

Anais do I Seminário de Estatística Matemática e Probabilidade

28 a 29 de Junho de 2017

Local: Instituto de Matemática e Estatística (UFG)

Sessões Orais
Trabalhos Completos
Resumos Simples



Anais do I Seminário de Estatística Matemática e Probabilidade

O I Seminário de Estatística Matemática e Probabilidade (SEM) é uma iniciativa conjunta dos docentes e discentes que integram o Projeto de Extensão Núcleo de Apoio em Estatística (NAE) vinculado ao curso de estatística da Universidade Federal de Goiás (UFG).

Este seminário têm como objetivo principal a difusão dos principais temas de pesquisa em estatística matemática e probabilidade desenvolvidos pelos professores do curso de estatística da Universidade Federal de Goiás, assim como pelos seus graduados que atualmente.

Um segundo objetivo deste seminário é motivar os discentes de graduação da UFG a participar ativamente nos projetos de pesquisa, dos professores do Instituto de Matemática e Estatística da UFG.

Comissão Organizadora

- Prof. Dr. Mario Ernesto Piscoya Díaz (IME - UFG)
- Prof. Dr. Helton Saulo Bezerra dos Santos (EST - UNB)
- Prof. Msc. David Henriques da Matta (IME - UFG)

Comissão Científica

- Prof. Dr. Mario Ernesto Piscoya Díaz (IME - UFG)
- Prof. Dr. Helton Saulo Bezerra dos Santos (EST - UNB)
- Prof. Msc. David Henriques da Matta (IME - UFG)
- Profa. Dra. Cynthia Arantes Vieira Tojeiro (IME - UFG)
- Prof. Dr. Fabiano Fortuanto Texeira dos Santos (IME - UFG)

Corpo Editorial

- Prof. Dr. Fabiano Fortuanto Texeira dos Santos (IME - UFG)
- Prof. Dr. Tiago Moreira Vargas (IME - UFG)

Sumário

Trabalhos Completos

As distribuições Beta Burr XII e Beta Weibull Exponenciada: uma abordagem bayesiana.....	01
Discriminante não linear para dados heterogêneos de extremos.....	24
Modelos de crescimento populacional estocásticos.....	42
A new asymmetric regression model for left-censored data.....	53
Modelos Gráficos Probabilísticos: Aplicações em biologia e ciências agrárias.	58
Correção tipo-Bartlett para a estatística gradiente em modelos lineares generalizados.....	81
Dispersão como estratégia de sobrevivência.....	94
Modelos espaço-temporal funcional misto.....	111

Resumos Simples

Improved hypothesis testing in a general multivariate elliptical model.....	111
---	-----



As Distribuições Beta Burr XII e Beta Weibull Exponenciada: uma abordagem Bayesiana

Alex Felipe Rodrigues Lima^{1*},

Orientadora: Prof. Dra. Cibele Queiroz da Silva²

¹Instituto Mauro Borges de Estatística e Estudos Socioeconômicos, Estado de Goiás, Brazil

²Department of Statistics, Universidade de Brasília, Brazil

Abstract

O interesse principal do trabalho está na proposta de uma abordagem Bayesiana adequada para a estimação dos parâmetros das distribuições Beta BurrXII (BBXII) e Beta Weibull Exponenciada (BEW). Essas distribuições pertencem à classe de distribuições Beta Generalizadas (Beta-G). Sendo a distribuição Beta uma parte integrante da distribuições BBXII e da BEW, constatou-se que a sua reparametrização, proposta por Ferrari e Cribari-Neto (2004), fornece vantagens computacionais para a convergência das estimativas Bayesianas. Foram propostas duas abordagens Bayesianas para a estimação dos parâmetros da distribuição BBXII e uma abordagem para a distribuição BEW. Para a BBXII, a primeira abordagem considera *prioris* e funções geradoras de candidatos Beta, para o parâmetro μ , e Qui-Quadrado para os demais parâmetros. A segunda abordagem considera transformações *logit* para μ e log para os demais parâmetros. As *prioris* e funções geradoras de candidatos adotadas foram Beta para μ e Gama para os demais parâmetros. Nessa abordagem, obtém-se um vetor de candidatos Gaussianos de acordo com uma adaptação da proposta Gaussiana de Gray(2001). Para a BEW, considerou-se as transformações, *prioris* e funções geradoras de candidatos equivalentes à segunda proposta da BBXII, diferindo na obtenção de candidatos, que somente puderam ser obtidos de forma univariada para cada parâmetro.

Palavras-chave: Beta BurrXII, Beta Weibull Exponenciada e Abordagem Bayesiana.

*Autor para correspondência: e-mail: afelipe7@hotmail.com.

1 Introdução

A família de distribuições Beta-G (Eugene et al., 2002) tem motivado um extenso número de pesquisas no desenvolvimento de novas distribuições de probabilidade desde sua criação. Eugene et al. (2002) postularam que, para a função distribuição acumulada G de uma dada variável aleatória (em que a função de densidade de probabilidade e acumulada são denotadas por $g(x)$ e $G(x)$, respectivamente), a função distribuição acumulada beta-G é definida por

$$F(x) = \frac{1}{B(a, b)} \int_0^{G(x)} w^{(a-1)}(1-w)^{(b-1)} dw \quad (1)$$

e, conseqüentemente, sua função de densidade de probabilidade é dada por

$$f(x) = \frac{d}{dx}F(x) = \frac{g(x)}{B(a, b)}G(x)^{a-1}(1-G(x))^{b-1}, \quad (2)$$

em que $B(a, b)$ é a função Beta.

Tahir e Nadarajah (2015) fazem uma compilação dos artigos de desenvolvimento relativos a distribuição Beta-G. Esta relação inclui, entre outras, a distribuição Beta Burr XII (Paranaíba, et. al., 2011) e a distribuição Beta Weibull Exponenciada (Cordeiro et al., 2013), que foram descritas ao considerar-se, respectivamente, as funções distribuição G relativas à distribuição Burr XII (Burr, 1942) e à distribuição Weibull (Weibull, 1951), que são de interesse central nesta dissertação.

Apesar de haver uma extensa literatura sobre as distribuições na família Beta-G e sobre os procedimentos frequentistas, via máxima verossimilhança, adotados na estimação dos parâmetros dos modelos propostos, são escassos os trabalhos que utilizam o paradigma Bayesiano no processo inferencial. Uma das vantagens do uso da abordagem Bayesiana é a obtenção de intervalos de credibilidade para os parâmetros do modelo beta-G em estudo sem a necessidade do uso de argumentos assintóticos.

Neste trabalho de dissertação objetiva-se o desenvolvimento de metodologia Bayesiana adequada para a estimação dos parâmetros dos modelos Beta BurrXII (BBXII) e Beta Weibull Exponenciada (BEW). Mas, primeiramente, será apresentado as definições dos modelos BBXII e BEW, além dos procedimentos frequentistas para a estimação dos parâmetros das distribuições BBXII e BEW e as formas funcionais das respectivas verossim-

ilhanças para o caso de dados censurados com censuras ao acaso.

2 Distribuição BBXII

A distribuição BBXII foi proposta por Paranaíba et.al (2011). Ela é uma composição da distribuição Beta generalizada com a distribuição Burr XII. A primeira foi proposta por Eugene et. al. (2002), que definem uma classe de distribuições beta generalizadas (beta-G), que consiste numa generalização da distribuição Beta. Considere $G(x)$ denotando uma função de distribuição acumulada (fda) de uma variável aleatória contínua e $a > 0$ e $b > 0$ (parâmetros de forma) da distribuição beta. Tem-se que a fda da beta-G é dada por

$$F(x) = \frac{1}{B(a, b)} \int_0^{G(x)} w^{(a-1)}(1-w)^{(b-1)} dw, \quad (3)$$

em que $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ é a função *beta* e $\Gamma(\cdot)$ é a função *Gama*. Pelas propriedades de probabilidade tem-se que a função densidade de probabilidade fdp associada a $F(x)$ é dada por:

$$f(x) = \frac{d}{dx}F(x) = \frac{g(x)}{B(a, b)}G(x)^{a-1}(1-G(x))^{b-1}, \quad (4)$$

onde $g(x) = dG(x)/dx$ é função densidade da distribuição de probabilidade (fdp) a ser definida.

Dessa forma, a distribuição BBXII é obtida ao assumir $G(x)$ e $g(x)$ sendo a fda e fdp da distribuição BurrXII, respectivamente. A distribuição BurrXII foi proposta por Burr (1942), possui 3 parâmetros ($s > 0, k > 0$ e $c > 0$) e tem a seguinte forma:

$$G(x; s, k, c) = 1 - \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k}, \quad x > 0, \quad (5)$$

e

$$g(x; s, k, c) = cks^{-c} \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k-1} x^{c-1}, \quad x > 0. \quad (6)$$

Diante disso, assumindo $G(x)$ a fda da BurrXII e substituindo na fda da beta-G,

tem-se que a fda da distribuição BBXII é dada por ($x > 0$)

$$F(x) = I_{\{1-[1+(x/s)^c]^{-k}\}}(a, b) = \frac{1}{B(a, b)} \int_0^{G(x)=1-[1+(x/s)^c]^{-k}} w^{a-1}(1-w)^{b-1} dw \quad (7)$$

e sua fdp é

$$f(x) = \frac{ckx^{c-1}}{s^c B(a, b)} [1 + (x/s)^c]^{-(kb+1)} \{1 - [1 + (x/s)^c]^{-k}\}^{a-1}. \quad (8)$$

Na Figura 1 ilustra-se algumas formas possíveis da distribuição BBXII.

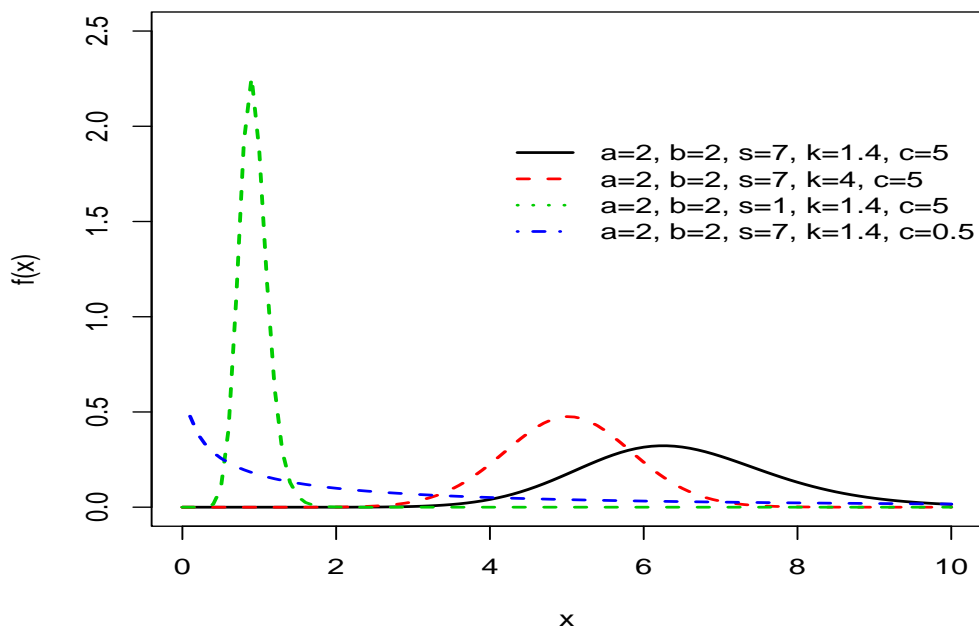


Figure 1: Gráfico da densidade BBXII

3 Distribuição BEW

A distribuição BEW foi proposto por Cordeiro et al (2011). Ela é uma composição da distribuição Weibull com a distribuição Exponencializada (ou Exponenciada) e a Beta generalizada. A primeira é uma distribuição muito conhecida e tem sido amplamente utilizada em estudos nas áreas médicas, engenharia e confiabilidade.

A distribuição Exponencializada é obtida quando a distribuição acumulada ($G(x)$) é elevada a um expoente arbitrário ($\alpha > 0$). Então a nova distribuição acumulada é dada por $F(x) = G(x)^\alpha$ e pode ser chamada de distribuição G exponencializada (ou exponenciada). Logo, a fda dessa distribuição é $f(x) = \alpha G(x)^{\alpha-1} g(x)$.

Assumindo que $G(x)$ seja a fda da distribuição Weibull, tem-se a fda da distribuição Weibull Exponencializada (EW), que foi proposta por Mudholkar e Srivastava (1993) e é dada por

$$G_{\lambda,\alpha,c}(x) = \{1 - \exp[-(\lambda x)^c]\}^\alpha, \quad x > 0, \quad (9)$$

onde $\alpha > 0$ e $c > 0$ são parâmetros de forma e $\lambda > 0$ é parâmetros de escala. Nota-se que um caso particular da distribuição EW é a distribuição Exponencial, quando $\alpha = c = 1$.

A distribuição Beta-G foi proposta por Eugene et al (2002) e foi definida em (3). Assumindo $G(x)$ dada em (9) e substituindo em (3) obtém-se a distribuição Beta Weibull Exponencializada (BEW) que possui a seguinte forma:

$$\begin{aligned} F(x) &= I_{1-(1-e^{-(\lambda x)^c})^\alpha}(a, b) \\ &= \frac{1}{B(a, b)} \int_0^{G_{\lambda,\alpha,c}(x)=\{1-\exp[-(\lambda x)^c]\}^\alpha} w^{a-1}(1-w)^{b-1} dw, w > 0, \end{aligned} \quad (10)$$

e sua fdp é dado por:

$$f(x) = \frac{\alpha c \lambda^c}{B(a, b)} x^{c-1} e^{-(\lambda x)^c} (1 - e^{-(\lambda x)^c})^{\alpha a - 1} \left\{ 1 - (1 - e^{-(\lambda x)^c})^\alpha \right\}^{b-1}. \quad (11)$$

Logo, a Função de Risco da distribuição BEW é dada por ($x > 0$):

$$h(x) = \frac{\lambda c \lambda^c x^{c-1} e^{-(\lambda x)^c} (1 - e^{-(\lambda x)^c})^{\alpha a - 1} \left\{ 1 - (1 - e^{-(\lambda x)^c})^\alpha \right\}^{b-1}}{B(a, b) I_{1-(1-e^{-(\lambda x)^c})^\alpha}}(a, b), \quad (12)$$

Na Figura 2 ilustra-se algumas formas possíveis da distribuição BEW.

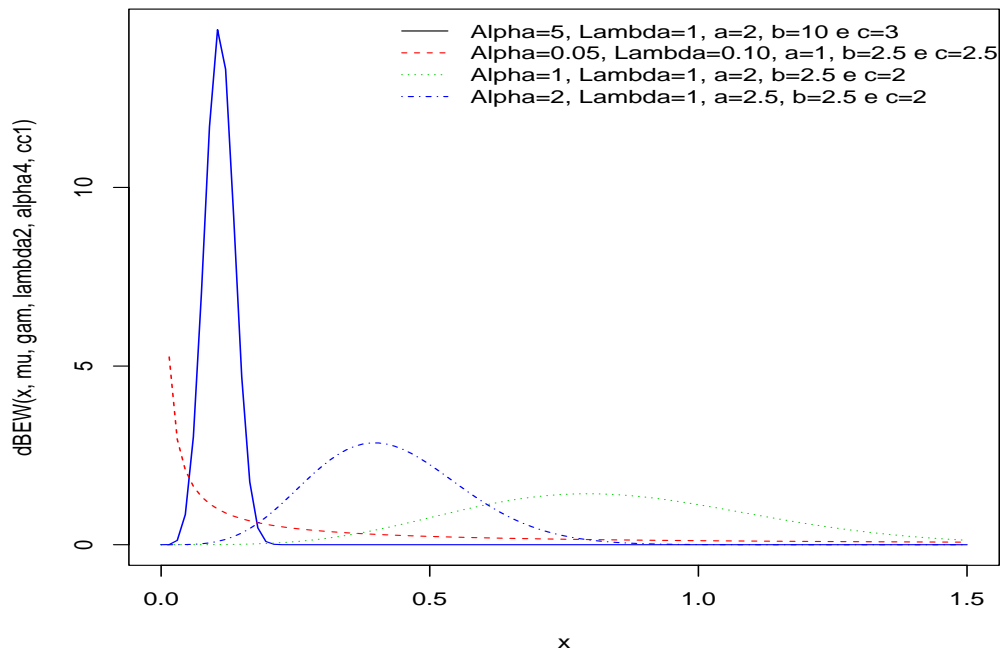


Figure 2: Gráfico da densidade BEW

4 A Abordagem Bayesiana

A seguir, apresenta-se todas as tentativas desenvolvidas para aprimorar o processo de estimação dos parâmetros da distribuição BBXII utilizando a abordagem Bayesiana. Em cada seção, apresenta-se uma descrição da metodologia a ser exposta, bem como os resultados obtidos e uma breve discussão sobre as limitações e vantagens do método em estudo.

Uma primeira implementação do problema de estimação Bayesiana da distribuição BBXII foi realizada utilizando-se os programas JAGS e WinBUGS, considerando-se *prioris Gama* independentes para todos os parâmetros do modelo. No entanto, não foi possível trabalhar com tais *prioris* devido a problemas numéricos. Então aplicou-se a trans-

formação \log nos parâmetros, na tentativa de evitar tais problemas numéricos. Com os parâmetros expressos na escala \log , trabalhou-se com *prioris* uniformes restritas ao intervalo $(-5, 5)$. O Código 1 ilustra a sintaxe de inserção dos dados no WinBUGS, a título de exemplo. Para a obtenção dos resultados, os dados foram gerados no software *R*.

Código 1: Comandos para inserção de dados no WinBUGS

```
1 Data
list(N=10, x=c( 0.7997404 ,0.7836458 ,0.6757031 ,0.8283034 ,0.7734923 ,
0.8068642 ,0.7907033,0.8234490 ,0.7720038 ,0.7909191))
```

Como já mencionamos, trabalhou-se com os parâmetros do modelo expressos na escala logarítmica, para evitar problemas numéricos. Além disso, utilizou-se *prioris* uniformes independentes. Achcar (2013) obtém estimativas bayesianas dos parâmetros de uma nova distribuição utilizando “zero’s trick” no WinBUGS e considerando *prioris* independentes uniformes.

Dentre todas as tentativas, o resultado mais coerente está apresentado na Tabela 1. Isso porque foi verificado uma grande sensibilidade na escolha dos hiperparâmetros das *prioris* uniformes. Nota-se que as estimativas dos parâmetros apresentadas na Tabela 1 não são “próximas” dos verdadeiros valores em todos os casos. Ainda foram observados problemas de autocorrelação nas estimativas.

Tabela 1: Estimativas do WinBUGS com transformação \log , *prioris* Uniformes e dados gerados com $a=16$, $b=8$, $k=6.0$, $s=1.0$, $c=5.0$

$n = 300$	Valor real	Média	Desvio Padrão	2,5%	50,0%	97,5%
a	16,0	1,702	0,864	0,584	1,502	4,013
b	8,0	10,280	5,102	2,259	9,965	19,460
c	6,0	23,760	8,017	12,780	22,300	44,510
k	5,0	0,524	0,250	0,099	0,522	0,965
s	1,0	0,835	0,024	0,800	0,831	0,889

Vale ressaltar que tentou-se implementar no *Winbugs* os cálculos de forma direta (sem o uso de transformações nos parâmetros), mas isso não foi possível na maioria dos casos. O programa indicava que os resultados não eram definidos, com a possibilidade de erros numéricos dos tipos *overflow* ou *underflow*. As razões da ocorrência de tais erros, conforme a mensagem no *software*, estão relacionados a seguir:

- Valores iniciais gerados a partir de uma distribuição a priori “vaga” podem ser numericamente extremos - especificar os valores iniciais adequados;
- Valores numericamente impossíveis, tais como registro de um número não-positivo;
- Dificuldades numéricas em amostragem. As soluções possíveis incluem: melhores valores iniciais; prioris mais informativas ou utilização de prioris uniformes, mas com a sua gama restrita a valores plausíveis; melhor parametrização para melhorar ortogonalidade; padronização de covariáveis a ter média 0 e desvio padrão 1.

Observa-se, de modo geral, que os resultados obtidos não foram satisfatórios, pois mesmo com valores iniciais semelhantes aos verdadeiros, as estimativas obtidas são totalmente diferentes dos verdadeiros, ainda com o aumento considerável do número de simulações. Isso pode estar relacionado com problemas numéricos. Além das *prioris Uniformes*, também foram investigadas *prioris Gama* para todos os parâmetros. No entanto, os resultados obtidos foram bastante insatisfatórios.

No JAGS, obedecendo a forma de geração de dados os resultados podem ser obtidos através do Código 2. Vale ressaltar que, primeiramente, deve ser instalado o pacote *runjags* no R. Utilizou-se 35.000 simulações, uma cadeia e *burn-in* de tamanho 4.000, que são as mesmas especificações utilizadas no WinBUGS.

Código 2: Comandos para inserção de dados e obtenção dos resultados no JAGS

```
require(runjags)
2 a<-16
  b<-8
4 k<-5
  s<-1
6 cc<-6
  x<-Ger.VA.BURRXII(a,b,k,s,cc,n)
8 N<-length(x)
  const <- 10000
10 z<-rep(0,N)
  dados.jags<-list('x' = x,'N' = N,'const'=const,'z'=z)
12 model <- read.jagsfile(string)
  results <- run.jags(string, n.chains=1,data=dados.jags,
14 method="rjags",sample=35000)
  summary(results$mcmc)
```

Os resultados obtidos no JAGS são apresentados na Tabela 2. Verifica-se que as estimativas não são próximas dos valores verdadeiros dos parâmetros em todos os casos, exceto para o parâmetro s .

Tabela 2: Estimativas do JAGS com Transformação \log , $prioris$ Uniformes.

$n = 300$	Valor real	Média	Desvio Padrão	2,5%	50,0%	97,5%
a	16,0	2,700	0,017	2,653	2,705	2,717
b	8,0	2,463	0,210	1,935	2,516	2,710
c	6,0	25,396	19,800	3,857	18,219	69,431
k	5,0	0,038	0,033	0,007	0,026	0,127
s	1,0	0,994	0,005	0,978	0,995	0,999

Diante das dificuldades, dos resultados insatisfatórios e de limitações dos *softwares* WinBUGS e JAGS, no que diz respeito à falta de flexibilidade para implementação de modelagens Bayesianas mais complexas, que fogem a certos padrões no uso desses programas, optou-se pela implementação de programas utilizando o software R.

Para isso utilizou-se *prioris* Gama para os parâmetros a , b , k e c e distribuição Gama inversa para o parâmetro s , uma vez que todos esses parâmetros são não-negativos. Dessa forma, as *prioris* descritas nas equações são tais que

$$a \sim \text{Gama}(a_1, b_1), \quad (13)$$

$$b \sim \text{Gama}(a_2, b_2), \quad (14)$$

$$s \sim \text{Gama} - \text{Inv}(a_3, b_3), \quad (15)$$

$$k \sim \text{Gama}(a_4, b_4), \quad (16)$$

$$c \sim \text{Gama}(a_5, b_5). \quad (17)$$

Tais *prioris* auxiliam na obtenção de propostas para a definição de funções geradoras de candidatos a serem utilizadas nos procedimentos MCMC, para cada um dos parâmetros acima.

Para tanto, a estratégia foi utilizar aproximações para as distribuições condicionais completas obtidas com as escolhas mencionadas de *prioris* que conduzam a distribuições conhecidas e fáceis de simular. Implementando-se os desenvolvimentos anteriores num algoritmo MCMC, verificou-se que as estimativas dos parâmetros k e s foram ruins, comparativamente às obtidas para os demais parâmetros, que foram, de fato, inaceitáveis.

A proposta também conduz a forte autocorrelação entre os valores gerados nas cadeias de MCMC. Os resultados são apresentados na Tabela 3

0.8

Tabela 3: Estimativas marginais a posteriori (MCMC) dos parâmetros da Distribuição BBXII com tamanho amostral 300

	Valor Real	Média	Desvio Padrão	2,5%	50,0%	97,5%	Taxa de Aceitação(%)
a	16,000	568,547	80,494	449,261	562,269	707,222	73,859
b	8,000	287,621	38,290	218,282	291,223	339,991	74,549
s	1,000	0,730	0,017	0,697	0,729	0,767	12,353
k	5,000	1,461	0,050	1,3591	1,453	1,568	4,906
c	6,000	1,195	0,095	1,015	1,193	1,375	4,226

Visando um tratamento Bayesiano mais satisfatório dos parâmetros, em especial dos parâmetros a e b associados à distribuição Beta, numa tentativa de eliminar os problemas apontados nas tentativas anteriores, propõe-se o uso de outra parametrização para a distribuição Beta. Ferrari e Cribari-Neto (2004) propõem outra parametrização da distribuição Beta, que consiste em tomar $\mu = a/(a + b)$ e $\gamma = a + b$. Dessa forma $a = \mu\gamma$ e $b = \gamma(1 - \mu)$. Então,

$$E(X) = \frac{a}{a + b} = \mu \quad (18)$$

e

$$V(X) = \frac{ab}{(a + b)^2(a + b + 1)} = \frac{\gamma(1 - \mu)}{1 + \gamma}. \quad (19)$$

Sendo assim, μ corresponde à média da variável resposta e γ é interpretado como um parâmetro de precisão, de modo que, dado μ fixo, quanto maior o valor de γ , menor será a variância de X .

Diante disso, espera-se que essa parametrização elimine a dependência nas estimativas dos parâmetros a e b que foram observadas em alguns estudos. Com essa nova parametrização, a densidade da variável aleatória Beta é dada por:

$$f_{Beta-G}(x) = \frac{1}{B(\mu\gamma, \gamma(1 - \mu))} x^{\mu\gamma-1} (1 - x)^{\gamma(1-\mu)-1}. \quad (20)$$

Nessa nova parametrização, a fdp da classe de distribuições Beta generalizada, descrita pela expressão (4) (Eugenne et. al., 2002) é expressa por:

$$f_{Beta-G}(x) = \frac{g(x)}{B(\mu\gamma, \gamma(1-\mu))} G(x)^{\mu\gamma-1} (1-G(x))^{\gamma(1-\mu)-1}, \quad (21)$$

em que $G(x)$ e $g(x)$ são descritos pelas expressões (5) e (6), respectivamente. Logo, a distribuição BBXII nessa nova parametrização, é dada por

$$f(x) = \frac{ckx^{c-1}}{s^c B(\mu\gamma, \gamma(1-\mu))} [1 + (x/s)^c]^{-(k\gamma(1-\mu)+1)} \{1 - [1 + (x/s)^c]^{-k}\}^{\mu\gamma-1}. \quad (22)$$

A princípio, as *prioris* foram definidas da seguinte maneira, exceto para o parâmetro s^\dagger ,

$$\mu \sim Beta(1, 1), \quad (23)$$

$$\gamma \sim Log - normal(0, \sigma_\gamma), \quad (24)$$

$$\log(k) \sim N(0, \sigma_k) \quad (25)$$

$$e \ c \propto \frac{1}{c}. \quad (26)$$

Definiu-se a função geradora de candidatos dos parâmetros do modelo BBXII da seguinte forma

$$P(\mu) \sim Beta(\mu^{(t-1)}\gamma^{(t-1)}, \gamma^{(t-1)}(1-\mu^{(t-1)})), \quad (27)$$

$$P(\gamma) \sim Log - Normal(\log(\gamma^{(t-1)}), \sigma_{\gamma_1}), \quad (28)$$

$$P(\log(k)) \sim Normal(\log(k^{(t-1)}), \sigma_{k_1}), \quad (29)$$

$$P(\log(c)) \sim Normal(\log(c^{(t-1)}), \sigma_{c_1}), \quad (30)$$

Implementando-se os desenvolvimentos das seções anteriores num algoritmo MCMC e com os parâmetros k e c da distribuição BBXII na transformação “log” foi realizado o estudo de simulação com $\mu = 0.66$, obtida por (18), e $\gamma = 24$, pois $\gamma = a + b$, onde $a = 16$, $b = 8$, $k = 5$ e $c = 6$ são iguais às definidas nas simulações anteriores.

Vale lembrar que, para a realização da simulação, definiu-se os valores dos hiper-

[†]Para evitar problemas numéricos, assume-se nesse caso $s = 1$

parâmetros da seguinte maneira: $\sigma_\gamma = 0.08$, $\sigma_{\gamma_1} = 2$, $\sigma_k = 3$, $\sigma_{k_1} = \sigma_{c_1} = 0.055$ e $\sigma_c = 3$. Os resultados das estimativas dos parâmetros são apresentados na Tabela 4. Nota-se que já ocorre uma “aproximação” entre as estimativas e os valores verdadeiros em todos os parâmetros. Embora, a taxa de aceitação de alguns parâmetros ainda não estejam em níveis aceitáveis.

0.8

Tabela 4: Estimativas marginais à posteriori da Distribuição BBXII

	Valor Real	Média	Desvio Padrão	2,5%	50,0%	97,5%	Taxa de Aceitação(%)
μ	0,666	0,677	0,005	0,666	0,6775	0,688	6,990
γ	24,000	21,178	1,246	18,774	21,145	23,712	61,031
k	5,000	5,730	0,314	5,140	5,718	6,391	34,436
c	6,000	6,405	0,252	5,910	6,399	6,902	26,648

Na tentativa de descrever *prioris* e funções geradoras de candidatos mais flexíveis, nessa seção as *prioris* dos parâmetros da distribuição BBXII foram definidas da seguinte maneira:

$$\mu \sim \text{Beta}(a_1, b_1), \quad (31)$$

$$\gamma \sim \text{Qui} - \text{quadrado}(a_2), \quad (32)$$

$$s \sim \text{Qui} - \text{quadrado}(a_3), \quad (33)$$

$$k \sim \text{Qui} - \text{quadrado}(a_4), \quad (34)$$

$$c \sim \text{Qui} - \text{quadrado}(a_5), \quad (35)$$

e as funções geradoras de candidatos dos parâmetros da seguinte forma

$$\mu \sim \text{Beta}(\mu^{(t-1)}\gamma^{(t-1)}, \gamma^{(t-1)}(1 - \mu^{(t-1)})), \quad (36)$$

$$\gamma \sim \text{Qui} - \text{quadrado}(\gamma^{(t-1)}), \quad (37)$$

$$k \sim \text{Qui} - \text{quadrado}(k^{(t-1)}), \quad (38)$$

$$c \sim \text{Qui} - \text{quadrado}(c^{t-1}). \quad (39)$$

A Tabela 5 apresenta os resultados das estimativas dos parâmetros. Observa-se que, para todos os parâmetros o intervalo HPD contempla o verdadeiro valor, exceto para o

parâmetro μ , embora o valor estimado deste parâmetro esteja próximo do valor verdadeiro. No entanto, o intervalo HPD para μ aparentemente exibe amplitude estreita demais, o que nos motiva a buscar outras possibilidades de efetuar a estimação Bayesiana da distribuição BBXII. Em relação a taxa de aceitação, não foi possível obter taxas em níveis entre 25% e 50% para todos os parâmetros.

0.8

Tabela 5: Estimativas marginais a posteriori da Distribuição BBXII

Parâmetro	Valor Real	Mediana	Média	HPD (2, 5%)	HPD(97, 5%)	Tx. de Aceitação(%)
μ	0,400	0,344	0,345	0,328	0,364	8,487
γ	18,000	15,070	16,050	7,010	28,163	30,470
k	5,000	5,209	5,784	2,273	10,712	7,385
c	6,000	6,782	6,859	4,797	9,102	3,471

Outra abordagem considerou a nova parametrização da distribuição BBXII e o parâmetro de escala $s=1$ fixo. Seja $\Theta = (\mu, \gamma, k, c) = (\theta_1, \theta_2, \theta_3, \theta_4)$ e admita que os θ_i 's são independentes *a priori*, com as seguintes distribuições:

$$\mu \sim \text{Beta}(a_1, b_1), \quad (40)$$

$$\gamma \sim \text{Gama}(a_2, b_2), \quad (41)$$

$$k \sim \text{Gama}(a_3, b_3), \quad (42)$$

$$c \sim \text{Gama}(a_4, b_4), \quad (43)$$

ou seja,

$$P(\mu) = \frac{\Gamma(a_1 + b_1)}{\Gamma(a_1)\Gamma(b_1)} \mu^{a_1-1} (1 - \mu)^{b_1-1}, \quad (44)$$

$$P(\gamma) = \frac{b_2^{a_2-1}}{\Gamma(a_2)} \gamma^{a_2-1} e^{-b_2\gamma}, \quad (45)$$

$$P(k) = \frac{b_3^{a_3-1}}{\Gamma(a_3)} k^{a_3-1} e^{-b_3k}, \quad (46)$$

$$P(c) = \frac{b_4^{a_4-1}}{\Gamma(a_4)} c^{a_4-1} e^{-b_4c}. \quad (47)$$

Sejam as transformações:

$$\eta_1 = \log\left(\frac{\mu}{1-\mu}\right), \quad \text{isto é,} \quad \mu = \frac{e^{\eta_1}}{1+e^{\eta_1}}, \quad (48)$$

$$\eta_2 = \log(\gamma), \quad \text{isto é,} \quad \gamma = e^{\eta_2}, \quad (49)$$

$$\eta_3 = \log(k), \quad \text{isto é,} \quad k = e^{\eta_3}, \quad (50)$$

$$\eta_4 = \log(c), \quad \text{isto é,} \quad c = e^{\eta_4}. \quad (51)$$

A distribuição a posteriori de $\eta = (\eta_1, \eta_2, \eta_3, \eta_4)$ é dada por

$$\pi_\eta(\eta_1, \eta_2, \eta_3, \eta_4|X) \propto \pi_\Theta(\theta_1, \theta_2, \theta_3, \theta_4|X) \times |J|, \quad (52)$$

em que $\theta_1 = \exp(\eta_1)/(1 + \exp(\eta_1))$, $\theta_2 = \exp(\eta_2)$, $\theta_3 = \exp(\eta_3)$ e $\theta_4 = \exp(\eta_4)$ e J é o Jacobiano da transformação.

