



Universidade de Brasília  
Instituto de Ciências Exatas  
Departamento de Matemática

# **Inferência via Máxima Verossimilhança para a Distribuição Pareto Generalizada**

BRUNO DANIEL MAZETO

Brasília  
2019

Universidade de Brasília  
Instituto de Ciências Exatas  
Departamento de Matemática

# Inferência via Máxima Verossimilhança para a Distribuição Pareto Generalizada

Bruno Daniel Mazeto

Dissertação apresentada como requisito parcial para  
a obtenção do título de Mestre em Matemática

Orientadora  
Prof<sup>a</sup>. Daniele da Silva Baratela Martins Neto

Brasília  
2019

Universidade de Brasília  
Instituto de Ciências Exatas  
Departamento de Matemática

# Inferência via Máxima Verossimilhança para a Distribuição Pareto Generalizada

por

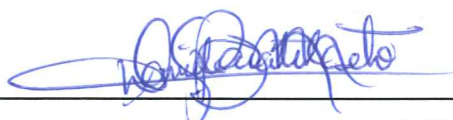
**Bruno Daniel Mazeto**

*Dissertação apresentada ao Departamento de Matemática da Universidade  
de Brasília, como parte dos requisitos para obtenção do grau de*

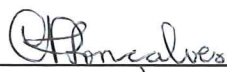
**MESTRE EM MATEMÁTICA**

Brasília, 15 de fevereiro de 2019.

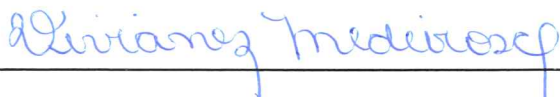
Comissão Examinadora:



Prof. Dra. Daniele da Silva Baratela Martins Neto - MAT/UnB (Orientador)



Prof. Dra. Cátia Regina Gonçalves – MAT/UnB (Membro)



Prof. Dra. Viviane Simioli Medeiros Campos – UFRN (Membro)

Ficha catalográfica elaborada automaticamente,  
com os dados fornecidos pelo(a) autor(a)

MM476i Mazeto, Bruno  
Inferência via Máxima Verossimilhança para a Distribuição  
Pareto Generalizada / Bruno Mazeto; orientador Daniele da  
Silva Baratela Martins Neto. -- Brasília, 2019.  
87 p.

Dissertação (Mestrado - Mestrado em Matemática) --  
Universidade de Brasília, 2019.

1. Distribuição Pareto Generalizada. 2. Estimador de  
máxima verossimilhança. 3. Inferência. 4. Distribuição de  
valor extremo generalizada. I. da Silva Baratela Martins  
Neto, Daniele, orient. II. Título.

## Resumo

Neste trabalho, fazemos um estudo baseado no método de máxima verossimilhança na distribuição Pareto generalizada (GPD). Apresentamos com detalhes os resultados teóricos de Castillo e Serra [1], Castillo e Daoudi [2] e Kozubowski [3], que demonstraram formalmente em quais regiões a função de verossimilhança admite um máximo global e que deram resultados matemáticos que fornecem argumentos precisos para explicar o comportamento anômalo da função de verossimilhança quando obtida da amostragem da distribuição GPD para valores positivos de  $\kappa$  (índice caudal). Simulações e uma aplicação deste modelo com dados reais serão apresentadas para ilustrar alguns dos resultados teóricos.

**Palavras-chave:** estimador de máxima verossimilhança, distribuição de valor extremo generalizada, distribuição Pareto generalizada.

## Abstract

In this work, we do a study based on the maximum likelihood method in the generalized Pareto distribution (GPD). We present, in detail, the theoretical results of Castillo and Serra [1], Castillo and Daoudi [2] and Kozubowski [3], who formally demonstrated in which regions the likelihood function admits a global maximum and that gave mathematical results that provide precise arguments to explain the anomalous behavior of the likelihood function when sampling from the GPD distribution for positive values of  $\kappa$  (tail index). Simulations and an application of this model with real data are presented to illustrate some of the theoretical results.

**Keywords:** maximum likelihood estimator, generalized extreme value distribution, generalized Pareto distribution.

# SUMÁRIO

|  |           |
|--|-----------|
| <b>Introdução</b>  | <b>1</b>  |
| <b>1 Teoria extremal clássica</b>  | <b>4</b>  |
| 1.1 Introdução . . . . .   | 4         |
| 1.2 Distribuições extremas clássicas . . . . .   | 5         |
| 1.3 Max-domínios de atração e distribuições max-estáveis . . . . .   | 8         |
| 1.3.1 Max-domínio de atração da Fréchet ( $D_{max}(\Phi_\alpha)$ ) . . . . .                                     | 9         |
| 1.3.2 Max-domínio de atração da Weibull ( $D_{max}(\Psi_\alpha)$ ) . . . . .                                     | 14        |
| 1.3.3 Max-domínio de atração da Gumbel ( $D_{max}(\Lambda)$ ) . . . . .  | 17        |
| <b>2 Distribuição Pareto generalizada</b>  | <b>20</b> |
| 2.1 Introdução . . . . .   | 20        |
| 2.2 Distribuição de valor extremo generalizada – GEV ( <i>Generalized Extreme Value Distribution</i> ) . . . . . | 21        |
| 2.3 Distribuição Pareto generalizada – GPD ( <i>Generalized Pareto Distribution</i> ) .                          | 26        |
| <b>3 Inferência via máxima verossimilhança</b>   | <b>37</b> |
| 3.1 Introdução . . . . .   | 37        |
| 3.2 Princípio da máxima verossimilhança . . . . .  | 38        |
| 3.3 Inferência para a distribuição Pareto generalizada . . . . .   | 42        |

---

|          |  |           |
|----------|--|-----------|
| 3.3.1    | Existência do estimador de máxima verossimilhança do parâmetro $\kappa$<br>(índice caudal) . . . . . | 43        |
| 3.3.2    | Matriz de informação de Fisher da GPD . . . . .  | 54        |
| 3.4      | Testes de hipóteses . . . . .  | 56        |
| 3.4.1    | Critério de informação Akaike (AIC) . . . . .  | 56        |
| 3.4.2    | Teste de seleção $\chi^2$ . . . . .  | 58        |
| <b>4</b> | <b>Simulações e aplicações práticas</b>  | <b>63</b> |
| 4.1      | Introdução . . . . .   | 63        |
| 4.2      | Simulações numéricas . . . . .   | 64        |
| 4.3      | Aplicação com dados reais . . . . .  | 73        |
|          | <b>Bibliografia</b>  | <b>79</b> |



# Introdução

São muitas as situações e diversas as áreas em que estão presentes problemas envolvendo eventos extremos. Como um exemplo, podemos citar a ocorrência de desastres naturais. Em climatologia, destacamos as inundações. O interesse aí consiste na previsão do nível máximo de água para facilitar o planejamento de ações preventivas. Outra área que vale ser mencionada, em que eventos extremos são observados, é a confiabilidade. Na manutenção de um equipamento constituído por componentes em paralelo, por exemplo, o tempo de falha é determinado pelo tempo máximo de vida desses componentes. Em Galambos [4] e Embrechts [5] são apresentadas outras situações envolvendo extremos (máximos ou mínimos).

Dessa forma, a modelagem matemática de problemas com essas características consiste em considerar variáveis aleatórias  $X_1, X_2, \dots, X_n$  independentes e identicamente distribuídas (i.i.d.), com funções de distribuição (f.d.) comum  $F$ , e analisar o comportamento de

$$Z_n = \max \{X_1, \dots, X_n\}$$

ou  $W_n = \min \{X_1, \dots, X_n\}$ . Como  $W_n = -\max \{-X_1, \dots, -X_n\}$ , o estudo restringe-se ao comportamento de  $Z_n$ .

A teoria extremal clássica estuda principalmente o comportamento assintótico do máximo,  $Z_n$ , estabilizado linearmente, ou seja, o comportamento limite quando  $n \rightarrow \infty$  de

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

em que  $a_n > 0$  e  $b_n$  são sequências de números reais.

Os principais resultados da teoria extremal clássica determinam condições sobre a função de distribuição  $F$  para a existência de constantes estabilizantes  $a_n > 0$  e  $b_n$  para as quais a distribuição (1) convirja para uma distribuição não-degenerada. Além disso, a teoria mostra que existem apenas três tipos possíveis de distribuições limites, as chamadas distribuições extremas clássicas: Fréchet ( $\Phi_\alpha$ ,  $\alpha > 0$ ), Gumbel ( $\Lambda$ ) e Weibull ( $\Psi_\alpha$ ,  $\alpha > 0$ ).

Dentre os trabalhos pioneiros no desenvolvimento desse estudo merecem destaque Fisher e Tippett [6], que formularam os três tipos de distribuições limites para o máximo de uma amostra aleatória; Fréchet [7], que estudou os possíveis limites de (1), embora tenha tido êxito somente com uma delas, conhecida como distribuição do tipo Fréchet; Gnedenko [8], que caracterizou completamente duas distribuições limites: aquela estudada por Fréchet, ou

seja, a distribuição de cauda pesada, e aquela estudada por Weibull, ou seja, a distribuição de cauda leve; e De Haan [9], que analisou as condições de limite para a distribuição do tipo Gumbel; entre outros.

Uma parte importante da teoria extremal consiste no estudo da distribuição de valor extremo generalizada (GEV – *Generalized Extreme Value distribution*), a qual combina três distribuições extremas. A principal característica da GEV é ter uma representação uniparamétrica para as três distribuições, com a introdução do parâmetro  $\xi$  ( $\xi = \alpha^{-1} > 0$ , para  $\Phi_\alpha$ ;  $\xi = 0$ , para  $\Lambda$ ; e  $\xi = -\alpha^{-1} < 0$ , para  $\Psi_\alpha$ ), cuja motivação é simplificar as implementações estatísticas. Dentro desse estudo, surge a distribuição Pareto generalizada (GPD – *Generalized Pareto distribution*), introduzida por Pickands [10]. Ela tem sido utilizada para modelar excedentes acima de um limiar. A GPD é dada por

$$G_{\xi,\beta}(x) = 1 - \left(1 + \xi \frac{x}{\beta}\right)^{-1/\xi}, \quad x \in D(\xi, \beta),$$

em que, para  $\beta > 0$ ,

$$D(\xi, \beta) = \begin{cases} [0, \infty), & \text{se } \xi \geq 0 \\ [0, -\beta/\xi], & \text{se } \xi < 0 \end{cases},$$

( $G_{0,\beta}$  é definida no limite com  $\xi \rightarrow 0$ ). Poderá ser conferido com detalhes neste trabalho que a GPD apresenta excelentes propriedades que contribuem para a inferência do modelo. Ressaltaremos a importância da GPD na análise estatística das funções de distribuição do excesso de uma variável acima de um limiar. Se  $X$  é uma variável aleatória com função de distribuição  $F$ , então para  $t < \sup\{x : F(x) < 1\}$ ,

$$F_t(x) = P(X - t \leq x | X > t), \quad x \geq 0,$$

é chamada função de distribuição do excesso de  $X$  acima do limiar  $t$ . Apesar das boas propriedades da GPD, na literatura são destacados alguns problemas quanto à estimação dos parâmetros dessa distribuição. Como exemplo, podemos citar casos em que o estimador de máxima verossimilhança (EMV) de  $\theta = (\kappa, \beta)$ , em que  $\kappa = -\xi$  (índice caudal) e  $\beta > 0$ , não existe, assim análises do EMV de  $\theta$  são feitas com base em técnicas numéricas, como pode ser visto em Castillo e Hadi [11] e Zhang e Stephens [12].

Neste trabalho, apresentaremos um estudo da GPD quanto à estimação de máxima verossimilhança do modelo, baseado em Castillo e Serra [1], Castillo e Daoudi [2] e Kozubowski [3]. Uma análise detalhada do EMV de  $\theta = (\kappa, \beta)$  será apresentada do ponto de vista teórico. Tendo como referências básicas os trabalhos mencionados, mostraremos, dentre outros resultados, que para  $\kappa < 1$  o EMV existe (Teorema 3.2), para  $k \leq 0$ , em que a GPD é

---

de cauda pesada, existe o EMV global (Lema 3.1) e, para  $\kappa \geq 1$ , a função de verossimilhança é ilimitada, portanto, o EMV não existe (Proposição 3.5).

Dessa forma, no Capítulo 1 introduziremos os conceitos e resultados preliminares da teoria extremal clássica para os fins deste trabalho. As principais referências desse capítulo são Galambos [4] e Embrechts [5].

No Capítulo 2, será apresentada a distribuição de valor extremo generalizada (GEV) e como ela se relaciona às três distribuições extremas clássicas, quando introduzida uma representação uniparamétrica, por intermédio da inserção do parâmetro  $\xi$ . Isso poderá ser conferido na Seção 2.2. Além disso, por meio de um dos resultados de caracterização do máximo domínio de atração da GEV, apresentaremos na Seção 2.3 a distribuição Pareto generalizada (GPD), juntamente com as principais propriedades dessa distribuição, relevantes para este trabalho.

No Capítulo 3, será aplicada a inferência via máxima verossimilhança na estimação dos parâmetros da GPD. Em especial, na Seção 3.3, serão analisadas as regiões onde o EMV de  $\kappa$  existe e onde a função de máxima verossimilhança é ilimitada. Os resultados deste capítulo são provenientes dos artigos Castillo e Serra [1], Castillo e Daoudi [2] e Kozubowski [3].

Finalmente, no Capítulo 4, serão apresentadas simulações numéricas e aplicações práticas, que ilustram alguns resultados da teoria extremal. Especificamente, na Seção 4.2, será apresentado o algoritmo proposto por Castillo e Serra [1] para o estimador de máxima verossimilhança dos parâmetros da GPD. Com ele, faremos alguns estudos sobre como dados extremos podem ser modelados, conforme a teoria vista nos capítulos precedentes. Este trabalho é concluído com a Seção 4.3, em que será apresentado um exemplo de um caso real, no qual são estimados o limiar  $t$  e os parâmetros  $\kappa$  e  $\beta$ .

## CAPÍTULO

# 1

# TEORIA EXTREMAL CLÁSSICA

## 1.1 Introdução

O estudo da teoria extremal consiste em analisar o comportamento de  $Z_n = \max\{X_1, \dots, X_n\}$  e  $W_n = \min\{X_1, \dots, X_n\}$ , em que, para cada  $n \in \mathbb{N}$ ,  $X_1, \dots, X_n$  são variáveis aleatórias independentes e identicamente distribuídas (i.i.d.), com função de distribuição (f.d.) comum  $F$ . Uma vez que  $W_n = -\max\{-X_1, \dots, -X_n\}$ , o estudo restringe-se ao comportamento de  $Z_n$ .

Sabemos que a f.d. de  $Z_n$  é dada por:

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

Uma questão de interesse nesse estudo é o comportamento limite de  $Z_n$ , quando  $n \rightarrow \infty$ . Como exemplo, podemos citar o volume de cheias numa barragem; seria de grande valia estimar os níveis máximos anuais de água.

Agora, observe que:

$$P(Z_n \leq x) \xrightarrow{n} \begin{cases} 0, & \text{se } x: F(x) < 1, \\ 1, & \text{se } x: F(x) = 1. \end{cases} \quad (1.1)$$

Se denotarmos o ponto extremo superior de  $F$  por:

$$\omega(F) = \sup \{x : F(x) < 1\} \leq +\infty, \quad (1.2)$$

então de (1.1) temos que  $Z_n \xrightarrow{P} \omega(F)$  e, como  $\{Z_n\}$  é uma sequência não-decrescente, temos também que  $Z_n \xrightarrow{q.c.} \omega(F)$ . Esse fato não fornece muita informação e, por isso, torna-se necessário normalizar  $Z_n$ . Assim sendo, o objetivo principal da teoria extremal clássica é analisar as possíveis distribuições limites para o máximo sob normalização linear:

$$Z_n^* = \frac{Z_n - b_n}{a_n},$$

em que  $\{a_n\}$  e  $\{b_n\}$  são sequências de números reais, com  $a_n > 0$ . Isso quer dizer que o objetivo é estabelecer condições sobre  $F$  que garantam que a distribuição de  $Z_n^*$  convirja para uma distribuição não-degenerada para escolhas apropriadas das constantes  $a_n > 0$  e  $b_n$ , ou seja,

$$P(Z_n \leq a_n x + b_n) = F^n(a_n x + b_n) \xrightarrow{n} H(x), \quad x \in C(H), \quad (1.3)$$

em que  $H$  é f.d. não-degenerada e  $C(H)$  é o conjunto de pontos de continuidade de  $H$ .

Fisher e Tippett [6] mostraram que existem apenas três tipos de distribuições que satisfazem o limite (1.3). Fréchet [7] estudou o limite (1.3) para um tipo de distribuição, que ficou conhecida como distribuição do tipo Fréchet. Gnedenko [8] apresentou, com rigor e precisão, as condições necessárias e suficientes para que as distribuições do tipo Fréchet e do tipo Weibull atendam ao limite (1.3). De Haan [9] analisou as condições de limite para a distribuição do tipo Gumbel. Esses trabalhos têm sido citados como as principais referências para esse problema desde então.

Neste capítulo, serão apresentados os principais resultados da teoria extremal clássica necessários para este trabalho. Na Seção 1.2, apresentaremos as possibilidades para o limite não-degenerado em (1.3). Na Seção 1.3, mostraremos condições suficientes sobre a função de distribuição  $F$  para cada uma dessas possibilidades.

Como referências para o que será tratado neste capítulo, podem-se citar Galambos [4] e Embrechts [5].

## 1.2 Distribuições extremais clássicas

Nesta seção, será apresentado sem demonstração o Teorema 1.1, que determina as possíveis distribuições limites não-degeneradas que satisfazem o limite (1.3). Ressaltamos desde já

que as sequências de constantes que estabilizam esse limite não são únicas. Antes, porém, serão introduzidos alguns conceitos e resultados preliminares.

**Definição 1.1.** *Uma função de distribuição não-degenerada,  $H$ , que é resultado do limite (1.3), é dita **distribuição de valor extremo** ou **distribuição extremal**.*

As sequências de constantes que estabilizam o limite de  $F^n(a_nx + b_n)$  não são únicas. Contudo, se, para  $a_n > 0$  e  $c_n > 0$ ,  $b_n$  e  $d_n$  constantes reais e  $T$  e  $G$  f.d.'s não-degeneradas, tais que

$$\lim_{n \rightarrow \infty} F^n(a_nx + b_n) = T(x)$$

e

$$\lim_{n \rightarrow \infty} F^n(c_nx + d_n) = G(x),$$

então é possível mostrar que existem constantes  $A > 0$  e  $B$  tais que

$$T(x) = G(Ax + B). \quad (1.4)$$

Nesse caso, as f.d.'s  $T$  e  $G$  são ditas **distribuições do mesmo tipo**.

Na verdade, se (1.3) é satisfeita, então mostra-se que, para  $m \geq 1$  inteiro arbitrário, existem constantes  $A_m > 0$  e  $B_m \in \mathbb{R}$  tais que

$$H^m(A_mx + B_m) = H(x). \quad (1.5)$$

Neste caso,  $H$  é dita **função de distribuição max-estável (linear)**.

Os resultados citados acima podem ser vistos com detalhes em Galambos [4].

**Definição 1.2** (Max-domínio de atração). *Seja  $H$  uma f.d. não-degenerada. A f.d.  $F$  é dita pertencer ao **max-domínio de atração (linear) de  $H$**  (denota-se  $F \in D_{\max}(H)$ ), se existem constantes  $\{a_n\}$  e  $\{b_n\}$ , com  $a_n > 0$ , tais que*

$$\lim_{n \rightarrow \infty} F^n(a_nx + b_n) = H(x), \quad \forall x \in C(H). \quad (1.6)$$

As constantes  $\{a_n\}$  e  $\{b_n\}$  são chamadas **constantes normalizadoras**.

O teorema a seguir é um resultado básico da Teoria Extremal Clássica, proposto por Fisher e Tippett [6] em 1928. Nele são apresentadas as possíveis distribuições max-estáveis. Vale ressaltar que existem apenas 3 tipos de distribuições limites para o máximo de variáveis aleatórias i.i.d., como pode ser conferido no Teorema 1.1.

**Teorema 1.1** (Fisher-Tippett). *Se existem sequências de constantes reais  $\{a_n\}$  e  $\{b_n\}$ , com  $a_n > 0$ , tais que*

$$\lim_{n \rightarrow \infty} P((Z_n - b_n)/a_n \leq x) = H(x), \quad x \in C(H),$$

*em que  $H$  é uma função de distribuição não-degenerada, então  $H$  é do tipo de uma das seguintes distribuições:*

$$\text{Fréchet: } \Phi_\alpha(x) = \begin{cases} 0, & x \leq 0 \\ \exp\{-x^{-\alpha}\}, & x > 0 \end{cases} \quad \alpha > 0, \quad (1.7)$$

$$\text{Weibull: } \Psi_\alpha(x) = \begin{cases} \exp\{-(-x)^\alpha\}, & x \leq 0 \\ 1, & x > 0 \end{cases} \quad \alpha > 0, \quad (1.8)$$

$$\text{Gumbel: } \Lambda(x) = \exp\{-e^{-x}\}, \quad x \in \mathbb{R}. \quad (1.9)$$

**Demonstração.** Ver Fisher e Tippett [6] ou Galambos [4].

**Observação 1.1.** As f.d.'s  $\Phi_\alpha$ ,  $\Psi_\alpha$  e  $\Lambda$ , apresentadas no Teorema 1.1, e suas f.d.'s do mesmo tipo são as **distribuições de valor extremo** ou **distribuições extremais**.

**Observação 1.2.** Note que as distribuições extremais são contínuas em  $\mathbb{R}$ , assim temos que:

$$\frac{Z_n - b_n}{a_n} \xrightarrow{d} H \iff \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = H(x), \quad \forall x \in \mathbb{R}.$$

( $\xrightarrow{d}$ : convergência em distribuição).

**Definição 1.3.** *Diz-se que uma função de distribuição  $F$  pertence à classe de distribuições de cauda leve à direita se para algum  $\varepsilon > 0$  tem-se que*

$$\limsup_{y \rightarrow \infty} \frac{1 - F(y)}{e^{-\varepsilon y}} < \infty.$$

**Definição 1.4.** *Diz-se que uma função de distribuição  $F$  pertence à classe de distribuições de cauda pesada à direita se a função geradora de momentos não é finita, ou seja,  $F^*(s) = \infty$ , para todo  $s > 0$ .*

**Observação 1.3.** Assim, mostra-se que a distribuição Fréchet pertence à classe de distribuições de cauda pesada. Da mesma forma, mostra-se que a distribuição Weibull pertence à classe distribuições de cauda leve.

**Exemplo 1.1.** (Máximo de variáveis aleatórias exponenciais)

Seja uma sequência  $\{X_n\}$  de variáveis aleatórias i.i.d. com distribuição exponencial de

parâmetro  $\lambda = 1$ . Tomando  $b_n = \ln n$  e  $a_n = 1$ , temos que

$$P(Z_n - \ln n \leq x) = \left(P(X_1 \leq x + \ln n)\right)^n = \left(1 - e^{-x - \ln n}\right)^n = \left(1 - \frac{e^{-x}}{n}\right)^n \\ \xrightarrow{n} \exp\{-e^{-x}\} = \Lambda(x), \quad x \in \mathbb{R}.$$

**Exemplo 1.2.** (Máximo de variáveis aleatórias uniforme)

Seja uma sequência  $\{X_n\}$  de variáveis aleatórias i.i.d. com distribuição uniforme  $U(0, 1)$ . Tomando  $b_n = 1$  e  $a_n = 1/n$ , temos que

$$P((Z_n - 1)n \leq x) = \left(P(X_1 \leq x/n + 1)\right)^n = \left(1 + \frac{x}{n}\right)^n, \quad -n < x \leq 0 \\ \xrightarrow{n} e^x = \Psi_1(x), \quad x \leq 0.$$

### 1.3 Max-domínios de atração e distribuições max-estáveis

Nesta seção, o interesse principal é apresentar as condições necessárias e suficientes sobre uma função de distribuição  $F$  que garantam a existência do limite (1.3) para cada um dos três tipos de distribuições de valor extremo apresentados no Teorema 1.1.

Para cada um dos três tipos de distribuições extremas, será exibida a demonstração da suficiência dessas condições. Isso poderá ser visto nos Teoremas 1.2, 1.3 e 1.4. Entretanto, antes, serão relacionadas algumas definições e resultados necessários.

**Proposição 1.1** (Caracterização do  $D_{\max}(H)$ ). *A f.d.  $F$  pertence ao max-domínio de atração de uma distribuição de valor extremo  $H$ , com constantes normalizadoras  $a_n > 0$  e  $b_n \in \mathbb{R}$ , se, e somente se,*

$$\lim_{n \rightarrow \infty} n\bar{F}(a_n x + b_n) = -\ln H(x), \quad x \in \mathbb{R}, \quad (1.10)$$

em que  $\bar{F} = 1 - F$ . Quando  $H(x) = 0$ , o limite é interpretado como infinito.

**Demonstração.** Ver Embrechts [5].

**Definição 1.5.** *Se  $F$  é uma f.d., dizemos que os números  $\gamma(F) = \inf\{x : F(x) > 0\}$  e  $\omega(F) = \sup\{x : F(x) < 1\}$  são respectivamente o **ponto extremo inferior** de  $F$  e o **ponto extremo superior** de  $F$ .*

**Definição 1.6** (Inversa generalizada). *Seja  $h$  uma função real não-decrescente. Denominamos a **inversa generalizada** de  $h$  a função*

$$h^{\leftarrow}(t) = \inf\{x \in \mathbb{R} : h(x) \geq t\}. \quad (1.11)$$



(Usa-se a convenção de que o ínfimo de um conjunto vazio é  $\infty$ .)

**Definição 1.7** (Função quantil). A inversa generalizada da f.d.  $F$

$$F^{\leftarrow}(t) = \inf \{x \in \mathbb{R} : F(x) \geq t\}, \quad 0 < t < 1, \quad (1.12)$$

é chamada a **função quantil** da f.d.  $F$ . A quantidade  $x_t = F^{\leftarrow}(t)$  define o **t-quantil** de  $F$ .

**Definição 1.8.** Uma função  $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  é dita **regularmente variante** (no infinito) com índice  $\alpha \in \mathbb{R}$  (ou  $\alpha$ -variante no infinito), e denotamos  $g \in R_\alpha$ , se

$$\lim_{t \rightarrow \infty} \frac{g(tx)}{g(t)} = x^\alpha, \quad \forall x > 0. \quad (1.13)$$

Se  $\alpha = 0$ , dizemos que  $g$  é **lentamente variante** (no infinito).

**Exemplo 1.3.** A distribuição exponencial de parâmetro  $\lambda$  tem função de distribuição  $F(x) = 1 - e^{-\lambda x}$ . Assim, aplicando a definição de regularmente variante, temos que

$$\lim_{t \rightarrow \infty} \frac{F(tx)}{F(t)} = \lim_{t \rightarrow \infty} \frac{1 - e^{-\lambda tx}}{1 - e^{-\lambda t}} = 1.$$

Sendo assim,  $F$  é lentamente variante ( $F(x) \in R_0$ ).

**Exemplo 1.4.** Para qualquer  $\beta \in \mathbb{R}$ , a função  $L(x) = \ln^\beta x$  é regularmente variante.

$$\lim_{t \rightarrow \infty} \frac{(\ln tx)^\beta}{(\ln t)^\beta} = \lim_{t \rightarrow \infty} \left( \frac{\ln t + \ln x}{\ln t} \right)^\beta = \lim_{t \rightarrow \infty} \left( 1 + \frac{\ln x}{\ln t} \right)^\beta = 1$$

Portanto,  $L(x) \in R_0$ .

