

UM NOVO MÉTODO PARA COMPOR DISTRIBUIÇÕES:
UMA ANÁLISE DA CLASSE G-POISSON

VITOR EMANUEL DE LYRA SANTOS NAVARRETE

Orientador: Prof. Dr. Gauss Moutinho Cordeiro

Área de Concentração: Probabilidade

Recife, fevereiro de 2013.

UM NOVO MÉTODO PARA COMPOR DISTRIBUIÇÕES:
UMA ANÁLISE DA CLASSE G-POISSON

VITOR EMANUEL DE LYRA SANTOS NAVARRETE

Orientador: Prof. Dr. Gauss Moutinho Cordeiro

Área de Concentração: Probabilidade

Dissertação submetida como requerimento parcial para obtenção do grau
de Mestre em Estatística pela Universidade Federal de Pernambuco

Recife, fevereiro de 2013.

Catálogo na fonte
Bibliotecária Monick Raquel Silvestre da Silva, CRB4-1217

Santos Navarrete, Vitor Emanuel de Lyra

Um novo método para compor distribuições: uma análise da classe G-Poisson / Vitor Emanuel de Lyra Santos Navarrete. - Recife: O Autor, 2013.

xiv, 63 f.: il., fig., tab.

Orientador: Gauss Moutinho Cordeiro.

Dissertação (mestrado) - Universidade Federal de Pernambuco. CCEN. Estatística, 2013.

Inclui bibliografia.

1. Probabilidades. 2. Distribuição de probabilidades I. Cordeiro, Gauss Moutinho (orientador). II. Título.

519.2

CDD (23. ed.)

MEI2013 – 030

Universidade Federal de Pernambuco
Pós-Graduação em Estatística

Recife, 18 de fevereiro de 2013.

Nós recomendamos que a dissertação de mestrado de autoria de

Vitor Emanuel de Lyra Santos Navarrete

Intitulada

“Um Novo Método para Compor Distribuições: Uma Análise da Classe G-Poisson”

Seja aceita como cumprimento parcial dos requerimentos para o grau de Mestre em Estatística.

Coordenador da Pós-Graduação em Estatística

Banca Examinadora:

Gauss Moutinho Cordeiro Orientador / UFPE

Leandro Chaves Rêgo UFPE

Antônio Eduardo Gomes UNB

Este documento será anexado à versão final da tese.

Ao muro de amizades que erguemos durante a vida.

Agradecimentos

Agradeço, inicialmente, à toda minha família, mãe, irmãs, avós, tios e tias, primos e primas, que são a base dessa fundação que chamamos de vida. As ligações genéticas são uma herança proveniente da engenhosidade da mãe natureza e estarão conosco por toda nossa vida.

Agradeço aos diversos professores que passaram por minha vida e acreditaram, ou não, em mim. Ao professor Gauss, por toda a paciência durante a orientação deste trabalho. A todos os professores do programa de pós-graduação da UFPE, que permitiram a construção deste trabalho. Ao professor Juvêncio, pelos muitos anos de paciência e conselhos. Ao professor João Maurício pela energia. Aos professores da graduação em Atuária da UFC, Iana, Alane, Sérgio, Ana e Alana que tanto me aturaram. E aos tantos outros professores, do ensino fundamental e médio, dos cursos de língua e de música. A todos estes, o meu eterno agradecimento.

Agradeço também aos funcionários que fazem o sistema de ensino andar nos trilhos. Sem o Emilson, a Valéria e o Jimmy eu não teria chegado até aqui.

Mas, também agradeço aos amigos, pois, se a família é a base que constroi o nosso lar, para o qual sempre podemos retornar, os amigos são o muro que cerca e protege, que nos acompanha por toda a jornada. No aproximado meio quarto de século que vivi, andei por vários lugares e, em cada um deles, coletei ao menos um amigo que se tornou um tijolo em tal muro.

Agradeço a todo e cada um destes tijolos. Ao Cleyton que é o primeiro de qual tenho lembrança. Ao Daniel pelas noites em claro, quando isso ainda era proibido. A Jayme, Neto, Pedro Silas e Thiciano pelas tardes preguiçosas. A Danilo, Daniel e o resto do grupo pelas aventuras. Aos sonhadores do CEFET que ainda mantêm seus ideais. A Greice pela confiança e a Nathália pela falta de juízo. A Sérgio e Fábio pela gozação. Ao pessoal das guildas, Deby, Paloma, André, Guilherme e todo o resto, pois as vezes trabalhar em grupo é mais divertido. À Natália por ser a pessoa mais paciente que conheço. A Maira e Robério, já que alguém tem que pagar a conta. Ao Eunílso pela confiança indubitável. Ao Mário que me ouviu. À Marilac, já que ninguém precisa fazer sentido. À Maristela pela imprevisibilidade. Ao Darsayev, David, Jonas e Ítalo, pois nada melhor para os

nervos que um estúdio. À Deborah por não ter estribeiras. Ao Gustavo pelos sete anos de tentativas de ficar rico. A Talles e Mel, pois piscina é sempre bom. À Roberta que estará sempre lá. Ao Márcio por aguentar minha idiossincrasia.

“Que espécie de homens somos nós, então, se não temos nosso próprio raciocínio?”

- Don Corleone.

O Poderoso Chefão, Mario Puzzo.

Resumo

Recentemente, diversos autores têm investigado novas distribuições de sobrevivência devido à sua flexibilidade para ajustar dados. Muitas dessas novas distribuições são obtidas através da composição de uma distribuição contínua com uma discreta. Neste trabalho um novo método para compor distribuições é proposto: a classe G-Poisson. Diversas propriedades dessa distribuição geral são estudadas. A distribuição Lomax-Poisson, pertencente a tal classe, também é investigada.

Palavras-chave: Análise de sobrevivência; Composição; Distribuição Poisson; Estimadores de máxima verossimilhança.

Abstract

Nowadays many authors investigate new survival distributions given their flexibility to model data. Many of those new distribution are obtained through the composition of a continuous and a discrete distribution. In this thesis a new method to compose distributions is proposed: the G-Poisson class. Many properties of this general distribution are studied. Also, the Lomax-Poisson distribution, contained in the proposed class, is investigated.

Keywords: Composition; Maximum likelihood estimators; Poisson distribution; Survival analysis.

Lista de Figuras

1.1	Densidade da distribuição Weibull exponencializada para $\alpha = 2, \beta = 2$ e alguns valores de a	3
1.2	Taxa de falha da distribuição Weibull exponencializada para alguns valores do parâmetros.	3
1.3	Densidade da distribuição exponencial exponencializada para $\beta = 2$ e alguns valores de a	4
1.4	Taxa de falha da distribuição exponencial exponencializada para $\beta = 2$ e alguns valores de a	4
1.5	Densidade da distribuição exponencial geométrica para alguns valores dos parâmetros.	7
1.6	Taxa de falha da distribuição exponencial geométrica para alguns valores dos parâmetros.	7
1.7	Densidade da distribuição exponencial Poisson para $\beta = 2$ e alguns valores de θ	8
1.8	Taxa de falha da distribuição exponencial Poisson para $\beta = 2$ e alguns valores de θ	8
1.9	Densidade da distribuição exponencial logarítmica para $\beta = 2$ e alguns valores de q	10
1.10	Taxa de falha da distribuição exponencial logarítmica para $\beta = 2$ e alguns valores de q	10