A log-posteriori de η é dada por:

$$\begin{aligned} \log(\pi(\eta|X)) &\propto \left[\frac{e^{\eta_1+\eta_2}}{1+e^{\eta_1}} - 1 \right] \log(G(x)) + \left[\frac{e^{\eta_2}}{1+e^{\eta_1}} - 1 \right] \log(1-G(x)) + \\ &+ (e^{\eta_4} - 1) \log(x) + (e^{\eta_3} + 1) \log(1 + (x)^{e^{\eta_4}}) - \\ &- \log\left(\text{Beta}\left[\frac{e^{\eta_1+\eta_2}}{1+e^{\eta_1}}, \frac{e^{\eta_2}}{1+e^{\eta_1}}\right]\right) + a_1\eta_1 + a_2\eta_2 + (a_3 + 1)\eta_3 + \\ &+ (a_4 + 1)\eta_4 - b_2[e^{\eta_2}] - b_3[e^{\eta_3}] - b_4[e^{\eta_4}] - \\ &- (a_1 + b_1) \log(1 + e^{\eta_1}). \end{aligned} \quad (53)$$

Implementando-se esses desenvolvimentos num algoritmo MCMC e considerando os dados gerados utilizando uma amostra de tamanho $n=200$ de uma distribuição BBXII com parâmetros $\mu = 0.4$, $\gamma = 18$, $k = 5$ e $c = 6$, com $s = 1$ fixo, adotou-se um cadeia de tamanho 100.000, *burn-in* de 10.000 amostras e amostra sistemática (*gap*) com período $k=100$.

A Tabela 6 apresenta os resultados das estimativas dos parâmetros. Observa-se que, em todos os casos, o valor verdadeiro de cada parâmetro está inserido no intervalo HPD. Nota-se que a taxa de aceitação do vetor de candidatos está em níveis aceitáveis. Dessa forma, aparentemente, a proposta desta seção é superior a seção ?? (com *prioris* Chi-quadrado).

0.8

Tabela 6: Estimativas marginais à posteriori da Distribuição BBXII.

Parâmetro	Valor Real	Mediana	Média	HPD (2,5%)	HPD(97,5%)	Tx. de Aceitação(%)
μ	0,400	0,561	0,557	0,369	0,775	32,700
γ	18,000	16,607	17,943	8,298	30,901	32,700
k	5,000	6,531	7,216	2,653	15,004	32,700
c	6,000	5,424	5,467	4,243	6,914	32,700

5 Aproximações Gaussianas para os parâmetros da distribuição BEW

Considere a distribuição BEW dada em (11) e a nova parametrização da distribuição Beta-G dada em (21). Logo a distribuição da BEW sob a nova parametrização é dada por

$$f_{BEW}(x) = \frac{\alpha c \lambda^c x^{c-1} e^{-(\lambda x)^c}}{B(\mu\gamma, \gamma(1-\mu))} (1 - e^{-(\lambda x)^c})^{\alpha\mu\gamma-1} \left\{ 1 - (1 - e^{-(\lambda x)^c})^\alpha \right\}^{\gamma(1-\mu)-1}. \quad (54)$$

Tome $\Theta = (\mu, \gamma, \alpha, \lambda, c) = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)$ e admita que os θ_i 's são independentes a *priori* com as seguintes distribuições:

$$\mu \sim Beta(a_1, b_1), \quad (55)$$

$$\gamma \sim Gama(a_2, b_2), \quad (56)$$

$$\alpha \sim Gama(a_3, b_3), \quad (57)$$

$$\lambda \sim Gama(a_4, b_4), \quad (58)$$

$$c \sim Gama(a_5, b_5), \quad (59)$$

ou seja,

$$P(\mu) = \frac{\Gamma(a_1 + b_1)}{\Gamma(a_1)\Gamma(b_1)} \mu^{a_1-1} (1 - \mu)^{b_1-1}, \quad (60)$$

$$P(\gamma) = \frac{b_2^{a_2-1}}{\Gamma(a_2)} \gamma^{a_2-1} e^{-b_2 \gamma}, \quad (61)$$

$$P(\alpha) = \frac{b_3^{a_3-1}}{\Gamma(a_3)} \alpha^{a_3-1} e^{-b_3 \alpha}, \quad (62)$$

$$P(\lambda) = \frac{b_4^{a_4-1}}{\Gamma(a_4)} \lambda^{a_4-1} e^{-b_4 \lambda}, \quad (63)$$

$$P(c) = \frac{b_5^{a_5-1}}{\Gamma(a_5)} c^{a_5-1} e^{-b_5 c}. \quad (64)$$

Considere as transformações

$$\eta_1 = \log\left(\frac{\mu}{1 - \mu}\right) \quad \text{isto é,} \quad \mu = \frac{e^{\eta_1}}{1 + e^{\eta_1}}, \quad (65)$$

$$\eta_2 = \log(\gamma) \quad \text{isto é,} \quad \gamma = e^{\eta_2}, \quad (66)$$

$$\eta_3 = \log(\alpha) \quad \text{isto é,} \quad \alpha = e^{\eta_3}, \quad (67)$$

$$\eta_4 = \log(\lambda) \quad \text{isto é,} \quad \lambda = e^{\eta_4}, \quad (68)$$

$$\eta_5 = \log(c) \quad \text{isto é,} \quad c = e^{\eta_5}. \quad (69)$$

Depois de alguns desenvolvimentos a distribuição a posteriori de η é dada por:

$$\begin{aligned} \pi(\eta|X) &= \frac{e^{\eta_3} e^{\eta_5} e^{\eta_4} e^{\eta_5}}{B\left(\frac{e^{\eta_1}}{1+e^{\eta_1}}, e^{\eta_2} \left(1 - \frac{e^{\eta_1}}{1+e^{\eta_1}}\right)\right)} x^{e^{\eta_5}-1} e^{-(e^{\eta_4} x) e^{\eta_5}} \left(1 - e^{-(e^{\eta_4} x) e^{\eta_5}}\right)^{e^{\eta_3} \frac{e^{\eta_1}}{1+e^{\eta_1}} e^{\eta_2}-1} \times \\ &\times \left\{1 - \left(1 - e^{-(e^{\eta_4} x) e^{\eta_5}}\right)^{e^{\eta_3}}\right\}^{e^{\eta_2} \left(1 - \frac{e^{\eta_1}}{1+e^{\eta_1}}\right)-1} \times \left[\frac{e^{\eta_1}}{(1+e^{\eta_1})}\right]^{a_1-1} \left[\frac{1}{(1+e^{\eta_1})}\right]^{b_1-1} \times \\ &\times [e^{\eta_2}]^{a_2-1} e^{-b_2 e^{\eta_2}} \times [e^{\eta_3}]^{a_3-1} e^{-b_3 e^{\eta_3}} \times [e^{\eta_4}]^{a_4-1} e^{-b_4 e^{\eta_4}} \times [e^{\eta_5}]^{a_5-1} e^{-b_5 e^{\eta_5}} \times \\ &\times \frac{\exp(\eta_1 + \eta_2 + \eta_3 + \eta_4 + \eta_5)}{(1 + e^{\eta_1})^2}. \end{aligned} \quad (70)$$

A log-posteriori de η é dada por

$$\begin{aligned}
\log(\pi(\eta|X)) &= \eta_3 + \eta_5 + \eta_4 e^{\eta_5} - \log\left(B\left(\frac{e^{\eta_1}}{1+e^{\eta_1}} e^{\eta_2}, e^{\eta_2}\left(1 - \frac{e^{\eta_1}}{1+e^{\eta_1}}\right)\right)\right) + \\
&+ (\eta_5 - 1)\log(x) - (e^{\eta_4} x)^{e^{\eta_5}} \left(\frac{e^{\eta_3}}{1+e^{\eta_1}} e^{\eta_2} - 1\right) \log\left[1 - e^{-(e^{\eta_4} x)^{e^{\eta_5}}}\right] + \\
&+ \left(e^{\eta_2}\left(1 - \frac{e^{\eta_1}}{1+e^{\eta_1}}\right) - 1\right) \log\left\{1 - \left(1 - e^{-(e^{\eta_4} x)^{e^{\eta_5}}}\right)^{e^{\eta_3}}\right\} + \\
&+ (a_1 - 1)\log\left(\frac{e^{\eta_1}}{1+e^{\eta_1}}\right) + (b_1 - 1)\log\left[\frac{1}{1+e^{\eta_1}}\right] + (a_2 - 1)\eta_2 - b_2\eta_2 + \\
&+ (a_3 - 1)\eta_3 - b_3 e^{\eta_3} + (a_4 - 1)\eta_4 - b_4 e^{\eta_4} + (a_5 - 1)\eta_5 - b_5 e^{\eta_5} + \\
&+ \log\left[\frac{\exp(\eta_1 + \eta_2 + \eta_3 + \eta_4 + \eta_5)}{(1+e^{\eta_1})^2}\right]. \tag{71}
\end{aligned}$$

Infelizmente a implementação do processo MCMC para o vetor η não apresentou resultados satisfatórios, uma vez que não obteve-se a convergência das cadeias. Implementou-se, então, uma estratégia baseada em propostas individuais para cada coordenada de η .

5.1 Resultados

Implementando-se os desenvolvimentos num algoritmo MCMC para a distribuição BEW e realizou-se um estudo para uma amostra gerada com $\mu = 0.66$, $\gamma = 24$, $\alpha = 3$, $\lambda = 5$ e $c = 6$ de tamanho 200. Adotou-se uma cadeia de tamanho 50.000, *burn-in* de 20.000 amostras e amostras sistemáticas com período 100.

Os valores iniciais foram escolhidos da seguinte maneira: primeiramente gerou-se 100 valores para cada parâmetro. Depois estimou-se o EMV para cada conjunto e avaliou se houve ou não a convergência no ponto estimado. Dentre todos os conjuntos que tiveram convergência, o ponto inicial escolhido foi a média desses valores iniciais.

A Tabela 7 apresenta os resultados das estimativas nesse desenvolvimento. Nota-se que, exceto pelo parâmetro c , os valores verdadeiros dos parâmetros estão inseridos no intervalo HPD. Em relação a taxa de aceitação, todos os parâmetros estão em níveis aceitáveis.

0.8

Tabela 7: Estimativas marginais à posteriori da Distribuição BEW utilizando a abordagem da Seção 5.

Parâmetro	Valor Real	Mediana	Média	HPD (2,5%)	HPD(97,5%)	Tx. de Aceitação(%)
μ	0,666	0,473	0,473	0,428	0,524	24,606
γ	24,000	18,670	18,670	14,191	23,584	45,518
α	3,000	2,382	2,394	2,099	2,780	24,860
λ	5,000	4,592	4,591	4,537	4,645	23,374
c	6,000	7,896	7,929	7,097	8,859	47,154

6 Discussão

Neste trabalho de dissertação objetivou-se o desenvolvimento de métodos Bayesianos para a estimação dos parâmetros da distribuições BBXII e BEW. O desenvolvimento de metodologia adequada para tal propósito revelou-se bastante desafiador, devido a complexidade de tais modelos.

Tentativas frustradas do uso de alguns pacotes computacionais destinados à estimação Bayesiana, tais como o WinBUGS e JAGS (Capítulo 4) motivaram o desenvolvimento de nossas próprias rotinas em R. Tais pacotes não são dotados de flexibilidade suficiente para permitir a descrição de métodos Bayesianos que não seguem determinada construção padrão.

Uma vez que os modelos BBXII e BEW incorporam elementos da distribuição Beta, inicialmente trabalhou-se com os parâmetros desta última distribuição em sua forma original. No entanto, em nosso estudo constatou-se que o uso da parametrização da distribuição Beta, como descrita em Ferrari e Cribari-Neto (2004), traz vantagens computacionais e conduz à convergência das cadeias de Markov geradas via método MCMC. No entanto, as distribuições geradoras de candidatos de alguns parâmetros ainda necessitavam de maiores desenvolvimentos de modo a robustificar o método.

De modo a obter estimativas mais precisas dos parâmetros e cobertura adequada nos intervalos de credibilidade, foram desenvolvidas métodos baseados em aproximações Gaussianas para a descrição de distribuições geradoras de candidatos do parâmetros, agora expressos ou na escala log ou logit e, portanto, definidos nos reais.

Alguns estudos de simulação foram empreendidos de modo a permitir a comparação

entre as metodologias Bayesianas que foram propostas nesta dissertação e as estimativas obtidas através de estimação por máxima verossimilhança. Em linhas gerais os métodos Bayesianos e de máxima verossimilhança conduziram a estimativas pontuais muito próximas.

No entanto, as estimativas por máxima verossimilhança conduzem a intervalos de confiança estreitos em demasia, o que nos causa a suspeita de que os mesmos são irrealisticamente otimistas, implicando em pequena probabilidade de cobertura. Na prática isso significa que, frequentemente os intervalos de confiança construídos não incluirão o verdadeiro valor do parâmetro.

Além dos aspectos acima, no cálculo dos intervalos de confiança frequentistas faz-se necessário a inversão da matriz de informação observada, que, muitas vezes, não é positiva definida, impossibilitando o cálculo do desvio-padrão das estimativas, por consequência, dos respectivos intervalos de confiança.

Por outro lado, na construção dos intervalos de credibilidade, para o método Bayesiano, utiliza-se simplesmente as amostras MCMC após a convergência da cadeia, pois não se utiliza nenhum argumento assintótico, tal como no caso frequentista.

Os desenvolvimentos em que se utiliza *prioris* Qui-Quadrado para vários parâmetros, nos parece promissor, no sentido de que o método proporciona boa varredura do espaço paramétrico.

No Capítulo 7 descreveu-se nossas análises Bayesianas para o modelo BBXII utilizando os dados de melanoma de Ibrahim et al. (2001). Utilizando o mesmo conjunto de dados, Paranaíba et. al. (2011), procedeu a análises Bayesianas para o modelo BBXII utilizando *prioris* gama(0.01,0.01) para todos os parâmetros (com a e b na sua parametrização original).

Os autores utilizaram distribuições geradores de candidatos gaussianas para todos os parâmetros (que são não-negativos), tomando-se o valor absoluto dos valores gerados. No entanto, tais distribuições geradoras de candidatos gaussianas foram parametrizadas com médias fixadas nos valores estimados dos parâmetros que foram obtidos pelo método de máxima verossimilhança. Adicionalmente, para o parâmetro de dispersão adotou-se um valor extremamente pequeno, conduzindo a pequena varredura do espaço paramétrico.

Acreditamos que os métodos desenvolvidos nesta dissertação conduzem a um avanço em relação a proposta Bayesiana de Paranaíba et. al.(2011) para a distribuição BBXII. Além disso, propusemos um método Bayesiano (que acreditamos seja o primeiro) para a

distribuição BEW.

Os trabalhos futuros envolvem a execução de mais simulações e a escrita de um artigo para ser submetido a um periódico internacional.

Bibliography

- [1] ACHCAR, J. A.; COELHO-BARROS, E. A.; CORDEIRO, G. M.. *Beta generalized distributions and related exponentiated models: a Bayesian approach*. Brazilian Journal of Probability and Statistics, 27(1), 1-19,2013.
- [2] BURR,I.W.. *Cumulative frequency functions*, Annals of Mathematical Statistics, 13, 215-232, 1942.
- [3] CORDEIRO, G.M.; GOMES, A.E.; DA SILVA, C.Q.; ORTEGA, E.M.M.. *The beta exponentiated Weibull distribution*. J Stat Comput Simul, 83: 114-138, 2013.
- [4] CORDEIRO, G.M.; SIMAS, A. B. ; STOSIC, B.. *Closed Form expressions for moments of the beta Weibull distribution*. Anais da Academia Brasileira de Ciências (Impresso), v. 83, p. 357-373, 2011.
- [5] COX, D. R.; SNELL, E. J. A general definition of residuals. Journal of the Royal Statistical Society B, London, v. 30, n. 2, p. 248-254, Mar. 1968.
- [6] DENWOOD, M. J..*runjags: An R package providing interface utilities, parallel computing methods and additional distributions for MCMC models in JAGS*. Journal of Statistical Software. Retrieved from <http://cran.r-project.org/web/packages/runjags/>.
- [7] EHLERS, R. S..*Inferência Bayesiana*. Disponível em:<http://conteudo.icmc.usp.br/pessoas/ehlers/bayes/> . Acesso em: Outubro de 2016.
- [8] EUGENE, N.; LEE, C.; FAMOYE, F..*Beta-normal distribution and its applications*. Commun. Statist. - Theory and Methods, 31, 497-512, 2002.
- [9] FERRARI, S.L.P.; CRIBARI-NETO, F.. *Beta regression for modelling rates and proportions*. Journal of Applied Statistics, 31(7), 799-815, 2004.

- [10] GEWEKE, J.; *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (with discussion)*. In Bayesian Statistics. (J.M. Bernardo, J. O. Berger, A. P. Dawid e A. F. M. Smith, eds), 169-193. Oxford University Press, Oxford, 1992.
- [11] GELFAND, A.; SMITH, A.. *Sampling - based approaches to calculating marginal densities*. Journal of the American Statistical Association, (85) : 398 - 409, 1990.
- [12] GELMAN, A.; RUBIN, D. B.. *Inference from iterative simulation using multiple sequences* *Statistical Science*, 7, 457-511, 1992.
- [13] GEMAN, S., GEMAN, D.. *Stochastic relaxation, gibbs distribution and the bayesian restoration of images*. IEEE Transactions on Pattern Analysis and Machine Intelligence, (6): 721-741, 1984.
- [14] GRAY, R.. *BIO 248 cd Advanced Statistical Computing*. Disponível em: <https://www.ida.liu.se/732A26/info/report.pdf> . Acesso em: Agosto de 2016.
- [15] HATINGS, W.. *Monte Carlo sampling methods using markov chains and their applications*. Biometrika, (57):97-109, 1970.
- [16] HOFF, P. D.. *A first course in Bayesian statistical methods*. Springer Science & Business Media, 2009.
- [17] LAWLESS, J. F.. *Statistical Models and Lifetime Data*. Ed. John Wiley & Sons, 1982.
- [18] LUNN, D.; JACKSON, C.; BEST, N., THOMAS, A.; SPIEGELHALTER, D.. *The BUGS book: A practical introduction to Bayesian analysis*. Boca Raton, FL: CRC Press, 2013.
- [19] KAPLAN, E. L.; MEIER, P.: *Nonparametric estimation from incomplete observations*. J. Amer. Statist. Assn. 53:457-481, 1958.
- [20] KRUSCHKE, J.; *Doing Bayesian Data Analysis A Tutorial with R, JAGS, and Stan*, Academic Press, 2 Edição, 2014.

- [21] McCORMIK, T. H.; RAFTERY, A. E.; MADIGAN, D.; BURD, R. S.. Dynamic logistic regression and dynamic model averaging for binary classification. *Biometrics*, 68(1), 23-30, 2012.
- [22] METROPOLIS, N.; ROSENBLUTH, A., ROSENBLUTH, M., TELLER, A., TELLER, E.. *Equations of state calculations by fast computing machines*. Journal of Chemical Physics, (21): 1087-1092, 1953.
- [23] MUDHOLKAR, G.S.; SRIVASTAVA, D.K.. *Exponentiated Weibull family for analyzing bathtub failure-real data*, IEEE Trans. Reliab. 42, pp. 299–302, 1993.
- [24] NADARAJAH, S.; KOTZ, S.. *The beta Gumbel distribution*. Mathematical Problems in Engineering, New York, v.10, p.323-332, 2004.
- [25] O'HAGAN, A.. *Bayesian Inference*. Kendall's Advanced Theory of Statistics, Vol. 2B. Arnold, London, 1994.
- [26] PARANAÍBA, P. F.. *Caracterização e extensões da distribuição Burr XII: propriedades e aplicações*. 2012. Tese (Doutorado em Estatística e Experimentação Agrônômica) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2012. Disponível em: <http://www.teses.usp.br/teses/disponiveis/11/11134/tde-29102012-085146/>. Acesso em: 2016-06-07.
- [27] PARANAÍBA, P. F.; ORTEGA, E. M.; CORDEIRO, G. M.; PESCIM, R. R.. *The beta Burr XII distribution with application to lifetime data*. Computational Statistics & Data Analysis, 55(2), 1118-1136, 2011.
- [28] R Core Team. *R: A language and environment for statistical computing* [Computer software manual]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>, 2013.
- [29] SPIEGELHALTER, D.J.; THOMAS, A.. *WinBUGS Version 1.2 User Manual*. MRC Biostatistics Unit, 1999.
- [30] UGRAY, Z.; LASDON, L.; PLUMMER, J.; GLOVER, F.; KELLY, J.. *R.: Scatter Search and Local NLP Solvers: A Multistart Framework for Global Optimization*. INFORMS Journal on Computing, 19, No. 3, 328-340, 2007.



- [31] TAHIR, M.H.; NADARAJAH, S.. *Parameter induction in continuous univariate distributions: Well-established G families*. Anais da Academia Brasileira de Ciências (Annals of the Brazilian Academy of Sciences), 87(2): 539-568, 2015.
- [32] WEIBULL, W. . Wide applicability. Journal of applied mechanics, 103, 33, 1951.
- [33] ZIMMER, W. J.; KEATS, J. B.; WANG, F.K.. *The Burr XII distribution in reliability analysis*. Journal of Quality Technology, Milwaukee, v. 30, p. 389-394, 1998.

Discriminante não linear para dados heterogêneos de extremos

Evelyn de Castro Cruvinel^{1*}, Cira Etheowalda Guevara Otiniano²

¹ Instituto Mauro Borges

² Universidade de Brasília

Resumo

Este trabalho apresenta o estudo de um discriminante não linear da mistura de duas distribuições de valor extremo generalizada, conhecidas como GEV, com o parâmetro de escala comum. Como o parâmetro de forma da GEV pode assumir valor positivo, negativo ou nulo foram considerados seis casos possíveis para mistura de duas distribuições GEV. Para amostras de tamanho $n = 100$ e $n = 50$ de uma população com distribuição mistura de duas GEV são obtidos os estimadores de máxima verossimilhança por meio de simulações Monte Carlo, bem como os erros de classificação da função de discriminante. Também é apresentada uma aplicação em dados reais que ilustra a eficiência da análise discriminante do modelo de mistura de duas distribuições GEV em situação bimodal.

Palavras-chave: Mistura; GEV; Discriminante;

1 Introdução

Mistura finita de densidades é uma abordagem matemática de modelos estatísticos com ampla aplicabilidade por ser tratar de uma ferramenta extremamente flexível. Tem sido utilizada principalmente onde há heterogenidade populacional, para modelar dados constituídos por uma densidade assimétrica e para aproximação de densidades complexas, como as de aspectos multimodais.

Esses modelos são aplicados em várias áreas da estatística: análise de agrupamento, análise discriminante, análise de sobrevivência, métodos não-paramétricos ou semi-paramétricos e processamento de imagens. Também são utilizados em várias outras áreas tais como

*Autor para correspondência: e-mail: evelyncruvinel@hotmail.com

astronomia, biologia, genética, medicina, psiquiatria, economia e engenharia. Teoria e aplicações de modelos de mistura finita podem ser encontrados em, por exemplo, McLachlan e Basford (1988), McLachlan e Peel (2000) e Titterington *et al.* (1985).

A mistura de k densidades (populações) é a combinação convexa de k densidades de certa família, sendo cada densidade denominada de componente da mistura e os coeficientes chamados de pesos da mistura. O vetor de parâmetros do modelo de mistura tem no mínimo $2k - 1$ parâmetros a serem estimados, em que $k - 1$ são referentes aos pesos e os outros k são os componentes da mistura. Esses parâmetros podem ser estimados, por exemplo, por máxima verossimilhança via algoritmo EM ou por inferência bayesiana.

O modelo de misturas finitas cujas componentes são a densidade família de distribuição de Valor Extremo Generalizada (GEV) abrangem diversos fenômenos. Escalante-Sandoval (2007) utiliza misturas das componentes GEV para avaliar dados sobre a frequência de fluídos de rios de uma região do Noroeste do México, estimando os parâmetros por máxima verossimilhança. No trabalho de Otiniano e Teixeira (2014) os estimadores dos parâmetros da mistura de duas componentes GEV são obtidos via algoritmo EM, e a aplicação foi para dados de energia.

Um problema interessante relacionado aos modelos de mistura de distribuições é o estudo do discriminante, pois ele permite classificar de qual população um valor amostral pertence. O estudo de discriminante para misturas foi introduzido por O'Neill (1978), que obteve a função discriminante para mistura de normais. Outros autores trabalharam com esse assunto, entre eles pode-se citar Amoh (1984) (para mistura de duas distribuições gaussianas inversas), Mahmoud e Moustafa (1993) (para uma mistura de duas distribuições gama), Ahmad e Abd-Elrahman (1994) (para uma mistura de duas distribuições Weibull) e, recentemente, Ahmad *et al.* (2010) (para uma mistura de duas distribuições Gumbel).

Como nesses trabalhos citados, neste trabalho utiliza-se o classificador de Bayes para a análise de discriminante não linear para a mistura de distribuições GEV.

Desta maneira, esse trabalho tem o interesse de descrever os três tipos de distribuições de valor extremo e a distribuição GEV, também são apontar as relações entre a GEV e essas três distribuições. Em seguida, aborda-se os conceitos básicos de mistura finita de distribuições, a propriedade de identificabilidade e a análise discriminante.

Por fim, realizar simulações de dados de distribuições GEV com intuito de estudar o comportamento e desempenho das estimativas por máxima verossimilhança dos parâmetros e das funções discriminantes. Além disso, retratar uma aplicação para dados

reais

2 Metodologia

2.1 Distribuição de valor extremo generalizada

O aquecimento global é uma das causas de eventos climáticos extremos, tais como grandes inundações ou secas extremas, altas temperaturas, furacões e tsunamis. Por esse motivo, torna-se importante compreender o comportamento dos valores extremos de determinadas grandezas ou de determinados processos. Por exemplo, estudar os níveis máximos e mínimos de precipitação pluviométrica ao longo de um ano em uma determinada região geográfica, ou as temperaturas máximas mensais em determinada cidade, velocidades máximas de ventos durante um ano em determinada localidade ou ainda os máximos diários dos níveis de maré em determinado trecho do litoral.

Pensando nesses casos é que desenvolveu-se a teoria do valor extremo, que especificamente, trata do estudo da distribuição limite das estatísticas do máximo ou mínimo normalizadas de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.). Note que, se X_1, X_2, \dots, X_n são variáveis aleatórias i.i.d. com função de distribuição comum $F(x)$, a função de distribuição de $X_{(n)} = \max \{X_1, X_2, \dots, X_n\}$ é dada por

$$F_{X_{(n)}}(x) = P(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x) = [F(x)]^n,$$

e

$$F^n(x) \rightarrow H(x) = \begin{cases} 0, & x < x_0 \\ 1, & x \geq x_0 \end{cases}, n \rightarrow \infty,$$

sendo x_0 o menor valor tal que $F(x) = 1$.

A função $H(x)$ representa uma função de distribuição de uma variável aleatória degenerada Y , tal que $P(Y = x_0) = 1$.

Desse modo, a fim de se obter uma distribuição limite não-degenerada de $X_{(n)}$, define-se uma normalização linear da variável $X_{(n)}$:

$$X_{(n)}^* = \frac{X_{(n)} - b_n}{a_n},$$

em que a_n e b_n são seqüências de constantes apropriadas. Se existem seqüências de constantes $a_n > 0$ e b_n tais que

$$P\left(\frac{X_{(n)} - b_n}{a_n} \leq x\right) \rightarrow G(x), \text{ quando } n \rightarrow \infty,$$

em que $G(x)$ é uma função de distribuição não degenerada, então $G(x)$ pertence à família de distribuições extremais.

Foi Jenkinson (1955) que obteve a distribuição GEV na forma padrão, da maneira como é apresentado a seguir:

$$F(x; \gamma, \mu, \sigma) = \begin{cases} \exp\left\{-\left[1 + \gamma\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\gamma}}\right\}, & \text{se } \gamma \neq 0 \text{ e } 1 + \gamma\left(\frac{x-\mu}{\sigma}\right) > 0, \\ \exp\left\{-\exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right\}, & \text{se } \gamma = 0. \end{cases}$$

A distribuição GEV possui três parâmetros: μ de localização, σ de escala e γ de forma. O parâmetro γ é denominado índice caudal e tem relação com o parâmetro de forma α dada por:

- $\gamma = 0$, tem-se a Distribuição Gumbel, com $F(x; \gamma, \mu, \sigma) = \Lambda(x; \mu, \sigma)$;
- $\gamma = \alpha^{-1}$, a Distribuição Fréchet, com $F(x; \gamma, \mu, \sigma) = \Phi(\sigma + \gamma x; \frac{1}{\gamma}, \gamma\mu, \sigma)$;
- $\gamma = -\alpha^{-1}$, a Distribuição Weibull, com $F(x; \gamma, \mu, \sigma) = \Psi(-\sigma - \gamma x; \frac{-1}{\gamma}, \gamma\mu, \sigma)$.

2.2 Mistura de distribuições

Misturas finitas de densidades são modelos de probabilidade que podem ser utilizados em uma ampla variedade de fenômenos. Tem sido utilizado principalmente onde há heterogeneidade populacional, para modelar dados constituídos por uma densidade assimétrica e para aproximação de densidades complexas, por exemplo, as com aspectos multimodais.

Um modelo de mistura finita é uma combinação convexa das funções de distribuição acumuladas. Mais precisamente, se \mathcal{F} é uma família de função de distribuição acumuladas e $F_1, F_2, \dots, F_k \in \mathcal{F}$, então

$$H(x; \Theta) = \sum_{i=1}^k \pi_i F_i(x; \theta_i) \quad (1)$$

é uma mistura finita de k componentes F_1, F_2, \dots, F_k com os pesos

$$\pi_1 > 0, \pi_2 > 0, \dots, \pi_k > 0, \text{ tal que } \sum_{i=1}^k \pi_i = 1,$$

sendo que $\Theta = (\theta_1, \theta_2, \dots, \theta_k)$ é o vetor de parâmetros do modelo.

A mistura (1) também pode ser caracterizada por

$$h(x; \Theta) = \sum_{i=1}^k \pi_i f_i(x; \theta_i), \quad (2)$$

sendo f_1, f_2, \dots, f_k as densidades correspondentes às F_1, F_2, \dots, F_k .

A classe de mistura finita da família \mathcal{H} é dada por:

$$\mathcal{H} = \left\{ H : H(x; \Theta) = \sum_{i=1}^k \pi_i F_i(x; \theta_i); F_i(x; \theta_i) \in \mathcal{F} \right\}. \quad (3)$$

Neste caso, as componentes F_i podem pertencer a uma mesma família \mathcal{F} , mas em geral \mathcal{F} pode ser a união de diferentes famílias de funções de distribuição.

O modelo de misturas finitas cujas componentes são as densidades da distribuição GEV tem ampla aplicabilidade. Escalante-Sandoval (2007) utiliza misturas das componentes GEV para avaliar dados sobre a frequência de fluídos de rios de uma região do Noroeste do México, estimando os parâmetros por máxima verossimilhança. No estudo de Otiniano e Teixeira (2014) os estimadores dos parâmetros da mistura de duas componentes GEV são obtidos via algoritmo EM, e também é apresentada uma aplicação para dados reais.

Na Figura 1 tem-se exemplos da mistura de GEVs para os seguintes casos:

1. $\gamma_1 = 0$ e $\gamma_2 = 0$;
2. $\gamma_1 > 0$ e $\gamma_2 > 0$;
3. $\gamma_1 < 0$ e $\gamma_2 < 0$;
4. $\gamma_1 > 0$ e $\gamma_2 = 0$;
5. $\gamma_1 < 0$ e $\gamma_2 = 0$;
6. $\gamma_1 > 0$ e $\gamma_2 < 0$;

Figura 1: Curvas de densidade da mistura de funções GEV: (1), (2), (3), (4), (5), (6)

2.2.1 Mistura de distribuições GEV

Seja, portanto, X uma variável aleatória com função de distribuição H , mistura de duas componentes GEV, com parâmetro de escala comum, cuja a função de densidade de probabilidade (f.d.p.) é dada por:

$$h(x; \Theta) = \frac{1}{\sigma} [p_1 g_1(x; \gamma_1, \mu_1, \sigma) + p_2 g_2(x; \gamma_2, \mu_2, \sigma)], p_1 + p_2 = 1, \quad (4)$$

sendo $g_i, i = 1, 2$, a i -ésima componente da mistura, dado por

$$g_i(x; \gamma_i, \mu_i, \sigma) = \begin{cases} \left[1 + \frac{\gamma_i}{\sigma}(x - \mu_i) \right]^{-\frac{1}{\gamma_i} - 1} \exp \left\{ - \left[1 + \frac{\gamma_i}{\sigma}(x - \mu_i) \right]^{-\frac{1}{\gamma_i}} \right\}, & \gamma_i \neq 0 \\ \exp \left\{ - \left(\frac{x - \mu_i}{\sigma} \right) - \exp \left[- \left(\frac{x - \mu_i}{\sigma} \right) \right] \right\}, & \gamma_i = 0 \end{cases}, \quad (5)$$

para $1 + \frac{\gamma_i}{\sigma}(x - \mu_i) > 0$.

O modelo (4) possui seis parâmetros: $\gamma_i \in \mathbb{R}$ de forma, $\mu_i \in \mathbb{R}$ de locação, $\sigma > 0$ de escala, $i=1,2$, p_1 e p_2 são as proporções das componentes g_1 e g_2 da mistura. Dessa maneira, o vetor de parâmetros de interesse é dado por $\Theta = (p_1, \gamma_1, \mu_1, \gamma_2, \mu_2, \sigma)$.