### 1.3.1 Max-domínio de atração da Fréchet ( $D_{\max}(\Phi_\alpha)$ )

Iniciaremos esta seção apresentando alguns resultados que serão utilizados na caracterização do max-domínio de atração de cada uma das distribuições extremas. Esses resultados poderão ser vistos pelos Lemas 1.1 e 1.2 e pelo Corolário 1.1.

Na sequência, serão apresentadas as condições para que uma função de distribuição pertença ao max-domínio de atração  $D_{\max}(\Phi_\alpha)$ .

**Lema 1.1.** Para qualquer  $0 < z < 1/2$ ,

$$e^{-nz} - (1-z)^n \{ \exp(2nz^2) - 1 \} < (1-z)^n \leq e^{-nz}, \quad n \in \mathbb{N}. \quad (1.14)$$

A desigualdade da direita vale também para  $0 \leq z \leq 1$ .

**Demonstração.** Para a demonstração do lado direito da desigualdade (1.14), observe primeiro que  $(1-z)^n = \exp\{n \ln(1-z)\}$ , para  $z \in [0, 1)$  e  $n \in \mathbb{N}$ . Daí, basta mostrar que  $\ln(1-z) \leq -z$ .

Seja a função  $g(z) = \ln(1-z) + z$ ,  $z \in [0, 1)$ . Temos que

$$g(0) = 0$$

e

$$g'(z) = \frac{-1}{1-z} + 1 \leq 0, \quad \forall z \in [0, 1).$$

Assim, temos que  $g(z) \leq 0$  para todo  $z \in [0, 1)$  e, dessa forma, segue  $(1-z)^n \leq e^{-nz}$ .

Para verificar o lado esquerdo da desigualdade (1.14), note primeiro que

$$\frac{1}{1-z} < 1 + 2z, \quad \forall z \in (0, 1/2). \quad (1.15)$$

Agora, integrando em  $(0, z)$  em ambos os lados de (1.15), temos

$$\ln(1-z) > -z - z^2$$

Daí, para  $z \in (0, 1/2)$ ,

$$(1-z) \ln(1-z) > -z$$

e, assim,

$$(1-z)^{n(1-z)} > e^{-nz}. \quad (1.16)$$

Agora, do lado direito da desigualdade (1.14) e de (1.16), temos

$$0 < e^{-nz} - (1-z)^n < (1-z)^{n-nz} - (1-z)^n = (1-z)^n [(1-z)^{-nz} - 1]. \quad (1.17)$$

Como  $-\ln(1-z) < 2z$ , para  $0 < z < 1/2$ , temos que

$$(1-z)^{-nz} = \exp\{-nz \ln(1-z)\} < \exp(2nz^2).$$

Por fim, daí e de (1.17), temos para  $z \in (0, 1/2)$

$$e^{-nz} - (1-z)^n < (1-z)^n \{\exp(2nz^2) - 1\}.$$

Portanto,

$$e^{-nz} - (1-z)^n \{ \exp(2nz^2) - 1 \} < (1-z)^n. \quad \square$$

**Lema 1.2.** *Sejam  $X_1, \dots, X_n$  variáveis aleatórias i.i.d. com distribuição comum  $F$ . Seja  $x$  tal que*

$$1 - F(x) < \frac{1}{2\sqrt{n}}. \quad (1.18)$$

Então, para  $n \in \mathbb{N}$ ,

$$T(x) - 4n[1 - F(x)]^2 F^n(x) < P(Z_n < x) < T(x),$$

em que  $T(x) = \exp \{ -n[1 - F(x)] \}$ .

**Demonstração.** Apliquemos o Lema 1.1 com  $z = 1 - F(x)$ . Assim, temos que

$$P(Z_n < x) = [F(x)]^n = (1-z)^n \leq e^{-nz} = e^{-n(1-F(x))} = T(x),$$

o que prova a segunda desigualdade.

Para a primeira parte da desigualdade, fazendo as substituições, usando o resultado elementar

$$|e^w - 1| < 2w, \quad 0 < w < 1/2,$$

temos, para  $w = 2nz^2$  e pelo Lema 1.1, que

$$\begin{aligned} T(x) - 4n[1 - F(x)]^2 F^n(x) &= e^{-nz} - 4nz^2(1-z)^n < \\ &< e^{-nz} - (1-z)^n [\exp(2nz^2) - 1] < (1-z)^n = P(Z_n < x). \quad \square \end{aligned}$$

**Corolário 1.1.** *Usando as notações do Lema 1.2, assuma que existam seqüências de números reais  $a_n > 0$  e  $b_n$  tais que, para todo  $x$ ,*

$$\lim_{n \rightarrow \infty} n [1 - F(a_n x + b_n)] = u(x). \quad (1.19)$$

Então existe

$$\lim_{n \rightarrow \infty} P(Z_n < a_n x + b_n) = \exp [-u(x)]. \quad (1.20)$$

**Demonstração.** Consideramos primeiramente o caso  $u(x) = +\infty$ . Fazendo  $z = 1 - F(a_n x + b_n)$ , temos do Lema 1.1

$$(1-z)^n = F(a_n x + b_n) = P(Z_n < a_n x + b_n) < \exp \left\{ -n [1 - F(a_n x + b_n)] \right\}.$$

De (1.19), temos que

$$\lim_{n \rightarrow \infty} P(Z_n < a_n x + b_n) < \exp \left\{ - \lim_{n \rightarrow \infty} \left[ n(1 - F(a_n x + b_n)) \right] \right\} = 0,$$

e, portanto, segue o resultado.

Agora, assumimos  $u(x) < +\infty$ .

De (1.19), segue que, para  $x \in \mathbb{R}$ ,

$$\lim_{n \rightarrow \infty} (1 - F(a_n x + b_n)) = 0,$$

e, sendo assim, para  $n$  suficientemente grande, temos que

$$1 - F(a_n x + b_n) < \frac{1}{2\sqrt{n}}.$$

Dessa maneira, pelo Lema 1.2, temos

$$\begin{aligned} \exp \left\{ -n[1 - F(a_n x + b_n)] \right\} - 4n[1 - F(a_n x + b_n)]^2 F^n(a_n x + b_n) < \\ < P(Z_n < a_n x + b_n) < \exp \left\{ -n[1 - F(a_n x + b_n)] \right\}. \end{aligned}$$

Novamente, de (1.19), temos que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \left\{ n[1 - F(a_n y + b_n)] \right\}^2 = 0,$$

e o resultado é obtido. □

No teorema a seguir, será vista a caracterização do max-domínio de atração da distribuição Fréchet,  $\Phi_\alpha$ , para  $\alpha > 0$ . Usaremos a notação  $\bar{F} = 1 - F$ .

**Teorema 1.2** (Max-domínio de atração da Fréchet). *Seja  $F$  uma f.d. e  $\alpha > 0$ . Então  $F \in D_{\max}(\Phi_\alpha)$  se, e somente se,*

$$\omega(F) = \infty \text{ e } \bar{F} \in R_{-\alpha} \text{ para alguma constante } \alpha > 0. \quad (1.21)$$

Neste caso, tem-se que  $b_n = 0$  e

$$\lim_{n \rightarrow \infty} P(Z_n \leq a_n x) = \lim_{n \rightarrow \infty} F^n(a_n x) = \Phi_\alpha(x), \quad (1.22)$$

em que  $a_n$  pode ser escolhido como

$$a_n = \inf \{x : 1 - F(x) \leq 1/n\} = F^{\leftarrow}(1 - 1/n).$$

**Demonstração** (Suficiência). Suponha que  $\omega(F) = \infty$  e  $\bar{F} \in R_{-\alpha}$  para alguma constante  $\alpha > 0$ . Pela Proposição 1.1 (caracterização do  $D_{\max}(H)$ ), em que  $H$ : f.d. extremal) e pelo Corolário 1.1, basta verificar que

$$\lim_{n \rightarrow \infty} n\bar{F}(a_n x) = \begin{cases} x^{-\alpha} & , x > 0 \\ \infty & , x < 0 \end{cases}.$$

Como  $\omega(F) = \infty$ , segue que  $a_n = \inf \{x : F(x) \geq 1 - \frac{1}{n}\} \xrightarrow{n} \infty$ . Se  $x < 0$ , temos que

$$a_n x \xrightarrow{n} -\infty \text{ e, portanto, } \lim_{n \rightarrow \infty} [1 - F(a_n x)] = 1.$$

Logo,

$$\lim_{n \rightarrow \infty} n\bar{F}(a_n x) = \infty,$$

e o resultado segue para  $x < 0$ .

Agora, para  $x > 0$ , temos, da hipótese que  $\bar{F} \in R_{-\alpha}$  e do fato  $a_n \rightarrow \infty$ , que

$$\begin{aligned} \lim_{n \rightarrow \infty} n\bar{F}(a_n x) &= \lim_{n \rightarrow \infty} n\bar{F}(a_n) \frac{n\bar{F}(a_n x)}{n\bar{F}(a_n)} \\ &= x^{-\alpha} \lim_{n \rightarrow \infty} n\bar{F}(a_n). \end{aligned}$$

Resta agora provar que

$$\lim_{n \rightarrow \infty} n\bar{F}(a_n) = 1.$$

Para isso, considere  $0 < \varepsilon < 1$  arbitrário. Como  $\bar{F}(x)$  é não-crescente e  $a_n(1 - \varepsilon) < a_n$ , então  $\bar{F}(a_n(1 - \varepsilon)) \geq \bar{F}(a_n)$ . Pela definição de  $a_n$ ,

$$\bar{F}(a_n) \leq \frac{1}{n} < \bar{F}(a_n(1 - \varepsilon)),$$

e, portanto,

$$\frac{\bar{F}(a_n)}{\bar{F}(a_n(1 - \varepsilon))} < n\bar{F}(a_n) \leq 1.$$

Novamente de  $\bar{F} \in R_{-\alpha}$ , temos que

$$(1 - \varepsilon)^\alpha \leq \liminf n\bar{F}(a_n) \leq \limsup n\bar{F}(a_n) \leq 1.$$

Portanto,

$$\lim_{n \rightarrow \infty} n\bar{F}(a_n) = 1,$$

e a suficiência está mostrada. □

(Necessidade) A prova será omitida. Para detalhes veja Galambos [4].

**Exemplo 1.5** (Distribuição Cauchy padrão).  $F(x) = \frac{1}{2} + \frac{\arctan x}{\pi}$ ,  $x \in \mathbb{R}$ .

Temos que  $\omega(F) = \infty$  e

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{\bar{F}(tx)}{\bar{F}(t)} &= \lim_{t \rightarrow \infty} \frac{\frac{1}{2} - \frac{1}{\pi} \arctan(tx)}{\frac{1}{2} - \frac{1}{\pi} \arctan(t)} = \lim_{t \rightarrow \infty} \frac{\frac{x}{(xt)^2 + 1}}{\frac{1}{t^2 + 1}} = \lim_{t \rightarrow \infty} \frac{x(t^2 + 1)}{(xt)^2 + 1} \\ &= \lim_{t \rightarrow \infty} \frac{x2t}{x^2 2t} = x^{-1}. \end{aligned}$$

Assim,  $\bar{F} \in R_{-1}$ . As hipóteses do Teorema 1.2 são satisfeitas com  $\alpha = 1$ . Logo,  $F \in D_{\max}(\Phi_1)$ , em que

$$\Phi_1(x) = \begin{cases} 0, & x \leq 0 \\ \exp\{-x^{-1}\}, & x > 0 \end{cases},$$

$$a_n = \inf\{x : \bar{F}(x) \leq 1/n\} = \tan\left(\frac{\pi}{2} - \frac{\pi}{n}\right)$$

**Exemplo 1.6** (Distribuição Pareto).  $F(x) = 1 - \left(\frac{k}{k+x}\right)^\alpha$ ,  $x > 0$  ( $\alpha > 0$  e  $k > 0$ ).

Temos,  $\omega(F) = \infty$  e, como  $\lim_{t \rightarrow \infty} \frac{k+t}{k+tx} = x^{-1}$ , também

$$\lim_{t \rightarrow \infty} \frac{\bar{F}(tx)}{\bar{F}(t)} = \lim_{t \rightarrow \infty} \frac{\left(\frac{k}{k+tx}\right)^\alpha}{\left(\frac{k}{k+t}\right)^\alpha} = \lim_{t \rightarrow \infty} \left(\frac{k+t}{k+tx}\right)^\alpha = x^{-\alpha}$$

Assim,  $\bar{F} \in R_{-\alpha}$ . Pelo Teorema 1.2, temos que  $F \in D_{\max}(\Phi_\alpha)$ , em que  $\Phi_\alpha$  é dada por

$$\Phi_\alpha(x) = \begin{cases} 0, & x \leq 0 \\ \exp\{-x^{-\alpha}\}, & x > 0 \end{cases}.$$

### 1.3.2 Max-domínio de atração da Weibull ( $D_{\max}(\Psi_\alpha)$ )

**Teorema 1.3** (Max-domínio de atração da Weibull). *Seja  $F$  uma f.d. e  $\alpha > 0$ . Então,  $F \in D_{\max}(\Psi_\alpha)$  se, e somente se,*

$$\omega(F) < \infty \text{ e } \bar{F}^* \in R_{-\alpha}, \quad (1.23)$$

para alguma constante  $\alpha > 0$  e para  $\bar{F}^*(x) = \bar{F}(\omega(F) - 1/x)$ .

Nesse caso, tem-se que

$$\lim_{n \rightarrow \infty} P(Z_n \leq a_n x + b_n) = \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = \Psi_\alpha(x), \quad (1.24)$$

em que  $a_n$  e  $b_n$  podem ser escolhidas como

$$a_n = \omega(F) - \inf \{x : \bar{F}(x) \leq 1/n\} = \omega(F) - F^{\leftarrow}(1 - 1/n) \text{ e } b_n = \omega(F).$$

**Demonstração** (Suficiência). Suponha que  $\omega(F) < \infty$  e  $\bar{F}^* \in R_{-\alpha}$ , para alguma constante  $\alpha > 0$ . Note que

$$\omega(F^*) = \sup \{x : F^*(x) < 1\} = \sup \{x : F(\omega(F) - 1/x) < 1\} = \infty.$$

Como, por hipótese,  $\bar{F}^* \in R_{-\alpha}$ , segue do Teorema 1.2 que

$$\lim_{n \rightarrow \infty} F^{*n}(a_n^* x) = \lim_{n \rightarrow \infty} F^n\left(\omega(F) - \frac{1}{a_n^* x}\right) = \Phi_\alpha(x), \quad x > 0,$$

com

$$\begin{aligned} a_n^* &= \inf \left\{ x : 1 - F^*(x) \leq 1/n \right\} = \inf \left\{ x : \bar{F}(\omega(F) - 1/x) \leq 1/n \right\} \\ &= \inf \left\{ \frac{1}{\omega(F) - s} : \bar{F}(s) \leq 1/n \right\} = \frac{1}{\omega(F) - \inf \left\{ s : \bar{F}(s) \leq 1/n \right\}}. \end{aligned}$$

Fazendo  $y = -1/x$ , temos, para  $x > 0$ , que

$$\lim_{n \rightarrow \infty} F^n\left(\omega(F) - \frac{1}{a_n^* x}\right) = \lim_{n \rightarrow \infty} F^n\left(\omega(F) + \frac{y}{a_n^*}\right) = \Phi_\alpha(-1/y) = e^{-(-y)^\alpha}, \quad y < 0.$$

Então,

$$\lim_{n \rightarrow \infty} F^n(a_n y + b_n) = \Psi_\alpha(y), \quad y < 0,$$

com  $b_n = \omega(F)$  e  $a_n = 1/a_n^*$ .

Finalmente, pela definição de ponto extremo,  $F^n(a_n x + b_n) \equiv 1, \forall x > 0$ , assim segue a conclusão do teorema.  $\square$

(Necessidade) A prova será omitida. Para detalhes veja Galambos [4].

**Exemplo 1.7** (Distribuição Uniforme). Considere a função de distribuição

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

Temos que  $\omega(F) = 1 < \infty$ . Além disso,

$$\bar{F}^*(x) = \bar{F}(\omega(F) - 1/x) = \bar{F}(1 - 1/x) = 1 - (1 - 1/x) = 1/x, \quad x > 1.$$

Assim,

$$\lim_{t \rightarrow \infty} \frac{\bar{F}^*(tx)}{\bar{F}^*(t)} = \lim_{t \rightarrow \infty} \frac{1/tx}{1/t} = x^{-1}, \quad x > 0,$$

ou seja,  $\bar{F}^* \in R_{-1}$ . Dessa forma, temos para  $a_n = 1 - \inf \{x : \bar{F}(x) \leq 1/n\} = 1/n$  e  $b_n = 1$ ,

$$\lim_{n \rightarrow \infty} F^n \left( 1 + \frac{x}{n} \right) = \Psi_1(x),$$

em que

$$\Psi_1(x) = \begin{cases} e^x, & x \leq 0 \\ 1, & x > 0 \end{cases}.$$

**Exemplo 1.8** (Distribuição Beta). Considere a função densidade

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1}, \quad 0 \leq x \leq 1, \quad a, b > 0.$$

Temos que  $\omega(F) = 1 < \infty$  e

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{\bar{F}^*(tx)}{\bar{F}^*(t)} &= \lim_{t \rightarrow \infty} \frac{\bar{F}(1 - 1/tx)}{\bar{F}(1 - 1/t)} = \lim_{t \rightarrow \infty} \frac{\int_{1-1/tx}^1 f(y) dy}{\int_{1-1/t}^1 f(y) dy} = \\ &\stackrel{\text{L'Hôpital}}{=} \lim_{t \rightarrow \infty} \frac{-f(1 - \frac{1}{tx}) \left( \frac{1}{t^2 x} \right)}{-f(1 - \frac{1}{t}) \left( \frac{1}{t^2} \right)} = \lim_{t \rightarrow \infty} \frac{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left( 1 - \frac{1}{tx} \right)^{a-1} \left( \frac{1}{tx} \right)^{b-1} \frac{1}{x}}{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left( 1 - \frac{1}{t} \right)^{a-1} \left( \frac{1}{t} \right)^{b-1}} \end{aligned}$$



$$= \frac{1}{x^b} \lim_{t \rightarrow \infty} \frac{\left(1 - \frac{1}{tx}\right)^{a-1}}{\left(1 - \frac{1}{t}\right)^{a-1}} = x^{-b}, \quad x > 0.$$

Daí, temos que  $\bar{F}^* \in R_{-b}$  e, pelo Teorema 1.3,  $F \in D_{\max}(\Phi_b)$ , em que  $\Phi_b$  é dada por

$$\Psi_b(x) = \begin{cases} \exp\{-(-x)^b\}, & x \leq 0 \\ 1, & x > 0 \end{cases}, \quad b > 0.$$

### 1.3.3 Max-domínio de atração da Gumbel ( $D_{\max}(\Lambda)$ )

**Teorema 1.4** (Max-domínio de atração da Gumbel). *Seja  $F$  uma f.d. Então,*

$$\lim_{n \rightarrow \infty} P(Z_n \leq a_n x + b_n) = \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = \Lambda(x), \quad (1.25)$$

ou seja,  $F \in D_{\max}(\Lambda)$  se, e somente se, para alguma  $a \in \mathbb{R}$ ,

$$\int_a^{\omega(F)} \bar{F}(y) dy < \infty \quad \text{e} \quad \lim_{t \rightarrow \omega(F)} \frac{\bar{F}(t + xR(t))}{\bar{F}(t)} = e^{-x}, \quad \forall x \in \mathbb{R}, \quad (1.26)$$

em que, para  $\gamma(F) < t < \omega(F)$ ,

$$R(t) = \bar{F}(t)^{-1} \int_t^{\omega(F)} \bar{F}(y) dy. \quad (1.27)$$

Neste caso,  $a_n$  e  $b_n$  podem ser escolhidas como

$$b_n = \inf\{x : \bar{F}(x) \leq 1/n\} \quad \text{e} \quad a_n = R(b_n).$$

**Demonstração** (Suficiência). Temos de forma direta que

$$b_n = \inf\{x : F(x) \geq 1 - 1/n\} \xrightarrow{n} \omega(F) = \sup\{x : F(x) < 1\}$$

De (1.26), temos que

$$\lim_{n \rightarrow \infty} \frac{\bar{F}(R(b_n)x + b_n)}{\bar{F}(b_n)} = \lim_{t \rightarrow \omega(F)} \frac{\bar{F}(R(t)x + t)}{\bar{F}(t)} = e^{-x}. \quad (1.28)$$

Agora, considerando  $a_n = R(b_n)$ , segue

$$\lim_{n \rightarrow \infty} n\bar{F}(a_n x + b_n) = \lim_{n \rightarrow \infty} n\bar{F}(b_n) \frac{\bar{F}(a_n x + b_n)}{\bar{F}(b_n)} = e^{-x} \lim_{n \rightarrow \infty} n\bar{F}(b_n).$$

Portanto, resta mostrar que

$$\lim_{n \rightarrow \infty} n\bar{F}(b_n) = 1.$$

Seja  $\varepsilon > 0$  arbitrário, então  $b_n - \varepsilon a_n < b_n$  e  $\bar{F}(b_n - \varepsilon a_n) > 1/n$ . Assim,

$$\bar{F}(b_n) \leq 1/n < \bar{F}(b_n - \varepsilon a_n),$$

e, portanto,

$$\frac{\bar{F}(b_n)}{\bar{F}(b_n - \varepsilon a_n)} < n\bar{F}(b_n) \leq 1.$$

De 1.28, segue que

$$\frac{1}{e^{-\varepsilon}} \leq \liminf n\bar{F}(b_n) \leq \limsup n\bar{F}(b_n) \leq 1.$$

Fazendo  $\varepsilon \rightarrow 0$ , obtemos  $\lim_{n \rightarrow \infty} n\bar{F}(b_n) = 1$ , e

$$\lim_{n \rightarrow \infty} n\bar{F}(a_n x + b_n) = e^{-x}$$

Segue então, pelo Corolário 1.1, que  $F \in D_{\max}(\Lambda)$ . □

(Necessidade) A prova será omitida. Para detalhes veja Galambos [4].

**Exemplo 1.9** (Distribuição Exponencial).

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \ (\lambda > 0) \\ 0, & x < 0 \end{cases}.$$

Temos que  $\omega(F) = \infty$  e

$$\int_a^{\omega(F)} \bar{F}(y) dy = \int_a^{\infty} e^{-\lambda y} dy = -\frac{1}{\lambda} e^{-\lambda y} \Big|_a^{\infty} = \frac{e^{-\lambda a}}{\lambda} < +\infty, \forall a \in \mathbb{R}.$$

Agora,

$$R(t) = \bar{F}(t)^{-1} \int_t^{\omega(F)} \bar{F}(y) dy = e^{\lambda t} \int_t^{\infty} e^{-\lambda y} dy = e^{\lambda t} \frac{e^{-\lambda t}}{\lambda} = \lambda^{-1},$$

e

$$\lim_{t \rightarrow \omega(F)} \frac{\bar{F}(t + xR(t))}{\bar{F}(t)} = \lim_{t \rightarrow \infty} \frac{e^{-\lambda(t+x\lambda^{-1})}}{e^{-\lambda t}} = e^{-x} \lim_{t \rightarrow \infty} \frac{e^{-\lambda t}}{e^{-\lambda t}} = e^{-x}.$$

Assim, pelo Teorema 1.4, temos  $F \in D_{max}(\Lambda)$ , em que  $\Lambda$  é dada por

$$\Lambda(x) = \exp\{-e^{-x}\} \quad x \in \mathbb{R}.$$

## CAPÍTULO

### 2

# DISTRIBUIÇÃO PARETO GENERALIZADA

## 2.1 Introdução

A distribuição de valor extremo generalizada (GEV – *Generalized extreme value distribution*), desenvolvida na teoria do valor extremo, combina as três distribuições estudadas no Capítulo 1: Gumbel, Fréchet e Weibull. A principal característica da GEV é ter uma representação uniparamétrica, com a introdução do parâmetro  $\xi$ , cuja motivação é simplificar as implementações estatísticas.

Quando se trabalha com as três distribuições de valor extremo separadamente, escolhe-se previamente a distribuição que modela os dados. Ocorre que, quando os parâmetros do modelo escolhido são desconhecidos, torna-se necessário estimá-los a partir da distribuição selecionada. Eventuais erros de escolha de modelo não são mais contestados. Uma das vantagens de se trabalhar com a GEV é que o tipo de distribuição também pode ser escolhido a partir dos dados. Dessa forma, a confiança na determinação do modelo pode ser avaliada pela estimativa do próprio parâmetro  $\xi$ .

A GEV é amplamente utilizada para o estudo de caudas das distribuições aplicadas nas áreas de seguro e de finanças. Neste caso, ela tem sido considerada um modelo para avaliar várias métricas financeiras, como o Value at Risk (VaR). Além disso, essa distribuição é utilizada em eventos extremos aplicados à hidrologia, para valores máximos de chuvas e de cheias em barragens ou rios.

Outra distribuição que tem grande influência na teoria de valor extremo é a distribuição Pareto generalizada (GPD – *Generalized extreme value distribution*). A GPD foi introduzida por Pickands [10] para modelar excendentes acima de um limiar. Ela aparece como limite de uma das caracterizações do max-domínio de atração (Teorema 2.1).

Assim, na Seção 2.2, será mostrado como o parâmetro  $\xi$  é utilizado para unir as três distribuições extremais em uma única, a GEV. Demonstraremos o Teorema 2.1, que mostra uma caracterização para o max-domínio da GEV e introduz a distribuição Pareto generalizada. As propriedades da GPD serão vistas na Seção 2.3, por intermédio do Teorema 2.2. O resultado de Pickands [10] será demonstrado na propriedade (b) do mesmo teorema.

## 2.2 Distribuição de valor extremo generalizada – GEV (*Generalized Extreme Value Distribution*)

Nesta seção, apresentamos a distribuição de valor extremo generalizada (GEV). Para isso, introduzimos uma família paramétrica  $(H_\xi)_{\xi \in \mathbb{R}}$  de distribuições que contém as distribuições de valor extremo padrão, dada por

$$H_\xi(x) = \begin{cases} \Phi_{1/\xi}(1 + \xi x) & \xi > 0, \\ \Lambda(x) & \xi = 0, \\ \Psi_{-1/\xi}(-1 - \xi x) & \xi < 0. \end{cases}$$

Dessa maneira, uma representação uniparamétrica dos três tipos de distribuições extremais pode ser dada por meio de uma única família de distribuições, introduzindo um parâmetro  $\xi$ , tal como

$$\begin{aligned} \xi = \alpha^{-1} > 0, & \quad \text{distribuição de Fréchet } \Phi_\alpha, \\ \xi = 0, & \quad \text{distribuição de Gumbel } \Lambda, \\ \xi = -\alpha^{-1} < 0, & \quad \text{distribuição de Weibull } \Psi_\alpha. \end{aligned}$$

**Definição 2.1** (Representação de Jenkinson–von Mises das distribuições de valores extremos – a distribuição de valor extremo generalizada (GEV)). *Seja  $H_\xi$  a função de distribuição*

dada por

$$H_{\xi}(x) = \begin{cases} \exp\{- (1 + \xi x)^{-1/\xi}\}, & \text{se } \xi \neq 0, \\ \exp\{- \exp\{-x\}\}, & \text{se } \xi = 0. \end{cases} \quad (2.1)$$

em que  $1 + \xi x > 0$ , então  $H_{\xi}$  é chamada a **distribuição de valor extremo generalizada (GEV)**.

**Observação 2.1.**

a)  $\xi = 0$  corresponde à interpretação  $\xi \rightarrow 0$ .

b) O suporte de  $H_{\xi}$  corresponde a

$$\begin{aligned} x &> -\xi^{-1}, \text{ para } \xi > 0, \\ x &< -\xi^{-1}, \text{ para } \xi < 0, \\ x &\in \mathbb{R} \quad , \text{ para } \xi = 0. \end{aligned}$$

c) A f.d.  $H_{\xi}$  é chamada GEV padrão.

d) Pode ser introduzida a família (locação–escala)  $H_{\xi,\mu,\beta}$ , substituindo  $x$  por  $\frac{x-\mu}{\beta}$ , em que  $\mu \in \mathbb{R}$  e  $\beta > 0$  são os parâmetros de locação e escala, respectivamente. Referimos a  $H_{\xi,\mu,\beta}$  como a f.d. GEV.

