



Universidade Estadual da Paraíba
Centro de Ciências e Tecnologia
Departamento de Estatística

Geovano Francisco Felipe

**A Distribuição Marshall-Olkin Exponencial Generalizada:
Uma Nova Generalização Para Distribuições Univariadas**

Campina Grande
Fevereiro de 2014.

Geovano Francisco Felipe

A Distribuição Marshall-Olkin Exponencial Generalizada: Uma Nova Generalização Para Distribuições Univariadas

Trabalho de Conclusão de Curso a ser apresentado como requisito para à conclusão do curso de bacharelado em Estatística do Departamento de Estatística do Centro de Ciência e Tecnologia da Universidade Estadual da Paraíba para obtenção do título de bacharel em Estatística.

Orientador:

Kleber Napoleão Nunes de Oliveira Barros

Campina Grande
Fevereiro de 2014.

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL - UEPB

F313d Felipe, Geovano Francisco.
A distribuição Marshall-olkin Exponencial Generalizada
[manuscrito] : uma nova generalização para distribuição univariadas /
Geovano Francisco Felipe. - 2014.
45 p. : il. color.

Digitado.
Trabalho de Conclusão de Curso (Graduação em Estatística) -
Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia,
2014.

"Orientação: Prof. Me. Kleber Napoleão Nunes de Oliveira
Barros, Departamento de Estatística".

1. Distribuição Generalizada. 2. Marshall-olkin Exponencial
Generalizada. 3. Estatística. I. Título.

21. ed. CDD 519.5

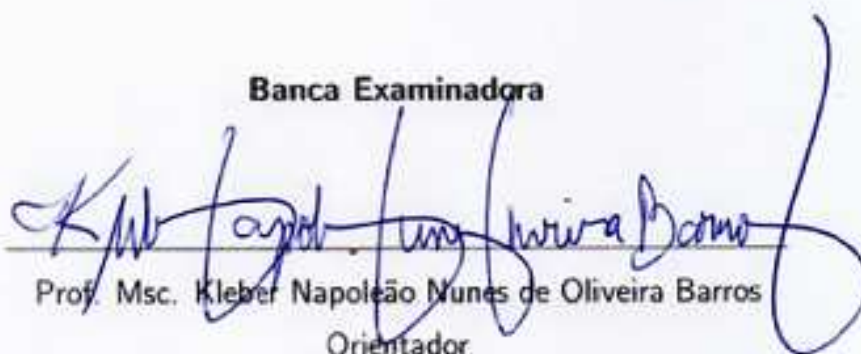
Geovano Francisco Felipe


A Distribuição Marshall-Olkin Exponencial Generalizada: Uma Nova Generalização Para Distribuições Univariadas


Trabalho de Conclusão de Curso a ser apresentado como requisito para à conclusão do curso de bacharelado em Estatística do Departamento de Estatística do Centro de Ciência e Tecnologia da Universidade Estadual da Paraíba para obtenção do título de bacharel em Estatística.

Aprovado em: 11 / 03 / 2014

Banca Examinadora


Prof. Msc. Kleber Napoleão Nunes de Oliveira Barros
Orientador.


Prof. Dr. Tiago Almeida de Oliveira
Universidade Estadual da Paraíba


Prof. Dr. Ricardo Alves de Olinda
Universidade Estadual da Paraíba

Agradecimentos

Agradeço ao Senhor Deus por ter iluminado meu caminho para que eu pudesse concluir mais uma estapa de minha vida.

Aos meus pais, Severino Francisco e Alzira Amaro, por todo amor e dedicação que sempre teve comigo, não deixando desistir e me mostrando que sou capaz de chegar onde desejo, sem dúvida foi quem me deu o maior incentivo para conseguir concluir esse trabalho.

Aos meus irmãos, Jardane e Jardoniél que sempre me apoiaram em todos os momentos, enfim por todos os conselhos e pela confiança, meu imenso agradecimento.

A todos os professores que me acompanharam durante a graduação, e em especial ao meu orientador Pro^o Msc. Kleber Napoleão Nunes de Oliveira Barros pelo ensinamento e dedicação no auxílio à concretização dessa monografia.

Campina Grande
Fevereiro de 2014.

Resumo

Distribuições de suporte positivo desempenham um papel central na análise de dados de vida, de sobrevivências e taxas de risco. É comum o uso da distribuição gama, log-normal ou de Weibull para dados de tempo de falhas. Estas distribuições têm, em geral, um ou dois parâmetros. O ajuste dos dados para essas distribuições é possível de ser realizado diretamente por um processo analítico como o método de estimação da máxima verossimilhança ou dos momentos. No entanto, para certos conjuntos de dados, o ajuste de tais distribuições não se adequa. Com o advento da computação de alto desempenho e de ferramentas de computação analítica muitos modelos novos vem sendo desenvolvidos. Em particular, nos últimos dez anos houve uma explosão de distribuições generalizadas. Neste contexto, distribuições mais flexíveis com a incorporação de um ou dois parâmetros vem sendo cada vez mais exploradas, denominando-a de família de distribuição Marshal-Olkin generalizada. Neste trabalho, pretende-se estudar o comportamento das família de distribuições que generaliza duas famílias concorrentes, estudando suas principais relações, tais como função de sobrevivência, de risco, densidade, momentos log-verossimilhança e matriz de informação de Fisher, implementando-os no *software* estatístico *R*. Em particular, obtém-se estas funções a partir da distribuição exponencial, a distribuição Marshal-Olkin exponencial generalizada (MOEG), uma distribuição tripamétrica competindo com a distribuição beta exponencial e distribuição gama generalizada e de Weibull generalizada. Os resultados revelam que a função de risco distribuição MOEG apresenta, por exemplo, comportamentos satisfatórios, tais como o formato de U e de U invertido, que são típicos de distribuições flexíveis.

Palavras-chave: Distribuições Generalizadas, Distribuições de Marshall-Olkin

Abstract

Support positive distributions play a central role in the life data analysis, employing and spreads. It is common to use the gamma distribution, Weibull or lognormal for failure time data. These distributions have, in General, one or two parameters. The fit of the data to these distributions can be accomplished directly by an analytical process as the method of maximum likelihood or moments. However, for certain data sets, the set of such distributions do not suit. With the advent of high performance computing and analytical computing tools many new models has been developed. In particular, over the last ten years there has been an explosion of generalized distributions. In this context, the most flexible distributions with the incorporation of one or two parameters has been increasingly exploited, calling it family of Marshall-General Olkin distribution. In this work, the aim is to get a new family of distributions that generalizes two competing families, studying its main relationships, such as survival, risk function, density, log-likelihood moments and Fisher information matrix, implementing them in Statistical *software* R. in particular, these functions from the exponential distribution, the Marshal-General exponential Olkin distribution (MOEG), a distribution triparamétrica competing with the exponential beta distribution and gamma distribution Weibull and generalized. The results show that the risk function MOEG distribution features, for example, satisfactory behaviour, such as the format of U and U inverted, that are typical of flexible distributions.

Keywords: Generalized distributions, distributions of Marshall-Olkin

Sumário

Lista de Tabelas	p. 10
1 Introdução	p. 11
2 Desenvolvimento	p. 12
2.1 Distribuição Generalizada Weibull	p. 12
2.2 Marshall-Olkin exponencial Generalizada	p. 13
2.3 Estimação de Máxima Verossimilhança	p. 20
2.4 Critério de informação de Akaike - AIC	p. 21
2.5 Critério de informação bayesiano - BIC	p. 21
2.6 Teste de Wald	p. 22
2.7 Teste da Razão de Verossimilhanças - TRV	p. 22
2.8 Aplicação	p. 23
3 Resultados e Discussão	p. 24
3.1 Expansão da Densidade	p. 25
3.2 Momentos e Função Geradora de Momentos	p. 26
3.3 Função Quantílica	p. 27
3.4 Mediana	p. 28
3.5 Desvio Médio e Desvio Mediano	p. 29
3.6 Aplicações	p. 29
Conclusão	p. 35
Referências Bibliográficas	p. 36

Apêndice A: Elementos da Matriz Observadas de Fisher	p. 39
Apêndice B: Código R utilizado para fazer os ajustes.	p. 40

Lista de Tabelas

- 3.1 Tamanho amostral(n), médis, variância, coeficiente de variação, mínimo, mediana e máximo dos conjuntos de dados trabalhados p. 30
- 3.2 Estimativas dos parâmetros, erros-padrão, estatística de Wald, MOGE1, MOGE2, MOGE3, gama, de Weibull e log-normal para o conjunto de dados de magnitudes de terremotos. p. 30
- 3.3 Critério da Informação de Akaike (AIC) e Bayesiano (BIC) para as distribuições estudadas ajustadas aos dados de magnitudes de terremotos. . . . p. 31
- 3.4 Log-verossimilhança e estatísticas do testa de razão de verossimilhanças para o modelo proposto contra seus submodelos. p. 32
- 3.5 Estimativas dos parâmetros, erros-padrões, estatística de Wald, MOGE1, MOGE2, MOGE3, gama, log-normal para o conjunto de dados de alumínio . p. 32
- 3.6 Critério da Informação de Akaike (AIC) e Bayesiano (BIC) para as distribuições estudadas ajustadas aos dados de tempos de falha de alumínio . . p. 33
- 3.7 Log-verossimilhança e estatísticas do testa de razão de verossimilhanças para o modelo proposto contra seus submodelos. p. 34

1 Introdução

A estatística paramétrica é o principal e mais promissor ramo da inferência estatística. Desde os tempos de De Moivre e Gauss, e mais recentemente Karl Pearson, Fisher e Weibull modelos probabilísticos e métodos de estimação tem sido produzidos de forma cada vez mais intensa. Nos últimos dez anos a quantidade de novos modelos paramétricos vem aumentando com o maior acesso à ferramentas computacionais. Distribuições como normal, exponencial, gama, beta, de Laplace, normal-inversa, de Rayleigh, de Weibull, Gumbel, Weibull estendida, Fréchet, entre outras vem sendo generalizadas pela adição de parâmetros que possibilitam uma maior flexibilidade. Em particular, para dados de sobrevivência (ou taxa de falha) a corrida tem sido grande, haja vista a grande aplicabilidade de distribuições que atendam tais características. Nomes como, Gupta, Kundu, Nadarajah, Kotz, Balakrishnan, Cordeiro, Ortega, Marshall, Olkin, Lemonte, Barreto-Souza, Jones, entre outros tem contribuído de forma significativa neste campo. Através da família de distribuição generalizada de Marshall-Olkin pretende-se estudar seus submodelos e comparar os resultados com as distribuição concorrentes, verificando os critérios de adequação de modelos, via dados reais, via dados simulados tem por objetivo obter as funções de distribuição de densidade de probabilidade, função de sobrevivência e a função de risco. Encontrar as funções geradora de momentos; obter as expressão do tempo de sobrevivência médio e tempo de vida real.