1.11	Densidade da distribuição Weibull geométrica para $\alpha = 2$, $\beta = 2$ e alguns valores de p	12
1.12	Taxa de falha da distribuição Weibull geométrica para alguns valores do parâmetros.	12
1.13	Densidade da distribuição Weibull Poisson para alguns valores dos parâmetros.	13
1.14	Taxa de falha da distribuição Weibull Poisson para alguns valores dos parâmetros.	13
1.15	Densidade da distribuição Weibull logarítmica para alguns valores dos parâmetros.	15
1.16	Taxa de falha da distribuição Weibull logarítmica para alguns valores dos parâmetros.	15
2.1	Densidade da distribuição MOEE para $\beta = 2$ e alguns valores de λ	19
2.2	Taxa de falha da distribuição MOEE para $\beta = 2$ e alguns valores de λ	19
2.3	Densidade da distribuição MOEW para $\alpha = 2$, $\beta = 2$ e alguns valores de λ	20
2.4	Taxa de falha da distribuição MOEW para alguns valores dos parâmetros.	20
2.5	Densidade da distribuição MOEL para alguns valores dos parâmetros.	22
2.6	Taxa de falha da distribuição MOEL para alguns valores dos parâmetros.	22
3.1	Densidade da exponencial-Poisson para alguns valores dos parâmetros.	34
3.2	Densidade da Weibull-Poisson para alguns valores dos parâmetros.	34
3.3	Densidade da gama-Poisson para alguns valores dos parâmetros.	35
3.4	Densidade da normal-Poisson padrão para alguns valores de θ	35
3.5	A função de risco da exponencial-Poisson para alguns valores de θ com $\lambda = 1$	38
3.6	Histograma dos sinistros e densidades ajustadas.	48
3.7	Histograma do tempo de sobrevivência e densidades ajustadas.	49
4.1	Densidade da distribuição LP para alguns valores dos parâmetros.	51
4.2	Taxa de falha da distribuição LP para alguns valores dos parâmetros.	51
4.3	Histograma do tempo de remissão (meses) e as densidades ajustadas.	59

Lista de Tabelas

3.1	EMVs dos parâmetros dos modelos e as estimativas dos erros-padrões para o primeiro conjunto de dados bem como as estatísticas AIC e BIC correspondentes.	47
3.2	EMVs dos parâmetros dos modelos e as estimativas dos erros-padrões para o segundo conjunto de dados bem como as estatísticas AIC e BIC correspondentes.	48
4.1	EMVs dos parâmetros dos modelos e as estimativas dos erros-padrões bem como as estatísticas AIC e BIC correspondentes.	58

Lista de Figuras	ix
Lista de Tabelas	xi
1 Introdução	1
1.1 Composições	5
1.1.1 Exponencial G	6
1.1.2 Weibull G	11
1.1.3 Outros métodos de composição	16
2 A classe Marshall-Olkin estendida G	17
2.1 A classe Marshall-Olkin	17
2.2 A distribuição Marshall-Olkin estendida Lomax	21
2.3 Algumas propriedades	23
2.3.1 Expansão para a distribuição	23
2.3.2 Momentos e Geratriz de Momentos	24
2.3.3 Entropia	26
2.3.4 Desvios Médios	28
2.3.5 Momentos Probabilisticamente Ponderados	29
2.3.6 Estatísticas de Ordem	30
2.4 Estimadores de Máxima Verossimilhança	30

3	A classe G-Poisson	32
3.1	A classe G-Poisson	32
3.2	Alguns casos especiais	33
3.2.1	Exponencial-Poisson	33
3.2.2	Weibull-Poisson	33
3.2.3	Gama-Poisson	36
3.2.4	Normal-Poisson	36
3.3	Algumas Propriedades	36
3.3.1	Expansão para a distribuição G-Poisson	36
3.3.2	Função taxa de falha	37
3.3.3	Moda	38
3.3.4	Simulação	39
3.3.5	Momentos	39
3.3.6	Função geratriz de momentos	41
3.3.7	Entropia	42
3.3.8	Desvios Médios	43
3.3.9	Momentos Probabilisticamente Ponderados	43
3.3.10	Estatísticas de ordem	44
3.4	Estimação	45
3.5	Aplicação	47
4	A distribuição Lomax-Poisson	50
4.1	A distribuição Lomax-Poisson	50
4.2	Propriedades	52
4.2.1	Expansão para a distribuição LP	52
4.2.2	Momentos e Geratriz de Momentos	53
4.2.3	Entropias	53
4.2.4	Desvios Médios	54
4.2.5	Momentos Probabilisticamente Ponderados	54
4.2.6	Estatísticas de Ordem	55
4.3	Estimadores de Máxima Verossimilhança	56
4.4	Aplicação	57

CAPÍTULO 1

Introdução

A análise de sobrevivência é a área da estatística que estuda a quantidade de tempo transcorrido até a efetivação de um evento de interesse. Tal evento pode ser tanto a morte como a cura de um paciente ou o tempo até um equipamento novo falhar. Dessa forma, a análise de sobrevivência tem aplicação em diversas áreas como medicina, engenharia, atuária e biologia, entre outras.

As distribuições exponencial, Weibull e gama são comumente empregadas em estudos de sobrevivência devido às suas interessantes propriedades. Essas distribuições, contudo, apresentam restrições, como formas limitadas das suas funções de risco, devido ao reduzido número de parâmetros.

A evolução da computação, entretanto, possibilita que distribuições mais complexas sejam ajustadas aos dados. Assim, diversos pesquisadores têm investigado novas maneiras de adicionar parâmetros às distribuições já conhecidas, tornando as distribuições geradas mais flexíveis.

Mudholkar e Srivastava (1993), por exemplo, propuseram a distribuição Weibull exponencializada. As distribuições exponencializadas são construídas da seguinte maneira: seja Y o tempo de falha de uma componente, interpretado como uma variável aleatória com função distribuição $G(\cdot)$. Considerando-se, então, um sistema com a componentes em paralelo, em que tal sistema falha se todas as a componentes falharem e os tempos de falha das componentes são independentes. Assim, o tempo de falha X do sistema é dado

por $X = \max\{Y_1, Y_2, \dots, Y_a\}$ e sua função distribuição por

$$H_a(x) = G(x)^a, \quad (1.1)$$

em que $H_a(x)$ é chamada de G-exponencializada, por vezes também referida por G-generalizada. Sua densidade é representada por

$$h_a(x) = a G(x)^{a-1} g(x). \quad (1.2)$$

A função distribuição definida por (1.1) e a densidade definida por (1.2) podem ser estendidas para qualquer a real positivo. A quantidade $a G(x)^{a-1}$ atua como um fator de correção na distribuição original, também, denominada de *baseline*. É fácil verificar que para $a = 1$, $H_a(x)$ é equivalente a *baseline* $G(x)$, i.e., $H_1(x) = G(x)$.

A função distribuição e a função densidade da distribuição Weibull (Weibull, 1951) são iguais a

$$G(x) = 1 - \exp[-(x/\beta)^\alpha] \quad \text{e} \quad g(x) = \frac{\alpha x^{\alpha-1}}{\beta^\alpha} \exp[-(x/\beta)^\alpha], \quad x > 0, \quad (1.3)$$

em que $\alpha > 0$ e $\beta > 0$ são os parâmetros de forma e escala, respectivamente. A distribuição Weibull exponencializada (WE), então, possui distribuição e densidade dadas por

$$H_a(x) = \{1 - \exp[-(x/\beta)^\alpha]\}^a \quad \text{e} \quad h_a(x) = \frac{a \alpha x^{\alpha-1}}{\beta^\alpha} \exp[-(x/\beta)^\alpha] \{1 - \exp[-(x/\beta)^\alpha]\}^{a-1}.$$

A função taxa de falha ou função de risco é definida como $r(x) = f(x)/\bar{F}(x)$, em que $\bar{F}(x) = 1 - F(x)$ é a função de sobrevivência. Para a distribuição Weibull exponencializada, a função de risco é dada por

$$r(x) = \frac{a \alpha x^{\alpha-1} \exp[-(x/\beta)^\alpha] \{1 - \exp[-(x/\beta)^\alpha]\}^{a-1}}{\beta^\alpha \{1 - \{1 - \exp[-(x/\beta)^\alpha]\}^a\}}.$$

As Figuras 1.1 e 1.2 mostram a densidade e a taxa de falha da distribuição Weibull exponencializada para alguns valores dos parâmetros. Na Figura 1.2, nota-se que a distribuição Weibull exponencializada possui função taxa de falha unimodal e em forma de U, enquanto a taxa de falha da distribuição Weibull apenas assume forma crescente, decrescente e constante.