**Teorema 2.1** (Caracterização do  $D_{\max}(H_{\xi})$ ). *Seja  $H_{\xi}$  uma GEV, com  $\xi \in \mathbb{R}$ , as seguintes afirmações são equivalentes:*

(a)  $F \in D_{\max}(H_{\xi})$ .

(b) Existe uma função positiva e mensurável,  $a(\cdot)$ , tal que, para  $1 + \xi x > 0$ ,

$$\lim_{t \uparrow \omega(F)} \frac{\bar{F}(t + xa(t))}{\bar{F}(t)} = \begin{cases} (1 + \xi x)^{-1/\xi} & \text{se } \xi \neq 0 \\ e^{-x} & \text{se } \xi = 0 \end{cases}. \quad (2.2)$$

(c) Para  $x, y > 0, y \neq 1$ ,

$$\lim_{s \rightarrow \infty} \frac{U(sx) - U(s)}{U(sy) - U(s)} = \begin{cases} \frac{x^{\xi} - 1}{y^{\xi} - 1} & \text{se } \xi \neq 0 \\ \frac{\ln x}{\ln y} & \text{se } \xi = 0 \end{cases}, \quad (2.3)$$

em que  $U(\cdot)$  é a função quantil da cauda, dada por  $U(s) = \inf\{x : F(x) \geq 1 - 1/s\}$ .

**Demonstração.** A seguir, damos as ideias principais para a demonstração de (a)  $\Leftrightarrow$  (b).

1. Caso  $\xi = 0$ . Como  $F \in D_{\max}(\Lambda(x))$ , considerando  $a(t) = R(t)$ , em que

$$R(t) = \frac{1}{\bar{F}(t)} \int_t^{\omega(F)} \bar{F}(y) dy, \quad \gamma(F) < t < \omega(F),$$

temos que  $a(\cdot)$  é função positiva, mensurável e pelo limite em (1.26) segue o resultado.

2. Caso  $\xi > 0$ . Temos que  $H_\xi(x) = \Phi_\alpha(1 + x/\alpha) = \exp\{-(1 + x/\alpha)^{-\alpha}\}$ , para  $\alpha = 1/\xi > 0$ , ou seja,

$$H_\xi(x) = \exp\{-(1 + \xi x)^{-1/\xi}\}, \quad 1 + \xi x > 0.$$

Pelo Teorema 1.2, temos que  $F \in D_{\max}(H_\xi) \iff \bar{F} \in R_{-\alpha}$  e  $\omega(F) = \infty$ , ou seja,

$$\lim_{t \rightarrow \infty} \frac{\bar{F}(t(1 + \xi x))}{\bar{F}(t)} = (1 + \xi x)^{-1/\xi}, \quad 1 + \xi x > 0.$$

Fazendo  $a(t) = \xi t$ , para  $1 + \xi x > 0$ , temos que

$$\lim_{t \uparrow \omega(F) = \infty} \frac{\bar{F}(t + xa(t))}{\bar{F}(t)} = (1 + \xi x)^{-1/\xi}.$$

3. Caso  $\xi < 0$ . Para  $\bar{F}^*(x) = \bar{F}(\omega(F) - 1/x)$ , vimos na demonstração do Teorema 1.3 que  $F(x) \in D_{\max}(\Psi_\alpha(-1 - \xi x)) \iff \bar{F}^* \in D_{\max}(\Phi_\alpha(1 + \xi x))$ . Dessa forma, segue do caso 2 que

$$\lim_{t \uparrow \omega(F^*) = \infty} \frac{\bar{F}^*(t + xa(t))}{\bar{F}^*(t)} = (1 + \xi x)^{-1/\xi}.$$

Dos casos 1,2 e 3 segue a equivalência (a)  $\iff$  (b). Para a ideia da demonstração de (b)  $\iff$  (c) veja Embrechts [5].  $\square$

Em seguida, será vista a aplicação da caracterização do  $D_{\max}(H_\xi)$  a algumas distribuições usuais.

**Exemplo 2.1.** Vamos verificar, por meio do Teorema 2.1, o max-domínio de atração da Cauchy padrão, que já vimos pertencer ao max-domínio de atração da Fréchet. Assim, temos

$$\lim_{u \uparrow \omega(F)} \frac{\bar{F}(u + xa(u))}{\bar{F}(u)} = \lim_{u \rightarrow \infty} \frac{\frac{1}{2} - \frac{1}{\pi} \arctan(u + xa(u))}{\frac{1}{2} - \frac{1}{\pi} \arctan(u)}.$$

Por L'Hôpital,

$$\lim_{u \rightarrow \infty} \frac{\frac{1+xa'(u)}{1+(u+xa(u))^2}}{\frac{1}{1+u^2}} = \lim_{u \rightarrow \infty} \frac{(1+xa'(u))(1+u^2)}{1+(u+xa(u))^2}.$$

Tomando a função  $a(u) = u$  e usando L'Hôpital novamente, temos

$$\lim_{u \rightarrow \infty} \frac{(1+x)(1+u^2)}{1+u^2(1+x)^2} \Rightarrow \lim_{u \rightarrow \infty} \frac{(1+x)2u}{2u(1+x)^2} = (1+x)^{-1}.$$

Portanto, a Cauchy padrão está no max-domínio de atração da  $H_\xi$ , com  $\xi = 1$ , conforme o esperado.

**Exemplo 2.2.** Vamos considerar agora a Uniforme em  $[0, 1]$  ( $U[0, 1]$ ), que está no max-domínio de atração da Weibull. Utilizando o Teorema 2.1, temos

$$\lim_{u \uparrow \omega(F)} \frac{\overline{F}(u+xa(u))}{\overline{F}(u)} = \lim_{u \uparrow 1} \frac{u+xa(u)}{u}.$$

Tomando a função  $a(u) = -u$ , temos

$$\lim_{u \uparrow 1} \frac{u-xu}{u} = \lim_{u \uparrow 1} 1-x = 1-x.$$

Portanto, a  $U[0, 1]$  está no max-domínio de atração da  $H_\xi$ , com  $\xi = -1$ , conforme o esperado.

**Exemplo 2.3.** Vamos determinar a qual max-domínio de atração a Normal padrão ( $N(0, 1)$ ) pertence. Utilizando o Teorema 2.1,

$$\lim_{u \uparrow \omega(F)} \frac{\overline{F}(u+xa(u))}{\overline{F}(u)} = \lim_{u \rightarrow \infty} \frac{\int_{u+xa(u)}^{\infty} e^{-\frac{t^2}{2}} dt}{\int_u^{\infty} e^{-\frac{t^2}{2}} dt}.$$

Por L'Hôpital,

$$\lim_{u \rightarrow \infty} \frac{e^{-\frac{(u+xa(u))^2}{2}}}{e^{-\frac{u^2}{2}}} = \lim_{u \rightarrow \infty} e^{-\frac{1}{2}[u^2+(xa)^2+2uxa-u^2]} = \lim_{u \rightarrow \infty} e^{-\frac{1}{2}[(xa)^2+2uxa]}.$$

Tomando a função  $a(u) = 1/u$ , temos que

$$\lim_{u \rightarrow \infty} e^{-\frac{1}{2}[(x/u)^2+2ux/u]} = \lim_{u \rightarrow \infty} e^{-\frac{1}{2}[(x/u)^2+2x]} = e^{-x}.$$

Portanto, a  $N(0, 1)$  está no max-domínio de atração de  $H_\xi$ , com  $\xi = 0$ .



**Observação 2.2.** Interpretação do item (b) do Teorema 2.1.

Seja  $X$  uma variável aleatória com f.d.  $F \in D_{\max}(H_\xi)$ . Então, (2.2) pode ser reescrita como

$$\lim_{t \uparrow \omega(F)} P\left(\frac{X-t}{a(t)} > x \mid X > t\right) = \begin{cases} (1 + \xi x)^{-1/\xi}, & \text{se } \xi \neq 0 \\ e^{-x}, & \text{se } \xi = 0 \end{cases}. \quad (2.4)$$

Dessa forma, (2.4) fornece uma distribuição aproximada para excessos acima de um limiar (alto)  $t$ .

**Definição 2.2** (Função de distribuição do excesso e função média do excesso).

Seja  $X$  uma variável aleatória com f.d.  $F$ . Para  $t < \omega(F)$  fixo,

$$F_t(x) = P(X - t \leq x \mid X > t), \quad x \geq 0, \quad (2.5)$$

é chamada **função de distribuição do excesso** da variável  $X$  acima do limiar  $t$ . A função definida por

$$e(t) = E(X - t \mid X > t) \quad (2.6)$$

é chamada **função média do excesso** de  $X$

Excessos acima de limiares desempenham um papel fundamental em diversas áreas e têm diferentes nomes de acordo com aplicações específicas. Podemos citar

- excesso de vida, em teoria da confiabilidade,
- tempo de vida residual, em análise de sobrevivência,
- excesso de perda, no contexto de seguros.

Podemos destacar que, se  $X$  é uma v.a. positiva com f.d.  $F$  e com esperança finita, então a função média do excesso pode ser calculada por

$$\begin{aligned} e(t) &= \int_t^{\omega(F)} \frac{(x-t)}{\bar{F}(t)} dF(x) = \frac{1}{\bar{F}(t)} \left[ (x-t)F(x) \Big|_t^{\omega(F)} - \int_t^{\omega(F)} F(x) dx \right] \\ &= \frac{1}{\bar{F}(t)} \left[ (\omega(F) - t)F(\omega(F)) - \int_t^{\omega(F)} F(x) dx \right] = \frac{1}{\bar{F}(t)} \left[ \int_t^{\omega(F)} dx - \int_t^{\omega(F)} F(x) dx \right] \\ &= \frac{1}{\bar{F}(t)} \int_t^{\omega(F)} \bar{F}(x) dx, \quad 0 < t < \omega(F). \end{aligned} \quad (2.7)$$

## 2.3 Distribuição Pareto generalizada – GPD (*Generalized Pareto Distribution*)

Há várias maneiras de justicar o uso da distribuição Pareto generalizada. Pickands [10], por exemplo, mostrou que ela é uma boa aproximação para a função de distribuição do excesso, como poderá ser visto pela propriedade (b) do Teorema 2.2. Uma outra razão é a estabilidade do limiar, descrita da seguinte forma: se  $Y$  é uma variável aleatória com distribuição Pareto generalizada, então a distribuição de  $Y$  acima de um limiar  $t > 0$  também segue uma Pareto generalizada (veja propriedade (c) do Teorema 2.2).

Outra propriedade de destaque é que se  $N$  tem uma distribuição de Poisson independente de  $Y_1, \dots, Y_n$ , variáveis aleatórias i.i.d. com distribuição comum Pareto generalizada, então o  $\max\{Y_1, \dots, Y_N\}$  segue uma distribuição de valor extremo generalizada.

Nesta seção, apresentaremos a distribuição Pareto generalizada (GPD), bem como suas propriedades mais importantes para efeito do que segue neste trabalho.

**Definição 2.3** (Distribuição Pareto generalizada (GPD)). *Defina a  $G_\xi$  por*

$$G_\xi(x) = \begin{cases} 1 - (1 + \xi x)^{-1/\xi}, & \text{se } \xi \neq 0 \\ 1 - e^{-x}, & \text{se } \xi = 0 \end{cases}, \quad (2.8)$$

em que

$$\begin{aligned} & x \geq 0, \text{ se } \xi \geq 0 \\ & 0 \leq x \leq -1/\xi, \text{ se } \xi < 0 \end{aligned}$$

é chamada **distribuição Pareto generalizada (GPD)**.

Podemos introduzir uma família de locação-escala,  $G_{\xi, \mu, \beta}$ , trocando o argumento  $x$  acima por  $(x - \mu)/\beta$  para  $\mu \in \mathbb{R}$  e  $\beta > 0$ . O suporte da função deve ser ajustado convenientemente.

A f.d.  $G_0$  pode ser interpretada como o limite de  $G_\xi$  com  $\xi \rightarrow 0$ . Além disso, temos interesse especial na GPD com parâmetros  $\kappa$  e  $\beta$ , ou seja, com  $\mu = 0$ . Assim, denotamos:

$$G_{\xi, \beta}(x) = 1 - \left(1 + \xi \frac{x}{\beta}\right)^{-1/\xi}, \quad x \in D(\xi, \beta), \quad (2.9)$$

em que

$$D(\xi, \beta) = \begin{cases} [0, \infty), & \text{se } \xi \geq 0 \\ [0, -\beta/\xi], & \text{se } \xi < 0 \end{cases}, \quad \beta > 0.$$

**Observação 2.3.**

- (1) Observe que, para  $\xi = 0$ , a GPD é uma distribuição exponencial com média  $\beta$ . Para  $\xi = -1$ , a GPD é a distribuição uniforme em  $[0, \beta]$ .
- (2) A GEV descreve as distribuições limites dos máximos normalizados. Por outro lado, a GPD aparece como a distribuição limite dos excessos sobre altos limiares. O ajuste pela GPD é um dos conceitos de maior utilidade para estatística de eventos extremos. A seguir veremos algumas propriedades importantes da GPD.

**Teorema 2.2** (Propriedades da GPD).

(a) Seja  $X$  uma v.a. com f.d.  $G_{\xi, \beta}$ . Então  $EX < \infty$  se, e somente se,  $\xi < 1$ . Neste caso,

$$E \left[ \left( 1 + \frac{\xi}{\beta} X \right)^{-r} \right] = \frac{1}{1 + \xi r}, \quad r > -1/\xi, \quad (2.10)$$

$$E \left[ \left( \ln \left( 1 + \frac{\xi}{\beta} X \right) \right)^k \right] = \xi^k k!, \quad k \in \mathbb{N}, \quad (2.11)$$

$$E \left[ X (\bar{G}_{\xi, \beta}(X))^r \right] = \frac{\beta}{(r+1-\xi)(r+1)}, \quad (r+1)/|\xi| > 0, \quad (2.12)$$

em que  $\bar{G}_{\xi, \beta} \equiv 1 - G_{\xi, \beta}$  e  $r \in \mathbb{N}$ .

Se  $0 < \xi < 1/r$ , com  $r \in \mathbb{N}$ , então,

$$E[X^r] = \frac{\beta^r}{\xi^{r+1}} \frac{\Gamma(\xi^{-1} - r)}{\Gamma(1 + \xi^{-1})} r!. \quad (2.13)$$

(b) Para todo  $\xi \in \mathbb{R}$ ,  $F \in D_{\max}(H_\xi)$  se, e somente se,

$$\lim_{t \uparrow \omega(F)} \sup_{0 < x < \omega(F) - t} |F_t(x) - G_{\xi, \beta(t)}(x)| = 0, \quad (2.14)$$

para alguma função positiva  $\beta(t)$ .

(c) Suponha  $x_i \in D(\xi, \beta)$ ,  $i = 1, 2$ , então,

$$\frac{\bar{G}_{\xi, \beta}(x_1 + x_2)}{\bar{G}_{\xi, \beta}(x_1)} = \bar{G}_{\xi, \beta + \xi x_1}(x_2). \quad (2.15)$$

(d) Seja  $N$  uma v.a. com distribuição Poisson( $\lambda$ ), independente da sequência de v.a.'s i.i.d. ( $X_n$ ) com f.d. comum  $G_{\xi,\beta}$ . Considere a v.a.  $Z_N = \max(X_1, \dots, X_N)$ . Então,

$$P(Z_N \leq x) = \exp \left\{ -\lambda \left( 1 + \xi \frac{x}{\beta} \right)^{-1/\xi} \right\} = H_{\xi,\mu,\psi}(x), \quad (2.16)$$

em que  $\mu = \beta \xi^{-1}(\lambda^\xi - 1)$  e  $\psi = \beta \lambda^\xi$ .

(e) Suponha que  $X$  seja uma v.a. com f.d.  $G_{\xi,\beta}$ , de modo que  $\xi < 1$  e  $\beta > 0$ . Então para  $t < \omega(F)$ ,

$$e(t) = E(X - t | X > t) = \frac{\beta + \xi t}{1 - \xi}, \quad \beta + t\xi > 0. \quad (2.17)$$

#### Observação 2.4.

- (1) A propriedade (c) pode ser reformulada da seguinte forma: a classe de GPDs é fechada com respeito à mudança de limiar. De fato, o lado esquerdo da equação em (c) é a probabilidade condicional de, dado que a variável excede  $x_1$ , então ela também excede o limiar  $x_1 + x_2$ . O lado direito diz que essa probabilidade é também do tipo GPD.
- (2) A propriedade (b) sugere que a GPD é uma aproximação apropriada para a função de distribuição do excesso ( $F_t$ ) para um limiar (alto)  $t$ . Esse resultado foi apresentado em Pickands [10] e é formulado da seguinte maneira: para alguma função  $\beta(t)$  a ser estimada dos dados observados e para  $t$  suficientemente grande,

$$\bar{F}_t(x) = P(X - t > x | X > t) \approx \bar{G}_{\xi,0,\beta(t)}(x), \quad x > 0,$$

ou então, considerando  $x > t$ ,

$$P(X > x | X > t) \approx \bar{G}_{\xi,t,\beta(t)}(x).$$

- (3) As propriedades (b) e (e) fornecem um método gráfico para escolher o limiar  $t$  suficientemente alto tal que uma aproximação para a função de distribuição do excesso  $F_t$  por uma GPD é adequada. Isso pode ser justificado da seguinte forma: dada uma amostra aleatória  $X_1, \dots, X_n$ , constrói-se, primeiramente, função média do excesso empírica,  $e_n(t)$ , como uma versão amostral da função média do excesso  $e(t)$ . De (e), temos que  $e(t)$  de uma GPD é linear. Assim, observando o gráfico de  $e_n(t)$ , verifica-se para quais valores de  $t$  o gráfico de  $e_n(t)$  é aproximadamente linear. Para tais valores de  $t$ , uma aproximação de  $F_t$  pela GPD é razoável.

(4) A propriedade (d) diz que, para um modelo no qual o número de excedentes é distribuído segundo uma Poisson e a f.d. do excesso é uma GPD, então o máximo desses excessos segue uma GEV.

(5) Aplicações dos resultados (b), (d) e (e) serão vistas no Capítulo 4.

### Demonstração do Teorema 2.2

**Propriedade (a).** De (2.9), temos, para  $x \in D(\xi, \beta)$ ,  $\xi \in \mathbb{R}$  e  $\beta > 0$ ,

$$g_{\xi, \beta}(x) = G'_{\xi, \beta}(x) = \frac{1}{\xi} \left(1 + \frac{\xi}{\beta}x\right)^{-\frac{1}{\xi}-1} \left(\frac{\xi}{\beta}\right) = \frac{1}{\beta} \left(1 + \frac{\xi}{\beta}x\right)^{-\frac{1}{\xi}-1}.$$

•  $\xi > 0$ . Neste caso,  $x \in [0, \infty)$  e

$$EX = \int_0^{\infty} \frac{x}{\beta} \left(1 + \frac{\xi}{\beta}x\right)^{-\frac{1}{\xi}-1} dx.$$

Tomando  $u = 1 + (\xi/\beta)x$ , temos que

$$EX = \int_1^{\infty} \frac{\beta}{\xi^2} (u-1) u^{-\frac{1}{\xi}-1} du = \frac{\beta}{\xi^2} \left( \frac{u^{-\frac{1}{\xi}+1}}{-\frac{1}{\xi}+1} - \frac{u^{-1/\xi}}{-1/\xi} \right) \Bigg|_1^{\infty}.$$

O limite acima será finito se, e somente se,  $1/\xi > 0$ , e  $-1/\xi + 1 < 0$ , ou seja,  $\xi < 1$ . Assim, para  $0 < \xi < 1$ ,

$$EX = \frac{\beta}{\xi^2} \left( \frac{\xi}{1-\xi} - \xi \right) = \frac{\beta}{1-\xi} < \infty.$$

•  $\xi = 0$ . Neste caso, a GPD é a distribuição exponencial com esperança finita e igual  $\beta$ .

•  $\xi < 0$ . Neste caso,  $x \in [0, -\beta/\xi]$  e

$$EX = \int_0^{-\beta/\xi} \frac{x}{\beta} \left(1 + \frac{\xi}{\beta}x\right)^{-\frac{1}{\xi}-1} dx.$$

Tomando  $u$  da mesma forma que no 1º caso, temos

$$\begin{aligned} EX &= - \int_0^1 (u-1) \frac{1}{\xi} u^{-\frac{1}{\xi}-1} \frac{\beta}{\xi} du = - \frac{\beta}{\xi^2} \int_0^1 \left( u^{-\frac{1}{\xi}} - u^{-\frac{1}{\xi}-1} \right) du = - \frac{\beta}{\xi^2} \left( \frac{1}{1-1/\xi} - \frac{1}{-1/\xi} \right) \\ &= - \frac{\beta}{\xi^2} \left( \frac{1}{1-1/\xi} + \xi \right) = \frac{\beta}{1-\xi} < \infty. \end{aligned}$$

Dos três casos, segue que  $EX < \infty$  se, e somente se,  $\xi < 1$ , além disso, sua expressão é dada por

$$EX = \frac{\beta}{1 - \xi}. \quad (2.18)$$

Assim sendo, no que segue da demonstração, consideraremos  $\xi < 1$ .

(A) Verificação de (2.10).

- $0 < \xi < 1$ . Para  $r > -1/\xi$ , temos que  $x \in [0, \infty)$  e

$$\begin{aligned} E \left[ \left( 1 + \frac{\xi}{\beta} X \right)^{-r} \right] &= \int_0^{\infty} \left( 1 + \frac{\xi}{\beta} x \right)^{-r} \frac{1}{\beta} \left( 1 + \frac{\xi}{\beta} x \right)^{-\frac{1}{\xi}-1} dx = \int_1^{\infty} \frac{1}{\xi} u^{-\frac{1}{\xi}-1-r} du \\ &= \frac{1}{\xi} \frac{u^{-\frac{1}{\xi}-r}}{-\frac{1}{\xi}-r} \Big|_1^{\infty} = \frac{1}{1 + \xi r}. \end{aligned}$$

- $\xi = 0$ . O resultado segue trivialmente.

- $\xi < 0$ . Neste caso, temos que  $x \in [0, -\beta/\xi]$  e

$$\begin{aligned} E \left[ \left( 1 + \frac{\xi}{\beta} X \right)^{-r} \right] &= \int_0^{-\beta/\xi} \left( 1 + \frac{\xi}{\beta} x \right)^{-r} \frac{1}{\beta} \left( 1 + \frac{\xi}{\beta} x \right)^{-\frac{1}{\xi}-1} dx = \frac{1}{\xi} \int_1^0 u^{-\frac{1}{\xi}-1-r} du \\ &= \frac{1}{\xi} \frac{u^{-\frac{1}{\xi}-r}}{-\frac{1}{\xi}-r} \Big|_1^0 = \frac{1}{1 + \xi r}. \end{aligned}$$

(B) Verificação de (2.11).

- $0 < \xi < 1$ . Para  $k \in \mathbb{N}$ , temos que

$$E \left[ \left( \ln \left( 1 + \frac{\xi}{\beta} X \right) \right)^k \right] = \int_0^{\infty} \left( \ln \left( 1 + \frac{\xi}{\beta} x \right) \right)^k \frac{1}{\beta} \left( 1 + \frac{\xi}{\beta} x \right)^{-\frac{1}{\xi}-1} dx.$$

Considerando as mudanças de variáveis, fazendo  $u = \ln \left( 1 + \frac{\xi}{\beta} x \right)$  e depois  $z = u/\xi$ , temos que

$$\begin{aligned} E \left[ \left( \ln \left( 1 + \frac{\xi}{\beta} X \right) \right)^k \right] &= \frac{1}{\xi} \int_0^{\infty} u^k e^{-u/\xi} du = \xi^k \int_0^{\infty} z^k e^{-z} dz \\ &= \xi^k \Gamma(k+1) = \xi^k k! \end{aligned}$$

- $\xi = 0$ . O resultado segue trivialmente.
- $\xi < 0$ .

$$E \left[ \left( \ln \left( 1 + \frac{\xi}{\beta} X \right) \right)^k \right] = \int_0^{-\beta/\xi} \left( \ln \left( 1 + \frac{\xi}{\beta} X \right) \right)^k \frac{1}{\beta} \left( 1 + \frac{\xi}{\beta} X \right)^{-\frac{1}{\xi}-1} dx.$$

Fazendo  $u = \ln \left( 1 + \frac{\xi}{\beta} x \right)$  e depois  $z = u/\xi$  na integral acima, obtemos

$$E \left[ \left( \ln \left( 1 + \frac{\xi}{\beta} X \right) \right)^k \right] = \xi^k \int_0^{\infty} z^k e^{-z} dz = \xi^k k!$$

(C) Verificação de (2.12). Observe que  $(r+1)/|\xi| > 0$ , logo  $\xi \neq 0$ .

