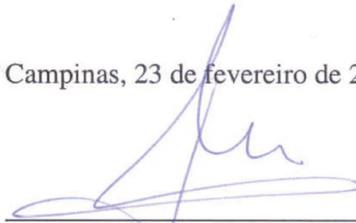


## INFERÊNCIA BAYESIANA PARA VALORES EXTREMOS

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Diego Fernando de Bernardini e aprovada pela comissão julgadora.

Campinas, 23 de fevereiro de 2010.



Prof. Dra.: Laura Leticia Ramos Rifo  
Orientadora

Banca Examinadora:

- 1 Profa. Dra.: Laura Leticia Ramos Rifo
- 2 Profa. Dra.: Verónica Andrea González-López
- 3 Prof. Dr.: Dani Gamerman

Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica, UNICAMP, como requisito parcial para obtenção do Título de MESTRE em ESTATÍSTICA.

**FICHA CATALOGRÁFICA ELABORADA PELA  
BIBLIOTECA DO IMECC DA UNICAMP**

Bibliotecária: Miriam Cristina Alves – CRB8 / 8059

Bernardini, Diego Fernando de

B456i Inferência Bayesiana para valores extremos/Diego Fernando de  
Bernardini -- Campinas, [S.P. : s.n.], 2010.

Orientadora : Laura Leticia Ramos Rifo

Dissertação (mestrado) - Universidade Estadual de Campinas,  
Instituto de Matemática, Estatística e Computação Científica.

1. Teoria Bayesiana de decisão estatística. 2. Teoria dos valores  
extremos. 3. Métodos de simulação. I. Sobrenome, Nome do Orientador.  
II. Universidade Estadual de Campinas. Instituto de Matemática,  
Estatística e Computação Científica. III. Título.

Título em inglês: Bayesian inference for extremes

Palavras-chave em inglês (Keywords): 1. Bayesian statistical decision theory. 2. Extreme  
value theory. 3. Simulation methods.

Área de concentração: Estatística

Titulação: Mestre em Estatística

Banca examinadora: Profa. Dra. Laura Leticia Ramos Rifo (IMECC-UNICAMP)  
Profa. Dra. Verónica Andrea González-López (IMECC-UNICAMP)  
Prof. Dr. Dani Gamerman (IM-UFRJ)

Data da defesa: 23/02/2010

Programa de Pós-Graduação: Mestrado em estatística

**Dissertação de Mestrado defendida em 23 de fevereiro de 2010 e aprovada**

**Pela Banca Examinadora composta pelos Profs. Drs.**



---

**Prof(a). Dr(a). LAURA LETICIA RAMOS RIFO**



---

**Prof(a). Dr(a). VERÓNICA ANDREA GONZÁLEZ-LÓPEZ**



---

**Prof(a). Dr(a). DANI GAMERMAN**

Aos meus pais, Flávio e Mariangela,  
e aos meus irmãos, Leilane e Alex.

# Agradecimentos

A Deus, pela família e amigos que tenho.

Aos meus pais e irmãos, que sempre me apoiaram diante dos desafios impostos pelas minhas escolhas e sempre contribuíram para minha formação.

Aos meus amigos, grandes companheiros, que também sempre contribuíram para minha formação.

À minha orientadora, que me acompanhou em todos os passos ao longo desta jornada.

À CAPES, que proporcionou o financiamento de minha bolsa de mestrado, viabilizando este trabalho.

Aos membros da Banca Examinadora, professores Dani Gamerman e Verónica Andrea González-López, que muito contribuíram com suas correções e sugestões.

# Resumo

Iniciamos o presente trabalho apresentando uma breve introdução à teoria de valores extremos, estudando especificamente o comportamento da variável aleatória que representa o máximo de uma sequência de variáveis aleatórias independentes e identicamente distribuídas.

Vemos que o Teorema dos Tipos Extremos (ou Teorema de Fisher-Tippett) constitui uma ferramenta fundamental no que diz respeito ao estudo do comportamento assintótico destes máximos, permitindo a modelagem de dados que representem uma sequência de observações de máximos de um determinado fenômeno ou processo aleatório, através de uma classe de distribuições conhecida como família de distribuições de Valor Extremo Generalizada (*Generalized Extreme Value* - GEV). A distribuição Gumbel, associada ao máximo de distribuições como a Normal ou Gama entre outras, é um caso particular desta família.

Torna-se interessante, assim, realizar inferência para os parâmetros desta família. Especificamente, a comparação entre os modelos Gumbel e GEV constitui o foco principal deste trabalho.

No Capítulo 1 estudamos, no contexto da inferência clássica, o método de estimação por máxima verossimilhança para estes parâmetros e um procedimento de teste de razão de verossimilhanças adequado para testar a hipótese nula que representa o modelo Gumbel contra a hipótese que representa o modelo completo GEV.

Prosseguimos, no Capítulo 2, com uma breve revisão em teoria de inferência Bayesiana obtendo inferências para o parâmetro de interesse em termos de sua distribuição a posteriori. Estudamos também a distribuição preditiva para valores futuros.

No que diz respeito à comparação de modelos, estudamos inicialmente, neste contexto bayesiano, o fator de Bayes e o fator de Bayes a posteriori.

Em seguida estudamos o *Full Bayesian Significance Test* (FBST), um teste de significância particularmente adequado para testar hipóteses precisas, como a hipótese que caracteriza o modelo Gumbel.

Além disso, estudamos outros dois critérios para comparação de modelos, o BIC (*Bayesian Information Criterion*) e o DIC (*Deviance Information Criterion*).

Estudamos as medidas de evidência especificamente no contexto da comparação entre os modelos Gumbel e GEV, bem como a distribuição preditiva, além dos intervalos de credibilidade e inferência a posteriori para os níveis de retorno associados a tempos de retorno fixos.

O Capítulo 1 e parte do Capítulo 2 fornecem os fundamentos teóricos básicos deste trabalho, e estão fortemente baseados em Coles (2001) e O'Hagan (1994).

No Capítulo 3 apresentamos o conhecido algoritmo de Metropolis-Hastings para simulação de distribuições de probabilidade e o algoritmo particular utilizado neste trabalho para a obtenção de amostras simuladas da distribuição a posteriori dos parâmetros de interesse.

No capítulo seguinte formulamos a modelagem dos dados observados de máximos, apresentando a função de verossimilhança e estabelecendo a distribuição a priori para os parâmetros.

Dois aplicações são apresentadas no Capítulo 5. A primeira delas trata das observações dos máximos trimestrais das taxas de desemprego nos Estados Unidos da América, entre o primeiro trimestre de 1994 e o primeiro trimestre de 2009. Na segunda aplicação estudamos os máximos semestrais dos níveis de maré em Newlyn, no sudoeste da Inglaterra, entre 1990 e 2007.

Finalmente, uma breve discussão é apresentada no Capítulo 6.

**Palavras-chave:** Teoria de valores extremos, inferência Bayesiana, algoritmo de Metropolis-Hastings.

# Abstract

We begin this work presenting a brief introduction to the extreme value theory, specifically studying the behavior of the random variable which represents the maximum of a sequence of independent and identically distributed random variables.

We see that the Extremal Types Theorem (or Fisher-Tippett Theorem) is a fundamental tool in the study of the asymptotic behavior of those maxima, allowing the modeling of data which represent a sequence of maxima observations of a given phenomenon or random process, through a class of distributions known as Generalized Extreme Value (GEV) family.

We are interested in making inference about the parameters of this family. Specifically, the comparison between the Gumbel and GEV models constitute the main focus of this work.

In Chapter 1 we study, in the context of classical inference, the method of maximum likelihood estimation for these parameters and likelihood ratio test procedure suitable for testing the null hypothesis associated to the Gumbel model against the hypothesis that represents the complete GEV model.

We proceed, in Chapter 2, with a brief review on Bayesian inference theory. We also studied the predictive distribution for future values.

With respect to the comparison of models, we initially study the Bayes factor and the posterior Bayes factor, in the Bayesian context.

Next we study the Full Bayesian Significance Test (FBST), a significance test particularly suitable to test precise hypotheses, such as the hypothesis characterizing the Gumbel model.

Furthermore, we study two other criteria for comparing models, the BIC (Bayesian Information Criterion) and the DIC (Deviance Information Criterion).

We study the evidence measures specifically in the context of the comparison between the Gumbel and GEV models, as well as the predictive distribution, beyond the credible intervals and posterior inference to the return levels associated with fixed return periods.

Chapter 1 and part of Chapter 2 provide the basic theoretical foundations of this

work, and are strongly based on Coles (2001) and O’Hagan (1994).

In Chapter 3 we present the well-known Metropolis-Hastings algorithm for simulation of probability distributions, and the particular algorithm used in this work to obtain simulated samples from the posterior distribution for the parameters of interest.

In the next chapter we formulate the modeling of the observed data of maximum, presenting the likelihood function and setting the prior distribution for the parameters.

Two applications are presented in Chapter 5. The first one deals with observations of the quarterly maximum for unemployment rates in the United States of America, between the first quarter of 1994 and first quarter of 2009. In the second application we studied the semiannual maximum of sea levels at Newlyn, in southwest of England, between 1990 and 2007.

Finally, a brief discussion is presented in Chapter 6.

**Key words:** Extreme value theory, Bayesian inference, Metropolis-Hastings algorithm.

# Lista de Figuras

1.1	Função de distribuição de $X_{(n)}$ , para $n = 1$ (linha tracejada), $n = 10$ (linha pontilhada) e $n = 50$ (linha contínua).	3
4.1	Gráfico de $J(\mu, \sigma, \xi)$ em função de $\xi$ próximo de zero, para $\mu \in R$ e $\sigma > 0$ fixos, obtido computacionalmente.	35
A.1	Máximos trimestrais das taxas de desemprego (%) nos EUA, entre o primeiro trimestre de 1994 e o primeiro trimestre de 2009.	52
A.2	Gráficos marginais simulados de $(\mu_i^1, g_i^1)$ , $(\sigma_i^1, g_i^1)$ e $(\xi_i^1, g_i^1)$ , para os dados das taxas de desemprego nos EUA.	53
A.3	Histogramas marginais a posteriori para $\mu$ , $\sigma$ e $\xi$ e gráficos de dispersão, obtidos por simulação, para os dados das taxas de desemprego nos EUA.	53
A.4	Primeira linha: níveis de retorno esperados a posteriori (linha tracejada), intervalos de credibilidade ( $\gamma = 0.95$ ) para os níveis de retorno (linhas contínuas) e níveis de retorno empíricos (pontos sólidos); segunda linha: função de distribuição preditiva (linha tracejada) e função de distribuição empírica (pontos sólidos); aproximações sob os modelos GEV e Gumbel, para os dados das taxas de desemprego nos EUA.	54
A.5	Máximos semestrais dos níveis de maré (metros acima de ODN) em Newlyn, sudoeste da Inglaterra, entre 1990 e 2007.	55
A.6	Gráficos marginais simulados de $(\mu_i^1, g_i^1)$ , $(\sigma_i^1, g_i^1)$ e $(\xi_i^1, g_i^1)$ , para os dados dos níveis de maré em Newlyn.	56
A.7	Histogramas marginais a posteriori para $\mu$ , $\sigma$ e $\xi$ e gráficos de dispersão, obtidos por simulação, para os dados dos níveis de maré em Newlyn.	56
A.8	Primeira linha: níveis de retorno esperados a posteriori (linha tracejada), intervalos de credibilidade ( $\gamma = 0.95$ ) para os níveis de retorno (linhas contínuas) e níveis de retorno empíricos (pontos sólidos); segunda linha: função de distribuição preditiva (linha tracejada) e função de distribuição empírica (pontos sólidos); aproximações sob os modelos GEV e Gumbel, para os dados dos níveis de maré em Newlyn.	57

# Lista de Tabelas

5.1	Medidas de evidência aproximadas para o teste de hipóteses $H_0$ : Modelo Gumbel contra $H_1$ : Modelo GEV, para os dados de taxa de desemprego nos EUA. . . . .	42
5.2	Medidas de evidência aproximadas para o teste de hipóteses $H_0$ : Modelo Gumbel contra $H_1$ : Modelo GEV, para os dados de níveis de maré em Newlyn, Inglaterra. . . . .	45

# SUMÁRIO

<b>1</b>	<b>Teoria de valores extremos</b>	<b>1</b>
1.1	Introdução . . . . .	1
1.2	Modelagem assintótica . . . . .	2
1.2.1	O teorema de Fisher-Tippett . . . . .	4
1.2.2	A Distribuição de Valor Extremo Generalizada . . . . .	8
1.3	Inferência clássica para a distribuição GEV . . . . .	10
1.3.1	Preliminares . . . . .	10
1.3.2	Estimação . . . . .	11
1.3.3	Níveis de retorno . . . . .	12
1.3.4	Teste de hipóteses . . . . .	12
<b>2</b>	<b>Inferência para valores extremos</b>	<b>15</b>
2.1	Preliminares . . . . .	15
2.2	Inferência a posteriori . . . . .	16
2.3	Fator de Bayes . . . . .	17
2.3.1	Sensibilidade . . . . .	17
2.3.2	Comparação de modelos . . . . .	17
2.3.3	Fator de Bayes a posteriori . . . . .	19
2.4	Distribuição preditiva . . . . .	20
2.5	<i>Full Bayesian Significance Test</i> - FBST . . . . .	21
2.6	Outras medidas de evidência . . . . .	23
2.7	Distribuição a posteriori e intervalos de credibilidade . . . . .	24

<b>3 Aspectos computacionais</b>	<b>26</b>
3.1 Introdução . . . . .	26
3.2 Metropolis-Hastings . . . . .	26
3.3 Algoritmo . . . . .	29
<b>4 Modelagem</b>	<b>31</b>
4.1 Preliminares . . . . .	31
4.2 Formulação do modelo . . . . .	32
4.2.1 Verossimilhança . . . . .	32
4.2.2 Distribuição a priori . . . . .	33
4.3 Aproximações . . . . .	36
<b>5 Aplicações</b>	<b>41</b>
5.1 Taxas de desemprego . . . . .	41
5.2 Níveis de maré . . . . .	44
<b>6 Conclusão</b>	<b>47</b>
<b>Referências</b>	<b>49</b>
<b>Bibliografia</b>	<b>51</b>
<b>A Figuras</b>	<b>52</b>
<b>B Programas</b>	<b>58</b>

# Capítulo 1

## Teoria de valores extremos

### 1.1 Introdução

Comumente nos deparamos com situações na natureza, nas quais se torna interessante investigar ou estudar o comportamento dos valores extremos de determinadas grandezas ou em determinados processos. Pode ser interessante, por exemplo, estudar os níveis máximos de precipitação pluviométrica ao longo de um ano em uma determinada região geográfica, ou as temperaturas máximas mensais em determinada cidade, ou ainda os máximos diários dos níveis de maré em determinado trecho do litoral.

Neste sentido, o estudo estatístico de valores extremos torna-se uma ferramenta muito útil, permitindo modelar tais situações.

Se consideramos uma sequência finita de variáveis aleatórias independentes e identicamente distribuídas, em que cada variável representa uma observação de um determinado fenômeno ou grandeza, é possível aproximar (segundo o teorema dos Tipos Extremos, para uma sequência suficientemente grande) a distribuição de uma certa padronização do máximo desta sequência por uma dentre três possíveis distribuições, chamadas distribuições de valor extremo. Tais distribuições podem ser combinadas, de modo a serem representadas através de uma única distribuição, chamada distribuição de valor extremo generalizada (*Generalized Extreme Value - GEV*). Torna-se interessante, então, realizar inferência para tal distribuição.

Especificamente, um dos objetivos principais deste trabalho é realizar a comparação entre o modelo GEV e o submodelo Gumbel. Constitui um segundo objetivo, porém não menos importante, entender e discutir a adequacidade de diversas metodologias de evidência para hipóteses precisas, em particular a medida de evidência associada ao FBST. Medidas estas que podem ser extendidas para diversos contextos com dimensões paramétricas maiores do que

a tratada neste trabalho.

Este capítulo está fortemente baseado em Coles (2001).

## 1.2 Modelagem assintótica

Nos concentraremos no estudo do comportamento da variável aleatória  $X_{(n)}$ , que representa o máximo de uma sequência de tamanho  $n$  de variáveis aleatórias,  $X_1, \dots, X_n$ , independentes e identicamente distribuídas com função de distribuição comum  $F(x)$ , onde  $X_{(n)} = \max\{X_1, \dots, X_n\}$ .

Na prática, em questões aplicadas, é natural assumir que cada uma das  $n$  observações de tal variável seja tomada em períodos regulares, em um determinado intervalo de tempo pré-definido.

Pode-se mostrar facilmente que a função de distribuição exata de  $X_{(n)}$  é  $[F(x)]^n$ . De fato,

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

e, como  $X_1, \dots, X_n$  são independentes e identicamente distribuídas, então

$$P(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x) = P(X_1 \leq x)P(X_2 \leq x) \dots P(X_n \leq x) = [F(x)]^n.$$

No que diz respeito ao comportamento assintótico da variável aleatória  $X_{(n)}$ , ocorre que tal variável converge para uma variável aleatória degenerada, quando  $n \rightarrow \infty$ . Especificamente, a distribuição caracterizada pela função de distribuição  $F^n$  converge para uma probabilidade pontual, concentrada em

$$x_0 = \sup\{x : F(x) < 1\} = \inf\{x : F(x) = 1\},$$

quando  $n \rightarrow \infty$ , se  $x_0$  existir. Observemos que para qualquer  $x < x_0$ , temos que

$$0 \leq F(x) < 1 \quad \Rightarrow \quad F^n(x) \rightarrow 0, \quad \text{quando } n \rightarrow \infty.$$

Por outro lado, para qualquer  $x \geq x_0$ ,

$$F(x) = 1 \quad \Rightarrow \quad F^n(x) \rightarrow 1, \quad \text{quando } n \rightarrow \infty,$$

de tal forma que

$$F^n(x) \rightarrow H(x), \quad \text{quando } n \rightarrow \infty,$$

onde

$$H(x) = \begin{cases} 0, & \text{se } x < x_0, \\ 1, & \text{se } x \geq x_0. \end{cases}$$

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

Como convergência em distribuição para uma constante implica convergência em probabilidade para esta constante, então

$$X_{(n)} \xrightarrow{P} x_0.$$

Mais ainda, Resnick (1987) afirma que  $X_{(n)}$  converge quase certamente a  $x_0$ .

Para ilustrar este resultado, apresentamos o exemplo seguinte.

**Exemplo 1** Consideremos uma sequência  $X_1, X_2, \dots$  de variáveis aleatórias independentes e identicamente distribuídas, com distribuição  $\mathcal{U}(0, 1)$ . Neste caso, para  $n \geq 1$  fixo, a função de distribuição do máximo  $X_{(n)}$  é dada por

$$F_{X_{(n)}}(x) = \begin{cases} x^n, & \text{se } x \in [0, 1], \\ 0, & \text{se } x \notin [0, 1]. \end{cases}$$

Observemos que o supremo do suporte de  $F$ ,  $x_0$ , é igual a 1.

Apresentamos na figura 1.1 os gráficos da função de distribuição  $F_{X_{(n)}}$ , para  $n = 1, 10, 50$ .

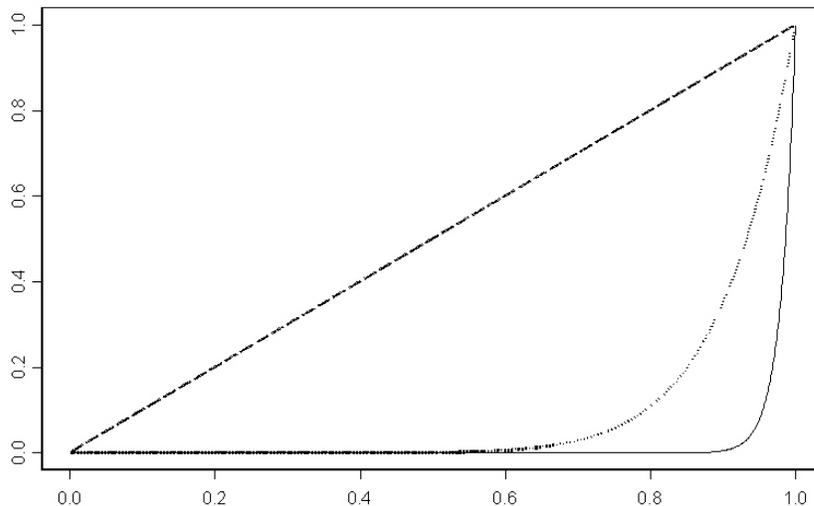


Figura 1.1: Função de distribuição de  $X_{(n)}$ , para  $n = 1$  (linha tracejada),  $n = 10$  (linha pontilhada) e  $n = 50$  (linha contínua).

