



EBEB 2014

Program and Abstract

12th Brazilian Meeting on Bayesian Statistics

**March 10-14, 2014
Atibaia – SP – Brazil**

PREFACE

The Brazilian Meeting on Bayesian Statistics (EBEB) is in its twelfth edition. This series of meetings aims at strengthening the research on Bayesian methods and widening their application. It also provides an environment where Brazilian and international researchers collaborate, present their most recent developments and discuss on open problems. EBEB also allows graduate students to make contacts with experienced researchers. This year's meeting has a particular focus on discussing recent developments in the many viewpoints of Bayesian statistics, such as computational, theoretical, methodological and applied views.

EBEB 2014 takes place in Atibaia, State of Sao Paulo – Brazil, from March 10 to 14, in a beautiful country-side hotel (<http://www.hotelfazendaatibaia.com.br/>) specialized in horse riding.

This document contains program and abstracts of the contributions to EBEB. Accepted full papers appear in the proceedings of the event, which is published by Springer in a book from the series “Proceedings in Mathematics & Statistics” (<http://www.springer.com/series/10533>).

EBEB has been organized by ISBrA – the Brazilian chapter of ISBA – since its 6th meeting in 2002, in the occasion of the First Latin American Bayesian Congress (Cobal I).

ORGANIZATION

Conference Committee Chair:

- Polpo, Adriano (UFSCar / Brazil)

Organizing Committee:

- Dias, Teresa Cristina Martins (UFSCar / Brazil)
- Lauretto, Marcelo (EACH-USP / Brazil)
- Louzada Neto, Francisco (ICMC-USP / Brazil)
- Pereira, Carlos Alberto de Bragana (IME-USP / Brazil)
- Polpo, Adriano (UFSCar / Brazil)
- Rifo, Laura Leticia Ramos (IMECC-UNICAMP / Brazil)
- Stern, Julio (IME-USP / Brazil)

Scientific Committee:

- Branco, Marcia D'Elia (IME-USP / Brazil)
- de Campos, Cassio (IDSIA / Switzerland)
- Diniz, Carlos A. R. (UFSCar / Brazil)
- Diniz, Marcio Alves (UFSCar / Brazil)
- Fossaluzza, Victor (IME-USP / Brazil)
- Gamerman, Dani (UFRJ / Brazil)
- Lauretto, Marcelo (EACH-USP / Brazil)
- Loschi, Rosangela H. (UFMG / Brazil)
- Louzada Neto, Francisco (ICMC-USP / Brazil)
- Pereira, Carlos Alberto de Bragana (IME-USP / Brazil)
- Polpo, Adriano (UFSCar / Brazil)
- Rifo, Laura Leticia Ramos (IMECC-UNICAMP / Brazil)
- Rogatko, Andre (Cedars-Sinai / USA)
- Schmidt, Alexandra M. (UFRJ / Brazil)
- Sinha, Debajyoti (FSU / USA)
- Stern, Julio (IME-USP / Brazil)

Executive Committee:

- Netto, Lourdes Vaz da Silva (IME-USP / Brazil)
- Takahashi, Sylvia Regina A. (IME-USP / Brazil)
- Bruno Borcado (Supremum Assessoria / Brazil)

Contact:

- isbra-eb@ime.usp.br

GUESTS

Speakers:

- Chakrabarty, Dalia
University of Leicester and University of Warwick – UK
- de Cooman, Gert
Ghent University – Belgium
- Dey, Dipak
University of Connecticut – USA
- Garcia, Nancy
Campinas State University – Brazil
- Lopes, Hedibert
INSPER – Brazil
- Rodrigues, Josemar
University of Sao Paulo – Brazil
- Ruggeri, Fabrizio
Consiglio Nazionale delle Ricerche – Italy
- Rogatko, Andre
Samuel Oschin Comprehensive Cancer Institute – USA
- Silva, Ricardo
University College London – UK
- Sinha, Debajyoti
Florida State University – USA

Session organizers:

- Branco, Marcia D'Elia
University of Sao Paulo – Brazil
- de Campos, Cassio P.
Dalle Molle Institute for Artificial Intelligence – Switzerland
- Gamerman, Dani
Federal University of Rio de Janeiro – Brazil
- Loschi, Rosangela H.
Federal University of Minas Gerais – Brazil
- Rogatko, Andre
Samuel Oschin Comprehensive Cancer Institute – USA
- Schmidt, Alexandra M.
Federal University of Rio de Janeiro – Brazil
- Sinha, Debajyoti
Florida State University – USA
- Stern, Julio M.
University of Sao Paulo – Brazil

EBEB 2014 – Program

	Monday - 10/03	Tuesday - 11/03	Wednesday - 12/03	Thursday - 13/03	Friday - 14/03					
07:00 – 9:00		Breakfast	Breakfast	Breakfast	Breakfast					
09:00 – 10:14	Registration	Oral 1A <i>*C.Campos</i>	Oral 1B <i>*V.Fossaluzza</i>	Oral 3A <i>*D.Gamerman</i>	Oral 3B <i>*L.Rifo</i>	Oral 4A <i>*A.Rogatko</i>	Oral 4B <i>*C.Diniz</i>	Oral 6A <i>*A.Schmidt</i>	Oral 6B <i>*M.Diniz</i>	
10:15 – 10:34		Coffe break		Coffe break		Coffe break		Coffe break		
10:35 – 11:04		ST1: C.Campos		ST3: D.Gamerman		ST4: D.Chakrabarty		ST6: A.Schmidt		
11:05 – 11:59		T1: G.Cooman <i>*C.Campos</i>		T3: D.Sinha <i>*D.Gamerman</i>		T4: A.Rogatko <i>*C.Pereira</i>		T6: F.Ruggeri <i>*A.Schmidt</i>		
11:30 – 11:59								Closure		
12:00 – 13:59	Lunch	Lunch	Lunch	Lunch	Lunch					
14:00 – 14:29	Opening	ST2: R.Loschi		Reunião ISBrA	ST5: Ricardo Silva					
14:30 – 14:59		T2: H.Lopes <i>*R.Loschi</i>			T5.1: J.Rodrigues <i>*M.Branco</i>					
15:00 – 15:24		T0: D.Dey <i>*D.Sinha</i>			Coffe break		Coffe break			
15:25 – 15:44		ST0: J.Stern			Oral 2A <i>*R.Loschi</i>		Oral 2B <i>*D.Sinha</i>		Oral 5A <i>*M.Branco</i>	
15:45 – 15:54							Oral 5B <i>*J.Stern</i>			
15:55 – 16:24							T5.2: N.Garcia <i>*J.Stern</i>			
16:25 – 16:59	Welcome Cocktail	Poster Session								
17:00 – 17:59										
18:00 – 18:59										
19:00 – 19:29										
19:30 – 22:00	Dinner	Dinner	Dinner	Banquet						

* Session Chair

Table 1: Contributions

Chair	Calendar	Session	Type	Authors	Title	Abstract
D.Sinha	10/Mar Mon 15:00 - 15:50	T 0	Talk	<i>Dipak Dey</i> , Gyuhyeong Goh	Bregman divergence measures in Bayesian modeling with applications	p. 18
C.Campos	11/Mar Tue 11:05 - 11:55	T 1	Talk	<i>Gert de Cooman</i> , Jasper De Bock, Marcio Diniz	Predictive inference under exchangeability, and the imprecise Dirichlet multinomial model	p. 19
R.Loschi	11/Mar Tue 14:30 - 15:20	T 2	Talk	<i>Hedibert Lopes</i>	Bayesian instrumental variables: priors and likelihoods	p. 20
D.Gamerman	12/Mar Wed 11:05 - 11:55	T 3	Talk	<i>Debajyoti Sinha</i>	Bayesian partial linear model for skewed longitudinal data	p. 21
4 C.Pereira	13/Mar Thu 11:05 - 11:55	T 4	Talk	<i>Andre Rogatko</i> , Galen Cook-Wiens, Mourad Tighiouart, Steven Piantadosi	FDA and innovative designs: case study of a missed opportunity	p. 22
M.Branco	13/Mar Thu 14:30 - 15:20	T 5.1	Talk	<i>Josemar Rodrigues</i>	An extended exponentiated-g-negative binomial family with threshold effect	p. 23
J.Stern	13/Mar Thu 17:00 - 17:50	T 5.2	Talk	<i>Nancy Garcia</i>	Aggregated functional data model for near-infrared spectroscopy: calibration and prediction	p. 24
A.Schmidt	14/Mar Fri 11:05 - 11:55	T 6	Talk	Jose Pablo Arias-Niolas, Jacinto R. Martin Jimenez, <i>Fabrizio Ruggeri</i> , Alfonso Suarez Llorens	Nondominated and least sensitive actions in modeling gas pipelines replacement	p. 25

Continues on the next page

Table 1 – *Contributions*

Chair	Calendar	Session	Type	Authors	Title	Abstract
C.Pereira	10/Mar Mon 15:55 - 16:20	ST 0	Short Talk	<i>Julio Stern</i> , Fabio Nakano	Affinity structure in reaction networks: minimum entropy or power dissipation	p. 26
G.Cooman	11/Mar Tue 10:35 - 11:00	ST 1	Short Talk	Giorgio Corani, <i>Cassio de Campos</i>	A maximum entropy approach to learn Bayesian networks from incomplete data	p. 27
H.Lopes	11/Mar Tue 14:00 - 14:25	ST 2	Short Talk	<i>Rosangela H. Loschi</i> , Caroline C. Vieira, Denise Duarte	Nonparametric mixtures based on skew-normal distributions: an application to density estimation	p. 28
D.Sinha	12/Mar Wed 10:35 - 11:00	ST 3	Short Talk	<i>Dani Gamerman</i> , Thiago R. Santos, Glauro C. Franco	Modeling reliability data using non-Gaussian state space models	p. 29
C.Pereira	13/Mar Thu 10:35 - 11:00	ST 4	Short Talk	<i>Dalia Chakrabarty</i>	Bayesian learning of density function by multiple sequential inversions of 2-D images in electron microscopy	p. 30
C.Campos	13/Mar Thu 14:00 - 14:25	ST 5	Short Talk	<i>Ricardo Silva</i>	Bayesian inference in cumulative distribution fields	p. 31
F.Ruggeri	14/Mar Fri 10:35 - 11:00	ST 6	Short Talk	<i>Alexandra Schmidt</i> , Patricia da Costa	A skew-normal, nonstationary spatio-temporal process	p. 32

Continues on the next page

Table 1 – *Contributions*

Chair	Calendar	Session	Type	Authors	Title	Abstract
C.Campos	11/Mar Tue 09:00 - 09:20	O 1A.1	Oral	Hunter Glanz, Luis Carvalho	A spanning tree hierarchical model for land cover classification	p. 33
C.Campos	11/Mar Tue 09:25 - 09:45	O 1A.2	Oral	Carlos A. de B. Pereira, Rafael Stern	Identification of differentially expressed genes in RNA-Seq platform: a Bayesian multivariate logistic-normal model	p. 34
C.Campos	11/Mar Tue 09:50 - 10:10	O 1A.3	Oral	Salimeh Yasaei Sekeh	Bayesian weighted information measures	p. 35
V.Fossaluza	11/Mar Tue 09:00 - 09:20	O 1B.1	Oral	Glaura Franco, Helio Migon, Marcos Prates	Poisson models for time series of counts: a comparison between observation and parameter driven models	p. 36
V.Fossaluza	11/Mar Tue 09:25 - 09:45	O 1B.2	Oral	Cristian Bayes, Jorge Luis B. Guzman, Luis Valdivieso	A binary regression mixed model using a skew-probit link	p. 37
V.Fossaluza	11/Mar Tue 09:50 - 10:10	O 1B.3	Oral	Jony Pinto Junior, Dani Gamerman, Marina Paez	Point process with spatio-temporal heterogeneity	p. 38

Continues on the next page

Table 1 – *Contributions*

Chair	Calendar	Session	Type	Authors	Title	Abstract	
R.Loschi	11/Mar Tue 15:45 - 16:05	O 2A.1	Oral	Victor Hugo Lachos, Aldo Garay, Heleno Bolfarine, Celso Romulo Barbosa Cabral	Bayesian analysis of censored linear regression models with scale mixtures of normal distributions	p. 39	
R.Loschi	11/Mar Tue 16:10 - 16:30	O 2A.2	Oral	Lijun Peng, Luis Carvalho	Bayesian ridge-regularized covariance selection with community behavior in latent Gaussian graphical Models	p. 40	
R.Loschi	11/Mar Tue 16:35 - 16:55	O 2A.3	Oral	Jhon Bernedo Gonzales, Vera Tomazella, Mario de Castro	Bayesian partitioning for survival data with a cure fraction	p. 41	
↵	D.Sinha	11/Mar Tue 15:45 - 16:05	O 2B.1	Oral	Vinicius Calsavara, Agatha Rodrigues	A flexible cure rate model with frailty term in latent risks: a Bayesian approach	p. 42
	D.Sinha	11/Mar Tue 16:10 - 16:30	O 2B.2	Oral	Leonardo Cella	Bayesian ordinal regression	p. 43
	D.Sinha	11/Mar Tue 16:35 - 16:55	O 2B.3	Oral	Mauricio Castro, Victor Lachos, Diana Galvis, Dipankar Bandyopadhyay	Bayesian semiparametric longitudinal data modeling using normal/independent densities	p. 44

Continues on the next page

Table 1 – *Contributions*

Chair	Calendar	Session	Type	Authors	Title	Abstract
D.Gamerman	12/Mar Wed 09:00 - 09:20	O 3A.1	Oral	Marcio Diniz, Jasper De Bock, Arthur Van Camp	Characterising Dirichlet priors	p. 45
D.Gamerman	12/Mar Wed 09:25 - 09:45	O 3A.2	Oral	Ian Johnston, Yang Jin, Luis Carvalho	Assessing a spatial boost model for quantitative trait GWAS	p. 46
D.Gamerman	12/Mar Wed 09:50 - 10:10	O 3A.3	Oral	Vinicius Diniz Mayrink, Joseph E. Lucas	Sparse latent factor models with interactions	p. 47
L.Rifo	12/Mar Wed 09:00 - 09:20	O 3B.1	Oral	Willian Oliveira, Carlos Diniz, Marinho Andrade	Bayesian variable selection in a class of zero inflated modified power series models	p. 48
∞ L.Rifo	12/Mar Wed 09:25 - 09:45	O 3B.2	Oral	Brian Melo, Luis Esteves	Hypotheses testing in finite populations	p. 49
L.Rifo	12/Mar Wed 09:50 - 10:10	O 3B.3	Oral	Natalia Oliveira, Marcio Diniz, Adriano Polpo	The homogeneity test for 2x2 contingency table	p. 50

