



Universidade de Brasília
IE – Departamento de Estatística
Estágio Supervisionado 2

SOBRE IDENTIFICABILIDADE PARA MISTURAS FINITAS

Yuri César Silva

Orientadora: Prof.^a Dr.^a Cira Etheowalda Guevara Otiniano

Brasília
Março de 2013



Universidade de Brasília
IE – Departamento de Estatística
Estágio Supervisionado 2

SOBRE IDENTIFICABILIDADE PARA MISTURAS FINITAS

Monografia apresentada junto ao curso de estatística da Universidade de Brasília como requisito parcial à obtenção do título de bacharel.

Banca examinadora:

- Prof.^a Dr.^a Cira Etheowalda Guevara Otiniano (Orientadora)
- Prof. Dr. Raul Yukihiro Matsushita
- Prof. Me. Démerson André Polli

Agradecimentos

Agradeço a todos que de alguma forma participaram no desenvolvimento do presente trabalho.

Resumo

No presente trabalho apresentamos conceitos, caracterizações e exemplos da identificabilidade de várias famílias de misturas finitas de distribuições. Apresentamos também a prova da identificabilidade das famílias de misturas finitas das distribuições logística, logística generalizada, beta generalizada do segundo tipo, gama generalizada, logarítmica e skew-normal.

O algoritmo EM é descrito e usado para obter fórmulas para a estimação dos parâmetros de misturas de duas componentes das distribuições logística, logística generalizada, beta generalizada do segundo tipo e gama generalizada. Adicionamos por fim um estudo de simulação para testar as estimativas da mistura de duas logísticas.

Palavras chave: modelo de misturas finitas, identificabilidade, algoritmo EM.

Abstract

In this paper we present concepts, characterizations and examples of identifiability for finite mixtures. We also present proof of the families of finite mixtures of the logistic, generalized logistic, generalized beta of the second kind, generalized gamma, logarithmic and skew-normal distributions.

The EM algorithm is described and used for obtaining formulas for estimation of parameters of mixtures of two components of the logistic, generalized logistic generalized beta and generalized gamma second type of distributions. Finally, a simulation study is used to test the algorithm considering the mixture of two logistcs.

Key words: model of finite mixtures, identifiability, EM algorithm.

Lista de ilustrações

1	Mistura de densidades normais	17
2	Misturas de densidades Weibull	18
3	Misturas de densidades lognormais	20
4	Misturas de massas binomial negativa	22
5	Misturas de densidades exponenciais	23
6	Misturas de densidades gama	24
7	Misturas de densidades logísticas	26
8	Mistura de densidades skew-normal	27
9	Mistura de densidades skew-t	29
10	Misturas estimadas e simuladas de densidades logísticas	62

Sumário

1	Introdução	5
2	Conceitos preliminares	7
2.1	Introdução	7
2.2	Misturas de densidades e distribuições	7
2.3	Algumas distribuições de probabilidade e resultados de interesse	9
2.3.1	Distribuição logística	9
2.3.2	Distribuição logística generalizada	10
2.3.3	Distribuição beta generalizada do segundo tipo	11
2.3.4	Distribuição gama generalizada	12
2.3.5	Distribuição logarítmica	12
2.3.6	Distribuição skew-normal	13
2.3.7	Distribuição Kumaraswamy	14
2.4	Alguns exemplos clássicos de misturas finitas	15
2.4.1	Mistura de densidades normais	15
2.4.2	Mistura de densidades Weibull	16
2.4.3	Mistura de densidades lognormal	17
2.4.4	Mistura de densidades Cauchy	18
2.4.5	Mistura de distribuições estáveis	19
2.4.6	Mistura de massas binomial negativa	20
2.4.7	Mistura de densidades gama	21
2.4.8	Mistura de densidades exponencial	22
2.4.9	Mistura de densidades logística	23
2.4.10	Mistura de densidades skew-normal	24
2.4.11	Mistura de densidades skew-t	25
3	Identificabilidade para classes de misturas	27
3.1	Introdução	27
3.2	Identificabilidade	27
3.3	Caracterizações da identificabilidade	30
3.4	Prova da identificabilidade para algumas famílias de distribuições	32
3.4.1	Mistura de distribuições normais	32

3.4.2	Mistura de distribuições Weibull	32
3.4.3	Mistura de distribuições lognormais	34
3.4.4	Mistura de distribuições Cauchy	34
3.4.5	Mistura de distribuições binomiais negativas	36
3.4.6	Mistura de distribuições gama	36
3.5	Novas provas de identificabilidade para algumas classes de misturas	37
3.5.1	Mistura de distribuições logísticas	38
3.5.2	Mistura de distribuições logísticas generalizadas	38
3.5.3	Mistura de distribuição beta generalizada do segundo tipo	39
3.5.4	Distribuição gama generalizada	40
3.5.5	Mistura de massas logarítmicas	41
3.5.6	Mistura de distribuições skew-normal	42
3.5.7	Mistura de distribuições Kumaraswamy	43
4	O algoritmo EM	45
4.1	Introdução	45
4.2	Estimador de máxima verossimilhança	45
4.3	O algoritmo EM básico	46
4.4	O algoritmo EM para misturas	49
4.5	Exemplos do algoritmo EM para misturas	53
4.5.1	Logística	53
4.5.2	Logística generalizada	54
4.5.3	Gama generalizada	56
4.6	Beta generalizada de segunda ordem	58
5	Simulação	60
5.1	Introdução	60
5.2	Escolha do valor inicial	60
5.3	Logística	60
6	Conclusão	63

Apêndice A - Álgebra Linear	74
Apêndice B - Programação no R	76

1 Introdução

Apesar da ênfase e tradição histórica no estudo de populações homogêneas em estatística, os esforços iniciais de Newcomb (1886) e Pearson (1894) abriram a possibilidade de aproximar populações heterogêneas por meio de misturas finitas de distribuições, com o primeiro fazendo uma aplicação de misturas de normais como modelos para outliers e o segundo decomposição de misturas de normais por meio do método de momentos.

Considerando a literatura existente, podemos notar um crescente número de pesquisas e estudos na área aparecendo na segunda metade do século 20, particularmente com o advento (e contínua evolução) da computação. Referências como Titterington et al. (1985), McLachlan e Peel (2000) e Murthy et al. (2004) apresentam mais de 100 aplicações distintas de misturas de distribuições de probabilidade na modelagem de fenômenos, sendo essa naturalmente uma pequena parte da importância dos modelos considerados nesse trabalho. Destacamos também que o tópico não é abordado na graduação em estatística na UnB e não é de nosso conhecimento a existência de um tomo em língua portuguesa que o aborde com profundidade, portanto, concluímos que é importante inicialmente que seja feito um estudo direcionado à revisão da literatura da área, bem como de suas definições e exemplos clássicos. Estudo que visa também cumprir o importante papel de agregar conhecimento e técnicas diferentes aumentando o escopo daquilo que já foi estudado.

A condição de identificabilidade, isto é, a existência de uma representação única para cada uma das classes dos modelos considerados, é de grande importância do ponto de vista estatístico. Por exemplo, quando a condição não é satisfeita, não é provável haver processos de estimação bem definidos (Titterington et al. 1985). Considerando que a ênfase do estudo está direcionada para identificabilidade com os modelos de misturas de distribuições, faz-se necessária (pelos mesmos motivos previamente abordados), também uma revisão direcionada aos principais resultados apresentados na literatura. Muitos artigos publicados investigam e fornecem resultados sobre a identificabilidade para misturas finitas. Desses destacamos aqueles que serão ex-

plorados nesse trabalho: Teicher (1963) e Yakowitz and Sprangins (1968) que apresentaram resultados com condições suficientes para provar a identificabilidade e Atienza et al. (2006) que apresenta ainda outra variação da condição inicial de Teicher (1963), procurando fazê-la aplicável a um número maior de famílias de distribuições.

Neste trabalho apresentamos provas inéditas de identificabilidade para classes de misturas finitas das distribuições logística, logística generalizada, beta generalizada do segundo tipo, gama generalizada, logarítmica, skew-normal e Kumaraswamy. Usando o algoritmo de "Expectation Maximization" (EM) são fornecidas as equações necessárias para estimar os parâmetros de misturas de dois componentes de distribuições logística, logística generalizada, beta generalizada de segunda ordem e gama generalizada. O algoritmo EM destaca-se no caso, pois é facilmente adaptável à modelos de misturas finitas de distribuições.

Além da presente introdução podemos destacar mais 4 capítulos. No capítulo 2 são abordados alguns conceitos fundamentais, que incluem: noções de álgebra linear, algumas distribuições de probabilidade e resultados e por fim o conceito formal de mistura de distribuição de probabilidade seguido de exemplos de algumas misturas apontando estudos e aplicações para cada caso.

No capítulo 3 é dada a definição formal da identificabilidade estendendo-a posteriormente a misturas de distribuições, com exemplos de misturas identificáveis. A segunda parte desse capítulo inclui provas inéditas para as classes de misturas de distribuições previamente citadas.

O capítulo 4 introduz estimação por máxima verossimilhança seguindo com o algoritmo EM, que é posteriormente entendido ao caso de misturas de distribuições. No mesmo capítulo apresentamos ainda as fórmulas obtidas para a estimação via algoritmo EM considerando certos modelos de misturas.

Por fim, no capítulo 5 apresentamos um estudo de simulação que considera a estimação dos parâmetros de uma mistura de duas distribuições logísticas, bem como considerações sobre o algoritmo em si e a escolha do valor inicial.

2 Conceitos preliminares

2.1 Introdução

Antes de entrar a fundo num estudo de identificabilidade destacamos no presente capítulo formalmente alguns conceitos fundamentais que devem ser estabelecidos, bem como algumas distribuições de interesse que serão estudadas ao longo do trabalho.

A seção 2.2 traz uma série de definições básicas sobre misturas de densidades e distribuições. As distribuições de probabilidades abordadas nas provas inéditas de identificabilidade (considerando misturas) bem como resultados referentes a estas são encontrados na seção 2.3, com os modelos de misturas para uma série de modelos sendo apresentados como exemplos na seção 2.4.

2.2 Misturas de densidades e distribuições

Definição 2.1 *Sejam f_1, \dots, f_k , funções de densidades e \mathcal{X} uma variável aleatória com densidade que toma a forma:*

$$h(x) = \sum_{i=1}^k \pi_i f_i = \pi_1 f_1 + \dots + \pi_k f_k \quad (2.2.1)$$

em que

$$\pi_i > 0, \forall i = 1, \dots, k; \sum_{i=1}^k \pi_i = 1$$

$$f_i(\cdot) \geq 0; \int_{-\infty}^{\infty} f_i(x) dx = 1, \forall i = 1, \dots, k.$$

Pode-se afirmar, sobre essas condições que $f(x)$ é uma mistura finita de densidades, com os π_1, \dots, π_k sendo os pesos, os f_1, \dots, f_k sendo os componentes da mistura e k representa o número de componentes na mistura.

Por outro lado se temos $f_1, \dots, f_k \in \mathcal{F} = \{f : f(x; \theta), x \in \mathbb{R}^d, \theta \in \Theta\}$, uma família de densidades num espaço paramétrico Θ , a mistura finita pode ser escrita como:

$$h(x; \Psi) = \sum_{i=1}^k \pi_i f_i(x; \theta_i) \quad (2.2.2)$$

Observamos que, nesse caso, cada componente $f_i(x; \theta_i)$ da mistura pertence a uma mesma família paramétrica (i.e. são parametrizadas por um $\theta_i \in \Theta$) e $\Psi = (\pi_1, \dots, \pi_{k-1}, \theta_1, \dots, \theta_k)$ é um resumo dos parâmetros da mistura.

Definição 2.2 *Segue ainda que a classe \mathcal{H} das misturas finitas de \mathcal{F} como previamente escrita pode ser definida como:*

$$\mathcal{H} = \left\{ h : \sum_{i=1}^k \pi_i f_i(x; \theta_i); f_i(x; \theta_i) \in \mathcal{F}; k = 1, 2, 3, 4, \dots \right\} \quad (2.2.3)$$

Alternativamente definindo \mathcal{F} como uma família d -dimensional de funções de distribuição, tal que $\mathcal{F} = \{F : F(x; \theta), x \in R^d, \theta \in \Theta\}$. É possível escrever a mistura de $F_1, \dots, F_k \in \mathcal{F}$, como:

$$H(x; \Psi) = \sum_{i=1}^k \pi_i F_i(x; \theta_i); \quad (2.2.4)$$

e a classe de misturas:

$$\mathcal{H} = \left\{ H : \sum_{i=1}^k \pi_i F_i(x; \theta_i); F_i(x; \theta_i) \in \mathcal{F}; k = 1, 2, 3, 4, \dots \right\} \quad (2.2.5)$$

Suponha que $F(x|\theta)$ seja função de distribuição acumulada de uma variável (vetor) aleatório, dado que $\Theta = \theta$, com $G(\theta)$ representando a função de distribuição acumulada da variável (vetor) aleatório Θ . Pode-se reescrever a mistura de F e G como:

$$H(x) = \int_{-\infty}^{\infty} F(x|\theta) dG(\theta) \quad (2.2.6)$$

Em que F é o "núcleo" e $G(\cdot)$ a "distribuição da mistura". Se a massa

de G fica limitada a um conjunto enúmeravel $\theta_1, \theta_2, \dots$ tal que as massas para cada θ_i seja, $G(\theta_i)$, podemos reescrever (2.2.6) como: $H = \sum_{i=1}^{\infty} G(\theta_i) F_i$. Em particular para um numero finito de pontos $\theta_1, \dots, \theta_k$, e fazendo $G(\theta_i) \equiv \pi_i$, obtemos (2.2.4).

2.3 Algumas distribuições de probabilidade e resultados de interesse

2.3.1 Distribuição logística

A densidade associada à distribuição logística é a seguinte:

$$f_L(x; \mu, \sigma) = \frac{\exp\left[-\frac{x-\mu}{\sigma}\right]}{\sigma(1 + \exp\left[-\frac{x-\mu}{\sigma}\right])^2}, \quad (2.3.1)$$

com função acumulada:

$$F_L(x; \mu, \sigma) = \frac{1}{(1 + \exp\left[-\frac{x-\mu}{\sigma}\right])}, \quad (2.3.2)$$

em que

$$x, \mu \in R \text{ e } \sigma > 0.$$

Seguindo Balakrishnan (1992, p.5), obtemos a seguinte forma para a função geradora de momentos (que considera $\mu = 0$ e $\sigma = 1$):

$$M_z(t) = B(1+t, 1-t).$$

Na expressão anterior podemos fazer $x = \sigma z + \mu$, obtendo por fim:

$$\begin{aligned} M_x(t) &= \exp[\mu t] M_z(\sigma t) \\ &= \exp[\mu t] B(1 + \sigma t, 1 - \sigma t) \\ &= \exp[\mu t] \Gamma(1 + \sigma t) \Gamma(1 - \sigma t) \end{aligned} \quad (2.3.3)$$

em que $-\frac{1}{\sigma} < t < \frac{1}{\sigma}$.

2.3.2 Distribuição logística generalizada

Uma generalização da distribuição logística é apresentada em Johnson et al. (1995), com a densidade:

$$f_{LG}(x; p, q) = \frac{1}{B(p, q)} \frac{e^{-qx}}{(1 + e^{-x})^{p+q}}, \quad (2.3.4)$$

em que

$$x \in R, p, q > 0$$

e

$$B(p, q) = \int_0^1 t^{p-1}(1-t)^{q-1} dt.$$

Podemos ainda fazer $x = \sigma y + \mu$ para adicionar parâmetros para locação e escala. No mesmo livro encontramos a seguinte fórmula para a função geradora de momentos:

$$M_z(t) = \frac{\Gamma(p+t)\Gamma(q-t)}{\Gamma(p)\Gamma(q)}$$

com

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-s} ds.$$

E para $x = \sigma z + \mu$:

$$\begin{aligned} M_x(t) &= \exp[\mu t] \varphi_y(\sigma t) \\ &= \exp[\mu t] \frac{\Gamma(p + \sigma t)\Gamma(q - \sigma t)}{\Gamma(p)\Gamma(q)} \end{aligned} \quad (2.3.5)$$

em que $-\frac{p}{\sigma} < t < \frac{q}{\sigma}$.

2.3.3 Distribuição beta generalizada do segundo tipo

A distribuição beta generalizada do segundo tipo tem a seguinte densidade:

$$f_{BG}(x; \alpha, \sigma, p, q) = \frac{\alpha (x)^{\alpha p - 1}}{\sigma^{\alpha p} B(p, q) \left[1 + \left(\frac{x}{\sigma}\right)^\alpha\right]^{p+q}}, \quad (2.3.6)$$

com função de distribuição:

$$F_{BG}(x; \alpha, \sigma, p, q) = I_z(p, q) \text{ e } z = \left(\frac{x}{b}\right)^\alpha, \quad (2.3.7)$$

em que

$$x, \alpha, \sigma, p, q > 0$$

e

$$B(p, q) = \int_0^1 t^{p-1} (1-t)^{q-1} dt, \quad I_z(p, q) = \frac{1}{B(p, q)} \int_0^z \frac{t^{p-1}}{(1+t)^{p+q}} dt.$$

O modelo beta generalizado do segundo tipo possui um número de aplicações práticas em economia sendo ainda a generalização de alguns outros modelos também usados na área econômica (Kleiber e Kotz 2003). Temos que: para $p = 1$ obtemos a distribuição Singh Maddala (ou Burr tipo XII), para $q = 1$ obtemos a distribuição Dagum (ou Burr inversa), obtemos a distribuição beta do segundo tipo, para $\alpha = 1$, a distribuição Fisk (log-logística) é obtida caso $p = q = 1$ e ainda quando $\alpha = p = 1$ obtemos a distribuição Lomax (Pareto tipo II).

Seguindo ainda Kleiber e Kotz (2003), temos a seguinte fórmula para o t -ésimo momento:

$$E(X^t) = \sigma^n \frac{B\left(1 - \frac{t}{\alpha q}, 1 + \frac{t}{\alpha p}\right)}{B(p, q)} \quad (2.3.8)$$

em que $-\alpha p < t < \alpha q$.

2.3.4 Distribuição gama generalizada

Associamos à distribuição gama generalizada a seguinte densidade:

$$f_{GG}(x; \alpha, \beta, \sigma) = \frac{\beta}{\sigma \Gamma(\alpha)} \left(\frac{x}{\sigma}\right)^{\alpha\beta-1} \exp\left[-\left(\frac{x}{\sigma}\right)^\beta\right], \text{ em que}$$

$$x, \alpha, \sigma > 0, \Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx.$$

O presente modelo foi proposto em Khodabin e Ahmadabadi (2010), que apresenta um estudo que inclui estimação pelo método de momentos e entropia de Shannon. Uma importância da distribuição gama generalizada é sua flexibilidade, no sentido de que podemos obter outras distribuições conhecidas, que podem ser resumidas como segue: distribuição gama para $\beta = 1$, Weibull para $\alpha = 1$, exponencial para $\alpha = \beta = 1$ e ainda, fazendo $\alpha \rightarrow \infty$ podemos obter a distribuição lognormal. Fazendo $\beta = 2$ obtemos ainda uma generalização da família normal, de onde obtemos ainda os seguintes casos especiais: distribuição 'half-normal' para $\alpha = \frac{1}{2}$ e $\sigma^2 = 2\lambda^2$, Rayleigh para $\alpha = 1$ e $\sigma^2 = 2\lambda^2$, Maxwell-Boltzmann fazendo $\alpha = \frac{3}{2}$ e por fim a distribuição qui para $\alpha = \frac{k}{2}$ e $k = 1, 2, 3, \dots$

Em Khodabin e Ahmadabadi (2010) encontramos também a fórmula do t -ésimo momento, como segue.

$$E(X^t) = \sigma^t \frac{\Gamma(\frac{t}{\beta} + \alpha)}{\Gamma(\alpha)} \quad (2.3.10)$$

em que $-\alpha\beta < t < +\infty$.

2.3.5 Distribuição logarítmica

A distribuição logarítmica baseia-se na conhecida identidade $\sum_{k=1}^{\infty} \frac{-1}{\ln(1-p)} \frac{p^k}{k} = 1$, para $0 < p < 1$. Seguindo Johnson et al. (2005) temos a massa da distribuição definida como:

$$f_{LGR}(x; p, k) = \frac{-1}{\ln(1-p)} \frac{p^k}{k}. \quad (2.3.11)$$

Com função de distribuição acumulada:

$$F_{LGR}(x; p, k) = 1 + \frac{B(p; k + 1, 0)}{\ln(1 - p)}, \quad (2.3.12)$$

em que

$$k \geq 1, 0 < p < 1$$

e

$$B(x; a, b) = \int_0^x t^{a-1}(1-t)^{b-1} dt.$$

.

Ainda seguindo Johnson et al. (2005) temos a função geradora de momentos como:

$$M_x(t) = \frac{\ln(1 - p \exp(t))}{\ln(1 - p)}. \quad (2.3.13)$$

em que $(-\infty < t < -\log(p))$

2.3.6 Distribuição skew-normal

Temos a densidade da skew-normal definida

$$f_{SN}(x; \mu, \sigma, \lambda) = \frac{2}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right) \Phi\left(\lambda \frac{x - \mu}{\sigma}\right), \quad (2.3.14)$$

com função acumulada:

$$F_{SN}(x; \mu, \sigma, \lambda) = \frac{2}{\sigma} \Phi\left(\frac{x - \mu}{\sigma}\right) - 2T\left(\frac{x - \mu}{\sigma}, \lambda\right), \quad (2.3.15)$$

em que

$$x, \mu, \lambda \in R, \sigma > 0$$

e

$$T(t, a) = \frac{1}{2\pi} \int_0^a \frac{\exp[-\frac{1}{2}t^2(1+x^2)]}{1+x^2} dx.$$

A função geradora de momentos, para y (com $\mu = 0$ e $\sigma = 1$) é dada em Azzalini (1985) como:

$$M_y(t) = 2\exp\left[\frac{t^2}{2}\right]\Phi(\delta t),$$

onde $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$. Para $x = \sigma y + \mu$, podemos obter a função geradora de momentos:

$$\begin{aligned} M_x(t) &= \exp[\mu t]M_y(\sigma t) \\ &= (2)\exp\left[\mu t + \frac{(\sigma t)^2}{2}\right]\Phi(\sigma\delta t) \end{aligned} \quad (2.3.16)$$

em que $-\infty < t < +\infty$

2.3.7 Distribuição Kumaraswamy

A distribuição Kumaraswamy foi inicialmente apresentada em relação com hidrologia (veja, Kumaraswamy 1980) e pode se dizer que havia sido pouco estudada, e certamente não havia sido investigada de forma sistemática até Jones (2009), que explora as aplicações práticas distribuição bem como apresenta alguns resultados teóricos.

