

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS - ICEX
DEPARTAMENTO DE ESTATÍSTICA

Uma nova abordagem da distribuição beta logística

Sérgio Luiz de Oliveira

Orientadora: **Profa. Lourdes C. Contreras Montenegro**

Co-Orientador: **Prof. Fredy W. Castellares Cáceres**

Belo Horizonte, Maio de 2012.

Agradecimentos

Manifesto os meus agradecimentos à minha esposa Márcia pelo apoio e paciência em todos os momentos, sobretudo nas minhas decisões pessoais e profissionais.

À minha mãe Maria Aparecida pelo incentivo e ajuda em todos os momentos.

À Prof^ª.Dr^ª Lourdes Coral Montenegro pelo comprometimento, incentivo, paciência, por todos os ensinamentos passados e por ter me dado a oportunidade de trabalharmos juntos.

Ao professor Fredy Castellares Cáceres pelos conhecimentos transmitidos e as valiosas sugestões neste estudo.

Aos professores e funcionários do Departamento de Estatística da UFMG pela oportunidade e por terem contribuído na minha formação.

Aos meus colegas e amigos da Pós-Graduação em Estatística, em especial, Spencer, Angélica, Cleide, Ronaldo e Luís por toda força e apoio nos momentos difíceis.

E finalmente agradeço a Deus por mais esta conquista.

Sumário

1	Introdução	11
1.1	Preliminares	13
1.1.1	Distribuição logística	14
1.1.2	Classe de distribuições beta generalizada	15
2	Distribuição beta logística	17
2.1	Definição da distribuição beta logística	17
2.2	Características da distribuição beta logística	22
2.2.1	Função geradora de momentos	22
2.2.2	Momentos	23
2.2.3	Esperança	24
2.2.4	Moda	24
2.2.5	Mediana	26

<i>SUMÁRIO</i>	4
2.2.6 Estatísticas de ordem	26
2.2.7 Entropia de Shanon	28
2.2.8 Entropia de Rényi	29
2.2.9 Entropia de Kullback-Leibler	29
2.2.10 Momentos-L	31
2.2.11 Desvio médio em relação à média	31
2.2.12 Desvio médio em relação à mediana	33
2.2.13 Curva de Lorenz	33
2.2.14 Curva de Bonferroni	36
2.3 Caracterização da beta logística	36
2.3.1 Características da beta logística	36
2.3.2 Relação e convergência para outras distribuições	37
3 Inferência	39
3.1 Inferência	39
4 Simulação e aplicação	44
4.1 Estudo de simulação	44
4.2 Dados de células cancerígenas	46
4.3 Dados de resistência de fibra de vidro	50

Resumo

Neste trabalho é proposta uma nova abordagem da distribuição beta logística pertencente à classe de distribuições beta generalizada. Esta distribuição já se encontra na literatura e é denominada de distribuição logística generalizada do tipo *IV*. Para essa nova abordagem, foi realizado o estudo da função densidade de probabilidade, função de distribuição acumulada e função de risco. Apresentaremos expressões para as principais propriedades desta distribuição, tais como, momentos, função geradora de momentos, medidas de tendência central e função densidade de probabilidade das estatísticas de ordem, entre outros. O estimador de máxima verossimilhança é proposto para estimar os parâmetros desta distribuição e a matriz de informação esperada é calculada. Para ilustrar a aplicação da distribuição beta logística utilizamos dois conjuntos de dados reais, mostrando que ela é mais flexível que outras distribuições existentes na literatura.

Palavras – chaves : Distribuição logística; Distribuição beta logística; Estimação de máxima verossimilhança.

Abstract

This work proposes a new approach to the beta logistic distribution belonging to the class of generalized beta distributions. This distribution is known in literature and is called the generalized logistic distribution of type IV. For this new approach, the study was conducted of the probability density function, cumulative distribution function and hazard function. We present expressions for the moments, moment generating function, measures of central tendency, probability density function of order statistics, average deviations, entropies, and Lorenz and Bonferroni curves. The estimator of maximum likelihood is proposed to estimate the parameters this distribution and the expected information matrix is calculated. To illustrate the application of the beta logistic distribution we used two real data sets showing that it is more flexible than other distributions in the literature.

Lista de Figuras

1.1	Gráficos da f.d.p. (1.1) e f.d.a. (1.3) da distribuição Lo para alguns valores dos parâmetros μ e σ	15
2.1	Gráficos da função de densidade de probabilidade da distribuição BLo para alguns valores dos parâmetros μ, σ, a e b	19
2.2	Gráficos da função de distribuição acumulada da distribuição BLo para alguns valores dos parâmetros μ, σ, a e b	20
2.3	Gráficos da função de risco da distribuição BLo para alguns valores dos parâmetros μ, σ, a e b	21
4.1	Gráfico da função densidade da distribuição BLo	45
4.2	Função de densidade estimada para dados de célula de câncer	47
4.3	Gráficos QQ das distribuições (a)BLo, (b)Lo, (c)BN e (d)BGL	48
4.4	Gráfico da função empírica e das funções de distribuição acumulada das distribuições BLo, Lo, BN e BGL.	50
4.5	Função de densidade estimada para dados fibra de vidro	52

4.6	Gráfico QQ das distribuições (a)BLo, (b)Lo, (c)BN, (d)BW e (e)BGL	55
4.7	Gráfico função empírica e função de distribuição acumulada das distribuições BLo, Lo, BN, BW e BGL	56

Lista de Tabelas

2.1	Valores esperados e moda para alguns valores dos parâmetros da distribuição BLo.	25
4.1	EMVs (erros padrão) dos parâmetros das distribuições BLo, Lo, BN e BGL para dados célula cancerígenas	46
4.2	Estatísticas AIC, BIC, CAIC e HQC das distribuições BLo, Lo, BN e BGL para dados célula de câncer	47
4.3	EMVs (erros padrão) dos parâmetros das distribuições BLo, Lo, BN, BW e BGL para dados de fibra de vidro.	51
4.4	Estatísticas AIC, BIC, CAIC e HQC das distribuições BLo, Lo, BN, BW e BGL para dados fibra de vidro	52

Capítulo 1

Introdução

O modelo logístico surgiu inicialmente para modelar estudos demográficos por Verhulst. Reed e Berkson chegaram a chamá-la de modelo de Verhulst, uma homenagem ao seu precursor. Outros autores aplicaram a distribuição logística para estimar o crescimento da população humana, por exemplo, Schultz (1930).

Generalizações da distribuição logística surgiram inicialmente com Perks (1932), que propôs uma distribuição que inclui todas as generalizações da distribuição logística conhecidas como Tipos *I*, *II*, *III* e *IV*. A distribuição logística generalizada tipo *I* tem recebido considerável atenção com respeito às caracterizações, tais como os momentos e estatísticas de ordem. A distribuição do tipo *II* inclui várias propriedades da distribuição do tipo *I*; por exemplo, se X é uma variável aleatória (v.a.) com distribuição logística generalizada do tipo *I*, então, $-X$ é uma v.a. que segue a distribuição generalizada logística do tipo *II*. A distribuição logística generalizada do tipo *III* é simétrica e extremamente útil como aproximação a outras distribuições simétricas, como por exemplo, a distribuição t de Student. A distribuição logística generalizada do tipo *IV* é a mais geral, pois inclui as distribuições do tipo *I* e *III*

como casos especiais.

Prentice (1975) propôs uma nova família de distribuições, que inclui algumas distribuições importantes como casos especiais, tais como as distribuições, normal, logística, valor extremo e logística generalizada. Distribuições de probabilidade com suporte nos reais positivos incluem as distribuições exponencial, log-normal, Weibull, gama, gama generalizada, log-logístico, qui-quadrado, t e F .

McDonald e Xu (1995) apresentam uma classe de distribuições beta com cinco parâmetros que inclui as distribuições beta generalizada e gama generalizada. Estas distribuições incluem mais de trinta distribuições como casos especiais. A distribuição beta generalizada gera a distribuição beta generalizada exponencial, o qual inclui formas generalizadas das distribuições logística, exponencial, Gompertz e Gumbel e a distribuição normal como caso especial.

Eugene et al. (2002) desenvolveram também um método de generalizar distribuições, introduzindo a classe beta generalizada (beta-G) de distribuições. Eles apresentaram a distribuição beta normal a partir da aplicação entre as distribuições beta e normal. Uma vantagem da distribuição beta normal em relação à distribuição normal, é que ela pode assumir a forma unimodal ou bimodal. Esta classe de distribuições generaliza uma distribuição padrão adicionando dois parâmetros de forma. Por isso, a forma generalizada proporciona maior flexibilidade na análise dos dados observados. Vários trabalhos surgiram na literatura nesta mesma linha de pesquisa. Estes trabalhos introduzem uma nova distribuição da classe beta-G, deduzindo algumas propriedades que caracterizam a distribuição e propõem uma aplicação para o modelo. Neste sentido, citamos algumas distribuições, tais como, beta Gumbel (Nadarajah e Kotz, 2004), beta exponencial (Nadarajah e Kotz, 2005), beta secante hiperbólico (Fischer e Vaughan, 2007), beta power (Cordeiro e Brito, 2010), beta

Weibull (Cordeiro et al, 2010) e beta logística generalizada (Morais et al., 2010), entre outros.

Neste contexto, o objetivo deste trabalho é propor uma nova abordagem da distribuição beta logística, pertencente à classe de distribuições beta generalizada. Determinamos expressões matemáticas para as propriedades que caracterizam a distribuição. Estudamos a teoria de inferência para a distribuição. Ainda, determinamos expressões para os estimadores de máxima verossimilhança dos parâmetros da distribuição e para verificar a flexibilidade da distribuição beta logística, desenvolvemos uma aplicação com dois conjuntos de dados reais, comparando com outras distribuições.

Esta dissertação está dividida em cinco capítulos e contém dois anexos. No Capítulo 2 apresentamos a distribuição beta logística explorando suas propriedades. No Capítulo 3 desenvolvemos a teoria de inferência da distribuição BLo. No Capítulo 4 um estudo de simulação é realizado e apresentamos duas aplicações da distribuição beta logística. Finalmente, considerações finais são apresentadas no Capítulo 5.

1.1 Preliminares

O objetivo desta seção é apresentar alguns conceitos para o desenvolvimento deste trabalho, assim como estabelecer a notação e terminologias pertinentes.

1.1.1 Distribuição logística

A função de densidade de probabilidade (f.d.p.) da distribuição logística de parâmetros μ e σ , denotada por $\text{Lo}(\mu, \sigma)$, é dada por

$$f(x) = \frac{e^{-\frac{(x-\mu)}{s}}}{s[1 + e^{-\frac{(x-\mu)}{s}}]^2}, \quad x, \mu \in \mathfrak{R}, \quad s > 0, \quad (1.1)$$

em que $s = \frac{\sigma\sqrt{3}}{\pi}$ e $\sigma > 0$.

Esta função de densidade de probabilidade da distribuição Lo é considerada mais geral, porque ela inclui a distribuição logística padrão $\text{Lo}(0,1)$. O termo s é utilizado apenas para simplificar as notações da distribuição Lo.

Uma outra forma alternativa para a f.d.p. pode ser escrita como

$$f(x) = \frac{1}{4s} \text{sech}^2 \left(\frac{x - \mu}{2s} \right). \quad (1.2)$$

Por consequência da equação (1.2), a função de distribuição acumulada (f.d.a.) da distribuição logística é denotada por

$$F(x) = \frac{1}{1 + e^{-\frac{(x-\mu)}{s}}}, \quad (1.3)$$

aqui, também podemos estabelecer de uma outra forma a f.d.a., dada por

$$F(x) = \frac{1}{2} \left[1 + \tanh \left(\frac{x - \mu}{2s} \right) \right].$$

Para as equações (1.2) e (1.3), a função de risco da distribuição Lo pode ser escrita como

$$h(x) = \frac{1}{s \left[1 + e^{-\frac{(x-\mu)}{s}} \right]}.$$

A Figura 1.1 apresenta os gráficos da função densidade de probabilidade e função de distribuição acumulada da distribuição Lo para alguns valores dos parâmetros, μ e σ .

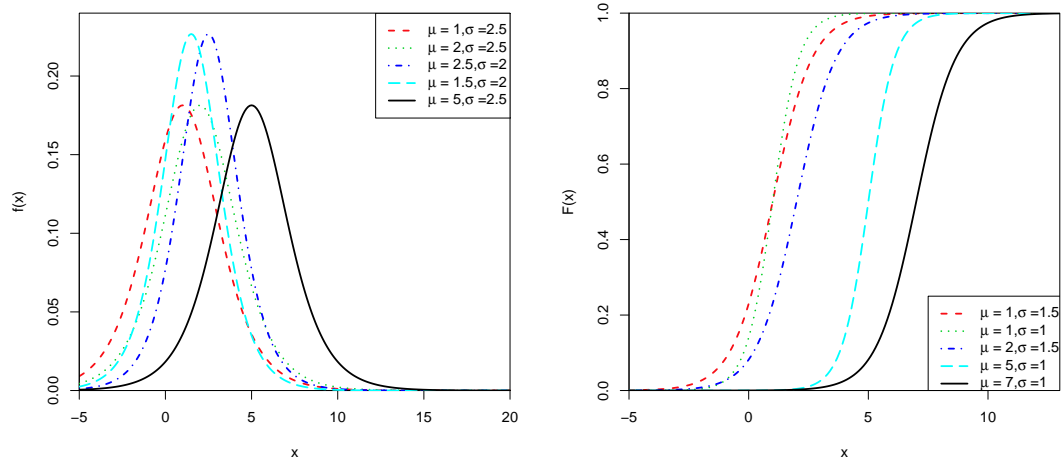


Figura 1.1: Gráficos da f.d.p. (1.1) e f.d.a. (1.3) da distribuição Lo para alguns valores dos parâmetros μ e σ

As principais propriedades que caracterizam a distribuição logística são apresentadas no Apêndice A.

