

Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Estatística

Dissertação de Mestrado

Mistura de Distribuições Extremais

por

Frederico Lara de Souza

Orientador: Prof. Dra. Cira Etheowalda Guevara Otiniano

Maio de 2010

Frederico Lara de Souza

Mistura de Distribuições Extremas

Dissertação apresentada ao Departamento de Estatística do Instituto de Ciências Exatas da Universidade de Brasília como requisito parcial à obtenção do título de Mestre em Estatística.

Universidade de Brasília
Brasília, Maio de 2010

A todo mundo que acompanhou

Agradecimentos

Agradeço a todos que participaram de alguma forma na produção deste trabalho.

Sumário

Lista de Figuras	4
Lista de Tabelas	5
Resumo	6
Abstract	7
Introdução	8
1 Distribuições Extremais	12
1.1 Introdução	12
1.2 Distribuições Maximais	13
1.3 Densidades Extremais e suas medidas descritivas	15
1.3.1 Fréchet	15
1.3.2 Weibull	16
1.3.3 Gumbel	16
2 Misturas Finitas	18
2.1 Introdução	18
2.2 Modelo de Mistura Finita	18
2.2.1 Mistura de densidades Normais	20
2.2.2 Mistura de densidades Fréchet	21
2.2.3 Mistura de densidades Weibull	21
2.2.4 Mistura de densidades Gumbel	22

3	Propriedades	24
3.1	Introdução	24
3.2	Medidas Descritivas de uma Mistura	24
3.2.1	Fréchet	24
3.2.2	Weibull	25
3.2.3	Gumbel	26
3.3	Identificabilidade	28
3.3.1	Caracterização de identificabilidade	30
4	Algoritmo EM básico	36
4.1	Introdução	36
4.2	Algoritmo EM Básico	37
4.2.1	EM para Mistura	39
4.2.2	Propriedades do algoritmo EM	42
4.2.3	Escolha do valor inicial	44
4.3	EM para Mistura de Extremais	45
4.3.1	Fréchet	45
4.3.2	Weibull	46
4.3.3	Gumbel	48
5	Simulação	49
5.1	Fréchet	49
5.2	Weibull	50
5.3	Gumbel	52
6	Conclusão	57
	Referências Bibliográficas	58
A	Newton-Raphson	62
A.1	Convergência do Método de Newton-Raphson	63
A.2	Interpretação geométrica	63
B	Mistura de Fréchet	65

C	Mistura de Weibull	69
D	Mistura de Gumbel	73

Lista de Figuras

1.1	Curvas de densidades das distribuições Fréchet, Weibull e Gumbel, onde $\mu = 0$, $\sigma = 1$, $e\alpha = 1$	17
3.1	Mistura de Fréchet: da esquerda pra direita: $h(0,2;1;2;2;3)$, $h(0,5;1;2;2;3)$ e $h(0,2;2,5;2;2;2,9)$	26
3.2	Mistura de Weibull: da esquerda pra direita: $h(0,2;1;2;2;3)$, $h(0,5;1;2;2;3)$ e $h(0,2;2,5;2;2;2,9)$	27
3.3	Mistura de Gumbel: da esquerda pra direita: $h(0,2;-3;3;1;2;2;3)$, $h(0,5;-1;0;1;2;2;3)$ e $h(0,25;2;2;2,5;2;2;2,9)$	28
5.1	Gráfico da mistura de Fréchet $h(0,2;2,5;1;2;2,9)$ juntamente com as estimativas	51
5.2	Gráfico da mistura de Weibull $h(0,2;2,5;1;5;2,9)$ juntamente com as estimativas	53
5.3	Gráfico da mistura de Gumbel $h(0,5;-3;3;2;3)$ juntamente com as estimativas	55
5.4	Gráfico da função $\varphi'(\sigma_\ell)$	55

Lista de Tabelas

3.1	Mistura de Fréchet	25
3.2	Mistura de Weibull	27
3.3	Mistura de Gumbel	28
5.1	EQM e estimativas para uma mistura de duas Fréchet	50
5.2	EQM e estimativas para uma mistura de duas Weibull	52
5.3	EQM e estimativas para uma mistura de duas Gumbel	54

Resumo

Neste trabalho estudamos propriedades da mistura finita de duas componentes extremas da mesma classe e utilizamos o algoritmo EM como método de estimação dos parâmetros das misturas.

Primeiro calculamos as medidas como média, mediana, variância e moda de cada mistura, em seguida provamos a identificabilidade da classe de distribuição Weibull e Gumbel, pelo mesmo método já utilizado para provar a identificabilidade da classe de distribuições Fréchet.

Descrevemos o algoritmo EM como método de estimação dos parâmetros das misturas e finalmente para testar o algoritmo com os modelos descritos, com simulações estimamos os parâmetros das misturas e calculamos o erro quadrático médio para as estimativas dos parâmetro de forma das misturas.

Palavras Chave: *modelo de mistura distribuição extremal, algoritmo EM.*

Abstract

We study properties of finite mixture of two extremal components of the same class and use the EM algorithm as a method to estimate the parameters of the mixtures.

First we calculate the measures as mean, median, variance and mode of each mixture, then we prove the identifiability of the class of Weibull and Gumbel distribution using the same method already used to prove the identifiability of the class of Fréchet distributions.

Describe the EM algorithm as a method of parameter estimation of mixtures and finally to test the algorithm with the models described, with simulations we estimate the parameters of mixtures and calculate the mean square error for the estimated shape parameter of the mixtures.

key words: *extremal mixture model distribution, EM algorithm.*

Introdução

Conforme *Mclachlan e Peel* (1981) misturas finitas de distribuições permitem construir modelos de probabilidade de uma variedade de fenômenos aleatórios. Devido a seu uso como um método extremamente flexível de modelagem, as misturas finitas de distribuições continuaram a receber atenção crescente ao longo dos anos, na parte teórica e na prática.

Os Livros *Frühwirth-Schnatter* (2006), *Mclachlan e Peel*(1981) e *Titterington* (1985) descrevem modelos de mistura finita, propriedades e aplicações em diversas áreas como: biologia, genética, medicina, psiquiatria, economia, engenharia e marketing, entre muitos outros campos da física e ciências sociais. Nestas aplicações, os modelos de mistura finita dão suporte a uma variedade de técnicas importantes na estatística, incluindo cluster, análise de classe latente, análise discriminante, análise de imagem e análise de sobrevivência, além de seu papel mais direto na análise de dados e inferência, fornecendo modelos descritivos de distribuição (vide *Mclachlan e Peel* (1981)).

Mas como distribuições contínuas podem ser arbitrariamente aproximadas por uma mistura finita de densidades normais, os modelos de mistura proporcionam um modelo semi-paramétrico para estimar uma densidade desconhecida.

A utilidade das distribuições de mistura na modelagem de dados heterogêneos em um contexto de análise de cluster é óbvio, bem como no ajuste de dados, onde a distribuição empírica indica mais de uma componente.

Um dos primeiros trabalhos envolvendo a utilização de modelos de mistura foi realizado por Karl Pearson. Em seu trabalho clássico, *Pearson* (1894), ajustou uma mistura de duas densidades normal com médias (μ_1, μ_2) e variâncias (σ_1^2, σ_2^2) com proporções (p_1, p_2) . Este trabalho pode ter sido o primeiro a defender a análise es-

estatística como o principal método para estudar problemas biológicos. O conjunto de dados analisados por *Pearson* (1894) consistia no comprimento da testa para o corpo de 1.000 caranguejos da Baía de Nápoles. *Pearson* (1984) utilizou o método dos momentos para ajustar este modelo de mistura, o que exigiu um esforço computacional muito grande para a época. Várias tentativas foram feitas para simplificar a abordagem de *Pearson* (1984) no modelo de mistura de duas normais.

Mistura de Fréchet, log-normais, inversas gaussianas, Weibull (exponencial e Rayleigh) são modelos usados em confiabilidade (vide *Bucar et al* (2004), *Al-Hussani e Sultan et al*(2001) e *Sultan et al* (2006)). Na área da biologia *Mclachlan* (1988) aplica mistura de normais junto com componentes principais. Em fenômenos naturais como abalos sísmicos, precipitação de chuvas *Chen e Huang* (1995) e *Tartaglia* (2006) utilizam a mistura de Gumbel para tentar prever esses eventos.

Mistura de distribuições extremas, em estudos de análise de sobrevivência e em problemas da engenharia hidráulica, apareceram na última década em diversas publicações como: *Escalante-Sandoval* (2007), *Jiang et al* (2001), *Jiang et al* (1999), *Silva e Zochi* (2008) e *Sultan et al* (2006), entre outras.

As distribuições extremas são distribuições limite do máximo normalizado de variáveis aleatórias independentes e identicamente distribuídas $\{X_n\}$, e são de um dos três tipos: Fréchet, Weibull e Gumbel, cujas densidades correspondentes são:

$$f_\phi(x) = \alpha\sigma^{-\alpha}(x - \mu)^{-1-\alpha}e^{-[\sigma(x-\mu)]^{-\alpha}}, \quad x > \mu, \quad (1)$$

$$f_\psi(x) = \frac{\alpha}{\sigma^\alpha}(-x + \mu)^{\alpha-1}e^{-(-\frac{x-\mu}{\sigma})^\alpha}, \quad x < \mu \quad (2)$$

$$f_\Lambda(x) = \frac{1}{\sigma} \exp\left[-\frac{x - \mu}{\sigma} - e^{-\frac{x-\mu}{\sigma}}\right], \quad x \in \mathbb{R}. \quad (3)$$

Dada uma família de densidades

$$\mathcal{F} = \{F(x|\theta), \theta \in \Omega, x \in \mathbb{R}^d\}, \quad (4)$$

uma mistura finita de $f_1, \dots, f_k \in \mathcal{F}$ é qualquer combinação convexa

$$h(x|\Theta) = \sum_{j=1}^M p_j f_j(x|\theta_j), \quad (5)$$

onde p_1, \dots, p_k são chamados de proporções e f_1, \dots, f_k são as componentes da mistura.

Neste trabalho tratamos de modelos de mistura (5), cujas componentes são do tipo (1) - (3). Mostramos a identificabilidade da classe de misturas Fréchet, Weibull e Gumbel, analisamos medidas descritivas como média, mediana, moda e variância, e usando o algoritmo “Expectation Maximization” (EM) para estimar os parâmetros dos modelos de mistura.

No capítulo 1, apresentamos inicialmente as distribuições extremas, a relação com a distribuição de valor extremal generalizado (GEV), as relações entre elas e da distribuição Weibull não negativa. Apresentamos também as medidas descritivas das distribuições.

No capítulo 2, baseado em *Al-Hussani e Sultan et al* (2001), *Sultan et al* (2006), *Chandra* (1977), *Teicher* (1954), *Teicher* (1960), *Teicher* (1963), descrevemos um modelo de mistura finita cujas componentes são Normal, Fréchet, Weibull e Gumbel, detalhando a aplicabilidade de tais modelos.

No capítulo 3, calculamos as medidas descritivas para os modelos de misturas referidos no capítulo anterior, bem como sua identificabilidade (uma das mais importantes propriedades) segundo *Chandra* (1977).

No capítulo 4, mostramos o Algoritmo EM, método muito utilizado para calcular estimativas de máxima verossimilhança dos parâmetros de um modelo de mistura. Este método se destaca dos convencionais, pela sua eficiência em situações de dados faltantes e pela fácil adequação às modelos de misturas. Esse algoritmo consiste em 2 passos:

1. Passo E (expectância): Avaliação de $Q(\Theta, \Theta^{(k-1)}) = E[\ln f(\mathcal{X}, \mathcal{Y}|\Theta)|\mathcal{X}, \Theta^{(k-1)}]$
2. Passo M (maximização): Avaliação de $\Theta^{(k)}$, o valor de Θ que maximiza $Q(\Theta, \Theta^{(k-1)})$

Os modelos de mistura de duas densidades Fréchet, duas Weibull e duas Gumbel, são, respectivamente:

$$h(x|\Theta) = \sum_{j=1}^M p_j \alpha_j \sigma_j^{-\alpha_j} x^{-1-\alpha_j} e^{-(\sigma_j x)^{-\alpha_j}} \quad x > 0 \quad (6)$$

$$h(x|\Theta) = \sum_{j=1}^M p_j \frac{\alpha_j}{\sigma_j^{\alpha_j}} (-x)^{\alpha_j-1} e^{-(-x/\sigma_j)^{\alpha_j}} \quad x < 0 \quad (7)$$

$$h(x|\Theta) = \sum_{j=1}^M \frac{p_j}{\sigma_j} \left[-\frac{x - \mu_j}{\sigma_j} - e^{-\frac{x - \mu_j}{\sigma_j}} \right] \quad x \in \mathcal{R}. \quad (8)$$

As estimativas do modelo (6), por exemplo, usando o algoritmo EM são obtidas calculando iterativamente

$$\sigma_\ell^{(k)} = \left[\frac{\sum_{i=1}^N f(\ell|x_i, \Theta^{(k)})}{\sum_{i=1}^N f(\ell|x_i, \Theta^{(k)}) x_i^{-\alpha_\ell}} \right]^{-1/\alpha_\ell}$$

e resolvendo a equação

$$\sum_{i=1}^N f(\ell|x_i, \Theta^{(k)}) \left[\frac{1}{\alpha_\ell} - \log \sigma_\ell - \log x_i - (\sigma_\ell x_i)^{-\alpha_\ell} \log(\sigma_\ell x_i) \right] = 0.$$

Para esta última equação usamos o algoritmo Newton-Raphson. Neste capítulo também escrevemos sobre as propriedades do EM como também a escolha de valores iniciais.

No capítulo 5, mostramos os resultados das estimativas dos parâmetros de (6) - (8), utilizando o algoritmo EM, em tabelas e figuras.

Finalizamos o trabalho com nossa conclusão. Adicionamos dois apêndices, um para o algoritmo Newton-Raphson e outro para os códigos em R dos programas utilizados para estimação dos parâmetros dos modelos (6) -(8).

Capítulo 1

Distribuições Extremais

1.1 Introdução

A Teoria de Valores Extremos (TVE) tem aplicações nas áreas de manutenção de equipamentos, finanças, economia, desastres naturais entre outras. O fato dessas áreas precisarem desta teoria vem do interesse de analisar as medidas extremas, como $X_{(n)} = \max\{X_1, \dots, X_n\}$ e $X_{(1)} = \min\{X_1, \dots, X_n\}$, onde X_1, \dots, X_n são variáveis aleatórias.

No estudo da TVE, encontram-se distribuições, chamadas de distribuições extremas, que possibilitam obter a probabilidade de eventos que ocorrem com pouca frequência. Isto é possível fazendo-se estimação das caudas de tais distribuições.

Apresentamos a seguir um estudo para a estatística do máximo. Para o mínimo o desenvolvimento é análogo, sabendo que $\min\{X_1, \dots, X_n\} = -\max\{-X_1, \dots, -X_n\}$.

Inicialmente, apresentamos na seção 1.2 as distribuições maximais Fréchet, Weibull e Gumbel, bem como a GEV. Na seção 1.3 apresentamos algumas relações entre as distribuições extremas, e na seção 1.4 apresentamos o resultado do cálculo das medidas descritivas das distribuições extremas.

1.2 Distribuições Maximais

Sejam X_1, X_2, \dots, X_n variáveis aleatórias i.i.d. com função distribuição comum F . Se existirem constantes $\{b_n\}$ e $\{a_n\}$ positivas, tais que

$$\lim_{n \rightarrow \infty} P \left(\frac{X_{(n)} - b_n}{a_n} \leq x \right) = H(x), \quad \forall x \in C(H)$$

ou equivalentemente

$$\frac{X_{(n)} - b_n}{a_n} \longrightarrow H(x), n \xrightarrow{d} \infty, \quad (1.1)$$

onde $H(x)$ é função distribuição não degenerada, a distribuição H é denominada de distribuição extremal (maximal). Se a distribuição F de X_1, \dots, X_n satisfaz a condição (1.1), diz-se que F pertence ao domínio de atração de H e denota-se

$$F \in D_{max}(H).$$

O limite (1.1) sugere que se $F \in D_{max}(H)$, o comportamento caudal de F é como de H .

Gnedenko (1943) ao determinar as constantes estabilizantes $\{b_n\}$ e $\{a_n\}$ de (1.1), mostrou que a distribuição limite de $X_{(n)}$ é de um dos três tipos abaixo

$$\begin{aligned} \phi_{\alpha, \sigma, \mu}(x) &= \begin{cases} 0 & , x < \mu \\ \exp \left\{ - [\sigma(x - \mu)]^{-\alpha} \right\} & , x > \mu \end{cases} & \text{(Fréchet)} \\ \psi_{\alpha, \sigma, \mu}(x) &= \begin{cases} \exp \left\{ - \left(-\frac{x - \mu}{\sigma} \right)^{\alpha} \right\} & , x < \mu \\ 1 & , x > \mu \end{cases} & \text{(Weibull)} \\ \Lambda_{\sigma, \mu}(x) &= \exp \left\{ - \exp \left(-\frac{x - \mu}{\sigma} \right) \right\}, \quad \forall x, & \text{(Gumbel)} \end{aligned}$$

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

Uma forma generalizada para essas três distribuições é chamada de Distribuição de Valor Extremal Generalizada (GEV) com parâmetro de forma γ , escala $\sigma > 0$ e locação μ ,

$$H_{\gamma, \sigma, \mu}(x) = \begin{cases} \exp \left\{ - \left(1 + \gamma \frac{x - \mu}{\sigma} \right)^{-1/\gamma} \right\} & \text{se } \gamma \neq 0, \\ \exp \left\{ - \exp \left(-\frac{x - \mu}{\sigma} \right) \right\} & \text{se } \gamma = 0, \end{cases}$$

onde $1 + \gamma \frac{x - \mu}{\sigma} > 0$. O parâmetro γ é denominado de índice caudal e sua relação com o parâmetro de forma α das distribuições extremas é a seguinte:

Se $\gamma = \alpha^{-1} > 0$, tem-se a distribuição Fréchet, ou seja, $H_{\gamma, \sigma, \mu}(x) = \phi_{1/\gamma, \sigma, \mu}(\sigma + \gamma x)$;

Se $\gamma = -\alpha^{-1} < 0$, tem-se a distribuição Weibull, ou seja, $H_{\gamma, \sigma, \mu} = \psi_{-1/\gamma, \sigma, \mu}(-\sigma - \gamma x)$;

Se $\gamma = 0$, tem-se a distribuição Gumbel, ou seja, $H_{\gamma, \sigma, \mu} = \Lambda_{\sigma, \mu}$.