Continues on the next page

Table 1 – *Contributions*

Chair	Calendar	Session	Type	Authors	Title	Abstract
A.Rogatko	13/Mar Thu 09:00 - 09:20	O 4A.1	Oral	Joao Daniel Nunes Duarte, Vinicius Diniz Mayrink	Factor analysis with mixture modeling to evaluate coherent patterns in mi- croarray data	p. 51
A.Rogatko	13/Mar Thu 09:25 - 09:45	O 4A.2	Oral	Victor Fossaluza, Marcelo Lauretto, Carlos Alberto de Braganca Pereira, Julio Stern	Combining optimization and random- ization approaches for the design of clinical trials	p. 52
A.Rogatko	13/Mar Thu 09:50 - 10:10	O 4A.3	Oral	Danilo Lopes, Danielle Ro- ges, Adriano Polpo	Implementation of censored models for survival analysis	p. 53
C.Diniz	13/Mar Thu 09:00 - 09:20	O 4B.1	Oral	Danilo Covaes Nogarott, Caio Lucidius N. Azevedo	Bayesian inference in beta and inflated beta regression models	p. 54
C.Diniz	13/Mar Thu 09:25 - 09:45	O 4B.2	Oral	Francisco Louzada, Adriano Suzuki, Luis Salasar, An- derson Ara, Jose Leite	A Bayesian approach to predicting football match outcomes considering time effect weight	p. 55
C.Diniz	13/Mar Thu 09:50 - 10:10	O 4B.3	Oral	Catalina Araya, Francisco Torres Aviles	Bayesian analysis of gamma regression models: state of the art and extensions	p. 56

Continues on the next page

Table 1 – *Contributions*

Chair	Calendar	Session	Type	Authors	Title	Abstract
M.Branco	13/Mar Thu 15:45 - 16:05	O 5A.1	Oral	Marcio Augusto Diniz, Carlos Alberto de Braganca Pereira, Adriano Polpo	Bayesian semi-parametric symmetric models for binary data	p. 57
M.Branco	13/Mar Thu 16:10 - 16:30	O 5A.2	Oral	Luca Martino, Victor Elvira, David Luengo, Jukka Corander	MCMC-driven adaptive multiple importance sampling	p. 58
M.Branco	13/Mar Thu 16:35 - 16:55	O 5A.3	Oral	Luis Valdivieso, Cristian Bayes	A zero variance differential geometric MCMC estimation approach for a beta 0-inflated mean regression model	p. 59
J.Stern	13/Mar Thu 15:45 - 16:05	O 5B.1	Oral	Plinio Andrade, Laura Rifo	A note on Bayesian inference for the long-range dependence of a stationary two-state process	p. 60
J.Stern	13/Mar Thu 16:10 - 16:30	O 5B.2	Oral	Fekadu Lemessa Bayisa	Spatial modelling of Malaria risk in Bayesian setting: a case study in Adama district of Oromiya regional state, Ethiopia	p. 61

Table 1 – *Contributions*

Chair	Calendar	Session	Type	Authors	Title	Abstract
A.Schmidt	14/Mar Fri 09:25 - 09:45	O 6A.2	Oral	Zaida Quiroz, Marcos Prates, Havard Rue	A Bayesian approach to estimate the biomass of anchovies in the coast of Peru	p. 62
A.Schmidt	14/Mar Fri 09:50 - 10:10	O 6A.3	Oral	Luiz Carvalho, Claudio Struchiner, Bastos, Leonardo	Bayesian inference of deterministic population growth models	p. 63
M.Diniz	14/Mar Fri 09:25 - 09:45	O 6B.2	Oral	Khader Khadraoui	Nonparametric Bayesian regression under combinations of local shape constraints	p. 64
M.Diniz	14/Mar Fri 09:50 - 10:10	O 6B.3	Oral	Orhan Sonmez, Taylan Cemgil	Annealed stochastic approximation EM for reinforcement learning	p. 65

Table 1 – *Contributions*

Chair	Calendar	Session	Type	Authors	Title	Abstract
#01	11/Mar Tue 17:00 - 19:00	P	Poster	Laura Rifo, Plinio Andrade	Bayesian inferences on the Rosenblatt distribution	p. 66
#02	11/Mar Tue 17:00 - 19:00	P	Poster	Renan X. Cortes, Thiago G. Martins, Marcos O. Prates, Braulio F. A. Silva	Inference on dynamic models for non-Gaussian random fields using INLA	p. 67
#03	11/Mar Tue 17:00 - 19:00	P	Poster	Caio Azevedo, Filidor Vilca-Labra, Cristian Lobos	Bayesian inference for Birnbaum-Saunders mixed models for censored responses	p. 68
#04	11/Mar Tue 17:00 - 19:00	P	Poster	Veronica Amparo Quispe, Vera Tomazella, Luis Salasar, Marcio Viola	Bivariate zero-inflated Poisson distribution: a Bayesian approach	p. 69
#05	11/Mar Tue 17:00 - 19:00	P	Poster	Bruna Turato Santos, Marcelo Lauretto	FBST-CT package: full Bayesian significance test for contingency tables	p. 70
#06	11/Mar Tue 17:00 - 19:00	P	Poster	Patricia Silva, Kamylla Rodrigues Leandro, Gabriela Peres	Bayesian's decision rule of investment	p. 71
#07	11/Mar Tue 17:00 - 19:00	P	Poster	George L. M. Pezzott, Luis Salasar, Jose Leite	Bayesian estimation of population size under closed capture-recapture model with heterogeneity among individuals	p. 72
#08	11/Mar Tue 17:00 - 19:00	P	Poster	Andressa Gigante, Danielle Loureiro Roges, Willian Oliveira	“Optimal” decision about sample sizes considering costs	p. 73

Continues on the next page

Table 1 – *Contributions*

Chair	Calendar	Session	Type	Authors	Title	Abstract
#09	11/Mar Tue 17:00 - 19:00	P	Poster	Eder Milani, Marinho Andrade	Autoregressive moving average processes with generalized normal distribution	p. 74
#10	11/Mar Tue 17:00 - 19:00	P	Poster	Cristian Bayes, Hugo Agurto	Bayesian inference in a semiparametric quantile regression model	p. 75
#11	11/Mar Tue 17:00 - 19:00	P	Poster	Amanda Gazon, Luis Salasar, Adriano Polpo	A Hierarchical Bayesian approach to human gait analysis	p. 76
#12	11/Mar Tue 17:00 - 19:00	P	Poster	Gabriel Hideki Vatanabe Brunello, Genaina Nunes Rodrigues, Eduardo Yoshio Nakano	Bayesian inference on proportional elections	p. 77
#13	11/Mar Tue 17:00 - 19:00	P	Poster	Katherine E. C. Zavaleta, Vicente G. Cancho, Francisco Louzada	A Bayesian detection zero-inflated model	p. 78
#14	11/Mar Tue 17:00 - 19:00	P	Poster	Jonatas Santo, Marcio Viola, Carlos Diniz	Bayesian extreme-value regression model assuming known dispersion parameter	p. 79
#15	11/Mar Tue 17:00 - 19:00	P	Poster	Dipankar Bandyopadhyay, Diana Milena Galvis Soto, Victor Hugo Lachos Davila	Augmented mixed models for clustered proportion data	p. 80

Continues on the next page

Table 1 – *Contributions*

Chair	Calendar	Session	Type	Authors	Title	Abstract
#16	11/Mar Tue 17:00 - 19:00	P	Poster	Yiqi Bao, Vicente Garibay Cancho, Francisco Louzada	On the Bayesian estimation and influence diagnostics for the Weibull-negative-binomial regression model with cure rate under latent failure causes	p. 81
#17	11/Mar Tue 17:00 - 19:00	P	Poster	Jose Augusto Fioruci, Yiqi Bao, Francisco Louzada, Vicente Garibay Cancho	Bayesian inference of the exponential-poisson-logarithmic distribution	p. 82
#18	11/Mar Tue 17:00 - 19:00	P	Poster	Bernardo Tormin, Vinicius Diniz Mayrink	Inuence of the spatial conguration on the efciciency of MCMC methods	p. 83
#19	11/Mar Tue 17:00 - 19:00	P	Poster	Guaraci Requena, Debora Goncalves, Carlos Diniz	Operational risk: tools for dependency modelling	p. 84
#20	11/Mar Tue 17:00 - 19:00	P	Poster	Carlos Montenegro, Marcia Branco	Bayesian inference of the logistic growth model in state space representation	p. 85
#21	11/Mar Tue 17:00 - 19:00	P	Poster	Rosineide Paz, Jorge Luis Bazan, Ricardo S. Ehlers	Modeling the votes of Brazilian political party through a Bayesian statistical model	p. 86
#22	11/Mar Tue 17:00 - 19:00	P	Poster	Jorge Luis Torrejon Matos, Julio Stern	Numerical approximation to Mellin convolution by mixtures of exponentials	p. 87

Continues on the next page

Table 1 – *Contributions*

Chair	Calendar	Session	Type	Authors	Title	Abstract
#23	11/Mar Tue 17:00 - 19:00	P	Poster	Alexandre Calixto, Erlandson Saraiva, Luis Aparecido Milan	A reversible-jump procedure for image analysis	p. 88
#24	11/Mar Tue 17:00 - 19:00	P	Poster	Erlandson Saraiva, Luis Aparecido Milan, Francisco Louzada, Adriano Suzuki	A predictive allocation sampler for analysis of mixture model	p. 89
#25	11/Mar Tue 17:00 - 19:00	P	Poster	Daniele C. T. Granzotto, Francisco Louzada	The transmuted log-logistic distribution: an application for time up to first calving of cows via Bayesian methods	p. 90
#26	11/Mar Tue 17:00 - 19:00	P	Poster	Marcia Aparecida Centanin Macera, Francisco Louzada	The exponential-Poisson regression model for recurrent events: a Bayesian approach	p. 91
#27	11/Mar Tue 17:00 - 19:00	P	Poster	Nikolai Kolev, Leandro Ferreira, Jayme Augusto	Bayesian analysis of the extended Marshall-Olkin model	p. 92
#28	11/Mar Tue 17:00 - 19:00	P	Poster	Jordania Oliveira	Bayesian approaches to identify active factors in factorial designs 2^k without replication: some comparisons.	p. 93
#29	11/Mar Tue 17:00 - 19:00	P	Poster	Elizabeth Gonzalez Patino, William G. Rojas	Bayesian estimation of Birnbaum-Saunders log-linear model	p. 94
#30	11/Mar Tue 17:00 - 19:00	P	Poster	Claudia Gallegos Perez, Maria Paz Casanova	Discerning among symmetric models based on Bayes factors	p. 95

Continues on the next page

Table 1 – *Contributions*

Chair	Calendar	Session	Type	Authors	Title	Abstract
#31	11/Mar Tue 17:00 - 19:00	P	Poster	Paulo C. Marques, Carlos A. de B. Pereira	Simple random densities	p. 96
#32	11/Mar Tue 17:00 - 19:00	P	Poster	Melaine Oliveira, Andressa Soreira, Victor Fossaluzza, Carlos A. de B. Pereira	Classifying the origin of archaeological fragments with Bayesian nets	p. 97
#33	11/Mar Tue 17:00 - 19:00	P	Poster	Lilian da Costa Fontes, Teresa Cristina Martins Dias	Reliability analysis: study of acceler- ated life testing under a Bayesian ap- proach	p. 98
#34	11/Mar Tue 17:00 - 19:00	P	Poster	Julio Stern, Anna Sikov	Application of Bayesian analysis to es- timation of parametric distributions of complex survey data under informative sampling	p. 99
#35	11/Mar Tue 17:00 - 19:00	P	Poster	Jean Paul Navarrete, Guillermo Ferreira, Mauri- cio Castro, Mario de Castro	Conditional predictive inference for beta regression model with autoregres- sive errors	p. 100
#36	11/Mar Tue 17:00 - 19:00	P	Poster	Viviana Giampaoli, Carlos A. de B. Pereira, Heleno Bolfarine, Julio Singer	Bayes factors for comparison of re- stricted regression coefficients	p. 101
#37	11/Mar Tue 17:00 - 19:00	P	Poster	Felipe de Gusmao, Danielle Roges, Kleber Barros, Eu- frazio Santos, Roseli Lean- dro, Edwin Ortega	Generalized Inverse Weibull distribu- tion: the classical and Bayesian esti- mation	p. 102

Continues on the next page

Table 1 – *Contributions*

Chair	Calendar	Session	Type	Authors	Title	Abstract
#38	11/Mar Tue 17:00 - 19:00	P	Poster	Nicholas Wagner Eugenio, Adriano Polpo	Regression via transform-both-sides model	p. 103
#39	11/Mar Tue 17:00 - 19:00	P	Poster	Laryssa Santos, Marcio Vi- ola	Using copulas in statistical quality con- trol	p. 104

Bregman divergence measures in Bayesian modeling with applications

Dipak K. Dey
University of Connecticut

Gyuhyeong Goh
University of Connecticut

Abstract

We bring the notion of various Bregman divergence measures and propose several applications in the statistical methodology under a Bayesian framework. We define three types of Bregman divergence measures; classical Bregman divergence, Total Bregman divergence, and functional Bregman divergence. The classical Bregman divergence, originally called Bregman divergence, provides a means to measure the dissimilarity between two vectors. Since the Bregman divergence includes a large number of useful loss functions, it has been widely used in machine learning applications. The functional Bregman divergence is a functional version of Bregman divergence, i.e., the functional Bregman divergence measures the dissimilarity between two functions. The Total Bregman divergence was recently developed as an alternative of the classical Bregman divergence due to the lack of robustness of the original Bregman divergence. Consequently the Total Bregman divergence could be applied in the presence of outliers. This presentation consists of three parts: first, we propose a general Bayesian model diagnostics method using functional Bregman divergence. Next, a new robust Bayesian model selection criterion will be proposed by using the Total Bregman divergence. Finally, an application of Bregman divergence will be discussed in semi-supervised learning problems based on a hierarchical Bayesian approach.