1.1.2 Classe de distribuições beta generalizada

A classe de distribuições beta generalizada é aplicada em análise de dados com distribuição assimétrica que não podem ser bem ajustados pelas distribuições existentes na literatura (Eugene et al, 2002). Assim, esta classe de distribuições pode ser definida da seguinte forma:

Sejam Z e Y duas variáveis aleatórias (v.a.), em que $Z \sim Beta(a, b)$ e Y com função de distribuição acumulada $G(y)$, chamada de distribuição primitiva. A aplicação $X = G_Y^{-1}(z)$ produz X seguindo a função de densidade de probabilidade da distribuição beta generalizada com expressão

$$f(x; \alpha, \beta, \gamma) = \frac{g(y)}{B(\alpha, \beta)} [G(y)]^{\alpha-1} [1 - G(y)]^{\beta-1}, \quad (1.4)$$

em que $B(\alpha, \beta)$ é a função beta, $g(y)$ é a f.d.p. de $G(y)$ e γ são os parâmetros da distribuição primitiva da v.a. Y

A f.d.a. da distribuição beta generalizada pode ser escrita como

$$F(x; \alpha, \beta, \gamma) = \frac{B_{G(y)}(\alpha, \beta)}{B(\alpha, \beta)}, \quad (1.5)$$

em que $B_{G(y)}(\alpha, \beta)$ é denominada função beta incompleta e é dada por

$$B_{G(y)}(\alpha, \beta) = \int_0^{G(y)} t^{\alpha-1} (1-t)^{\beta-1} dt.$$

Quando $\alpha = \beta = 1$ a função de distribuição acumulada da distribuição beta generalizada coincide com a f.d.a. $G(y)$ da distribuição primitiva.

A adição dos parâmetros de forma, α e β , proporciona maior flexibilidade à distribuição, pois determinam assimetria e variabilidade ao peso das caudas.

Capítulo 2

Distribuição beta logística

Neste capítulo é proposta uma distribuição com quatro parâmetros, denotada por distribuição beta logística (BLo). Apresentamos a função densidade de probabilidade (f.d.p.), função de distribuição acumulada (f.d.a.), função de risco, função geradora de momentos (f.g.m.), momentos, média, moda, mediana, estatística de ordem, entropia, momentos-L, desvios médios em relação à média e à mediana e as curvas de Lorenz e Bonferroni.

2.1 Definição da distribuição beta logística

Sejam Z e Y duas variáveis aleatórias (v.a.), $Z \sim Beta(a, b)$ e $Y \sim Lo(\mu, \sigma)$. Como forma de combinar as duas variáveis, definimos X como

$$X = s \log \left[\frac{z}{1-z} \right] + \mu.$$

em que $s = \frac{\sigma\sqrt{3}}{\pi}$. Desta forma, a variável aleatória X segue uma distribuição beta logística, denotada por, $X \sim BLo(\mu, \sigma, a, b)$.

Utilizando a equação (1.4), a função densidade de probabilidade da distribuição beta logística pode ser escrita como

$$f(x) = \frac{[e^{-\frac{(x-\mu)}{s}}]^b}{B(a, b)s[1 + e^{-\frac{(x-\mu)}{s}}]^{a+b}}. \quad (2.1)$$

A função de distribuição acumulada da distribuição BLo é obtida a partir da equação (1.5), isto é,

$$F_X(x) = \sum_{n=0}^{b-1} d_n \left[\frac{1}{1 + e^{-\frac{(x-\mu)}{s}}} \right]^{a+n}, \quad (2.2)$$

em que $d_n = \binom{b-1}{n} \frac{(-1)^n}{B(a, b)(a+n)}$, $n \geq 0$.

Das equações (2.1) e (2.2), a função de risco da distribuição BLo pode ser escrita como

$$h(x) = \frac{[e^{-\frac{(x-\mu)}{s}}]^b}{s \left[1 + e^{-\frac{(x-\mu)}{s}} \right]^{a+b} \left[B(a, b) - \int_0^{\frac{1}{1+e^{-\frac{(x-\mu)}{s}}}} t^{a-1}(1-t)^{b-1} dt \right]}.$$

As Figuras 2.1, 2.2 e 2.3 apresentam os gráficos das funções densidade de probabilidade, função de distribuição acumulada e da função de risco para alguns valores dos parâmetros μ , σ , a e b .

Distribuições com maiores valores dos parâmetros a e b apresentam maior curtose. Se $a > b$ a curva da distribuição BLo é assimétrica à esquerda e se $a < b$ a curva é assimétrica à direita.

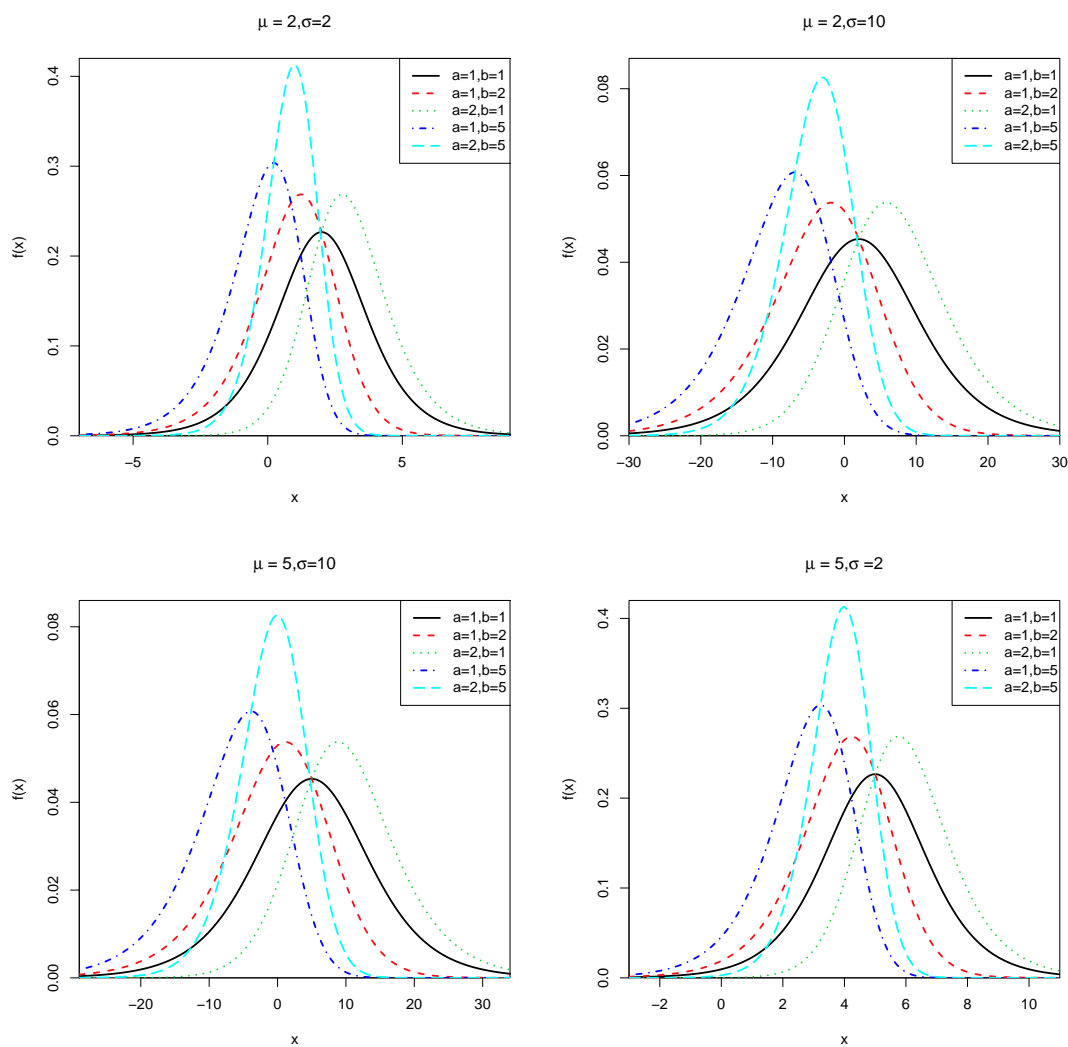


Figura 2.1: Gráficos da função de densidade de probabilidade da distribuição BLo para alguns valores dos parâmetros μ , σ , a e b

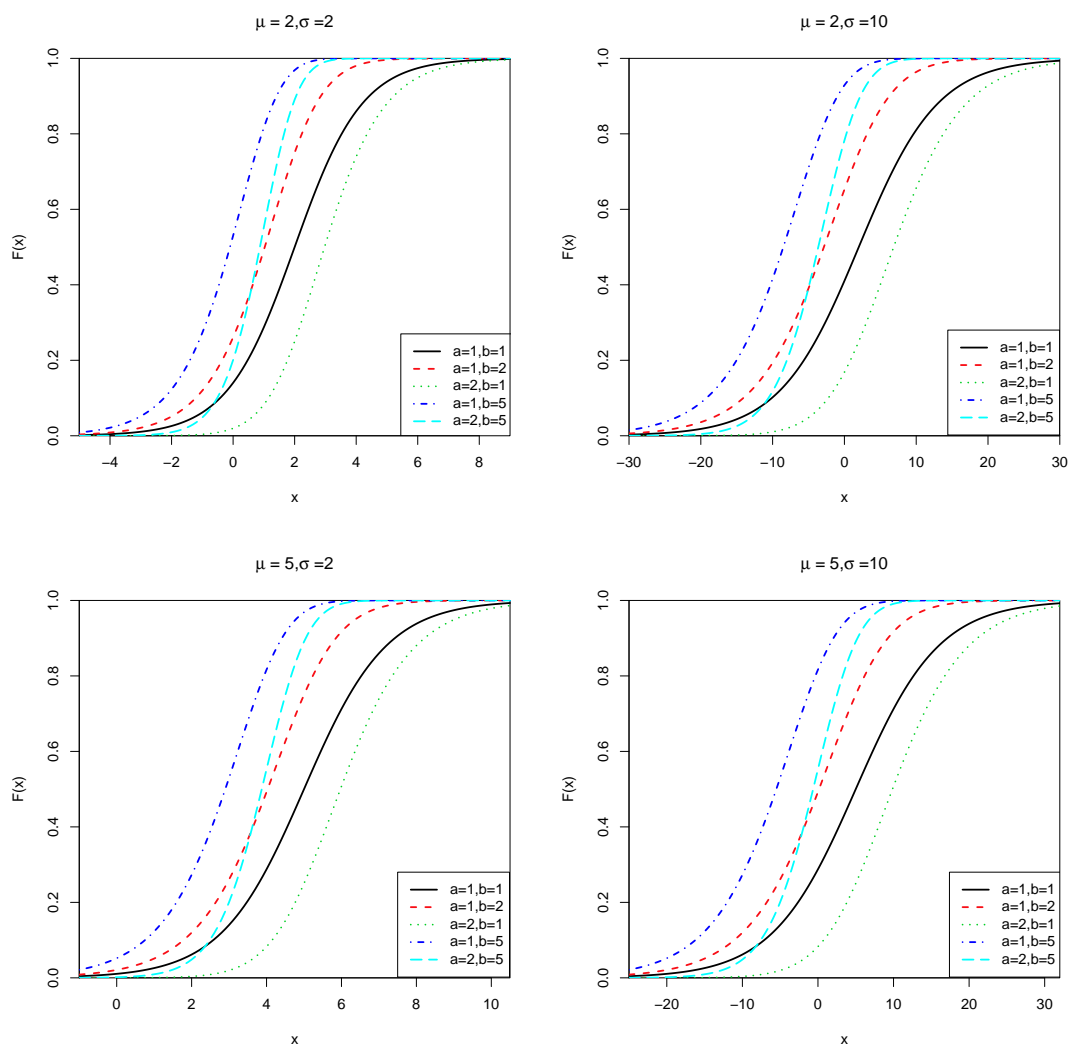


Figura 2.2: Gráficos da função de distribuição acumulada da distribuição BLo para alguns valores dos parâmetros μ , σ , a e b