Observação:

(i) Embora, para fins de modelagem, as distribuições Λ , ϕ_α e ψ_α são muito diferentes, do ponto de vista matemático há uma correspondência entre elas. Suponha que X seja uma variável aleatória estritamente positiva, então

$$X \stackrel{d}{=} \phi_\alpha \Leftrightarrow \ln X^\alpha \stackrel{d}{=} \Lambda \Leftrightarrow -X^{-1} \stackrel{d}{=} \psi_\alpha,$$

(veja *Embrechts (1997)*).

(ii) A distribuição ϕ_α de uma variável aleatória X Fréchet é também chamada de inversa Weibull (*Sultan(2006)* e *Gusmão(2008)*). Isto acontece porque a distribuição Fréchet é a distribuição da variável aleatória $1/Y$, onde Y tem densidade

$$g(y; \alpha, \sigma) = \frac{\alpha}{\sigma^\alpha} y^{\alpha-1} e^{-\left(\frac{y}{\sigma}\right)^\alpha}, \quad (1.2)$$

comumente conhecida na estatística, em particular em análise de sobrevivência, como densidade de uma variável Weibull. Para tal, calculando o jacobiano de $Y = 1/X$, temos que:

$$f(x) = g\left(\frac{1}{x}\right) \frac{1}{x^2}$$

então

$$f(x) = \frac{\alpha}{\sigma^\alpha} \left(\frac{1}{x}\right)^{\alpha+1} e^{-\left(\frac{1}{x\sigma}\right)^\alpha}$$

é a densidade Fréchet mostrada posteriormente, com parâmetro de locação $\mu = 0$.

(iii) De (ii) concluímos que a distribuição Weibull extremal, $\Psi_{\alpha, \sigma, \mu}$ é distinta da distribuição Weibull associada a (1.2), porém, como

$$g(x) = f_\psi(-x), \quad x > 0 \quad (1.3)$$

onde $f_\psi(x)$ é dada por (1.5), a distribuição Weibull extremal também é chamada de Weibull negativa.

1.3 Densidades Extremas e suas medidas descritivas

1.3.1 Fréchet

A densidade associada a distribuição Fréchet ϕ_α é dada por

$$f_\phi(x) = \alpha\sigma^{-\alpha}(x - \mu)^{-1-\alpha}e^{-[\sigma(x-\mu)]^{-\alpha}}, \quad x > \mu, \quad (1.4)$$

onde $\mu \in R$ (parâmetro de localização), $\alpha > 0$ (parâmetro de forma) e $\sigma > 0$ (parâmetro de escala).

Se X é uma variável aleatória (v.a.) com distribuição Fréchet, a média e a variância, para $\mu = 0$, de X são

$$EX = \int_0^{+\infty} x f_\phi(x) dx = \frac{1}{\sigma} \Gamma(1 - 1/\alpha), \quad \alpha > 1$$

e

$$\text{var}(X) = \int_0^{+\infty} (x - EX)^2 f_\phi(x) dx = \frac{1}{\sigma^2} [\Gamma(1 - 2/\alpha) - \Gamma^2(1 - 1/\alpha)], \quad \alpha > 2$$

onde

$$\Gamma(r) = \int_0^{+\infty} x^{r-1} e^{-x} dx, \quad r > 0$$

é a função Gamma.

A moda de X , para $\mu = 0$, é obtida ao resolver a equação

$$f'_\phi(x) = 0,$$

e neste caso temos

$$\text{Moda} = \left(\frac{\alpha + 1}{\alpha} \right)^{1/\alpha}.$$

Finalmente, a mediana, de X , para $\mu = 0$, é a solução da equação

$$\int_0^M x f_\phi(x) dx = 0,5,$$

que é:

$$\text{Mediana} = \frac{(\ln 2)^{-1/\alpha}}{\sigma}.$$

1.3.2 Weibull

Denotamos a densidade associada a distribuição Weibull ψ_α por

$$f_\psi(x) = \frac{\alpha}{\sigma^\alpha}(-x + \mu)^{\alpha-1} e^{-\left(\frac{x-\mu}{\sigma}\right)^\alpha}, \quad x < \mu \quad (1.5)$$

onde $\mu \in \mathbb{R}$ (parâmetro de locação), $\alpha > 0$ (parâmetro de forma) e $\sigma > 0$ (parâmetro de escala).

Fazendo contas semelhantes as utilizadas na distribuição Fréchet também considerando $\mu = 0$, se X é uma v.a. com densidade Weibull,

$$E(X) = -\sigma\Gamma(1 + 1/\alpha),$$

$$\text{var}(X) = \sigma^2[\Gamma(1 + 2/\alpha) + \Gamma^2(1 + 1/\alpha)],$$

$$\text{Moda} = \sigma \left(\frac{\alpha - 1}{\alpha} \right)^{1/\alpha}$$

e

$$\text{Mediana} = (\ln 2)^{1/\alpha} \sigma.$$

1.3.3 Gumbel

Da mesma maneira, denotamos a densidade associada a distribuição Gumbel Λ por

$$f_\Lambda(x) = \frac{1}{\sigma} \exp \left[-\frac{x - \mu}{\sigma} - e^{-\frac{x-\mu}{\sigma}} \right], \quad x \in \mathbb{R} \quad (1.6)$$

onde $\mu \in \mathbb{R}$ (parâmetro de locação) e $\sigma > 0$ (parâmetro de escala).

Se X é uma v.a. com densidade Gumbel as medidas descritivas associadas, são:

$$E(X) = \mu - \xi\sigma,$$

onde $\xi = 0,543\dots$ (constante de Euler),

$$\text{var}(X) = \frac{1}{6}\sigma^2\pi^2;$$

$$\text{Moda} = 0$$

e

$$\text{Mediana} = \mu - \sigma[\ln(\ln 2)].$$

A figura 1.1 mostra as densidades extremas com parâmetros $\mu = 0$, $\sigma = 1$ e $\alpha = 1$.

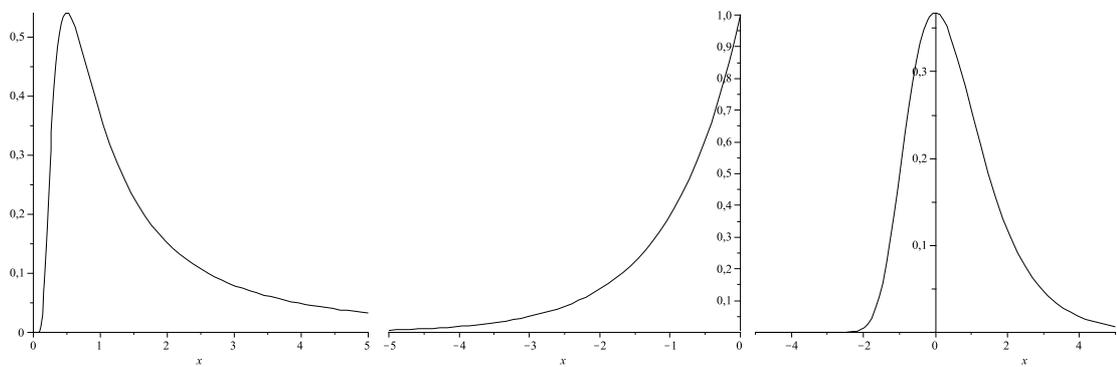


Figura 1.1: Curvas de densidades das distribuições Fréchet, Weibull e Gumbel, onde $\mu = 0$, $\sigma = 1$, e $\alpha = 1$

Capítulo 2

Misturas Finitas

2.1 Introdução

Devido a sua ampla utilidade na estatística e em outras áreas, os modelos de mistura finita receberam atenção crescente ao longo dos anos, tanto do ponto de vista prático quanto teórico. Domínios em que a mistura de modelos têm sido aplicados com sucesso incluem astronomia, biologia, genética, medicina, psiquiatria, economia, engenharia e marketing, entre outros.

Neste capítulo, na seção 2.2 apresentamos o modelo de mistura finita de densidades, que será utilizado nos capítulos seguintes. Como exemplos, na seção 2.3, escrevemos os modelos de mistura de Normal, Fréchet, Weibull e Gumbel, junto a resumo de aplicações.

2.2 Modelo de Mistura Finita

Suponha que $F(x|\theta)$ representa a função distribuição de uma variável aleatória X dado $\Theta = \theta$ e $G(\theta)$ representa a função distribuição da variável aleatória Θ . A distribuição marginal de (X, Θ) denotada por

$$H(x) = \int_{-\infty}^{+\infty} F(x|\theta)dG(\theta) \quad (2.1)$$

foi chamada por Fisher (1936) de distribuição composta de F e G . Teicher (1960) chamou H de mistura de F e G . $F(x|\theta)$ é conhecida como núcleo e G a distribuição de mistura.

Se toda a massa correspondente a G está concentrada em um número enumerável de pontos $\theta_1, \theta_2, \dots$ e as massas em $\theta_j (j = 1, 2, \dots)$ são $G(\theta_j)$, então (2.1) assume a forma

$$H(x) = \sum_{j=1}^{\infty} F(x|\theta_j)G(\theta_j). \quad (2.2)$$

Neste caso, $H(x)$ é uma função distribuição de mistura com componentes enumeráveis.

Se toda a massa correspondente a G está concentrada em um número finito de pontos $\theta_1, \theta_2, \dots, \theta_M$, então (2.1) se torna uma mistura finita de M componentes, cuja função distribuição é dada por

$$H(x) = \sum_{j=1}^M F(x|\theta_j)G(\theta_j). \quad (2.3)$$

Para simplificar a notação, $p_j \equiv G(\theta_j)$ e $F_j(x) = F(x|\theta_j)$, então (2.3) pode ser escrito como

$$H(x) = \sum_{j=1}^M p_j F_j(x) \quad (2.4)$$

onde $p_j \geq 0$, para $j = 1, 2, \dots$, e

$$\sum_{j=1}^M p_j = \sum_{j=1}^{\infty} G(\theta_j) = 1.$$

Em (2.4) p_j é conhecido como a j -ésima componente de proporção, $F_j(x)$ a j -ésima componente de mistura e M o número de componentes. Pode-se ter a impressão que, para todos os valores de j , todas as componentes da mistura pertencem a mesma família de distribuições. Porém isto nem sempre acontece, a única exigência é que $F_j(x)$ seja uma função distribuição.

Se X for contínua ou discreta, pode-se escrever $f(x|\theta)$ para representar a densidade condicional de X dado θ , e $h(x)$ para representar a densidade marginal de X . Se, em (2.1), $G(\theta)$ é absolutamente contínua, a densidade $g(\theta)$ existe de tal forma que $g(\theta) \equiv G'(\theta)$ e se $h(x)$ e $f(x|\theta)$ são as densidades correspondentes as distribuições $H(x)$ e $F(x|\theta)$, a partir de (2.1), tem-se

$$h(x) = \int_{-\infty}^{+\infty} f(x|\theta)g(\theta)d\theta. \quad (2.5)$$

Analogamente, as expressões (2.2) e (2.3) podem, respectivamente, ser escritas como

$$h(x) = \sum_{j=1}^{\infty} p_j f_j(x) \quad (2.6)$$

e

$$h(x) = \sum_{j=1}^M p_j f_j(x) \quad (2.7)$$

onde $f_j(x)$ é a j -ésima componente de mistura correspondente a distribuição $F_j(x)$

Neste trabalho será usada a seguinte notação para um modelo de mistura de M componentes

$$h(x|\Theta) = \sum_{j=1}^M p_j f_j(x|\theta_j), \quad (2.8)$$

onde $\sum_{j=1}^M p_j = 1$, $\Theta = (p_1, \dots, p_{M-1}, \theta_1, \dots, \theta_M)$ e θ_j são os parâmetros da j -ésima componente f_j , para $j = 1, \dots, M$.

Um modelo clássico de mistura de densidades é mostrado a seguir.

2.2.1 Mistura de densidades Normais

O modelo de mistura de M componentes normais com $\Theta = (p_j, \mu_j, \sigma_j)$ é:

$$h(x|\Theta) = \sum_{j=1}^M p_j \frac{1}{\sigma_j \sqrt{2\pi}} e^{-\left(\frac{x-\mu_j}{\sigma_j}\right)^2}.$$

A mistura de componentes normais foi um dos primeiros modelos a serem abordados para o estudo das propriedades e conceitos de mistura de distribuições. Sabe-se que um artigo pioneiro é de *Pearson* (1894) onde foi ajustada uma mistura de duas normais com médias e variâncias distintas para analisar dados de carangueijos, utilizando o método dos momentos como forma de estimação dos parâmetros. *Day*(1969) estudou as estimativas dos parâmetros da mistura de normais utilizando o método dos momentos e de máxima verossimilhança tanto para os casos onde as duas componentes possuíam matrizes de covariância iguais ou não. Nem sempre os estimadores de máxima verossimilhança apresentam boa consistência ou simplesmente não convergem para uma resposta devido a falta de restrições da função de verossimilhança. Para contornar este problema outros tipos de estimadores já foram utilizados como o algoritmo EM (*Redner e Walker*, 1984) e também os métodos quasi-Newton Raphson (*Fowlks*, 1979). *Hathaway* (1985) propôs restrições simples para um problema de

otimização que apresentaram uma solução fortemente consistente para os estimadores de máxima verossimilhança. *Al-Hussani e Sultan* (2001) utilizaram a mistura de Normal e outros tipos de misturas mostradas na área de confiabilidade, mostrando o comportamento tanto da taxa de falha quanto da função de sobrevivência.

2.2.2 Mistura de densidades Fréchet

O modelo de mistura de M densidades Fréchet de (1.4) com $\Theta = (p_j, \sigma_j, \alpha_j)$ é:

$$h(x|\Theta) = \sum_{j=1}^M p_j \alpha_j \sigma_j^{-\alpha_j} x^{-1-\alpha_j} e^{-(\sigma_j x)^{-\alpha_j}}.$$

A mistura de componentes Fréchet é aplicada, principalmente, em modelos dentro da área de análise de sobrevivência, *Sultan et al.* (2006) estudaram a taxa de falha e estimam os parâmetros do modelo através do algoritmo EM. *Jiang et al.* (2001) estudam três modelos incluindo a mistura, envolvendo duas distribuições de inversa Weibull estudando as formas das densidades e a função de taxa de falha e métodos gráficos para determinar qual deles ajusta melhor um certo banco de dados. *Jiang et al.* (1999) mostram que quando os componentes de um sistema seguem uma distribuição Weibull ou distribuição de inversa Weibull com parâmetro de forma comum, então o sistema pode ser representado por um modelo de mistura Weibull ou inversa Weibull onde as componentes de proporção podem assumir valores negativos.

2.2.3 Mistura de densidades Weibull

O modelo de mistura de M densidades Weibull extremal com $\Theta = (p_j, \sigma_j, \alpha_j)$ é:

$$h(x|\Theta) = \sum_{j=1}^M p_j \frac{\alpha_j}{\sigma_j^{\alpha_j}} (-x)^{\alpha_j-1} e^{-(x/\sigma_j)^{\alpha_j}}, \quad x < 0.$$

A mistura de M componentes Weibull não negativa com $\Theta = (p_j, \sigma_j, \alpha_j)$ é:

$$h(x|\Theta) = \sum_{j=1}^M p_j \frac{\alpha_j}{\sigma_j^{\alpha_j}} x^{\alpha_j-1} e^{-(x/\sigma_j)^{\alpha_j}}, \quad x > 0.$$

A mistura de componentes Weibull apresentada desta forma tem aplicabilidade principalmente na área de confiabilidade, *Bucar et al.* (2004) estudaram tempo de vida de sistemas que podem ser modelados por mistura finita de densidades Weibull com

proporções de mistura positivas, sem conhecer a estrutura do sistema. Para estimação dos parâmetros utilizaram quatro algoritmos: o algoritmo EM, um algoritmo alternativo, o algoritmo minimax e uma regressão multivariada. *Nagode e Fajdiga* (2000) utilizaram o modelo de mistura de Weibull ao invés do modelo com uma componente Weibull para a análise de dados de espectro de fadiga, pois a extrapolação dos dados e a previsão de dispersão são mais fáceis. Para estimar os parâmetros é mostrado um novo algoritmo que não possui a resolução de sistema de equações complicados. *Nagode et al* (2001) utilizaram dois tipos de modelos: a mistura entre uma Weibull e uma Normal e uma mistura de normais multivariadas bem como dois algoritmos para estimativa de parâmetros: o algoritmo EM e um algoritmo desenvolvido por Nagode e Fajdiga para estimar dados em matrizes de precipitação de chuva comparando-se estas estimativas.