2.3 Identificabilidade

Identificabilidade é uma propriedade relativa ao modelo que garante a unicidade dos parâmetros. Desta maneira, antes de discutir problemas de estimação é importante investigar a propriedade de identificabilidade da mistura. Em termos matemáticos, a classe de funções $\mathcal{F} = \{F(\cdot; \Theta) : F \text{ é distribuição de probabilidade}\}$ é identificável se para $F \in \mathcal{F}$, $\Theta \neq \Theta^*$ tem-se que $F(x; \Theta) \neq F(x; \Theta^*)$ para $\forall x$.

A propriedade de identificabilidade para mistura de distribuições é definida a seguir. A classe \mathcal{H} , definida (3), de todas a misturas finitas de \mathcal{F} é identificável se, e somente se,

para $H, H' \in \mathcal{H}$,

$$H = \sum_{i=1}^k \pi_i F_i, \quad H' = \sum_{j=1}^{k'} \pi'_j F'_j,$$

a igualdade $H = H'$ implica $k = k'$ e $(\pi_1, F_1), (\pi_2, F_2), \dots, (\pi_k, F_k)$ são permutações de $(\pi'_1, F'_1), (\pi'_2, F'_2), \dots, (\pi'_k, F'_k)$.

A identificabilidade da classe de mistura de distribuições GEV ainda não é conhecida, porém a identificabilidade da classe de mistura de distribuições extremais, que possui relação estreita com distribuição GEV, foi provada no trabalho de Otiniano *et al.* (2017).

2.4 Análise Discriminante

A análise discriminante é uma técnica utilizada para classificação de elementos de uma amostra ou população em grupos predefinidos, ou seja, conhecidos a priori. Esse conhecimento permite a elaboração de uma função matemática chamada regra de classificação ou discriminação, que será fundamentada na teoria de probabilidade (Mingoti, 2005).

No livro de Mingoti (2005) são citadas muitas situações nas quais o interesse está na discriminação de grupos, como na agricultura, em que é importante identificar as áreas de maior potencial para plantação de certa cultura; em marketing, na identificação de mercados potenciais e não potenciais para determinado produto; em finanças, diferenciar se uma pessoa que está pleitando um empréstimo bancário será inadimplente ou não, no ensino, discriminar os estudantes com maior potencial para concluir um determinado curso e os com menor potencial.

No entanto, este trabalho trata da análise discriminante no contexto de misturas de distribuições. O estudo de discriminante para misturas foi introduzido por O'Neill (1978), que obteve a função discriminante para mistura de normais.

2.4.1 Função discriminante para mistura de distribuições

A função discriminante de uma mistura de distribuições (2) é extremamente útil para classificar de qual população (componente) um valor amostral pertence. Suponha que há duas populações π_1 e π_2 com as correspondentes densidades $f_i(x), i = 1, 2$. Ahmad e Abd-Elrahman (1994), Ahmad *et al.* (2010) e Sultan e Al-Moisheer (2013) utilizaram a

função discriminante para uma mistura de distribuições com duas componentes, descrita a seguir:

A probabilidade do indivíduo x de origem desconhecida pertencer a π_i , $i = 1, 2$ é dada por:

$$P(x \in \pi_i) = \frac{p_i f_i(x)}{p_1 f_1(x) + p_2 f_2(x)}. \quad (6)$$

A função de probabilidade a posteriori, (6), é expressa em termos da função logística, com interesse de simplificar a função discriminante. Desta maneira, tem-se:

$$P(x \in \pi_1) = \frac{1}{1 + \exp(a - bz)} \quad (7)$$

e

$$P(x \in \pi_2) = \frac{\exp(a - bz)}{1 + \exp(a - bz)}, \quad (8)$$

sendo a , b constantes e z uma função de x . Se z for uma função linear de x tem-se uma função discriminante linear, caso z seja função não linear de x , a função discriminante é não linear.

Para definir de qual população provém um indivíduo x de origem desconhecida utiliza-se uma regra de classificação. Assim, classifica-se x em π_1 se $P(x \in \pi_1) > P(x \in \pi_2)$ ou $a - bz < 0$, (9), pois as seguintes desigualdades são equivalentes.

$$\begin{aligned} P(x \in \pi_1) &> P(x \in \pi_2) \\ \frac{1}{1 + \exp(a - bz)} &> \frac{\exp(a - bz)}{1 + \exp(a - bz)} \\ \exp(a - bz) &< 1 \\ a - bz &< 0. \end{aligned} \quad (9)$$

De forma análoga, uma observação é classificada em π_2 se $a - bz > 0$. Desse modo, a função discriminante de uma mistura de distribuições é definida por:

$$NL(x) = a - bz, \quad (10)$$

que permite classificar x em π_1 se $NL(x) < 0$ e em π_2 se $NL(x) \geq 0$. Além disso, pode-se observar que x será classificado em π_1 se $P(x \in \pi_1) > \frac{1}{2}$, pois

$$\begin{aligned} P(x \in \pi_1) > \frac{1}{2} &\Leftrightarrow \{1 + \exp\{NL(x)\}\}^{-1} > \frac{1}{2} \\ &\Leftrightarrow 1 + \exp\{NL(x)\} < 2 \\ &\Leftrightarrow \exp\{NL(x)\} < 1 \\ &\Leftrightarrow NL(x) < 0. \end{aligned}$$

Analogamente, x será classificado em π_2 se $P(x \in \pi_2) \leq \frac{1}{2}$.

2.4.2 Erros de Classificação

Todo processo de tomada de decisões traz consigo um possível erro de decisão. Portanto, é importante definir o erro de classificar uma observação em uma população quando, na realidade, ela pertence a outra.

Classifica-se uma observação x em π_1 , primeira população (componente), se a função discriminante é menor que zero, $NL(x) < 0$. Desta maneira, a probabilidade de classificar erroneamente uma observação x para π_1 , utilizando a função discriminante $NL(x)$, é dada por:

$$e_1 = P(NL(x) > 0 | x \in \pi_1).$$

Analogamente, a probabilidade classificar de forma errada uma observação x para π_2 , por meio da função discriminante $NL(x)$ é obtido por:

$$e_2 = P(NL(x) < 0 | x \in \pi_2).$$

Para caso de mistura de GEV os erros de classificação são obtidos empiricamente, pois, para calcular de forma analítica são necessárias muitas restrições.

Forma empírica

Seja X uma variável aleatória com função de densidade de probabilidade $h(x; \Theta)$ dada em (4). Com uma amostra aleatória de tamanho n de X , X_1, X_2, \dots, X_n , pode-se calcular

empiricamente e_{1k} e e_{2k} , $k = o, m, c$:

$$\widehat{e}_{1k} = \frac{1}{n} \sum_{j=1}^n I_{(NL_k(x_j) \leq 0 | x_j \in \pi_1) > 0}(x_j), \quad (11)$$

e

$$\widehat{e}_{2k} = \frac{1}{n} \sum_{j=1}^n I_{(NL_k(x_j) < 0 | x_j \in \pi_2)}(x_j). \quad (12)$$

Sendo que

$$I_A(x) = \begin{cases} 1, & \text{se } x \in A \\ 0, & \text{se } x \notin A \end{cases}. \quad (13)$$

O erro de classificação global ponderado pela proporção de mistura, \widehat{e}_k é dado por

$$\widehat{e}_k = p_1 \widehat{e}_{1k} + p_2 \widehat{e}_{2k}. \quad (14)$$

Também pode ser calculado o viés absoluto relativo amostral,

$$B(\widehat{e}_m) = \frac{|\widehat{e}_m - \widehat{e}_o|}{\widehat{e}_o}. \quad (15)$$

3 Resultados

3.1 Simulações

Com o interesse de mostrar a utilidade das técnicas propostas neste trabalho, são realizados experimentos de simulação. Os principais objetivos são:

- Testar o comportamento dos estimadores de máxima verossimilhança para dois tamanhos de amostra, $n = 50$ e 100 ;
- Investigar o desempenho da função discriminante não linear $NL_i(x)$, $i=m$ e o , comparando os casos amostral (m) e ótimo (o) por meio do erro de classificação.

Para obter os estimadores de máxima verossimilhança do modelo de mistura (4) são utilizadas amostras de tamanho $n = 50$ e $n = 100$. Como na Seção 2.2 a estimação

dos parâmetros contempla os seis casos possíveis para mistura de GEV, são escolhidos seis conjuntos de parâmetros de tal forma a considerar um conjuntos para cada caso. Na Tabela 1, são apresentados esses parâmetros. Sendo que, $\Theta_1, \Theta_2, \Theta_3, \Theta_4, \Theta_5$ e Θ_6 corresponde aos casos (1), (2), (3), (4), (5) e (6), respectivamente.

Tabela 1: Parâmetros para simulações

Θ	p_1	p_2	γ_1	γ_2	σ	μ_1	μ_2
Θ_1	0.6	0.4	0	0	1	6	8
Θ_2	0.5	0.5	0.5	1	1	1	3
Θ_3	0.4	0.6	-0.5	-0.5	1	-1	1
Θ_4	0.6	0.4	0.5	0	1	1	3
Θ_5	0.5	0.5	-0.5	0	1	2	0
Θ_6	0.4	0.6	0.5	-0.5	1	-1	1

Para a estimação utiliza-se a função `newtonRaphson(.)` do software R versão 3.3.1, de acordo com os seguintes procedimentos:

1. Gera-se uma amostra dos tamanhos $n = 100$ e $n = 50$ para cada escolha do vetor Θ .
2. A amostra aleatória da variável X cuja densidade é a mistura (5) é gerada da seguinte maneira:
 - (a) Gerar duas variáveis uniformes u_1 e u_2 no intervalo $(0, 1)$;
 - (b) Se $u_1 < p_1$, então usa-se u_2 para gerar um valor de x da variável aleatória X , sendo que $x = F_1^{-1}(u_2)$ e F_1 é a distribuição acumulada de f_1 dada por (5);
 - (c) Se $u_1 \geq p_1$, então usa-se u_2 para gerar um valor de x da variável aleatória X , sendo que $x = F_2^{-1}(u_2)$ e F_2 é a distribuição acumulada de f_2 dada por (5).
3. Calcula-se iterativamente os estimadores de Θ utilizando as derivadas parciais do logaritmo de verossimilhança em relação aos parâmetros do modelo.
4. Obtém-se estimativas de Θ utilizando 100 amostras dos tamanhos $n = 100$ e $n = 50$, então calcula-se a média e o erro do quadrado médio (EQM) de Θ por Monte Carlo.

Os resultados das estimações e do EQM são apresentados nas Tabelas 2 e 3, respectivamente.

Com base nos resultados apresentados na Tabela 2 nota-se que as médias das estimativas para amostra estão próximas aos seus verdadeiros valores. Ainda é possível verificar uma pequena tendência de sobrestimar ou subestimar os parâmetros γ_i , $i = 1, 2$. Na maior parte dos casos, a sobrestimação ocorre quando o valor de γ_i ($i = 1, 2$) é positivo, como nos casos 2, 4 e 6 (apenas primeira componente). Já a subestimação é verificada quando o valor para γ_i ($i = 1, 2$) é negativo, como nos casos 3, 5 e 6 (segunda componente).

A Tabela 3 certifica os bons resultados alcançados pelo algoritmo, pois, verifica-se que o EQM é pequeno para todos os casos. As estimativas de p_i , σ e μ_i são mais precisas do que as de γ_i por apresentarem valores de EQM menores, para $i = 1, 2$.

Também, nota-se que os valores do EQM para todos os parâmetros aumentam para amostra de tamanho menor, entretanto essas diferenças são pequenas, o que mostra que há pouca alteração no resultado para amostra menor.

Tabela 2: Média de $\hat{\Theta}$

$\hat{\Theta}$	n	\hat{p}_1	\hat{p}_2	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\sigma}$	$\hat{\mu}_1$	$\hat{\mu}_2$
$\hat{\Theta}_1$	50	0.6032	0.3968	-	-	1.0020	6.0052	8.0057
$\hat{\Theta}_1$	100	0.5976	0.4024	-	-	1.0057	5.9978	8.0126
$\hat{\Theta}_2$	50	0.5081	0.4919	0.5707	1.0626	0.9904	1.0160	3.0268
$\hat{\Theta}_2$	100	0.5007	0.4993	0.5318	1.0065	0.9936	1.0099	3.0064
$\hat{\Theta}_3$	50	0.3952	0.6048	-0.5064	-0.8226	0.9985	-1.0164	1.0018
$\hat{\Theta}_3$	100	0.3945	0.6055	-0.5047	-0.5377	0.9919	-0.9943	0.9822
$\hat{\Theta}_4$	50	0.6036	0.3964	0.5430	-	0.8045	1.0234	3.0114
$\hat{\Theta}_4$	100	0.5986	0.4014	0.5211	-	0.8251	1.0086	3.0176
$\hat{\Theta}_5$	50	0.5081	0.4919	-0.7388	-	0.9921	1.9818	0.0585
$\hat{\Theta}_5$	100	0.5019	0.4981	-0.5918	-	0.9956	1.9918	0.0269
$\hat{\Theta}_6$	50	0.4002	0.5998	0.7153	-0.6239	0.9920	-0.9607	0.9900
$\hat{\Theta}_6$	100	0.3968	0.6032	0.5153	-0.5408	0.9984	-0.9852	1.0038

Os gráficos representados na Figura (2) comparam a curva de densidade $h(x; \Theta)$ com as curvas de densidade $h(x; \hat{\Theta})$ de tamanho $n = 50$ e $n = 100$. Em cada gráfico a curva contínua representa a densidade $h(x; \Theta)$, os valores de Θ estão na Tabela (1). Já a curva tracejada em negrito representa a densidade da estimação da amostra de tamanho 100.

Tabela 3: Erro quadrático médio de $\hat{\Theta}$

$\hat{\Theta}$	n	\hat{p}_1	\hat{p}_2	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\sigma}$	$\hat{\mu}_1$	$\hat{\mu}_2$
$\hat{\Theta}_1$	50	0.0536	0.0536	-	-	0.0774	0.1749	0.1516
$\hat{\Theta}_1$	100	0.0303	0.0303	-	-	0.0618	0.1351	0.0836
$\hat{\Theta}_2$	50	0.0555	0.0555	0.1881	0.3407	0.0720	0.1362	0.0769
$\hat{\Theta}_2$	100	0.0344	0.0344	0.0945	0.0890	0.0355	0.0761	0.0322
$\hat{\Theta}_3$	50	0.0533	0.0533	0.1612	1.9412	0.0358	0.0866	0.1178
$\hat{\Theta}_3$	100	0.0363	0.0363	0.0274	0.1258	0.0278	0.0658	0.0823
$\hat{\Theta}_4$	50	0.0523	0.0523	0.2553	-	0.1124	0.1255	0.1634
$\hat{\Theta}_4$	100	0.0328	0.0328	0.1237	-	0.0966	0.0993	0.0871
$\hat{\Theta}_5$	50	0.0491	0.0491	1.0284	-	0.0608	0.1206	0.1929
$\hat{\Theta}_5$	100	0.0366	0.0366	0.3872	-	0.0420	0.0699	0.1269
$\hat{\Theta}_6$	50	0.0440	0.0440	1.7505	0.5246	0.0696	0.1632	0.1304
$\hat{\Theta}_6$	100	0.0360	0.0360	0.1684	0.0975	0.0410	0.1198	0.0680

E a curva pontilhada representa a densidade da estimação da amostra de tamanho 50, os valores estimados estão na Tabela (2).

Nota-se pela Figura (2) que no Caso 1 para ambos os tamanhos de amostra as curvas estimadas obtém um bom ajuste, assim como para o Caso 2.

Ainda verifica-se que para os casos 3, 5 e 6 a estimativa com amostra de tamanho maior, $n = 100$, apresenta um ajuste melhor do que amostra menor, $n = 50$. Uma possível explicação para esse comportamento é subestimação do parâmetro γ_2 no Caso 3 e do γ_1 no Caso 5. Já no Caso 6 houve sobrestimação do parâmetro γ_1 . Como pode ser verificado nas Tabelas (2) e (3).

Já para o Caso 4, observa-se que as curvas estimadas possuem comportamento semelhante, sendo que nas regiões modais elas se distanciam da curva de densidade dos parâmetros.

Figura 2: Gráfico das densidades $h(x; \Theta)$ (curva contínua), $h(x; \hat{\Theta})$ $n = 100$ (curva tracejada e em negrito) e $h(x; \hat{\Theta})$ $n = 50$ (curva pontilhada) para $\Theta_1, \Theta_2, \Theta_3, \Theta_4, \Theta_5$ e Θ_6 .

Agora com interesse de investigar o desempenho da função discriminante não linear $NL_i(x)$, $i=m$ e o , de acordo com o tamanho da amostra são calculados os erros de classificação. São usados dois tamanhos de amostra $n = 50$ e $n = 100$.

A Tabela 4 apresenta as estimativas dos erros de classificação \widehat{e}_{1k} , \widehat{e}_{2k} , \widehat{e}_k e o cálculo do viés absoluto relativo, $B(\widehat{e}_m)$, para cada conjuntos de parâmetros de mistura de GEV, para $k = o$ e m . Esses erros são calculados empiricamente conforme é apresentado na Seção 2.4.2, nas Equações (11), (12), (14) e (15).

Tabela 4: Erros de Classificação de acordo com as Equações (11), (12), (14) e (15).

Θ	n	\widehat{e}_{1o}	\widehat{e}_{1m}	\widehat{e}_{2o}	\widehat{e}_{2m}	\widehat{e}_o	\widehat{e}_m	$B(\widehat{e}_m)$
$\widehat{\Theta}_1$	50	0.2170	0.2187	0.1520	0.1588	0.1910	0.1950	0.0684
$\widehat{\Theta}_1$	100	0.2240	0.2256	0.1519	0.1486	0.1952	0.1946	0.0145
$\widehat{\Theta}_2$	50	0.3146	0.3194	0.0059	0.0123	0.1602	0.1683	0.1664
$\widehat{\Theta}_2$	100	0.3187	0.3211	0.0052	0.0093	0.1620	0.1654	0.0975
$\widehat{\Theta}_3$	50	0.1204	0.1146	0.1509	0.1484	0.1387	0.1350	0.0762
$\widehat{\Theta}_3$	100	0.1222	0.1270	0.1512	0.1480	0.1396	0.1397	0.0044
$\widehat{\Theta}_4$	50	0.2524	0.2439	0.1395	0.1742	0.2072	0.2163	0.1633
$\widehat{\Theta}_4$	100	0.2609	0.2505	0.1321	0.1550	0.2094	0.2121	0.0706
$\widehat{\Theta}_5$	50	0.1570	0.1543	0.2255	0.2248	0.1912	0.1890	0.0354
$\widehat{\Theta}_5$	100	0.1592	0.1593	0.2298	0.2266	0.1945	0.1929	0.0356
$\widehat{\Theta}_6$	50	0.2748	0.2636	0.0821	0.0833	0.1592	0.1555	0.0685
$\widehat{\Theta}_6$	100	0.2907	0.2949	0.0872	0.0894	0.1686	0.1710	0.0671

Na maior parte dos exemplos, tanto o erro de classificação e_{1k} , quanto o e_{2k} , $i=m$ e o , obtém-se resultados próximos para os casos ótimo e amostral. Do mesmo modo, nota-se que os erros de classificação total apresentam valores próximos para os casos ótimo e amostral, as maiores diferenças são verificadas quando se utiliza o tamanho de amostra $n = 50$, conforme mostra os resultados do cálculo dos viés absoluto relativo. Apontando que ao utilizar as estimativas dos parâmetros, apresentados na Tabela (2), para calcular os erros de classificação não houve grandes perdas.

Nota-se que no geral os erros de classificação são menores para a amostra de tamanho 50, para os erros e_{1k} , $i=m$ e o . Enquanto, na maior parte dos casos, os erros e_{2k} , $i=m$ e o , são menores na amostra de tamanho 100. Sendo que a diferença é muito pequena, demonstrando que o tamanho da amostra não tem grande influência no valor

do erro de classificação.

3.2 Aplicação

O Índice *Standard & Poor* 500 (S&P 500) é um índice de 500 ações norte-americanas de maior representatividade no mercado de capitais, pois capta cerca de 80% de cobertura de capitalização de mercado disponível. É considerado o melhor indicador individual do mercado acionário americano de grande capitalização. Fora coletado para análise, o maior índice registrado de cada mês, no período de 01/01/2010 a 31/12/2016, somando portanto 84 observações. No histograma da Figura 3.2 vê-se o comportamento bimodal nos dados. Nota-se na Figura 3.2, por meio do esboço das curvas de distribuição GEV, que os dados adequam-se ao caso da mistura de GEV com um gama positivo e outro negativos.

Figura 3: Histograma dos maiores valores mensais do índice S&P 500 e esboços das curvas de distribuição GEV em cada região modal.

As estimativas da função mistura de GEV, explicitadas na Tabela 5, concebem uma curva com características muito próximas do histograma dos dados, como pode ser visto na Figura 4. Com isso, a função estimada apresenta um ajuste muito bom, dado que seu comportamento é muito semelhante ao histograma, principalmente representando as duas regiões modais com exatidão.

Tabela 5: Estimativa e desvio padrão dos parâmetros da função mistura de GEV para os maiores valores mensais do índice S&P 500.

\hat{p}_1	\hat{p}_2	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\sigma}$	$\hat{\mu}_1$	$\hat{\mu}_2$
0,55	0,45	0,11	-0,52	136,06	1289	2011,51
0,0710	0,0710	0,1102	0,8273	13,5744	22,9349	27,7996

Figura 4: Histograma dos maiores valores mensais do índice S&P 500 e sua curva estimada

Também, foi realizado o teste de Kolmogorov-Smirnov para verificar a hipótese nula de que os dados seguem a distribuição em questão, obtendo um p-valor de 0.8466, ou seja, não rejeita-se a hipótese nula. Portanto, os dados tem distribuição da mistura de GEV estudada.

Tanto o QQ-plot quanto a distribuição acumulada da função estimada, Figura 5, apresentam bons resultados, com um bom alinhamento dos pontos em ambas situações.

Figura 5: (a) QQ-plot e (b) Função de distribuição acumulada da mistura de GEV estimada.

Nesta aplicação, como as populações não estavam previamente separadas, não é possível calcular o erro de classificação. Mas, pode-se comparar a proporção que foi classificada para cada população por meio da função discriminante com os valores estimados para os parâmetros p_1 e p_2 . Assim, há 84 observações, das quais 45 são classificadas na primeira população e 38 na segunda população, obtendo proporções de 0,55 e 0,45, respectivamente. Sendo que, $\hat{p}_1 = 0,59$ e $\hat{p}_2 = 0,41$, logo são valores próximos das proporções obtidas pela função discriminante.

Ainda, destaca-se que ao aplicar o Teste Ljung-Box, a estatística de teste rejeitou a hipótese nula, indicando que os dados podem ser autocorrelacionados.

4 Conclusões

Neste trabalho foram apresentados os métodos para estimação dos parâmetros de uma mistura de duas distribuições GEV e suas respectivas funções de discriminante. Foram analisados os comportamentos das estimativas do modelo e as estimativas da função discriminante. Além disso, como o parâmetro de forma da GEV pode assumir valor positivo, negativo ou nulo, foram analisados os seis modelos possíveis de mistura de duas distribuições GEV, quando se combina cada componente com os diferentes sinais de γ_i .

Inicialmente, o interesse foi o de comparar o comportamento dos estimadores de máxima verossimilhança para dois tamanhos de amostra. Os experimentos de simulação mostraram que as estimativas de p_i , σ_i e μ_i são mais precisas do que as de γ_i por apresentarem valores de EQM menores, para $i = 1, 2$. Também, nota-se que os valores do EQM para todos os parâmetros aumentam para amostra de tamanho menor, entretanto essas diferenças são pequenas, o que mostra que há pouca alteração no resultado para amostra menor.

Outro objetivo era investigar o desempenho da função discriminante não linear, comparando os casos amostral e ótimo por meio do erro de classificação. Com base nos

resultados dos experimentos de simulação verificou-se que, no geral, mesmo havendo diferença nas estimativas do modelo quando as amostras são de tamanho 50 e 100, os erros de classificação para esses diferentes tamanhos de amostras são bastante próximos. Demonstrando que o uso de uma amostra menor não tem grande influência no valor do erro de classificação. Ainda, na maior parte dos exemplos, o erro de classificação e_{1k} obteve resultados próximos para os casos ótimo e amostral, e o mesmo foi visto para o erro de classificação e_{2k} . Em relação aos erros de classificação global, de modo geral, as maiores diferenças foram verificadas quando se utilizou o tamanho de amostra $n=50$. Apontando que ao utilizar as estimativas dos parâmetros, apresentados na Tabela (2), para calcular os erros de classificação não houve grandes perdas.

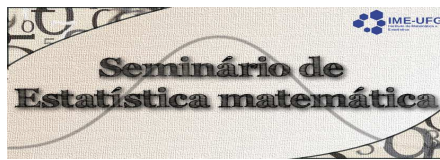
Destaca-se que houve dificuldade para a obtenção de expressões fechadas para os erros de classificação, dado que teriam que ser feitas muitas restrições. Desta maneira, optou-se por calcular os erros de classificação de forma empírica.

A aplicação em dados reais ilustraram a eficiência do modelo de mistura de duas distribuições GEV para situações bimodais.

Referências

- Ahmad, K. E.; Abd-Elrahman, A. M. Updating a nonlinear discriminant function estimated from a mixture of two weibull distributions. *Mathematical and Computer Modelling*, v. 19, n. 11, p. 41–51, 1994.
- Ahmad, K. E.; Jaheen, Z. F.; Modhesh, A. A. Estimation of a discriminant function based on small sample size from a mixture of two gumbel distributions. *Communications in Statistics - Simulation and Computation*, v. 39, n. 4, p. 713–725, 2010.
- Amoh, R. K. Estimation of a discriminant function from a mixture of two inverse gaussian distributions when the sample size is small. *Journal of Statistical Computation and Simulation*, v. 20, n. 4, p. 275–286, 1984.
- Escalante-Sandoval, C. A Mixed distribution with EV1 and GEV components for analyzing heterogeneous samples. *Ingeniería Investigación y Tecnología*, v. 8, n. 3, p. 123–133, 2007.
- Jenkinson, A. F. The frequency distribution of the annual maximum (or minimum)

- values of meteorological elements. *Quarterly Journal of the Royal Meteorology Society*, v. 87, p. 145–158, 1955.
- Mahmoud, M.A.W. and Moustafa, H.M. Estimation of a discriminant function from a mixture of two gamma distributions when the sample size is small. *Mathematical and Computer Modelling*, v. 18, n. 5, p. 87–95, 1993.
- McLachlan, G. J.; Basford, K. E. *Mixture models*. New York: Marcel Dekker, 1988.
- McLachlan, G. J.; Peel, D. *Finite mixture models*. New York: Wiley, 2000.
- Mingoti, S. A. *Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada*. Belo Horizonte: Editora UFMG, 2005.
- O’Neill, T. J. Normal discrimination with unclassified observations. *Journal of the American Statistical Association*, v. 73, p. 821–826, 1978.
- Otiniano, C. E. G. and Gonçalves, C. R. and Dorea, C. C. Y. Mixture of extreme value distributions: identifiability and estimation. *Communications in Statistics - Theory and Methods*, v. 46, n. 13, p. 6528–6542, 2017.
- Otiniano, C.E. G.; Teixeira, E.C.M. Estimação dos Parâmetros da Mistura de Duas Componentes GEV via Algoritmo EM. *Tendências em Matemática Aplicada e Computacional*, v. 15, n. 1, p. 59–71, 2014.
- Sultan, K. S.; Al-Moisheer, A.S. Estimation of a discriminant function from a mixture of two inverse weibull distributions. *Journal of Statistics Computation and Simulation*, v. 83, n. 3, p. 405–416, 2013.
- Sultan, K. S.; Ismail, M. A.; Al-Moisheer, A. S. Mixture of two inverse Weibull distributions: Properties and estimation. *Computational Statistics & Data Analysis*, v. 51, p. 5377–5387, 2007.
- Titterington, D. M.; Smith, A. M. F.; Makov, U. E. *Statistical analysis of finite mixture distributions*. New York: Wiley, 1985.



Modelos de Crescimento Populacional Estocásticos

Fabiano F. T. dos Santos^{1*}

¹Departamento de Matemática e Estatística, Universidade Federal de Goiás

Resumo

O problema de modelar fenômenos econômicos, físicos ou biológicos pode ser realizado por meio de modelos matemáticos determinísticos ou estocásticos. Neste trabalho consideraremos duas versões estocásticas do modelo geométrico, que é um problema de valor inicial associado ao crescimento populacional.

Palavras-chave: Modelos de crescimento populacional; Equações diferenciais estocásticas; Movimento Browniano.

1 Introdução

Quando se fala em modelos de crescimento populacional, podemos abordar a questão sob o ponto de vista dos modelos determinísticos ou sob a luz dos modelos estocásticos. Dentre os modelos determinísticos mais conhecidos estão os modelos exponencial, de Verhulst e de Gompertz (ver Boyce (2001)). Em cada um desses modelos matemáticos podemos introduzir aleatoriedade com o objetivo de deixar o modelo mais realista. Essa imprevisibilidade pode ser acoplada ao problema de formas diferentes; como cada modelo é um problema de valor inicial (PVI), o objeto aleatório, que pode ser uma variável aleatória ou um processo estocástico, pode estar na equação, governando-a, como um termo multiplicativo ou pode ser uma condição inicial. Dependendo de como a imprevisibilidade é colocada na equação, as técnicas para se determinar a solução podem ser muito avançadas. De uma forma geral, o objetivo deste trabalho é apresentar duas maneiras de trabalhar com modelos estocásticos. Especificamente, o modelo exponencial será transformado em um modelo estocástico, primeiramente multiplicando um termo da equação diferencial que gera o problema por uma variável aleatória contínua. Num segundo momento, a multiplicação será por um processo estocástico muito importante, o movimento Browniano.

*Autor para correspondência: e-mail: fabianoftds@yahoo.com.

Em ambos os casos a solução será apresentada e, quando for possível, suas propriedades serão exploradas. Na próxima seção, alguns modelos determinísticos clássicos e dois correspondentes estocásticos de um deles serão apresentados.

2 Metodologia

Nesta seção, os modelos determinísticos mais comuns serão apresentados, bem como suas limitações. Fixando a atenção no modelo exponencial, estudaremos duas classes de PVI estocásticos. O modelo exponencial prevê que a taxa de variação da população num instante t é proporcional ao tamanho da população neste instante. Matematicamente, se $P(t)$ é o tamanho da população no instante t , temos o PVI:

$$\frac{dP(t)}{dt} = kP(t), \quad P(0) = P_0, \quad (1)$$

onde a constante k é chamada taxa de crescimento ou declínio, se k é positiva ou negativa, respectivamente. A solução deste problema é dada por:

$$P(t) = P_0 e^{kt}, \quad t \geq 0.$$

Na Figura 1 são apresentados os gráficos de $P(t) = P_0 e^{kt}$ para valores positivos e negativos de k e $P_0 = 1$. Observe que as curvas crescem, para $k > 0$, sem limitação superior, ou seja, a população cresce indefinidamente; isso pode não ser muito realístico.

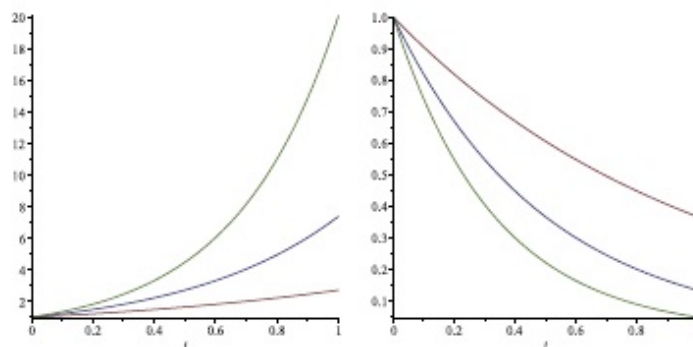


Figura 1: Gráfico de $P_0 e^{kt}$ com $P_0 = 1$ e $k > 0$ (gráficos à esquerda) e $k < 0$ (gráficos à direita)

Este modelo é razoável se considerarmos períodos de tempo limitados, pois o cres-

cimento não pode perdurar indefinidamente; em algum momento, as limitações sobre o espaço, suprimento de alimentos ou outros recursos reduzirá a taxa de crescimento e acabará inibindo o crescimento. O modelo Logístico ou de Verhulst leva em conta o fato de que a taxa de crescimento ou declínio depende da população; isto pode ser feito substituindo a constante k do modelo exponencial por uma função de P , digamos $h(P)$. Dessa forma, o modelo exponencial transforma-se em

$$\frac{dP(t)}{dt} = h(P)P(t), \quad P(0) = P_0. \quad (2)$$

Como o objetivo dessa modelagem é proporcionar a h a possibilidade de ficar próxima de $k > 0$ quando a população for pequena, decrescer quando P crescer e tornar-se negativa quando P for suficientemente grande (suprindo assim as deficiências do modelo exponencial), a escolha mais simples de uma função com essas propriedades é $h(P) = k - aP$, onde $a > 0$ também é uma constante. Posto isso, obtemos o seguinte PVI:

A constante k é chamada de taxa de crescimento intrínseco e a solução deste problema é dada por:

$$P(t) = \frac{P_0 \left(\frac{a}{k}\right)}{P_0 + \left(\frac{a}{k} - P_0\right)e^{-kt}}, \quad t \geq 0.$$

Na Figura 2, estão expostos três gráficos da solução (2) para diferentes valores de a .

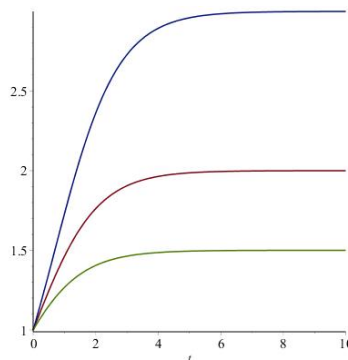


Figura 2: Gráficos de (2) com $P_0 = 1$, $k = 1$ e $a = 1; 1, 5; 2$

Observe que

$$\lim_{t \rightarrow \infty} P(t) = a,$$

ou seja, a população tende à capacidade suporte da população, nome pela qual é conhecida a constante a . O modelo de Verhulst descreve bem o crescimento de tumores sólidos quando a angiogênese ainda não ocorreu (ver Forys (2003)).