- $0 < \xi < 1$ . Para  $(r+1)/|\xi| > 0$ , temos que

$$E \left[ X (\overline{G}_{\xi, \beta}(X))^r \right] = \int_0^{\infty} x \left( 1 + \frac{\xi x}{\beta} \right)^{-\frac{r}{\xi}} \frac{1}{\beta} \left( 1 + \frac{\xi x}{\beta} \right)^{-\frac{1}{\xi}-1} dx = \int_0^{\infty} \frac{1}{\beta} x \left( 1 + \frac{\xi x}{\beta} \right)^{\frac{1+r}{-\xi}-1} dx.$$

Fazendo  $u = \left( 1 + \frac{\xi}{\beta} x \right)$  na integral acima, obtemos

$$\begin{aligned} E \left[ X (\overline{G}_{\xi, \beta}(X))^r \right] &= \int_1^{\infty} \frac{1}{\beta} (u-1) \frac{\beta}{\xi} u^{\frac{1+r}{-\xi}-1} \frac{\beta}{\xi} du = \int_1^{\infty} \frac{\beta}{\xi^2} (u-1) u^{\frac{1+r}{-\xi}-1} du \\ &= -\frac{\beta}{\xi^2} \left[ \frac{u^{\frac{1+r}{-\xi}}}{-\frac{1+r}{-\xi}} - \frac{u^{\frac{1+r}{-\xi}+1}}{-\frac{1+r}{-\xi}+1} \right] \Big|_1^{\infty} = -\frac{\beta}{\xi^2} \left[ \frac{\xi}{1+r} - \frac{\xi}{1+r-\xi} \right] \\ &= \frac{\beta}{(r+1-\xi)(r+1)}. \end{aligned}$$

- $\xi < 0$ .

$$\begin{aligned} E \left[ X (\overline{G}_{\xi, \beta}(X))^r \right] &= \int_0^{-\beta/\xi} x \left( 1 + \frac{\xi x}{\beta} \right)^{-\frac{r}{\xi}} \frac{1}{\beta} \left( 1 + \frac{\xi x}{\beta} \right)^{-\frac{1}{\xi}-1} dx \\ &= \int_0^{-\beta/\xi} \frac{1}{\beta} x \left( 1 + \frac{\xi x}{\beta} \right)^{\frac{1+r}{-\xi}-1} dx. \end{aligned}$$

Fazendo  $u = \left(1 + \frac{\xi}{\beta}x\right)$  na integral acima, obtemos

$$\begin{aligned} E\left[X(\overline{G}_{\xi,\beta}(X))^r\right] &= \int_1^0 \frac{1}{\beta}(u-1)\frac{\beta}{\xi}u^{\frac{1+r}{-\xi}-1}\frac{\beta}{\xi}du \\ &= -\int_0^1 \frac{\beta}{\xi^2}(u-1)u^{\frac{1+r}{-\xi}-1}du = \frac{\beta}{\xi^2}\left[\frac{u^{\frac{1+r}{-\xi}}}{\frac{1+r}{-\xi}} - \frac{u^{\frac{1+r}{-\xi}+1}}{\frac{1+r}{-\xi}+1}\right]\Bigg|_0^1 \\ &= \frac{\beta}{\xi^2}\left[\frac{\xi}{1+r} - \frac{\xi}{1+r-\xi}\right] = \frac{\beta}{(r+1-\xi)(r+1)}. \end{aligned}$$

(D) Verificação de (2.13). Como  $0 < \xi < 1/r$  e  $r \in \mathbb{N}$ , então  $0 < \xi < 1$ , ou seja,  $x \in [0, \infty)$ . Calculando  $E[X^r]$  pela definição, temos

$$E[X^r] = \int_0^\infty \frac{x^r}{\beta}\left(1 + \frac{\xi}{\beta}x\right)^{-\frac{1}{\xi}-1}dx = \lim_{a \rightarrow \infty} \int_0^a \frac{x^r}{\beta}\left(1 + \frac{\xi}{\beta}x\right)^{-\frac{1}{\xi}-1}dx.$$

Tomando  $G(x) = -\left(1 + \frac{\xi}{\beta}x\right)^{-1/\xi}$  e  $g(x) = G'(x)$ , integrando por partes, temos

$$\begin{aligned} E[X^r] &= \lim_{a \rightarrow \infty} \int_0^a x^r g(x) dx = \lim_{a \rightarrow \infty} \left[ x^r G(x) \Big|_0^a - \int_0^a r x^{r-1} G(x) dx \right] = \\ &= \lim_{a \rightarrow \infty} \left[ x^r G(x) \Big|_0^a + \int_0^a r x^{r-1} \left(1 + \frac{\xi}{\beta}x\right)^{-1/\xi} dx \right] = \\ &= \lim_{a \rightarrow \infty} \left[ a^r G(a) + \int_0^a r x^{r-1} \left(1 + \frac{\xi}{\beta}x\right)^{-1/\xi} dx \right]. \end{aligned}$$

Agora, como  $r < 1/\xi$ , temos que

$$\lim_{a \rightarrow \infty} a^r G(a) = \lim_{a \rightarrow \infty} \left[ -a^r \left(1 + \frac{\xi}{\beta}a\right)^{-1/\xi} \right] = 0,$$

Assim,

$$E[X^r] = \int_0^\infty r x^{r-1} \left(1 + \frac{\xi}{\beta}x\right)^{-1/\xi} dx. \quad (2.19)$$



Fazendo  $-1/\xi = -1/\xi^* - 1$  e  $\beta/\xi = \beta^*/\xi^*$ , temos

$$\xi^* = \frac{1}{\frac{1}{\xi} - 1} \text{ e } \beta^* = \frac{\beta}{1 - \xi}.$$

Denote  $E[X^r] = E_{\xi, \beta}[X^r]$  e  $E_{\xi^*, \beta^*}[X^r]$  a esperança de  $X$ , para  $(\xi^*, \beta^*)$ . De (2.19), temos a seguinte fórmula de recorrência

$$E[X^r] = E_{\xi, \beta}[X^r] = \frac{r\beta^*}{\beta^*} \int_0^\infty x^{r-1} \left(1 + \frac{\xi^*}{\beta^*}x\right)^{-\frac{1}{\xi^*}-1} dx = r\beta^* E_{\xi^*, \beta^*}[X^{r-1}]. \quad (2.20)$$

Agora, por indução, temos por (2.18) que

$$EX = \frac{\beta}{1 - \xi}.$$

Por outro lado,

$$\frac{\beta^1}{\xi^{1+1}} \frac{\Gamma(-1 + 1/\xi)}{\Gamma(1 + 1/\xi)} 1! = \frac{\beta}{\xi^2} \frac{\Gamma(-1 + 1/\xi)}{(1/\xi)\Gamma(1/\xi)} = \frac{\beta}{\xi} \frac{\Gamma(-1 + 1/\xi)}{(-1 + 1/\xi)\Gamma(-1 + 1/\xi)} = \frac{\beta}{1 - \xi}.$$

Sendo assim, temos que (2.13) é válida para  $r = 1$ . Supondo agora que (2.13) seja válida para  $r - 1$ , de (2.20), temos

$$\begin{aligned} E[X^r] &= E_{\xi, \beta}[X^r] = r\beta^* E_{\xi^*, \beta^*}[X^{r-1}] = r\beta^* \frac{\beta^{*r-1}}{(\xi^*)^r} \frac{\Gamma(-r + 1 + 1/\xi^*)}{\Gamma(1 + 1/\xi^*)} (r-1)! = \\ &= r! \frac{\beta^{*r}}{(\xi^*)^r} \frac{\Gamma(-r + 1 + 1/\xi^*)}{\Gamma(1 + 1/\xi^*)} = \left(\frac{\beta^*}{\xi^*}\right)^r \frac{\Gamma(-r + 1/\xi)}{\Gamma(1/\xi)} r! = \\ &= \left(\frac{\beta}{\xi}\right)^r \frac{(1/\xi)\Gamma(-r + 1/\xi)}{(1/\xi)\Gamma(1/\xi)} r! = \frac{\beta^r}{\xi^{r+1}} \frac{\Gamma(-r + 1/\xi)}{\Gamma(1 + 1/\xi)} r! \end{aligned}$$

**Propriedade (b).** Pela caracterização do  $D_{\max}(H_\xi)$ , Teorema 2.1, temos que  $F \in D_{\max}(H_\xi)$ , se, e somente se, existe uma função positiva e mensurável,  $a(t)$ , tal que, para  $1 + \xi x > 0$ ,

$$\lim_{t \uparrow \omega(F)} \frac{\overline{F}(t + xa(t))}{\overline{F}(t)} = \begin{cases} (1 + \xi x)^{-1/\xi}, & \text{se } \xi \neq 0 \\ e^{-x}, & \text{se } \xi = 0 \end{cases}.$$

Por outro lado, se  $X$  é uma v.a. com f.d.  $F$ , então,

$$\lim_{t \uparrow \omega(F)} \frac{\overline{F}(t + xa(t))}{\overline{F}(t)} = \lim_{t \uparrow \omega(F)} P\left(\frac{X-t}{a(t)} > x \mid X > t\right),$$

em que  $t < \omega(F)$  e  $xa(t) < \omega(F) - t$ .

Assim, para  $\xi \neq 0$  e  $1 + \xi x > 0$ , temos

$$F \in D_{\max}(H_{\xi}) \iff \lim_{t \uparrow \omega(F)} \left| P\left(\frac{X-t}{a(t)} > x \mid X > t\right) - (1 + \xi x)^{-1/\xi} \right| = 0.$$

Fazendo  $y = xa(t)$  e, como  $t + xa(t) < \omega(F)$ , então  $0 < y < \omega(F) - t$  e

$$\begin{aligned} F \in D_{\max}(H_{\xi}) &\iff \lim_{t \uparrow \omega(F)} \left| P(X - t > y \mid X > t) - \left(1 + \xi \frac{y}{a(t)}\right)^{-1/\xi} \right| = 0 \\ &\iff \lim_{t \uparrow \omega(F)} \left| F_t(y) - G_{\xi, \beta(t)}(y) \right| = 0, \end{aligned}$$

em que  $\beta(t) = a(t)$ . Para  $\xi = 0$ , temos

$$F \in D_{\max}(H_{\xi}) \iff \lim_{t \uparrow \omega(F)} \left| P\left(\frac{X-t}{a(t)} > x \mid X > t\right) - e^{-x} \right| = 0.$$

Da mesma maneira, fazendo  $y = xa(t)$  e, como  $t + xa(t) < \omega(F)$ , então  $0 < y < \omega(F) - t$  e

$$\begin{aligned} F \in D_{\max}(H_{\xi}) &\iff \lim_{t \uparrow \omega(F)} \left| P(X - t > y \mid X > t) - e^{-y/a(t)} \right| = 0 \\ &\iff \lim_{t \uparrow \omega(F)} \left| F_t(y) - G_{0, \beta(t)}(y) \right| = 0, \end{aligned}$$

em que  $\beta(t) = a(t)$ . Como a distribuição GPD é contínua, segue pela uniformidade da convergência

$$\lim_{t \uparrow \omega(F)} \sup_{0 < y < \omega(F) - t} \left| F_t(y) - G_{\xi, \beta(t)}(y) \right| = 0.$$

**Propriedade (c).** Observe que, para  $\xi \neq 0$ , temos  $\bar{G}_{\xi, \beta}(x) = \left(1 + (\xi/\beta)x\right)^{-1/\xi}$ , em que  $x \in D(\xi, \beta)$ . Para  $x_i \in D(\xi, \beta)$ ,  $i = 1, 2$ , então

$$\begin{aligned} \frac{\bar{G}_{\xi, \beta}(x_1 + x_2)}{\bar{G}_{\xi, \beta}(x_1)} &= \frac{\left(1 + \xi \frac{x_1 + x_2}{\beta}\right)^{-1/\xi}}{\left(1 + \xi \frac{x_1}{\beta}\right)^{-1/\xi}} = \left(\frac{1 + \xi \frac{x_1}{\beta}}{1 + \xi \frac{x_1}{\beta} + \frac{\xi x_2}{\beta}}\right)^{-1/\xi} = \\ &= \left(1 + \frac{\xi x_2}{\beta + \xi x_1}\right)^{-1/\xi} = \bar{G}_{\xi, \beta + \xi x_1}(x_2). \end{aligned}$$

Para  $\xi = 0$ ,  $\bar{G}_{0,\beta}(x) = e^{-x/\beta}$ ,  $x \geq 0$  e (2.15) segue trivialmente.

**Propriedade (d).** Das hipóteses consideradas, temos que

$$\begin{aligned}
 P(Z_N \leq x) &= \sum_{n=0}^{\infty} P(Z_N \leq x | N = n) P(N = n) = \sum_{n=0}^{\infty} P(\max\{X_1, \dots, X_n\} \leq x) P(N = n) = \\
 &= \sum_{n=0}^{\infty} [P(X_1 \leq x)]^n P(N = n) = \sum_{n=0}^{\infty} \left(1 - \left(1 + \frac{\xi x}{\beta}\right)^{-1/\xi}\right)^n \frac{e^{-\lambda} \lambda^n}{n!} = \\
 &= e^{-\lambda} \sum_{n=0}^{\infty} \left[\lambda \left(1 - \left(1 + \frac{\xi x}{\beta}\right)^{-1/\xi}\right)\right]^n \frac{1}{n!} = \\
 &= e^{-\lambda} \exp\left\{\lambda \left(1 - \left(1 + \frac{\xi x}{\beta}\right)^{-1/\xi}\right)\right\} = \exp\left\{-\lambda \left(1 + \frac{\xi x}{\beta}\right)^{-1/\xi}\right\}.
 \end{aligned}$$

Note que

$$\begin{aligned}
 \exp\left\{-\lambda \left(1 + \frac{\xi x}{\beta}\right)^{-1/\xi}\right\} &= \exp\left\{-\left(1 - 1 + \lambda^{-\xi} + \lambda^{-\xi} \frac{\xi x}{\beta}\right)^{-1/\xi}\right\} \\
 &= \exp\left\{-\left(1 + \frac{-\beta \lambda^{\xi} + \beta + \xi x}{\beta \lambda^{\xi}}\right)^{-1/\xi}\right\} \\
 &= \exp\left\{-\left(1 + \xi \frac{x - \xi^{-1} \beta (\lambda^{\xi} - 1)}{\beta \lambda^{\xi}}\right)^{-1/\xi}\right\} \\
 &= H_{\xi, \mu, \psi},
 \end{aligned}$$

em que  $\mu = \beta \xi^{-1} (\lambda^{\xi} - 1)$  e  $\psi = \beta \lambda^{\xi}$ , portanto, segue o resultado.

**Propriedade (e).** Para  $X$  uma v.a. com f.d. Pareto generalizada,  $G_{\xi, \beta}$ , com parâmetros  $\xi < 1$  e  $\beta > 0$ , a função média do excesso, em  $t < \omega(G_{\xi, \beta})$ , é dada por (2.7) como

$$e(t) = \frac{1}{\bar{G}_{\xi, \beta}(t)} \int_t^{\omega(G_{\xi, \beta})} \bar{G}_{\xi, \beta}(x) dx.$$

Para  $0 < \xi < 1$ ,  $x \in [0, \infty)$  e  $\omega(G_{\xi, \beta}) = \infty$ . Assim, para  $t > 0$  e tomando  $u = 1 + \frac{\xi}{\beta}x$ , temos que

$$\begin{aligned} e(t) &= \frac{1}{\overline{G}_{\xi, \beta}(t)} \int_{1+\frac{\xi t}{\beta}}^{\infty} u^{-1/\xi} \left(\frac{\beta}{\xi}\right) du = \frac{\beta}{\xi \overline{G}_{\xi, \beta}(t)} \frac{u^{-\frac{1}{\xi}+1}}{-\frac{1}{\xi}+1} \Bigg|_{1+\frac{\xi t}{\beta}}^{\infty} = \frac{\beta}{-\xi \left(1 + \frac{\xi t}{\beta}\right)^{-\frac{1}{\xi}}} \frac{\left(1 + \frac{\xi t}{\beta}\right)^{1-\frac{1}{\xi}}}{1 - \frac{1}{\xi}} \\ &= \frac{\beta}{1-\xi} \left(1 + \frac{\xi t}{\beta}\right) = \frac{\beta + \xi t}{1-\xi}, \end{aligned}$$

em que  $\beta + \xi t > 0$ .

Se  $\xi = 0$ ,  $x \in [0, \infty)$  e  $\omega(G_{0, \beta}) = \infty$ . Assim, para  $t > 0$ , temos

$$e(t) = \frac{1}{e^{-t/\beta}} \int_t^{\infty} e^{-x/\beta} dx = e^{t/\beta} (-\beta) e^{-x/\beta} \Bigg|_t^{\infty} = e^{t/\beta} (-\beta) (-e^{-t/\beta}) = \beta.$$

Por último, se  $\xi < 0$ ,  $x \in [0, -\beta/\xi]$  e  $\omega(G_{\xi, \beta}) = -\beta/\xi$ . Para  $t > 0$  e tomando  $u = 1 + \frac{\xi}{\beta}x$ , temos

$$\begin{aligned} e(t) &= \frac{1}{\overline{G}_{\xi, \beta}(t)} \int_{1+\frac{\xi t}{\beta}}^0 u^{-1/\xi} \left(\frac{\beta}{\xi}\right) du = \frac{\beta}{\xi \overline{G}_{\xi, \beta}(t)} \int_{1+\frac{\xi t}{\beta}}^0 u^{-1/\xi} du = \\ &= \frac{\beta}{-\xi \left(1 + \frac{\xi t}{\beta}\right)^{-\frac{1}{\xi}}} \frac{\left(1 + \frac{\xi t}{\beta}\right)^{1-\frac{1}{\xi}}}{1 - \frac{1}{\xi}} = \frac{\beta}{1-\xi} \left(1 + \frac{\xi t}{\beta}\right) = \frac{\beta + \xi t}{1-\xi}, \end{aligned}$$

em que  $\beta + \xi t > 0$ . □

## CAPÍTULO

### 3

# INFERÊNCIA VIA MÁXIMA VEROSSIMILHANÇA

## 3.1 Introdução

A distribuição Pareto generalizada foi introduzida por Pickands [10] para modelar excedentes acima de um limiar. Contudo, muitos são os autores que destacam certos problemas quanto à estimação dos parâmetros dessa distribuição, por exemplo, há situações em que o estimador de máxima verossimilhança não existe. Muito do que se conhece vem de técnicas numéricas, como pode ser conferido em Castillo e Hadi [11] e Zhang e Stephens [12].

Neste capítulo, trataremos da inferência dos parâmetros da GPD,  $\kappa = -\xi$  (índice caudal) e  $\beta$ . Na Seção 3.2 serão apresentados conceitos preliminares acerca do princípio de máxima verossimilhança. Na Seção 3.3, mostraremos um estudo da GPD quanto à estimação dos parâmetros do modelo, baseados nos trabalhos de Castillo e Serra [1], Castillo e Daoudi [2] e Kozubowski [3]. Um análise detalhada do estimador de máxima verossimilhança de  $\theta = (\kappa, \beta)$  será apresentada do ponto de vista teórico. Por fim, na Seção 3.4, apresentaremos dois testes que serão utilizados nas aplicações vistas no Capítulo 4. São eles: o critério

de informação Akaike (AIC), para seleção de modelos, e o Teste  $\chi^2$ , para testar se uma distribuição de frequências difere de uma distribuição teórica.

## 3.2 Princípio da máxima verossimilhança

**Definição 3.1.** *Uma amostra aleatória de tamanho  $n$  de uma f.d.  $F$  é dada por  $\mathbf{X} = (X_1, X_2, \dots, X_n)$ , em que  $X_1, X_2, \dots, X_n$  são variáveis aleatórias i.i.d com f.d. comum  $F$ .*

*Se  $X_1, X_2, \dots, X_n$  são variáveis aleatórias i.i.d com densidade comum  $f$ , podemos dizer que  $\mathbf{X}$  é uma amostra aleatória de  $f$ .*

Para as definições seguintes, considere  $X_1, X_2, \dots, X_n$  uma amostra aleatória com densidade  $f(x; \theta)$ , tal que  $\theta$  é um parâmetro pertencente ao espaço  $\Theta \subset \mathbb{R}^r$ ,  $r \in \mathbb{N}$ . Assumimos que o valor de  $\theta$  é desconhecido e, dessa forma, tem-se como objetivo estimá-lo.

**Definição 3.2.** *Um estimador de  $\theta$  é uma estatística  $\hat{\theta}_n = W(X_1, \dots, X_n)$ , com  $W : \mathbb{R}^n \rightarrow \Theta$ .*

**Definição 3.3.** *A função de verossimilhança de um parâmetro  $\theta$  baseada nas observações  $x_1, \dots, x_n$  da amostra  $X_1, \dots, X_n$  é dada por*

$$L(\theta; x_1, \dots, x_n) = \prod_{i=1}^n f(x_i; \theta). \quad (3.1)$$

**Definição 3.4.** *O estimador de máxima verossimilhança (EMV) de um parâmetro  $\theta$  é o valor  $\hat{\theta}$  que maximiza a função de verossimilhança  $L(\theta) = L(\theta; x_1, \dots, x_n)$ , caso exista, isto é, é um valor  $\hat{\theta} = \hat{\theta}_n = \hat{\theta}(x_1, \dots, x_n) \in \Theta$  tal que*

$$L(\hat{\theta}) = \arg \max_{\theta \in \Theta} L(\theta; x_1, \dots, x_n) = \arg \max_{\theta \in \Theta} \prod_{i=1}^n f(x_i; \theta), \quad (3.2)$$

em que  $(x_1, \dots, x_n)$  são observações da amostra  $X_1, \dots, X_n$ .

### Observação 3.1.

- (1) É possível que  $\hat{\theta}$  não exista ou pode existir mas não ser único.
- (2) A função logarítmica é monótona estritamente crescente. Dessa forma, com o intuito de simplificar o processo para obter o EMV de  $\theta$ , ele pode ser calculado maximizando a função log-verossimilhança de  $\theta$ , dada por

$$l(\theta) = \ln L(\theta) = \sum_{i=1}^n \ln f(x_i; \theta). \quad (3.3)$$

Se existir, o valor que maximiza a função  $l(\theta)$  será o mesmo que maximiza  $L(\theta)$ , ou seja, ele será o estimador de máxima verossimilhança desejado.

(3) O valor do EMV de  $\theta$  pode ser obtido resolvendo as equações de verossimilhança

$$\frac{\partial l(\theta)}{\partial \theta_i} = 0, \quad i = 1, \dots, r, \quad (3.4)$$

desde que  $l(\theta)$  seja diferenciável em  $\Theta$  e  $\hat{\theta}$  seja um máximo local.

**Exemplo 3.1** (Distribuição exponencial). Seja uma amostra  $X_1, \dots, X_n$  de variáveis aleatórias com distribuição comum exponencial, com o parâmetro  $\lambda$  desconhecido. Assim, para estimá-lo, foram observados os seguintes valores dessa amostra:  $x_1, \dots, x_n$ . Portanto, temos  $f(x; \theta) = f(x_i; \lambda) = \lambda e^{-\lambda x_i}$ ,  $x_i > 0$ ,  $i = 1, \dots, n$ . De (3.3), a função log-verossimilhança é dada por

$$l(\theta) = l(\lambda) = \sum_{i=1}^n \ln f(x_i; \lambda) = \sum_{i=1}^n \ln (\lambda e^{-\lambda x_i}) = n \ln \lambda - \lambda \sum_{i=1}^n x_i. \quad (3.5)$$

Já, de (3.4), a equação de verossimilhança é resolvida da seguinte forma

$$\frac{\partial l(\lambda)}{\partial \lambda} = 0 \implies \frac{n}{\lambda} - \sum_{i=1}^n x_i = 0 \implies \frac{1}{\lambda} = \frac{\sum_{i=1}^n x_i}{n} = \bar{x} \implies \lambda = (\bar{x})^{-1}.$$

Agora,

$$\left. \frac{\partial^2 l(\lambda)}{\partial \lambda^2} \right|_{\lambda=\hat{\lambda}} = \left( -\frac{n}{\lambda^2} \right) \Big|_{\lambda=\hat{\lambda}} = -\frac{n}{\bar{x}^2} < 0.$$

Assim,  $\hat{\lambda} = (\bar{x})^{-1}$  é o EMV de  $\lambda$ .

A inferência de parâmetros pelo método de máxima verossimilhança tem grande importância devido às suas propriedades matemáticas, sobretudo quando  $n$  é suficientemente grande. A seguir, enunciamos as principais delas.

**Definição 3.5.** Um estimador  $\hat{\theta}_n$  é dito um **estimador consistente** de  $\theta$  se a sequência  $\{\hat{\theta}_n\}$  converge em probabilidade para  $\theta$ , isto é, se para todo  $\varepsilon > 0$  arbitrário temos que

$$\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| < \varepsilon) = 1. \quad (3.6)$$

**Exemplo 3.2** (Distribuição exponencial). Seja a amostra  $X_1, \dots, X_n$  de variáveis aleatórias dada com distribuição comum exponencial de parâmetro  $\lambda$ . Temos

$$\{\hat{\theta}_n\} = \{\hat{\lambda}_n\} = \frac{n}{\sum_{i=1}^n X_i} = (\bar{X})^{-1}.$$

Agora, para  $\varepsilon > 0$ ,

$$\lim_{n \rightarrow \infty} P(|\hat{\lambda}_n - \lambda| < \varepsilon) = \lim_{n \rightarrow \infty} P(|(\bar{X})^{-1} - \lambda| < \varepsilon).$$

Pela Lei Forte dos Grandes Números para variáveis aleatórias i.i.d., temos que

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\text{q.c.}} \frac{1}{\lambda},$$

e também  $(\bar{X})^{-1} \xrightarrow{\text{q.c.}} \lambda$ , em que  $\xrightarrow{\text{q.c.}}$  denota a convergência quase certa.

Por outro lado, a convergência quase certa implica a convergência em probabilidade (denotada por  $\xrightarrow{P}$ ), então

$$(\bar{X})^{-1} \xrightarrow{P} \lambda.$$

Portanto,

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n x_i},$$

em que  $x_i$   $i = 1, \dots, n$  são observações da amostra  $X_1, \dots, X_n$ , é um estimador consistente de  $\lambda$ .

**Definição 3.6** (Vício). A diferença entre o valor esperado de  $\hat{\theta}$  e  $\theta$ , dada por

$$E(\hat{\theta}) - \theta,$$

é chamado o **vício** do estimador  $\hat{\theta}$  do parâmetro  $\theta$ .

**Definição 3.7.** A matriz quadrada de ordem  $r$ ,  $I(\theta) = [I_{jk}(\theta)]_{r \times r}$ , cujos termos são dados por

$$I_{jk}(\theta) = \text{cov} \left[ \frac{\partial}{\partial \theta_j} \ln f(X, \theta), \frac{\partial}{\partial \theta_k} \ln f(X, \theta) \right], \quad j, k = 1, \dots, r, \quad (3.7)$$

e nos casos em que

$$E \left( \frac{\partial}{\partial \theta_j} \ln f(X, \theta) \right) = 0, \quad \forall j = 1, \dots, r$$

e

$$E \left( \frac{\partial}{\partial \theta_j} \ln f(X, \theta) \cdot \frac{\partial}{\partial \theta_k} \ln f(X, \theta) \right) = -E \left( \frac{\partial^2}{\partial \theta_j \partial \theta_k} \ln f(X, \theta) \right),$$



então,

$$I_{jk} = -E\left(\frac{\partial^2}{\partial\theta_j\partial\theta_k} \ln f(X, \theta)\right) \quad (3.8)$$

é chamada a **matriz de informação de Fisher** de  $X$ .

**Exemplo 3.3** (Distribuição exponencial). Tomando  $X_1, \dots, X_n$  variáveis aleatórias i.i.d. com distribuição comum exponencial de parâmetro  $\lambda$ , como no Exemplo 3.1, temos que  $\theta = \lambda$  e, de (3.8), segue que

$$\begin{aligned} I(\theta) = I(\lambda) &= -E\left(\frac{\partial^2}{\partial\lambda^2} \ln f(X_1, \lambda)\right) = -E\left(\frac{\partial^2}{\partial\lambda^2} \ln(\lambda e^{-\lambda X_1})\right) \\ &= -E\left(\frac{\partial^2}{\partial\lambda^2} (\ln(\lambda) - \lambda X_1)\right) = -E\left(-\frac{1}{\lambda^2}\right) = \frac{1}{\lambda^2}. \end{aligned}$$

Um estimador **assintoticamente normal** é um estimador consistente cuja distribuição em torno do verdadeiro parâmetro  $\theta$  se aproxima de uma distribuição normal.

Se  $\hat{\theta}_n$  é o EMV do vetor paramétrico  $\theta$ , obedecendo algumas condições de regularidade (veja Lehmann [14]), então mostra-se que ele é assintoticamente normal e satisfaz

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N\left(\mathbf{0}, \frac{1}{I(\theta)}\right), \quad (3.9)$$

ou seja,  $\sqrt{n}(\hat{\theta}_n - \theta)$  converge em distribuição para um vetor aleatório com distribuição normal multivariada de vetor média zero,  $\mathbf{0} \in \mathbb{R}^r$ , e matriz de covariância  $[I(\theta)]^{-1} \in \mathbb{R}^{r \times r}$ , em que  $r$  é o número de parâmetros que compõem  $\theta$ .