A densidade Kumaraswamy tem a forma:

$$f_k(x; \alpha, \beta) = \alpha\beta x^{\alpha-1}(1-x^\alpha)^{\beta-1}, \quad (2.3.17)$$

com função de distribuição:

$$F_k(x; \alpha, \beta) = 1 - (1-x^\alpha)^\beta. \quad (2.3.18)$$

em que

$$x \in [0, 1], \alpha, \beta > 0$$

e

$$B(p, q) = \int_0^1 t^{p-1}(1-t)^{q-1} dt$$

Em Jones (2009) encontramos ainda a seguinte expressão para o t-ésimo momento:

$$E(X^t) = \beta B \left(1 + \frac{t}{\alpha}, \beta \right), \quad (2.3.19)$$

em que $-\alpha < t < +\infty$

2.4 Alguns exemplos clássicos de misturas finitas

No presente capítulo as expressões para a mistura de k -componentes e as classes de misturas de cada uma das distribuições abordadas podem ser obtidas usando respectivamente (2.2.2) e (2.2.3).

2.4.1 Mistura de densidades normais

Sejam $f_1, \dots, f_k \in \mathcal{F}_N$ tal que,

$$\mathcal{F}_N = \left\{ f : f(x; \mu_i, \sigma_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp \left[-\frac{1}{2} \left(\frac{x - \mu_i}{\sigma_i} \right)^2 \right] \right\} \quad (2.4.1)$$

em que

$$x, \mu_i \in R, \sigma_i > 0, \forall i = 1, 2, \dots, k.$$

Um dos primeiros estudos direcionados a um modelo de mistura de normais pode ser encontrado em Pearson (1884) que usou o método de momentos para estimar os parâmetros de uma mistura finita de normais com duas componentes normais na mistura ($k=2$) e variâncias desiguais. Partindo desse marco histórico essa é uma distribuição que foi continuamente estudada e aplicada, algumas referências notáveis são: Behboodian (1970, 1972), Quandt e Ramsey (1978), Titterington et al. (1985) e McLachlan e Peel (2000).

McLachlan e Peel (2000) é uma referência particularmente didática sobre o assunto, propondo inicialmente a discussão sobre misturas de k componentes normais como parte de uma introdução às misturas de distribuições como um todo. Fazendo um apanhado da literatura existente são apresentadas algumas características descritivas das misturas de normais homo e heterocedásticas (em particular quando $k = 2$) considerando principalmente

a modalidade (a referência direta aqui é o capítulo 5 de Titterington et al. (1985)), que apresenta um bom estudo sobre o tema). Também são discutidos a importância desse tipo de mistura na modelagem de dados assimétricos, a importância histórica em estudos genéticos bem como algumas aplicações diretas em medicina.

2.4.2 Mistura de densidades Weibull

Sejam $f_1, \dots, f_k \in \mathcal{F}_W$ tal que,

$$\mathcal{F}_W = \left\{ f : f(x; c_i, \alpha_i) = \frac{c_i}{\alpha_i} \left(\frac{x}{\alpha_i} \right)^{c_i-1} \exp \left[- \left(\frac{x}{\alpha_i} \right)^{c_i} \right] \right\} \quad (2.4.2)$$

em que

$$x \in R; c_i > 0, \alpha_i > 0, \forall i = 1, 2, \dots, k.$$

O modelo Weibull é de particular interesse pois contém as densidades exponencial (para $c_i = 1$) e Rayleigh (para $c_i = 2$). As primeiras aplicações para modelos de misturas Weibull começam a aparecer no final da década de 60, com destaque para o trabalho de Kao (1959), que estudou a distribuição referente à vida de tubos de elétrons. Murthy et al (2004) apresenta uma tabela com nada menos que 30 outras aplicações de misturas de modelos Weibull em confiabilidade (reliability), com mais 13 aplicações em diversos campos como: agricultura, pluviometria, biologia e etc.

Também em Murthy et al (2004), que é um livro exclusivamente dedicado à modelos Weibull, encontramos organizados muitos resultados de interesse, bem como considerações sobre outros tipos de modelos de misturas para Weibull (Weibull inversa e modelos híbridos). São apresentados considerações para a modalidade, resultados sobre momentos e o Weibull probability plot (WPP) com estudos de caso especiais considerando duas componentes na mistura (k=2).

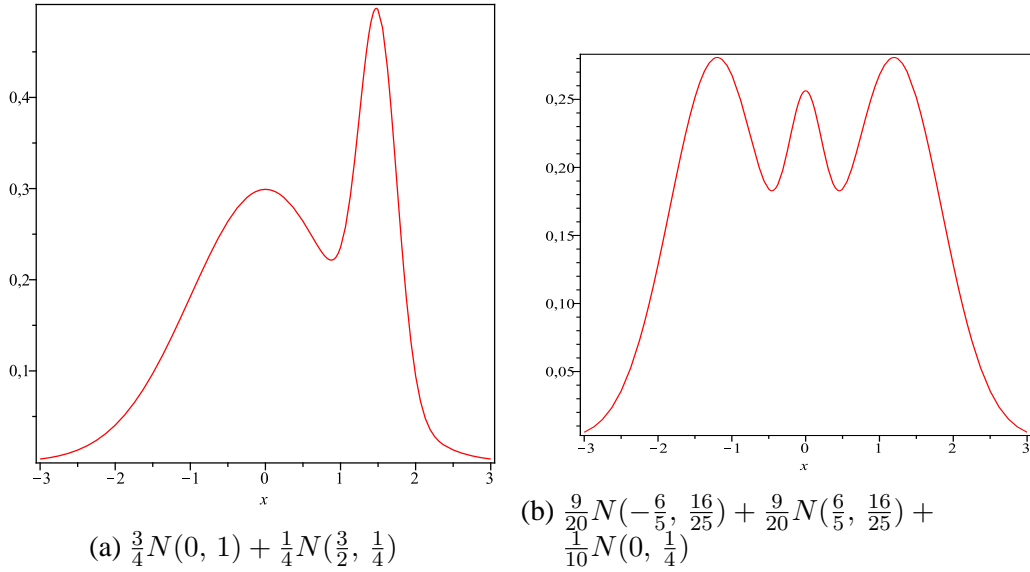


Figura 1: Mistura de densidades normais

2.4.3 Mistura de densidades lognormal

Sejam $f_1, \dots, f_k \in \mathcal{F}_{LN}$ tal que,

$$\mathcal{F}_{LN} = \left\{ f : f(x; \mu_i, \sigma_i) = \frac{1}{\sqrt{2\pi\sigma^2x^2}} \exp \left[-\frac{1}{2} \left(\frac{\log x - \mu_i}{\sigma_i} \right)^2 \right] \right\} \quad (2.4.3)$$

em que

$$x > 0; \mu_i \in R, \sigma_i > 0, \forall i = 1, 2, \dots, k.$$

Podemos encontrar em AL-Hussaini et al. (2001) 10 aplicações de modelos de misturas de componentes lognormais, divididas entre ciência atmosférica, geologia, duração de telefonemas e confiabilidade stress-strength. Algumas outras aplicações e resultados podem ser encontrados em McLachlan e Peel (2000).

Considerações teóricas sobre a distribuição em questão são encontradas organizadas em Crow e Shimuzu (1988), livro dedicado exclusivamente a essa distribuição. Sultan (1992) fez um estudo de estimação considerando misturas com duas componentes lognormais na mistura ($k = 2$) e ainda AL-Hussaini et al. (1997) fez um estudo de estimação para a confiabilidade

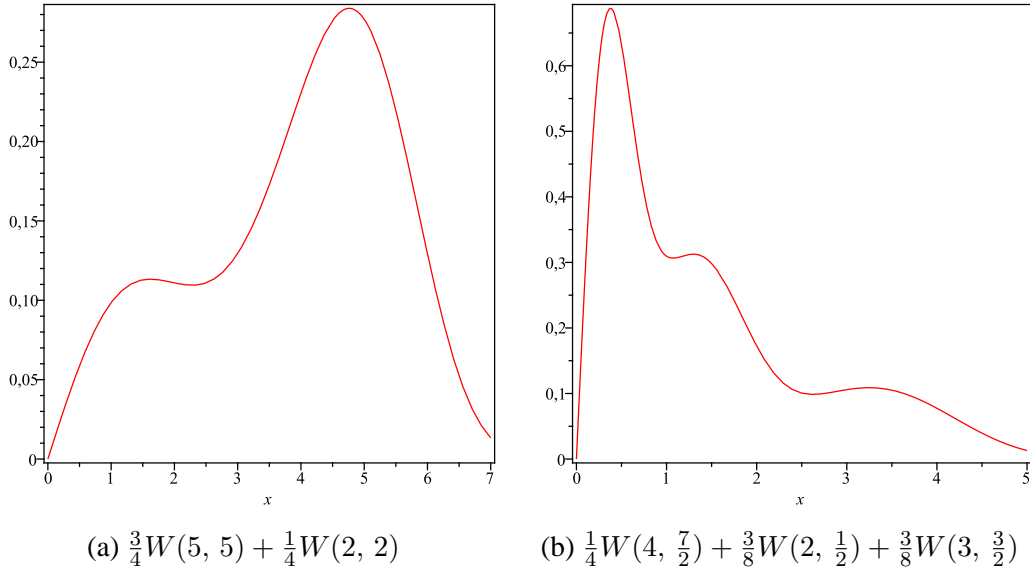


Figura 2: Misturas de densidades Weibull

stress strength $P(X < Y)$ para os casos paramétricos e não-paramétricos, onde tanto X quanto Y são misturas de lognormais independentes entre si.

2.4.4 Mistura de densidades Cauchy

Sejam $f_1, \dots, f_k \in \mathcal{F}_C$ tal que,

$$\mathcal{F}_C = \left\{ f : f(x; u_i, r_i) = \frac{1}{\pi} \left(\frac{r_i}{r_i^2 + (x - u_i)^2} \right) \right\} \quad (2.4.4)$$

em que

$$x \in R; r_i > 0, u_i \in R, \forall i = 1, 2, \dots, k$$

A aplicação clássica de modelos de misturas com componentes Cauchy aparece em Wilkins (1948), trabalho no qual o autor identifica no artigo dois casos especiais no problema da abertura de spray nos quais se aplicam a mistura de componentes Cauchy truncadas.

Uma outra aplicação de misturas de densidades Cauchy foi apresentada em Gower e Hawksford (2011), que contém ainda considerações sobre estimação. Uma fonte completa de resultados teóricos e considerações sobre a história da distribuição podem ser encontrados em Johnson et al (2005).

Podemos destacar ainda Lindsay (1995) que introduz o capítulo sobre máxima verossimilhança não-paramétrica por meio de um exemplo que considera densidades Cauchy apenas com o parâmetro de locação (a escala é igual a 1).

2.4.5 Mistura de distribuições estáveis

Antes de passarmos às considerações sobre misturas em si, alguns conceitos preliminares sobre as ditribuições estáveis são necessários. Uma variável aleatória \mathcal{X} é estável se e somente se sua função característica toma a forma:

$$\int_{-\infty}^{+\infty} e^{itx} f(x; \Psi) dx = \varphi(t; \Psi) = \begin{cases} \exp [it\mu - \sigma^\alpha |t|^\alpha \{1 - i\beta s(t) \tan(\frac{\pi\alpha}{2})\}] & \text{se } \alpha \neq 1; \\ \exp[it\mu - \sigma |t| \{1 + i\beta s(t) (\frac{\pi}{2}) \ln|t|\}] & \text{se } \alpha = 1 \end{cases}$$

em que $x \in R$ e $\Psi = (\alpha, \beta, \mu, \sigma)$ é um resumo dos parâmetros, tal que: $\alpha \in (0, 2]$ é o parâmetro de estabilidade, $\beta \in [-1, 1]$ é o parâmetro de assimetria, $\mu \in R$ é o parâmetro de locação, $\sigma > 0$ é o parâmetro de escala e $s(t)$ indica o sinal de t .

Podemos obter a densidade para estáveis como a transformação inversa de Fourier:

$$f(x; \Psi) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{(itx)} \varphi(t; \Psi) dt$$

A integral acima só tem solução analítica para algumas combinações de valores nos parâmetros: com $\alpha = 2$, $\beta = 0$ temos a função característica da distribuição normal, com $\alpha = 1$, $\beta = 0$ a função característica da distribuição Cauchy e com $\alpha = \frac{1}{2}$, $\beta = 1$ temos a função característica da distribuição Lévy. Portanto, dois exemplos de classes de misturas finitas de distribuições estáveis foram definidas nos exemplos (2.4.1) e (2.4.4).

Sejam $f_1, \dots, f_k \in \mathcal{F}_E$ tal que,

$$\mathcal{F}_E = \left\{ f : f(x; \alpha_i, \beta_i, \mu_i, \sigma_i) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{itx} \varphi(t; \Psi) dt \right\} \quad (2.4.5)$$

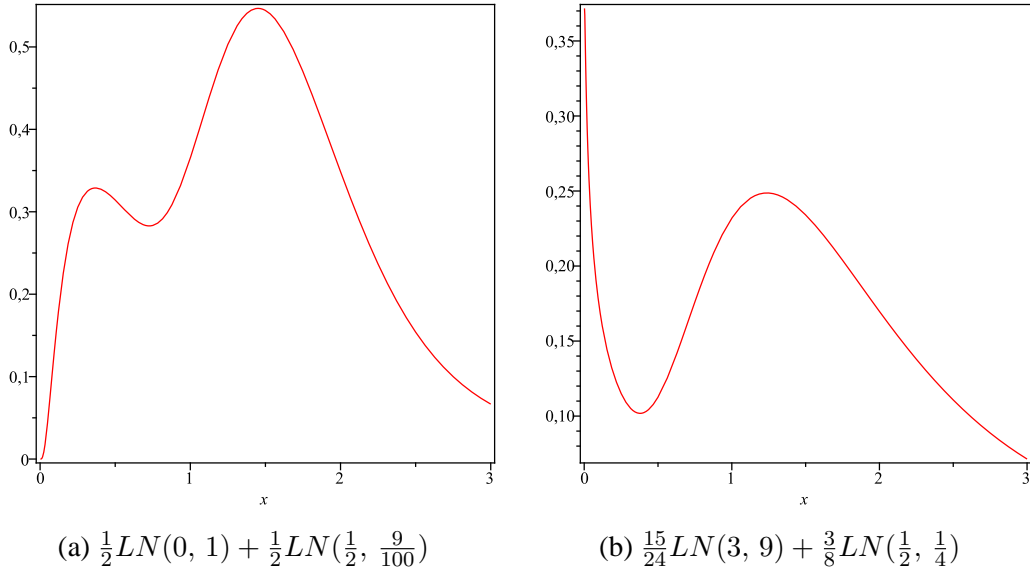


Figura 3: Misturas de densidades lognormais

em que

$$x \in R, \alpha_i \in (0, 2], \beta_i \in [-1, 1], \mu_i \in R, \sigma_i > 0, \forall i = 1, 2, \dots, k$$

Referenciais teóricos interessantes sobre misturas de estáveis, contendo considerações teóricas, métodos de estimação num contexto bayesiano bem como aplicações tanto em dados de simulação quanto em dados reais (atividade enzimática no sangue humano, acidez numa amostra de 155 lagos e velocidade de 82 galáxias na via láctea) são: Salas-Gonzales et al. (2009 e 2010). Outros exemplos de aplicações podem ser observados nos capítulos sobre misturas finitas de componentes normais e misturas finitas de componentes Cauchy.

2.4.6 Mistura de massas binomial negativa

Sejam $f_1, \dots, f_k \in \mathcal{F}_{BN}$ tal que,

$$\mathcal{F}_{BN} = \left\{ f : f(x; p_i, r_i) = \binom{x-1}{r_i-1} p_i^{r_i} (1-p_i)^{x-r_i} \right\} \quad (2.4.6)$$

em que

$$x = r_i, r_i + 1, \dots; r_i > 0, 0 < p_i < 1, \forall i = 1, 2, \dots, k$$

A primeira referência encontrada em relação à mistura abordada é o trabalho de Paull (1978) que usou a Poisson compostas (compound poisson) e binomiais negativas (como um caso particular da última) para modelar dados referentes a comportamento de consumidores quanto a compra.

Teoria e considerações históricas sobre a distribuição em si podem ser encontradas em detalhe no capítulo 5 de Johnson et al. (2005). Resultados de interesse, informações gerais, métodos de estimação, e uma aplicação (para dados clínicos de pacientes com feridas de queimadura) podem ser encontrados organizados em Tan et al. (2010).

2.4.7 Mistura de densidades gama

Sejam $f_1, \dots, f_k \in \mathcal{F}_G$ tal que,

$$\mathcal{F}_G = \left\{ f : f(x; \alpha_i, \sigma_i) = \frac{1}{\Gamma(\alpha_i)\sigma_i^{\alpha_i}} x^{\alpha_i-1} \exp\left[-\frac{x}{\sigma_i}\right] \right\} \quad (2.4.7)$$

onde

$$x, \alpha_i, \sigma_i > 0, \forall i = 1, 2, \dots, k$$

e

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$$

Assim como no exemplo Weibull, o modelo gama também é a generalização de outros modelos. Para $\alpha_i = 1$ e $\sigma_i > 0$ obtêm-se a densidade exponencial, fazendo $\alpha_i = \frac{n}{2}$, $n = 1, 2, 3, \dots$ e $\sigma_i = 2$ obtêm-se a densidade qui-quadrado (com n graus de liberdade) e por fim com $\alpha_i = k$, $k = 1, 2, 3, \dots$ e $\sigma_i > 0$ obtêm-se a densidade Erlang de ordem k. Outras considerações sobre o modelo gama poder ser encontrados em Johnson et al. (1994).

Titterton et al. (1985) cita como primeira referência de aplicação de misturas de duas densidades gama para dados de artrite por idade (mais detalhes em Masuyama (1977)). Outras aplicações incluem padrões de reconhecimento de alvos por radares em Webb (2000), estudos na área médica

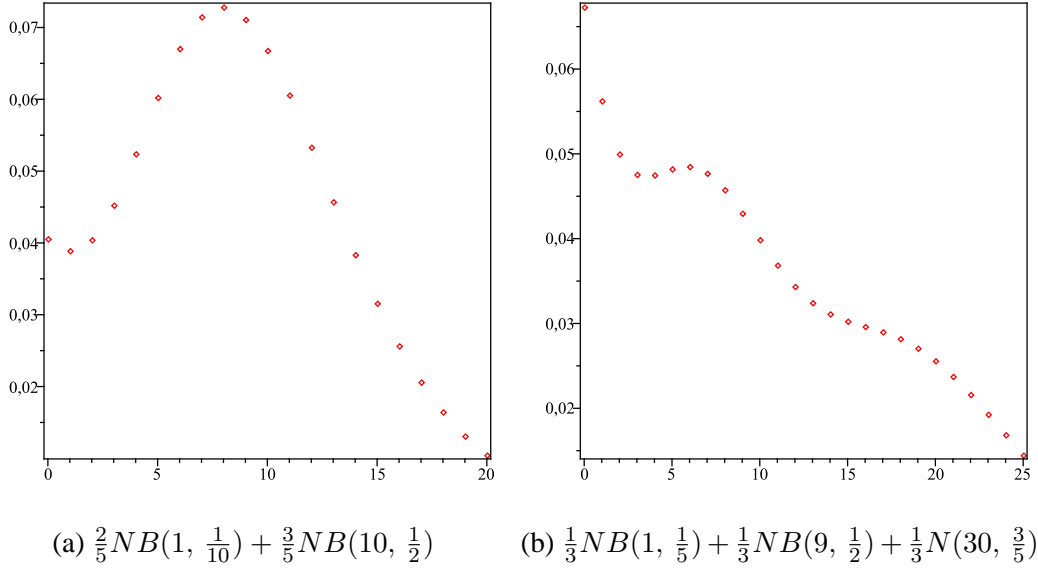


Figura 4: Misturas de massas binomial negativa

sobre características moleculares das células em Mayrose et al (2005) e por fim um sobre inferencia bayesiana em Venturini et al. (2008). Resultados teóricos sobre a mistura (em particular sobre estimação) podem ser encontrados em Titterington et al (1985).

2.4.8 Mistura de densidades exponencial

Sejam $f_1, \dots, f_k \in \mathcal{F}_{EX}$ tal que,

$$\mathcal{F}_{EX} = \left\{ f : f(x; \sigma_i) = \frac{1}{\sigma_i} \exp\left(-\frac{x}{\alpha_i}\right) \right\} \quad (2.4.8)$$

em que

$$x, \sigma_i > 0, \forall i = 1, 2, \dots, k.$$

Algumas aplicações de misturas de densidades exponencial podem ser encontradas Titterington et al. (1985) que lista pelo menos 9 exemplos de estudos que abrangem áreas bem distintas: medicina, geologia, tempos de falha e radioatividade.

