



Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Estatística

Dissertação de Mestrado

Distribuição Gumbel Bimodal: propriedades e estimação

por

Eduarda Bahiense Machado da Silva

Brasília, 19 de novembro de 2021

Distribuição Gumbel Bimodal: propriedades e estimação

por

Eduarda Bahiense Machado da Silva

Dissertação apresentada ao Departamento de Estatística da Universidade de Brasília como requisito parcial para obtenção do título de Mestre em Estatística.

Orientadora: Profa. Dra. Cira Etheowalda Guevara Otiniano

Brasília, 19 de novembro de 2021

Dissertação submetida ao Programa de Pós-Graduação em Estatística do Departamento de Estatística da Universidade de Brasília como parte dos requisitos para a obtenção do grau de Mestre em Estatística.

Texto aprovado por:

Profa. Dra. Cira Etheowalda Guevara Otiniano
Orientador, EST/UnB

Prof. Dr. Raul Yukihiro Matsushita
EST/UnB

Prof. Dr. Eduardo Nakano
EST/UnB

Prof. Dr. Fernando Ferraz do Nascimento
EST/UFPI

É necessário sair da ilha para ver a ilha, não nos vemos se não saímos de nós.

(José Saramago)

Para os meus meninos.

Meus mais profundos agradecimentos à professora Cira Etheowalda Guevara Otiniano por todo apoio, paciência e compreensão.

Agradeço também a toda minha família e a todos que de alguma forma me deram forças pra chegar até o fim.

Resumo

A modelagem de eventos extremos, observações máximas ou mínimas, é altamente relevante em diversas áreas do conhecimento, como hidrologia, finanças e atuária. A distribuição Gumbel, uma das três distribuições que surge na teoria de valores extremos, é a mais utilizada para modelar esse tipo de dado. Diversas generalizações dessa distribuição buscam modelar os dados com mais de uma moda, dentre elas as misturas de Gumbel. Assim, com o objetivo de modelar dados de eventos extremos com comportamento heterogêneo bimodal, neste trabalho é proposta uma nova generalização da distribuição Gumbel com apenas três parâmetros, denominada Gumbel Bimodal. As principais medidas descritivas, como moda, quantis e momentos, foram estudadas. Também foi realizado um estudo inferencial, com utilização de simulação e posterior aplicação em dados reais.

Palavras-Chave: Valores extremos; distribuição Gumbel; Gumbel Bimodal.

Abstract

The modeling of extreme events, maximum and minimum observations, is highly relevant in several areas of knowledge, such as hydrology, finance and actuarial. The Gumbel distribution, one of the three distributions that appears in the theory of extreme values, is the most used to model this type of data. Several generalizations of this distributions can be found in the literature, including the mixtures of Gumbel in order to model data with more than one mode. Thus, in order to model data from extreme events with heterogeneous bimodal behavior, this work proposes a new generalization of the Gumbel distribution with only three parameters, called Gumbel Bimodal. The main descriptive measures, such as mode, quantiles and moments, were studied. Also an inferential study was carried out, using simulation and subsequent application to real data.

Palavras-Chave: Extreme values; Gumbel distribution; Bimodal Gumbel.

Conteúdo

1	Introdução	1
2	Distribuição Gumbel Bimodal	5
2.1	Conceitos Preliminares	5
2.2	Modelo Gumbel Bimodal	7
2.3	Estudo gráfico	9
2.4	Moda	10
2.5	Momentos	11
2.5.1	Função tipo geratriz de momentos	12
2.5.2	Função tipo geradora de cumulantes	14
2.6	Quantis	17
3	Estimação	19
3.1	Estimação por máxima verossimilhança	19
4	Simulação	23
4.1	Simulação por Monte Carlo	23
5	Aplicação	31
5.1	Metodologia	31
5.2	Análise Descritiva	32

5.3	Blocos Máximos	36
5.4	Cálculo do VaR	39
6	Conclusão	45

Lista de Tabelas

4.1	Estimativas via Monte Carlo (n=50)	25
4.2	Estimativas via Monte Carlo (n=100)	26
4.3	Estimativas via Monte Carlo (n=1000)	27
5.1	Resumo dos dados - Dólar	36
5.2	Resumo dos dados - Petróleo	36
5.3	Estimativa dos parâmetros - Petróleo	38
5.4	Estimativa dos parâmetros - Dólar	38
5.5	Estimativa dos parâmetros - Petróleo	41
5.6	Estimativa dos parâmetros - Dólar	43

Lista de Figuras

2.1	Densidade de $X \sim F_G(\cdot; \mu, \sigma)$	7
2.2	Densidade de $X \sim F_{GB}(\cdot; 0, 1, \delta)$	9
2.3	Densidade de $X \sim F_{GB}(\cdot; \mu, 1, \delta)$	10
2.4	Densidade de $X \sim F_{GB}(\cdot; 0, \sigma, \delta)$	10
4.1	Comparativo para $X \sim F_{GB}(\cdot; \mu = -1, \sigma = 1, \delta = 0)$	28
4.2	Comparativo para $X \sim F_{GB}(\cdot; \mu = -1, \sigma = 1, \delta = 2)$	28
4.3	Comparativo para $X \sim F_{GB}(\cdot; \mu = -1, \sigma = 1, \delta = 4)$	28
4.4	Comparativo para $X \sim F_{GB}(\cdot; \mu = 0, \sigma = 1, \delta = 0)$	29
4.5	Comparativo para $X \sim F_{GB}(\cdot; \mu = 0, \sigma = 1, \delta = 2)$	29
4.6	Comparativo para $X \sim F_{GB}(\cdot; \mu = 0, \sigma = 1, \delta = 4)$	29
4.7	Comparativo para $X \sim F_{GB}(\cdot; \mu = 1, \sigma = 1, \delta = 0)$	30
4.8	Comparativo para $X \sim F_{GB}(\cdot; \mu = 1, \sigma = 1, \delta = 2)$	30
4.9	Comparativo para $X \sim F_{GB}(\cdot; \mu = 1, \sigma = 1, \delta = 4)$	30
5.1	Série histórica - variação do preço da ação do petróleo em dólar	33
5.2	Série histórica - variação do preço da ação do petróleo em dólar	33
5.3	Distribuição de log-retornos - Dólar	34
5.4	Distribuição de log-retornos - Petróleo	34
5.5	Histograma de log-retornos - Dólar	35

5.6	Histograma de log-retornos - Petróleo	35
5.7	Histogramas de sub-amostras - Petróleo	37
5.8	Sub-amostras - Petróleo	37
5.9	Histogramas de sub-amostras - Dólar	38
5.10	Sub-amostras - Dólar	38
5.11	Densidade ajustada para os blocos - Petróleo	39
5.12	Densidade ajustada para os blocos - Dólar	39
5.13	Bimodalidade das caudas: $\alpha = 5\%$ - Petróleo	41
5.14	Bimodalidade das caudas: $\alpha = 2\%$ - Petróleo	41
5.15	Densidade ajustada - Petróleo	42
5.16	Bimodalidade das caudas: $\alpha = 5\%$ - Dólar	42
5.17	Densidade ajustada - Dólar	43

Abreviações, Siglas e Símbolos

FC	função característica
$E(X)$	esperança de X
FDP	função de densidade de probabilidade
FDA	função de distribuição acumulada
FGM	função de geratriz de momento
IID	independentes e identicamente distribuídas
MLE	método da máxima verossimilhança
VA	variável aleatória
Var	variância
VaR	Value at Risk
$\Psi(\cdot)$	função Digamma
$\Gamma(\alpha)$	função Gamma
$\gamma(a; x)$	função Gamma incompleta inferior
$\Gamma(a; x)$	função Gamma incompleta superior
$\varphi_X(t)$	função Geratriz
$\text{sign}(\cdot)$	função Sinal

Capítulo 1

Introdução

Em teoria de probabilidade e estatística, busca-se obter uma explicação da realidade através da análise de dados, normalmente amostrais. Eventos raros ou que ocorram com pouca frequência possuem enorme relevância em diversas áreas da ciência e suas estimativas se tornam imprescindíveis para o planejamento de atividades sujeitas às adversidades desses eventos. Esse tipo de estudo é modelado através da teoria de valores extremos (Sansigolo, 2008).

A teoria consiste em calcular o limite da estatística de máximo (ou mínimo, se for o caso) de uma amostra aleatória, que tende a convergir para uma distribuição assintótica de extremos normalizados. Essa distribuição assintótica poderá ser de três tipos: Fréchet, Weibull negativa e Gumbel, que são, em verdade, variações da distribuição Generalizada de Valor Extremo (GEV) a partir alteração dos valores do parâmetro ξ .

A distribuição Weibull positiva, simétrica com a Weibull negativa, assim denominada devido ao engenheiro e matemático Waloddi Weibull, possui vasto campo de aplicação, particularmente para a engenharias, administração e biologia. Segundo Rinne (2008), juntamente com as distribuições normal, exponencial, t, F e quiquadrado, a Weibull é uma das mais populares e aplicáveis da estatística moderna.

Já distribuição Gumbel, também conhecida como distribuição de valor extremo ou distribuição Fisher-Tippet tipo I, é usada para modelar a distribuição dos valores máximos de de-

terminada amostra. Recebeu essa nomenclatura em homenagem ao matemático alemão Emil Julius Gumbel, após a publicação de seu livro sobre valores extremos em 1958, aperfeiçoando o trabalho de Leonard Tippett e Ronald Fisher (Beirland e Segers, 2005).

A distribuição Gumbel é um caso particular da distribuição de valores extremos e pode ser obtida por meio do logaritmo de uma variável aleatória com distribuição Weibull positiva. É amplamente utilizada em áreas como hidrologia, finanças, atuária e análise de sobrevivência. Em hidrologia, por exemplo, é utilizada para obter informações sobre riscos e tempo de retorno de determinados níveis extremos em rios ou bacias, estudo esse fundamental para a previsão de enchentes ou predição das chances de ocorrência de desabastecimentos (Naghetini, 2007).

A distribuição Gumbel possui diversas variações e generalizações, tais como: Distribuição Gumbel exponencializada desenvolvida por Nadarajah (2006) e inspirada na distribuição exponencial; Distribuição valor extremo transmutada, definida pelos autores Shaw e Buckley (2009), com a finalidade de obter distribuições assimétricas e com curtoses distintas da normal; Distribuição Kumaraswamy Gumbel, criada a partir da distribuição Kumaraswamy por Cordeiro (2012); Distribuição beta Gumbel concebida por Dubey (1969); dentre outras.

Uma revisão completa da distribuição Gumbel e suas generalizações foi realizada por Pinheiro (1999). Dentre essas alterações, a autora ressalta a mistura de dois componentes de distribuição valor extremo máximo, que teve o objetivo de analisar a frequência de inundações e gerou uma distribuição bimodal (Pinheiro, 1999).

Otiniano (2014), e também Escalante-Sandoval (2007), mostraram que a mistura de distribuições Gumbel é um ajuste adequado para dados de comportamento heterogêneo em hidrologia e finanças. A desvantagem em utilizar esse tipo de modelagem reside no fato de o vetor de parâmetros consistir em diversas variáveis, aumentando o custo computacional de estimação.

O mercado financeiro, por exemplo, pode se utilizar amplamente da teoria de valores extremos e seus diversos desdobramentos. As transações comerciais envolvendo clientes e fornecedores em países distintos são inerentes ao mundo globalizado e a variação do preço do dólar possui alto impacto nas políticas sociais e econômicas, além de afetar diretamente o consumo

do país. Sendo os Estados Unidos da América o país com uma das economias mais fortes e influentes do mundo, o dólar tornou-se a moeda mais relevante, sendo utilizada como base para grande parte das operações comerciais, especialmente no âmbito do comércio exterior (Thomé, 2014).

Nesse sentido, as operações de importação e exportação apresentam uma dinâmica em que a desvalorização da moeda local frente ao dólar favorece a exportação de produtos. No caso do mercado interno brasileiro, a moeda americana também atua no poder de compra dos residentes, considerando que os produtos importados ficam mais caros. O mesmo ocorre com algumas matérias-primas e *commodities*, que também respondem diretamente ao aumento do preço do dólar. Assim, o item final do bem adquirido pelo consumidor também possui seu preço final atrelado ao dólar.

Em especial em relação ao petróleo, a Petrobrás é frequentemente associada à essa *commodity*, sendo empresa de capital aberto cujo o acionista majoritário é o Estado brasileiro. A sua criação está ligada ao processo de desenvolvimento da industrialização brasileira, e sua história é marcada por vultuosos investimentos em pesquisa e desenvolvimento. Essa característica culminou no incremento tecnológico e humano, assim como permitiu à Petrobrás o pioneirismo da exploração de petróleo em águas profundas (Sentyurina, 2019).

Negociados na bolsa de valores brasileira sob os códigos PETR3 e PETR4, esses ativos, juntamente com os papéis da gigante mineradora Vale apresentam um dos maiores níveis de liquidez e volume negociados por dia na B3 (Sauer, 2016).

Desta forma, neste trabalho, propõe-se uma generalização do modelo Gumbel, denominado modelo Gumbel Bimodal, que possui apenas três parâmetros e é bastante flexível, de tal forma que pode ser utilizado para modelar dados extremos de comportamento heterogêneo.

Além disso, objetiva avaliar a variação cambial do dólar e o preço das ações da Petrobrás ao longo dos últimos anos. A partir dos dados será modelado o comportamento das variáveis e suas caudas, sendo proposta, ainda, uma nova fórmula de cálculo do VaR ou *Value at Risk*.

Inicialmente será apresentada uma breve revisão da distribuição Gumbel. Posteriormente,

nos capítulos seguintes, será apresentada a nova distribuição Gumbel Bimodal e suas aplicações.