Gupta e Kundu (1999) investigaram a distribuição exponencial exponencializada. A distribuição exponencial possui função distribuição e função densidade iguais a

$$G(x) = 1 - \exp(-x/\beta) \quad \text{e} \quad g(x) = \frac{1}{\beta} \exp(-x/\beta), \quad x > 0, \quad (1.4)$$

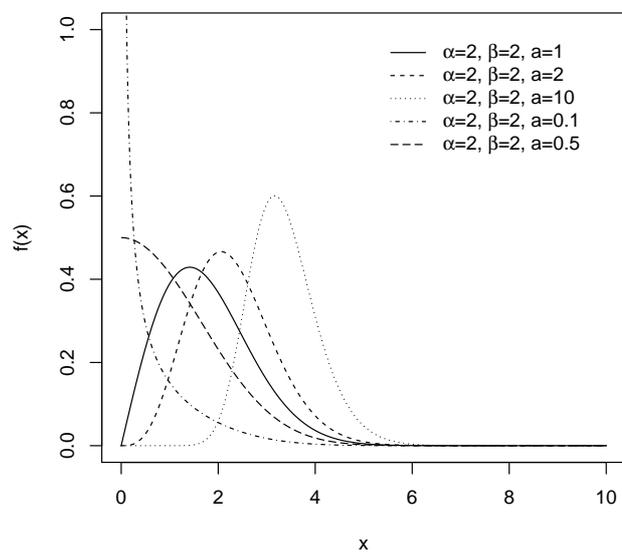


Figura 1.1: Densidade da distribuição Weibull exponencializada para $\alpha = 2$, $\beta = 2$ e alguns valores de a .

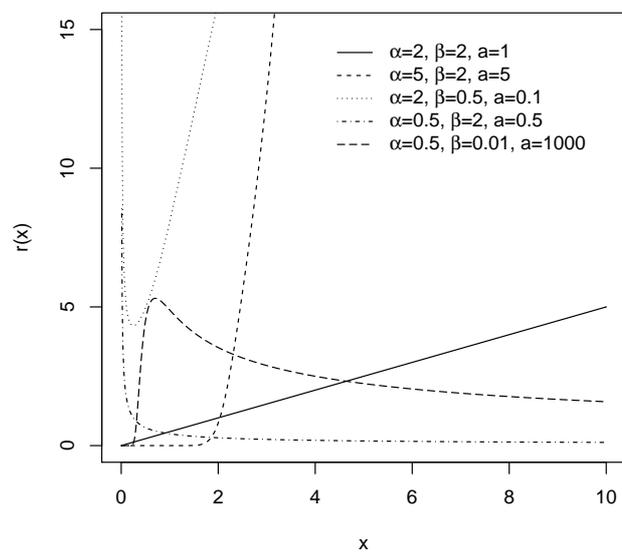


Figura 1.2: Taxa de falha da distribuição Weibull exponencializada para alguns valores dos parâmetros.

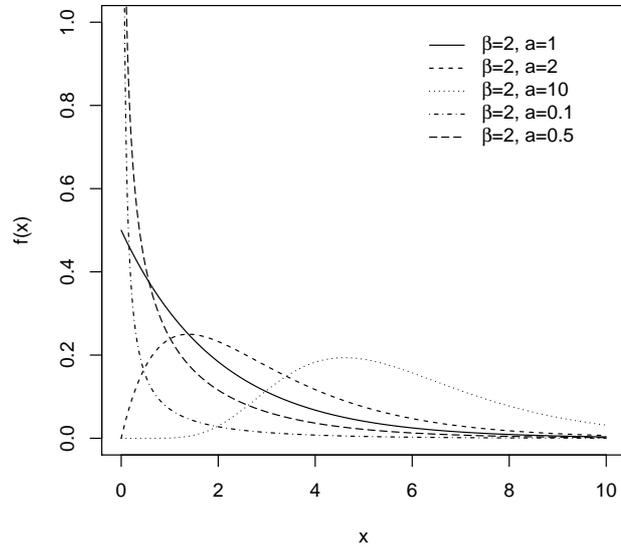


Figura 1.3: Densidade da distribuição exponencial exponencializada para $\beta = 2$ e alguns valores de a .

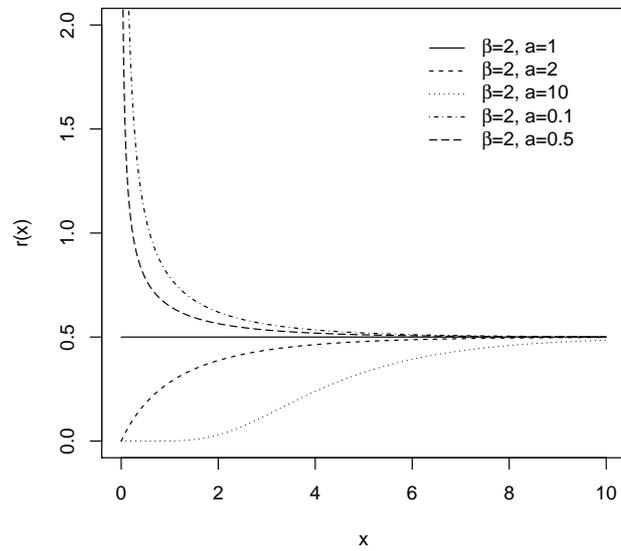


Figura 1.4: Taxa de falha da distribuição exponencial exponencializada para $\beta = 2$ e alguns valores de a .

respectivamente, em que $\beta > 0$ é o parâmetro de escala. A distribuição Weibull, para $\alpha = 1$, é equivalente a distribuição exponencial. A distribuição e densidade da exponencial exponencializada são

$$H_a(x) = \{1 - \exp(-x/\beta)\}^a \quad \text{e} \quad h_a(x) = \frac{a}{\beta} \exp(-x/\beta) \{1 - \exp(-x/\beta)\}^{a-1},$$

respectivamente. Da mesma forma, para $\alpha = 1$, a distribuição Weibull exponencializada é equivalente a distribuição exponencial exponencializada.

A função taxa de falha para a distribuição exponencial exponencializada é

$$r(x) = \frac{a \exp(-x/\beta) [1 - \exp(-x/\beta)]^{a-1}}{\beta \{1 - [1 - \exp(-x/\beta)]^a\}}.$$

As Figuras 1.3 e 1.4 mostram a densidade e a taxa de falha da distribuição exponencial exponencializada para $\beta = 2$ e alguns valores de a . Na Figura 1.4, nota-se que esta distribuição tem função taxa de falha com forma crescente e decrescente, enquanto a taxa de falha da distribuição exponencial é sempre constante.