Existe um grande número de resultados teóricos apresentados para misturas de densidades exponencial, com o primeiro talvez sendo Mendehall e

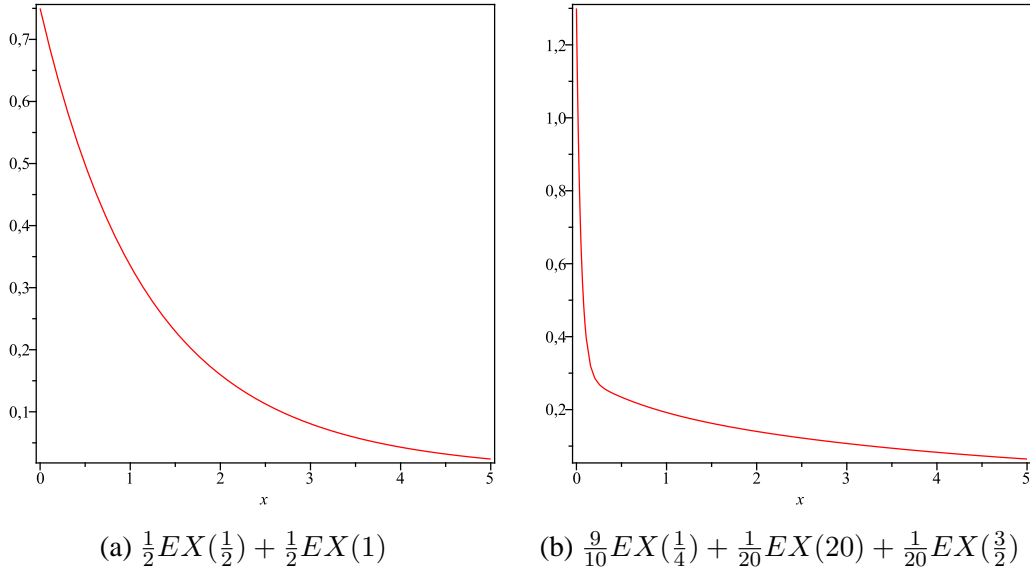


Figura 5: Misturas de densidades exponenciais

Hader (1958) um trabalho que foca em estimação para misturas de exponenciais com 2 componentes ($k = 2$). Outros exemplos de estudos (que abordam diferentes métodos de estimação, inferência bayesiana, testes estatísticos e etc) são: Rider (1961), Kleyde e Dahiya (1975), McClean (1986), Roy e Mukherjee (1988), Mccullagh (1994) e Ciuperca (2002). Mais referências de resultados e exemplos podem ser encontrados em: Balakrishnan and Basu (1995) (livro exclusivamente sobre a exponencial com um capítulo dedicado a misturas), Lindsay (1995) e McLachlan e Peel (2000).

2.4.9 Mistura de densidades logística

Sejam $f_1, \dots, f_k \in \mathcal{F}_L$ tal que,

$$\mathcal{F}_L = \left\{ f : f(x; \mu_i, \sigma_i) = \frac{\exp[-\frac{x-\mu_i}{\sigma_i}]}{\sigma_i(1 + \exp[-\frac{x-\mu_i}{\sigma_i}])^2} \right\} \quad (2.4.9)$$

em que

$$x, \mu_i \in R, \sigma_i > 0, \forall i = 1, 2, \dots, k.$$

Alternativamente, a logística pode ser expressa em termos da secante hiperbólica tal que: $f(x; \mu_i, \sigma_i) = \frac{1}{4\sigma_i} \operatorname{sech}^2(\frac{x-\mu_i}{2\sigma_i})$.

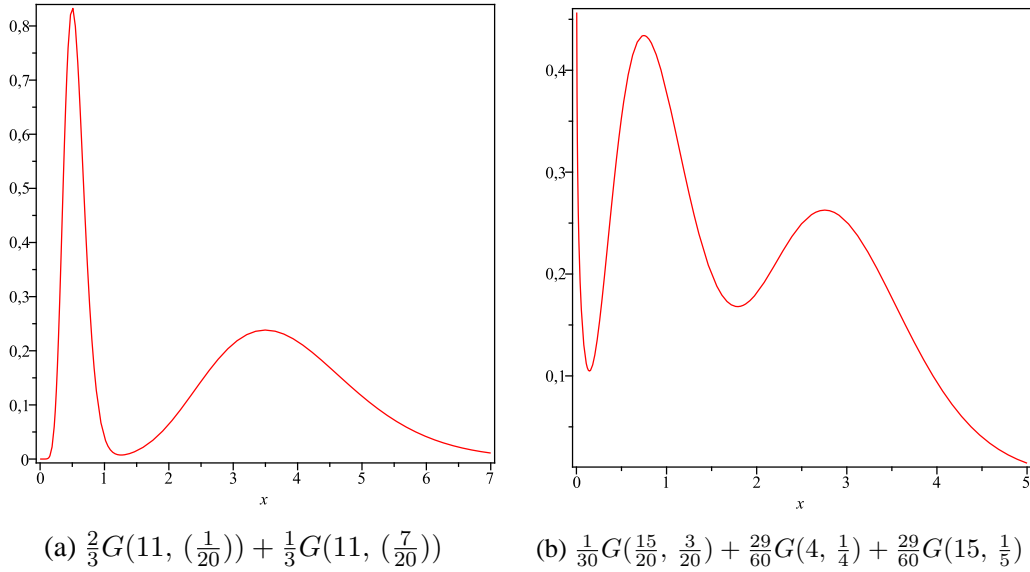


Figura 6: Misturas de densidades gama

Johnson et al. (1995) tem um capítulo dedicado à distribuição logística e Balakrishnan (1992) livro sobre a distribuição, são ótimas fontes para encontrar definições, propriedades, resultados teóricos e história da distribuição. Alguns trabalhos que envolvem misturas de logística são: Al-Hussaini e Ateya (2003 e 2005) e Ateya et al. (2011).

2.4.10 Mistura de densidades skew-normal

Sejam $f_1, \dots, f_k \in \mathcal{F}_{SN}$ tal que,

$$\mathcal{F}_{SN} = \left\{ f : f(x; \mu_i, \sigma_i, \lambda_i) = \frac{2}{\sigma_i} \phi\left(\frac{x - \mu_i}{\sigma_i}\right) \Phi\left(\lambda_i \frac{x - \mu_i}{\sigma_i}\right) \right\} \quad (2.4.10)$$

em que

$$x, \mu_i, \lambda_i \in R, \sigma_i > 0, \forall i = 1, 2, \dots, k$$

Observamos aqui que $\phi(\cdot)$ e $\Phi(\cdot)$ são respectivamente a densidade e função de distribuição acumulada da normal padrão, de modo que, podemos ainda

representar a densidade em questão como:

$$f(x; \mu_i, \sigma_i, \lambda_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{(x - \mu_i)^2}{2\sigma_i^2}\right] \int_{-\infty}^{\lambda_i \frac{(x - \mu_i)}{\sigma_i}} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{t^2}{2}\right] dt$$

Temos que μ_i é o parâmetro de locação, σ_i o de escala e λ_i o de forma, sendo que esse último regula a assimetria da distribuição e em particular $\lambda_i = 0$ a densidade se reduz à normal (i.e. a skew-normal é uma generalização da normal). Fazendo $z = \frac{x - \mu_i}{\sigma_i}$ reescrevemos a densidade skew-normal como: $f(z; \lambda) = 2\phi(z)\Phi(\lambda z)$, que aparece em grande parte dos textos sobre assunto.

Apresentada inicialmente em Azzalini (1985) a distribuição skew-normal tem sido foco de muitos estudos e aplicações práticas em anos recentes¹, e isso inclui modelos de misturas finitas com componentes skew-normal. Uma referencia teórica interessante é Lin et al. (2007), que no âmbito das misturas de skew-normal apresenta métodos de estimação, tópicos sobre modelagem bayesiana além de exemplos de aplicações em dados reais. Outras aplicações práticas e estudos (que não se limitam ao caso univariado aqui apresentado) são: Kim (2008), Lin (2009), Kumar e Anusree (2011) e Kim e Genton (2011).

2.4.11 Mistura de densidades skew-t

Sejam $f_1, \dots, f_k \in \mathcal{F}_{ST}$ tal que,

$$\mathcal{F}_{ST} = \left\{ f : f(x; \xi_i, \omega_i, \lambda_i, v_i) = \frac{2}{\omega_i} t\left(\frac{x - \xi_i}{\omega_i}; v_i\right) T\left(\lambda_i \left(\frac{x - \xi_i}{\omega_i}\right) \sqrt{\frac{v_i + 1}{v_i + \left(\frac{x - \xi_i}{\omega_i}\right)^2}}; v_i + 1\right) \right\} \quad (2.4.11)$$

em que

$$x, \xi_i, \lambda_i \in R, \omega_i^2 > 0, v_i \in Z_+^*, \forall i = 1, 2, \dots, k$$

Observe aqui que $t(\cdot; v)$ e $T(\cdot; v)$ representam respectivamente a densidade e função de distribuição acumulada da t de student padrão com v graus de liberdade. Temos que ξ_i é o parâmetro de locação, ω_i o de escala e λ_i o de forma, sendo que esse último regula a assimetria da distribuição e em particular $\lambda_i = 0$ a densidade definida se reduz à t de student (i.e. a skew-t é uma generalização da t de student). A notação apresentada segue Azzalini e

¹Para uma lista bibliográfica bem completa veja: <http://azzalini.stat.unipd.it/SN/list-publ.pdf>

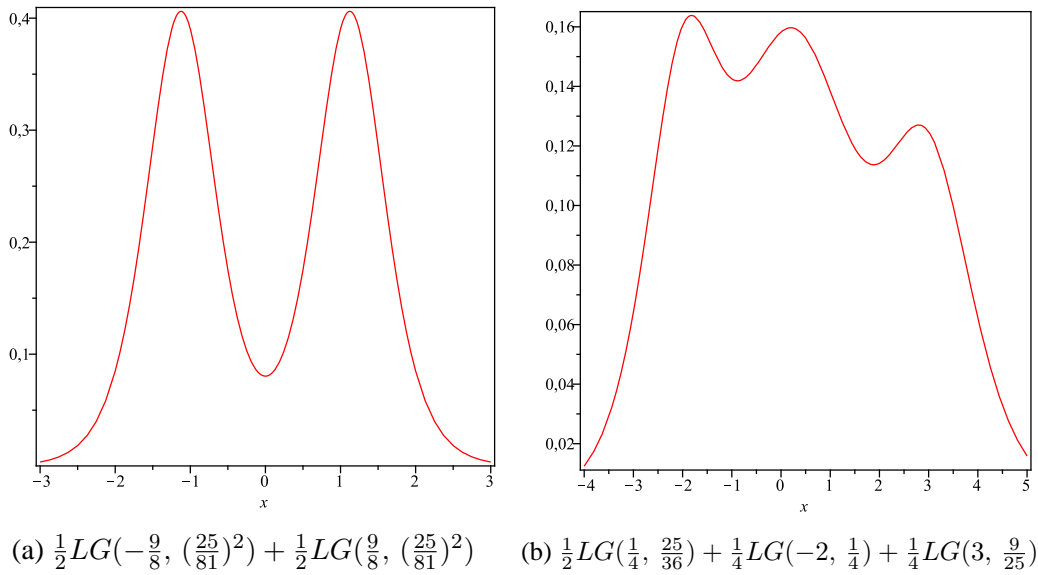


Figura 7: Misturas de densidades logísticas

Capitanio (2003), outras formas da densidade skew-t foram consideradas em Jones e Faddy (2003), Sahu et al. (2003) e Ma e Genton (2004). Branco e Dey (2001) e Kim e Mallick (2003) apresentam resultados de interesse sobre a distribuição em questão.

Considerando o modelo de misturas finitas de componentes skew-t um trabalho notável, que apresenta um método de estimação bem como um exemplo ilustrativo de aplicação do modelo, é Lin et al. (2007). Outras aplicações práticas e estudos (que não se limitam ao caso univariado aqui apresentado) são: Lin (2010) e Vrbik e McNicholas (2012).

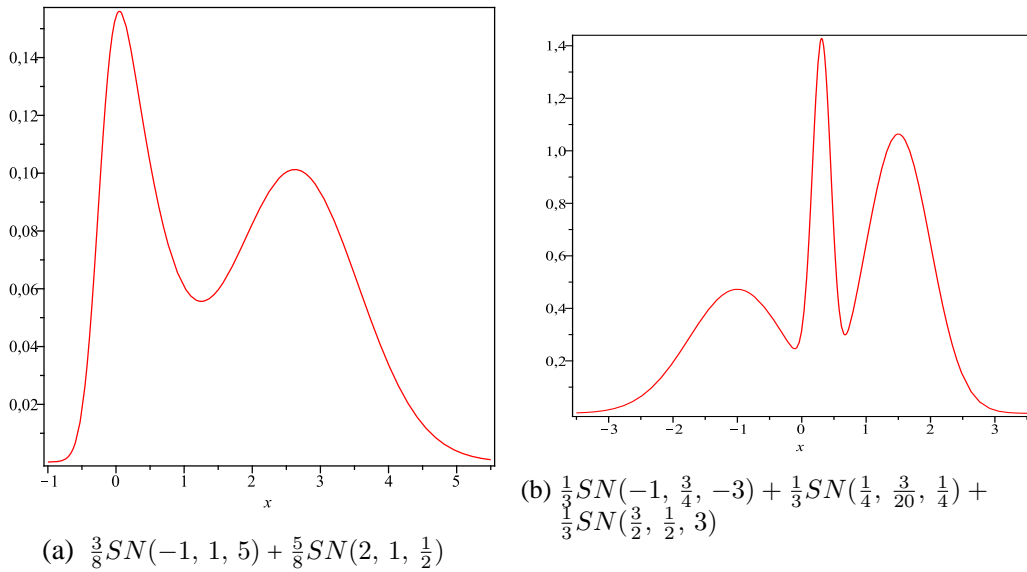


Figura 8: Mistura de densidades skew-normal

3 Identificabilidade para classes de misturas

3.1 Introdução

A identificabilidade, uma das questões centrais do presente trabalho e um importante ponto a se considerar referente à estimação, é apresentada formalmente neste capítulo na seção 3.2.

Na parte 3.3 são dadas as caracterizações de identificabilidade considerando misturas de distribuições, baseadas nos trabalhos de Teicher (1963), Yakowitz e Spragins (1968) e Atienza (2005).

As seções 3.4 e 3.5 correspondem às aplicações das caracterizações, a primeira reproduz alguns exemplos clássicos de classes de misturas de distribuições já provadas como identificáveis e a segunda apresenta provas inéditas de identificabilidade para algumas classes de misturas.

3.2 Identificabilidade

Iniciaremos esta sessão com a definição para identificabilidade considerando famílias de distribuições.

Definição 3.1 Um parâmetro θ para uma família de distribuições \mathcal{F} tal que $\mathcal{F} = \{f : f(x; \theta), x \in R^d, \theta \in \Theta\}$ é identificável se distintos valores de θ correspondem a funções de densidades ou funções de massa distintas. Isto é, $\theta \neq \theta'$ implica que $f(x; \theta)$ não é a mesma função que $f(x; \theta')$.

É importante que fique claro que identificabilidade é uma propriedade relativa ao modelo e não a algum método específico de estimação, porém se um modelo não é identificável a inferência pode ser dificultada. Por exemplo, considerando uma observação x , teríamos $f(x; \theta) = f(x; \theta')$, onde seria impossível distinguir θ de θ' , em particular ambos parâmetros dariam um mesmo valor para a função de verossimilhança. Segue um exemplo ilustrativo de família não identificável.

Exemplo 3.1 Considere a família $\mathcal{F} = \{p(x; \theta) = \binom{2}{x} \theta^x (1 - \theta)^{(2-x)}\}$ com $x = 0, 1, 2$ e $\theta \in (0, 1)$. Podemos facilmente mostrar que essa família não é identificável, basta notar que:

$$\begin{aligned} p(0; \theta) &= p'(2; 1 - \theta) = (1 - \theta)^2, \\ p(1; \theta) &= p'(1; 1 - \theta) = 2(\theta)(1 - \theta), \\ p(2; \theta) &= p'(0; 1 - \theta) = (\theta)^2 \end{aligned}$$

Temos assim que $p = p'$ quando $\theta' = 1 - \theta$, caracterizando a não identificabilidade.

A identificabilidade para misturas de distribuições é definida de forma ligeiramente diferente, para perceber por que isso acontece considere uma mistura f de densidades tal que, $f(x; \Psi) = \pi f(x; \theta_1, \theta_2) + (1 - \pi) f(x; \theta_2, \theta_1)$, em que $\Psi = (\pi, \theta_1, \theta_2)$. Note que fazendo $\Psi' = (1 - \pi, \theta_2, \theta_1)$, temos que $\Psi \neq \Psi'$ e $f(x; \Psi) = f(x; \Psi')$. Ainda mais, se todas as g componentes pertencem a uma mesma família paramétrica, a mistura $f(x; \Psi)$ não varia sobre $g!$ permutações dos índices de Ψ . Esse problema de identificabilidade em Ψ , conhecido como *label-switching*, leva à seguinte definição para misturas de densidades (para mais detalhes veja McLachlan e Peel(2000)).

Definição 3.2 Sejam $f(x; \Psi)$ e $f'(x; \Psi')$ dois membros de uma classe de

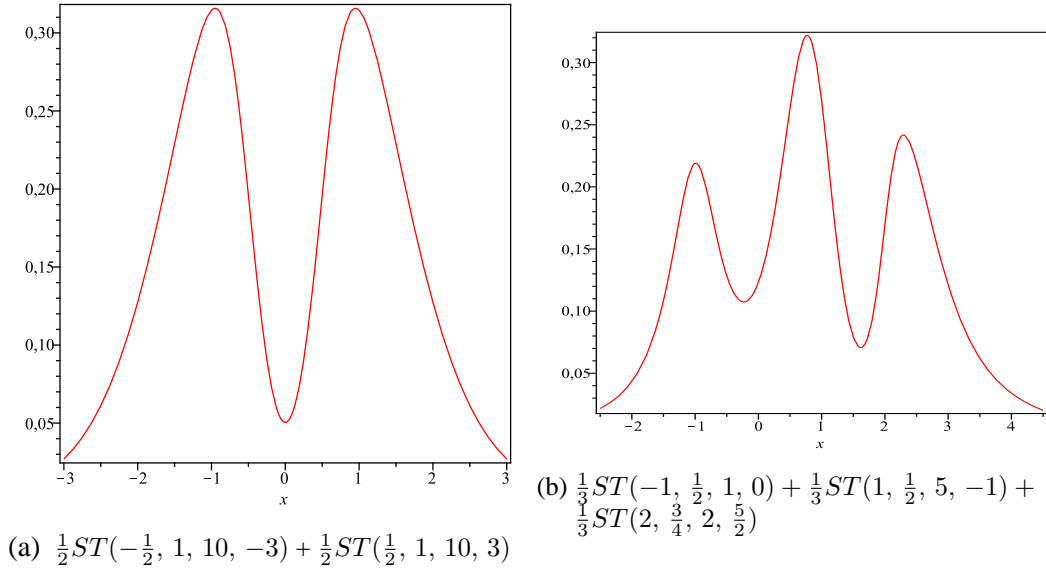


Figura 9: Mistura de densidades skew-t

mistura \mathcal{H} como em (2.2.3),

$$f(x; \Psi) = \sum_{i=1}^k \pi_i f_i(x; \theta_i)$$

e

$$f(x; \Psi') = \sum_{i=1}^{k'} \pi'_i f_i(x; \theta'_i)$$

dizemos que a classe \mathcal{H} é identificável com respeito a Ψ se: $f(x; \Psi) \equiv f'(x; \Psi')$ implica que $k = k'$ e podemos permutar os índices dentro das respectivas somas tal que tal que $\pi_i = \pi'_i$ e $f(x; \theta_i) = f'(x; \theta'_i)$, em que \equiv implica a igualdade das densidades de quase todo x . De forma equivalente, uma classe de mistura \mathcal{H} como em (2.2.5) será identificável com respeito a θ se: $F(x; \Psi) \equiv F'(x; \Psi')$ implica que $k = k'$ e podemos permutar os índices dentro das respectivas somas tal que $\pi_i = \pi'_i$ e $F(x; \theta_i) = F'(x; \theta'_i)$. Com $F(x; \Psi)$ sendo a função de distribuição relativa à densidade $f(x; \Psi)$.

Note que a definição 3.2 implica que cada mistura pertencente a uma classe \mathcal{H} de misturas deve ter uma representação única para que a classe seja identificável, segue um exemplo ilustrativo.

Exemplo 3.2 Considere $\mathcal{F} = \{p(x; \theta) = \binom{2}{x}\theta^x(1-\theta)^{2-x}, \theta \in (0, 1) \text{ e } x = 0, 1, 2\}$. Podemos definir a classe \mathcal{H} como as misturas de dois membros de \mathcal{F} onde temos:

$$f(x; \theta_1, \theta_2, \pi) = \pi p(x, \theta_1) + (1 - \pi)p(x, \theta_2) \text{ e}$$

$$f'(x; \theta'_1, \theta'_2, \pi') = \pi' p(x, \theta'_1) + (1 - \pi')p(x, \theta'_2), x = 0, 1, 2.$$

Note que:

$$f(0; \theta_1, \theta_2, \pi) = f'(2; 1 - \theta_1, 1 - \theta_2, \pi)$$

$$f(1; \theta_1, \theta_2, \pi) = f'(1; 1 - \theta_1, 1 - \theta_2, \pi)$$

$$f(2; \theta_1, \theta_2, \pi) = f'(0; 1 - \theta_1, 1 - \theta_2, \pi)$$

Assim, temos que $f = f'$ quando $\theta'_1 = 1 - \theta_1$ e $\theta'_2 = 1 - \theta_2$ caracterizando a classe \mathcal{H} como não identificável.

3.3 Caracterizações da identificabilidade

Nessa serão apresentados resultados referentes à identificabilidade de misturas finitas de distribuições. A caracterização a seguir foi demonstrado em Teicher (1963) e apresenta condições suficientes para que seja garantida a identificabilidade.