Predictive inference under exchangeability, and the imprecise Dirichlet multinomial model

Gert de Cooman

Ghent University, SYSTeMS Research Group

Jasper De Bock

Ghent University, SYSTeMS Research Group

Marcio A. Diniz

Statistics Department, Universidade Federal de S. Carlos

Abstract

Coherent reasoning under uncertainty can be represented in a very general manner by coherent sets of desirable gambles. In this framework, and for a given finite category set, coherent predictive inference under exchangeability can be represented using Bernstein coherent cones of multivariate polynomials on the simplex generated by this category set. This is a powerful generalisation of de Finetti's Representation Theorem allowing for both imprecision and indecision. We define an inference system as a map that associates a Bernstein coherent cone of polynomials with every finite category set. Many inference principles encountered in the literature can then be interpreted, and represented mathematically, as restrictions on such maps. We discuss two important inference principles: representation insensitivity and a strengthened version of Walley's representation invariance and specificity. We show that there is an infinity of inference systems that satisfy these two principles, amongst which we discuss in particular the inference systems corresponding to (a modified version of) Walley and Bernard's Imprecise Dirichlet Multinomial Models (IDMMs) and the Haldane inference system.

Bayesian instrumental variables: priors and likelihoods

Hedibert Lopes
INSPER

Abstract

Instrumental variable (IV) regression provides a number of statistical challenges due to the shape of the likelihood. We review the main Bayesian literature on instrumental variables and highlight these pathologies. We discuss Jeffreys priors, the connection to the errors-in-the-variables problems and more general error distributions. We propose, as an alternative to the inverted Wishart prior, a new Cholesky-based prior for the covariance matrix of the errors in IV regressions. We argue that this prior is more exible and more robust than the inverted Wishart prior since it is not based on only one tightness parameter and therefore can be more informative about certain components of the covariance matrix and less informative about others. We show how prior-posterior inference can be formulated in a Gibbs sampler and compare its performance in the weak instruments case for synthetic as well as two illustrations based on well-known real data.

Bayesian partial linear model for skewed longitudinal data

Debajyoti Sinha
FSU

Abstract

Current statistical models and methods focusing on mean response are not appropriate for longitudinal studies with heavily skewed continuous response. For such longitudinal response, we present a novel model accommodating a partially linear median regression function and within subject association. We develop a semiparametric Bayesian estimation procedure using a Dirichlet process mixture prior for the skewed error distribution. We provide merits for our methods including theoretical investigation of the prior, asymptotic properties of the posterior and also simulation studies of finite sample properties. Ease of computational implementation via available MCMC tools and other additional advantages of our method compared to those of existing methods are illustrated via analysis of a cardiotoxicity study of children of HIV infected mothers.

FDA and innovative designs: case study of a missed opportunity

Andre Rogatko

Biostatistics and Bioinformatics Research Center
Cedars-Sinai Medical Center

Galen Cook-Wiens

Biostatistics and Bioinformatics Research Center
Cedars-Sinai Medical Center

Mourad Tighiouart

Biostatistics and Bioinformatics Research Center
Cedars-Sinai Medical Center

Steven Piantadosi

Biostatistics and Bioinformatics Research Center
Cedars-Sinai Medical Center

Abstract

The standard 3+3 or “modified Fibonacci” up-and-down (MF-UD) method of dose escalation is by far the most used design in dose-finding cancer early trials. MF-UD was the state-of-the-art in 1971 when it was first used in cancer clinical studies. Since then, over one hundred methodological papers proposing or evaluating dose escalation designs have been published and MF-UD has always shown inferior performance when compared with its competitors regarding maximizing number of patients treated at optimal doses and minimizing number of patients under or overdosed. A consequence of using less effective designs is that more patients are treated with doses outside the therapeutic window. We present a case study where the FDA rejected the proposal to use Escalation with Overdose Control (EWOC), an established dose-finding method which has been extensively used in FDA-approved first in human trials and imposed a sub-optimal design, a variation of the MF-UD, known as accelerated titration (AT) design. We show through extensive simulation studies that the AT design has poor operating characteristics relative to two versions of EWOC under several practical scenarios.

An extended exponentiated-g-negative binomial family with threshold effect

Josemar Rodrigues
ICMC-USP

Abstract

In this paper, we formulate a very flexible family of distributions which unifies most recent lifetime distributions. The main idea is to obtain a cumulative distribution function in the unit interval to transform the baseline distribution with an activation mechanism characterized by a latent threshold variable. The new family has a strong biological interpretation from the competitive risks point of view and an elegant solution thorough the Box-Cox transformation. Several structural properties of the new model are investigated. A Bayesian analysis using Markov Chain Monte Carlo is developed to illustrate with a numerical example the usefulness of the proposed family.

Aggregated functional data model for near-infrared spectroscopy: calibration and prediction

Nancy Garcia
UNICAMP

Abstract

Calibration and prediction for NIR spectroscopy data are performed based on a functional interpretation of the Beer-Lambert formula. Considering that, for each chemical sample, the resulting spectrum is a continuous curve obtained as the summation of overlapped absorption spectra from each analyte plus a Gaussian error, we assume that each individual spectrum can be expanded as a linear combination of B-splines basis.

Nondominated and least sensitive actions in modeling gas pipelines replacement

Jose Pablo Arias-Nicolas
University of Extremadura

Jacinto R. Martin Jimenez
University of Extremadura

Fabrizio Ruggeri
CNR IMATI

Alfonso Suarez Llorens
University of Cadiz

Abstract

The paper performs a detailed sensitivity analysis about replacement policy of gas pipelines in a city network. Stemming from qualitative judgments expressed by company's experts, we present different approaches, within a context of Bayesian robustness analysis, to rank propensity to fail of different classes of gas pipelines. The novelty of the paper is about the use of recent relative sensitivity measures and the interest for nondominated sets, providing a sounder approach with respect to previous works.

Affinity structure in reaction networks: minimum entropy or power dissipation

Julio M. Stern

IME-USP - Institute of Mathematics and Statistics of the
University of São Paulo

Fabio Nakano

EACH-USP - School of Arts, Sciences and Humanities of the
University of São Paulo

Abstract

This article presents a simple derivation of optimization models for reaction networks leading to a generalized form of mass-action kinetics, and compares the formal structure of Minimum Information Divergence, Quadratic Programming and Kirchhoff type network models. These optimization models are used in related articles to develop and illustrate the operation of ontology alignment algorithms and to discuss closely connected issues concerning the epistemological and statistical significance of sharp or precise hypotheses in empirical science.

A maximum entropy approach to learn Bayesian networks from incomplete data

Giorgio Corani

Dalle Molle Institute for Artificial Intelligence

Cassio P. de Campos

Dalle Molle Institute for Artificial Intelligence

Abstract

This paper addresses the problem of estimating the parameters of a Bayesian network from incomplete data. This is a hard problem, which for computational reasons cannot be effectively tackled by a full Bayesian approach. The workaround is to search for the estimate with maximum posterior probability. This is usually done by selecting the highest posterior probability estimate among those found by multiple runs of Expectation-Maximization with distinct starting points. However, many local maxima characterize the posterior probability function, and several of them have similar high probability. We argue that high probability is necessary but not sufficient in order to obtain good estimates. We present an approach based on maximum entropy to address this problem and describe a simple and effective way to implement it. Experiments show that our approach produces significantly better estimates than the most commonly used method.

Nonparametric mixtures based on skew-normal distributions: an application to density estimation

Rosangela H. Loschi
UFMG

Caroline C. Vieira
UFES

Denise Duarte
UFMG

Abstract

This work addresses the density estimation problem using non-parametric Bayesian approach. It is considered hierarchical mixture models where the uncertainty about the mixing measure is modeled using the Dirichlet process. The main goal is to build more flexible models for density estimation. Extensions of the normal mixture model via Dirichlet process previously introduced in the literature are twofold. Firstly, Dirichlet mixtures of skew-normal distributions are considered, say, in the first stage of the hierarchical model, the normal distribution is replaced by the skew-normal one. We also assume a skew-normal distribution as the center measure in the Dirichlet mixture of normal distributions. Some important results related to Bayesian inference in the location-scale skew-normal family are introduced. In particular, we obtain the stochastic representations for the full conditional distributions of the location and skewness parameters. The algorithm introduced by MacEachern and M'ller in 1998 is used to sample from the posterior distributions. The models are compared considering simulated data sets. Finally, the well-known Old Faithful Geyser data set is analyzed using the proposed models and the Dirichlet mixture of normal distributions. The model based on Dirichlet mixture of skew-normal distributions captured the data bimodality and skewness shown in the empirical distribution.

Modeling reliability data using non-gaussian state space models

Dani Gamerman
UFRJ

Thiago R. Santos
UFMG

Glaura C. Franco
UFMG

Abstract

This paper presents a likelihood structure, encountered in several reliability models, that consists on a decomposition of data information in stages or times, thus leading to latent parameters. Piecewise exponential models, proportional hazards models and a few software reliability models are among the models in this unifying framework. Latent states are related across stages through a non-Gaussian state space model framework. A distinctive feature of the models is their mathematical tractability, allowing for the exact computation of the marginal likelihood function, despite the non-Gaussianity of the state-space. Real-life examples illustrate the approach and reveal some of its advantages.

Bayesian learning of density function by multiple sequential inversions of 2-D images in electron microscopy

Dalia Chakrabarty

Department of Mathematics, University of Leicester
and Department of Statistics, University of Warwick

Abstract

We present a novel application of Bayesian density estimation, achieved by multiple inversions of images achieved with electron microscopes. The imaging technique renders an image datum as the projection of the convolution $\rho * \eta$ of the unknown material density function $\rho(x, y, z)$ and the unknown kernel function $\eta(x, y, z)$, onto the centre of an identified three dimensional region. Thus, it is a sequence of three contractive orthogonal projections of $\rho * \eta$ that results in a datum which implies that learning the material density and kernel functions individually would require multiple inversions of the image data; this is a harder than usual inverse problem. In the Bayesian context, such learning involves writing the posterior of the unknowns given the image data. We resort to expanding the availability of data by imaging at different values of a parameter E that allows for information over different sub-surface depths to be compressed. This new design of imaging technique is in acknowledgement of the severe logistical difficulties involved in multi-angle imaging. In the small noise limit, the convolution of the unknown density and the unknown kernel is uniquely recovered, but the ill-posed problem of learning the individual functions from the learnt convolution is addressed by developing priors on the material density, such that these priors adapt to the sparsity inherent in the density function. The priors on the kernel function are elicited from the microscopy literature, resulting in more and less informed models. The likelihood is defined in terms of the distance between the convolution of the unknown functions and the image data. We write the posterior probability density of the unknowns given the data, and sample from this posterior density using an adaptive Metropolis-within-Gibbs inference scheme. Applications to various data sets will be presented.

Bayesian Inference in Cumulative Distribution Fields

Ricardo Silva
University College London

Abstract

One approach for constructing copula functions is by multiplication. Given that products of cumulative distribution functions (CDFs) result also in CDFs, an adjustment to this multiplication will result in a copula model, as discussed by Liebscher (J Mult Analysis, 2008). A model parameterization via products of CDFs has some advantages, both from the copula perspective (e.g., it is well-defined for any dimensionality) and from general multivariate analysis (e.g., provides models where small dimensional marginal distributions can be easily read-off from the parameters). Independently, Huang and Frey (J Mach Learn Res, 2011) showed the connection between certain sparse graphical models and products of CDFs, as well as message-passing (dynamic programming) schemes for computing the likelihood function of such models. Such schemes allows models to be estimated with likelihood-based methods. We discuss and demonstrate MCMC approaches for estimating such models in a Bayesian context, their application in copula modelling, and how message-passing can be strongly simplified. Importantly, our view of message-passing opens up possibilities to scaling up such methods, given that even dynamic programming is intractable for calculating likelihood functions in many models.

A skew-normal, nonstationary spatio-temporal process

Alexandra Schmidt
IM-UFRJ

Patricia L.V. da Costa
UFF & UFRJ

Abstract

In the analysis of most spatio-temporal processes in environmental studies, observations present skewed distributions. Usually, data transformations are used to approximate normality, and stationary Gaussian processes are assumed to model the transformed data. The choices of distribution and underlying covariance structure are key for spatial interpolation and temporal prediction.

We propose a spatio-temporal model for skew-normal data that precludes the use of data transformation. The process is decomposed as the sum of a purely temporal structure with two independent components that are considered to be partial realizations from independent spatial Gaussian processes, for each time t . We discuss the use of flexible covariance structures for these components. In particular, we explore the use of covariate information in the covariance structure and investigate how this affects the resultant covariance structure of the process. We analyze daily average temperature during the autumn and winter of 2006 in the South of Brazil.

A Spanning tree hierarchical model for land cover classification

Hunter Glanz
Boston University

Luis Carvalho
Boston University

Abstract

Image segmentation persists as a major statistical problem, with the volume and complexity of data expanding alongside new technologies. Land cover classification, one of the largest problems in Remote Sensing, provides an important example of image segmentation whose needs transcend the choice of a particular classification method. That is, the challenges associated with land cover classification pervade the analysis process from data pre-processing to estimation of a final land cover map. Multispectral, multitemporal data with inherent spatial relationships have hardly received adequate treatment due to the large size of the data and the presence of missing values. In this article we propose a novel, concerted application of methods which provide a unified way to estimate model parameters, impute missing data, reduce dimensionality, and classify land cover. This comprehensive analysis adopts a Bayesian approach which incorporates prior subject matter knowledge to improve the interpretability, efficiency and versatility of land cover classification. We explore a parsimonious, parametric model whose structure allows for a natural application of principal components analysis to isolate important spectral characteristics while preserving temporal information. Moreover, it allows us to impute missing data and estimate parameters via expectation-maximization. We employ a spanning tree approximation to a Potts prior to incorporate spatial relationships in a judicious way and more efficiently access the posterior distribution of the pixel labels. We achieve exact inference of the labels via the centroid estimator. We demonstrate this series of analysis on a set of MODIS data centered on Montreal, Canada.