2.2.4 Mistura de densidades Gumbel

O modelo de mistura de M densidades Gumbel de (1.6) com $\Theta = (p_j, \mu_j, \sigma_j)$ é:

$$h(x|\Theta) = \sum_{j=1}^M \frac{p_j}{\sigma_j} \left[-\frac{x - \mu_j}{\sigma_j} - e^{-\frac{x - \mu_j}{\sigma_j}} \right].$$

O modelo de mistura de componentes Gumbel é encontrado na literatura com aplicações no ajuste das ocorrências de fenômenos naturais como inundações, terremotos e precipitações pluviométricas. *Escalante-Sandoval* (2007) compara o ajuste a dados de cheia em estações de água no México utilizando uma mistura entre pares de distribuições extremas. *Chen e Huang* (1995) aplicaram uma mistura de componentes Gumbel para a análise do risco sísmico do continente chinês. Neste estudo, dividiu-se o continente chinês em quatorze zonas sísmicas com dados relativos a cinco anos. Ao mesmo tempo, a fim de testar a aplicabilidade e a confiabilidade das misturas de distribuições de valores extremos na previsão do risco sísmico do continente chinês, os dados de 1964 a 1993 de cada zona foram divididos em seis intervalos de acordo com um período de cinco anos intervalo. Os resultados da previsão das quatorze zonas nos seis intervalos foram testados a posteriori, a precisão das previsões combinadas e os limites para cada zona sísmica foram obtidos. *Tartaglia* (2006) também utilizou a mistura de Gumbel para modelar eventos hidrológicos extremos, no caso modelagem

de alturas máximas de precipitação pluviométrica utilizando o método L-momentos para estimação dos parâmetros. As expressões foram testadas na série temporal na Toscana, Itália. *Silva e Zocchi* (2008) mostraram ajustes da distribuição Gumbel e de misturas de duas distribuições Gumbel com parâmetros de escala iguais e diferentes valores em dados de velocidades máximas de vento dos meses de abril, maio, agosto e setembro em Piracicaba, SP, avaliando a qualidade de ajuste destes modelos.

Capítulo 3

Propriedades

3.1 Introdução

Iniciamos este capítulo, na seção 3.2, com os resultados da média, variância, moda e mediana da mistura de duas Fréchet, duas Weibull e duas Gumbel.

Na seção 3.3 apresentamos a propriedade da identificabilidade. Aplicamos a caracterização de *Chandra* (1977), para provar a identificabilidade das classes de mistura das distribuições Fréchet, Weibull e Gumbel.

3.2 Medidas Descritivas de uma Mistura

3.2.1 Fréchet

Considerando a expressão (1.4) com os parâmetros de escala σ e forma α (onde $\mu = 0$), a densidade da Fréchet é

$$f(x) = \alpha\sigma^{-\alpha}x^{-1-\alpha}e^{-(\sigma x)^{-\alpha}}, \quad x > 0$$

e o modelo (2.8) torna-se

$$h(x|\Theta) = \sum_{j=1}^M p_j \alpha_j \sigma_j^{-\alpha_j} x^{-1-\alpha_j} e^{-(\sigma_j x)^{-\alpha_j}}, \quad x < 0 \quad (3.1)$$

As média e variância da densidade $h(x|\Theta)$ são:

$$E(X) = \frac{p_1}{\sigma_1} \Gamma(1 - 1/\alpha_1) + \frac{p_2}{\sigma_2} \Gamma(1 - 1/\alpha_2), \text{ para } \alpha_1, \alpha_2 > 1; \quad (3.2)$$

$$\begin{aligned} \text{var}(X) = & \frac{p_1}{\sigma_1^2} [\Gamma(1 - 2/\alpha_1) - p_1\Gamma^2(1 - 1/\alpha_1)] + \\ & \frac{p_2}{\sigma_2^2} [\Gamma(1 - 2/\alpha_2) - p_2\Gamma^2(1 - 1/\alpha_2)] + \frac{p_1p_2}{\sigma_1^2\sigma_2^2} [\Gamma(1 - 1/\alpha_1)\Gamma^2(1 - 1/\alpha_1)], \\ & \alpha_1, \alpha_2 > 0 \end{aligned} \quad (3.3)$$

A moda é a solução, em x , de:

$$\sum_{j=1}^2 p_j \alpha_j \sigma_j^{-\alpha_j} x^{-2-\alpha_j} e^{-(x/\sigma_j)^{-\alpha_j}} [-1 - \alpha_j + (\sigma_j/x)^{-\alpha_j}] = 0, \quad (3.4)$$

e a mediana é a solução, em x , de:

$$\sum_{j=1}^2 p_j \exp\{-(\sigma_j/x)^{-\alpha_j}\} = 0,5. \quad (3.5)$$

Na tabela 3.1 mostramos os valores da média, variância, moda e a mediana da mistura de duas componentes Fréchet. Os valores utilizados foram escolhidos de modo a se obter casos unimodais e bimodais. Para resolver as equações não lineares (3.4) e (3.5) utilizamos o algoritmo Newton Raphson, apresentado no Apêndice A deste trabalho.

$h(p, \sigma_1, \sigma_2, \alpha_1, \alpha_2)$	Média	Variância	Moda	Mediana
$h(0,2;1;2;2;3)$	0,814	0,851	0,4591	0,6289
$h(0,5;1;2;2;3)$	0,945	0,856	0,4787	0,8045
$h(0,2;2,5;2;2;2,9)$	0,862	0,901	0,3265 e 0,8862	1,0397

Tabela 3.1: Mistura de Fréchet

3.2.2 Weibull

Considerando a expressão (1.5) com os parâmetros de escala σ e forma α (onde $\mu = 0$), a densidade da Weibull é

$$f(x) = \frac{\alpha}{\sigma^\alpha} (-x)^{\alpha-1} e^{-(-\frac{x}{\sigma})^\alpha}, \quad x < 0$$

e o modelo (2.8) torna-se

$$h(x|\Theta) = \sum_{j=1}^M p_j \frac{\alpha_j}{\sigma_j^{\alpha_j}} (-x)^{\alpha_j-1} e^{-(-x/\sigma_j)^{\alpha_j}}, \quad x < 0 \quad (3.6)$$

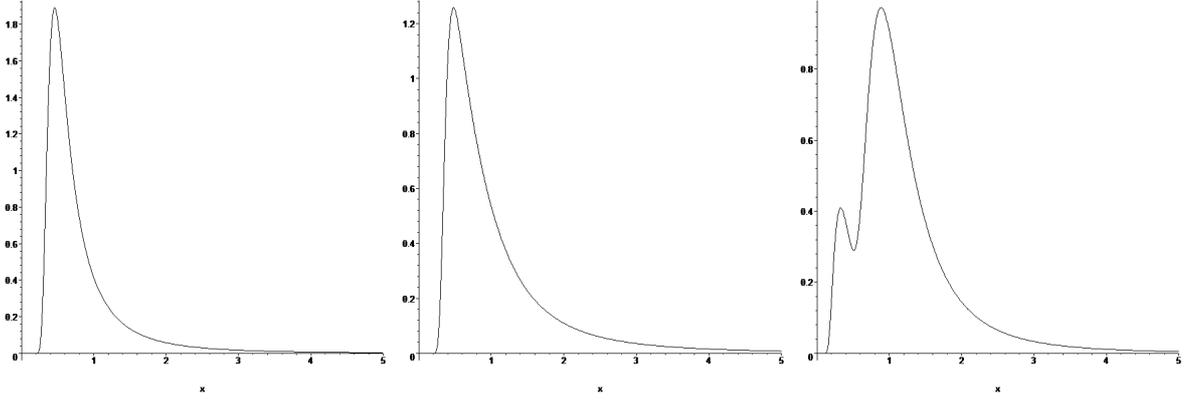


Figura 3.1: Mistura de Fréchet: da esquerda pra direita: $h(0,2;1;2;2;3)$, $h(0,5;1;2;2;3)$ e $h(0,2;2,5;2;2;2.9)$

A média e a variância da densidade $h(x|\Theta)$ são:

$$E(X) = -p_1\sigma_1\Gamma(1 + 1/\alpha_1) - p_2\sigma_2\Gamma(1 + 1/\alpha_2) \quad (3.7)$$

$$\begin{aligned} \text{var}(X) = & p_1\sigma_1^2[\Gamma(1 + 2/\alpha_1) + p_1\Gamma^2(1 + 1/\alpha_1)] + p_2\sigma_2^2[\Gamma(1 + 2/\alpha_2) + \\ & p_2\Gamma^2(1 + 1/\alpha_2)] + p_1p_2\sigma_1^2\sigma_2^2[\Gamma(1 + 1/\alpha_1)\Gamma^2(1 + 1/\alpha_2)] \end{aligned} \quad (3.8)$$

A moda é a solução, em x , de:

$$\text{Moda} : \sum_{j=1}^2 p_j \frac{\alpha_j}{\sigma_j} x^{\alpha_j-2} e^{-(x/\sigma_j)^{\alpha_j}} \left[1 - \alpha_j + \alpha_j \left(\frac{x}{\sigma_j} \right)^{\alpha_j} \right] = 0 \quad (3.9)$$

e a mediana é a solução, em x , de:

$$\text{Mediana} : \sum_{j=1}^2 p_j \exp\{-(-\sigma_j x)^{\alpha_j}\} = 0,5. \quad (3.10)$$

Na tabela 3.2 como na anterior também aparecem os valores da média, variância, moda e a mediana da mistura de duas componentes Weibull. Os valores também foram escolhidos de modo a se obter casos unimodais e bimodais e as equações não lineares (3.9) e (3.10) foi utilizado o algoritmo de Newton Raphson.

3.2.3 Gumbel

Considerando a expressão dada em (1.6) com os parâmetros de locação μ e escala σ , a densidade da Gumbel é

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

$h(p, \sigma_1, \sigma_2, \alpha_1, \alpha_2)$	Média	Variância	Moda	Mediana
$h(0,2;1;2;2;3)$	-1,606	5,613	-1,6909	-1,5899
$h(0,5;1;2;2;3)$	-1,336	4,005	-0,8905	-1,2428
$h(0,2;2,5;2;2;2,9)$	-1,869	9,201	-0,8662 e -2,7734	-0,9922

Tabela 3.2: Mistura de Weibull

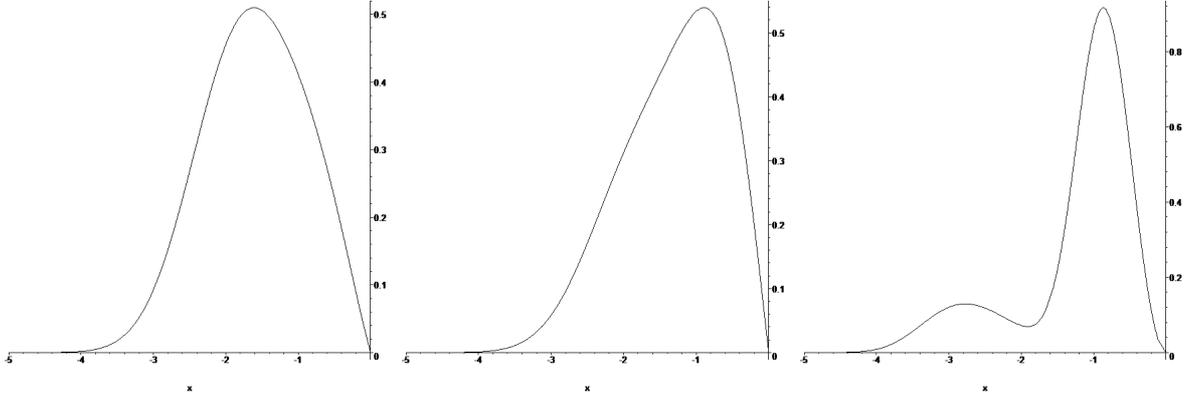


Figura 3.2: Mistura de Weibull: da esquerda pra direita: $h(0,2;1;2;2;3)$, $h(0,5;1;2;2;3)$ e $h(0,2;2,5;2;2;2,9)$

e o modelo (2.8) torna-se

$$h(x|\Theta) = \sum_{j=1}^M \frac{p_j}{\sigma_j} \left[-\frac{x - \mu_j}{\sigma_j} - e^{-\frac{x - \mu_j}{\sigma_j}} \right]. \quad (3.11)$$

A média e variância da densidade $h(x|\Theta)$ são:

$$E(X) = p_1(\mu_1 - \xi\sigma_1) + p_2(\mu_2 - \xi\sigma_2) \quad (3.12)$$

$$\begin{aligned} \text{var}(X) = & p_1 \left\{ \left[\frac{\pi^2 \sigma_1^2}{6} + (\mu_1 - \sigma_1 \xi) \right] - p_1(\mu_1 - \sigma_1 \xi)^2 \right\} + \\ & p_2 \left\{ \left[\frac{\pi^2 \sigma_2^2}{6} + (\mu_2 - \sigma_2 \xi) \right] - p_2(\mu_2 - \sigma_2 \xi)^2 \right\} + p_1 p_2 (\mu_1 - \sigma_1 \xi)(\mu_2 - \sigma_2 \xi) \end{aligned} \quad (3.13)$$

A moda é a solução, em x , de:

$$\sum_{j=1}^2 \frac{p_j}{\sigma_j^2} \left[-1 + e^{-(x - \mu_j)/\sigma_j} \right] \exp \left[-\frac{x}{\sigma_j} + e^{-(x - \mu_j)/\sigma_j} \right] = 0 \quad (3.14)$$

e a mediana é a solução, em x , de:

$$\sum_{j=1}^2 p_j \exp \left[-\exp \left(\frac{x - \mu_j}{\sigma_j} \right) \right] = 0,5 \quad (3.15)$$

Como anteriormente, mostra-se os valores da média, variância, moda e a mediana da mistura de duas componentes Gumbel na tabela 3.3. Também foram escolhidos valores para caracterizar tanto os casos unimodais quanto os bimodais e foram encontradas as soluções para as equações não lineares da moda e da mediana utilizando o algoritmo de Newton Raphson.

$h(p, \mu_1, \mu_2, \sigma_1, \sigma_2)$	Média	Variância	Moda	Mediana
$h(0,2;-3;3;2;3)$	0,279	10,67	-2,99 e 2,98	3,0419
$h(0,5;-1;0;2;3)$	-1,314	2,326	-0,8688	-0,1337
$h(0,25;2;2;2;2,9)$	0,547	12,48	2	2,4

Tabela 3.3: Mistura de Gumbel

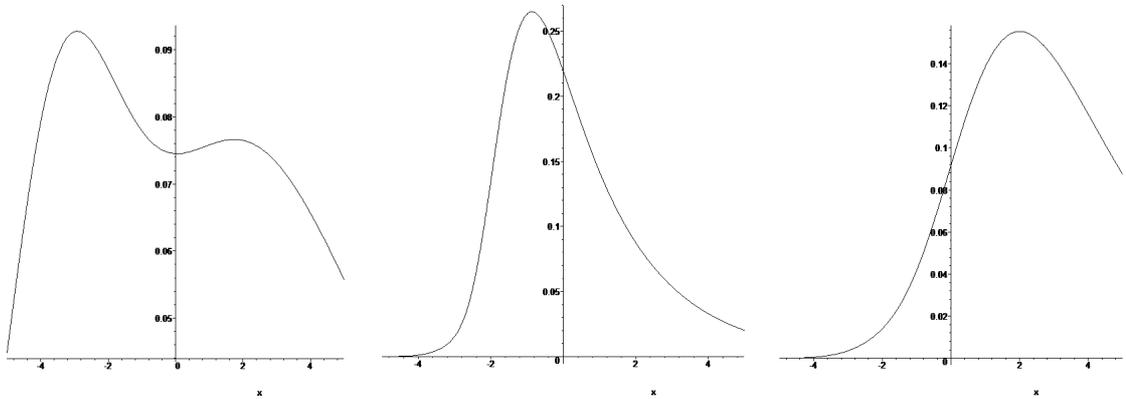


Figura 3.3: Mistura de Gumbel: da esquerda pra direita: $h(0,2;-3;3;1;2;2;3)$, $h(0,5;-1;0;1;2;2;3)$ e $h(0,25;2;2;2;2,5;2;2,9)$

3.3 Identificabilidade

A estimação de Θ baseado nas observações x só é relevante se Θ for identificável, pois a falta de identificabilidade pode ocasionar problemas especialmente no caso de uso de técnicas bayesianas onde simulações são feitas a partir da distribuição a posteriori obtida e utilizadas para fazer inferências do modelo de mistura.

Em síntese, dizemos que uma família paramétrica de densidades é identificável para Θ se valores de parâmetros distintos determinam membros distintos de uma

família de densidades, isto é,

$$f(x|\Theta) = f(x|\Theta') \forall x,$$

se e somente se $\Theta = \Theta'$.

Para este trabalho consideramos misturas finitas do tipo (2.8) onde as componentes pertencem a mesma família ou classe de distribuições.

Para definir a identificabilidade, primeiro consideramos a família

$$\mathcal{F} = \{F(x|\theta), \theta \in \Omega, x \in \mathbb{R}^d\} \quad (3.16)$$

chamada de classe de funções de distribuição, onde Ω é o espaço de parâmetros correspondente. Daí a classe

$$\mathcal{H} = \left\{ H(x) : H(x) = \sum_{j=1}^M p_j F_j(x|\theta_j), F_j(\cdot|\theta_j) \in \mathcal{F}, \forall j, M \in \mathbb{Z}^+ \right\} \quad (3.17)$$

é a classe de misturas de M componentes de \mathcal{F} .

Definição 3.3.1 (Identificabilidade). *Sejam*

$$h(x|\Theta) = \sum_{j=1}^M p_j f_j(x|\theta_j)$$

e

$$h'(x|\Theta') = \sum_{j=1}^{M'} p'_j f'_j(x|\theta'_j)$$

quaisquer dois membros da classe \mathcal{H} . Esta classe de misturas finitas é dita identificável para $\Theta \in \Omega$ se

$$h(x|\Theta) = h'(x|\Theta')$$

se e somente se $M = M'$, $p_j = p'_j$ e $f_j(x|\theta_j) = f'_j(x|\theta'_j)$ $j = 1, \dots, M$

Em síntese, dizemos que uma família H definida em (3.17) é identificável se qualquer membro desta família pode ser obtida de forma única.

Um exemplo clássico de classe de distribuições não identificável apresentado em *Titterington(1985)* é descrito a seguir.

Exemplo 3.1 (Mistura de duas binomiais). *Seja \mathcal{F} a classe das distribuições Binomial(2, θ), com $0 < \theta < 1$, e seja \mathcal{H} a classe de mistura de duas componentes de \mathcal{F} , com densidade*

$$h(x) = \sum_{i=1}^2 p_i \binom{2}{x} \theta_i^x (1 - \theta_i)^{2-x}, \quad x = 0, 1, 2$$

então

$$h(0) = p(1 - \theta_1)^2 + (1 - p)(1 - \theta_2)^2$$

$$h(1) = 2p\theta_1(1 - \theta_1) + 2(1 - p)\theta_2(1 - \theta_2)$$

Se considerarmos $\theta_1 = \theta_2 = \theta$

$$h(0) = (1 - \theta)^2$$

$$h(1) = 2p\theta(1 - \theta)$$

Substituindo o valor de $h(0)$ em $h(1)$

$$h(1) = 2\theta h(0)^{1/2}$$

ou seja, a solução não é única portanto a mistura não é identificável.