Caracterização 1 *Seja \mathcal{F} uma família de funções de distribuição univariadas com transformações $\phi(t)$, tal que t pertence a um domínio de definição $S(\phi)$ e a aplicação $M : F \rightarrow \phi$ é linear. Suponha que existe uma ordem total em \mathcal{F} , denotada por \prec , de modo que $F_1 \prec F_2$ implica:*

(i) *Os domínios $S(\phi_1)$ e $S(\phi_2)$ são tais que:*

$$S(\phi_1) \subseteq S(\phi_2) \tag{3.3.1}$$

(ii) *Existe $t_1 \in \bar{S}(\phi_1)$ (onde $\bar{S}(\phi_1)$ é o complemento de $S(\phi_1)$) com t_1 independente de $S(\phi_2)$ tal que*

$$\lim_{t \rightarrow t_1} \frac{\phi_2(t)}{\phi_1(t)} = 0 \quad (3.3.2)$$

Portanto, a classe \mathcal{H} de todas as misturas finitas de \mathcal{F} é identificável.

O próximo corolário foi demonstrado em Yakowitz e Spragins (1968), e apresenta uma condição suficiente e necessária para a identificabilidade de misturas finitas de distribuições e suas respectivas classes.

Caracterização 2 *Uma classe \mathcal{H} de misturas é identificável se e somente se a imagem de \mathcal{F} sob qualquer isomorfismo em $\langle \mathcal{F} \rangle$ seja linearmente independente no espaço da imagem, onde $\langle \mathcal{F} \rangle$ representa todas as combinações lineares dos elementos de \mathcal{F} .*

O corolário a seguir foi apresentado em Atienza et al. (2006), e fornece condições suficientes para que seja estabelecida a identificabilidade considerando uma mistura finita de distribuições de uma mesma família paramétrica. Para o seguinte corolário, A' representa o conjunto dos pontos de acumulação de $A \subset R^d$.

Caracterização 3 *Seja \mathcal{F} uma família de funções de distribuição. Seja M uma aplicação linear que transforma qualquer $F \in \mathcal{F}$ numa função real $\phi(t)$, com domínio $S(\phi)$. Seja*

$$S_0(\phi) = \{t \in S(\phi) : \phi(t) \neq 0\} \quad (3.3.3)$$

e suponha que exista t_1 satisfazendo

$$t_1 \in \left[\bigcap_{1 \leq i \leq k} S_0(\phi_i) \right]' \quad (3.3.4)$$

para qualquer coleção finita de distribuições $F_1, \dots, F_k \in \mathcal{F}$. Se a ordem

$$F_1 \prec F_2 \text{ se e somente se } \lim_{t \rightarrow t_1} \frac{\phi_2(t)}{\phi_1(t)} = 0 \quad (3.3.5)$$

é uma ordem total em \mathcal{F} , então a classe \mathcal{H} de todas as misturas finitas de distribuições de \mathcal{F} é identificável.

3.4 Prova da identificabilidade para algumas famílias de distribuições

3.4.1 Mistura de distribuições normais

Proposição 3.4.1 *A classe \mathcal{H}_N de todas as misturas finitas de \mathcal{F}_N (definida em 2.4.1), é identificável.*

Prova: Sejam $F_1 = N(x; \mu_1, \sigma_1^2)$ e $F_2 = N(x; \mu_2, \sigma_2^2)$ duas funções de distribuição normal em uma família \mathcal{F}_N , tal que $F_i = \int_{-\infty}^x f(x; \mu_i, \sigma_i^2)$ (tendo em vista a densidade em 2.4.1). Considere a aplicação linear M que transforma $F \in \mathcal{F}_N$ na função geradora de momentos que assume a forma $\phi(t) = \exp[\frac{t^2\sigma^2}{2} + \mu t]$, com domínio $S(\phi) = (-\infty, +\infty)$.

Para provar a identificabilidade usando a caracterização 3 observe inicialmente que por (3.3.3) temos $S_0(\phi) = \{t \in S(-\infty, +\infty) : \phi(t) \neq 0\}$ de onde, por (3.3.4): $t_1 \in \left[\bigcap_{1 \leq i \leq k} S_0(\phi_i) \right]' = [-\infty, +\infty)$. Para $t_1 = +\infty$ temos:

$$\lim_{t \rightarrow +\infty} \frac{\phi_2(t)}{\phi_1(t)} = \lim_{t \rightarrow +\infty} \frac{\exp[\frac{t^2\sigma_2^2}{2} + \mu_2 t]}{\exp[\frac{t^2\sigma_1^2}{2} + \mu_1 t]} = \lim_{t \rightarrow +\infty} \exp \left[t(\mu_2 - \mu_1) \exp\left[\frac{t^2}{2}(\sigma_2^2 - \sigma_1^2)\right] \right]$$

Temos assim que (3.3.5) é satisfeita quando: $F_1 \prec F_2 \Leftrightarrow \lim_{t \rightarrow +\infty} \frac{\phi_2(t)}{\phi_1(t)} \sim 0 \Leftrightarrow [\mu_1 > \mu_2, \sigma_1 = \sigma_2]$ ou $[\mu_1 = \mu_2, \sigma_1 > \sigma_2]$ que são ordens total em \mathcal{F}_N . Assim, pela caracterização 3 (Atienza 2005) a classe \mathcal{H}_N das misturas de distribuições normais é identificável.

3.4.2 Mistura de distribuições Weibull

Proposição 3.4.2 (Atienza 2005) *A classe \mathcal{H}_W de todas as misturas finitas de \mathcal{F}_W (definida em 2.4.2), é identificável.*

Prova: Sejam $F_1 = W(x; c_1, \alpha_1)$ e $F_2 = W(x; c_2, \alpha_2)$ duas funções de distribuição weibull em um família \mathcal{F}_W , tal que $F_i = \int_{-\infty}^x f_i(x; c_i, \alpha_i)$ (tendo em vista a densidade em 2.4.2). Considere a aplicação linear M que

transforma $F \in \mathcal{F}_W$ na função geradora de momentos de $\log X$, que tem domínio $S(\phi) = (-c, +\infty)$ e forma:

$$\phi(t) = E[e^{\log(x)t}] = E[X^t] = \left(\frac{1}{\alpha}\right)^t \Gamma\left(\frac{t}{c} + 1\right)$$

Usando a fórmula de Stirling temos que $\Gamma(z+1) \sim \sqrt{2\pi z} \left(\frac{z}{e}\right)^z$ para $z \rightarrow +\infty$ (i.e. assintoticamente o quociente converge para 1). Portanto:

$$\phi(t) \sim (\alpha)^t \sqrt{2\pi \frac{t}{c}} \left(\frac{t}{c}\right)^{\frac{t}{c}} e^{-\frac{t}{c}}$$

Iremos utilizar a caracterização 3 de Atienza na presente prova. Observe que por (3.3.3) temos $S_0(\phi) = \{t \in S(-c_i, \infty) : \phi(t) \neq 0\}$ de onde, por (3.3.4): $t_1 \in \left[\bigcap_{1 \leq i \leq k} S_0(\phi_i) \right]' = [-\mathcal{C}, +\infty)$, onde $\mathcal{C} = \min_{1 \leq i \leq k} (c_i)$, tal que:

$$\begin{aligned} \frac{\phi_2(t)}{\phi_1(t)} &\sim \frac{(\alpha_2)^t \sqrt{2\pi \frac{t}{c_2}} \left(\frac{t}{c_2}\right)^{\frac{t}{c_2}} e^{-\frac{t}{c_2}}}{(\alpha_1)^t \sqrt{2\pi \frac{t}{c_1}} \left(\frac{t}{c_1}\right)^{\frac{t}{c_1}} e^{-\frac{t}{c_1}}} \\ &\sim \left(\frac{\alpha_2}{\alpha_1}\right)^t \sqrt{\frac{c_1}{c_2} \frac{c_1^{\left(\frac{t}{c_2}\right)}}{c_2^{\left(\frac{t}{c_1}\right)}}} \left(\frac{t}{c_1 c_2}\right)^{\frac{t(c_1 - c_2)}{(c_1 c_2)}} \exp\left[t \frac{-(c_1 - c_2)}{c_1 c_2}\right]. \end{aligned}$$

Portanto, podemos fazer $t_1 = +\infty$ tal que:

$$\begin{aligned} \lim_{t \rightarrow +\infty} \frac{\phi_2(t)}{\phi_1(t)} &\sim \lim_{t \rightarrow +\infty} \left(\frac{\alpha_2}{\alpha_1}\right)^t \sqrt{\frac{c_1}{c_2}} \exp\left\{ [(c_1 \ln(c_1) - c_2 \ln(c_2)) - (c_1 - c_2)] \frac{t}{c_1 c_2} \right. \\ &\quad \left. + \frac{t(c_1 - c_2)}{(c_1 c_2)} \ln\left(\frac{t}{c_1 c_2}\right) \right\} \end{aligned}$$

Temos assim que (3.3.5) é satisfeita quando: $F_1 \prec F_2 \Leftrightarrow \lim_{t \rightarrow +\infty} \frac{\phi_2(t)}{\phi_1(t)} \sim 0 \Leftrightarrow [c_1 < c_2]$ ou $[c_1 = c_2, \alpha_2 < \alpha_1]$ que são ordens total em \mathcal{F}_W .

Assim, pela caracterização 3 (Atienza 2005) a classe \mathcal{H}_W das misturas de distribuição Weibull é identificável. Podemos afirmar ainda que esse resultado prova também a identificabilidade em (2.4.8), sendo a Weibull uma

generalização da exponencial como previamente afirmado.

3.4.3 Mistura de distribuições lognormais

Proposição 3.4.3 (Atienza 2005) *A classe \mathcal{H}_{LN} de todas as misturas finitas de \mathcal{F}_{LN} (definida em 2.4.3), é identificável.*

Prova: Sejam $F_1 = N(x; \mu_1, \sigma_1^2)$ e $F_2 = N(x; \mu_2, \sigma_2^2)$ duas funções de distribuição lognormal em uma família \mathcal{F}_{LN} , tal que $F_i = \int_{-\infty}^x f_i(x; \mu_i, \sigma_i^2)$ (tendo em vista a densidade em 2.4.3). Considere a aplicação linear M que transforma $F \in \mathcal{F}_{LN}$ na função geradora de momentos de $\log X$, que tem a forma $\phi(t) = \exp\left[\frac{t^2\sigma^2}{2} + \mu t\right]$ com domínio $S(\phi) = (-\infty, +\infty)$.

Iremos utilizar a caracterização 3 de Atienza na presente prova. Observe que por (3.3.3) temos $S_0(\phi) = \{t \in S(-\infty, \infty) : \phi(t) \neq 0\}$ e onde, por (3.3.4): $t_1 \in \left[\bigcap_{1 \leq i \leq k} S_0(\phi_i)\right]' = (-\infty, +\infty)$. Portanto, podemos fazer $t_1 = +\infty$ tal que:

$$\lim_{t \rightarrow +\infty} \frac{\phi_2(t)}{\phi_1(t)} = \lim_{t \rightarrow +\infty} \frac{\exp\left[\frac{t^2\sigma_2^2}{2} + \mu_2 t\right]}{\exp\left[\frac{t^2\sigma_1^2}{2} + \mu_1 t\right]} = \lim_{t \rightarrow +\infty} \exp[t(\mu_2 + \mu_1)] \exp\left[\frac{t^2}{2}(\sigma_2^2 - \sigma_1^2)\right]$$

Podemos dispor os casos em que (3.3.5) é satisfeita da seguinte forma: $F_1 \prec F_2 \Leftrightarrow \lim_{t \rightarrow +\infty} \frac{\phi_2(t)}{\phi_1(t)} = 0 \Leftrightarrow [\sigma_2^2 < \sigma_1^2]$ ou $[\sigma_2^2 = \sigma_1^2, \mu_2 < \mu_1]$, que são ordens totais em \mathcal{F}_{LN} .

Portanto pela caracterização 3 (Atienza 2006) a classe \mathcal{H}_{LN} das misturas de distribuição lognormal é identificável.

3.4.4 Mistura de distribuições Cauchy

Proposição 3.4.4 (Yakowitz e Spragins 1968) *A classe \mathcal{H}_C de todas as misturas finitas de \mathcal{F}_C (definida em 2.4.4), é identificável.*

Prova: Seja um elemento $f \in \mathcal{F}_C$ como definidas em (2.4.4). Inicialmente vamos considerar a função característica de f , dada por:

$$\phi(t; u, r) = \exp[uit - r|t|]$$

No conjunto das funções características geradas por \mathcal{F}_C uma segunda aplicação linear é feita de modo que:

$$\psi(t; u, r) = \begin{cases} \phi(t; u, r) & \text{se } t \geq 0; \\ 0 & \text{se } t < 0; \end{cases}$$

Por fim, considerando as funções compondo a imagem de ψ (note que teremos r distintas funções características, i.e. $(r) \exp[iut - rt]$), a transformação de Laplace pode ser obtida como:

$$\begin{aligned} \zeta(t; u, r) &= r \int_0^{\infty} \exp[-ts] \exp[iut - rt] \\ &= r \int_0^{\infty} \exp[-t(s + b)] \\ &= \frac{r}{s + b} \end{aligned}$$

onde $b = (r - ui)$ e s é uma variável num conjunto $D(r)$ de números complexos cuja parte real é maior que $-r$ (pois, $s + r > 0$). Observe também que $\zeta(t; u, r)$ é um isomorfismo em \mathcal{F}_C . Podemos expressar a imagem de um relação linear como:

$$0 = \sum_{j=1}^k \frac{\pi_j}{(s + b_j)}; \quad s \in D = \bigcap_{j=1}^k D(r_j)$$

onde, para que sejam atendidas as condições de identificabilidade pela caracterização 2 devemos verificar que $\pi_1 = \pi_2 = \dots = \pi_k = 0$. Suponha agora que $r_1 \leq r_j, j > 1$. Para a transformação multiplicando a expressão anterior por $(s + b_1)$ nos dois lados, com $s \rightarrow -b_1$, ainda em D , temos:

$$|\pi_1| \leq \lim_{s \rightarrow -b_1} |(s + b_1)| \sum_{i=2}^k \left| \frac{\pi_i}{(s + b_i)} \right| = 0.$$

De forma similar, os outros escalares podem ser mostrados como 0, de forma que a imagem de \mathcal{F}_C é linearmente independente, portanto, pela caracterização 2 a classe \mathcal{H}_C (Yakowitz e Spragins 1968) das misturas de distribuição

Cauchy é identificável.

3.4.5 Mistura de distribuições binomiais negativas

Proposição 3.4.5 (*Yakowitz e Spragins 1968*) *A classe \mathcal{H}_{BN} de todas as misturas finitas de \mathcal{F}_{BN} (definida em 2.4.6), é identificável.*

Prova: Definindo $F_1 = BN(x; p_1, r_1)$ e $F_2 = BN(x; p_2, r_2)$ duas funções de distribuição binomial negativa duas funções de distribuição normal em uma família \mathcal{F}_{BN} , tal que $F_i = \sum_{x_j \leq x} f(x_j; p_i, r_i)$ (tendo em vista a massa em 2.4.6). Considere a aplicação linear M que transforma $F \in \mathcal{F}_{BN}$ na função geradora de momentos, de forma $\phi(t) = (\frac{p}{1-qt})^r$ com domínio $q = (1-p)$ e domínio $S(\phi) = ((q)^{-1}, +\infty)$.

Para provar a identificabilidade usando a caracterização 1, devemos escrever uma ordem total em \mathcal{F}_N e mostrar que as condições (3.3.1) e (3.3.2) são verificadas. Considere assim a ordem total $F_1 \prec F_2$ caso $p_2 > p_1$ ou $p_2 = p_1, r_2 < r_1$, com: $S(\phi) = (q^{-1}, +\infty)$, temos:

- (i) $S(\phi_1) = (q_1^{-1}, +\infty) \subset (q_2^{-1}, +\infty) = S(\phi_2)$
- (ii) $\bar{S}(\phi_1) = (-\infty, q_1^{-1}]$, portanto podemos fazer $t_1 = (q_1)^{-1}$, independente de $S(\phi_2)$ tal que:

$$\lim_{t \rightarrow q_1^{-1}} \frac{\phi_2(t)}{\phi_1(t)} = \frac{(\frac{p_1}{1-q_1 t})^{r_1}}{(\frac{p_2}{1-q_2 t})^{r_2}} = \lim_{t \rightarrow q_1^{-1}} (1 - q_1 t)^{r_1} \frac{(p_2)^{r_2}}{(p_1)^{r_1} (1 - q_2 t)^{r_2}} = 0$$

Portanto, pela caracterização 1 (Teicher 1963) a classe \mathcal{H}_{BN} das misturas de distribuição binomial negativa é identificável.

3.4.6 Mistura de distribuições gama

Proposição 3.4.6 (*Teicher 1963*) *A classe \mathcal{H}_G de todas as misturas finitas de \mathcal{F}_G (definida em 2.4.7), é identificável.*

Prova: Sejam $F_1 = G(x; \alpha_1, \beta_1)$ e $F_2 = G(x; \alpha_2, \beta_2)$ duas funções de distribuição Weibull em um família \mathcal{F}_G , tal que $F_i = \int_{-\infty}^x f_i(x; \alpha_i, \beta_i)$ (tendo em vista a densidade em 2.4.7). Considere a aplicação linear M como a

transformação de Laplace leva $F \in \mathcal{F}_G$ para a forma $\phi(t) = (1 + \frac{t}{\beta})^{-\alpha}$ com domínio $S(\phi) = t : (-\beta, +\infty)$. A transformação citada pode ser obtida da seguinte maneira:

$$\begin{aligned}\phi(t) &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \exp[-tx] x^{\alpha-1} \exp[-x\beta] dx \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} \exp[-x(t + \beta)] dx \\ &= \left(1 + \frac{t}{\beta}\right)^{-\alpha}\end{aligned}$$

O resultado é obtido pois na integral da segunda linha temos o núcleo de uma gama com parâmetros α e $t + \beta$. Para usar agora a caracterização 1 devemos escrever uma ordem total em \mathcal{F}_G e mostrar que esta verifica as condições (3.3.1) e (3.3.2). Considere como ordem total, $F_1 \prec F_2$ caso $\beta_1 < \beta_2$ ou $\beta_1 = \beta_2, \alpha_1 > \alpha_2$, de onde:

- (i) $S(\phi_1) = (-\beta_1, +\infty) \subset (-\beta_2, +\infty) = S(\phi_2)$
- (ii) $\bar{S}(\phi_1) = (-\infty, -\beta_1]$, portanto podemos fazer $t_1 = -\beta_1$, independente de $S(\phi_2)$ tal que:

$$\lim_{t \rightarrow -\beta_1} \frac{\phi_2(t)}{\phi_1(t)} = \lim_{t \rightarrow -\beta_1} \frac{(1 + \frac{t}{\beta_2})^{-\alpha_2}}{(1 + \frac{t}{\beta_1})^{-\alpha_1}} = \lim_{t \rightarrow -\beta_1} \frac{(1 + \frac{t}{\beta_1})^{\alpha_1}}{(1 + \frac{t}{\beta_2})^{\alpha_2}} = 0$$

Portanto, pela caracterização 1 (Teicher 1963) a classe \mathcal{H}_G das misturas de distribuição gama é identificável.

3.5 Novas provas de identificabilidade para algumas classes de misturas

As proposições do presente capítulo são alguns de nossos resultados para este trabalho. Em todas as provas foi utilizado a caracterização 3 que foi demonstrada em Atienza (2005).

3.5.1 Mistura de distribuições logísticas

Proposição 3.5.1 *A classe \mathcal{H}_L de todas as misturas finitas de \mathcal{F}_L (definida em 2.4.9), é identificável.*

Prova: Sejam $F_1 = G(x; \mu_1, \sigma_1)$ e $F_2 = G(x; \mu_2, \sigma_2)$ duas funções de distribuição logística (veja 2.3.2) em um família \mathcal{F}_L . Considere a aplicação linear M que transforma $F \in \mathcal{F}_L$ na função geradora de momentos que assume a forma $\phi(t) = \exp[\mu t] \Gamma(1 + \sigma t) \Gamma(1 - \sigma t)$ apresentada em (2.3.3), com domínio $S(\phi) = (-\frac{1}{\sigma}, \frac{1}{\sigma})$.

Neste caso, usando (3.3.3) temos que: $S_0(\phi) = \{t \in (-\frac{1}{\sigma_i}, \frac{1}{\sigma_i}) : \phi(t) \neq 0\}$ de onde, por (3.3.4) $\left[\bigcap_{1 \leq i \leq k} S_0(\phi_i) \right]' = (-\mathcal{C}, \mathcal{C})$, em que $\mathcal{C} = \min_{1 \leq i \leq k} \left(\frac{1}{\sigma_i} \right)$. Portanto, podemos fazer $t_1 = \frac{1}{\sigma_1}$ tal que:

$$\lim_{t \rightarrow \frac{1}{\sigma_1}} \frac{\phi_2(t)}{\phi_1(t)} = \lim_{t \rightarrow \frac{1}{\sigma_1}} \exp[(\sigma_2 - \sigma_1)t] \frac{\Gamma(1 + \sigma_2 t) \Gamma(1 - \sigma_2 t)}{\Gamma(1 + \sigma_1 t) \Gamma(1 - \sigma_1 t)}$$

O caso em que (3.3.5) é satisfeita é o seguinte: $F_1 \prec F_2 \Leftrightarrow \lim_{t \rightarrow \frac{1}{\sigma_1}} \frac{\phi_2(t)}{\phi_1(t)} = 0 \Leftrightarrow [\sigma_2 < \sigma_1]$ que é uma ordem total em \mathcal{F}_L . Portanto pela caracterização 3 (Atienza 2005) a classe \mathcal{H}_L de misturas finitas de distribuição logística é identificável.