**Exemplo 3.4** (Distribuição exponencial). Tomando o estimador de máxima verossimilhança de  $\theta = \lambda$  para uma amostra aleatória com distribuição comum exponencial, temos

$$\sqrt{n}(\hat{\theta}_n - \theta) = \sqrt{n}((\bar{X})^{-1} - \lambda).$$

Expandindo a função  $h(x) = x^{-1} - \lambda$  em série de Taylor no ponto  $x_0 = \lambda^{-1}$ , temos

$$h(x) = h(x_0) + h'(x_0)(x - x_0) + r_2(x), \quad x \neq 0,$$

em que  $\lim_{x \rightarrow x_0} \frac{r_2(x)}{(x - x_0)^2} = 0$ . Assim,

$$h(x) = -\lambda^2(x - \lambda^{-1}) + r_2(x), \quad x \neq 0,$$

e

$$h(\bar{X}) = -\lambda^2(\bar{X} - \lambda^{-1}) + r_2(\bar{X}).$$

Como  $\bar{X} \xrightarrow{q.c.} \lambda^{-1}$ , temos que

$$\frac{r_2(\bar{X})}{(\bar{X} - \lambda^{-1})^2} \xrightarrow{q.c.} 0. \quad (3.10)$$

Agora, denotando por  $\mu = 1/\lambda$  a esperança da distribuição exponencial, por  $\sigma^2 = 1/\lambda^2$  sua variância e por  $S_n = \sum_{i=1}^n X_i$ , temos que

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n - \theta) &= -\sqrt{n}\lambda^2\left(\bar{X} - \frac{1}{\lambda}\right) + \sqrt{nr}r_2(\bar{X}) = -\frac{1}{\sigma}\left(\frac{S_n - \mu n}{\sigma\sqrt{n}}\right) + \sqrt{nr}r_2(\bar{X}) \\ &= -\frac{1}{\sigma}\left(\frac{S_n - \mu n}{\sigma\sqrt{n}}\right) + \sqrt{n}\left(\frac{\bar{X} - \lambda^{-1}}{\lambda^{-1}}\right)\lambda^{-1}(\bar{X} - \lambda^{-1})\frac{r_2(\bar{X})}{(\bar{X} - \lambda^{-1})^2} \\ &= -\frac{1}{\sigma}\left(\frac{S_n - \mu n}{\sigma\sqrt{n}}\right) + \left(\frac{S_n - \mu n}{\sigma\sqrt{n}}\right)\lambda^{-1}(\bar{X} - \lambda^{-1})\frac{r_2(\bar{X})}{(\bar{X} - \lambda^{-1})^2}. \end{aligned}$$

Do Teorema do Limite Central e de (3.10), segue que

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, 1/\sigma^2) = N(0, [I(\theta)]^{-1}).$$

Portanto, o EMV para o parâmetro  $\lambda$  da distribuição exponencial tem distribuição assintoticamente normal.

Existem condições que garantem a consistência e a normalidade assintótica dos estimadores de máxima verossimilhança. Como referências básicas para esses estudos podemos citar Cramer [13] e Lehmann [14].

### 3.3 Inferência para a distribuição Pareto generalizada

Como pode ser visto na Seção 2.3, a GPD inclui a distribuição exponencial, quando  $\xi = 0$ , e, nesse caso, o EMV do parâmetro  $\beta$  existe, é consistente e tem distribuição normal assintótica. Em outros casos, quando  $\xi \neq 0$ , os estimadores de máxima verossimilhança para  $\kappa = -\xi$  e  $\beta$ , podem existir ou não, ou ainda não possuem as propriedades assintóticas.

Nesse sentido, nesta seção apresentaremos resultados decorrentes do estudo acerca da existência dos estimadores de máxima verossimilhança para os parâmetros da GPD, especialmente para  $\kappa$ . Mostraremos também uma análise das limitações desses estimadores na inferência dos parâmetros da GPD.

Seja a reparametrização da GPD considerando  $\kappa = -\xi$ . Assim,

$$G_{\kappa, \beta}(x) = 1 - (1 - \kappa x / \beta)^{1/\kappa}, x \in D(\kappa, \beta), \quad (3.11)$$

$$\text{em que, } D(\kappa, \beta) = \begin{cases} [0, \infty), & \text{se } \kappa \leq 0 \\ [0, \beta/\kappa], & \text{se } \kappa > 0 \end{cases} \quad (\beta > 0).$$

A densidade correspondente é dada por

$$g(x; \kappa, \beta) = \frac{1}{\beta} \left(1 - \frac{\kappa}{\beta} x\right)^{1/\kappa - 1}. \quad (3.12)$$

**Observação 3.2.** Note que  $\kappa$  é o índice caudal da GPD e

- (1) para  $\kappa = 0$ , a GPD é a distribuição exponencial com média  $\beta$ ;
- (2) para  $\kappa = 1$ , a GPD é a distribuição uniforme em  $[0, \beta]$ ;
- (3) para  $\kappa < 0$ , a GPD é uma distribuição de cauda pesada;
- (4) para  $\kappa > 0$ , a GPD é uma distribuição com cauda leve.

Sejam  $x_1, \dots, x_n$  valores amostrados de  $X$ , v.a. com distribuição  $G_{\kappa, \beta}$ . A log-verossimilhança de  $(\kappa, \beta)$  é dada por

$$l(\kappa, \beta) = -n \ln \beta + \left(\frac{1}{\kappa} - 1\right) \sum_{i=1}^n \ln \left(1 - \kappa \frac{x_i}{\beta}\right), \quad (3.13)$$

em que  $x_i \geq 0$ ,  $i = 1, \dots, n$ , se  $\kappa \leq 0$ , e  $x_{(n)} \leq \beta/\kappa$ , se  $\kappa \geq 0$ , em que  $x_{(n)} = \max\{x_1, \dots, x_n\}$ .

### 3.3.1 Existência do estimador de máxima verossimilhança do parâmetro $\kappa$ (índice caudal)

Nesta subseção, serão apresentados os resultados mais importantes deste capítulo. Trata-se de uma sequência de teoremas e proposições que avaliam a existência do estimador de máxima verossimilhança de  $\theta = (\kappa, \beta)$  para a inferência dos parâmetros  $\kappa$  e  $\beta$  da distribuição Pareto generalizada. Este estudo foi baseado nos trabalhos de Castillo e Serra [1], Castillo e Daoudi [2] e Kozubowski [3].

Conforme visto no início desta seção, a GPD tem características importantes quanto ao índice caudal. Para  $\kappa = 0$ , a GPD é uma distribuição exponencial. Pode ser visto na literatura uma vasta exploração desse modelo. Talvez uma explicação para esse fato se deve às qualidades do EMV nesse caso.

Entretanto, a GPD se apresenta de forma bem diferente nas regiões: (a)  $\kappa < 0$ , em que a distribuição tem cauda pesada; e (b)  $\kappa > 0$ , em que a distribuição tem cauda leve. Assim, são nessas regiões que reside o interesse no estudo do modelo.

Para  $\kappa = 1$ , em que a GPD é a distribuição uniforme, o modelo não foi suficientemente explorado na literatura, segundo afirmam Castillo e Serra [1].

O resultado a seguir mostra o papel da distribuição uniforme no estudo de caudas.

**Proposição 3.1.** *Se uma variável aleatória absolutamente contínua com f.d  $F$  e densidade  $f(x)$ , com suporte em  $[0, \omega(F)]$ , contínua à esquerda em  $\omega(F)$  e  $0 < f(\omega(F)) < \infty$ , então  $F \in D_{\max}(H_\xi)$ , com  $\xi = -1$  ( $\kappa = 1$ ).*

**Demonstração.** Pelo Teorema 2.2 (b), já demonstrado que:

$$\lim_{t \rightarrow \omega(F)} \sup_{0 \leq x < \omega(F) - t} |F_t(x) - G_{\kappa, \beta(t)}(x)| = 0,$$

para alguma função positiva  $\beta(t)$ . Portanto, resta mostrar que  $\kappa = 1$ .

Fazendo  $\beta(t) = \omega(F) - t$ , temos que

$$G_{1, \beta(t)}(x) = \frac{x}{\omega(F) - t} \quad (3.14)$$

é a distribuição uniforme em  $[0, \omega(F) - t]$  e

$$F_t(x) = \frac{F(x+t) - F(t)}{1 - F(t)} \quad (3.15)$$

é a função de distribuição dos excessos em  $t > 0$ .

Dado  $\varepsilon_0$ , encontraremos  $t$  tal que  $\sup_{0 \leq x < \omega(F) - t} |F_t(x) - G_{1, \beta(t)}(x)| < \varepsilon_0$ .

Seja  $c = f(\omega(F))$ . Como  $f$  é contínua pela esquerda em  $\omega(F)$ , se  $\varepsilon < \min\{c\varepsilon_0/4, c/2\}$ , existe  $t$  tal que

$$c - \varepsilon < f(s) < c + \varepsilon, \text{ para todo } t \leq s \leq \omega(F).$$

Fazendo  $s = t + x$ , com  $x \geq 0$ , pelo Teorema Fundamental do Cálculo em  $[t, s]$ ,

$$x(c - \varepsilon) < F(t+x) - F(t) < x(c + \varepsilon). \quad (3.16)$$

Em particular para  $s = \omega(F)$ ,

$$(\omega(F) - t)(c - \varepsilon) < 1 - F(t) < (\omega(F) - t)(c + \varepsilon). \quad (3.17)$$

De (3.16) e (3.17), temos

$$\frac{x}{\omega(F) - t} \left( \frac{c - \varepsilon}{c + \varepsilon} \right) < \frac{F(t+x) - F(t)}{1 - F(t)} < \frac{x}{\omega(F) - t} \left( \frac{c + \varepsilon}{c - \varepsilon} \right). \quad (3.18)$$

Assim, com  $x < \omega(F) - t$ ,

$$\left| F_t(x) - G_{1, \beta(t)}(x) \right| < \left( \frac{c + \varepsilon}{c - \varepsilon} - 1 \right) \frac{x}{\omega(F) - t} \leq \frac{2\varepsilon}{c - \varepsilon} < 2 \frac{c\varepsilon_0/4}{c - c/2} = 4\varepsilon_0/4 = \varepsilon_0.$$

□

Considere agora a função log-verossimilhança dada em (3.13), ou seja,

$$l(\kappa, \beta) = -n \ln \beta + \left( \frac{1}{\kappa} - 1 \right) \sum_{i=1}^n \ln \left( 1 - \frac{\kappa}{\beta} x_i \right),$$

em que  $x_1, \dots, x_n$  são valores amostrados da  $G_{\kappa, \beta}$ , com  $x_i \geq 0$ ,  $i = 1, \dots, n$ , se  $\kappa \leq 0$ , e  $x_{(n)} \leq \beta/\kappa$ , se  $\kappa > 0$ . Daí, temos que

$$\frac{\partial l}{\partial \beta} = -\frac{n}{\beta} + \left( \frac{1}{\kappa} - 1 \right) \sum_{i=1}^n \frac{\frac{\kappa}{\beta^2} x_i}{1 - \frac{\kappa}{\beta} x_i} = -\frac{n}{\beta} \left[ 1 - \left( \frac{1}{\kappa} - 1 \right) \frac{1}{n} \sum_{i=1}^n \frac{\frac{\kappa}{\beta} x_i}{1 - \frac{\kappa}{\beta} x_i} \right], \quad (3.19)$$

para  $\kappa \neq 0$ .

Ao resolver  $\frac{\partial l}{\partial \beta} = 0$ , encontramos

$$-\frac{n}{\beta} \left[ 1 - \left( \frac{1}{\kappa} - 1 \right) \frac{1}{n} \sum_{i=1}^n \frac{\frac{\kappa}{\beta} x_i}{1 - \frac{\kappa}{\beta} x_i} \right] = 0 \implies 1 = \left( \frac{1}{\kappa} - 1 \right) \frac{1}{n} \sum_{i=1}^n \frac{\frac{\kappa}{\beta} x_i}{1 - \frac{\kappa}{\beta} x_i},$$

com  $\kappa \neq 1$ . Dessa forma, para  $\kappa \neq 0$  e  $\kappa \neq 1$ ,

$$\frac{\kappa}{1 - \kappa} = \frac{1}{n} \sum_{i=1}^n \frac{\frac{\kappa}{\beta} x_i}{1 - \frac{\kappa}{\beta} x_i}, \quad (3.20)$$

e, assim, a primeira forma para  $\frac{\partial l}{\partial \beta} = 0$  é dada por

$$\frac{1}{1 - \kappa} = \frac{1}{n} \sum_{i=1}^n \frac{x_i/\beta}{1 - \frac{\kappa}{\beta} x_i} = \frac{1}{n} \sum_{i=1}^n \frac{x_i}{\beta - \kappa x_i}. \quad (3.21)$$

Se adicionarmos 1 em cada termo de (3.20), temos a seguinte igualdade

$$\frac{\kappa}{1-\kappa} = \frac{1}{n} \sum_{i=1}^n \left[ \frac{\frac{\kappa}{\hat{\beta}} x_i}{1 - \frac{\kappa}{\hat{\beta}} x_i} + 1 \right] - \frac{n}{n},$$

e, portanto, a segunda forma para  $\frac{\partial l}{\partial \beta} = 0$  é dada por

$$\frac{1}{1-\kappa} = \frac{1}{n} \sum_{i=1}^n \frac{1}{1 - \frac{\kappa}{\hat{\beta}} x_i}, \quad (3.22)$$

em que  $\kappa \neq 0$  e  $\kappa \neq 1$ .

**Definição 3.8** (Função de log-verossimilhança profile). *Dada a função log-verossimilhança da família GPD (3.13), a função de log-verossimilhança profile de  $\kappa$  é definida por*

$$l_p(\kappa) = l(\kappa, \hat{\beta}(\kappa)) = -n \ln \hat{\beta}(\kappa) + \left( \frac{1}{\kappa} - 1 \right) \sum_{i=1}^n \ln \left( 1 - \frac{x_i \kappa}{\hat{\beta}(\kappa)} \right), \quad (3.23)$$

em que  $\hat{\beta}(\kappa)$  é resultado de (3.22) para um  $\kappa$  fixo ( $\kappa \neq 0$  e  $\kappa \neq 1$ ).

**Proposição 3.2.** *Sejam  $x_1, \dots, x_n$  observações de variáveis aleatórias i.i.d. com f.d. comum GPD de parâmetros  $\kappa$  e  $\beta$ . Se  $\kappa < 1$  fixo, então existe uma única solução para  $\frac{\partial l}{\partial \beta} = 0$ . Essa solução é dada por  $\hat{\beta}(\kappa) = \arg \max_{\beta} \{l(\kappa, \beta)\}$ .*

**Demonstração.**

• Caso  $\kappa = 0$ . Aqui, temos que a GPD é a distribuição exponencial com média  $\beta$  e, como já visto antes (Exemplo 3.1), temos que

$$\hat{\beta}(0) = \arg \max_{\beta} \{l(0, \beta)\} = \bar{x},$$

em que  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ .

• Caso  $\kappa < 1$  e  $\kappa \neq 0$ , fixo. Neste caso, de (3.19), temos que

$$\frac{\partial l}{\partial \beta} = \frac{1}{\beta} \left[ -n + (1-\kappa) \sum_{i=1}^n \frac{x_i}{\beta - \kappa x_i} \right]$$

e, da equação de verossimilhança, (3.21), ou seja,  $\frac{\partial l}{\partial \beta} = 0$ , obtemos a equação dada por

$$-n + (1 - \kappa) \sum_{i=1}^n \frac{x_i}{\beta - \kappa x_i} = 0.$$

Agora, calculando

$$\left. \frac{\partial^2 l}{\partial \beta^2} \right|_{\beta=\hat{\beta}(\kappa)} = \left[ \frac{1-\kappa}{\beta} \sum_{i=1}^n \frac{-x_i}{(\beta - \kappa x_i)^2} - \frac{1}{\beta^2} \left( -n + (1-\kappa) \sum_{i=1}^n \frac{x_i}{\beta - \kappa x_i} \right) \right] \Bigg|_{\beta=\hat{\beta}(\kappa)},$$

em que  $\hat{\beta}(\kappa)$  é a solução de (3.21), obtemos

$$\frac{\partial^2 l}{\partial \beta^2}(\kappa, \hat{\beta}(\kappa)) = \frac{1-\kappa}{\hat{\beta}(\kappa)} \sum_{i=1}^n \frac{-x_i}{(\hat{\beta}(\kappa) - \kappa x_i)^2} < 0.$$

Portanto,  $\hat{\beta}(\kappa) = \arg \max_{\beta} \{l(\kappa, \beta)\}$  existe para  $\kappa < 1$  fixo.  $\square$

Os resultados a seguir (Teoremas 3.1 e 3.2 e Proposições 3.3, 3.4 e 3.5) referem-se à GPD de cauda leve, ou seja, quando  $\kappa > 0$ . O caso  $\kappa < 0$  será tratado no Lema 3.1.

**Teorema 3.1.** *Sejam  $x_1, \dots, x_n$  observações de variáveis aleatórias i.i.d. com f.d. comum GPD de parâmetros  $\kappa$  e  $\beta$ . Se  $\kappa > 0$ , então todo máximo local de  $l(\kappa, \beta)$  que satisfaz as equações de verossimilhança para a GPD ocorre para  $0 < \kappa < 1$ .*

**Demonstração.** Seja  $\kappa > 0$ . De (3.11), temos que  $x_i \in [0, \beta/\kappa] \forall i = 1, \dots, n$ . Assim, temos

$$\frac{1}{1 - \kappa x_i / \beta} > 0, \quad \forall i = 1, \dots, n,$$

e

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{1 - \kappa x_i / \beta} > 0.$$

Daí, de (3.22), temos que

$$\frac{\partial l}{\partial \beta} = 0 \iff \frac{1}{1 - \kappa} > 0,$$

ou seja,  $\kappa < 1$ .  $\square$

**Proposição 3.3.** *Seja  $\hat{\beta}(\kappa)$  a função resultante da equação de verossimilhança  $\frac{\partial l}{\partial \beta} = 0$ . A função  $\hat{\beta}(\kappa)$  é contínua, estritamente crescente e diferenciável para  $0 < \kappa < 1$ . Além disso,*

$\bar{x} \leq \widehat{\beta}(\kappa) \leq x_{(n)}$ , e a função estende-se continuamente na fronteira, com

$$\lim_{\kappa \rightarrow 0} \widehat{\beta}(\kappa) = \bar{x} \text{ e } \lim_{\kappa \rightarrow 1} \widehat{\beta}(\kappa) = x_{(n)}. \quad (3.24)$$

**Demonstração.** A continuidade e a diferenciabilidade da função  $\widehat{\beta}(\kappa)$  são obtidas pelo Teorema da Função Implícita aplicado a  $g(\kappa, \beta) = 1/(1 - \kappa) - \frac{1}{n} \sum 1/(1 - \kappa x_i/\beta) = 0$ , definida em  $(0, 1) \times (\bar{x}, x_{(n)})$ .

Devemos provar que as derivadas parciais, dadas abaixo, são diferentes de zero para cada  $\kappa \in (0, 1)$ . São elas

$$g_1(\kappa, \widehat{\beta}(\kappa)) = \frac{\partial g}{\partial \kappa}(\kappa, \widehat{\beta}(\kappa)) = \frac{1}{\kappa} \left( \frac{\kappa}{(1 - \kappa)^2} - \frac{1}{n} \sum_{i=1}^n \frac{\kappa x_i / \widehat{\beta}(\kappa)}{(1 - \kappa x_i / \widehat{\beta}(\kappa))^2} \right)$$

e

$$g_2(\kappa, \widehat{\beta}(\kappa)) = \frac{\partial g}{\partial \beta}(\kappa, \widehat{\beta}(\kappa)) = \frac{1}{\widehat{\beta}(\kappa)^2} \left( \frac{1}{n} \sum_{i=1}^n \frac{\kappa x_i}{(1 - \kappa x_i / \widehat{\beta}(\kappa))^2} \right).$$

É fácil ver que a função  $g_2$  é estritamente positiva para todo  $\kappa \geq 0$ . Devemos agora provar que  $g_1 < 0$ .

Seja  $a$  um real qualquer. Então, desenvolvendo  $(1 - a)^2$ , obtemos a seguinte equivalência

$$\frac{2a}{(1 - a)^2} = \left( \frac{1}{1 - a} \right)^2 + \left( \frac{a}{1 - a} \right)^2 - 1. \quad (3.25)$$

Por outro lado, de (3.22),

$$\frac{1}{1 - \kappa} = \frac{1}{n} \sum_{i=1}^n \frac{1}{1 - \kappa x_i / \widehat{\beta}(\kappa)} \Rightarrow \left( \frac{1}{1 - \kappa} \right)^2 < \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{1 - \kappa x_i / \widehat{\beta}(\kappa)} \right)^2 \quad (3.26)$$

e

$$\frac{\kappa}{1 - \kappa} = \frac{1}{n} \sum_{i=1}^n \frac{\kappa x_i / \widehat{\beta}(\kappa)}{1 - \kappa x_i / \widehat{\beta}(\kappa)} \Rightarrow \left( \frac{\kappa}{1 - \kappa} \right)^2 < \frac{1}{n} \sum_{i=1}^n \left( \frac{\kappa x_i / \widehat{\beta}(\kappa)}{1 - \kappa x_i / \widehat{\beta}(\kappa)} \right)^2. \quad (3.27)$$



Ambas as igualdades são consequências diretas de  $g(\kappa, \hat{\beta}(\kappa)) = 0$ . Então, substituindo (3.26) e (3.27) em (3.25), temos

$$\begin{aligned} 2 \frac{1}{n} \sum_{i=1}^n \frac{\kappa x_i / \hat{\beta}(\kappa)}{(1 - \kappa x_i / \hat{\beta}(\kappa))^2} &= \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{1 - \kappa x_i / \hat{\beta}(\kappa)} \right)^2 + \frac{1}{n} \sum_{i=1}^n \left( \frac{\kappa x_i / \hat{\beta}(\kappa)}{1 - \kappa x_i / \hat{\beta}(\kappa)} \right)^2 - 1 \\ &> \left( \frac{1}{1 - \kappa} \right)^2 + \left( \frac{\kappa}{1 - \kappa} \right)^2 - 1 = \frac{1 + \kappa^2 - 1 - \kappa^2 + 2\kappa}{(1 - \kappa)^2} = \\ &= \frac{2\kappa}{(1 - \kappa)^2}. \end{aligned}$$

Assim,

$$\frac{1}{n} \sum_{i=1}^n \frac{\kappa x_i / \hat{\beta}(\kappa)}{(1 - \kappa x_i / \hat{\beta}(\kappa))^2} > \frac{\kappa}{(1 - \kappa)^2}.$$

Logo,  $g_1 < 0$ , para todo  $\kappa \geq 0$  e  $\kappa \neq 1$ .

Verificaremos agora que  $\hat{\beta}(\kappa)$  é estritamente crescente. Para isso, derivando (3.22) em relação a  $\kappa$ , temos

$$\frac{-1}{(1 - \kappa)^2} = \frac{1}{n} \sum_{i=1}^n \frac{x_i}{(1 - \kappa x_i / \hat{\beta}(\kappa))^2} \frac{\hat{\beta}(\kappa) - \kappa \hat{\beta}'(\kappa)}{\hat{\beta}(\kappa)^2}.$$

Assim,

$$\frac{1}{(1 - \kappa)^2} - \frac{1}{n \hat{\beta}(\kappa)} \sum_{i=1}^n \frac{x_i}{(1 - \kappa x_i / \hat{\beta}(\kappa))^2} = -\frac{\hat{\beta}'(\kappa)}{n} \sum_{i=1}^n \frac{\kappa x_i / \hat{\beta}(\kappa)^2}{(1 - \kappa x_i / \hat{\beta}(\kappa))^2}. \quad (3.28)$$

Mas o lado esquerdo da equação (3.28) é  $g_1$ , já o lado direito é  $-\hat{\beta}'(\kappa)g_2$ . Logo,

$$\hat{\beta}'(\kappa) = -\frac{g_1(\kappa, \hat{\beta}(\kappa))}{g_2(\kappa, \hat{\beta}(\kappa))} > 0,$$

para todo  $\kappa$ , e a função  $\hat{\beta}(\kappa)$  é crescente para todo  $0 \leq \kappa \leq 1$ .

Por fim, vamos mostrar que os limites em (3.24) existem e concluir que  $\hat{\beta}(\kappa)$  é contínua no conjunto  $[0, 1]$ .

• Cálculo do  $\lim_{\kappa \rightarrow 0} \hat{\beta}(\kappa)$ . Tomando o limite nos dois lados da equação em (3.21), temos que

$$\lim_{\kappa \rightarrow 0} \frac{1}{1 - \kappa} = \lim_{\kappa \rightarrow 0} \frac{1}{n} \sum_{i=1}^n \frac{x_i}{\hat{\beta}(\kappa) - \kappa x_i},$$

então,

$$1 = \frac{1}{n} \sum_{i=1}^n \frac{x_i}{\lim_{\kappa \rightarrow 0} \widehat{\beta}(\kappa)}$$

e, portanto,

$$\lim_{\kappa \rightarrow 0} \widehat{\beta}(\kappa) = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}.$$

Por outro lado, temos que  $\widehat{\beta}(0) = \arg \max_{\beta} \{l(0, \beta)\} = \bar{x}$ . Sendo assim, a função  $\widehat{\beta}(\kappa)$  é contínua em 0.

• Cálculo do  $\lim_{\kappa \rightarrow 1} \widehat{\beta}(\kappa)$ . Tomando o limite em (3.22), temos

$$\lim_{\kappa \rightarrow 1} \frac{1}{1 - \kappa} = \lim_{\kappa \rightarrow 1} \frac{1}{n} \sum_{i=1}^n \frac{1}{1 - \kappa x_i / \widehat{\beta}(\kappa)},$$

ou seja,

$$\lim_{\kappa \rightarrow 1} \frac{1}{n} \sum_{i=1}^n \frac{1}{1 - \kappa x_i / \widehat{\beta}(\kappa)} = \infty,$$

Assim,

$$\lim_{\kappa \rightarrow 1} 1 - \frac{\kappa}{\widehat{\beta}(\kappa)} x_{i_0} = 0, \text{ para algum } i_0 \in \{1, \dots, n\},$$

ou seja,

$$\lim_{\kappa \rightarrow 1} \frac{\kappa}{\widehat{\beta}(\kappa)} x_{i_0} = 1$$

e, portanto,

$$\lim_{\kappa \rightarrow 1} \widehat{\beta}(\kappa) = x_{i_0}.$$

Mas  $x_i \leq \beta/\kappa$  para todo  $i \in \{1, \dots, n\}$ , ou seja,

$$x_{i_0} \leq \max\{x_1, \dots, x_n\} = x_{(n)} \leq \lim_{\kappa \rightarrow 1} \frac{\widehat{\beta}(\kappa)}{\kappa} = x_{i_0}.$$

Portanto,

$$\lim_{\kappa \rightarrow 1} \widehat{\beta}(\kappa) = x_{(n)}.$$

Sendo assim, como a função  $\widehat{\beta}(\kappa)$  é contínua, estritamente crescente para todo  $\kappa \geq 0$  e  $\kappa \neq 1$ ,  $\bar{x} \leq \widehat{\beta}(\kappa) \leq x_{(n)}$ , para  $0 \leq \kappa \leq 1$ , e  $\lim_{\kappa \rightarrow 1} \widehat{\beta}(\kappa) = x_{(n)}$ , podemos estender continuamente a função  $\widehat{\beta}(\kappa)$  em  $0 \leq \kappa \leq 1$ , tomando  $\widehat{\beta}(1) = x_{(n)}$ .