O problema de identificabilidade tem sido amplamente investigado e foi o foco de interesse nas últimas décadas. Os livros de *Everitt e Hand* (1981), *Mclanchlan e Basford* (1988), *Mclanchlan e Peel* (2000) são as mais populares referências sobre mistura de distribuições os quais abordam o problema da identificabilidade.

O método mais famoso para a identificabilidade de mistura de distribuições é dado pelo teorema de Teicher de 1963.

3.3.1 Caracterização de identificabilidade

Existem muitos estudos sobre identificabilidade usando o Teorema de Teicher tais como *Chandra* (1977), *Essam e Khalaf* (1981), *Khalaf* (1988), *Henna* (1994), *Atienza et al* (2006), *Karlis e Meligkotsilov* (2007) e *Shamilov e Asma* (2008 e 2009).

Neste trabalho usaremos o resultado de *Chandra* (1977). Esse resultado baseado no Teorema de Teicher (1963) dá um método para investigar a identificabilidade de uma classe de mistura de distribuições cujas componentes são da mesma família paramétrica de distribuições.

Atienza et al (2006) investigaram a identificabilidade de uma mistura de distribuições cujas componentes pertencem a diferentes famílias paramétricas.

Teorema:(Chandra, 1977) Suponha que para cada $F_j \in \mathcal{F}$ existe uma transformação φ_j com domínio definido por D_{φ_j} tal que a relação $M : F_j \rightarrow \varphi_j$ é linear. Suponha que também exista uma relação de ordem de F tal que

- (i) se $F_1 \leq F_2$, então $D_{\varphi_1} \subseteq D_{\varphi_2}$, e
- (ii) existe algum $t_1 \in T_1 = \{t : \varphi_1(t) \neq 0\}$ tal que

$$\lim_{t \rightarrow t_1} \frac{\varphi_2(t)}{\varphi_1(t)} = 0 \quad (3.18)$$

para cada $F_1 < F_2$. Então a classe \mathcal{H} das distribuições de \mathcal{F} é identificável.

Prova:

Suponha

$$\sum_{j=1}^M p_j F_j(\cdot) = 0, \quad F_j \in \mathcal{H}, \quad 1 \leq j \leq M \quad (3.19)$$

e que exista uma ordem onde $F_i < F_j$ para $i < j$. Então aplicando a transformação φ_j a (3.19) fica assim

$$\sum_{j=1}^M p_j \varphi_j(t) = 0.$$

Assim, para cada $t \in T_1 = \{t \in D_{\varphi_1} : \varphi_1(t) \neq 0\}$ (um conjunto não vazio por definição) e $j > 1$

$$p_1 + \sum_{j=2}^M p_j (\varphi_j(t)/\varphi_1(t)) = 0.$$

Fazendo $t \rightarrow t_1$, tem-se por (3.18) que $p_1 = 0$, então

$$\sum_{j=2}^M p_j F_j = 0.$$

Repetindo o argumento acima, conclui-se que

$$p_j = 0, \quad 1 \leq j \leq M \quad (3.20)$$

Exemplos da caracterização são apresentados abaixo para uma mistura de duas normais (*Teicher,1963*) e outro para uma mistura de duas Fréchet (*Sultan,2006*).

Exemplo 3.2. *Seja X uma v.a. associada a \mathcal{H} , a classe de misturas, de duas componentes de \mathcal{F} , onde a classe seja de densidades Normal.*

$$h(x) = \frac{p}{\sigma_1 \sqrt{2\pi}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} + \frac{(1-p)}{\sigma_2 \sqrt{2\pi}} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}.$$

A função distribuição j -ésima componente

$$F_j(x|\theta_j) = \Phi\left(\frac{x - \mu_j}{\sigma_j}\right).$$

Se considerarmos

$$F_1 < F_2 \quad \text{se } \sigma_1 > \sigma_2 \text{ e } \mu_1 = \mu_2$$

ou

$$F_1 < F_2 \quad \text{se } \sigma_1 = \sigma_2 \text{ e } \mu_1 < \mu_2$$

obtem-se uma relação de ordem das funções distribuições e sabendo que a transformada de Laplace bilateral da j -ésima componente é dada por

$$\varphi_j(t|\mu_j, \sigma_j^2) = \mathbb{E}(e^{-tX}) = \exp\{\sigma_j^2 t^2/2 - \mu_j t\}, \quad j = 1, 2,$$

então $D_{\varphi_j} = (-\infty, +\infty)$ para $j = 1, 2$ e podemos considerar $t_1 = +\infty$. Portanto da primeira relação estabelecida entre os parâmetros, tem-se que $D_{\varphi_1} \subseteq D_{\varphi_2}$.

Aplicando limite

$$\lim_{t \rightarrow t_1} \frac{\varphi_2(t)}{\varphi_1(t)} = \lim_{t \rightarrow +\infty} \frac{\exp\{\sigma_2^2 t^2/2 - \mu_2 t\}}{\exp\{\sigma_1^2 t^2/2 - \mu_1 t\}} = 0$$

Portanto a classe \mathcal{H} de mistura de duas normais é identificável.

Exemplo 3.3. Seja X uma v.a. associada a \mathcal{H} , a classe de misturas, de duas componentes de \mathcal{F} , onde a classe seja de densidades Fréchet com $\mu = 0$.

$$h(x) = p\alpha_1\sigma_1^{-\alpha_1}x^{-1-\alpha_1}\exp\{-(\sigma_1x)^{-\alpha_1}\} + (1-p)\alpha_2\sigma_2^{-\alpha_2}x^{-1-\alpha_2}\exp\{-(\sigma_2x)^{-\alpha_2}\}.$$

A função distribuição da j -ésima componente é

$$F_j(x|\theta_j) = \exp\left\{-\left(\frac{x}{\sigma_j}\right)^{-\alpha_j}\right\}, \quad x > 0.$$

Se considerarmos

$$F_1 < F_2 \quad \text{quando } \alpha_1 = \alpha_2 \text{ e } \sigma_1 > \sigma_2$$

ou

$$F_1 < F_2 \quad \text{quando } \sigma_1 = \sigma_2 \text{ e } \alpha_1 < \alpha_2.$$

obtem-se uma relação de ordem das funções distribuições e sabendo que a o t -ésimo momento da j -ésima componente de uma mistura de Fréchet é dado por

$$\varphi_j(t) = E(X^t) = \sigma_j^{-t} \Gamma \left(1 - \frac{t}{\alpha_j} \right), \text{ para } t < \alpha_j, j = 1, 2$$

então $D_{\varphi_j}(t) = (-\infty, \alpha_j)$ para $j = 1, 2$ e podemos considerar $t_1 = \alpha_1$. Portanto da primeira relação entre as F_j estabelecida, tem-se que $D_{\varphi_1}(t) \subseteq D_{\varphi_2}(t)$.

Aplicando limite,

$$\lim_{t \rightarrow \alpha_1} \varphi_1(t) = \lim_{t \rightarrow \alpha_1} \sigma_1^{-t} \Gamma \left(1 - \frac{\alpha_1}{\alpha_1} \right) = \infty$$

e

$$\lim_{t \rightarrow \alpha_1} \varphi_2(t) = \lim_{t \rightarrow \alpha_1} \sigma_1^{-t} \Gamma \left(1 - \frac{\alpha_1}{\alpha_2} \right) < \infty$$

Portanto

$$\lim_{t \rightarrow \alpha_1} \frac{\varphi_2(t)}{\varphi_1(t)} = 0.$$

O que verifica (3.18) para a classe \mathcal{H} de misturas de duas densidades de Fréchet.

Para os exemplos a seguir provamos a identificabilidade da classe de mistura de duas componentes Gumbel e para uma mistura de duas componentes Weibull, usando o mesmo teorema dos exemplos anteriores. No caso Gumbel a função $\varphi_j(t)$ é a função geradora de momentos e para o caso Weibull é utilizada $\varphi_j(t)$ como o t -ésimo momento.

Exemplo 3.4. Seja X uma v.a. associada a \mathcal{H} , a classe de misturas, de duas componentes de \mathcal{F} , onde a classe seja de densidades Weibull com $\mu = 0$.

$$h(x) = p \frac{\alpha_1}{\sigma_1^{\alpha_1}} (-x)^{\alpha_1-1} e^{-(x/\sigma_1)^{\alpha_1}} + (1-p) \frac{\alpha_2}{\sigma_2^{\alpha_2}} (-x)^{\alpha_2-1} e^{-(x/\sigma_2)^{\alpha_2}}$$

A função distribuição da j -ésima componente

$$F_j(x|\theta_j) = \exp \left\{ - \left(-\frac{x}{\sigma} \right)^\alpha \right\}.$$

Se considerarmos

$$F_1 < F_2 \text{ para } \alpha_1 < \alpha_2 \text{ se } \sigma_1 = \sigma_2$$

ou

$$F_1 < F_2 \text{ para } \sigma_1 > \sigma_2 \text{ se } \alpha_1 = \alpha_2$$

obtem-se uma relação de ordem das funções distribuições e sabendo que a o t -ésimo momento da j -ésima componente de uma mistura de Weibull é dado por

$$\varphi_j(t) = E(X^t) = (-1)^t \sigma_j^t \Gamma\left(1 + \frac{t}{\alpha_j}\right) \quad j = 1, 2,$$

então $D_{\varphi_j}(s) = (-\alpha_j, \infty)$ para $j = 1, 2$ e se considerar $t_1 = -\alpha_1$. Portanto da primeira relação estabelecida entre os parâmetros, tem-se que $D_{\varphi_1}(t) \subseteq D_{\varphi_2}(t)$.

Aplicando limite

$$\lim_{t \rightarrow -\alpha_1} \varphi_1(t) = \lim_{t \rightarrow -\alpha_1} (-1)^t \sigma_1^t \Gamma\left(1 - \frac{\alpha_1}{\alpha_1}\right) = \infty$$

e

$$\lim_{t \rightarrow -\alpha_1} \varphi_2(t) = \lim_{t \rightarrow -\alpha_1} (-1)^t \sigma_2^t \Gamma\left(1 - \frac{\alpha_1}{\alpha_2}\right) < \infty.$$

Portanto

$$\lim_{t \rightarrow -\alpha_1} \frac{\varphi_2(t)}{\varphi_1(t)} = 0$$

O que verifica (3.18) provando a identificabilidade da classe \mathcal{H} de misturas de duas densidades Weibull.

Exemplo 3.5. Seja X uma v.a. associada a \mathcal{H} , a classe de misturas, de duas componentes de \mathcal{F} , onde a classe seja de densidades Gumbel.

$$h(x) = \frac{p}{\sigma_1} \exp\left[-\frac{x - \mu_1}{\sigma_1} - e\left(-\frac{x - \mu_1}{\sigma_1}\right)\right] + \frac{1 - p}{\sigma_2} \exp\left[-\frac{x - \mu_2}{\sigma_2} - e\left(-\frac{x - \mu_2}{\sigma_2}\right)\right]$$

A função distribuição da j -ésima componente

$$F_j(x|\theta_j) = \exp\left\{-\exp\left(-\frac{x - \mu_j}{\sigma_j}\right)\right\}.$$

Então

$$F_1 < F_2 \quad \text{para } \sigma_1 > \sigma_2, \mu_1 = \mu_2$$

ou

$$F_1 < F_2 \quad \text{para } \sigma_1 = \sigma_2, \mu_1 > \mu_2$$

obtem-se uma relação de ordem das funções distribuições e sabendo que a função geradora de momentos da j -ésima componente de uma mistura de Gumbel é dado por

$$\varphi_j(t) = \Gamma(1 - \sigma_j t) e^{\mu_j t}, \quad j = 1, 2,$$

então $D_{\varphi_j}(t) = (-\infty, 1/\sigma_j)$ para $j = 1, 2$ e se considerar $t_1 = 1/\sigma_1$. Portanto da primeira relação entre os parâmetros estabelecida, tem-se que $D_{\varphi_1}(t) \subseteq D_{\varphi_2}(t)$.

Aplicando o limite

$$\lim_{t \rightarrow 1/\sigma_1} \varphi_1(t) = \lim_{t \rightarrow 1/\sigma_1} \Gamma(1 - \sigma_1/\sigma_1) e^{\frac{\mu_1}{\sigma_1}} = \infty$$

e

$$\lim_{t \rightarrow 1/\sigma_1} \varphi_2(t) = \lim_{t \rightarrow 1/\sigma_1} \Gamma(1 - \sigma_2/\sigma_1) e^{\frac{\mu_1}{\sigma_1}} < \infty.$$

Portanto

$$\lim_{t \rightarrow 1/\sigma_1} \frac{\varphi_2(t)}{\varphi_1(t)} = 0.$$

O que verifica (3.18) provando a identificabilidade da classe \mathcal{H} de misturas de duas densidades Gumbel.

Capítulo 4

Algoritmo EM básico

4.1 Introdução

O algoritmo EM é um método iterativo usado para calcular estimativas de máxima verossimilhança em situações que podem ser considerados como problemas de dados incompletos.

Seja X uma variável aleatória com densidade de probabilidade $f(x|\Theta)$, onde Θ é o vetor de parâmetros desconhecidos. Seja X_1, \dots, X_N uma amostra aleatória de tamanho N de X , cujos valores são dados por $\mathcal{X} = \{x_1, \dots, x_N\}$. Então a função de verossimilhança da amostra é a função de Θ

$$L(\Theta|\mathcal{X}) = \prod_{i=1}^N f(x_i|\Theta).$$

O estimador de máxima verossimilhança (EMV), Θ^* , de Θ , é o valor que maximiza $L(\Theta|\mathcal{X})$

$$\Theta^* = \arg \max_{\Theta} L(\Theta|\mathcal{X}),$$

isto é, o EMV é o valor de Θ que maximiza a probabilidade de ocorrência dos resultados amostrais.

Dependendo da forma de $f(x|\Theta)$ pode ser fácil ou difícil achar as estimativas de máxima verossimilhança de Θ . Em alguns casos essas estimativas não tem forma fechada, isto é, é necessário a utilização de algoritmos iterativos para solução de equações não-lineares, como por exemplo o método de Newton-Raphson. Em situações

onde dados incompletos comprometem a utilização de métodos iterativos conhecidos, o algoritmo EM é utilizado.

Parte da teoria do algoritmo EM foi mostrada em alguns trabalhos como de *Hartley* (1958), *Sundberg* (1976).

O algoritmo EM foi apresentado por *Dempster et al* (1977) como um algoritmo de grande generalidade e utilidade na situação de dados incompletos, cujo nome deve-se às duas fases iterativas do algoritmo: a fase E (Expectância) e a fase M (maximização).

Neste capítulo, primeiramente na seção 4.2 mostramos a forma básica do algoritmo, uma forma geral de se obter estimativas de máxima verossimilhança, mostrando suas propriedades e problemas com o valor inicial. Na seção 4.3 mostramos o algoritmo aplicado nas misturas de distribuições extremas.

4.2 Algoritmo EM Básico

Existem duas aplicações mais utilizadas do algoritmo EM. A primeira ocorre quando o conjunto de dados possui muitos valores faltantes, referente a problemas ou limitações na coleta de dados. A segunda acontece quando otimizar a função de verossimilhança é analiticamente intratável, porém a função de verossimilhança é simplificada adicionando valores para os dados faltantes. Esta última aplicação é mais comum em ciências da computação, reconhecimento de padrões, genética, etc.

Se assumirmos que \mathcal{X} é observado e gerado por alguma distribuição paramétrica $f(x|\Theta)$, chamamos \mathcal{X} de dados incompletos, estes dados são completados com $\mathcal{Y}(\{y_1, \dots, y_N\})$, então $\mathcal{Z} = (\mathcal{X}, \mathcal{Y})$ são os dados completos cuja densidade conjunta é

$$f(z|\Theta) = f(x, y|\Theta) = f(y|x, \Theta)f(x|\Theta) \quad (4.1)$$

Esta nova densidade conjunta é obtida utilizando a densidade marginal $f(x|\Theta)$ e a suposição de variáveis faltantes.

Com esta nova densidade, definimos a função de verossimilhança de dados completos

$$L(\Theta|z) = L(\Theta|x, y) = f(\mathcal{X}, \mathcal{Y}|\Theta). \quad (4.2)$$

A função $L(\Theta|x) = f(\mathcal{X}|\Theta)$ é chamada de função de verossimilhança de Θ dado os dados incompletos \mathcal{X} , e $l(\Theta|x) = \ln f(\mathcal{X}|\Theta)$ a log-verossimilhança de Θ dado \mathcal{X} .

Uma descrição detalhada do algoritmo EM pode ser encontrada em *Dempster et al* (1977), *Bilmes* (1998), *Da Silva* (2002), entre outros.

Baseados nestas referências, descrevemos o algoritmo EM como segue.