3.5.2 Mistura de distribuições logísticas generalizadas

Proposição 3.5.2 *Sejam $f_1, \dots, f_k \in \mathcal{F}_{LG}$ tal que,*

$$\mathcal{F}_{LG} = \left\{ f : f(x; \mu_i, \sigma_i, p_i, q_i) = \frac{1}{B(p_i, q_i)} \frac{e^{-q_i \left(\frac{x - \mu_i}{\sigma_i} \right)}}{\left(1 + e^{\left(\frac{x - \mu_i}{\sigma_i} \right)} \right)^{p_i + q_i}} \right\}$$

em que

$$x, \mu_i \in R, \sigma_i, p, q > 0, \forall i = 1, 2, \dots, k$$

e

$$B(p, q) = \int_0^1 t^{p-1} (1-t)^{q-1} dt.$$

A classe \mathcal{H}_{LG} de todas as misturas finitas de \mathcal{F}_{LG} , é identificável.

Prova: Sejam $F_1 = F(x; \mu_1, \sigma_1, p_1, q_1)$ e $F_2 = F(x; \mu_2, \sigma_2, p_2, q_2)$ duas funções de distribuição logística generalizada em um família \mathcal{F}_{LG} , tal que $F_i = \int_{-\infty}^x f_i(x; \mu_i, \sigma_i, p_i, q_i)$. Considere a aplicação linear M que transforma $F \in \mathcal{F}_{LG}$ no t -ésimo momento (função geradora de momentos de $\log(x)$) de fórmula $\phi(t) = \exp[\mu t] \frac{\Gamma(p+\sigma t)\Gamma(q-\sigma t)}{\Gamma(p)\Gamma(q)}$ com domínio $S(\phi) = (-\frac{p_i}{\sigma_i}, \frac{q_i}{\sigma_i})$.

Considerando a caracterização 3 note que, no caso, usando (3.3.3) temos $S_0(\phi) = \{t \in (-\frac{p_i}{\sigma_i}, \frac{q_i}{\sigma_i}) : \phi(t) \neq 0\}$ e por (3.3.4) $\left[\bigcap_{1 \leq i \leq k} S_0(\phi_i) \right]' = (-\mathcal{C}, \mathcal{D})$, em que $\mathcal{C} = \min_{1 \leq i \leq k} \left(\frac{p_i}{\sigma_i} \right)$ e $\mathcal{D} = \min_{1 \leq i \leq k} \left(\frac{q_i}{\sigma_i} \right)$ Portanto, podemos fazer $t = \frac{q_1}{\sigma_1}$ tal que:

$$\lim_{t \rightarrow \frac{q_1}{\sigma_1}} \frac{\phi_2(t)}{\phi_1(t)} = \lim_{t \rightarrow \frac{q_1}{\sigma_1}} \exp[(\mu_2 - \mu_1)t] \frac{\Gamma(p_2 + \sigma_2 t)\Gamma(q_2 - \sigma_2 t)}{\Gamma(p_1 + \sigma_1 t)\Gamma(q_1 - \sigma_1 t)} \frac{\Gamma(p_1)\Gamma(q_1)}{\Gamma(p_2)\Gamma(q_2)}$$

Podemos dispor os casos em que (3.3.5) é satisfeita da seguinte forma: $F_1 \prec F_2 \Leftrightarrow \lim_{t \rightarrow \frac{q_1}{\sigma_1}} \frac{\phi_2(t)}{\phi_1(t)} = 0 \Leftrightarrow [\frac{q_2}{\sigma_2} > \frac{q_1}{\sigma_1}]$ ou $[q_2 > q_1, \sigma_2 = \sigma_1]$ que são ordens totais em \mathcal{F}_{LG} .

Portanto pela caracterização 3 (Atienza 2005) a classe \mathcal{H}_{LG} de misturas de distribuição logística generalizada é identificável.

3.5.3 Mistura de distribuição beta generalizada do segundo tipo

Proposição 3.5.3 *Sejam $f_1, \dots, f_k \in \mathcal{F}_{BG}$ tal que,*

$$\mathcal{F}_{BG} = \left\{ f : f(x; \alpha, \sigma, p, q) = \frac{\alpha(x)^{\alpha p - 1}}{\sigma^{\alpha p} B(p, q) \left[1 + \left(\frac{x}{\sigma} \right)^\alpha \right]^{p+q}} \right\}$$

em que

$$x, \alpha_i, \sigma_i, p_i, q_i > 0, \forall i = 1, 2, \dots, k$$

e

$$B(p, q) = \int_0^1 t^{p-1} (1-t)^{q-1} dt$$

A classe \mathcal{H}_{BG} de todas as misturas finitas de \mathcal{F}_{BG} , é identificável.

Prova: Sejam $F_1 = F(x; \alpha_1, \beta_1)$ e $F_2 = F(x; \alpha_2, \beta_2)$ duas funções de distribuição beta generalizada segundo tipo (veja 2.3.7) em um família \mathcal{F}_{BG} . Considere a aplicação linear M que transforma $F \in \mathcal{F}_{BG}$ no t -ésimo momento (função geradora de momentos de $\log(x)$) de fórmula $\phi(t) = \sigma^t \frac{B(1-\frac{t}{\alpha q}, 1+\frac{t}{\alpha p})}{B(p, q)}$ apresentada anteriormente em (2.3.8), com domínio $S(\phi) = (-\alpha_i q_i, \alpha_i p_i)$.

Note que, neste caso por (3.3.3) temos que $S_0(\phi) = \{t \in (-\alpha_i p_i, \alpha_i q_i) : \phi(t) \neq 0\}$ usando agora (3.3.4) obtemos: $t_1 \in \left[\bigcap_{1 \leq i \leq k} S_0(\phi_i) \right]' = t_1 \in (-\mathcal{A}, \mathcal{B})$, em que $\mathcal{B} = \min_{1 \leq i \leq k} \alpha_i p_i$ e $\mathcal{A} = \min_{1 \leq i \leq k} (\alpha_i q_i)$. Portanto, podemos fazer $t_1 = \alpha_1 q_1$ tal que:

$$\lim_{t \rightarrow \alpha_1 q_1} \frac{\phi_2(t)}{\phi_1(t)} = \lim_{t \rightarrow \alpha_1 q_1} \frac{B(p_1, q_1)}{B(p_2, q_2)} \left(\frac{\sigma_2}{\sigma_1} \right)^t \frac{B\left(1 - \frac{t}{\alpha_2 q_2}, 1 + \frac{t}{\alpha_2 p_2}\right)}{B\left(1 - \frac{t}{\alpha_1 q_1}, 1 + \frac{t}{\alpha_1 p_1}\right)}$$

Podemos assim dispor os casos em que (3.3.5) é satisfeita da seguinte forma: $F_1 \prec F_2 \Leftrightarrow \lim_{t \rightarrow \alpha_1 q_1} \frac{\phi_2(t)}{\phi_1(t)} = 0 \Leftrightarrow [\alpha_2 q_2 > \alpha_1 q_1]$ ou $[\alpha_2 > \alpha_1, q_2 = q_1]$ ou $[q_2 > q_1, \alpha_2 = \alpha_1]$ que são ordens totais em \mathcal{F}_{BG} . Portanto pela caracterização 3 (Atienza 2005) a classe \mathcal{H}_{BG} de misturas finitas de distribuição beta generalizada do segundo tipo é identificável.

3.5.4 Distribuição gama generalizada

Proposição 3.5.4 *Sejam $f_1, \dots, f_k \in \mathcal{F}_{GG}$ tal que,*

$$\mathcal{F}_{GG} = \left\{ f : f(x; \alpha_i, \beta_i, \sigma_i) = \frac{\beta_i}{\sigma_i \Gamma(\alpha_i)} \left(\frac{x}{\sigma} \right)^{\alpha_i - 1} \exp \left[- \left(\frac{x}{\sigma} \right)^{\beta_i} \right] \right\}$$

em que

$$x, \alpha_i, \beta_i, \sigma_i > 0, \forall i = 1, 2, \dots, k$$

e

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx.$$

A classe \mathcal{H}_{GG} de todas as misturas finitas de \mathcal{F}_{GG} , é identificável.

Prova: Sejam $F_1 = F(x; \alpha_1, \beta_1)$ e $F_2 = F(x; \alpha_2, \beta_2)$ duas funções de distribuição gama generalizada em uma família \mathcal{F}_{GG} . Considere a aplicação linear M que transforma $F \in \mathcal{F}_{GG}$ no t -ésimo momento (função geradora de momentos de $\log(x)$) de fórmula $\phi(t) = \sigma^t \frac{\Gamma(\frac{t}{\beta} + \alpha)}{\Gamma(\alpha)}$, apresentada anteriormente em (2.3.10), com domínio $S(\phi) = (-\alpha\beta_i, +\infty)$.

Por (3.3.3) temos que $S_0(\phi) = \{t \in (-\alpha_i\beta_i, \infty) : \phi(t) \neq 0\}$. Usando agora (3.3.4) obtemos: $t_1 \in \left[\bigcap_{1 \leq i \leq k} S_0(\phi_i) \right]' = t_1 \in (-\mathcal{C}, +\infty)$, em que $\mathcal{C} = \min_{1 \leq i \leq k} (\alpha_i\beta_i)$. Portanto, podemos fazer $t_1 = -\alpha_i\beta_i$ tal que:

$$\lim_{t \rightarrow -\alpha_i\beta_i} \frac{\phi_2(t)}{\phi_1(t)} = \left(\frac{\sigma_2}{\sigma_1} \right)^t \frac{\Gamma(\alpha_1) \Gamma(\frac{t}{\beta_2} + \alpha_2)}{\Gamma(\alpha_2) \Gamma(\frac{t}{\beta_1} + \alpha_1)}.$$

Podemos assim dispor os casos em que (3.3.5) é satisfeita da seguinte forma: $F_1 \prec F_2 \Leftrightarrow \lim_{t \rightarrow -\alpha_i\beta_i} \frac{\phi_2(t)}{\phi_1(t)} = 0 \Leftrightarrow [\alpha_2\beta_2 > \alpha_1\beta_1]$ ou $[\alpha_2 > \alpha_1, \beta_2 = \beta_1]$ ou $[\beta_2 > \beta_1, \alpha_2 = \alpha_1]$ que são ordens totais em \mathcal{F}_{GG} . Portanto pela caracterização 3 (Atienza 2005) a classe \mathcal{H}_{BG} de misturas finitas de distribuição gama generalizada é identificável.

3.5.5 Mistura de massas logarítmicas

Proposição 3.5.5 *Sejam $f_1, \dots, f_k \in \mathcal{F}_{LGR}$ tal que,*

$$\mathcal{F}_{LGR} = \left\{ f : f(x; p) = \frac{-1}{\ln(1-p)} \frac{p_i^x}{x} \right\}$$

em que

$$x \in 1, 2, 3, \dots, 0 \leq p_i \leq 1, \forall i = 1, 2, \dots, k$$

A classe \mathcal{H}_{LR} de todas as misturas finitas de \mathcal{F}_{LGR} , é identificável.

Prova: Sejam $F_1 = F(x; p_1, k_1)$ e $F_2 = F(x; p_2, k_2)$ duas funções de distribuição logarítmica (veja 2.3.12) em um família \mathcal{F}_{LR} . Considere a aplicação linear M que transforma $F \in \mathcal{F}_{LR}$ na função geradora de momentos de fórmula $\phi(t) = \frac{\ln(1-p \exp(t))}{\ln(1-p)}$ com domínio $S(\phi) = (-\infty, -\log(p_i))$.

Note que, neste caso por (3.3.3): $S_0(\phi) = \{t \in (-\infty, -\log(p_i)) : \phi(t) \neq 0\}$ e usando (3.3.4) $\left[\bigcap_{1 \leq i \leq k} S_0(\phi_i) \right]' = (-\infty, -\mathcal{C})$, onde $\mathcal{C} = \min_{1 \leq i \leq k} \log(p_i)$. Portanto, podemos fazer $t = -\log(p_1)$ tal que:

$$\lim_{t \rightarrow -\log(p_1)} \frac{\phi_2(t)}{\phi_1(t)} = \lim_{t \rightarrow -\log(p_1)} \frac{\ln(1 - p_2 \exp(t)) \ln(1 - p_1)}{\ln(1 - p_1 \exp(t)) \ln(1 - p_2)}$$

Podemos assim dispor os casos em que (3.3.5) é satisfeita da seguinte forma: $F_1 \prec F_2 \Leftrightarrow \lim_{t \rightarrow -\log(p_1)} \frac{\phi_2(t)}{\phi_1(t)} = 0 \Leftrightarrow [p_2 > p_1]$ que é uma ordem total em \mathcal{F}_{LR} . Portanto pela caracterização 3 (Atienza 2005) a classe de misturas de distribuição logarítmica é identificável.

3.5.6 Mistura de distribuições skew-normal

Proposição 3.5.6 *A classe \mathcal{H}_{SN} de todas as misturas finitas de \mathcal{F}_{SN} (definida em 2.3.10), é identificável.*

Prova: Sejam $F_1 = G(x; \mu_1, \sigma_1)$ e $F_2 = G(x; \mu_2, \sigma_2)$ duas funções de distribuição skew-normal (veja 2.3.15) pertencentes a uma família \mathcal{F}_{SN} . Considere a aplicação linear M que transforma $F \in \mathcal{F}_{SN}$ na função geradora de momentos de fórmula $\phi(t) = (2)\exp\left[\mu t + \frac{(\sigma t)^2}{2}\right] \Phi(\sigma \delta t)$ (veja 2.3.16) e domínio $S(\phi) = (-\infty, \infty)$.

Note que, neste caso por (3.3.3): $S_0(\phi) = \{t \in (-\infty, \infty) : \phi(t) \neq 0\}$ e usando (3.3.4) $\left[\bigcap_{1 \leq i \leq k} S_0(\phi_i) \right]' = (-\infty, \infty)$. Portanto, podemos fazer $t = +\infty$ tal que:

$$\begin{aligned} \lim_{t \rightarrow +\infty} \frac{\phi_2(t)}{\phi_1(t)} &= \lim_{t \rightarrow +\infty} \frac{\exp[(\mu_2)t + (\sigma_2^2) \frac{t^2}{2}] \Phi(\sigma_2 \delta_2 t)}{\exp[(\mu_1)t + (\sigma_1^2) \frac{t^2}{2}] \Phi(\sigma_1 \delta_1 t)} \\ &= \lim_{t \rightarrow +\infty} \exp[(\mu_2 - \mu_1)t + (\sigma_2^2 - \sigma_1^2) \frac{t^2}{2}] \frac{\Phi(\sigma_2 \delta_2 t)}{\Phi(\sigma_1 \delta_1 t)} \end{aligned}$$

Podemos assim dispor os casos em que (3.3.5) é satisfeita da seguinte forma: $F_1 \prec F_2 \Leftrightarrow \lim_{t \rightarrow +\infty} \frac{\phi_2(t)}{\phi_1(t)} = 0 \Leftrightarrow [\sigma_2 < \sigma_1, \mu_2 = \mu_1, \delta_1 \geq 0]$ ou $[\sigma_2 = \sigma_1, \mu_2 < \mu_1, \delta_1 \geq 0]$ ambas ordens totais em \mathcal{F}_{SN} . Portanto, pela

caracterização 3 (Atienza 2005) a classe \mathcal{H}_{SN} de misturas finitas de distribuição logística é identificável. Caso $\delta_1 < 0$ temos que, considerando $[\sigma_2 < \sigma_1, \mu_2 = \mu_1]$ ou $[\sigma_2 = \sigma_1, \mu_2 < \mu_1]$, o limite que segue gera uma indefinição.

$$\lim_{t \rightarrow +\infty} \frac{\phi_2(t)}{\phi_1(t)} = \frac{\exp[(\mu_2 - \mu_1)t + (\sigma_2^2 - \sigma_1^2) \frac{t^2}{2}] \Phi(\sigma_2 \delta_2 t)}{\Phi(\sigma_1 \delta_1 t)}$$

Portanto aplicaremos a regra de L'Hôpital, considere $B = [(\mu_2 - \mu_1)t + (\sigma_2^2 - \sigma_1^2) \frac{t^2}{2}]$, sua derivada $B' = [(\mu_2 - \mu_1) + (\sigma_2^2 - \sigma_1^2)t]$. Considere ainda $\varphi(\sigma \delta t) = \frac{|\sigma \delta|}{\sqrt{2\pi}} e^{-\frac{(\sigma \delta t)^2}{2}}$ a densidade normal no caso. Portanto temos:

$$\begin{aligned} \lim_{t \rightarrow +\infty} \frac{\phi_2(t)}{\phi_1(t)} &= \lim_{t \rightarrow +\infty} \frac{B' \exp[B] \Phi(\sigma_2 \delta_2 t) + \exp[B] \varphi(\sigma_2 \delta_2 t)}{\varphi(\sigma_1 \delta_1 t)} \\ &= \lim_{t \rightarrow +\infty} \frac{B' \exp[B] \Phi(\sigma_2 \delta_2 t) + \exp[B] \frac{|\sigma_2 \delta_2|}{\sqrt{2\pi}} \exp[-(\sigma_2 \delta_2)^2 \frac{t^2}{2}]}{\frac{|\sigma_1 \delta_1|}{\sqrt{2\pi}} \exp[-(\sigma_1 \delta_1)^2 \frac{t^2}{2}]} \\ &= \lim_{t \rightarrow +\infty} \frac{(C + S [(\sigma_2^2 - \sigma_1^2)t]) \Phi(\sigma_2 \delta_2 t) + D \exp[-(\sigma_2 \delta_2)^2 \frac{t^2}{2}]}{\exp[(\mu_1 - \mu_2)t + (\sigma_1^2 - ((\sigma_1 \delta_1)^2 + \sigma_2^2)) \frac{t^2}{2}]} \end{aligned}$$

$$\text{com } C = \frac{(\mu_2 - \mu_1)\sqrt{2\pi}}{|\sigma_1 \delta_1|}, S = \frac{\sqrt{2\pi}}{|\sigma_1 \delta_1|} \text{ e } D = \frac{|\sigma_2 \delta_2|}{|\sigma_1 \delta_1|}.$$

Assim, para $\delta_1 < 0$ (3.3.5) é satisfeita da seguinte forma: $F_1 \prec F_2 \Leftrightarrow \lim_{t \rightarrow +\infty} \frac{\phi_2(t)}{\phi_1(t)} = 0 \Leftrightarrow [\sigma_2 = \sigma_1, \mu_2 < \mu_1, |\delta_1| < |\delta_2|]$ que é uma ordem total em \mathcal{F}_{SN} . Concluímos assim que pela caracterização 3 (Atienza 2005) a classe de misturas finitas de distribuição skew-normal é identificável também para este caso.

3.5.7 Mistura de distribuições Kumaraswamy

Proposição 3.5.7 *Sejam $f_1, \dots, f_k \in \mathcal{F}_K$ tal que,*

$$\mathcal{F}_{LGR} = \{f : f(x; \alpha_i, \beta_i) = \alpha_i \beta_i x^{\alpha_i - 1} (1 - x_i^\alpha)^{\beta_i - 1}\}$$

em que

$$x \in [0, 1], \alpha_i, \beta_i > 0, \forall i = 1, 2, \dots, k$$

A classe \mathcal{H}_K das misturas finitas de \mathcal{F}_K é identificável.

Prova: Sejam $F_1 = F(x; \alpha_1, \beta_1)$ e $F_2 = F(x; \alpha_2, \beta_2)$ duas funções de distribuição Kumaraswamy (veja 2.3.18) em uma família \mathcal{F}_K . Considere a aplicação linear M que transforma $F \in \mathcal{F}_K$ no t -ésimo momento (função geradora de momentos de $\log(x)$) de fórmula e com domínio $S(\phi) = (-\alpha_i, +\infty)$.

Note que, neste caso por (3.3.3): $S_0(\phi) = \{t \in (-\alpha_i, \infty) : \phi(t) \neq 0\}$ e usando (3.3.4) $\left[\bigcap_{1 \leq i \leq k} S_0(\phi_i) \right]' = (-\mathcal{C}, +\infty)$ onde $\mathcal{C} = \min_{1 \leq i \leq k} (\alpha_i)$ Portanto, podemos fazer $t_1 = -\alpha_1$ tal que:

$$\lim_{t \rightarrow -\alpha_1} \frac{\phi_2(t)}{\phi_1(t)} = \lim_{t \rightarrow -\alpha_1} \left(\frac{\beta_2}{\beta_1} \right) \frac{B\left(1 + \frac{t}{\alpha_2}, \beta_2\right)}{B\left(1 + \frac{t}{\alpha_1}, \beta_1\right)}$$

Podemos assim dispor os casos em que (3.3.5) é satisfeita da seguinte forma: $F_1 \prec F_2 \Leftrightarrow \lim_{t \rightarrow -\alpha_1} \frac{\phi_2(t)}{\phi_1(t)} = 0 \Leftrightarrow [\alpha_2 > \alpha_1]$ uma ordem total em \mathcal{F}_K . Portanto pela caracterização 3 (Atienza 2005) a classe de misturas da distribuição Kumaraswamy é identificável.

4 O algoritmo EM

4.1 Introdução

Desenvolvido inicialmente o problema de dados incompletos, o algoritmo EM é um método iterativo no qual o objetivo é que sejam obtidas estimativas de máxima verossimilhança.

Na parte 4.2 é feita uma breve introdução ao estimador de máxima verossimilhança, que é um preliminar à forma básica do algoritmo EM mostrado na parte 4.3.

O algoritmo EM é estendido ao caso de misturas de distribuições na parte 4.4 seguindo de perto o apresentado em Bilmes (1998). Os conceitos de 4.4 são posteriormente aplicados em 4.5, parte na qual apresentamos fórmulas para a obtenção das estimativas dos parâmetros das misturas de duas componentes de densidades logísticas, logísticas generalizadas, gama generalizada e beta generalizada de segunda ordem.