Segue, portanto, o resultado. □

**Proposição 3.4.** A função de verossimilhança-profile (3.23) é contínua em  $\{(\kappa, \widehat{\beta}(\kappa)) : 0 \leq \kappa \leq 1\}$ , ou seja,  $l_p(\kappa) = l(\kappa, \widehat{\beta}(\kappa))$  é contínua em  $0 \leq \kappa \leq 1$ . De fato, os limites para  $\kappa \rightarrow 0$  e  $\kappa \rightarrow 1$  correspondem aos valores da log-verossimilhança para as distribuições exponencial e uniforme, respectivamente, e são dados por

$$\lim_{\kappa \rightarrow 0} l_p(\kappa) = -n \ln(\bar{x}) - n = l(0, \bar{x})$$

e

$$\lim_{\kappa \rightarrow 1} l_p(\kappa) = -n \ln(x_{(n)}) = l(1, x_{(n)}).$$

**Demonstração.** Da Proposição 3.3, temos que a função  $\widehat{\beta}(\kappa)$  é contínua em  $\kappa \in [0, 1]$ . Portanto  $l_p(\kappa)$  dada por (3.23) é uma composição de funções contínuas em  $\kappa \in (0, 1)$ . Logo, temos que  $l_p(\kappa)$  é contínua em  $\kappa \in (0, 1)$ . Desenvolvendo (3.23), temos

$$l_p(\kappa) = -n \ln(\widehat{\beta}(\kappa)) + \sum_{i=1}^n \ln \left( 1 - \frac{x_i \kappa}{\widehat{\beta}(\kappa)} \right)^{1/\kappa} - \sum_{i=1}^n \ln \left( 1 - \frac{x_i \kappa}{\widehat{\beta}(\kappa)} \right).$$

• Cálculo do  $\lim_{\kappa \rightarrow 0} l_p(\kappa)$ . Da Proposição 3.3,  $\widehat{\beta}(0) = \lim_{\kappa \rightarrow 0} \widehat{\beta}(\kappa)$ , então

$$\begin{aligned} \lim_{\kappa \rightarrow 0} l_p(\kappa) &= \lim_{\kappa \rightarrow 0} \left[ -n \ln(\widehat{\beta}(\kappa)) + \sum_{i=1}^n \ln \left( 1 - \frac{x_i \kappa}{\widehat{\beta}(\kappa)} \right)^{1/\kappa} - \sum_{i=1}^n \ln \left( 1 - \frac{x_i \kappa}{\widehat{\beta}(\kappa)} \right) \right] = \\ &= -n \ln(\widehat{\beta}(0)) + \sum_{i=1}^n \left( \frac{-x_i}{\widehat{\beta}(0)} \right) = -n \ln \bar{x} - n \frac{\bar{x}}{\bar{x}} = \\ &= -n \ln \bar{x} - n = \\ &= l(0, \bar{x}). \end{aligned}$$

• Cálculo do  $\lim_{\kappa \rightarrow 1} l_p(\kappa)$ . Da Proposição 3.3,  $\widehat{\beta}(1) = \lim_{\kappa \rightarrow 1} \widehat{\beta}(\kappa) = x_{(n)}$ , então

$$\begin{aligned} \lim_{\kappa \rightarrow 1} l_p(\kappa) &= -n \ln(\widehat{\beta}(1)) + \sum_{i=1}^n \ln \left( 1 - \frac{x_i}{\widehat{\beta}(1)} \right) - \sum_{i=1}^n \ln \left( 1 - \frac{x_i}{\widehat{\beta}(1)} \right) = \\ &= -n \ln(x_{(n)}) = \\ &= l(1, x_{(n)}). \end{aligned}$$

□

**Teorema 3.2.** Sejam  $x_1, \dots, x_n$  observações de variáveis aleatórias i.i.d com f.d. comum Pareto generalizada de parâmetros  $\kappa$  e  $\beta$ . Se  $0 \leq \kappa \leq 1$ , então existe o EMV.

**Demonstração.**

Da Proposição 3.2, a função  $\widehat{\beta}(\kappa)$  define o EMV para  $\beta$  para cada  $\kappa$  fixado em  $(0, 1)$ . Dada uma amostra aleatória, para  $\kappa = 0$  e  $\kappa = 1$ , o EMV para o parâmetro  $\beta$  é  $\bar{x}$  e  $x_{(n)}$ , respectivamente. A função  $\widehat{\beta}(\kappa)$  é contínua no conjunto fechado  $[0, 1]$ . O conjunto  $\{(\kappa, \widehat{\beta}(\kappa)) | 0 \leq \kappa \leq 1\}$  é um subconjunto fechado de  $[0, 1] \times [\bar{x}, x_{(n)}]$ , sendo portanto um conjunto compacto.

Pela Proposição 3.4, a função de verossimilhança é contínua em  $\{(\kappa, \widehat{\beta}(\kappa)) | 0 \leq \kappa \leq 1\}$ , logo, existe o EMV global.  $\square$

**Proposição 3.5.** *Sejam  $x_1, \dots, x_n$  observações de variáveis aleatórias i.i.d com f.d. Pareto generalizada de parâmetros  $\kappa$  e  $\beta$ . Se  $\kappa > 1$ , então a função de verossimilhança,  $l(\kappa, \beta)$ , é ilimitada.*

**Demonstração.** Considere  $\kappa > 1$  fixo. Temos que a função de verossimilhança,  $l(\kappa, \beta)$ , é dada por

$$l(\kappa, \beta) = -n \ln \beta + \left(\frac{1}{\kappa} - 1\right) \sum_{i=1}^n \ln \left(1 - \frac{\kappa}{\beta} x_i\right).$$

Agora, note que  $x_{(n)} = \max\{x_1, \dots, x_n\} = x_{i_0}$ , para algum  $i_0 \in \{1, \dots, n\}$ . Sendo assim,

$$\lim_{\beta \rightarrow \kappa x_{(n)}} \left(1 - \frac{\kappa}{\beta} x_{i_0}\right) = 0.$$

Por outro lado,

$$\lim_{\beta \rightarrow \kappa x_{(n)}} (-n \ln \beta) = -n \ln(\kappa x_{(n)})$$

e

$$|-n \ln(\kappa x_{(n)})| < +\infty.$$

Além disso,

$$\lim_{\beta \rightarrow \kappa x_{(n)}} \sum_{i=1}^n \ln \left(1 - \frac{\kappa}{\beta} x_i\right) = -\infty.$$

Portanto, para  $\kappa > 1$ ,

$$\lim_{\beta \rightarrow \kappa x_{(n)}} l(\kappa, \beta) = +\infty,$$

ou seja,  $l(\kappa, \beta)$  é ilimitada.  $\square$

Os resultados vistos referem-se a  $\kappa \geq 0$ . O caso em que  $\kappa < 0$  foi tratado por Kozubowski [3]. A seguir, apresentamos um resultado desse estudo.

**Lema 3.1.** *Sejam  $x_1, \dots, x_n$  observações de variáveis aleatórias i.i.d com f.d. comum Pareto generalizada de parâmetros  $\kappa$  e  $\beta$ . A função log-verossimilhança profile,  $l_p(\kappa)$ , é contínua em  $(-\infty, 0)$  e*

$$\lim_{\kappa \rightarrow -\infty} l_p(\kappa) = -\infty \text{ e } \lim_{\kappa \rightarrow 0^-} l_p(\kappa) = -n(1 + \ln \bar{x}). \quad (3.29)$$

**Demonstração.** A função log-verossimilhança profile é composta de funções contínuas em  $\kappa < 0$ , portanto, é contínua.

• Cálculo do  $\lim_{\kappa \rightarrow -\infty} l_p(\kappa)$ . Da Proposição 3.2, temos que, para  $\kappa < 0$ ,  $\hat{\beta}(\kappa)$  existe e é positivo e, portanto,  $\lim_{\kappa \rightarrow -\infty} \hat{\beta}(\kappa) > 0$ .

Além disso, de (3.22), segue que

$$\lim_{\kappa \rightarrow -\infty} \frac{\hat{\beta}(\kappa) - \kappa x_i}{\hat{\beta}(\kappa)} = +\infty, \quad \forall i = 1, \dots, n.$$

Portanto,

$$\begin{aligned} \lim_{\kappa \rightarrow -\infty} l_p(\kappa) &= \lim_{\kappa \rightarrow -\infty} \left[ -n \ln \hat{\beta}(\kappa) + \left( \frac{1}{\kappa} - 1 \right) \sum_{i=1}^n \ln \left( \frac{\hat{\beta}(\kappa) - \kappa x_i}{\hat{\beta}(\kappa)} \right) \right] = \\ &= -\infty. \end{aligned}$$

• Cálculo do  $\lim_{\kappa \rightarrow 0^-} l_p(\kappa)$ . Tomando o limite de ambos os lados em (3.21), obtemos

$$1 = \frac{1}{n} \sum_{i=1}^n \frac{x_i}{\lim_{\kappa \rightarrow 0^-} \hat{\beta}(\kappa)},$$

ou seja,

$$\lim_{\kappa \rightarrow 0^-} \hat{\beta}(\kappa) = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}.$$

Agora,

$$\begin{aligned} \lim_{\kappa \rightarrow 0^-} l_p(\kappa) &= \lim_{\kappa \rightarrow 0^-} \left[ -n \ln \hat{\beta}(\kappa) + \frac{1}{\kappa} \sum_{i=1}^n \ln \left( 1 - \frac{\kappa x_i}{\hat{\beta}(\kappa)} \right) - \sum_{i=1}^n \ln \left( 1 - \frac{\kappa x_i}{\hat{\beta}(\kappa)} \right) \right] \\ &= \lim_{\kappa \rightarrow 0^-} \left[ -n \ln \hat{\beta}(\kappa) + \sum_{i=1}^n \ln \left( 1 - \frac{\kappa x_i}{\hat{\beta}(\kappa)} \right)^{1/\kappa} \right] \\ &= -n \ln \bar{x} + \sum_{i=1}^n \ln e^{-x_i/\bar{x}} = -n \ln \bar{x} - \sum_{i=1}^n \frac{x_i}{\bar{x}} = -n \ln \bar{x} - n \frac{\bar{x}}{\bar{x}} \\ &= -n \ln \bar{x} - n. \end{aligned}$$

□

**Observação 3.3.**

- (1) Portanto fica claro que sempre existe um máximo global da função de log-verossimilhança profile de  $\theta = (\kappa, \beta)$ , que ocorre para algum  $\kappa \leq 0$ .
- (2) Observe que  $\lim_{\kappa \rightarrow 0^-} \hat{\beta}(\kappa) = \lim_{\kappa \rightarrow 0^+} \hat{\beta}(\kappa) = \bar{x}$ . Além disso,  $\hat{\beta}(0) = \bar{x}$ . Portanto, a função  $\hat{\beta}(\kappa)$  é contínua em  $\kappa = 0$ .

**3.3.2 Matriz de informação de Fisher da GPD**

Não faz parte dos propósitos deste trabalho analisar as propriedades de consistência e normalidade assintótica dos estimadores de máxima verossimilhança dos parâmetros da GPD. Entretanto, nesta subseção, apontaremos algumas referências que tratam do assunto e apresentaremos os cálculos da matriz de informação de Fisher da GPD.

Na literatura, podemos citar Drees [15], que demonstra a consistência e normalidade assintótica dos estimadores de máxima verossimilhança dos parâmetros da GPD quando  $\kappa < 1/2$ ; e Zhou [16], que demonstra a existência e a consistência dos EMVs de  $\kappa$  e  $\beta$  da GPD quando  $\kappa < 1$ .

Em seguida, vamos calcular a matriz de informação. Maiores detalhes podem ser consultados em Arnold [17] e Kozubowski [3].

$I(\theta) = I(\kappa, \beta)$  é dada por

$$I_{ij} = -E \left( \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln g(X, \theta) \right), \quad i, j = 1, 2,$$

em que  $\theta = (\theta_1, \theta_2) = (\kappa, \beta)$  e  $g$  é a função densidade da distribuição Pareto generalizada, dada por

$$g(x; \kappa, \beta) = \frac{1}{\beta} \left( 1 - \frac{\kappa}{\beta} x \right)^{1/\kappa - 1}, \quad x \in D(\kappa, \beta),$$

$$\text{em que, } D(\kappa, \beta) = \begin{cases} [0, \infty), & \text{se } \kappa \leq 0, \\ [0, \beta/\kappa], & \text{se } \kappa > 0. \end{cases}$$

Assim, temos

$$\ln g(x) = \ln \left( \frac{1}{\beta} \left( 1 - \frac{\kappa x}{\beta} \right)^{\frac{1}{\kappa} - 1} \right) = -\ln \beta - \left( 1 - \frac{1}{\kappa} \right) \ln \left( 1 - \frac{\kappa x}{\beta} \right).$$

Portanto,

$$\frac{\partial \ln g(x)}{\partial \kappa} = -\frac{1}{\kappa^2} \ln \left(1 - \frac{\kappa x}{\beta}\right) + \left(1 - \frac{1}{\kappa}\right) \frac{x}{\beta - \kappa x},$$

e

$$\frac{\partial \ln g(x)}{\partial \beta} = -\frac{1}{\beta} - \left(1 - \frac{1}{\kappa}\right) \frac{\kappa x}{\beta^2 - \kappa \beta x}.$$

Assim,

$$\begin{aligned} (i) \quad \frac{\partial^2 \ln g(x)}{\partial \kappa^2} &= \frac{2}{\kappa^3} \ln \left(1 - \frac{\kappa x}{\beta}\right) + \frac{2}{\kappa^2} \frac{x}{\beta - \kappa x} + \left(1 - \frac{1}{\kappa}\right) \frac{x^2}{(\beta - \kappa x)^2} \\ (ii) \text{ e } (iii) \quad \frac{\partial^2 \ln g(x)}{\partial \beta \partial \kappa} &= \frac{\partial^2 \ln g(x)}{\partial \kappa \partial \beta} = -\frac{x}{\kappa(\beta^2 - \kappa \beta x)} - \left(1 - \frac{1}{\kappa}\right) \frac{x}{(\beta - \kappa x)^2} \\ (iv) \quad \frac{\partial^2 \ln g(x)}{\partial \beta^2} &= \frac{1}{\beta^2} + \left(1 - \frac{1}{\kappa}\right) \frac{\kappa x(2\beta - \kappa x)}{(\beta^2 - \kappa \beta x)^2}. \end{aligned}$$

Calculando a esperança, em (i) a (iv), obtemos

$$E \left[ \frac{\partial^2 \ln g(X)}{\partial \kappa^2} \right] = E \left[ \frac{2}{\kappa^3} \ln \left(1 - \frac{\kappa X}{\beta}\right) + \frac{2}{\kappa^2} \frac{X}{\beta - \kappa X} + \left(1 - \frac{1}{\kappa}\right) \frac{X^2}{(\beta - \kappa X)^2} \right].$$

Calculando termo a termo, temos:

• 1º termo

$$\begin{aligned} E \left[ \frac{2}{\kappa^3} \ln \left(1 - \frac{\kappa X}{\beta}\right) \right] &= \frac{2}{\kappa^3} \int_0^\infty \ln \left(1 - \frac{\kappa x}{\beta}\right) \frac{1}{\beta} \left(1 - \frac{\kappa x}{\beta}\right)^{\frac{1}{\kappa}-1} dx \\ &= \frac{2}{\kappa^3} \int_1^\infty \ln u \frac{1}{\beta} u^{\frac{1}{\kappa}-1} \frac{(-\beta)}{\kappa} du = -\frac{2}{\kappa^3} \int_1^\infty \frac{1}{\kappa} u^{\frac{1}{\kappa}-1} \ln u du \\ &= -\frac{2}{\kappa^3} \left[ u^{1/\kappa} \ln u - \kappa u^{1/\kappa} \right] \Big|_1^\infty \stackrel{\kappa \leq 0}{=} -\frac{2}{\kappa^2}. \end{aligned}$$

O cálculo para o caso  $0 < \kappa < 1/2$  é feito de forma similar.

• Semelhantemente os dois outros termos são

$$E \left[ \frac{2}{\kappa^2} \frac{X}{\beta - \kappa X} \right] = \frac{2}{\kappa^2(1 - \kappa)}$$

e

$$E \left[ \left(1 - \frac{1}{\kappa}\right) \frac{X^2}{(\beta - \kappa X)^2} \right] = -\frac{2}{\kappa(1 - 2\kappa)}.$$

Finalmente,

$$\begin{aligned} E \left[ \frac{\partial^2 \ln g(X)}{\partial \kappa^2} \right] &= -\frac{2}{\kappa^2} + \frac{2}{\kappa^2(1-\kappa)} - \frac{2}{\kappa(1-2\kappa)} = -\frac{2\kappa^2}{\kappa^2(1-\kappa)(1-2\kappa)} \\ &= -\frac{2}{(1-\kappa)(1-2\kappa)}. \end{aligned}$$

Procedendo de maneira similar, temos

$$E \left[ \frac{\partial^2 \ln g(x)}{\partial \kappa \partial \beta} \right] = E \left[ \frac{\partial^2 \ln g(x)}{\partial \beta \partial \kappa} \right] = -\frac{1}{\beta(1-\kappa)(1-2\kappa)}.$$

$$E \left[ \frac{\partial^2 \ln g(x)}{\partial \beta^2} \right] = -\frac{1}{\beta^2(1-2\kappa)}.$$

Logo a matriz de informação é dada por

$$I(\kappa, \beta) = \left[ -E \left( \frac{\partial^2 \ln g(X)}{\partial \theta_i \partial \theta_j} \right) \right]_{i,j=1,2} = \frac{1}{(1-\kappa)(1-2\kappa)\beta^2} \begin{bmatrix} 2\beta^2 & \beta \\ \beta & 1-\kappa \end{bmatrix}. \quad (3.30)$$

Conforme Drees [15], o estimador de máxima verossimilhança de  $\theta = (\kappa, \beta)$  da distribuição Pareto generalizada é assintoticamente normal quando  $\kappa < 1/2$ . Assim, para  $\kappa < 1/2$ , temos que

$$\sqrt{n}(\hat{\theta} - \theta) \rightarrow N(\mathbf{0}, [I(\theta)]^{-1}),$$

em que  $N$  representa a distribuição normal bivariada e  $\mathbf{0}$  é o vetor nulo do  $\mathbb{R}^2$ .

## 3.4 Testes de hipóteses

Nesta seção, apresentaremos dois testes de hipóteses que serão utilizados no Capítulo 4 para verificar alguns resultados deste trabalho. Serão vistos o teste de seleção de modelos pelo critério de informação Akaike (AIC - *Akaike information criterion*) e o teste  $\chi^2$  para seleção de hipóteses.

### 3.4.1 Critério de informação Akaike (AIC)

Para uma breve descrição do critério de informação Akaike (AIC - *Akaike information criterion*), suponha que certa densidade  $f$  denote a verdade, ou realidade completa da



distribuição de uma variável aleatória em estudo. No entanto,  $f$  é desconhecida, portanto pretende-se aproximar a distribuição  $f$  por um modelo. Suponha que essa aproximação será feita por  $g(\cdot; \theta)$ , com  $\theta = (\theta_1, \dots, \theta_k)$ , então o valor da divergência, conhecida como divergência de Kullback-Leibler (K-L), para um modelo  $g(\cdot; \theta)$ , em relação ao modelo “real” representado por  $f$ , é

$$I(f, g; \theta) = \int f(x) \ln \frac{f(x)}{g(x; \theta)} dx. \quad (3.31)$$

Esta divergência está relacionada à informação perdida por se usar um modelo aproximado em vez do “real”. Assim, o objetivo aqui é encontrar o modelo que perde menos informação, ou seja, o objetivo é minimizar  $I(f, g)$  com relação a  $g$ . Uma vez que  $f$  é fixa, pois é a densidade verdadeira que se quer aproximar, ela não varia nem mesmo com o tamanho da amostra. Por outro lado,  $g$  varia conforme  $\theta$ .

Um cálculo direto resulta em

$$I(f, g) = \int f(x) \ln f(x) dx - \int f(x) \ln g(x; \theta) dx,$$

ou seja,

$$I(f, g) = E(\ln f(X)) - E(\ln g(X; \theta)).$$

Mas, embora desconhecida, a esperança  $E(\ln f(X))$  é fixa.

A estimativa do AIC para um determinado modelo é dada pela definição a seguir (maiores detalhes sobre o AIC podem ser vistos em Akaike [18]).

**Definição 3.9.** *O critério de informação Akaike é dado por*

$$AIC = -2 \ln(\hat{L}) + 2r, \quad (3.32)$$

em que  $\hat{L}$  é o valor máximo da função de verossimilhança de  $\theta$  e  $r$  é o número de parâmetros.

O modelo com o menor valor de AIC é considerado o modelo de melhor ajuste.

**Exemplo 3.5.** Seja uma amostra aleatória  $x_1, \dots, x_n$  observada de uma distribuição desconhecida. Suponha que se queira modelar por uma distribuição exponencial com parâmetro  $\lambda$  desconhecido. Primeiramente, encontramos o estimador de máxima verossimilhança de  $\lambda$ , dado por  $\hat{\lambda} = 0,0102$ . Substituindo esse valor na função de log-verossimilhança (3.5) e calculando o AIC, de (3.32), para o modelo exponencial temos que

$$AIC_E = -2 \ln(\hat{L}_E) + 2K = -2 \times (-111,68) + 2 \times 1 = 225,36.$$

Agora, deseja-se modelar os mesmos dados pela distribuição Pareto generalizada. Dessa maneira, formulamos as hipóteses:

- $H_0$ : a amostra observada vem da distribuição exponencial com parâmetro  $\hat{\lambda}$ ;
- $H_1$ : a amostra segue uma distribuição Pareto generalizada com parâmetros  $\hat{\kappa}$  e  $\hat{\beta}$ .

Estimando os parâmetros da distribuição Pareto generalizada via máxima verossimilhança, para o mesmo conjunto de dados, obtemos  $\hat{\kappa} = 0,56$  e  $\hat{\beta} = 51,56$ . Assim, de (3.13), obtemos o valor da função log-verossimilhança calculada em  $(\hat{\kappa}, \hat{\beta})$ , ou seja, no ponto de máximo. Daí temos o AIC para o modelo Pareto generalizada,

$$\begin{aligned} AIC_{GPD} &= -2 \ln(\hat{L}_{GPD}) + 2K = -2 \times (-110,06) + 2 \times 2 = 224,12 \\ &< 225,36 = AIC_E. \end{aligned}$$

Portanto, pelo teste AIC rejeita-se  $H_0$ , ou seja, ele favorece o modelo da distribuição Pareto generalizada para esse conjuntos de dados.

**Observação 3.4.** Os dados utilizados no Exemplo 3.5 referem-se aos tempos de construção de usinas eólicas, levantados pela ANEEL (Agência Nacional de Energia Elétrica). Eles serão apresentados no Capítulo 4.

### 3.4.2 Teste de seleção $\chi^2$

Nesta subseção, introduziremos o teste  $\chi^2$  para variáveis aleatórias contínuas. Para maiores detalhes, veja Rohatgi [19].

Seja  $X$  uma variável aleatória contínua com distribuição desconhecida. Suponha que tenha sido observada uma amostra aleatória,  $x_1, \dots, x_n$ , dessa distribuição. O objetivo é introduzir uma distribuição e testar se os dados provêm dela. Mais formalmente, considere  $f$  uma densidade e a hipótese nula formulada da seguinte forma:

$H_0$ : a amostra observada provém da distribuição  $f$ .

Considere os intervalos  $I_1, \dots, I_k$  uma partição do conjunto dos reais. Sejam  $p_j = P_f(X \in I_j)$ ,  $j = 1, \dots, k$ , e  $\hat{p}_j$  a proporção observada de valores da amostra que pertencem a  $I_j$ . Tem-se que, sob  $H_0$ , para  $n$  (tamanho amostral) suficientemente grande, a estatística

$$Q = \sum_{j=1}^k \frac{(n\hat{p}_j - np_j)^2}{np_j} \quad (3.33)$$

tem aproximadamente uma distribuição  $\chi^2$  com  $k - 1$  graus de liberdade. Dessa maneira, o teste consiste em rejeitar  $H_0$ , ao nível de significância  $\alpha$ ,  $0 < \alpha < 1$ , se o valor observado de

$Q$ ,  $q_0$ , é maior do que  $\chi_{k-1, \alpha}^2$ , valor este que satisfaz  $P(Z > \chi_{k-1, \alpha}^2) = \alpha$ , em que  $Z$  é uma v.a. com distribuição  $\chi^2$  com  $k - 1$  graus de liberdade.

Tem-se também que, dada  $H_0$ ,  $P_{valor} = P(Q \geq q_0)$ , que é o menor nível de significância com que se rejeita  $H_0$ , ou seja, rejeita-se  $H_0$  se  $P_{valor} < \alpha$ .

**Exemplo 3.6** (Processo de Poisson). Suponha que seja dada uma amostra observada  $(y_1, \dots, y_n)$  de uma distribuição desconhecida e que desejamos testar se essa amostra provém de um Processo de Poisson com taxa  $\lambda = 5$ . Sabemos que o Processo de Poisson é um processo de contagem que consiste de números de ocorrências de um determinado evento, no qual os tempos entre essas ocorrências são variáveis aleatórias i.i.d. com distribuição exponencial de parâmetro  $\lambda$ . Dessa maneira, consideremos  $n = 100$  valores observados de um processo de contagem (ver Figura 3.1), em que  $Y$  denota a variável número de ocorrências,  $X_1$  o tempo do primeiro evento e  $X_n$ ,  $n > 1$ , o tempo entre o  $(n - 1)$ -ésimo e  $n$ -ésimo evento. Da amostra observada de  $Y$ , extraímos os valores observados  $x_1, \dots, x_n$  de  $X$  (tempo entre ocorrências) e formulamos o teste considerando a seguinte hipótese nula:

- $H_0$ : a amostra observada,  $x_1, \dots, x_n$ , provém de uma distribuição exponencial com parâmetro  $\lambda = 5$ .

Como  $P(X > 0) = 1$ , escolhemos os intervalos  $I_1, \dots, I_k$  como uma partição do intervalo  $(0, +\infty)$ . Essa escolha foi feita de maneira que:

$$k = 8 \text{ e } p_j = P(X \in I_j) = 1/8, \quad \forall j = 1, \dots, 8.$$

(para uma discussão sobre o número de classes, veja Rohatgi [19]).

Assim, para a distribuição exponencial, em que  $P(X \leq x) = 1 - e^{-\lambda x}$  e  $\lambda = 5$ , temos

$$P(X \leq x) = 1 - e^{-5x} = \frac{i}{8} \Rightarrow x = -\frac{\ln(1 - \frac{i}{8})}{5}, \quad i = 1, \dots, 7,$$

em que  $i$  define o limite de cada intervalo, conforme apresentado na Tabela 3.1.

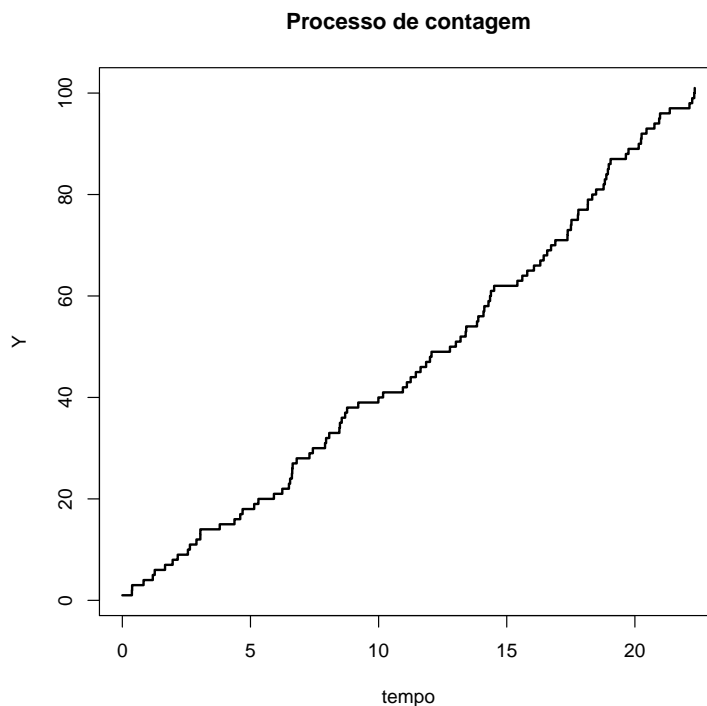
Em seguida, encontra-se o número de observações em cada intervalo e calcula-se  $q_0$ , conforme (3.33). Todos os resultados estão detalhados na Tabela 3.1.

O valor de  $q_0$  é 3,52, e há  $k - 1 = 8 - 1 = 7$  graus de liberdade. Calculando o  $P_{valor}$ , temos, sob  $H_0$ , que

$$P(Q \geq 3,52) = 0,83$$

Portanto, com uma probabilidade de 83%, é bem razoável concluir que não se rejeita  $H_0$ .