Notemos que a função de distribuição tende a se concentrar em  $x_0 = 1$  à medida que aumentamos o valor  $n$ , como era de se esperar.

■

Torna-se necessário, desta forma, procurar uma normalização de  $X_{(n)}$ , digamos

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

segundo sequências de constantes  $\{a_n > 0\}_{n=1}^{\infty}$  e  $\{b_n\}_{n=1}^{\infty}$  convenientemente escolhidas, de tal forma que a distribuição de  $X_{(n)}^*$  possa convergir para uma distribuição não-degenerada.

Assim será possível aproximar, para  $n$  suficientemente grande, a distribuição de  $X_{(n)}^*$ .

### 1.2.1 O teorema de Fisher-Tippett

Uma ferramenta fundamental no estudo do comportamento assintótico de  $X_{(n)}$  é o teorema a seguir, conhecido como Teorema dos Tipos Extremos (COLES, 2001), ou Teorema de Fisher-Tippett, o qual afirma que a convergência em distribuição da variável  $X_{(n)}^*$ , quando ocorrer, se dará para uma dentre três possíveis distribuições.

**Teorema 2 (Fisher-Tippett)** *Seja  $X_{(n)} = \max\{X_1, \dots, X_n\}$ , onde  $X_1, \dots, X_n$  são variáveis aleatórias independentes e identicamente distribuídas. Se existirem sequências de constantes  $(a_n > 0)_{n=1}^{\infty}$  e  $(b_n)_{n=1}^{\infty}$  tais que*

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

onde  $G$  é uma função de distribuição não-degenerada, então  $G$  irá pertencer a uma das três seguintes famílias:

$$I : G(x) = \exp\left\{-\exp\left[-\left(\frac{x-b}{a}\right)\right]\right\}, \quad -\infty < x < \infty; \quad (1.1)$$

$$II : G(x) = \begin{cases} 0, & \text{se } x \leq b, \\ \exp\left\{-\left(\frac{x-b}{a}\right)^{-\alpha}\right\}, & \text{se } x > b; \end{cases} \quad (1.2)$$

$$III : G(x) = \begin{cases} \exp\left\{-\left[-\left(\frac{x-b}{a}\right)^\alpha\right]\right\}, & \text{se } x < b, \\ 1, & \text{se } x \geq b, \end{cases} \quad (1.3)$$

onde  $a > 0$ ,  $b \in \mathbb{R}$  e  $\alpha > 0$ .

O Teorema 2 afirma, portanto, que quando ocorrer a convergência em distribuição da variável aleatória  $X_{(n)}^*$ , a distribuição limite  $G$  irá pertencer a uma das famílias de distribuições dos Tipos I, II ou III, também conhecidas como famílias Gumbel, Fréchet e Weibull, respectivamente. Nos referimos a tais famílias como distribuições de valor extremo.

Notemos que a família Gumbel é indexada pelos parâmetros  $a > 0$  e  $b$ , ao passo que as outras duas famílias são indexadas por  $a > 0$ ,  $b$  e  $\alpha > 0$ . Em todas as famílias,  $a$  e  $b$  são chamados os parâmetros de escala e locação, respectivamente, e nas duas últimas  $\alpha$  é o chamado parâmetro de forma.

Coles (2001) apresenta uma ideia da demonstração do Teorema 2. Para tanto, o autor enuncia a definição e o teorema seguintes.

**Definição 3 (Coles, 2001)** *Uma função de distribuição  $G$  é dita ser max-estável se, para todo  $n = 2, 3, \dots$ , existem constantes  $\alpha_n > 0$  e  $\beta_n \in \mathbb{R}$  tais que*

$$[G(\alpha_n x + \beta_n)]^n = G(x).$$

Notemos que, segundo a definição, se a função de distribuição do máximo de uma sequência de variáveis aleatórias, independentes e identicamente distribuídas, difere da função de distribuição de uma variável desta sequência apenas por parâmetros de locação e escala, então a distribuição da sequência original é max-estável.

**Teorema 4 (Coles, 2001)** *Uma distribuição é max-estável se, e somente se, esta distribuição for uma distribuição de valor extremo, ou seja, dos Tipos I, II ou III.*

Prosseguimos com a ideia da demonstração do Teorema 2. Denotemos por  $X_{(nk)}$  o máximo de uma sequência de  $n \times k$  variáveis aleatórias independentes e identicamente distribuídas, que pode ser visto também como o máximo de  $k$  variáveis aleatórias, em que cada uma delas representa, por sua vez, o máximo de  $n$  variáveis aleatórias independentes e identicamente distribuídas. Denotemos por  $X_{(n)}$  o máximo usual de  $n$  variáveis aleatórias independentes e identicamente distribuídas.

Supondo que exista uma função de distribuição limite não-degenerada para  $\frac{X_{(n)} - b_n}{a_n}$ , digamos  $G$ , então teremos que, para  $n$  suficientemente grande,

$$P\left(\frac{X_{(n)} - b_n}{a_n} \leq x\right) \approx G(x).$$

Se  $n$  for suficientemente grande, então para qualquer  $k \in \mathbb{N}$  fixo,  $nk$  também será suficientemente grande, de tal forma que, analogamente,

$$P\left(\frac{X_{(nk)} - b_{nk}}{a_{nk}} \leq x\right) \approx G(x). \quad (1.4)$$

Como  $X_{(nk)}$  pode ser visto como o máximo de  $k$  variáveis aleatórias  $X_{(n)}$  independentes e com a mesma distribuição, então

$$P(X_{(nk)} \leq x) = P(X_{(n)} \leq x)^k,$$

de tal forma que

$$P\left(\frac{X_{(nk)} - b_n}{a_n} \leq x\right) = P\left(\frac{X_{(n)} - b_n}{a_n} \leq x\right)^k \approx G^k(x). \quad (1.5)$$

Da relação (1.4) obtemos

$$P(X_{(nk)} \leq x) \approx G\left(\frac{x - b_{nk}}{a_{nk}}\right),$$

e da relação (1.5) obtemos

$$P(X_{(nk)} \leq x) \approx G^k\left(\frac{x - b_n}{a_n}\right),$$

de modo que  $G$  e  $G^k$  diferem apenas por parâmetros de locação e escala.

Assim a distribuição limite  $G$  é max-estável, e portanto uma distribuição de valor extremo.

A seguir apresentamos alguns exemplos, extraídos de Coles (2001), que ilustram a convergência apresentada pelo Teorema 2, segundo a escolha de constantes  $(a_n > 0)_{n=1}^\infty$  e  $(b_n)_{n=1}^\infty$  particulares.

**Exemplo 5 (Coles, 2001)** *Seja  $X_1, X_2, \dots$  uma sequência de variáveis aleatórias independentes e identicamente distribuídas, com distribuição exponencial de parâmetro 1. A função de distribuição acumulada de um elemento desta sequência é*

$$F(x) = \begin{cases} 0, & \text{se } x \leq 0, \\ 1 - e^{-x}, & \text{se } x > 0. \end{cases}$$

Se  $a_n = 1$  e  $b_n = \log n$ , para  $n = 1, 2, \dots$ , então para cada  $x \in \mathbb{R}$  fixo, teremos

$$\begin{aligned} P\left(\frac{X_{(n)} - b_n}{a_n} \leq x\right) &= F^n(x + \log n) = [1 - e^{-(x+\log n)}]^n \\ &= \left[1 - \frac{e^{-x}}{n}\right]^n \longrightarrow \exp\{-e^{-x}\}, \end{aligned}$$

quando  $n \rightarrow \infty$ . Neste caso, a distribuição limite corresponde à distribuição Gumbel. ■

**Exemplo 6 (Coles, 2001)** *Seja  $X_1, X_2, \dots$  uma sequência de variáveis aleatórias independentes com distribuição Fréchet padrão, cuja função de distribuição é dada por*

$$F(x) = \begin{cases} 0, & \text{se } x \leq 0, \\ e^{-\frac{1}{x}}, & \text{se } x > 0. \end{cases}$$

Se tomamos  $a_n = n$  e  $b_n = 0$ , então para todo  $x > 0$  fixo,

$$P\left(\frac{X_{(n)} - b_n}{a_n} \leq x\right) = F^n(nx) = \left[e^{-\frac{1}{nx}}\right]^n = e^{-\frac{1}{x}},$$

quando  $n \rightarrow \infty$ .

Por outro lado, para cada  $x \leq 0$  fixo,

$$P\left(\frac{X_{(n)} - b_n}{a_n} \leq x\right) = F^n(nx) = 0,$$

de modo que a distribuição limite é também Fréchet padrão. ■

**Exemplo 7 (Coles, 2001)** Sejam  $X_1, X_2, \dots$  variáveis aleatórias independentes com distribuição  $\mathcal{U}(0, 1)$ . A função de distribuição de uma destas variáveis é dada por

$$F(x) = \begin{cases} 0, & \text{se } x < 0, \\ x, & \text{se } 0 \leq x \leq 1, \\ 1, & \text{se } x > 1. \end{cases}$$

Para  $a_n = \frac{1}{n}$ ,  $b_n = 1$  e todo  $x < 0$  fixo, temos

$$\begin{aligned} P\left(\frac{X_{(n)} - b_n}{a_n} \leq x\right) &= F^n\left(\frac{x}{n} + 1\right) \\ &= \left(\frac{x}{n} + 1\right)^n \rightarrow e^x, \end{aligned}$$

quando  $n \rightarrow \infty$ , pois quando  $n > -x$ , teremos  $0 \leq 1 + \frac{x}{n} \leq 1$ .

Por outro lado, para todo  $x \geq 0$ ,

$$P\left(\frac{X_{(n)} - b_n}{a_n} \leq x\right) = F^n\left(\frac{x}{n} + 1\right) = 1,$$

pois  $\left(\frac{x}{n} + 1\right) > 1$ .

Assim a distribuição limite será Weibull. ■

Tais exemplos ilustram a convergência da distribuição da padronização do máximo  $X_{(n)}$  para determinadas amostras aleatórias de distribuições específicas, e segundo a escolha de certas constantes normalizadoras ( $a_n$ ) e ( $b_n$ ) particulares em cada caso. Eventualmente, outras escolhas destas constantes, em um mesmo problema, poderão levar à convergência para uma outra distribuição não-degenerada, que deverá diferir das distribuições limite encontradas nos exemplos, possivelmente, apenas por parâmetros de locação e escala.

## 1.2.2 A Distribuição de Valor Extremo Generalizada

Vimos então que a distribuição limite de uma certa padronização do máximo  $X_{(n)}$  deverá pertencer a uma de três possíveis famílias distintas. Neste sentido, uma alternativa para a modelagem de um determinado fenômeno ou processo em questão seria a aproximação da distribuição de  $X_{(n)}^*$  por uma distribuição dentro de uma destas famílias.

Surgem, no entanto, basicamente dois grandes problemas no que diz respeito a este procedimento. Primeiramente, torna-se necessário adotar algum critério razoável para a escolha de uma destas possíveis famílias. Depois desta escolha, assume-se que a família adotada é a correta para os dados disponíveis, e ignora-se a incerteza correspondente a esta escolha ao se realizar a inferência dentro desta família selecionada a priori.

No sentido de contornar estes problemas, podemos combinar as três famílias de modo a representá-las através de uma única família (COLES, 2001). Uma distribuição pertencente a esta última família será chamada Distribuição de Valor Extremo Generalizada (GEV). Desta maneira, nos concentramos na distribuição GEV para realizar as inferências desejadas.

Especificamente, a distribuição GEV é caracterizada pela função de distribuição

$$G(x|\mu, \sigma, \xi) = \exp \left\{ - \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]_+^{-1/\xi} \right\}, \quad x \in \mathbb{R},$$

onde  $z_+ = \max\{z, 0\}$ , e  $\mu$ ,  $\sigma$  e  $\xi$  são os parâmetros de locação, escala e forma, respectivamente, e tais que  $-\infty < \mu < \infty$ ,  $\sigma > 0$  e  $-\infty < \xi < \infty$ .

O parâmetro  $\xi$  determina o tipo da distribuição. Os casos em que  $\xi > 0$  e  $\xi < 0$  caracterizam, respectivamente, os Tipos II (Fréchet) e III (Weibull), enquanto que o caso  $\xi = 0$ , que na realidade corresponde ao limite da distribuição GEV quando  $\xi \rightarrow 0$ , caracteriza o Tipo I (Gumbel),

$$G(x|\mu, \sigma, \xi = 0) = \exp \left\{ - \exp \left[ - \left( \frac{x - \mu}{\sigma} \right) \right] \right\}, \quad x \in \mathbb{R}.$$

Assim, a inferência a respeito de  $\xi$  na distribuição GEV fornece informação sobre qual dos três tipos de valor extremo se ajusta melhor aos dados disponíveis.

Desta forma, torna-se conveniente aproximar a distribuição de  $X_{(n)}^*$ , para  $n$  suficientemente grande, pela distribuição GEV. No entanto, a determinação das sequências  $(a_n)$  e  $(b_n)$  que compõem  $X_{(n)}^*$ , de modo que sua convergência ocorra, pode ser uma tarefa bastante penosa. Mas, tratando-se de uma aproximação, pode-se também aproximar a distribuição de  $X_{(n)}$  por um elemento da família GEV, não fazendo-se necessária a determinação de  $(a_n)$  e  $(b_n)$ .

Para um  $n$  fixo e suficientemente grande, pode-se escrever

$$P \left( \frac{X_{(n)} - b_n}{a_n} \leq x \right) \approx G(x),$$

em que  $G(x)$  é uma função de distribuição pertencente a família GEV.

Portanto,

$$P(X_{(n)} \leq xa_n + b_n) \approx G(x),$$

de modo que

$$P(X_{(n)} \leq y) \approx G\left(\frac{y - b_n}{a_n}\right) = G^*(y),$$

onde  $G^*(y)$  também pertence à família GEV, pois difere de  $G(y)$  somente por uma mudança nos parâmetros de locação e escala. Precisamente, se  $a$  e  $b$  são os parâmetros de escala e locação de  $G(y)$ , então  $a_n a$  e  $b_n + a_n b$  serão os parâmetros de escala e locação de  $G^*(y)$ .

Assim, torna-se natural aproximar a distribuição de  $X_{(n)}$  por uma distribuição GEV.

Como comentado anteriormente, na prática é natural assumir que a sequência de variáveis aleatórias, independentes e identicamente distribuídas, que representam as observações do processo ou grandeza de interesse são obtidas em intervalos regulares de tempo, em um período especificado. É comum, por exemplo, encontrar situações nas quais as variáveis  $X_1, \dots, X_n$  são observadas em intervalos regulares ao longo de um ano, ou seja, a cada ano são obtidas  $n$  observações, a partir das quais o máximo é calculado. Nestes casos, a variável  $X_{(n)}$  constitui o máximo anual da grandeza estudada.

Seja  $q_p$  tal que  $G(q_p) = 1 - p$ , ou seja,  $q_p$  é o quantil  $1 - p$  da distribuição GEV que modela o máximo anual de uma certa variável. Então é possível escrever este quantil em função dos parâmetros da distribuição, simplesmente invertendo sua função de distribuição, o que nos leva a

$$q_p = \begin{cases} \mu - \frac{\sigma}{\xi} [1 - \{-\log(1 - p)\}^{-\xi}], & \text{para } \xi \neq 0, \\ \mu - \sigma \log\{-\log(1 - p)\}, & \text{para } \xi = 0. \end{cases} \quad (1.6)$$

Notemos que  $p = 1 - G(q_p)$  é a probabilidade de que o máximo anual supere a quantia  $q_p$  em um ano qualquer, segundo esta modelagem. Neste sentido, se consideramos uma variável aleatória que representa o número de anos até que o máximo anual supere  $q_p$ , é razoável assumir que a distribuição desta variável seja Geométrica com parâmetro  $p$ . Assim, é esperado que  $q_p$  seja superado pelo máximo  $X_{(n)}$  em média a cada  $\frac{1}{p}$  anos. Segundo Coles (2001), a quantia  $q_p$  é chamada *nível de retorno* associado ao *tempo de retorno*  $\frac{1}{p}$ .

Através de uma análise gráfica relativamente simples utilizando os quantis  $q_p$ , é possível avaliar qual dos três tipos de distribuição de valor extremo se ajusta melhor aos dados disponíveis. Se  $y_p = -\log(1 - p)$ , então

$$q_p = \begin{cases} \mu - \frac{\sigma}{\xi} [1 - y_p^{-\xi}], & \text{para } \xi \neq 0, \\ \mu - \sigma \log y_p & \text{para } \xi = 0. \end{cases} \quad (1.7)$$

Assim, podemos construir o gráfico de  $q_p$  em função de  $\log y_p$ , de tal forma que, se o gráfico apresentar um comportamento linear, então haverá evidência de que  $\xi = 0$ . Se o gráfico for, por outro lado, côncavo ou convexo, então haverá evidência de que  $\xi < 0$  ou  $\xi > 0$ , respectivamente.

## 1.3 Inferência clássica para a distribuição GEV

### 1.3.1 Preliminares

Levando em conta a aproximação citada anteriormente, da distribuição do máximo  $X_{(n)}$  pela distribuição GEV, torna-se natural ajustar a distribuição GEV aos dados de máximos disponíveis em aplicações e problemas práticos.

Uma questão importante nestes problemas diz respeito ao tamanho  $n$  que deve ser considerado para as amostras a partir das quais são calculados os máximos  $X_{(n)}$ . Como comentado anteriormente, é comum trabalhar com o máximo anual, ou seja, considerar o máximo das observações ao longo de um ano.

Trabalhando com o máximo anual contorna-se, por exemplo, o efeito da sazonalidade nas observações, de modo que se torna razoável assumir que os máximos de cada ano sejam identicamente distribuídos, apesar de não ser satisfeita a condição do Teorema dos Tipos Extremos de que as observações ao longo de um ano sejam identicamente distribuídas. No entanto, a definição deste período está claramente vinculada à natureza dos dados. A suposição de independência deve ser abordada em cada problema específico.

Denotemos por  $Z_1, \dots, Z_m$  a sequência de máximos, cuja distribuição é assumida ser GEV. Assumimos ainda que tais variáveis são independentes. Nosso interesse reside agora na estimação dos parâmetros desta distribuição. Para tanto, nos concentraremos no método de estimação de máxima verossimilhança.

No entanto, a distribuição GEV não satisfaz as condições de regularidade necessárias para que os estimadores de máxima verossimilhança apresentem as propriedades assintóticas usuais, pois o suporte da distribuição depende dos parâmetros, por exemplo.

Segundo Smith (1985), no caso em que  $\xi > -0.5$ , os estimadores de máxima verossimilhança serão regulares, apresentando as referidas propriedades.

### 1.3.2 Estimação

Seja  $Z_1, \dots, Z_m$  uma sequência de variáveis aleatórias independentes e identicamente distribuídas com distribuição GEV, cuja função de distribuição é dada por

$$G(x) = \exp \left\{ - \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]_+^{-1/\xi} \right\}, \quad x \in \mathbb{R},$$

onde  $\mu \in \mathbb{R}$ ,  $\sigma > 0$  e  $\xi \in \mathbb{R}$ .

Diferenciando a função de distribuição, obtemos a função de densidade

$$g(x) = \frac{1}{\sigma} \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^{-(1+\frac{1}{\xi})} \exp \left\{ - \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\},$$

se  $1 + \xi \left( \frac{x - \mu}{\sigma} \right) > 0$ , para  $x \in \mathbb{R}$ . Quando  $x \in \mathbb{R}$  é tal que  $1 + \xi \left( \frac{x - \mu}{\sigma} \right) \leq 0$  a função de densidade é igual a zero.

No caso em que  $\xi \neq 0$ , a função de log-verossimilhança é dada por

$$l(\mu, \sigma, \xi) = \log \left[ \prod_{i=1}^m g(x_i) \right], \quad x_i \in \mathbb{R}, \quad i \in \{1, \dots, m\}.$$

Desta forma,

$$l(\mu, \sigma, \xi) = -m \log \sigma - \left( 1 + \frac{1}{\xi} \right) \sum_{i=1}^m \log \left[ 1 + \xi \left( \frac{x_i - \mu}{\sigma} \right) \right] - \sum_{i=1}^m \left[ 1 + \xi \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}}, \quad (1.8)$$

se  $\mu \in \mathbb{R}$ ,  $\sigma > 0$ ,  $\xi \in \mathbb{R}$  ( $\xi \neq 0$ ) satisfazem

$$1 + \xi \left( \frac{x_i - \mu}{\sigma} \right) > 0, \quad \text{para todo } i \in \{1, \dots, m\}. \quad (1.9)$$

Por outro lado, se

$$1 + \xi \left( \frac{x_i - \mu}{\sigma} \right) \leq 0, \quad \text{para pelo menos um } i \in \{1, \dots, m\}, \quad (1.10)$$

então  $l(\mu, \sigma, \xi) = 0$ .