Capítulo 2

Distribuição Gumbel Bimodal

2.1 Conceitos Preliminares

A teoria de valores extremos consiste no estudo da distribuição assintótica de extremos normalizados. Ou seja, da distribuição limite das estatísticas $\frac{X_{(n)}-a_n}{b_n}$ ou $\frac{X_{(1)}-a_n}{b_n}$, em que $X_{(n)} = \max\{X_1, \dots, X_n\}$ e $X_{(1)} = \min\{X_1, \dots, X_n\}$, para X_1, \dots, X_n v.a's i.i.d.

O papel do teorema para máximos é similar ao do teorema central do limite para médias, exceto pelo fato de este se aplica à média de uma amostra a partir de qualquer distribuição com variância finita, enquanto aquele, também conhecido como teorema de Fisher-Tippett-Gnedenko, afirma que se a distribuição de um máximo normalizado converge, então o limite deve pertencer a uma Distribuição Generalizada de Valores Extremos (GEV), que será, conforme variação de

um de seus parâmetros, uma das três seguintes classes particulares de distribuições:

$$\begin{aligned}
 \text{Fréchet: } \Phi_{\alpha, \sigma, \mu}(x) &= \begin{cases} \exp\left(-\left(\frac{x-\mu}{\sigma}\right)^{-\alpha}\right), & x \geq \mu, \\ 0, & x < \mu, \end{cases} \\
 \text{Weibull Negativa: } \Psi_{\alpha, \sigma, \mu}(x) &= \begin{cases} \exp\left(-\left(-\frac{x-\mu}{\sigma}\right)^{\alpha}\right), & x \geq \mu, \\ 0, & x < \mu, \end{cases} \\
 \text{Gumbel: } \Lambda_{\sigma, \mu}(x) &= \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right), \forall x \in \mathbb{R}
 \end{aligned} \tag{2.1}$$

cujos parâmetros $\alpha > 0$, $\sigma > 0$ e $\mu \in \mathbb{R}$ são de forma, escala e locação, respectivamente.

Neste trabalho será desenvolvido um modelo de probabilidade que generaliza a distribuição Gumbel dada em (2.1).

A função densidade de probabilidade (fdp) se uma variável aleatória (v.a) X com distribuição Gumbel, $X \sim F_G(\cdot; \mu, \sigma)$, é dada por

$$f_G(x; \mu, \sigma) = \frac{1}{\sigma} \exp\left\{-\left(\frac{x-\mu}{\sigma}\right) - \exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right\}, \quad x \in \mathbb{R} \tag{2.2}$$

Os efeitos dos parâmetros $\mu \in \mathbb{R}$ e $\sigma > 0$ na densidade de f_G são apresentado na Figura (2.1).

Na seção a seguir é utilizada a distribuição Gumbel com $\mu = 0$, cujas equações da função densidade (fdp) e da função de distribuição acumulada (fda) são dadas, respectivamente, por

$$f_G(x; \sigma) = \frac{1}{\sigma} \exp\left\{-\left(\frac{x}{\sigma}\right) - \exp\left[-\left(\frac{x}{\sigma}\right)\right]\right\}, \tag{2.3}$$

$$F_G(x; \sigma) = \exp\left\{-\exp\left[-\left(\frac{x}{\sigma}\right)\right]\right\}. \tag{2.4}$$

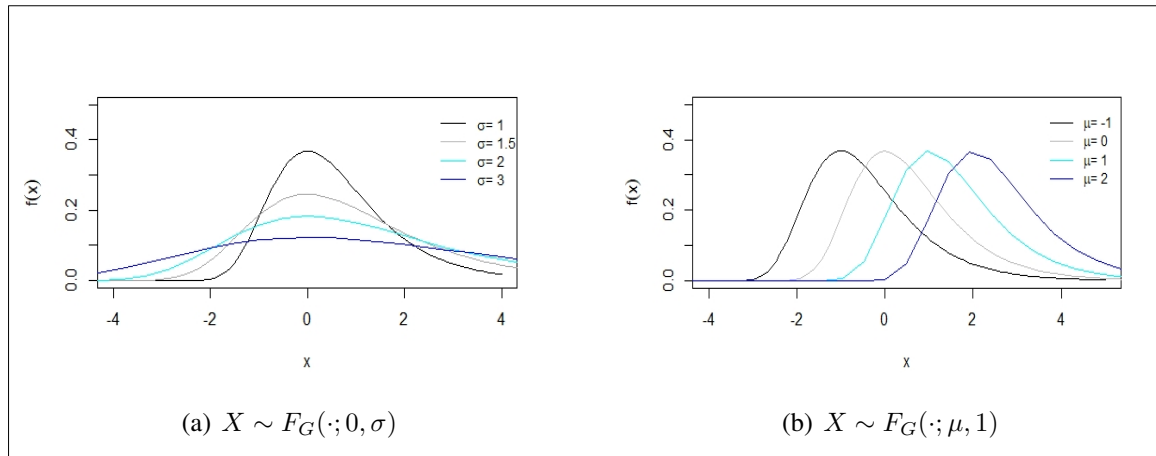


Figura 2.1: Densidade de $X \sim F_G(\cdot; \mu, \sigma)$

2.2 Modelo Gumbel Bimodal

Nesta etapa é definida uma generalização da distribuição Gumbel, chamada da distribuição Gumbel Bimodal, a partir da equação (2.3), e construída por meio de uma transformação inversível. É realizado, ainda, um estudo gráfico do novo modelo, bem como o estudo das principais medidas descritivas, como moda, quantis e momentos.

A distribuição Gumbel Bimodal é construída a partir das equações (2.3) e (2.4) e a Transformada $T_{\mu, \delta}$ definida por

$$T_{\mu, \delta}(x) = x|x|^\delta + \mu, \quad x \in \mathbb{R}^*, \quad \delta > -1, \quad \mu \in \mathbb{R}, \quad (2.5)$$

cujas primeiras derivadas são dadas por

$$T'_{\mu, \delta}(x) = (\delta + 1)|x|^\delta, \quad (2.6)$$

$$T''_{\mu, \delta}(x) = \text{sign}(x) (\delta + 1)\delta|x|^{\delta-1}, \quad (2.7)$$

sendo $\text{sign}(x)$ a função sinal.

Então, a k -ésima derivada é dada por

$$T_{\mu,\delta}^{(k)}(x) = [\text{sign}(x)]^{k-1} \left[\prod_{i=-1}^{k-2} (\delta - i) \right] |x|^{\delta-(k-1)}, \quad k \geq 2. \quad (2.8)$$

Para $\delta > -1$, a Transformada T é não decrescente, então a função inversa, denotada por $T_{\mu,\delta}^{-1}$, existe e é dada por

$$T_{\mu,\delta}^{-1}(x) = \begin{cases} (x - \mu)^{\frac{1}{1+\delta}}, & x > \mu \\ -(\mu - x)^{\frac{1}{1+\delta}}, & x < \mu. \end{cases} \quad (2.9)$$

Definição 2.2.1. Uma variável aleatória X possui distribuição Gumbel Bimodal, $X \sim F_{\text{GB}}(\cdot; \mu, \sigma, \delta)$, se sua função densidade de probabilidade (fdp) é definida por

$$\begin{aligned} f_{\text{GB}}(x) &= f_G(T_{\mu,\delta}(x); \sigma) T'_{\mu,\delta}(x) \\ &= \frac{1}{\sigma} (\delta + 1) |x|^\delta \exp \left\{ - \left(\frac{(x|x|^\delta) + \mu}{\sigma} \right) - \exp \left[- \left(\frac{(x|x|^\delta) + \mu}{\sigma} \right) \right] \right\}, \end{aligned} \quad (2.10)$$

sendo $\delta > -1$ e $\mu \in \mathbb{R}$ parâmetros de forma e $\sigma > 0$ parâmetro de escala. T e T' são dados conforme as equações (2.5) a (2.6).

A função distribuição acumulada (fda) de uma variável aleatória Gumbel Bimodal, $X \sim F_{\text{GB}}(\cdot; \mu, \sigma, \delta)$, calculada a partir de (2.4), é dada por

$$\begin{aligned} F_{\text{GB}}(x) &= F_G(T_{\mu,\delta}(x); \sigma) \\ &= \exp \left\{ - \exp \left[- \left(\frac{(x|x|^\delta) + \mu}{\sigma} \right) \right] \right\}, x \in \mathbb{R}. \end{aligned} \quad (2.11)$$

Em geral, a modelagem estatística de um conjunto de dados se inicia com uma análise descritiva dos elementos por meio das medidas de posição e medidas de dispersão, capazes

de fornecer informações que descrevam o modelo em estudo. Dessa forma, a seguir, serão apresentadas as principais medidas da nova distribuição proposta, como a moda, momentos e quantis.

2.3 Estudo gráfico

Inicialmente, foi realizado o estudo gráfico da nova distribuição, a partir da equação (2.10), variando seus parâmetros.

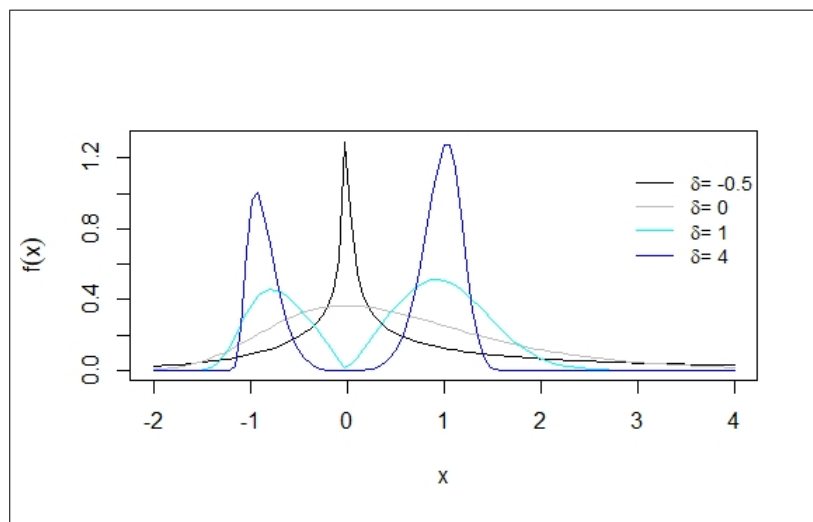


Figura 2.2: Densidade de $X \sim F_{GB}(\cdot; 0, 1, \delta)$

Na Figura 2.2 pode-se notar que ao realizar a variação em δ ocorre alteração no comportamento da distribuição, indicando tratar-se de um parâmetro de forma. A função aparenta ser bimodal com o aumento de δ . Para $\delta = 0$, a função torna-se uma distribuição Gumbel tradicional.

A Figura 2.3 mostra a densidade da Gumbel Bimodal quando é realizada uma variação no parâmetro μ . Percebe-se que com essa variação, μ aparenta também ser um parâmetro de forma.

Do mesmo modo, ao realizar a variação em σ , é possível notar, na Figura 2.4, que trata-se de um parâmetro de escala. As variações no seu valor, mantidos constantes os demais parâmetros,

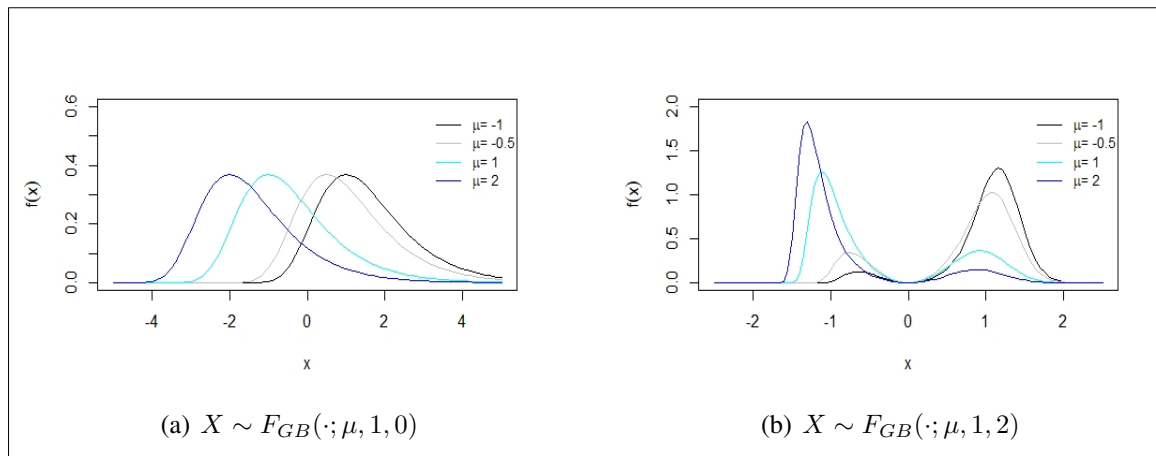


Figura 2.3: Densidade de $X \sim F_{GB}(\cdot; \mu, 1, \delta)$

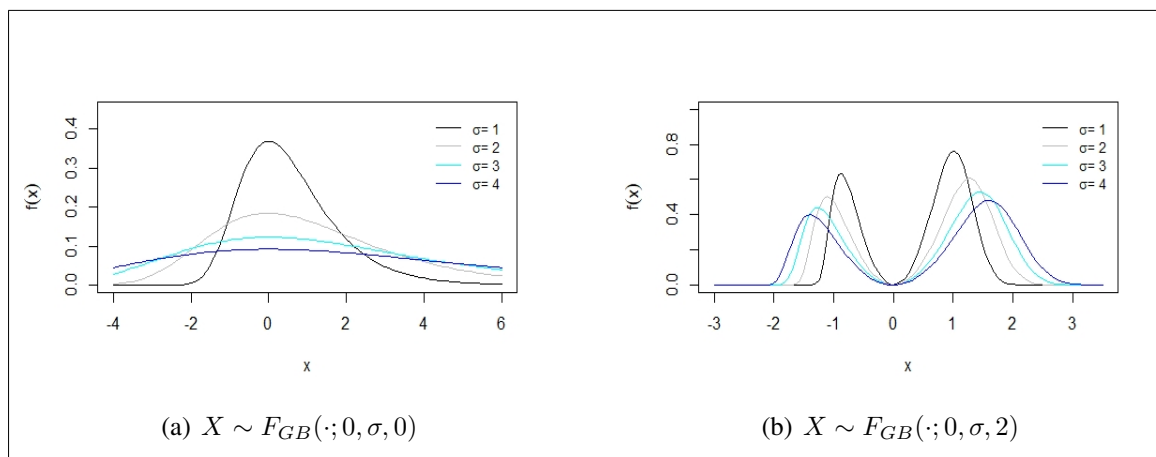


Figura 2.4: Densidade de $X \sim F_{GB}(\cdot; 0, \sigma, \delta)$

causam apenas compressão ou expansão do gráfico.