4.2 Estimador de máxima verossimilhança

Considere uma variável aleatória X com distribuição $f(x|\Psi)$, onde Ψ é um vetor de parâmetros com espaço paramétrico Θ . Suponha agora uma amostra aleatória de X , com valores $\mathcal{X} = (x_1, \dots, x_k)$, podemos definir a função de verossimilhança como:

$$\mathcal{L}(\Psi|x) = \prod_{i=1}^k f(x_i|\Psi) = f(\mathcal{X}|\Psi).$$

Nesse contexto uma estimativa de máxima verossimilhança é o valor de $\hat{\Psi}$ que maximiza $\mathcal{L}(\Psi|x)$ dentro do espaço amostral,

$$\hat{\Psi} = \operatorname{argmax}_{\Psi \in \Theta} \mathcal{L}(\Psi|x)$$

de forma equivalente,

$$\hat{\Psi} = \operatorname{argmax}_{\Psi \in \Theta} l(\Psi|x)$$

sendo $l(\Psi|x) = \log(\mathcal{L}(\Psi|x))$.

Frequentemente na prática a função de verossimilhança (ou seu respectivo log) não podem ser maximizadas analiticamente, nesses casos uma alternativa que possibilita a estimação por máxima verossimilhança de forma iterativa é o algoritmo EM (do inglês *expectation maximization*) que será apresentado a seguir.

4.3 O algoritmo EM básico

O algoritmo EM é um método de se computar a estimativa de máxima verossimilhança em situações em que o vetor de dados observados \mathcal{X} é dado como incompleto (Dempster et al. 1977). Note que a noção de dados incompletos inclui tanto a situação convencional de dados perdidos quanto o caso em que é possível simplificar uma função de verossimilhança analiticamente intratável supondo que os dados observados podem adicionados de dados 'faltantes'.

Mais uma vez vamos considerar \mathcal{X} um vetor de dados observados, gerados por uma distribuição $f(x|\Psi)$, chamaremos \mathcal{X} de dados incompletos. Vamos assumir a existência de um vetor $\mathcal{Y} = (y_1, \dots, y_k)$, que complete os dados de forma que $\mathcal{Z} = (\mathcal{X}, \mathcal{Y})$ com distribuição conjunta:

$$f(z|\Psi) = f(x, y|\Psi) = f(x, y|\Psi)f(y|x, \Psi)f(x|\Psi), \quad (4.3.1)$$

com função de verossimilhança:

$$\mathcal{L}(\Psi|z) = \prod_{i=1}^k f(z_i|\Psi) = f(\mathcal{Z}|\Psi) = f(\mathcal{X}, \mathcal{Y}|\Psi). \quad (4.3.2)$$

e $l(\Psi|z) = \log(\mathcal{L}(\Psi|z))$. Note aqui que em pelo menos um dos casos expostos estamos assumindo a existência de dados que estariam disponíveis num experimento hipotético e sendo nesse sentido não observáveis (McLachlan e Krishnan 1997). A seguinte exposição do algoritmo EM (e subsequente extensão ao caso das misturas finitas) segue de perto o apresentado em Dempster et al (1977) e Bilmes (1998). Considerando (4.3.1) a solução usual para

encontrar o estimador seria:

$$\frac{\partial l(\Psi|z)}{\partial \Psi} = 0.$$

O algoritmo EM aproxima o mesmo problema de uma forma distinta, procedendo iterativamente em termos da função de log-verossimilhança dos dados completos, como veremos a seguir. Perceba inicialmente que:

$$l(\Psi|x) = \log(f(\mathcal{X}|\Psi)) = \log\left(\int f(\mathcal{X}, y|\Psi)dy\right).$$

Temos ainda que:

$$f(\mathcal{X}|\Psi) = \frac{f(\mathcal{X}, \mathcal{Y}|\Psi)}{f(\mathcal{Y}|\mathcal{X}, \Psi)}$$

tirando o log dos dois lados na igualdade anterior obtemos por fim:

$$l(\Psi|x) = \log(f(\mathcal{X}, \mathcal{Y}|\Psi)) - \log(f(\mathcal{Y}|\mathcal{X}, \Psi)). \quad (4.3.3)$$

Considerando $\Psi^{(0)}$ como um valor inicial para Ψ e usando \mathcal{X} conhecido podemos tirar a esperança condicional dos dois lados em (4.3.3), obtendo:

$$l(\Psi|x) = E[\log(f(\mathcal{X}, \mathcal{Y}|\Psi))|\mathcal{X}, \Psi^{(0)}] - E[\log(f(\mathcal{Y}|\mathcal{X}, \Psi))|\mathcal{X}, \Psi^{(0)}] \quad (4.3.4)$$

Note que no último passo ao tomar a esperança condicionada é possível retirar do problema o vetor não observável \mathcal{Y} . Perceba ainda que:

$$E[l(\Psi|x)|\mathcal{X}, \Psi^{(0)}] = \int l(\Psi|x)f(\mathcal{Y}|\mathcal{X}, \Psi^{(0)})dy = l(\Psi|x).$$

Definiremos aqui,

$$Q(\Psi;\Psi^{(0)}) = E[\log(f(\mathcal{X}, \mathcal{Y}|\Psi))|\mathcal{X}, \Psi^{(0)}]$$

e

$$H(\Psi;\Psi^{(0)}) = E[\log(f(\mathcal{Y}|\mathcal{X}, \Psi))|\mathcal{X}, \Psi^{(0)}]$$

sendo por fim

$$l(\Psi|x) = Q(\Psi, \Psi^{(0)}) - H(\Psi, \Psi^{(0)}). \quad (4.3.5)$$

O algoritmo EM como apresentado em Dempster et al (1977) consiste de dois passos E (expectância) e M (maximização). O passo E é definido como o calculo de $Q(\Psi; \Psi^{(0)})$ e o passo M requer que seja determinado o $\Psi^{(1)}$ que maximiza $Q(\Psi; \Psi^{(0)})$ E então os passos E e M são repetidos, mas dessa vez com $\Psi^{(1)}$ substituindo $\Psi^{(0)}$.

Para entender porque é suficiente trabalhar iterativamente na maximização da função $Q(\Psi; \Psi^{(0)})$, seguiremos um resultado apresentado em Dempster et al (1977). Partindo de (4.3.5), temos:

$$\begin{aligned} l(\Psi^{(s)}|x) - l(\Psi^{(s-1)}|x) &= [Q(\Psi^{(s)}, \Psi^{(s-1)}) - Q(\Psi^{(s-1)}, \Psi^{(s-1)})] \\ &\quad + [H(\Psi^{(s)}, \Psi^{(s-1)}) - H(\Psi^{(s-1)}, \Psi^{(s-1)})] \end{aligned} \quad (4.3.6)$$

Pela desigualdade de Jensen temos que para qualquer $\Psi \in \Theta$:

$$\begin{aligned} H(\Psi, \Psi^{(s)}) - H(\Psi^{(s)}, \Psi^{(s)}) &= E \left[\log \left(\frac{f(\mathcal{Y}|\mathcal{X}, \Psi)}{f(\mathcal{Y}|\mathcal{X}, \Psi^{(s)})} \right) | \mathcal{X}, \Psi^{(s)} \right] \\ &\leq \log \left(E \left[\frac{f(\mathcal{Y}|\mathcal{X}, \Psi)}{f(\mathcal{Y}|\mathcal{X}, \Psi^{(s)})} | \mathcal{X}, \Psi^{(s)} \right] \right) \\ &= \log \left(\int f(\mathcal{Y}|\mathcal{X}, \Psi) dy \right) = 0. \end{aligned} \quad (4.3.7)$$

Aplicando (4.3.7) em (4.3.6) e tendo em mente que $\Psi^{(s)}$ satisfaz $Q(\Psi^{(s)}; \Psi^{(s-1)}) \geq Q(\Psi; \Psi^{(s-1)})$ obtemos por fim:

$$l(\theta^{(s)}|x) - l(\theta^{(s-1)}|x) \geq Q(\Psi^{(s)}; \Psi^{(s-1)}) - Q(\Psi; \Psi^{(s-1)}) \geq 0. \quad (4.3.8)$$

Assim, de (4.3.8) concluímos que escolher $\Psi^{(s)}$ que melhore $Q(\Psi; \Psi^{(s-1)})$ sobre $Q(\Psi^{(s-1)}; \Psi^{(s-1)})$ tem um efeito igual ou maior sobre $l(\Psi^{(s)}|x)$ e $l(\Psi^{(s-1)}|x)$. Ou ainda, como colocado em Dempster et al (1977), a sequencia $\{\Psi^{(s)}\}$ converge para o ponto crítico de $l(\Psi|x)$ e assim de (4.3.8) a sequencia que maximiza $Q(\Psi|x)$ também maximiza $l(\Psi|x)$. Podemos concluir também que depois de uma iteração do algoritmo a diferença $l(\theta^{(s)}|x) - l(\theta^{(s-1)}|x)$ nunca

será negativa, isto é, uma sequência limitada de valores da verossimilhança $\{l(\Psi^{(s)})\}$ converge monotonamente para algum L^* .

Podemos assim reproduzir sintetizar o algoritmo com relação aos dados observados, faltantes (\mathcal{X} e \mathcal{Y} respectivamente) e um vetor Ψ de parâmetros, como:

1. Passo E: Cálculo de $Q(\Psi; \Psi^{(s-1)}) = E [l(\Psi|x)|\mathcal{X}, \Psi^{(s-1)}]$.
2. Passo M: Determinação do $\Psi^{(s)}$, que maximiza $Q(\Psi; \Psi^{(s-1)})$.

Os passos são alternadamente repetidos até que a diferença

$$l(\Psi^{(s)}|x) - l(\Psi^{(s-1)}|x)$$

mude por um valor arbitrariamente pequeno, considerando o caso de convergência da sequência de valores $l(\Psi^{(s)}|x)$.

4.4 O algoritmo EM para misturas

O problema de estimação no contexto de mistura de densidades é provavelmente uma das principais aplicações do algoritmo EM. Considere inicialmente o modelo:

$$h(x|\Psi) = \sum_{i=1}^k \pi_i f_i(x|\theta_i),$$

sendo $\Psi = (\pi_1, \dots, \pi_{k-1}, \theta_1, \dots, \theta_k)$ um resumo dos parâmetros da mistura e $\sum_{i=1}^k \pi_i = 1$. Nesse contexto a função de log-verossimilhança para os dados incompletos $\mathcal{X} = (x_1, \dots, x_n)$ é:

$$l(\Psi|x) = \log \left(\prod_{i=1}^n f(x_i|\Psi) \right) = \sum_{i=1}^n \log \left(\sum_{j=1}^k \pi_j f_j(x_i|\theta_j) \right),$$

que seria difícil de otimizar devido ao log da soma. Porém, podemos considerar \mathcal{X} como dados incompletos e supor a existência de dados faltantes $\mathcal{Y} = (y_1, \dots, y_n)$. Note que estamos assumindo que para cada i temos y_i tal que $y_i \in (1, \dots, k)$, com $y_i = j$ se o i -ésimo dado foi gerado pela j -ésima

componente da mistura. Se sabemos os valores de \mathcal{Y} temos,

$$l(\Psi|x, y) = \log \left(\prod_{i=1}^n f(x_i, y_i|\Psi) \right) = \sum_{i=1}^n \log (f(x_i|y_i, \Psi) f(y_i|\Psi)) \quad (4.4.1)$$

fazendo $f(y_i|\Psi) = P(Y = y_i) = \pi_{y_i}$ e $f(x_i|y_i, \Psi) = f_{y_i}(x_i|\theta_{y_i})$ em (4.4.1), obtemos por fim:

$$l(\Psi|x, y) = \sum_{i=1}^n \log (\pi_{y_i} f_{y_i}(x_i|\theta_{y_i})),$$

que dependendo da componente densidade pode ser otimizada usando várias técnicas diferentes.

Considerando que \mathcal{Y} é não observável, para proceder devemos assumi-lo como aleatório. Inicialmente encontraremos uma expressão para a distribuição dos dados faltantes. Sejam $\Psi^{(s)} = (\pi_1^{(s)}, \dots, \pi_k^{(s)}, \theta_1^{(s)}, \dots, \theta_k^{(s)})$ apropriados para $\mathcal{L}(\Psi|\mathcal{X}, \mathcal{Y})$. Dado $\Psi^{(s)}$ procederemos a calcular $f_j(x_i|\theta_j^{(s)})$ para cada i e j . Além disso, vamos considerar aqui as proporções π_j como probabilidades a priori para cada componente da mistura $\pi_j = p(j\text{-ésimo componente})$. Assim, usando o teorema de Bayes:

$$f(y_j|x_j, \Psi^{(s)}) = \frac{\pi_{y_j}^{(s)} f_{y_j}(x_j|\theta_{y_j}^{(s)})}{h(x_j|\Psi^{(s)})} = \frac{\pi_{y_j}^{(s)} f_{y_j}(x_j|\theta_{y_j}^{(s)})}{\sum_{l=1}^k \pi_l^{(s)} f_l(x_j|\theta_l^{(s)})}$$

e

$$f(\mathcal{Y}|\mathcal{X}, \Psi^{(s)}) = \prod_{i=1}^n f(y_i|x_i, \Psi^{(s)}), \quad (4.4.2)$$

sendo $\mathcal{Y} = (y_1, \dots, y_n)$ é um dado faltante com um espaço paramétrico Ω . Escolhendo valores iniciais para os parâmetros da distribuição e assumindo a existência de dados é possível encontrar a densidade marginal em (4.4.2).

Podemos assim fazer:

$$\begin{aligned}
Q(\Psi; \Psi^{(s)}) &= \sum_{\mathcal{Y} \in \Omega} l(\Psi | \mathcal{X}, \mathcal{Y}) f(\mathcal{Y} | \mathcal{X}, \Psi^{(s)}) \\
&= \sum_{\mathcal{Y} \in \Omega} \sum_{i=1}^n \log(\pi_{y_i} f_{y_i}(x_i | \theta_{y_i})) \prod_{j=1}^n f(y_j | x_j, \Psi^{(s)}) \\
&= \sum_{y_1=1}^k \sum_{y_2=1}^k \cdots \sum_{y_n=1}^k \sum_{i=1}^n \log(\pi_{y_i} f_{y_i}(x_i | \theta_{y_i})) \prod_{j=1}^n f(y_j | x_j, \Psi^{(s)}) \\
&= \sum_{y_1=1}^k \cdots \sum_{y_n=1}^k \sum_{i=1}^n \sum_{q=1}^k \delta_{q, y_i} \log(\pi_q f_q(x_i | \theta_q)) \prod_{j=1}^n f(y_j | x_j, \Psi^{(s)}) \\
&= \sum_{q=1}^k \sum_{i=1}^n \log(\pi_q f_q(x_i | \theta_q)) \sum_{y_1=1}^k \cdots \sum_{y_n=1}^k \delta_{q, y_i} \prod_{j=1}^n f(y_j | x_j, \Psi^{(s)})
\end{aligned} \tag{4.4.3}$$

Nessa forma $Q(\Psi; \Psi^{(s)})$ parece complicada, mas ainda sim pode ser simplificada. Para $q \in (1, \dots, k)$ temos:

$$\begin{aligned}
&\sum_{y_1=1}^k \cdots \sum_{y_n=1}^k \delta_{q, y_i} \prod_{j=1}^n f(y_j | x_j, \Psi^{(s)}) = \\
&= \left(\sum_{y_1=1}^k \cdots \sum_{y_{i-1}=1}^k \sum_{y_{i+1}=1}^k \cdots \sum_{y_n=1}^k \prod_{j=1, j \neq i}^n f(y_j | x_j, \Psi^{(s)}) \right) f(q | x_i, \Psi^{(s)}) \\
&= \prod_{j=1, j \neq i}^n \left(\sum_{y_j=1}^k f(y_j | x_j, \Psi^{(s)}) \right) f(q | x_i, \Psi^{(s)}) = f(q | x_i, \Psi^{(s)}), \tag{4.4.4}
\end{aligned}$$

considerando que $\sum_{i=1}^k f(i | x_j, \Psi^{(s)}) = 1$. Usando (4.12) podemos escrever

(4.11) como:

$$\begin{aligned}
Q(\Psi; \Psi^{(s)}) &= \sum_{q=1}^k \sum_{i=1}^n \log(\pi_q f_q(x_i|\theta_q)) f(q|x_i, \Psi^{(s)}) \\
&= \sum_{q=1}^k \sum_{i=1}^n \log(\pi_q) f(q|x_i, \Psi^{(s)}) + \sum_{q=1}^k \sum_{i=1}^n \log(f_q(x_i|\theta_q)) f(q|x_i, \Psi^{(s)})
\end{aligned} \tag{4.4.5}$$

Podemos aqui maximizar os dois termos em (4.4.5) de forma independente, considerando que eles não são relacionados. Considerando inicialmente o termo contendo π_q , usaremos o multiplicador de Lagrange λ com restrição $\sum_{q=1}^k \pi_q = 1$. Resolvendo a equação:

$$\frac{\partial}{\partial \pi_q} \left[\sum_{q=1}^k \sum_{i=1}^n \log(\pi_q) f(q|x_i, \Psi^{(s)}) + \lambda \left(\sum_{q=1}^k \pi_q - 1 \right) \right] = 0$$

ou

$$\sum_{i=1}^n \frac{1}{\pi_q} f(q|x_i, \Psi^{(s)}) + \lambda = 0$$

Somando os dois lados sobre q temos $\lambda = -n$, resultando por fim em:

$$\pi_q = \frac{1}{n} \sum_{i=1}^n f(q|x_i, \Psi^{(s)})$$

A expressão anterior é referente a atualização da componente π_q referentes às proporção da mistura. A expressão para atualizar os parâmetros θ_q é encontrada maximizando a segunda expressão em (4.4.5) em relação à θ_q .

Podemos, portanto sumarizar os passos do algoritmo EM para misturas da seguinte forma:

1. Passo E: Calcular

$$Q(\Psi; \Psi^{(s)}) = \sum_{q=1}^k \sum_{i=1}^n \log(\pi_q f_q(x_i|\theta_q)) f(q|x_i, \Psi^{(s)}) \tag{4.4.6}$$

em que

$$f(q|x_i, \Psi^{(s)}) = \frac{\pi_q^{(s)} f_q(x_i|\theta_q^{(s)})}{\sum_{j=1}^k \pi_j^{(s)} f_j(x_i|\theta_j^{(s)})}$$

2. Passo M: Determinar os estimadores atualizados $\pi_q^{(s+1)}$ e $\theta_q^{(s+1)}$ utilizando respectivamente:

$$\pi_q^{(s)} = \frac{1}{n} \sum_{i=1}^n f(q|x_i, \Psi^{(s)}) \quad (4.4.7)$$

e

$$\frac{\partial}{\partial \theta_q^{(s)}} \sum_{q=1}^k \sum_{i=1}^n \log(f_q(x_i|\theta_q)) f(q|x_i, \Psi^{(s)}) = 0 \quad (4.4.8)$$

4.5 Exemplos do algoritmo EM para misturas

4.5.1 Logística

Considere inicialmente a densidade que corresponde à mistura de duas logísticas:

$$h(x; \Psi) = \sum_{q=1}^2 \frac{\pi_q \exp\left[-\frac{x-\mu_q}{\sigma_q}\right]}{\sigma_q \left(1 + \exp\left[-\frac{x-\mu_q}{\sigma_q}\right]\right)^2} \quad (4.5.1)$$

com $\Psi = (\pi_q, \mu_q, \sigma_q)$, $q = 1, 2$.

Usando a fórmula em (4.4.6) podemos escrever o passo E como:

$$Q(\Psi; \Psi^{(s)}) = \sum_{q=1}^k \sum_{i=1}^n \left[-\frac{x_i - \mu_q}{\sigma_q} - \log(\sigma_q) - 2 \log\left(1 + \exp\left[-\frac{x_i - \mu_q}{\sigma_q}\right]\right) \right] f(q|x_i, \Psi^{(s)}) \quad (4.5.2)$$

com

$$f(q|x_i, \Psi^{(s)}) = \frac{\pi_q^{(s)} \exp\left[-\frac{x_i - \mu_q}{\sigma_q}\right] \left\{ \sigma_q \left(1 + \exp\left[-\frac{x_i - \mu_q}{\sigma_q}\right]\right)^2 \right\}^{-1}}{\sum_{j=1}^2 \pi_j \exp\left[-\frac{x_i - \mu_j}{\sigma_j}\right] \left\{ \sigma_j \left(1 + \exp\left[-\frac{x_i - \mu_j}{\sigma_j}\right]\right)^2 \right\}^{-1}}$$

Usando as fórmulas em (4.4.7) e (4.4.8), podemos compor o passo M.

Temos que a proporção $\pi_q^{(s)}$ é atualizada como segue.

$$\pi_q^{(s)} = \frac{1}{n} \sum_{i=1}^n \frac{\pi_q^{(s)} \exp \left[-\frac{x_i - \mu_q}{\sigma_q} \right] \left\{ \sigma_q \left(1 + \exp \left[-\frac{x_i - \mu_q}{\sigma_q} \right] \right)^2 \right\}^{-1}}{\sum_{j=1}^2 \pi_j \exp \left[-\frac{x_i - \mu_j}{\sigma_j} \right] \left\{ \sigma_j \left(1 + \exp \left[-\frac{x_i - \mu_j}{\sigma_j} \right] \right)^2 \right\}^{-1}} \quad (4.5.3)$$

Para obter as atualizações de $\mu_q^{(s)}$ devemos resolver a seguinte equação:

$$\sum_{i=1}^n \frac{f(q|x_i, \Psi^{(s)})}{\sigma_q} \left[1 - 2 \left(\frac{\exp \left[-\frac{x_i}{\sigma_q} \right]}{\exp \left[-\frac{\mu_q}{\sigma_q} \right] - \exp \left[-\frac{x_i}{\sigma_q} \right]} \right) \right] = 0. \quad (4.5.4)$$

Como a equação não é linear usaremos o método de Newton-Raphson.