Identification of differentially expressed genes in RNA-Seq platform: a Bayesian multivariate logistic-normal model.

Carlos A. de B. Pereira
IME-USP

Rafael Stern
CMU

Abstract

We consider the problem of diagnosing a patient according to his digital gene expression profile; i. e., classifying the patient into one of r (≥ 1) possible health conditions. It is assumed that each patient suffers from one of these r conditions. The patient is to be classified according to his observed vector $z = (z_1, z_2, \dots, z_k)$ of frequencies associated to the k most differentially expressed genes. The choice of which k counts are to be considered for diagnostic must be made beforehand and based on vectors of frequencies of patients who had their condition previously found. These data vectors, however, have a number of components, g , which is much larger than k . We therefore aim to identify the k most valuable “tags” for diagnostic. Our method consists in using the multivariate normal logistic distribution for each patient, a meta-analysis to build the models for all of the r groups of patients and then calculate, for each pair of conditions, the probability of one group be more expressed than the other for each gene of the whole set of individual libraries. This will allow us to determine the group of most differentially expressed genes. Finally we define an index to be used in the classification of a new patient. This index is based in a linear combination of the parameters representing gene expression.

Bayesian weighted information measures

Salimeh Yasaei Sekeh
Departamento de Estatística - UFSCar

Abstract

In the context of neurobiology, the occurrence of an event removes a double uncertainty: the quantitative one related to its probability, and the qualitative one related to its utility, where the utility of an event is independent of its probability.

In this situation one can measure the information for the occurrence of an event with probability p and utility u through a quantity $I = I(u, p)$ depending on both variables.

In agreement with Belis and Guiasu (1968) and Di Crescenzo and Longobardi (2006) we present some measures of uncertainty based on the notion of the weighted entropy and weighted information measures.

Indeed, the main purpose of this work is to introduce weighted information measures for Bayesian analysis. We focus also on dynamic cases, the lifetime distributions where the support is truncated at the current age $t > 0$ and notion of dynamic Bayesian Weighted Information Measure is presented. Further, some remarks and examples of interest in reliability theory on use of weighted expected information is given.

Poisson models for time series of counts: a comparison between observation and parameter driven models

Glaura C. Franco

Departamento de Estatística, Universidade Federal de Minas Gerais

Hélio S. Migon

Departamento de Métodos Estatísticos, Universidade Federal do Rio de Janeiro

Marcos O. Prates

Departamento de Estatística, Universidade Federal de Minas Gerais

Abstract

In the time series literature there are two basic approaches for dealing with count data, specially those following a Poisson distribution, namely the observation and parameter driven models. Although there are some works on this subject, an extensive simulation study comparing the two procedures has not yet been performed. The objective of this paper is to present some of the most used models under the two approaches, based upon conditional Poisson distribution, pointing out their main differences and similarities, concerning parameter estimation, model fitting and forecasting. The models are fitted under the Bayesian framework, using a Poisson generalized linear model with a latent $AR(p)$ process in the mean, which accounts for autocorrelation in the data. In the parameter driven case, the INLA approach is used, while for the observation driven models addressed here, an MCMC procedure is developed to estimate the parameters. Concerning the covariates, coefficient estimates turn out to be remarkably similar across the different models. By the other hand, estimates for the autoregressive coefficients and the mean function and forecasts of future values depend heavily on the underlying process which generates the data. The results also show that when the parameters of the autoregressive process approach the non-stationary region, the parameter driven models present a much better performance.

A Binary Regression Mixed Model using a Skew-Probit Link

Jorge Luis B. Guzman

ICMC - Universidade de So Paulo

Cristian Bayes

Pontificia Universidad Católica del Perú

Luis Valdivieso

Pontificia Universidad Católica del Perú

Abstract

We present in this paper a new mixed model for binary response data by considering a skew probit link function. This model, that includes as a special case the Probit mixed model, is appropriate for the analysis of several types of correlated structures, like clustered, longitudinal and multilevel data. A Bayesian approach is adopted for inference using a Markov Chain Monte Carlo (MCMC) method that can be easily implemented throughout an augmented likelihood technique. Model comparison criteria and model assessment are discussed. We include two applications with real data, showing that the skew probit link outperforms other links proposed in the literature.

Point process with spatio-temporal heterogeneity

Jony Arrais Pinto Junior

Instituto de Matemática e Estatística, Universidade Federal Fluminense

Dani Gamerman

Instituto de Matemática, Universidade Federal do Rio de Janeiro

Marina Silva Paez

Instituto de Matemática, Universidade Federal do Rio de Janeiro

Abstract

This article proposes a modelling approach for handling spatio-temporal heterogeneity present in the study of the geographical pattern point. Our model is a generalization of log-Gaussian Cox process, obtained after allowing effects of all components involving individual covariates to vary over space and time. A simulation study will be conducted to verify the ability of the model to estimate the parameters and hyperparameters. Our model is able to capture and highlight important data information that would not be noticed otherwise, providing information that is required for appropriate decision making, for example, in public health where the study of occurrences of a particular disease in time and space are extremely important.

Bayesian analysis of censored linear regression models with scale mixtures of Normal distributions

Víctor Hugo Lachos

Departamento de Estatística, Universidade Estadual de Campinas

Aldo M. Garay

Departamento de Estatística, Universidade de São Paulo

Heleno Bolfarine

Departamento de Estatística, Universidade de São Paulo

Celso Rómulo Barbosa Cabral

Departamento de Estatística, Universidade Federal do Amazonas

Abstract

As is the case of many studies, the data collected are limited and an exact value is recorded only if it falls within an interval range. Hence, the responses can be either left, interval or right censored. Linear (and nonlinear) regression models are routinely used to analyze these types of data and are based on normality assumptions for the errors terms. However, those analyses might not provide robust inference when the normality assumptions are questionable. In this article, we develop a Bayesian framework for censored linear regression models by replacing the Gaussian assumptions for the random errors with scale mixtures of normal (SMN) distributions. The SMN is an attractive class of symmetric heavy-tailed densities that includes the normal, Student-t, Pearson type VII, slash and the contaminated normal distributions, as special cases. Using a Bayesian paradigm, an efficient Markov chain Monte Carlo (MCMC) algorithm is introduced to carry out posterior inference. A new hierarchical prior distribution is suggested for the degrees of freedom parameter in the Student-t distribution. The likelihood function is utilized to compute not only some Bayesian model selection measures but also to develop Bayesian case-deletion influence diagnostics based on the q -divergence measures. The proposed Bayesian methods are implemented in the R package BayesCR. The newly developed procedures are illustrated with an application and simulated.

Bayesian ridge-regularized covariance selection with community behavior in latent Gaussian graphical models

Lijun Peng

Department of Mathematics and Statistics, Boston University

Luis Carvalho

Department of Mathematics and Statistics, Boston University

Abstract

Gaussian graphical models have been extensively used to model conditional independence via the concentration matrix of a random vector. They are particularly relevant to incorporate structure when the length of the vector is large and naive methods lead to unstable estimation of the concentration matrix. In covariance selection, we have a latent network among vector components such that two components are not connected if they are conditionally independent, that is, if their corresponding entry in the concentration matrix is zero. In this work, we expect that, in addition, vector components show a block dependency structure that represents community behavior in the context of biological and social applications, that is, connections between nodes from different blocks are sparse while connections within nodes of the same block are dense. Thus, to identify the latent network and detect communities, we propose a Bayesian approach with a hierarchical prior in two levels: a spike-and-slab prior on each off-diagonal entry of the concentration matrix for variable selection; and a degree-corrected stochastic blockmodel to capture the community behavior. To conduct inference, we develop an efficient routine based on ridge regularization and MAP estimation. Finally, we demonstrate the proposed approach in a meta-genomic dataset of complex microbial biofilms from dental plaque and show how bacterial communities can be identified.

Bayesian partitioning for survival data with a cure fraction

Jhon Franky Bernedo Gonzales

Department of Statistics - Federal University of São Carlos

Vera Lúcia Damasceno Tomazella

Department of Statistics - Federal University of São Carlos

Mário de Castro

ICMC, Universidade de São Paulo

Abstract

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

A flexible cure rate model with frailty term in latent risks: a Bayesian approach

Vinicius Fernando Calsavara
IME - Universidade de São Paulo

Agatha Sacramento Rodrigues
IME - Universidade de São Paulo

Abstract

The occurrence of an event of interest in survival analysis may be given by one or many competitive risks. In competitive risks scenario the number of competitive factors and the lifetimes associated with its causes are unobservable, which leads to so called latent competitive risks. This scenario has the assumption of independence among risk factors. However, this assumption is not very realistic, since such latent time are random variables of the same sampling unit. In this paper, we propose a flexible cure rate model with a frailty term in latent risks, which extends Cancho et al. (2011) models by incorporating a frailty term in risk function of latent factors. We consider a cure rate model based on the negative binomial distribution for the number of competitive risks and power variance function for the frailty variable, encompassing several special models. Parameters estimation is based on Bayesian analysis considering Markov chain Monte Carlo (MCMC) methods with non-informative priori. The practical relevance and applicability of the proposed model is demonstrated in a real data set.

Bayesian ordinal regression

Leonardo Cella
UnB

Abstract

This work presents inferences of ordinal regression models considering the Logit link function and the multinomial likelihood approach. A new reparametrization was proposed for the regression model. The inferences were performed in a bayesian scenario, using the MCMC (Markov Chain Monte Carlo) technics. Point estimates of the parameters and their respective HPD credibility intervals are presented, as well a Full Bayesian Significance Test (FBST) for the regression parameters. This methodology was applied on simulated data and illustrated in a genetic problem which was to verify the influence of certain radiation on the occurrence of cellular damage. The multinomial likelihood approach combined with the model reparametrization is easy to treat due the increasing computing power and the advancement of MCMC methods. Moreover, the FBST proved being a simple and useful procedure for testing the significance of regression coefficients, thus motivating the use of a bayesian approach in ordinal data modeling.

Bayesian semiparametric longitudinal data modeling using normal/independent densities

Luis M. Castro
Universidad de Concepcion

Victor H. Lachos
UNICAMP

Diana M. Galvis
UNICAMP

Dipankar Bandyopadhyay
Division of Biostatistics, University of Minnesota

Abstract

In this paper, we consider normal/independent distributions as a tool for robust modeling of semiparametric linear mixed-effects models from a Bayesian perspective. The normal/independent distributions is an attractive class of symmetric heavy-tailed distributions that includes the normal, Student-t, slash and the contaminated normal distribution as special cases, providing an appealing robust alternative to the routine use of normal distributions for longitudinal data. Using a Bayesian paradigm, an efficient Markov chain Monte Carlo (MCMC) algorithm is introduced to carry out posterior inference. In order to examine the robust aspects of this flexible class against outlying and influential observations we present a Bayesian case deletion influence diagnostics based on the q-divergence measures. Further, some discussions on Bayesian model selection criteria are given. The new method is exemplified through simulated data and a real data set of AIDS/HIV infected patients.

Characterising Dirichlet priors

Marcio A. Diniz

Universidade Federal de S. Carlos, Statistics Department

Jasper De Bock

Ghent University, SYSTeMS Research Group

Arthur Van Camp

Ghent University, SYSTeMS Research Group

Abstract

The characterisation of prior distributions is a problem that has been discussed since Bayes published his paper. Bayes himself tried to justify his prior following a principle that became known as principle of insufficient reason. In the last century, several applied researchers used conjugated priors mainly for mathematical convenience while others still tried to justify their choices based on reasonable principles of inference and/or key features, depending on the problem at hand. This work justifies the use of Dirichlet priors based on a principle called partition invariance, alongside with other requirements that are already well known from the literature.

Assessing a spatial boost model for quantitative trait GWAS

Ian Johnston
Boston University

Yang Jin
Boston University

Luis Carvalho
Boston University

Abstract

Bayesian variable selection provides a principled framework for incorporating prior information to regularize parameters in high-dimensional large-p-small-n regression models such as genome-wide association studies. Although these models produce more informative results, researchers often disregard them in favor of simpler models because of their high computational cost. We explore our recently proposed Spatial Boost model for GWAS on quantitative traits to assess the computational efficiency of a more representative model. The Spatial Boost model is a Bayesian hierarchical model that exploits spatial information on the genome to uniquely define prior probabilities of association of genetic markers based on their proximities to relevant genes. We propose analyzing large data sets by first applying an Expectation-Maximization filter to reduce the dimensionality of the space and then applying an efficient Gibbs sampler on the remaining markers. Finally we conduct a thorough simulation study based on real genotypes provided by the Wellcome Trust Case Control Consortium and compare our model to single association tests.

Sparse latent factor models with interactions

Vinicius Diniz Mayrink
Universidade Federal de Minas Gerais

Joseph E. Lucas
Duke University

Abstract

Sparse latent multi-factor models have been used in many exploratory and predictive problems with high-dimensional multivariate observations. Because of concerns with identifiability, the latent factors are almost always assumed to be linearly related to measured feature variables. Here we explore the analysis of multi-factor models with different structures of interactions between latent factors, including multiplicative effects as well as a more general framework for non-linear interactions introduced via Gaussian Process. We utilize sparsity priors to test whether the factors and interaction terms have significant effect. The performance of the models is evaluated through the analysis of simulated and real gene expression data. Variation in the number of copies of regions of the genome is a well known and important feature of most cancers. We examine interactions between factors directly associated with different chromosomal regions detected with copy number alteration in breast cancer data. In this context, significant interaction effects for specific genes suggest synergies between duplications and deletions in different regions of the chromosome.