2.4 Moda

A moda de uma distribuição de probabilidade contínua é o valor x em que a fdp possui seu valor máximo. Ela não necessariamente é única, e no caso da distribuição Gumbel Bimodal espera-se que possua um ou dois valores modais. Assim, para o cálculo dessa propriedade, é necessário avaliar a monotonicidade da função, ou seja, estudar o sinal da primeira derivada da densidade, $f'_{GB}(x; \mu, \sigma, \delta)$.

Proposição 2.4.1. Seja $X \sim F_{GB}(\cdot; \mu, \sigma, \delta)$. Então x_m é um ponto de moda de X se for solução da equação diferencial:

$$\frac{T''_{\mu,\delta}(x)}{[T'_{\mu,\delta}(x)]^2} = \frac{1}{\sigma} - \frac{e^{-\frac{T_{\mu,\delta}(x)}{\sigma}}}{\sigma}. \quad (2.12)$$

Demonstração. Sabe-se que $f_{GB}(x; \mu, \sigma, \delta) = f_G(T_{\mu,\delta}(x); \mu) T'_{\mu,\delta}(x)$. Utilizando a regra da cadeia, é possível chegar a seguinte expressão:

$$f'_{GB}(x) = f'_G(T_{\mu,\delta}(x))T'_{\mu,\delta}(x)T'_{\mu,\delta}(x) + f_G(T_{\mu,\delta}(x))T''_{\mu,\delta}(x) = 0, \quad (2.13)$$

com f_G dada em (2.3) e $T'_{\mu,\delta}$ e $T''_{\mu,\delta}$ dadas por (2.6) e (2.7), respectivamente.

O resultado é obtido ao substituir a função

$$f'_G(x; \sigma) = \frac{1}{\sigma} f_G(x; \sigma) \left\{ \exp\left[-\left(\frac{x}{\sigma}\right)\right] - 1 \right\}$$

na equação (2.13).

□

2.5 Momentos

Em um estudo descritivo de um modelo de probabilidade, é também relevante definir os momentos de uma variável aleatória. Essa análise, no entanto, pode ainda ser utilizada, posteriormente, em seu estudo inferencial.

Em estatística, a expressão da esperança de uma variável aleatória, ou seu momento de ordem k , é dada por:

$$E(X^k) = \int x^k f(x) dx.$$

Para a função Gumbel Bimodal, corresponde ao cálculo de:

$$\begin{aligned} E(X^k) &= \int_{-\infty}^{\infty} x^k (f_{GB}(x)) dx & (2.14) \\ &= \int_{-\infty}^{\infty} x^k \frac{(\delta + 1)}{\sigma} |x|^\delta \exp \left\{ - \left(\frac{(x|x|^\delta) + \mu}{\sigma} \right) - \exp \left[- \left(\frac{(x|x|^\delta) - \mu}{\sigma} \right) \right] \right\} dx. \end{aligned}$$

O cálculo da integral (2.14) não é trivial, e, portanto, deverá ser calculada numericamente. Por isso, outra alternativa para o cálculo dos momentos de uma v.a. $X \sim F_{GB}(\cdot; \mu, \sigma, \delta)$ é através de uma função tipo geratriz de momentos.

2.5.1 Função tipo geratriz de momentos

Proposição 2.5.1. Seja X uma variável aleatória com distribuição Gumbel Bimodal, $X \sim F_{GB}(\cdot; \mu, \sigma, \delta)$. Então, a função tipo geratriz de momentos $\varphi_X(t) = E(e^{tX^{\delta+1}})$ é dada por:

$$\varphi_X(t) = e^{-t\mu} \Gamma(1 - t\sigma; e^{\frac{\mu}{\sigma}}) + e^{t\mu(-1)^{\delta+1}} \gamma(1 + (-1)^{\delta+1} t\sigma; e^{\frac{\mu}{\sigma}}). \quad (2.15)$$

Demonstração. Por definição,

$$\begin{aligned} \varphi_X(t) &= \int_{-\infty}^{\infty} e^{tx^{1+\delta}} f(T_{\mu,\delta}(x); \sigma) T'_{\sigma,\delta}(x) dx \\ &= \int_{-\infty}^{\infty} \frac{e^{tx^{1+\delta}}}{\sigma} e^{-\frac{T(x)}{\sigma}} e^{-e^{-\frac{T(x)}{\sigma}}} T'(x) dx. \end{aligned} \quad (2.16)$$

Realizando as substituições de $y = e^{-\frac{T(x)}{\sigma}}$ e $dy = e^{-\frac{T(x)}{\sigma}} (-T'(x)) dx$ e considerando, para ajustar os limites, que $T(x) = x|x|^\delta + \mu$, (2.16) pode ser reescrita por

$$\begin{aligned} \varphi_X(t) &= \int_{+\infty}^0 e^{tx^{1+\delta}} e^{-y} (-dy) \\ &= \int_0^{\infty} e^{tx^{1+\delta}} e^{-y} dy. \end{aligned} \quad (2.17)$$

A partir do cálculo de y ,

$$\begin{aligned} y &= e^{\frac{-T(x)}{\sigma}} \\ \ln y &= \frac{-T(x)}{\sigma} \\ T(x) &= -\sigma \ln y \\ x &= T^{-1}(-\sigma \ln y) \\ x &= T^{-1}(\ln y^{-\sigma}); \end{aligned}$$

e da função inversa de T , dada por (2.9),

$$T^{-1}(\ln y^{-\sigma}) = \begin{cases} (\ln y^{-\sigma} - \mu)^{\frac{1}{1+\delta}}, & y < e^{\frac{\mu}{\sigma}} \\ -(\mu - \ln y^{-\sigma})^{\frac{1}{1+\delta}}, & y > e^{\frac{\mu}{\sigma}}, \end{cases}$$

as equações (2.16) e (2.17) são reescritas em função de y . Então a função geratriz é dada por:

$$\begin{aligned} \varphi_X(t) &= \int_0^{e^{\frac{\mu}{\sigma}}} e^{t(\ln y^{-\sigma} - \mu)^{\frac{1+\delta}{1+\delta}}} e^{-y} dy + \int_{e^{\frac{\mu}{\sigma}}}^{\infty} e^{t(-1)^{1+\delta}(\mu - \ln y^{-\sigma})^{\frac{1+\delta}{1+\delta}}} e^{-y} dy \\ &= \int_0^{e^{\frac{\mu}{\sigma}}} e^{t(-\sigma \ln y - \mu)} e^{-y} dy + \int_{e^{\frac{\mu}{\sigma}}}^{\infty} e^{(-1)^{1+\delta}t(\mu + \sigma \ln y)} e^{-y} dy \quad (2.18) \\ &= I + II. \end{aligned}$$

De (2.18) é possível notar que o cálculo de $\varphi_X(t)$ deve ser realizado a partir da solução de duas integrais, I e II. Com alguma manipulações algébricas calcula-se a integral I.

$$I = \int_0^{e^{\frac{\mu}{\sigma}}} e^{t(-\sigma \ln y - \mu)} e^{-y} dy$$

$$\begin{aligned}
I &= \int_0^{e^{\frac{\mu}{\sigma}}} e^{-t\sigma \ln y} e^{-t\mu} e^{-y} dy \\
I &= e^{-t\mu} \int_0^{e^{\frac{\mu}{\sigma}}} e^{-t\sigma \ln y} e^{-y} dy \\
I &= e^{-t\mu} \int_0^{e^{\frac{\mu}{\sigma}}} y^{-t\sigma} e^{-y} dy.
\end{aligned} \tag{2.19}$$

De fomar análoga, calcula-se o valor da integral II,

$$\begin{aligned}
II &= \int_{e^{\frac{\mu}{\sigma}}}^{\infty} e^{(-1)^{1+\delta}t(\mu+\sigma \ln y)} e^{-y} dy \\
II &= e^{(-1)^{1+\delta}(t\mu)} \int_{e^{\frac{\mu}{\sigma}}}^{\infty} e^{(-1)^{1+\delta}(t\sigma \ln y)} e^{-y} dy \\
II &= e^{(-1)^{1+\delta}(t\mu)} \int_{e^{\frac{\mu}{\sigma}}}^{\infty} y^{t\sigma(-1)^{1+\delta}} e^{-y} dy.
\end{aligned} \tag{2.20}$$

Por fim, a partir de (2.19) e (2.20), pode ser obtido o valor da função geratriz:

$$\begin{aligned}
\varphi_X(t) &= e^{-t\mu} \int_0^{e^{\frac{\mu}{\sigma}}} y^{-t\sigma} e^{-y} dy + e^{(-1)^{1+\delta}(t\mu)} \int_{e^{\frac{\mu}{\sigma}}}^{\infty} y^{t\sigma(-1)^{1+\delta}} e^{-y} dy \\
&= e^{-t\mu} \Gamma(1 - t\sigma; e^{\frac{\mu}{\sigma}}) + e^{t\mu(-1)^{\delta+1}} \gamma(1 + (-1)^{\delta+1}t\sigma; e^{\frac{\mu}{\sigma}}),
\end{aligned}$$

sendo $\gamma(a; x) = \int_x^{\infty} t^{a-1} e^{-t} dt$ a função Gamma incompleta inferior e, complementarmente, $\Gamma(a; x) = \int_0^x t^{a-1} e^{-t} dt$, a função Gamma incompleta superior. Essas funções estão implementadas em diversos *softwares* matemáticos e estatísticos.

□

2.5.2 Função tipo geradora de cumulantes

Baseando-se no logaritmo da função $\varphi_X(t) = e^{tX^{\delta+1}}$, pode ser obtida $C_X(t)$, uma função tipo geradora de cumulantes $\ln(\varphi_X(t))$. Ao considerar a série de Taylor

$$\ln(Z) = (Z - 1) - \frac{(Z - 1)^2}{2} + \frac{(Z - 1)^3}{3} - \dots$$

para a função $E(e^{tX^{\delta+1}})$, tem-se a função do tipo cumulantes, dada por:

$$C_X(t) = \ln(E(e^{tX^{\delta+1}})) = (E(e^{tX^{\delta+1}}) - 1) - \frac{(E(e^{tX^{\delta+1}}) - 1)^2}{2} + \frac{(E(e^{tX^{\delta+1}}) - 1)^3}{3} - \dots \quad (2.21)$$

As derivadas de $C_X(t)$ em $t = 0$ permitem que sejam obtidos momentos da v.a. X . Em particular, da primeira e segunda derivadas, tem-se:

$$\frac{\partial}{\partial t} C_X(t) = (E(e^{tX^{\delta+1}} X^{\delta+1})) - 2 \frac{(E(e^{tX^{\delta+1}}) - 1)}{2} E(e^{tX^{\delta+1}} X^{\delta+1}) + \dots$$

$$\frac{\partial^2}{\partial t^2} C_X(t) = (E(e^{tX^{\delta+1}} X^{2(\delta+1)})) - [(E(e^{tX^{\delta+1}} X^{\delta+1}))]^2 - [E(e^{tX^{\delta+1}}) - 1] + \dots$$

E, quando $t = 0$, as derivadas são dadas por

$$\frac{\partial}{\partial t} C_X(t) |_{t=0} = E(X^{\delta+1}) \quad (2.22)$$

$$\frac{\partial^2}{\partial t^2} C_X(t) |_{t=0} = E(X^{2(\delta+1)}) - [E(X^{\delta+1})]^2. \quad (2.23)$$

Proposição 2.5.2. Seja X uma variável aleatória com distribuição Gumbel Bimodal, $X \sim$

$F_{GB}(\cdot; \mu, \sigma, \delta)$, então:

$$C_X(t) = \ln[e^{-t\mu} \Gamma(1 - t\sigma; e^{\frac{\mu}{\sigma}}) + e^{t\mu(-1)^{\delta+1}} \gamma(1 + (-1)^{\delta+1} t\sigma; e^{\frac{\mu}{\sigma}})] \quad (2.24)$$

Demonstração. A prova segue diretamente das equações (2.15) e (2.21). □

Corolário 2.5.1. Seja $X \sim F_{GB}(\cdot; \mu, \sigma, 0)$. Então:

$$E(X) = \sigma\gamma - \mu \quad (2.25)$$

$$Var(X) = \sigma^2 \frac{\pi^2}{6} \quad (2.26)$$

Demonstração. Ao considerar $\delta = 0$, a equação (2.24) pode ser reescrita da seguinte forma:

$$\begin{aligned} C_X(t) &= \ln[e^{-t\mu}\Gamma(1 - t\sigma; e^{\frac{\mu}{\sigma}}) + e^{-t\mu}\gamma(1 - t\sigma; e^{\frac{\mu}{\sigma}})] \\ &= \ln[e^{-t\mu}\Gamma(1 - t\sigma)]. \end{aligned}$$

Pela propriedade $\Gamma(x; a) + \gamma(x; a) = \Gamma(x)$, obtém-se $C_X(t) = \ln[e^{-t\mu}\Gamma(1 - t\sigma)]$, cujas derivadas são dadas por:

$$\frac{\partial}{\partial t} C_X(t) = -\mu - \sigma\Psi(1 - t\sigma) \quad (2.27)$$

e

$$\frac{\partial^2}{\partial t^2} (C_X(t)) = \sigma^2\Psi'(1 - t\sigma), \quad (2.28)$$

sendo $\Psi(\cdot)$ a função Digamma, definida por $\Psi(x) = \frac{\partial}{\partial x} \ln(\Gamma(x)) = \frac{\Gamma'(x)}{\Gamma(x)}$, com $\Psi(1) = \gamma$, $\gamma =$ constante de Euler (0,5722) e $\Psi'(1) = \frac{\pi^2}{6}$. Ao igualar as equações (2.22) com (2.27) e as equações (2.23) com (2.28), obtém-se a média e a variância de X , dados, respectivamente, por

$$E(X) = \sigma\gamma - \mu$$

$$Var(X) = \sigma^2 \frac{\pi^2}{6}$$

□

Observação. Um caso particular de $X \sim F_{GB}$ ocorre quando $\delta = 0$, caso em que $F_{GB}(\cdot; \mu, \sigma, 0) = F_G(\cdot; -\mu, \sigma)$, em que F_G é a função de distribuição acumulada da Gumbel, cuja média e variância coincidem com (2.25) e (2.26).