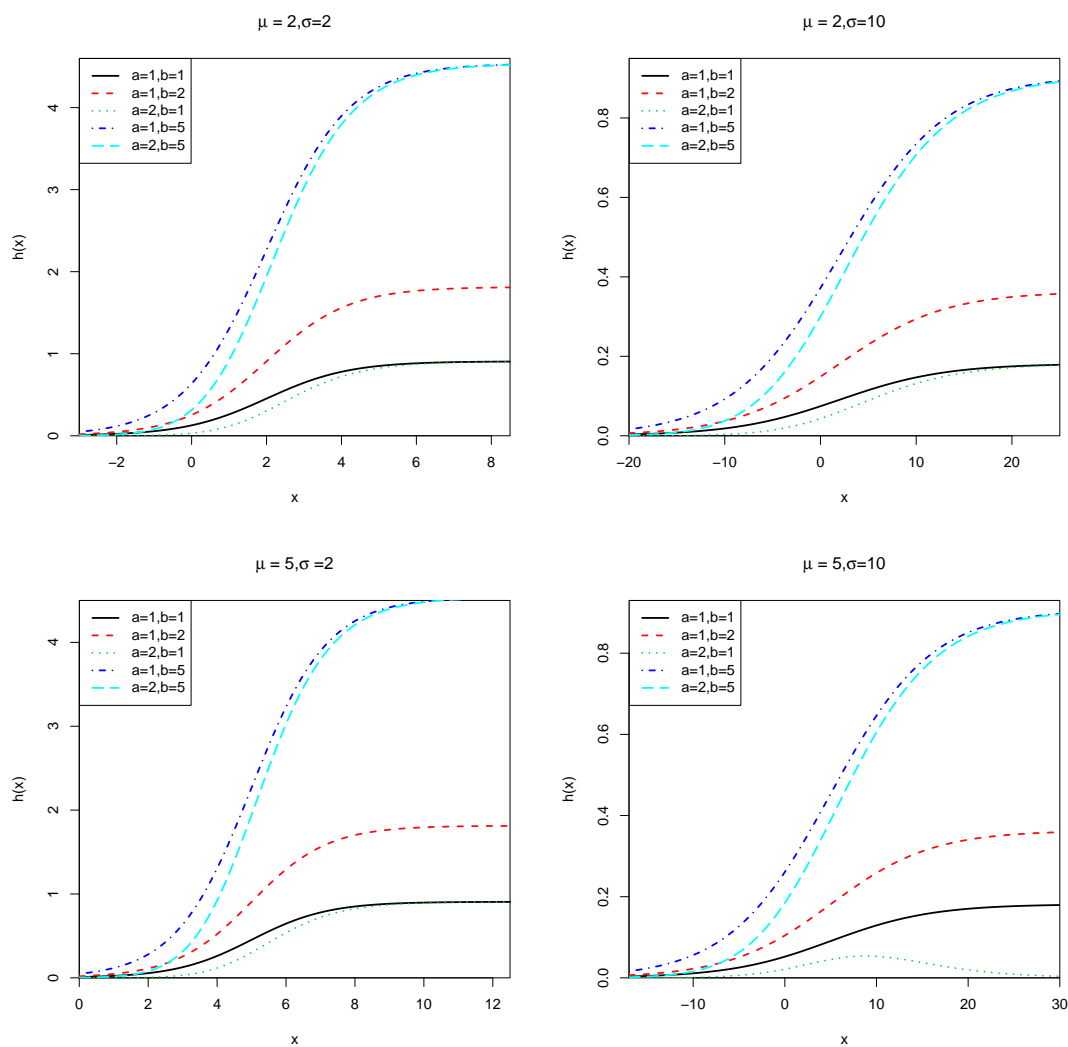


Figura 2.3: Gráficos da função de risco da distribuição BLo para alguns valores dos parâmetros μ , σ , a e b

2.2 Características da distribuição beta logística

Nesta seção desenvolvemos expressões formais para função geradora de momentos, momentos, esperança, média, moda, mediana, estatística de ordem, entropias de Renyi, Shanon e Kullback, momentos L, foram também desenvolvidos desvios médios em relação a média e a mediana. Finalmente, as curvas de Lorenz e Bonferroni.

2.2.1 Função geradora de momentos

A função de geração de momentos é utilizada para determinar os momentos de uma distribuição de probabilidade.

Utilizando a função de densidade de probabilidade da distribuição BLo, equação (2.1), obtemos a função geradora de momentos (f.g.m.), dada por

$$M(t) = \int_{-\infty}^{\infty} e^{tx} \frac{[e^{-\frac{(x-\mu)}{s}}]^b}{B(a, b)s(1 + e^{-\frac{(x-\mu)}{s}})^{a+b}} dx.$$

Substituindo $v = \frac{x-\mu}{s}$, temos

$$M(t) = \frac{e^{t\mu}}{B(a, b)} \int_{-\infty}^{\infty} e^{-(1-ts)v} \left(\frac{1}{1 + e^{-v}} \right)^{a+b} dv,$$

fazendo $m = \frac{1}{1+e^{-v}}$, obtemos

$$\begin{aligned} M(t) &= \frac{e^{t\mu}}{B(a, b)} \int_0^1 m^{(ts+a-1)} (1-m)^{(b-ts-1)} dm \\ &= \frac{e^{t\mu}}{B(a, b)} B(ts+a, b-ts) \\ &= \frac{e^{t\mu}}{B(a, b)} \frac{\Gamma(ts+a)\Gamma(b-ts)}{\Gamma(a+b)}, \end{aligned} \tag{2.3}$$

em que $ts < b$.

2.2.2 Momentos

Os momentos de uma distribuição de probabilidade são importantes para determinar algumas características fundamentais, como por exemplo, esperança, variância, assimetria e curtose.

A partir da definição do n -ésimo momento da distribuição beta-G, podemos facilmente obter o n -ésimo momento da distribuição BLo após uma mudança de variáveis.

O n -ésimo momento da distribuição beta-G pode ser escrito como

$$\tau_n = \int_{-\infty}^{\infty} x^n \frac{g(x)}{B(a, b)} G(x)^{a-1} [1 - G(x)]^{b-1} dx. \quad (2.4)$$

Fazendo $G(x) = u$, obtemos $x = G^{-1}(u)$, isto é,

$$x = s \log \left(\frac{u}{1-u} \right) + \mu = G^{-1}(u).$$

Aplicando esta relação na equação (2.4), temos

$$\begin{aligned} \tau_n &= \int_0^1 [G^{-1}(u)]^n \frac{1}{B(a, b)} u^{a-1} (1-u)^{b-1} du \\ &= \int_0^1 \left[s \log \left(\frac{u}{1-u} \right) + \mu \right]^n \frac{1}{B(a, b)} u^{a-1} (1-u)^{b-1} du \\ &= \sum_{k=0}^n \binom{n}{k} s^k \mu^{n-k} \int_0^1 [(\log u - \log(1-u))]^k \frac{1}{B(a, b)} u^{a-1} (1-u)^{b-1} du \\ &= \sum_{k=0}^n \sum_{v=0}^k \binom{n}{k} \binom{k}{v} s^k \mu^{n-k} (-1)^v \int_0^1 (\log u)^{k-v} [\log(1-u)]^v \frac{1}{B(a, b)} u^{a-1} (1-u)^{b-1} du. \end{aligned} \quad (2.5)$$

Usando a relação

$$\int_0^1 (\log x)^{(k-v)} [\log(1-x)]^v x^{\beta-1} (1-x)^{\alpha-1} dx = \frac{\partial^k}{\partial \beta^{k-v} \partial \alpha^v} B(\beta, \alpha),$$

em (2.5), o n -ésimo momento da distribuição BLo pode ser escrito como

$$\tau = \sum_{k=0}^n \sum_{v=0}^k \binom{n}{k} \binom{k}{v} s^k \mu^{n-k} \frac{(-1)^v}{B(a, b)} \frac{\partial^k B(a, b)}{\partial a^{k-v} \partial b^v}.$$

2.2.3 Esperança

A esperança da distribuição BLo pode ser facilmente encontrada a partir da função geradora de momentos, calculando o primeiro momento.

Derivando a equação (2.3) em relação a t , temos

$$\frac{dM(t)}{dt} = \mu + s \frac{\Gamma'(ts + a)}{\Gamma(ts + a)} - s \frac{\Gamma'(b - ts)}{\Gamma(b - ts)}, \quad (2.6)$$

e tomando $t = 0$ na equação (2.6), obtemos a esperança da distribuição BLo

$$E(X) = \mu + s[\Psi(a) - \Psi(b)],$$

em que $\Psi(\cdot)$ denota a função digama.

2.2.4 Moda

Uma medida de tendência central, chamada moda, é o valor ou valores mais frequentes em uma distribuição de frequência. A moda é obtida pela maximização do logaritmo da f.d.p. da equação (2.1), isto é,

$$\frac{\partial \log f(x)}{\partial x} = -\frac{b}{s} - (a + b) \frac{1}{1 + e^{-\frac{(x-\mu)}{s}}} \left[e^{-\frac{(x-\mu)}{s}} \left(-\frac{1}{s} \right) \right].$$

Derivando a expressão acima e igualando a 0, temos

$$-\frac{b}{s} + \frac{(a + b)}{s} \left[\frac{e^{-\frac{(x-\mu)}{s}}}{1 + e^{-\frac{(x-\mu)}{s}}} \right] = 0.$$

Assim, desenvolvendo a equação anterior, obtemos

$$e^{-\frac{(x-\mu)}{s}} = \frac{b}{a},$$

logo,

$$\tilde{X}_{mo} = \mu + s(\log a - \log b).$$

em que \tilde{X}_{mo} é a moda da distribuição BLo.

Na Tabela 2.1 apresentamos valores para a esperança e moda da distribuição beta logística para alguns valores dos parâmetros μ , σ , a e b .

Tabela 2.1: Valores esperados e moda para alguns valores dos parâmetros da distribuição BLo.

Parâmetros				E(X)	Mo
μ	σ	a	b		
2	2	1	1	2	2
1	2	3	1	2,654	3,197
2	1	3	1	2,827	3,100
0,1	10	1	4	-7,543	-10,008
10	0,4	1	4	9,694	9,600
1	2	10	0,4	1,710	2,061
3	1	0,1	40	1,679	-0,110
3	1	0,2	0,5	2,800	2,267
3	1	40	25	3,104	3,105

2.2.5 Mediana

A mediana de uma distribuição contínua é definida como

$$P(X \leq v) = P(X \geq v) = \int_{-\infty}^v f(x)dx = \frac{1}{2}.$$

Podemos mostrar que a distribuição BLo não têm forma fechada para a mediana, isto é, a equação

$$\int_{-\infty}^v \frac{[e^{-\frac{(x-\mu)}{s}}]^b}{B(a, b)s[1 + e^{-\frac{(x-\mu)}{s}}]^{a+b}} dx = \frac{1}{2}$$

não pode ser resolvida analiticamente.

Fazendo a mudança de variáveis $t = \frac{x-\mu}{s}$, obtemos

$$\int_{-\infty}^{\frac{v-\mu}{s}} \frac{e^{-bt}}{B(a, b)(1 + e^{-t})^{a+b}} dt = \frac{1}{2}.$$

Substituindo $m = \frac{1}{1+e^{-t}}$, temos

$$\int_0^{\frac{1}{1+e^{-\frac{(v-\mu)}{s}}}} m^{a-1}(1-m)^{b-1} dm = \frac{1}{2}.$$

A solução desta equação não linear resulta no valor da mediana da distribuição beta logística.

2.2.6 Estatísticas de ordem

Estatísticas de ordem são consideradas de fundamental importância na estatística. Importantes casos especiais das estatísticas de ordem são consideradas o mínimo e o máximo valor de uma amostra.

Considerando uma amostra aleatória de tamanho n de uma distribuição de probabilidade contínua, a expressão da função densidade da k -ésima estatística de ordem $X_{k:n}$ é dada por

$$f_{k:n}(x) = \frac{f(x)}{B(k, n-k+1)} F(x)^{k-1} (1-F(x))^{n-k},$$

para $k = 1, \dots, n$. Expandindo em binômio de Newton, temos

$$f_{k:n}(x) = \frac{f(x)}{B(k, n-k+1)} \sum_{m=0}^{n-k} \binom{n-k}{m} (-1)^m F(x)^{m+k-1}. \quad (2.7)$$

Substituindo a equação (2.2) em (2.7), obtemos

$$f_{k:n}(x) = \frac{f(x)}{B(k, n-k+1)} \sum_{m=0}^{n-k} \binom{n-k}{m} (-1)^m \left[\sum_{t=0}^{b-1} d_t G(x)^{t+a} \right]^{m+k-1}. \quad (2.8)$$

Utilizando a identidade $(\sum_{k=0}^{\infty} d_k x^k)^n = \sum_{k=0}^{\infty} c_{k,n} x^k$ (Gradshteyn e Ryzhik, 2000) onde

$$c_{0,n} = d_0^n \text{ e } c_{k,n} = (kd_0)^{-1} \sum_{l=0}^k (nl - k + l) d_l c_{k-l,n},$$

na equação (2.8), temos

$$f_{k:n}(x) = \frac{f(x)}{B(k, n-k+1)} \sum_{m=0}^{n-k} \binom{n-k}{m} (-1)^m \sum_{t=0}^{\infty} c_{t+a, m+k-1} G(x)^{t+a}, \quad (2.9)$$

em que

$$c_{t+a, m+k-1} = [d_0(t+a)]^{-1} \sum_{l=0}^{t+a} [(m+k-1)l - (t+a) + l] d_l c_{t+a-l, m+k-1}.$$

Substituindo a f.d.p. e a f.d.a. da distribuição Lo dadas pelas equações (1.1) e (1.3), respectivamente, em (2.9) temos

$$\begin{aligned} f_{k:n}(x) &= \frac{1}{B(k, n-k+1)} \sum_{m=0}^{n-k} \binom{n-k}{m} \frac{(-1)^m [e^{-\frac{(x-\mu)}{s}}]^b}{B(a, b)s \left[1 + e^{-\frac{(x-\mu)}{s}}\right]^{a+b}} \\ &\times \sum_{t=0}^{\infty} c_{t+a, m+k-1} \left[\frac{1}{1 + e^{-\frac{(x-\mu)}{s}}} \right]^{t+a}. \end{aligned}$$

em que $f_{k:n}(x)$ é a função densidade da estatística de ordem da distribuição BLo.

2.2.7 Entropia de Shanon

O conceito de entropia de Shannon (Shannon, 1948) vem do campo da Teoria da Informação. Ela mede o grau de incerteza que existe em um sistema termodinâmico em evolução.