2 Desenvolvimento

2.1 Distribuição Generalizada Weibull

MUDHOLKAR *et al.* (1995) propuseram a distribuição Weibull exponencializada. GUPTA e KUNDU (1999) obtiveram as relações para um caso particular, denominando-o de distribuição exponencial generalizada (EG) como alternativas às distribuições gama e de Weibull. Por sua vez, diversos trabalhos em análise de sobrevivência foram produzidos com a EG desde então RAQAB (2002), RAQAB e AHSANULLAH (2001), ZHENG (2002), SARHAN (2007), GUPTA e KUNDU (2008), ACHCAR e BOLETA (2009). NADARAJAH (2011) revisa a distribuição EG acrescentando novos resultados. Seguindo a mesma linha, diversos outros trabalhos obtiveram importantes generalizações de distribuições de tempos de vida pela adição de parâmetros. O princípio de tais distribuições é simples, dada uma distribuição com função de distribuição contínua $G(y)$, sua generalização ou exponencialização é obtida por

$$F(y; \alpha) = G(y)^\alpha, \quad y \in \mathbb{R}, \alpha > 0. \quad (2.1)$$

Quando $G(y) = 1 - e^{-\lambda y}$ e $G(y) = 1 - e^{-(\lambda y)^\gamma}$, tem-se a distribuição exponencial generalizada e Weibull generalizada, respectivamente.

As distribuições da família beta foram obtidas por NADARAJAH e KOTZ (2006) com a inserção da função de distribuição acumulada da distribuição de exponencial na integral da distribuição beta, formando-se a distribuição beta exponencial (BE). Assim, seja uma distribuição com função de distribuição contínua $G(y)$, a distribuição beta-G será

$$F(y; a, b) = \frac{1}{B(a, b)} \int_0^{G(y)} w^{a-1} (1-w)^{b-1} dw, \text{ com } y \in \mathbb{R}, a > 0 \text{ e } b > 0, \text{ em que}$$

$$B(a, b) = \int_0^1 w^{a-1} (1-w)^{b-1} dw.$$

Esta classe de distribuições tem se desenvolvido bastante nos últimos anos com os trabalhos de JONES (2004) e CORDEIRO *et al.* (2011). Um inconveniente das distribuições beta-G é

que elas são funções da distribuição beta incompleta, dada por

$$B(a, b)_y = \int_0^y w^{a-1} (1-w)^{b-1} dw,$$

que não tem forma analítica fechada, sendo portanto necessário algum método numérico para obtenção dos resultados. Para contornar esta situação, uma alternativa JONES (2008), CORDEIRO e CASTRO (2011) é utilizar a distribuição de Kumaraswamy, definindo-se as distribuições K-G por

$$F(y; a, b) = 1 - (1 - G(y))^a)^b, \quad y \in \mathbb{R}, a > 0, b > 0,$$

pois esta leva a uma função de distribuição analiticamente tratável para uma dada distribuição primitiva G .

Por outra linha, Marshall e Olkin (1997) propõem o seguinte método de adição de parâmetros para uma dada distribuição G ,

$$F(y; \beta) = \frac{G(y)}{G(y) + \beta(1 - G(y))}, \quad y \in \mathbb{R}, \beta > 0. \quad (2.2)$$

PREDA *et al.* (2011) utilizam a família de Marshall e Olkin para obter, de forma condicional, a distribuição exponencial-Poisson modificada. CORDEIRO e LEMONTE (2012) utilizam de uma extensão da distribuição de Weibull, pela introdução de um parâmetro, para obter a distribuição Marshall-Olkin Weibull Estendida, com

$$G(y) = 1 - e^{-\lambda y^\delta}, \quad y \in \mathbb{R}, \lambda > 0, \delta > 0.$$

2.2 Marshall-Olkin exponencial Generalizada

No presente trabalho, introduzimos um segundo parâmetro à família Marshall-Olkin, ficando com a forma mais geral

$$F(y; \alpha, \beta) = \frac{G(y)^\alpha}{G(y)^\alpha + \beta(1 - G(y)^\alpha)}, \quad y \in \mathbb{R}, \alpha > 0, \beta > 0. \quad (2.3)$$

Naturalmente, esta nova classe engloba a classe de distribuições de Marshall-Olkin (2.2) quando $\alpha = 1$ e a família de distribuições generalizadas (2.1) quando $\beta = 1$, o que é bastante conveniente a fim de comparação das propriedades. Se $G(y) = 1 - e^{-\lambda y}$, definimos a distribuição Marshall-Olkin exponencial generalizada (MOEG), como uma distribuição tri-paramétrica competindo com a distribuição beta exponencial e distribuição Weibull generalizada.

Seja a Equação (2.3) da família de distribuições de Marshall-Olkin generalizada, obtida pela adição de um expoente $\alpha > 0$ em cada termo $G(y)$ na Equação (2.2). Escrevendo-se $G(y) = 1 - e^{-\lambda y}$, com $y > 0, \lambda > 0$, obtém-se a seguinte função

$$F(y; \alpha, \beta, \lambda) = \frac{(1 - e^{-\lambda y})^\alpha}{(1 - e^{-\lambda y})^\alpha + \beta [1 - (1 - e^{-\lambda y})^\alpha]}, \quad (2.4)$$

que é a função de distribuição para a distribuição de probabilidade Marshall-Olkin exponencial generalizada. Na Figura 2.1 alguns gráficos para a função de distribuição são construídos com α, β e λ variando de $1/2$ para 2. Mesmo para este conjunto pequeno de possibilidades, é possível observar curvaturas negativas (para $\alpha = 0,5, \beta = 0,5, \lambda = 0,5$, por exemplo) e mudanças de concavidade (para $\alpha = 2, \beta = 0,5, \lambda = 0,5$, por exemplo).

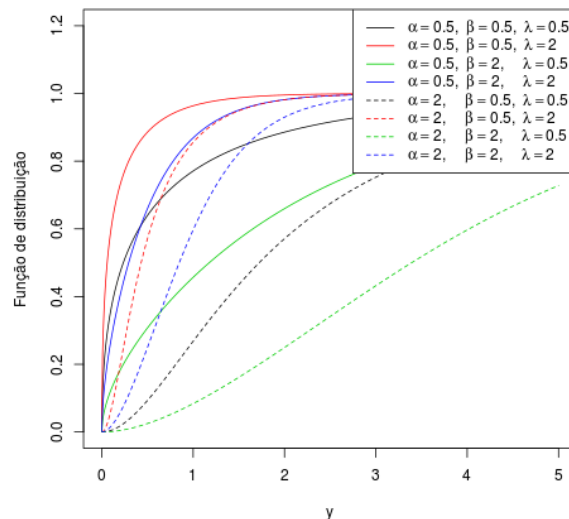


Figura 2.1: Alguns gráficos para a função de distribuição Marshall-Olkin exponencial generalizada.

Derivando a Equação (2.3) em relação a y , obtém-se a respectiva função densidade de probabilidade

$$f(y; \alpha, \beta) = \frac{\alpha \beta G(y)^{\alpha-1} g(y)}{\{1 - \beta [1 - G(y)^\alpha]\}^2}$$

Escrevendo $G(y) = 1 - \exp(-\lambda y)$ e $g(y) = \lambda \exp(-\lambda y)$, com $y > 0, \lambda > 0$, obtém-se a função densidade de probabilidade da distribuição Marshall-Olkin exponencial generalizada

$$f(y; \alpha, \beta, \lambda) = \frac{\alpha \beta \lambda (1 - e^{-\lambda y})^{\alpha-1} e^{-\lambda y}}{\{1 - \beta [1 - (1 - e^{-\lambda y})^\alpha]\}^2}. \quad (2.5)$$

O resultado (2.5) pode ser obtido diretamente pela derivada de (2.4) em relação a y . Na Figura

2.2 é possível se observar que para $\alpha < 1$ se tem um comportamento típico da distribuição exponencial para a distribuição MOEG. Para $\alpha \geq 1$ temo um comportamento unimodal, tal como as distribuições gama e de Weibull.