□

2.6 Quantis

Proposição 2.6.1. Seja $X \sim F_{GB}(\cdot; \mu, \sigma, \delta)$. Os quantis da distribuição são dados por

$$x_q = \begin{cases} \left(-\sigma \ln(-\ln(q)) - \mu \right)^{\frac{1}{1+\delta}}, & q > e^{-e^{-\frac{\mu}{\sigma}}} \\ -\left(\mu + \sigma \ln(-\ln(q)) \right)^{\frac{1}{1+\delta}}, & q < e^{-e^{-\frac{\mu}{\sigma}}} \end{cases} \quad (2.29)$$

Demonstração.

$$q = P(X \leq x_q) = F_{GB}(x_q) \Rightarrow x_q = F_{GB}^{-1}(q).$$

A partir da equação (2.11), sabe-se o valor da $F_{GB}(\cdot)$, que pode ser substituído na equação acima. Assim,

$$\begin{aligned} q &= \exp \left\{ -\exp \left[-\frac{T_{\mu, \delta}(x_q)}{\sigma} \right] \right\} \\ \ln(q) &= -\exp \left[-\frac{T_{\mu, \delta}(x_q)}{\sigma} \right] \\ -\sigma \ln[-\ln(q)] &= T_{\mu, \delta}(x_q) \\ x_q &= T_{\mu, \delta}^{-1} \left[-\sigma \ln(-\ln(q)) \right] \end{aligned}$$

Da equação (2.9) são obtidas as fórmulas de cálculo da função da Transformada Inversa, repetidas abaixo. Assim, é possível utilizar esse resultado para realizar a substituição e encontrar a expressão para x_q .

$$T_{\mu,\delta}^{-1}(x) = \begin{cases} (x - \mu)^{\frac{1}{1+\delta}}, & x > \mu \\ -(\mu - x)^{\frac{1}{1+\delta}}, & x < \mu \end{cases}$$

$$x_q = \begin{cases} (-\sigma \ln(-\ln(q)) - \mu)^{\frac{1}{1+\delta}}, & -\sigma \ln(-\ln(q)) > \mu \\ -(\mu + \sigma \ln(-\ln(q)))^{\frac{1}{1+\delta}}, & -\sigma \ln(-\ln(q)) < \mu \end{cases}$$

$$x_q = \begin{cases} (-\sigma \ln(-\ln(q)) - \mu)^{\frac{1}{1+\delta}}, & q > e^{-e^{-\frac{\mu}{\sigma}}} \\ -(\mu + \sigma \ln(-\ln(q)))^{\frac{1}{1+\delta}}, & q < e^{-e^{-\frac{\mu}{\sigma}}} \end{cases}$$

□

Capítulo 3

Estimação

Neste capítulo utilizou-se o método de máxima verossimilhança para estimar os parâmetros da densidade $f_{GB}(\cdot; \mu, \sigma, \delta)$ dada em (2.10). O estimador do parâmetro de escala μ possui fórmula fechada, já os estimadores de σ e δ são obtidos ao resolver numericamente duas equações.

3.1 Estimação por máxima verossimilhança

O método da máxima verossimilhança (MLE) consiste em estimar os parâmetros do modelo de estudo utilizando as estimativas que tornam máximo o valor da função verossimilhança. É usual utilizar a aplicação da função logaritmo no cálculo das estimações. Assim, os valores estimados dos parâmetros podem ser encontrados igualando as primeiras derivadas da função a zero.

Assim, a partir de valores amostrais x_1, x_2, \dots, x_n de uma população X com fdp $f(\cdot; \theta)$, o objetivo da metodologia é estimar o valor do parâmetro Θ que maximiza a função de verossimilhança (L), dada por

$$L(\Theta; x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i | \Theta), \quad (3.1)$$

ou, ainda, maximizar a função logarítmo da verossimilhança, denotada por:

$$l(\Theta; x) = \log(L(\Theta; x_1, x_2, \dots, x_n)) = \sum_{i=1}^n \log(f(x_i | \Theta)). \quad (3.2)$$

Então, para valores amostrais x_1, x_2, \dots, x_n de $X \sim F_{GB}(\cdot; \mu, \sigma, \delta)$, cuja função é dada pela equação (2.10), e que o vetor de parâmetros é $\Theta = (\mu, \sigma, \delta)$, a função verossimilhança é dada por:

$$\begin{aligned} l(\Theta; x) &= \sum_{i=1}^n \left\{ \ln(\delta + 1) - \ln(\sigma) + \delta \ln(|x_i|) + \left[- \left(\frac{(x_i |x_i|^\delta) + \mu}{\sigma} \right) - e^{-\left(\frac{(x_i |x_i|^\delta) + \mu}{\sigma} \right)} \right] \right\} \\ &= n \ln(\delta + 1) - n \ln(\sigma) + \delta \sum_{i=1}^n \ln(|x_i|) - \frac{\sum_{i=1}^n (x_i |x_i|^\delta) + \mu}{\sigma} - \sum_{i=1}^n e^{-\left(\frac{(x_i |x_i|^\delta) + \mu}{\sigma} \right)} \end{aligned} \quad (3.3)$$

Dessa forma, a partir de (3.3), é possível encontrar os estimadores dos parâmetros para a distribuição Gumbel Bimodal, por meio da derivação em relação a cada parâmetro.

(a)

$$\begin{aligned} \frac{\partial l(\Theta; x)}{\partial \mu} &= -\frac{n}{\sigma} - \left(\frac{-1}{\sigma} \right) \sum_{i=1}^n e^{-\left(\frac{(x_i |x_i|^\delta) + \mu}{\sigma} \right)} = 0 \\ 0 &= -\frac{n}{\sigma} + \frac{1}{\sigma} \sum_{i=1}^n \left(e^{-\frac{x_i |x_i|^\delta}{\sigma}} e^{-\frac{\mu}{\sigma}} \right) \\ \frac{n}{\sigma} &= \frac{e^{-\frac{\mu}{\sigma}}}{\sigma} \sum_{i=1}^n e^{-\frac{x_i |x_i|^\delta}{\sigma}} \\ e^{-\frac{\mu}{\sigma}} &= \frac{n}{\sum_{i=1}^n e^{-\frac{x_i |x_i|^\delta}{\sigma}}} \\ \hat{\mu} &= \ln \left(\frac{n}{\sum_{i=1}^n e^{-\frac{x_i |x_i|^\delta}{\sigma}}} \right)^{-\sigma} \\ \therefore \hat{\mu} &= \sigma \ln \left(\frac{\sum_{i=1}^n e^{-\frac{x_i |x_i|^\delta}{\sigma}}}{n} \right) \end{aligned} \quad (3.4)$$

(b)

$$\begin{aligned} \frac{\partial l(\Theta; x)}{\partial \sigma} &= -\frac{n}{\sigma} - \left[\left(\sum_{i=1}^n (-1) \sigma^{-2} ((x_i | x_i |^\delta) + \mu) \right) \right] - \\ &\quad \left[\sum_{i=1}^n e^{-((x_i | x_i |^\delta) + \mu) \sigma^{-1}} \left(\sigma^{-2} ((x_i | x_i |^\delta) + \mu) \right) \right] = 0 \quad (3.5) \\ \frac{n}{\sigma} &= \left[\left(\sum_{i=1}^n \sigma^{-2} ((x_i | x_i |^\delta) + \mu) \right) \right] - \left[\sum_{i=1}^n e^{-((x_i | x_i |^\delta) + \mu) \sigma^{-1}} \left(\sigma^{-2} ((x_i | x_i |^\delta) + \mu) \right) \right] \end{aligned}$$

(c)

$$\begin{aligned} \frac{\partial l(\Theta; x)}{\partial \delta} &= \frac{n}{(\delta + 1)} + \sum_{i=1}^n \ln(|x_i|) - \left[\frac{\sum_{i=1}^n (x_i | x_i |^\delta)}{\sigma} \ln \left(\frac{\sum_{i=1}^n (x_i | x_i |^\delta)}{\sigma} \right) \right] \\ &\quad - \left[e^{\frac{\mu}{\sigma}} \sum_{i=1}^n e^{-\frac{(x_i | x_i |^\delta)}{\sigma}} (-1) \frac{\sum_{i=1}^n (x_i | x_i |^\delta)}{\sigma} \ln \left(\frac{\sum_{i=1}^n (x_i | x_i |^\delta)}{\sigma} \right) \right] = 0 \quad (3.6) \end{aligned}$$

Com o objetivo de encontrar os estimadores de máxima verossimilhança para cada um dos parâmetros da distribuição Gumbel Bimodal, é utilizado o método da otimização de funções, através do *software* estatístico R.

Neste trabalho, a estratégia para escolha dos valores iniciais dos parâmetros foi considerar algumas estatísticas amostrais baseada em uma amostra aleatória de X . Destarte, optou-se pela média amostral para μ , o desvio padrão para σ e a moda para δ , após geração de amostra aleatória utilizando a distribuição uniforme.

Capítulo 4

Simulação

Com o objetivo de investigar a performance dos estimadores de máxima verossimilhança apresentados no Capítulo 3, neste capítulo são definidos 27 vetores de parâmetros e obtidas suas estimativas médias, erro quadrático médio (EQM), viés e erro padrão através da simulação de Monte Carlo.

4.1 Simulação por Monte Carlo

Existem diversos procedimentos e algoritmos para gerar amostras a partir de uma distribuição. Um deles é o método da transformada inversa. Assim, para obter uma observação x de uma variável aleatória X com distribuição F , primeiramente deve-se gerar uma variável q com distribuição uniforme $(0,1)$, onde q refere-se ao quantil da distribuição em questão (nesse caso à equação (2.29)). Em seguida, a partir de $X = F^{-1}(q)$, onde F^{-1} é a inversa de F , gera-se a amostra da distribuição (Nelsen, 2006).

Assim, foram realizadas simulações de amostras aleatórias de $X \sim F_{GB}(\cdot; \mu, \sigma, \delta)$ por meio do método da transformada inversa. Foram realizadas $M = 100$ replicações de Monte Carlo para três tamanhos de amostra: $n = 50$, $n = 100$ e $n = 1000$.

Para cada amostra gerada os parâmetros μ , σ e δ foram estimados pelo método da má-

ximaverossimilhança, com o uso das equações (3.4) a (3.6). Dessa forma, foram obtidos 100 estimadores para cada parâmetro e calculado o estimador de Monte Carlo. Esse estimador é representado pelo valor esperado dos estimadores encontrados, ou seja $\hat{\theta} = \frac{1}{M} \sum_{i=1}^M \hat{\theta}_i$.

O parâmetro σ , por tratar-se de um parâmetro de escala, foi mantido constante e igual a 1 em todas as combinações. Os demais parâmetros variaram do seguinte modo: o parâmetro μ apresentou os valores -1, 0 e 1 e o parâmetro δ os valores 0, 2 e 4, totalizando 27 vetores de parâmetros.

Também foi calculado o viés entre o estimador e o valor real do parâmetro, $vies(\hat{\theta}) = (\hat{\theta} - \theta)$.

Ainda, para avaliar a qualidade do estimador, foi calculado o erro quadrático médio, dado pela expressão $EQM = \frac{1}{M} \sum_{i=1}^M (\hat{\theta}_i - \theta)^2$, e o erro padrão.

Pode-se notar que as estimativas geradas estão próximas aos valores reais dos parâmetros. Com o aumento no número de amostras, o erro quadrático médio e o erro padrão reduzem drasticamente. Assim, por apresentarem valores muito baixos e próximos a zero, indicam que os estimadores são eficientes.

A partir da análise gráfica, compara-se a curva da distribuição com os parâmetros reais e a curva dos parâmetros estimados. Assim, os gráficos apresentados contêm a variação dos valores estimados; a curva da distribuição com os parâmetros com seus verdadeiros valores, plotada na linha de cor preta; e a curva da distribuição com os parâmetros estimados, plotada com gráficos de pontos, em azul.

É possível notar que, por as estimativas obtidas serem muito próximas aos valores reais, as curvas se sobrepõem.