A entropia de Shanon é definida como

$$h[f] = \int_{-\infty}^{\infty} f(x) \log f(x) dx. \quad (2.10)$$

Substituindo a f.d.p. da BLo, equação (2.1), em (2.10) obtemos

$$h[f] = \int_{-\infty}^{\infty} \frac{[e^{-\frac{(x-\mu)}{s}}]^b}{B(a, b)s[1 + e^{-\frac{(x-\mu)}{s}}]^{a+b}} \log \left\{ \frac{[e^{-\frac{(x-\mu)}{s}}]^b}{B(a, b)s[1 + e^{-\frac{(x-\mu)}{s}}]^{a+b}} \right\} dx.$$

Fazendo $t = \frac{x-\mu}{s}$, temos

$$\begin{aligned} h[f] &= \int_{-\infty}^{\infty} \frac{(e^{-t})^b}{B(a, b)(1 + e^{-t})^{a+b}} \log \left[\frac{(e^{-t})^b}{B(a, b)s(1 + e^{-t})^{a+b}} \right] dt \\ &= \int_{-\infty}^{\infty} -bt \left[\frac{e^{-bt}}{B(a, b)(1 + e^{-t})^{a+b}} \right] dt - \int_{-\infty}^{\infty} \log B(a, b) \left[\frac{e^{-bt}}{B(a, b)(1 + e^{-t})^{a+b}} \right] dt \\ &\quad - \int_{-\infty}^{\infty} \log \left[s \frac{e^{-bt}}{B(a, b)(1 + e^{-t})^{a+b}} \right] dt \\ &\quad - (a + b) \int_{-\infty}^{\infty} \log \left[(1 + e^{-t}) \frac{e^{-bt}}{B(a, b)(1 + e^{-t})^{a+b}} \right] dt \\ &= -bE_{BLo}(T) - \log B(a, b) - \log(s) - (a + b)E_{BLo}[\log(1 + e^{-T})] \\ &= -\log[sB(a, b)] - bs[\Psi(a) - \Psi(b)] + (a + b)[\Psi(a + b) - \Psi(a)], \end{aligned}$$

em que E_{BLo} é a esperança da distribuição BLo.

2.2.8 Entropia de Rényi

A entropia de Rényi (Rényi, 1961) é uma generalização da entropia Shannon.

A entropia de Rényi de ordem γ de uma distribuição contínua é definida como

$$r[f] = \frac{1}{1-\gamma} \log \left[\int_{-\infty}^{\infty} f^\gamma(x) dx \right], \quad (2.11)$$

em que $\gamma > 0$ e $\gamma \neq 1$.

Substituindo a equação (2.1) na equação (2.11), obtemos que

$$r[f] = \frac{1}{1-\gamma} \log \int_{-\infty}^{\infty} \left\{ \frac{\left[e^{-\frac{(x-\mu)}{s}} \right]^b}{B(a,b)s \left[1 + e^{-\frac{(x-\mu)}{s}} \right]^{a+b}} \right\}^\gamma dx.$$

Fazendo $t = \frac{x-\mu}{s}$, temos

$$r[f] = \frac{1}{1-\gamma} \log \left[\frac{1}{B^\gamma(a,b)s^{\gamma-1}} \int_{-\infty}^{\infty} \frac{e^{-bt\gamma}}{(1+e^{-t})^{\gamma(a+b)}} dt \right].$$

Substituindo $m = \frac{1}{1+e^{-t}}$, obtemos

$$\begin{aligned} r[f] &= \frac{1}{1-\gamma} \log \left[\frac{1}{B^\gamma(a,b)s^{\gamma-1}} \int_0^1 m^{a\gamma-1} (1-m)^{b\gamma-1} dm \right] \\ &= \frac{1}{1-\gamma} \log \left[\frac{1}{B^\gamma(a,b)s^{\gamma-1}} B(a\gamma, b\gamma) \right] \\ &= \frac{1}{1-\gamma} [\log B(a\gamma, b\gamma) - \gamma \log B(a, b) - (\gamma-1) \log s]. \end{aligned}$$

2.2.9 Entropia de Kullback-Leibler

Na teoria de probabilidade a entropia de Kullback-Leibler é uma medida da diferença das funções densidade f e g definida por

$$I[F, G] = \int_{-\infty}^{\infty} \log \frac{f(x)}{g(x)} f(x) dx, \quad (2.12)$$

em que $I[F, G] \geq 0$.

Considerando $f(x)$ a f.d.p. da distribuição BLo e $g(x)$ a f.d.p. da distribuição Lo, substituindo as equações (2.1) e (1.1) em (2.12), obtemos

$$\begin{aligned}
I[F, G] &= \int_{-\infty}^{\infty} \log \left\{ \frac{[e^{-\frac{(x-\mu)}{s}}]^b}{B(a, b)s[1 + e^{-\frac{(x-\mu)}{s}}]^{a+b}} \left[\frac{s(1 + e^{-\frac{(x-\mu)}{s}})^2}{e^{-\frac{(x-\mu)}{s}}} \right] \right\} \\
&\quad \times \frac{[e^{-\frac{(x-\mu)}{s}}]^b}{B(a, b)s[1 + e^{-\frac{(x-\mu)}{s}}]^{a+b}} dx \\
&= \int_{-\infty}^{\infty} \log \left\{ \frac{[e^{-\frac{(x-\mu)}{s}}]^{b-1}}{B(a, b)[1 + e^{-\frac{(x-\mu)}{s}}]^{a+b-2}} \right\} \frac{[e^{-\frac{(x-\mu)}{s}}]^b}{B(a, b)s[1 + e^{-\frac{(x-\mu)}{s}}]^{a+b}} dx \\
&= \int_{-\infty}^{\infty} \frac{(b-1)(x-\mu)}{s} \left[\frac{e^{-b\frac{(x-\mu)}{s}}}{B(a, b)s(1 + e^{-\frac{(x-\mu)}{s}})^{a+b}} \right] dx \\
&\quad - \log B(a, b) - \int_{-\infty}^{\infty} (a+b-2) \log [1 + e^{-\frac{(x-\mu)}{s}}] \\
&\quad \times \left[\frac{e^{-b\frac{(x-\mu)}{s}}}{B(a, b)s(1 + e^{-\frac{(x-\mu)}{s}})^{a+b}} \right] dx.
\end{aligned}$$

Fazendo a mudança de variáveis $t = \frac{x-\mu}{s}$, obtemos

$$\begin{aligned}
I[F, G] &= -(b-1) \int_{-\infty}^{\infty} t \frac{e^{-bt}}{B(a, b)s(1 + e^{-t})^{a+b}} s dt - \log B(a, b) \\
&\quad - (a+b-2) \int_{-\infty}^{\infty} \log(1 + e^{-t}) \frac{e^{-bt}}{B(a, b)s(1 + e^{-t})^{a+b}} s dt \\
&= -(b-1)E_{BLo}(T) - \log B(a, b) - (a+b-2)E_{BLo}(\log(1 + e^{-T})) \\
&= -\log B(a, b) - (b-1)\{\mu + s[\Psi(a) - \Psi(b)]\} + (a+b-2)[\Psi(a+b) - \Psi(a)],
\end{aligned}$$

Se $a = b = 1$ temos que $I[F, G] = 0$

2.2.10 Momentos-L

Os momentos-L são estatísticas usadas para resumir a forma de uma distribuição de probabilidade. Eles são análogos aos momentos convencionais, mas são definidos como combinações lineares das estatísticas de ordem. Eles são definidos por Hosking (1990) como

$$\lambda_{r+1} = r(r+1)^{-1} \sum_{k=1}^r \frac{(-1)^k}{k} E(X_{(r+1-k;r+1)})$$

para $r = 1, \dots$

Os quatro primeiros momentos-L de uma distribuição de probabilidade são

$$\lambda_1 = E(X_{1:1}), \quad \lambda_2 = \frac{1}{2}E(X_{2:2} - X_{1:2}), \quad \lambda_3 = \frac{1}{3}E(X_{3:3} - 2X_{2:3} + X_{1:3}) \text{ e}$$

$$\lambda_4 = \frac{1}{4}E(X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4}).$$

A partir das expansões dos momentos das estatísticas de ordem dadas anteriormente, podemos obter expansões para os momentos-L da distribuição BLo como combinações lineares ponderadas das esperanças de estatísticas de ordem da distribuição beta logística.

2.2.11 Desvio médio em relação à média

O desvio médio em relação ao valor esperado de uma distribuição é dado por

$$\delta_1 = \int_{-\infty}^{\infty} |x - \mu| f(x) dx.$$

Desenvolvendo a integral da expressão acima temos

$$\begin{aligned}\delta_1 &= \int_{-\infty}^{\mu} (\mu - x)f(x)dx + \int_{\mu}^{\infty} (x - \mu)f(x)dx = 2 \int_{-\infty}^{\mu} (\mu - x)f(x)dx \\ &= 2\mu F(\mu) - 2 \int_{-\infty}^{\mu} xf(x)dx.\end{aligned}\quad (2.13)$$

Substituindo a f.d.p. da distribuição BLo, equação (2.1), em (2.13), obtemos

$$\delta_1 = \int_{-\infty}^{\mu} x \frac{[e^{-\frac{(x-\mu)}{s}}]^b}{B(a, b)s[1 + e^{-\frac{(x-\mu)}{s}}]^{a+b}} dx.$$

Fazendo $t = \frac{x-\mu}{s}$ temos

$$\begin{aligned}\delta_1 &= \int_{-\infty}^0 (\mu + st) \frac{(e^{-t})^b}{B(a, b)(1 + e^{-t})^{a+b}} dt \\ &= \mu F(0) + s \int_{-\infty}^0 t \frac{(e^{-t})^b}{B(a, b)(1 + e^{-t})^{a+b}} dt.\end{aligned}$$

Substituindo $m = \frac{1}{1+e^{-t}}$, obtemos

$$\delta_1 = \frac{s}{B(a, b)} \left[\int_0^{\frac{1}{2}} \log m (1 - m)^{b-1} m^{a-1} dm + \int_0^{\frac{1}{2}} \log (1 - m) (1 - m)^{b-1} m^{a-1} dm \right].$$

Portanto,

$$\begin{aligned}\delta_1 &= 2\mu F(\mu) - 2\mu F(0) \\ &+ 2 \left\{ \frac{\log 0,5 \times s}{B(a, b)} \left[\sum_{j=0}^{b-1} \binom{b-1}{j} (-1)^j \frac{1}{(a+j)2^{a+j}} \right] \right\} \\ &+ 2 \left\{ \frac{\log 0,5 \times s}{B(a, b)} \left[\sum_{j=0}^{b-1} \sum_{k=0}^{a+j-1} \binom{b-1}{j} \binom{a+j-1}{k} \frac{1}{(k+1)2^{k+1}} \left(\frac{1}{2} - \frac{1}{(k+1)} \right) \right] \right\}.\end{aligned}$$

2.2.12 Desvio médio em relação à mediana

O desvio médio em relação a mediana de uma distribuição é dado por

$$\delta_2 = \int_{-\infty}^{\infty} |x - M| f(x) dx.$$

Desenvolvendo a integral da expressão acima, obtemos

$$\begin{aligned} \delta_2 &= \int_{-\infty}^M (M - x) f(x) dx + \int_M^{\infty} (x - M) f(x) dx \\ &= 2 \int_{-\infty}^M M f(x) dx - 2 \int_M^{\infty} x f(x) dx + E(x) - M \\ &= 2MF(M) + E(x) - M - 2 \int_M^{\infty} x f(x) dx. \end{aligned}$$

De forma análoga ao desvio médio em relação à média, podemos concluir que o desvio médio em relação à mediana da distribuição beta logística é expresso por

$$\begin{aligned} \delta_2 &= 2MF(M) + E(x) - M - 2\mu F\left(\frac{M - \mu}{s}\right) \\ &+ 2 \left\{ \frac{s}{B(a, b)} \log \left[\frac{1}{1 + e^{-\frac{(M-\mu)}{s}}} \right] \left[\sum_{j=0}^{b-1} \binom{b-1}{j} (-1)^j \frac{1}{a+j} \left(\frac{1}{1 + e^{-\frac{(M-\mu)}{s}}} \right)^{a+j} \right] \right\} \\ &+ 2 \left\{ \sum_{j=0}^{b-1} \sum_{k=0}^{a+j-1} \binom{b-1}{j} \binom{a+j-1}{k} \frac{\left(1 + e^{-\frac{(M-\mu)}{s}}\right)^{-(k+1)}}{(k+1)} \left[\log \frac{1}{1 + e^{-\frac{(M-\mu)}{s}}} - \frac{1}{(k+1)} \right] \right\}. \end{aligned}$$

2.2.13 Curva de Lorenz

A curva de Lorenz é um gráfico utilizado para representar a distribuição relativa de uma variável em um domínio determinado. O domínio pode ser o conjunto de pessoas de uma região ou país, por exemplo. A variável cuja distribuição se estuda pode ser a renda das pessoas. A curva é traçada considerando-se a percentagem

acumulada de pessoas no eixo das abscissas e a percentagem acumulada de renda no eixo das ordenadas. Ela foi desenvolvida pelo economista Max O. Lorenz em 1905 para representar a distribuição de renda.

Cada ponto da curva é lido como percentagem cumulativa das pessoas. A curva parte da origem (0,0) e termina no ponto (100,100). Se a renda estivesse distribuída de forma perfeitamente equitativa, a curva coincidiria com a linha de 45 graus que passa pela origem (por exemplo, 30% da população recebe 30% da renda).

Para uma variável aleatória X com função inversa da função distribuição acumulada dada por $F^{-1}(X)$ (ou função de quantis de X). A curva de Lorenz é definida por

$$L(p) = \frac{1}{\mu} \int_0^p F^{-1}(x) dx, \quad (2.14)$$

em que $\mu = E(X)$.

A inversa da função beta incompleta, $I_z^{-1}(a, b)$, pode ser encontrada no sítio <http://functions.wolfram.com/06.23.06.0004.01>.

$$\begin{aligned} I_z^{-1}(a, b) &= t + \frac{b-1}{a+1} t^2 + \frac{(b-1)(a^2 + 3ab - a + 5b - 4)}{2(a+1)^2(a+2)} t^3 \\ &+ \left\{ \frac{(b-1)[a^4 + (6b-1)a^3 + (b+2)(8b-5)a^2]}{3(a+1)^3(a+2)(a+3)} \right\} \\ &+ \left\{ \frac{(b-1)[(33b^2 - 30b + 4)a + b(31b - 47) + 18]}{3(a+1)^3(a+2)(a+3)} \right\} t^4 + O(z^{\frac{5}{a}}), \end{aligned}$$

em que $t = [az B(a, b)]^{\frac{1}{a}}$, para $a > 0$ e $b > 0$.