No caso particular em que  $\xi = 0$ , a função de log-verossimilhança é obtida a partir da função de densidade correspondente à família Gumbel:

$$l(\mu, \sigma) = -m \log \sigma - \sum_{i=1}^m \left( \frac{x_i - \mu}{\sigma} \right) - \sum_{i=1}^m \exp \left\{ - \left( \frac{x_i - \mu}{\sigma} \right) \right\}, \quad \mu \in \mathbb{R}, \quad \sigma > 0. \quad (1.11)$$

Segundo o método de estimação de máxima verossimilhança, os estimadores para  $(\mu, \sigma, \xi)$  são obtidos maximizando-se a função de log-verossimilhança. Não há solução analítica para este problema de maximização, de modo que se faz necessário o uso de procedimentos computacionais para a obtenção dos estimadores. Em uma vizinhança em torno de  $\xi = 0$  deve-se maximizar  $l(\mu, \sigma)$  ao invés de  $l(\mu, \sigma, \xi)$ .

### 1.3.3 Níveis de retorno

Como vimos anteriormente, se  $G(q_p) = 1 - p = \exp\{-y_p\}$ , onde  $G$  é a função de distribuição GEV, então

$$q_p = \begin{cases} \mu - \frac{\sigma}{\xi}(1 - y_p^{-\xi}), & \text{para } \xi \neq 0, \\ \mu - \sigma \log y_p & \text{para } \xi = 0 \end{cases}$$

é o nível de retorno associado ao tempo de retorno  $\frac{1}{p}$ .

Portanto, pelo Princípio de Invariância dos estimadores de máxima verossimilhança, temos que, para  $0 < p < 1$ ,

$$\hat{q}_p = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}}(1 - y_p^{-\hat{\xi}}), & \text{para } \hat{\xi} \neq 0, \\ \hat{\mu} - \hat{\sigma} \log y_p & \text{para } \hat{\xi} = 0, \end{cases}$$

será o estimador de máxima verossimilhança de  $q_p$ , uma vez que  $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$  são os estimadores de máxima verossimilhança de  $(\mu, \sigma, \xi)$ .

### 1.3.4 Teste de hipóteses

Consideremos um modelo paramétrico com parâmetro  $\theta \in \Omega$  desconhecido e de dimensão  $d$ , para a amostra aleatória  $(x_1, \dots, x_n)$ . Suponhamos ainda que este parâmetro possa ser particionado como

$$\theta = (\theta_0, \theta_1),$$

com  $\theta_0$  de dimensão  $k$  e  $\theta_1$  de dimensão  $d - k$ .

Nosso interesse aqui reside em realizar inferência para  $\theta_0$ . Seja  $M_1$  o modelo indexado pelo vetor  $\theta$  e  $M_0$  o submodelo obtido assumindo-se que  $\theta_0 = 0$ , ou seja, indexado por  $\theta_1$  apenas. Especificamente, nosso interesse está em testar as hipóteses

$$H_0 : \text{ o modelo } M_0 \text{ é adequado.} \quad \text{contra} \quad H_1 : \text{ o modelo } M_1 \text{ é adequado.}$$

Para abordar este problema seguimos a estratégia de Coles (2001), que utiliza a estatística *deviance*

$$D = 2\{l_1 - l_0\},$$

em que  $l_1$  representa a log-verossimilhança do modelo  $M_1$  maximizada, e  $l_0$  representa a log-verossimilhança do modelo  $M_0$  maximizada. O teorema seguinte, descreve um teste da razão de verossimilhança, constituindo um procedimento para testar as hipóteses acima.

**Teorema 8 (Coles, 2001)** *Suponhamos que o modelo  $M_0$  indexado pelo parâmetro  $\theta_1$  é o submodelo de  $M_1$  indexado pelo parâmetro  $\theta = (\theta_0, \theta_1)$ , sob a suposição de que o subvetor  $\theta_0$  de dimensão  $k$  é igual a zero. Sejam  $l_0$  e  $l_1$  as log-verossimilhanças maximizadas sob os modelos  $M_0$  e  $M_1$ , respectivamente. Um procedimento de teste para a validade do modelo  $M_0$  contra o modelo  $M_1$  ao nível de significância  $\alpha$  é*

$$\text{rejeitar } M_0 \text{ em favor de } M_1 \text{ se } D > c_\alpha,$$

onde  $c_\alpha$  é o quantil  $(1-\alpha)$  da distribuição  $\chi_k^2$ .

Equivalentemente ao procedimento descrito pelo teorema anterior, a fim de testar as mesmas hipóteses, podemos calcular o *p-valor*

$$\text{p-valor} = P(W > d_0),$$

em que  $W \sim \chi_k^2$  e  $d_0$  é o valor observado da estatística  $D$ . Neste caso rejeitamos  $H_0$  em favor de  $H_1$  ao nível de significância  $\alpha$  se *p-valor*  $< \alpha$ .

No contexto da inferência para a distribuição GEV, que é indexada pelo vetor de parâmetros  $\theta = (\mu, \sigma, \xi)$ , vimos que o modelo Gumbel corresponde, na realidade, a um submodelo obtido quando  $\xi = 0$  e indexado apenas por  $\theta_1 = (\mu, \sigma)$ .

Neste sentido, o teste descrito acima constitui uma ferramenta frequentista conveniente para testar a hipótese do modelo Gumbel contra o modelo mais geral, GEV.

Podemos representar o espaço paramétrico associado à distribuição GEV por

$$\Omega = \{\theta = (\mu, \sigma, \xi) \in \mathbb{R}^3 : \sigma > 0\},$$

de modo que a família Gumbel acaba por representar uma hipótese precisa, digamos

$$\Omega_0 = \{(\mu, \sigma, \xi) \in \Omega : \xi = 0\}.$$

A estatística *deviance* fica, neste contexto,

$$D = 2(l_1 - l_0),$$

onde  $l_1 = \max\{l(\mu, \sigma, \xi) : (\mu, \sigma, \xi) \in \Omega\}$  e  $l_0 = \max\{l(\mu, \sigma) : (\mu, \sigma) \in \Omega_0\}$ , de modo que rejeitaremos  $\Omega_0$  ao nível de significância  $\alpha$ , em favor de  $\Omega$ , se  $D > c_\alpha$ , onde  $c_\alpha$  representa o quantil  $(1 - \alpha)$  da distribuição  $\chi_1^2$ .

# Capítulo 2

## Inferência para valores extremos

### 2.1 Preliminares

Uma das diferenças fundamentais entre as teorias frequentista e bayesiana de inferência estatística reside no fato de que, no enfoque desta última, admite-se uma distribuição de probabilidade para o parâmetro de interesse que indexa uma determinada família de distribuições.

A teoria de inferência Bayesiana está fundamentada basicamente em torno do teorema de Bayes, cuja formulação mais simples é dada em termos de eventos. Se consideramos uma partição<sup>1</sup>  $\{A_i\}_{i=1}^n$  do espaço amostral  $\Omega$  associado a um experimento aleatório  $\epsilon$ , então a probabilidade de um dos elementos da partição,  $A_i$  digamos, condicional a um evento  $B$  será dada, segundo o teorema de Bayes, por

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^n P(B|A_i)P(A_i)}.$$

De forma mais geral, tratando-se de variáveis aleatórias, o teorema pode ser reformulado, e enunciado como

$$f(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)f(\theta)}{\int_{\Theta} f(\mathbf{x}|\theta)f(\theta)d\theta},$$

onde  $f(\mathbf{x}|\theta)$  é a função de verossimilhança de  $\theta$  dado a amostra  $\mathbf{x}$ ,  $f(\theta)$  representa a densidade a priori para  $\theta$ , enquanto que  $f(\theta|\mathbf{x})$  representa a densidade a posteriori para  $\theta$ . As funções envolvidas no teorema podem ser funções de densidade absolutamente contínuas ou discretas.

Ainda, o denominador

$$f(\mathbf{x}) = \int_{\Theta} f(\mathbf{x}|\theta)f(\theta)d\theta$$

---

<sup>1</sup> $\{A_i\}_{i=1}^n$  constitui uma sequência exaustiva e mutuamente exclusiva de subconjuntos de  $\Omega$ .

é constante em  $\theta$ , de tal forma que podemos escrever

$$f(\theta|\mathbf{x}) \propto f(\mathbf{x}|\theta)f(\theta).$$

No caso em que admite-se distribuição discreta para  $\theta$  (a priori e a posteriori), a integral no denominador é naturalmente substituída por uma somatória em  $\theta$ .

Notemos que, segundo o teorema, a distribuição a posteriori fornece informação a respeito de  $\theta$ , combinando duas fontes distintas de informação: a informação a priori, fornecida pela distribuição a priori, e a informação contida nos dados, fornecida pela verossimilhança. Assim, o teorema pode ser visto como um mecanismo de atualização da distribuição do parâmetro  $\theta$ : a informação a respeito de  $\theta$  fornecida por sua distribuição a priori é atualizada através da informação a respeito de  $\theta$  fornecida pelos dados disponíveis, ou seja, pela verossimilhança, obtendo-se assim a distribuição a posteriori.

Parte deste capítulo está fortemente baseada em O'Hagan (1994).

## 2.2 Inferência a posteriori

Segundo o chamado método bayesiano de estimação, nos concentramos na utilização do Teorema de Bayes para determinação da distribuição a posteriori, que então permitirá a formulação de estimativas para o parâmetro de interesse e funções dele.

Primeiramente devemos estabelecer uma distribuição a priori para o parâmetro. Diversos critérios podem ser levados em conta para sua formulação, tais como o conhecimento prévio do pesquisador ou especialista a respeito do parâmetro, ou critérios teóricos do fenômeno em estudo. O importante é que esta distribuição deve ser estabelecida antes da obtenção da amostra.

Para uma amostra dada, devemos estabelecer a função de verossimilhança, que será, como no contexto frequentista, a distribuição conjunta da amostra vista como função do parâmetro.

O Teorema de Bayes permite então combinar estas duas últimas funções, fornecendo a distribuição a posteriori.

A distribuição a posteriori, por sua vez, fornece toda a informação a respeito do parâmetro. Certas características da distribuição a posteriori podem fornecer informação sobre o parâmetro de forma mais específica.

É possível obter inferências para o parâmetro comparáveis àquelas obtidas no contexto clássico segundo os métodos de estimação pontual, por intervalo e de testes de hipóteses. Medidas de locação da distribuição a posteriori, por exemplo, fornecerão informação pontual acerca do parâmetro.

Podemos determinar também intervalos no espaço paramétrico cuja probabilidade a posteriori seja algum valor de interesse, de modo que um tal intervalo fica definido como um intervalo de credibilidade para o parâmetro. Além disso, se houver interesse em estudar uma determinada hipótese a respeito do parâmetro, como por exemplo se ele pertence a um certo subconjunto do espaço paramétrico, então a probabilidade a posteriori deste subconjunto será uma ferramenta capaz de auxiliar no procedimento de decisão sobre a hipótese.

## 2.3 Fator de Bayes

### 2.3.1 Sensibilidade

O'Hagan (1994) chama de *inferências sobre  $\theta$*  ao conjunto das conclusões (quantitativas ou não) obtidas a respeito de  $\theta$  após a realização de um procedimento inferencial segundo o método bayesiano.

Suponhamos que é sugerida, por exemplo pelo pesquisador ou especialista do estudo, uma densidade a priori específica para o parâmetro de interesse  $\theta$ , digamos  $f_0(\theta)$ , pertencente a uma classe  $\Gamma$  de possíveis densidades a priori para  $\theta$ . Além disso, suponhamos também uma específica função de verossimilhança sugerida para  $\theta$ , digamos  $f_0(x|\theta)$ , pertencente à classe  $\Lambda$  de possíveis verossimilhanças. Segundo o teorema de Bayes, tais priori e verossimilhança podem ser combinadas de modo a gerar uma densidade a posteriori  $f_0(\theta|x)$ . Denotemos por  $\Gamma^*$  a classe de todas as possíveis densidades a posteriori, obtidas segundo a combinação dos elementos em  $\Gamma$  e  $\Lambda$ .

Como vimos anteriormente, pode-se obter inferências sobre  $\theta$  a partir da densidade a posteriori  $f_0(\theta|x)$ . Se as inferências obtidas a partir das outras densidades a posteriori no conjunto  $\Gamma^*$  diferirem significativamente daquelas obtidas de  $f_0(\theta|x)$ , diremos que existe sensibilidade. Se as inferências forem razoavelmente próximas, diremos que a inferência é robusta. Ou seja, dizemos que existe sensibilidade se variações na especificação de  $f_0(\theta)$  e  $f_0(x|\theta)$  provocarem grandes variações nas inferências a posteriori obtidas.

### 2.3.2 Comparação de modelos

Consideremos dois possíveis modelos (dentre os quais um deles deve ser escolhido como correto para um determinado conjunto de dados) caracterizados por uma função de verossimilhança  $f_i(x|\theta_i)$  e uma função de densidade a priori  $f_i(\theta_i)$ ,  $i = 1, 2$ .

Para cada um dos modelos, podemos calcular a distribuição marginal de  $x$ ,

$$m_i(x) = \int f_i(x|\theta_i)f_i(\theta_i)d\theta_i, \quad i = 1, 2,$$

a partir da qual O'Hagan (1994) define o fator de Bayes para o modelo 1 contra o modelo 2,

$$\frac{m_1(x)}{m_2(x)}.$$

Se  $p$  for a probabilidade a priori de que o modelo adequado é o modelo 1, por exemplo, então a razão  $\frac{p}{1-p}$  será a razão de chances a priori do modelo 1 contra o modelo 2, e desta forma a razão de chances a posteriori será o produto da razão de chances a priori pelo fator de Bayes.

O fator de Bayes assim definido poderá então ser utilizado para comparação dos modelos propostos. Um valor do fator de Bayes maior de que 1 fornecerá evidência em favor do modelo 1, ao passo que um valor menor do que 1 fornecerá evidência em favor do modelo 2.

No contexto da comparação entre os modelos GEV e Gumbel, o fator de Bayes contra  $\Omega_0$  fica definido como

$$B_{10} = \frac{f(\mathbf{x}|\Omega)}{f(\mathbf{x}|\Omega_0)} = \frac{\int_{\Omega} f(\theta|\Omega)f(\mathbf{x}|\theta)d\theta}{\int_{\Omega_0} f(\theta|\Omega_0)f(\mathbf{x}|\theta)d\theta} = \frac{\int_{\Omega} f(\mu, \sigma, \xi|\Omega)l(\mu, \sigma, \xi)d(\mu, \sigma, \xi)}{\int_{\Omega_0} f(\mu, \sigma|\Omega_0)l(\mu, \sigma)d(\mu, \sigma)}. \quad (2.1)$$

Jeffreys (1961) sugere a utilização de  $2 \log B_{10}$  para quantificar a evidência contra  $\Omega_0$ , e estabelece as seguintes classes para esta quantia e a força da evidência associada:

- se  $2 \log B_{10}$  está entre 0 e 2, há pouca evidência contra  $\Omega_0$ ;
- se  $2 \log B_{10}$  está entre 2 e 6, há evidência positiva contra  $\Omega_0$ ;
- se  $2 \log B_{10}$  está entre 6 e 10, há forte evidência contra  $\Omega_0$ ;
- se  $2 \log B_{10}$  é maior do que 10, há muito forte evidência contra  $\Omega_0$ .

O'Hagan (1994) também apresenta uma definição do fator de Bayes destinado a comparação de distribuições a priori. Para tanto, o autor define o conjunto  $\Gamma$  de possíveis densidades a priori como

$$\Gamma = \{f_{\phi}(\theta) : \phi \in \Phi\},$$

onde  $\phi$  denota um hiperparâmetro da distribuição a priori, e ainda admite uma distribuição a priori para o hiperparâmetro.

### 2.3.3 Fator de Bayes a posteriori

Aitkin (1991) define o fator de Bayes a posteriori, uma ferramenta análoga ao fator de Bayes convencional, para realizar a comparação entre dois ou mais modelos.

Como no contexto abordado até o momento, consideremos apenas dois modelos, caracterizados por uma função de verossimilhança  $f_i(x|\theta_i)$  e uma função de densidade a priori  $f_i(\theta_i)$ ,  $i = 1, 2$ . Como sabemos, a cada modelo está associada uma função de densidade a posteriori, digamos  $f_i(\theta_i|x)$ ,  $i = 1, 2$ .

O fator de Bayes a posteriori para o modelo 1 contra o modelo 2 fica definido como

$$FBP = \frac{\int f_1(x|\theta_1)f_1(\theta_1|x)d\theta_1}{\int f_2(x|\theta_2)f_2(\theta_2|x)d\theta_2}. \quad (2.2)$$

Notemos que tanto o numerador quanto o denominador nesta definição podem ser vistos como esperanças a posteriori das respectivas funções de verossimilhança sob cada modelo, ao passo que no fator de Bayes convencional poderiam ser vistos como esperanças a priori das mesmas quantidades.

A principal vantagem da utilização do fator de Bayes a posteriori sobre o convencional, citada pelo autor, é o fato de que este último apresenta maior sensibilidade à especificação da distribuição a priori para os parâmetros, uma vez que apresenta em sua construção a média a priori da função de verossimilhança. Naturalmente, seria razoável utilizar critérios de comparação menos sensíveis à especificação de priors.

O autor sugere que o FBP poderia ser utilizado na comparação de modelos gerais (ou arbitrários), e não somente na comparação entre modelos e submodelos. Sugere ainda que valores do FBP maiores do que 20, 100 e 1000 forneceriam evidência forte, muito forte e opressiva contra o modelo 2 e em favor do modelo 1.

A definição do FBP promoveu uma vasta discussão a respeito de sua coerência e validade como medida de comparação de modelos. Diversos pesquisadores apresentam sérias críticas ao FBP. Talvez a principal delas resida na ideia de que esta medida de evidência possa estar utilizando, em sua avaliação, a mesma amostra duas vezes: primeiramente a amostra é utilizada para a obtenção da distribuição a posteriori, e depois novamente utilizada ao se calcular a esperança a posteriori da função de verossimilhança.

Devemos ressaltar a opinião do autor, de que a afirmação de uma das contradições, decorrente dessa suposta dupla utilização da amostra, não passa de uma questão de cunho filosófico dos críticos do método.

Neste trabalho não entramos no mérito desta discussão especificamente, e nos limitamos apenas a utilizar o FBP como mais uma medida para comparação de modelos.

No contexto da comparação entre os modelos GEV e Gumbel, o fator de Bayes a posteriori contra  $\Omega_0$  (modelo Gumbel) fica dado por

$$FBP = \frac{\int_{\Omega} f(\mathbf{x}|\theta)f(\theta|\mathbf{x}, \Omega)d\theta}{\int_{\Omega_0} f(\mathbf{x}|\theta)f(\theta|\mathbf{x}, \Omega_0)d\theta} = \frac{\int_{\Omega} l(\mu, \sigma, \xi)f(\mu, \sigma, \xi|\mathbf{x}, \Omega)d(\mu, \sigma, \xi)}{\int_{\Omega_0} l(\mu, \sigma)f(\mu, \sigma|\mathbf{x}, \Omega_0)d(\mu, \sigma)}.$$

## 2.4 Distribuição preditiva

Até o presente momento temos denotado por  $x$  os dados observados em um determinado experimento ou processo. Denotemos por  $y$  os dados que ainda serão observados, ou seja, as observações futuras.

Temos interesse em estudar a distribuição destas observações futuras, dadas as observações atuais. Tal distribuição é chamada distribuição preditiva e está definida pela função de densidade

$$f(y|x) = \int f(\theta|x)f(y|\theta, x)d\theta, \quad (2.3)$$

onde  $f(\theta|x)$  é a densidade a posteriori de  $\theta$  e  $f(y|\theta, x)$  é a função de verossimilhança dos dados futuros. Note que, quando  $x$  e  $y$  são condicionalmente independentes dado  $\theta$ ,  $f(y|\theta, x) = f(y|\theta)$ .

Neste contexto, nosso interesse reside na obtenção de inferência para as observações futuras.

O'Hagan (1994) cita ainda que o problema da inferência para estas observações futuras pode ser abordado do ponto de vista de um problema de decisão, apresentando uma função de perda associada.

Notemos ainda que a função de densidade preditiva, como definida em (2.3), pode ser vista como uma esperança da função de verossimilhança dos dados futuros  $f(y|\theta, x)$  (ou  $f(y|\theta)$ ) sob a distribuição a posteriori de  $\theta$ . Esta característica permitirá o cálculo de uma aproximação para a densidade preditiva, como veremos mais adiante.