Tabela 4.1: Estimativas via Monte Carlo (n=50)

	Verdadeiro Valor	Valor Esperado	Viés	EQM	Erro Padrão
μ	-1	-1.00952	-0.00952	0.02163	0.1475
σ	1	0.97492	-0.02507	0.02765	0.1653
δ	0	0.01114	0.01114	0.01554	0.1248
μ	-1	-1.03142	-0.03142	0.03419	0.1831
σ	1	1.02366	0.02366	0.02592	0.1600
δ	2	2.05863	0.05863	0.12062	0.3440
μ	-1	-0.99788	0.00211	0.02748	0.1666
σ	1	0.96513	-0.03486	0.02763	0.1634
δ	4	3.91971	-0.08028	0.34339	0.5834
μ	0	0.00145	0.00145	0.02113	0.1461
σ	1	0.99965	-0.00034	0.01646	0.1289
δ	0	0.03047	0.03047	0.01316	0.1111
μ	0	0.00339	0.00339	0.02363	0.1544
σ	1	1.00572	0.00572	0.01401	0.1188
δ	2	2.14749	0.14749	0.14933	0.3589
μ	0	-0.02841	-0.02841	0.02129	0.1438
σ	1	0.97091	-0.02908	0.01374	0.1141
δ	4	4.00093	0.00093	0.26488	0.5172
μ	1	0.95417	-0.04582	0.02859	0.1636
σ	1	0.98530	-0.01469	0.01918	0.1384
δ	0	0.00560	0.00560	0.01242	0.1118
μ	1	0.96207	-0.03792	0.02218	0.1447
σ	1	1.04081	0.04081	0.02338	0.1481
δ	2	2.10659	0.10659	0.17084	0.4013
μ	1	0.93990	-0.06009	0.05954	0.2376
σ	1	0.97762	-0.02237	0.12103	0.3489
δ	4	3.96419	-0.03580	0.91265	0.9594

Tabela 4.2: Estimativas via Monte Carlo (n=100)

	Verdadeiro Valor	Valor Esperado	Viés	EQM	Erro Padrão
μ	-1	-0.98025	0.01974	0.01280	0.1120
σ	1	0.98180	-0.01819	0.01146	0.1060
δ	0	-0.00898	-0.00898	0.00780	0.0883
μ	-1	-0.99734	0.00265	0.01532	0.1243
σ	1	0.99027	-0.00972	0.01188	0.1091
δ	2	2.01504	0.01504	0.06759	0.2608
μ	-1	-0.99937	0.00062	0.01558	0.1254
σ	1	0.97695	-0.02304	0.01185	0.1069
δ	4	3.9113	-0.08860	0.13284	0.3553
μ	0	-0.02087	-0.02087	0.01232	0.1095
σ	1	0.99485	-0.00514	0.00724	0.0854
δ	0	0.02235	0.02235	0.00722	0.0824
μ	0	-0.01413	-0.01413	0.01182	0.1083
σ	1	0.98941	-0.01058	0.00694	0.0830
δ	2	2.06691	0.06691	0.07194	0.2610
μ	0	-0.00765	-0.00765	0.01211	0.1103
σ	1	0.99801	-0.00198	0.00713	0.0848
δ	4	4.05904	0.05904	0.16391	0.4025
μ	1	0.94706	-0.05293	0.01350	0.1039
σ	1	0.98666	-0.01333	0.00964	0.0977
δ	0	-0.02825	-0.0282	0.00588	0.0717
μ	1	0.96052	-0.03947	0.014831	0.1157
σ	1	1.01628	0.01628	0.01126	0.1053
δ	2	1.99650	-0.00349	0.06319	0.2526
μ	1	0.98981	-0.01018	0.01102	0.1050
σ	1	0.99397	-0.00602	0.00709	0.0844
δ	4	4.02093	0.02093	0.11929	0.3464

Tabela 4.3: Estimativas via Monte Carlo (n=1000)

	Verdadeiro Valor	Valor Esperado	Viés	EQM	Erro Padrão
μ	-1	-0.99837	0.00162	0.00149	0.0388
σ	1	0.99413	-0.00586	0.00104	0.0319
δ	0	-0.00209	-0.00209	0.00072	0.0269
μ	-1	-0.99906	0.00093	0.00130	0.0362
σ	1	1.00507	0.00507	0.00088	0.0293
δ	2	2.01457	0.01457	0.00540	0.0724
μ	-1	-0.99417	0.00582	0.00110	0.0328
σ	1	0.99574	-0.00425	0.00091	0.03016
δ	4	3.99378	-0.00621	0.01433	0.1201
μ	0	-0.00232	-0.00232	0.00134	0.0367
σ	1	1.00226	0.00226	0.00069	0.02643
δ	0	0.00303	0.00303	0.00076	0.0277
μ	0	-0.00322	-0.00322	0.00094	0.0307
σ	1	0.99633	-0.00366	0.00062	0.0248
δ	2	2.01715	0.01715	0.00587	0.0751
μ	0	-0.00824	-0.00824	0.00114	0.0330
σ	1	1.00379	0.00379	0.00065	0.0254
δ	4	4.01962	0.01962	0.01519	0.1223
μ	1	0.99310	-0.00689	0.00136	0.0365
σ	1	0.99649	-0.00350	0.00107	0.0327
δ	0	-0.00116	-0.00116	0.00080	0.0284
μ	1	0.95857	-0.04142	0.00253	0.0286
σ	1	0.99516	-0.00483	0.00104	0.0320
δ	2	1.96530	-0.03469	0.00929	0.0904
μ	1	0.99638	-0.00361	0.00133	0.0365
σ	1	1.00135	0.00135	0.00108	0.0330
δ	4	4.00819	0.00819	0.01701	0.1308

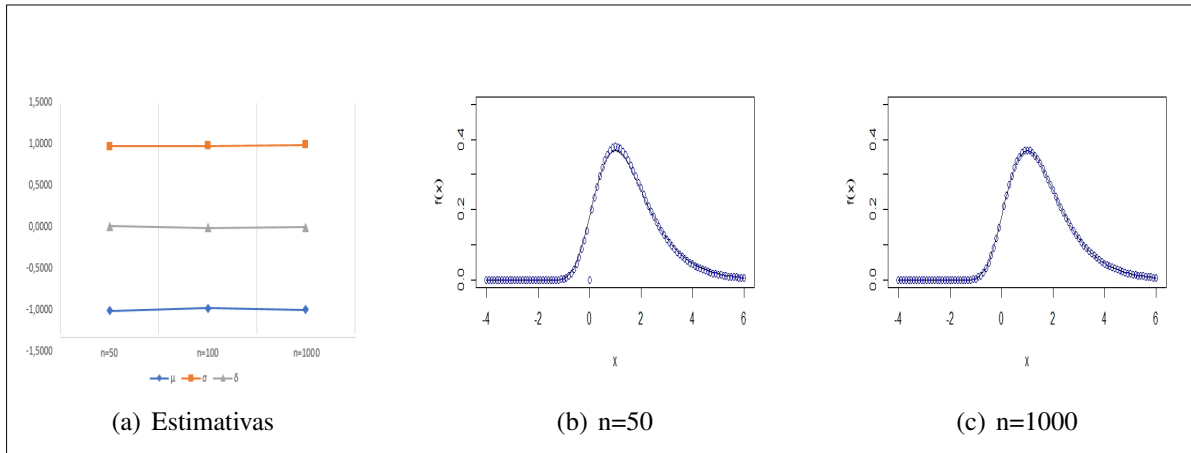


Figura 4.1: Comparativo para $X \sim F_{GB}(\cdot; \mu = -1, \sigma = 1, \delta = 0)$

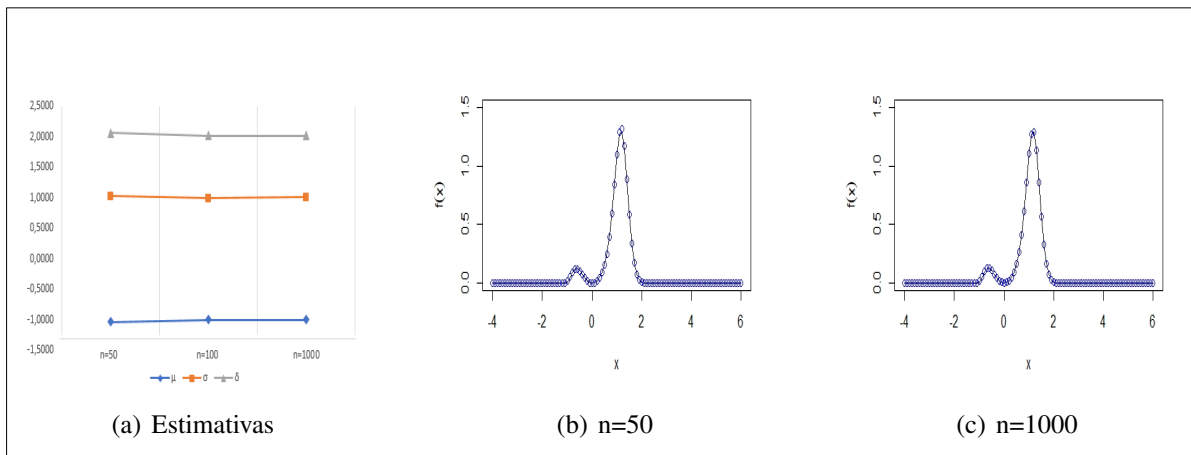


Figura 4.2: Comparativo para $X \sim F_{GB}(\cdot; \mu = -1, \sigma = 1, \delta = 2)$

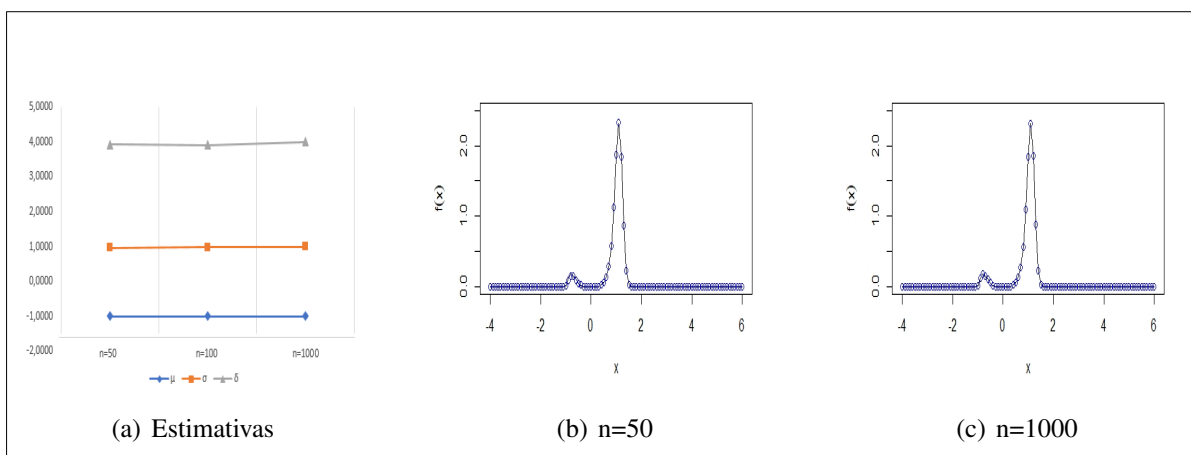


Figura 4.3: Comparativo para $X \sim F_{GB}(\cdot; \mu = -1, \sigma = 1, \delta = 4)$

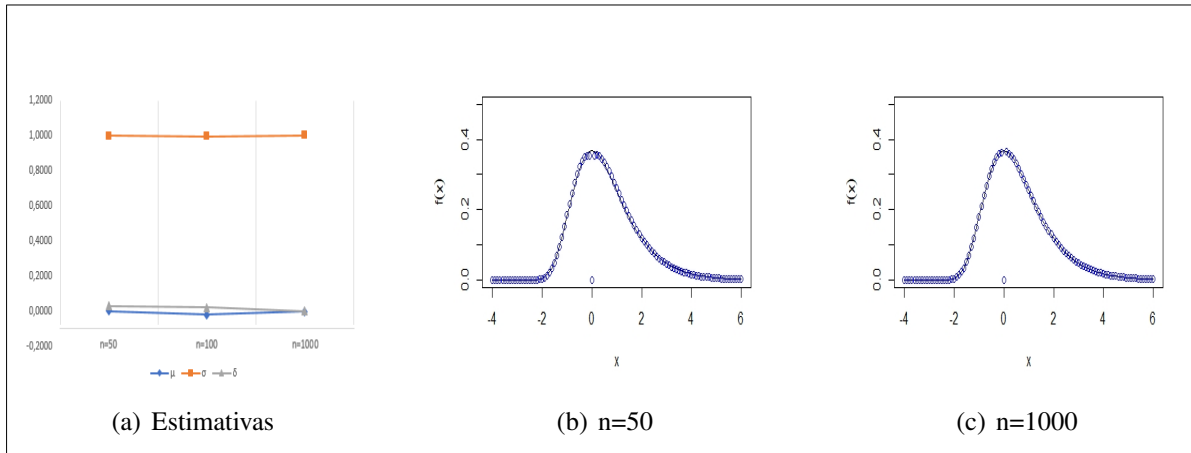


Figura 4.4: Comparativo para $X \sim F_{GB}(\cdot; \mu = 0, \sigma = 1, \delta = 0)$

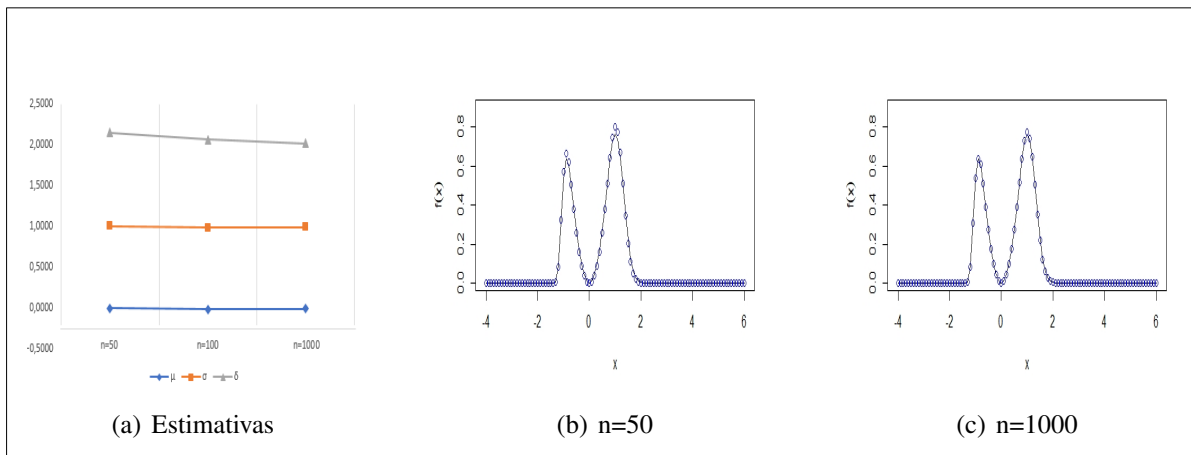


Figura 4.5: Comparativo para $X \sim F_{GB}(\cdot; \mu = 0, \sigma = 1, \delta = 2)$