1.1 Composições

Além da classe G-exponencializada, outro método popular para incluir novos parâmetros em distribuições conhecidas envolve a composição de distribuições. Neste trabalho será considerado apenas a composição de distribuições contínuas com distribuições discretas. Seja $F(x|Z = z)$ a função distribuição condicional de X dado que a variável aleatória discreta Z teve como valor observado z . Então, a distribuição incondicional de X , $F(x)$, é dada por

$$F(x) = \sum_{z=0}^{\infty} F(x|Z = z)P(Z = z).$$

Considerando-se um sistema que possui um número aleatório de componentes, Z , e que o tempo de falha de cada componente, W , é independente de Z e identicamente distribuído, com função distribuição $G(\cdot)$. Dessa forma, caso o sistema pare de funcionar assim que o primeiro componente falhar (sistema em série), tem-se

$$X = \min\{W_1, W_2, \dots, W_Z\} \quad \text{e} \quad \bar{F}(x|Z = z) = \bar{G}(x)^z. \quad (1.5)$$

Por outro lado, caso o sistema pare apenas quando a última componente falhar (sistema em paralelo), tem-se

$$X = \max\{W_1, W_2, \dots, W_Z\} \quad \text{e} \quad F(x|Z = z) = G(x)^z. \quad (1.6)$$

Também é possível definir $F(x|Z = z)$ de diversas outras formas. Contudo, o modelo (1.5) é o mais comumente utilizado para construir distribuições compostas. Nas próximas seções serão abordadas distribuições contruídas usando essa metodologia.

1.1.1 Exponencial G

A distribuição geométrica tem função de probabilidade dada por

$$P(Z = z) = (1 - p)p^{z-1}, \quad p \in (0, 1), \quad z = 1, 2, \dots \quad (1.7)$$

Dessa forma, Adamidis e Loukas (1998), usando (1.5) para compor (1.4) com (1.7), introduziram a distribuição exponencial geométrica que possui função distribuição e densidade dadas por

$$F(x) = \frac{1 - \exp(-x/\beta)}{1 - p \exp(-x/\beta)} \quad \text{e} \quad f(x) = \frac{(1 - p) \exp(-x/\beta)}{\beta[1 - p \exp(-x/\beta)]^2},$$

respectivamente. Para $p = 0$ a distribuição exponencial geométrica se reduz à distribuição exponencial.

A função taxa de risco, para a distribuição exponencial geométrica, é dada por

$$r(x) = \{\beta[1 - p \exp(-x/\beta)]\}^{-1}.$$

As Figuras 1.5 e 1.6 mostram a densidade e a taxa de falha da distribuição exponencial geométrica para alguns valores dos parâmetros. Na Figura 1.6, nota-se que a distribuição exponencial geométrica possui função taxa de falha com forma decrescente, enquanto a taxa de falha da distribuição exponencial é sempre constante.

A distribuição Poisson, truncada em zero, tem função de probabilidade dada por

$$P(Z = z) = \theta^z [z!(e^\theta - 1)]^{-1}, \quad \theta > 0, \quad z = 1, 2, \dots \quad (1.8)$$

Kuş (2007), utilizando (1.5), com $G(x)$ dado por (1.4) e Z por (1.8), investigou a distribuição exponencial Poisson. A distribuição e densidade dela são iguais a

$$F(x) = \frac{e^{\theta \exp(-x/\beta)} - e^\theta}{1 - e^\theta} \quad \text{e} \quad f(x) = \frac{\theta e^{-\theta - x/\beta + \theta \exp(-x/\beta)}}{\beta(1 - e^{-\theta})},$$

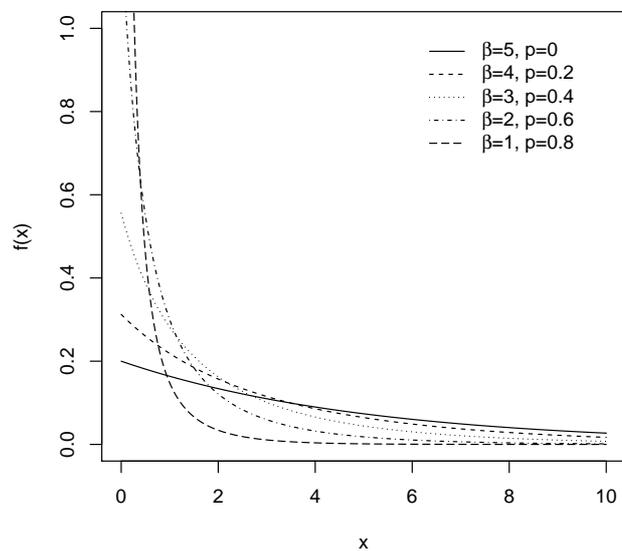


Figura 1.5: Densidade da distribuição exponencial geométrica para alguns valores dos parâmetros.

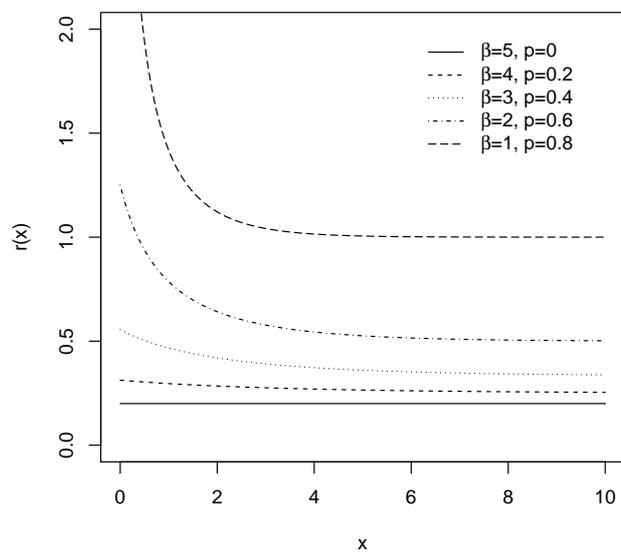


Figura 1.6: Taxa de falha da distribuição exponencial geométrica para alguns valores dos parâmetros.

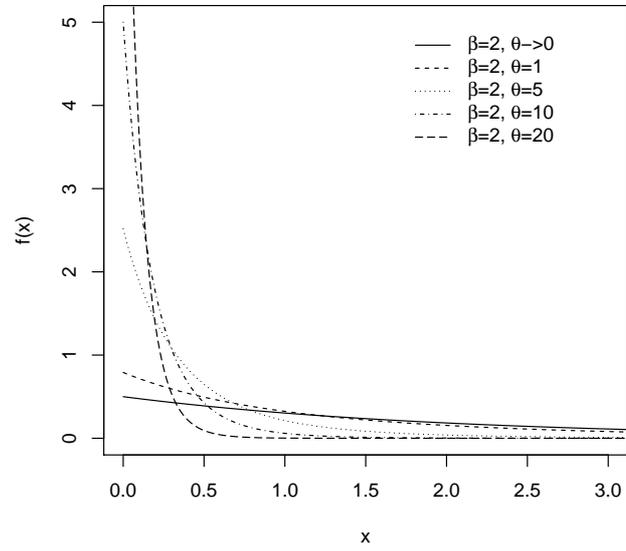


Figura 1.7: Densidade da distribuição exponencial Poisson para $\beta = 2$ e alguns valores de θ .

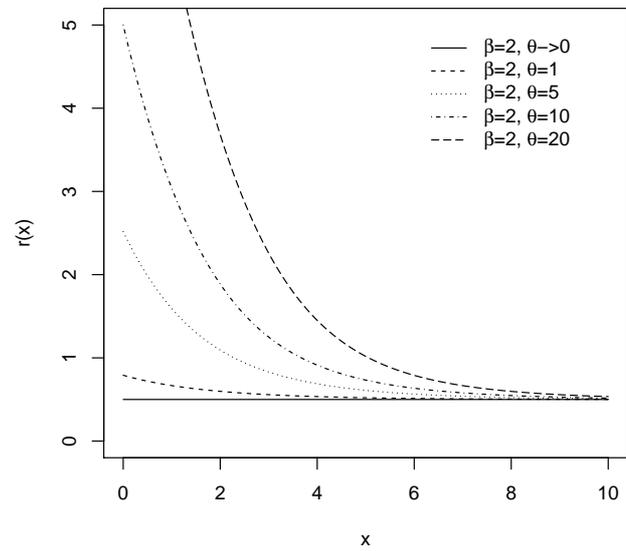


Figura 1.8: Taxa de falha da distribuição exponencial Poisson para $\beta = 2$ e alguns valores de θ .

respectivamente. Quando $\theta \rightarrow 0$, a distribuição exponencial Poisson converge para a distribuição exponencial.