O modelo de Gompertz utiliza uma taxa de inibição da variável de estado, proporcional ao logaritmo desta variável; isto significa que a taxa de crescimento é grande no início do processo, mudando rapidamente para um crescimento mais lento. Formalmente, o PVI que governa esse modelo é:

$$\frac{dP(t)}{dt} = cP(t) \ln \left[\frac{k}{P(t)} \right], \quad P(0) = P_0, \quad (3)$$

onde $c > 0$ e $k > 0$ são constantes. A solução deste problema é dada por:

$$P(t) = ke^{\ln(\frac{P_0}{k})e^{-ct}}, \quad t \geq 0.$$

Na Figura 3 apresentamos os gráficos da solução do modelo de Gompertz para $P_0 = 1$ e $k = c$ iguais a 2,3 e 4.

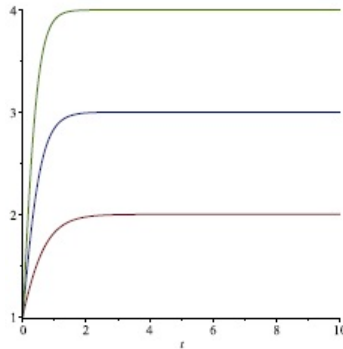


Figura 3: Gráficos de (3) com $P_0 = 1$, $k = c = 2$, $k = c = 3$ e $k = c = 4$

Note que, assim como no modelo de Verhulst,

$$\lim_{t \rightarrow \infty} P(t) = c.$$

No estudo do crescimento de tumores sólidos, a constante c faz o papel da capacidade

suporte do ambiente; ou seja, o tumor alcançará, no máximo, a quantidade c de células, com os nutrientes disponíveis (ver Enderling (2014)).

Vejamos agora duas maneiras distintas de transformar os modelos determinísticos em modelos estocásticos.

Os PVI's apresentados anteriormente podem tornar-se mais realísticos se puderem carregar alguma parcela de aleatoriedade. Isso pode ser feito substituindo-se a constante k por uma variável aleatória ou por um processo estocástico. No caso de substituirmos k por uma variável aleatória, poderemos estudar o problema utilizando a estratégia adotada em Dorini (2011). Substituindo k por um processo estocástico, a escolha natural é o movimento Browniano (Brzezniak (2005), Kannan (1979), Parzen (1962)) e se justifica por conta de toda a teoria envolvendo as equações diferenciais estocásticas (EDEs) governadas por este processo (que é um processo de segunda ordem). A teoria citada é conhecida como Cálculo de Itô (Arnold (1974), Klebaner (2005), Oksendal (1998)) e tem aplicações, por exemplo, na Economia, Física e Biologia (Klebaner (2005), Cont (2004)).

Fixemos nosso estudo no modelo exponencial (1). Precisaremos de algum aparato matemático para não deixar o trabalho muito informal. Considere um espaço de probabilidade (Ω, \mathcal{F}, P) e uma variável aleatória $K : \Omega \rightarrow \mathbb{R}$ que só assume valores positivos e que possui função de densidade f e função de distribuição acumulada F . Substituindo k por K em (1), obtemos a seguinte EDE:

$$\frac{dP(t)}{dt} = KP(t), \quad P(0) = P_0. \quad (4)$$

Fixado w , o problema (4) torna-se o problema de valor inicial determinístico

$$\frac{dP(t, w)}{dt} = K(w)P(t, w), \quad P(0, w) = P_0, \quad (5)$$

cujas soluções são dadas por

$$P(t, w) = P_0 e^{K(w)t}, \quad t \geq 0, w \in \Omega.$$

Algumas trajetórias desse processo solução estão expostas na Figura 4.

Podemos agora, calcular o tamanho médio da população no tempo t , bastando para isso determinar a função de densidade de $P(t)$, ao qual chamaremos de g . Note que se G é a função de distribuição acumulada de $P(t)$, então para $p > P_0$,

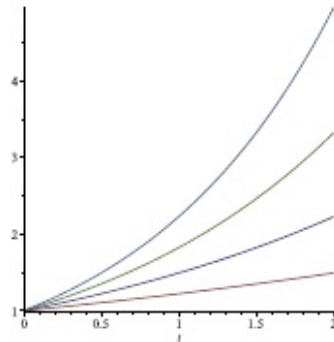


Figura 4: trajetórias do processo solução (5), com $P_0 = 1$ e $K = 0, 2; 0, 4; 0, 6; 0, 8$

$$G(p, t) = F \left(\frac{1}{t} \ln \left(\frac{p}{P_0} \right) \right),$$

e então

$$g(p, t) = \frac{P_0}{pt} f \left(\frac{1}{t} \ln \left(\frac{p}{P_0} \right) \right).$$

Posto isso, o valor esperado de $P(t)$ é dado por:

$$E[P(t)] = \frac{P_0}{t} \int_{-\infty}^{\infty} f \left(\frac{1}{t} \ln \left(\frac{p}{P_0} \right) \right) dp, \quad p > P_0.$$

Se, por exemplo, K tiver distribuição uniforme no intervalo $(0, 1)$, teremos

$$E[P(t)] = \frac{P_0^2(e^t - 1)}{t}, \quad t > 0.$$

O estudo agora poderia tomar a direção abordada em (ver Dorini (2011)) e analisar o problema (4) substituindo K por $E(K)$ e depois comparar o valor esperado da solução (5) dado por (6). O objetivo é analisar o comportamento das duas soluções em um mesmo intervalo. Se K segue o modelo uniforme no intervalo $(0, 1)$, então $E(K) = 1/2$ e assim a solução de (4) é dada por:

$$P(t) = P_0 e^{t/2}, \quad t \geq 0.$$

Os gráficos dessas duas funções são apresentados na Figura 5. Veja que as funções estão próximas no intervalo $(0, 1)$. Do ponto de vista computacional, é menos trabalhoso

estudar o problema em que substituímos $E(K)$ por K .

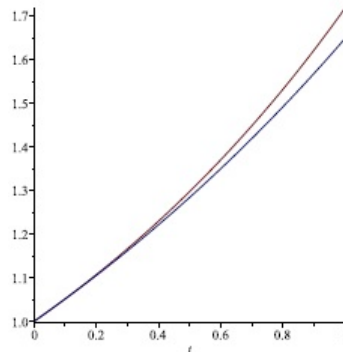


Figura 5: Gráficos de $E[P(t)]$ (em vermelho) e de $P(t)$ (em azul) com $P_0 = 1$ e $K = 1/2$.

A partir de agora, a abordagem será via Cálculo de Itô. Algumas ferramentas necessárias para se aprofundar nesse estudo são o conceito de convergência em média quadrática e o conhecimento de processos estocásticos a tempo contínuo. Aqui, fixaremos nosso estudo no movimento Browniano, mas poderíamos trabalhar com qualquer processo que tenha segundo momento finito.

Uma alternativa para introduzir aleatoriedade no modelo (1), consiste em substituir a constante K pelo movimento Browniano $B(t)$. Vale lembrar que um processo estocástico $\{B(t); t \geq 0\}$ é chamado de movimento Browniano se

- i) $B(0) = 0$ com probabilidade 1;
- ii) todo incremento $B(t) - B(s)$ é normalmente distribuído com média zero e variância $\sigma^2|t - s|$, para σ fixado;
- iii) para $0 = t_0 < t_1 < t_2 < \dots < t_n < \infty$, os incrementos $B(t_i) - B(t_{i-1})$; $1 \leq i \leq n$, são independentes e distribuídos como em (ii).

Posto isso, o modelo exponencial estocástico apresenta-se da forma

$$\frac{dP(t)}{dt} = B(t)P(t), \quad P(0) = P_0. \quad (6)$$

A solução deste problema de valor inicial pode ser encontrada utilizando-se a consagrada fórmula de Itô (Klebaner (2005), Oksendal (1998), Protter (2005)). O processo solução é dado por

$$P(t) = P_0 e^{-t/2+B(t)}, \quad t \geq 0.$$

Este processo estocástico é conhecido como movimento Browniano geométrico e tem aplicações importantes na Economia (Cont (2004), Klebaner (2005)). A Figura 6 apresenta trajetórias típicas desse processo.

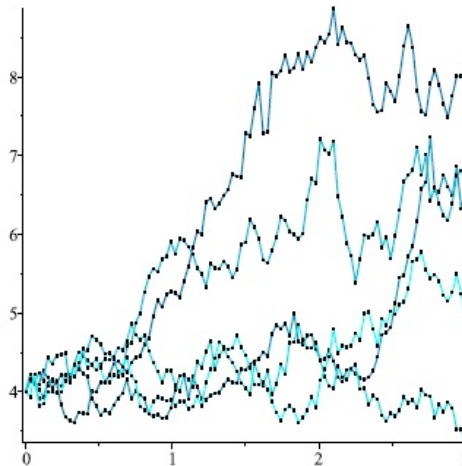


Figura 6: Trajetórias de $P(t) = P_0 e^{-t/2+B(t)}$ com $P_0 = 4$

Na Figura 7 vemos duas populações que crescem segundo trajetórias que não são suaves; isto ampara o argumento de que trajetórias como as do movimento Browniano geométrico se ajustam melhor do que aquelas provenientes de modelos determinísticos.

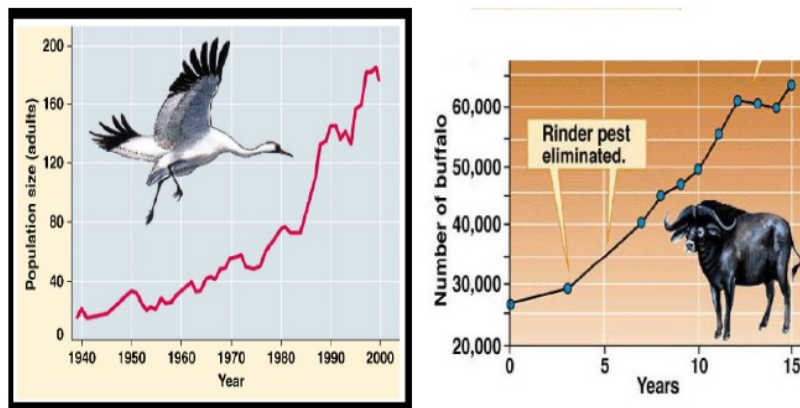


Figura 7: Imagens obtidas em <http://sky.scnu.edu.cn/life/class/ecology/chapter/Chapter11.htm>

Pode-se mostrar que a solução (9) goza das seguintes propriedades:

- i) $E[P(t)] = E(P_0)$, desde que P_0 e $B(t)$ sejam independentes.
- ii) $P(t) \rightarrow 0$ quando $t \rightarrow \infty$, quase certamente.

A propriedade (i) significa que o valor esperado do processo solução é o valor esperado da variável aleatória que é a condição inicial do PVI (8). No estudo do crescimento populacional $P(0) = P_0$ é uma constante e é igual ao tamanho da população no instante $t = 0$, logo $E[P(0)] = E(P_0) = P_0$. A propriedade (ii) significa que para t suficientemente grande, a população se extinguirá com probabilidade 1. Em geral, a EDE:

$$\frac{dP(t)}{dt} = rP(t)dt + \alpha P(t)dB(t), \quad P(0) = P_0, \quad (7)$$

onde r e α são constantes, é tal que sua solução é dada por:

$$P(t) = P_0 e^{\left(r - \frac{\alpha^2}{2}\right)t + \alpha B(t)}, \quad t \geq 0.$$

O modelo exponencial estocástico apresentado anteriormente pode ser generalizado, substituindo-se o movimento Browniano por um processo de Lévy ou uma semimartingale, porém, a teoria de Itô poderá não ser aplicada (se o processo envolvido não tiver segundo momento finito) e, nesse caso, a determinação de soluções para a equação torna-se mais árdua (Cont (2004), Klebaner (2005), Protter (2005), Santos (2011)).

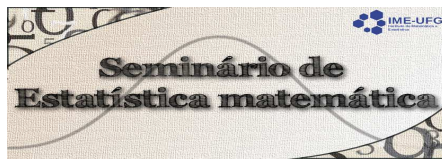
3 Conclusões

Os modelos estocásticos são mais realistas do que os modelos determinísticos, pois a imprevisibilidade está presente nos fenômenos do mundo real. A literatura a respeito das aplicações de tais modelos na Demografia, Economia, Física, Biologia e Medicina é riquíssima e o aparato matemático necessário para entender tais modelos torna-se tão sofisticado quanto mais realista for a equação que governa o fenômeno associado.

Outros aspectos importantes associados a tais modelos são a estimação de parâmetros relacionados às EDEs e o estudo de equações diferenciais parciais estocásticas (Ferrante (2000), Holden (2010)).

Referências

- ARNOLD, L. Stochastic Differential Equations: Theory and Applications. John Wiley Sons, New York, 1974.
- BOYCE, W. E. Equações Diferenciais Elementares e Problemas de Valores de Contorno. LTC, Rio de Janeiro, 2001.
- BRZEZNIAK, Z. Basic Stochastic Processes. Springer, UK, 2005.
- CONT, R. Financial Modelling with Jump Processes. Chapman Hall/CRC, Florida, 2004.
- DORINI, F. A.; CUNHA, M. C. C.; OLIVEIRA, S. P. Soluções de Problemas Envolvendo Equações Diferenciais Sujeitas a Incertezas. TEMA, v. 12, n. 2, p. 111-123, 2011.
- ENDERLING, H.; CHAPLAIN, M. A. J. Mathematical Modeling of Tumor Growth and Treatment. Current Pharmaceutical Design, v. 20, p. 1-7, 2014.
- FERRANTE, L. et al. Parameter Estimation in a Gompertzian Stochastic Model for Tumor Growth. Biometrics, v. 56, n. 4, p. 1076-1081, 2000.
- FORYS, U.; CZOCHRA, A. M. Logistic Equations in Tumour Growth Modelling. Int. J. Appl. Math. Comput. Sci, v. 13, n. 3, p. 317-325, 2003.
- HOLDEN, H. et al. Stochastic Partial Differential Equations - A Modeling, White Noise Functional Approach. Springer, New York, 2010.
- KANNAN, D. An introduction to Stochastic Processes. Elsevier North Holland, New York, 1979.
- KLEBANER, F. C. Introduction to Stochastic Calculus with Applications. Imperial College Press, London, 2005.
- OKSENDAL, B. Stochastic Differential Equations An Introduction with Applications. Springer-Verlag, London, 1998.
- PARZEN, E. Stochastic Processes. Holden-Day, USA, 1962.



PROTTER, P. E. Stochastic Integration and Differential Equations. Springer-Verlag, Berlin, 2005.

SANTOS, F. F. T. Classes de Soluções para a Equação de Langevin Generalizada. Tese (Doutorado) - UnB, Brasil, Brasília, BR, 2011.

A new asymmetric regression model for left-censored data

Mário F. Desousa¹, Helton Saulo^{2*}, Manoel Santos-Neto³, Víctor Leiva⁴

¹Department of Statistics, Universidade Estadual de Campinas, Brazil

²Department of Statistics, Universidade de Brasília, Brazil

³Department of Statistics, Universidade Federal de Campina Grande, Brazil

⁴School of Industrial Engineering, Pontificia Universidad Católica de Valparaíso, Chile

Abstract

Most of the models designed for modeling censored data rely on the assumption of normality for the error distribution. It is well known that not all applications are well modeled by this distribution. Some efforts have relaxed the normality assumption by considering more flexible distributions such as t and log-alpha-power. Nevertheless, these models do not consider partial observations from the assumed distribution which potentially leads to biased inference. We have explored a real data example of measles vaccine in Haiti and confirmed both the possibility of partial observation and asymmetry problems. Then, to solve such problems, we propose a mixture model consisting of the Birnbaum-Saunders and Bernoulli distributions. We discuss estimation of the model parameters based on the maximum likelihood method. We then carry out a Monte Carlo simulation study to evaluate the performance of the maximum likelihood estimators. We use the R software in all computations and the results favor the proposed methodology.

Palavras-chave: Mixture model; Censoring; Birnbaum-Saunders Distribution.

1 Introduction

The determination of antibody concentration by quantitative assays is a very important topic of research, because there is always a concentration value (\mathcal{T}) below which an exact measurement cannot be obtained regardless of the employed technique. Nevertheless, this antibody concentration value (\mathcal{T}) is a function of the associated assay. When

*Autor para correspondência: e-mail: heltonsaulo@gmail.com.

left-censoring is present in data from an assay, the lower detection limit (LDL) can be used to substitute a value for the censored observation, namely, the value \mathcal{T} . In special, this substitution is applied in a safety and immunogenicity study of measles vaccine in Haiti presented by Moulton and Halsey (1995), an example explored in this paper.

A two-part model for the situation of zero excess was considered by Cragg (1971). The Cragg model considers the possibility of having observations from the assumed distribution f and from the point mass distribution. In this model, the log-normal distribution was considered for the independent variable. The Cragg model, however, does not consider the existence of both a lower limit and some observations below this limit. Moulton and Halsey (1995) proposed a straightforward generalization of the two-part model, called Bernoulli/lognormal model, by considering the possibility of some limiting responses resulting from interval censoring associated with f . The generalized two-part model allows the possibility of a observation i , if located below \mathcal{T} to be either a partial observation from f or a realization of the point mass distribution.

The main objective of this paper is to propose a regression model for censored data based on the mixture between the Birnbaum-Saunders (BS) and Bernoulli distributions, that is, a censored continuous distribution and a point mass distribution located below the detection limit. The Birnbaum-Saunders (BS) distribution is positively skewed and has a failure rate with upside-down bathtub shape and a close relation with the normal distribution; see Birnbaum and Saunders (1969) and Johnson et al. (1995). The proposed model extends to the BS case the Moulton and Halsey (1995)'s Bernoulli/lognormal model.

2 The Bernoulli/BS mixture model

2.1 Formulation

We propose a mixture model between the Bernoulli and BS distributions (Bernoulli/BS), that is,

$$g(y_i) = [p + (1 - p) \Phi(\zeta_{i2}^c)] I_i + (1 - p) \left[\frac{c_1}{\alpha} \cosh\left(\frac{y - \mu_1}{2}\right) \exp\left(-\frac{2}{\alpha^2} \sinh^2\left(\frac{y - \mu_1}{2}\right)\right) \right] (1 - I_i), \quad (1)$$

where $c_1 = 1/\sqrt{2\pi}$, $\mu_1 = \mathbf{x}_{(1)}^\top \boldsymbol{\beta}_{(1)}$, $\zeta_{i2}^c = (2/\alpha) \sinh((y_0 - \mathbf{x}_{(1)}^\top \boldsymbol{\beta}_{(1)})/2)$,

$$I_i = \begin{cases} 1, & \text{if } y \leq y_0, \\ 0, & \text{if } y > y_0, \end{cases} \quad (2)$$

and $\Phi(\cdot)$ is the CDF of the standard normal distribution, with $\mathbf{x}_{(1)}$ being the covariates associated with $\boldsymbol{\beta}_{(1)}$. We assume a logit link for the random variable D , thus it is possible to include covariates as follows

$$\text{logit} [\mathbb{P}(D = 1 | \mathbf{x}_{(2)})] = \mathbf{x}_{(2)}^\top \boldsymbol{\beta}_{(2)}, \quad (3)$$

where $\mathbf{x}_{(2)}$ are the covariates associated with $\boldsymbol{\beta}_{(2)}$. The formulation of the logit link defined in Equation (3) becomes,

$$\tau_i = 1 - p_i = \frac{\exp(\mathbf{x}_{(2)}^\top \boldsymbol{\beta}_{(2)})}{1 + \exp(\mathbf{x}_{(2)}^\top \boldsymbol{\beta}_{(2)})}, \quad (4)$$

where the vector $\mathbf{x}_{(2)}$ has a dimension q , which can be, and usually is, different from the dimension of the vector $\mathbf{x}_{(1)}$.

The log-likelihood function for the mixture Bernoulli/BS is given by

$$\begin{aligned} \ell(\boldsymbol{\theta}) = & -(n - m) \log(2) - (n - m) \frac{\log(2\pi)}{2} + \\ & \sum_{i=1}^m I_i [\log(1 + \exp(\mathbf{x}_{(2)}^\top \boldsymbol{\beta}_{(2)})) [\Phi(\zeta_{i2}^c) - 1] - \log(1 + \exp(\mathbf{x}_{(2)}^\top \boldsymbol{\beta}_{(2)}))] \\ & + \sum_{m+1}^n (1 - I_i) [\mathbf{x}_{(2)}^\top \boldsymbol{\beta}_{(2)} + \log(\zeta_{i1}) - \frac{1}{2} \zeta_{i2}^2 - \log(1 + \exp(\mathbf{x}_{(2)}^\top \boldsymbol{\beta}_{(2)}))], \end{aligned} \quad (5)$$

where ζ_{i2}^c is defined in Equation (1) and ζ_{i1} and ζ_{i2} are defined in Equation (6),

$$\zeta_{i1} = \frac{2}{\alpha} \cosh\left(\frac{y_i - \mathbf{x}_{(1)i}^\top \boldsymbol{\beta}_{(1)}}{2}\right), \quad \zeta_{i2} = \frac{2}{\alpha} \sinh\left(\frac{y_i - \mathbf{x}_{(1)i}^\top \boldsymbol{\beta}_{(1)}}{2}\right). \quad (6)$$

3 Application

We analyse a data set provided by Moulton and Halsey (1995) from a study of measles vaccines. Neutralization antibody levels were collected from 330 children at 12 months of age. The LDL was 0.1 international units (IU) or -2.306 in logarithm scale. Around 86

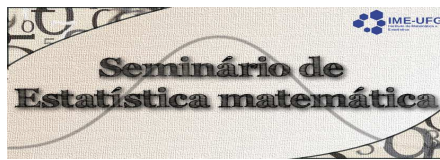
(26.1%) of the observations fell below the LDL and then recorded as 0.1. The following covariates were considered: X_1 indicates the type of vaccine used (0 if Schwartz and 1 if Edmonston-Zagreb); X_2 is the level of the dosage (0 if medium and 1 if high); and X_3 is the gender where 0 is male and 1 is female.

We here present the estimation results for the Bernoulli/BS model along with those of the standard tobit, tobit-BS (Chapter 2) and Bernoulli/LPN Martínez-Flórez et al. (2013) models. The Bernoulli/BS and Bernoulli/LPN have both a logit link. The covariates EZ and HI were used only in the logit component, and covariate FEM entered only in the continuous component of the models. Table 1 shows the ML estimates, Akaike information criterion (AIC) values and standard errors for the considered models. A glance at the results indicates that, in the Bernoulli/BS model, the receiver of Edmonston-Zagreb strain does not contribute to the odds ratio of being above the detection limit, however, the receiver of a high dose impacts $\exp(1.499) = 4.472$ in the odds of being above the detection limit. Moreover, the Bernoulli/BS model suggests that girls have $\exp(-0.078) - 1 = -0.075$ less concentration of measles antibody concentration than boys. We observe that the Bernoulli/LPN and Bernoulli/BS models do not agree on the sign of the coefficient corresponding to the FEM variable, while the first model indicates that girls have a higher measles antibody concentration than boys, the other indicates the opposite. Also from Table 1, we note that the Bernoulli/BS model provides a better fit compared to the other models based on the AIC values.

Table 1: ML estimates (with SE in parentheses) and AIC values for the indicated models with the measles vaccine data

Model	AIC	α	Continuous component			
			INT	EZ	HI	FEM
tobit	1299.27	0.945*** (0.047)	0.597** (0.288)	0.225 (0.297)	-0.228 (0.295)	0.271 (0.296)
tobit-BS	1168.60	1.545*** (0.048)	-0.910*** (0.105)	0.188* (0.111)	0.074 (0.109)	0.121 (0.110)
Bernoulli/LPN	976.48	8.918** (3.922)	-2.869*** (0.582)			0.222* (0.134)
Bernoulli/BS	760.64	1.560*** (0.108)	0.123*** (<0.001)			-0.078*** (<0.001)

Obs: Rejects H_0 at *10% of significance, ** 5% of significance and ***1% of significance.



4 Concluding remarks

The paper presents a mixture model where the continuous part follows a Birnbaum-Saunders distribution. We performed estimation based on the maximum likelihood approach. The simulation study showed the good performance of the maximum likelihood estimators. An application to measles vaccine data showed that the Bernoulli/BS models fits the data better than the standard tobit, tobit-BS and Bernoulli/LPN models.

Acknowledgments

This study was partially supported by CAPES from the Brazilian government.

References

- BIRNBAUM, Z.W.; SAUNDERS, S.C. A new family of life distributions. *Journal of Applied Probability*, 6, 319-327, 1969.
- CRAGG, J.G. Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica*, 39, 829-844, 1971.
- JOHNSON, N.; KOTZ, S.; BALAKRISHNAN, N. *Continuous Univariate Distributions*, volume 2. Wiley, New York, US, 1995.
- MARTÍNEZ-FLÓREZ, G.; BOLFARINE, H.; AND GÓMEZ, H.W. Asymmetric regression models with limited responses with an application to antibody response to vaccine. *Biometrical Journal*, 55, 156-172, 2013.
- MOULTON, L.H.; HALSEY, N.A. A mixture model with detection limits for regression analyses of antibody response to vaccine. *Biometrics*, 51, 1570-1578, 1995.

Modelos Gráficos Probabilísticos: Aplicações em biologia e ciências agrárias.

Renato Rodrigues Silva^{1*}

¹ Instituto de Matemática e Estatística, Universidade Federal de Goiás, Brasil

1 Introdução

Segundo Huang (2008), um modelo gráfico é um modelo probabilístico definido em termo de um grafo no qual os vértices representam variáveis aleatórias e as arestas descrevem relações de dependências entre essas variáveis. Um caso particular de modelo gráfico é a rede Bayesiana. Rede Bayesiana é um modelo gráfico cujo o conjunto de variáveis aleatórias e suas dependências condicionais são representados por meio de um grafo acíclico direcionado. Convém ressaltar que já existem diversas aplicações de Redes Bayesianas em ciências agrárias e em biologia em geral. Entre as diversas aplicações de redes Bayesianas, pode-se citar a modelagem de interação proteína-proteína. Rhodes (2005) utilizaram redes Bayesianas para predizer interações proteína-proteína em seres humanos usando um conjunto de dados integrado composto por dados de domínios de proteína, de expressão gênica e de genômica funcional. Felicia et al. (2016) modelaram uma rede Bayesiana para dados de fatores de transcriptomas oriundas de células hematopoiéticas de ratos. Através da rede foi possível distinguir interações diretas das indiretas entre os fatores de transcriptomas. Assim, o objetivo desse trabalho é apresentar como se obtém cálculos de probabilidades em modelos gráficos e mostrar resultados preliminares de algumas aplicações. Na seção metodologia será apresentado um exemplo hipotético de aplicação de redes Bayesianas na modelagem de interação proteína-proteína e aspectos teóricos da representação e dos algoritmos para obtenção de probabilidades em modelos gráficos. Na seção posterior, serão apresentados alguns resultados preliminares do projeto de pesquisa que temos aqui na UFG na área de modelos gráficos e, por fim, serão feitas algumas considerações finais sobre o tema.

*Autor para correspondência: e-mail: renato.rrsilva@gmail.com, Fone: +55-62-3521-1123.

2 Exemplo hipotético: Rede de regulação gênica

Nessa seção será apresentado um exemplo hipotético que servirá de base para explicação da teoria de representação e de obtenção de probabilidades em redes Bayesianas. Além de mostrar, hipoteticamente, uma das principais aplicações de modelos gráficos em biologia.

O exemplo a seguir trata-se de utilizar as redes Bayesianas na modelagem de uma rede de regulação gênica. Gene é a unidade fundamental da hereditariedade formado por sequências específicas de DNA, ver Griffiths et al. (2012). Segundo Alberts et al. (2012), o dogma central da biologia afirma que existem 3 processos vitais para o funcionamento das células: replicação do DNA (duplicação da molécula de DNA), transcrição do DNA (formação do RNA mensageiro a partir da molécula de DNA) e tradução (síntese de proteína a partir do RNA mensageiro). Uma rede regulatória gênica é uma coleção de reguladores moleculares que interagem entre si e com outras substâncias na célula para governar níveis de expressão gênica, ver Karlebach e Shamir. (2008). Por uma questão didática, o sistema de regulação gênica será simplificado. Convém ressaltar que esse exemplo é uma adaptação do exemplo apresentado por Koller e Friedman (2009) para um contexto biológico.

Considere uma determinada rede de regulação gênica formada por 8 genes. C, D, I, G, S, L, J, H . O objetivo será representar essa rede de regulação gênica com uma rede Bayesiana e calcular probabilidades do gene J está ativo ou suprimido. Assume-se que C, D, I, S, L, J e H tem dois níveis de expressão e G apresenta três.

3 Representação de Modelos Gráficos

3.1 Redes Bayesianas

Inicialmente, será considerado uma versão reduzida da rede de regulação gênica do exemplo 1 com apenas 5 genes D, I, G, S, L . Considere ainda que essa rede pode ser representada pelo seguinte grafo acíclico direcionado:

Nesse exemplo, os genes são representados por meio dos vértices do grafo e os arcos (setas) representam as relações de dependências e ou casualidade entre eles.

Um grafo direcionado (dígrafo) é um par ordenado $G = (V, A)$ tal que: V é um conjunto

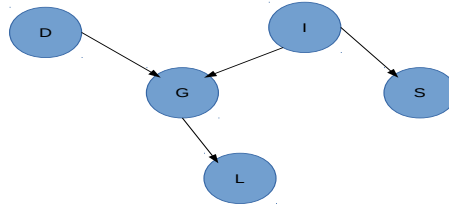


Figura 1: Modelo Gráfico do Exemplo Hipotético de Regulação Gênica.

chamado conjunto dos vértices ou nodos; A é um conjunto de pares ordenados $a = (v, w)$, v e $w \in V$ chamados de arcos (setas) do grafo, ver Feofiloff, et al. (2004). Observe que a relação entre os vértices não é simétrica, assim tem-se duas definições importante para o entendimento dos dígrafos.

- Sucessor: um vértice w é sucessor de v caso exista um arco que parte de v e chegue em w ;
- Antecessor: um vértice v é um antecessor de w caso exista um arco que parte de v e chegue em w .

Por exemplo, considerando a (figura 1) D e I são antecessores de G e L é sucessor de G .

Observe que o grafo por si só não nos permite o cálculo de probabilidades entre os elementos da rede de regulação gênica. Para isso, considere que cada vértice do grafo seja uma variável aleatória, v.a. Assim, denota-se X_1 como uma v.a que representa o nível de expressão gênica do gene D ; X_2, X_3, X_4 e X_5 representam os níveis de expressão dos genes I, G, S e L , respectivamente.

Dessa forma, podemos reescrever o grafo da figura 1 conforme mostra a figura 2.

Observe que uma das maneiras de se escrever a distribuição conjunta desse exemplo

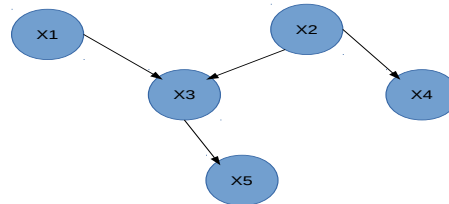


Figura 2: Modelo Gráfico do Exemplo Hipotético de Regulação Gênica Revisitado.

seria:

$$\begin{aligned}
 P(X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4, X_5 = x_5) &= P(X_1 = x_1)P(X_2 = x_2|X_1 = x_1) \times \\
 P(X_3 = x_3|X_2 = x_2, X_1 = x_1)P(X_4 = x_4|X_3 = x_3, X_2 = x_2, X_1 = x_1) &\times \\
 P(X_5 = x_5|X_4 = x_4, X_3 = x_3, X_2 = x_2, X_1 = x_1) &
 \end{aligned}$$

No entanto, sob o ponto de vista computacional essa forma de escrever a distribuição conjunta não é eficiente dado que haveria um alto consumo de memória do computador para armazenar algumas das tabelas de distribuição. Além disso, estaria sendo ignorado as informações do grafo que pode ter sido construído a partir do conhecimento de um especialista do problema.

Por outro lado, nota-se algumas declarações de independências implícitas no grafo. De uma maneira informal, duas variáveis X_i e X_j são independentes condicional à X_k se o conhecimento sobre X_i não fornece nenhuma informação extra sobre X_j dado que haja o conhecimento sobre X_k . Isto é, uma vez que X_k é conhecido, X_i não se adiciona nada ao que já se sabe sobre X_j .

No nosso exemplo de regulação gênica tem-se as seguintes declarações de independências:

- $X_5 \perp X_1, X_2, X_3, X_4 | X_3$;

- o que significa que o nível de expressão do gene L é independente do genes I, D, S condicional ao nível de expressão do gene G

- $X_4 \perp X_1, X_3, X_5 | X_2$;
 - o que significa que o nível de expressão do gene S é independente dos genes D, G, L dado um nível de expressão do gene I
- $X_3 \perp X_4 | X_1, X_2$;
 - o que significa que o nível de expressão do gene G é independente do gene S dado os níveis de expressão dos genes I, D .
- $X_1 \perp X_2$;

Assim, podemos reescrever a distribuição conjunta de uma forma mais compacta, isto é:

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4, X_5 = x_5) = P(X_1 = x_1)P(X_2 = x_2) \times P(X_3 = x_3 | X_1 = x_1, X_2 = x_2)P(X_4 = x_4 | X_2 = x_2)P(X_5 = x_5 | X_3 = x_3)$$

Nesse contexto, segundo Koller e Friedman (2009), pode-se definir uma rede Bayesiana como sendo :

- Um grafo direcionado acíclico os quais os vértices representam variáveis aleatórias X_1, \dots, X_n
- Para cada vértice X_i associa-se uma distribuição condicional $P(X_i = x_i | Pa_{X_i} = pa_{x_i})$

cuja distribuição conjunta definida por meio de

$$P(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n P(X_i = x_i | Pa_{X_i} = pa_{x_i})$$

em que Pa_{X_i} representa os antecessores de X_i .