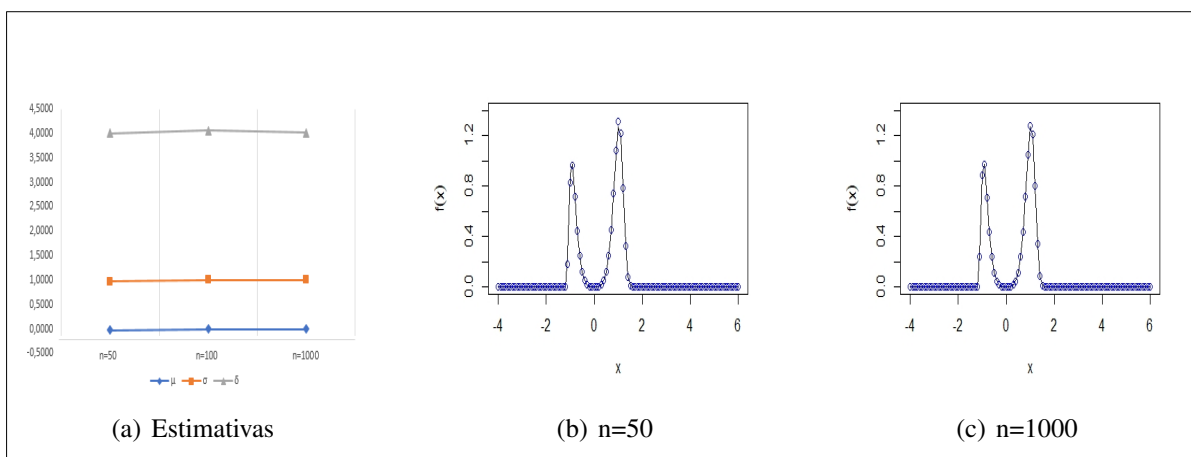


Figura 4.6: Comparativo para $X \sim F_{GB}(\cdot; \mu = 0, \sigma = 1, \delta = 4)$

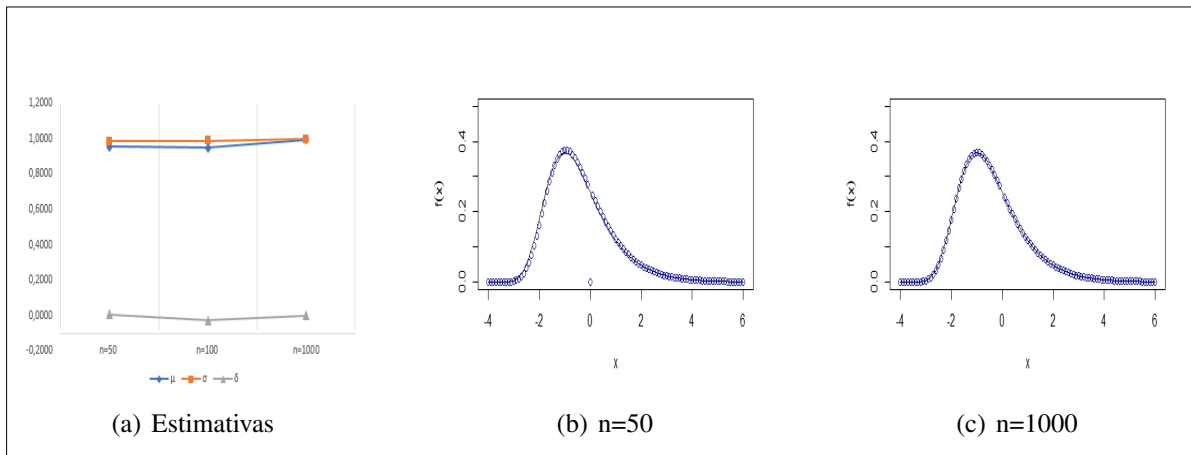


Figura 4.7: Comparativo para $X \sim F_{GB}(\cdot; \mu = 1, \sigma = 1, \delta = 0)$

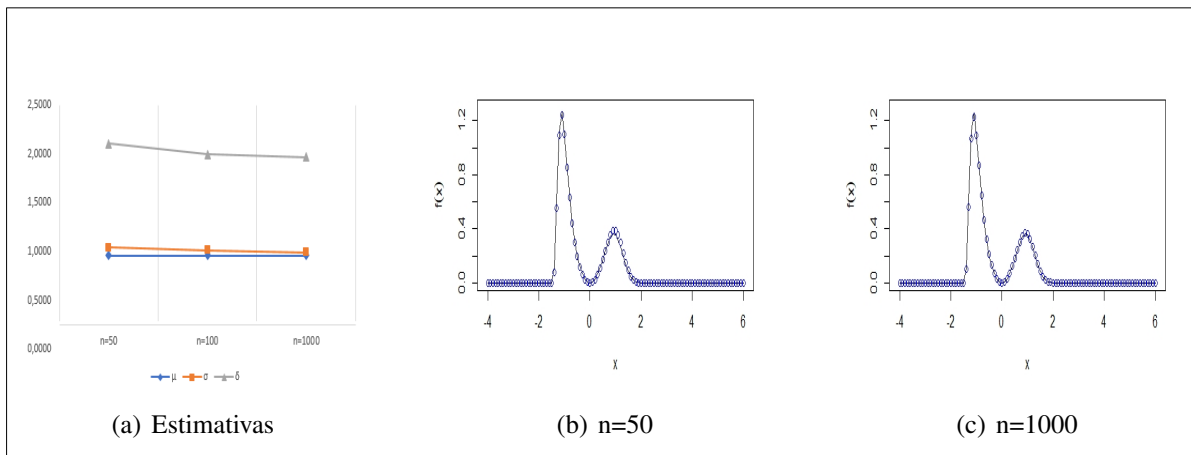


Figura 4.8: Comparativo para $X \sim F_{GB}(\cdot; \mu = 1, \sigma = 1, \delta = 2)$

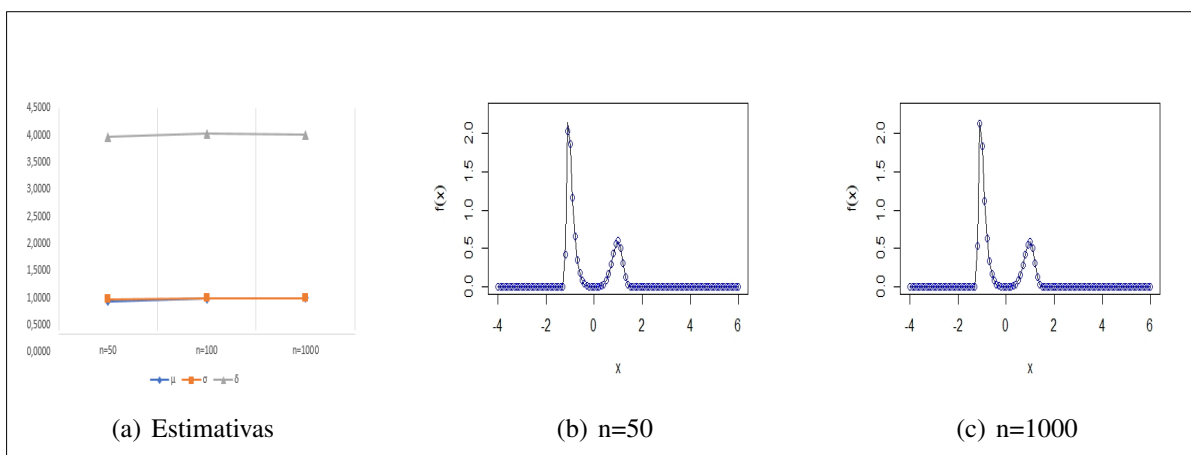


Figura 4.9: Comparativo para $X \sim F_{GB}(\cdot; \mu = 1, \sigma = 1, \delta = 4)$

Capítulo 5

Aplicação

Neste capítulo, para ilustrar a aplicabilidade do modelo Gumbel Bimodal, serão utilizados dois conjuntos de dados de finanças. As séries utilizadas foram analisadas no *software* R e obtidas no site <https://finance.yahoo.com>.

Os dados iniciais correspondem às ações da Petrobrás (PETR4) cotadas em dólar no período de 03/01/2000 a 01/10/2021, totalizando total de 5465 observações. No caso dos dados do dólar, a série histórica se inicia em 01/12/2003 e finaliza em 15/10/2021, totalizando 4223 observações.

O principal objetivo é avaliar os dados sob a luz da teoria de valores extremos e determinar o VaR, *Value at Risk*, dos retornos máximos e mínimos das variáveis.

5.1 Metodologia

Em finanças, usualmente, calcula-se o VaR correspondente a uma das caudas da distribuição de retornos, aquela que for de maior interesse do pesquisador. O VaR é usualmente definido como o menor valor das perdas acumuladas que ocorre com uma probabilidade α na cauda esquerda e $1-\alpha$ na cauda direita, dentro de um horizonte de tempo. Então, se F_X é a distribuição acumulada dos retornos X e $1 - F_X$ é a função sobrevivência, então o VaR_α de X é definido

por

$$VaR_\alpha(X) = \inf\{x \in \mathbb{R} : F_X(x) \geq \alpha\} \quad (5.1)$$

ou, equivalentemente,

$$VaR_\alpha(X) = \inf\{x \in \mathbb{R} : \bar{F}_X(x) \leq 1 - \alpha\}. \quad (5.2)$$

Neste trabalho, p_t será denotado o valor do preço de uma ação no tempo t e os retornos logarítmicos serão dados por

$$X_t = \ln(p_{t+1} - p_t). \quad (5.3)$$

Para retornos X_1, \dots, X_N de um investimento X , em que a variância de X é finita, os retornos extremos (máximos ou mínimos) são naturalmente modelados por uma distribuição Gumbel, $X \sim F_G$. Neste caso, para uma subamostra de máximos, $\alpha\%$ dos retornos máximos estão abaixo do VaR_α , dentro do horizonte de tempo no qual foi construída a sub-amostra.

Como $X \sim F_{GB}$ é uma distribuição que generaliza a distribuição Gumbel tradicional, neste capítulo ela será utilizada para modelar os dados de máximos e mínimos, ao mesmo tempo. assim, pode-se calcular e interpretar o VaR dos retornos mínimos e máximos a partir da F_{GB} .

5.2 Análise Descritiva

Observando a variação cambial do dólar, na Figura 5.1 é possível verificar que a moeda possui variação constante, com tendências a maior crescimento ao longo do tempo.

Em 2008, devido a recessão econômica mundial, causada principalmente pelos altos gastos dos Estados Unidos e as inicialmente baixas taxas de juros que financiaram a bolha imobiliária-

ria, pode-se notar um grande salto no valor do dólar frente ao real. Outro salto considerável pode ser observado no ano 2020, possivelmente devido à pandemia do novo coronavírus e as instabilidades econômicas decorrentes.



Figura 5.1: Série histórica - variação do preço da ação do petróleo em dólar

Da mesma forma, a variação do preço das ações do petróleo também está em constante mutação, sendo possível notar uma oscilação ainda maior que a da moeda americana. Além disso, é possível notar que a crise econômica de 2008 também afetou a indústria petrolífera. Os movimentos especulativos de escala global fizeram que o preço do produto subisse muito no início do ano, sofrendo, posteriormente, severas quedas após a quebra do banco americano Lehman Brothers.



Figura 5.2: Série histórica - variação do preço da ação do petróleo em dólar

Até o ano de 2019, a variação do dólar parece ter mais influência no preço das ações da Petrobrás. Em 2020, as preocupações com o novo coronavírus afetaram bolsas pelo mundo todo e as ações da Petrobrás despencaram 57% devido a baixa demanda de derivados do produto, menor desempenho de vendas no Brasil, e a disputa entre Arábia Saudita e Rússia.

Ainda, os dados ainda foram analisados a partir dos retornos logarítmicos, conforme Figuras (5.3) e (5.4), dos valores de fechamento das cotações diárias do dólar em relação ao real e das ações da Petrobrás em dólar.

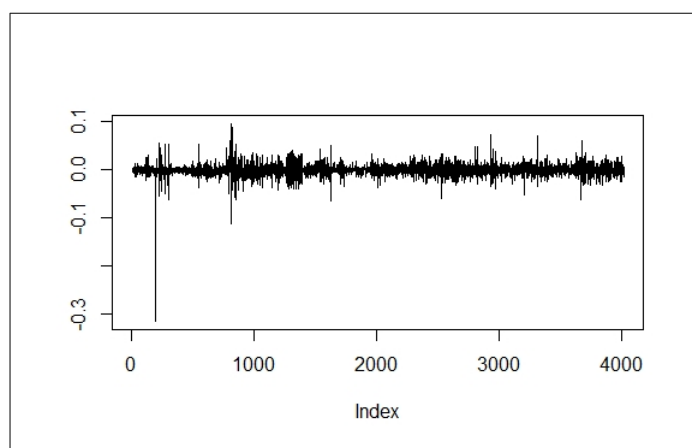


Figura 5.3: Distribuição de log-retornos - Dólar

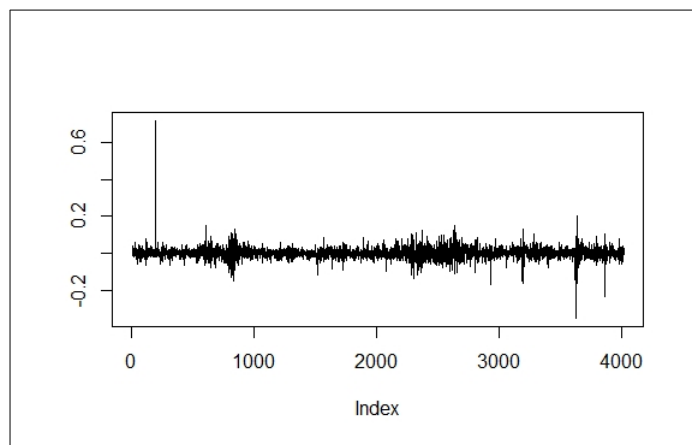


Figura 5.4: Distribuição de log-retornos - Petróleo

Ainda, foram elaborados os histogramas dos retornos para melhor visualização dos dados,

conforme Figuras (5.5) e (5.6) a seguir.