A partir da determinação da função de densidade  $f(\theta|x)$ , poderemos obter a função de distribuição preditiva para uma observação futura  $y$ , sob  $\Omega$ ,

$$G(y|x) = \int_{\Omega} \exp \left\{ - \left[ 1 + \xi \left( \frac{y - \mu}{\sigma} \right) \right]_+^{-1/\xi} \right\} f(\mu, \sigma, \xi|x)d(\mu, \sigma, \xi), \quad y \in \mathbb{R}. \quad (2.4)$$

Por outro lado, sob  $\Omega_0$ , a função de distribuição preditiva de  $y$  será

$$G(y|x) = \int_{\Omega_0} \exp \left\{ - \exp \left[ - \left( \frac{y - \mu}{\sigma} \right) \right] \right\} f(\mu, \sigma, \xi|x)d(\mu, \sigma), \quad y \in \mathbb{R}. \quad (2.5)$$

## 2.5 Full Bayesian Significance Test - FBST

O FBST (*Full Bayesian Significance Test*) é um procedimento de teste originalmente desenvolvido no sentido de construir um procedimento bayesiano para testar hipóteses precisas (PEREIRA et al., 2008).

Denotemos por  $\chi$  e  $\Theta$  os espaços amostral e paramétrico, respectivamente, de modo que  $x \in \chi \subset \mathbb{R}^n$  e  $\theta \in \Theta \subset \mathbb{R}^k$ , e consideremos a partição  $\Theta = \Theta_0 \cup \Theta_1$ ,  $\Theta_0 \cap \Theta_1 = \emptyset$ .

Uma hipótese precisa,  $H_0$ , será uma afirmação a respeito de  $\theta$ , digamos  $H_0 : \theta \in \Theta_0$ , onde  $\dim(\Theta_0) < \dim(\Theta)$ . Por exemplo, no caso uniparamétrico  $\Theta \subset \mathbb{R}$ , uma hipótese precisa seria  $H_0 : \theta = \theta_0$ ,  $\theta_0 \in \Theta$ .

Considerando o caso em que a distribuição a posteriori para  $\theta$  é absolutamente contínua, naturalmente a probabilidade a posteriori de uma hipótese precisa será nula. No sentido de obter uma medida de evidência em favor de  $H_0$ , os autores definem um conjunto  $T$  ‘tangente’ a  $\Theta_0$  como o subconjunto do espaço paramétrico  $\Theta$  contendo todos os valores de  $\theta$  cuja densidade a posteriori é maior do que o supremo desta densidade restrito ao conjunto  $\Theta_0$ , ou seja,

$$T = \{\theta \in \Theta : f(\theta|x) > s\},$$

onde

$$s = \sup_{\Theta_0} f(\theta|x).$$

Em outras palavras, o conjunto tangente  $T$  considera todos os pontos do espaço paramétrico que são, de certa forma, ‘mais prováveis’ do que  $\Theta_0$ , segundo a distribuição a posteriori.

Pereira et al. (2008) definem a medida de evidência (*e-valor*) em favor da hipótese  $H_0$  como a probabilidade a posteriori de  $\theta$  não pertencer ao conjunto tangente, ou seja,

$$\text{e-valor} = P(\theta \in T^C|x) = 1 - P(\theta \in T|x) = 1 - \int_T f(\theta|x)d\theta.$$

Portanto, se o conjunto tangente possuir probabilidade a posteriori alta, então a evidência em favor de  $\Theta_0$  será baixa. Por outro lado, naturalmente, se possuir probabilidade a posteriori baixa, então a evidência contra  $\Theta_0$  será baixa.

O FBST será o procedimento que rejeita  $H_0$  se *e-valor* for pequeno.

No contexto do teste para a hipótese precisa  $\Omega_0$ , que representa o modelo Gumbel, para obter o conjunto tangente  $T$  precisamos calcular o supremo, em  $\Omega_0$ , da função de densidade a posteriori.

Dado que a função de verossimilhança de  $\theta$  é contínua em  $\xi = 0$  e que existe uma representação da função de densidade a priori de  $\theta$  tal que

$$\lim_{\xi \rightarrow 0} f(\mu, \sigma, \xi) = f(\mu, \sigma, 0),$$

então haverá uma representação da função de densidade a posteriori tal que

$$\lim_{\xi \rightarrow 0} f(\mu, \sigma, \xi|x) = f(\mu, \sigma, 0|x).$$

Mais adiante apresentaremos uma ideia da prova destas afirmações a respeito da função de verossimilhança e densidade a priori.

Pela continuidade da função de densidade a posteriori de  $\theta$ , o supremo  $s$  pode ser determinado como

$$s = \lim_{\epsilon \rightarrow 0} [\sup\{f(\theta|\mathbf{x}) \mid \theta \in V_{\xi=0}(\epsilon)\}], \quad (2.6)$$

onde  $V_{\xi=0}(\epsilon) = \{\theta \in \Omega \mid \text{dist}(\theta, \Omega_0) < \epsilon\}$  é uma vizinhança de  $\Omega_0$ . A distância  $\text{dist}(\cdot, \cdot)$  representa a distância euclidiana. Esta característica permitirá obter, mais adiante, uma aproximação para o *e-valor*.

A determinação de um teste de hipóteses baseado no *e-valor* pode ser vista dentro da teoria de decisão. De fato, consideremos um espaço de ações, digamos

$$\mathcal{D} = \{a_1, a_0\}, \text{ onde } a_1 = \text{"Rejeitar } \Omega_0\text{" e } a_0 = \text{"Não rejeitar } \Omega_0\text{"},$$

e uma função de perda  $l : \mathcal{D} \times \Omega \rightarrow \mathbb{R}$ .

Madruga et al. (2001) definem a função de perda

$$l(a_1, \theta) = w_0(1 - \mathbf{1}_T(\theta)), \quad l(a_0, \theta) = w_1 + c\mathbf{1}_T(\theta),$$

$w_0, w_1, c > 0$ , e mostram que a regra de decisão caracterizada por

$$\text{Rejeitar } \Omega_0 \quad \Leftrightarrow \quad \text{e-valor} < \frac{w_1 + c}{w_0 + c} \quad (2.7)$$

minimiza a função de perda  $l$  esperada sob a distribuição a posteriori,  $E[l(a, \theta)|\mathbf{x}]$ . Prosseguimos com uma ideia da demonstração.

Temos que

$$\begin{aligned} E[l(a_0, \theta)|\mathbf{x}] &= \int_{\Omega} l(a_0, \theta) f(\theta|\mathbf{x}) d\theta \\ &= \int_{\Omega} [w_1 + c\mathbf{1}_T(\theta)] f(\theta|\mathbf{x}) d\theta \\ &= w_1 \int_{\Omega} f(\theta|\mathbf{x}) d\theta + c \int_T f(\theta|\mathbf{x}) d\theta \\ &= w_1 + c(1 - \text{e-valor}). \end{aligned}$$

Por outro lado,

$$\begin{aligned}
 E[l(a_1, \theta)|\mathbf{x}] &= \int_{\Omega} l(a_1, \theta) f(\theta|\mathbf{x}) d\theta \\
 &= \int_{\Omega} [w_0(1 - \mathbf{1}_T(\theta))] f(\theta|\mathbf{x}) d\theta \\
 &= w_0 \int_{\Omega} f(\theta|\mathbf{x}) d\theta - w_0 \int_T f(\theta|\mathbf{x}) d\theta \\
 &= w_0 - w_0(1 - e\text{-valor}) \\
 &= w_0 e\text{-valor}.
 \end{aligned}$$

Então devemos tomar a decisão de não rejeitar  $\Omega_0$  ( $a_0$ ) se, e somente se,

$$E[l(a_0, \theta)|\mathbf{x}] < E[l(a_1, \theta)|\mathbf{x}],$$

ou seja,

$$w_1 + c(1 - e\text{-valor}) < w_0 e\text{-valor} \quad \Leftrightarrow \quad e\text{-valor} > \frac{w_1 + c}{w_0 + c}.$$

■

Desta forma, a decisão pela rejeição de  $\Omega_0$  ocorrerá para valores pequenos do *e-valor*, segundo a escolha de uma certa função de perda.

Os coeficientes  $w_0$  e  $w_1$  representam, respectivamente, as perdas ao se tomar as decisões de Rejeitar  $\Omega_0$  e Não rejeitar  $\Omega_0$ . A constante  $c$  representa a perda adicional ao Não rejeitar  $\Omega_0$  quando, de fato,  $\theta \in T$  possui densidade a posteriori maior do que qualquer valor em  $\Omega_0$ .

## 2.6 Outras medidas de evidência

Além do fator de Bayes, do fator de Bayes a posteriori e do *p-valor* associado ao teste da razão de verossimilhanças (estatística *deviance*), já discutidos anteriormente, também calcularemos algumas outras medidas de evidência para testar a hipótese  $\Omega_0$ .

Spiegelhalter et al. (2002) propõem o DIC (*Deviance Information Criterion*), que pode ser utilizado como uma medida para comparação de modelos, e é definido como

$$DIC = D(\bar{\theta}) + 2p_D,$$

onde  $D(\theta)$  é o desvio (*Bayesian deviance*) dado por

$$D(\theta) = -2 \log f(\mathbf{x}|\theta) + 2 \log g(\mathbf{x}).$$

A função  $g(\mathbf{x})$  constitui um termo normalizador completamente especificado e é uma função somente da amostra. Algumas propostas para a escolha de  $g(\mathbf{x})$  são apresentadas por

Spiegelhalter et al. (2002). Para a comparação de modelos é suficiente assumir  $g(\mathbf{x}) = 1$ , e este foi o valor adotado nas aplicações apresentadas mais adiante.

A esperança a posteriori de  $\theta$ ,  $\bar{\theta} = E(\theta|\mathbf{x})$ , e o termo  $p_D$  estão relacionados por

$$p_D = \overline{D(\theta)} - D(\bar{\theta}),$$

onde  $\overline{D(\theta)} = E[D(\theta)|\mathbf{x}]$ .

Desta maneira, pode-se reescrever

$$DIC = 2\overline{D(\theta)} - D(\bar{\theta}) = 2E[D(\theta)|\mathbf{x}] - D[E(\theta|\mathbf{x})],$$

ou ainda

$$DIC = -4E[\log f(\mathbf{x}|\theta)|\mathbf{x}] + 2\log f[\mathbf{x}|E(\theta|\mathbf{x})] + 2\log g(\mathbf{x}).$$

O valor de DIC deve ser calculado sob os diferentes modelos a serem comparados, e o menor DIC indica o melhor modelo. Em nosso contexto devemos calculá-lo sob os modelos Gumbel e GEV.

Outro critério para comparação de modelos é o BIC (*Bayesian Information Criterion*), definido por Schwarz (1978) como

$$BIC_k = -2\log L_k + j_k \log n,$$

para o modelo  $k$ ,  $k = 0, 1$ , onde  $j_k$  é a dimensão do espaço paramétrico  $\Omega_k$ ,  $n$  é o tamanho da amostra e  $L_k = \max\{f(\mathbf{x}|\theta) : \theta \in \Omega_k\}$ , assumindo  $\Omega_1 = \Omega$ .

Aqui também, quanto menor o valor de BIC, melhor será o modelo.

## 2.7 Distribuição a posteriori e intervalos de credibilidade

Como anteriormente, denotamos por  $f(\theta|\mathbf{x})$  a função de densidade a posteriori do vetor de parâmetros  $\theta = (\mu, \sigma, \xi)$ .

Assim, uma vez determinada  $f(\theta|\mathbf{x})$ , podemos obter a probabilidade a posteriori de um subconjunto de  $\Omega$ , digamos

$$P(\xi > 0|\mathbf{x}) = \int_{\{\theta \in \Omega: \xi > 0\}} f(\theta|\mathbf{x}) d\theta = \int_{\Omega} \mathbf{1}_{\{\theta \in \Omega: \xi > 0\}}(\theta) f(\theta|\mathbf{x}) d\theta.$$

Além disso, podemos determinar a esperança a posteriori do nível de retorno  $q_p$ , para um período de retorno  $\frac{1}{p}$  fixo, como

$$E(q_p|\mathbf{x}) = \int_{\Omega} \left\{ \mu - \frac{\sigma}{\xi} \left[ 1 - (-\log(1-p))^{-\xi} \right] \right\} f(\theta|\mathbf{x}) d\theta,$$

sob o modelo GEV. Similarmente, para o modelo Gumbel, a esperança a posteriori denotada por  $E_0(q_p|\mathbf{x})$  é dada por

$$E_0(q_p|\mathbf{x}) = \int_{\Omega_0} \{\mu - \sigma \log[-\log(1-p)]\} f(\theta|\mathbf{x}) d\theta.$$

Podemos determinar uma região de credibilidade  $\gamma$  ( $\gamma \in (0, 1)$ ) para o nível de retorno  $q_p$  associado ao período de retorno  $\frac{1}{p}$  fixo, sob  $\Omega$ :

$$RC_\gamma(q_p) = \left\{ q_p = \mu - \frac{\sigma}{\xi} \left[ 1 - (-\log(1-p))^{-\xi} \right] : a_1 < q_p < a_2 \right\}, \quad (2.8)$$

tal que  $P(q_p < a_1|\mathbf{x}, \Omega) = P(q_p > a_2|\mathbf{x}, \Omega) = (1 - \gamma)/2$ .

Analogamente, podemos determinar uma região de credibilidade  $\gamma$  para  $q_p$  sob  $\Omega_0$ ,

$$RC_\gamma(q_p) = \{q_p = \mu - \sigma \log[-\log(1-p)] : b_1 < q_p < b_2\}, \quad (2.9)$$

tal que  $P(q_p < b_1|\mathbf{x}, \Omega_0) = P(q_p > b_2|\mathbf{x}, \Omega_0) = (1 - \gamma)/2$ .

Dado  $\gamma \in (0, 1)$  constante, podemos ainda obter o conjunto de maior densidade a posteriori, HPD (*highest posterior density region*), (O'HAGAN, 1994), para  $\theta$  com probabilidade a posteriori  $\gamma$ , sob  $\Omega_k$ , como

$$HPD_\gamma^k = \{\theta \in \Omega_k \mid f(\theta|\mathbf{x}) \geq c_\gamma\}, \quad k = 0, 1, \quad \Omega_1 = \Omega,$$

onde  $P(\theta \in HPD_\gamma^k|\mathbf{x}) = \gamma$ , de tal forma que o conjunto  $HPD_\gamma^k$  fica determinado por uma superfície de nível da função de densidade a posteriori  $f(\theta|\mathbf{x})$  em  $\Omega_k$ .

# Capítulo 3

## Aspectos computacionais

### 3.1 Introdução

Os métodos MCMC (*Markov chain Monte Carlo*) constituem uma ferramenta poderosa para simulação de variáveis aleatórias com uma determinada distribuição, que pode ser multivariada e complexa.

Basicamente, um método MCMC constitui um procedimento (algoritmo) a partir do qual se produz uma sequência de variáveis aleatórias (pseudo-aleatórias) que deve representar uma cadeia de Markov aperiódica e irredutível, cuja distribuição estacionária é exatamente a distribuição da qual originalmente se deseja simular. Neste sentido, ao simular tal cadeia por um longo período, é razoável assumir que as variáveis obtidas (após este longo período) apresentam uma distribuição muito próxima da distribuição estacionária, ou distribuição alvo.

Tais métodos têm se tornado cada vez mais disseminados, especialmente em inferência bayesiana, quando o objetivo é simular de uma determinada distribuição a posteriori.

Dentre os métodos MCMC, um dos mais conhecidos é o algoritmo de Metropolis-Hastings, que foi desenvolvido por Metropolis et al. (1953) e em seguida generalizado por Hastings (1970).

### 3.2 Metropolis-Hastings

Suponhamos que nosso interesse seja simular variáveis aleatórias de uma determinada distribuição (possivelmente multivariada) identificada por uma função de densidade  $\pi(x)$ . Suponhamos ainda que tal densidade tenha constante normalizadora desconhecida  $K$ , ou seja,

$$\pi(x) = \frac{g(x)}{K},$$

com  $g$  conhecida.

No contexto específico da inferência bayesiana, denotando por  $f(\theta)$  uma função de densidade a priori e por  $f(x|\theta)$  a função de verossimilhança do parâmetro  $\theta$ , a função de densidade a posteriori  $\pi$  assume a forma

$$\pi(\theta|x) = \frac{f(x|\theta)f(\theta)}{\int f(x|\theta)f(\theta)d\theta} = \frac{f(x|\theta)f(\theta)}{K},$$

com o denominador  $K$  constante em  $\theta$ . O objetivo pode ser a simulação desta densidade a posteriori, onde de fato, em geral, a contante  $K$  é desconhecida.

O algoritmo de Metropolis-Hastings permite a simulação de tais densidades, produzindo uma cadeia de Markov cuja distribuição estacionária é  $\pi$ .

Suponhamos que seja possível simular um valor  $y$  de uma determinada distribuição conhecida, segundo algum método simples. Denotemos por  $q(x, y)$  a função de densidade desta distribuição, que pode depender do estado atual da cadeia,  $x$ . Por exemplo,  $y$  pode ser uma simulação da distribuição  $\mathcal{U}(-x, x)$ , ou da distribuição  $N(0, 1)$ . Notemos que estas distribuições (Uniforme e Normal) são facilmente simuladas computacionalmente.

Basicamente, no algoritmo de Metropolis-Hastings, estas candidatas  $y$  são geradas a partir da densidade  $q(x, y)$ , e então a cadeia irá se movimentar do valor (estado) atual  $x$  para  $y$ , ou irá permanecer em  $x$ , iterativamente.

Chib e Greenberg (1995) mostram que, uma vez que a cadeia apresente o valor  $x$  no instante atual, ela deverá se movimentar para o valor candidato  $y$ , gerado a partir da densidade  $q(x, y)$ , com probabilidade

$$\alpha(x, y) = \begin{cases} \min\left(\frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}, 1\right), & \text{se } \pi(x)q(x, y) > 0 \\ 1, & \text{caso contrário.} \end{cases}$$

chamada probabilidade de movimento, de tal forma que as transições desta cadeia irão satisfazer a condição de reversibilidade, e possuir  $\pi$  como distribuição estacionária.

Se a cadeia não se movimentar para o candidato  $y$ , então ela permanece no estado  $x$ , ou seja, a cadeia assume no instante seguinte o mesmo valor do instante atual.

Desta forma,  $\pi$  é a distribuição limite da cadeia e  $q(\cdot, \cdot)$  representa o núcleo de transição da cadeia.

Após simular esta cadeia por um longo período de tempo, espera-se que as observações possuam distribuição identificada por  $\pi$ , aproximadamente. Diversas propostas para a seleção dos valores gerados pelo método para compor a amostra desejada são encontradas na literatura. Dentre elas, sugere-se por exemplo iniciar a cadeia várias vezes a partir de diferentes valores

iniciais e compor a amostra com os valores finais de cada cadeia simulada por longos períodos, ou ainda iniciar uma única cadeia e descartar uma porção inicial gerada, compondo a amostra com os valores simulados depois disso. Esta última foi a proposta adotada neste trabalho. Boas referências dentro deste assunto são Gelman et al. (2004) e Gamerman (1996).

Note que o desconhecimento da constante normalizadora  $K$  não influencia o método, uma vez que a probabilidade de movimento é definida em termos de uma razão, que apresenta a mesma constante no numerador e denominador.

Lembremos que a irreduzibilidade e a aperiodicidade são condições para que a cadeia gerada pelo procedimento possua uma distribuição estacionária. Segundo Chib e Greenberg (1995), tais condições serão geralmente satisfeitas quando a distribuição geradora de candidatas, caracterizada por  $q(x, y)$ , possuir suporte limitado ou estiver definida no mesmo suporte da distribuição alvo  $\pi$ .

Os mesmos autores resumem o algoritmo da seguinte forma:

1. Inicializar um valor arbitrário  $x^{(0)}$  (que pode ser um vetor);

Para  $j \in \{0, 1, 2, \dots, N\}$ , repetir os passos 2 e 3:

2. Gerar  $y$  a partir de  $q(x^{(j)}, \cdot)$  e  $u$  a partir de  $\mathcal{U}(0, 1)$ ;
3. Se  $u \leq \alpha(x^{(j)}, y)$ , então  $x^{(j+1)} = y$ ,  
senão  $x^{(j+1)} = x^{(j)}$ .
4. Obter  $x^{(1)}, x^{(2)}, \dots, x^{(N)}$ .