Podemos reescrever $I_z^{-1}(a, b)$ como

$$\begin{aligned}
I_z^{-1}(a, b) &= [az B(a, b)]^{\frac{1}{a}} + \left(\frac{b-1}{a+1}\right) [az B(a, b)]^{\frac{2}{a}} \\
&\quad + \left[\frac{(b-1)(a^2 + 3ab - a + 5b - 4)}{2(a+1)^2(a+2)}\right] [az B(a, b)]^{\frac{3}{a}} + \dots \\
&= z^{\frac{1}{a}} [az B(a, b)]^{\frac{1}{a}} + z^{\frac{2}{a}} [az B(a, b)]^{\frac{2}{a}} + z^{\frac{3}{a}} [az B(a, b)]^{\frac{3}{a}} + \dots \\
&= d_1 z^{\frac{1}{a}} + d_2 z^{\frac{2}{a}} + d_3 z^{\frac{3}{a}} + \dots \\
&= \sum_{j=1}^{\infty} d_j z^{\frac{j}{a}},
\end{aligned}$$

em que $d_j = c_j [a B(a, b)]^{\frac{j}{a}}$, $j = 1, 2, 3, \dots$ e $c_j = \frac{(b-1)^{j-1}}{(a+1)^{j-1}}$

Para a distribuição BLo, a função inversa da função distribuição acumulada é dada por

$$F^{-1}(x) = I_{\frac{1}{1+e^{-\frac{(x-\mu)}{s}}}}^{-1}(a, b) = \sum_{j=1}^{\infty} d_j \left[\frac{1}{1+e^{-\frac{(x-\mu)}{s}}} \right]^{\frac{j}{a}}. \quad (2.15)$$

Aplicando a equação (2.15) em (2.14), obtemos

$$L(p) = \frac{1}{\mu} \int_0^p \sum_{j=1}^{\infty} d_j \left[\frac{1}{1+e^{-\frac{(x-\mu)}{s}}} \right]^{\frac{j}{a}} dx.$$

Substituindo $u = \frac{1}{1+e^{-\frac{(x-\mu)}{s}}}$, temos

$$\begin{aligned}
L(p) &= \frac{1}{\mu} \sum_{j=1}^{\infty} d_j \int_{\frac{1+e^{-\frac{\mu}{s}}}{1+e^{-\frac{(p-\mu)}{s}}}}^{\frac{1+e^{-\frac{(p-\mu)}{s}}}{1+e^{-\frac{\mu}{s}}}} u^{\frac{j}{a}} \frac{s}{u(1-u)} du \\
&= \frac{s}{\mu} \sum_{j=1}^{\infty} d_j \int_{\frac{\mu}{s}}^{\frac{(p-\mu)}{s}} \sum_{k=1}^{\infty} u^{k+\frac{j}{a}-1} du
\end{aligned}$$

Desenvolvendo a integral acima, a expressão da curva de Lorenz da distribuição beta logística é dada por

$$L(p) = \frac{s}{\mu} \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} d_j \frac{\left[1+e^{-\frac{(p-\mu)}{s}}\right]^{k+\frac{j}{a}} - \left[1+e^{-\frac{\mu}{s}}\right]^{k+\frac{j}{a}}}{k+\frac{j}{a}}.$$

2.2.14 Curva de Bonferroni

Embora do ponto de vista formal a curva de Bonferroni represente a desigualdade em um forma equivalente à curva de Lorenz, a informação que produzem é diferente. Os valores de $L(p)$ da curva de Lorenz são frações da renda total, enquanto os valores de $Q(p)$, da curva de Bonferroni, referem-se os níveis de rendimento relativos. A curva de Bonferroni é definida por

$$Q(p) = \frac{1}{p\mu} \int_0^p F^{-1}(x) dx,$$

A curva de Bonferroni da distribuição BLo é análoga à curva de Lorenz. Portanto,

$$Q(p) = \frac{s}{p\mu} \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} d_j \frac{\left[1 + e^{-\frac{(p-\mu)}{s}}\right]^{k+\frac{j}{a}} - \left[1 + e^{\frac{\mu}{s}}\right]^{k+\frac{j}{a}}}{k + \frac{j}{a}}.$$

2.3 Caracterização da beta logística

Nesta seção apresentamos algumas características e relações da distribuição beta logística com outras distribuições existentes na literatura, que poderão ser úteis para os estudos de diversas áreas do conhecimento.

2.3.1 Características da beta logística

Uma das principais características da distribuição BLo é dada por:

Seja X uma variável aleatória com distribuição $BLo(\mu, \sigma, a, b)$, considerando $\mu = 0$ e $\sigma = \frac{\pi}{\sqrt{3}}$, a função de densidade de probabilidade de uma variável com distribuição

$BLo(0, \frac{\pi}{\sqrt{3}}, a, b)$ é representada por

$$\begin{aligned}
 f(x) &= \frac{e^{-bx}}{B(a, b)s(1 + e^{-x})^{a+b}} \\
 &= \frac{1}{B(a, b)s} \frac{e^{-bx}}{(1 + e^{-x})^{a+b}} \\
 &= \frac{1}{B(a, b)s} \left[\frac{e^{-x}}{(1 + e^{-x})} \right]^b \left[\frac{1}{(1 + e^{-x})} \right]^a \\
 &= \frac{1}{B(b, a)s} \left[\frac{1}{(1 + e^x)} \right]^b \left[\frac{e^x}{(1 + e^x)} \right]^a \\
 &= \frac{e^{ax}}{B(a, b)s(1 + e^x)^{a+b}}.
 \end{aligned}$$

o que determina a f.d.p. de uma v.a. $-X \sim BLo(0, \frac{\pi}{\sqrt{3}}, b, a)$. Assim podemos concluir que se $X \sim BLo(0, \frac{\pi}{\sqrt{3}}, a, b)$, então, $-X \sim BLo(0, \frac{\pi}{\sqrt{3}}, b, a)$.

A segunda característica da distribuição BLo é verificada se X é uma variável aleatória $BLo \sim (\mu = 0, \sigma = \frac{\pi}{\sqrt{3}}, a, b)$, então, a f.d.p. é dada por

$$\begin{aligned}
 f(x) &= \frac{e^{-bx}}{B(a, b)(1 + e^{-x})^{a+b}} \\
 &= \frac{1}{B(a, b)} \frac{e^{-bx}}{(1 + e^{-x})^{a+b}} \\
 &= \frac{1}{B(a, b)} \left[\frac{e^{-x}}{(1 + e^{-x})} \right]^b \left[\frac{1}{(1 + e^{-x})} \right]^a \\
 &= \frac{1}{B(a, b)} [1 - G(x)]^b [G(x)]^a,
 \end{aligned}$$

em que $G(x)$ é a f.d.a. da distribuição $Lo(\mu = 0, \sigma = \frac{\pi}{\sqrt{3}})$.

2.3.2 Relação e convergência para outras distribuições

Considerando a variável aleatória $X \sim BLo(\mu, \sigma, a, b)$. Para valores específicos destes parâmetros é possível estabelecer relações ou mesmo verificar a convergência da distribuição beta logística para outras distribuições apresentadas na literatura. Algumas destas relações são listadas a seguir:

- 1- Se $b = 1$ e $a = \alpha$, então $X \sim$ Distribuição logística generalizada tipo I.
- 2- Se $a = b = \alpha$, então $X \sim$ Distribuição logística generalizada tipo III.
- 3- Se $a = b = 1$, $\mu = 0$ e $\sigma = \frac{\pi}{\sqrt{3}}$, então $X \sim Lo\left(0, \frac{\pi}{\sqrt{3}}\right)$.
- 4- Se $S = \log\left(\frac{T}{1-T}\right)$ e $T \sim Beta(a, b)$, então $S \sim X$.
- 5- Se $S = -\log V$ e $V \sim F(2p, 2q)$, então $S \xrightarrow{d} X$ (onde "d" denota a convergência em distribuição de S em X).
- 6- Se $a \rightarrow \infty$, então $X \sim Lognormal(\mu, \sigma^2)$.
- 7- Se $a = 1$ e $b \rightarrow \infty$, então $X \sim Weibull(\lambda, k)$.
- 8- Se $a \rightarrow \infty$ e $b \rightarrow \infty$, então $X \sim Normal(\mu, \sigma)$.

Capítulo 3

Inferência

Nesta seção apresentamos os estimadores aproximados de máxima verossimilhança dos parâmetros da distribuição beta logística e deduzimos a matriz de informação esperada. Posteriormente, desenvolvemos a teoria assintótica referente à distribuição BLo, utilizamos alguns critérios de comparação de modelos estatísticos e finalmente verificamos a bondade do ajuste proposto.

3.1 Inferência

Seja x o valor observado de uma variável aleatória X que possui distribuição beta logística (BLo), com o vetor de parâmetros $\boldsymbol{\theta}=(\mu, \sigma, a, b)^T$. A função de log-verossimilhança $l(\boldsymbol{\theta})$ é dada pela seguinte expressão:

$$l(\boldsymbol{\theta}) = \log(\pi) - \log(\sigma) - \frac{1}{2} \log(3) - \log[B(a, b)] - \frac{b\pi}{\sqrt{3}} \left(\frac{x - \mu}{\sigma} \right) - (a+b) \log \left[1 + e^{-\frac{\pi}{\sqrt{3}} \left(\frac{x - \mu}{\sigma} \right)} \right].$$

Substituindo $t = \frac{\pi}{\sqrt{3}} \left(\frac{x - \mu}{\sigma} \right)$, temos

$$l(\boldsymbol{\theta}) = \log(\pi) - \log(\sigma) - \frac{1}{2} \log(3) - \log[B(a, b)] - bt - (a + b) \log [1 + e^{-t}]$$

As derivadas parciais de t em relação μ e σ , são dadas, respectivamente, por $\frac{\partial t}{\partial \mu} = -\frac{\pi}{\sigma\sqrt{3}}$ e $\frac{\partial t}{\partial \sigma} = -\frac{\pi}{\sqrt{3}} \left(\frac{x-\mu}{\sigma^2}\right)$.

Os componentes do vetor escore $\mathbf{U} = \left(\frac{\partial l}{\partial \mu}, \frac{\partial l}{\partial \sigma}, \frac{\partial l}{\partial a}, \frac{\partial l}{\partial b}\right)^T$, são obtidos por diferenciação, isto é,

$$\begin{aligned}\frac{\partial l}{\partial \mu} &= \frac{b\pi}{\sigma\sqrt{3}} - (a+b) \left[\frac{\pi e^{-t}}{\sigma\sqrt{3}(1+e^{-t})} \right], \\ \frac{\partial l}{\partial \sigma} &= \frac{bt}{\sigma} - (a+b) \left[\frac{te^{-t}}{\sigma(1+e^{-t})} \right], \\ \frac{\partial l}{\partial a} &= -\psi(a) + \psi(a+b) - \log(1+e^{-t})\end{aligned}$$

e

$$\frac{\partial l}{\partial b} = -\psi(b) + \psi(a+b) - \log(1+e^{-t}),$$

em que $\psi(\cdot)$ é a função digama.

Para uma amostra aleatória x_1, \dots, x_n de tamanho n , o logaritmo da função de verossimilhança é denotado por $l_n = l_n(\mu, \sigma, a, b) = \sum_{i=1}^n l^{(i)}$, em que $l^{(i)}$ é o logaritmo da função de verossimilhança para a i -ésima observação, $i = 1, \dots, n$. A função escore é denotada por $\mathbf{U}_n = \sum \mathbf{U}^{(i)}$, em que cada $\mathbf{U}^{(i)}$ é definido acima. O estimador de máxima verossimilhança (EMV) $\hat{\boldsymbol{\theta}}$ de $\boldsymbol{\theta}$ é a solução de equações não lineares $\mathbf{U}_n = \mathbf{0}$. Neste trabalho a solução é obtida numericamente.

As segundas derivadas da função de log-verossimilhança são apresentadas no Apêndice B.

A matriz de informação esperada é utilizada para desenvolver a estimação por

intervalo e testes de hipóteses sobre os parâmetros em questão. É denotada por

$$\mathbf{K} = \mathbf{K}(\boldsymbol{\theta}) = \begin{bmatrix} k_{\mu,\mu} & k_{\mu,\sigma} & k_{\mu,a} & k_{\mu,b} \\ k_{\sigma,\mu} & k_{\sigma,\sigma} & k_{\sigma,a} & k_{\sigma,b} \\ k_{a,\mu} & k_{a,\sigma} & k_{a,a} & k_{a,b} \\ k_{b,\mu} & k_{b,\sigma} & k_{b,a} & k_{b,b} \end{bmatrix}.$$

Os elementos da matriz de informação esperada $K(\boldsymbol{\theta})$ para os parâmetros μ, σ, a e b da distribuição beta logística são

$$\begin{aligned} k_{\mu,\mu} &= \frac{\pi^2(a+b)}{3\sigma^2} E \left[\frac{e^{-T}}{(1+e^{-T})^2} \right], \\ k_{\mu,\sigma} &= \frac{-b\pi}{\sqrt{3}\sigma^4} E \left[\frac{e^{-T}}{1+e^{-T}} \right] + \frac{b\pi}{\sqrt{3}\sigma^4} E \left[\frac{T e^{-T}}{(1+e^{-T})^2} \right], \\ &\quad - \frac{\pi(a+b)}{\sqrt{3}\sigma^2} E \left[\frac{e^{-T}}{1+e^{-T}} \right] + \frac{\pi(a+b)}{\sqrt{3}\sigma^2} E \left[\frac{T e^{-T}}{(1+e^{-T})^2} \right], \\ k_{\mu,a} &= \frac{\pi}{\sqrt{3}\sigma} E \left(\frac{e^{-T}}{1+e^{-T}} \right), \\ k_{\mu,b} &= \frac{\pi}{\sqrt{3}\sigma} E \left(\frac{e^{-T}}{(1+e^{-T})} \right), \\ k_{\sigma,\sigma} &= -\frac{1}{\sigma^2} + \frac{b}{\sigma^2} E(T) - \frac{2\pi(a+b)}{\sqrt{3}\sigma^2} E \left[\frac{T e^{-T}}{1+e^{-T}} \right], \\ &\quad - \frac{\pi(a+b)}{\sqrt{3}\sigma^2} E \left[\frac{T}{(1+e^{-T})^2} \right], \\ k_{\sigma,a} &= \frac{1}{\sigma} E \left[\frac{T e^{-T}}{1+e^{-T}} \right], \\ k_{\sigma,b} &= \frac{1}{\sigma} E \left[\frac{T e^{-T}}{1+e^{-T}} \right], \\ k_{a,a} &= \psi'(a) - \psi'(a+b), \\ k_{b,b} &= \psi'(b) - \psi'(a+b), \\ k_{a,b} &= -\psi'(a+b). \end{aligned}$$

Sob certas condições que são satisfeitas para os parâmetros no interior do espaço paramétrico, exceto na fronteira. A distribuição assintótica $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ de uma distribuição de probabilidade é expressa por $N_4(0, \mathbf{K}(\boldsymbol{\theta})^{-1})$. A estimativa do vetor

de parâmetros da distribuição assintótica normal multivariada $N_4(0, n^{-1}\mathbf{K}(\hat{\boldsymbol{\theta}}^{-1}))$, $\hat{\boldsymbol{\theta}}$, pode ser aplicado para determinar os parâmetros do intervalo de confiança da distribuição beta logística e compará-la com outros através de alguns testes estatísticos assintóticos, tais como, estatística da razão de verossimilhanças (RV), escore(s) e a estatística de Wald (W).