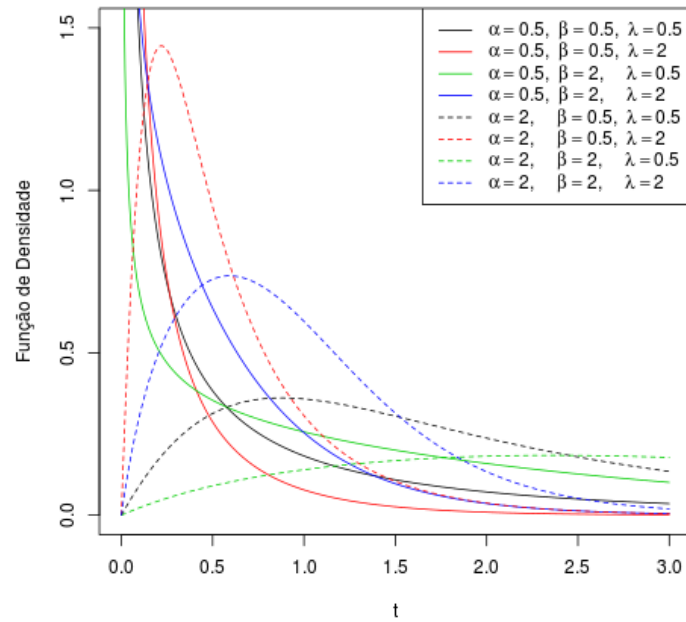
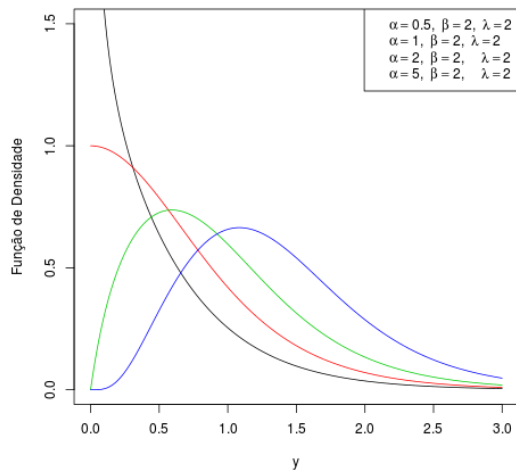
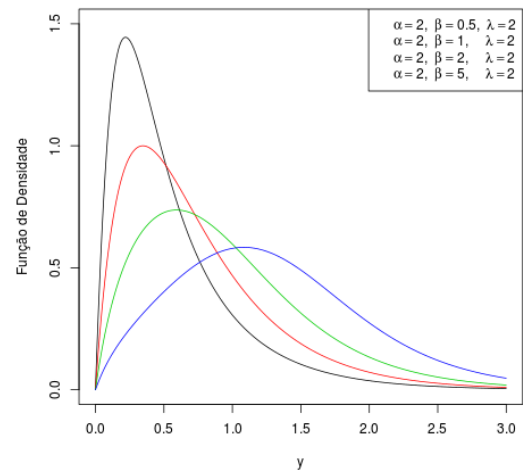


Figura 2.2: Alguns gráficos para a função densidade de probabilidade da distribuição Marshall-Olkin exponencial generalizada.

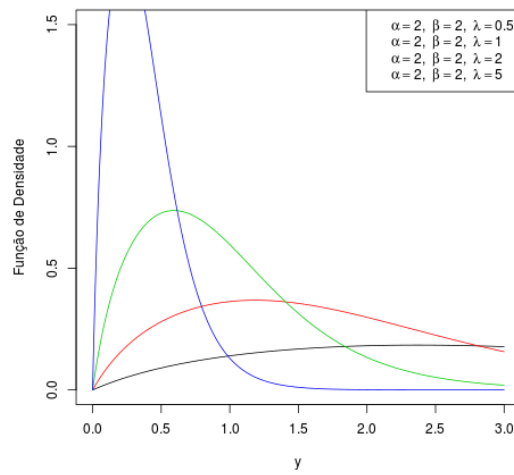
Na Figura 2.3.a são construídas as função densidade de probabilidade da distribuição MOEG fixando $\beta = \lambda = 2$ e variando $\alpha = 1/2, 1, 2, 5$. Assim, α é um parâmetro que modifica a posição da média. O parâmetro α também altera a forma da distribuição, conforme já mencionado. Tem-se, ainda, que para $\alpha = 1$ (linha vermelha) a distribuição fica reduzida ao sub-modelo: Marshal-Olkin exponencial. O parâmetro β também altera a posição da média, embora não altere a forma da distribuição, conforme Figura 2.3.b, em que se fixou $\alpha = \lambda = 2$ e variou-se $\beta = 1/2, 1, 2, 5$. Quando $\beta = 1$ (linha vermelha) a distribuição fica reduzida ao sub-modelo: exponencial generalizada. O parâmetro λ altera significativamente a dispersão da densidade, conforme se observa na Figura 2.3.c, na qual se variou $\lambda = 1/2, 1, 2, 5$ e se fixou os demais parâmetros.



(a)



(b)



(c)

Figura 2.3: Alguns gráficos para a função densidade de probabilidade da distribuição MOEG (a) variando α , (b) variando β e (c) variando λ .

A função de sobrevivência é uma função muito utilizada em Teoria da Confiabilidade e Análise de Sobrevivência. Sua expressão é obtida pelo complemento da acumulada, isto é, $S(t) = 1 - F(t)$, $t \geq 0$. Assim, para a classe de distribuições Marshal-Olkin generalizada

$$S(t) = \frac{\beta \overline{G(t)}^\alpha}{G(y)^\alpha + \beta \overline{G(t)}^\alpha}, \quad (2.6)$$

em que $\overline{X(y)} = 1 - X(y)$, $\alpha > 0$ e $\beta > 0$. Esta expressão é, exatamente, a expressão (1) encontrada em (CORDEIRO e LEMONTE (2012) substituindo $G(t)^\alpha$ por $G(t)$ e β por α , tal como notação empregada. Fazendo $G(t) = 1 - e^{-\lambda t}$, $\lambda > 0$, chega-se a seguinte expressão

$$S(t; \alpha, \beta, \lambda) = \frac{\beta [1 - (1 - e^{-\lambda t})^\alpha]}{(1 - e^{-\lambda t})^\alpha + \beta [1 - (1 - e^{-\lambda t})^\alpha]}. \quad (2.7)$$

Na Figura 2.4 gráficos para a função de sobrevivência são construídos com α , β e λ variando de 1/2 para 2. É perceptível que o parâmetro α tem uma relação direta com a taxa de sobrevivência, isto é, se t representa o tempo até a falha de componentes (ou a morte por câncer de um paciente), então se α for crescente, a porcentagem de componentes funcionando tende a ser grande, em contrapartida se α for decrescente a mesma porcentagem deve ser baixa. As mesmas conclusões são obtidas para o parâmetro β fixando-se um valor de α . No entanto, λ tem uma relação inversa com a taxa de sobrevivência. Assim, para α e β fixos, o aumento de λ acarreta na diminuição no número de sobreviventes.

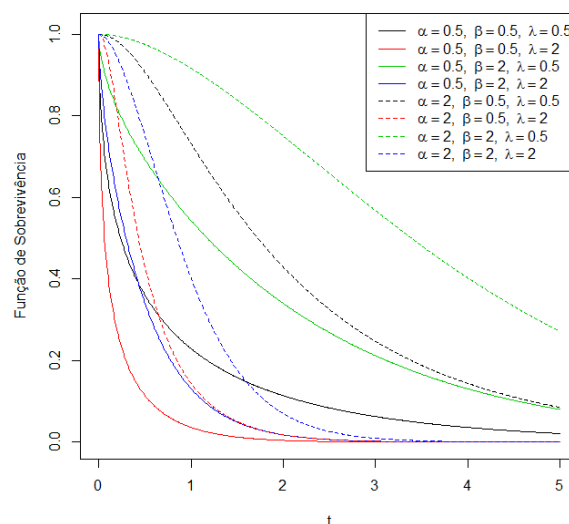


Figura 2.4: Alguns gráficos para a função de sobrevivência da distribuição Marshall-Olkin exponencial generalizada.

A função de risco ou função de taxa de falha é uma função muito importante em análise

de sobrevivência, pois um gráfico crescente em relação ao tempo desta função indica que a probabilidade de morte dos indivíduos aumenta com o passar do tempo, ao passo que um gráfico decrescente indica que o risco de morte diminui. Ela é definida como a razão entre a função densidade de probabilidade e a função de sobrevivência da distribuição em questão. Para a distribuição MOEG a função de risco é a razão entre (2.5) e (2.7), isto é,

$$h(t; \alpha, \beta, \lambda) = \frac{\alpha \lambda (1 - e^{-(\lambda t)})^{\alpha-1} e^{-(\lambda t)}}{\{(1 - e^{-(\lambda t)})^\alpha + \beta [1 - (1 - e^{-(\lambda t)})^\alpha]\} [1 - (1 - e^{-(\lambda t)})^\alpha]}. \quad (2.8)$$

em que foi substituído y por $t > 0$ na Equação (2.5). Na Figura 2.5a gráficos para a função de sobrevivência são construídos com α , β e λ variando entre $1/2, 1$ e 2 . É perceptível que o parâmetro λ é a assíntota da função de risco, isto é, quando t aumenta (os dados envelhecem) o risco de falha tende para λ . A função de risco da distribuição MOEG tem formas crescentes, decrescentes e constante. Nos seres humanos, em geral, o risco de morte é alto na infância, aproximadamente constante na fase adulta e alto para indivíduos idosos. Este comportamento decrescente-constante-crescente é conhecido na literatura como *curva da banheira* (COLOSIMO; GIOLO, 2006). Assim, é importante que as distribuições de sobrevivência consigam gerar tal comportamento. Na Figura 2.5b é possível observar que a distribuição MOEG é capaz de simular esta situação da curva de banheira para, por exemplo $\alpha = 0,6; \beta = 40$ e $\lambda = 8$ (curva preta), bem como o comportamento de banheira invertida: $\alpha = 1,2; \beta = 0,4$ e $\lambda = 4$ (curva vermelha). Estes resultados mostram que a distribuição Marshal-Olkin exponencial generalizada é bastante flexível.

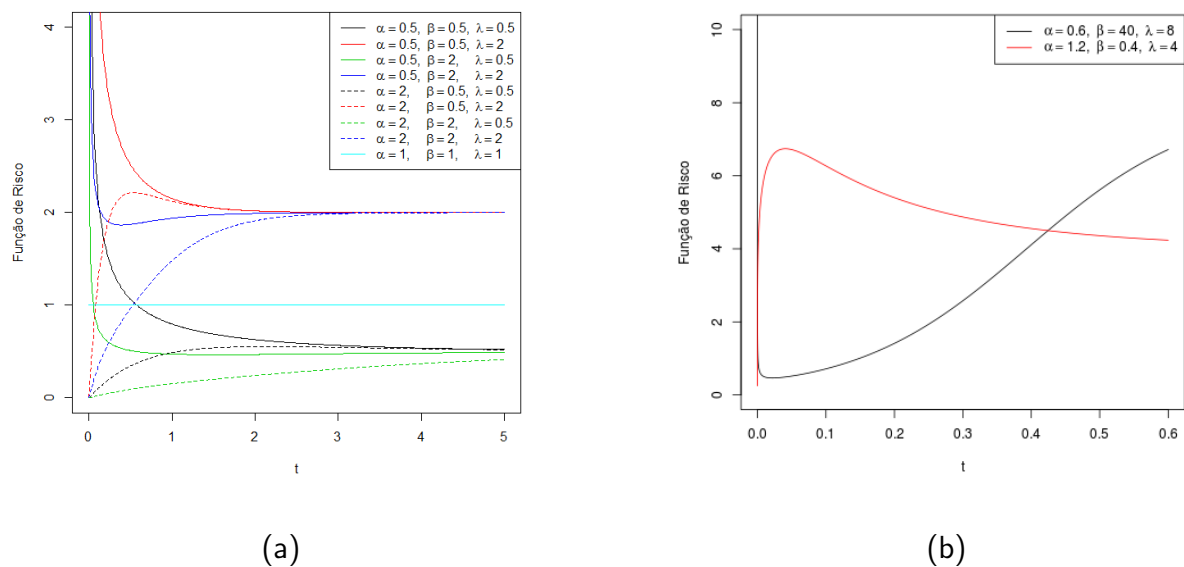


Figura 2.5: Alguns gráficos para a função de risco da distribuição MOEG (a); comportamento de banheira e de banheira invertida (b).