Para atualizar $\sigma_q^{(s)}$ a expressão a ser resolvida é a seguinte:

$$\sum_{i=1}^n f(q|x_i, \Psi^{(s)}) \left[\frac{x_i - \mu_q}{(\sigma_q)^2} \left\{ 1 - 2 \left(\frac{\exp \left[-\frac{x_i}{\sigma_q} \right]}{\exp \left[-\frac{\mu_q}{\sigma_q} \right] - \exp \left[-\frac{x_i}{\sigma_q} \right]} \right) \right\} - \frac{1}{\sigma_q} \right] = 0. \quad (4.5.5)$$

Mais uma vez a equação não é linear mais uma vez usaremos o método de Newton-Raphson.

4.5.2 Logística generalizada

Considere inicialmente a densidade que corresponde à mistura de duas logísticas generalizadas:

$$h(x; \Psi) = \sum_{q=1}^2 \frac{\pi_q \exp \left[-\left(q_q \right) \frac{x - \mu_q}{\sigma_q} \right]}{B(p_q, q_q) \left(1 + \exp \left[-\frac{x - \mu_q}{\sigma_q} \right] \right)^{p_q + q_q}} \quad (4.5.6)$$

com $\Psi = (\pi_q, \mu_q, \sigma_q, p_q, q_q)$, $q = 1, 2$.

Usando a fórmula em (4.4.6) podemos escrever o passo E como:

$$Q(\Psi; \Psi^{(s)}) = \sum_{q=1}^k \sum_{i=1}^n \left[- (q_q) \frac{x_i - \mu_q}{\sigma_q} \right] - \log (B(p_q, q_q)) - \\ - (p_q + q_q) \log \left(1 + \exp \left[- \frac{x_i - \mu_q}{\sigma_q} \right] \right) f (q|x_i, \Psi^{(s)}) \quad (4.5.7)$$

com

$$f (q|x_i, \Psi^{(s)}) = \frac{\pi_q^{(s)} \exp \left[- (q_q) \frac{x_i - \mu_q}{\sigma_q} \right] \left\{ B(p_q, q_q) \left(1 + \exp \left[- \frac{x_i - \mu_q}{\sigma_q} \right] \right)^{p_q + q_q} \right\}^{-1}}{\sum_{j=1}^2 \pi_j \exp \left[- (q_j^{(s)}) \frac{x_i - \mu_j^{(s)}}{\sigma_j^{(s)}} \right] \left\{ B(p_j, q_j) \left(1 + \exp \left[- \frac{x_i - \mu_j}{\sigma_j} \right] \right)^{p_j + q_j} \right\}^{-1}}$$

Usando as fórmulas em (4.4.7) e (4.4.8), podemos compor o passo M. Temos que a proporção $\pi_q^{(s)}$ é atualizada da seguinte forma, em que $f (q|x_i, \Psi^{(s)})$ é como foi descrita logo acima.

$$\pi_q^{(s+1)} = \frac{1}{n} \sum_{i=1}^n f (q|x_i, \Psi^{(s)}) . \quad (4.5.8)$$

Todas as fórmulas que seguem são não-lineares e devem ser resolvidas utilizando métodos iterativos. Para obter as atualizações de $\mu_q^{(s)}$ devemos resolver a seguinte equação:

$$\sum_{i=1}^n f (q|x_i, \Psi^{(s)}) \left(\frac{q_q}{p_q} \right) \left[1 - \left(\frac{q_q}{p_q} + 1 \right) \left(\frac{\exp \left[- \frac{x_i}{\sigma_q} \right]}{\exp \left[- \frac{\mu_q}{\sigma_q} \right] - \exp \left[- \frac{x_i}{\sigma_q} \right]} \right) \right] = 0. \quad (4.5.9)$$

As atualizações de $\sigma_q^{(s)}$ são obtidas resolvendo a equação:

$$\sum_{i=1}^n f (q|x_i, \Psi^{(s)}) \left[\frac{(q_q) x_i - \mu_q}{(\sigma)^2} \times \right. \\ \left. \times \left\{ 1 - \left(\frac{q_q}{p_q} + 1 \right) \left(\frac{\exp \left[- \frac{x_i}{\sigma_q} \right]}{\exp \left[- \frac{\mu_q}{\sigma_q} \right] - \exp \left[- \frac{x_i}{\sigma_q} \right]} \right) \right\} \right] = 0. \quad (4.5.10)$$

Podemos obter as atualizações para $p_q^{(s)}$ resolvendo a seguinte equação:

$$\sum_{i=1}^n f(q|x_i, \Psi^{(s)}) \left[\psi(p_q + q_q) - \psi(p_q) - (q_q) \log \left(1 + \exp \left[-\frac{x_i - \mu_q}{\sigma_q} \right] \right) \right] = 0. \quad (4.5.11)$$

Por fim, para obter atualizações de $q_q^{(s)}$ devemos resolver a equação:

$$\sum_{i=1}^n f(q|x_i, \Psi^{(s)}) \left[\psi(p_q + q_q) - \psi(q_q) - \left\{ (p_q) \log \left(1 + \exp \left[-\frac{x_i - \mu_q}{\sigma_q} \right] \right) + \frac{x_i - \mu_q}{\sigma_q} \right\} \right] = 0. \quad (4.5.12)$$

4.5.3 Gama generalizada

Considere inicialmente a densidade que corresponde à mistura de duas gamas generalizadas:

$$h(x; \Psi) = \sum_{q=1}^2 \frac{\pi_q \beta_q}{\sigma_q \Gamma(\alpha_q)} \left(\frac{x}{\sigma_q} \right)^{\alpha_q \beta_q - 1} \exp \left[- \left(\frac{x}{\sigma_q} \right)^{\beta_q} \right], \quad (4.5.13)$$

com $\Psi = (\pi_q, \alpha_q, \beta_q, \sigma_q)$, $q = 1, 2$.

Usando (4.4.6) o passo E é escrito como:

$$Q(\Psi; \Psi^{(s)}) = \sum_{q=1}^k \sum_{i=1}^n \log(\beta_q) - (\alpha_q \beta_q) \log(\sigma_q) - \log(\Gamma(\alpha_q)) + (\alpha_q \beta_q - 1) \log(x_i) - \left(\frac{x_i}{\sigma_q} \right)^{\beta_q} f(q|x_i, \Psi^{(s)}) \quad (4.5.14)$$

em que

$$f(q|x_i, \Psi^{(s)}) = \frac{\frac{\pi_q^{(s)} \beta_q}{\sigma \Gamma(\alpha_q)} \left(\frac{x_i}{\sigma_q}\right)^{\alpha_q \beta_q - 1} \exp\left[-\left(\frac{x_i}{\sigma_q}\right)^{\beta_q}\right]}{\sum_{j=1}^2 \frac{\pi_j \beta_j}{\sigma \Gamma(\alpha_j)} \left(\frac{x_i}{\sigma_j}\right)^{\alpha_j \beta_j - 1} \exp\left[-\left(\frac{x_i}{\sigma_j}\right)^{\beta_j}\right]}.$$

Seguindo as fórmulas em (4.4.7) e (4.4.8), podemos compor as fórmulas do passo M como segue. Obtemos as atualizações da proporção $\pi_q^{(s)}$ usando a seguinte expressão:

$$\pi_q^{(s+1)} = \frac{1}{n} \sum_{i=1}^n f(q|x_i, \Psi^{(s)}) = \frac{\frac{\pi_q^{(s)} \beta_q}{\sigma \Gamma(\alpha_q)} \left(\frac{x_i}{\sigma_q}\right)^{\alpha_q \beta_q - 1} \exp\left[-\left(\frac{x_i}{\sigma_q}\right)^{\beta_q}\right]}{\sum_{j=1}^2 \frac{\pi_j \beta_j}{\sigma \Gamma(\alpha_j)} \left(\frac{x_i}{\sigma_j}\right)^{\alpha_j \beta_j - 1} \exp\left[-\left(\frac{x_i}{\sigma_j}\right)^{\beta_j}\right]}. \quad (4.5.15)$$

As atualizações de $\alpha_q^{(s)}$ podem ser obtidas resolvendo a seguinte equação:

$$\sum_{i=1}^n f(q|x_i, \Psi^{(s)}) [-\psi(\alpha_q) - \beta_q \log(\sigma_q) + \beta_q \log(x_i)] = 0,$$

que podemos reduzir à seguinte forma:

$$\psi(\alpha_q) = \frac{\sum_{i=1}^n f(q|x_i, \Psi^{(s)}) \left[\beta_q \log\left(\frac{x_i}{\sigma_q}\right)\right]}{\sum_{i=1}^n f(q|x_i, \Psi^{(s)})}. \quad (4.5.16)$$

Para $\beta_q^{(s)}$ a equação a ser resolvida é a que segue:

$$\sum_{i=1}^n f(q|x_i, \Psi^{(s)}) \left[\frac{1}{\beta_q} + \log\left(\frac{x_i}{\alpha_q}\right) \left\{\alpha_q - \left(\frac{x_i}{\sigma_q}\right)^{\beta_q}\right\}\right] = 0. \quad (4.5.17)$$

A equação (4.5.17) não é linear, e isso deve ser considerado quando resolvendo-a para $\beta_q^{(s)}$.

Por fim, para atualizar $\sigma_q^{(s)}$ devemos resolver a seguinte equação:

$$\sum_{i=1}^n f(q|x_i, \Psi^{(s)}) \left[-\frac{\alpha_q \beta_q}{\sigma_q} + \beta_q \sigma_q^{-(\beta_q+1)} x_i^{\beta_q}\right] = 0,$$

que pode ainda ser reduzida para a forma fechada:

$$\sigma_q^{(s+1)} = \left[\frac{\sum_{i=1}^n f(q|x_i, \Psi^{(s)}) \alpha_q}{\sum_{i=1}^n f(q|x_i, \Psi^{(s)}) x_i^{\beta_q}} \right]^{-\frac{1}{\beta_q}} \quad (4.5.18)$$

4.6 Beta generalizada de segunda ordem

No caso, a mistura de duas densidades da neta generalizada de segunda ordem toma a seguinte forma:

$$h(x; \Psi) = \sum_{q=1}^2 \frac{\pi_q \alpha_q (x)^{\alpha_q p_q - 1}}{\sigma_q^{(\alpha_q p_q)} B(p_q, q_q) \left[1 + \left(\frac{x}{\sigma_q} \right)^{\alpha_q} \right]^{p_q + q_q}} \quad (4.5.19)$$

com $\Psi = (\pi_q, p_q, q_q, \alpha_q, \beta_q, \sigma_q)$, $q = 1, 2$.

Usando (4.4.6) o passo E é escrito como:

$$\begin{aligned} Q(\Psi; \Psi^{(s)}) &= \sum_{q=1}^k \sum_{i=1}^n \log(\alpha_q) + (\alpha_q p_q - 1) \log(x) - \log(B(p_q, q_q)) - \\ &\quad - (\alpha_q p_q) \log(\sigma_q) - (p_q + q_q) \log \left(1 + \left(\frac{x}{\sigma_q} \right)^{\alpha_q} \right) (f(q|x_i, \Psi^{(s)})) \end{aligned} \quad (4.5.20)$$

em que

$$f(q|x_i, \Psi^{(s)}) = \frac{\pi_q^{(s)} \alpha_q (x)^{\alpha_q p_q - 1} \left\{ \sigma_q^{(\alpha_q p_q)} B(p_q, q_q) \left[1 + \left(\frac{x}{\sigma_q} \right)^{\alpha_q} \right]^{p_q + q_q} \right\}^{-1}}{\sum_{j=1}^2 \pi_j \alpha_j (x)^{\alpha_j p_j - 1} \left\{ \sigma_j^{(\alpha_j p_j)} B(p_j, q_j) \left[1 + \left(\frac{x}{\sigma_j} \right)^{\alpha_j} \right]^{p_j + q_j} \right\}^{-1}}$$

Seguindo as fórmulas em (4.4.7) e (4.4.8), podemos compor as fórmulas do passo M como segue. Obtemos as atualizações da proporção $\pi_q^{(s)}$ usando

a seguinte expressão:

$$\pi_q^{(s+1)} = \frac{1}{n} \sum_{i=1}^n \frac{\pi_q^{(s)} \alpha_q(x)^{\alpha_q p_q - 1} \left\{ \sigma_q^{(\alpha_q p_q)} B(p_q, q_q) \left[1 + \left(\frac{x}{\sigma_q} \right)^{\alpha_q} \right]^{p_q + q_q} \right\}^{-1}}{\sum_{j=1}^2 \pi_j \alpha_j(x)^{\alpha_j p_j - 1} \left\{ \sigma_j^{(\alpha_j p_j)} B(p_j, q_j) \left[1 + \left(\frac{x}{\sigma_j} \right)^{\alpha_j} \right]^{p_j + q_j} \right\}^{-1}} \quad (4.5.21)$$

Todas as seguintes fórmulas para estimação de parâmetros são não-lineares e portanto métodos iterativos devem ser usados para solucioná-las. As atualizações de $\alpha_q^{(s)}$ são obtidas resolvendo a equação que segue:

$$\sum_{i=1}^n f(q|x_i, \Psi^{(s)}) \left[\frac{1}{\alpha_q} + (p_q) \log \left(\frac{x_i}{\sigma_q} \right) \left\{ 1 - \left(1 + \frac{q_q}{p_q} \right) \left(\frac{\left(\frac{x_i}{\sigma_q} \right)^{\alpha_q}}{1 + \left(\frac{x_i}{\sigma_q} \right)^{\alpha_q}} \right) \right\} \right] = 0. \quad (4.5.22)$$

Para $\sigma_q^{(s)}$ a equação que deve ser resolvida é a seguinte:

$$\sum_{i=1}^n f(q|x_i, \Psi^{(s)}) (\alpha_q) \left[(p_q + q_q) \left(\frac{\left(\frac{x_i}{\sigma_q} \right)^{\alpha_q}}{1 + \left(\frac{x_i}{\sigma_q} \right)^{\alpha_q}} \right) - p_q \right] = 0. \quad (4.5.23)$$

Considerando agora $p_q^{(s)}$ a equação a ser resolvida é:

$$\sum_{i=1}^n f(q|x_i, \Psi^{(s)}) \left[\psi(p_q + q_q) - \psi(p_q) + (\alpha_q) \log \left(\frac{x_i}{\sigma_q} \right) - (q_q) \log \left(1 + \left(\frac{x}{\sigma_j} \right)^{\alpha_j} \right) \right] = 0. \quad (4.5.24)$$

Por fim, para $q_q^{(s)}$ devemos resolver a seguinte equação:

$$\sum_{i=1}^n f(q|x_i, \Psi^{(s)}) \left[\psi(p_q + q_q) - \psi(q_q) - (p_q) \log \left(1 + \left(\frac{x}{\sigma_j} \right)^{\alpha_j} \right) \right] = 0. \quad (4.5.25)$$

5 Simulação

5.1 Introdução

O objetivo do presente capítulo é, por meio de uma simulação, avaliar numericamente os estimadores apresentados para a mistura de duas componentes logísticas, considerando tanto a consistência quanto o viés. Na seção 5.2 são feitas algumas considerações sobre a escolha do valor inicial e na seção 5.3 apresentamos os parâmetros estimados para a mistura logística com base em amostras geradas por simulação.

5.2 Escolha do valor inicial

Um importante ponto a ser abordado é a escolha do valor inicial para o algoritmo, que pode influenciar de forma significativa a velocidade de convergência do algoritmo e sua capacidade de localizar o máximo global (Karlis e Xekalaki 2002).

Na literatura pode encontrar alguns estudos que tocam o tema, sugerindo diferentes formas de abordar o problema. Como exemplo destacamos: Laird (1978) propõe o uso de um "grid" de valores para selecionar o melhor inicial, Furman e Lindsay (1994a, b) propõe o uso de estimativas obtidas com outros métodos, no caso o método de momentos.

Em Karlis e Xekalaki (2002) temos um estudo de simulação direcionado a comparar os diferentes métodos de escolha do valor inicial, uma das possibilidades exploradas é usar o próprio valor utilizado para simular uma mistura finita como valor inicial do algoritmo EM. No presente capítulo seguiremos a mesma ideia para as misturas consideradas.

5.3 Logística

As amostras de uma classe de misturas com duas componentes logísticas foram geradas da seguinte forma:

1. São geradas duas variáveis aleatórias, u_1 e u_2 , com distribuição $U(0, 1)$

2. Caso $u_1 \geq \pi_1$, é utilizado u_2 para gerar x , tal que $x = F_2^{-1}(u_2)$

3. Caso $u_1 < \pi_1$, é utilizado u_2 para gerar x , tal que $x = F_1^{-1}(u_2)$

em que F_1^{-1} é a função acumulada inversa, de fórmula:

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

Na estimação dos parâmetros Ψ do modelo (4.5.1) serão consideradas 100 amostras simuladas usando as expressões (4.5.2), (4.5.3) e (4.5.4) sendo que para as duas ultimas (μ_q e σ_q respectivamente, para $q = 1$ e 2) as soluções para os parâmetros foram encontrados resolvendo a equação pelo método de Newton-Raphson. A seguir temos duas tabelas: a primeira mostra contém parâmetros simulados considerando, o valor esperado para as estimativas bem como o erro quadrático médio (EQM) para as estimativas de μ e σ sempre considerando o tamanho amostral $n = 100$ enquanto a segunda apresenta os erros quadráticos médios e tempos de computação do algoritmo para tamanhos de amostra diferentes, considerando uma mistura com os parâmetros $h(0.5, 2, 1, 3, 2)$.

$(p, \mu_1, \mu_2, \sigma_1, \sigma_2)$	Est. Médias	EQM
(.5, 2, 1, 3, 2)	(0.5002, 2.0490, 1.0330, 2.3930, 1.287)	(0.2570, 0.1124, 0.3936, 0.5165)
(.5, -5, 4, 2, 2.75)	(0.4967, -4.9882, 4.1000, 2.5833, 2.6820)	(0.0237, 0.3013, 0.3320, 0.0046)
(.75, 3, -4, 2, 1)	(0.7490, 3.234, -3.963, 1.946, 1.051)	(2.0639, 0.0907, 0.0029, 0.3532)
(.5, -3, 3.5, 4, 4)	(0.4981, -2.998, 3.512, 5.732, 3.646)	(0.00007, 0.0062, 3.0769, 0.1293)
(.5, 2, 6, 3, 5)	(0.4972, 2.0360, 6.084, 2.432, 4.852)	(0.1653, 0.1010, 0.3282, 0.0224)
(.75, 2, 3, 2, 3)	(0.7490, 2.028, 3.066, 1.808, 2.761)	(0.1311, 0.3883, 0.0372, 0.0601)

Tabela 1: Estimativas e EQM para misturas de logísticas

n	EQM	T. de Comp.
15	(0.7707, 1.8306, 1.0684, 0.8961)	0.72 seg.
50	(0.4795, 0.2394, 0.5778, 0.6620)	1.26 seg.
100	(0.2570, 0.1124, 0.3936, 0.5165)	1.99 seg.
500	(0.0354, 0.0227, 0.3241, 0.4989)	7.54 seg.

Tabela 2: EQM e t. de computação para $h(0.5, 2, 1, 3, 2)$

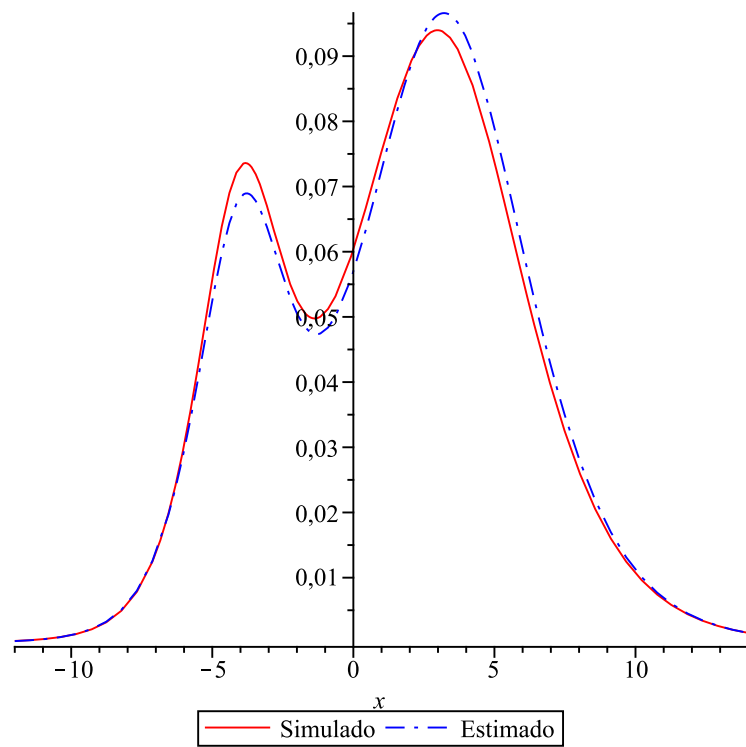


Figura 10: Gráfico da mistura logística $h(0.75, 3, -4, 2, 1)$ para $n = 100$

6 Conclusão

Neste trabalho fizemos uma revisão dos conceitos à cerca de mistura finita de distribuições e identificabilidade, mostramos que as classes de misturas finitas das distribuições logística, logística generalizada, beta generalizada do segundo tipo, gama generalizada, logarítmica e skew-normal são identificáveis e utilizamos o algoritmo EM para obter fórmulas para a estimação dos parâmetros de misturas de duas componentes das distribuições logística, logística generalizada, beta generalizada do segundo tipo e gama generalizada.

Tínhamos como objetivo, na concepção do presente trabalho, a revisão dos conceitos sobre o tema e procurar provas para novas classes de misturas de distribuições, ambos objetivos foram concluídos e ainda foi possível entrar satisfatoriamente na questão de estimação de parâmetros apresentando resultados. A simulação feita não apresentou problemas e o algoritmo montado no R poderia ser ajustado para contemplar outras distribuições. A exemplo da distribuição skew-t, existem ainda outras classes de misturas de distribuições para as quais a identificabilidade não foi formalmente demonstrada que poderiam ser foco de trabalhos futuros.

Referências bibliográficas

- [1] AL-HUSSAINI, E. K.; MOUSA, M. A. M. A.; SULTAN, K. S. Parametric and nonparametric estimation of $P(Y < X)$ for finite mixtures of lognormal components. **Communications in Statistics - Theory and Method**, v. 26, p. 1269-1289, 1997.