A função taxa de risco, para a distribuição exponencial Poisson, é dada por

$$r(x) = \frac{\theta(1 - e^{-\theta})e^{-\theta - x/\beta + \theta \exp(-x/\beta)}}{\beta(1 - e^{-\theta})(1 - e^{\theta \exp(-x/\beta)})}.$$

As Figuras 1.7 e 1.8 mostram a densidade e a taxa de falha da distribuição exponencial Poisson para $\beta = 2$ e alguns valores de θ . Na Figura 1.8, nota-se que a distribuição exponencial Poisson possui função taxa de falha com forma decrescente, enquanto a taxa de falha da distribuição exponencial é sempre constante.

A distribuição logarítmica tem função de probabilidade dada por

$$P(Z = z) = \frac{(1 - q)^z}{-z \log(q)}, \quad q \in (0, 1), \quad z = 1, 2, \dots \quad (1.9)$$

Partindo do modelo (1.5), Tahmasbi e Rezaei (2008) construíram a distribuição exponencial logarítmica, em que $G(x)$ é dado por (1.4) e $P(Z = z)$ por (1.9). A função distribuição e função densidade da distribuição exponencial logarítmica são dadas por

$$F(x) = 1 - \frac{\log[1 - (1 - q)e^{-x/\beta}]}{\log(q)} \quad \text{e} \quad f(x) = \frac{(1 - q)e^{-x/\beta}}{-\beta \log(q)[1 - (1 - q)e^{-x/\beta}]},$$

respectivamente. A distribuição exponencial logarítmica se aproxima da distribuição exponencial com parâmetro de escala β quando $q \rightarrow 1$.

A função taxa de risco correspondente é dada por

$$r(x) = \frac{-(1 - q)e^{-x/\beta}}{\beta[1 - (1 - q)e^{-x/\beta}] \log[1 - (1 - q)e^{-x/\beta}]}.$$

As Figuras 1.9 e 1.10 mostram a densidade e a taxa de falha da distribuição exponencial logarítmica para $\beta = 2$ e alguns valores de q . Na Figura 1.10, nota-se que a distribuição exponencial logarítmica possui função taxa de falha com forma decrescente, enquanto a taxa de falha da distribuição exponencial é sempre constante.

Adicionalmente, Chahkandi e Ganjali (2009) propuseram a classe exponencial série de potências, que generaliza as distribuições apresentadas nesta seção, além de possuir diversos outros submodelos. Diz-se que uma distribuição discreta pertence à classe série de potências, truncada em zero, se sua densidade for escrita da seguinte forma

$$P(Z = z) = \frac{c_z \phi^z}{C(\phi)}, \quad z = 1, 2, \dots, \quad (1.10)$$

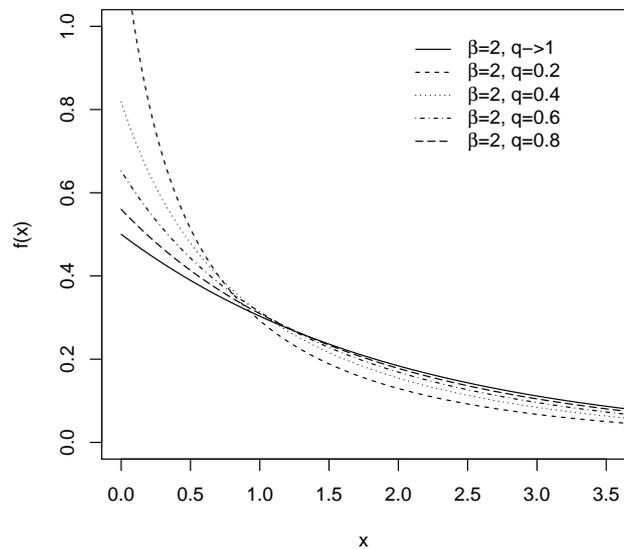


Figura 1.9: Densidade da distribuição exponencial logarítmica para $\beta = 2$ e alguns valores de q .

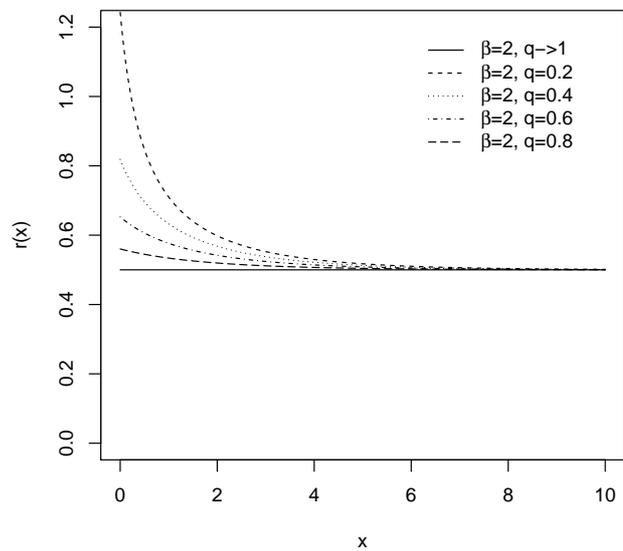


Figura 1.10: Taxa de falha da distribuição exponencial logarítmica para $\beta = 2$ e alguns valores de q .

em que $c_z \geq 0$ depende apenas de z , $C(\phi) = \sum_{z=1}^{\infty} c_z \phi^z$ e $\phi \in (0, s)$, com s escolhido de forma que $C(\phi)$ seja sempre finito.

As distribuições geométrica, Poisson truncada em zero, logarítmica e outras, como a binomial truncada em zero e a binomial negativa truncada em zero, são casos especiais de (1.10). Dessa forma, seguindo (1.5), a distribuição e densidade da família exponencial série de potências, respectivamente, são dadas por

$$F(x) = 1 - \frac{C(\phi e^{-x/\beta})}{C(\phi)} \quad \text{e} \quad f(x) = \frac{\phi e^{-x/\beta}}{\beta C(\phi)} \frac{\partial}{\partial u} [C(u)]_{u=\phi e^{-x/\beta}},$$

e tem como casos especiais as distribuições exponencial geométrica, exponencial Poisson e exponencial logarítmica.

1.1.2 Weibull G

Barreto-Souza et al. (2011) estudaram a distribuição Weibull geométrica. Partindo do modelo (1.5), com $P(Z = z)$ dado por (1.7), obtém-se a distribuição e a densidade da Weibull geométrica como

$$F(x) = \frac{1 - \exp[-(x/\beta)^\alpha]}{1 - p \exp[-(x/\beta)^\alpha]} \quad \text{e} \quad f(x) = \frac{\alpha(1-p)x^{\alpha-1} \exp[-(x/\beta)^\alpha]}{\beta^\alpha \{1 - p \exp[-(x/\beta)^\alpha]\}^2},$$

respectivamente. Para $p = 0$, a distribuição Weibull geométrica se reduz à distribuição Weibull e, para $\alpha = 1$, tem-se que a Weibull geométrica é equivalente à exponencial geométrica. Adicionalmente, Barreto-Souza et al. (2011) notaram que, mesmo para valores de $p < 0$, a densidade da Weibull geométrica continua sendo uma densidade.

A função taxa de risco, para a distribuição Weibull geométrica, é dada por

$$r(x) = \frac{\alpha x^{\alpha-1}}{\beta^\alpha \{1 - p \exp[-(x/\beta)^\alpha]\}}.$$

As Figuras 1.11 e 1.12 mostram a densidade e a taxa de falha da distribuição Weibull geométrica para alguns valores dos parâmetros. Na Figura 1.12, nota-se que a distribuição Weibull geométrica possui função taxa de falha com forma crescente-decrescente-crescente, enquanto a taxa de falha da distribuição Weibull é sempre monótona.