Para determinar um intervalo de confiança assintótico, com coeficiente de confiança $1-\gamma$, para cada um dos parâmetros θ_i é dado por

$$ICA(\theta_i, (1 - \gamma)) = \left(\hat{\theta}_i - z_{\frac{\gamma}{2}} \sqrt{\hat{k}^{\theta_i, \theta_i}}, \hat{\theta}_i + z_{\frac{\gamma}{2}} \sqrt{\hat{k}^{\theta_i, \theta_i}} \right)$$

em que $\hat{k}^{\theta_i, \theta_i}$ é o i -ésimo elemento da diagonal da matriz $n^{-1}\mathbf{K}(\boldsymbol{\theta})^{-1}$ calculado em $\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}$ para $i = 1, \dots, 4$ e $z_{\frac{\gamma}{2}}$ é o quantil $1 - \frac{\gamma}{2}$ da distribuição normal padrão.

A partição $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$ determina os testes de hipóteses do tipo $H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^{(0)}$ contra $H_A : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_1^{(0)}$. A estatística de RV é expressa por $w = 2 \left[l(\hat{\boldsymbol{\theta}}) - l(\tilde{\boldsymbol{\theta}}) \right]$ em que $\hat{\boldsymbol{\theta}}$ e $\tilde{\boldsymbol{\theta}}$ são os EMVs de $\boldsymbol{\theta}$ sob H_0 e H_A , respectivamente. Sob a hipótese nula, $H_0 : w \rightarrow X_q^2$, em que q é a dimensão do vetor $\boldsymbol{\theta}_1$. O teste da RV rejeita H_0 se $w > \zeta_\gamma$, em que ζ_γ denota o quantil 100γ da distribuição X_q^2 .

Ao compararmos dois modelos estatísticos, podemos utilizar critérios de seleção de modelos, tais como o critério de informação de Akaike - AIC (Akaike, 1974), critério de informação de Schwarz - BIC (Schwarz, 1978), critério de informação de Akaike corrigido -CAIC (Liang e Zou, 2007) e critério de informação Hannan-Quinn - HQC (1979). Estes critérios são baseados no valor do logaritmo da função de verossimilhança do modelo e dependem dos números de observações e parâmetros.

O critério AIC é definido por $AIC = -2l(\boldsymbol{\theta}) + 2p$, em que p representa o número de parâmetros do modelo.

O critério de informação de Schwarz, é obtido por $BIC = -2l(\boldsymbol{\theta}) + p \log(n)$,

em que n denota o número de observações do conjunto de dados.

O critério de informação Hannan-Quinn é dado por $HQC = -2l(\boldsymbol{\theta}) + 2p \log[\log(n)]$.

Liang e Zou (2007) propuseram um AIC melhorado (CAIC) dado por

$$CAIC = AIC + \frac{2p(p+1)}{n-p-1}.$$

Valores menores de AIC, BIC, HQC e CAIC indicam modelos mais adequados.

Após uma análise paramétrica, propomos um estudo não paramétrico cujo enfoque é comparar a função empírica com algumas distribuições. Neste contexto utilizamos o teste de Kolmogorov-Smirnov, denotado por K-S.

O teste não paramétrico (KS) é utilizado para determinar se duas distribuições de probabilidade diferem uma da outra ou se uma das distribuições de probabilidade difere da distribuição em teste, em qualquer dos casos com base em amostras finitas. As hipótese do teste são $H_0 : F = F_n$ vs $H_A : F \neq F_n$, em que F é a f.d.a. da distribuição em estudo e F_n é a f.d.a. da distribuição empírica.

A estatística do teste Kolmogorov-Smirnov é expressa por

$$D = \sup |F_n(x) - F(x)|$$

em que $F_n(i) = \frac{i-0.5}{n}$.

Capítulo 4

Simulação e aplicação

Neste capítulo realizamos um estudo de simulação da distribuição BLo. Posteriormente são apresentados dois conjuntos de dados relacionados às ciências biomédicas e às ciências dos materiais, respectivamente. Fazemos um estudo paramétrico nos dados, obtendo de forma numérica as estimativas de máxima verossimilhança (EMVs) e os respectivos erros padrão, as estatísticas de comparação entre modelos (AIC, BIC, CAIC e HQC). Ajustamos os dados com as distribuições beta logística, logística, beta normal, beta Weibull e beta logística generalizada. Além disso, construímos gráficos QQ para verificarmos o ajuste da distribuição beta logística com as distribuições e comparamos os modelos através de testes de hipóteses. Após esta análise, realizamos um estudo de bondade do ajuste utilizando o teste K-S.

4.1 Estudo de simulação

Para simular amostras da distribuição beta logística com f.d.p. dada pela equação (2.1), realizamos os seguintes passos:

Passo 1: Seja Y uma variável aleatória que segue uma distribuição beta com parâmetros a e b , isto é, $Y \sim Beta(a, b)$. Seja $G(X)$ a f.d.a. da distribuição $Lo(\mu, \sigma)$.

Passo 2: Consideramos $X = G^{-1}(Y)$. Esta aplicação determina a distribuição BLo, ou seja, $X \sim BLo(\mu, \sigma, a, b)$.

Por meio de simulação, geramos uma amostra de tamanho 60 da variável aleatória da distribuição BLo com parâmetros ($\mu = 4; \sigma = 3; a = 1,5; b = 2$).

A Figura 4.1 apresenta o gráfico da função densidade da distribuição BLo comparado com o histograma dos dados simulados para os valores dos parâmetros mencionados anteriormente.

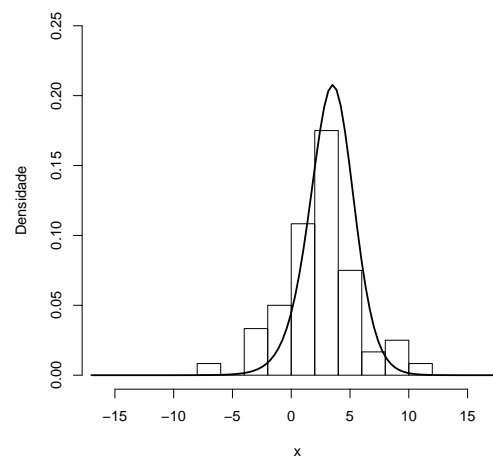


Figura 4.1: Gráfico da função densidade da distribuição BLo

Para o desenvolvimento da simulação foi utilizado o software estatístico R (R Development Core Team, 2012)

4.2 Dados de células cancerígenas

O conjunto de dados de cDNA consiste de várias observações medidas em 60 células humanas cancerosas (NC160). Mais detalhes sobre este conjunto de dados pode ser encontrado no sítio <http://discover.nci.nih.gov/datasetsNature2000.jsp>. O conjunto de dados contém as medidas de 1375 níveis de expressão gênica nas 60 células cancerosas e é chamado de Matrix-T com dimensão 1375×60 . Cada coluna da matriz representa uma célula de cancer. Para esta análise de dados, consideramos uma das colunas desta matriz chamada de CO:KM12.

Na Tabela 4.1 apresentamos os EMVs e os correspondentes erros padrão (que estão entre parênteses) para os parâmetros das distribuições beta logística (BLo), logística (Lo), beta normal (BN) e beta logística generalizada (BGL).

Tabela 4.1: EMVs (erros padrão) dos parâmetros das distribuições BLo, Lo, BN e BGL para dados célula cancerígenas

Parâmetros				
Distribuições	μ	σ	a	b
BLo	0,56 (0,37)	0,88 (0,45)	0,39 (0,33)	0,97 (0,57)
Lo	-0,30 (26,69)	1,45 (30,04)	1 -	1 -
BN	2,72 (0,31)	1,62 (1,62)	0,46 (0,71)	6,44 (0,49)
	p	q	a	b
BGL	4,08 (11,41)	0,53 (44,70)	0,26 (148,18)	4,42 (4,56)

Podem ser vistos os valores das estatísticas AIC, BIC, HQC e CAIC para as distribuições BLo, Lo, BN e BGL. Estes resultados indicam que a distribuição *BLo* tem o menor valor em cada estatística entre as distribuições ajustadas. Portanto, esta distribuição apresenta o melhor ajuste, conforme a Tabela 4.2.

Tabela 4.2: Estatísticas AIC, BIC, CAIC e HQC das distribuições BLo, Lo, BN e BGL para dados célula de câncer

Distribuições	AIC	BIC	CAIC	HQC
BLo	4715,51	4736,34	4715,54	4723,31
Lo	4824,01	4844,84	4824,04	4823,91
BN	4746,56	4767,39	4746,59	4754,36
BGL	4733,79	4754,62	4733,82	4741,59

A Figura 4.2 mostra o ajuste das funções de densidade das distribuições BLo, Lo, BN e BGL. Podemos observar que o melhor ajuste é apresentado pela distribuição beta logística.

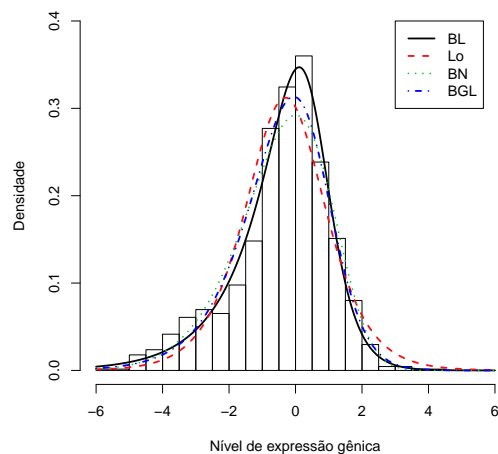


Figura 4.2: Função de densidade estimada para dados de célula de câncer

Os gráficos QQ das distribuições BLo, Lo, BN e BGL são apresentados na Figura 4.3. Podemos observar que há um melhor ajuste na distribuição BLo quando comparada com as outras distribuições.

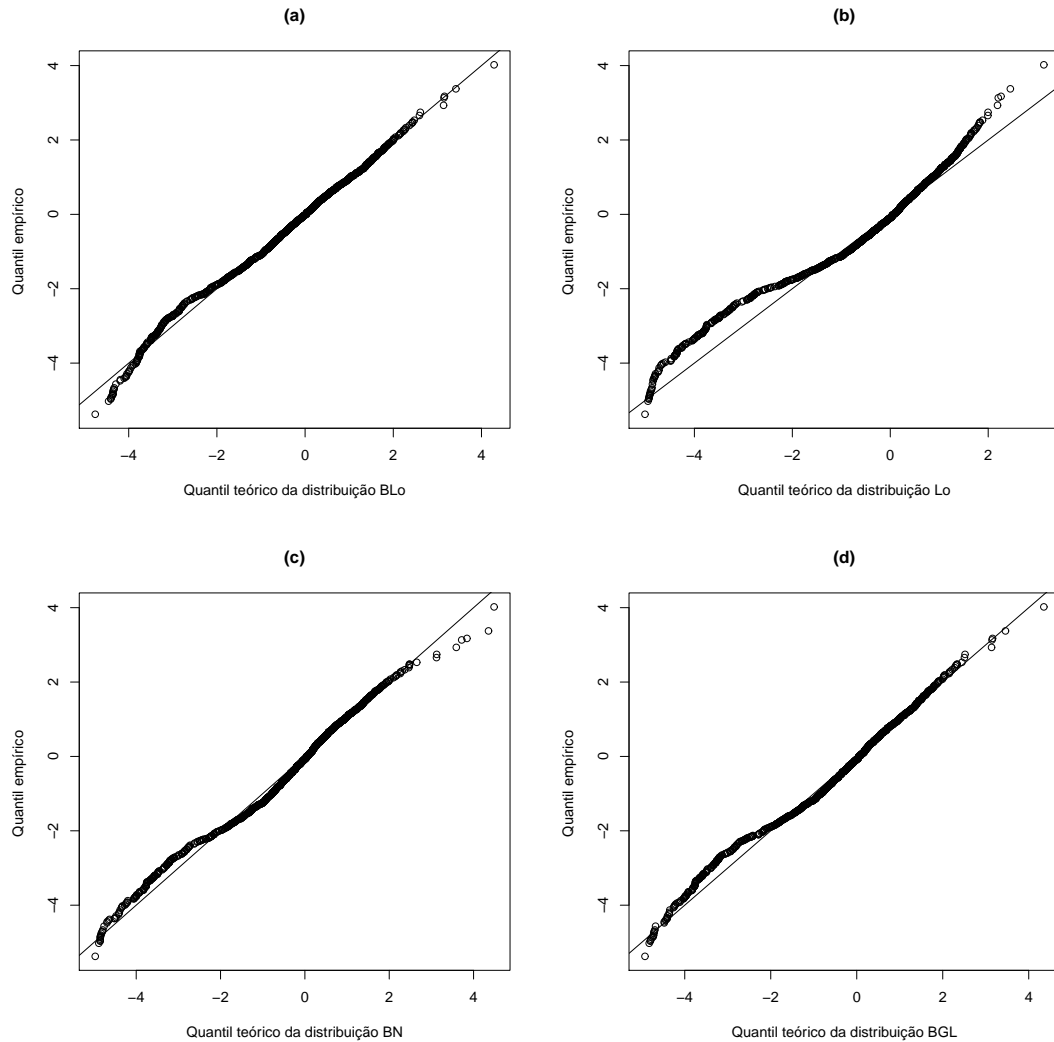


Figura 4.3: Gráficos QQ das distribuições (a)BLo, (b)Lo, (c)BN e (d)BGL

Utilizamos a estatística da razão de verossimilhanças (RV) para comparar distribuições encaixados $H_0 : Lo$ vs $H_A : BLo$, cujo resultado é dado por $w = 108,722$ ($p < 0,0001$), com 2 graus de liberdade. Assim, a distribuição BLo ajusta os dados significativamente melhor do que a distribuição Lo.