De (4.2), a log-verossimilhança dos dados incompletos pode ser escrita como

$$l(\Theta|x) = \ln f(\mathcal{X}|\Theta) = \ln \left(\int f(\mathcal{X}, y|\Theta) dy \right),$$

de onde se tem a igualdade

$$\ln \left(\frac{f(\mathcal{X}, \mathcal{Y}|\Theta)}{f(\mathcal{Y}|\mathcal{X}, \Theta)} \right) = \ln \left(\frac{f(\mathcal{X}, \mathcal{Y}|\Theta)f(\mathcal{X}|\Theta)}{f(\mathcal{X}, \mathcal{Y}|\Theta)} \right) = \ln f(\mathcal{X}|\Theta).$$

Com isso,

$$\ln \left(\frac{f(\mathcal{X}, \mathcal{Y}|\Theta)}{f(\mathcal{Y}|\mathcal{X}, \Theta)} \right) = \ln \left(\int f(\mathcal{X}, y|\Theta) dy \right)$$

e

$$l(\Theta|x) = \ln f(\mathcal{X}, \mathcal{Y}|\Theta) - \ln f(\mathcal{Y}|\mathcal{X}, \Theta) \quad (4.3)$$

Ao assumir um valor $\Theta^{(0)}$ de Θ e mantendo \mathcal{X} fixo,

$$l(\Theta|x) = \int l(\Theta|x) f(y|\mathcal{X}, \Theta^{(0)}) dy = E [l(\Theta|x)|\mathcal{X}, \Theta^{(0)}]. \quad (4.4)$$

Como queremos maximizar a log-verossimilhança dos dados incompletos $l(\Theta|x)$, é necessário eliminar os dados faltantes \mathcal{Y} . Uma maneira de fazer isto é tomando o valor esperado de $l(\Theta|x)$ com respeito a $f(\mathcal{Y}|\mathcal{X}, \Theta)$ como na equação (4.4). Usando (4.4) em (4.3)

$$l(\Theta|x) = E [\ln f(\mathcal{X}, \mathcal{Y}|\Theta)|\mathcal{X}, \Theta^{(0)}] - E [\ln f(\mathcal{Y}|\mathcal{X}, \Theta)|\mathcal{X}, \Theta^{(0)}].$$

Para simplificar a notação, defina

$$Q(\Theta, \Theta^{(0)}) = E[\ln f(\mathcal{X}, \mathcal{Y}|\Theta)|\mathcal{X}, \Theta^{(0)}] \quad (4.5)$$

e

$$H(\Theta, \Theta^{(0)}) = E[\ln f(\mathcal{Y}|\mathcal{X}, \Theta)|\mathcal{X}, \Theta^{(0)}], \quad (4.6)$$

então

$$l(\Theta|x) = Q(\Theta, \Theta^{(0)}) - H(\Theta, \Theta^{(0)}). \quad (4.7)$$

Veremos que para encontrar Θ que maximiza $l(\Theta|x)$ é suficiente maximizar o termo $Q(\Theta, \Theta^{(0)})$. Utilizando a desigualdade de Jensen, temos que $\forall \Theta, \bar{\Theta} \in \Omega$,

$$\begin{aligned}
H(\Theta, \bar{\Theta}) - H(\bar{\Theta}, \bar{\Theta}) &= \int \ln \left(\frac{f(\mathcal{Y}|\mathcal{X}, \Theta)}{f(\mathcal{Y}|\mathcal{X}, \bar{\Theta})} \right) f(\mathcal{Y}|\mathcal{X}, \bar{\Theta}) dy \\
&= \text{E} \left[\ln \left(\frac{f(\mathcal{Y}|\mathcal{X}, \Theta)}{f(\mathcal{Y}|\mathcal{X}, \bar{\Theta})} \right) | \mathcal{X}, \bar{\Theta} \right] \\
&\leq \ln \text{E} \left[\left(\frac{f(\mathcal{Y}|\mathcal{X}, \Theta)}{f(\mathcal{Y}|\mathcal{X}, \bar{\Theta})} \right) | \mathcal{X}, \bar{\Theta} \right] \\
&= \ln \int f(\mathcal{Y}|\mathcal{X}, \Theta) dy = 0,
\end{aligned} \tag{4.8}$$

Então

$$\begin{aligned}
&l(\Theta^{(k)}|x) - l(\Theta^{(k-1)}|x) \\
&= Q(\Theta^{(k)}, \Theta^{(k-1)}) - H(\Theta^{(k)}, \Theta^{(k-1)}) - Q(\Theta^{(k-1)}, \Theta^{(k-1)}) + H(\Theta^{(k-1)}, \Theta^{(k-1)}) \\
&= \{Q(\Theta^{(k)}, \Theta^{(k-1)}) - Q(\Theta^{(k-1)}, \Theta^{(k-1)})\} + \{H(\Theta^{(k-1)}, \Theta^{(k-1)}) - H(\Theta^{(k)}, \Theta^{(k-1)})\} \\
&\geq \{Q(\Theta^{(k)}, \Theta^{(k-1)}) - Q(\Theta^{(k-1)}, \Theta^{(k-1)})\} \geq 0.
\end{aligned} \tag{4.9}$$

Portanto, como a sequência $\{\Theta^{(k)}\}$ converge para o ponto crítico de $l(\Theta|x)$ (*Dempster et al.* 1977), de (4.9) temos que a sequência que maximiza $Q(\Theta|x)$, maximiza $l(\Theta|x)$.

Em síntese se \mathcal{X} denota os dados observados e \mathcal{Y} os dados faltantes, o parâmetro Θ estimado através do algoritmo EM é obtido através dos seguintes passos:

1. Passo E (expectância): Avaliação de $Q(\Theta, \Theta^{(k-1)}) = \text{E}[\ln f(\mathcal{X}, \mathcal{Y}|\Theta) | \mathcal{X}, \Theta^{(k-1)}]$
2. Passo M (maximização): Avaliação de $\Theta^{(k)}$, o valor de Θ que maximiza $Q(\Theta, \Theta^{(k-1)})$.

4.2.1 EM para Mistura

A estimação dos parâmetros da mistura de densidades é uma das aplicações mais utilizada do algoritmo EM. Neste caso consideramos a seguinte densidade

$$h(x) = \sum_{j=1}^M p_j f_j(x|\theta_j) \tag{4.10}$$

onde $\Theta = (p_1, \dots, p_{M-1}, \theta_1, \dots, \theta_M)$ é o vetor dos parâmetros, $\sum_{j=1}^M p_j = 1$ e cada componente da mistura f_j é uma densidade com parâmetro θ_j . A expressão da log-

verossimilhança incompleta de Θ dado $\mathcal{X} = \{x_1, \dots, x_N\}$ é

$$\log(L(\Theta|\mathcal{X})) = \log \prod_{i=1}^N h(x_i) = \sum_{i=1}^N \log \left(\sum_{j=1}^M p_j f_j(x_i|\theta_j) \right)$$

que é difícil de otimizar porque contém o log de uma soma de termos. Se considerarmos \mathcal{X} como incompleto e supondo a existência de dados faltantes $\mathcal{Y} = \{y_i\}_{i=1}^N$ cujos valores nos informam qual densidade componente gerou cada dado, a expressão da verossimilhança é simplificada. Isto é, assumimos que $y_i \in \{1, \dots, M\}$ para cada i , e $y_i = j$ se o i -ésimo dado foi gerado pela j -ésima componente da mistura.

Assim,

$$\begin{aligned} \log(L(\Theta|x, y)) &= \log(f(\mathcal{X}, \mathcal{Y}|\Theta)) \\ &= \log \left(\prod_{i=1}^N f(x_i, y_i|\Theta) \right) \\ &= \sum_{i=1}^N \log(f(x_i|y_i, \Theta)f(y_i|\Theta)) \\ &= \sum_{i=1}^N \log(p_i f_{y_i}(y_i|\theta_{y_i})) \end{aligned}$$

onde

$$f(y_i|\theta) = P(Y = y_i) = p_{y_i} \text{ e } f(x_i|y_i, \theta) = f_{y_i}(y_i|\theta_{y_i})$$

que, dependendo de qual for a componente densidade, pode ser otimizada utilizando várias técnicas.

Para contornar o fato da não existência de \mathcal{Y} que será assumido como sendo um vetor aleatório para proceder.

Primeiro devemos encontrar uma expressão para a distribuição dos dados faltantes. Assumimos que os parâmetros da mistura $\Theta^{(k)} = (p_1^{(k)}, \dots, p_k^{(k)}, \theta_1^{(k)}, \dots, \theta_k^{(k)})$ são apropriados para a verossimilhança $L(\Theta^{(k)}|\mathcal{X}, \mathcal{Y})$. Dado $\Theta^{(k)}$, podemos calcular $f_j(x_i|\theta_j^{(k)})$ para cada i e j . As proporções da mistura, p_j , podem ser consideradas como sendo probabilidades a priori para cada componente da mistura. Então, usando a regra de Bayes, podemos calcular

$$f(y_i|x_i, \Theta^{(k)}) = \frac{p_{y_i}^{(k)} f_{y_i}(x_i|\theta_{y_i}^{(k)})}{h(x_i|\Theta^{(k)})} = \frac{p_{y_i}^{(k)} f_{y_i}(x_i|\theta_{y_i}^{(k)})}{\sum_{j=1}^k p_j^{(k)} f_j(x_i|\theta_j^{(k)})}$$

e

$$f(\dagger|\mathcal{X}, \Theta^{(k)}) = \prod_{i=1}^N f(y_i|x_i, \Theta^{(k)})$$

onde $\dagger = (y_1, \dots, y_N)$ é um exemplo de dado faltante com espaço paramétrico Υ . Com isso foi possível encontrar a densidade marginal assumindo a existência de variáveis e escolhendo valores iniciais para os parâmetros da distribuição.

Neste caso a equação (4.5) fica assim

$$\begin{aligned} Q(\Theta, \Theta^{(k)}) &= \sum_{\mathbf{y} \in \Upsilon} \log(L(\Theta|\mathcal{X}, \dagger)) f(\dagger|\mathcal{X}, \Theta^{(k)}) \\ &= \sum_{\mathbf{y} \in \Upsilon} \sum_{i=1}^N \log(p_{y_i} f_{y_i}(x|\theta_{y_i})) \prod_{j=1}^N f(y_j|x_j, \Theta^{(k)}) \\ &= \sum_{y_i=1}^M \dots \sum_{y_N=1}^M \sum_{i=1}^N \log(p_{y_i} f_{y_i}(x|\theta_{y_i})) \prod_{j=1}^N f(y_j|x_j, \Theta^{(k)}) \\ &= \sum_{y_i=1}^M \dots \sum_{y_N=1}^M \sum_{i=1}^N \sum_{\ell=1}^k \delta_{\ell, y_i} \log(p_{\ell} f_{\ell}(x_i|\theta_{\ell})) \prod_{j=1}^N f(y_j|x_j, \Theta^{(k)}) \\ &= \sum_{\ell=1}^M \sum_{i=1}^N \log(p_{\ell} f_{\ell}(x_i|\theta_{\ell})) \sum_{y_i=1}^M \dots \sum_{y_N=1}^M \delta_{\ell, y_i} \prod_{j=1}^N f(y_j|x_j, \Theta^{(k)}) \end{aligned} \quad (4.11)$$

Nesta forma, $Q(\Theta, \Theta^{(k)})$ parece difícil, mas pode ser simplificada. Primeiro note que para $\ell \in \{1, \dots, M\}$,

$$\begin{aligned} &\sum_{y_i=1}^M \dots \sum_{y_N=1}^M \delta_{\ell, y_i} \prod_{j=1}^N f(y_j|x_j, \Theta^{(k)}) \\ &= \left(\sum_{y_1=1}^M \dots \sum_{y_{i-1}=1}^M \sum_{y_{i+1}=1}^M \dots \sum_{y_N=1}^M \prod_{j=1, j \neq i}^N f(y_j|x_j, \Theta^{(k)}) \right) f(\ell|x_i, \Theta^{(k)}) \\ &= \prod_{j=1, j \neq i}^N \left(\sum_{y_j=1}^M f(y_j|x_j, \Theta^{(k)}) \right) f(\ell|x_i, \Theta^{(k)}) = f(\ell|x_i, \Theta^{(k)}), \end{aligned} \quad (4.12)$$

pois $\sum_{j=1}^M f(j|x_i, \Theta^{(k)}) = 1$. Usando (4.12), podemos reescrever (4.11)

$$\begin{aligned} Q(\Theta, \Theta^{(k)}) &= \sum_{\ell=1}^M \sum_{i=1}^N \log(p_{\ell} f_{\ell}(x_i|\theta_{\ell})) f(\ell|x_i, \Theta^{(k)}) \\ &= \sum_{\ell=1}^M \sum_{i=1}^N \log(p_{\ell}) f(\ell|x_i, \Theta^{(k)}) + \sum_{\ell=1}^M \sum_{i=1}^N \log(f_{\ell}(x_i|\theta_{\ell})) f(\ell|x_i, \Theta^{(k)}) \end{aligned} \quad (4.13)$$

Podemos maximizar esta expressão em termos de p_{ℓ} e em termos de θ_{ℓ} independentemente.

Para encontrar a expressão de p_ℓ , será usado o multiplicador de Lagrange λ_p considerando a restrição $\sum_{\ell=1}^M p_\ell = 1$ e resolvendo a equação

$$\frac{\partial}{\partial p_\ell} \left[\sum_{\ell=1}^M \sum_{i=1}^N \log(p_\ell) f(\ell|x_i, \Theta^{(k)}) + \lambda_p \left(\sum_{\ell=1}^k p_\ell - 1 \right) \right] = 0$$

ou

$$\sum_{i=1}^N \frac{1}{p_\ell} f(\ell|x_i, \Theta^{(k)}) + \lambda_p = 0$$

Somando ambos os lados com relação a ℓ , temos que $\lambda_p = -N$ resultando em

$$p_\ell = \frac{1}{N} \sum_{i=1}^N f(\ell|x_i, \Theta^{(k)}).$$

Esta expressão resulta na atualização da componente de proporção da mistura (4.10). Para encontrar as atualizações dos parâmetros θ_ℓ da mistura basta maximizar a expressão (4.13) com relação θ_ℓ .

Em síntese, o algoritmo para uma mistura de distribuições é:

- Passo E, calcular

$$Q(\Theta, \Theta^{(k)}) = \sum_{\ell=1}^k \sum_{i=1}^N \log(p_\ell f_\ell(x_i|\theta_\ell)) f(\ell|x_i, \Theta^{(k)}) \quad (4.14)$$

onde

$$f(\ell|x_i, \Theta^{(k)}) = \frac{p_\ell^{(k)} f_\ell(x_i|\theta_\ell^{(k)})}{\sum_{j=1}^M p_j^{(k)} f_j(x_i|\theta_j^{(k)})}$$

- Passo M, determinar os estimadores de p_ℓ e θ_ℓ , usando

$$p_\ell^{(k)} = \frac{1}{N} \sum_{i=1}^N f(\ell|x_i, \Theta^{(k)}) \quad (4.15)$$

$$\sum_{\ell=1}^k \sum_{i=1}^N \frac{\partial}{\partial \theta_\ell^{(k)}} \log(f_\ell(x_i|\theta_\ell)) f(\ell|x_i, \Theta^{(k)}) = 0 \quad (4.16)$$

4.2.2 Propriedades do algoritmo EM

Alguns fatos observados no algoritmo EM, como razões para estabilidade numérica, convergência quase certa, facilidade em programar, bem como problemas como a convergência e intratabilidade no passo E, do algoritmo EM, podem ser encontradas em *Mclachlan e Krishnan (1997)*.

O algoritmo EM tem várias propriedades atraentes em relação a outros algoritmos iterativos, como Newton-Raphson e o método de escore de Fisher para encontrar estimadores de máxima verossimilhança. Algumas de suas vantagens em relação aos seus concorrentes são as seguintes:

- O algoritmo EM é numericamente estável, com cada iteração do EM aumentando a verossimilhança (exceto em um ponto fixo do algoritmo).
- Sob condições bastante gerais, o algoritmo EM alcança a convergência. Isto é, a partir de um ponto arbitrário $\Theta^{(0)}$ no espaço paramétrico, o algoritmo quase sempre encontra um máximo local, com exceção de má escolha do ponto inicial $\Theta^{(0)}$ ou alguma patologia na função de verossimilhança.
- O algoritmo EM é geralmente é fácil de programar e de ser implementado de forma fácil, pois se baseia em cálculos relacionados com os dados completos: passo E de cada iteração envolve apenas o cálculo dos valores esperados das distribuições marginais dos dados completos e o passo M de cada iteração requer apenas a estimação de máxima verossimilhança dos dados completos, que muitas vezes é na forma fechada.
- O algoritmo EM é leve computacionalmente falando, ou seja, requer pouco espaço de armazenamento.
- Quando a existe um padrão no uso do algoritmo EM, o passo M pode ser adaptado em pacotes estatísticos quando os estimadores de máxima verossimilhança dos dados completos não existem na forma fechada. Extensões do algoritmo EM, algoritmo EM generalizado (GEM) e o algoritmo maximização de expectância condicional (ECM) muitas vezes permitem que o passo M possa ser implementado iterativamente, de forma mais simples que no EM. Além disso, essas extensões do algoritmo também convergem de forma monótona e estável.
- O trabalho analítico necessário é muito mais simples do que com outros métodos, pois só é necessário maximizar o valor esperado condicional da log-verossimilhança dos dados completos.

- O custo por iteração é geralmente baixo, compensando o maior número de iterações necessárias para o algoritmo EM convergir em comparação com outros processos concorrentes.
- Ao observar o aumento na verossimilhança (se for facilmente calculável) durante as iterações, é fácil de controlar a convergência e erros de programação.
- O algoritmo EM pode ser usado para fornecer os valores estimados de dados missing.

Algumas das críticas do algoritmo EM são as seguintes:

- Ao contrário do método de escore de Fisher, o algoritmo EM não tem um processo integrado para produzir uma estimativa da matriz de covariância das estimativas de parâmetros. No entanto, esta desvantagem pode ser facilmente removida usando metodologia adequada associada com o algoritmo EM.
- O algoritmo EM pode convergir lentamente, mesmo em alguns problemas aparentemente simples e nos problemas que há muita informação incompleta.
- O algoritmo EM, como os métodos de Newton não garante a convergência para o máximo global quando existem vários máximos locais. Além disso, neste caso, a estimativa obtida depende do valor inicial. Mas, em geral, nenhum algoritmo de otimização é garantido a convergência a um máximo local ou global, e o algoritmo EM, não é diferente neste sentido. Existem outros procedimentos, como *simulated annealing* para enfrentar tais situações. No entanto, estas são complicadas de se aplicar.
- Em alguns problemas, o passo E pode ser analiticamente intratável, uma possibilidade de efetivá-lo seria de uma abordagem via Monte Carlo.