Lu e Shi (2012) investigaram a distribuição Weibull Poisson. Utilizando o modelo (1.5) e escolhendo como variável discreta a distribuição Poisson com parâmetro θ , obtém-se a distribuição e a densidade da Weibull Poisson

$$F(x) = \frac{e^{\theta \exp[-(x/\beta)^\alpha]} - e^\theta}{1 - e^\theta} \quad \text{e} \quad f(x) = \frac{\theta \alpha x^{\alpha-1} e^{-\theta - (x/\beta)^\alpha + \theta \exp[-(x/\beta)^\alpha]}}{\beta^\alpha (1 - e^{-\theta})},$$

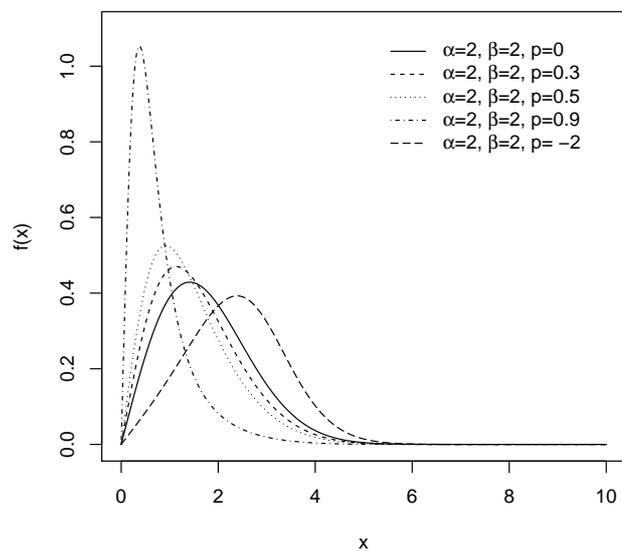


Figura 1.11: Densidade da distribuição Weibull geométrica para $\alpha = 2$, $\beta = 2$ e alguns valores de p .

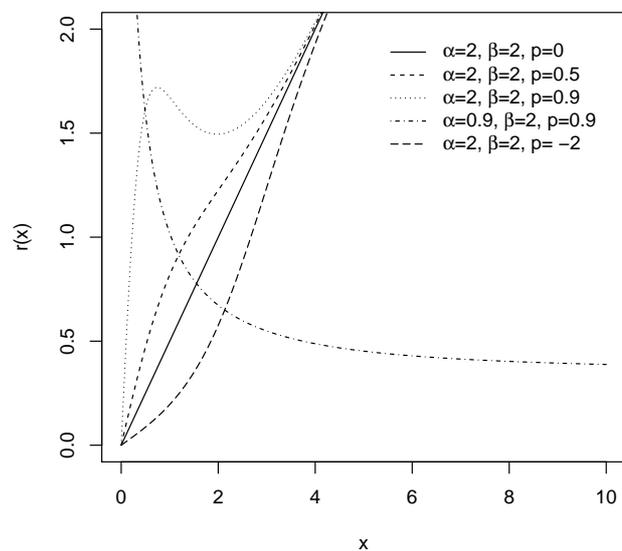


Figura 1.12: Taxa de falha da distribuição Weibull geométrica para alguns valores do parâmetros.

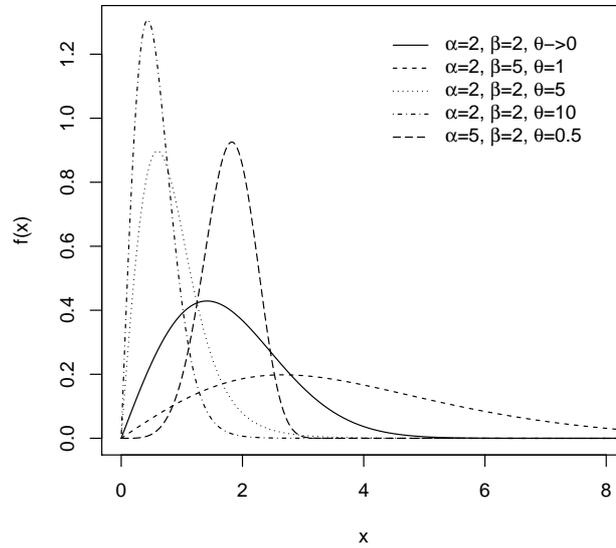


Figura 1.13: Densidade da distribuição Weibull Poisson para alguns valores dos parâmetros.

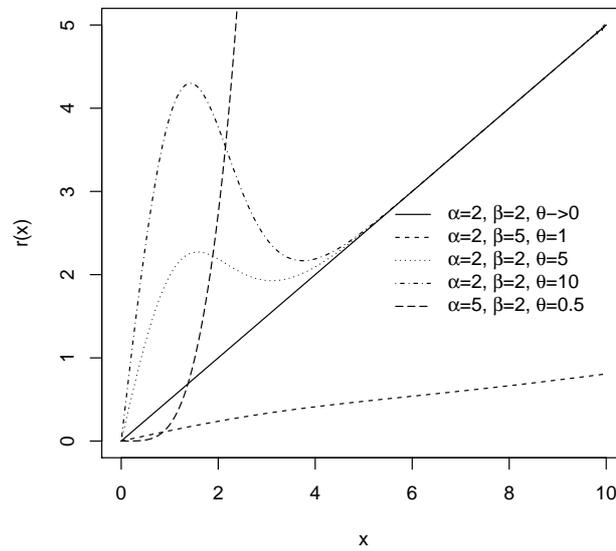


Figura 1.14: Taxa de falha da distribuição Weibull Poisson para alguns valores dos parâmetros.

respectivamente. Quando $\theta \rightarrow 0$, a distribuição Weibull Poisson converge para a distribuição Weibull e, caso $\alpha = 1$, a distribuição Weibull Poisson é equivalente à distribuição exponencial Poisson.

A função taxa de risco, para a distribuição Weibull Poisson, é dada por

$$r(x) = \frac{\theta \alpha x^{\alpha-1} e^{-(x/\beta)^\alpha + \theta \exp[-(x/\beta)^\alpha]}}{\beta^\alpha (e^{\theta \exp[-(x/\beta)^\alpha]} - 1)}.$$

As Figuras 1.13 e 1.14 mostram a densidade e a taxa de falha da distribuição Weibull Poisson para alguns valores dos parâmetros. Na Figura 1.14, nota-se que a distribuição Weibull Poisson possui função taxa de falha com forma crescente-decrescente-crescente, enquanto a taxa de falha da distribuição Weibull é sempre monótona.

Ciumara e Preda (2009) introduziram a distribuição Weibull logarítmica, cuja distribuição e densidade são

$$F(x) = 1 - \frac{\log[1 - (1 - q)e^{-(x/\beta)^\alpha}]}{\log(q)} \quad \text{e} \quad f(x) = \frac{\alpha(1 - q)x^{\alpha-1}e^{-(x/\beta)^\alpha}}{-\beta^\alpha \log(q)[1 - (1 - q)e^{-(x/\beta)^\alpha}]},$$

respectivamente. Esta distribuição é obtida ao se compor (1.3) com (1.9) através de (1.5). Quando $q \rightarrow 1$, a distribuição Weibull logarítmica se aproxima da distribuição Weibull e, para $\alpha = 0$, ela é idêntica à distribuição exponencial logarítmica.