Ou seja, cada variável aleatória é independente dos seus não descendentes dado os pais. $X_i \perp X_{\text{não descendentes de } i} | Pa_{X_i}$.

O Dígrafo G pode ser visto de duas maneiras distintas:

- Como uma estrutura de dados que proporciona o esqueleto para representação compacta de uma distribuição conjunta em uma forma fatorada;
- Como uma representação compacta de um conjunto de pressuposições de independência

Dada uma distribuição $P(X_1 = x_1, \dots, X_n = x_n)$ denota-se o mapa de independência $I(P)$ como sendo todas as afirmações da forma $X_i \perp X_j | X_k$ para todo $i, j, k \subseteq \{1, \dots, n\}$

3.2 Separação d

O objetivo dessa seção é entender quando pode-se garantir que uma afirmação de independência condicional $X_i \perp X_j | X_k$ é válida para uma rede Bayesiana associada a um grafo G . Separação d é um procedimento para verificar, dado um grafo causal, se X_i é independente de X_j dado X_k . Esse procedimento é baseado no conceito de “trilho ativo” e é descrito por Pearl (1998).

Pode-se dizer que X_i e X_j estão d separados por X_k , $d - sep_G(X_i; X_j | X_k)$, se não há nenhum trilho ativo entre eles. A definição formal de trilho ativo segue:

Seja G uma grafo de uma rede Bayesiana, $X_1 - \dots - X_n$ um trilho e \mathbf{x} um subconjunto de variáveis observadas. O trilho está ativo dado \mathbf{x} se toda vez que existir uma estrutura em v tal que $X_{l-1} \rightarrow X_l \leftarrow X_{l+1}$, X_l ou um dos seus descendentes estão em \mathbf{x} , nenhum outro vertice ao longo do trilho está em \mathbf{x} , ver Koller e Friedman (2009) para maiores detalhes.

Como exemplo de ilustração, considere 4 possíveis redes Bayesianas representados na figura 3 com três vértices, duas arestas e 3 variáveis aleatórias X_i , X_j e X_k .

Observe que os trilhos a,c e d da figura 3 aparecem no grafo do exemplo de regulação gênica representado na figura 2. Nota-se que $X_1 \rightarrow X_3 \rightarrow X_5$; $X_2 \leftarrow X_3 \rightarrow X_4$; e $X_1 \rightarrow X_3 \leftarrow X_2$; formam os padrões a,c e d.

Segundo Pearl (1998) pode-se afirmar o seguinte:

- Em (a, b, c), o trilho entre X_i e X_j está ativo se X_k não for observado.
- Em d, o trilho entre X_i e X_j está ativo se X_k ou alguns dos seus descendentes forem observados;

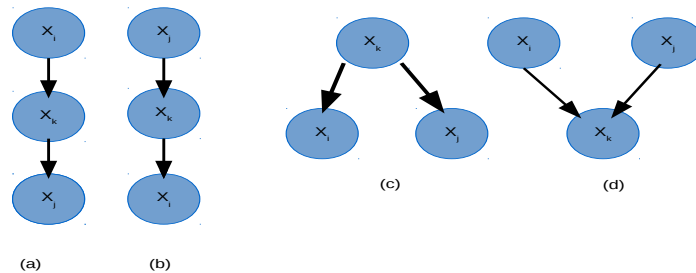


Figura 3: Quatro possíveis trilhos ativos de X_i até X_j via X_k .

Assim, pode-se dizer que os trilhos (a, b, c) são equivalentes.

O conjunto de independências definido por separação d é denotado por $I(G)$. É possível demonstrar que o método separação d tem duas propriedades: solidez e completude, em que solidez significa que se for determinado que X_i e X_j são separados por um conjunto \mathbf{x} , confirmar que de fato essas duas v.a são independentes condicionais na distribuição de probabilidade $P(X_1 = x_1, \dots, X_n = x_n)$. Ou seja, $I(G) \subseteq I(P)$. Embora o contrário nem sempre é verdadeiro, isto é, existem independências em $I(P)$ que não estão em $I(G)$. Por sua vez, completude significa confirmar que X_i e X_j são independentes dado X_k então essas v.a são d separadas.

3.3 Mapas Equivalentes e Mapa Minimal

Pode-se afirmar que dois grafos G_1 e G_2 são equivalentes se $I(G_1) = I(G_2)$ Nota-se que na seção anterior foi dito que os trilhos a,b e c são equivalentes. Agora, porque esse conceito é importante? É importante porque existem alguns aspectos dos modelos gráficos que são identificáveis. Em outras palavras, não é possível afirmar sem um conhecimento prévio do problema em questão se a direção da dependência é $X_i \rightarrow X_k \rightarrow X_j$ ou $X_j \rightarrow X_k \rightarrow X_i$, ou mesmo se é $X_i \leftarrow X_k \rightarrow X_j$. Porque todos esses trilhos contém as mesmas afirmações de independência condicional.

Mapa minimal é um mapa de independência sem arcos redundantes. A importância desse conceito reside no fato que um mapa minimal é mais parcimonioso, ou seja, tem menos parâmetros e, portanto, é mais informativo.

3.4 Redes de Markov

Embora esse não seja o foco desse texto, é necessário apresentar sucintamente o que venha a ser as redes de Markov para facilitar o entendimento dos algoritmos Eliminação de Variáveis e propagação da crença (“Belief Propagation”).

Uma rede de Markov pareada é um grafo não direcionado o quais os vértices X_1, \dots, X_n são associados com fatores.

Seja \mathbf{X} um conjunto de variáveis aleatórias. Define-se um fator ϕ uma função $\text{Val}(\mathbf{X}) \Rightarrow \mathbb{R}$ em que $\text{Val}(\mathbf{X})$ seja os possíveis valores que \mathbf{X} possa assumir. O conjunto \mathbf{X} é denominado de escopo do fator. Aqui serão usados apenas fatores não negativos $\phi : \text{Val}(\mathbf{X}) \Rightarrow \mathbb{R}$.

A seguir serão apresentados algumas operações com fatores, sendo que muitas delas se assemelham com o que é feito na teoria da probabilidade.

Seja X_1, X_2 e X_3 três variáveis aleatórias e seja $\phi_1(X_1, X_2)$ e $\phi(X_2, X_3)$ dois fatores. Define-se um produto de fatores $\phi_1 \times \phi_2$ como um fator $\psi : \text{Val}(X_1, X_2, X_3) \Rightarrow \mathbb{R}$

Um exemplo prático é mostrado da figura 4.

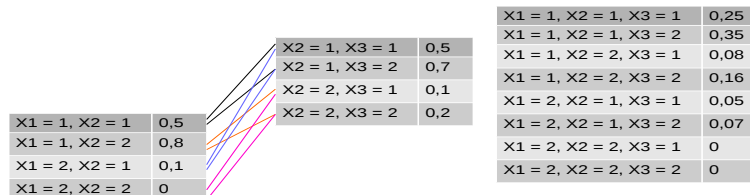


Figura 4: Produto de fatores.

Adicionalmente, tem-se a marginalização de fatores, que é feita de modo análoga a marginalização de função massa de probabilidade conjunta, e a redução de fatores que será definido a seguir

Seja um fator $\phi(X_1, X_2, X_3)$ e considere, por exemplo, $X_3 = x_3$. Nesse caso, um fator reduzido $\phi[x_3](X_1, X_2)$ teria como escopo as variáveis X_1 e X_2 . A figura 5 ilustra um exemplo que o fator $\phi(X_1, X_2, X_3)$ foi reduzido para $\phi[1](X_1, X_2)$.

$\Phi(x_1, x_2, x_3)$		$x_3 = 1; \Phi[1](x_1, x_2)$	
$x_1 = 1, x_2 = 1, x_3 = 1$	0,25	$x_1 = 1, x_2 = 1, x_3 = 1$	0,25
$x_1 = 1, x_2 = 1, x_3 = 2$	0,35	$x_1 = 1, x_2 = 2, x_3 = 1$	0,08
$x_1 = 1, x_2 = 2, x_3 = 1$	0,08	$x_1 = 2, x_2 = 1, x_3 = 1$	0,05
$x_1 = 1, x_2 = 2, x_3 = 2$	0,16	$x_1 = 2, x_2 = 2, x_3 = 1$	0
$x_1 = 2, x_2 = 1, x_3 = 1$	0,05		
$x_1 = 2, x_2 = 1, x_3 = 2$	0,07		
$x_1 = 2, x_2 = 2, x_3 = 1$	0		
$x_1 = 2, x_2 = 2, x_3 = 2$	0		

Figura 5: Redução de fatores.

4 Modelos Gráficos: Obtenção de Probabilidades

Considere um modelo gráfico definido por meio de um grafo direcionado ou não direcionado cujo os vértices sejam um conjunto de variáveis aleatórias \mathbf{X} . Considere ainda que o conjunto \mathbf{X} pode ser repartido em alguns subconjuntos tais como: o conjunto evidência \mathbf{X}_E que define as variáveis que serão observadas, o conjunto \mathbf{X}_Y que serão as variáveis as quais tem-se o interesse em calcular probabilidade e o conjunto $\mathbf{X}_W = \mathbf{X} - \mathbf{X}_E - \mathbf{X}_Y$ que são as variáveis aleatórias remanescentes. Portanto, tem-se o interesse em obter $P(\mathbf{X}_Y = \mathbf{x}_Y)$ ou $P(\mathbf{X}_Y = \mathbf{x}_Y | \mathbf{X}_E = \mathbf{x}_E)$. Entretanto, esse são problemas NP-difíceis. Sendo assim, é necessário utilizar algoritmos para que essas probabilidades possam ser solucionadas na maioria dos casos. Os algoritmos mais utilizados para esta finalidade são:

- Eliminação de Variáveis;
- Propagação de crença (“Belief Propagation”);
- MCMC;
- Amostragem por importância.

Nesse texto, trataremos dos dois primeiros.

4.1 Algoritmo Eliminação de Variáveis (EV)

O algoritmo Eliminação de Variáveis é baseado na programação dinâmica. Por simplicidade, o algoritmo será apresentado para a obtenção de $P(\mathbf{X}_Y = \mathbf{x}_Y)$, maiores detalhes podem ser encontrados em Koller e Friedman (2009).

Considere X um conjunto de variáveis aleatórias Φ um conjunto de fatores tais que para cada $\phi \in \Phi$, $\text{Escopo}(\phi) \subseteq \mathbf{X}$. Seja $\mathbf{X}_Y \subset \mathbf{X}$ o conjunto de variáveis de interesse em obter probabilidades e seja $\mathbf{X}_Z = \mathbf{X} - \mathbf{X}_Y$. Então para cada ordem \prec de \mathbf{X}_Z o algoritmo soma produto eliminação de variáveis retorna o fator $\phi(\mathbf{X}_Y)$ tal que

$$\phi(\mathbf{X}_Y) = \sum_{\mathbf{X}_Z} \prod_{\phi \in \Phi} \phi$$

Para exemplo de aplicação desse algoritmo, será extendido o exemplo de rede regulatória gênica para uma rede com 8 genes conforme mostra figura 6.

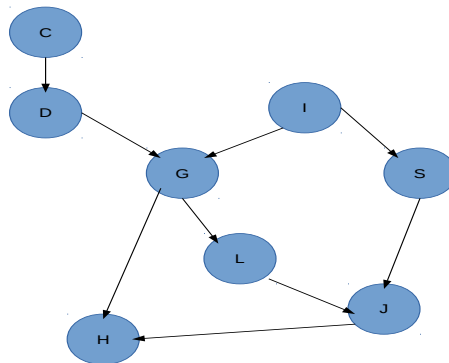


Figura 6: Grafo da rede de regulação gênica extendida.

Assim, redefine-se a rede Bayesiana, agora com 8 variáveis aleatórias sendo que X_1 representa o nível de expressão gênica do gene D ; $X_2, X_3, X_4, X_5, X_6, X_7$ e X_8 representam os níveis de expressão dos genes I, G, S, L, C, J e H , respectivamente (figura 7).

O objetivo é obter $P(X_7)$ utilizando o algoritmo de eliminação de variáveis:

A distribuição da rede Bayesiana é dada por:

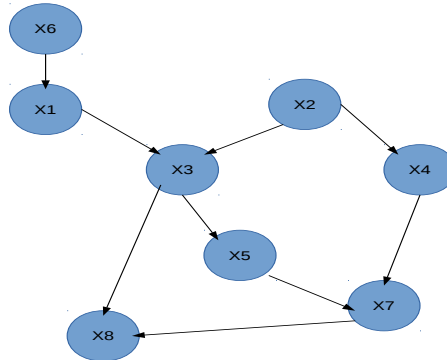


Figura 7: Rede Bayesiana para representar a rede de regulação gênica estendida.

$$\begin{aligned}
 P(X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4, X_5 = x_5, X_6 = x_6, X_7 = x_7, X_8 = x_8) &= P(X_6 = x_6) \times \\
 &P(X_1 = x_1 | X_6 = x_6) P(X_2 = x_2) P(X_3 = x_3 | X_1 = x_1, X_2 = x_2) P(X_4 = x_4 | X_2 = x_2) \\
 &P(X_5 = x_5 | X_3 = x_3) P(X_7 = x_7 | X_5 = x_5, X_4 = x_4) P(X_8 = x_8 | X_3 = x_3, X_7 = x_7) \\
 &= \phi(X_6) \phi(X_1, X_6) \phi(X_2) \phi(X_3, X_1, X_2) \phi(X_4, X_2) \phi(X_5, X_3) \\
 &\quad \phi(X_7, X_5, X_4) \phi(X_8, X_3, X_7)
 \end{aligned}$$

O algoritmo Eliminação de Variáveis será aplicado com a seguinte ordem de eliminação $X_6, X_1, X_2, X_8, X_3, X_4, X_5$

1- Eliminar X_6 :

$$\begin{aligned}
 \psi_1(X_1, X_6) &= \phi(X_6) \phi(X_1, X_6) \\
 \tau_1(X_1) &= \sum_{X_6} \psi_1(X_1, X_6)
 \end{aligned}$$

2- Eliminar X_1 :

$$\begin{aligned}\psi_2(X_1, X_2, X_3) &= \tau_1(X_1)\phi(X_3, X_1, X_2) \\ \tau_2(X_2, X_3) &= \sum_{X_1} \psi_2(X_1, X_2, X_3)\end{aligned}$$

3- Eliminar X_2 :

$$\begin{aligned}\psi_3(X_2, X_3, X_4) &= \tau_2(X_2, X_3)\phi(X_4, X_2)\phi(X_2) \\ \tau_3(X_3, X_4) &= \sum_{X_2} \psi_3(X_2, X_3, X_4)\end{aligned}$$

4- Eliminar X_8 :

$$\begin{aligned}\psi_4(X_8, X_3, X_7) &= \phi(X_8, X_3, X_7) \\ \tau_4(X_3, X_7) &= \sum_{X_8} \psi_4(X_8, X_3, X_7)\end{aligned}$$

4- Eliminar X_3 :

$$\begin{aligned}\psi_5(X_3, X_4, X_5, X_7) &= \tau_4(X_3, X_7)\tau_3(X_3, X_4)\phi(X_5, X_3) \\ \tau_5(X_4, X_5, X_7) &= \sum_{X_3} \psi_5(X_3, X_4, X_5, X_7)\end{aligned}$$

4- Eliminar X_4 :

$$\begin{aligned}\psi_6(X_4, X_5, X_7) &= \tau_5(X_4, X_5, X_7)\phi(X_7, X_5, X_4) \\ \tau_6(X_5, X_7) &= \sum_{X_4} \psi_6(X_4, X_5, X_7)\end{aligned}$$

4- Eliminar X_5 :

$$\begin{aligned}\psi_7(X_5, X_7) &= \tau_6(X_5, X_7) \\ \tau_6(X_7) &= \sum_{X_5} \psi_7(X_5, X_7)\end{aligned}$$

que nesse caso é igual a $P(X_7 = x_7)$.

Percebe-se que o algoritmo soma os fatores de dentro para fora de acordo com a ordem definida pelo usuário.

No entanto, um ponto crítico desse algoritmo é a ordem das variáveis que devem ser eliminadas. Dependendo da ordem escolhida, o algoritmo gera fatores intermediários com um escopo muito grande de variáveis.

Existem alguns algoritmos disponíveis na literatura para determinar a ordem ótima denominados “greedy search” (busca gulosa). Esses algoritmos consistem em cada ponto eliminar o vértice com menor custo. Possíveis funções custos:

- vizinho mínimo: vértice com menor número de vizinhos no atual grafo;
- mínimo peso: vértice que gera fatores intermediários com menor número de valores;
- mínimo preenchimento: vértice que geram menor adição de novas arestas durante o algoritmo.

Maiores detalhes sobre esses algoritmos podem ser encontrados em Koller e Friedman (2009).

4.2 Grafos de grupos e Árvores de cliques

Observa-se que o algoritmo eliminação de variáveis tem uma propriedade de induzir um novo grafo a partir do original. Mais adiante, será visto que essa propriedade será base para o entendimento do algoritmo soma produto propagação de crença.

No contexto de modelos gráficos probabilísticos, um grafo induzido pelo algoritmo EV denotado por $I_{\Phi, \prec}$ é um grafo não direcionado sobre um subconjunto $\mathbf{X}_Z \subset \mathbf{X}$, em que cada par de v.a X_i e X_j são conectadas por uma aresta se ambas aparecem em algum fator intermediário gerado pelo algoritmo eliminação de variáveis usando alguma \prec como ordem de eliminação.

Usando $I_{\Phi, \prec}$ é possível demonstrar que:

- O escopo de cada fator gerado durante o algoritmo de eliminação de variáveis é um clique em $I_{\Phi, \prec}$;

- Todo clique maximal em $I_{\Phi, \prec}$ é o escopo de algum fator intermediário usado na computação do algoritmo;
- Todo grafo induzido $I_{\Phi, \prec}$ é um grafo cordal.

Define-se um clique de um grafo como sendo qualquer subconjunto de vértices tal que para cada dois vértices desse subconjunto, existe uma aresta conectando-os. Por sua vez, um clique maximal é um clique que não está estritamente contido dentro de outro clique. Um grafo cordal é um grafo no qual todos os ciclos de quatro ou mais vértices tem uma corda, maiores detalhes podem ser encontrados em Koller e Friedman (2009); Leonard (2006); Feofiloff, et al. (2004).

A figura 8 apresenta o grafo induzido pelo algoritmo eliminação de variáveis para o exemplo de regulação gênica.

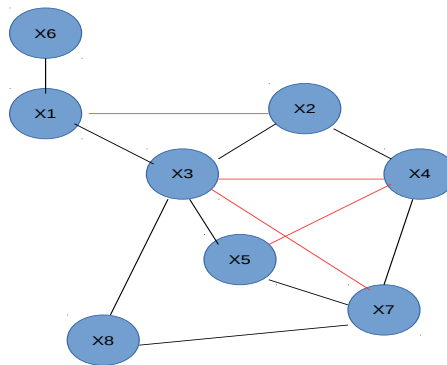


Figura 8: Grafo induzido referente ao exemplo da rede de regulação gênica estendida.

Nota-se algumas diferenças entre o grafos da figuras 7 e 8. O grafo induzido da figura 8 e é não direcionado e contem algumas arestas adicionais. Essas arestas foram adicionadas porque algumas v.a apareceram em algum fator intermediário do algoritmo de eliminação de váriáveis. Por exemplo a aresta que liga X_1 e X_2 foi adicionada devido a existência do fator intermediário $\psi_2(X_1, X_2, X_3)$.

Ainda analisando a figura 8, observa-se a formação de 5 cliques maximais

- $C_1 = \{X_1, X_6\}$;
- $C_2 = \{X_1, X_2, X_3\}$;

- $C_3 = \{X_2, X_3, X_4\}$;
- $C_4 = \{X_3, X_7, X_8\}$;
- $C_5 = \{X_3, X_4, X_5, X_7\}$;

Por meio desses cliques podemos formar uma estrutura de dados chamada de grafos de grupos.

Um grafo de grupos é um grafo não direcionado tal que cada vértice é formado por um conjunto de variáveis aleatórias pertencentes a \mathbf{X} (grupo C_i). Além disso, o grafo de grupos apresenta duas propriedades:

- Dado um conjunto de fatores Φ , atribui-se cada ϕ_k a um grupo $C_{\alpha(k)}$, isto é, $\text{Escopo}[\phi_k] \subseteq C_{\alpha(k)}$ e para cada fator $\phi_k \in \Phi$, existe um grupo C_i tal que $\text{Escopo}[\phi_k] \subseteq C_j$.
- Para cada par de grupo C_i e C_j e uma variável aleatória $X \in C_i \cap C_j$ existe um único caminho entre C_i e C_j para o qual todos os grupos e variáveis “sepsets” contém X .

sendo uma variável “sepsets” um conjunto de variáveis aleatórias comum a C_i e C_j .

Convém ressaltar que o grafo de grupos podem ser formados a partir de qualquer grafo cordal não necessariamente pelo grafo induzido pelo algoritmo EV, descrito por Koller e Friedman (2009).

Entretanto, caso o grafo cordal usado para gerar o grafo de grupos seja o grafo induzido pelo algoritmo EV, tem-se uma árvore de cliques. Ou seja, uma árvore de cliques é um caso particular de grafos de grupos

Segundo Szwarcfiter (1988), uma árvore é um grafo não direcionado, acíclico e conexo e em que cada vértice possui no máximo apenas um antecessor direto. Por sua vez, segundo Koller e Friedman (2009), uma árvore de cliques é uma árvore cujo os vértices são cliques maximais oriundos do grafo induzido pelo EV.

Um exemplo de árvore de cliques aplicada a rede de regulação gênica é apresentado na figura 9

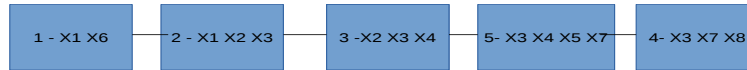


Figura 9: Árvore de clique referente ao exemplo da rede de regulação gênica extendida.

4.3 Algoritmo da propagação da crença

Dado um grafo de grupos, o algoritmo de soma e produto propagação da crença é definido por meio de:

- Atribuir cada fator $\phi_k \in \Phi$ a um clique $C_{\alpha(k)}$
- Construir fatores potenciais iniciais $\psi_i(C_i) = \prod_{k:\alpha(k)=i} \phi_k$;
- Inicializar todas mensagens iguais a 1;
- Repetir até a convergência:
 - Selecionar arestas (i, j) e passar a mensagem

$$\delta_{i \rightarrow j}(\mathbf{X}_{S_{ij}}) = \sum_{C_i - \mathbf{X}_{S_{ij}}} \psi \prod_{k \in (N_i - \{j\})} \delta_{k \rightarrow i}$$

em que $\mathbf{X}_{S_{ij}}$ são as variáveis “sepsets” e N_i são os cliques vizinhos do i -ésimo clique exceto o j -ésimo clique.

- Computar $\beta_i(C_i) = \psi_i \prod_{k \in N_i} \delta_{k \rightarrow i}$

Algumas observações importantes da aplicação desse algoritmo é que as vezes ele pode não convergir. Além disso, para definir o grafo de grupos, o grafo da rede Bayesiana ou da rede de Markov tem que ser transformado em um grafo cordal.

Entretanto, quando o grafo de grupos trata-se de uma árvore de cliques, o algoritmo sempre converge e, além disso, a implementação do mesmo é mais simples.

Antes de apresentar o algoritmo de propagação da crença baseado em árvores de cliques é importante que se faça alguns comentários.

Observa-se que o algoritmo EV usa os fatores intermediários τ_i no máximo um vez e que durante a execução do algoritmo, o fluxo da “mensagem” vai para um determinado vértice (grupo), que é denominado raiz da árvore de clique.

A aplicação do algoritmo de propagação de crença com árvores de cliques será apresentado por meio de um exemplo.

Considere a árvore de cliques da figura 9 e considere ainda que se deseja obter $P(X_7)$. Inicialmente, para executar o algoritmo de propagação de crença no contexto de arvores de cliques, deve-se definir qual clique será a raiz. O clique raiz pode ser qualquer clique que contém a variável de interesse, nesse caso a variável X_7 . Assim, o clique 5 da figura 9 será escolhido como raiz.

Definido a raiz, o próximo passo é definir os fatores potenciais iniciais de cada clique por exemplo:

A atribuição dos fatores aos respectivos cliques é feita *ad hoc*. A única regra é: cada clique recebe os fatores que contem as váriáveis do clique. Por exemplo, $\phi(X_2)$ pode-se ser atribuído ao clique 2 ou 3. A seguir um exemplo dessa atribuição:

1- Fator Potencial Inicial do Clique 1 $\{X_1, X_6\}$:

$$\psi_{C_1}(X_1, X_6) = \phi(X_6)\phi(X_1, X_6)$$

2- Fator Potencial Inicial do Clique 2 $\{X_1, X_2, X_3\}$:

$$\psi_{C_2}(X_1, X_2, X_3) = \phi(X_3, X_1, X_2)$$

3- Fator Potencial Inicial do Clique 3 $\{X_2, X_3, X_4\}$

$$\psi_{C_3}(X_2, X_3, X_4) = \phi(X_4, X_2)\phi(X_2)$$

4- Fator Potencial Inicial do Clique 4 $\{X_3, X_7, X_8\}$.

$$\psi_{C_4}(X_8, X_3, X_7) = \phi(X_8, X_3, X_7)$$

5- Fator Potencial Inicial do Clique 5 $\{X_3, X_4, X_5, X_7\}$

$$\psi_{C_5}(X_3, X_4, X_5, X_7) = \phi(X_3, X_5)\phi(X_7, X_5, X_4)$$

Agora a ideia é passar a mensagem partindo de todos os cliques não raiz até chegar no clique raiz. Observando a figura 9 nota-se que o fluxo das mensagens deve seguir a seguinte ordem $1 \rightarrow 2$, $2 \rightarrow 3$, $3 \rightarrow 5$ e $5 \leftarrow 4$

Ou seja,

1- Passar Mensagem do clique 1 para o clique 2

$$\delta_{1 \rightarrow 2}(X_1) = \sum_{X_6} \psi_{C_1}(X_1, X_6)$$

2- Passar Mensagem do clique 2 para o clique 3

$$\delta_{2 \rightarrow 3}(X_2, X_3) = \sum_{X_1} \delta_{1 \rightarrow 2}(X_1) \psi_{C_2}(X_1, X_2, X_3)$$

3- Passar Mensagem do clique 3 para o clique 5

$$\delta_{3 \rightarrow 5}(X_3, X_4) = \sum_{X_2} \delta_{2 \rightarrow 3}(X_2, X_3) \psi_{C_3}(X_2, X_3, X_4)$$

4- Passar Mensagem do clique 4 para o clique 5

$$\delta_{4 \rightarrow 5}(X_3, X_7) = \sum_{X_4, X_8} \delta_{3 \rightarrow 5}(X_3, X_4) \psi_{C_4}(X_8, X_3, X_7)$$

Quando todas as mensagens chegarem no clique raiz, multiplica-se cada mensagem com o fator potencial inicial do clique raiz formando um fator denominado crença

$$\beta(X_3, X_4, X_5, X_7) = \delta_{4 \rightarrow 5}(X_3, X_7) \delta_{3 \rightarrow 5}(X_3, X_4) \psi_{C_5}(X_3, X_4, X_5, X_7)$$

O passo final do algoritmo é dado por:

$$\tilde{P}(X_7 = x_7) = \sum_{X_3, X_4, X_5} \beta(X_3, X_4, X_5, X_7)$$

observe que se os fatores ϕ não forem distribuição de probabilidade, no final deve-se normalizar $\tilde{P}(X_7 = x_7)$.

A figura 10 apresenta o algoritmo propagação de crença utilizando árvores de cliques em pseudo linguagem de computador

Algorithm 10.1 Upward pass of variable elimination in clique tree

```

Procedure CTree-SP-Upward (
     $\Phi$ , // Set of factors
     $T$ , // Clique tree over  $\Phi$ 
     $\alpha$ , // Initial assignment of factors to cliques
     $C_r$  // Some selected root clique
)
1 Initialize-Cliques
2 while  $C_r$  is not ready
3   Let  $C_i$  be a ready clique
4    $\delta_{i \rightarrow p(i)}(S_{i,p(i)}) \leftarrow$  SP-Message( $i, p(i)$ )
5    $\beta_r \leftarrow \psi_r \cdot \prod_{k \in \text{Nb}_{C_r}} \delta_{k \rightarrow r}$ 
6   return  $\beta_r$ 

Procedure Initialize-Cliques (
)
1 for each clique  $C_i$ 
2    $\psi_i(C_i) \leftarrow \prod_{\phi_j : \alpha(\phi_j) = i} \phi$ 
3

Procedure SP-Message (
     $i$ , // sending clique
     $j$  // receiving clique
)
1  $\psi(C_i) \leftarrow \psi_i \cdot \prod_{k \in (\text{Nb}_i - \{j\})} \delta_{k \rightarrow i}$ 
2  $\tau(S_{i,j}) \leftarrow \sum_{C_l - S_{i,j}} \psi(C_l)$ 
3 return  $\tau(S_{i,j})$ 

```

Figura 10: Pseudocódigo do algoritmo de propagação de crença descrito em Koller e Friedman (2009).

5 Aplicações

A seguir serão descritos os dois projetos de iniciação científica referentes ao tema e alguns resultados preliminares.

5.1 Modelos Gráficos e Casualidade: Predição de estrutura secundária de proteínas

Propõe-se utilizar redes Bayesianas para prever estrutura secundária de proteínas globular. Proteínas globulares são aquelas que apresentam uma ou mais cadeias polipeptídicas organizadas em uma forma final aproximadamente esférica, ver Nelson e Cox (2011) para maiores detalhes. Ou seja, a ideia do projeto consiste em considerando uma sequência linear de aminoácidos classificar essa sequência em alfa hélice, folha beta, e estrutura não regular (“random coil”). A predição da estrutura secundária é importante para o estudo das propriedades biofísicas das proteínas.

Esse projeto está na fase inicial, mais precisamente na fase de representação da rede Bayesiana. A figura 11 apresenta um possível modelo gráfico para esse problema. A ideia é que as estruturas secundárias possam assumir apenas 3 estados não observáveis ou latentes $\{H, B, C\}$, em que H representa a estrutura α -hélice, B representa folha β , C representa a estrutura não regular e AA_i representam a sequencias de aminoácidos observados.

5.2 Modelos Gráficos e Casualidade: Redes Bayesianas para identificação de grãos de trigo

Os dados que serão usados no projeto estão armazenados no banco de dados no repositório “UCI Machine Learning Repository” cujo o website está no endereço <http://archive.ics.uci.edu/ml/datasets/seeds>. Os detalhes serão descrito a seguir: O conjunto de dados refere-se três variedades diferentes de trigo: Kama, Rosa and Canadian. Um total de 70 indivíduos foram selecionados aleatoriamente de cada variedade. Visualização da estrutura interna da semente foi detectada usando tecnologia de raio X. As imagens foram gravada em placas KODAK 13×18 cm. As características morfológicas do grão foram: área, perímetro, compacidade, comprimento do grão, largura do grão, coeficiente

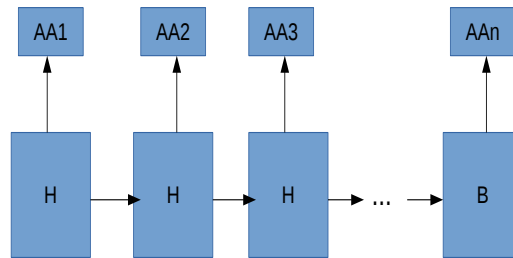


Figura 11: Representação do modelo gráfico para predição de estruturas de proteínas.

de assimetria e comprimento do sulco do grão. O objetivo desse trabalho é utilizar uma rede Bayesiana para classificar um grão de acordo com sua variedade baseado nas suas características morfológicas.

Resultados preliminares desse projeto mostra a representação gráfica da rede Bayesiana que futuramente será ajustada aos dados (figura 12).

Nesse grafo T representa as possíveis variedades de grãos e assume três valores possíveis; L, C, A, P, C e SG são v.a contínuas e representam largura, comprimento, área, perímetro, compacidade e sulco de grão respectivamente.

6 Considerações Finais

A partir dos resultados obtidos nesse presente projeto, espera-se uma maior difusão dos modelos gráficos entre o grupo de estatística do instituto de matemática e estatística

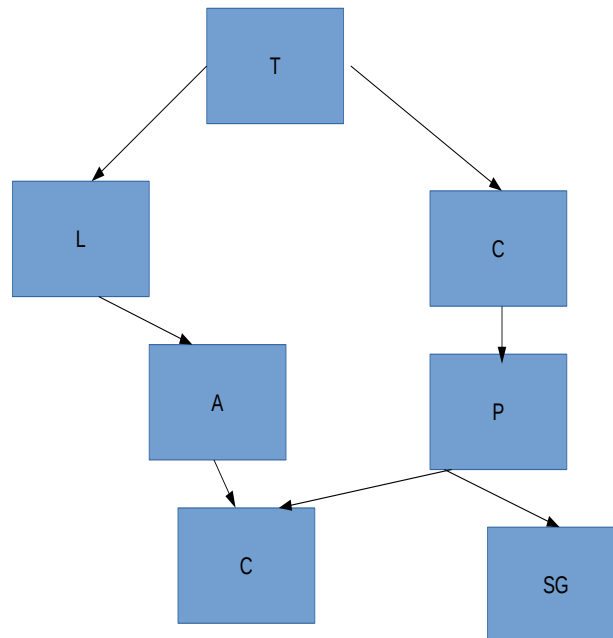


Figura 12: Representação do modelo gráfico para identificação de grãos de trigo.

da universidade federal de Goiás. Pretende-se ainda incentivar os alunos de graduação a participar de pesquisas científicas, e, conseqüentemente, de elaboração de publicações em revistas científicas; além de participações em congressos e ou seminários.