4.2.3 Escolha do valor inicial

Segundo *Karlis e Xekalaki* (2002), no algoritmo EM a escolha dos valores iniciais é de grande importância, pois ele pode influenciar significativamente a velocidade de convergência do algoritmo e sua capacidade de localizar o máximo global. *Laird* (1978) propôs uma pesquisa em diferentes tipos de valores para definir os valores

iniciais. *Leroux* (1992) sugeriu o uso de informações suplementares, a fim de formar grupos cujos recursos foram utilizados como valores iniciais. *McLachlan* (1988) propôs o uso da análise de componentes principais para a seleção de valores iniciais para o caso de misturas de distribuições multivariadas. Outra idéia é de *cluster* descrito por *Woodward et al.* (1984).

Finch et al. (1989) propôs que as componentes de proporção da mistura de duas componentes normais, deveria-se atribuir valores iniciais, então o resto dos parâmetros são calculados automaticamente com base nesses valores.

Outra escolha natural é utilizar as estimativas obtidas por outros métodos de estimação, como o método dos momentos, como em *Furman e Lindsay* (1994a, b) e *Lindsay e Basak* (1993) que consideram como valores iniciais para os parâmetros das misturas de componentes normais. *Fowlkes* (1979) propôs alguns gráficos e métodos *ad hoc* para a escolha de valores iniciais, no caso das misturas de normais. *Seidel et al.* (2000a, b, c), examinou algumas outras escolhas para os valores iniciais para misturas finitas de exponenciais. A prática tem mostrado que é preferível começar a partir de vários valores iniciais, a fim de garantir que a convergência seja obtida.

Nesta dissertação para determinar as estimativas dos valores iniciais pelo método dos momentos seria necessário obter o inverso da função gamma, portanto não é um método direto de conseguir estas estimativas, porém o método dos momentos modificado como o de momentos ponderados dado *Smith* (1990) pode ser aplicado. *Karlis e Xekalaki* (2002) sugerem utilizar como valores iniciais os valores utilizados para simular uma mistura de M componentes normais. Neste trabalho foi utilizado esta mesma idéia para misturas de extremas.

4.3 EM para Mistura de Extremas

4.3.1 Fréchet

Considerando a mistura de duas Fréchet, cuja densidade é dada por

$$h(x) = \sum_{j=1}^2 p_j \alpha_j \sigma_j^{-\alpha_j} x^{-1-\alpha_j} \exp\{-(\sigma_j x)^{-\alpha_j}\}, > 0$$

com parâmetros $\Theta = (p_\ell, \alpha_\ell, \sigma_\ell)$, $j = 1, 2$, temos nesse caso

$$\log(f_\ell(x|\theta_\ell)) = \log \alpha_\ell - \alpha_\ell \log \sigma_\ell + (-1 - \alpha_\ell) \log x - (\sigma_\ell x)^{-\alpha_\ell}.$$

O algoritmo para a mistura de Fréchet fica assim:

Passo E: usando (4.14) temos que

$$Q(\Theta, \Theta^{(k)}) = \sum_{\ell=1}^k \sum_{i=1}^N \log(p_\ell \alpha_\ell \sigma_\ell^{-\alpha_\ell} x^{-1-\alpha_\ell} \exp\{-(\sigma_\ell x)^{-\alpha_\ell}\}) f(\ell|x_i, \Theta^{(k)}), \quad (4.17)$$

onde

$$f(\ell|x_i, \Theta^{(k)}) = \frac{p_\ell^{(k)} \alpha_\ell \sigma_\ell^{-\alpha_\ell} x^{-1-\alpha_\ell} \exp\{-(\sigma_\ell x)^{-\alpha_\ell}\}}{\sum_{j=1}^2 p_j \alpha_j \sigma_j^{-\alpha_j} x^{-1-\alpha_j} \exp\{-(\sigma_j x)^{-\alpha_j}\}}.$$

Passo M: usando a equação (4.15), a componente de proporção é atualizada usando

$$p_\ell^{(k)} = \frac{1}{N} \sum_{i=1}^N \frac{p_\ell^{(k)} \alpha_\ell \sigma_\ell^{-\alpha_\ell} x^{-1-\alpha_\ell} \exp\{-(\sigma_\ell x)^{-\alpha_\ell}\}}{\sum_{j=1}^2 p_j \alpha_j \sigma_j^{-\alpha_j} x^{-1-\alpha_j} \exp\{-(\sigma_j x)^{-\alpha_j}\}}. \quad (4.18)$$

Para maximizar σ_ℓ utilizamos a equação (4.16)

$$\sum_{i=1}^N f(\ell|x_i, \Theta^{(k)}) \left[\frac{-\alpha_\ell}{\sigma_\ell} - \alpha_\ell \sigma_\ell^{-1-\alpha_\ell} x_i^{-\alpha_\ell} \right] = 0$$

e resolvendo esta equação obtemos a forma fechada para a atualização de σ_ℓ

$$\sigma_\ell^{(k)} = \left[\frac{\sum_{i=1}^N f(\ell|x_i, \Theta^{(k)})}{\sum_{i=1}^N f(\ell|x_i, \Theta^{(k)}) x_i^{-\alpha_\ell}} \right]^{-1/\alpha_\ell}. \quad (4.19)$$

Agora para maximizar α_ℓ , novamente utilizamos (4.16) e encontramos suas atualizações resolvendo a equação

$$\sum_{i=1}^N f(\ell|x_i, \Theta^{(k)}) \left[\frac{1}{\alpha_\ell} - \log \sigma_\ell - \log x_i - (\sigma_\ell x_i)^{-\alpha_\ell} \log(\sigma_\ell x_i) \right] = 0. \quad (4.20)$$

Como esta equação é não linear usamos o método de Newton-Raphson para determinar as atualizações de α_ℓ .

4.3.2 Weibull

Considerando a mistura de duas Weibull, cuja densidade é

$$h(x) = \sum_{j=1}^2 p_j \frac{\alpha_j}{\sigma_j^{\alpha_j}} (-x)^{\alpha_j-1} e^{-(x/\sigma_j)^{\alpha_j}}, x < 0$$

com parâmetros $\Theta = (p_\ell, \alpha_\ell, \sigma_\ell)$, $j = 1, 2$, tem-se

$$\log(f_\ell(x|\theta_\ell)) = \log \alpha_\ell - \alpha_\ell \log \sigma_\ell + (\alpha_\ell - 1) \log(-x) - \left(\frac{-x}{\sigma_\ell}\right)^{\alpha_\ell}.$$

O algoritmo para a mistura de Weibull fica assim

Passo E: usando (4.14) temos que

$$Q(\Theta, \Theta^{(k)}) = \sum_{\ell=1}^k \sum_{i=1}^N \log(p_\ell \frac{\alpha_\ell}{\sigma_\ell^{\alpha_\ell}} (-x)^{\alpha_\ell-1} e^{-(-x/\sigma_\ell)^{\alpha_\ell}}) f(\ell|x_i, \Theta^{(k)}), \quad (4.21)$$

onde

$$f(\ell|x_i, \Theta^{(k)}) = \frac{p_\ell^{(k)} \frac{\alpha_\ell}{\sigma_\ell^{\alpha_\ell}} (-x_i)^{\alpha_\ell-1} e^{-(-x_i/\sigma_\ell)^{\alpha_\ell}}}{\sum_{j=1}^2 p_j \frac{\alpha_j}{\sigma_j^{\alpha_j}} (-x_i)^{\alpha_j-1} e^{-(-x_i/\sigma_j)^{\alpha_j}}}.$$

Passo M: usando a equação (4.15), a componente de proporção é atualizada usando

$$p_\ell^{(k)} = \frac{1}{N} \sum_{i=1}^N \frac{p_\ell^{(k)} \frac{\alpha_\ell}{\sigma_\ell^{\alpha_\ell}} (-x_i)^{\alpha_\ell-1} e^{-(-x_i/\sigma_\ell)^{\alpha_\ell}}}{\sum_{j=1}^2 p_j \frac{\alpha_j}{\sigma_j^{\alpha_j}} (-x_i)^{\alpha_j-1} e^{-(-x_i/\sigma_j)^{\alpha_j}}}. \quad (4.22)$$

Para maximizar σ_ℓ utilizamos a equação (4.16)

$$\sum_{i=1}^N f(\ell|x_i, \Theta^{(k)}) \left[-\frac{\alpha_\ell}{\sigma_\ell} + \frac{\alpha_\ell (-x_i)^{\alpha_\ell}}{\sigma_\ell^{\alpha_\ell-1}} \right] = 0$$

e resolvendo esta equação obtemos a forma fechada para a atualização de σ_ℓ

$$\sigma_\ell^{(k)} = \left[\frac{\sum_{i=1}^N f(\ell|x_i, \Theta^{(k)}) (-x_i)^{\alpha_\ell}}{\sum_{i=1}^N f(\ell|x_i, \Theta^{(k)})} \right]^{1/\alpha_\ell}. \quad (4.23)$$

Agora para maximizar α_ℓ , novamente utilizamos (4.16) e encontramos suas atualizações resolvendo a equação

$$\sum_{i=1}^N f(\ell|x_i, \Theta^{(k)}) \left[\frac{1}{\alpha_\ell} - \log \sigma_\ell - \log(-x_i) - \left(\frac{-x_i}{\sigma_\ell}\right)^{\alpha_\ell} \log\left(\frac{x_i}{\sigma_\ell}\right) \right] = 0. \quad (4.24)$$

Como esta equação é também não linear usamos o método de Newton-Raphson para determinar as atualizações de α_ℓ .

4.3.3 Gumbel

Considerando a mistura de duas Gumbel cuja densidade é

$$h(x) = \sum_{j=1}^2 \frac{p_j}{\sigma_j} \exp \left[-\frac{x - \mu_j}{\sigma_j} - e^{-\left(\frac{x - \mu_j}{\sigma_j}\right)} \right]$$

com parâmetros $\Theta = (p_\ell, \mu_\ell, \sigma_\ell)$, $j = 1, 2$, tem-se

$$\log(f_\ell(x|\theta_\ell)) = -\log \sigma - \frac{x - \mu_\ell}{\sigma_\ell} - \exp \left\{ -\frac{x - \mu_\ell}{\sigma_\ell} \right\}.$$

O algoritmo para a mistura de duas Gumbel fica assim:

Passo E: usando (4.14) para $Q(\Theta, \Theta^{(k)})$

$$Q(\Theta, \Theta^{(k)}) = \sum_{\ell=1}^k \sum_{i=1}^N \log \left(\frac{p_\ell}{\sigma_\ell} \exp \left[-\frac{x_i - \mu_\ell}{\sigma_\ell} - e^{-\left(\frac{x_i - \mu_\ell}{\sigma_\ell}\right)} \right] \right) f(\ell|x_i, \Theta^{(k)}), \quad (4.25)$$

onde

$$f(\ell|x_i, \Theta^{(k)}) = \frac{\frac{p_\ell^{(k)}}{\sigma_\ell} \exp \left[-\frac{x_i - \mu_\ell}{\sigma_\ell} - e^{-\left(\frac{x_i - \mu_\ell}{\sigma_\ell}\right)} \right]}{\sum_{j=1}^2 \frac{p_j}{\sigma_j} \exp \left[-\frac{x_i - \mu_j}{\sigma_j} - e^{-\left(\frac{x_i - \mu_j}{\sigma_j}\right)} \right]}.$$

Passo M, usando a equação (4.15) a componente de proporção é atualizada usando

$$p_\ell^{(k)} = \frac{1}{N} \sum_{i=1}^N \frac{\frac{p_\ell^{(k)}}{\sigma_\ell} \exp \left[-\frac{x_i - \mu_\ell}{\sigma_\ell} - e^{-\left(\frac{x_i - \mu_\ell}{\sigma_\ell}\right)} \right]}{\sum_{j=1}^2 \frac{p_j}{\sigma_j} \exp \left[-\frac{x_i - \mu_j}{\sigma_j} - e^{-\left(\frac{x_i - \mu_j}{\sigma_j}\right)} \right]}. \quad (4.26)$$

Para maximizar μ_ℓ utilizamos a equação (4.16)

$$\sum_{i=1}^N f(\ell|x_i, \Theta^{(k)}) \left[\frac{1}{\sigma_\ell} - \frac{e^{-\left(\frac{x_i - \mu_\ell}{\sigma_\ell}\right)}}{\sigma_\ell} \right] = 0.$$

E resolvendo esta equação para obter a forma fechada para as atualizações de μ_ℓ , obtemos

$$\mu_\ell^{(k)} = \sigma_\ell \log \left[\frac{\sum_{i=1}^N f(\ell|x_i, \Theta^{(k)})}{\sum_{i=1}^N f(\ell|x_i, \Theta^{(k)}) e^{-x_i/\sigma_\ell}} \right]. \quad (4.27)$$

E agora para maximizar σ_ℓ , novamente utilizamos (4.16) e encontramos suas atualizações resolvendo a equação por Newton-Raphson,

$$\sum_{i=1}^N f(\ell|x_i, \Theta^{(k)}) \left[-\frac{1}{\sigma_\ell} + \frac{x_i - \mu_\ell}{\sigma_\ell^2} - e^{-\left(\frac{x_i - \mu_\ell}{\sigma_\ell}\right)} \left(\frac{x_i - \mu_\ell}{\sigma_\ell^2} \right) \right] = 0. \quad (4.28)$$

Capítulo 5

Simulação

Neste capítulo será mostrada a análise numérica dos estimadores para as classes de mistura de distribuições extremas e avaliá-los tanto em consistência quanto em viés. Para tal, foram geradas amostras de tamanho $n = (25, 50, 75, 100)$ para cada mistura de duas componentes. Na seção 5.1 geramos amostras aleatórias da mistura de duas Fréchet e estimamos seus parâmetros. Para a mistura de duas Weibull e para a mistura de duas Gumbel é apresentado na seção 5.2 e 5.3, respectivamente

5.1 Fréchet

Para gerar as amostras de uma classe de mistura de duas componentes Fréchet utilizamos o seguinte algoritmo:

- (1) Gera-se duas variáveis aleatórias u_1 e u_2 com distribuição $U(0, 1)$.
- (2) Se $u_1 < p$, então utilize u_2 e gere $x = F_1^{-1}(u_2)$ onde $F_1(x)$ é a função distribuição da variável x
- (3) Se $u_1 \geq p$, então utilize u_2 e gere $x = F_2^{-1}(u_2)$.

onde

$$F_j^{-1}(u) = \frac{(-\ln(u))^{1/\alpha_j}}{\sigma_j}, \quad j = 1, 2.$$

Para estimar os parâmetros Θ do modelo (3.1), utilizamos as expressões (4.18), (4.19) e (4.20) de modo iterativo, onde para encontrar a solução para o parâmetro α_ℓ foi

utilizado o método de Newton-Raphson (ver apêndice A). A tabela 5.1 mostra o valor esperado das estimativas dos parâmetros baseadas nos valores da mistura selecionada para a geração da amostra. Como critério de parada o algoritmo EM termina quando

$$\log L(\Theta^{(k+1)}) - \log L(\Theta^{(k)})$$

for menor que $n \times 10^{-5}$.

$h(p, \sigma_1, \sigma_2, \alpha_1, \alpha_2)$	n	E(<i>Estimativas</i>)					EQM	
		\hat{p}	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$
h(0,2;1;2;2;3)	25	0,239	0,932	2,045	3,578	3,684	8,75	1,923
	50	0,217	0,9914	2,019	2,734	3,281	3,17	0,494
	75	0,208	0,991	2,0065	2,537	3,1901	1,368	0,2409
	100	0,206	0,988	2,004	2,335	3,1498	0,681	0,177
h(0,5;1;2;2;3)	25	0,5076	0,971	2,02	2,575	3,979	1,562	3,774
	50	0,5033	0,988	2,012	2,252	3,4351	0,3802	1,0461
	75	0,4998	0,989	1,999	2,154	3,269	0,212	0,4471
	100	0,5023	0,988	2,007	2,1363	3,2102	0,1558	0,3401
h(0,2;2,5;1;2;2,9)	25	0,174	2,61	0,937	7,28	3,44	53,11	2,364
	50	0,1809	2,683	1,005	3,926	3,059	25,852	0,361
	75	0,184	2,661	1,001	2,687	2,99	2,062	0,161
	100	0,1878	2,621	1,002	2,467	2,947	1,1251	0,1162

Tabela 5.1: EQM e estimativas para uma mistura de duas Fréchet

Na figura 5.1 é apresentado o gráfico da mistura de duas Fréchet com parâmetros $\Theta = (0, 2; 2, 5; 1; 2; 2, 9)$ e o gráfico da mistura estimada, com os valores estimados dos parâmetros encontrados utilizando o algoritmo EM na amostra gerada.

5.2 Weibull

Para fazer o estudo da mistura de duas Weibull, o processo foi similar ao caso Fréchet, então foram geradas amostras para uma classe de mistura Weibull como se segue:

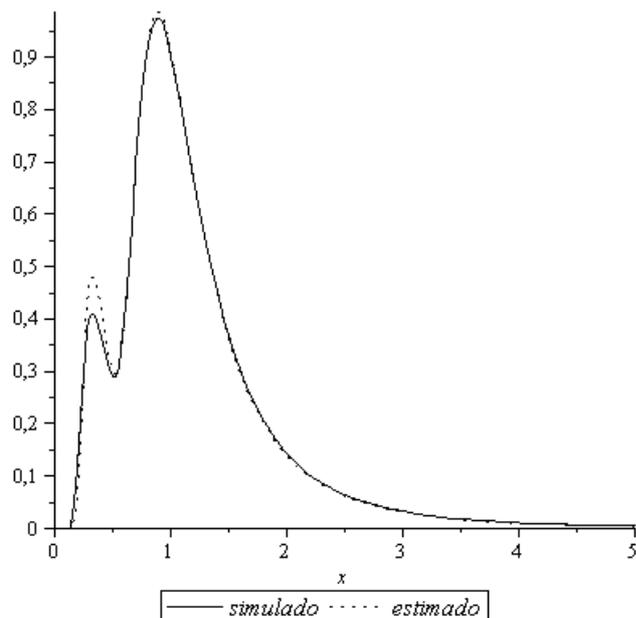


Figura 5.1: Gráfico da mistura de Fréchet $h(0, 2; 2, 5; 1; 2; 2, 9)$ juntamente com as estimativas

- (1) Gera-se duas variáveis aleatórias u_1 e u_2 com distribuição $U(0, 1)$.
- (2) Se $u_1 < p$, então utilize u_2 e gere $x = F_1^{-1}(u_2)$ onde $F_1(x)$ é a função distribuição da variável x
- (3) Se $u_1 \geq p$, então utilize u_2 e gere $x = F_2^{-1}(u_2)$.

onde

$$F_j^{-1}(u) = -\sigma_j(-\ln(u))^{1/\alpha_j}, \quad j = 1, 2.$$

Similar ao caso Fréchet utiliza-se as expressões (4.22), (4.23) e (4.24) referentes ao modelo (3.6), sendo também necessário estimar o parâmetro α_ℓ utilizando o método Newton-Raphson. Na tabela 5.2 é mostrado igualmente ao caso Fréchet os valores dos parâmetros utilizados para gerar as amostras, os valores esperados das estimativas e os EQM relacionados ao parâmetro de forma.