**Observação 3.5.** Os dados do Exemplo 3.6 foram gerados no software R. Justamente foi gerado um processo de Poisson com uma taxa  $\lambda = 5$ . Contudo, nem sempre o parâmetro da



**Figura 3.1:** Exemplo de processo de contagem.

**Tabela 3.1:** Foram escolhidos 8 intervalos para um total de  $n=100$  medidas.

| Intervalos                | $n\hat{p}_j$ | $np_j$ | $\frac{(n\hat{p}_j - np_j)^2}{n\hat{p}_j}$ |
|---------------------------|--------------|--------|--|
| $I_1$ [0; 0,0267)         | 11           | 12,5   | 0,18                                       |
| $I_2$ [0,0267; 0,0575)    | 14           | 12,5   | 0,18                                       |
| $I_3$ [0,0575; 0,0940)    | 14           | 12,5   | 0,18                                       |
| $I_4$ [0,0940; 0,1386)    | 11           | 12,5   | 0,18                                       |
| $I_5$ [0,1386; 0,1962)    | 12           | 12,5   | 0,02                                       |
| $I_6$ [0,1962; 0,2773)    | 16           | 12,5   | 0,98                                       |
| $I_7$ [0,2773; 0,4159)    | 14           | 12,5   | 0,18                                       |
| $I_8$ [0,4159; $\infty$ ) | 8            | 12,5   | 1,62                                       |
| Soma                      | 100          | 100    | $q_0 = 3,52$                               |

distribuição é conhecido. Na verdade, nem mesmo a distribuição, que é o objeto do teste, é

conhecida. Assim, o caso de maior interesse não é se os dados podem ser modelados por uma distribuição específica, mas se eles podem ser modelados por uma família de distribuições.

Seja  $X$  uma v.a. com distribuição definida pelo parâmetro  $\theta$  (desconhecido). O objetivo é verificar se para algum valor de  $\theta$  a distribuição se ajusta bem. Uma maneira é encontrar  $\theta$  que minimiza a função  $Q = Q(\theta)$ .

**Definição 3.10.** *Seja  $\hat{\theta}$  o valor de  $\theta$  para o qual  $Q(\theta)$  é mínimo. Então,  $\hat{\theta}$  é chamado estimador mínimo de  $\theta$  para a  $\chi^2$ , denotado por  $\hat{\theta} = \arg \min_{\theta} \{Q(\theta)\}$ .*

Temos o seguinte resultado (para mais detalhes, ver Rohatgi [19]).

**Teorema 3.3.** *Seja  $\tilde{\theta}$  o estimador mínimo de  $\theta$ . Então, sob certas condições de regularidade e para o número de valores observados,  $n$ , grande, o valor mínimo  $Q(\tilde{\theta})$  de  $Q(\theta)$  dado por*

$$Q(\tilde{\theta}) = \sum_{j=1}^k \frac{(n_j - np_j(\tilde{\theta}))^2}{np_j(\tilde{\theta})}, \quad (3.34)$$

tem aproximadamente a distribuição  $\chi^2$  com  $k - 1 - r$  graus de liberdade, em que  $n_j$  é o número de valores observados em cada intervalo,  $I_j$ ,  $j = 1, \dots, k$ ,  $\sum_{j=1}^k n_j = n$ ,  $p_j(\tilde{\theta}) = P(X \in I_j)$  e  $r$  é a dimensão de  $\theta$ .

**Observação 3.6.** As condições de regularidade não especificadas no Teorema 3.3 são satisfeitas para a maioria das distribuições. Contudo, não é simples estimar  $\theta$  minimizando  $Q(\theta)$ . Por outro lado, é possível mostrar que o Teorema 3.3 é válido para o estimador de máxima verossimilhança (para mais detalhes, veja Rohatgi [19]).

**Exemplo 3.7.** Considerando o Exemplo 3.6, temos interesse em encontrar um valor para  $\lambda$  a partir dos dados, em vez de utilizar um valor fixo como foi feito naquele exemplo. Dessa forma, a hipótese será reformulada da seguinte maneira:

- $H_0$ : a amostra observada do tempo entre ocorrências segue uma distribuição exponencial de parâmetro  $\lambda = \hat{\lambda}$ ,

em que  $\hat{\lambda}$  é o estimador de máxima verossimilhança de  $\lambda$ .

Conforme a Observação 3.6, o Teorema 3.3 é válido para o estimador de máxima verossimilhança. Assim, temos que o estimador de máxima verossimilhança para  $\lambda$  é  $\hat{\lambda} = 5,36$ . Portanto, é possível construir a Tabela 3.2 com o mesmo número de intervalos do Exemplo 3.6, utilizando, agora, o valor de  $\hat{\lambda}$ , ou seja,

$$P(X \leq x) = 1 - e^{-5,36x} = \frac{i}{8} \Rightarrow x = -\frac{\ln(1 - \frac{i}{8})}{5,36}, \quad i = 1, \dots, 7.$$

Assim, temos que  $k - 1 - r = 8 - 1 - 1 = 6$  (graus de liberdade). Dessa forma, o  $P_{valor}$ , sob  $H_0$ , é

$$P(Q \geq 2,88) = 0,82,$$

em que  $q_0 = 2,88$  é o valor observado de  $Q(\hat{\theta})$ . Portanto, o teste  $\chi^2$  aceita  $H_0$  com 82% de confiança.

**Tabela 3.2:** Foram escolhidos 8 intervalos para um total de  $n=100$  medidas.

| Intervalos                | $n_j$ | $np_j(\hat{\theta})$ | $\frac{(n_j - np_j(\hat{\theta}))^2}{np_j(\hat{\theta})}$ |
|---------------------------|-------|----------------------|---|
| $I_1$ [0; 0,0249)         | 11    | 12,5                 | 0,18  |
| $I_2$ [0,0249; 0,0537)    | 14    | 12,5                 | 0,18  |
| $I_3$ [0,0537; 0,0878)    | 10    | 12,5                 | 0,5   |
| $I_4$ [0,0878; 0,1294)    | 13    | 12,5                 | 0,02  |
| $I_5$ [0,1294; 0,1831)    | 11    | 12,5                 | 0,18  |
| $I_6$ [0,1831; 0,2588)    | 17    | 12,5                 | 1,62  |
| $I_7$ [0,2588; 0,3882)    | 13    | 12,5                 | 0,02  |
| $I_8$ [0,3882; $\infty$ ) | 11    | 12,5                 | 0,18  |
| Soma                      | 100   | 100                  | $q_0 = \mathbf{2,88}$                                     |

**Observação 3.7.** Na prática o mais comum é definir previamente o nível de confiança  $\alpha$ , em geral 5%. Assim, para este valor, no Exemplo 3.7,

$$\chi_{6;5\%}^2 = 12,59 > 2,88 = q_0.$$

Dado que  $q_0$  é menor, então a hipótese  $H_0$  deve ser aceita.

**Observação 3.8.** O teste da  $\chi^2$  é usado para comparar um primeiro modelo para um determinado conjunto de dados. Observe que, uma vez que os dados foram gerados por uma rotina no software R, obteve-se um valor de 82%, que é um bem mais alto do que se obtém, em geral, com dados reais.

É costume o uso de níveis mais baixos de confiança, em torno de 5%. Portanto, esse teste costuma ser favorável à distribuição escolhida. Melhores modelos podem ser comparados posteriormente entre si com outros testes, como, por exemplo, o AIC, visto na Seção 3.4.1.

## CAPÍTULO

### 4

# SIMULAÇÕES E APLICAÇÕES PRÁTICAS

## 4.1 Introdução

O objetivo deste capítulo é apresentar alguns resultados numéricos obtidos do estudo teórico nos capítulos anteriores. Além disso, será feita uma aplicação prática com dados reais.

Na Seção 4.2, será estudado o estimador de máxima verossimilhança para os parâmetros  $\kappa$  e  $\beta$  da distribuição Pareto generalizada. As simulações foram feitas no software R (versão x64 3.5.1), utilizando o algoritmo de estimação proposto por Castillo e Serra [1].

Em seguida, serão geradas amostras para analisar a consistência, a normalidade assintótica e os vícios dos parâmetros  $\kappa$  e  $\beta$ . Por fim, será estudado o comportamento da função log-verossimilhança profile para  $\kappa \leq 0$ , onde já foi visto (Lema 3.1) que os estimadores de máxima verossimilhança de  $\kappa$  e  $\beta$  sempre existem.

Na Seção 4.3, será visto um exemplo prático da utilização da teoria de valor extremo. Trata-se do tempo de construção de usinas eólicas no Brasil. Os dados foram levantados pela Superintendência de Fiscalização da Geração da ANEEL (Agência Nacional de Energia Elétrica). Assim, serão inferidos três parâmetros da distribuição Pareto generalizada: o limiar

$t$ , os parâmetros de forma  $\kappa$  e de escala  $\beta$ . Esse estudo pode ser utilizado na previsão de construção para novas usinas do setor elétrico. Será feita também a discussão de seleção de parâmetros por meio do critério de informação Akaike (AIC), definido na Seção 3.4.1. Por fim, será analisada a propriedade (d) do Teorema 2.2 por meio do teste da  $\chi^2$ .

## 4.2 Simulações numéricas

Nesta seção, serão avaliados o estimadores de máxima verossimilhança dos parâmetros  $\kappa$  e  $\beta$ , calculados por meio do algoritmo 4.1, proposto no artigo Castillo e Serra [1].

Em (3.13) foi apresentada a função log-verossimilhança em função dos parâmetros  $\kappa$  e  $\beta$ , entretanto, do ponto de vista computacional e para efeito de simulações, é mais eficiente utilizar o parâmetro  $\sigma = \beta/\kappa$ . Portanto, denotaremos a nova função de log-verossimilhança por  $l^*$ , de modo que  $l^*(\kappa, \sigma) = l(\kappa, \kappa\sigma)$ , em que  $\sigma < 0$ , para  $\kappa < 0$  e  $\sigma > x_{(n)}$ , para  $\kappa > 0$ .

Com essa nova parametrização, obtemos a log-verossimilhança profile encontrando  $\hat{\kappa} = \hat{\kappa}(\sigma)$  das equações de verossimilhança. Assim, de (3.13),  $l^*(\kappa, \sigma)/\partial\kappa = 0$  é dada por

$$\begin{aligned} \frac{\partial l^*(\kappa, \sigma)}{\partial \kappa} &= \frac{\partial}{\partial \kappa} l(\kappa, \kappa\sigma) = \frac{\partial}{\partial \kappa} \left[ -n \ln(\kappa\sigma) + \left(\frac{1}{\kappa} - 1\right) \sum_{i=1}^n \ln\left(1 - \frac{x_i}{\sigma}\right) \right] = 0 \\ &= -\frac{n}{\kappa} - \frac{1}{\kappa^2} \sum_{i=1}^n \ln\left(1 - \frac{x_i}{\sigma}\right) = 0. \end{aligned}$$

Assim,

$$n = -\frac{1}{\kappa} \sum_{i=1}^n \ln\left(1 - \frac{x_i}{\sigma}\right).$$

Dessa forma,

$$\hat{\kappa}(\sigma) = -\frac{1}{n} \sum_{i=1}^n \ln\left(1 - \frac{x_i}{\sigma}\right). \quad (4.1)$$

Substituindo (4.1) em (3.13), obtemos

$$l_p(\sigma) = n \left( -\ln(\hat{\kappa}(\sigma)\sigma) + \hat{\kappa}(\sigma) - 1 \right). \quad (4.2)$$

O Algoritmo 4.1 a seguir foi apresentado por Castillo e Serra [1] e é baseado na log-verossimilhança profile. Ele será utilizado para as simulações e para obtenção dos estimadores dos parâmetros  $\kappa$  e  $\beta$ .



**Algoritmo 4.1.**

```
#to estimate the maximum likelihood (MLE) of a sample x by GPD(k,psi)
eGPD<-function(x){
  fk<-function(sigma) -mean(log(1-x/sigma))
  fp<-function(sigma) length(x)*(-log(fk(sigma)*sigma)+fk(sigma)-1)
  int<-c(-100*max(x),100*max(x))
  sigma<-optimize(fp,interval=int,maximum=T)$maximum
  list(k=fk(sigma),psi=fk(sigma)*sigma)}
```

**Observação 4.1.**

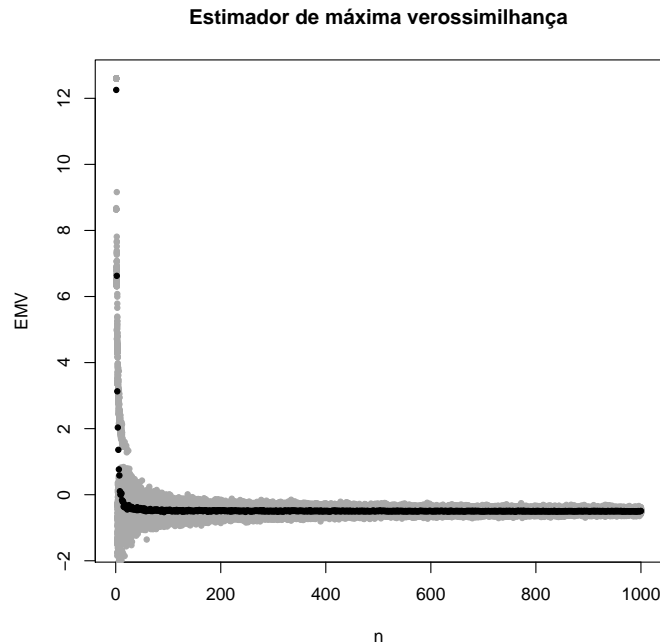
- Em Castillo e Serra [1], o parâmetro  $\beta$  é chamado  $\psi$ .
- As funções  $fk$  e  $fp$  são (4.1) e (4.2), respectivamente.
- A função *optimize()* faz parte do pacote *stats* do software R.
- Para se utilizar esse algoritmo, pressupõe-se o conhecimento do limiar  $t$

**Aplicação 4.1** (Consistência do EMV). Com o intuito de observar a consistência do EMV para o parâmetro  $\kappa$  da GPD (lembrando que a consistência é definida como a convergência em probabilidade de  $\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| < \varepsilon) = 1$ ), foram geradas 100 amostras para cada tamanho de 1 a 1.000, com  $\kappa = -1/2 \in (-\infty, 1/2)$ , região em que o EMV de  $\kappa$  é consistente e assintoticamente normal. O resultado está apresentado na Figura 4.1.

Observa-se que o erro para amostras muito pequenas é grande, porém o EMV de  $\kappa$  se aproxima do valor verdadeiro,  $\kappa = -1/2$ , para  $n$  suficientemente grande. Os pontos em preto são a média, e, em cinza, a dispersão para cada uma das 100 amostras.

**Aplicação 4.2** (Estimativa do parâmetro  $\kappa$  da GPD). Nesta aplicação, será testado o Algoritmo 4.1 para diferentes valores de  $\kappa$ . Para isso, foram geradas 1.000 amostras de tamanho 50, com  $\kappa = \{-6; -6 + 0,01; -6 + 0,02; \dots; +6\}$ , semelhantemente ao que Castillo e Serra [1] fizeram. O resultado está apresentado na Figura 4.2.

A escolha do tamanho da amostra foi feita baseada em um valor comum na prática. Quando se trabalha com valores extremos, os dados são mais raros. Assim, o objetivo desta aplicação é verificar a validade da estimação dos parâmetros no caso em que a amostra tem um tamanho médio.



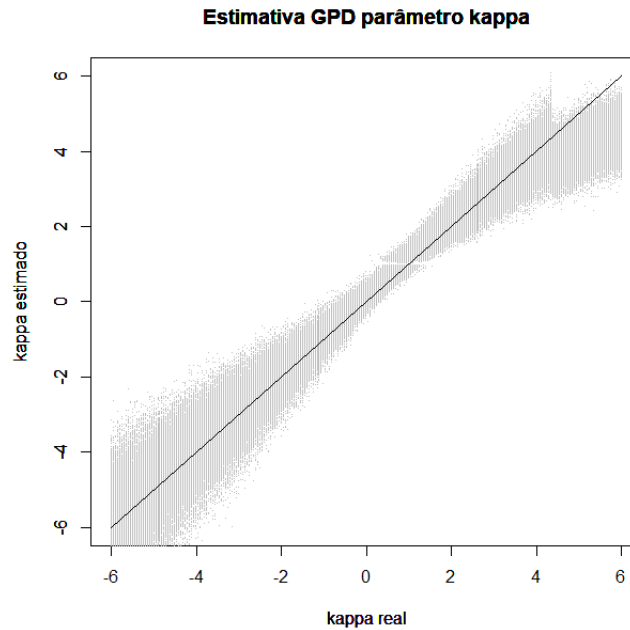
**Figura 4.1:** Foram geradas 100 amostras para cada tamanho de 1 a 1.000 e calculado o EMV do parâmetro  $\kappa$ .

Note que há problemas na estimação de  $\kappa$  em torno de 1. Isso ocorre pois a GPD é muito heterogênea em torno de  $\kappa = 1$ , onde o suporte da GPD é finito, ou seja, região onde o ponto extremo superior é  $\omega(G_{\kappa,\beta}(x)) = \beta/\kappa$ . Quando  $0 < \kappa < 1$ , a função densidade calculada no ponto extremo é  $g(\beta/\kappa) = 0$ . Em  $\kappa = 1$ , ou seja, quando a GPD é a distribuição uniforme, a densidade tem valor constante igual a 1. Quando  $\kappa > 1$ , a densidade da Pareto tende a infinito. Esse comportamento pode ser conferido na Figura 4.3, em que é apresentada a função densidade da GPD, (3.12), para  $\kappa = \beta$ , ou seja,  $\sigma = 1$ , assim o suporte é  $x \in [0, \beta/\kappa] = [0, 1]$ .

Além disso, como visto no Capítulo 3, os estimadores de máxima verossimilhança dos parâmetros  $\kappa$  e  $\beta$  não existem quando  $\kappa \geq 1$ . Porém, o Algoritmo 4.1, baseado no princípio de máxima verossimilhança, é útil para identificar, em uma amostra, quando  $\kappa \geq 1$ , para então, a partir de outro método de inferência, estimar os parâmetros  $\kappa$  e  $\beta$ .

Assim, Castillo e Serra [1] propõem que seja classificada previamente a região de  $\kappa$  em três modelos distintos para estudo do EMV: **A**, em que  $\kappa \leq 0$ ; **B**, em que  $0 \leq \kappa \leq 1$ ; e **C**, em que  $\kappa \geq 1$ .

**Aplicação 4.3** (Vício dos estimadores dos parâmetros  $\kappa$  e  $\beta$  da GPD). Nesta aplicação, será analisado o vício dos estimadores de máxima verossimilhança dos parâmetros  $\kappa$  e  $\beta$  da GPD, com base na inferência pelo Algoritmo 4.1. Foram geradas 10.000 amostras de tamanho 500 e 50, com parâmetros  $\beta = 1$  e  $\kappa = \{-4; -4 + 0,1; -4 + 0,2; \dots; 4\}$ . Embora



**Figura 4.2:** No eixo horizontal é representado o parâmetro  $\kappa$  utilizado para o gerar os valores da distribuição Pareto generalizada. Foram geradas 1.000 amostras de tamanho 50,  $\kappa = \{-6; -6 + 0,01; -6 + 0,02; \dots; +6\}$ . No eixo vertical são representados os valores estimados.

amostras de tamanho elevado não sejam comuns em modelos de valor extremo, o propósito desta aplicação é observar o comportamento do Algoritmo 4.1 para inferência via máxima verossimilhança. Por outro lado, amostras de tamanho 50 são mais comuns e podem ser comparadas em termos de vício conforme a Figura 4.4.

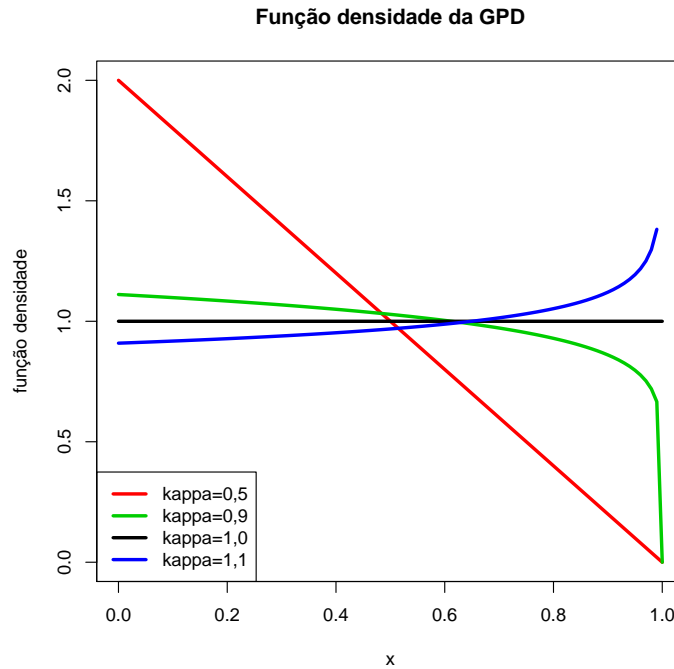
A aplicação a seguir é realizada para verificar as condições do Lema 3.1 para  $\kappa \leq 0$ . Porém, como já discutido, do ponto de vista computacional, é mais eficiente utilizar a parametrização da GPD com  $\sigma = \beta/\kappa$ . Dessa forma, os limites (3.29) são dados por

$$\lim_{\sigma \rightarrow 0^-} l_p(\sigma) = -\infty \text{ e } \lim_{\sigma \rightarrow -\infty} l_p(\sigma) = -n(1 + \ln \bar{x}). \quad (4.3)$$

em que  $l_p(\sigma)$  é a log-verossimilhança profile com parâmetro  $\sigma$ , dada por (4.2).

**Aplicação 4.4.** Foram geradas 2 amostras de tamanho 10.000; uma para a GPD com parâmetros  $\kappa = -1/2$  e  $\beta = 1$ , e outra para a distribuição exponencial com parâmetro  $\lambda = \frac{1}{\beta} = 1$ , lembrando que essa distribuição é um caso particular da GPD com  $\kappa = 0$ . Assim, os limites (4.3) teóricos para a GPD são

$$\lim_{\sigma \rightarrow 0^-} l_p(\sigma) = -\infty$$



**Figura 4.3:** Função densidade da GPD para valores de  $\kappa > 0$ , em torno a 1. Foi tomado  $\kappa = \beta$ , ou seja,  $\sigma = 1$ , portanto,  $x \in [0, 1]$

e, conforme o Teorema 2.2 (a),

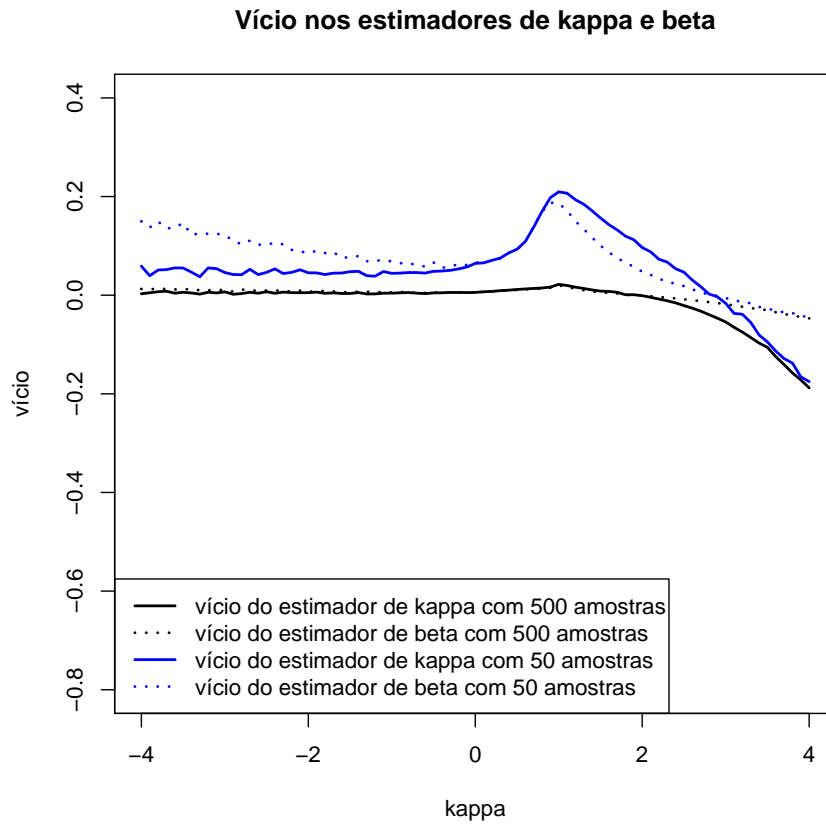
$$\lim_{\sigma \rightarrow -\infty} l_p(\sigma) = -n \left( 1 + \ln \frac{\beta}{1 + \kappa} \right) = -10.000(1 + \ln 2) = -16.931.$$

No caso da exponencial,

$$\lim_{\sigma \rightarrow 0^-} l_p(\sigma) = -\infty \text{ e } \lim_{\sigma \rightarrow -\infty} l_p(\sigma) = -10.000.$$

Como o objetivo é avaliar a teoria para valores altos de  $n$ , nota-se na Figura 4.5 que a função log-verossimilhança profile assume valores bem próximos aos limites teóricos (4.3). Porém, para amostras menores, a diferença entre os valores teóricos e os valores calculados pode ser maior.

No caso da distribuição exponencial, pode ser observado que a função log-verossimilhança profile é sempre menor que  $-n(1 + \ln \bar{x})$  e tende a esse valor quando  $\sigma$  tende a  $-\infty$ . No caso da GPD com  $\kappa < 0$ , a função assume valores maiores do que  $-n(1 + \ln \bar{x})$ , e da mesma maneira que a exponencial, tende a esse valor quando  $\sigma$  tende a  $-\infty$ . Entretanto, o máximo da log-verossimilhança profile ocorre para um  $\sigma$  finito. No caso da Aplicação 4.4, esse valor é  $\sigma = \beta/\kappa = 1/(-1/2) = -2$ .



**Figura 4.4:** Vício na inferência pelo Algoritmo 4.1 dos parâmetros  $\kappa$  e  $\beta$ . Foram geradas 10.000 amostras de tamanho 500 e 50, com parâmetros  $\beta = 1$  e  $\kappa = \{-4; -4 + 0,1; -4 + 0,2; \dots; 4\}$ .

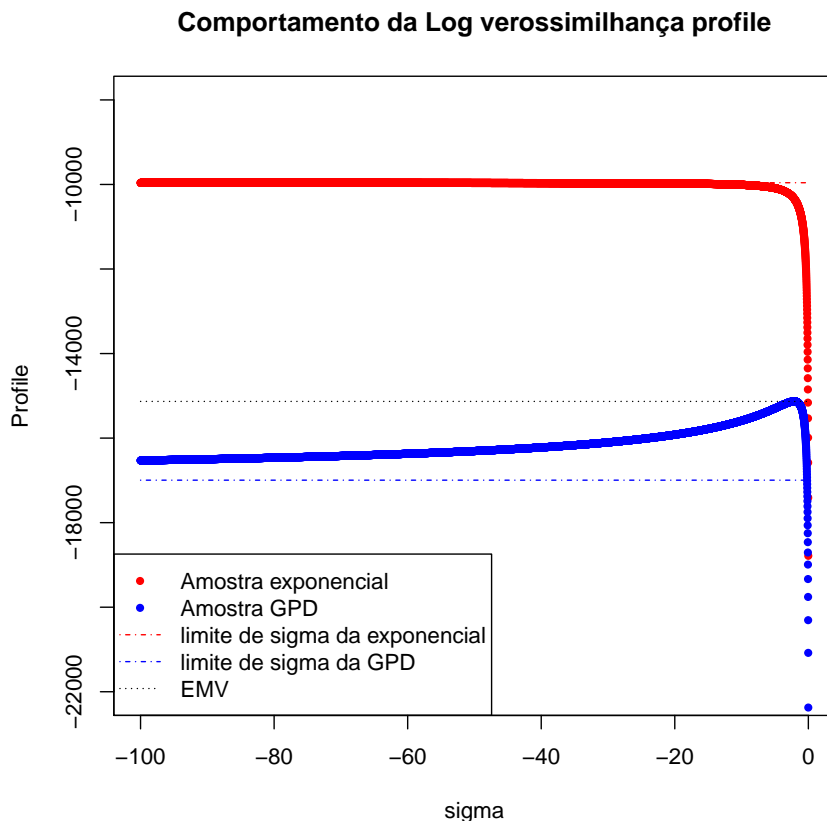
**Aplicação 4.5** (Propriedade assintótica da GPD). Lembrando a definição de normalidade assintótica, que é válida para a GPD com  $\kappa < 1/2$ , temos, de (3.9), que

$$Y_n = \sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} Z,$$

em que  $Z$  é uma variável aleatória com distribuição  $N(0, [I(\theta)]^{-1})$  e  $\theta = (\kappa, \beta)$ .