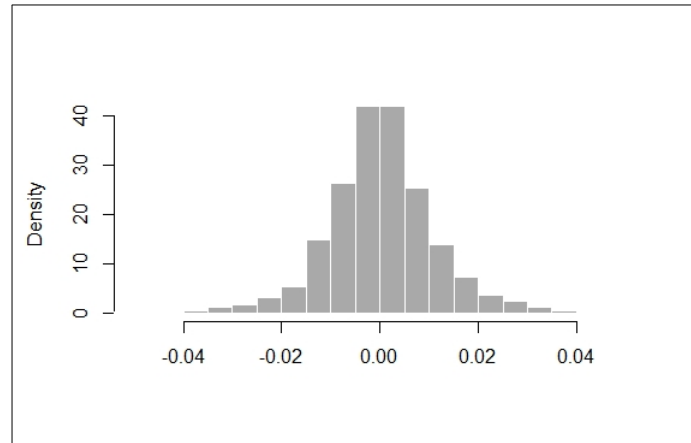


Figura 5.5: Histograma de log-retornos - Dólar

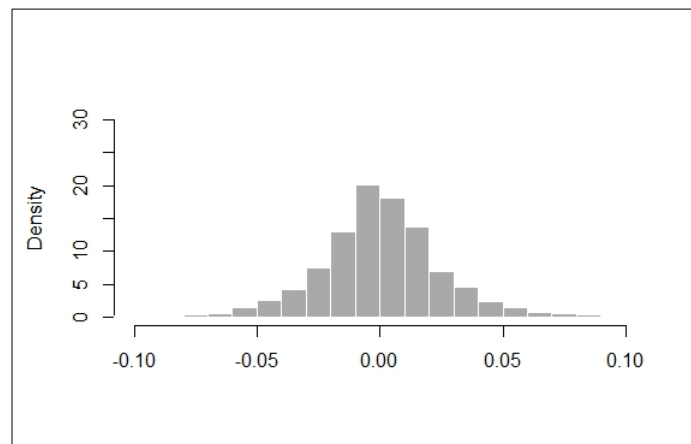


Figura 5.6: Histograma de log-retornos - Petróleo

Observa-se que para ambas variáveis há uma tendência aos valores se comportarem de forma aleatória em torno de uma média e que, de fato, em alguns momentos, como no início de 2020, houve uma variação mais extrema dos valores da cotação de ambos.

Pelo resumo dos dados, é possível notar que a média dos valores dos log-retornos é bem próxima de zero.

Tabela 5.1: Resumo dos dados - Dólar

Mínimo	1ª Quartil	Mediana	Média	3ª Quartil	Máximo
-0.3148314	-0.0056746	0.0000000	0.0001515	0.0060800	0.0966945

Tabela 5.2: Resumo dos dados - Petróleo

Mínimo	1ª Quartil	Mediana	Média	3ª Quartil	Máximo
-0.3523667	-0.0137843	0.0000000	0.0003036	0.0138190	0.7203695

A seguir serão apresentadas duas abordagens para aplicação da Distribuição Gumbel Bimodal. A primeira utilizando a teoria de Blocos Máximos para convergir os valores máximos e mínimos para a distribuição, e a segunda realizando um corte nos dados para que apenas as caudas (α e $1 - \alpha$) sejam utilizadas para realizar um cálculo alternativo para o VaR, também utilizando a nova distribuição.

5.3 Blocos Máximos

No trabalho também é utilizada uma técnica chamada de blocos máximos, que por sua vez possui o intuito de extrair uma sub-amostra dos retornos extremos. Como não é possível fazer a suposição de que os dados são independentes e identicamente distribuídos, foi realizado esse estudo dos blocos máximos e, a partir dele, é possível satisfazer a hipótese de independência dos dados.

O estudo de blocos máximos consiste em dividir as observações em blocos, como uma sub-amostra. Portanto, os valores da primeira sub-amostra, seguindo o ordenamento temporal, corresponderão ao primeiro grupo (M_1), e assim sucessivamente até a ocorrência do último bloco. Então, dessas sub-amostras, o dado de maior valor é selecionado, obtendo-se uma amostra de máximos. Da mesma forma pode ser feito para obter uma amostra de mínimos.

Para os valores de X_1, \dots, X_{5465} foi utilizado o teste de Ljung-Box com intuito de determinar o tamanho do bloco dos quais é retirada a sub-amostra de valores máximos. Então, para os dados da PETR4, o teste é aceito a um nível de confiança de 95% para o tamanho de bloco

20. Assim, a cada 20 retornos é extraído um valor máximo. Dessa forma, pode-se considerar que M_1, \dots, M_T é uma amostra aleatória de uma população $M \sim F_{GB}$. O mesmo ocorre para quando dos blocos são extraídos os valores mínimos.

Assim, a Figura 5.7 representa os histogramas dos mínimos e máximos, respectivamente, ajustados por uma estatística não paramétrica da curva. Já a partir da Figura 5.8, pode-se notar que M possui densidade bimodal.

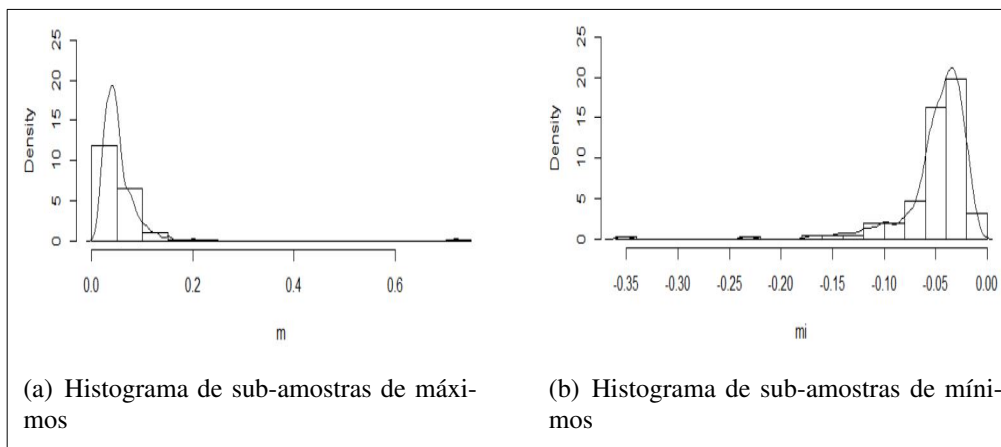


Figura 5.7: Histogramas de sub-amostras - Petróleo

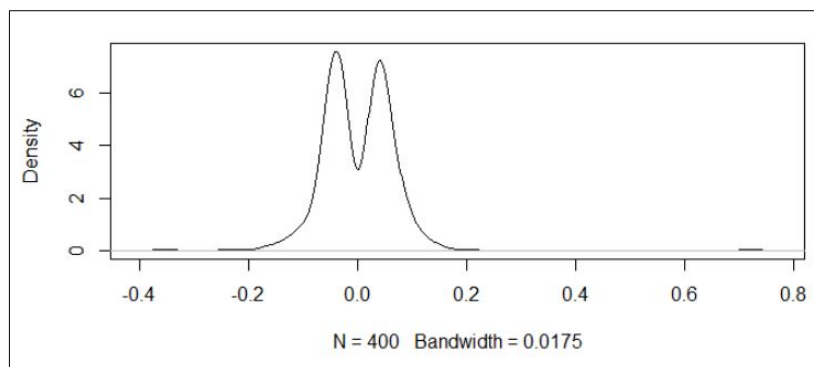


Figura 5.8: Sub-amostras - Petróleo

Analogamente, para a cotação do dólar o tamanho do bloco foi de $T = 10$. Os histogramas das sub-amostras de retornos extremos do dólar, assim como o gráfico 5.10, também indicam uma distribuição bimodal.

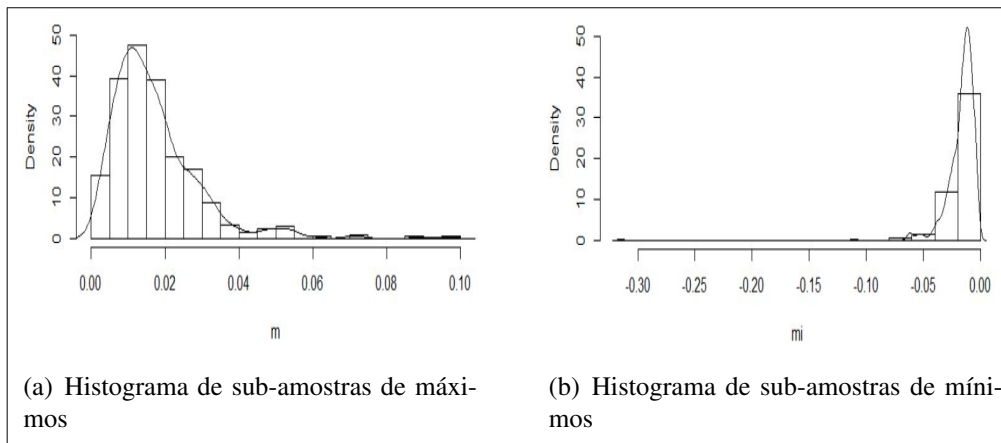


Figura 5.9: Histogramas de sub-amostras - Dólar

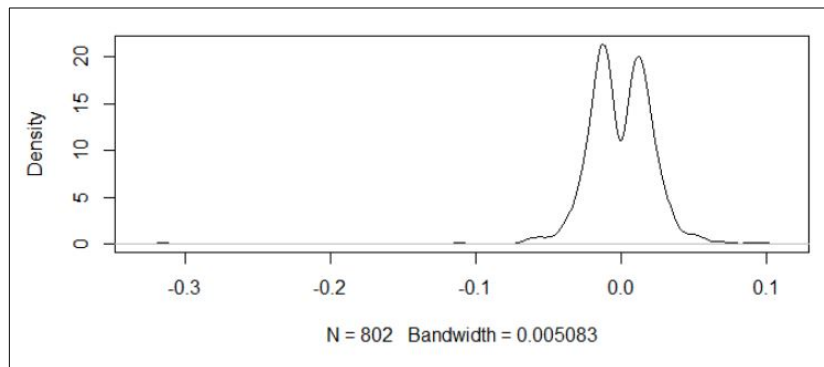


Figura 5.10: Sub-amostras - Dólar

Então, utilizando o método da máxima verossimilhança para estimação dos parâmetros e realizando o ajuste das subamostras conforme seus histogramas, é possível notar um bom ajuste dos dados pela distribuição Gumbel Bimodal.

Tabela 5.3: Estimativa dos parâmetros - Petróleo

$\hat{\mu}$	$\hat{\sigma}$	$\hat{\delta}$
0.000009962	0.000089877	2.246255323

Tabela 5.4: Estimativa dos parâmetros - Dólar

$\hat{\mu}$	$\hat{\sigma}$	$\hat{\delta}$
0.000016486	0.000099908	1.31295954

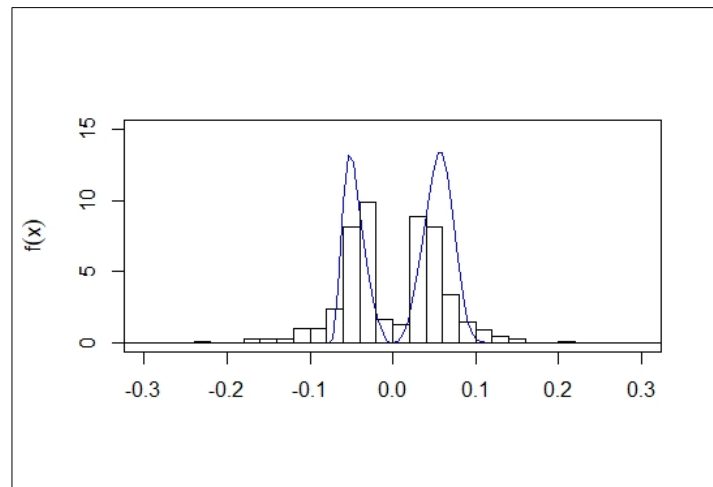


Figura 5.11: Densidade ajustada para os blocos - Petróleo

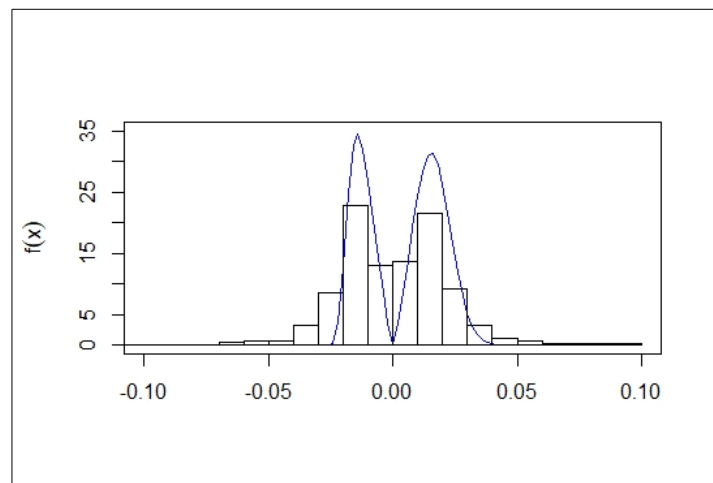


Figura 5.12: Densidade ajustada para os blocos - Dólar

5.4 Cálculo do VaR

Dentro da Teoria de Valores Extremos torna-se relevante realizar o cálculo do chamado *Value at Risk*, ou VaR. Essa estatística é um quantil que em finanças refere-se, por exemplo, ao valor do retorno máximo de um ativo com nível de confiança pré-determinado.