2.3 Estimação de Máxima Verossimilhança

Sejam Y_1, Y_2, \dots, Y_n denotando uma amostra aleatória de tamanho n de distribuição $Y \sim \mathcal{D}(\theta)$. A função de máxima verossimilhança é dada por

$$L(\theta) = \prod_{i=0}^n f(y_i; \theta).$$

Em particular, se Y_1, Y_2, \dots, Y_n tem distribuição $Y \sim MOEG(\alpha, \beta, \lambda)$, com $\theta = \langle \alpha, \beta, \lambda \rangle$ a função de máxima verossimilhança é dada por:

$$L(\alpha, \beta, \lambda) = \frac{\alpha^n \beta^n \lambda^n \prod_{i=0}^n (1 - e^{-\lambda y_i})^{\alpha-1} e^{-\lambda y_i}}{\prod_{i=0}^n \{1 - \bar{\beta} [1 - (1 - e^{-\lambda y})^\alpha]\}^2}, \quad (2.9)$$

em que $\bar{\beta} = 1 - \beta$. O logaritmo da função de máxima verossimilhança, denominado log-verossimilhança, é uma expressão computacionalmente menos complexa. Assim, para a função de densidade (2.5) com verossimilhança (2.9), a log-verossimilhança é dada por

$$\begin{aligned} \ell(\alpha, \beta, \lambda) &= n \log(\alpha \beta \lambda) - \lambda \sum_{i=0}^n y_i + (\alpha - 1) \sum_{i=0}^n \log(1 - e^{-\lambda y_i}) \\ &\quad - 2 \sum_{i=0}^n \log\{1 - \bar{\beta} [1 - (1 - e^{-\lambda y})^\alpha]\} \end{aligned} \quad (2.10)$$

O vetor escore é definido como $\mathbf{U}(\theta) = \langle \partial \ell(\theta) / \partial \theta_1, \partial \ell(\theta) / \partial \theta_2, \dots, \partial \ell(\theta) / \partial \theta_k \rangle$. Para a função (2.10), com $\ell = \ell(\alpha, \beta, \lambda)$, tem-se

$$\mathbf{U}(\theta) = \langle U_\alpha, U_\beta, U_\lambda \rangle$$

em que, por exemplo, $U_\alpha = \frac{\partial \ell}{\partial \alpha}$. As funções U_α , U_β e U_λ são

$$U_\alpha = -2 \sum_{i=1}^n (1 - \beta) u_i z_i + \sum_{i=1}^n z_i + \frac{n}{\alpha},$$

$$U_\beta = \frac{n}{\beta} - 2 \sum_{i=1}^n v_i$$

e

$$U_\lambda = -2 \sum_{i=1}^n \alpha (1 - \beta) u_i w_i + (\alpha - 1) \sum_{i=1}^n w_i - \sum_{i=1}^n y_i + \frac{n}{\lambda},$$

respectivamente, com as seguintes expressões: $z_i = \log(1 - e^{-\lambda y_i})$, $u_i = \frac{(1 - e^{-\lambda y_i})^\alpha}{1 - (1 - \beta)(1 - (1 - e^{-\lambda y_i})^\alpha)}$,

$$w_i = \frac{y_i e^{-\lambda y_i}}{1 - e^{-\lambda y_i}} \text{ e } v_i = \frac{1 - (1 - e^{-\lambda y_i})^\alpha}{1 - (1 - \beta)(1 - (1 - e^{-\lambda y_i})^\alpha)}.$$

Fazendo $\mathbf{U} = 0$, tem-se o conhecido estimador da função de máxima verossimilhança, com as estimativas $\hat{\alpha}$, $\hat{\beta}$ e $\hat{\lambda}$.

Uma outra quantidade muito importante é matriz de informação observada de Fisher dada por

$$I = - \left[\begin{array}{ccc} \frac{\partial^2 \ell}{\partial \alpha^2} & \frac{\partial^2 \ell}{\partial \alpha \beta} & \frac{\partial^2 \ell}{\partial \alpha \partial \lambda} \\ \cdot & \frac{\partial^2 \ell}{\partial \beta^2} & \frac{\partial^2 \ell}{\partial \beta \partial \lambda} \\ \cdot & \cdot & \frac{\partial^2 \ell}{\partial \lambda^2} \end{array} \right] \Bigg|_{(\hat{\alpha}, \hat{\beta}, \hat{\lambda})},$$

cujas expressões podem ser encontradas no **Apêndice A**. Com a matriz de informação de Fisher é possível se fazer estimativas para α, β e λ por métodos iterativos. Intervalos de confiança e testes de hipóteses para os parâmetros podem ser calculados com o seguinte resultado assintótico

$$\langle \hat{\alpha}, \hat{\beta}, \hat{\lambda} \rangle \sim N(\langle \alpha, \beta, \lambda \rangle, I^{-1}).$$

2.4 Critério de informação de Akaike - AIC

Akaike utilizou a Informação de Kullback-Leibler para testar se um dado modelo é adequado. Ele informou que o viés é dado assintoticamente por p , em que p é o número de parâmetros a serem estimados no modelo, e definiu seu critério de informação como:

$$AIC = -2 \log L(\theta) + 2(p)$$

Dentre todos os modelos testados, deve-se preferir aquele que tem o menor AIC.

2.5 Critério de informação bayesiano - BIC

O Critério de Informação Bayesiano(BIC), proposto por Schwarz (1978) é dado por:

$$BIC = -2 \log L(\theta) + p(\log n)$$

em que $L(\theta)$ é a verossimilhança, o modelo escolhido, p é o número de parâmetros a serem estimados e n é o número de observações da amostra. Tal como o AIC, este critério seleciona entre todos os modelos testados aquele que tem o menor BIC.

2.6 Teste de Wald

O teste de Wald é gerado pela razão entre a estimativa da função de máxima verossimilhança do parâmetro ($\hat{\theta}$) e a sua respectiva estimativa do erro padrão. Sob a hipótese $H_0: \theta = 0$, tem distribuição normal padrão. A estatística do teste Wald é então definida por

$$W = \frac{\hat{\theta}}{\widehat{DP}(\hat{\theta})}.$$

Esta estatística é comparada com o valor crítico da normal padrão, que ao nível de confiança de 95% é, aproximadamente, igual a 2. Assim, se a estimativa é duas vezes maior que o erro padrão, em valores absolutos, a estimativa é dita significativa. Hauck e Donner (1977) ao examinar o teste de Wald descobriram anomalias em determinadas situações. O teste, por exemplo, não rejeitava a hipótese nula quando o coeficiente é significativo. Recomendaram então, a utilização do teste da razão de verossimilhança, como alternativa, quando não se fosse rejeitada a hipótese nula no teste de Wald para uma determinada estimativa.

2.7 Teste da Razão de Verossimilhanças - TRV

O Teste da Razão de Verossimilhanças é um teste bastante geral, podendo ser utilizado para testar um vetor paramétrico (WASSERMAN, 2004). Suponha que deseje-se testar:

$$H_0: \theta \in \Theta_0 \text{ vs } H_1: \theta \notin \Theta_0.$$

O **Teste da Razão de Verossimilhanças** é definido por

$$D = 2 \log \left(\frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)} \right) = -2\ell(\hat{\theta}_0) + 2\ell(\hat{\theta}) \sim \chi_{p-q}^2$$

em que $\ell(\hat{\theta})$ é a log-verossimilhança obtida no seu ponto máximo (estimador de máxima verossimilhança) sob o modelo proposto com p parâmetros e $\ell(\hat{\theta}_0)$ é a log-verossimilhança obtida no seu ponto máximo (estimador de máxima verossimilhança) sob o modelo reduzido (sub-modelo) com q parâmetros. Como Λ têm distribuição qui-quadrado, para se proceder o teste deve-se testar a estatística Λ contra o quantil desejado. Suponha que o modelo completo tem três parâmetros, enquanto seu sub-modelo (modelo proposto sob H_0) têm um parâmetro, se $D \geq \chi_2^2 = 5,99$ rejeita-se H_0 ao nível de confiança de 95%, isto é, o modelo proposto é melhor.

2.8 Aplicação

Como parte dos resultados, serão obtidas a esperança e a variância da distribuição MOEG a partir de expansões da densidade (2.5), uma vez que a forma analítica simples dessas integrais não pode ser determinada. Será utilizado o software Mathematica 7 (Wolfram Research, Inc., 2008) para implementar estes cálculos a partir do qual se estudará a acurácia dessas expansões truncadas e o tempo de execução das mesmas. Serão ajustadas as densidade MOEG com parametrização completa, denominada MOEG1, a densidade MOEG quando $\alpha = 1$ (o submodelo Marshall-Olkin Exponencial), denominada MOEG2, a densidade MOEG quando $\beta = 1$ (o submodelo Exponencial Generalizada), além das densidades gama, de Weibull e log-normal (frequentemente utilizadas em análise de sobrevivência) a dois conjuntos de dados. O primeiro sendo proveniente de magnitudes de terremotos, na escala de magnitude de momentos, na falha da Anatólia (RISTIĆ; NADARAJAH, 2012), o segundo provêm de um estudo de tempos até a falha de alumínio submetidos a uma grande pressão (MORAIS; BARRETO-SOUZA, 2011). Os dados são mostrados no Apêndice A e um resumo com as principais estatísticas encontra-se no próximo capítulo. Na otimização será utilizado o algoritmo BFGS implementado no software *R* (R Development Core Team, 2011).