Para esta aplicação, tomaremos  $\kappa = -1/2$  e  $\beta = 4$ . Dessa forma, de (3.30), a matriz de informação da GPD é dada por

$$I(\kappa, \beta) = \frac{1}{(1 - \kappa)(1 - 2\kappa)\beta^2} \begin{bmatrix} 2\beta^2 & \beta \\ \beta & 1 - \kappa \end{bmatrix} = \frac{1}{48} \begin{bmatrix} 32 & 4 \\ 4 & 1,5 \end{bmatrix}$$



**Figura 4.5:** Função log-verossimilhança profile.

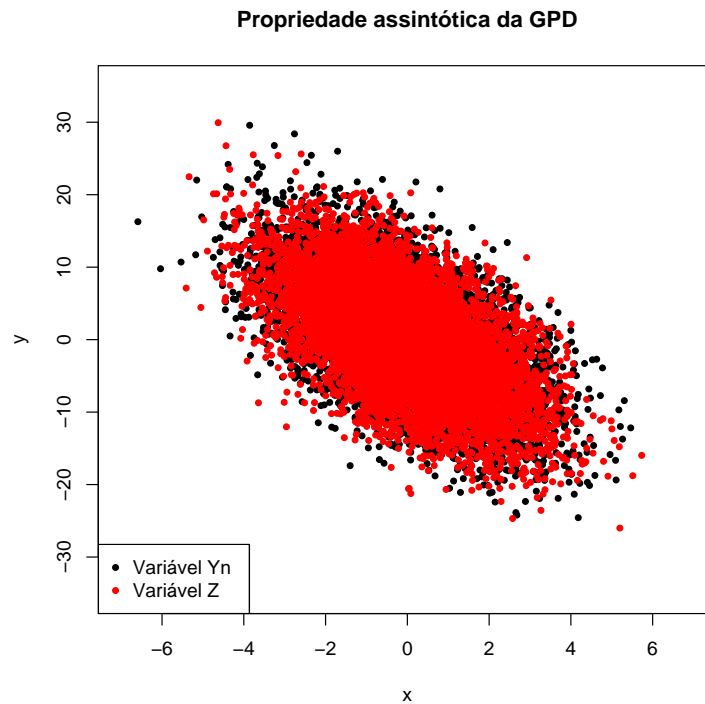
Assim, foram geradas 10.000 amostras para a variável aleatória  $Y_n = \sqrt{n}(\hat{\theta}_n - \theta)$ , de tamanho  $n = 10.000$  e  $\theta = (\kappa, \beta) = (-1/2, 4)$ , e 10.000 amostras para a variável aleatória  $Z$ . O resultado está apresentado na Figura 4.6.

Para efeitos de visualização, a fim de observar a diferença entre as variáveis aleatórias  $Y_n$  e  $Z$ , foi gerado a partir dessas variáveis o gráfico das densidades respectivas aproximadas pela função *kde2d* do pacote *MASS* do software R. O resultado está apresentado na Figura 4.7.

A diferença entre as duas densidades está apresentada na Figura 4.8.

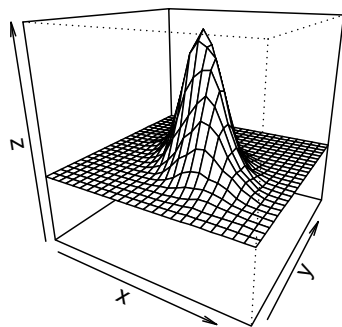
Nota-se que, embora possa ser percebido que a distribuição de  $Y_n$  está próxima da distribuição de  $Z$ , há certa discrepância mesmo com amostras grandes de tamanho 10.000.

**Aplicação 4.6.** Com o intuito de estudar a proposta de Castillo e Serra [1] de dividir  $\kappa$  nas regiões **A**:  $\kappa \leq 0$ ; **B**:  $0 \leq \kappa \leq 1$ ; e **C**:  $\kappa \geq 1$ , apresentamos também, na Tabela 4.1, as porcentagens de classificação de cada região para valores próximos às fronteiras em  $\kappa = 0$  e

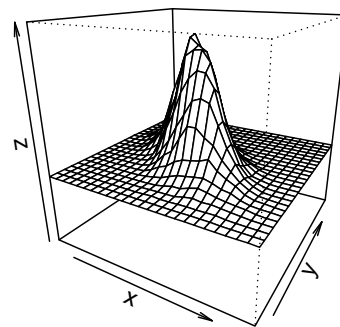


**Figura 4.6:** Propriedade assintótica da GPD, com  $\sqrt{(n)}(\hat{\kappa} - \kappa)$  no eixo  $x$  e  $\sqrt{(n)}(\hat{\beta} - \beta)$  no eixo  $y$ .

**Densidade da variável aleatória  $Y_n$**

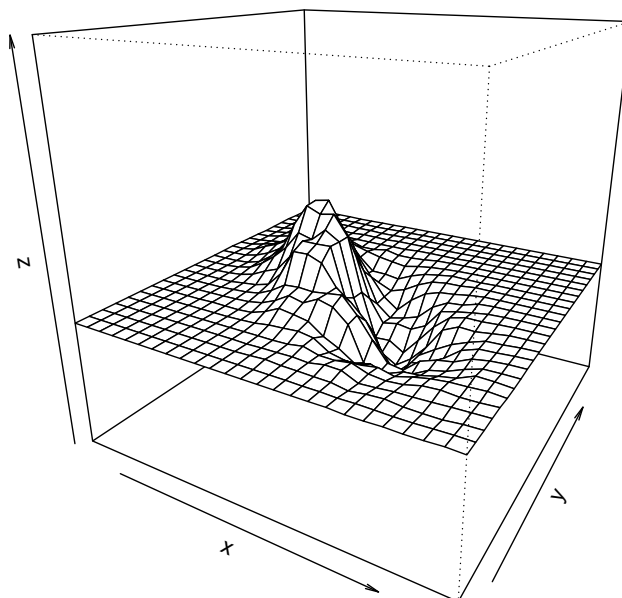


**Densidade da variável aleatória  $Z$**



**Figura 4.7:** Densidade das variáveis  $Y_n = \sqrt{n}(\hat{\theta}_n - \theta)$  e  $Z = N(0, [I(\theta)]^{-1})$ .

$\kappa = 1$ , para alguns tamanhos de amostras especificados. Foram geradas 1.000 amostras para cada percentagem.

Diferença entre as variáveis aleatórias  $Y_n$  e  $Z$ 

**Figura 4.8:** Diferença entre as densidades de  $Y_n = \sqrt{n}(\hat{\theta}_n - \theta)$  e  $Z = N(0, [I(\theta)]^{-1})$ .

**Tabela 4.1:** Porcentagem de classificação de cada região para alguns tamanhos de amostras e para  $\kappa$  próximo às fronteiras entre cada região, ou seja, **A**, para  $\kappa \leq 0$ ; **B**, para  $0 \leq \kappa \leq 1$ ; e **C**, para  $\kappa \geq 1$ . Foram geradas 1.000 amostras para obtenção de cada porcentagem. O valor em negrito corresponde à correta classificação.

|     | A           |      |     | B    |             |      |     |             |      | C   |     |             |
|-----|-------------|------|-----|------|-------------|------|-----|-------------|------|-----|-----|-------------|
|     | -0,1        |      |     | 0,1  |             |      | 0,9 |             |      | 1,1 |     |             |
| n   | A           | B    | C   | A    | B           | C    | A   | B           | C    | A   | B   | C           |
| 15  | <b>40,5</b> | 50,4 | 9,1 | 18,1 | <b>66,8</b> | 15,1 | 0,1 | <b>11,6</b> | 88,3 | 0,0 | 4,2 | <b>95,8</b> |
| 25  | <b>47,5</b> | 51,4 | 0,9 | 15,8 | <b>82,0</b> | 2,2  | 0,0 | <b>16,1</b> | 83,9 | 0,0 | 5,1 | <b>94,9</b> |
| 30  | <b>53,9</b> | 46,0 | 0,1 | 17,4 | <b>82,3</b> | 0,3  | 0,0 | <b>20,3</b> | 79,7 | 0,0 | 5,7 | <b>94,3</b> |
| 50  | <b>60,6</b> | 39,4 | 0,0 | 16,2 | <b>83,8</b> | 0,0  | 0,0 | <b>30,1</b> | 69,9 | 0,0 | 5,3 | <b>94,7</b> |
| 100 | <b>75,6</b> | 24,4 | 0,0 | 8,0  | <b>92,0</b> | 0,0  | 0,0 | <b>50,3</b> | 49,7 | 0,0 | 3,9 | <b>96,1</b> |
| 500 | <b>96,8</b> | 3,2  | 0,0 | 0,4  | <b>99,6</b> | 0,0  | 0,0 | <b>96,0</b> | 4,0  | 0,0 | 0,6 | <b>99,4</b> |



Observa-se que o valor correto do modelo escolhido tende a 100% conforme o aumento do tamanho da amostra. Por outro lado, o objetivo é fazer a inferência para os parâmetros da GPD a partir de valores extremos, entretanto, é difícil conseguir elevado número de dados. Por exemplo, se o objetivo for a máxima anual de chuva, para a obtenção de 25 dados extremos são necessários 25 anos de medições. Assim, a importância do estudo dessa teoria é grande, porém, na prática, a dificuldade reside nos casos com amostras pequenas e médias.

### 4.3 Aplicação com dados reais

Será apresentado um caso real a ser modelado pela teoria de valores extremos. Os dados referem-se ao tempo de construção de usinas eólicas e foram levantados pela Superintendência de Fiscalização dos Serviços de Geração da ANEEL.

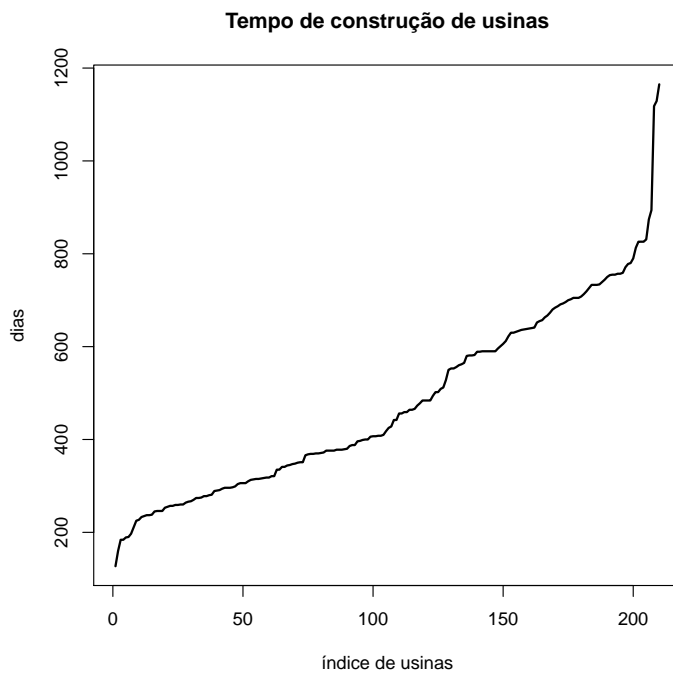
Algumas usinas levam um tempo consideravelmente maior para serem construídas que a média histórica do setor. Por outro lado, não há um limiar já determinado do que é um tempo longo para construção. Assim, tem-se interesse em estimar primeiramente esse limiar, para então verificar a modelagem da GPD.

A importância de uma abordagem estatística para esses dados se dá na sua utilização para realizar previsões para novos empreendimentos eólicos. A previsão do tempo de construção de usinas é utilizada para o planejamento de novos Leilões de Energia Elétrica e para cálculo do Custo Marginal de Operação (CMO), cujo dado é fundamental para operação do Sistema Elétrico Brasileiro.

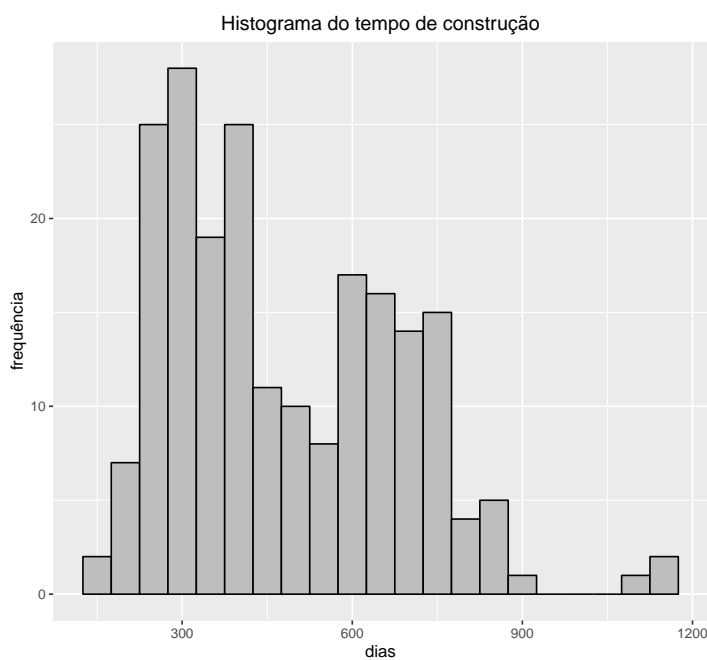
Na Figura 4.9, temos o tempo de construção de 210 usinas eólicas, e na Figura 4.10 temos o histograma do tempo de construção das mesmas usinas. A mais rápida concluiu a construção em torno de 100 dias. Já a mais lenta levou quase 1.200 dias para terminar.

Com o intuito de obter uma primeira aproximação do limiar para a GPD, utilizaremos a função média do excesso (2.6). Conforme vimos no Teorema 2.2 (e),  $e(t)$  é linear em  $t$  (2.17). Assim na Figura 4.11, pode ser observado que o valor limiar, que é aquele considerado a partir do qual a função média do excesso tende a ser linear, é 750 dias. Observamos que esse valor, num primeiro momento, pode ser subjetivo. Assim, é importante analisar os dados em torno desse limiar.

Com uma variação relativamente pequena, de 5 dias, observa-se, pela Figura 4.11, que há grande variação entre as retas azul e vermelha, obtidas da expressão teórica (2.17). Isso se dá em grande parte devido à pequena quantidade de valores extremos obtidos dos dados. Como já salientado nos capítulos precedentes, a importância e a dificuldade dessa modelagem ocorrem para amostras de tamanhos médios e pequenos.

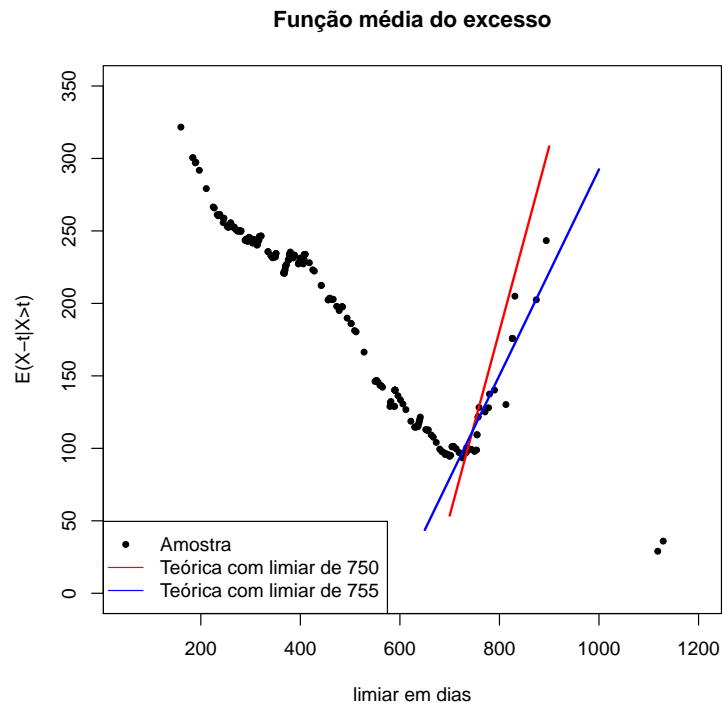


**Figura 4.9:** Tempo de construção de usinas eólicas. Há 210 no total.

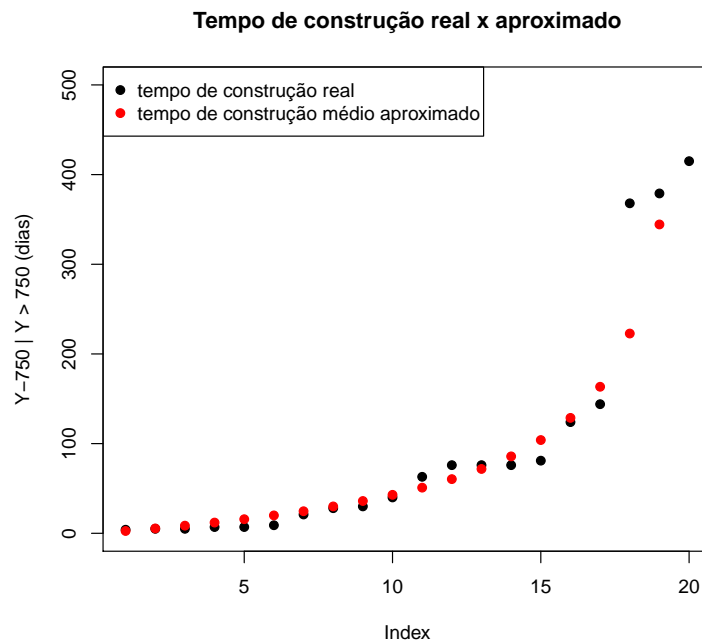


**Figura 4.10:** Histograma do tempo de construção das 210 usinas eólicas.

Na Figura 4.12, é apresentada a comparação entre os dados reais e a aproximação pela GPD com limiar  $t = 750$  e parâmetros  $\kappa = -0,560219$  e  $\beta = 51,56165$ , inferidos pelo Algoritmo 4.1.



**Figura 4.11:** Função média do excesso para o tempo de construção de usinas eólicas.



**Figura 4.12:** Tempo de construção real e aproximado pela GPD.

Daí pode ser observado que, apesar das limitações, a aproximação é razoavelmente boa entre o modelo teórico e os dados reais.

A Tabela 4.2 mostra os testes de seleção para a modelagem dos dados de usinas eólicas. O valor encontrado pela análise gráfica para o limiar foi 750 dias. Assim, foram feitos testes para valores próximos de 750, conforme a coluna limiar na tabela. A coluna  $n$ , na Tabela 4.2, corresponde ao tamanho da amostra utilizada na modelagem para cada limiar escolhido. Assim, se o limiar for por exemplo 755, há 17 valores acima desse limiar, considerados os pontos extremos. Os parâmetros  $\kappa$  e  $\beta$  foram inferidos pelo Algoritmo 4.1 para cada limiar escolhido. Conforme a Definição 3.9, foram calculados o  $AIC_{GPD}$  e  $AIC_E$ , critérios de informação Akaike relativos à GPD e exponencial, respectivamente. Lembrando, de (3.32),  $AIC = -2 \ln(\hat{L}) + 2r$ , em que  $r$  é o número de parâmetros da distribuição. Temos que  $r = 2$  no caso da GPD, e  $r = 1$  no caso da exponencial. Por fim, a coluna *Preferência*, na Tabela 4.2, é o resultado final do teste.

Conforme discutido na Seção 3.4.1, escolhe-se o modelo com menor  $AIC$ . Assim, na Tabela 4.2, o  $AIC$  foi calculado como  $AIC = AIC_{GPD} - AIC_E$ , ou seja, se  $AIC < 0$ , então  $AIC_{GPD} < AIC_E$  e, então, escolhe-se a distribuição Pareto generalizada. Caso contrário, se  $AIC > 0$ , escolhe-se a distribuição exponencial.

**Tabela 4.2:** Seleção de modelo para EOL.

| limiar | n  | $\kappa$ | $\beta$ | $AIC_E$ | $AIC_{GPD}$ | AIC   | Preferência |
|--------|----|----------|---------|---------|-------------|-------|-------------|
| 740    | 22 | -0,39    | 63,58   | 248,06  | 247,81      | -0,25 | GPD         |
| 750    | 20 | -0,56    | 51,56   | 225,36  | 224,12      | -1,24 | GPD         |
| 755    | 17 | -0,42    | 69,17   | 195,62  | 196,17      | 0,55  | Exp         |
| 760    | 14 | -0,13    | 110,78  | 165,68  | 167,57      | 1,88  | Exp         |
| 770    | 14 | -0,36    | 79,18   | 163,39  | 164,60      | 1,21  | Exp         |

Por fim, o Teorema 2.2 (d) diz que, se  $N$  segue uma distribuição Poisson de parâmetro  $\lambda$ , independente da sequência de variáveis aleatórias i.i.d.  $\{X_n\}$  com GDP de parâmetros  $\xi$  e  $\beta$ , e  $Z_N = \max(X_1, \dots, X_N)$ , então,

$$P(Z_N \leq x) = \exp \left\{ -\lambda \left( 1 - \kappa \frac{x}{\beta} \right)^{1/\kappa} \right\} = H_{\kappa, \mu, \psi}(x),$$

em que  $\mu = \beta \xi^{-1} (\lambda \xi - 1)$  e  $\psi = \beta \lambda \xi$ .

Para modelar os dados *tempos de construção de usinas eólicas* utilizando a GPD, não foi verificado se  $N$  segue uma distribuição Poisson. Vamos terminar esse capítulo verificando

essa propriedade. Da teoria de processos estocásticos, temos que um processo de contagem de um determinado evento é um processo de Poisson com taxa  $\lambda > 0$  se, e somente se, os tempos entre ocorrências são variáveis aleatórias i.i.d. com distribuição exponencial de parâmetro  $\lambda$  (média  $1/\lambda$ ). Utilizaremos essa propriedade e o teste  $\chi^2$  para verificar se  $N$  pode ser considerada uma variável aleatória com distribuição Poisson, semelhantemente ao que foi feito no Exemplo 3.7.

Utilizando o limiar 750 e a hipótese

- $H_0$ : a amostra observada segue a distribuição exponencial com parâmetro  $\lambda = \hat{\lambda}$ .

**Tabela 4.3:** Foram escolhidos 5 intervalos para um total de 19 medidas.

| Intervalos             | $n_j$ | $np_j(\hat{\lambda})$ | $\frac{(n_j - np_j(\hat{\lambda}))^2}{np_j(\hat{\lambda})}$ |
|------------------------|-------|-----------------------|---|
| $I_1$ [0, 14)          | 10    | 3,8                   | 10,12   |
| $I_2$ [14, 32)         | 4     | 3,8                   | 0,01  |
| $I_3$ [32, 57)         | 2     | 3,8                   | 0,85  |
| $I_4$ [57, 100)        | 2     | 3,8                   | 0,85  |
| $I_5$ [100, $\infty$ ) | 1     | 3,8                   | 2,06  |
| Soma                   | 19    | 19                    | $q_0 = \mathbf{13,89}$                                      |

O valor encontrado para  $q_0$  foi 13,89.

O valor de  $\chi_{k-1-r;\alpha}^2$  é obtido conforme a Seção 3.4. Assim, com 5 intervalos e 1 parâmetro, há  $5 - 1 - 1 = 3$  graus de liberdade. Com nível de significância,  $\alpha$ , de 5%, encontra-se  $\chi_{3;5\%}^2 = 7,815$ , e, portanto, menor que  $q_0$ . Assim, o teste  $\chi^2$  rejeita a hipótese  $H_0$ .

Por outro lado, analisando os dados, observa-se que um dos pontos ocorre num tempo muito superior aos demais. Dessa forma, refazendo a mesma análise sem considerar esse ponto, obtém-se o resultado apresentado na Tabela 4.4.

Neste caso,  $q_0 = 4,67 < 7,815 = \chi_{3;5\%}^2$ . Situação em que o teste  $\chi^2$  aceita a hipótese  $H_0$ .

Uma análise estatística mais rigorosa faz-se necessária para verificar os limites da aplicação do modelo. No entanto, isso não faz parte do escopo deste trabalho, ficando isso para um estudo futuro. O objetivo desse exemplo era utilizar as ferramentas teóricas em um caso prático e verificar como os resultados se aplicam a esses dados.

**Tabela 4.4:** Foram escolhidos 5 intervalos para um total de 18 medidas.

| Intervalos    | $n_j$ | $np_j(\hat{\lambda})$ | $\frac{(n_j - np_j(\hat{\lambda}))^2}{np_j(\hat{\lambda})}$ |
|---------------|-------|-----------------------|---|
| $I_1$ [0,6)   | 7     | 3,6                   | 3,21  |
| $I_2$ [6,14)  | 3     | 3,6                   | 0,1   |
| $I_3$ [14,25) | 3     | 3,6                   | 0,1   |
| $I_4$ [25,45) | 2     | 3,6                   | 0,71  |
| $I_5$ [45,∞)  | 5     | 3,6                   | 0,54  |
| Soma          | 18    | 18                    | $q_0 = 4,67$  |

# BIBLIOGRAFIA

- [1] J. Castillo and I. Serra. Likelihood inference for generalized pareto distribution. *Computational Statistics and Data Analysis*, 83:116–128, 2015.
- [2] J. Castillo and J. Daoudi. Estimation of the generalized pareto distribution. *Statist. Probab. Lett.*, 79:684–688, 2009.
- [3] T.J. et al. Kozubowski. Testing exponentiality versus pareto distribution via likelihood ratio. *Communications in Statistics*, 38:1:118–139, 2009.
- [4] J. Galambos. *The Asymptotic Theory of Extreme Order Statistics*. John Wiley & Sons, 1978.
- [5] P. Embrechts, C. Kluppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, 1997.
- [6] R. A. Fisher and L.H.C. Tippett. Limiting forms of the frequency distributions of the largest or smallest number of a sample. *Proc. Cambridget Philos. Soc.*, 24:180–190, 1928.
- [7] M. Fréchet. Sur la loi de probabilité de l'écart maximum. *Ann. de la Soc. polonaise de Math*, 6:93, 1927.
- [8] B.V. Gnedenko. Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.*, 44:423–453, 1943.
- [9] L. De Haan. A form of regular variation and its application to the domain of attraction of double exponential distribution. *Z. Wahrscheinlichkeitstheorie verw. Geb.*, 17:241–258, 1971.
- [10] J. Pickands. Statistical inference using extreme order statistics. *The Annals of Statistics*, pages 119–131, 1975.

- 
- [11] E. Castillo and A.S. Hadi. Taylor & francis. *ASTIN Bullets*, 92:1609–1620, 1997.
- [12] J. Zhang and M.A. Stephens. A new and efficient estimation method for the generalized pareto distribution. *Technometrics*, 51:316–325, 2009.
- [13] H. Cramer. *Mathematical methods of statistics*. Princeton University Press, Princeton, 1946.
- [14] E. L. Lehmann. *Elements of Large-Sample Theory*. Springer, New York, 1983.
- [15] H. Drees, A. Ferreira, and L. Haan. On maximum likelihood estimation of the extreme value index. *The annals of applied probability*, 14:1179–1201, 2004.
- [16] C. Zhou. Existence and consistency of the maximum likelihood estimator for the extreme value index. *Journal of Multivariate Analysis*, 100:794–815, 2009.
- [17] B. C. Arnold. *Pareto Distributions*. Taylor & Francis Group, 2 edition, 2015.
- [18] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–723, 1974.
- [19] V. K. Rohatgi. *Statistical Inference*. Dover Publications, Inc., 2003.