Bayesian variable selection in a class of zero inflated modified power series models

Willian Luís de Oliveira

Department of Statistics - Federal University of São Carlos

Carlos Alberto Ribeiro Diniz

Department of Statistics - Federal University of São Carlos

Marinho G. Andrade

Institute of Mathematics and Computer Science - University of São Paulo - USP

Abstract

It is common to find situations in several areas involving discrete data which amount of zero values observed is high, with an incidence of zeros greater than that supported by the distribution. In this context, zero inflated regression models have gained prominence. In this paper we consider a class of zero inflated modified power series regression models, well explored class of model in the statistical literature. Our aim is to develop a variable selection procedure for this model using both a classical and a Bayesian approaches. Several methods are considered for this class of model. We have adapted to the zero inflated modified power series model the Aikiake Information Criterion (AIC) and Bayesian Information Criterion (BIC) classical methods and the Conditional Predictive Ordinate (CPO), Deviance Information Criterion (DIC), Expected Aikiake Information Criterion (EAIC), Expected Bayesian Information Criterion (EBIC) and L measure Bayesian methods. Using OX and R programming language, several simulated datasets are used to illustrate the methods and to compare their performance.

Bayesian hypothesis testing in finite population

Brian Alvarez R. de Melo

Institute of Mathematic and Statistics, University of São Paulo

Luís Gustavo Esteves

Institute of Mathematic and Statistics, University of São Paulo

Abstract

In this paper, we consider Bayesian hypothesis testing for the (operational) parameter of interest of a finite population composed of N units. More precisely, we develop test procedures for the relevant parameter under the predictivistic perspective, introduced by Bruno de Finetti, to contrast with the usual development for superpopulation models. The comparison between these approaches, exemplified in a simple scenario of majority elections, shows considerable differences between the corresponding results for large sampling fractions.

The homogeneity test for 2x2 contingency table

Natalia Oliveira

Departamento de Estatística - UFSCar

Marcio Diniz

Departamento de Estatística - UFSCar

Adriano Polpo

Departamento de Estatística - UFSCar

Abstract

Using the likelihood ratio statistics, we developed a significance index, called P-value, to test the hypothesis of homogeneity in 2x2 contingency tables. The P-value does not depend on asymptotic distributions, and is based on the elimination of the nuisance parameter. That is, we proposed a way to obtain the exact distribution of the likelihood ratio statistics. This procedure is compatible with the likelihood principle. To a better understanding of significance indices to test homogeneity, we performed a simulation study comparing some frequentist indices (likelihood ratio test, chi-square test) and the full Bayesian significance test (FBST). The comparative study showed an interesting relation between all the indices studied, Bayesian and frequentists.

Factor analysis with mixture modeling to evaluate coherent patterns in microarray data

Joao Daniel Nunes Duarte

Departamento de Estatística, ICEX, UFMG

Vinicius Diniz Mayrink

Departamento de Estatística, ICEX, UFMG

Abstract

The computational advances over the last decades have allowed the use of complex models to analyze large data sets. The development of simulated-based methods, such as the MCMC, has contributed to an increased interest in the Bayesian framework as an alternative to work with factor models. Many studies have applied the factor analysis to explore gene expression data with results often outperforming traditional methods for estimating and identifying patterns and metagene groups related to the underlying biology. In this work, we present a Sparse Latent Factor Model (SLFM) using a mixture prior (sparsity prior) to evaluate the significance of each factor loading; when the loading is significant the effect of the corresponding factor is detected through patterns displayed along the samples. The SLFM is applied to investigate simulated and real microarray data. The real data sets represent the gene expression for different types of cancer; this includes breast, brain, ovarian and lung tumors. The proposed model can indicate how strong is the observed expression pattern allowing the measurement of the evidence of presence/absence of the gene activity. Finally, we compare the SLFM with two simpler gene detection methods available in the literature. The results suggest that the SLFM outperforms the traditional methods.

Combining optimization and randomization approaches for the design of clinical trials

Victor Fossaluzza
IME - USP

Marcelo de Souza Lauretto
EACH - Universidade de So Paulo

Carlos Alberto de Bragança Pereira
IME - Universidade de So Paulo

Julio Michael Stern
IME - Universidade de So Paulo

Abstract

Intentional sampling methods are non-randomized procedures that select a group of individuals for a sample with the purpose of meeting specific prescribed criteria. In this paper we extend previous works related to intentional sampling, and address the problem of sequential allocation for clinical trials with few patients. Roughly speaking, patients are enrolled sequentially, according to the order in which they start the treatment at the clinic or hospital. The allocation problem consists in assigning each new patient to one, and only one, of the alternative treatments (arms). The main requisite is that the profiles in the alternative arms remain similar with respect to some relevant patients attributes (age, gender, disease, symptom severity and others). We perform numerical experiments based on a real case study and discuss how to conveniently set up perturbation parameters, in order to yield a suitable balance between optimality the similarity among the relative frequencies of patients in the several categories for both arms, and decoupling the absence of a tendency to allocate each pair of patients consistently to the same arm.

Implementation of censored models for survival analysis

Danilo Lopes

Department of Statistics - Federal University of São Carlos

Danielle L. Roges

Department of Statistics - Federal University of São Carlos

Adriano Polpo

Department of Statistics - Federal University of São Carlos

Abstract

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

Bayesian inference in beta and inflated beta regression models

Danilo Covaes Nogarott
Química Amparo Ltda

Caio Lucidius N. Azevedo
Departamento de Estatística, IMECC, UNICAMP

Abstract

In the present work we developed Bayesian tools, concerning parameter estimation and diagnostics, for noninflated, zero inflated, one inflated and zero-one inflated beta regression models. Due to the impossibility of obtaining the posterior distributions of interest, analytically, all these tools were developed through MCMC algorithms. For the regression and precision parameters we exploited the using of prior distributions commonly considered in regression models as well as Jeffreys and independence Jeffreys priors. For the parameters related to the discrete components, we considered conjugate prior distributions. We performed simulation studies, considering some situations of practical interest, in order to compare the Bayesian and frequentist estimates as well as to evaluate the sensitivity of the models to the prior choice. A psychometric real data set was analyzed to illustrate the performance of the developed tools. The results indicated that there is an overall improvement in using models that consider the inflated observations compared to transforming these observations in order to use noninflated models.

A Bayesian approach to predicting football match outcomes considering time effect weight

Francisco Louzada

Instituto de Ciências Matemáticas e de Computação - ICMC,
Universidade de São Paulo

Adriano K. Suzuki

Instituto de Ciências Matemáticas e de Computação - ICMC,
Universidade de São Paulo

Luis Ernesto B. Salasar

Department of Statistics - Federal University of São Carlos

Anderson Ara

Department of Statistics - Federal University of São Carlos

José Galvão Leite

Department of Statistics - Federal University of São Carlos

Abstract

In this paper we propose a simulation-based method for predicting football match outcomes. We adopt a Bayesian perspective modeling the number of goals of two opposing teams as a Poisson distribution whose mean is proportional to the relative technical level of opponents. FIFA ratings were taken as the measure of technical level of teams saw well as experts opinions on the scores of the matches were taken in account to construct the prior distributions of the parameters. Tournament simulations were performed in order to estimate probabilities of winning the tournament assuming different values for the weight attached to the experts information and different choices for the sequence of weights attached to the previous observed matches. The methodology is illustrated on the 2010 Football World Cup.

Bayesian analysis of gamma regression models: state of the art and extensions

Catalina Araya

Universidad de Santiago de Chile

Francisco Torres-Avilés

Universidad de Santiago de Chile

Abstract

This paper presents a review of the inference in gamma regression models from a Bayesian perspective with emphasis on mixed models. The work begins presenting the usual construction of this class of regression models, to later define some extensions of these ones. We discuss the choice of the link function, the elicitation of prior distributions, inclusion of random effects and model selection through a simulation study. Finally, the methodology is illustrated using real data from the public health area.

Bayesian semi-parametric symmetric models for binary data

Márcio Augusto Diniz

Institute of Mathematics and Statistics - University of São Paulo

Carlos Alberto de Bragança Pereira

Institute of Mathematics and Statistics - University of São Paulo

Adriano Polpo

Department of Statistics - Federal University of São Carlos

Abstract

This work proposes a general Bayesian semi-parametric model to binary data. It is considered symmetric prior probability curves as an extension for discussed ideas from Basu and Mukhopadhyay (2000) using the Blocked Gibbs sampler which is more general than the Polya Urn Gibbs sampler. The semi-parametric approach allows to incorporate the uncertainty around the F distribution of latent data and modeling asymmetric probability curves. The implementation of Blocked Gibbs Sampler overcomes the limitation of conjugate distributions to define the mixture of distributions which composes F. Then, this framework opens several possibilities to deal with binary data in Bayesian perspective.

MCMC-driven adaptive multiple importance sampling

Luca Martino

Dep. of Mathematics and Statistics, University of Helsinki

Víctor Elvira

Dep. of Signal Theory and Communic., Universidad Carlos III de Madrid

David Luengo

Dep. of Circuits and Systems Engineering, Universidad Politecnica de Madrid

Jukka Corander

Dep. of Mathematics and Statistics, University of Helsinki

Abstract

Monte Carlo (MC) methods are widely used for statistical inference and stochastic optimization. A well-known class of MC methods is composed of importance sampling (IS) and its adaptive extensions (such as adaptive multiple IS and population Monte Carlo). In this work, we introduce an iterated batch importance sampler using a population of proposal densities, which are adapted according to an MCMC technique over the population of location parameters. The novel algorithm provides a global estimation of the variables of interest iteratively, using all the generated samples weighted according to the so-called deterministic mixture scheme.

Compared with a traditional multiple IS scheme with the same number of samples, the performance is substantially improved at the expense of a moderate increase in the computational cost due to the additional MCMC steps. Furthermore, the dependence on the choice of the cloud of proposals is sensibly reduced, and the proposal density in the MCMC method can be adapted in order to optimise the performance. Numerical results show the advantages of the proposed sampling scheme in terms of mean absolute error.

A zero variance differential geometric MCMC estimation approach for a beta 0-inflated regression model

Cristian Bayes

Pontificia Universidad Católica del Perú

Luis Valdivieso

Pontificia Universidad Católica del Perú

Abstract

We study in this article the zero variance differential geometric MCMC estimation approach to obtain the parameters of a Beta-inflated regression model that searches to explain a non-negative bounded response variable with a 0 value as a possible outcome. From a classical point of view, this model is estimated using a likelihood two-part methodology. This paper explores and compares some MCMC Bayesian estimation alternatives including the use of the zero variance differential geometric MCMC approach.

A note on Bayesian inference for the long-range dependence of a stationary two-state process

Plinio Andrade
IME, University of São Paulo

Laura Rifo
Institute of Mathematics and Statistics, University of Campinas

Abstract

In this work we propose a Bayesian approach for selecting the dependence range of a stationary process with two states. The analysis is based on approximate posterior distributions on the Hurst index obtained from a likelihood-free method. Our simulations shows that the main advantage of our approach, besides of simplicity, is that we can obtain an approximate sample of the posterior distribution on the Hurst index, providing better estimates. Furthermore, there is no need for Gaussian and/or asymptotic assumptions.

Spatial modelling of malaria risk in Bayesian setting: a case study in Adama district of Oromiya Regional State, Ethiopia

Fekadu Lemessa Bayisa
Adama Science and Technology University

Abstract

Background: Malaria is a location specific and spatially varying infectious disease. It is responsible for one million worldwide deaths every year. Geo-spatially varying climatic and environmental factors play a great role in malaria transmission. The research intended to develop Bayesian spatial model for malaria risk investigation. The research tested the existence of spatial clustering and association of spatially varying climatic and environmental factors with malaria prevalence.

Methods: Using 80% of the data, Generalized Linear Mixed Model (GLMM) was fitted to estimate GLMM parameters. Spatial autocorrelation was assessed using spatial statistic Global Moran's I index. GLMM and spatial models were also fitted in Bayesian setting. The mixing up of MCMC simulations and Estimated Potential Scale Reduction confirmed simulation convergence. Using 20% of the data, Chi-square test and Kullback-Leibler distance were used to select the best model.

Result: The research finding showed that malaria prevalence exhibits spatial clustering. Minimum temperature and NDVI were positively associated with malaria risk. Elevation, maximum temperature and rainfall were negatively associated. Spatial stationary model was selected as the best model.

Conclusion: Spatial modelling of malaria risk provides spatial variation of malaria risk in the form of map. The concerned body may use high malaria risk areas, characterized by high vegetation cover, high minimum temperature, low elevation, not consistently high rainfall and not consistently high maximum temperature, as a starting point for malaria control activities. Malaria risk map can serve as a milestone for optimizing malaria interventions, limited financial and human resources.

A Bayesian approach to estimate the biomass of anchovies in the coast of Perú

Zaida Quiroz

Department of Statistics, Universidade Federal de Minas Gerais

Marcos O. Prates

Department of Statistics, Universidade Federal de Minas Gerais

Havard Rue

Department of Mathematical Sciences, Norwegian University for Science and Technology

Abstract

The Northern Humboldt Current System (NHCS) is the world most productive ecosystem in terms of fish. In particular, Peruvian anchovy (*Engraulis ringens*) is the major prey of the principal top predators, like mammals, seabirds, fish and fishers. In this context, it is important to understand the dynamics of the anchovy distribution to preserve it as well as to explore its economical capacities. Using the data collected by the Instituto del Mar del Per (IMARPE), during a scientific survey in 2005, we present a statistical analysis that has as main goals: (i) adapt to the characteristics of the sampled data, such as spatial dependence, high proportions of zeros and big samples size, (ii) provide important insights on the dynamics of the anchovy population and propose a model for estimation and prediction of anchovy biomass in the NHCS of Per. The data is analyzed in a Bayesian framework using the Integrated Nested Laplace Approximation (INLA) methodology. Finally, are performed models comparison to select the best model and predictive checks to study the predictive power of each model. Moreover, a Bayesian spatial influence diagnostic is performed for the preferred model.