O valor em risco (VaR) é uma estatística que quantifica a extensão das possíveis perdas financeiras em uma empresa em um período de tempo específico. Essa métrica é mais comumente usada por bancos comerciais e de investimento para determinar a extensão e as probabilidades

de perdas potenciais em suas carteiras institucionais.

Existem diversas maneiras de se calcular o VaR. Por exemplo, pode-se utilizar o método histórico, que observa o comportamento de valores anteriores e os ordena das piores perdas aos maiores ganhos, partindo da premissa de que as informações anteriores predirão os resultados futuros.

Também é muito comum utilizar o método de variância-covariância. Em vez de presumir que o passado informará o futuro, esse método presume que os ganhos e as perdas são normalmente distribuídos. Dessa forma, as perdas potenciais podem ser enquadradas em termos de eventos de desvio padrão da média.

O problema da maioria dessas abordagens é utilizar o valor de perda (ou de ganho) e espelhar para o oposto, trabalhando com valor absoluto, ou seja, pressupondo a simetria dos dados. Dessa forma, obtém-se os mesmos valores para ganhos e perdas, sem considerar o que pode de fato ocorrer com os dois extremos da distribuição.

Assim, sugere-se a utilização da distribuição Gumbel Bimodal para modelar a distribuição das caudas direita e esquerda, demonstrando a particularidade de cada uma delas. Em outras palavras, a modelagem é utilizada para obtenção do VaR e realizada apenas com uma distribuição com a utilização de seus parâmetros para demonstrar a assimetria da caudas.

Para os dados do Petróleo, a Figura 5.13 demonstra a bimodalidade das caudas ao se utilizar apenas 5% dos dados em cada cauda. Reduzindo ainda mais o percentual de corte, para 2% em cada cauda, é possível notar que a bimodalidade dos extremos é mantida, conforme Figura 5.14.

Assim, utilizando a distribuição Gumbel Bimodal para modelar essas caudas com $\alpha = 5\%$ e, posteriormente, calcular o risco, foi realizada a estimativa dos parâmetros pelo método da máxima verossimilhança utilizando, novamente, como chutes iniciais a média, o desvio padrão e a moda para os parâmetros μ , σ e δ , respectivamente.

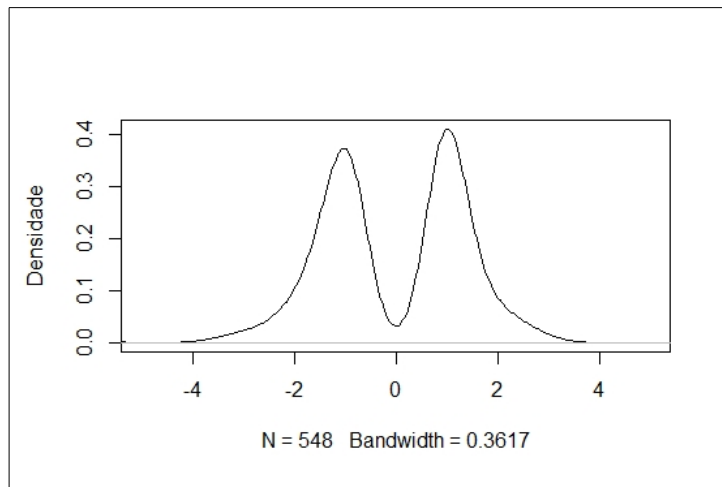


Figura 5.13: Bimodalidade das caudas: $\alpha = 5\%$ - Petróleo

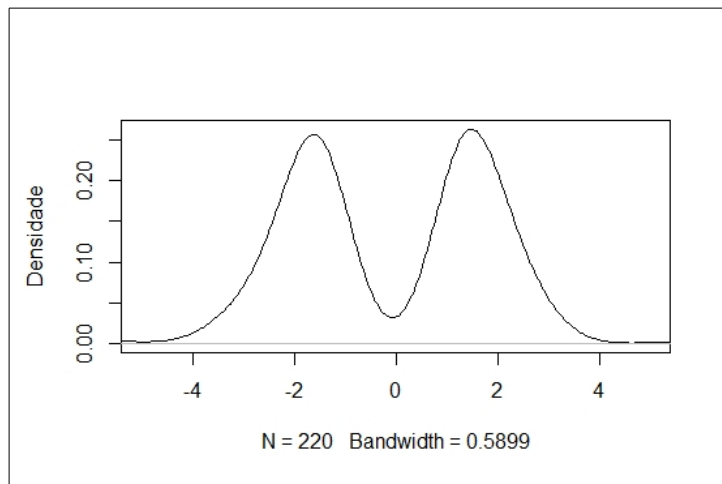


Figura 5.14: Bimodalidade das caudas: $\alpha = 2\%$ - Petróleo

Tabela 5.5: Estimativa dos parâmetros - Petróleo

$\hat{\mu}$	$\hat{\sigma}$	$\hat{\delta}$
1.0945884	3.5328335	2.3645777904

Para visualização do ajuste da distribuição aos dados, a Figura 5.15 apresenta o histograma das caudas juntamente com a densidade Gumbel Bimodal ajustada para os dados extremos.

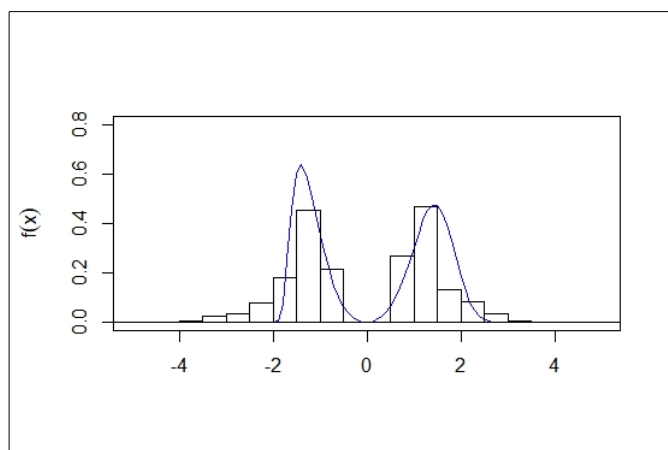


Figura 5.15: Densidade ajustada - Petróleo

Analogamente, para os dados da variação cambial, é possível observar a bimodalidade das caudas ao realizar o corte em 90% dos dados, deixando 5% em cada cauda.

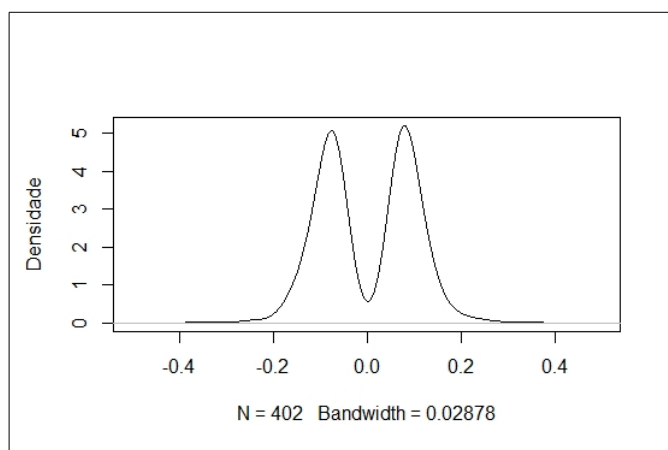


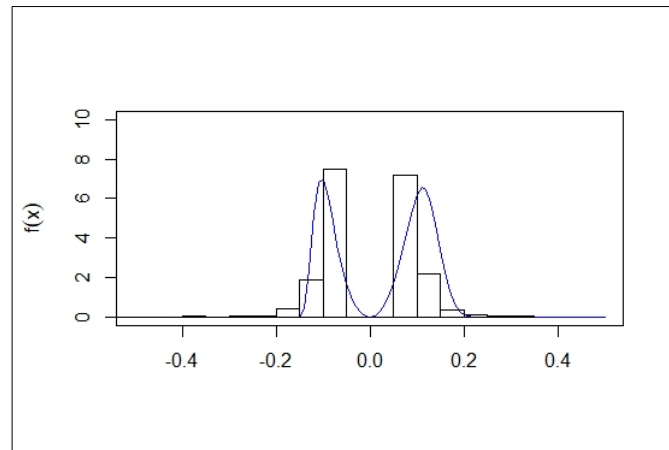
Figura 5.16: Bimodalidade das caudas: $\alpha = 5\%$ - Dólar

Percebe-se, novamente utilizando o método da máxima verossimilhança, que as estimativas do parâmetro se adequam bem aos dados, explicando o comportamento das caudas e, consequentemente, dos valores extremos da distribuição.

A densidade que ajustar os histogramas corresponde à distribuição de excessos acima do limiar u , ou seja, $P(X - u \leq x | |X| > u)$. Na prática, ao considerar os dados truncados com 90%, o limiar será $u = -0,5$.

Tabela 5.6: Estimativa dos parâmetros - Dólar

$\hat{\mu}$	$\hat{\sigma}$	$\hat{\delta}$
0.000132706	0.000830183	2.246012200

**Figura 5.17:** Densidade ajustada - Dólar

Nas Figuras 5.15 e 5.17 mostra-se que a densidade Gumbel Bimodal é uma distribuição dos excessos acima de 0,5 e abaixo de -0,5. Isto é, a densidade das Figuras explica perdas e ganhos abaixo e acima de 5%. Esse limiar pode ser escolhido ou estimado de acordo com o interesse do especialista.

Capítulo 6

Conclusão

A distribuição Gumbel Bimodal, proposta neste trabalho, buscou modelar dados extremos que apresentam bimodalidade.

Inicialmente, foram apresentadas as principais fórmulas da nova distribuição, assim como algumas de suas medidas descritivas. Em seguida foi realizada uma simulação de Monte Carlo com o objetivo de avaliar os estimadores, em que foi possível observar cada um de seus comportamentos, que se aproximaram dos verdadeiros valores da distribuição, sendo, então, eficientes e adequados.

Por fim, foi realizada uma aplicação em dados reais considerando as ações da Petrobrás e a variação cambial do dólar frente ao real. As informações foram analisadas sob a luz da técnica de blocos máximos e também utilizadas para aplicação da nova metodologia de cálculo do VaR proposta.

Conclui-se, de forma geral que, com base nas técnicas apresentadas, os dados de variação cambial do dólar frente ao real e o a variação do preço das ações da Petrobrás podem ser modeladas por essa nova distribuição.

Trata-se de um trabalho em aberto, com intuito futuro de investigar a relação entre a distribuição de máximos e mínimos e a distribuição dos excessos acima e abaixo de um limiar.

Bibliografia

- Beirland Jan e Goegebeur, Yuri e Teugels Jozef e Segers, Johan (ago. de 2005). *Statistics of Extremes: Theory and Applications*. ISBN: 0-471-97647-4. DOI: 10.1002/0470012382.
- Cordeiro G.M., Nadarajah S. e Ortega E.M.M (2012). “The Kumaraswamy Gumbel distribution”. *Stat Methods Appl* 21, 139 à 168.
- Dubey, Satya D. (1969). “A new derivation of the logistic distribution”. *Naval Research Logistics Quarterly* 16.1, pp. 37–40. DOI: <https://doi.org/10.1002/nav.3800160103>.
- Escalante-Sandoval, C. (set. de 2007). “A Mixed distribution with EV1 and GEV components for analyzing heterogeneous samples”. en. *Ingeniería, investigación y tecnología* 8, pp. 123–133. ISSN: 1405-7743.
- Nadarajah, Saralees (2006). “The exponentiated Gumbel distribution with climate application”. *Environmetrics* 17.1, pp. 13–23. DOI: <https://doi.org/10.1002/env.739>.
- Naghetini Mauro e Pinto, Éber (2007). *Hidrologia estatística*. CPRM.
- Nelsen, Roger B. (2006). *An Introduction to Copulas*. Springer. ISBN: 0-387-28659-4.
- Otiniano C.E.G. e Teixeira, E.C.M. (abr. de 2014). “Estimação dos parâmetros da mistura de duas componentes GEV via Algoritmo EM”. pt. *TEMA (São Carlos)* 15, pp. 59–71. ISSN: 2179-8451.
- Pinheiro, Eliane C (1999). “Contribuições em inferência e modelagem de valores extremos”. Tese de doutoramento. São Paulo: Universidade de São Paulo, Instituto de Matemática e Estatística.

Rinne, Horst (2008). *The Weibull Distribution. A handbook*. CRC Press.

Sansigolo, Clóvis Angeli (set. de 2008). “Distribuição de extremos de precipitação diária, temperatura máxima e mínima e velocidade do vento em Piracicaba, SP (1917-2006)”. pt. *Revista Brasileira de Meteorologia* 23, pp. 341 –346. ISSN: 0102-7786.

Sauer Ildo L. e Rodrigues, Larissa (dez. de 2016). “Pré-sal e Petrobras além dos discursos e mitos: disputas, riscos e desafios”. pt. *Estudos Avançados* 30, pp. 185 –229. ISSN: 0103-4014.

Sentyurina, Kristina (2019). “Política estatal petrolífera da Rússia e do Brasil no século XXI: os casos da Rosneft e da Petrobras”. pt.

Shaw, William e Buckley, Ian (fev. de 2009). “The alchemy of probability distributions: beyond Gram-Charlier expansions, and a skew-kurtotic-normal distribution from a rank transmutation map”.

Thomé Octavio e Bordeaux-Rego, Ricardo (jun. de 2014). “Avaliação de investimentos sob variação cambial e seus impactos sobre o retorno dos ativos no mercado de capitais: o caso da Petrobras”. pt. *Relatórios de Pesquisa em Engenharia de Produção* 14, pp. 105 –115.