Para se testar hipóteses acerca dos parâmetros serão utilizados os testes de Wald e Teste da Razão de Verossimilhanças, descritos anteriormente. Os códigos referentes aos ajustes e testes se encontram no **Apêndice B**.

3 Resultados e Discussão

Fazendo $G_\lambda(t) = 1 - e^{-(\lambda t)}$ na equação (2.3) definimos a função de distribuição da variável Marshall-Olkin exponencial generalizada $T \sim MOEG(\lambda, a, b)$ pela expressão

$$F(t) = \frac{(1 - e^{-(\lambda t)})^a}{\{1 - \bar{b} [1 - (1 - e^{-(\lambda t)})^a]\}}, \quad (3.1)$$

cuja função densidade de probabilidade (fdp) é

$$f(t) = \frac{ab (1 - e^{-(\lambda t)})^{a-1} \lambda e^{-(\lambda t)}}{\{1 - \bar{b} [1 - (1 - e^{-(\lambda t)})^a]\}^2}, \quad (3.2)$$

sendo $\lambda > 0$ o parâmetro de escala, $a > 0$ e $b > 0$ os parâmetros adicionais. Se $b = 1$, temos a distribuição de exponencial generalizada (EG) e se $a = 1$, temos a distribuição de Marshall-Olkin exponencial (MOE). Por outro lado, se ambos a e b são unitários, obtemos a distribuição exponencial.

Para demonstrar que a função (3.2) é uma fdp, uma vez que $g(t)$ e $G(t) \geq 0$ o numerador é positivo. Já o denominador é majorado por b (quando $G(t) = 0$) que é positivo, logo $f(t) \geq 0$. Fazendo a substituição $u = 1 - \bar{b} [1 - (1 - e^{\lambda t})^a]$ na integral $\int_0^\infty f(t) dt$ temos

$$\frac{b}{1-b} \int_b^1 \frac{1}{u^2} du = 1,$$

logo $f(t)$ é uma densidade.

A função de sobrevivência toma a forma

$$S(t) = \frac{b [1 - (1 - e^{-\lambda t})^a]}{\{1 - \bar{b} [1 - (1 - e^{-\lambda t})^a]\}}.$$

E assim, a função de risco é dada por

$$h(t) = \frac{a(1 - e^{-\lambda t})^{a-1} \lambda e^{-\lambda t}}{\{1 - \bar{b}[1 - (1 - e^{-\lambda t})^a]\} [1 - (1 - e^{-\lambda t})^a]} \quad (3.3)$$

3.1 Expansão da Densidade

Nesta sessão desenvolve-se resultados interessantes, que serão utilizados nas sessões futuras. Diretamente do teorema binomial para $|z| < 1$ e $r > 0$, tem-se que

$$(1 - z)^r = \sum_{i=0}^{\infty} (-1)^i \binom{r}{i} z^i.$$

Adicionalmente, para $|z| < 1$ e $s > 0$, tem-se

$$(1 - z)^{-s} = \sum_{j=0}^{\infty} \frac{\Gamma(s + j)}{\Gamma(s) j!} z^j,$$

em que $\binom{r}{i}$ é o número binomial e $\Gamma(\cdot)$ é função gamma. Particularmente, para $s = 2$,

$$(1 - z)^{-2} = \sum_{j=0}^{\infty} (j + 1) z^j.$$

Aplicando as expansões acima repetidas vezes para $b \in (0, 1)$, pode-se obter

$$\begin{aligned} \left\{1 - \bar{b} \left[1 - (1 - e^{-\lambda t})^a\right]\right\}^{-2} &= \sum_{j=0}^{\infty} (j + 1) \bar{b}^j \left[1 - (1 - e^{-\lambda t})^a\right]^j \\ &= \sum_{j=0}^{\infty} (j + 1) \bar{b}^j \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} (-1)^{k+l} \binom{j}{k} \binom{ak}{l} e^{-(l\lambda t)} \end{aligned}$$

e

$$\left(1 - e^{-\lambda t}\right)^{a-1} = \sum_{m=0}^{\infty} (-1)^m \binom{a-1}{m} e^{-(m\lambda t)}.$$

Note que o somatório em l se estende ao infinito, uma vez que a é real positivo. Substituindo as duas expressões em (3.2), obtemos

$$f(t) = \sum_{j,k,l,m} w_{j,k,l,m} \cdot g_{\lambda_{l,m}}(t) \quad (3.4)$$

em que $\sum_{j,k,l,m}$ significa $\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \sum_{m=0}^{\infty}$, $g_{\lambda_{l,m}}(t)$ representa a fdp da distribuição exponencial com parâmetro $\lambda_{l,m} = \frac{\lambda}{(1+l+m)}$, e

$$w_{j,k,l,m} = \frac{(-1)^{k+l+m}(j+1)}{1+l+m} \binom{j}{k} \binom{ak}{l} \binom{a-1}{m} ab\bar{b}^j$$

para $b \in (0, 1)$. Se pelo contrário $b > 1$, substitui $w_{j,k,l,m}$ por

$$v_{j,k,l,m} = \frac{(-1)^{j+k+l+m}(j+1)}{1+l+m} \binom{j}{k} \binom{ak}{l} \binom{a-1}{m} a \sum_{s=0}^{\infty} \binom{s}{j} (1-1/b)^s.$$

Com o auxílio do software algébrico, como MATHEMATICA ou MAPLE, é possível concluir que $\sum_{j,k,l,m} w_{j,k,l,m} = \sum_{j,k,l,m} v_{j,k,l,m} = 1$, tal que a distribuição MOEG pode ser expressa por combinações lineares da distribuição exponencial através de (3.4). A partir daqui, utiliza-se $w_{j,k,l,m}$ nas equações expandidas. Para o caso $b > 1$, deve-se substituir $w_{j,k,l,m}$ por $v_{j,k,l,m}$.

3.2 Momentos e Função Geradora de Momentos

Os momentos são características muito importante das distribuição, os momentos dão ideia da tendência central, dispersão e assimetria de uma distribuição.

Teorema 1. *Seja $T \sim MOEG(\lambda, a, b)$, o r -ésimo momento é dado por*

$$\mu'_r = \Gamma(1+r) \sum_{j,k,l,m} w_{j,k,l,m} \lambda_{l,m}^r. \quad (3.5)$$

Prova

Seja $X \sim exp(\lambda)$ uma variável aleatória da distribuição exponencial com parâmetro λ . Uma vez que seu r -ésimo momento é $E[X^r] = \frac{\Gamma(1+r)}{\lambda^r}$, segue diretamente que

$$\begin{aligned} \mu'_r &= E[T^r] = \sum_{j,k,l,m} w_{j,k,l,m} \int_0^{\infty} t^r g_{\lambda_{l,m}}(t) dt \\ &= \sum_{j,k,l,m} w_{j,k,l,m} \frac{\Gamma(1+r)}{\lambda_{l,m}^r} \end{aligned}$$

(c.q.d)

Corolário 1. *Seja $T \sim MOEG(\lambda, a, b)$, a esperança é*

$$\mu = E[T] = \sum_{j,k,l,m} \frac{w_{j,k,l,m}}{\lambda_{l,m}}.$$

Corolário 2. A variância de T é

$$\sigma^2 = E[(T - \mu)^2] = \sum_{j,k,l,m} \frac{w_{j,k,l,m}}{\lambda_{l,m}^2}.$$

Corolário 3. A assimetria de T é dada por

$$\begin{aligned} \gamma_1 &= E \left[\left(\frac{T - \mu}{\sigma} \right)^3 \right] = \sqrt{2} \sum_{j,k,l,m} w_{j,k,l,m} \\ &= \sqrt{2} \end{aligned} \quad (3.6)$$

em que $\sigma = \sqrt{\sigma^2}$ é o desvio-padrão da variável aleatória T .

Corolário 4. A curtose de T é

$$\begin{aligned} \gamma_2 &= E \left[\left(\frac{T - \mu}{\sigma} \right)^4 \right] = 3 \sum_{j,k,l,m} w_{j,k,l,m} \\ &= 3 \end{aligned} \quad (3.7)$$

Assim, a assimetria e a curtose da distribuição MOEG são constantes e iguais as da distribuição exponencial.

Os cumulantes (κ_p) são obtidos diretamente da fórmula recursiva

$$\kappa_p = \mu_p - \sum_{k=1}^p \binom{p-1}{k-1} \kappa_k \mu'_{p-k}.$$

Da equação recursiva acima seguem as conhecidas fórmulas alternativas:

- esperança: $\mu = \kappa_1$
- variância: $\sigma^2 = \kappa_2$
- assimetria: $\gamma_1 = \frac{\kappa_3}{\kappa_2^{3/2}}$
- curtose: $\gamma_2 = \frac{\kappa_4}{\kappa_2^2}$

3.3 Função Quantílica

Invertendo a fd (3.1) obtem-se a função quantílica da distribuição MOEG

$$t = Q(p) = \frac{1}{\lambda} \left\{ -\log \left[1 - \left(\frac{pb}{1-p+pb} \right)^{1/a} \right] \right\}. \quad (3.8)$$

Com a função quantílica (3.8) é possível simular números pseudo-aleatórios da distribuição $T \sim MOEG(\lambda, a, b)$ a partir de $p \sim U(0, 1)$, isto é, p simulado de uma distribuição uniforme no intervalo zero-um. Na Figura 3.1 foi gerada uma amostra de tamanho 10000 da distribuição T , pela função (3.8) com parâmetros fixados em $\lambda = 2$, $a = 5$ e $b = 3$. Também foi ajustada, em azul, a fdp (3.2) para os mesmos valores dos parâmetros, isto é, a densidade verdadeira.