Também apresentam algumas sugestões para a escolha da distribuição  $q(x, y)$ , geradora de candidatas. Dentre as propostas, citam a escolha de  $q(x, y)$  dentro de uma família de distribuições caracterizadas por

$$q(x, y) = q_1(y - x),$$

em que  $q_1$  representa uma função de densidade multivariada. Neste caso, a candidata  $y$  é gerada a partir da atribuição de um ruído  $z$  ao valor atual da cadeia,  $x$ , de tal forma que

$$y = x + z,$$

onde  $z$  possui distribuição identificada por  $q_1$ . A cadeia obtida através desta escolha é conhecida como cadeia de passeio aleatório.

### 3.3 Algoritmo

Apresentamos agora o algoritmo de Metropolis-Hastings implementado para a simulação da distribuição a posteriori  $\pi(\theta|\mathbf{x})$ . Lembremos que  $\mathbf{x} = (x_1, \dots, x_n)$  denota a amostra de máximos observados.

Em nosso contexto, sob o modelo GEV para a amostra  $\mathbf{x}$ , temos interesse em obter uma amostra da distribuição a posteriori do vetor de parâmetros  $\theta = (\mu, \sigma, \xi)$ ,

$$\pi(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)f(\theta)}{K} \propto f(\mathbf{x}|\theta)f(\theta),$$

em que  $f(\mathbf{x}|\theta)$  denota a função de verossimilhança de  $\theta$  e  $f(\theta)$  denota a função de densidade a priori de  $\theta$ .

Utilizamos como distribuição geradora de estados possíveis a proposta pela cadeia de passeio aleatório, ou seja, dado o estado atual da cadeia,  $w$ , geramos um estado candidato  $y$  a partir da função de densidade

$$q(w, y) = q_1(y - w) = q_1(z).$$

Neste caso,  $w$ ,  $y$  e  $z$  são vetores de dimensão 3. Especificamente, geramos um ruído  $z = (z_1, z_2, z_3)$ , em que  $z_1$ ,  $z_2$  e  $z_3$  são variáveis aleatórias independentes, uniformemente distribuídas (não necessariamente identicamente) e simétricas em torno de zero, ou seja,

$$z_i \sim \mathcal{U}(-\delta_i, \delta_i), \quad i = 1, 2, 3,$$

e então somamos este ruído ao estado atual, gerando assim uma candidata,  $y = w + z$ .

Notemos que, como a distribuição de  $z$  é simétrica em torno de zero, então  $q_1(z) = q_1(-z)$ , ou seja,  $q(w, y) = q(y, w)$ , de tal forma que a probabilidade de movimento anteriormente definida se resume a

$$\alpha(w, y) = \min\left(\frac{\pi(y)}{\pi(w)}, 1\right).$$

A seguir apresentamos o algoritmo implementado neste trabalho:

1. Inicializar  $\theta^{(0)} = (\mu^{(0)}, \sigma^{(0)}, \xi^{(0)})$  arbitrário tal que  $\pi(\theta^{(0)}|\mathbf{x}) > 0$ .
2. Para  $j = 0, 1, 2, \dots, N$ :
  - Gerar  $z^{(j)} = (z_1^{(j)}, z_2^{(j)}, z_3^{(j)})$ , onde  $z_i^{(j)} \sim \mathcal{U}(-\delta_i, \delta_i)$ ,  $i = 1, 2, 3$ .
  - Gerar  $u \sim \mathcal{U}(0, 1)$ .

- Fazer  $y = \theta^{(j)} + z^{(j)}$ .
- Se  $u \leq \alpha(\theta^{(j)}, y)$ , então  $\theta^{(j+1)} = y$ , senão  $\theta^{(j+1)} = \theta^{(j)}$ .

Rodamos o programa algumas vezes, obtendo seqüências  $\theta^{(1)}, \dots, \theta^{(N+1)}$  distintas a partir de diferentes pontos iniciais. Com isso notamos um comportamento regular do procedimento, através da comparação descritiva destas seqüências geradas. Desta forma, para uma seqüência  $\theta^{(1)}, \dots, \theta^{(N+1)}$ , descartamos uma porção inicial de tamanho  $m$ , e tomamos

$$\theta^{(m+1)}, \dots, \theta^{(N+1)}$$

como uma amostra de  $\pi(\cdot|\mathbf{x})$ .

Os parâmetros  $\delta_i$  da distribuição do incremento  $z$  são escolhidos de tal forma que a taxa de aceitação das candidatas geradas esteja em torno de 40% a 50%. A ideia é obter uma amostra que não se concentre em uma região restrita do espaço paramétrico.

Sob o modelo Gumbel ( $\xi = 0$ ), o algoritmo é implementado de maneira análoga, no sentido de proporcionar a simulação da distribuição a posteriori de  $\theta = (\mu, \sigma)$ .

Nos exemplos estudados, adotamos  $N = 15000$  e  $m = 5000$ . A seleção destes valores foi determinada por critérios gráficos. Foram comparados os histogramas marginais dos valores simulados  $\mu$ ,  $\sigma$  e  $\xi$ , com os respectivos gráficos simulados de  $(\mu, g)$ ,  $(\sigma, g)$  e  $(\xi, g)$ , onde  $g$  representa a função de densidade a posteriori não normalizada avaliada nos valores simulados. A similaridade destes gráficos indica que a densidade a posteriori é bem representada pela amostra simulada.

# Capítulo 4

## Modelagem

### 4.1 Preliminares

Como vimos anteriormente, o teorema de Fisher-Tippett pode ser utilizado para se aproximar a distribuição do máximo de uma sequência de variáveis aleatórias independentes e identicamente distribuídas,

$$X_k = \max\{X_1, \dots, X_k\},$$

por uma das distribuições de valor extremo, cuja função de distribuição é dada por

$$G(x | \mu, \sigma, \xi) = \exp \left\{ - \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]_+^{-1/\xi} \right\}, \quad x \in \mathbb{R} \quad (4.1)$$

onde  $z_+ = \max\{z, 0\}$ .

Lembremos que a distribuição GEV é indexada por três parâmetros:  $\mu, \xi \in \mathbb{R}$  e  $\sigma > 0$ . Ainda, o parâmetro de forma,  $\xi$ , determina o tipo da distribuição. O caso em que  $\xi = 0$ , que corresponde ao limite da função de distribuição (4.1) quando  $\xi \rightarrow 0$ , representa a distribuição Gumbel, cuja função de distribuição é dada por

$$G(x | \mu, \sigma) = \exp \left\{ - \exp \left[ - \left( \frac{x - \mu}{\sigma} \right) \right] \right\}, \quad x \in \mathbb{R}.$$

O comportamento assintótico do máximo de variáveis aleatórias independentes com distribuição pertencente a famílias como a Normal ou a Gama está associado à família Gumbel.

Desta forma, podemos representar o espaço paramétrico associado à distribuição GEV por

$$\Omega = \{\theta = (\mu, \sigma, \xi) \in \mathbb{R}^3 : \sigma > 0\},$$

de modo que a família Gumbel acaba por representar uma hipótese precisa, digamos

$$\Omega_0 = \{(\mu, \sigma, \xi) \in \Omega : \xi = 0\}.$$

Na realização da inferência para o parâmetro  $\theta$  através do procedimento de teste de hipóteses no contexto bayesiano, que pode ser visto como um procedimento para seleção de modelos, calculamos o *e-valor*, a medida de evidência estabelecida pelo FBST, entre outras medidas de evidência usuais capazes de testar tais hipóteses, no contexto clássico e bayesiano.

## 4.2 Formulação do modelo

### 4.2.1 Verossimilhança

Como comentado anteriormente, é natural assumir na modelagem que os máximos são observados em períodos fixos de tempo, como meses, semestres ou anos, por exemplo.

Consideremos então uma amostra, de tamanho  $n$ , de máximos observados durante tais períodos fixos. Denotemos esta amostra por  $\mathbf{x} = (x_1, \dots, x_n)$ . Assumimos que os elementos desta amostra são independentes e identicamente distribuídos, com distribuição GEV.

Desta forma, a função de verossimilhança para  $\theta = (\mu, \sigma, \xi) \in \Omega$ , dada a amostra  $\mathbf{x}$ , se  $\xi \neq 0$ , é dada pelo produtório da função de densidade da distribuição GEV avaliada em cada  $x_i \in \mathbb{R}$ ,  $i = 1, \dots, n$ ,

$$f(\mathbf{x}|\theta) = \sigma^{-n} \prod_{i=1}^n \exp \left\{ - \left[ 1 + \xi \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-1/\xi} \right\} \left[ 1 + \xi \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-(\xi+1)/\xi}.$$

se  $\mu \in \mathbb{R}$ ,  $\sigma > 0$ ,  $\xi \in \mathbb{R}$  ( $\xi \neq 0$ ) satisfazem

$$1 + \xi \left( \frac{x_i - \mu}{\sigma} \right) > 0, \text{ para todo } i \in \{1, \dots, n\}. \quad (4.2)$$

Por outro lado, se

$$1 + \xi \left( \frac{x_i - \mu}{\sigma} \right) \leq 0, \text{ para pelo menos um } i \in \{1, \dots, n\}, \quad (4.3)$$

então  $f(\mathbf{x}|\theta) = 0$ .

Para  $\theta \in \Omega_0$  ( $\xi = 0$ ), dado a amostra  $\mathbf{x}$ , a função de verossimilhança fica, segundo o modelo Gumbel, dada por

$$f(\mathbf{x}|\theta) = \sigma^{-n} \prod_{i=1}^n \exp \left[ - \exp \left( - \frac{x_i - \mu}{\sigma} \right) \right] \exp \left( - \frac{x_i - \mu}{\sigma} \right), \quad \mu \in \mathbb{R}, \quad \sigma > 0.$$

Conforme havíamos comentado, a função de verossimilhança de  $\theta \in \Omega$  é contínua em  $\xi = 0$ . De fato, sabemos que a função de densidade da distribuição GEV é dada por

$$f(x|\mu, \sigma, \xi) = \frac{1}{\sigma} \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^{-(1+\frac{1}{\xi})} \exp \left\{ - \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\},$$

para  $1 + \xi \left(\frac{x-\mu}{\sigma}\right) > 0$ , e é igual a zero no caso contrário.

Sejam  $x, \mu \in \mathbb{R}$  e  $\sigma > 0$  fixos. Então

$$\lim_{\xi \rightarrow 0} \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} = \exp \left\{ - \left( \frac{x - \mu}{\sigma} \right) \right\}.$$

De modo que

$$\lim_{\xi \rightarrow 0} f(x|\mu, \sigma, \xi) = \frac{1}{\sigma} \exp \left\{ - \left( \frac{x - \mu}{\sigma} \right) \right\} \exp \left\{ - \exp \left[ - \left( \frac{x - \mu}{\sigma} \right) \right] \right\} = f(x|\mu, \sigma),$$

onde  $f(x|\mu, \sigma)$  é a função de densidade Gumbel.

Para  $x_1, \dots, x_n \in \mathbb{R}$  fixos, temos

$$\lim_{\xi \rightarrow 0} f(x_i|\mu, \sigma, \xi) = f(x_i|\mu, \sigma), \quad i = 1, \dots, n,$$

de tal forma que

$$\lim_{\xi \rightarrow 0} \prod_{i=1}^n f(x_i|\mu, \sigma, \xi) = \prod_{i=1}^n f(x_i|\mu, \sigma),$$

ou seja, a função de verossimilhança do modelo GEV converge para a função de verossimilhança do modelo Gumbel, quando  $\xi \rightarrow 0$ . ■

## 4.2.2 Distribuição a priori

Denotemos por  $q_p$  o quantil  $(1-p)$  da distribuição GEV que modela o máximo, definido pela relação  $G(q_p) = 1 - p$ .

Vimos que

$$q_p = \begin{cases} \mu - \frac{\sigma}{\xi} [1 - \{-\log(1-p)\}^{-\xi}], & \text{para } \xi \neq 0, \\ \mu - \sigma \log\{-\log(1-p)\}, & \text{para } \xi = 0, \end{cases}$$

representa o nível de retorno associado ao tempo de retorno  $\frac{1}{p}$ , e espera-se que  $q_p$  seja superado pelo máximo uma vez a cada  $\frac{1}{p}$  períodos (meses, semestres, anos, etc.) em média.

Ao invés de estabelecer uma distribuição a priori diretamente para  $\mu$ ,  $\sigma$  e  $\xi$ , seguimos a proposta apresentada por Coles e Tawn (1996), na qual uma distribuição a priori é estabelecida para as diferenças

$$\tilde{q}_1 = q_{p_1} - e_1, \quad \tilde{q}_2 = q_{p_2} - q_{p_1}, \quad \tilde{q}_3 = q_{p_3} - q_{p_2},$$

para valores fixos  $p_1 > p_2 > p_3$  pequenos. A constante  $e_1$  corresponde a um limitante inferior natural para o processo ou grandeza sob estudo. Nos exemplos, foram usados  $p_i = 10^{-i}$ , para  $i = 1, 2, 3$ .

Especificamente, se denotarmos  $y_p = -\log(1-p)$ , no caso em que  $\xi \neq 0$  temos

- $\tilde{q}_1 = \mu - \frac{\sigma}{\xi}(1 - y_{p_1}^{-\xi}) - e_1,$
- $\tilde{q}_2 = \frac{\sigma}{\xi}(y_{p_2}^{-\xi} - y_{p_1}^{-\xi}),$
- $\tilde{q}_3 = \frac{\sigma}{\xi}(y_{p_3}^{-\xi} - y_{p_2}^{-\xi}).$

Desta maneira, há uma transformação  $T$  que leva  $\theta = (\mu, \sigma, \xi)$  em  $(\tilde{q}_1, \tilde{q}_2, \tilde{q}_3)$ ,  $T(\theta) = (\tilde{q}_1, \tilde{q}_2, \tilde{q}_3)$ .

Coles e Tawn (1996) assumem que as quantidades  $\tilde{q}_i$ ,  $i = 1, 2, 3$ , são independentes a priori, cada uma com distribuição da forma

$$\tilde{q}_i \sim \text{Gamma}(\alpha_i, \beta_i), \quad i = 1, 2, 3.$$

Assumimos que  $\tilde{q}_1$ ,  $\tilde{q}_2$  e  $\tilde{q}_3$  são quantidades aleatórias independentes com distribuição a priori quase não-informativa, ou seja, assumimos que são independentes com distribuição uniforme em intervalos suficientemente grandes:

$$\tilde{q}_i \sim \mathcal{U}(0, b_i), \quad i = 1, 2, 3,$$

sendo  $b_i$ ,  $i = 1, 2, 3$ , suficientemente grandes.

Em questões aplicadas, os valores de  $b_i$ ,  $i = 1, 2, 3$ , podem ser escolhidos levando em consideração algum conhecimento acerca da grandeza estudada, bem como a natureza dos dados. Foi isso que levamos em conta ao estabelecer estes valores nas aplicações apresentadas mais adiante.

Assim, a distribuição a priori para  $\theta$ , induzida pela transformação  $T$ , é dada por

$$f(\theta) = J(\theta) \mathbf{1}_{\prod_{i=1}^3 (0, b_i)} T(\theta) \frac{1}{b_1 b_2 b_3} \propto J(\theta) \mathbf{1}_{\prod_{i=1}^3 (0, b_i)} T(\theta),$$

onde  $J(\theta)$  é o Jacobiano da transformação de  $(\tilde{q}_1, \tilde{q}_2, \tilde{q}_3)$  em  $\theta$ , apresentado por Stephenson e Tawn (2004),

$$J(\theta) = \frac{\sigma}{\xi^2} \left| \sum_{\substack{i, j \in \{1, 2, 3\} \\ i < j}} (-1)^{i+j} (y_i y_j)^{-\xi} \log(y_j / y_i) \right|, \quad y_i = -\log(1 - p_i).$$

Como comentado anteriormente, existe uma representação da função de densidade a priori de  $\theta \in \Omega$  tal que

$$\lim_{\xi \rightarrow 0} f(\mu, \sigma, \xi) = f(\mu, \sigma, 0). \quad (4.4)$$

Lembremos que a função de densidade a priori de  $\theta \in \Omega$  pode ser reescrita como

$$f(\theta) = J(\theta)\mathbf{1}_{(0,b_1)}(\tilde{q}_1)\mathbf{1}_{(0,b_2)}(\tilde{q}_2)\mathbf{1}_{(0,b_3)}(\tilde{q}_3)\frac{1}{b_1b_2b_3}.$$

Para  $\mu \in \mathbb{R}$  e  $\sigma > 0$  fixos, usando a Regra de L'Hospital, temos

$$\lim_{\xi \rightarrow 0} \tilde{q}_1 = \lim_{\xi \rightarrow 0} \left\{ \mu - \frac{\sigma}{\xi} (1 - y_{p_1}^{-\xi}) - e_1 \right\} = \lim_{\xi \rightarrow 0} \left\{ \mu - \sigma (y_{p_1}^{-\xi} \log y_{p_1}) - e_1 \right\} = \mu - \sigma \log y_{p_1} - e_1,$$

e

$$\lim_{\xi \rightarrow 0} \tilde{q}_2 = \lim_{\xi \rightarrow 0} \left\{ \frac{\sigma}{\xi} (y_{p_2}^{-\xi} - y_{p_1}^{-\xi}) \right\} = \lim_{\xi \rightarrow 0} \left\{ \sigma (y_{p_1}^{-\xi} \log y_{p_1} - y_{p_2}^{-\xi} \log y_{p_2}) \right\} = \sigma (\log y_{p_1} - \log y_{p_2}).$$

Analogamente,

$$\lim_{\xi \rightarrow 0} \tilde{q}_3 = \sigma (\log y_{p_2} - \log y_{p_3}).$$

O limite do Jacobiano  $J(\mu, \sigma, \xi)$  quando  $\xi \rightarrow 0$ , para  $\mu \in \mathbb{R}$  e  $\sigma > 0$  fixos, obtido computacionalmente, foi

$$-\frac{1}{2} \log[y_{p_3}/y_{p_2}] \log[y_{p_3}/y_{p_1}] \log[y_{p_2}/y_{p_1}].$$

O gráfico de  $J(\mu, \sigma, \xi)$ , para  $\mu \in \mathbb{R}$  e  $\sigma > 0$  fixos, e  $\xi$  próximo de zero, também obtido computacionalmente, está representado na Figura 4.1.

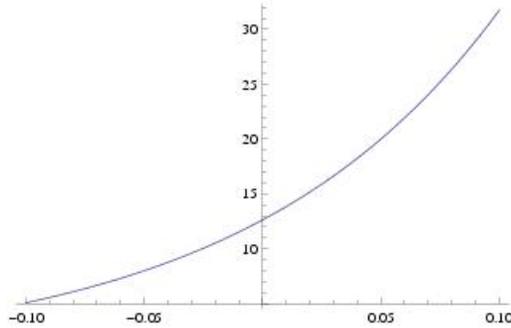


Figura 4.1: Gráfico de  $J(\mu, \sigma, \xi)$  em função de  $\xi$  próximo de zero, para  $\mu \in \mathbb{R}$  e  $\sigma > 0$  fixos, obtido computacionalmente.

Desta forma,

$$\lim_{\xi \rightarrow 0^+} f(\mu, \sigma, \xi) = \lim_{\xi \rightarrow 0^-} f(\mu, \sigma, \xi),$$

de modo que existe uma representação da função de densidade a priori em  $\Omega$  tal que (4.4) é satisfeito.

■

Analogamente, no caso em que  $\theta \in \Omega_0$ , ou seja,  $\xi = 0$ , que corresponde ao modelo Gumbel, estabelecemos uma distribuição a priori para as diferenças

$$\tilde{q}_1 = q_{p_1} - e_1, \quad \tilde{q}_2 = q_{p_2} - q_{p_1},$$

com  $p_i = 10^{-i}$ ,  $i = 1, 2$ .

Assumimos novamente que

$$\tilde{q}_i \sim \mathcal{U}(0, b_i), \quad i = 1, 2,$$

com  $b_i$ ,  $i = 1, 2$ , suficientemente grandes.

Neste caso, a transformação  $T$  é tal que  $T(\mu, \sigma) = (\tilde{q}_1, \tilde{q}_2)$ , de modo que

- $\tilde{q}_1 = \mu - \sigma \log y_{p_1} - e_1$ ,
- $\tilde{q}_2 = \sigma(\log y_{p_1} - \log y_{p_2})$ .

Então a distribuição a priori para  $\theta = (\mu, \sigma) \in \Omega_0$  induzida pela transformação de  $(\tilde{q}_1, \tilde{q}_2)$  em  $(\mu, \sigma)$  é dada por

$$f(\mu, \sigma) = J(\mu, \sigma) \mathbf{1}_{\prod_{i=1}^2 (0, b_i)} T(\mu, \sigma) \frac{1}{b_1 b_2} \propto \mathbf{1}_{\prod_{i=1}^2 (0, b_i)} T(\mu, \sigma),$$

com  $J(\mu, \sigma) = |\log y_{p_1} - \log y_{p_2}|$ .