Bayesian Inference of deterministic population growth models

Luiz Max F. de Carvalho

Program for Scientific Computing (PROCC) - Oswaldo Cruz
Foundation

Claudio J. Struchiner

Program for Scientific Computing (PROCC) - Oswaldo Cruz
Foundation

Leonardo S. Bastos

Program for Scientific Computing (PROCC) - Oswaldo Cruz
Foundation

Abstract

Deterministic mathematical models play an important role in our understanding of population growth dynamics. In particular, the effect of temperature on the growth of disease-carrying insects. In this paper we propose a modified Verhulst logistic growth equation with temperature-dependent parameters. Namely, the growth rate r and the carrying capacity K are given by thermodynamic functions of temperature T , $r(T)$ and $K(T)$. Our main concern is with the problem of learning about unknown parameters of these deterministic functions from observations of population time series $P(t, T)$. We propose a strategy to estimate parameters of $r(T)$ and $K(T)$ by treating the model output $P(t, T)$ as a realization of a Gaussian process (GP) with fixed variance and mean function given by the analytic solution to the modified Verhulst equation. We then use Gibbs sampling, implemented using the recently developed `rstan` package of the R statistical computing environment, to approximate the posterior distribution of the parameters of interest. In order to evaluate the performance of our algorithm, we run a Monte Carlo study on a simulated example, where bias and coverage were calculated. We then proceed to apply the approach described herein to laboratory data of the Chagas disease vector, *Rhodnius prolixus*. Analysis of this data shows that the growth rate for the insect population under study achieves its maximum around 26°C and the carrying capacity is maximum around 25°C , suggesting that *R. prolixus* populations may thrive even in temperate climates.

Nonparametric Bayesian regression under combinations of local shape constraints

Khader Khadraoui

Laval University, Department of Mathematics and Statistics

Abstract

A nonparametric Bayesian method for regression under combinations of local shape constraints is proposed. The shape constraints considered include monotonicity, concavity (or convexity), unimodality and in particular combinations of several types of range-restricted constraints. By using a B-spline basis, the support of the prior distribution is included in the set of piecewise polynomial functions. It is shown that the choice of the number of interior spline knots and their positions have a profound effect on the smoothness of the estimated regression function. Furthermore, thanks to the local support property of B-splines, many combinations of constraints can easily be considered by identifying B-splines whose support intersects with each constrained region. Simulations from the posterior distribution are obtained by an MCMC sampling algorithm.

Annealed stochastic approximation EM for reinforcement learning

Orhan Snmez
Bogazici Universitesi

Taylan Cemgil
Bogazici Universitesi

Abstract

Instead of adopting the Bellman equations, we approach the reinforcement learning problem with a probabilistic point of view and transform it into an equivalent probabilistic inference problem. Then, we propose an annealed version of stochastic approximation EM for model-free reinforcement learning to estimate the value function. By annealing the received rewards, we aim to struggle with the exploration-exploitation dilemma. When the annealing process is finalized, our method actually converges to an optimistic policy iteration method whose convergence is guaranteed for some cases.

Approximate Bayesian inference for the Rosenblatt distribution

Laura Rifo

Institute of Mathematics and Statistics, University of Campinas

Plinio Andrade

IME, University of São Paulo

Abstract

The Rosenblatt distribution is a one-parameter family arising from a non-Central limit theorem for long-range dependent random variables. This family includes the standard normal distribution, the standardized chi-squared distribution, and weighted sums of chi-squared variates. Its analytical form is not manageable, and its moments, cumulants and empirical distribution have just recently been numerically studied. We apply a Bayesian likelihood-free methodology to obtain inferences for that family, comparing the performance of some statistics.

Inference on dynamic models for near-Gaussian random fields using INLA

Renan X. Cortes

Departamento de Estatística, Universidade Federal de Minas Gerais

Thiago G. Martins

Department of Mathematical Sciences, Norwegian University of Science and Technology

Marcos O. Prates

Departamento de Estatística, Universidade Federal de Minas Gerais

Bráulio F. A. Silva

Departamento de Sociologia, Universidade Federal de Minas Gerais

Abstract

Robust time series analysis is an important subject in statistical modeling. Models based on Gaussian distribution are sensitive to outliers, which may imply in a significant degradation in estimation performance as well as in prediction accuracy. State-space models, also referred as Dynamic Models, is a very useful way to describe the evolution of a time series variable through a structured latent evolution system. Integrated Nested Laplace Approximation (INLA) is a recent approach proposed to perform fast Bayesian inference in Latent Gaussian Models which naturally comprises Dynamic Models. We present how to perform fast and accurate non-Gaussian dynamic modeling with INLA and show how these models can provide a more robust time series analysis when compared with standard dynamic models based on Gaussian distributions. We formalize the framework used to fit complex non-Gaussian space-state models using the R package INLA and illustrate our approach with a brazilian homicide rate dataset.

Bayesian inference for Birnbaum-Saunders mixed models for censored responses

Caio Lucidius Nabrezny Azevedo

Departamento de Estatística, IMECC, UNICAMP

Filidor Vilca-Labra

Departamento de Estatística, IMECC, UNICAMP

Cristian Lobos

Departamento de Ciências Exatas, ESALQ, USP

Abstract

In this work we propose Bayesian inference mechanisms for the class of Birnbaum-Saunders mixed models for censored responses developed by Villegas et al. (2012). Such class of regression models is very suitable for analyzing longitudinal data in fields like reliability and survival analysis. We study the using of the so-called half Cauchy prior, developed by Gelman (2006), to model the variance of the random effects. Due to intractable form of the joint posterior distribution, we obtained numerical approximations for the marginal posterior distributions, through MCMC algorithms implemented in the R program. We conducted simulation studies in order to compare the Bayesian estimates with the marginal maximum likelihood ones, which were developed by Villegas et al. (2012). Also, we performed a sensitivity study concerning the prior choice for the variance of the random effects. Bayesian mechanisms of model comparison are discussed. Finally, a real data set of reliability field is analyzed.

Bivariate zero-inflated Poisson distribution: a Bayesian approach

Veronica Amparo Quispe
UFSCar

Vera Lúcia Damasceno Tomazella
UFSCar

Luis Ernesto Bueno Salasar
UFSCar

Marcio L. L. Viola
UFSCar

Abstract

Bivariate count data occurs in several fields such as epidemiology, medicine, industry, among others. One of the most applied distribution in such instances is the Holgate bivariate Poisson distribution. Nonetheless, data-sets may present a large number of zeros, which can not be accommodated by that bivariate distribution. Thus, in this paper we propose a zero-inflated version of the Holgate bivariate Poisson distribution under a bayesian perspective. The prior distributions of the model parameters are assumed to be in a given conjugated class. The calculations from the posterior distributions are obtained by the Gibbs sampling algorithm with the inclusion of latent data, which simplifies the expressions of the conditional distributions of the parameters given the augmented data. The proposed methods are illustrated on a real data-set concerning the number of accidents involving 708 bus drivers occurred in two consecutive periods in north of Ireland.

FBST-CT Package: full Bayesian significance test for contingency tables

Bruna Turato Santos

EACH - Universidade de São Paulo

Marcelo Lauretto

EACH - Universidade de São Paulo

Abstract

The FBST (Full Bayesian Significance Test) was proposed as a coherent and intuitive Bayesian approach which does not assign positive probabilities to zero measure sets when testing sharp hypotheses. FBST have shown a good comparative performance in several empirical studies, including problems related to contingency tables. This work introduces the FBST-CT Package, an open source R Package which implements the FBST for five classical tests on contingency tables: proportion, homogeneity, independence, symmetry and genetic (Hardy-Weinberg) equilibrium. Besides the standard evidence measure (e-value), FBST-CT also implements two approaches for computing the threshold for acceptance/rejection of a hypothesis H . The first is based on an asymptotic approximation of the evidence against H , given that H is true, whilst the second approach computes an "empirical p-value", via Monte Carlo Simulation. FBST-CT is easy to install and to use, it is adjustable to any table size and explores multi-core architectures when computing the empirical p-value. These features are intended to make the package a handy tool for helping researchers and practitioners to take advantage of FBST for tests on contingency tables.

Bayesians decision rule of investment

Patrcia Viana da Silva

Faculdade de Matemática, Universidade Federal de Uberlândia

Kamylla Rodrigues Leandro

Faculdade de Matemática, Universidade Federal de Uberlândia

Gabriela Golçalves Peres

Faculdade de Matemática, Universidade Federal de Uberlândia

Abstract

The stock market has indispensable function to the economy's development and also to the society's development. Investments in financial assets involve decisions about where and how to apply the patrimony in accordance with the market risk. Thus, the investment depends on the behavior of the investor, in other words, the level of risk you want to take to obtain the expected return. Whether for the speculator, whether for the small investor, the ensuring of preservation or evolution of patrimony will depend on the assumed risk. And a way to define this risk is to check how is the behavior of the variation of returns of this asset over time. In this paper is presented a problem in which you want to choose an investment portfolio using economic indicators. The model used is based on Decision Theory and Bayesian Inference in which the loss is proposed as a function of the expected gain by the investor. And the assets are studied Scores Electricity and considering IBOVESPA data April 2011 to March 2013.

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

George L. M. Pezzott

Departamento de Estatística - UFSCar

Luis E. B. Salasar

Departamento de Estatística - UFSCar

José G. Leite

Departamento de Estatística - UFSCar

Abstract

In this paper, we consider Bayesian estimation of unknown number of elements in a closed population through a capture-recapture model. Let N be the number of individuals in the population which are sampled in t different occasions, $t > 1$. We shall assume initially that each individual i is captured in the j occasion, independently from the other individuals and other occasions, with probability p_i , that is, the capture probabilities varies with individuals but not with occasions. Furthermore, we suppose the capture probabilities p_1, \dots, p_N are i.i.d random variables with discrete distribution function F which takes values $\theta_1, \dots, \theta_k$ with probabilities $\omega_1, \dots, \omega_k$, respectively, $k > 1$. For identifiability purpose, we shall adopt the restriction $\theta_1 < \dots < \theta_k$. Thus, we interpret the parameters $\theta_1, \dots, \theta_k$ as different levels of catchability for the individuals and $\omega_1, \dots, \omega_k$ as the proportions of individuals in the population pertaining to each level. The estimation process was performed in the Bayesian paradigm and we have adopted independent prior distributions N , $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ and $\boldsymbol{\omega} = (\omega_1, \dots, \omega_k)$. For the interest parameter N we assumed the improper discrete prior pmf $\pi(N) = N^{-r} I_{N>0}$ with $r = 0$ (uniform) or $r = 1$ (Jeffreys), for $\boldsymbol{\omega}$ a vague Dirichlet distribution and for $\boldsymbol{\theta}$ an uniform pdf over its restricted space. The existence of the posterior distribution for the parameters are discussed for the proposed prior distributions. The inference was carried through MCMC Gibbs sampling algorithm. To determine the conditional posterior distributions for the nuisance parameters, we make use of artificial variables in the likelihood function. The proposed method is applied to two real data sets. The first one is related to an animal trapping study performed in 18 consecutive days and the second concerns a software review designed to identify the number of errors in a specific feature of AT&T 5 ESS switch.

Optimal decision about sample sizes considering costs

Andressa Gigante

Departamento de Estatística - UFSCar

Danielle Loureiro Roges

Departamento de Estatística - UFSCar

Willian Oliveira

Departamento de Estatística - UFSCar

Abstract

It is common in many problems from many sector, propose statistical models such as problem solving tool. An important part of these cases is the fit, the estimation of the model assumed. The estimation problem can be viewed as a decision problem in which the decision to be made will be the estimated values of the parameters of interest. As a decision problem, the problem of estimating the need exists for choosing a loss function as the optimal decision will be directly related to that choice.

Statistical before choosing for a decision have the opportunity to observe the value of one or more random variables, related to a particular experiment which will help

you make the best decision. In this case the decision to be made tend to be more accurate as the number of observations increases. However, in general, there is a cost associated in obtaining the observations. Thus, it is desirable to determine the number of good observations in such a way that the overall risk involving the expected loss and expected cost is minimal.

In this context, the main objective of this study was to consider some statistical models and determine the size of great required sample that results in a decision - estimation - lower risk for some loss functions - quadratic loss and the absolute loss - taking into account the cost of each observation. Different cost functions have been considered and the results were compared.

Autoregressive moving average processes with generalized normal distribution

Eder A. Milani

Department of Statistics - Federal University of São Carlos

Marinho G. Andrade

Institute of Mathematics and Computer Science - University of São Paulo - USP

Abstract

It is common to find situations in several areas involving discrete data which amount of zero values observed is high, with an incidence of zeros greater than that supported by the distribution. In this context, zero inflated regression models have gained prominence. In this paper we consider a class of zero inflated modified power series regression models, well explored class of model in the statistical literature. Our aim is to develop a variable selection procedure for this model using both a classical and a Bayesian approaches. Several methods are considered for this class of model. We have adapted to the zero inflated modified power series model the Aikiake Information Criterion (AIC) and Bayesian Information Criterion (BIC) classical methods and the Conditional Predictive Ordinate (CPO), Deviance Information Criterion (DIC), Expected Aikiake Information Criterion (EAIC), Expected Bayesian Information Criterion (EBIC) and L measure Bayesian methods. Using OX and R programming language, several simulated datasets are used to illustrate the methods and to compare their performance.

Bayesian inference in a semiparametric quantile regression model

Cristian Bayes

Pontificia Universidad Católica del Perú

Hugo Agurto

Hugo Agurto Meja Facultad de Ciencias Económicas y
Empresariales, Campus Lima

Abstract

This article proposes a Semiparametric Quantile Regression Model. We use the methodology suggested by Crainiceanu, Ruppert, and Wand (2005), for a semiparametric model in the context of a quantile regression model. A Bayesian inference approach is adopted using Markov Chain MonteCarlo algorithms(MCMC). We obtain closed forms for the full conditional distributions so the Gibbs sampler algorithm can be easily implemented. A simulation study is carried out to illustrate the Bayesian approach to estimate the parameters in the model. The method developed is illustrated using a real data set.

A hierarchical Bayesian approach to human gait analysis

Amanda B. Gazon

Universidade Federal de S. Carlos, Statistics Department

Luis Ernesto B. Salasar

Universidade Federal de S. Carlos, Statistics Department

Adriano Polpo

Universidade Federal de S. Carlos, Statistics Department

Abstract

The analysis of human motion is fundamental to studies that seek to understand the altered/regular movement and to propose preventive or rehabilitation of movement disorders programs. In one of the pioneering studies in which statistically models human gait data, Olshen et al (1989) propose a trigonometric regression model with random coefficients to obtain confidence bands for the curves of the rotation angles using the bootstrap resampling method. The objective of this work is to obtain a characteristic curve of a given population and build credibility bands. We propose a hierarchical Bayesian approach and use simulation algorithms of Markov Chain Monte Carlo (MCMC) for the estimation process.