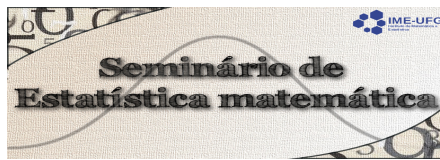
Agradecimentos

Esse trabalho não teve qualquer apoio financeiro. Agradecemos a todos envolvidos na organização desse evento pela oportunidade de participar.

Referências

KOLLER, D; FRIEDMAN, N. *Probabilistic Graphical Models: Principles and Techniques*. edited by . MIT Press.4. 2009.

- RHODES, D.R. et al. Probabilistic model of the human protein-protein interaction network, *Nat. Biotechnology*, 23:951-959, 2005.
- FELICIA, S.L. Ng et. al. A graphical model approach visualizes regulatory relationships between genome-wide transcription factor binding profiles. *Briefings in Bioinformatics*, 1-12. 2016. doi: 10.1093/bib/bbw102
- JORDAN, M.I. *Learning in Graphical Models*. The MIT Press, Cambridge, MA, 1998.
- HUANG, R. *Graphical models for object segmentation*. Ph.D. Department of Computer Science of Graduate School-New Brunswick Rutgers, The State University of New Jersey. 2008.
- GRIFFITHS, A.J.F.; WESSLER, S.R.; LEWONTIN, R.C.; GELBART, W.M.; SUZUKI, D.T.; MILLER J.H.; DOEBLEY J.; *Introduction to genetic analysis*. 10th edition. New York: W. H. Freeman and Co. 832 p, 2012.
- ALBERTS, B.; JOHNSON, A.; LEWIS, J.; RAFF, M.; ROBERTS, K.; WALTER, P. *Molecular Biology of the Cell*, 4th edition New York: Garland Science; 2002.
- KARLEBACH, G.; SHAMIR, R.; *Nature Reviews Molecular Cell Biology*. 2008 Oct;9(10):770-80. doi: 10.1038/nrm2503.
- NELSON, DAVID L.; COX, MICHAEL M. *Princípios de bioquímica de Lehninger*. Porto Alegre: Artmed, 2011. 6. ed.
- FEOFILOFF, P.; KOHAYAKAWA, Y.; WAKABAYASHI, Y. *Uma Introdução Sucinta à Teoria dos Grafos*, 2004.
- PEARL, J. *Probabilistic Reasoning in Intelligent Systems*. Morgan and Kaufman, San Mateo, 1988.
- FARIA, R.C. *Redes probabilísticas: aprendendo estruturas e atualizando probabilidades*. Dissertação (Mestrado) - Programa em Estatística. Universidade de São Paulo, Brasil, São Paulo, BR, 2014.
- SZWARCFITER, J. L. *Grafos e algoritmos computacionais*. Rio de Janeiro: Campus. 1998.



Correção tipo-Bartlett para a estatística gradiente em modelos lineares generalizados

Tiago Moreira Vargas

Instituto de Matemática e Estatística, Universidade Federal de Goiás, Brasil

Resumo

Apresentamos aqui a expansão até ordem $o(n^{-1})$ para a estatística gradiente, onde n é o tamanho da amostra, em modelos lineares generalizados com dispersão desconhecida. A partir desta expansão, apresentamos uma correção tipo-Bartlett para a estatística gradiente em modelos lineares generalizados e comparamos, através de simulações de Monte Carlo, o desempenho do teste baseado nesta estatística com o teste baseado na estatística gradiente não-coriigida e com os testes baseados nas estatísticas da razão de verossimilhanças, score, suas respectivas formas corrigidas e o teste de Wald.

Palavras-chave: Teste gradiente; Correção de Bartlett; Correção tipo Bartlett; Modelos lineares generalizados.

1 Introdução

Os modelos lineares generalizados (MLG), definidos por Nelder e Wedderburn (1972) formam uma classe de modelos com ampla utilização na análise de dados de diversas áreas. O diferencial do MLG está na flexibilidade da ligação entre a média da variável resposta e o componente linear do modelo. A inferência em MLG baseia-se principalmente no método da máxima verossimilhança e os métodos de testes de hipóteses tomam como base principalmente os testes da razão de verossimilhanças, score e Wald. A utilização destes testes em MLG podem ser vistos em Pregibon (1982), Davison e Tsai (1990), Cook e Weisberg, S. (1983) e outros.

De forma pioneira, Cordeiro (1983, 1987) apresentou formas matriciais para o fator de correção de Bartlett (Bartlett 1953a, 1953b) do teste baseado na estatística da razão de verossimilhanças em MLG com dispersão conhecida e desconhecida, respectivamente.

Mais tarde, a partir da obtenção do fator de correção tipo-Bartlett para a estatística escore em Cordeiro e Ferrari (1991), Cordeiro, Ferrari e Paula (1993) e Cribari-Neto e Ferrari (1995) obtiveram quantidades matriciais para este fator de correção em MLG com dispersão conhecida e desconhecida, respectivamente.

O teste baseado na estatística gradiente foi proposto por Terrell (2002) e é uma alternativa à *hole trinity* de testes baseados nas estatísticas da razão de verossimilhanças, escore e Wald. A estatística gradiente também está na classe das estatísticas que convergem em distribuição para a distribuição qui-quadrado. Rao (2005) disse " *The suggestion by Terrell is attractive as it is simple to compute. It would be of interest to investigate the performance of the [gradient] statistic.*" A partir daí vários estudos tomando como base o teste baseado na estatística gradiente foram apresentados. Lemonte e Ferrari (2012a) compararam o poder local do teste baseado na estatística gradiente com os testes da razão de verossimilhanças, escore e Wald e concluíram que nenhum dos testes é uniformemente mais poderoso que o outro. Lemonte (2013) fez um estudo da estatística gradiente em modelos com especificação incorreta. Outros trabalhos envolvendo o teste baseado na estatística gradiente podem ser vistos em Lemonte (2011), Lemonte (2012), Lemonte e Ferrari (2011), Lemonte e Ferrari (2012b) e Lemonte e Ferrari (2012c).

Recentemente, Vargas, Ferrari e Lemonte (2013) derivaram a expansão sob a hipótese nula da função de distribuição acumulada da estatística gradiente até ordem n^{-1} . Esta expansão é função de momentos do logaritmo da função de verossilhança para o modelo escolhido. A partir daí, os autores propuseram uma estatística gradiente corrigida por um fator de correção tipo-Bartlett. Esta estatística possui distribuição χ_q^2 , onde q é o número de restrições da hipótese nula. A diferença desta estatística, com a versão não corrigida, está na velocidade de convergência. A estatística corrigida, converge para a distribuição χ_q^2 a partir de um tamanho de amostra menor do que a estatística não-corrigida. Sendo assim, o teste baseado na estatística corrigida tende a ser mais acurado em amostras pequenas. Estes resultados são muito gerais, visto que não estão atrelados à modelos específicos.

Com base nos resultados de Vargas, Ferrari e Lemonte (2013), foi obtida uma expansão para a função de distribuição da estatística gradiente sob a hipótese nula até ordem n^{-1} , onde n é o tamanho da amostra, para modelos lineares generalizados com parâmetro de dispersão desconhecido. As quantidades envolvidas nesta expansão estão em notação matricial. A partir disso fazemos um estudo da estatística gradiente corrigida por um fator

de correção tipo-Bartlett, que possui distribuição χ_q^2 de referência, onde q é o número de restrições impostas na hipótese nula. Estas quantidades foram obtidas em Vargas, Ferrari e Lemonte (2014). Estudos de simulação são exibidos com o objetivo de verificar a performance dessa estatística corrigida para testar hipóteses compostas em tamanhos de amostra pequenos. Comparações com o teste gradiente não corrigido e com as estatísticas da razão de verossimilhanças e escore bem como suas respectivas versões corrigidas, e com o teste baseado na estatística de Wald.

2 Correção tipo-Bartlett para MLG com dispersão desconhecida

Considere $y = (y_1, \dots, y_n)^\top$ variáveis independentes, onde cada componente y_l tem densidade na família exponencial da forma

$$f(y, \theta_l, \phi) = \exp[\phi\{y\theta_l - b(\theta_l) + c(y)\} + a(y, \phi)], \quad (1)$$

em que $a(\cdot, \cdot)$, $b(\cdot)$ e $c(\cdot)$ são conhecidas. Aqui, θ_l e ϕ_l são possíveis parâmetros desconhecidos. Para o modelo (1) temos $E[y_l] = \mu_l = b'(\theta_l)$, $Var[y_l] = V_l/\phi$, onde ϕ^{-1} é o parâmetro de dispersão e $V = V(\mu) = d\mu/d\theta$ é a função de variância em que $\theta = \int V^{-1}d\mu = q(\mu)$. Suponha x_1, \dots, x_p , onde $x_j = (x_{1j}, \dots, x_{nj})^\top$, determinam uma forma linear dada por $\eta = X\beta$, sendo $X = (x_1, \dots, x_p)$ a matriz de planejamento do modelo conhecida e de posto completo sendo $\beta = (\beta_1, \dots, \beta_p)^\top$ um vetor de parâmetros desconhecidos. Um modelo linear generalizado (MLG) é definido por (1) e por uma função de ligação $d(\mu) = \eta$, estritamente monótona e duas vezes diferenciável, que relaciona o vetor $\mu = (\mu_1, \dots, \mu_n)$, em que μ_l é a média de cada observação com o vetor $\eta = (\eta_1, \dots, \eta_n)$ e η_l é o preditor linear do modelo. Consideramos a partição $\beta = (\beta_1^\top, \beta_2^\top)^\top$ em que $\beta_1 = (\beta_1, \dots, \beta_q)^\top$ e $\beta_2 = (\beta_{q+1}, \dots, \beta_p)^\top$ com $q \leq p$, o que induz a partição $X = (X_1 \ X_2)$ para a matriz do modelo. O interesse é o teste de hipóteses composto: $\mathcal{H}_0: \beta_1 = \beta_{10}$ versus $\mathcal{H}_a: \beta_1 \neq \beta_{10}$, onde β_2 é o vetor de parâmetros de perturbação. Assumimos até esse momento que o parâmetro de dispersão ϕ^{-1} é conhecido. Seja L o logaritmo da função de verossimilhança total, $K = E[UU^\top]$ a matriz de informação de Fisher e $U = \partial L/\partial \beta$ a função escore total para β que pode ser escrita como $U = U(\beta) = \phi X^\top W^{1/2} V^{-1/2} (y - \mu)$. A matriz de informação

de Fisher total pode ser escrita como $K = \phi X^T W X$, onde $V = \text{diag}\{V_1, \dots, V_n\}$, $W = \text{diag}\{w_1, \dots, w_n\}$ com $w_l = (d\mu_l/d\eta_l)^2/V_l$. Seja $\hat{\beta} = (\hat{\beta}_1^T, \hat{\beta}_2^T)^T$ o estimador de máxima verossimilhança irrestrito para β e $\tilde{\beta}_2$ o estimador de máxima verossimilhança restrito de β_2 sob \mathcal{H}_0 . As funções avaliadas em $\tilde{\beta} = (\beta_{10}^T, \tilde{\beta}_2^T)$ são indicadas pela adição de um 'til' sobrescrito. A função escore total é particionada de acordo com a partição de β como $U(\beta) = (U_1(\beta)^T, U_2(\beta)^T)^T$. A matriz K é particionada de acordo com a partição de β nas submatrizes $K_{11} = \phi X_1^T W X_1$, $K_{22} = \phi X_2^T W X_2$ e $K_{12} = K_{21} = \phi X_1^T W X_2$. A inversa da matriz de informação de Fisher particionada é dada por

$$K^{-1} = \begin{bmatrix} K^{11} & K^{12} \\ K^{21} & K^{22} \end{bmatrix},$$

onde $K^{11} = \phi^{-1}(R^T W R)^{-1}$, $K^{12} = K^{21^T} = -\phi^{-1}(R^T W R)^{-1} C^T$ e $K^{22} = \phi^{-1}\{(X_2^T W X_2)^{-1} + C(R^T W R)^{-1} C^T\}$, em que $R = X_1 - X_2 C$ e $C = (X_2^T W X_2)^{-1} X_2^T W X_1$. Seja $\hat{K}^{11} = \phi^{-1}(\hat{R}^T \hat{W} \hat{R})^{-1}$ a submatriz correspondente à partição da matriz inversa da informação de Fisher observada correspondente a β_1 .

Assim, a gradiente é dada por

$$S_G = U_1(\tilde{\beta})^T (\hat{\beta}_1 - \beta_{10}) \quad (2)$$

em que $s = \phi^{1/2} V^{-1/2} (y - \mu)$. Esta estatística possui distribuição assintótica χ_q^2 sob a hipótese nula. Sejam $U_i = \partial L / \partial \beta_i$, $U_{ij} = \partial^2 L / \partial \beta_i \partial \beta_j$, etc derivativas do logaritmo da função de verossimilhança. No que se segue, todos os índices variam de 1 a p . Definimos

$$\kappa_{ij} = E(U_{ij}), \quad \kappa_{ijk} = E(U_{ijk})$$

$$\kappa_{ijk} = E(U_{ijk}), \quad \kappa_{ij}^{(r)} = \partial \kappa_{ij} / \partial \beta_r, \quad \kappa_{ij}^{(kr)} = \partial^2 \kappa_{ij} / \partial \beta_k \partial \beta_r,$$

e assim por diante, momentos e derivadas de momentos do logaritmo da função de verossimilhança total. Consideremos a seguinte expansão para a função de distribuição da estatística gradiente S_G sob a hipótese nula:

$$\begin{aligned} \mathbb{P}(S_G \leq c) = & G_q(c) + \frac{1}{24} \{ A_3 G_{q+6}(c) (A_2 - 3A_3) G_{q+4}(c) + (A_1 - 2A_2 + 3A_3) G_{q+2}(c) \\ & + (A_2 - A_1 - A_3) G_q(c) \} + o(n^{-1}), \end{aligned} \quad (3)$$

onde $G_r(c) = \mathbb{P}(\chi_r^2 \leq c)$. As quantidades A_1, A_2 e A_3 em (3), são funções de momentos do logaritmo da função de verossimilhança. Sua forma geral é dada em Vargas, Ferrari e Lemonte (2013). Sejam

$$a = \frac{A_1 - A_2 + A_3}{12q}, b = \frac{A_2 - 2A_3}{12q(q+2)}, c = \frac{A_3}{12q(q+2)(q+4)}.$$

Consideramos a estatística S_G^* , que tem distribuição assintótica χ_q^2 sob a hipótese nula até ordem n^{-1} .

$$S_G^* = S_G(1 - c - bS_G - aS_G^2) \quad (4)$$

que corresponde à estatística gradiente corrigida pelo fator de correção tipo-Bartlett (Vargas, Ferrari e Lemonte, 2013).

Aqui, apresentamos fórmulas matriciais para os coeficientes da correção da estatística gradiente em MLG com ϕ desconhecido.

Definimos $(r)_l = \partial\eta_l/\partial\beta_r = x_{lr}$, $(rs)_l = (\partial\eta_l/\partial\beta_r)(\partial\eta_l/\partial\beta_s) = x_{lr}x_{ls}$, etc. Também

$$f = \frac{1}{V} \frac{d\mu}{d\eta} \frac{d^2\mu}{d\eta^2}, \quad f - g = \frac{1}{V^2} \frac{dV}{d\mu} \left(\frac{d\mu}{d\eta} \right)^3, \quad \lambda_1 = \frac{1}{V^2} \frac{dV}{d\mu} \left(\frac{d\mu}{d\eta} \right)^2 \frac{d^2\mu}{d\eta^2}, \quad \lambda_2 = \frac{1}{V} \left(\frac{d^2\mu}{d\eta^2} \right)^2,$$

$$\lambda_3 = \frac{1}{V} \frac{d\mu}{d\eta} \frac{d^3\mu}{d\eta^3}, \quad \lambda_4 = \frac{1}{V^3} \left(\frac{dV}{d\mu} \right)^2 \left(\frac{d\mu}{d\eta} \right)^4, \quad \lambda_5 = \frac{1}{V^2} \frac{d^2V}{d\mu^2} \left(\frac{d\mu}{d\eta} \right)^4,$$

$$t = -9\lambda_1 + 3\lambda_2 + 3\lambda_3 + 4\lambda_4 - 2\lambda_5,$$

$$d = -5\lambda_1 + 2\lambda_2 + 2\lambda_3 + 2\lambda_4 - \lambda_5, \quad e = -12\lambda_1 + 3\lambda_2 + 4\lambda_3 + 6\lambda_4 - 3\lambda_5.$$

Apresentamos a seguir alguns momentos de derivadas do logaritmo da função de verossimilhanças e algumas relações:

$$\kappa_{ijk} = -\phi \sum_{l=1}^n (f + 2g)_l (ijk)_l, \quad \kappa_{ijk}^{(r)} = -\phi \sum_{l=1}^n t_l (ijk_r)_l, \quad \kappa_{ij}^{(kr)} = -\phi \sum_{l=1}^n d_l (ijk_r)_l,$$

$$\kappa_{ijk_r} = -\phi \sum_{l=1}^n e_l (ijk_r)_l.$$

Adicionalmente, seja

$$Z = X(X^\top W X)^{-1} X^\top, \quad Z_2 = X_2(X_2^\top W X_2)^{-1} X_2^\top, \quad Z_d = \text{diag}(z_{11}, \dots, z_{nn}),$$

$$Z_{2d} = \text{diag}(Z_{211}, \dots, z_{2nn}), \quad F = \text{diag}(f_1, \dots, f_n), \quad G = \text{diag}(g_1, \dots, g_n),$$

$$T = \text{diag}(t_1, \dots, t_n), \quad E = \text{diag}(e_1, \dots, e_n), \quad d = \text{diag}(d_1, \dots, d_n)$$

e seja \otimes designando o produto de Hadamard entre duas matrizes, $Z^{(2)} = Z \otimes Z$, $Z^3 = Z \otimes Z \otimes Z$, etc.

Consideramos $\Psi = (\beta_1^\top, \beta_2^\top, \phi)^\top$ um vetor de parâmetros $(p+1)$ -dimensional. O objetivo aqui é testar $\mathcal{H}_a : \beta_1 \neq \beta_{10}$, em que $(\beta_2^\top, \phi)^\top$ é o vetor de parâmetros de perturbação. Teremos então três tipos de momentos do logaritmo da função de verossimilhança: momentos que envolvem somente o vetor de parâmetros β , que envolvem os parâmetros β e ϕ e os que envolvem somente o parâmetro ϕ , ou seja, $\kappa_{\phi\phi} = E(\partial^2 L / \partial \phi^2)$, $\kappa_{\phi r s} = E(\partial^3 L / \partial \phi \partial \beta_r \partial \beta_s)$, etc. Um fato importante é que parâmetros β e ϕ são globalmente ortogonais (Cox e Reid (1987)), ou seja,

$$K = \begin{bmatrix} K_\beta & 0 \\ 0 & K_\phi \end{bmatrix}, \quad K^{-1} = \begin{bmatrix} K_\beta^{-1} & 0 \\ 0 & K_\phi^{-1} \end{bmatrix},$$

onde $K_\phi = -\kappa_{\phi\phi} = -E(\partial^2 L / \partial \phi^2)$. Para a hipótese nula considerada, temos:

$$K = \begin{bmatrix} K_{11} & K_{12} & 0 \\ K_{21} & K_{22} & 0 \\ 0 & 0 & \kappa_{\phi,\phi} \end{bmatrix}, \quad K^{-1} = \begin{bmatrix} K^{11} & K^{12} & 0 \\ K^{21} & K^{22} & 0 \\ 0 & 0 & \kappa_{\phi,\phi}^{-1} \end{bmatrix}, \quad A = \begin{bmatrix} A_\beta & 0 \\ 0 & \kappa_{\phi,\phi}^{-1} \end{bmatrix}, \quad M = \begin{bmatrix} M_\beta & 0 \\ 0 & 0 \end{bmatrix},$$

onde $M_\beta = K_\beta^{-1} - A_\beta$. Aqui, $m^{r\phi}$ e $a^{r\phi}$ correspondem aos respectivos elementos da posição $(r, p+1)$ de M e A , e $m^{\phi\phi}$ e $a^{\phi\phi}$ os elementos da posição $(p+1, p+1)$ de M e A respectivamente. Temos $m^{r\phi} = m^{\phi r} = m^{\phi\phi} = a^{r\phi} = a^{\phi r} = 0$, $(r = 1, \dots, p)$ com $a^{\phi\phi} = -\kappa_{\phi\phi}^{-1}$.

Para a família exponencial biparamétrica de posto completo com parâmetros canônicos ϕ e $\phi\theta$, podemos reescrever o termo $a(y, \phi)$ da equação (1) como $a(y, \phi) = d_1(\phi) + d_2(y)$.

Seja $d_{(2)} = \phi^2 d_1''(\phi)$ e $d_{(3)} = \phi^3 d_1'''(\phi)$. Assim,

$$\kappa_{\phi\phi} = nd_1''(\phi), \quad \kappa_{\phi\phi i} = 0, \quad \kappa_{ij\phi} = \kappa_{ij}^{(\phi)} = - \sum_{l=1}^n w_l(ij)_l, \quad \kappa_{ij\phi}^{(\phi)} = 0. \quad (5)$$

Outros momentos e funções de momentos de derivadas do logaritmo da função de verossimilhanças podem ser obtidos através destes, utilizando relações de Bartlett (Bartlett (1953a), Bartlett (1953b)).

Podemos mostrar que quando ϕ é desconhecido,

$$\begin{aligned} A_1 = & 12\phi^{-1}1^\top(F+G)\{Z_d Z Z_d - Z_{2d} Z_2 Z_{2d} + Z^{(3)} - Z_2^{(3)}\}(F+G)1 \\ & - 6\phi^{-1}1^\top(F+2G)\{(Z+Z_2) \otimes (Z^{(2)} - Z_2^{(2)})\}(F+G)1 \\ & - 6\phi^{-1}1^\top\{(F+2G)(Z-Z_2)_d(Z Z_d + Z_2 Z_{2d})\}(F+G)1 \\ & - 12\phi^{-1}1^\top(F+2G)\{Z_{2d}(Z Z_d - Z_2 Z_{2d})\}(F+G)1 \\ & - 12\phi^{-1}1^\top(F+2G)\{Z_2^{(2)} \otimes (Z - Z_2)\}(F+G)1 \\ & + 6\phi^{-1}1^\top T\{(Z - Z_2)_d Z_d + 2(Z - Z_2)_d Z_{2d}\}1 \\ & + 6\phi^{-1}1^\top(T - E)\{(Z - Z_2)_d Z_{2d}\}1 \\ & + 6\phi^{-1}1^\top(F+2G)\{(Z - Z_2)_d Z_2 Z_{2d}\}(F+2G)1 \\ & + 6\phi^{-1}1^\top(F+2G)\{Z_2^{(2)} \otimes (Z - Z_2)\}(F+2G)1 \\ & + 3\phi^{-1}1^\top(F+2G)\{Z_{2d}(Z - Z_2) Z_{2d}\}(F+2G)1 \\ & + 3\phi^{-1}1^\top(F+2G)\{Z_{2d}(Z - Z_2)(Z - Z_2)_d\}(F+2G)1 \\ & - 12\phi^{-1}1^\top D(Z_d^{(2)} - Z_{2d}^{(2)})1 + \frac{6q\{d_{(3)} + (2-p+q)d_{(2)}\}}{nd_{(2)}^2}, \end{aligned} \quad (6)$$

$$\begin{aligned}
A_2 = & 6\phi^{-1}\mathbf{1}^\top(F + 2G)\{(Z - Z_2) \otimes (Z^{(2)} - Z_2^{(2)})\}(F + G)\mathbf{1} \\
& + 6\phi^{-1}\mathbf{1}^\top(F + 2G)\{(Z - Z_2)_d(ZZ_d - Z_2Z_{2d})\}(F + G)\mathbf{1} \\
& - \frac{9}{4}\phi^{-1}\mathbf{1}^\top(F + 2G)\{(Z - Z_2)_d(Z - Z_2)(Z - Z_2)_d\}(F + 2G)\mathbf{1} \\
& - \frac{3}{2}\phi^{-1}\mathbf{1}^\top(F + 2G)\{(Z - Z_2)^{(3)}\}(F + 2G)\mathbf{1} \\
& - 3\phi^{-1}\mathbf{1}^\top(F + 2G)\{Z_{2d}(Z - Z_2)(Z - Z_2)_d\}(F + 2G)\mathbf{1} \\
& - 3\phi^{-1}\mathbf{1}^\top(F + 2G)\{(Z - Z_2)_dZ_2(Z - Z_2)_d\}(F + 2G)\mathbf{1} \\
& - 6\phi^{-1}\mathbf{1}^\top(F + 2G)\{Z_2 \otimes (Z - Z_2)^{(2)}\}(F + 2G)\mathbf{1} \\
& - 3\phi^{-1}\mathbf{1}^\top(2T - E)(Z - Z_2)_d^{(2)}\mathbf{1} + \frac{3q(q + 2)}{nd_{(2)}}, \tag{7}
\end{aligned}$$

$$\begin{aligned}
A_3 = & \frac{3}{4}\phi^{-1}\mathbf{1}^\top(F + 2G)\{(Z - Z_2)_d(Z - Z_2)(Z - Z_2)_d\}(F + 2G)\mathbf{1} \\
& + \frac{1}{2}\phi^{-1}\mathbf{1}^\top(F + 2G)\{(Z - Z_2)^{(3)}\}(F + 2G)\mathbf{1}, \tag{8}
\end{aligned}$$

em que $\mathbf{1}$ é um vetor de 1's. Assim, as quantidades A_1 , A_2 e A_3 que são utilizados na correção do teste gradiente em modelos lineares generalizados quando o parâmetro de dispersão é desconhecido, são dados pelas quantidades (6),(7) e (8) substituindo ϕ por sua estimativa de máxima verossimilhança sob \mathcal{H}_0 , ou seja, $\tilde{\phi}$. Quando os A 's envolvem outros parâmetros desconhecidos, substituímos estes por seus estimadores de máxima verossimilhança sob \mathcal{H}_0 .

3 Simulações

Como ilustração à simulações, tomamos como base o modelo de regressão com variável resposta gama, na classe de modelos lineares generalizados. Neste caso, cada y_l em (1) tem densidade gama, com $\phi c(y) + a(y, \phi) = (\phi - 1) \log(y\phi) + \log \phi - \log \Gamma(\phi)$ em que $\Gamma(\phi)$ é a função gama, $b(\theta) = -\log(-\theta)$, $\mu = E(y) = -1/\theta$ e $V(\mu) = \mu^2$. Assim, $\mu = \exp\left\{\sum_{j=1}^p \beta_j x_j\right\}$, tomando a função de ligação logaritmica. Com base nesse modelo, as quantidades A_1 , A_2 e A_3 são facilmente calculadas. Para as simulações de tamanho

e poder, consideramos a primeira coluna da matriz de covariadas como um vetor de 1's e as demais covariadas geradas como valores absolutos da distribuição normal padrão. Para o método da máxima verossimilhança foi utilizado o método quasi-Newton. As simulações foram conduzidas na linguagem matricial Ox (Doornik , 2011). Comparações entre os testes baseados nas estatísticas de Wald (S_W), razão de verossimilhanças (S_{LR}), escore (S_R), gradiente (S_G), e as versões corrigidas dos testes baseados na estatísticas da razão de verossimilhanças (S_{LR}^*) (Bartlett, 1953a), escore (S_R^*) (Cordeiro e Ferrari, 1991) e gradiente (S_G^*) (Vargas, Ferrari e Lemonte , 2013) são apresentadas.

3.1 Tamanho dos testes

Aqui, foram consideradas amostras de tamanho 10, 15 e 20 com dimensão de β dada por $p = 5$ com $q = 3, 2, 1$ parâmetros testados na hipótese nula. As taxas de rejeição consideradas para a hipótese nula foram de 10% e 5%, com 100000 réplicas e Monte Carlo. Os resultados da simulação encontram-se descritos na Tabela 1.

Podemos observar que o teste gradiente corrigido teve um bom desempenho frente aos outros testes com comportamento semelhante ao teste escore corrigido. O teste gradiente corrigido teve uma distorção menor que o teste da razão de verossimilhança corrigido. As estatísticas corrigidas tiveram comportamento mais liberal, principalmente no caso da estatística de Wald.

3.2 Estudo de poder

Aqui, foi considerado o tamanho de amostra $n = 20$ com dimensão de β dada por $p = 5$ com $q = 2$ e taxa de rejeição da hipótese nula 5%. Os testes foram analisados sob a hipótese alternativa $\mathcal{H}_1 : \beta_1 = \beta_2 = \delta$ com $\delta = \{-3, -2, -1, -0,5, 0,5, 1, 2, 3\}$. Foram consideradas 100000 réplicas de Monte Carlo. Os resultados da simulação encontram-se descritos na Tabela 2. De modo geral, podemos concluir que não há significante diferença entre os testes em termos de poder.

Tabela 1: Taxas de rejeição de \mathcal{H}_0 no modelo de regressão gama com dispersão desconhecida para as estatísticas de Wald (S_W), LR (S_{LR}), escore (S_R), gradiente (S_G), e versões corrigidas: (S_{LR}^* , S_R^* , S_G^*); $p = 5$.

n	α	S_W	S_{LR}	S_R	S_G	S_{LR}^*	S_R^*	S_G^*
$q = 3$								
10	10	50,69	31,87	7,33	13,82	16,85	9,25	9,49
	5	42,52	22,27	1,71	3,77	9,75	3,79	3,23
15	10	35,95	22,10	10,00	12,48	12,52	10,74	9,97
	5	27,69	13,71	4,80	5,11	6,65	5,56	4,63
20	10	29,64	17,63	9,84	11,84	11,88	10,65	10,23
	5	21,70	10,56	5,04	5,28	5,87	5,28	4,98
$q = 2$								
10	10	43,27	30,10	14,07	20,34	16,86	11,53	11,46
	5	35,12	21,04	5,95	9,69	9,93	5,59	5,08
15	10	30,34	21,08	11,87	15,52	12,5	10,78	10,44
	5	22,28	13,06	5,80	7,50	6,60	5,59	5,13
20	10	24,83	17,26	10,75	13,60	11,39	10,33	10,32
	5	17,45	10,15	5,18	6,69	5,99	5,17	5,17
$q = 1$								
10	10	32,38	26,02	18,47	22,31	16,07	11,58	11,57
	5	24,80	17,82	10,35	13,06	9,41	5,94	5,96
15	10	23,23	18,76	13,15	16,63	12,45	10,42	10,61
	5	16,25	11,65	7,03	9,12	6,64	5,38	5,44
20	10	19,88	15,71	11,75	14,24	11,20	10,21	10,25
	5	13,04	9,25	5,85	7,60	5,82	5,12	5,17

Tabela 2: Taxas de rejeição não-nulas no modelo de regressão gama com dispersão desconhecida para as estatísticas de Wald (S_W), LR (S_{LR}), escore (S_R), gradiente (S_G), e versões corrigidas: (S_{LR}^* , S_R^* , S_G^*); $p = 5$, $q = 2$ e $n = 20$, $\alpha = 5\%$

δ	S_W	S_{LR}	S_R	S_G	S_{LR}^*	S_R^*	S_G^*
-3,00	100,00	100,00	93,15	100,00	100,00	99,79	100,00
-2,00	99,99	99,70	67,71	99,16	99,26	78,10	98,83
-1,00	62,24	52,48	39,92	45,23	42,15	42,28	40,72
-0,50	47,14	36,01	17,72	26,71	26,06	21,21	23,20
0,50	42,84	35,55	32,62	31,34	26,76	34,03	27,88
1,00	64,45	52,54	33,11	43,35	41,74	35,91	38,40
2,00	99,99	99,94	94,69	99,84	99,83	97,96	99,80
3,00	100,00	100,00	99,98	100,00	100,00	100,00	100,00

4 Conclusão

A obtenção de fórmulas matriciais para o fator de correção tipo-Bartlett para a estatística gradiente corrigida em modelos lineares generalizados com dispersão desconhecida é feita de forma simples a partir dos resultados de Cordeiro (1983) e Cordeiro (1987). Em amostras pequenas, a distribuição qui-quadrado nem sempre é uma boa aproximação para a distribuição das estatísticas da razão de verossimilhanças, escore, Wald e gradiente, portanto, os testes baseados nestas estatísticas apresentam resultados distorcidos. As correções de Bartlett e tipo-Bartlett para as estatísticas da razão de verossimilhanças, escore, Wald e gradiente se justificam nesse sentido. Para os MLG com dispersão desconhecida, o teste gradiente corrigido apresentou-se como uma alternativa competitiva em relação ao teste escore corrigido, e apresentou menos distorção que o teste da razão de verossimilhança corrigido. Em todos os casos, as correções de Bartlett e tipo-Bartlett foram efetivas.

Agradecimentos

O autor agradece ao NAE/UFG pela oportunidade de apresentação do trabalho.

Referências

- BARTLETT, M.S . Approximate confidence intervals. *Biometrika*, 40, 12-19, 1953.
- BARTLETT, M.S . Approximate confidence intervals II. More than one unknown parameter. *Biometrika*, 40, 306-317, 1953.
- COOK, R.D e WEISBERG, S. Diagnostics for heterocedasticity in regression. *Biometrika*, 70, 1-10, 1983.
- CORDEIRO, G.M., FERRARI, S.L.P. A modified score test statistic having chi-squared distribution to order n^{-1} . *Biometrika*, 78, 573-582, 1991.
- CORDEIRO, G.M., FERRARI, S.L.P, PAULA, G.M. Improved score tests for generalized linear models. *Journal of the Royal Statistical Society B*, 53, 629-643, 1993.