A função taxa de falha é dada por

$$r(x) = \frac{-\alpha(1 - q)x^{\alpha-1}e^{-(x/\beta)^\alpha}}{\beta^\alpha [1 - (1 - q)e^{-(x/\beta)^\alpha}] \log[1 - (1 - q)e^{-(x/\beta)^\alpha}]}$$

As Figuras 1.15 e 1.16 mostram a densidade e a taxa de falha da distribuição Weibull logarítmica para alguns valores dos parâmetros. Na Figura 1.16 nota-se que a distribuição Weibull logarítmica possui função taxa de falha com forma crescente-decrescente-crescente, enquanto a taxa de falha da distribuição Weibull é sempre monótona.

Morais e Barreto-Souza (2011) definiram a classe Weibull série de potências, cuja distribuição e densidade são representadas por

$$F(x) = 1 - \frac{C(\phi e^{-(x/\beta)^\alpha})}{C(\phi)} \quad \text{e} \quad f(x) = \frac{\phi \alpha x^{\alpha-1} e^{-(x/\beta)^\alpha}}{\beta^\alpha C(\phi)} \frac{\partial}{\partial u} [C(u)]_{u=\phi e^{-(x/\beta)^\alpha}},$$

respectivamente, com $C(\cdot)$ dado por (1.10). As distribuições Weibull geométrica, Weibull Poisson e Weibull logarítmica são casos especiais da classe Weibull série de potências.

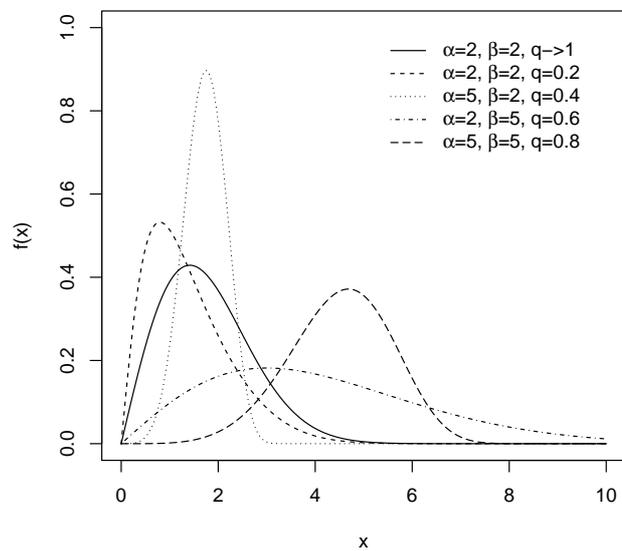


Figura 1.15: Densidade da distribuição Weibull logarítmica para alguns valores dos parâmetros.

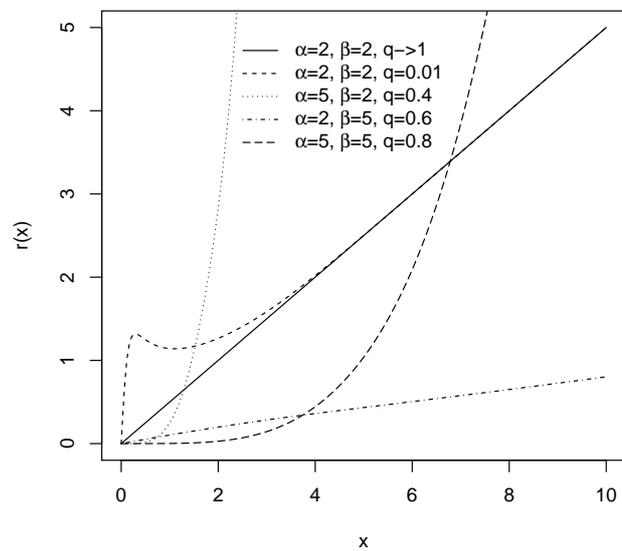


Figura 1.16: Taxa de falha da distribuição Weibull logarítmica para alguns valores dos parâmetros.

1.1.3 Outros métodos de composição

Além das composições abordadas anteriormente, diversos autores investigaram outras distribuições compostas. Barreto-Souza e Bakouchb (2011), por exemplo, estudaram a distribuição exponencial Poisson-Lindley. Esta distribuição é obtida ao se compor uma distribuição exponencial com a distribuição Poisson-Lindley, na qual $P(Z = z) = [\gamma^2(2 + \gamma + z)][(1 + 3\gamma + \gamma^2)(1 + \gamma)^z]^{-1}$, $\gamma > 0$, $z = 1, 2, \dots$, utilizando o método (1.5). A distribuição exponencial Poisson-Lindley possui função taxa de falha decrescente.

A distribuição exponencial exponencializada foi investigada em sua versão composta com a distribuição Poisson por Ristić e Nadarajah (2010) e, quando composta com a distribuição binomial, por Bakouch et al. (2012). Pertencente a classe Weibull série de potências, a distribuição Weibull binomial negativa foi estudada por Rodrigues et al. (2011).

Nos próximos capítulos desta dissertação, serão abordados outros métodos de composição. No Capítulo 2, a classe de distribuições proposta por Marshall e Olkin (1997) é investigada e diversas propriedades da distribuição Marshall-Olkin estendida Lomax são deduzidas pela primeira vez. No Capítulo 3, uma nova classe de distribuições que generaliza os trabalhos de Kuş (2007) e Lu e Shi(2012), além de possuir como casos especiais diversos outros modelos, é estudada. No Capítulo 4, a distribuição Lomax-Poisson é estudada utilizando a teoria desenvolvida no Capítulo 3. No Capítulo 5, são estabelecidas algumas considerações finais.

A classe Marshall-Olkin estendida G

Neste capítulo revisa-se a classe de distribuições proposta por Marshall e Olkin (1997) e algumas de suas propriedades. Diversas propriedades matemáticas inéditas da distribuição Marshall-Olkin Lomax, inicialmente abordada por Ghitany et al. (2007), também são apresentadas.

2.1 A classe Marshall-Olkin

O trabalho de Marshall e Olkin (1997) propôs um novo método para adicionar um parâmetro de forma a uma variável aleatória conhecida. Seja $\overline{G}(x)$ uma função de sobrevivência qualquer. A nova distribuição terá sobrevivência dada por

$$\overline{F}(x) = \frac{\lambda \overline{G}(x)}{1 - \overline{\lambda} \overline{G}(x)} = \frac{\lambda \overline{G}(x)}{G(x) - \lambda \overline{G}(x)}, \quad 0 < \lambda < \infty, \quad (2.1)$$

em que $\overline{\lambda} = 1 - \lambda$ e $G(x) = 1 - \overline{G}(x)$ é a função de distribuição da *baseline*. Para $\lambda = 1$, tem-se que $\overline{F}(x) = \overline{G}(x)$.

É possível motivar (2.1) da seguinte forma:

- $0 < \lambda < 1$: Usando o modelo (1.5), com $Z \sim \text{Geométrica}(\lambda)$, em que $P(Z = z)$ é dado por (1.7), então X terá distribuição dada por (2.1); e
- $\lambda > 1$: Usando o modelo (1.6), com $Z \sim \text{Geométrica}(\lambda^{-1})$, em que $P(Z = z)$ é dado por (1.7), então X terá distribuição dada por (2.1).