Para comparar modelos não encaixados, tais como, as distribuições BN e BGL, utilizamos a estatística da razão de verossimilhanças generalizada (RVG). Considere dois modelos não encaixados F_θ e G_λ com f.d.p. $f(y_i; x_i, \theta)$ e $g(y_i; x_i, \lambda)$, respectivamente Vuong (1989) propõe a estatística

$$T_{LR,NN} = \left\{ \frac{1}{\sqrt{n}} \sum_1^n \log \frac{f(y_i; x_i, \theta)}{g(y_i; x_i, \lambda)} \right\} \\ \times \left\{ \frac{1}{n} \sum_1^n \left(\log \frac{f(y_i; x_i, \theta)}{g(y_i; x_i, \lambda)} \right)^2 - \left(\frac{1}{n} \sum_1^n \log \frac{f(y_i; x_i, \theta)}{g(y_i; x_i, \lambda)} \right)^2 \right\}^{-1}$$

A distribuição da estatística $T_{LR,NN}$ converge sob a hipótese nula, se os modelos são equivalentes, para a distribuição normal padrão. A hipótese nula não é rejeitada se $T_{LR,NN} \leq z_{\frac{\alpha}{2}}$ (nível de significância= α). E rejeita-se a hipótese nula se $T_{LR,NN} > z_\alpha$.

Para comparar distribuições não encaixados, tais como, beta normal (BN) e beta logística generalizada (BGL), cada uma destas distribuições versus a distribuição beta logística (BLo). Por exemplo, podemos considerar que $f(y_i; x_i, \theta)$ e $g(y_i; x_i, \lambda)$ são as densidades da distribuição BN e BLo, respectivamente. Consideramos as hipóteses $H_0 : BN \text{ vs } H_A : BLo$ e $H_0 : BGL \text{ vs } H_A : BLo$ as estatísticas do teste RVG assumem os valores 10,584 ($p < 0,0001$) e 20,395 ($p < 0,0001$), respectivamente. Assim, podemos concluir que a distribuição BLo apresenta o melhor ajuste que as outras distribuições em estudo.

Após fazer um estudo paramétrico nos dados de célula de câncer, percebemos que a distribuição BLo é mais adequada no ajuste dos dados se compararmos com as outras distribuições. Para comprovar esta tendência, foi realizado um teste de bondade do ajuste, enfocando a aplicação da função empírica e o teste Kolmogorov-Smirnov(KS) nos dados.

A Figura 4.4 mostra a função distribuição empírica e as funções de distribuição

acumulada das distribuições beta logística, logística, beta normal e beta logística generalizada.

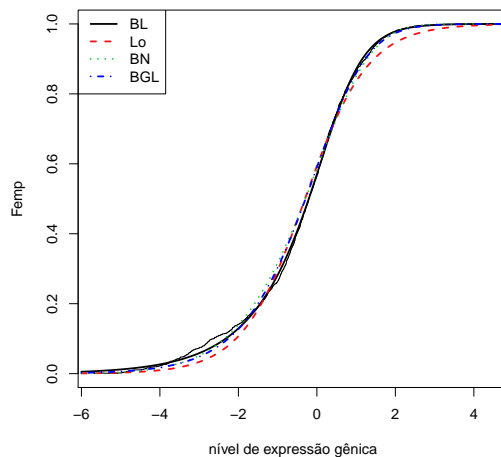


Figura 4.4: Gráfico da função empírica e das funções de distribuição acumulada das distribuições BLo, Lo, BN e BGL.

Para verificar a bondade do ajuste, propormos o teste não paramétrico K-S e obtemos as estatísticas do teste para cada distribuição, BLo ($KS=0,0239$; $p=0,4227$), Lo ($KS=0,0505$; $p=0,0021$), BN ($KS=0,0558$; $p=0,0004$) e BGL ($KS=0,0443$; $p=0,0009$). Ao nível de 5% a única com ajuste satisfatório é a distribuição BLo.

4.3 Dados de resistência de fibra de vidro

Fibra de vidro é um material composto da aglomeração de finíssimos filamentos de vidro, que não são rígidos e são altamente flexíveis. Quando adicionado resina poliéster (ou outro tipo de resina), transforma-se em um composto popularmente conhecido como fibra de vidro, mas na verdade o nome correto é PRFV, ou seja, "Polímero Reforçado com Fibra de Vidro".

O conjunto de dados corresponde a resistência à ruptura de $n = 63$ fibras de vidro de 1,5 cm de comprimento, originalmente obtidos por trabalhadores no Laboratório Nacional de Física do Reino Unido (Smith e Naylor, 1987).

Na Tabela 4.3 , apresentamos os EMVs e os erros-padrão para cada distribuição.

Tabela 4.3: EMVs (erros padrão) dos parâmetros das distribuições BLo, Lo, BN, BW e BGL para dados de fibra de vidro.

Distribuições	Parâmetros			
	μ	σ	a	b
BLo	1,69 (34,34)	0,05 (166,76)	0,10 (77,92)	0,21 (28,45)
Lo	1,54 (0,19)	0,31 (0,19)	1 -	1 -
BN	2,27 (0,94)	0,36 (3,47)	0,45 (0,64)	8,18 (0,62)
	λ	q	a	b
BW	0,94 (0,63)	6,00 (0,58)	0,66 (1,08)	0,07 (0,04)
	p	q	a	b
BGL	14,64 (2,95)	2,27 (1,17)	2,13 (1,33)	6,00 (2,31)

As estatísticas AIC, BIC, HQC e CAIC para as distribuições BLo, Lo, BN, BW e BGL são apresentadas na Tabela 4.4. Estes resultados indicam que a distribuição *BLo* tem o menor valor em cada estatística entre as distribuições ajustadas, assim, esta distribuição apresenta o melhor ajuste.

Tabela 4.4: Estatísticas AIC, BIC, CAIC e HQC das distribuições BLo, Lo, BN, BW e BGL para dados fibra de vidro

Distribuições	AIC	BIC	CAIC	HQC
BLo	31,11	51,95	31,15	38,92
Lo	40,04	60,87	40,07	39,94
BN	39,17	60,00	39,20	46,97
BW	36,62	57,45	36,65	44,42
BGL	45,05	53,62	45,74	42,74

A Figura 4.5 mostra o ajuste das funções de densidade das distribuições BLo, Lo, BN, BW e BGL. Observamos que o melhor ajuste é apresentado pela distribuição BLo.

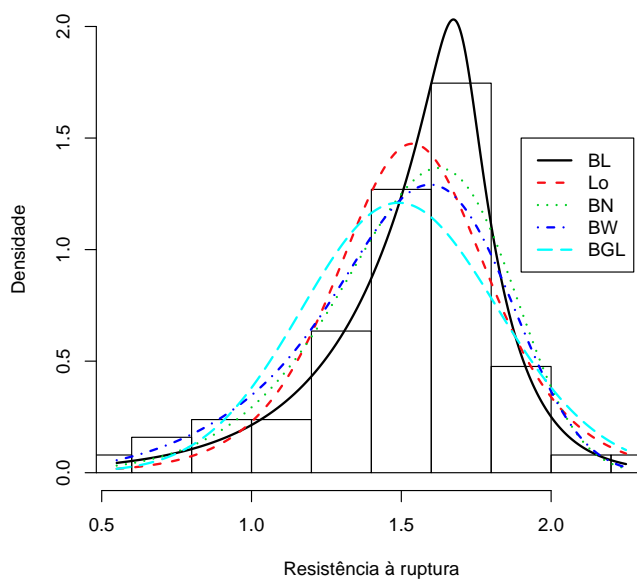


Figura 4.5: Função de densidade estimada para dados fibra de vidro

Os gráficos QQ das distribuições beta logística, logística, beta normal, beta Weibull e beta logística generalizada são apresentados na Figura 4.6. Podemos observar que há um melhor ajuste na distribuição *BLo* do que comparadas com as outras distribuições.

Para comparar distribuições, utilizamos a estatística da razão de verossimilhanças (RV) para modelos encaixados e estatística da razão de verossimilhança generalizada (RVG) para comparar modelos não encaixados.

Considerando as hipóteses $H_0 : Lo \text{ vs } H_A : BLo$, a estatística de RV assume o valor $w = 10,04$ ($p = 0,0001$), assim, a distribuição *BLo* é significativamente melhor nos ajuste dos dados.

Para comparar distribuições beta normal (BN), beta Weibull (BW) beta logística generalizada (BGL) cada uma delas versus a distribuição beta logística (*BLo*). Consideramos as hipóteses $H_0 : BN \text{ vs } H_A : BLo$, $H_0 : BW \text{ vs } H_A : BLo$ e $H_0 : BGL \text{ vs } H_A : BLo$, respectivamente. As estatísticas da RVG são dadas por: 5,50 ($p < 0,0001$), 3,25 ($p = 0,0006$) e 4,54 ($p < 0,0001$), respectivamente. Portanto, a distribuição *BLo* apresenta o melhor ajuste que as outras distribuições em estudo.

Para verificar que a distribuição *BLo* é mais adequada no ajuste dos dados, realizamos um teste de bondade do ajuste para ratificar este comportamento.

A Figura 4.7 mostra a função empírica e as funções de distribuição acumulada das distribuições *BLo*, *Lo*, *BN*, *BW* e *BGL*.

Para verificar quais modelos ajustam melhor os dados, aplicamos o teste não paramétrico K-S. As estatísticas do teste para cada distribuição são: *BLo* (KS=0,0987; $p = 0,5719$), *Lo* (KS=0,1305; $p = 0,2332$), *BN* (KS=0,1187; $p = 0,3375$), *BW* (KS=0,1559; $p = 0,0935$) e *BGL* (KS=0,191; $p = 0,0202$). Ao nível de 5% a única com ajuste insa-

tisfatório é a distribuição BGL.