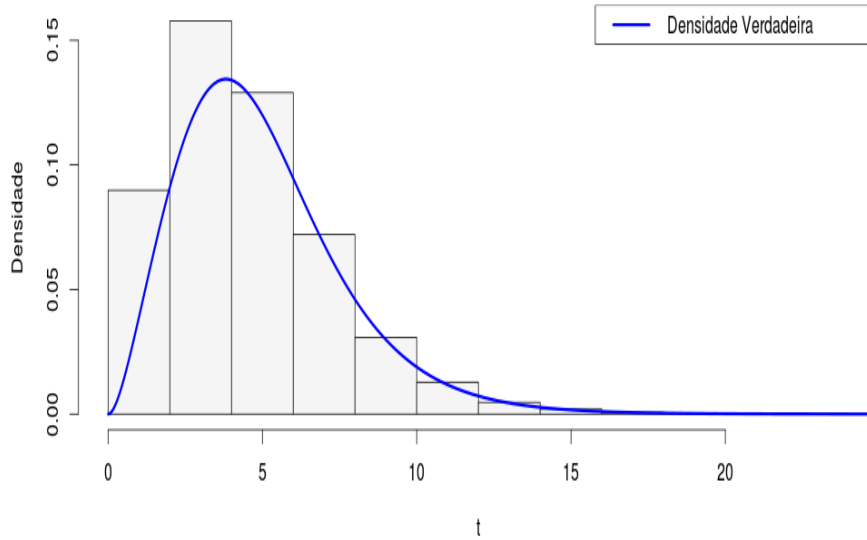


Figura 3.1: Histograma de 10000 números pseudo-aleatórios gerados pela função (3.8) para os valores $\lambda = 2$, $a = 5$ e $b = 3$ com respectiva densidade ajustada.

3.4 Mediana

Uma importante medida de tendência central que utilizaremos a seguir é a mediana. A mesma tem utilidade em Robustez Estatística, pois a mesma é menos sensível a extremos que a média.

Teorema 2. A mediana da distribuição $MOEG(\lambda, a, b)$ é

$$m = \frac{1}{\lambda} \left\{ -\log \left[1 - \left(\frac{b}{1+b} \right)^{1/a} \right] \right\} \quad (3.9)$$

Prova Substituindo $u = 1 - \bar{b}[1 - (1 - e^{\lambda t})^a]$ nas equações $\int_0^m f(t) dt$ e $\int_m^\infty f(t) dt$ temos

$$\frac{b}{1-b} \int_b^{m'} \frac{1}{u^2} du = \frac{b-m'}{(b-1)m'} \quad \text{e} \quad \frac{b}{1-b} \int_{m'}^1 \frac{1}{u^2} du = \frac{b(m'-1)}{(b-1)m'}$$

em que $m' = 1 - \bar{b} \left[1 - \left(1 - e^{-\lambda m} \right)^a \right]$. Igualando as equações acima e resolvendo em m , chegamos ao resultado desejado.

(c.q.d)

3.5 Desvio Médio e Desvio Mediano

Duas estatísticas que medem a dispersão da variável T são o desvio em relação à média e o desvio em relação à mediana definidos por

$$\delta_1(T) = \int_0^{\infty} |t - \mu| f(t) dt \quad \text{e} \quad \delta_2(T) = \int_0^{\infty} |t - m| f(t) dt,$$

respectivamente, em que μ calculado por (3.5) com $r = 1$ é a média de T e m dado por (3.9) é a mediana de T . As medidas $\delta_1(t)$ e $\delta_2(t)$ são expressas em (CORDEIRO; LEMONTE, 2012) por

$$\delta_1(T) = 2\mu G(\mu) - 2J(\mu) \quad \text{e} \quad \delta_2(T) = \mu - 2J(m), \quad (3.10)$$

em que $G(q)$ é obtido de (3.1) e $J(q) = \int_0^q t f(t) dt$. Considerando o caso $b \in (0, 1)$, podemos escrever

$$\begin{aligned} J(q) &= \sum_{j,k,l,m}^{\infty} w_{j,k,l,m} \int_0^q t g_{\lambda_{l,m},\theta} \\ &= \sum_{j,k,l,m}^{\infty} w_{j,k,l,m} \frac{1 - (1 - q \lambda_{l,m}) e^{-q \lambda_{l,m}}}{\lambda_{l,m}}. \end{aligned} \quad (3.11)$$

Se $b > 1$, substituímos $w_{j,k,l,m}$ por $v_{j,k,l,m}$ em (3.11).

3.6 Aplicações

Passamos agora ao ajuste do modelo à dados reais e o confronto do modelo proposto com seus submodelos e as distribuições gama, de Weibull e log-normal, comumente empregadas nas teorias de Análise de Sobrevivência e Confiabilidade. A Tabela 3.1 traz um resumo das principais estatísticas dos dados testados.

Aplicação 1

Para o primeiro ajuste foram utilizadas 25 observações provenientes de magnitudes de

Tabela 3.1: Tamanho amostral(n), médis, variância, coeficiente de variação, mínimo, mediana e máximo dos conjuntos de dados trabalhados

	n	Média	Variância	Coef. variação	Mínimo	Mediana	Máximo
Dados 1	25	6,6	0,27	7,96	6,00	6,50	7,90
Dados 2	100	134,11	490,30	16,51	70,00	133,50	212,00

terremotos, na escala logarítmica dos momentos, na falha da Anatólia próxima ao Mar Morto (RISTIĆ; NADARAJAH, 2012).

Pode-se observar por meio da Tabela 3.2 mostra as estimativas dos parâmetros, erros-padrões, estatística de Wald, para as distribuições simbolizadas por MOEG1 (distribuição proposta), MOEG2 (submodelo quando $\alpha = 1$), MOEG3 (submodelo quando $\beta = 1$), gama, Weibull e log-normal.

Tabela 3.2: Estimativas dos parâmetros, erros-padrão, estatística de Wald, MOGE1, MOGE2, MOGE3, gama, de Weibull e log-normal para o conjunto de dados de magnitudes de terremotos.

	Parâmetro	estimativa	erro-padrão	Wald
MOEG1	α	83,032	7,371	11,265
	β	0,013	0,005	2,600
	λ	0,553	0,025	22,120
MOEG2	β	543,192	46,552	11,668
	λ	1,418	0,0203	69,852
MOEG3	α	4,452	0,180	24,733
	λ	0,283	0,005	56,600
Weibull	α	0,207	0,001	270,000
	β	10,673	0,232	46,004
Gama	α	28,859	1,283	22,493
	β	6,246	0,280	22,307
Log-normal	μ	1,527	0,002	763,500
	σ	0,085	0,002	42,500

Com base no teste Wald, se afirma que os parâmetros α , β e λ do modelo proposto MOEG1 são significativos ao nível de confiança de 95%. O mesmo teste indica que os parâmetros de todas as distribuições subsequentes são significativos para este conjunto de dados. Na Tabela 3.3 encontram os valores dos AIC's e BIC's para cada distribuição ajustada ao conjunto de dados de magnitudes de terremotos.

Tabela 3.3: Critério da Informação de Akaike (AIC) e Bayesiano (BIC) para as distribuições estudadas ajustadas aos dados de magnitudes de terremotos.

	AIC	BIC
MOEG1	964,532	979,256
MOEG2	2258,721	2268,537
MOEG3	3390,707	3400,523
Weibull	1270,478	1280,293
Gama	1750,725	1760,540
Log-normal	980,125	982,563

Tomando como base o menor AIC e BIC, a distribuição proposta é a mais adequada. A seguir na Figura 3.2 são plotados, num mesmo gráfico, o histograma dos dados e as densidades das distribuições aqui analisadas.

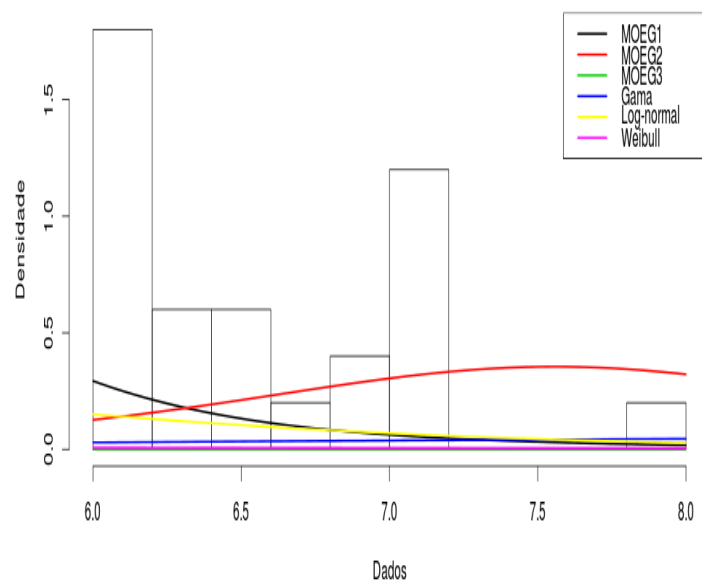


Figura 3.2: Histograma dos dados de magnitudes de terremotos e o confronto do modelo proposto com seus submodelos e as distribuições gama, de Weibull e log-normal

Como se pode observar no gráfico a acima, poucos modelos tiveram um bom ajuste ao gráfico. O modelo proposto MOEG1, e a distribuição Log-Normal são aqueles que visualmente apresentam os melhores ajustes aos dados.

Na Tabela 3.4 se podem observar o teste da razão de verossilhanças. Ao nível de significância de 5% se rejeita a hipótese nula, ou seja os modelos reduzidos devem ser descartados.

Tabela 3.4: Log-verossimilhança e estatísticas do testa de razão de verossimilhanças para o modelo proposto contra seus submodelos.

Modelo	$\ell(\theta)$	$2 \ell(\theta)$	TRV
MOEG1	-479,266	-958,532	-
MOEG2	-1127,361	-2254,722	1296,190
MOEG3	-1693,354	-3386,708	2428,176

Aplicação 2

Para a próxima aplicação, tomou-se os dados de (BIRNBAUM; SAUNDERS, 1969) referentes aos tempos de vida de componentes de alumínio expostos a a uma pressão de 31000 psi.

A Tabela 3.5 mostra as estimativas dos parâmetros, erros-padrões, estatística de Wald, para as distribuições simbolizadas por MOEG1 (distribuição proposta), MOEG2 (submodelo quando $\alpha = 1$), MOEG3 (submodelo quando $\beta = 1$), gama e log-normal ajustadas para os dados descritos acima.