### 4.3 Aproximações

Consideremos uma amostra simulada de tamanho  $N$  da distribuição a posteriori  $f(\theta|\mathbf{x})$  em  $\Omega$ , obtida segundo o procedimento de simulação MCMC descrito anteriormente, levando em consideração o descarte de uma determinada porção inicial gerada pelo procedimento (conhecido como período de *burn-in*) necessário para se atingir a estacionariedade da cadeia gerada. Denotemos então por  $\theta_i^1 = (\mu_i^1, \sigma_i^1, \xi_i^1)$ ,  $i = 1, \dots, N$ , esta amostra simulada de  $f(\theta|\mathbf{x})$  em  $\Omega$ .

Analogamente, consideremos uma amostra simulada, também de tamanho  $N$ , da distribuição a posteriori  $f(\theta|\mathbf{x})$  em  $\Omega_0$ ,  $\theta_i^0 = (\mu_i^0, \sigma_i^0)$ ,  $i = 1, \dots, N$ .

Denotemos ainda por  $g_i^k$  o valor da função de densidade a posteriori de  $\theta \in \Omega_k$  não normalizada avaliada em  $\theta_i^k$ , ou seja,

$$g_i^k \propto f(\theta_i^k|\mathbf{x}),$$

e por  $L_i^k$  a função de verossimilhança de  $\theta \in \Omega_k$  avaliada em  $\theta_i^k$ ,  $f(\mathbf{x}|\theta_i^k)$ ,  $i = 1, \dots, N$ ,  $k = 0, 1$ , e  $\Omega_1 = \Omega$ .

A seleção do período de *burn-in* foi determinada por critérios gráficos. Foram comparados os histogramas marginais dos valores simulados  $\theta_i^1 = (\mu_i^1, \sigma_i^1, \xi_i^1)$  com os respectivos gráficos de  $(\mu_i^1, g_i^1)$ ,  $(\sigma_i^1, g_i^1)$  e  $(\xi_i^1, g_i^1)$ . O mesmo foi feito sob  $\Omega_0$ . A similaridade destes gráficos indica que a densidade a posteriori é bem representada pela amostra simulada.

Podemos ordenar os vetores  $(\mu_i^1, \sigma_i^1, \xi_i^1)$  e os valores  $g_i^1$  e  $L_i^1$  segundo os valores de  $\xi_i^1$ , e obter os valores  $g_i^1$  em uma vizinhança  $V_{\xi=0}(\epsilon)$ , como definida anteriormente.

Podemos então aproximar o supremo  $s$ , segundo (2.6),

$$s = \sup_{\Omega_0} f(\theta|\mathbf{x}) \approx \max\{f(\mu_i^1, \sigma_i^1, \xi_i^1|\mathbf{x}) \mid \theta_i^1 \in V_{\xi=0}(\epsilon)\}.$$

Como desconhecemos a constante normalizadora em  $f(\theta|\mathbf{x})$ , definimos

$$g_0 = \max\{g_i^1 \mid \theta_i^1 \in V_{\xi=0}(\epsilon)\},$$

e aproximamos o *e-valor* por

$$\text{e-valor} \approx 1 - \frac{1}{N} \left[ \#\{g_i^1 : g_i^1 > g_0\} \right],$$

onde  $\#A$  denota a cardinalidade do conjunto  $A$ . Nos exemplos, consideramos  $\epsilon = 10^{-3}$ .

Para determinar o fator de Bayes contra  $\Omega_0$ ,  $B_{10}$ , dado em (2.1), podemos aproximar a função de densidade marginal de  $\mathbf{x}$  sob  $\Omega_k$ ,  $f(\mathbf{x}|\Omega_k)$ , por

$$f_k(\mathbf{x}) = \left( \frac{1}{N} \sum_{i=1}^N (L_i^k)^{-1} \right)^{-1},$$

conforme sugerido por Kass e Raftery (1995), utilizando, desta forma, a amostra simulada da distribuição a posteriori em  $\Omega_k$ ,  $k = 0, 1$ .

Assim, podemos aproximar  $2 \log B_{10}$  por

$$2 \log B_{10} \approx 2 \log \frac{f_1(\mathbf{x})}{f_0(\mathbf{x})}.$$

No fator de Bayes a posteriori contra  $\Omega_0$ , FBP, a integral

$$\int_{\Omega_k} f(\theta|\mathbf{x}, \Omega_k) f(\mathbf{x}|\theta) d\theta$$

pode ser vista como a esperança sob a distribuição a posteriori em  $\Omega_k$  da função de verossimilhança  $f(\mathbf{x}|\theta)$ , que pode naturalmente ser aproximada pela média amostral da função de verossimilhança de  $\theta \in \Omega_k$  avaliada nos valores simulados da distribuição a posteriori em  $\Omega_k$ ,

$$\frac{1}{N} \sum_{i=1}^N L_i^k, \quad k = 0, 1.$$

Por sua vez, o máximo da função de verossimilhança de  $\theta \in \Omega_k$  pode ser aproximado pelo valor máximo da própria função de verossimilhança de  $\theta \in \Omega_k$  avaliada na amostra simulada em  $\Omega_k$ , ou seja, pode ser aproximado por

$$\hat{L}_k = \max_{i \in \{1, \dots, N\}} \{L_i^k\}, \quad k = 0, 1,$$

de tal forma que a estatística *deviance* pode ser aproximada por

$$\hat{D} = 2(\log \hat{L}_1 - \log \hat{L}_0),$$

bem como o BIC para o modelo  $k$ ,  $BIC_k$ ,

$$BIC_k \approx -2 \log \hat{L}_k + d_k \log n, \quad k = 0, 1,$$

onde  $d_k$  é a dimensão de  $\Omega_k$  e  $n$  é o tamanho da amostra de máximos observados. Lembremos que o modelo  $k = 0$  corresponde ao modelo Gumbel, enquanto que o modelo  $k = 1$  corresponde ao GEV.

No que diz respeito ao DIC, precisamos aproximar

$$E[ \log f(\mathbf{x}|\theta) \mid \mathbf{x} ] \text{ e } f[ \mathbf{x} \mid E(\theta|\mathbf{x}) ],$$

sob os dois modelos.

Já a constante  $2 \log g(\mathbf{x})$ , conforme comentamos, sob o propósito da comparação entre dois modelos, não será determinada, pois é a mesma sob os dois modelos.

A esperança a posteriori  $E[\log f(\mathbf{x}|\theta) \mid \mathbf{x}]$ , sob o modelo  $k$ , pode ser aproximada pela média amostral

$$\frac{1}{N} \sum_{i=1}^N \log f(\mathbf{x}|\theta_i^k), \quad k = 0, 1,$$

enquanto que a  $j$ -ésima componente de  $E(\theta|\mathbf{x})$ , sob o modelo  $k$ , analogamente pode ser aproximada por

$$\left( E(\widehat{\theta|\mathbf{x}})_k \right)_j = \left( \frac{1}{N} \sum_{i=1}^N \theta_i^k \right)_j, \quad k = 0, 1, \quad j = 1, 2, 3. \quad (4.5)$$

Assim,  $f[\mathbf{x} \mid E(\theta|\mathbf{x})]$  será aproximada, sob o modelo  $k$ , por

$$f[\mathbf{x} \mid E(\widehat{\theta|\mathbf{x}})_k], \quad k = 0, 1.$$

A esperança a posteriori do nível de retorno  $q_p$  associado ao período de retorno  $\frac{1}{p}$ ,  $E(q_p|\mathbf{x})$ , sob  $\Omega$ , também pode ser aproximada por uma média amostral como

$$E(q_p|\mathbf{x}) \approx \frac{1}{N} \sum_{i=1}^N \left\{ \mu_i^1 - \frac{\sigma_i^1}{\xi_i^1} \left[ (1 - (-\log(1-p))^{-\xi_i^1}) \right] \right\}.$$

Sob  $\Omega_0$ , a esperança a posteriori de  $q_p$  pode ser aproximada por

$$E(q_p|\mathbf{x}) \approx \frac{1}{N} \sum_{i=1}^N \left\{ \mu_i^0 - \sigma_i^0 \log[-\log(1-p)] \right\}.$$

A integral que define a função de distribuição preditiva sob  $\Omega$ , (2.4), pode ser vista como a esperança a posteriori de

$$\exp \left\{ - \left[ 1 + \xi \left( \frac{y - \mu}{\sigma} \right) \right]_+^{-1/\xi} \right\},$$

e portanto pode ser aproximada pela média amostral

$$G(y|\mathbf{x}) \approx \frac{1}{N} \sum_{i=1}^N \exp \left\{ - \left[ 1 + \xi_i^1 \left( \frac{y - \mu_i^1}{\sigma_i^1} \right) \right]_+^{-1/\xi_i^1} \right\},$$

para cada  $y \in \mathbb{R}$ , sob a amostra simulada da distribuição a posteriori em  $\Omega$ .

Analogamente, a função de distribuição preditiva sob  $\Omega_0$ , (2.5), poderá ser aproximada por

$$G(y|\mathbf{x}) \approx \frac{1}{N} \sum_{i=1}^N \exp \left\{ - \exp \left[ - \left( \frac{y - \mu_i^0}{\sigma_i^0} \right) \right] \right\},$$

para cada  $y \in \mathbb{R}$ .

Por fim, os limitantes dos intervalos de credibilidade  $\gamma$  para  $q_p$  sob  $\Omega$  e  $\Omega_0$ , (2.8) e (2.9), podem ser aproximados. Suponhamos  $\gamma = 0.95$ . Para cada tempo de retorno  $\frac{1}{p}$  fixo, ordenamos (em ordem crescente) os valores de  $q_p$  obtidos a partir de cada  $\theta_i^k = (\mu_i^k, \sigma_i^k, \xi_i^k)$  simulado, e então tomamos como aproximações para os limitantes inferior e superior dos intervalos de credibilidade  $\gamma$ , respectivamente, os quantis empíricos 2.5% e 97.5% dos valores de  $q_p$ .

Temos interesse também em obter intervalos de credibilidade  $\gamma$  para os parâmetros  $\mu$ ,  $\sigma$  e  $\xi$  sob  $\Omega$ . Um intervalo de credibilidade  $\gamma$  para  $\xi$ , por exemplo, é dado por

$$IC_\gamma(\xi) = \{ \xi : a_1 < \xi < a_2 \text{ e } P(\xi < a_1|\mathbf{x}) = P(\xi > a_2|\mathbf{x}) = (1 - \gamma)/2 \}.$$

Os intervalos para  $\mu$  e  $\sigma$  são dados de maneira análoga.

Supondo novamente  $\gamma = 0.95$ , podemos então aproximar os limitantes destes intervalos de credibilidade  $\gamma$ , para  $\xi$  por exemplo: ordenamos os valores  $\xi_i^1$  obtidos por simulação em  $\Omega$ , e então tomamos como aproximações para os limitantes inferior e superior, respectivamente, os quantis empíricos 2.5% e 97.5% dos valores de  $\xi_i^1$ . Os intervalos para  $\mu$  e  $\sigma$  são aproximados analogamente.

Também temos interesse em determinar as correlações amostrais a posteriori entre os parâmetros  $\mu$ ,  $\sigma$  e  $\xi$  sob  $\Omega$ . A correlação amostral a posteriori entre  $\sigma$  e  $\xi$ , por exemplo, é dada

por

$$\text{corr}(\sigma, \xi) = \frac{\sum_{i=1}^N (\sigma_i^1 - \bar{\sigma}^1)(\xi_i^1 - \bar{\xi}^1)}{(N-1)s_{\sigma^1}s_{\xi^1}},$$

onde  $\bar{\sigma}^1$  e  $\bar{\xi}^1$  são as médias amostrais dos valores simulados de  $\sigma$  e  $\xi$ , e  $s_{\sigma^1}$  e  $s_{\xi^1}$  são os desvios padrões amostrais dos valores simulados de  $\sigma$  e  $\xi$ .

O desvio padrão amostral de  $\xi$ , por exemplo, é dado por

$$s_{\xi^1} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (\xi_i^1 - \bar{\xi}^1)^2}.$$

Além disso, quando  $\xi < 0$  a distribuição GEV possui um limitante superior finito, dado por  $\mu - \frac{\sigma}{\xi}$ , de modo que o máximo estudado não supera este valor. Assim torna-se interessante estimar esta quantia. Se  $\hat{\mu}_{MV}$ ,  $\hat{\sigma}_{MV}$  e  $\hat{\xi}_{MV}$  são as estimativas de máxima verossimilhança de  $\mu$ ,  $\sigma$  e  $\xi$  aproximadas computacionalmente, então pelo Princípio de Invariância dos estimadores de máxima verossimilhança, a quantia

$$\hat{\mu}_{MV} - \frac{\hat{\sigma}_{MV}}{\hat{\xi}_{MV}}$$

será a estimativa de máxima verossimilhança aproximada para o limitante superior da distribuição.

Podemos ainda obter outra estimativa pontual para este limitante superior, baseada na média a posteriori, e aproximada por

$$\frac{\sum_{\xi_i^1 < 0} \left( \mu_i^1 - \frac{\sigma_i^1}{\xi_i^1} \right)}{N^*},$$

onde  $N^* = \#\{\xi_i^1 : \xi_i^1 < 0\}$ .

Para obter um intervalo de credibilidade de 95% aproximado para este limitante superior, ordenamos os valores  $\mu_i^1 - \frac{\sigma_i^1}{\xi_i^1}$  obtidos por simulação em  $\Omega$  e restritos a  $\xi_i^1 < 0$ , e então tomamos como aproximações para os limitantes inferior e superior do intervalo, respectivamente, os quantis empíricos 2.5% e 97.5% destes valores.

# Capítulo 5

## Aplicações

### 5.1 Taxas de desemprego

Nesta primeira aplicação estudamos os máximos trimestrais das taxas de desemprego (em %) nos Estados Unidos da América (EUA), no período entre o primeiro trimestre de 1994 e o primeiro trimestre de 2009. Os dados foram obtidos na *homepage* do Banco Central do Brasil, <http://www4.bcb.gov.br/?SERIESTEMP>, e são apresentados na figura A.1.

Neste problema pode ser discutida a adequacidade do modelo GEV como distribuição limite para os máximos trimestrais, dado que há uma aparente e natural correlação entre as observações, além de serem considerados os máximos em intervalos de tempo relativamente pequenos. No entanto, nada nos impede de testar o ajuste da distribuição GEV aos dados disponíveis. Isto é o que faremos.

Como descrevemos anteriormente, distribuições a priori Uniformes em intervalos positivos suficientemente grandes foram estabelecidas para as quantidades assumidas independentes  $\tilde{q}_i$ ,  $i = 1, 2, 3$  sob o modelo GEV e  $i = 1, 2$  sob o modelo Gumbel, induzindo uma distribuição a priori para os parâmetros de interesse sob cada um dos modelos.

Através do algoritmo de Metropolis-Hastings descrito anteriormente, obtivemos amostras simuladas das distribuições a posteriori de  $\theta$  em  $\Omega$  e em  $\Omega_0$ , ambas de tamanho  $N = 10000$ , após considerar períodos de *burn-in* de tamanho 5000. A seleção destes valores foi determinada por critérios gráficos. Foram comparados os histogramas marginais dos valores simulados de  $\mu$ ,  $\sigma$  e  $\xi$ , com os respectivos gráficos simulados  $(\mu_i^1, g_i^1)$ ,  $(\sigma_i^1, g_i^1)$  e  $(\xi_i^1, g_i^1)$ , Figura A.2. A similaridade entre os gráficos e os respectivos histogramas indica que a densidade a posteriori é bem representada pela amostra simulada.

Na figura A.3 apresentamos os histogramas marginais dos valores simulados de  $(\mu, \sigma, \xi)$

sob a distribuição a posteriori em  $\Omega$ , bem como os gráficos de dispersão. Notemos que a maior parte da massa marginal a posteriori (simulada) em  $\xi$  se concentra na região  $\xi > 0$ . Além disso, o máximo da distribuição marginal a posteriori (simulada) em  $\xi$  parece ser atingido próximo de  $\xi = 0$ .

A estimativa média a posteriori, aproximada por  $E(\widehat{\theta}|\mathbf{x})_1$ , (4.5), para os parâmetros da distribuição GEV, é

$$(\hat{\mu}, \hat{\sigma}, \hat{\xi})_{\text{média}} = (4.89, 0.70, -0.0962),$$

enquanto que a estimativa de máxima verossimilhança destes parâmetros é

$$(\hat{\mu}, \hat{\sigma}, \hat{\xi})_{MV} = (4.92, 0.669, -0.0380).$$

Estas estimativas pontuais aproximadas para  $\xi$  aparentemente fornecem alguma evidência de que o modelo Gumbel seria adequado para ajustar os dados dos máximos trimestrais das taxas de desemprego, no sentido de que estas estimativas se apresentam próximas de zero. As outras medidas de evidência aproximadas, apresentadas na Tabela 5.1, também apontam para a mesma direção. O p-valor apresentado na tabela está associado ao teste da Razão de Verossimilhanças (estatística Deviance). A função  $g(\mathbf{x})$  utilizada no cálculo do DIC é assumida ser igual a 1.

Tabela 5.1: Medidas de evidência aproximadas para o teste de hipóteses  $H_0$ : Modelo Gumbel contra  $H_1$ : Modelo GEV, para os dados de taxa de desemprego nos EUA.

e-valor	$2 \log B_{10}$	FBP	$P(\xi \leq 0 \mathbf{x})$	p-valor	BIC	DIC
0.898	-6.46	0.461	0.251	0.663	151.39 ( $\Omega$ )	146.65 ( $\Omega$ )
					147.48 ( $\Omega_0$ )	142.92 ( $\Omega_0$ )

Todas estas medidas de evidência concordam no sentido de sugerir que o modelo Gumbel (estabelecido pela hipótese nula) é mais adequado que o GEV para ajustar as observações em questão.

Ressaltamos que a probabilidade a posteriori  $P(\xi \leq 0|\mathbf{x})$  é aproximada simplesmente pela proporção amostral (segundo a amostra simulada da distribuição a posteriori de  $\theta$  em  $\Omega$ ) de pontos  $(\mu_i, \sigma_i, \xi_i)$  tais que  $\xi_i \leq 0$ .

Em particular, com respeito à regra de decisão (2.7), deveríamos rejeitar a hipótese nula ( $\Omega_0$ ) se, e somente se,

$$0.898(w_0 + c) < w_1 + c \iff w_0 < 1.11w_1 + 0.114c.$$

Especificamente, se  $w_1 \ll c$ , e  $c$  é maior do que 9 vezes  $w_0$  (aproximadamente), então rejeitamos  $\Omega_0$ .

Na figura A.4 apresentamos, na primeira linha, sob os modelos GEV e Gumbel respectivamente, os gráficos dos níveis de retorno esperados a posteriori aproximados em função dos períodos de retorno (linha tracejada). As linhas contínuas limitam os intervalos de credibilidade ( $\gamma = 0.95$ ) aproximados para os níveis de retorno para cada período de retorno fixo. Os pontos mostram os níveis de retorno empíricos em função dos (inversos dos) quantis empíricos. Na segunda linha apresentamos, também sob os modelos GEV e Gumbel respectivamente, a função de distribuição preditiva aproximada para uma observação futura (linha tracejada) e a função de distribuição empírica dos dados (pontos sólidos).

Comparando estes gráficos, entre os modelos, também obtemos suporte para as conclusões obtidas segundo as medidas de evidência encontradas.

Nos gráficos dos níveis de retorno, por exemplo, os intervalos de credibilidade aproximados parecem cobrir quase todos os valores empíricos sob ambos os modelos, sugerindo a escolha do modelo Gumbel como a decisão mais parsimoniosa.

Quanto aos gráficos da função de distribuição preditiva para uma observação futura, a função de distribuição empírica aparentemente se aproxima muito da preditiva, para os dois modelos, sugerindo a mesma escolha.

Apresentamos a seguir os intervalos de credibilidade 95% aproximados para os parâmetros da distribuição GEV.

$$IC_{0.95}(\mu) = (4.66, 5.08),$$

$$IC_{0.95}(\sigma) = (0.56, 0.87),$$

$$IC_{0.95}(\xi) = (-0.11, 0.41).$$

Notemos que o  $IC_{0.95}(\xi)$  fornece evidência de que  $\xi > -0.5$ , de modo que os estimadores de máxima verossimilhança serão regulares.

Além disso, apresentamos as correlações amostrais a posteriori, que concordam com os gráficos de dispersão apresentados na figura A.3,

$$corr(\mu, \sigma) = 0.40,$$

$$corr(\mu, \xi) = -0.35,$$

$$corr(\sigma, \xi) = -0.03.$$

Devemos destacar a baixa correlação a posteriori obtida entre  $\sigma$  e  $\xi$ .