Bayesian inference on proportional elections

Gabriel Hideki Vatanabe Brunello

Genáina Nunes Rodrigues

UnB

Eduardo Yoshio Nakano

UnB

Abstract

In Brazil, the elections for president, governors and mayors use the majority system, where the candidate with the absolute majority of the votes is elected. On the proportional system, however, the absolute majority of the votes do not guarantee the election of the candidate. The proportional scenario is the kind of election that deputies (federal, state and district) as well as members of the city council are elected from. The problem of proportional elections is the difficulty to evaluate the precise number of seats (vacancy) that each party won. Since there is no guarantee that the ratio between the number of votes and the number of seats is an integer, an approximation and redistribution system must take place. Brazil defines the total number of valid votes divided by the number of seats as electoral quotient. Each party has its votes divided by the electoral quotient to obtain the party quotient. The integer part of this quotient corresponds to the number of seats reserved to the party. The remaining seats are then allocated using the DHondt method. These peculiarities of the proportional elections make the classic statistical inference not viable. However, the same inference can be easily carried out using Bayesian inference combined with Monte Carlo simulation methods. In this context, the purpose of this paper is to perform a Bayesian inference on proportional elections considering the Brazilian system of seats distribution. More specifically, we developed a methodology to allow answering the probability that a given party will have representation (at least one seat) on the deputies chamber. The inferences were made on a Bayesian scenario using the Monte Carlo simulation technique. The calculation and simulations were made using the R software. The methodology developed was applied on data from the 2010 election for Federal and State deputies chamber.

A Bayesian detection zero-inflated model

Katherine E. C. Zavaleta

Department of Statistics, Universidade Federal de S. Carlos

Vicente G. Cancho

Institute of Mathematics and Computer Science - University of São Paulo - USP

Francisco Louzada

Institute of Mathematics and Computer Science - University of São Paulo - USP

Abstract

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

Bayesian extreme-value regression model assuming known dispersion parameter

Jonatas Silva Espirito Santo

Department of Statistics, Federal University of São Carlos

Márcio Luis Lanfredi Viola

Department of Statistics, Federal University of São Carlos

Carlos Alberto Ribeiro Diniz

Department of Statistics, Federal University of São Carlos

Abstract

The extreme-value distribution (Gumbel distribution) is very useful in predicting the probability that an extremal event will occur. The use of this model is very common in insurance and finance problems as well as in natural phenomena problems such as rainfall, floods, wind gusts and air pollution. Barreto-Souza & Vasconcellos (2011) proposed an extreme-value regression model assuming that the location and dispersion parameters vary across observations through nonlinear regression models. In this work, we show Bayesian extreme-value regression model supposing known dispersion parameter. Moreover, the location parameter is defined by a linear regression model with one covariate.

Augmented mixed models for clustered proportion data

Dipankar Bandyopadhyay

Division of Biostatistics, University of Minnesota

Diana Milena Galvis Soto

UNICAMP

Victor Hugo Lachos Dvila

UNICAMP

Abstract

Continuous (clustered) proportion data often arise in various domains of medicine and public health where the response variable of interest is a proportion quantifying disease status for the cluster units, ranging between zero and one. However, due to the presence of relatively disease-free as well as highly diseased subjects in any study, the proportion values can lie in the interval $[0, 1]$. While the widely used beta, simplex and beta rectangular regressions can be adapted for assessing covariate effects here, its versatility is often challenged due to the presence/excess of zeros and ones because the support of these distributions lies in the interval $(0, 1)$. To circumvent this problem, we first introduced a unified family of distributions with support in the interval $(0, 1)$, called General Proportional Data (GPD), that contains the beta, simplex and beta rectangular densities as particular cases. Then, we augmented the probabilities of zero and one with the GPD density controlling for the clustering effect, and hence the augmented-GPD model is defined. Our approach is Bayesian with the ability to borrow information across various stages of the complex model hierarchy, and produces a computationally convenient framework amenable to available freeware. We develop Bayesian case-deletion influence diagnostics based on q -divergence measure. Both simulation studies and application to a real dataset from a clinical periodontology study quantify the gain in model fit and parameter estimation over other ad-hoc alternatives, and provide quantitative insight into assessing the true covariate effects on the proportion responses.

On the Bayesian estimation and influence diagnostics for the Weibull-Negative-Binomial regression model with cure rate under latent failure causes

Yiqi Bao

Department of Statistics - Federal University of São Carlos

Vicente Garibay Cancho

Instituto de Ciências Matemáticas e de Computação - ICMC,
Universidade de São Paulo

Francisco Louzada

Instituto de Ciências Matemáticas e de Computação - ICMC,
Universidade de São Paulo

Abstract

The purpose of this paper is to develop a Bayesian approach for the Weibull-Negative-Binomial (WNB) regression model with cure rate under latent failure causes and presence of randomized activation mechanisms. We assume the number of competing causes of the event of interest follows a Negative Binomial (NB) distribution while the latent lifetimes are assumed to follow a Weibull distribution. Markov chain Monte Carlo (MCMC) methods are used to develop the Bayesian procedure. Model selection to compare the fitted models are discussed. Moreover, we develop case deletion influence diagnostics for the joint posterior distribution based on the ψ -divergence, which has several divergence measures as particular cases. The developed procedures are illustrated with a real data set.

Bayesian approach of the exponential Poisson logarithmic model

José Augusto Fioruci

Department of Statistics - Federal University of São Carlos

Yiqi Bao

Department of Statistics - Federal University of São Carlos

Francisco Louzada

Instituto de Ciências Matemáticas e de Computação - ICMC,
Universidade de São Paulo

Vicente Garibay Cancho

Instituto de Ciências Matemáticas e de Computação - ICMC,
Universidade de São Paulo

Abstract

Recently a new three parameter lifetime distribution motivated mainly by lifetime issues has been proposed by the authors. In this paper, we consider the Bayesian analysis for this new distribution and compare their performances with the classic ones. The approximate Bayes estimators obtained by Markov chain Monte Carlo (MCMC) methods under the assumptions of non-informative priors are compared with the maximum likelihood estimators by simulation. Finally, the model is fitted to a real data set and it is compared with several models.

Influence of the spatial conguration on the efficiency of MCMC methods

Bernardo Tormink

Universidade Federal de Minas Gerais

Vinicius Diniz Mayrink

Universidade Federal de Minas Gerais

Abstract

Bayesian spatial models for areal data often consider Gaussian Markov Random Fields (GMRF) in their prior specification of parameters. This stochastic component induces correlation between elements of the model and influences computational tools required to approximate the posterior distribution. Other areas of statistics using GMRF priors are dynamic and state-space models applied to time series analysis. In this case, the results show that the number of neighbors in the spatial structure affects the computational efficiency of the MCMC methods used to sample from the posterior distributions of the target parameters. In our study, we evaluate and compare the efficiency of an MCMC algorithm applied to spatial structures with different number of neighbors per region. The efficiency is measured with respect to the autocorrelation of the chains; the larger the autocorrelation, the larger should be the generated chain so that the level of information is equivalent to that of a random sample. We consider a CAR spatial model for areal data following a local linear GMRF. The spatial structure is specified through the prior precision matrix for state parameters, which can represent different neighborhood configurations. The MCMC method is applied to evaluate the simulated data generated under distinct spatial structures; ranging from the simplest case with 2 neighbors per regions to the more general case with a large number of neighbors per regions.

Operational risk: tools for dependency modelling

Guaraci Requena

Department of Statistics - Federal University of São Carlos

Débora Gonçalves

Department of Statistics - Federal University of São Carlos

Carlos Alberto Ribeiro Diniz

Department of Statistics - Federal University of São Carlos

Abstract

The main objective of this work is the allocation of regulatory capital associated with operational risk. The identification and measurement of operational risk is a new issue in the banking sector and, for this reason, the most financial institutions use a conservative method to calculate the capital to be allocated. The method, proposed by the Basel Committee, assumes that all losses of risk units (type of loss and line of business event) have perfect positive dependence. Thus, the dependency relationships among such losses are not considered in the model, making the final allocated capital overestimated and conservative. In this paper we propose a new method for the calculation of regulatory capital required for operational risk. The method is based on copula theory and we discuss Bayesian methods to calculate the estimates of the copula parameters. A simulation study was performed to compare both the conservative and the proposed methods.

Bayesian inference of the logistic growth model in state space representation

Carlos Montenegro

IME - Universidade de São Paulo

Márcia Branco

IME - Universidade de São Paulo

Abstract

We study the logistic population growth model using state-space approach to update the knowledge of a population of marine shrimp from the Chilean coast. The unobserved state is the annual biomass of the shrimp's population and the observation is the mean annual fishing yield (mean catch per unit of effort). The observation equation is linear, with one parameter (the catchability coefficient) and the state equation is nonlinear with two parameters (carrying capacity and the intrinsic growth rate of the population). The models include normal, lognormal, t-student, skew-normal and skew-t distributions. The posterior distribution of states and parameters of the models are approximated using Markov Chain Monte Carlo methods (MCMC) and for model selection we use DIC (Deviance Information Criterion) and posterior predictive distribution. For the five adjusted models which presents the lowest DIC is the one that considers skew-t distribution for observation error and normal process error. The second best model is one that considers skew-normal distribution for observation error and normal process error. Furthermore, the posterior predictive distribution of the autocorrelation of observation errors and the χ^2 goodness of fit statistic show that the largest evidence points to these models as the best for the analyzed dataset.

Modeling the votes of Brazilian political party through a Bayesian statistical model

Rosineide F. da Paz

Department of Statistics - Federal University of São Carlos

Jorge Luis Bazan

Instituto de Ciências Matemáticas e de Computação - ICMC,
Universidade de São Paulo

Ricardo S. Ehlers

Instituto de Ciências Matemáticas e de Computação - ICMC,
Universidade de São Paulo

Abstract

Statistical modeling in Political Analysis is used recently to describe electoral behavior of political party. In this paper we propose a Weibull mixture model that describe the votes obtained for a political party in Brazilian general elections. Were considered the votes obtained by Workers' Party in the five elections since 1994 to 2010. A Bayesian approach consider with a random walk Metropolis algorithm within Gibbs sampling was implemented. In addition, Bayes factor was consider to the choose of the Number of Components of mixture. In addition the probability of get the 50 percent of the votes in the first round was estimated. The results show that the only few component are need to describe the votes obtained in this election. In addition we found that the probability of get the 50 percent of the votes in the first round is increasing to long time. Future development are discussed.

Numerical approximation to Mellin convolution by mixtures of exponentials

Jorge Luis Torrejón Matos

Department of Applied Mathematics, Institute of Mathematics and Statistics - University of São Paulo

Julio Michael Stern

Department of Applied Mathematics, Institute of Mathematics and Statistics - University of São Paulo

Abstract

The purpose of this work is to calculate the compositional models of FBST (the Full Bayesian Significance Test) studied by Stern (The rules of logic composition for the Bayesian epistemic e-Values - 2007). The objective of this work is to find an approximation method numerically efficient that can replace the condensation methods described by Kaplan. Two techniques are compared: First, the approximation of Mellin convolution using discretization and condensation described by Kaplan (An Improved Condensation Procedure in Discrete Probability Distribution Calculation - 1987), second, the approximation of Mellin convolution using mixtures of exponentials described by Dufresne (Fitting combinations of exponentials to probability distributions - 2007) to calculate the Fourier convolution and then to apply the operator described by Collins (The relationship between Fourier and Mellin transforms, with applications to probability and stochastic processes - 2011) to transform the usual convolution to Mellin convolution.

A Reversible-jump procedure for image analysis

Alexandre Calixto

Federal University of Grande Dourados and Department of
Statistics - Federal University of São Carlos

Erlandson F. Saraiva

Federal University of Mato Grosso do Sul and Department of
Statistics - Federal University of São Carlos

Luis Aparecido Milan

Department of Statistics - Federal University of São Carlos

Abstract

In this work, we propose a Bayesian approach for image analysis. In the proposed model we consider associated to each pixel a latent discrete label where the set of pixels with the same label is referred to as a component, which can consist of disjoint clusters of pixels. Conditional on latent labels, the observed data is supposed to be from a normal distribution. For latent labels is assumed a Gibbs prior distribution governed by a spatial regularization parameter b . In order to sample from the joint posterior distribution we propose a reversible-jump procedure, using a split-merge strategy that change the value of the number of state M of a Markov random field in the neighborhood $M-1$ and $M+1$. Conditional on M , the normal parameters are updated via Gibbs sampling, the b is update via Metropolis-Hastings and the latent labels via Swendsen-Wang Algorithm. We illustrate the performance of the methodology on some simulated data sets and a real data set.

A Predictive allocation sampler for Analysis of Mixture Model

Erlandson F. Saraiva

Federal University of Mato Grosso do Sul and Department of
Statistics - Federal University of São Carlos

Luis Aparecido Milan

Department of Statistics - Federal University of São Carlos

Francisco Louzada

Instituto de Ciências Matemáticas e de Computação - ICMC,
Universidade de São Paulo

Adriano K. Suzuki

Instituto de Ciências Matemáticas e de Computação - ICMC,
Universidade de São Paulo

Abstract

We propose a Bayesian approach for mixture model where the number of components is treated as unknown and inferred from the data using a Markov chain Monte Carlo algorithm, the predictive allocation sampler (PAS). The PAS is a Metropolis-Hastings within Gibbs sampling developed on a state space consisting of the number of components and the latent allocation variables. The Gibbs sampling is performed to update the latent allocation variables and the Metropolis-Hastings is performed to change the number of components using the pair of ejection/absorption moves. We verify the performance of the PAS on artificial data as well as analyze two publicly available datasets.