A figura 5.2 mostra o gráfico da mistura entre duas Weibull com parâmetros $\Theta = (0, 2; 2, 5; 1; 5; 2, 9)$ e o gráfico da mistura estimada com as estimativas dos parâmetros encontradas utilizando o algoritmo EM nos dados simulados, com os parâmetros propostos.

$h(p, \sigma_1, \sigma_2, \alpha_1, \alpha_2)$	n	E(<i>Estimativas</i>)					EQM	
		\hat{p}	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$
h(0,2;1;2;2;3)	25	0,1867	1,3832	2,0451	2,098	2,3305	4,525	4,893
	50	0,1713	1,48	2,098	2,638	3,472	5,682	0,5007
	75	0,162	1,499	2,025	2,546	3,394	3,684	0,346
	100	0,1523	1,521	2,0156	2,30	3,349	2,027	0,358
h(0,5;1;2;2;3)	25	0,413	1,244	2,068	3,254	4,06	7,94	1,712
	50	0,403	1,630	1,9406	2,5854	3,011	3,78	0,251
	75	0,449	1,43	1,935	2,398	2,95	2,26	0,137
	100	0,402	1,533	1,828	2,298	2,83	1,76	0,09
h(0,2;2,5;1;5;2,9)	25	0,508	2,811	0,977	8,02	3,133	48,47	1,49
	50	0,5035	2,785	0,975	6,89	3,018	4,47	0,45
	75	0,508	2,885	0,983	6,233	2,75	2,62	0,108
	100	0,502	2,832	0,997	5,69	3,168	1,634	0,185

Tabela 5.2: EQM e estimativas para uma mistura de duas Weibull

5.3 Gumbel

Foram geradas amostras para uma classe de mistura Gumbel utilizando o algoritmo similar ao utilizado no caso Fréchet e Weibull:

- (1) Gera-se duas variáveis aleatórias u_1 e u_2 com distribuição $U(0, 1)$.
- (2) Se $u_1 < p$, então utilize u_2 e gere $x = F_1^{-1}(u_2)$ onde $F_1(x)$ é a função distribuição da variável x
- (3) Se $u_1 \geq p$, então utilize u_2 e gere $x = F_2^{-1}(u_2)$.

onde

$$F_j^{-1}(u) = \mu_j - \sigma_j \ln(-\ln u)$$

Utilizando o modelo (3.11) e as expressões (4.26) e (4.27) encontra-se as estimativas via algoritmo EM para uma mistura entre duas Gumbel. Pelo fato de se não encontrar convergência do parâmetro de escala (σ_ℓ), foram somente estimados os parâmetros $\Theta = (p, \mu_j)$. Aplicando-se o mesmo princípio apresentado nos casos

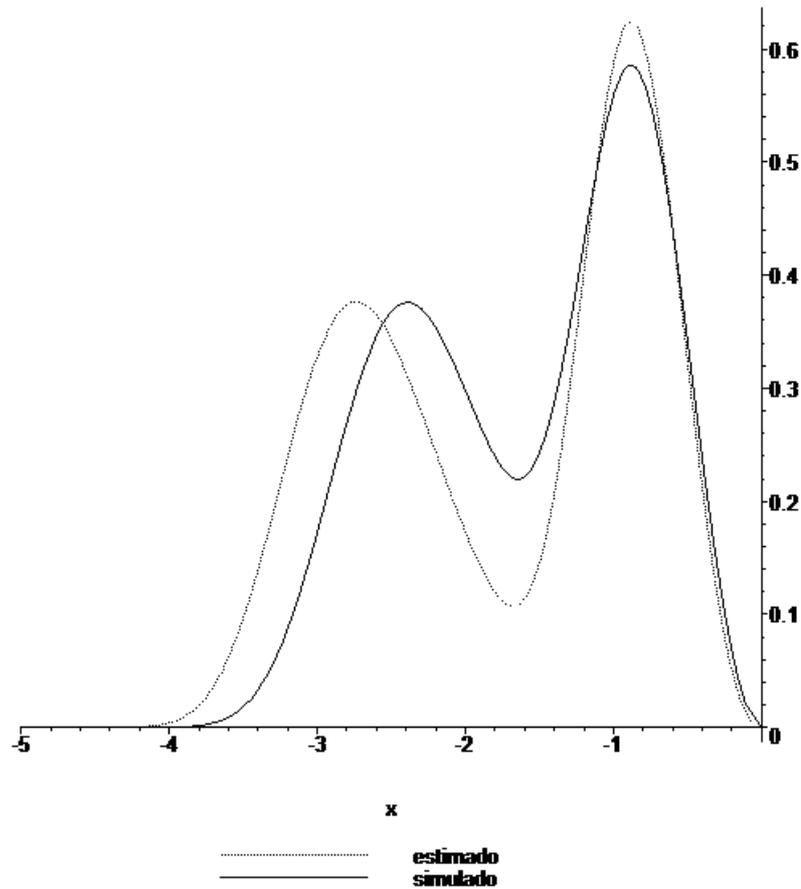


Figura 5.2: Gráfico da mistura de Weibull $h(0, 2; 2, 5; 1; 5; 2, 9)$ juntamente com as estimativas

anteriores obtém-se a tabela 5.3 com os valores utilizados para gerar as amostras, bem como os valores esperado das estimativas e o EQM com relação ao parâmetro de locação.

O gráfico 5.3 mostra a mistura entre duas Gumbel com parâmetros $\Theta = (0, 2; -3; 3; 2; 3)$ e o gráfico da mistura estimada com as estimativas dos parâmetros encontradas utilizando o algoritmo EM nos dados simulados, com os parâmetros propostos.

Observação:

- (1) Na tentativa de estimar o parâmetro σ_ℓ pelo algoritmo Newton-Raphson encontramos alguns problemas para convergência da equação (4.28), ao verificar as condições de convergência pelo Newton-Raphson. Seguindo a equação (A.8) a expressão de

$$\varphi(\sigma_\ell) = \sigma_\ell - \frac{S_1(\sigma_\ell)}{S_2(\sigma_\ell)}$$

$h(p, \mu_1, \mu_2, \sigma_1, \sigma_2)$	n	E(<i>Estimativas</i>)			EQM	
		\hat{p}	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\mu}_1$	$\hat{\mu}_2$
h(0,5;-3;3;2;3)	25	0,5065	-2,886	3,35	0,549	2,393
	50	0,4977	-2,997	3,122	0,248	1,18
	75	0,4945	-2,99	3,05	0,1748	0,8314
	100	0,4945	-2,99	3,023	0,139	0,5809
h(0,5;-1;0;2;3)	25	0,593	-0,9272	0,3214	0,6497	1,2771
	50	0,5591	-0,9681	0,1576	0,4175	0,653
	75	0,5439	-0,9816	0,1379	0,299	8,5104
	100	0,5372	-0,96768	0,1223	0,223	0,331
h(0,5;2;2;2;2,9)	25	0,5898	2,015	2,291	0,8656	1,2545
	50	0,553	1,967	2,1672	0,491	0,692
	75	0,5339	1,961	2,1418	0,3537	0,4853
	100	0,5341	1,9768	2,1218	0,2637	0,3523

Tabela 5.3: EQM e estimativas para uma mistura de duas Gumbel

onde

$$S_1(\sigma_\ell) = \sum_{i=1}^n f(\ell|x_i, \Theta^{(k)}) e^{-\frac{x_i - \mu_\ell}{\sigma_\ell}} - e^{-\frac{x_i - \mu_\ell}{\sigma_\ell}} \left(-\sigma_\ell^{-1} + \frac{x_i - \mu_\ell}{\sigma_\ell^2} - (x_i - \mu_\ell) e^{-\frac{x_i - \mu_\ell}{\sigma_\ell}} \sigma_\ell^{-2} \right)$$

e

$$S_2(\sigma_\ell) = \sum_{i=1}^n f(\ell|x_i, \Theta^{(k)}) \left(\sigma_\ell^{-2} - 2 \frac{x_i - \mu_\ell}{\sigma_\ell^3} + \frac{2(x_i - \mu_\ell) e^{-\frac{x_i - \mu_\ell}{\sigma_\ell}}}{\sigma_\ell^3} - \frac{(x_i - \mu_\ell)^2 e^{-\frac{x_i - \mu_\ell}{\sigma_\ell}}}{\sigma_\ell^4} \right)$$

e como a função tem a restrição, $|\varphi'(\sigma_\ell)| < 1$, para a convergência do método, calculamos $\varphi'(\sigma_\ell)$ para o conjunto de parâmetros Θ , cujo gráfico 5.4 mostra como esse função se comporta.

Observando o gráfico concluímos que $|\varphi'(\sigma_\ell)|$ não é menor que 1, portanto outro método numérico para solução de (4.27) deve ser utilizado.

- (2) Se considerarmos a mistura de Gumbel com $\Theta = (p_\ell, \mu_\ell)$, sobre o parâmetro de locação e a componente de proporção para estimação, as estimativas de p_ℓ são obtidos por (4.26) e μ_ℓ por (4.27).

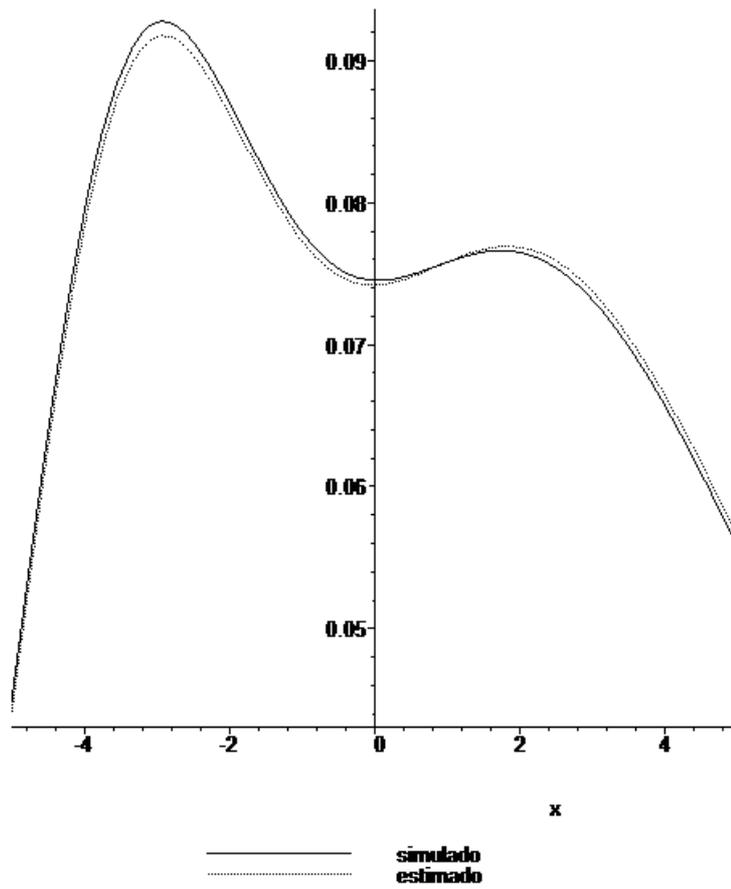


Figura 5.3: Gráfico da mistura de Gumbel $h(0, 5; -3; 3; 2; 3)$ juntamente com as estimativas

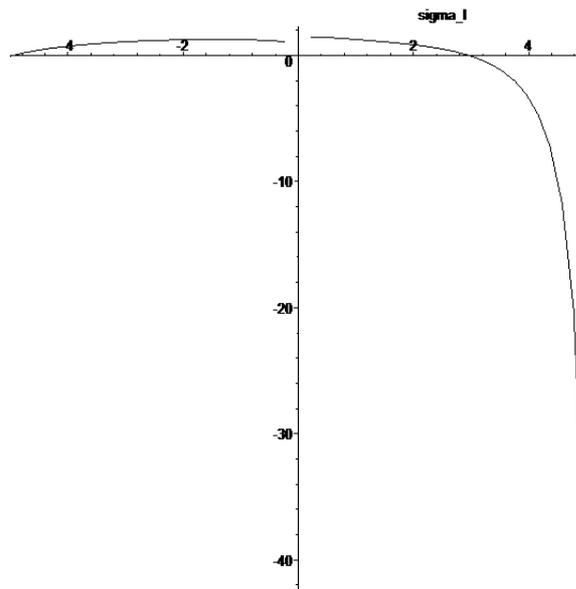


Figura 5.4: Gráfico da função $\varphi'(\sigma_\ell)$

(3) Técnicas alternativas de estimação dos parâmetros da mistura de duas Gumbel foram aplicadas. *Sandoval* (2007) encontra estimativas de máxima verossimilhança para a mistura de Gumbel e para o cálculo das soluções é utilizado o método de Rosenbreck restrito multivariado descrito em *Kustede* 1973. *Silva e Zocchi* (2009) estimam os parâmetros da mistura de duas Gumbel utilizando o algoritmo ECM (Expectation-conditional-maximization), uma extensão do algoritmo EM, aqui estudado e na resolução das equações lineares usam o método numérico "BFGS" descrito por *Noedal e Wright* (1999).

Capítulo 6

Conclusão

Neste trabalho mostramos a média, mediana, moda, variância e identificabilidade da mistura de duas componentes extremas da mesma classe. Utilizamos o algoritmo EM como método de estimação dos parâmetros das misturas.

Não se teve dificuldades nas simulações feitas e nem na montagem do algoritmo EM no software R. O tempo de processamento foi de aproximadamente 5 minutos quando se fazia a estimação de uma amostra de tamanho 100 no caso bimodal. O problema que ocorreu foi na estimação do parâmetro da escala (σ_1 e σ_2) da mistura de duas distribuições Gumbel, onde suas expressões não convergiram ocasionando uma não resposta para este parâmetro da mistura. Entretanto, dentro das referências lidas encontra-se em *Escalante-Sandoval* (2007) e *Silva e Zocchi* (2008) soluções para este problema onde eles utilizam com sucesso a utilização da máxima verossimilhança e do algoritmo ECM (uma variante do EM) na estimação dos parâmetros de uma mistura de duas Gumbel.

Na tentativa de estender o trabalho para mistura de classes diferentes, constatou-se que o algoritmo EM não alcança as estimativas de pelo menos um dos parâmetros da mistura de uma Fréchet com uma Gumbel e da mistura de uma Gumbel com uma Weibull.

Referências Bibliográficas

- [1] E. K. Al-Hussaini and K. S. Sultan. Reliability hazard based on finite mixture models. *Elsevier Science*, 20:139–183, 2001.
- [2] J. A. Bilmes. A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. *International Computer Science Institute*, pages 1–13, 1998.
- [3] T. Bucar, M. Nagode, and M. Fajdiga. Reliability approximation using finite weibull mixture distributions. *Reliability Engeneering and System Safety*, 84:241–251, 2004.
- [4] S. Chandra. On the mixtures of probability distributions. *Scandinavian Journal of Statistics*, 4:105–112, 1977.
- [5] H. Chen and Z.-X. Huang. Application of gumbel mixture extreme theory and maximum likelihood to estimate the seismic risk of the chinese mainland. *Acta Seismologica Sinica*, 8:325–331, 1995.
- [6] C. Q. da Silva. Tutorial sobre modelos markovianos com estados latentes-hmm. In *XXV CONGRESSO NACIONAL DE MATEMÁTICA APLICADA E COMPUTACIONAL CNMAC*, 2002.
- [7] N. Day. Estimating the components of a mixture of normal distributions. *Biometrika*, 56:463–474, 1969.
- [8] A. Dempster, N. Laird, and D. Rubin. Maximum-likelihood from incomplete data via the em algorithm. *J. Royal Statist. Soc. Ser. B*, 39:1–38, 1977.

- [9] P. Embrechts, C. Kluppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. SpringerVerlag, 1997.
- [10] C. Escalante-Sandoval. A mixed distribution with ev1 and gev components for analyzing heterogeneous samples. *Ingeniería investigación y Tecnología*, 3:123–133, 2007.
- [11] S. Finch, N. Mendell, and H. Thode. Probabilistic measures of adequacy of a numerical search for a global maximum. *Journal of American Statistical Association*, 84:10201023, 1989.
- [12] E. Fowlkes. Some methods for studying the mixture of two normal (lognormal) distributions. *Journal of American Statistical Association*, 74:561–575, 1979.
- [13] S. Frühwirth-Schnatter. *Finite Mixture and Markov Switching Models*. Springer, 2006.
- [14] W. Furman and B. Lindsay. Measuring the relative effectiveness of moment estimators as starting values in maximizing likelihoods. *Computational Statistics and Data Analysis*, 17:493–508, 1994.
- [15] W. Furman and B. Lindsay. Testing for the number of components in a mixture of normal distributions using moment estimators. *Computational Statistics and Data Analysis*, 17:473–492, 1994.
- [16] F. R. S. D. Gusmão. Uma abordagem bayesiana para distribuição weibull inversa generalizada. Master’s thesis, Universidade Federal Rural de Pernambuco, 2008.
- [17] R. Jiang, D. Murthy, and P. Ji. Models involving two inverse weibull distributions. *Reliability Engineering and System Safety*, 73:73–81, 2001.
- [18] R. Jiang, M. Zuo, and H. Li. Weibull and inverse weibull mixture models allowing negative weights. *Reliability Engineering and System Safety*, 66:227–234, 1999.
- [19] D. Karlis and E. Xekalaki. Choosing initial values for the em algorithm for finite mixtures. *Computational Statistics & Data Analysis*, 41:577–590, 2002.