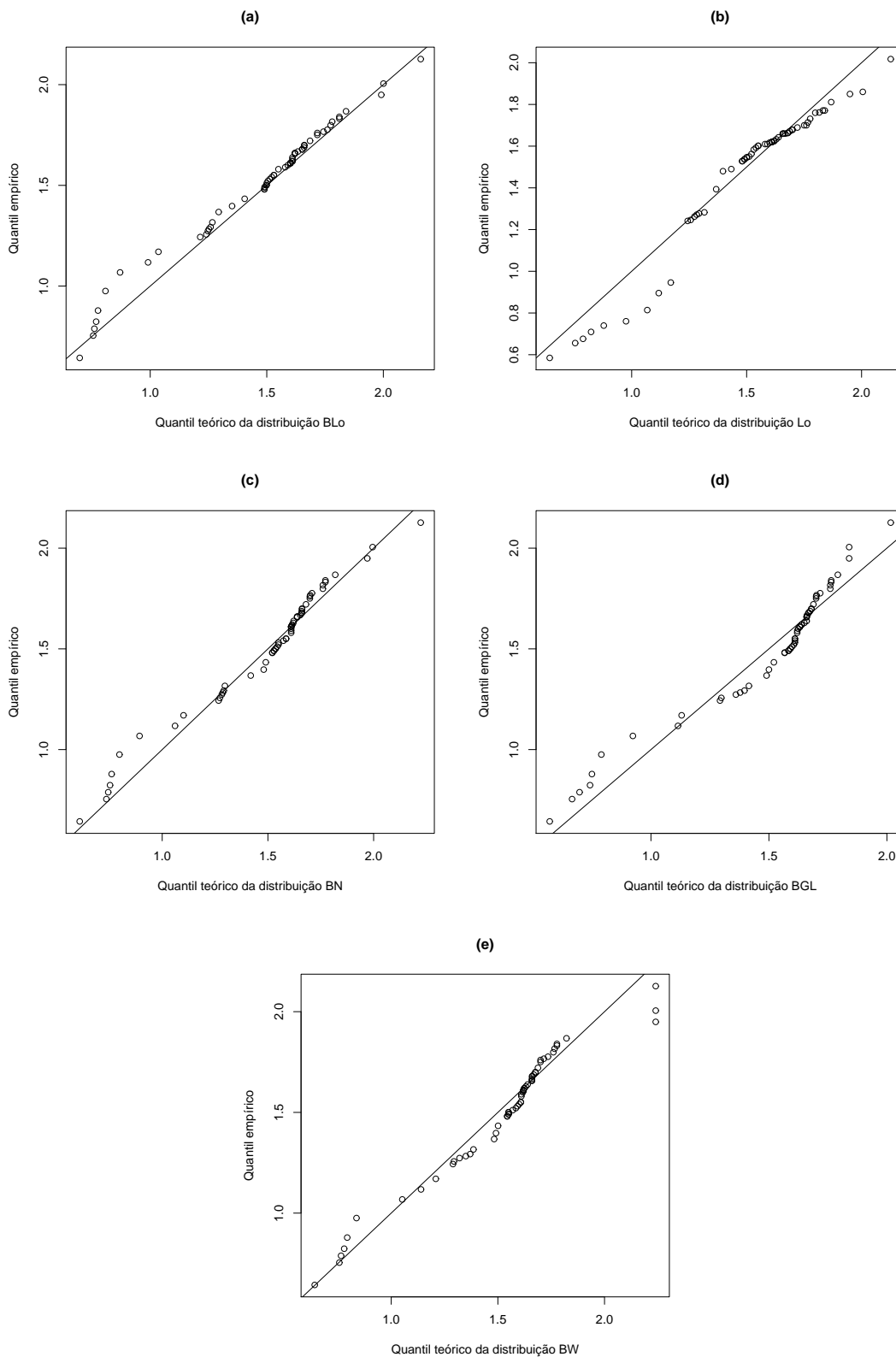


Figura 4.6: Gráfico QQ das distribuições (a)BLo, (b)Lo, (c)BN, (d)BW e (e)BGL

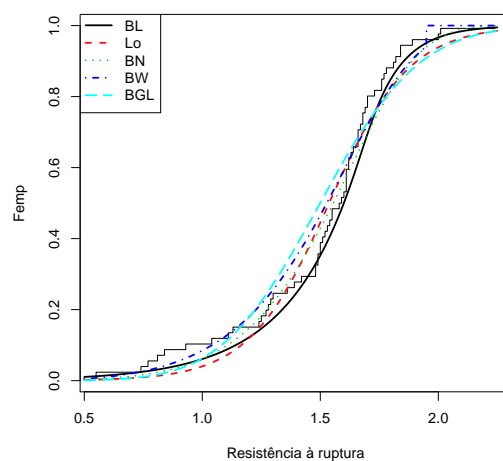


Figura 4.7: Gráfico função empírica e função de distribuição acumulada das distribuições BLo, Lo, BN, BW e BGL

Capítulo 5

Considerações finais

Neste trabalho foi proposta a distribuição beta logística a partir do trabalho de Eugene et al. (2002). Determinamos expressões matemáticas para algumas propriedades, tais como, função densidade de probabilidade, função de distribuição acumulada, função de risco, momentos, distribuições das estatísticas de ordem, medidas de tendência central, função geradora de momentos, entropias, os desvios médios, as curvas de Lorenz e Bonferroni. Examinamos o processo de estimação dos parâmetros aplicando o método de máxima verossimilhança e determinamos a matriz de informação esperada.

Duas aplicações da distribuição beta logística foram realizadas para mostrar a flexibilidade na análise dos dados. Foram realizados testes de razão de verossimilhanças e o teste da razão de verossimilhanças generalizado em modelos encaixados e não encaixados, respectivamente. Mostramos que a distribuição beta logística apresenta melhor ajuste se comparada com outras distribuições. Para ratificar esta afirmação aplicamos um teste de bondade do ajuste nos dados que comprova que a distribuição beta logística é mais adequada no ajuste do conjunto de dados.

Em termos de pesquisa futura, pretendemos propor a distribuição beta logística em modelos de regressão, caso seja possível.

Apêndice A

Propriedades da distribuição logística

Momentos

O n -ésimo momento da distribuição logística (Lo) pode ser obtido pela seguinte expressão

$$\tau = \int_{-\infty}^{\infty} x^n \frac{e^{-\frac{(x-\mu)}{s}}}{s[1 + e^{-\frac{(x-\mu)}{s}}]^2} dx.$$

Substituindo $t = \frac{x-\mu}{s}$, temos

$$\begin{aligned} \tau &= \int_{-\infty}^{\infty} (\mu + st)^n \left[\frac{e^{-t}}{s(1 + e^{-t})^2} \right] \\ &= \int_{-\infty}^{\infty} \sum_{j=0}^n \binom{n}{j} s^j t^j \mu^{n-j} \left[\frac{e^{-t}}{s(1 + e^{-t})^2} \right] dt \\ &= \sum_{j=0}^n \binom{n}{j} s^j \mu^{n-j} 2 \cdot \Gamma(j+1) \cdot [1 - 2^{-(j-1)}] \zeta(j), \end{aligned}$$

em que $\zeta(j) = \sum_{r=1}^{\infty} r^{-j}$ e $j > 1$ é chamada de função zeta de Riemann.

Esperança

A esperança de uma distribuição de probabilidade é definida por

$$E[x] = \int_{-\infty}^{\infty} x f(x) dx, \quad (5.1)$$

substituindo a equação (1.2) em (5.1), obtemos

$$E[x] = \int_{-\infty}^{\infty} \frac{x}{4s} \operatorname{sech}^2 \left(\frac{x-\mu}{2s} \right) dx$$

fazendo $t = \frac{x-\mu}{2s}$, temos

$$E[x] = \frac{1}{2}\mu \int_{-\infty}^{\infty} \operatorname{sech}^2 t \, dt + s \int_{-\infty}^{\infty} t \operatorname{sech}^2 t \, dt.$$

Portanto, a esperança da distribuição logística (Lo) é dada por

$$E(x) = \mu$$

.

Mediana

A mediana de uma distribuição de probabilidade é definida por

$$P(X \leq m) = P(X \geq m) = \int_m^{\infty} f(x) dx = 0,5$$

em que $f(x)$ é a função de probabilidade de uma distribuição contínua. No caso da distribuição Lo , temos

$$\int_m^{\infty} \frac{1}{4s} \operatorname{sech}^2 \left(\frac{x-\mu}{2s} \right) dx = 0,5$$

substituindo $t = \frac{x-\mu}{2s}$, obtemos

$$\int_{\frac{m-\mu}{2s}}^{\infty} \operatorname{sech}^2 t \, dt = 1.$$

Logo, a mediana da distribuição Lo é representada por

$$m = \mu.$$

Moda

A moda de uma distribuição é obtida pela maximização do logaritmo da função densidade (1.1), então

$$f'(x) = -\frac{1}{s} + \frac{2e^{-\frac{(x-\mu)}{s}}}{s[1 + e^{-\frac{(x-\mu)}{s}}]}, \quad (5.2)$$

onde $f'(x)$ é a primeira derivada da f.d.p. da distribuição Lo .

Igualando a equação (5.2) a zero, temos a moda da distribuição Lo .

$$x = \mu.$$

Momento Central

O n -ésimo momento central pode ser expresso em termos da função quantil, isto é,

$$E[(x - \mu)^n] = \int_{-\infty}^{\infty} (x - \mu)^n dF(x). \quad (5.3)$$

A função quantil da distribuição Lo é dado por

$$F^{-1}(p; \mu, s) = \mu + s \log \left(\frac{p}{1-p} \right). \quad (5.4)$$

Introduzindo a equação (5.4) em (5.3), obtemos

$$\begin{aligned} E[(x - \mu)^n] &= \int_0^1 [F^{-1}(p) - \mu]^n dp \\ &= s^n \int_0^1 \left\{ \log \left[\left(\frac{p}{1-p} \right) \right] \right\}^n dp \\ &= s^n \pi^n (2^n - 2) |B_n|, \end{aligned} \quad (5.5)$$

onde B_n são números de Bernoulli.

Variância

Aplicando a equação (5.5) para encontrar a variância da distribuição Lo obtemos

$$E[(x - \mu)^2] = s^2 \pi^2 (2^2 - 2) |B_2| = \frac{\pi^2 s^2}{3}.$$

Assimetria

A assimetria de distribuição Lo é dada por

$$E \left[\frac{(x - \mu)}{\sigma} \right]^3 = 0,$$

porque os B_n ímpares são nulos, se $n \neq 1$.

Curtose

A curtose da distribuição Lo é representada como

$$\begin{aligned} E \left[\frac{(x - \mu)}{\sigma} \right]^4 - 3 &= \frac{s^4 \pi^4 (2^4 - 2) |B_4|}{\left(\frac{\pi^2 s^2}{3}\right)^2} - 3 \\ &= \frac{s^4 \pi^4 (2^4 - 2) \left(\frac{1}{30}\right)}{\frac{\pi^4 s^4}{9}} - 3 = \frac{6}{5}. \end{aligned}$$

Apêndice B

Derivadas parciais de segunda ordem em relação aos parâmetros da distribuição beta logística

$$\begin{aligned} \frac{\partial^2 l}{\partial \mu^2} &= -\frac{\pi^2(a+b)e^{-t}}{3\sigma^2(1+e^{-t})^2}, \\ \frac{\partial^2 l}{\partial \mu \partial \sigma} &= -\frac{b\pi}{\sigma^2\sqrt{3}} - \frac{\pi(a+b)}{\sqrt{3}} \left\{ -\frac{e^{-t}}{\sigma^2(1+e^{-t})} + \frac{t}{\sigma^2} \left[\frac{e^{-t}}{(1+e^{-t})^2} \right] \right\}, \\ \frac{\partial^2 l}{\partial \mu \partial a} &= -\frac{\pi}{\sigma\sqrt{3}} \left[\frac{e^{-t}}{(1+e^{-t})} \right], \\ \frac{\partial^2 l}{\partial \mu \partial b} &= -\frac{\pi}{\sigma\sqrt{3}} \left[\frac{e^{-t}}{(1+e^{-t})} \right], \\ \frac{\partial^2 l}{\partial \sigma^2} &= \frac{1}{\sigma^2} - \frac{bt}{\sigma^2} - \frac{\pi(a+b)}{\sqrt{3}} \left[-\frac{2te^{-t}}{\sigma^2(1+e^{-t})} - \frac{\sqrt{3}}{\sigma^2} \left(\frac{t}{1+e^{-t}} \right)^2 \right], \\ \frac{\partial^2 l}{\partial \sigma \partial a} &= -\frac{t}{\sigma} \left(\frac{e^{-t}}{1+e^{-t}} \right), \\ \frac{\partial^2 l}{\partial \sigma \partial b} &= -\frac{t}{\sigma} \left(\frac{e^{-t}}{1+e^{-t}} \right), \\ \frac{\partial^2 l}{\partial a^2} &= -\psi'(a) + \psi'(a+b), \\ \frac{\partial^2 l}{\partial b^2} &= -\psi'(b) + \psi'(a+b), \\ \frac{\partial^2 l}{\partial a \partial b} &= \psi'(a+b). \end{aligned}$$

Referências

AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716-722, 1974.

CORDEIRO, G.M.; BRITO, R.S. The beta power distribution. *Brazilian Journal of Probability and Statistics*, **26**, 88-112, 2012.

CORDEIRO, G.M.; SIMAS, A.B.; STOSIC, B.D. The beta Weibull distribution. Universidade Federal Rural de Pernambuco, 2010.

EUGENE, N; LEE, C; FAMOYE, F. The beta-normal distribution and its applications. *Communication in Statistics - Theory and Methods*, **31**, 497-512, 2002.

FISCHER, M.; VAUGHAN, D. The beta-hyperbolic secant distribution. University of Erlangen, 2004.

GRADSHTEYN, I.S.; RYZHIK, I.M. Table of integrals, series, and products. San Diego: Academic Press, 2004.

HANNAN, E.J.; QUINN, B.G. The determination of order of an autoregression. *Journal of the Royal Statistical Society, B*, **41**, 190-195, 1979.

HOSKING, J.R.M. L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society, Series B*, Oxford, **52**, 105-124, 1990.

LIANG, H.; ZOU, G. Improved AIC selection strategy for survival analysis. *Computational Statistics and Data analysis*, **52**, 2538-2548, 2007.

McDONALD, J.B.; XU, Y.J. A generalization of the beta distribution with applications. *Journal of econometrics*, **66**, 133-152. 1995.

MORAIS, A.L.; CORDEIRO, G.M.; CYSNEIROS, A.H.M.A. The beta generalized logistic distribution. Submitted, *Brazilian Journal of Probability and Statistics*, 2010.

NADARAJAH, S; KOTZ, S. The beta Gumbel distribution. *Mathematical Problems in Engineering*, **10**, 323-332, 2004.

NADARAJAH, S; KOTZ, S. The beta exponential distribution. *Reliability Engineering and System Safety*, **91**, 689-697, 2005.

PERKS, W. On some experiments in the graduation of mortality statistics. *J. Inst. Actuaries*, **63**, 12-57, 1932.

PRENTICE, R. L. Discrimination among some parametric models. *Biometrika*, **62**, 607-614, 1975.

R Development Core Team, *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012.

REED, L.J.; BERKSON, J. The application of the logistic function to experimental data. *Journal of Physical Chemistry*, **33**, 760-779, 1929.

RENYI, A. On measures of information and entropy. *Statistics and Probability*, 547-561, 1960.

SCHULTZ, H. The standard error of a forecast from a curve. *Journal of the American Statistical Association*, **25**, 139-185, 1930.

SCHWARZ, G. E. Estimating the dimension of a model. *Annals of Statistics*, **6**, 461-464, 1978.

SHANNON, C.E. A mathematical theory of communication. *Bell System Technical Journal*, **27**(III), 379-423, 623-656, 1948.

SMITH, R.L.; NAYLOR, J.C. A comparison of maximum likelihood and Bayesian estimators for three parameter Weibull distribution. *Applied Statistics*, **36**, 358-369, 1987.

VUONG, Q.H. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, **57**, 307-333, 1989.