The transmuted log-logistic distribution: an application for time up to first calving of cows via Bayesian methods

Daniele C. T. Granzotto

Universidade Estadual de Maringá-UEM

Francisco Louzada

Instituto de Ciências Matemáticas e de Computação - ICMC,
Universidade de São Paulo

Abstract

The literature on proposing new survival distributions is growing quickly and researchers are concerned with derivation of statistical survival probability models (or simply survival distributions of real world lifetime phenomena), that can represent more consistently the random behavior of experimental observations. Several approaches for construct lifetime distributions can be mentioned such as transformation of variables, transformations of distribution, mixtures of two or more lifetime distributions, compound distributions, and probability integral transforms. According to Shaw and Buckley (2007) an alternative procedure to the above ones is the transmutation map which is a convenient way of constructing new distributions. So, we propose a new lifetime distribution by using a quadratic rank transmutation map in order to add a new parameter to the log-logistic distribution and leave it more flexible. Also, we provide a comprehensive description of the properties of the proposed distribution along with its reliability study. The usefulness of the transmuted log-logistic distribution for modeling reliability data is illustrated on a polled Tabapua race time up to rst calving data by using Bayesian methods. The results were obtained via Monte Carlo Markov Chains methods.

The Exponential-Poisson regression model for recurrent events: a Bayesian approach

Márcia Aparecida Centanin Macera
Department of Statistics, Federal University of São Carlos

Francisco Louzada
ICMC, University of São Paulo

Vicente G. Cancho
ICMC, University of São Paulo

Abstract

In this paper we introduce a new regression model for recurrent event data, in which the time of each recurrence is associated to one or multiple latent causes and no information is provided about the cause responsible for the event occurrence. This model is characterized by a rate function fully parametric and it is based on the Exponential-Poisson distribution. The time of each recurrence is then given by the minimum lifetime value among all latent causes. Inference aspects of the proposed model are discussed via Bayesian inference by using Markov Chain Monte Carlo method. A simulation study investigates the frequentist properties of a posteriori estimators for different sample sizes. A real data application demonstrates the use of the proposed model.

Bayesian analysis of the extended Marshall-Olkin model

Nikolai Kolev
IME-USP

Leandro Ferreira
IME-USP

Jayme Augusto
IME-USP

Abstract

Consider the stochastic representation $[\min(X,Z), \min(Y,Z)]$, where the non-negative random variables X and Y are dependent, but being independent of Z . This is the construction of the so-called Extended Marshall-Olkin model. We will provide a Bayesian analysis of the model defined.

Bayesian approaches to identify active factors in factorial designs 2^k without replication: some comparisons.

Jordania Oliveira
UFRN

Abstract

Designs 2^k factorial are widely used in industry and for reasons of time and economy the majority of these experiments are performed without replicas. The usual methods of analysis for these experiments have been the normal probability plot and semi-normal where, due to the subjectivity of methods, two researchers analyzing the same data may reach different conclusions. Others also well known frequentist methods are the methods of Lenth (1989) and step-down Lenth, proposed by Ye et al. (2001). However, Bayesian methods have been proposed in recent decades. The objective of this study is to compare some Bayesian methods: - Box and Meyer (1986), Chipman, Hamada and Wu (1997) and Bergquist, Vanhatalo and Nordenvaad (2011), along with the above frequentist methods. In order to identify differences and similarities between the methods and the evaluate the performance of these methods in Identifying important factors reanalyzed two experiments found in the literature and some of the simulations. The method of Box and Meyer (1986) uses the principle for factorial experiments , the scattered effects, since the methods of Chipman, Hamada and Wu (1997) and Bergquist, Vanhatalo and Nordenvaad (2011) use the three principles used in the trial, que They are: the principles of sparsity, hierarchy, and heredity. The method of Bergquist, Vanhatalo and Nordenvaad (2011) depending on the priors used can be very similar to the Normal probability plot and semi-normal. The method of Box and Meyer (1986) in turn can lead to a similar Conclusion Lenth method. The Chipman, Hamada and Wu (1997) as well as find the active factors also give an estimate of the parameters. The Bayesian methods are effective in Identifying important factors.

Bayesian Estimation of Birnbbaum-Saunders Log-linear Model

Elizabeth Gonzalez Patino
IME-USP

Abstract

The Birnbbaum-Saunders distribution was derived to model failure times of materials subjected to fluctuating stresses and strains. Motivated by applications in the characterizations of materials, in 1991 Rieck and Nedelman proposed a log-linear model for the Birnbbaum-Saunders distribution. This model has many applications, for instance, to compare the median time life of several populations or to assess the effect of covariates on accelerated life testing. In addition to the model studied under the classical approach, we considered Markov chain Monte Carlo (MCMC) and we made an implementation in **WinBUGS** to get a Bayesian approach under non-informative priori distribution. Similar results for both classic and Bayesian approaches were obtained. This implementation was also adapted for censoring and we assessed the influence of different percentages of censored data.

Discerning among symmetric models based on Bayes factor

Claudia Gallegos Perez
Banco Santander, Chile

Maria Paz Casanova Laudien
Universidad de Concepcion

Abstract

This project addresses the problem of the discrimination among probability models and it is presented in the framework of measurement errors models, assuming symmetric distributions, under the Bayesian approach. Bayes Factor and Pseudo Bayes Factor are presented as tools for statistical decision; models to be compared include Normal, Student-t and Double Exponential. Results are applied both, to an example presented in the literature, and to simulated data sets. The importance of this study is related to the need to consider non-Normal models, and to assess their adequacy, in the context of Bayesian analysis. To illustrate the methodology, a set of data modeled in the literature as Normal or Student-t, was alternatively fitted by using a Double Exponential distribution. In order to evaluate the proposed methodology, a simulation study is presented.

Simple Random Densities

Paulo C. Marques F.
IME-USP

Carlos A. de B. Pereira
IME-USP

Abstract

We construct a prior distribution over densities which is suitable for Bayesian nonparametric inference. Conditional modeling of a set of observables using this prior gives us a tractable solution to the problem of Bayesian density estimation. Stochastic simulations show a consistent asymptotic behavior of the posterior distribution. Extensions of the nonparametric model are presented, including a decision theoretic smoothing of its Bayes estimates.

Classifying the origin of archaeological fragments with Bayesian nets

Melaine Oliveira
IME - USP

Andressa Soreira
IME - USP

Victor Fossaluzza
IME - USP

Carlos Alberto de Bragança Pereira
IME - USP

Abstract

The objective of the present work is to classify an archeological fragment as coming from one of a predetermined set of possible origins, Indian tribes: Guarani, Jacadigo and Kadiweu. A database with information from ceramic fragments found in different sites, located in Lalima village, Mato Grosso do Sul, were collected. A database with information about different technological traditions on many ceramic fragments from Lalima was considered. The research's expertise provided precise classification of about 760 pieces into the above three traditions. A supervised model of classification was created in order to infer the origin, among the above three, of other set of 2100 fragments. Bayesian nets were the basis for defining our model. Bayesian networks are directed acyclic graphs that properly represent the dependency between a set of random covariates with their joint probability distribution. We consider that our approach provides a robust classification technique since it provides the probability of a new fragment was originated from one of the three archeological communities with its group of technological tradition. Also, if the probability of technological tradition indicates low probabilities for all three groups, there could be an indication of the presence of an additional community. Comparison with alternative methods to build the networks was also presented.

Reliability analysis: study of accelerated life testing under a Bayesian approach

Lilian da Costa Fontes
UFSCar

Teresa Cristina Martins Dias
UFSCar

Abstract

Currently, studies related to product reliability are very important to the industrial area, because of the rapid growth of competition and quality requirements requested by the buyers. Thus, the consumers buy not only the product, but the quality, practicality and durability, among other features. The reliability study involves the analysis of failure times obtained under normal operating conditions in order to quantify the life characteristics of a particular product, system or component. Obtaining the records of the times of failure can be very difficult. For example, delays in product failure under study. In these situations, accelerated life tests can be used to anticipate the failure of the product, increasing or decreasing the stress load of the system. In this paper, we analyze a data set of spring failures. Assuming a Weibull model for failure times of the springs, Bayesian estimates were obtained using MCMC simulation methods.

Application of Bayesian analysis to estimation of parametric distributions of complex survey data under informative sampling

Julio Stern
IME-USP

Anna Sikov
IME-USP

Abstract

In this research we develop and apply the Bayesian approach to estimation of a population model under informative probability sampling. We assume that the sample selection probabilities are known to the analyst and are defined by design variables. In most practical situations these variables are correlated with the values of the outcome variable, which results in a difference between the distribution corresponding to the units in the population and the distribution of sampled units. Following the methods, proposed in the literature we assume a model for the outcome variable, holding in the population, and expectations for the sample inclusion probabilities given the outcome and the auxiliary variables. In this research we propose to incorporate the defined models into a Bayesian model by specifying appropriate prior distributions for the unknown parameters. Using the Bayesian approach allows solving an important problem of hypothesis testing, which may be complicated under the frequentist approach due to complexity of the models, resulted by the informative sampling. In this research we also address the problem of estimation of a population model and hypothesis testing in the cases where the parametric assumptions lead to nonidentifiable sample models.

Conditional predictive inference for beta regression model with autoregressive errors

Jean Navarrete

Department of Statistics, Universidad de Concepcion

Guillermo Ferreira

Department of Statistics, Universidad de Concepcion

Mauricio Castro

Department of Statistics, Universidad de Concepcion

Mario de Castro

University of Sao Paulo

Abstract

In this paper, we developed a beta regression model, introduced by Ferrari and Cribari-Neto (2004), in order to propose a partially linear model with autoregressive errors from the Bayesian point of view. Our proposal provides a useful generalization of autoregressive beta processes, studied by Casarin et al. (2012). In this context the linear predictor of the beta regression model $g(\mu_t)$ incorporates an unknown smooth function for the auxiliary covariate t and a sequence of autoregressive errors τ_t , i.e.

$$g(\mu_t) = x_t^T \beta + f(t) + \tau_t$$

for $t = 1, \dots, T$, where x_t is a $k \times 1$ vector of non-stochastic explanatory variable values and β is a $k \times 1$ parameter vector. Furthermore, these models have a convenient hierarchical representation allowing to us an easily implementation of a Markov chain Monte Carlo (MCMC) scheme. In this paper, we propose to modify the CPO to obtain what we call the autoregressive CPO, which is computed for each new observation using only the data from previous time periods. The proposed methodology is applied to the *Cuprum* Company monthly returns.

References

- [1] Casarin, R., L. Dalla Valle, and F. Leisen (2012). Bayesian model selection for beta autoregressive processes. *Bayesian Analysis*, 7, 1-26.
- [2] Ferrari, S. and F. Cribari-Neto (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31, 799-815.

Bayes factors for comparison of restricted regression coefficients

Viviana Giampaoli
IME-USP

Carlos A. de B. Pereira
IME-USP

Heleno Bolfarine
IME-USP

Julio Singer
IME-USP

Abstract

This work compares two slopes in the regression lines that are restricted to an order constraint and to a proper subset of parameter space. Two methods of Bayes Factor calculations are discussed. The motivation is a real practical example related to the residual tax of the dental plaque index. In fact, this example originates this paper. The results indicated that the approach based in the restricted parameter space is more informative than the ones with unrestricted parameter space.

Generalized inverse Weibull distribution: the classical and Bayesian estimation

Felipe R. S. de Gusmao, Danielle L. Roges
Universidade Federal de Sao Carlos

Kleber N. N. O. Barros, Eufrazio de S. Santos
Universidade Federal Rural de Pernambuco

Roseli Aparecida Leandro, Edwin M. M. Ortega
Universidade de Sao Paulo

Abstract

The inverse Weibull distribution has the ability to model failure rates which are quite common in reliability and biological studies. A new three-parameter generalized inverse Weibull distribution with decreasing and unimodal failure rate is studied. It was produced an estimate of the parameters of the generalized inverse Weibull distribution through of the classic method and of the bayesian method. It was also done an application with agrarian data see Colosimo and Giolo et. al. (2006) and and it was also compared the results of the estimates classic and bayesian.

Regression via transform-both-sides model

Nicholas Wagner Eugenio
UFSCar

Adriano Polpo
UFSCar

Abstract

A parametric regression model for data with domain in real numbers with a linear median regression function and a transformation in both response and regression parts is presented. This model, called TBS (Transform-Both-Sides), is an adaptation of the original TBS model for survival right-censored data. It has a parameter that handles the potential data asymmetry and supposes that the errors distribution is unimodal and symmetric about zero. The density, distribution, survival and quantile functions of the model will be shown as a Bayesian study of the estimation procedure with five errors distributions (normal, double-exponential, Students t, Cauchy and logistic). These procedures are implemented in TBSSurvival, an open-source fully documented R package. The use of the package with the necessary adaptations is illustrated and the performance of the model is analyzed using simulated data sets.

Using copulas in statistical quality control

Laryssa Santos
UFSCar

Marcio Viola
UFSCar

Abstract

In many situations it is of great interest to study the dependence between two or more random variables. Therefore, it is necessary to know the function of the joint distribution of the random variables involved.

Overall, individually, modeling of random variables is easier than conjoint modeling. In the case where the random variables are independent, it is possible to obtain the combined distribution function through the marginal distribution functions. On the other hand, when such variables have dependency, the bond between the joint distribution function and its marginal distribution functions is via a function called copula.

In simple terms, a copula is a function that models the dependency between two or more random variables.

The basic theory of copulas emerged in the 1940 and 1950, with the work of Hoeffding (1940) and Sklar (1959), but remained forgotten for a few years, until it was rediscovered in the 90s. Since the work of Joe (1997) and Nelsen (2006) there has been a growing interest in the use of copulas in several areas, including: Actuary Sciences, which are used in modeling mortality and losses; Finance, used in the credit rating and risk modeling, in biomedical research, where the copula are used in the modeling of correlated events and competitive risks, and in engineering, used to control and multivariate hydrological modeling process.

In multivariate statistical process control, uses a methodology in which the variables involved in the study are modeled using multivariate Normal, according to Montgomery (2004). However, there are situations in which the variables may not have multivariate normal distribution. In this context, in particular, this work aims to use the theory of bivariate copulas in multivariate statistical control.

Promotion and Support:



Programa Interinstitucional
de Pós-graduação em Estatística
UFSCar/ICMC-USP



PÓS-GRADUAÇÃO EM
PROBABILIDADE E
ESTATÍSTICA **IME-USP**

