a linear regression model explaining arousal and valence, a widely employed scale to measure emotion content, from acoustical features extracted from audio signals. Variational Inference will be applied to our ARD model in order to reduce computational costs. In both approaches, samples will be drawn from their respective distributions and used to obtain the predictive distribution for arousal and valence for a given song. Joint with Hugo Carvalho (UFRJ).

19. Nicholas Eugenio, University of São Paulo

    Title: *Bayesian inference in stochastic process to identify mortality attributed to a specific factor*

    Abstract: In this work, we present the progress of developing bayesian approach in stochastic process to estimate the population attributable fraction (PAF) of a specific factor's mortality at Brazilian medical centers. That is, the fraction of the total deaths in the inpatients population that would not have occurred if the effect associated with this factor were absent. Most popular procedures to calculate PAF takes into account incidences, prevalences, mortalities and morbidities (sample summary measures for a specified time interval), but do not consider individual evolutions'. Our proposal is to incorporate daily and personal data in PAF's calculation to understand factor possible processes and outcomes. For that, it is in course a case-control study of inpatients' observational records, where 30 medical centers are participating and, for each one, 100 inpatients are observed up to a maximum of 28 days: 50 for case group (death) and 50 for control group (hospital discharge). The aim of the study is to identify what are the differences between the two groups with regard to factors' presence or absence in PAF of factors' mortality. We consider, until now, homogeneity for the transition probabilities matrix and absorbing Markov Chain theory. Joint with Victor Fossaluza (USP).

20. Ricardo Pedroso, Federal University of Minas Gerais

    Title: *A Bayesian multiparametric change point model*

    Abstract: We propose an extension of the Product Partition Model to detect structural changes in sequences of observations where groups of structural parameters are subject to multiple changes at different times, with both the number and locations of the change points unknown. Additionally to the original goals in the use of product partition models in change point problems, that is, the estimation of the random partition and its block parameters, we want to identify in which parameter the change occurred. We develop a partially collapsed Gibbs scheme to sample from partitions and parameters in the case of an univariate sequence of conditional independent Normal observations subject to regime changes in the mean and variance. We consider two random partitions, one for the mean and other to the variance, such that it is possible to identify the regime changes and the specific parameter where the changes occurred. The scheme proposed may be extended to other observational models with two or more parameters, with one partition for each parameter.

21. Ricardo Souza, Fluminense Federal University

    Title: *Evaluating spatial models for disease data: a study about tuberculosis*

    Abstract: When modeling spatial data it is necessary to address the inherent spatial dependence. A frequent solution to deal with this problem is the use of hierarchical models, under a Bayesian framework, with prior distributions that can account for the spatial autocorrelation present in the data. The inventory of this class of prior distributions is rich, with many popular choices being well documented in the literature. The present work proposes a collating of some of these prior distributions, comparing the Intrinsic CAR, BYM, Leroux among other models. The proposed methodology was then applied to a tuberculosis notifications in the state of Rio de Janeiro, considering various socioeconomic covariates and different prior distributions to spatial effects. The performance of these models was evaluated by using the Deviance Information Criterion (DIC) and the Watanabe-Akaike Information Criterion(WAIC). Joint with Jony Pinto Jr (UFF) and Patricia Silva (UFU).

22. Rodrigo Bulhões, Federal University of Bahia
Title: *Anisotropy in spatiotemporal models with multiple responses*

Abstract: In this work we propose a multivariate model to fit two or more response variables, measured at different points in time and distinct locations in space. Here we relax the hypothesis of isotropy in space because this assumption is not reasonable in many real situations. We treat the anisotropy of the spatial process using the concept of spatial deformation, in a Bayesian framework. This development is motivated by an application to an environmental problem, where the aim is to propose a joint space-time model for three different air pollutants (PM10, PM2:5 and total amount of particles in suspension (TSP)), measured between 2001 and 2015 in some monitoring stations located in the Metropolitan region of Rio de Janeiro, Brazil. Joint with Marina Paez (UFRJ).