- CORDEIRO, G.M., CRIBARI-NETO, F. On Bartlett and Bartlett-type corrections. *Econometric Reviews*, 15, 339-367, 1996.
- CORDEIRO, G.M. Improved likelihood ratio statistics for generalized linear models. *Journal of the Royal Statistical Society B*, 45, 404-413, 1983.
- CORDEIRO, G.M.. On the corrections to the likelihood ratio statistics. *Biometrika*, 74, 265-274, 1987
- COX, D.R, REID, N. Parameter orthogonality and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society B*, 40, 1-39, 1987
- CRIBARI-NETO, F., FERRARI, S.L.P. Second order asymptotics for score tests in generalised linear models. *Biometrika*, 82, 426-432, 1995.
- DAVIDSON, A., TSAI, T.L. Regression Model Diagnostics *Statistical Review / Revue Internationale de Statistique*, 60, 337-353, 1990.
- J.A.DOORNIK. Ox: an object-oriented matrix programming language, 2011. *Timberlake Consultants and Oxford*.
- LEMONTE, A.J. (2011). Local power of some tests in exponential family nonlinear models. *Journal of Statistical Planning and Inference*, 141, 1981-1989, 2011.
- LEMONTE, A.J. Local power properties of some asymptotic tests in symmetric linear regression models. *Journal of Statistical Planning and Inference*, 142, 1178-1188, 2012.
- LEMONTE, A.J., FERRARI, S.L.P. Size and power properties of some tests in the Birnbaum–Saunders regression model. *Computational Statistics and Data Analysis*, 55, 1109-1117, 2011.
- LEMONTE, A.J., FERRARI, S.L.P. The local power of the gradient test. *Annals of the Institute of Statistical Mathematics*, 64, 373-381, 2012.
- LEMONTE, A.J., FERRARI, S.L.P. A note on the local power of the LR, Wald, score and gradient tests. *Electronic Journal of Statistics*, 6, 421-434, 2012.
- LEMONTE, A.J., FERRARI, S.L.P. Local power and size properties of the LR, Wald, score and gradient tests in dispersion models. *Statistical Methodology*, 9, 537-554, 2012.



- LEMONTE, A.J. On the gradient statistic under model misspecification. *Statistics & Probability Letters*, 83, 390-398, 2013.
- MCCULLAGH, P., NELDER, J.A (1989) Generalized Linear Models. *Chapmann and Hall*, 511p. *Monographs on Statistics and Applied Probability*, 1989.
- NELDER, J.A., WEDDERBURN, R.W.M. (1972). Generalised linear models *Journal of Royal Statistical Society A*, 135, 370-384, 1972.
- PREGIBON, D.(1982). Score tests in GLIM with applications . *GILCHRIST, R., ed. GLIM 82: Proceedings of the International Conference on Generalized Linear* ,1, 87-97, 1982.
- RAO, C.R. Score test: historical review and recent developments. In *Advances in Ranking and Selection, Multiple Comparisons, and Reliability*, N. Balakrishnan, N. Kannan and H. N. Nagaraja, eds. Birkhuser, Boston, 2005.
- TERRREL, G.R. (2002). The gradient statistic. *Computing Science and Statistics*,34, 206-215, 2002.
- VARGAS, T.M., FERRARI, S.L.P, LEMONTE, A.J. (2013). Gradient statistitc: Higher-order asymptotics and Bartlett-type correction. *Electronic Journal of Statistics*, 7, 43-61, 2013.
- VARGAS, T.M., FERRARI, S.L.P, LEMONTE, A.J. Improved likelihood inference in generalized linear models. *Computational Statistics & Data Analysis (Print)*, 74, 110-124, 2014.

Dispersão como Estratégia de Sobrevivência

Valdivino Vargas Junior^{1*}, Fabio Prates Machado², Alejandro Roldan Correa¹

¹ Instituto de Matemática e Estatística, Universidade Federal de Goiás, Brasil

² Instituto de Matemática e Estatística, Universidade Federal de São Paulo, Brasil

³ Instituto de Matematicas, Universidad de Antioquia, Medellin, Colômbia

Resumo

Aplicações da teoria de probabilidade em biologia matemática tem sido alvo de estudo por diversos probabilistas. Existem diversos modelos nesta área com importantes aplicações. Dentre as aplicações destacam-se os modelos estocásticos de evolução populacional. Neste texto são apresentados alguns modelos de evolução populacional de espécies que vivem em colônias as quais podem em momentos aleatórios sofrerem colapsos. Diversas estratégias podem ser tomadas por uma população de sobreviventes perante um colapso. O principal objetivo desse texto consiste em apresentar variantes de um modelo estocástico populacional e analisar a dispersão como estratégia de sobrevivência.

Palavras-chave: Dinâmica Populacional; Colonização; Colapso; Processos de Ramificação; Dispersão.

1 Introdução

Nos dias atuais tem-se intensificado o estudo de modelos estocásticos cujas motivações iniciais são problemas encontrados na Física Estatística e Biologia Matemática. A teoria de probabilidade aliada a métodos estatísticos e computacionais permite o estudo e avaliação de diversos modelos estocásticos com aplicações relevantes em uma quantidade muito extensa de áreas. No âmbito da modelagem estocástica via sistemas de partículas interagentes a ideia é entender fenômenos globais decorrentes de inúmeras interações microscópicas entre agentes cuja dinâmica segue regras probabilísticas ou mesmo

*Autor para correspondência: e-mail: vvjunior@gmail.com, Fone: +55-62-3521-1208.

determinísticas. O foco deste trabalho é a dinâmica de colonização de localidades por parte de populações de uma dada espécie.

Populações biológicas estão sujeitas a catástrofes que podem causar extinção em massa ou eliminação parcial dos indivíduos. Quando ocorre uma catástrofe os indivíduos sobreviventes podem agir de diferentes maneiras. Uma estratégia adotada por algumas populações é a dispersão. Nesse caso, os indivíduos tentam criar novas colônias numa outra localidade, podendo ou não haver competição. Basicamente, pode-se imaginar dinâmicas populacionais com o seguinte funcionamento. Indivíduos iniciam uma colônia, esta é povoada e após algum tempo há um colapso. Alguns indivíduos sobrevivem ao colapso e migram para novas localidades visando formar novas colônias. Nesse tipo de dinâmica há algumas questões a serem consideradas. De que modo a população cresce? Qual o tempo de colonização até o instante do colapso? Qual tipo de efeito catastrófico ocorre no colapso? Há ou não dispersão?

Além disso, caso haja dispersão é preciso considerar as seguintes questões. Qual o tipo de meio considerado? De que modo os indivíduos sobreviventes agem para se dispersar? Há concorrência por espaço?

Encontra-se na literatura artigos com diversas abordagens diferentes. Artalejo et al. (2007) consideram modelos sem dispersão. Neste trabalho o tamanho populacional evolui segundo um processo de nascimento e morte que incorpora a possibilidade de redução populacional de tamanho arbitrário no momento de um colapso. Os autores consideram dois tipos de catástrofes, a saber binomial e geométrica. Brockwell et al. (2016) consideram a evolução populacional através de um processo de Markov a tempo contínuo que inclui em seu gerador taxas de nascimento, imigração e catástrofe. Os autores consideram três tipos de catástrofes. Além das catástrofes binomial e geométrica, consideram catástrofes do tipo uniforme. Neste trabalho os autores consideram também variantes onde o crescimento populacional entre catástrofes se dá de forma determinística. Cairns et al. (2016) modela o crescimento populacional de acordo com o processo de nascimento, morte e catástrofes. Porém em seu modelo as taxas de transição dependem do tamanho populacional de modo completamente arbitrário. No caso supercrítico desse modelo, o tamanho populacional pode atingir tamanho infinito em tempo finito.

Schinazi (2015) e Machado et al. (2017) consideram modelos onde há dispersão e esta é uma boa estratégia para sobrevivência. Formalmente falando, um modelo estocástico de evolução populacional sobrevive se em qualquer instante de tempo existir colônia em

povoamento. Caso contrário, diz-se que o processo se extingue ou morre. Schinazi (2015) usa cadeias de nascimento e morte em meios aleatórios para modelar o crescimento populacional. Machado et al. (2017) consideram um modelo onde a população de cada colônia cresce de acordo com um processo de Yule e os indivíduos criam colônias sobre os vértices de um grafo conexo não orientado de grau localmente limitado.

Neste texto é apresentado um conjunto de modelos estudados por Junior et al. que ao serem comparados mostram que a dispersão nem sempre é a melhor estratégia de sobrevivência. Basicamente, isto depende de três fatores, a saber, o tipo de catástrofe, as restrições espaciais do meio e a probabilidade que cada indivíduo tem de sobreviver ao colapso. Neste texto, as catástrofes binomial e geométrica são misturadas num mesmo processo de modo que no instante de colapso, a catástrofe é de cada um dos tipos com uma certa probabilidade. Tanto em Machado et al. (2017) como nesse texto, cada colônia cresce durante um tempo exponencial de média 1 até o instante do colapso. Para avaliar a dispersão são considerados três tipos de processos. De forma análoga ao processo apresentado em Artalejo et al. (2007), no processo sem dispersão a população cresce de acordo com o processo de nascimento, morte e catástrofes. Nos outros dois processos onde há dispersão cada colônia tem crescimento populacional dado por um processo de Poisson. No modelo onde há restrição espacial as colônias ocupam vértices de um grafo regular.

2 Os Modelos

2.1 O Modelo de Schinazi

Na natureza é comum encontrar espécies que vivem em colônia durante um certo tempo até que essa sofra algum tipo de colapso. Ocorrendo colapso, eventuais sobreviventes podem dispersar e procurar localidades para formar colônias novas. Schinazi (2015) propôs um modelo estocástico não espacial para modelar tal tipo de dinâmica populacional. Em seu modelo, mostra que a dispersão aleatória é uma boa estratégia para a sobrevivência da população. A seguir seguem os principais resultados do trabalho de Schinazi (2015).

Teorema 2.1 (Teorema 1 em Schinazi (2015)). *Considere um processo onde os indivíduos estão divididos em um número aleatório de colônias independentes. Cada colônia é um processo de nascimento e morte, onde a taxa de nascimento é amostrada de uma distribuição fixada μ e a taxa de morte é 1. Cada colônia é associada a um tempo aleatório T*

amostrado de uma distribuição fixada ν . No tempo T cada indivíduo na colônia começa uma nova colônia independente com um novo Λ e um novo T . Esta população sobrevive se e somente se $\mathbb{E}[\exp((\Lambda - 1)\tau)] > 1$, onde Λ tem distribuição μ e τ tem distribuição ν .

Nesse modelo os indivíduos não morrem no colapso, o que ocorre é a destruição do habitat onde a colônia se encontra. Assim, os indivíduos migram procurando novos locais. Analisando o resultado, Schinazi (2015) conclui que comparando com o modelo de ambiente fixo (sem colapsos), segue que a dispersão aleatória ajuda na sobrevivência. Por outro lado, não é claro se a sobrevivência se torna mais provável pela multiplicação de ambientes aleatórios (dispersão) ou simplesmente devida à mudança do ambiente. Para dar continuidade a essa análise ele introduz um outro modelo sem dispersão e com mudanças globais do ambiente para toda população. O modelo é descrito no seguinte Teorema.

Teorema 2.2 (Teorema 2 em Schinazi (2015)). *Considere uma cadeia de nascimento e morte em ambientes aleatórios para a qual a taxa de morte é 1 e a taxa de nascimento muda nos tempos $T_k = \tau_1 + \tau_2 + \dots + \tau_k$ onde τ_1, τ_2, \dots são variáveis aleatórias independentes e identicamente distribuídas. As taxas de nascimento são amostradas de uma distribuição fixada μ . Assuma que $\mathbb{E}(\tau_1) < \infty$ e $\mathbb{E}(\Lambda) < \infty$ onde Λ tem distribuição μ . Assuma que as taxas de nascimento e os tempos τ_i são independentes. Então, o processo sobrevive se e somente se $\mathbb{E}(\Lambda) > 1$.*

A conclusão do teorema é que apenas mudanças globais no ambiente não favorecem à sobrevivência. Dessa forma, Schinazi (2015) conclui que é de fato a mistura da dispersão com a mudança do ambiente o que ajuda à sobrevivência.

2.2 Modelo de Brockwell

Brockwell et al. (2016) considera um modelo não espacial com uma única colônia. Cada indivíduo dá nascimentos a taxa $\lambda > 0$ e morre a taxa $\mu > 0$. Além disso, catástrofes (colapsos) ocorrem a taxa $a > 0$. Quando uma catástrofe ocorre todo indivíduo na colônia tem probabilidade p de sobreviver e $1 - p$ de morrer, independente do que ocorre a outros indivíduos. Brockwell et al. (2016) mostra que a sobrevivência (isto é, em todos os instantes de tempo existe pelo menos um indivíduo na colônia) tem probabilidade positiva

de ocorrer se e somente se $\lambda > \mu$ e

$$\mu - a \log p < \lambda.$$

Isso significa que, existe um valor crítico para p ,

$$p_1 = \exp\left(-\frac{\lambda - \mu}{a}\right).$$

Nesse caso, o modelo sobrevive se e somente se $p > p_1$.

2.3 O Modelo de Roldan- Machado-Schinazzi

Trata-se de um modelo estocástico de colonização e colapso em um grafo \mathcal{G} com conjunto de vértices V , publicado em Machado et al. (2017). Em cada vértice de \mathcal{G} pode haver uma colônia ou não. Cada colônia começa com um único indivíduo. O número de indivíduos na colônia no tempo t é determinado por um processo de nascimento puro (processo de Yule) de taxa $\lambda \geq 0$. A cada colônia é associada um tempo exponencial de média 1. Quando o tempo exponencial ocorre, a colônia colapsa e o vértice da colônia se torna vazio. No tempo do colapso cada indivíduo da colônia sobrevive com probabilidade $p > 0$ ou morre com probabilidade $q = 1 - p$. Cada sobrevivente pula aleatoriamente para um dos vértices vizinhos. Se o vértice já está ocupado o indivíduo morre, se o vértice está vazio o indivíduo inicia nele uma nova colônia. Quando mais de um indivíduo salta para um mesmo vértice vazio, somente um é responsável pela formação da colônia nova; os demais morrem. Este modelo é denotado por $CC(\mathcal{G}, \lambda, p)$. Trata-se de um processo de Markov a tempo contínuo com espaço de estados \mathbb{N}^V e cuja evolução (estado no tempo t) é denotada por η_t . Para cada vértice $x \in V$, $\eta_t(x) = i$ significa que no tempo t existem i indivíduos no vértice x . Em Junior et al. (2017) é apresentada uma versão desse modelo onde o grafo é uma árvore homogênea, cada colônia cresce segundo um processo de contagem genérico e o tempo de duração de uma colônia é uma variável aleatória contínua não-negativa genérica.

Definição 2.3. *Seja η_t um processo $CC(\mathcal{G}, \lambda, p)$ começando com um número finito de*

colônias e $|\eta_t| := \sum_{x \in V} \eta_t(x)$. Diz-se que η_t sobrevive (globalmente) se

$$\mathbb{P}(|\eta_t| \geq 1 \text{ para todo } t \geq 0) > 0.$$

Caso contrário, é dito que o processo morre (globalmente).

Definição 2.4. Seja η_t um processo $CC(\mathcal{G}, \lambda, p)$. Diz-se que η_t morre localmente se para todo vértice $x \in V$ existe um tempo (aleatório) finito T tal que $\eta_t(x) = 0$ para todo $t > T$. Caso contrário, é dito que o processo sobrevive localmente.

Note que morte local significa que todo vértice é colonizado somente um número finito de vezes.

Definição 2.5. Seja η_t um processo $CC(\mathcal{G}, \lambda, p)$ com $0 < p < 1$. Diz-se que η_t exibe transição de fase (em λ) se $0 < \lambda_c(p, \mathcal{G}) < \infty$.

Observe que o número de indivíduos por vértice não é limitado. Assim, é concebível pensar que o modelo sobreviva em um grafo finito. O resultado a seguir mostra que isso não ocorre e portanto não existe transição de fase do modelo em grafos finitos.

Proposição 2.6. Para $p < 1$ e todo grafo finito \mathcal{G} , o processo de colonização e colapso $CC(\mathcal{G}, \lambda, p)$ morre.

No seguinte teorema são apresentadas condições suficientes para extinção (global e local) no modelo de colonização e colapso sobre grafos infinitos.

Teorema 2.7. Seja \mathcal{G} um grafo m -regular, η_t um processo $CC(\mathcal{G}, \lambda, p)$ e

$$\mu(m) = m - \frac{m}{\lambda} \sum_{k \geq 1} B\left(1 + \frac{1}{\lambda}, k\right) \left(1 - \frac{p}{m}\right)^k,$$

onde $B(a, b) = \int_0^1 u^{a-1}(1-u)^{b-1} du$ é a função Beta.

(i) Se $\mu(m) \leq 1$ e $|\eta_0| < \infty$, então η_t morre local e globalmente.

(ii) Seja $\mathcal{G} = \mathbb{Z}^d, d \geq 1$. Se $\mu(2d) < 1$ e $\eta_0(x) \leq 1$ para todo x em \mathbb{Z}^d então η_t morre localmente.

(iii) Seja $\mathcal{G} = \mathbb{T}^d, d \geq 1$. Se $\mu(d+1) < 1/d$ e $\eta_0(x) \leq 1$ para cada x em \mathbb{T}^d então η_t morre localmente.

No resultado a seguir são apresentadas condições suficientes para a sobrevivência do processo de colonização e colapso sobre alguns grafos infinitos.

Teorema 2.8. Para $p > 0$ e $\lambda := \lambda(p, \mathcal{G}) > 0$ suficientemente grande, o processo $CC(\mathcal{G}, \lambda, p)$ com $\mathcal{G} = \mathbb{Z}^d$ ou \mathbb{T}^d sobrevive global e localmente.

Os autores introduzem uma versão não espacial do modelo e comparam com o modelo de catástrofe descrito por Brockwell et al. (2016). Considere um modelo para o qual todo indivíduo dá nascimentos a taxa λ e morre a taxa μ . O processo inicia com um único indivíduo e portanto com uma única colônia. Quando uma colônia colapsa seus indivíduos sobrevivem com probabilidade p e morrem com probabilidade $1 - p$, independentemente do que ocorre com outros indivíduos. Todo indivíduo sobrevivente gera uma nova colônia a qual eventualmente colapsa. Colônias colapsam independentemente de outras a taxa $a > 0$. Tem-se a seguinte propriedade deste modelo de múltiplas colônias.

Proposição 2.9. O modelo de múltiplas colônias sobrevive com probabilidade positiva se e somente se

$$p\mathbb{E}[\exp((\lambda - \mu)T)] > 1,$$

onde T tem distribuição exponencial com taxa a .

A conclusão é que é mais provável para o modelo com múltiplas colônias sobreviver do que o modelo com uma única colônia. Isto é, morar em diversas colônias pequenas é uma estratégia melhor do que morar em uma única grande colônia. Note que esta conclusão não é óbvia. O modelo com uma única colônia tem taxa de catástrofe a enquanto o modelo com múltiplas colônias tem taxa de catástrofe na se existem n colônias. Além disso, uma catástrofe tem maior probabilidade de acabar com uma colônia menor do que com uma maior. Por outro lado, múltiplas colônias dão várias possibilidades para sobrevivência e isso acaba sendo uma importante vantagem do modelo com múltiplas colônias em relação ao modelo com uma única colônia.

2.4 Modelo de Artalejo

O modelo apresentado em Artalejo et al. (2007) é um modelo estocástico de uma única colônia. A colônia dá nascimento a um novo indivíduo a taxa $\lambda > 0$, e colapsos ocorrem a taxa μ . Se na ocorrência de um colapso o tamanho da população é i , esta se reduz a um tamanho j com probabilidade μ_{ij} . O valor μ_{ij} é determinado pelo tipo de colapso. São considerados dois tipos de colapsos descritos a seguir.

- *Catástrofe Binomial*: Os indivíduos são expostos ao efeito catastrófico simultaneamente. Cada indivíduo sobrevive com probabilidade $p < 1$ (morre com probabilidade $q = 1 - p$), independentemente de qualquer outra coisa. Isto é,

$$\mu_{ij}^B = \binom{i}{j} p^j q^{i-j}, \quad 0 \leq j \leq i.$$

O modelo de catástrofe binomial assume que nos tempos de catástrofe cada indivíduo sobrevive com probabilidade $p \in (0, 1)$, independentemente de qualquer outra coisa. Modelos com catástrofes binomiais foram introduzidas por Brockwell et al. (2016) e têm sido bastante estudados, veja por exemplo, Artalejo et al. (2007), Kapodistria et al. (2016) e as referências contidas nesses artigos.

- *Catástrofe Geométrica*: Os indivíduos são expostos ao efeito catastrófico sequencialmente e o declínio na população para no primeiro indivíduo que sobrevive, ou quando toda a população torna-se extinta. Se cada indivíduo sobrevive com probabilidade p , então

$$\mu_{ij}^G = \begin{cases} q^i, & j = 0 \\ pq^{i-j}, & 1 \leq j \leq i. \end{cases}$$

A catástrofe geométrica corresponde aos casos onde o declínio da população cessa imediatamente após algum indivíduo sobreviver a catástrofe. Isto ocorre em algumas formas de catástrofes epidêmicas e quando a catástrofe tem propagação sequencial semelhante a modelos presa-predador. Os autores consideram os modelos com catástrofe binomial e catástrofe geométrica de forma isolada. Em outras palavras, apresentam uma versão do processo para cada tipo de catástrofe.

3 Resultados

Nessa seção são apresentados resultados de Junior et al. (2016). Os autores analisam a dispersão como estratégia de dispersão. Para isso comparam três modelos descritos nas subseções que seguem.

3.1 Modelo sem restrições Espaciais

Em Artalejo et al. (2007) os autores consideram os modelos com catástrofe binomial e catástrofe geométrica de forma isolada. Em outras palavras, apresentam uma versão do processo para cada tipo de catástrofe. Dado que é possível pensar que diferentes tipos de colapsos podem afetar uma mesma população, considera-se aqui um modelo que mistura os dois tipos (catástrofes binomiais e geométricas). Isto é, nos tempos de colapso, com probabilidade r o colapso é do tipo catástrofe geométrica e com probabilidade $1 - r$ o colapso é do tipo catástrofe binomial. Assim,

$$\mu_{ij} := r\mu_{ij}^G + (1 - r)\mu_{ij}^B.$$

Assuma também que $\mu = 1$. Formalmente falando, o tamanho da população (número de indivíduos na colônia) no tempo t é um processo de Markov a tempo contínuo $\{X(t) : t \geq 0\}$ com gerador infinitesimal $(q_{ij})_{i,j \geq 0}$ dado por

$$q_{ij} = \begin{cases} \lambda, & j = i + 1, i \geq 0, \\ \mu_{ij}, & 0 \leq j < i, \\ -(\lambda + \sum_{j=0}^{i-1} \mu_{ij}), & i = j, \\ 0 & \text{em outro caso.} \end{cases}$$

Assuma $X(0) = 1$, denote por $MC^1(r, \lambda, p)$ o modelo descrito pelo processo $\{X(t) : t \geq 0\}$. Fazendo $r = 0$ e $r = 1$, obtém-se os modelos considerados em Artalejo et al. (2007).

Teorema 3.1. *Seja $X(t)$ um processo $MC^1(r, \lambda, p)$, com $\lambda > 0$ e $0 < p < 1$. Então, a*

extinção (isto é, $X(t) = 0$ para algum $t > 0$) ocorre com probabilidade

$$\rho_1(r) = \begin{cases} 1 & , \text{ se } r < 1 \\ \min \left\{ \frac{1-p}{\lambda p}, 1 \right\} & , \text{ se } r = 1. \end{cases}$$

Além disso, se $r < 1$, ou $r = 1$ e $\lambda p < 1 - p$, o tempo até a extinção do processo tem média finita.

Observação 3.2. O resultado do Teorema 3.1 já tinha sido provado em Artalejo et al. (2007) nos casos $r = 0$ e $r = 1$. A partir deste resultado pode-se ver que a sobrevivência é possível apenas quando as catástrofes são puramente geométricas ($r = 1$). A razão para isso é bastante clara: se $r < 1$ as catástrofes binomiais acontecem a taxa $(1 - r) > 0$, assim mesmo se considerar $p = 1$ quando as catástrofes são geométricas, a população vai sofrer extinção como provado em Artalejo et al. ? para o caso $r = 0$.

3.2 Modelo com dispersão e sem restrições espaciais

Considere uma população de indivíduos divididos em colônias independentes. Cada colônia começa com um indivíduo. O número de indivíduos na colônia aumenta segundo um processo de Poisson de taxa $\lambda > 0$. De forma independente, a cada colônia é associado um tempo exponencial de média 1. Quando o tempo exponencial ocorre, a colônia colapsa (ocorre uma catástrofe binomial ou geométrica) e cada um dos sobreviventes ao colapso começa uma nova colônia de forma independente. Após o colapso da colônia, ela desaparece. Denote esta dinâmica por $MC^2(r, \lambda, p)$ iniciando com um único indivíduo (uma colônia).

No resultado a seguir são estabelecidas condições necessárias e suficientes para a sobrevivência (existência de colônias em todo instante de tempo) no modelo $MC^2(r, \lambda, p)$.

Teorema 3.3. O modelo $MC^2(r, \lambda, p)$ sobrevive com probabilidade positiva se e somente se

$$\frac{p(\lambda + 1)^2 r}{\lambda p + 1} + p(\lambda + 1)(1 - r) > 1. \quad (1)$$

O Teorema 3.3 mostra que, ao contrário do que acontece no $MC^1(r, \lambda, p)$, em $MC^2(r, \lambda, p)$ a população pode sobreviver mesmo quando catástrofes binomiais são possíveis ($r < 1$).

Veja o exemplo 3.5. Em particular, se $r = 0$ (só catástrofes binomiais) o processo sobrevive com probabilidade positiva quando $p(\lambda + 1) > 1$.

No teorema a seguir é apresentado o modo de calcular a probabilidade de extinção, ou seja, a probabilidade de que em algum instante de tempo o sistema fique sem colônias.

Teorema 3.4. *Seja $\rho_2(r)$ a probabilidade de extinção no modelo $MC^2(r, \lambda, p)$. Então $\rho_2(r)$ é a menor solução não negativa de*

$$\phi(s) := \frac{1}{1 + \lambda p} \left[q + \frac{r(\lambda + 1)ps}{1 + \lambda - \lambda s} + \frac{(1 - r)(\lambda + 1)ps}{1 + \lambda p - \lambda ps} \right] = s \quad (2)$$

Exemplo 3.5. *Considere $MC^2(r, 1, 2/5)$. Neste caso,*

$$\phi(s) = \frac{3}{7} + \frac{4rs}{14 - 7s} + \frac{20(1 - r)s}{49 - 14s}.$$

Além disso, a menor solução não negativa de $\phi(s) = s$, é dada por

$$\rho_2(r) = \begin{cases} 1 & , r \leq 7/12 \\ \frac{12r + 49 - \sqrt{144r^2 + 1176r + 49}}{28} & , r > 7/12. \end{cases}$$

Observação 3.6. *Nos casos $r = 0$ (apenas catástrofes binomiais) e $r = 1$ (apenas catástrofes geométricas) a solução de (2) é simples:*

$$\rho_2(0) = \min \left\{ \frac{q}{\lambda p}, 1 \right\} \quad e \quad \rho_2(1) = \min \left\{ \frac{q(\lambda + 1)}{\lambda(1 + \lambda p)}, 1 \right\}.$$

Observe ainda que $\rho_2(0) \geq \rho_2(1)$ onde a desigualdade é estrita quando $(1 + \lambda + \lambda^2)^{-1} < p < 1$. Mais ainda,

- Se $p < \frac{1}{1 + \lambda + \lambda^2}$ então $\rho_2(0) = \rho_2(1) = 1$.
- Se $\frac{1}{1 + \lambda + \lambda^2} < p < \frac{1}{1 + \lambda}$ então $\rho_2(0) = 1$ e $\rho_2(1) = \frac{q(\lambda + 1)}{\lambda(1 + \lambda p)}$.
- Se $p > \frac{1}{1 + \lambda}$ então $\rho_2(0) = \frac{q}{\lambda p}$ e $\rho_2(1) = \frac{q(\lambda + 1)}{\lambda(1 + \lambda p)}$.

Note que, assim como no modelo sem dispersão, a catástrofe binomial é mais severa que a catástrofe geométrica.

Observação 3.7. Observe que $\rho_1(r) \geq \rho_2(r)$. Em particular, se $r < 1$ a desigualdade é estrita quando (1) ocorre. Além disso, $\rho_1(1) > \rho_2(1)$ para $\lambda(1 + \lambda p) > q(\lambda + 1)$. Portanto, quando não existem restrições espaciais a dispersão é uma boa estratégia para a sobrevivência da população. Isto coincide com os resultados encontrados no modelo de Schinazi ? e no modelo de colonização e colapso de Roldan-Machado-Schinazzi.

3.3 Modelo com dispersão e restrições espaciais

Seja \mathcal{G} um grafo m -regular. Cada vértice de \mathcal{G} pode estar vazio ou ocupado por um número finito de colônias. Cada colônia começa com um indivíduo. O número de indivíduos na colônia aumenta segundo um processo de Poisson de taxa $\lambda > 0$. Cada colônia é associada a um tempo exponencial de média 1. Quando o tempo exponencial ocorre, a colônia colapsa (ocorre uma catástrofe binomial ou geométrica). Cada um dos sobreviventes ao colapso tenta criar uma nova colônia em um dos vértices vizinhos escolhido aleatoriamente (mesmo que esse vértice já esteja ocupado). Entre os sobreviventes que migram para um mesmo vértice tentando criar uma nova colônia, somente um tem sucesso; os demais morrem. Assim, neste processo quando uma colônia colapsa, esta é substituída por $0, 1, \dots$ ou m colônias. Denote este modelo por $MC_m^3(r, \lambda, p)$ e assumamos que o processo começa com um único indivíduo em um vértice de \mathcal{G} .

Teorema 3.8. O modelo $MC_m^3(r, \lambda, p)$ sobrevive com probabilidade positiva, se e somente se,

$$\frac{mp(1 + \lambda)^2 r}{(m + \lambda)(\lambda p + 1)} + \frac{mp(1 + \lambda)(1 - r)}{m + \lambda p} > 1.$$

O Teorema 3.8 dá uma condição necessária e suficiente para a sobrevivência da população no modelo $MC_m^3(r, \lambda, p)$. O próximo teorema mostra uma forma para calcular a probabilidade de extinção do processo. Esta é uma das raízes de um polinômio de grau m .

Teorema 3.9. Seja $\rho_3(r)$ a probabilidade de extinção no modelo $MC_m^3(r, \lambda, p)$. Então $\rho_3(r)$ é a menor solução não negativa de

$$\psi(s) := r\psi_G(s) + (1 - r)\psi_B(s) = s,$$

onde

$$\psi_B(s) := \frac{q}{1 + \lambda p} + \frac{m(1 + \lambda)}{\lambda} \sum_{k=1}^m \binom{m}{k} \left[\frac{-\lambda p s}{m(1 + \lambda p)} \right]^k \sum_{j=0}^k \binom{k}{j} \frac{(-1)^j j^k}{m(1 + \lambda p) - \lambda p j},$$

$$\psi_G(s) := \frac{q}{1 + \lambda p} + \frac{(1 + \lambda) p s}{\lambda p + 1} \sum_{k=1}^m \binom{m}{k} \left[\frac{-\lambda s}{m(1 + \lambda)} \right]^{k-1} \sum_{j=0}^k \binom{k}{j} \frac{(-1)^{j-1} j^k}{m(1 + \lambda) - \lambda j}.$$

Exemplo 3.10. Considere $MC_3^3(r, 1, 2/3)$. Então

$$\psi(s) = \left(\frac{126r}{3575} + \frac{32}{715} \right) s^3 + \left(\frac{138r}{3575} + \frac{144}{715} \right) s^2 + \left(\frac{36}{65} - \frac{24r}{325} \right) s + \frac{1}{5}.$$

Logo, a menor solução não negativa de $\psi(s) = s$, é dada por

$$\rho_3(r) = \frac{-440 - 132r + \sqrt{22(14000 + 9375r + 792r^2)}}{2(80 + 63r)}.$$

4 Conclusão

A fim de avaliar a dispersão como estratégia de sobrevivência, defina

$$\lambda_i^r(p) := \inf\{\lambda : \mathbb{P}[MC^i(r, \lambda, p) \text{ sobrevive}] > 0\}, \quad \text{para } i = 1, 2,$$

$$\text{e } \lambda_3^r(p, m) := \inf\{\lambda : \mathbb{P}[MC_m^3(r, \lambda, p) \text{ sobrevive}] > 0\}.$$

Observe que para $i = 1, 2, 3$ e qualquer m , quando $0 < \lambda_i^r(p) < \infty$ para $0 < p < 1$, o gráfico da função $\lambda_i^r(p)$ divide o espaço dos parâmetros $\lambda \times p$ em duas regiões. Para valores de (λ, p) acima da curva $\lambda_i^r(p)$ o modelo $MC^i(r, \lambda, p)$ sobrevive com probabilidade positiva, e para valores de (λ, p) abaixo da curva $\lambda_i^r(p)$ o modelo $MC^i(r, \lambda, p)$ torna-se extinto com probabilidade 1.

No próximo resultado são estabelecidas algumas propriedades de $\lambda_2^r(p)$ e $\lambda_3^r(p, m)$.