- [20] J. Kuester and J. Mize. *Optimization Techniques with FORTRAN*. McGraw-Hill, 1973.
- [21] N. Laird. Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association*, 73:805–811, 1978.
- [22] B. Leroux. Consistent estimation of a mixing distribution. *Annals of Statistics*, 20:1350–1360, 1992.
- [23] B. Lindsay and P. Basak. Multivariate normal mixtures: a fast consistent method of moments. *Journal of American Statistical Association*, 88:468475, 1993.
- [24] G. McLachlan. On the choice of initial values for the em algorithm in fitting mixture models. *The Statistician*, 37:417–425, 1988.
- [25] G. McLachlan and D. Peel. *Finite Mixture Models*. Wiley, 1981.
- [26] G. J. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. Wiley, 1997.
- [27] M. Nagode and M. Fajdiga. An improved algorithm for parameter estimation suitable for mixed weibull distributions. *International Journal of Fatigue*, 22:75–80, 2000.
- [28] M. Nagode, J. Klemenc, and M. Fajdiga. Parametric modelling and scatter prediction of rainflow matrices. *International Journal of Fatigue*, 23:525–532s, 2001.
- [29] W. Seidel, K. Mosler, and M. Alker. A cautionary note on likelihood ratio tests in mixture models. *Annals of the Institute of Statistical Mathematics*, 52:481487, 2000.
- [30] W. Seidel, K. Mosler, and M. Alker. Likelihood ratio tests based on subglobal optimisation: a power comparison in exponential mixture models. *Statist. Hefte*, 41:85–98, 2000.

- [31] W. Seidel, H. Sevcikova, and M. Alker. On the power of different versions of the likelihood ratio test for homogeneity in an exponential mixture model. *Dept. Stat. und Quantit. Ökon. Universität der Bundeswehr Hamburg*, 92, 2000.
- [32] R. R. Silva and S. S. Zochi. Misturas de duas distribuições de gumbel para análise dos dados de velocidade máxima do vento em piracicaba, sp. *Revista Brasileira de Biometria*, 26:49–79, 2008.
- [33] R. L. Smith. *Extreme value theory*. Wiley, Chichester, 1990.
- [34] K. S. Sultan, M. A. Ismail, and A. S. Al-Moisheer. Mixture of two inverse weibull distributions: Properties and estimation. *Computational Statistics and Data Analysis*, 51:5377–5387, 2006.
- [35] V. Tartaglia, E. Caporali, E. Cavigli, and A. Moro. L-moments based assessment of a mixture model from frequency analysis of rainfall extremes. *Advanced in Geosciences*, 2:331–334, 2006.
- [36] H. Teicher. On the convolution of distributions. *The Annals of Mathematical Statistics*, 25:775–778, 1954.
- [37] H. Teicher. On the mixture of distributions. *The Annals of Mathematical Statistics*, 31:55–73, 1960.
- [38] H. Teicher. Identifiability of mixtures. *The Annals of Mathematical Statistics*, 32:244–248, 1961.
- [39] H. Teicher. Identifiability of finite mixtures. *The Annals of Mathematical Statistics*, 34:1265–1269, 1963.
- [40] D. M. Titterton, A. F. M. Smith, and U. E. Makov. *Statistical Analysis of Finite Mixture Models*. Wiley, 1985.
- [41] W. Woodward, W. Parr, R. Schucany, and H. Lindsey. A comparison of minimum distance and maximum likelihood estimation of a mixture proportion. *Journal of American Statistical Association*, 79:590–598, 1984.

Apêndice A

Newton-Raphson

O método de iteração linear consiste em estimar a raiz de uma função $f(x)$ usando o processo iterativo

$$x_{n+1} = \varphi(x_n) \tag{A.1}$$

Podemos escrever uma forma geral para a função de iteração

$$\varphi(x) = x + A(x).f(x) \tag{A.2}$$

pois, para x igual à raiz de $f(x)$, tem-se $f(x) = 0$, ou seja $x = \varphi(x)$ para qualquer $A(x) \neq 0$.

Para a convergência do método é preciso que $|\varphi'(x)| < 1$ em um intervalo $[a, b]$ que contem a raiz de $f(x)$. Portanto, a ideia no método de Newton-Raphson é escolher uma função $\varphi(x)$ tal que $\varphi(\alpha) = 0$ onde α é a raiz de $f(x)$ e $\alpha \in [a, b]$. Com isso, teremos $|\varphi'(x)| < 1$ desde que não nos afastemos muito do valor de α durante o processo de resolução do problema.

Derivando $\varphi(x)$ dada pela expressão (A.2) em relação a x , temos

$$\varphi'(x) = 1 + A'(x).f(x) + A(x).f'(x) \tag{A.3}$$

Calculando (A.3) em $x = \alpha$, segue que

$$\varphi'(\alpha) = 1 + A(\alpha).f'(\alpha) \tag{A.4}$$

Exigindo que $\varphi'(x) = 0$, tem-se

$$A(\alpha).f'(\alpha) = -1 \tag{A.5}$$

ou seja

$$A(\alpha) = \frac{-1}{f'(\alpha)} \quad (\text{A.6})$$

Escolhendo

$$A(x) = \frac{-1}{f'(x)} \quad (\text{A.7})$$

segue de (A.2) que

$$\varphi(x) = x - \frac{f(x)}{f'(x)} \quad (\text{A.8})$$

O método de Newton-Raphson consiste em usar o processo iterativo $x_{n+1} = \varphi(x_n)$ e como função de iteração a expressão (A.8)

A.1 Convergência do Método de Newton-Raphson

Apesar de obtermos a forma da função $\varphi(x)$ procurando garantir a convergência do processo iterativo, esta não está sempre garantida para este método (mas quase sempre). A convergência no método de Newton-Raphson está sempre garantida para um certo intervalo $[a, b]$ que contem a raiz de $f(x)$, desde que $f(x)$ e $f'(x)$ sejam contínuas nesse intervalo e que $f'(\alpha) \neq 0$, onde α é a raiz de $f(x)$. Portanto, se utilizarmos uma estimativa inicial x_0 tal que $x_0 \in [a, b]$, a convergência estará garantida. Em outras palavras, para o método de Newton-Raphson convergir, é preciso que nossa estimativa inicial esteja próxima da raiz de $f(x)$. A proximidade exigida para a convergência vai depender de caso a caso e nem sempre é simples de determinar.

A.2 Interpretação geométrica

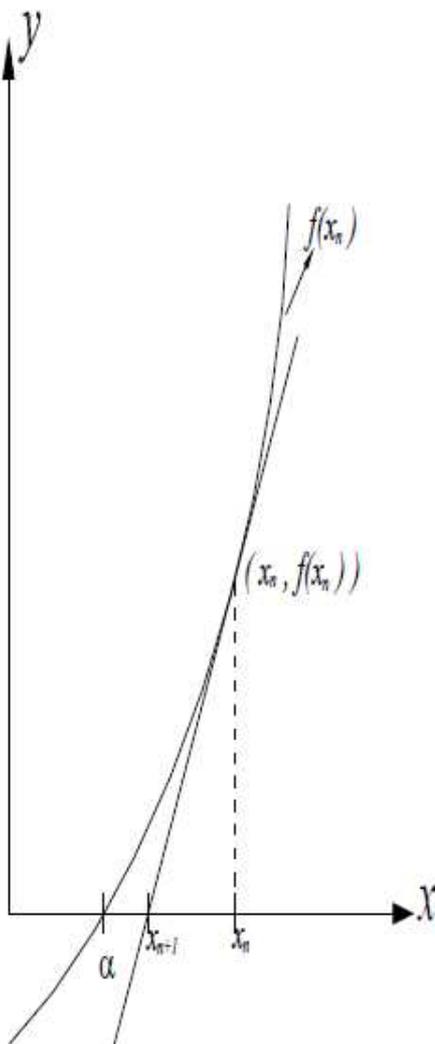
Dado x_n , o ponto x_{n+1} será obtido pela intercessão da reta tangente a $f(x)$ em x_n com o eixo x (a abscissa). Podemos ilustrar isso matematicamente. A reta tangente a $f(x)$ em x_n é dada por:

$$f'(x_n) = \frac{f(x_n) - 0}{x_n - x_{n+1}} \quad (\text{A.9})$$

A partir dessa expressão, obtemos a fórmula de Newton-Raphson, ou seja

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \quad (\text{A.10})$$

Portanto, a cada iteração do nosso processo, nos aproximamos cada vez mais da raiz de $f(x)$ através da tangente (ou seja, da derivada) da função $f(x)$. A figura a seguir ilustra essa interpretação geométrica do Método de Newton-Raphson.



Apêndice B

Mistura de Fréchet

```
#Funcao acumulada inversa
fmenos1=function(u,a,s){
  (-log(u))^(1/a)*(1/s)}
#mistura de duas frechet
mfrechet=function(x,p,s,a){
  mf=p*a[1] *s[1] ^(-a[1] )*x^(-1-a[1] )*exp(-(s[1] *x)^(-a[1] ))
  mf=mf+(1-p)*a[2] *s[2] ^(-a[2] )*x^(-1-a[2] )*exp(-(s[2]*x)^(-a[2] ))
  return(mf)
}
#funcao f(ell)
tau=function(x,p,s,a,k){
  pj=p[k]
  sj=s[k]
  aj=a[k]
  tf=pj*aj*sj^(-aj)*x^(-1-aj)*exp(-(sj*x)^(-aj))
  tf=tf/mfrechet(x,p[1],s,a)
  return(tf)
}
#algoritmo N-R para alpha
NR.fas<-function(erro,maxiter,j,x,pk,sk,ak){
```

```

atual<-ak[j]
valor<-1
i<-0
while(valor>erro && i<=maxiter){
score1= 1/atual-log(sk[j])-log(x)+log(sk[j]*x)*(sk[j]*x)^(-atual)
invJ1= -1/atual^2-(sk[j]*x)^(-atual)*log(sk[j]*x)*log(sk[j]*x)
score2=tau(x,pk,sk,ak,j)*score1
invJ2=tau(x,pk,sk,ak,j)*invJ1
score=sum(score2)
invJ=1/sum(invJ2)
ai= atual -invJ*score
valor= abs(ai-atual)
atual=ai
i=i+1
#if(i>maxiter) cat(i," a= ",atual," erro= ",valor,"\n")
}
return(atual)
}

#valores dos parametros
a=c(2,3)
s=c(1,2)
p=c(.2,.8)
n=100
###
piis=c(0,0)
siis=c(0,0)
aiis=c(0,0)
for(ii in 1:500){
#simulacao
set.seed(9643)
u1=runif(n)
u2=runif(n)

```

```

x=numeric(n)
for(i in 1:n){
  if(u1[i]<p[1]) x[i]= fmenos1(u2[i],a[1],s[1])
  else x[i]=fmenos1(u2[i],a[2],s[2])
}

erro=0.0001
pis=c(0,0)
sis=c(0,0)
ais=c(0,0)
pk=p
sk=s
ak=a
pi=numeric(2)
si=numeric(2)
ai=numeric(2)
#algoritmo EM
for(w in 1:15){
  for(j in 1:2){
    pi[j]=mean(tau(x,pk,sk,ak,j))

    tau.s=sum(tau(x,pk,sk,ak,j))
    tau.s.x=sum(tau(x,pk,sk,ak,j)*(x^(-ak[j])))
    si[j]=(tau.s/tau.s.x)^(-1/ak[j])

    ai[j]=NR.fas(erro,10,j,x,pk,sk,ak)

  }
  veros=prod(mfrechet(x,pk[1],sk,ak))
  n.veros=prod(mfrechet(x,pi[1],si,ai))
  valor=abs(n.veros-veros)
  pis=rbind(pis,pi)
}

```

```
sis=rbind(sis,si)
ais=rbind(ais,ai)
pk=pi
sk=si
ak=ai
}
piis=rbind(piis,pis[16,])
siis=rbind(siis,sis[16,])
aiis=rbind(aiis,ais[16,])
}
```

Apêndice C

Mistura de Weibull

```
#Funcao acumulada inversa
fmenoslw=function(u,s,a){
-s*(-log(u))^(1/a)
}

#mistura de duas weibull
mweibull=function(x,p,s,a){
h=p*(a[1]/s[1]^a[1])*(-x)^(a[1]-1)*exp(-(-x/s[1])^a[1])
h=h+(1-p)*(a[2]/s[2]^a[2])*(-x)^(a[2]-1)*exp(-(-x/s[2])^a[2])
return(h)
}

#funcao f(ell)
tauw=function(x,p,s,a,k){
pj=p[k]
sj=s[k]
aj=a[k]
t=pj*(aj/sj^aj)*(-x)^(aj-1)*exp(-(-x/sj)^aj)
t=t/mweibull(x,p[1],s,a)
return(t)
}

#algoritmo NR para alpha
NR.was=function(erro,maxiter,x,p,s,a,j){
```

```

atual=a[j]
valor=1
i=1
while(valor> erro && i<=maxiter){
score1=1/atual-log(s[j])+log(-x)-log(-x/s[j])*(-x/s[j])^atual
score2=tauw(x,p,s,a,j)*score1
score=sum(score2)
invJ1=-1/atual^2-(-x/s[j])^(atual)*log(-x/s[j])*log(-x/s[j])
invJ2=tauw(x,p,s,a,j)*invJ1
invJ=1/sum(invJ2)
ai= atual - score*invJ
valor=abs(atual-ai)
atual=ai
i=i+1
}
#if(i>maxiter) cat(i," a= ",atual," erro= ",valor,"\n")
return(atual)
}

#valores dos parametros
a=c(2,2.9)
s=c(2.5,2)
p=c(.2,.8)
n=100
piis=c(0,0)
siis=c(0,0)
aiis=c(0,0)
for(ii in 1:200){
x=numeric(n)
#simulacao
set.seed(ii+4444)
u1=runif(n)
u2=runif(n)

```

```

for(r in 1:n){
if(u1[r]<p[1]) x[r]=fmenos1w(u1[r],s[1],a[1])
else x[r]=fmenos1w(u2[r],s[2],a[2])
}
erro=0.0001
pis=c(0,0)
sis=c(0,0)
ais=c(0,0)
pk=p
sk=s
ak=a
pi=numeric(2)
si=numeric(2)
ai=numeric(2)
#algoritmo EM
for(w in 1:15){
for(j in 1:2){
pi[j]=mean(tauw(x,pk,sk,ak,j))

s.tau=sum(tauw(x,pk,sk,ak,j))
s.tau.x=sum(tauw(x,pk,sk,ak,j)*(-x)^ak[j])
si[j]=(s.tau.x/s.tau)^(1/ak[j])

ai[j]=NR.was(erro,10,x,pk,sk,ak,j)
}
veros=prod(mweibull(x,pk[1],sk,ak))
n.veros=prod(mweibull(x,pi[1],si,ai))
valor=abs(n.veros-veros)
pis=rbind(pis,pi)
sis=rbind(sis,si)
ais=rbind(ais,ai)
pk=pi

```

```
sk=si
ak=ai
}
piis=rbind(piis,pis[w,])
siis=rbind(siis,sis[w,])
aiis=rbind(aiis,ais[w,])
}
```

Apêndice D

Mistura de Gumbel

```
#Funcao acumulada inversa
fmenoslg=function(u,m,s){
m-s*log(-log(u))
}

#mistura de duas gumbel
mgumbel=function(x,p,m,s){
y1=(x-m[1])/s[1]
y2=(x-m[2])/s[2]
f=(p[1]/s[1])*exp(-y1-exp(-y1))
f=f+(p[2]/s[2])*exp(-y2-exp(-y2))
return(f)
}

#funcao f(ell)
taug=function(x,p,m,s,j){
pk=p[j]
mk=m[j]
sk=s[j]
yk=(x-mk)/sk
t=(pk/sk)*exp(-yk-exp(-yk))
t=t/mgumbel(x,p,m,s)
return(t)
```

```

}
#valores dos parametros
m=c(-1,0)
s=c(2,3)
p=c(0.5,0.5)
n=25
piis=c(0,0)
miis=c(0,0)
#siis=c(0,0)
for(ii in 1:50){
#simulacao
set.seed(-ii+90649)
u1=runif(n)
u2=runif(n)
x=numeric(n)
pop=rep(2,n)
for(i in 1:n){
if(u1[i]>p[1]) x[i]=fmenos1g(u2[i],m[1],s[1])
else x[i]=fmenos1g(u2[i],m[2],s[2])
}
erro=0.0001
pii=numeric()
#si=numeric()
mi=numeric()
pis=c(0,0)
#mis=c(0,0)
sis=c(0,0)
#utilizando como valor inicial
#os simulados
pk=p
mk=m
sk=s

```

```

#algoritmo EM
for(w in 1:15){
for(j in 1:2){
pii[j]=mean(taug(x,pk,mk,sk,j))

s.tau=sum(taug(x,pk,mk,sk,j))
s.tau.e=sum(taug(x,pk,mk,sk,j)*exp(-x/sk[j]))
mi[j]=sk[j]*log(s.tau/s.tau.e)
#como seria calculado sigma,
#em caso de convergencia
#si[j]=NR.gs(erro,10,x,pk,mk,sk,j)
}
n.veros=prod(mgumbel(x,pii,mi,si))
veros=prod(mgumbel(x,pk,mk,sk))
pis=rbind(pis,pii)
mis=rbind(mis,mi)
#sis=rbind(sis,si)
pk=pii
mk=mi
#sk=si
}
piis=rbind(piis,pis[w,])
miis=rbind(miis,mis[w,])
#siis=rbind(siis,siis[w,])
}

```