23. Sergio Britto, Federal University of Rio de Janeiro
Title: *Incorporating temporal dependence in survival models through Bayesian nonparametric autoregression*

Abstract: Semiparametric survival models, such as the Cox proportional-hazards and the accelerated failure time model (AFT), split the model into a parametric and a nonparametric part. The nonparametric part, which can be represented by the baseline hazard, cumulative hazard, or survival function, is modeled as flexibly as possible, but it is usually considered to be static even in applications where the survival chances of the population under study change over time Particularly, the AFT is a log-linear model with errors that follow a nonparametric distribution. In this work we propose extending this model by allowing the distributionof the errors to vary smoothly over time according to the Bayesian nonparametric autoregressive models proposed by Di Lucca et al. (2013). This specification is based on a dependent Dirichlet process prior on a family of random probability measures indexed by lagged covariates. Joint with Marina Paez (UFRJ) and Mariane Alves (UFRJ).

24. Victor Freguglia, Federal University of Minas Gerais
Title: *Bayesian selection of variable length Markov chain models via reversible jump*

Abstract: Inference for High-order Markov chains is a difficult problem, even on finite state-spaces, as the number of parameters grows exponentially as the order of dependence increases. Variable Length Markov Chain (VLMC) models reduce the parameter dimension by performing a tree-based clustering for conditional probabilities. The original algorithm for model (tree) selection relies on multiple likelihood-ratio tests with levels properly chosen for good asymptotic properties. We propose a Bayesian approach for tree selection using a hyerarchical model with random trees and a Reversible Jump Monte-Carlo Markov Chain to sample from models with different trees, and consequently, different parameter dimension given an observed chain.

25. Victor Silva, Federal University of Rio de Janeiro

    Title: *Bayesian factor models for multivariate categorical data obtained from questionnaires*

    Abstract: Factor analysis is a flexible technique for assessment of multivariate dependence and codependence. Besides being an exploratory tool used to reduce the dimensionality of multivariate data, it allows estimation of common factors that often have an interesting theoretical interpretation in real problems. However, standard factor analysis is only applicable when the variables are scaled, which is often inappropriate, for example, in data obtained from questionnaires in the field of psychology, where the variables are often categorical. In this framework, we propose a factor model for the analysis of multivariate ordered and non-ordered polychotomous data. The inference procedure is done under the Bayesian approach via Markov chain Monte Carlo methods. Two Monte-Carlo simulation studies are presented to investigate the performance of this approach in terms of estimation bias, precision and assessment of the number of factors. We also illustrate the proposed method to analyze participants' responses to the Motivational State Questionnaire dataset, developed to study emotions in laboratory and field settings. Joint with Kelly Gonçalves (UFRJ) and João Pereira (UFRJ).

26. Pedro Ramos, University of São Paulo at São Carlos

    Title: *Objective Bayesian inference in power law distributions*

    Abstract: The power-law distribution plays an important role in complex networks as well as different applied sciences. This distribution is usually used to describe the distribution of degrees related to complex networks. The parameter estimation is usually conducted under the classical approach using the maximum likelihood estimators. On the other hand, the obtained estimator for the parameter is biased for small samples which may lead to biased conclusions compromising the decision to be made. For instance, different values for the parameters lead to different classification as a scale-free distribution which is an important concept in a complex network. In this work, we consider the Bayesian approach to perform inference in the model parameter. To achieve that we considered the use of an objective reference prior that leads to efficient estimator in terms of Bias and mean square error. Moreover, the obtained posterior is invariance under one-to-one transformations, consistent under marginalization and with good coverage probabilities. A simulation study was conducted using the R statistical software to check the efficiency of our proposed approach. The theoretical results are confirmed where it was verified that the proposed method returns almost non-biased estimates with smaller mean square errors, even for small samples being more accurate than those obtained by the maximum likelihood method. These results have not been discussed earlier and can have an important impact on many areas of complex networks where the parameter to be estimated plays is important to check whether or not a network is scaled-free.

27. Nivea Bispo, Federal University of Bahia

Title: *Robust Bayesian modeling using finite mixtures of Student t distributions*

Abstract: Statistical modeling based on finite mixtures of distributions is a research area in growing development and recently have been used to model the errors distribution in univariate and multivariate linear regression models. This work consider a methodology based on finite mixtures of Student-t distributions to model the errors distribution in linear regression models. The proposed approach contemplates a finite mixture in two levels considering in its specification separate structures for multimodality/skewness and tail behavior modeling, estimating the tail structure of the model without estimating degree of freedom parameters. The inference is performed via Markov chain Monte Carlo and simulation studies are conducted for to evaluate the performance of the proposed approach. Results from the analysis of a real data set are too presented. Joint with Marcos Prates (UFMG) and Flávio Gonçalves (UFMG).