Proposição 4.1. Seja $0 \leq r \leq 1$ e $0 < p < 1$. Então,

- (i) $0 < \lambda_2^r(p) < \lambda_3^r(p, m + 1) < \lambda_3^r(p, m) < \infty$, para todo $m \geq 2$. Além disso, $\lambda_3^r(p, 1) = \infty$.

$$(ii) \lim_{m \rightarrow \infty} \lambda_3^r(p, m) = \lambda_2^r(p).$$

A partir de agora assuma $0 < p < 1$. Do Teorema 3.1 segue que se $r < 1$ então $\lambda_1^r(p) = \infty$, e pela Proposição 4.1 obtemos que

$$\lambda_2^r(p) < \lambda_3^r(p, m) < \lambda_1^r(p),$$

para todo $m \geq 2$. Assim, quando catástrofes binomiais são possíveis ($r < 1$), a dispersão é sempre uma boa estratégia para a sobrevivência da população, com ou sem restrições espaciais.

Na ausência de catástrofes binomiais ($r = 1$), ou seja, quando os colapsos são apenas do tipo catástrofes geométricas, é simples calcular $\lambda_1^1(p)$, $\lambda_2^1(p)$ e $\lambda_3^1(p, m)$. Dos Teoremas 3.1, 3.3 e 3.8, tem-se que

$$\lambda_1^1(p) = \frac{1-p}{p},$$

$$\lambda_2^1(p) = \sqrt{\frac{1}{4} + \frac{1-p}{p}} - \frac{1}{2},$$

$$\lambda_3^1(p, m) = \frac{1 - mp + \sqrt{(1 - mp)^2 + 4m(m - 1)p(1 - p)}}{2p(m - 1)}.$$

Nesse caso, vale que $\lambda_2^1(p) < \lambda_1^1(p)$. No entanto, é possível ver que a dispersão nem sempre é a melhor estratégia de sobrevivência (veja a Figura 1). Observe que

$$\lambda_3^1(p, m) \leq \lambda_1^1(p) \iff p \leq 1 - \frac{1}{m-1}.$$

Portanto, quando os colapsos são sempre do tipo catástrofe geométrica, a existência de vantagem ou não da dispersão como estratégia de sobrevivência depende das restrições espaciais (m) e da probabilidade (p) de um indivíduo exposto a um colapso sobreviver. Veja Figura 2.

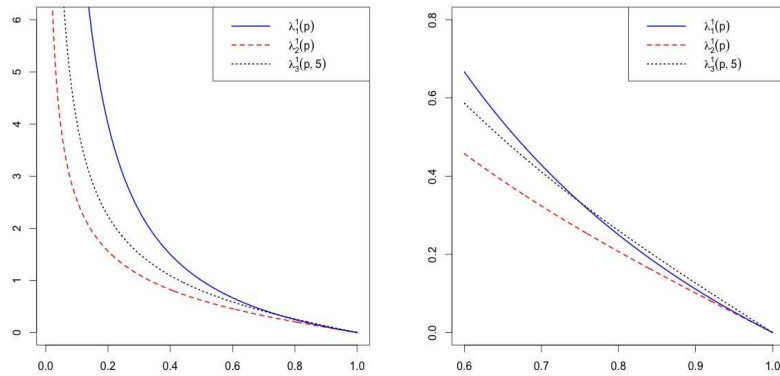


Figura 1: Gráficos de $\lambda_1^1(p)$, $\lambda_2^1(p)$, $\lambda_3^1(p, 5)$

Agradecimentos

Os autores são gratos a Rinaldo Schinazi e Elcio Lebensztayn por discussões úteis sobre os modelos.

Referências

- ARTALEJO, J.R., ECONOMOU, A., LOPEZ-HERRERO, M.J. *Evaluating growth measures in an immigration process subject to binomial and geometric catastrophes*. Math. Biosci. Eng. 4(4), 573-594 (2007).
- BROCKWELL, P.J., GANI, J., RESNICK, S.I. *Birth, immigration and catastrophe processes*. J Adv. Appl. Prob. 14, 709-731 (1982).

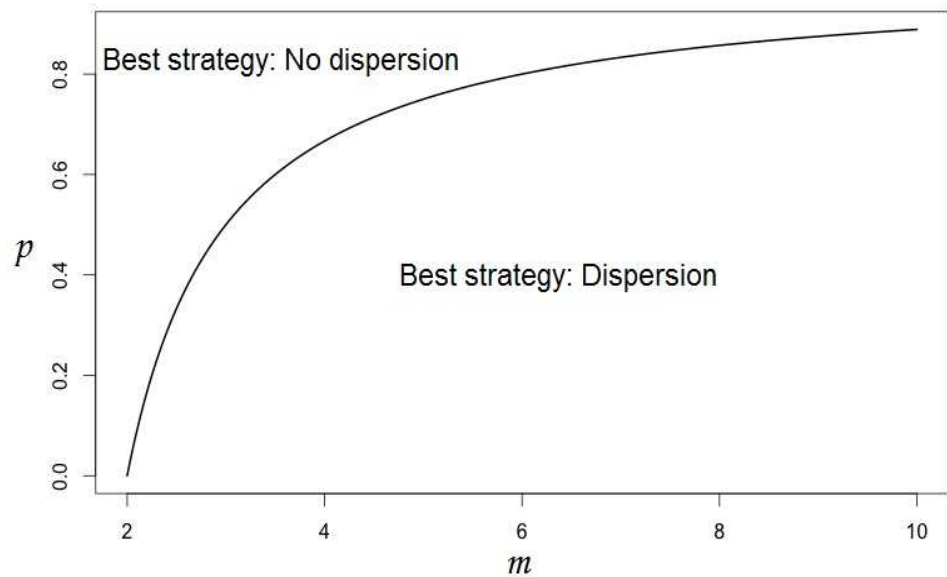


Figura 2: Curva $p = 1 - (m - 1)^{-1}$. Melhor estratégia para sobrevivência no caso $r = 1$ segundo as restrições espaciais (m) e a probabilidade individual de sobreviver a um colapso (p).

CAIRNS, B., POLLET, P.K. *Evaluating Persistence Times in Populations that are Subject to Local Catastrophes*. MODSIM 2003 International Congress on Modelling and Simulation (ed. D.A. Post), Modelling and Simulation Society of Australia and New Zealand, 747752 (2003).

KAPODISTRIA, S., PHUNG-DUC, T., RESING, J. *Linear birth/immigration-death process with binomial catastrophes*. J Probab. Eng. Inf. Sci. 30(1), 79-111 (2016).



JUNIOR,V.V.; MACHADO, F.P. ; ROLDAN-CORREA, A. *Dispersion as a Survival Strategy*. Journal of Statistical Physics August 2016, Volume 164, Issue 4, pp 937-951.

JUNIOR,V.V.; MACHADO, F.P. ; ROLDAN-CORREA, A. *Colonization and collapse on Homogeneous Trees*. arXiv:1612.06408 .

MACHADO, F.P., ROLDÁN-CORREA, A., SCHINAZI, R. *Colonization and Collapse*. arXiv:1510.02704 (2017).

SCHINAZI, R. *Does random dispersion help survival?*. Journal of Statistical Physics 159(1), 101-107 (2015).

Modelo Espaço-Temporal Funcional Misto

David Henriques da Matta^{1*}, Nancy Lopes Garcia² Mariana Rodrigues
Motta³

¹Instituto de Matemática e Estatística, Universidade Federal de Goiás, Brazil

²Departamento de Estatística, Universidade Estadual de Campinas, Brazil

³Departamento de Estatística, Universidade Estadual de Campinas, Brazil

Abstract

Nesse trabalho será apresentado a estruturação de um Modelo Espacial-Temporal Funcional Misto, que deverá ser aplicado na análise de dados espaço-temporais obtidos através de registros de índices pluviométricos coletados em diversas regiões.

Nossa proposta é apresentar esse modelo dentro de uma abordagem de dados funcionais (ADF), a qual é mais adequada para dados cuja cada observação é representada por uma função real contínua no tempo e no espaço.

No decorrer do presente trabalho será abordada além da estrutura do modelo sugerido, uma análise de simulação para a elucidação de prioris e propostas para a avaliação de alguns dos parâmetros de interesse, dentro da perspectiva Bayesiana, através de métodos de Monte Carlo (veja Sorensen e Gianola (2004)).

Palavras-chave: Modelos Mistos; Dados Funcionais; Bases B-splines.

1 Introdução

O objetivo deste trabalho é desenvolver um modelo para a análise de índices pluviométricos ao longo do tempo em uma certa região. Nosso conjunto de dados consiste de índices pluviométricos coletados pela Agência Meteorológica do Japão (JMA), registrados em cerca de 79 estações meteorológicas (localizações geográficas) ao longo do tempo (mensalmente) durante vários anos (veja <http://www.jma.go.jp/jma/indexe.html>).

Neste caso, apesar dos dados serem medidos somente em algumas coordenadas geográficas e mensalmente, a variável a ser modelada é uma variável contínua no tempo e

*Autor para correspondência: e-mail: dhmatta@ufg.br.

no espaço, sendo assim modelos multivariados não são adequados à análise de tais conjuntos de dados. Nossa proposta é utilizar uma abordagem de dados funcionais (ADF), a qual é adequada para dados cuja a j -ésima observação é representada por uma função real, $y_j(\mathbf{x}, t)$, $\mathbf{x} \in \mathbf{D}_x \subset \mathbb{R}^2$, $t \in D_T$, $j = 1, \dots, J$, isto é, cada ponto y_j é um elemento em algum espaço de funções \mathcal{F} (veja Ramsay e Silverman (2002), Rice (2004)).

Nesse trabalho vamos considerar que as curvas medidas ano a ano sejam interpretadas como medidas repetidas. Nossa abordagem será baseada em expansões por bases, isto é, cada observação $y_j(\mathbf{x}, t)$, $\mathbf{x} \in \mathbf{D}_x \subset \mathbb{R}^2$, $t \in D_T$, $j = 1, \dots, J$, pode ser muito bem aproximada por combinações lineares de produtos tensoriais de bases B-splines cúbicos avaliadas no tempo e no espaço (veja De Boor, C. (2001)). É importante, ainda, observar que a terminologia funcional é referência à estrutura pressuposta para os dados que na prática são coletados de forma discreta.

No decorrer do presente trabalho será abordada a estrutura do modelo sugerido, bem como uma análise de simulação para a elucidação de prioris e propostas para a avaliação de alguns dos parâmetros de interesse, dentro da perspectiva Bayesiana, através de métodos de Monte Carlo (veja Sorensen e Gianola (2004)).

2 Modelo Proposto

Considere o seguinte modelo para $y_j(\mathbf{x}, t)$, a j -ésima medida repetida avaliada no tempo t e na localização $\mathbf{x} = (x_1, x_2)$, onde x_1 é a latitude e x_2 a longitude,

$$y_j(\mathbf{x}, t) = \mu(\mathbf{x}, t) + \beta_j(\mathbf{x}) + \Gamma_j(t) + \epsilon_j(\mathbf{x}, t), \quad (1)$$

com $\mathbf{x} \in \mathbf{D}_x = \mathbf{D}_{x_1} \times \mathbf{D}_{x_2} \subset \mathbb{R}^2$, $t \in \mathbf{D}_T$. Aqui, $\epsilon_j(\mathbf{x}, t)$ representam erros aleatórios independentes com média zero e variância ω^2 . A média $\mu(\mathbf{x}, t)$ é independente de j e representa uma função média geral na localização e no tempo. Vamos supor que μ pode ser bem aproximada por funções base da seguinte forma

$$\mu(\mathbf{x}, t) = \sum_{k_1=1}^K \sum_{k_2=1}^K \sum_{k_3=1}^K c_{k_1 k_2 k_3} B_{k_1}^{(E_1)}(x_1) B_{k_2}^{(E_2)}(x_2) B_{k_3}^{(T)}(t), \quad (2)$$

onde $B_{k_1}^{(E_1)}(\cdot)$, $B_{k_2}^{(E_2)}(\cdot)$ e $B_{k_3}^{(T)}(\cdot)$ são bases B-splines cúbicos avaliadas em \mathbf{D}_{x_1} , \mathbf{D}_{x_2} e

\mathbf{D}_T , respectivamente. A componente $\beta_j(\mathbf{x})$ representa o efeito espacial comum a todos os tempos e todas as repetições, e será representada pelo produto tensorial de bases B-splines cúbicos avaliadas em \mathbf{D}_{x_1} e \mathbf{D}_{x_2} , tal que

$$\beta_j(\mathbf{x}) = \sum_{k_4=1}^K \sum_{k_5=2}^K \theta_{k_4 k_5}^{(j)} B_{k_4}^{(E_1)}(x_1) B_{k_5}^{(E_2)}(x_2), e \quad (3)$$

onde $\Theta^{(j)} = (\theta_{11}^{(j)}, \dots, \theta_{1K}^{(j)}, \dots, \theta_{K1}^{(j)}, \dots, \theta_{KK}^{(j)}) \sim \mathbf{N}_{K^2}(\mathbf{0}, \sigma_\theta^2 \Sigma_{\Theta_{K^2 \times K^2}}(\delta))$ com $\delta > 0$, $\sigma_\theta^2 > 0$ e

$$\begin{pmatrix} 1 & -e^{-\delta} & e^{-2\delta} & \dots & (-e^{-\delta})^{(K-1)} \\ -e^{-\delta} & 1 & -e^{-\delta} & \dots & (-e^{-\delta})^{(K-2)} \\ e^{-2\delta} & -e^{-\delta} & 1 & \dots & (-e^{-\delta})^{(K-3)} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ (-e^{-\delta})^{(K-1)} & (-e^{-\delta})^{(K-2)} & (-e^{-\delta})^{(K-3)} & \dots & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & -e^{-\delta} & e^{-2\delta} & \dots & (-e^{-\delta})^{(K-1)} \\ -e^{-\delta} & 1 & -e^{-\delta} & \dots & (-e^{-\delta})^{(K-2)} \\ e^{-2\delta} & -e^{-\delta} & 1 & \dots & (-e^{-\delta})^{(K-3)} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ (-e^{-\delta})^{(K-1)} & (-e^{-\delta})^{(K-2)} & (-e^{-\delta})^{(K-3)} & \dots & 1 \end{pmatrix}. \quad (4)$$

Por outro lado, $\Gamma_j(t)$ acomoda o efeito temporal, na repetição j , comum a todas localizações, e será representado por

$$\Gamma_j(t) = \sum_{k_6=1}^K \vartheta_{k_6}^{(j)} B_{k_6}^{(T)}(t), \quad (5)$$

onde $\boldsymbol{\vartheta}^{(j)} = (\vartheta_1^{(j)}, \dots, \vartheta_K^{(j)}) \sim \mathbf{N}_K(\mathbf{0}, \sigma_\vartheta^2 \Sigma_{\vartheta_{K \times K}}(\rho))$ com $\rho > 0$, $\sigma_\vartheta^2 > 0$ e

$$\Sigma_{\vartheta} = \begin{pmatrix} 1 & -e^{-\rho} & e^{-2\rho} & \dots & (-e^{-\rho})^{(K-1)} \\ -e^{-\rho} & 1 & -e^{-\rho} & \dots & (-e^{-\rho})^{(K-2)} \\ e^{-2\rho} & -e^{-\rho} & 1 & \dots & (-e^{-\rho})^{(K-3)} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ (-e^{-\rho})^{(K-1)} & (-e^{-\rho})^{(K-2)} & (-e^{-\rho})^{(K-3)} & \dots & 1 \end{pmatrix}. \quad (6)$$

Note que os coeficientes $\Theta^{(j)}$ e $\vartheta^{(j)}$ são efeitos aleatórios independentes para $j = 1, \dots, J$, a fim de acomodar a correlação positiva entre as medidas repetidas.

Seja $\mathbf{B}_{\tau \times K}^{(T)}$, $\mathbf{B}_{n \times K}^{(E1)}$ e $\mathbf{B}_{n \times K}^{(E2)}$ as matrizes contendo as K bases B-splines cúbicas avaliadas em $\{t_1, \dots, t_\tau\}$, $\{\mathbf{x}_{11}, \dots, \mathbf{x}_{1n}\}$ e $\{\mathbf{x}_{21}, \dots, \mathbf{x}_{2n}\}$ respectivamente. Assim, a vetorização de $\mathbf{y} := \mathbf{y}(\mathbf{x}^{(n)}, \mathbf{T})_{nJ\tau \times 1} = (\mathbf{y}^T(\mathbf{x}_1, \mathbf{T}), \dots, \mathbf{y}^T(\mathbf{x}_n, \mathbf{T}))^T$ pode ser representada por

$$\mathbf{y}_{nJ\tau \times 1} = \mathbf{M}_{nJ\tau \times K^3} \mathbf{c}_{K^3 \times 1} + \mathbf{P}_{nJ\tau \times JK^2} \Theta_{JK^2 \times 1} + \mathbf{N}_{nJ\tau \times JK} \vartheta_{JK \times 1} + \mathbf{E}_{nJ\tau \times 1}(\mathbf{x}^{(n)}, \mathbf{T}), \quad (7)$$

onde $\mathbf{c} = (c_{111}, \dots, c_{11K}, c_{121}, \dots, c_{12K}, \dots, c_{KK1}, \dots, c_{KKK})$,

$$\mathbf{M} = \begin{pmatrix} \mathbf{1}_{J \times 1} \otimes \{[(B_1^{(E1)}(x_{11}), \dots, B_K^{(E1)}(x_{11})) \otimes (B_1^{(E2)}(x_{12}), \dots, B_K^{(E2)}(x_{12}))] \otimes \mathbf{B}^{(T)}\} \\ \mathbf{1}_{J \times 1} \otimes \{[(B_1^{(E1)}(x_{21}), \dots, B_K^{(E1)}(x_{21})) \otimes (B_1^{(E2)}(x_{22}), \dots, B_K^{(E2)}(x_{22}))] \otimes \mathbf{B}^{(T)}\} \\ \mathbf{1}_{J \times 1} \otimes \{[(B_1^{(E1)}(x_{31}), \dots, B_K^{(E1)}(x_{31})) \otimes (B_1^{(E2)}(x_{32}), \dots, B_K^{(E2)}(x_{32}))] \otimes \mathbf{B}^{(T)}\} \\ \dots \\ \mathbf{1}_{J \times 1} \otimes \{[(B_1^{(E1)}(x_{n1}), \dots, B_K^{(E1)}(x_{n1})) \otimes (B_1^{(E2)}(x_{n2}), \dots, B_K^{(E2)}(x_{n2}))] \otimes \mathbf{B}^{(T)}\} \end{pmatrix}, \quad (8)$$

$\mathbf{N}_{nJ\tau \times JK} = [\mathbf{1}_{n \times 1} \otimes (\mathbf{I}_{J \times J} \otimes \mathbf{B}^{(T)})] \vartheta$ e

$$\mathbf{P} = \begin{pmatrix} \mathbf{I}_{J \times J} \otimes \{\mathbf{1}_{\tau \times 1} \otimes [(B_1^{(E1)}(x_{11}), \dots, B_K^{(E1)}(x_{11})) \otimes (B_1^{(E2)}(x_{12}), \dots, B_K^{(E2)}(x_{12}))]\} \\ \mathbf{I}_{J \times J} \otimes \{\mathbf{1}_{\tau \times 1} \otimes [(B_1^{(E1)}(x_{21}), \dots, B_K^{(E1)}(x_{21})) \otimes (B_1^{(E2)}(x_{22}), \dots, B_K^{(E2)}(x_{22}))]\} \\ \dots \\ \mathbf{I}_{J \times J} \otimes \{\mathbf{1}_{\tau \times 1} \otimes [(B_1^{(E1)}(x_{n1}), \dots, B_K^{(E1)}(x_{n1})) \otimes (B_1^{(E2)}(x_{n2}), \dots, B_K^{(E2)}(x_{n2}))]\} \end{pmatrix}. \quad (9)$$

Assume-se que $\Sigma_{\mathbf{D}_\Theta}(\phi)$ e $\Sigma_{\mathbf{D}_\vartheta}(\eta)$ tem estrutura de decaimento AR(1) para os di-

ferentes j 's. Logo, $\Theta_{JK^2 \times 1} = (\Theta^{(1)}, \dots, \Theta^{(J)}) \sim \mathbf{N}_{JK^2}(\mathbf{0}, [\Sigma_{\mathbf{D}_{\Theta_{J \times J}}(\phi)} \otimes \sigma_{\theta}^2 \Sigma_{\Theta_{K^2 \times K^2}}(\delta)])$ e $\vartheta = (\vartheta^{(1)}, \dots, \vartheta^{(J)}) \sim \mathbf{N}_{JK}(\mathbf{0}, [\Sigma_{\mathbf{D}_{\vartheta_{J \times J}}(\eta)} \otimes \sigma_{\vartheta}^2 \Sigma_{\vartheta_{K \times K}}(\rho)])$. Considerando ainda que $\mathbf{E}(\mathbf{x}^{(n)}, \mathbf{T}) \sim \mathbf{N}_{nJ\tau}(\mathbf{0}, \omega^2 \mathbf{I}_{nJ\tau \times nJ\tau})$, segue que $\mathbf{y} \sim \mathbf{N}_{nJ\tau}(\mathbf{Mc}, \mathbf{P}^T [\Sigma_{\mathbf{D}_{\Theta_{J \times J}}(\phi)} \otimes \sigma_{\theta}^2 \Sigma_{\Theta_{K^2 \times K^2}}(\delta)]) \mathbf{P} + \mathbf{N}^T [\Sigma_{\mathbf{D}_{\vartheta_{J \times J}}(\eta)} \otimes \sigma_{\vartheta}^2 \Sigma_{\vartheta_{K \times K}}(\rho)] \mathbf{N} + \omega^2 \mathbf{I}$.

A estimação via máxima verossimilhança (EMV) não se mostrou robusta devido à complexidade estrutural e ao grande número de parâmetros a serem estimados. Sendo assim, optou-se pelo paradigma Bayesiano.

3 Inferência Bayesiana e Simulação

Considere o vetor de parâmetros $\Psi = (\mathbf{c}, \omega^2, \sigma_{\theta}^2, \sigma_{\vartheta}^2, \delta, \phi, \rho, \eta)$. Seja $\Pi(\Psi|\mathbf{H})$ a distribuição a priori com \mathbf{H} sendo um vetor conhecido que representa os hiperparâmetros. Desta forma, a distribuição conjunta a posteriori de Ψ aumentada em Θ e ϑ é proporcional a

$$p(\mathbf{c}, \Theta, \vartheta, \omega^2, \sigma_{\theta}^2, \sigma_{\vartheta}^2, \delta, \phi, \rho, \eta | \mathbf{y}, \mathbf{H}) \propto p(\mathbf{y} | \mathbf{c}, \Theta, \vartheta, \omega^2) p(\Theta | \sigma_{\theta}^2, \delta, \phi) p(\vartheta | \sigma_{\vartheta}^2, \rho, \eta) \Pi(\Psi | \mathbf{H}), \quad (10)$$

onde $p(\mathbf{y} | \mathbf{c}, \Theta, \vartheta, \omega^2)$ é a densidade de uma distribuição normal com média $(\mathbf{Mc} + \mathbf{P}\Theta + \mathbf{N}\vartheta)$ e variância $(\omega^2 \mathbf{I}_{nJ\tau})$, $p(\Theta | \sigma_{\theta}^2, \delta, \phi)$ é a densidade de uma distribuição normal com média zero e variância $[\Sigma_{\mathbf{D}_{\Theta_{J \times J}}(\phi)} \otimes \sigma_{\theta}^2 \Sigma_{\Theta_{K^2 \times K^2}}(\delta)]$ e $p(\vartheta | \sigma_{\vartheta}^2, \rho, \eta)$ é a densidade de uma distribuição normal com média zero e variância $[\Sigma_{\mathbf{D}_{\vartheta_{J \times J}}(\eta)} \otimes \sigma_{\vartheta}^2 \Sigma_{\vartheta_{K \times K}}(\rho)]$.

O objetivo do estudo de simulação é a elucidação das prioris e propostas mais adequadas na obtenção de amostras das distribuições a posteriori dos parâmetros responsáveis pelas estruturas de correlação temporal e espacial $(\rho, \eta, \delta$ e $\phi)$. Sendo assim, o primeiro passo foi o estudo marginal de cada um dos parâmetros considerando-se que os demais parâmetros e também efeitos aleatórios sejam conhecidos. Na geração dos dados considerou-se $\tau = 12$, $J = 20$, $n = 35$ e $\mathbf{D}_{\mathbf{x}} = (0, 2) \times (0, 2)$. As distribuições a priori adotadas são vagas porém próprias, e seguem descritas na Tabela 1.

Variável	Dist. Priori
δ, ρ	Uniforme(a, b), $a = 0.1$ e $b = 4.6$
η, ϕ	Uniforme(c, d), $c = 0.3$ e $d = 3$

Table 1: Distribuições a priori

Os valores a , b , c e d apresentados na Tabela 1 excluem os casos independente e altamente correlacionados. Considere que $\tilde{\Theta} = [\mathbf{P}^T \mathbf{P} + (\sigma_\theta^2 / \omega^2 \Sigma_{\mathbf{D}_\theta} \otimes \Sigma_\Theta)^{-1}]^{-1}$, $\tilde{\vartheta} = [\mathbf{N}^T \mathbf{N} + (\sigma_\vartheta^2 / \omega^2 \Sigma_{\mathbf{D}_\vartheta} \otimes \Sigma_\Theta)^{-1}]^{-1}$ e $\tilde{\omega}^2 = (\mathbf{y} - \mathbf{M}\mathbf{c} - \mathbf{P}\Theta - \mathbf{N}\vartheta)^T (\mathbf{y} - \mathbf{M}\mathbf{c} - \mathbf{P}\Theta - \mathbf{N}\vartheta)$. As distribuições marginais condicionadas nos demais parâmetros e também nos efeitos aleatórios são apresentadas na Tabela 2.

Parâmetro	Dist. Marginal Condicional
δ	$ \Sigma_\Theta ^{-0.5J} \exp\left(\frac{\Theta^T (\Sigma_{\mathbf{D}_\theta} \otimes \Sigma_\Theta)^{-1} \Theta}{2\sigma_\theta^2}\right) 1_{(0,1 < \delta < 4.6)}$
ϕ	$ \Sigma_{\mathbf{D}_\theta} ^{-0.5K^2} \exp\left(\frac{\Theta^T (\Sigma_{\mathbf{D}_\theta} \otimes \Sigma_\Theta)^{-1} \Theta}{2\sigma_\theta^2}\right) 1_{(0,3 < \phi < 3)}$
ρ	$ \Sigma_\vartheta ^{-0.5J} \exp\left(\frac{\vartheta^T (\Sigma_{\mathbf{D}_\vartheta} \otimes \Sigma_\vartheta)^{-1} \vartheta}{2\sigma_\vartheta^2}\right) 1_{(0,1 < \rho < 4.6)}$
η	$ \Sigma_{\mathbf{D}_\vartheta} ^{-0.5K} \exp\left(\frac{\vartheta^T (\Sigma_{\mathbf{D}_\vartheta} \otimes \Sigma_\vartheta)^{-1} \vartheta}{2\sigma_\vartheta^2}\right) 1_{(0,3 < \eta < 3)}$

Table 2: Distribuições marginais condicionais

As distribuições marginais condicionais, apresentadas na Tabela 2, não possuem formas conhecidas, daí a necessidade de se utilizar Metropolis-Hastings (veja Gamerman e Lopes (2006)). Com o auxílio da rotina *coda* implementada no software estatístico *R* foi possível verificar que um *burn-in* de 5×10^5 e espaçamento de 100 são suficientes para garantir convergência das cadeias e que as amostras sejam independentes.

Para avaliar as estimativas a posteriori dos parâmetros a partir de diferentes distribuições propostas, considere os resultados apresentados na Tabela 3. Os resultados foram obtidos a partir de 50 conjuntos de dados independentes (gerados com $\delta = 0.35$, $\phi = 0.4$, $\rho = 0.9$ e $\delta = 0.6$), utilizando-se como estimativas a média, a mediana e a moda, as quais são mais adequadas para resumir a informação dependendo da função perda utilizada.

Na Tabela 3, verifica-se que os resultados são bastante próximos para todas as propostas utilizadas no algoritmo Metropolis-Hastings. Entre todos os casos supracitados, cerca de 96% dos 50 intervalos de credibilidade gerados incluíram os respectivos verdadeiros valores dos parâmetros, sendo que as taxas de aceitação variaram em torno de 35%. Finalmente, nota-se que para as estimativas de δ e ϕ a mediana (função perda modular) é a melhor escolha, já para ρ e η temos a moda (função perda 0-1).

Par.	Dist. Proposta	Média/Var(Média)	Média/Var(Mediana)	Média/Var(Moda)
$\delta = 0.35$	Lognormal	0.348/0.0002	0.348/0.0001	0.345/0.0002
	Normal Truncada	0.347/0.0002	0.347/0.0001	0.345/0.0002
	Gama	0.347/0.0002	0.347/0.0002	0.345/0.0002
$\phi = 0.4$	Lognormal	0.401/0.0009	0.399/0.0009	0.396/0.0010
	Normal Truncada	0.402/0.0009	0.399/0.0009	0.396/0.0010
	Gama	0.402/0.0009	0.399/0.0009	0.397/0.0009
$\rho = 0.9$	Lognormal	1,092/0.131	0,995/0,085	0,883/0,041
	Normal Truncada	1,093/0,132	0,994/0,085	0,881/0,044
	Gama	1,092/0,131	0,994/0,086	0,875/0,043
$\eta = 0.6$	Lognormal	0.641/0.009	0.619/0.007	0.586/0.006
	Normal Truncada	0.642/0.009	0.621/0.008	0.589/0.006
	Gama	0.642/0.009	0.620/0.007	0.590/0.005

Table 3: Medidas resumo dos resultados do estudo de simulação baseados em 50 conjuntos de dados independentes.

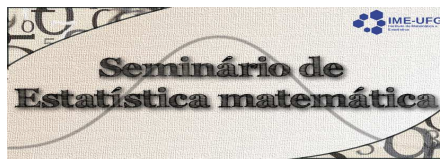
4 Conclusões e Estudos Futuros

Nesse trabalho foi posto um modelo funcional baseado na expansão em funções bases B-splines cúbicos, para dados espaço-temporais com medidas repetidas. Um estudo de simulação permitiu avaliar que a abordagem Bayesiana mostra-se eficaz na obtenção das estimativas dos parâmetros de interesse. Observa-se também que as escolhas das propostas utilizadas no algoritmo de Metropolis-Hastings não impactou o processo de amostragem. Por outro lado, as escolhas das diferentes funções perda afetam o processo de estimação.

Baseados nos resultados obtidos no estudo de simulação, pretende-se realizar a estimação conjunta de todos os parâmetros do modelo. Mais ainda, será introduzido um novo termo no modelo para acomodar uma possível interação espaço-temporal.

Referências

- [1] De Boor, C. (1978) *A Practical Guide to Splines*, New York: Springer-Verlag.
- [2] Garmeman, D. e Lopes, H. F. (2006) *Markov Chain Monte Carlo : Stochastic Simulation for Bayesian Inference*, Boca Raton: Chapman and Hall.
- [3] Ramsay, J. O. e Silverman, B. W. (2002) *Applied Functional Data Analysis: Methods and Case Studies*, New York: Springer.



- [4] Rice, J. A. (2004) *Functional and longitudinal data analysis: perspectives on smoothing*. *Statistica Sinica*, 14, 613-629.
- [5] Sorensen, D. e Gianola, D (2004) *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics*, New York: Springer-Verlag.
- [6] The R Project for Statistical Computing, <http://www.rproject.org>.

Improved hypothesis testing in a general multivariate elliptical model

Tatiane Ferreira do Nascimento Melo da Silva,¹ Silvia Lopes de Paula Ferrari
e Alexandre Galvão Patriota²

¹Instituto de Matemática e Estatística, Universidade Federal de Goiás, Brasil

²Instituto de Matemática e Estatística, Universidade de São Paulo, Brasil

Resumo

Likelihood inference is usually based on the first order asymptotic theory, which can lead to inaccurate inference when the sample is small. In general, this is the case of the signed likelihood ratio test, whose statistic has asymptotic standard normal distribution under the null hypothesis, with an error of order $n^{-1/2}$, where n is the size sample. In order to improve this approach, Barndorff-Nielsen (1986) proposed a new test statistic, that under the null hypothesis is asymptotically standard normal distributed with error of order $n^{-3/2}$. Barndorff-Nielsen's adjustment is applied when the parameter of interest is scalar. Skovgaard (2001) developed an extension of this adjustment for the multidimensional case. These adjustments require a suitable ancillary statistic such that, in conjunction with the maximum likelihood estimator, is a sufficient statistic for the model. It is difficult or even impossible to find an appropriate ancillary for some statistical models. In this paper, we obtain Barndorff-Nielsen's and Skovgaard's adjustments in a general multivariate elliptical model using an approximate ancillary statistic. A general multivariate elliptical model was introduced by Lemonte and Patriota (2011). It considers that the mean vector and the dispersion matrix are indexed by the same vector of parameters. Special cases of this general multivariate elliptical model: multiple linear regressions; multivariate nonlinear regressions; mixed-effects models (Verbeke and Molenberghs, 2000); errors-in-variables models (Cheng and Van Ness, 1999); log-symmetric regression models (Vanegas and Paula, 2015). The elliptical family of distributions includes the multivariate normal as well as many other important distributions such as the multivariate Student t , power exponential, contaminated normal, logistic distributions.

*Autor para correspondência: e-mail: tafename@gmail.com, Fone: +55-62-3521-1208.



References:

- Barndorff-Nielsen, O.E. (1986). Inference on full or partial parameters, based on the standardized signed log likelihood ratio. *Biometrika*, 73, 307–322.
- Cheng, C.L. and Van Ness, J.W. (1999). *Statistical Regression with Measurement Error*. Oxford University Press, London.
- Lemonte, A.J. and Patriota, A.G. (2011). Multivariate elliptical models with general parameterization. *Statistical Methodology*, 8, 389–400.
- Skovgaard, I.M. (2001). Likelihood asymptotics. *Scandinavian Journal of Statistics*, 28, 3–32.
- Vanegas, L.H. and Paula, G.A. (2015). A semiparametric approach for joint modeling of median and skewness. *Test*, 24, 110–135.
- Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer, New York.