Tabela 3.5: Estimativas dos parâmetros, erros-padrões, estatística de Wald, MOEG1, MOEG2, MOEG3, gama, log-normal para o conjunto de dados de alumínio

	Parâmetro	estimativa	erro-padrão	Wald
	α	209,330	264,361	0,792
MOEG1	β	190,295	272,043	0,700
	λ	0,079	0,007	12,201
MOEG2	β	268,542	72,675	3,695
	λ	0,044	0,002	19,883
MOEG3	α	268,542	72,675	3,695
	λ	0,044	0,002	19,883
Gama	α	36,549	5,142	7,108
	β	0,273	0,039	7,059
Log-normal	μ	4,885	0,017	291,864
	σ	0,028	0,004	7,097

Outra vez, com base no teste Wald, se afirma que os parâmetros α e β do modelo proposto

MOEG1 são não significativos ao nível de confiança de 95%, com λ significativo. O teste ainda indica que os parâmetros de todas as distribuições subsequentes são significativos para este conjunto de dados. Na Tabela 3.6 se encontram os valores dos AIC's e BIC's para cada distribuição ajustada ao conjunto de dados reais de tempos de falha da alumínios submetidos a pressão de 31000 psi.

Tabela 3.6: Critério da Informação de Akaike (AIC) e Bayesiano (BIC) para as distribuições estudadas ajustadas aos dados de tempos de falha de alumínio

	AIC	BIC
MOGE1	905,392	913,208
MOGE2	950,567	955,778
MOGE3	950,567	955,778
Gama	905,677	910,888
Log-normal	907,247	912,458

De acordo com a esta tabela, a distribuição proposta MOEG1 é o melhor modelo testado uma vez que os menores valores de AIC e BIC são para esta distribuição. A distribuição gama, desta vez está em segundo lugar. A seguir na Figura 3.3 são plotados, num mesmo gráfico, o histograma dos dados e as densidades das distribuições aqui analisadas.

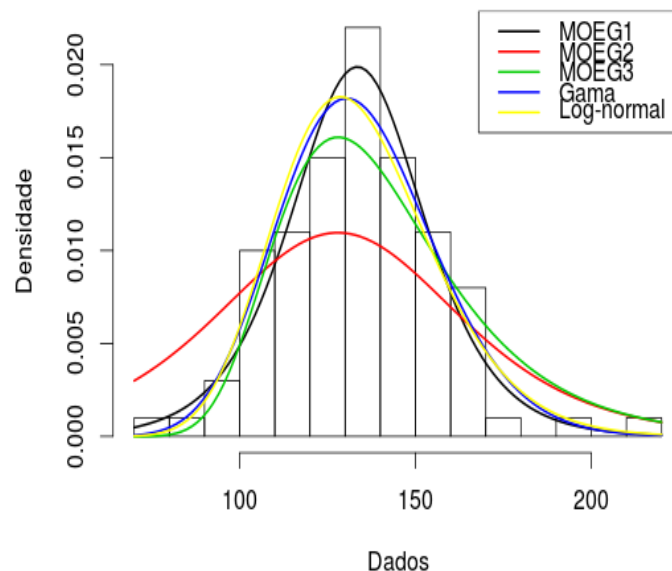


Figura 3.3: Histograma dos dados de alumínio e o confronto do modelo proposto com seus submodelos e as distribuições gama e log-normal

Como se pode observar no gráfico a acima, o modelo proposto MOEG, com parâmetros

α , β e λ livres, apresentado pela linha preta se ajusta bem aos dados. Outras distribuições que se ajustam bem visualmente são as distribuições gama e log-normal.

Na Tabela 3.7 se podem observar o teste da razão de verossilhanças. Ao nível de significância de 5% se rejeita a hipótese nula, ou seja os modelos reduzidos devem ser descartados.

Tabela 3.7: Log-verossilhança e estatísticas do testa de razão de verossilhanças para o modelo proposto contra seus submodelos.

Modelo	$\ell(\theta)$	$2 \ell(\theta)$	TRV
MOEG1	-449,696	-899,392	-
MOEG2	-473,284	-946,568	47,176
MOEG3	-473,284	-946,568	47,176

Assim, embora o teste de Wald determinasse a não significância dos parâmetros da distribuição triparamétrica proposta, pelo Teste da Razão de Verossilhanças e pelos critérios de AIC e BIC observou-se que a distribuição se ajusta bem a dados de análise de sobrevivência e pode competir com distribuições clássicas da área.

Conclusão

Conclui-se que ao introduzir um segundo parâmetro à família Marshall-Olkin, obtemos uma distribuição Marshall-Olkin Exponencial Generalizada (MOEG), uma distribuição triparmétrica podendo competir com a distribuição beta exponencial e Weibull generalizada. Com base no gráfico da Figura 3.4 observar-se que o parâmetro alfa tem uma relação direta com a taxa de sobrevivência. Na Figura 3.5b é possível observar que a função de risco tem um comportamento da forma crescente-decrescente-constante ou decrescente-constante-crescente, que denominamos curva de banheira ou banheira invertida, demonstrando assim o quanto a distribuição Marshall-Olkin é flexível.

Tomando como base o menor coeficiente de Akaike(AIC) em todas as aplicações, o modelo proposto é o mais indicado.

Referências Bibliográficas

- ACHCAR, J. A.; BOLETA, J. Distribuição exponencial generalizada: uso de métodos bayesianos. *Revista Brasileira de Biometria*, 2009.
- BIRNBAUM, Z.; SAUNDERS, S. A new family of life distributions. *Journal of Applied Probability*, v. 6, p. 319–327, 1969.
- COLOSIMO, E. A.; GIOLO, S. R. *Análise de Sobrevida Aplicada*. [S.l.]: Associação Brasileira de Estatística., 2006.
- CORDEIRO, G. M.; CASTRO, M. A new family of generalized distributions. *Journal of Statistical Computation and Simulation*, v. 81, n. 7, p. 883–898, 2011.
- CORDEIRO, G. M.; LEMONTE, A. On the marshall–olkin extended weibull distribution. *Statistical Papers*, 2012.
- CORDEIRO, G. M.; SIMAS, A. B.; STOSIC, B. D. Closed form expressions for moments of the beta weibull distribution. *Annals of the Brazilian Academy of Sciences*, v. 83, p. 357 – 373, 2011.
- GUPTA, R. D.; KUNDU, D. Generalized exponential distributions. *Australian and New Zealand Journal of Statistics*, 1999.
- GUPTA, R. D.; KUNDU, D. Generalized exponential distribution: Bayesian inference. *Comput. Stat. Data Anal.*, v. 52, n. 4, p. 1873–1883, 2008.
- HAUCK, W. W. J.; DONNER, A. Wald's test as applied to hypotheses in logit analysis. *J. Am. Stat. Assoc.*, v. 72, p. 851–853, 1977.
- JONES, M. C. Families of distributions arising from distributions of order statistics. *Test*, v. 13, p. 1 – 43, 2004.
- JONES, M. C. Kumaraswamy's distribution: A beta-type distribution with some tractability advantages. *Statistical Methodology*, v. 6, p. 70 – 81, 2008.
- LEIVA, V.; BARROS, M.; PAULA, G. A. *Generalized Birnbaum-Saunders Models using R*. [S.l.]: Ass, 2009.
- MARSHALL, A. W.; OLKIN, I. A new method for adding a parameter to a family of distributions with application to the exponential and weibull families. *Biometrika*, v. 84, n. 3, p. 641–652, 1997.
- MORAIS, A. L.; BARRETO-SOUZA, W. A compound class of weibull and power series distributions. *Computational Statistics & Data Analysis*, Elsevier, v. 55, n. 3, p. 1410–1425, 2011.

- MUDHOLKAR, G. S.; SRIVASTAVA, D. K.; M., F. The exponentiated weibull family. *Technometrics*, v. 37, p. 436–45, 1995.
- N., E.; C., L.; F., F. Beta-normal distribution and its applications. *Commun Stat-Theor M.*, v. 31, p. 497 – 512, 2002.
- NADARAJAH, S. The exponentiated exponential distribution: a survey. *AStA Adv Stat Anal*, 2011.
- NADARAJAH, S.; KOTZ, S. The beta exponential distribution. *Reliability Engineering and System Safety*, 2006.
- PREDA, V.; PANAITESCU, E.; CIUMARA, R. The modified exponential-poisson distribution. *Proceedings of the Romanian Academy*, v. 12, n. 1, p. 22–29, 2011.
- R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2011. ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org/>>.
- RAQAB, M. Z. Inferences for generalized exponential distribution based on record statistics. *J. Stat. Plann. Infer.*, v. 104, p. 339–350, 2002.
- RAQAB, M. Z.; AHSANULLAH, M. Estimation of the location and scale parameters of generalized exponential distribution based on order statistics. *J. Stat. Comput. Simul.*, v. 69, p. 109–124, 2001.
- RISTIĆ, M. M.; NADARAJAH, S. A new lifetime distribution. *Journal of Statistical Computation and Simulation*, Taylor & Francis, n. ahead-of-print, p. 1–16, 2012.
- SARHAN, A. M. Analysis of incomplete, censored data in competing risks models with generalized exponential distributions. *IEEE Trans. Reliab.*, v. 56, p. 132–138, 2007.
- WASSERMAN, L. *All of Statistics: a concise course in statistics inference*. 2. ed. Heidelberg: Springer, 2004.
- Wolfram Research, Inc. *Mathematica Edition: Version 7.0*. Champaign, Illinois: Wolfram Research, Inc., 2008.
- ZHENG, G. Fisher information matrix in type-ii censored data from exponentiated exponential family. *Biometrical J.*, v. 44, p. 353–357, 2002.