## 5.2 Níveis de maré

Estudamos agora os dados apresentados na figura A.5, que correspondem às observações dos máximos semestrais dos níveis de maré em Newlyn, no sudoeste da Inglaterra, medidos em metros acima de uma constante chamada ODN, entre 1990 e 2007. ODN (*Ordnance Datum Newlyn*) é uma constante baseada na média dos níveis de maré em Newlyn entre 1915 e 1921. Os dados foram fornecidos pelo *British Oceanographic Data Centre*.

Assim como fizemos na primeira aplicação, distribuições a priori Uniformes em intervalos positivos suficientemente grandes foram estabelecidas para as quantidades assumidas independentes  $\tilde{q}_i$ ,  $i = 1, 2, 3$  sob o modelo GEV e  $i = 1, 2$  sob o modelo Gumbel, induzindo uma distribuição a priori para os parâmetros de interesse sob cada um dos modelos.

Novamente utilizando o algoritmo de Metropolis-Hastings, simulamos amostras de tamanho  $N = 10000$  a partir das distribuições a posteriori de  $\theta$  em  $\Omega$  e  $\Omega_0$ , levando em consideração períodos de *burn-in* de tamanho 5000 em ambos os casos.

Apresentamos os histogramas marginais dos valores simulados de  $(\mu, \sigma, \xi)$  sob a distribuição a posteriori em  $\Omega$ , bem como os gráficos de dispersão destes valores, na Figura A.7.

Notemos aqui, bem como havia ocorrido na aplicação anterior, o caráter unimodal das distribuições a posteriori marginais, como pode ser observado nestes histogramas.

No entanto, ao contrário do que havia acontecido na primeira aplicação, a maior parte da massa marginal a posteriori (simulada) em  $\xi$  se concentra na região  $\xi \leq 0$ , como podemos observar na Tabela 5.2 o valor aproximado de  $P(\xi \leq 0|\mathbf{x})$ , ou mesmo no histograma marginal a posteriori dos valores simulados de  $\xi$ .

Observamos, também na Figura A.7, que a distribuição marginal a posteriori (simulada) em  $\xi$  no ponto  $\xi = 0$  não assume densidade próxima do máximo, e nem próxima de zero.

A estimativa média a posteriori, aproximada por  $E(\widehat{\theta}|\mathbf{x})_1$  (4.5), para os parâmetros  $(\mu, \sigma, \xi)$  da distribuição GEV, é

$$(\hat{\mu}, \hat{\sigma}, \hat{\xi})_{\text{média}} = (6.02, 0.171, -0.184),$$

Já a estimativa de máxima verossimilhança destes parâmetros é

$$(\hat{\mu}, \hat{\sigma}, \hat{\xi})_{MV} = (6.03, 0.165, -0.367).$$

Agora as estimativas pontuais de  $\xi$  já não se aproximam tanto de zero como anteriormente, especialmente a estimativa de máxima verossimilhança.

Corroborando esta observação, vemos que as medidas de evidência aproximadas fornecem decisões de certa forma discrepantes, conforme apresentado na Tabela 5.2. Novamente, a função  $g(\mathbf{x})$  utilizada no cálculo do DIC é assumida ser igual a 1.

Tabela 5.2: Medidas de evidência aproximadas para o teste de hipóteses  $H_0$ : Modelo Gumbel contra  $H_1$ : Modelo GEV, para os dados de níveis de maré em Newlyn, Inglaterra.

e-valor	$2 \log B_{10}$	FBP	$P(\xi \leq 0   \mathbf{x})$	p-valor	BIC	DIC
0.483	-4.57	23.8	0.87	0.00455	<b>-20.14</b> ( $\Omega$ ) -15.73 ( $\Omega_0$ )	<b>-34.24</b> ( $\Omega$ ) -29.38 ( $\Omega_0$ )

Os critérios BIC e DIC, assim como o p-valor e o fator de Bayes a posteriori (FBP) contra a hipótese nula fornecem forte evidência contra o modelo Gumbel, em favor do modelo GEV. Tanto o BIC quanto o DIC são menores sob o modelo GEV. No contexto clássico, um p-valor igual a 0.00455 entrega forte evidência contra a hipótese nula a qualquer nível de significância maior que 0.00455.

Já a medida  $2 \log B_{10}$ , baseada no fator de Bayes  $B_{10}$ , não fornece evidência contra  $\Omega_0$ , o modelo Gumbel.

O e-valor fornecido pelo FBST, por sua vez, apresenta uma evidência moderada em favor do modelo Gumbel. Especificamente, de acordo com (2.7), deveríamos rejeitar a hipótese nula (modelo Gumbel) se, e somente se,

$$w_0 < 2.07w_1 + 1.07c.$$

Neste caso, se  $w_1 \ll c$ , e  $c$  é maior do que  $w_0$ , então rejeitamos  $\Omega_0$ .

Lembremos que a distribuição marginal a posteriori simulada em  $\xi = 0$  não é tão baixa, e este fato está refletido no e-valor obtido.

Na Figura A.8 apresentamos, na primeira linha, sob os modelos GEV e Gumbel respectivamente, os gráficos dos níveis de retorno esperados a a posteriori aproximados em função dos períodos de retorno (linha tracejada), os intervalos de credibilidade ( $\gamma = 0.95$ ) aproximados para os níveis de retorno para cada período de retorno fixo (linhas contínuas) e os níveis de retorno empíricos em função dos períodos de retorno empíricos (pontos sólidos), para os dados dos níveis de maré em Newlyn. Na segunda linha apresentamos, também sob os modelos GEV e Gumbel respectivamente, a função de distribuição preditiva aproximada para uma observação futura (linha tracejada) e a função de distribuição empírica dos dados (pontos sólidos).

Podemos observar, nos gráficos dos níveis de retorno, que os intervalos de credibilidade aproximados não cobrem todos os valores empíricos sob ambos os modelos, e de forma mais destacada sob o modelo Gumbel, sugerindo neste sentido, a escolha do modelo GEV. No entanto, se considerarmos que é pequena a quantia de pontos empíricos fora desta região de credibilidade, o modelo Gumbel ainda poderia ser levado em conta. Aparentemente o e-valor e o fator de Bayes traduzem esta possibilidade neste problema.

Na figura A.6 apresentamos os gráficos simulados de  $(\mu_i^1, g_i^1)$ ,  $(\sigma_i^1, g_i^1)$  e  $(\xi_i^1, g_i^1)$ . A similaridade destes gráficos com os respectivos histogramas marginais simulados indica que a densidade a posteriori é bem representada pela amostra simulada. Novamente, este foi o critério utilizado para selecionar o tamanho da amostra simulada e o período de *burn-in*.

Apresentamos agora os intervalos de credibilidade (95%) aproximados para os parâmetros  $\mu$ ,  $\sigma$  e  $\xi$  da distribuição GEV.

$$IC_{0.95}(\mu) = (5.95, 6.07),$$

$$IC_{0.95}(\sigma) = (0.13, 0.22),$$

$$IC_{0.95}(\xi) = (-0.44, 0.30).$$

Aqui, assim como havia ocorrido na primeira aplicação, o  $IC_{0.95}(\xi)$  fornece evidência de que  $\xi > -0.5$ , garantindo a regularidade dos estimadores de máxima verossimilhança.

Também obtivemos as correlações amostrais a posteriori entre os parâmetros da distribuição GEV, que novamente concordam com os respectivos gráficos de dispersão,

$$corr(\mu, \sigma) = -0.04,$$

$$corr(\mu, \xi) = -0.32,$$

$$corr(\sigma, \xi) = 0.09.$$

Além disso, segundo as estimativas pontuais para  $\xi$  e a probabilidade a posteriori  $P(\xi < 0 | \mathbf{x})$  aproximada, que é igual a 0.87, temos evidência de que  $\xi < 0$ , de modo que neste caso a distribuição do máximo apresenta um limitante superior finito, cuja estimativa de máxima verossimilhança aproximada é igual a 6.48. Assim o máximo semestral do nível de maré em Newlyn, segundo este modelo e esta estimativa, não deve superar 6.48 metros acima de ODN.

Já a estimativa média a posteriori para este limitante superior é igual a 8.37, enquanto que o seu intervalo de credibilidade 95% aproximado é dado por (6.44, 14.02).

# Capítulo 6

## Conclusão

No sentido de buscar um procedimento bayesiano para realizar o teste para a hipótese precisa do modelo Gumbel no contexto de valores extremos, estudamos o *Full Bayesian significance test*, o qual fornece uma medida de evidência bayesiana em favor da hipótese nula, a qual chamamos *e-valor*.

De um ponto de vista exploratório pensamos que o *e-valor* fornecido pelo FBST pode ser interessante para testar modelos a custos relativamente baixos. Para calcular o *e-valor* não é necessário obter uma densidade a priori e nem a posteriori em  $\Omega_0$ , como é o caso ao se calcular o fator de Bayes, por exemplo. Para o *e-valor* somente é necessária a distribuição a posteriori em todo o espaço paramétrico  $\Omega$ .

A partir da simulação computacional da distribuição a posteriori, podemos obter o *e-valor* para qualquer hipótese precisa, desde que a função de densidade seja contínua no subespaço correspondente.

Como vimos, um critério de decisão baseado no *e-valor* é um procedimento bayesiano, no sentido de que existe uma função de perda cujo valor esperado sob a distribuição a posteriori é minimizado pelo FBST. Em um problema de decisão real, podemos e devemos levar em conta a função de perda associada ao problema, no sentido de realizar o teste adequadamente. Isso nos proporciona interpretar a força da evidência fornecida.

Também estudamos outras medidas de evidência para o mesmo teste de hipóteses, nos contextos clássico e bayesiano, proporcionando assim uma comparação descritiva entre as regras de decisão estabelecidas pelos diferentes procedimentos, como discutido nas aplicações apresentadas.

Devemos ressaltar que no processo de seleção de modelos, de forma geral, devem ser considerados a parsimônia do modelo e a precisão nas previsões.

No que diz respeito à possibilidade de estudos futuros, podemos sugerir um estudo

mais profundo que leve em conta aspectos preditivistas, ou a precisão nas previsões, dos modelos comparados através do teste, como critério de seleção levando em conta dados de fato observados a partir do fenômeno ou processo em estudo, após a coleta da amostra original. Neste sentido, supõe-se que poderíamos comparar tais valores futuros (de fato observados) com valores preditos segundo a própria distribuição preditiva, estabelecendo assim algum tipo de medida de precisão.

Também podemos sugerir um estudo que leve em consideração uma análise de sensibilidade, estabelecendo algumas distribuições a priori distintas na modelagem.

# Referências

- AITKIN, M. **Posterior Bayes Factors**.(With discussion). J. R. Statist. Soc. B, [S.I.], v. 53, n. 1, p.111-142, 1991.
- CHIB, S. ; GREENBERG, E. **Understanding the Metropolis-Hastings Algorithm**. The American Statistician, [S.I.], v. 49, n. 4, p.327-335, 1995.
- COLES, S. **An Introduction to Statistical Modeling of Extreme Values**. [S.I.]: Springer, 2001.
- COLES, S. G. ; TAWN, J. A. **A Bayesian Analysis of Extreme Rainfall Data**. Appl. Statist., [S.I.], v. 45, n. 4, p.463-478, 1996.
- GAMERMAN, D. **Simulação Estocástica Via Cadeias de Markov**. 1. ed. São Paulo: Associação Brasileira de Estatística, 1996. 196p.
- GELMAN, A. et al. **Bayesian Data Analysis**. 2. ed. [S.I.]: Chapman & Hall/CRC, 2004. 668p.
- HASTINGS, W. K. **Monte Carlo Sampling Methods Using Markov Chains and Their Applications**. Biometrika, [S.I.], v. 57, p.97-109, 1970.
- JEFFREYS, H. **Theory of Probability**. 3. ed. Oxford, U.K.: Oxford University Press, 1961.
- KASS, R. E. ; RAFTERY, A. E. **Bayes Factors**. Journal of the American Statistical Association, [S.I.], v. 90, n. 430, p.773-795, 1995.
- MADRUGA, M. R.; ESTEVES, L. G.; WECHSLER, S. **On the Bayesianity of Pereira-Stern tests**. Test, [S.I.], v. 10, n. 2, p.291-299, 2001.
- METROPOLIS, N. et al. **Equations of State Calculations by Fast Computing Machines**. Journal of Chemical Physics, [S.I.], v. 21, p.1087-1092, 1953.
- O'HAGAN, A. **Bayesian Inference**. [S.I.:s.n.], 1994. (Kendall's Advanced Theory of Statistics, 2B).
- PEREIRA, C. A. B.; STERN, J.; WECHSLER, S. **Can a Significance Test be Genuinely Bayesian?**. Bayesian Analysis, [S.I.], v. 3, p.79-100, 2008.
- RESNICK, S. I. **Extreme Values, Regular Variation, and Point Processes**. [S.I.]: Springer, 1987.
- SCHWARZ, G. **Estimating the Dimension of a Model**. The Annals of Statistics, [S.I.], v. 6, n. 2, p.461-464, 1978.
- SMITH, R. L. **Maximum likelihood estimation in a class of non-regular cases**. Biometrika, [S.I.], v. 72, p.67-90, 1985.
- SPIEGELHALTER, D. J. *et al.* **Bayesian measures of model complexity and fit**. J. R. Statist. Soc. B, [S.I.], v. 64, Part 4, p.583-639, 2002.

STEPHENSON, A.; TAWN, J. **Bayesian Inference for Extremes: Accounting for the Three Extremal Types**. *Extremes*, [S.I.], v. 7, p.291-307, 2004.

# Bibliografia

CASELLA, G. ; GEORGE, E. I. **Explaining the Gibbs Sampler**. The American Statistician, [S.I.], v. 46, n. 3, p.167-174, 1992.

GUMBEL, E. J. **Statistics of Extremes**. [S.I.]: Columbia Univ. Press, 1958.

HÄGGSTRÖM, O. **Finite Markov Chains and Algorithmic Applications**. [S.I.]: Cambridge Univ. Press, 2002.

LEADBETTER, M. R.; LINDGREN, G.; ROOTZÉN, H. **Extremes and Related Properties of Random Sequences and Processes**. [S.I.]: Springer, 1983.

ROBERT, C. P.; CASELLA, G. **Monte Carlo Statistical Methods**. 2. ed.. [S.I.]: Springer, 2004.

# Apêndice A

## Figuras

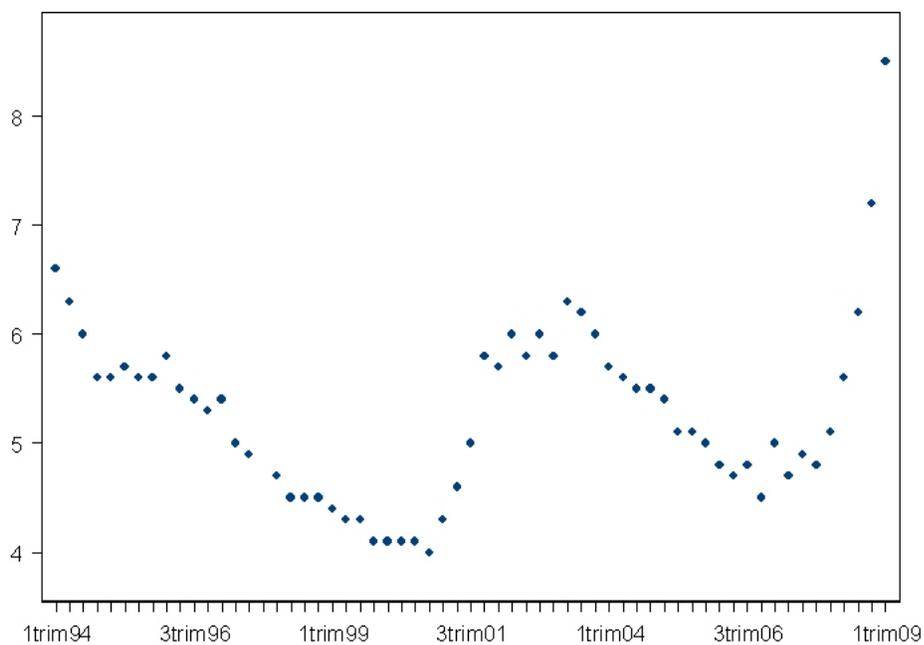


Figura A.1: Máximos trimestrais das taxas de desemprego (%) nos EUA, entre o primeiro trimestre de 1994 e o primeiro trimestre de 2009.

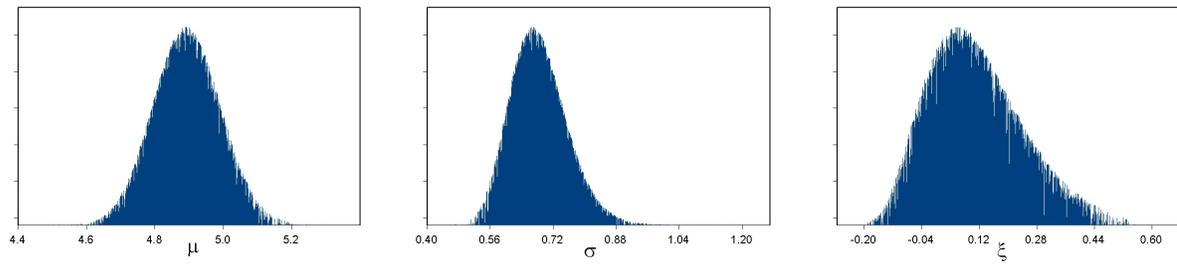


Figura A.2: Gráficos marginais simulados de  $(\mu_i^1, g_i^1)$ ,  $(\sigma_i^1, g_i^1)$  e  $(\xi_i^1, g_i^1)$ , para os dados das taxas de desemprego nos EUA.

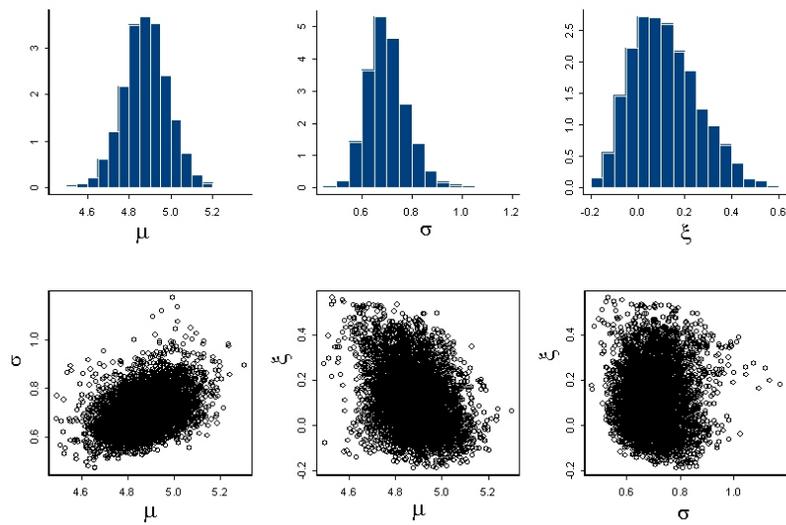


Figura A.3: Histogramas marginais a posteriori para  $\mu$ ,  $\sigma$  e  $\xi$  e gráficos de dispersão, obtidos por simulação, para os dados das taxas de desemprego nos EUA.

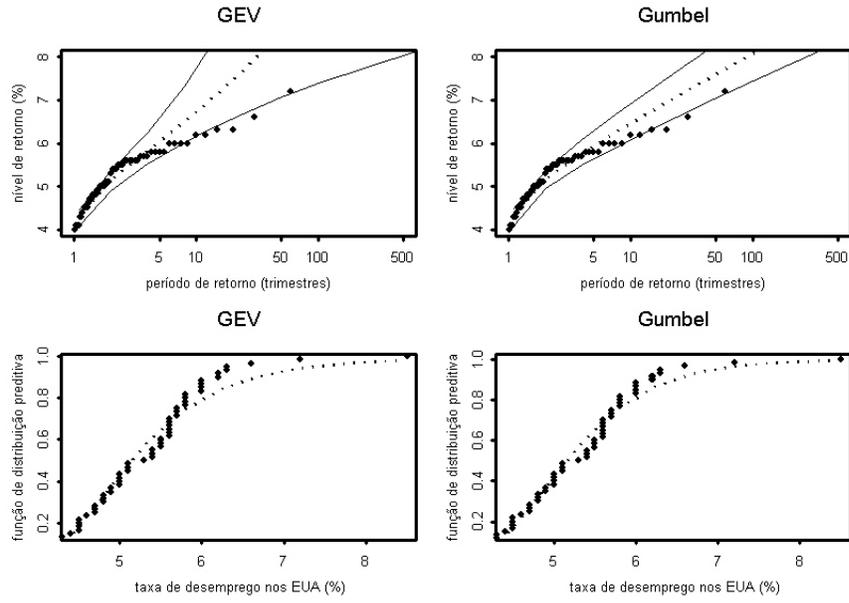


Figura A.4: Primeira linha: níveis de retorno esperados a posteriori (linha tracejada), intervalos de credibilidade ( $\gamma = 0.95$ ) para os níveis de retorno (linhas contínuas) e níveis de retorno empíricos (pontos sólidos); segunda linha: função de distribuição preditiva (linha tracejada) e função de distribuição empírica (pontos sólidos); aproximações sob os modelos GEV e Gumbel, para os dados das taxas de desemprego nos EUA.