- [2] AL-HUSSAINI, E. K.; SULTAN, K. S. Reliability and hazard based on finite mixture models. **Handbook of Statistics**, v. 20, p. 139-183, 2001.

- [3] AL-HUSSAINI, E. K.; ATEYA, S. F. Maximum likelihood estimations under a mixture of truncated type I generalized logistic components model. **Pioneer Journal of Theoretical and Applied Statistics**, v. 2, issue 1, p. 47-60, 2003.

- [4] AL-HUSSAINI, E. K.; ATEYA, S. F. Bayes estimations under a mixture of truncated type I generalized logistic components model. **Pioneer Journal of Theoretical and Applied Statistics**, v. 4, issue 2, p. 183-208, 2005.

- [5] ATEYA, S. F.; RIZK, M. M.; EL-ADLL, M. E. Estimation under finite mixture of truncated type I generalized logistic components model based on censored data via EM algorithm. **International Mathematical Forum**, v. 6, no. 67, p. 3323-3341, 2011.

- [6] ATIENZA, N.; GARCIA-HERAS, J.; MUÑOZ-PICHARDO, J.M. A new condition for identifiability of finite mixture distributions. **Metrika**, v. 63, p. 215-221, 2006.

- [7] AZZALINI, A. A class of distributions which Includes the normal ones. **Scandinavian Journal of Statistics**, v. 12, p. 171-178, 1985.
- [8] AZZALINI, A.; CAPITANIO, A. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. **Journal of the Royal Statistical Society, B**, v. 65, p. 367-389, 2003.
- [9] BALAKRISHNAN, N. **The Handbook oh the Logistic Distribution**. New York: Marcel Dekker, 1992.
- [10] BALAKRISHNAN, N.; BASU, A.P. **The Exponential Distribution: Theory, methods and Applications**. New York: Gordon and Breach, 1995.
- [11] BEHBOODIAN, J. On a mixture of two normal distributions. **Biometrika** v. 57, p. 215-217, 1970.
- [12] BEHBOODIAN, J. Information matrix for a mixture of two normal distributions. **Journal of Statistical Computation and Simulation**, v. 1, p. 295-314, 1972.
- [13] BILMES, J. A. A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. **International Computer Science Institute**, p. 1-13, 1998.
- [14] BRANCO, M. D.; DEY, D. K. A general class of multivariate skew elliptical distributions. **Journal of Multivariate Analysis**, v. 79, p. 99-113, 2001.

- [15] CASELLA, G.; BERGER, R. L. **Statistical Inference**. 2 ed., California: Duxbury Press, 2002.
- [16] CHANDRA, S. On the mixtures of probability distributions. **Scandinavian Journal of Statistics**, v. 4, p. 105-112, 1977.
- [17] CIUPERCA, G. Likelihood ratio statistic for exponential mixtures. **Annals of the Institute of Statistical Mathematics**, v. 54, p. 585-594, 2002.
- [18] CROW, E. L.; SHIMIZU, K. **Lognormal distributions - Theory and Applications**. New York: Marcel Dekker, 1988.
- [19] DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the em algorithm **Journal of the Royal Statistical Society**, Series B, v. 39, p. 1-38, 1977.
- [20] FISH, P. R. The Graduation of Income Distributions **Econometrica**, v. 29, p. 171-185, 1961.
- [21] FURMAN, W.; LINDSAY, B. Measuring the relative effectiveness of moment estimators as starting values in maximizing likelihoods. **Computational Statistics and Data Analysis**, v. 17, p. 493-508, 1994.
- [22] FURMAN, W.; LINDSAY, B. Testing for the number of components in a mixture of normal distributions using moment estimators. **Computational Statistics and Data Analysis**, v. 17, p. 473-492, 1994.

- [23] HAWSFORD, M. O. J.; GOWER, E. S. Learning an overcomplete dictionary using a cauchy mixture model for sparse decay. **World Academy of Science, Engineering and Technology**, v. 75, p. 1227-1234, 2011.
- [24] JOHNSON, Norman L.; KOTZ, Samuel; BALAKRISHNAN, N. **Continuous Univariate Distributions**. Vol. 1. 2 ed. New York: John Wiley & Sons Inc., 1994.
- [25] JOHNSON, Norman L.; KOTZ, Samuel; BALAKRISHNAN, N. **Continuous Univariate Distributions**. Vol. 2. 2 ed. New York: John Wiley & Sons Inc., 1995.
- [26] JOHNSON, Norman L.; KEMP, Adrienne W.; KOTZ, Samuel. **Univariate Discrete Distributions**. 3 ed. New York: John Wiley & Sons Inc., 2005.
- [27] JONES, M. C.; FADDY, M. J. A skew extension of the t distribution, with applications. **Journal of the Royal Statistical Society, B**, v. 65, p. 159-174, 2003.
- [28] JONES, M. C. Kumaraswamy's distribution: A beta-type distribution with some tractability advantages. **Statistical Methodology** v. 6, p. 70-81, 2009.
- [29] KAO, J. H. K. A graphical estimation of mixed Weibull parameters in life testing of electron tubes. **Technometrics**, v. 1, p. 389-407, 1959.
- [30] KARLIS, D.; XEKALAKI, E. Choosing initial values for the em algorithm for finite mixtures. **Computational Statistics & Data Analysis**, v. 41, p. 577-590, 2002.

- [31] KHODABIN, M.; AHMADABADI, A. Some properties of generalized gamma distribution. **Mathematical Sciences Quarterly Journal**, v. 4, no. 1, p. 9-28, 2010.
- [32] KIM, H.-M.; MALLICK, B. K. Moments of random vectors with skew t distribution and their quadratic forms. **Statistics and Probability Letters**, v. 63, p. 417-423, 2003.
- [33] KIM, H.-M.; A note on scale mixtures of skew normal distribution. **Statistics and Probability Letters**, v. 78, p. 1694-1701, 2008.
- [34] KIM, H.-M.; GENTON, M. Characteristic functions of scale mixtures of multivariate skew-normal distributions. **Journal of Multivariate Analysis**, v. 102, p. 1105-1117, 2011.
- [35] KLEIBER, C.; KOTZ, S. **Statistical size distributions in economics and actuarial sciences**. 1 ed. New York: John Wiley & Sons, Inc., 2003.
- [36] KLEYDE, R.M.; DAHIYA R.C. Estimation of parameters of mixed failure time distribution from censored samples. **Communications in Statistics - Theory and Methods**, v. 4, p. 873-407, 1975.
- [37] KUMAR, C. S.; ANUSREE M. R. On a generalized mixture of standard normal and skew normal distributions. **Statistics and Probability Letters**, v. 81, p. 1813-1821, 2011.
- [38] KUMARASWAMY, P. A generalized probability density function for double-bounded random processes. **Journal of Hydrology**, v. 46, p. 79-88, 1980.

- [39] LAIRD, N. Nonparametric maximum likelihood estimation of a mixing distribution. **Journal of the American Statistical Association**, v. 73, p. 805-811, 1978.
- [40] LANG, S. **Linear Algebra**. 3 ed. New York: Springer Inc., 1987.
- [41] LIN, T. I.; JACK C. L.; HSIEH, W. J. Robust mixture modeling using the skew t distribution. **Statistics and Computing**, v. 17, p. 81-92, 2007.
- [42] LIN, T. I.; JACK C. L.; SHU, Y. Y. Finite mixture modelling using the skew normal distribution. **Statistica Sinica**, v. 17, p. 909-927, 2009.
- [43] LIN, T. I. Maximum likelihood estimation for multivariate skew normal mixture models. **Journal of Multivariate Analysis**, v. 100, p. 257-265, 2009.
- [44] LIN, T. I. Robust mixture modeling using multivariate skew t distributions. **Statistics and Computing**, v. 20, p. 343-356, 2010.
- [45] LINDSAY, B. G. **Mixture Models: Theory, Geometry, and Applications**. Hayward: Institute of Mathematical Statistics, v. 1, p. 389-407, 1995.
- [46] MA, Y.; GENTON, G. G. A flexible class of skew-symmetric distributions. **Scandinavian Journal of Statistics**, v. 31, p. 459-468, 2004.
- [47] MAGALHÃES, Marcos Nascimento. **Probabilidade e variáveis aleatórias**. 2 ed. São Paulo: EDUSP, 2006.

- [48] MAHMOUD, M. A. W.; MOUSTAFA, H. M. Estimation of a discriminant function from a mixture of two gamma distributions when the sample size is small. **Mathematical and Computer Modelling**, v. 18, p. 87-95, 1993.
- [49] MASUYAMA, M. A mixture of two gamma distributions applied to rheumatoid arthritis. **Reports of Statistical Application Research of the Union of Japanese Scientists and Engineers**, v. 24, p. 28-31, 1977.
- [50] MAYROSE, I.; FRIEDMAN, N.; PUPKO, T. A gamma mixture model better accounts for among site rate heterogeneity. **Bioinformatics**, v. 31, Suppl., p. 151-158, 2005.
- [51] MCCLEAN, S. Estimation for the mixed exponential distribution using grouped follow-up data. **Journal of Applied Statistics**, v. 15 p. 31-37, 1986.
- [52] MENDEHALL, W.; HADER, R. J. Estimation of parameters of mixed exponentially distributed failure time distributions. **Biometrika**, v. 45, p. 504-520, 1958.
- [53] MCCULLAGH, P. Exponential mixtures and quadratic exponential families. **Biometrika**, v. 81, p. 721-729, 1994.
- [54] MCLACHLAN, G.J. PEEL, D. **Finite mixture models**. New York: Wiley, 413 p. 2000.
- [55] MCLACHLAN, G.J. KRISHNAN, T. **The EM Algorithm and Extensions**. New York: Wiley, 274 p. 1997.

- [56] MURTHY, D.N.P.; XIE M; JIANG R. **Weibull Models**. John Wiley and Sons, 2004.
- [57] NEWCOMB, S. A generalized theory of the combination of observations so as to obtain the best result. **American Journal of Mathematics**, v. 8, p. 343-366, 1886.
- [58] PAULL, A. E. A generalized compound poisson model for consumer purchase panel data analysis. **Journal of the American Statistical Association**, v. 73, p. 706-713, 1978.
- [59] PEARSON, K. Contributions to the mathematical theory of evolution. **Philosophical Transactions**, A, v. 185, p. 71-110, 1894.
- [60] QUANDT, R. E. E RAMSEY, J. B. Estimating mixture of normal distributions and switching regression (with discussion). **Journal of the American Statistical Association**, v. 73, p.730-738, 1978.
- [61] RIDER, P. R. The method of moments applied to a mixture of two exponential distributions. **Annals of Mathematical Statistics**, v. 32, p.143-147, 1961.
- [62] ROY, D.; MURKHERJEE, S. P. Generalised mixtures of exponential distributions. **Journal of Applied Probability**, v. 25, p.510-518, 1988.
- [63] SAHU, S.K.; DEY, D. K.; BRANCO, M. A new class of multivariate skew distributions with applications to bayesian regression models. **Canadian Journal of Statistics**, v. 31, p. 129-150, 2003.

- [64] SALAS-GONZALEZ, D.; KURUOGLU, E. E.; RUIZ, D. P. Finite mixture of α -stable distributions. **Digital Signal Processing**, v. 19, p. 250-264, 2009.
- [65] SALAS-GONZALEZ, D.; KURUOGLU, E. E.; RUIZ, D. P. Modelling with mixture of symmetric stable distributions using Gibbs sampling. **Signal Processing**, v. 90, p. 774-783, 2010.
- [66] SHOUKRI, M. M.; MIAN, I. U. M.; TRACY, D. S. Sampling Properties of Estimators of the Log-Logistic Distribution with Application to Canadian Precipitation Data. **The Canadian Journal of Statistics**, v. 16, p. 223-236, 1988.
- [67] SULTAN, K. S. Identifiability of finite mixtures. **Annals of Mathematical Statistics**, v. 34, p.1265-1269, 1963.
- [68] TADIKAMALLA, P. R.; JOHNSON, N. L. Systems of Frequency Curves Generated by Transformations of Logistic Variables. **Biometrika**, v. 69, p.461-465, 1982.
- [69] TEICHER, H. Identifiability of finite mixtures. **The Annals of Mathematical Statistics**, v. 32, p.244-248, 1961.
- [70] TAN, F.; RAYNER, G. J.; WANG, X.; PENG, H. A full likelihood procedure of exchangeable negative binomials for modelling correlated and overdispersed count data. **Journal of Statistical Planning and Inference**, v. 140, p.2849-2959, 2010.
- [71] TEICHER, H. Identifiability of finite mixtures. **The Annals of Mathematical Statistics**, v. 34, p.1265-1269, 1963.

- [72] TITTERINGTON, D.M.; SMITH, A. F. M.; MAKOV U. E. **Statistical analysis of finite mixture distributions**. New York: Wiley, p. 246, 1985.
- [73] VENTURINI, S; DOMINICI, F.; PARMIGIANI, G. Gamma shape mixtures for heavy-tailed distributions. **The Annals of Applied Statistics**, v.2, p. 756-776, 2008.
- [74] VRBIK, I.; MCNICHOLAS M. R. Analytic calculations for the EM algorithm for multivariate skew-t mixture models. **Statistics and Probability Letters**, v. 82, p. 1169-1174, 2012.
- [75] WEBB, A. R. Gamma mixture models for target recognition. **Pattern Recognition**, v.33, p. 2045-2054, 2000.
- [76] WILKINS, C. A. A problem concerned with the weighting of distributions. **Journal of the American Statistical Association**, v.56, p. 281-292, 1948.
- [77] YAKOWITZ, S. J.; SPARGINS, J. D. On the identifiability of finite mixtures. **Journal Of Mathematical Statistics**, v.39, p. 209-214, 1968.

Apêndice A

Algebra linear

Para que um estudo eficiente de identificabilidade seja feito é necessário alguns conceitos introdutórios de algebra linear, que serão apresentados a seguir. Todas as definições do presente capítulo são apresentadas seguindo Lang (1983).

Definição 6.1 *Sejam U e V dois espaços vetoriais. Uma aplicação (ou transformação) $F : U \rightarrow V$ é linear se satisfaz as duas seguintes propriedades:*

(i) *para qualquer $u, v \in U$ temos:*

$$F(u + v) = F(u) + F(v)$$

(ii) *para todo $c \in R$ e $v \in U$ temos:*

$$F(cv) = cF(v).$$

Considere agora os espaços vetoriais U , V e W . Sejam ainda $F : U \rightarrow V$ e $G : V \rightarrow W$ duas aplicações lineares, é fácil mostrar que a composição $G \circ F$ é também uma aplicação linear. Sendo u e v elementos de U , temos:

$$(G \circ F)(u + v) = G(F(u + v)) = G(F(u) + F(v)) = G(F(u)) + G(F(v)).$$

Por outro lado, para $c \in R$, temos:

$$(G \circ F)(cu) = G(F(cu)) = G(cF(u)) = cG(F(u)).$$

Verificando portanto que valem as duas condições da definição 2.1.

Definição 6.2 *Seja $F : U \rightarrow V$ uma aplicação linear. A imagem de F é definida como o conjunto:*

$$Im(F) = \{t \in V \mid \text{existe } u \in U \text{ tal que } F(u) = t\}.$$

Segue por fim uma definição que diz respeito a isomorfismos.

Definição 6.3 *Uma aplicação linear $F : U \rightarrow V$ que possui uma inversa (ou inversível) $H : V \rightarrow U$ é chamada de isomorfismo.*

Apêndice B

Programação no R

```
# Parâmetros 1
m=c(-5,4)
s=c(2,2.75)
p=c(.5,.5)
# Parâmetros 2
m=c(2,1)
s=c(3,2)
p=c(.5,.5)
# Parâmetros 3
m=c(3,-4)
s=c(2,1)
p=c(.75,.25)
# Parâmetros 4
m=c(-3,3.5)
s=c(4,4)
p=c(.5,.5)
# Parâmetros 5
m=c(2,6)
s=c(3,5)
p=c(.5,.5)
# Parâmetros 6
m=c(2,3)
s=c(2,3)
p=c(.75,.25)
# Função acumulada Inversa
finv = function(u,m,s){
  m - (s*log((u)^-1 - 1))
}
```

```

# Mistura de 2 logísticas
mlog=function(x){
fl=(p[1]/s[1])*(exp(-(x-m[1])/s[1])/(1 + exp(-(x-m[1])/s[1]))^ 2)
f=fl + (p[2]/s[2])*(exp(-(x-m[2])/s[2])/(1 + exp(-(x-m[2])/s[2]))^ 2)
return(f)
}
# Expressão do Passo E
fpost = function(x,p,m,s,j){
fp1 = (p[j]/s[j])*exp(-(x-m[j])/s[j])/(1 + exp(-(x-m[j])/s[j]))^ 2
fp2 = mlog(x)
fp = fp1/fp2
return(fp) }
# Newton Raphson para média
NR.log<-function(erro,maxiter,x,p,m,s,j){
a = m[j]
errops = 1
i = 0
while(errops> erro && i<=maxiter){
dev1= - (2/s[j]) *(exp(-(x + m[j])/s[j]))/(exp(-x/s[j]) + exp(-a/s[j]))^ 2)
dev2=(fpost(x,p,m,s,j)/s[j])*dev1
dev = 1/sum(dev2)
score1= 1 - 2 *(exp(-x/s[j])/(exp(-x/s[j]) + exp(-a/s[j])))
score2=(fpost(x,p,m,s,j)/s[j])*score1
score=sum(score2)
a = a - score*dev
errops = abs(a-m[j])
i=i+1
}
return(a)
}
# Newton Raphson para variancia
NRS.log<-function(erro,maxiter,x,p,m,s,j){
a = s[j]

```

```

errobs = 1
i = 0
while(errobs > erro && i <= maxiter){
dev1 = ((2*(x-m[j]/a^3))*(-1 + ((exp(-(x)/a))/(exp(-(x)/a) + exp(-(m[j])/a))*
(2 - (x/a) + ((m[j]/a*exp(m[j]/a) + x/a*exp(x/a))/((exp(-(x)/a) + exp(-
(m[j])/a))))))) + 1/a^2
dev2 = (fpost(x,p,m,s,j)/a)*dev1
dev = 1/sum(dev2)
score1 = ((x-m[j]/a^2)*(1 - 2*(exp(-(x)/a))/(exp(-(x)/a) + exp(-(m[j])/a)))
- 1/a
score2 = (fpost(x,p,m,s,j)/a)*score1
score = sum(score2)
a = a - score*dev
errobs = abs(a-s[j])
i=i+1
}
return(a)
}
# Declarando as funções para as rotinas automaticas de otimização do R
library(maxLik)
mu1 = function(m){
sum(fpost(x,1)*(1 - 2*(exp(-x/s[j])/((exp(-x/s[j]) + exp(-m/s[j]))))),na.rm=T)
}
mu2 = function(m){
sum(fpost(x,2)*(1 - 2*(exp(-x/s[j])/((exp(-x/s[j]) + exp(-m/s[j]))))),na.rm=T)
}
sig1 = function(s){
sum(fpost(x,1)*((x-m[1]/s^2)*(1 - 2*(exp(-(x)/s))/(exp(-(x)/s) + exp(-(m[1])/s))))
- 1/s,na.rm=T)
}
sig2 = function(s){
sum(fpost(x,2)*((x-m[2]/s^2)*(1 - 2*(exp(-(x)/s))/(exp(-(x)/s) + exp(-(m[2])/s))))
- 1/s,na.rm=T)
}

```

```

}
# Seed e amostra
set.seed(9643)
n=25
# Simulação
# Gerando as amostras e estimando via newton raphson
n=50
#system.time(
for(ii in 1:100){
set.seed(-ii+123456)
u1=runif(n)
u2=runif(n)
x=numeric(n)
for(i in 1:n){
if(u1[i]< p[1]) x[i]= finv(u2[i],m[1],s[1])
if(u1[i]>= p[1]) x[i]= finv(u2[i],m[2],s[2])
}
erro = 0.001
for(j in 1:2){
pj[j] = mean(fpost(x,j))
mj[j] = NR.log(0.001,10,p,m,s,j)
sj[j] = NRS.log(0.001,10,p,mk,s,j)
}
#s1 = maxBFGSR(sig1,start=(1),tol=1e-5,iterlim = 10,print.level=0)
#s2 = maxBFGSR(sig2,start=(2.5),tol=1e-5,iterlim = 10,print.level=0)
#m1 = maxBFGSR(mu1,start=(-1),tol=1e-5,iterlim = 10,print.level=0)
#m2 = maxBFGSR(mu2,start=(2),tol=1e-5,iterlim = 10,print.level=0)
#mj[1]= m1$estimate
#mj[2]= m2$estimate
#sj[1]= s1$estimate
#sj[2]= s2$estimate
pis=rbind(pis,pj)
mis=rbind(mis,mj)

```

```

sis=rbind(sis,sj)
mjs=rbind(mjs,mi)
}
#)
warnings()
#Cálculo do EQM e viés
mis=mis[-1,]
pis=pis[-1,]
sis=sis[-1,]
summary(pis[,1])
summary(mis[,1])
summary(mis[,2])
summary(sis[,1])
summary(sis[,2])
m1=as.matrix(mis[,1])
m2=as.matrix(mis[,2])
s1=as.matrix(sis[,1])
s2=as.matrix(sis[,2])
vies_m1 = apply((m[1]-m1),2,mean,na.rm=T)
vies_m2 = apply((m[2]-m2),2,mean,na.rm=T)
eqm_m1 = apply((m[1]-m1)^2,2,mean,na.rm=T)
eqm_m2 = apply((m[2]-m2)^2,2,mean,na.rm=T)
eqm_s1 = apply((s[1]-s1)^2,2,mean,na.rm=T)
eqm_s2 = apply((s[2]-s2)^2,2,mean,na.rm=T)
eqm_m1
eqm_m2
eqm_s1
eqm_s2

```