APÊNDICE

Apêndice A: Elementos da Matriz Observadas de Fisher

Abaixo são relacionadas todas as derivadas segundas da informação observadas de Fisher para a distribuição Marshall-Olkin Generalizada.

$$\frac{\partial^2 \ell}{\partial \alpha^2} = 2 \sum_{i=1}^n ((1-\beta)v_i z_i^2 - (1-\beta)^2 v_i^2 z_i^2) + \frac{n}{\alpha^2}$$

$$\frac{\partial^2 \ell}{\partial \alpha \beta} = -2 \sum_{i=1}^n v_i z_i ((1-\beta)u_i + 1)$$

$$\frac{\partial^2 \ell}{\partial \alpha \partial \lambda} = 2 \sum_{i=1}^n (1-\beta)v_i w_i (1 - e^{-\lambda y_i}) (-\alpha(1-\beta)v_i z_i + \alpha z_i + 1) - \sum_{i=1}^n w_i$$

$$\frac{\partial^2 \ell}{\partial \beta^2} = -2 \sum_{i=1}^n u_i^2$$

$$\frac{\partial^2 \ell}{\partial \beta \partial \lambda} = -2 \sum_{i=1}^n \alpha v_i w_i (1-\beta)$$

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \lambda^2} &= -2 \sum_{i=1}^n (\alpha(1-\beta)v_i w_i y_i + (\alpha(1-\beta)v_i w_i)^2 + (1-\alpha)\alpha(1-\beta)v_i w_i^2) \\ &\quad - (1-\alpha) \sum_{i=1}^n w_i (w_i + y_i) + \frac{n}{\lambda^2} \end{aligned}$$

Sendo:

$$\begin{aligned} z_i &= \log(1 - e^{-\lambda y_i}) \\ u_i &= \frac{(1 - e^{-\lambda y_i})^\alpha}{1 - (1-\beta)(1 - (1 - e^{-\lambda y_i})^\alpha)} \\ w_i &= \frac{y_i e^{-\lambda y_i}}{1 - e^{-\lambda y_i}} \\ v_i &= \frac{1 - (1 - e^{-\lambda y_i})^\alpha}{1 - (1-\beta)(1 - (1 - e^{-\lambda y_i})^\alpha)} \end{aligned}$$

Apêndice B: Código R utilizado para fazer os ajustes.

```

#x=round(rgamma(50,2,1)+rweibull(50,5,1),1)
x=c(1.8, 1.4, 1.7, 1.8, 4.9, 1.1, 1.8, 1.3, 3.2, 3.3, 5.1,
3.6, 1.7, 2.1, 2.2, 3.9, 3.4, 2.1, 1.2, 4.2, 1.2, 0.9, 2.4,
2.6, 6.0, 1.4, 2.5, 1.9, 2.2, 0.8, 2.3, 2.1, 9.8, 1.3, 2.5,
1.4, 4.1, 3.7, 2.0, 1.5, 1.8, 5.6, 4.6, 2.1, 2.6, 4.1, 5.6, 1.1, 2.0, 4.1)

# Tempos de vida de componentes de alumínio expostos a 31000 psi
#x=c(70,90,97,99,100,103,104,104,105,
107,108,108,108,109,109,112,112,113,
114,114,114,116,119,120,120,120,121,
121,123,124,124,124,124,124,128,128,
129,129,130,130,130,131,131,131,131,
131,132,132,132,133,134,134,134,134,
134,136,136,137,138,138,138,139,139,
141,141,142,142,142,142,142,142,144,
144,145,146,148,148,149,151,151,152,
155,156,157,157,157,157,158,159,162,
163,163,164,166,166,168,170,174,196,
212)

# Tempos de reparo em (horas) de um trasceptor aéreo

#x=c(0.2, 0.3, 0.5, 0.5, 0.5, 0.5, 0.6, 0.6,
0.7, 0.7, 0.7, 0.8, 0.8, 1.0, 1.0, 1.0,
1.0,1.1, 1.3, 1.5, 1.5, 1.5, 1.5, 2.0,
2.0, 2.2, 2.5, 2.7, 3.0, 3.0, 3.3, 3.3,

```



```

    4.0, 4.0, 4.5, 4.7, 5.0, 5.4, 5.4, 7.0,
    7.5, 8.8, 9.0, 10.3, 22.0, 24.5)
n=length(x)

##### MOEG #####

ll <- function(z){
  a <- z[1]
  b <- z[2]
  l <- z[3]
  n <- length(x)
  ll= n*log(a*b*l)-l*sum(x) + (a-1)*sum( log( 1-exp(-l*x) ) ) -
  2*sum( log( 1 - (1-b)*( 1- ( 1-exp(-l*x) )^a ) ) )
}

par<-c(.1,.1,.1)
fit<-optim(par, ll, method = "BFGS" ,
          control=list(fnscale=-1),hessian=TRUE)

est<-fit$par
lik<-fit$value
est
estim1=est
erro1=sqrt(diag(solve(-fit$hessian)))

lik
aic1=2*length(par)-2*lik
bic1=-2*lik+length(par)*log(n)

hist(x,breaks=12,freq=FALSE,
     ylab='Densidade',xlab='Dados',main='')
a<-est[1]
b<-est[2]

```

```

l<-est[3]
density2<-function(x){
  a*b*1*(1-exp(-1*x))^(a-1)*exp(-1*x)/(1 - (1-b)*( 1- ( 1-exp(-1*x) )^a ))^2}
curve(density2(x),add=TRUE,col=1,lwd=2)

##### MOEG a=1 #####
a=1
ll <- function(z){
  b <- z[1]
  l <- z[2]
  n <- length(x)
  ll= n*log(a*b*1)-l*sum(x) + (a-1)*sum( log( 1-exp(-1*x) ) ) -
  2*sum( log( 1 - (1-b)*( 1- ( 1-exp(-1*x) )^a ) ) )
}

#x=round(rgamma(50,2,1)+rweibull(50,5,1),1)
#n=length(x)

par<-c(.1,.1)
#fit<-optim(par, ll, gr = grad, method = "BFGS" ,
#control=list(fnscale=-1),hessian=TRUE)
fit<-optim(par, ll, method = "BFGS" ,
          control=list(fnscale=-1),hessian=TRUE)

est<-fit$par
lik<-fit$value
est
estim2=est
erro2=sqrt(diag(solve(-fit$hessian)))

lik

```

```

aic2=2*length(par)-2*lik
bic2=-2*lik+length(par)*log(n)

#hist(x,breaks=12,freq=FALSE)
#a<-est[1]
#b<-est[1]
#l<-est[2]
density1<-function(x){
  a*b*l*(1-exp(-l*x))^(a-1)*exp(-l*x)/( 1- (1-b)* (1-exp(-a*l*x)) )^2}
curve(density2(x),add=TRUE,col=2,lwd=2)

##### MOEG b=1 #####
b=1
ll <- function(z){
  a <- z[1]
  l <- z[2]
  n <- length(x)
  ll= n*log(a*b*l)-l*sum(x) + (a-1)*sum( log( 1-exp(-l*x) ) ) -
  2*sum( log( 1 - (1-b)*( 1- ( 1-exp(-l*x) )^a ) ) )
}

par<-c(.1,.1)
#fit<-optim(par, ll, gr = grad, method = "BFGS" ,
#control=list(fnscale=-1),hessian=TRUE)
fit<-optim(par, ll, method = "BFGS" ,
          control=list(fnscale=-1),hessian=TRUE)

est<-fit$par
lik<-fit$value
est

```

```

estim3=est
erro3=sqrt(diag(solve(-fit$hessian)))

lik
aic3=2*length(par)-2*lik
bic3=-2*lik+length(par)*log(n)

#hist(x,breaks=12,freq=FALSE)
a<-est[1]
l<-est[2]
density1<-function(x){
  a*b*1*(1-exp(-l*x))^(a-1)*exp(-l*x)/(1-(1-b)*(1-exp(-a*l*x)))^2}
curve(density2(x),add=TRUE,col=3,lwd=2)

##### Gama #####

ll <- function(z){
  a <- z[1]
  b <- z[2]
  n <- length(x)
  ll= n*a*log(b)-b*sum(x) - n*log(gamma(a)) + (a-1)*sum(log(x))
}

par<-c(1,1)
fit<-optim(par, ll, method = "BFGS" ,
           control=list(fnscale=-1),hessian=TRUE)

est<-fit$par
lik<-fit$value
estim4=est
erro4=sqrt(diag(solve(-fit$hessian)))
lik

```

```

aic4=2*length(par)-2*lik
bic4=-2*lik+length(par)*log(n)

#hist(x,breaks=12,freq=FALSE)
a<-est[1]
b<-est[2]
density2<-function(x){
  dgamma(x,a,b)
}
curve(density2(x),add=TRUE,col=4,lwd=2)

##### Log-normal #####

ll <- function(z){
  mu <- z[1]
  s <- z[2]
  n <- length(x)
  ll= -n/2*log(2*pi*s)-sum( (log(x)-mu)^2/(2*s) )-sum(log(x))
}

par<-c(.1,.1)
fit<-optim(par, ll, method = "BFGS" ,
           control=list(fnscale=-1),hessian=TRUE)

est<-fit$par
lik<-fit$value
estim5=est
erro5=sqrt(diag(solve(-fit$hessian)))
lik
aic5=2*length(par)-2*lik
bic5=-2*lik+length(par)*log(n)

```

```
#hist(x,breaks=12,freq=FALSE)
mu<-est[1]
s<-est[2]
density2<-function(x){
  1/(x*sqrt(2*pi*s))*exp(-(log(x)-mu)^2/(2*s))
}
curve(density2(x),add=TRUE,col=7,lwd=2)
legend('topright',c('MOEG1',expression(MOEG2),'MOEG3','Gama','Log-normal'),
col=c(1,2,3,4,7),lwd=2)
```

```
#Tabelas
```

```
#Estimativas
```

```
estim1
```

```
estim2
```

```
estim3
```

```
estim4
```

```
estim5
```

```
#Erros Padrão
```

```
erro1
```

```
erro2
```

```
erro3
```

```
erro4
```

```
erro5
```

```
# Estatisticas de Wald
```

```
estim1/erro1
```

```
estim2/erro2
```

```
estim3/erro3
```

```
estim4/erro4
```

```
estim5/erro5
```

```
#AIC (Critério de Informação de Akaike)
```

```
aic1
```

```
aic2
```

```
aic3
```

```
aic4
```

```
aic5
```

```
#BIC
```

```
bic1
```

```
bic2
```

```
bic3
```

```
bic4
```

```
bic5
```

```
mean(x)
```

```
sd(x)
```

```
sd(x)/mean(x)
```