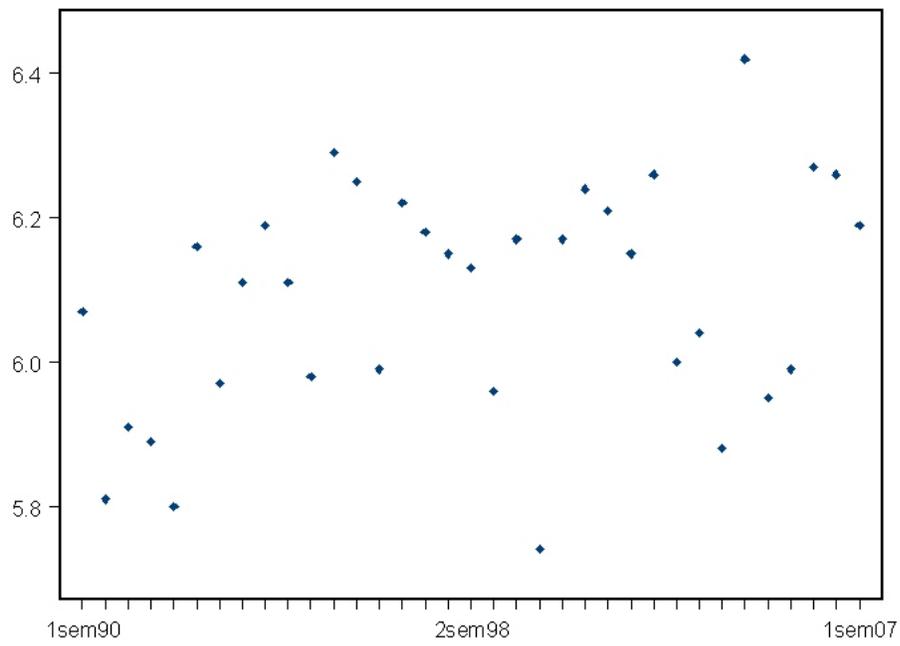


Figura A.5: Máximos semestrais dos níveis de maré (metros acima de ODN) em Newlyn, sudoeste da Inglaterra, entre 1990 e 2007.

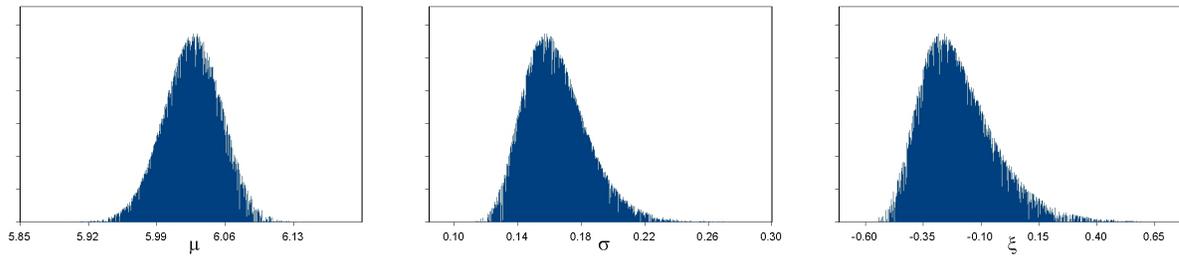


Figura A.6: Gráficos marginais simulados de  $(\mu_i^1, g_i^1)$ ,  $(\sigma_i^1, g_i^1)$  e  $(\xi_i^1, g_i^1)$ , para os dados dos níveis de maré em Newlyn.

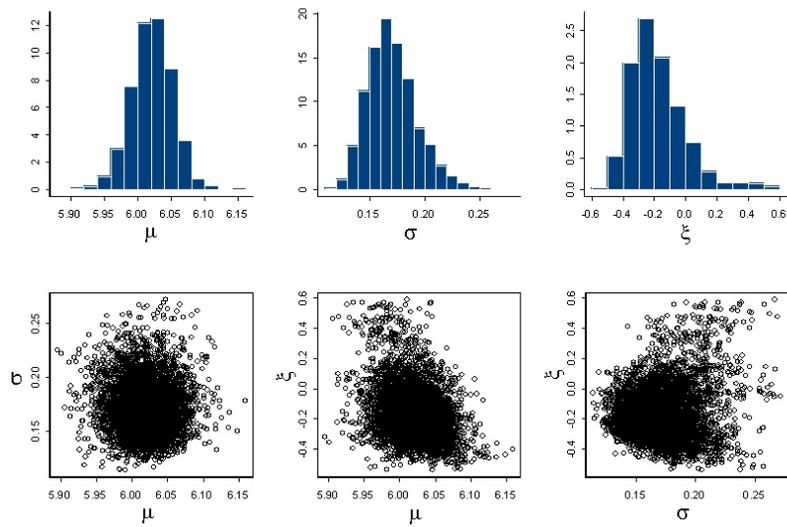


Figura A.7: Histogramas marginais a posteriori para  $\mu$ ,  $\sigma$  e  $\xi$  e gráficos de dispersão, obtidos por simulação, para os dados dos níveis de maré em Newlyn.

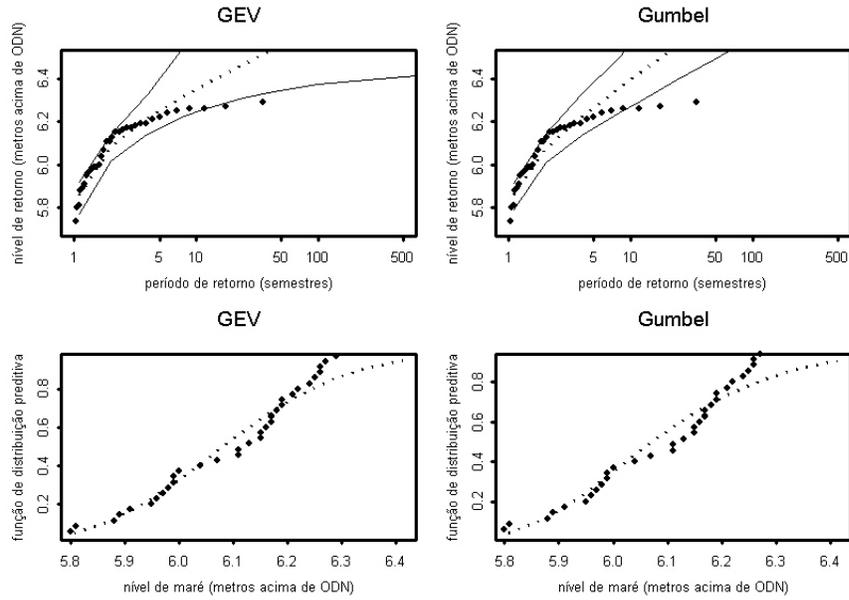


Figura A.8: Primeira linha: níveis de retorno esperados a posteriori (linha tracejada), intervalos de credibilidade ( $\gamma = 0.95$ ) para os níveis de retorno (linhas contínuas) e níveis de retorno empíricos (pontos sólidos); segunda linha: função de distribuição preditiva (linha tracejada) e função de distribuição empírica (pontos sólidos); aproximações sob os modelos GEV e Gumbel, para os dados dos níveis de maré em Newlyn.

# Apêndice B

## Programas

PROGRAMA 1 - Priori, Verossimilhança e Posteriori

```
UKdata<- (dados de máximos);
lUKdata<-length(data);

y1<--log(.9);y2<--log(.99);y3<--log(.999);
log21<-log(y2/y1);log31<-log(y3/y1);log32<-log(y3/y2);

jacob<-function(b,c){
soma<- -(y1*y2)^(-c)*log21+(y1*y3)^(-c)*log31-
(y2*y3)^(-c)*log32;
abs(soma)*b/c^2}

b1<- (definir em cada problema);
b2<- (definir em cada problema);
b3<- (definir em cada problema);

prioriUK<-function(a,b,c){
qtil1<- a+b*(y1^(-c)-1)/c;
qtil2<- b*(y2^(-c)-y1^(-c))/c;
qtil3<- b*(y3^(-c)-y2^(-c))/c;pri<-0;
if (0<=qtil1 && qtil1<=b1 && 0<=qtil2 && qtil2<=b2 &&
0<=qtil3 && qtil3<=b3){
```

```

pri<-jacob(b,c); }
pri
}

likeliUK<-function(a,b,c){
fact<-0;
for (i in 1:lUKdata){
ppi<-max(0,1+c*(UKdata[i]-a)/b);
fact[i]<-ppi^(-(c+1)/c)*exp(-ppi^(-1/c)) }
likeli<-b^(-lUKdata)*prod(fact);
likeli
}

posteriorUK<-function(a,b,c){
prioriUK(a,b,c)*likeliUK(a,b,c)
}

b10<- (definir em cada problema);
b20<- (definir em cada problema);

prioriUK0<-function(a,b){
qtil10<- a-b*log(y1);
qtil20<- b*(log(y1)-log(y2)); pri0<-0;
if (0<=qtil10 && qtil10<=b10 && 0<=qtil20 && qtil20<=b20){
pri0<-1;
}
pri0
}

likeliUK0<-function(a,b){
fact0<-0;
for (i in 1:lUKdata){
ppi0<--(UKdata[i]-a)/b;
fact0[i]<-exp(-exp(ppi0))*exp(ppi0); }

```

```

likeli0<-b^(-1UKdata)*prod(fact0);
likeli0
}

```

```

posteriorUK0<-function(a,b){
prioriUK0(a,b)*likeliUK0(a,b)}

```

PROGRAMA 2 - Metropolis-Hastings para posterior em  $\Theta=(\mu,\sigma,\xi)$  (GEV)

```

mu0<-definir;
sig0<-definir;
xi0<-definir;

```

```

alpha<- function(a,b,c,d,e,f){
R<-posteriorUK(d,e,f)/posteriorUK(a,b,c);
min(1,R)
}

```

```

UKMHmu<-0;
UKMHsig<-0;
UKMHxi<-0;
N<- ;
delta1<- ; delta2<- ; delta3<- ;
UKMHmu[1]<-mu0; UKMHsig[1]<-sig0; UKMHxi[1]<-xi0;
cont<-0;
alph<-0;
for(i in 1:N){
z1<-runif(1,min=-delta1,max=delta1);
z2<-runif(1,min=-delta2,max=delta2);
z3<-runif(1,min=-delta3,max=delta3);
u<-runif(1,min=0,max=1);
yc1<- UKMHmu[i]+z1;
yc2<- UKMHsig[i]+z2;

```

```

yc3<- UKMHxi[i]+z3;
alph<-alpha(UKMHmu[i],UKMHsig[i],UKMHxi[i],yc1,yc2,yc3);
if (is.finite(alph) && u <= alph){
UKMHmu[i+1]<- yc1;
UKMHsig[i+1]<- yc2;
UKMHxi[i+1]<- yc3;
cont<-cont+1; }
else{
UKMHmu[i+1]<- UKMHmu[i];
UKMHsig[i+1]<- UKMHsig[i];
UKMHxi[i+1]<- UKMHxi[i];
}
}

burn<- ;
mu<-0; sig<-0; xi<-0;
for(i in 1:(N+1-burn)){
mu[i]<- UKMHmu[burn+i];
sig[i]<- UKMHsig[burn+i];
xi[i]<- UKMHxi[burn+i];
}

par(mfrow=c(2,2))
hist(mu, plot=T, probability=T)
hist(sig, plot=T, probability=T)
hist(xi, plot=T, probability=T)

post<-0;
for(i in 1:(N+1-burn)){
post[i]<- posteriorUK(mu[i],sig[i],xi[i]);
}

c(mean(mu), mean(sig), mean(xi))

```

```

like<-0;
for(i in 1:(N+1-burn)){
like[i]<- likeliUK(mu[i],sig[i],xi[i]);
}

wh<- which(like==max(like));
c(mu[wh], sig[wh], xi[wh]);
ML1<- like[wh];

B1<- mean(like);

ord2<- order(xi);
post2<- post[ord2];
xi2<- xi[ord2];
I0<- which(abs(xi2)<0.001);
f0<- max(post2[I0]);
evalor <- 1-length(post2[post2>f0])/(N+1-burn);

```

PROGRAMA 3 - Metropolis-Hastings para posterior em  $\Theta=(\mu,\sigma)$  (Gumbel)

```

Gmu0<- ;
Gsig0<- ;

alpha0<- function(a,b,d,e){
R<-posteriorUK0(d,e)/posteriorUK0(a,b);
min(1,R)
}

GUKMHmu<-0;
GUKMHsig<-0;
GN<- ;
Gdelta1<- ; Gdelta2<- ;
GUKMHmu[1]<-Gmu0; GUKMHsig[1]<-Gsig0;

```

```

Gcont<-0;
Galph<-0;
for(i in 1:GN){
z1<-runif(1,min=-Gdelta1,max=Gdelta1);
z2<-runif(1,min=-Gdelta2,max=Gdelta2);
u<-runif(1,min=0,max=1);
yc1<- GUKMHmu[i]+z1;
yc2<- GUKMHsig[i]+z2;
Galph<-alpha0(GUKMHmu[i],GUKMHsig[i],yc1,yc2);
if (is.finite(Galph) && u <= Galph){
GUKMHmu[i+1]<- yc1;
GUKMHsig[i+1]<- yc2;
Gcont<-Gcont+1;
}
else{
GUKMHmu[i+1]<- GUKMHmu[i];
GUKMHsig[i+1]<- GUKMHsig[i];
}
}

Gburn<- ;
Gmu<-0; Gsig<-0;
for(i in 1:(GN+1-Gburn)){
Gmu[i]<- GUKMHmu[Gburn+i];
Gsig[i]<- GUKMHsig[Gburn+i];
}

Gpost<-0;
for(i in 1:(GN+1-Gburn)){
Gpost[i]<- posteriorUK0(Gmu[i],Gsig[i]);
}

Glike<-0;
for(i in 1:(GN+1-Gburn)){

```

```

Glike[i]<- likeliUK0(Gmu[i],Gsig[i]);
}

ML0<- max(Glike);

B0<- mean(Glike);

PROGRAMA 4 - Evidências

# Fator de Bayes a Posteriori contra H0

FBP<- B1/B0;

# P-valor

dev<- 2*(log(ML1)-log(ML0));
pvalor<- 1-pchisq(dev,1);

# BIC

BIC0 <- -2*log(ML0)+2*log(1UKdata);
BIC1 <- -2*log(ML1)+3*log(1UKdata);

#DIC

llike<-0;
for(i in 1:(N+1-burn)){
llike <- log(like[i]);
}

D1 <- mean(llike);
D2 <- 2*log(likeliUK(mean(mu), mean(sig), mean(xi)));
DIC1 <- -4*D1+D2;

```

```

Gllike <-0;
for(i in 1:(N+1-burn)){
Gllike <- log(Glike[i]);
}

GD1 <- mean(Gllike);
GD2 <- 2*log(likeliUK0(mean(Gmu), mean(Gsig)));
DICO <- -4*GD1+GD2;

# Fator de Bayes contra H0

invlike<-0;
for(i in 1:(N+1-burn)){
invlike[i]<- like[i]^(-1);
}

minvlike <- mean(invlike);
fbc1 <- minvlike^(-1);

Ginvlike<-0;
for(i in 1:(N+1-burn)){
Ginvlike[i]<- Glike[i]^(-1);
}

Gminvlike <- mean(Ginvlike);
fbc0 <- Gminvlike^(-1);

lFB <- 2*log(fbc1/fbc0);

```

PROGRAMA 5 - Tempos e Níveis de Retorno

```

emppf<- function(y){

```

```

(1-(y/lUKdata))(-1)
}

emprp<-0;
for(y in 1:(lUKdata-1)){
emprp[y]<- emppf(y);
}

ord<-order(UKdata);
UKord<-UKdata[ord];
empqp<-UKord[1:(lUKdata-1)];

tamMH<- length(mu21);

rp<- c(1.1, 2, 4, 8, 12, 24, 48, 100, 1000);
lrp<- length(rp);

MHRetLevel<-function(x){
cumi<-0;
for (i in 1:tamMH){
ai<-mu[i]; bi<-sig[i]; ci<-xi[i];
cumi[i] <- ai-((bi/ci)*(1-(-log(1-x(-1)))(-ci))));
}
cumiord<- order(cumi);
cumiordered<- cumi[cumiord];
c(mean(cumi),cumiordered[251],cumiordered[9750]);
}

qpmean<-0; qpmin<-0; qpmax<-0;
for (i in 1:lrp){
zp<-rp[i];
teste<- MHRetLevel(zp)
qpmean[i]<- teste[1]
qpmin[i]<- teste[2]
}

```

```

qpmax[i]<- teste[3]
}

plot(rp,qpmean,log="x",type="l",xlim=c(1,500),ylim=c(4,8),lty=2,lwd=4,cex=1.5);
lines(rp,qpmin,type="l")
points(emprp,empqp,pch=16);
lines(rp,qpmax,type="l")

GtamMH<-length(Gmu21); GtamMH

GMHRetLevel<-function(x){
cumi<-0;
for (i in 1:GtamMH){
ai<-Gmu[i]; bi<-Gsig[i];
cumi[i]<- ai-(bi*log(-log(1-x^(-1)))));
}
cumiord<- order(cumi);
cumioordered<- cumi[cumiord];
c(mean(cumi),cumioordered[251],cumioordered[9750]);
}

Gqpmean<-0; Gqpmin<-0; Gqpmax<-0;
for (i in 1:lrp){
zp<-rp[i];
teste<-GMHRetLevel(zp)
Gqpmean[i]<- teste[1]
Gqpmin[i]<- teste[2]
Gqpmax[i]<- teste[3]
}

plot(rp,Gqpmean,log="x",type="l",xlim=c(1,500),ylim=c(4,8),lty=2,lwd=4,cex=1.5);
lines(rp,Gqpmin,type="l")
points(emprp,empqp,pch=16);
lines(rp,Gqpmax,type="l")

```

## PROGRAMA 6 - Distribuição preditiva

```
acumgev<-function(y,a,b,c){  
pp<-max(0,1+c*(y-a)/b);  
acum <- exp(-(pp^(-1/c)));  
acum  
}
```

```
acumgumbel<-function(y,a,b){  
acum0<-exp(-exp(-((y-a)/b)));  
acum0  
}
```

```
FDacumemp <- 0;  
for(i in 1:LUKdata){  
FDacumemp[i]<- i/1UKdata;  
}
```

```
grade <- function(n,vetor){  
a1<-min(vetor, na.rm=T); a2<-max(vetor, na.rm=T);  
zp<-0;  
for (i in 1:n){  
zp[i] <- a1 + (a2 - a1) * i * .99 / n  
}  
zp  
}
```

```
grad<- 10;  
obsfut<- grade(grad, UKord);
```

```
acumu<-0;  
acpredic<-0;
```

```

for(i in 1:length(obsfut)){
for(j in 1:length(mu)){
acumu[j]<- acumgev(obsfut[i],mu[j],sig[j],xi[j]);
}
acpredic[i]<- mean(acumu);
}

plot(obsfut, acpredic, type="l",lty=2,lwd=4,cex=1.5);
points(UKord, FDacumemp, pch=16);

acumu0<-0;
acpredic0<-0;
for(i in 1:length(obsfut)){
for(j in 1:length(mu)){
acumu0[j]<- acumgumbel(obsfut[i],Gmu[j],Gsig[j]);
}
acpredic0[i]<- mean(acumu0);
}

plot(obsfut, acpredic0, type="l",lty=2,lwd=4,cex=1.5);
points(UKord, FDacumemp, pch=16);

```

#### PROGRAMA 7 - Intervalos de credibilidade e correlação

```

muord<- order(mu);
muordered<- mu[muord];
c(muordered[251],muordered[9750]);

sigord<- order(sig);
sigordered<- sig[sigord];
c(sigordered[251],sigordered[9750]);

xiord<- order(xi);

```

```
xiordered<- xi[xiord];  
c(xiordered[251],xiordered[9750]);
```

```
# Correlações #
```

```
cor(mu,sig);  
cor(mu,xi);  
cor(sig,xi);
```