A densidade da classe Marshall-Olkin estendida G, abreviada como MOEG, em que G faz referência à distribuição original, é

$$f(x) = \frac{\lambda g(x)}{[1 - \bar{\lambda} \bar{G}(x)]^2}. \quad (2.2)$$

Sua função de risco é dada por

$$r(x) = \frac{r_G(x)}{1 - \bar{\lambda} \bar{G}(x)}, \quad (2.3)$$

em que $r_G(x) = g(x)/\bar{G}(x)$ é a função de risco da *baseline*. Partindo de (2.3), Marshall e Olkin (1997) mostraram que

$$r_G(x) \leq r(x) \leq r_G(x)/\lambda, \quad 0 < \lambda \leq 1, \text{ e}$$

$$r_G(x)/\lambda \leq r(x) \leq r_G(x), \quad \lambda \geq 1.$$

A distribuição Marshall-Olkin estendida exponencial (MOEE), obtida de (2.1) e (1.4), tem distribuição e densidade dadas por

$$F(x) = 1 - \frac{1}{e^{x/\beta} - \bar{\lambda}} \text{ e } f(x) = \frac{\lambda e^{x/\beta}}{\beta(e^{x/\beta} - \bar{\lambda})^2},$$

respectivamente. A sua função taxa de falha é

$$r(x) = \frac{e^{x/\beta}}{\beta(e^{x/\beta} - \bar{\lambda})}.$$

As Figuras 2.1 e 2.2 mostram a densidade e a taxa de falha da distribuição MOEE para $\beta = 2$ e alguns valores de λ . Na Figura 2.2, nota-se que essa distribuição tem função taxa de falha com forma crescente ou decrescente, enquanto a taxa de falha da distribuição exponencial é sempre constante.

A distribuição Marshall-Olkin estendida Weibull (MOEW), obtida de (2.1) e (1.3), tem distribuição e densidade dadas por

$$F(x) = 1 - \frac{\lambda e^{-(x/\beta)^\alpha}}{1 - \bar{\lambda} e^{-(x/\beta)^\alpha}} \text{ e } f(x) = \frac{\lambda \alpha x^{\alpha-1} e^{-(x/\beta)^\alpha}}{\beta^\alpha [1 - \bar{\lambda} e^{-(x/\beta)^\alpha}]^2}.$$

Sua função taxa de falha é

$$r(x) = \frac{\alpha x^{\alpha-1}}{\beta^\alpha [1 - \bar{\lambda} e^{-(x/\beta)^\alpha}]}.$$

As Figuras 2.3 e 2.4 mostram a densidade e a taxa de falha da distribuição MOEW para alguns valores dos parâmetros. Na Figura 2.4, nota-se que essa distribuição tem função taxa de falha em forma crescente-decrescente-crescente, enquanto a taxa de falha da distribuição Weibull é sempre monótona.

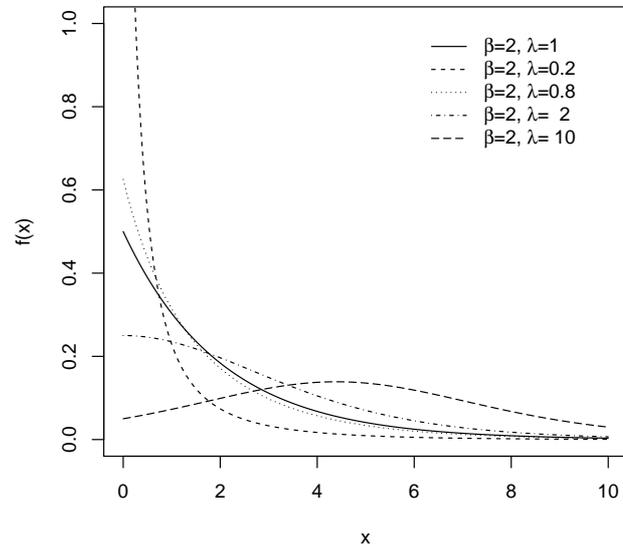


Figura 2.1: Densidade da distribuição MOEE para $\beta = 2$ e alguns valores de λ .

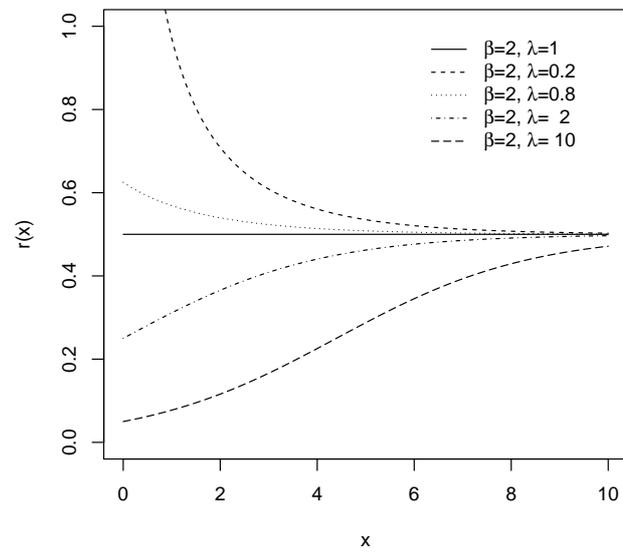


Figura 2.2: Taxa de falha da distribuição MOEE para $\beta = 2$ e alguns valores de λ .

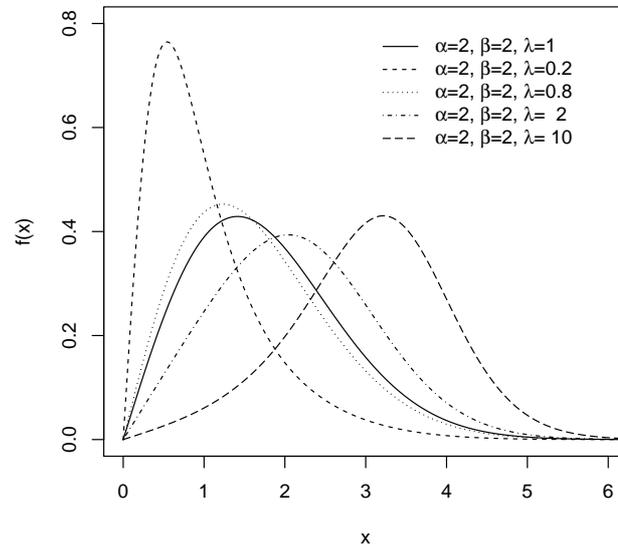


Figura 2.3: Densidade da distribuição MOEW para $\alpha = 2, \beta = 2$ e alguns valores de λ .

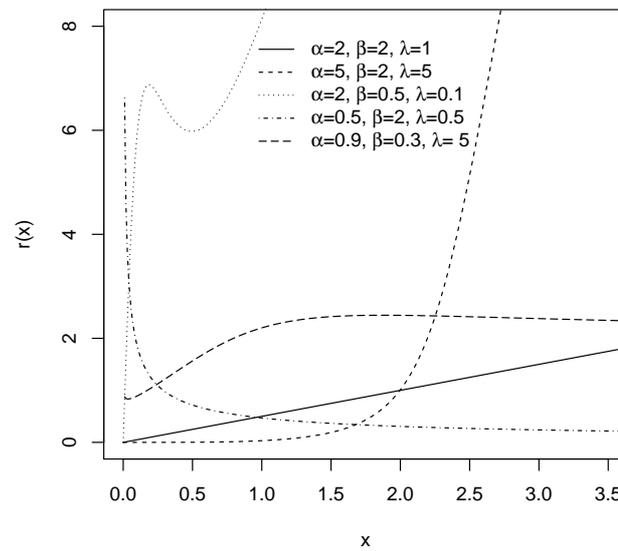


Figura 2.4: Taxa de falha da distribuição MOEW para alguns valores dos parâmetros.

2.2 A distribuição Marshall-Olkin estendida Lomax

A distribuição Lomax (Lomax, 1954), por vezes, também, referida como distribuição Pareto Tipo II, tem função de distribuição e densidade, respectivamente, dadas por

$$G(x) = 1 - (1 + x/\beta)^{-\alpha} \quad \text{e} \quad g(x) = \alpha\beta^{-1}(1 + x/\beta)^{-(\alpha+1)}, \quad x > 0, \quad (2.4)$$

em que $\alpha > 0$ e $\beta > 0$ são os parâmetros de forma e de escala, respectivamente. Ghitany et al. (2007) investigaram a distribuição Marshall-Olkin estendida Lomax (MOEL), obtida de (2.1) e (2.4). Ela tem função distribuição dada por

$$F(x) = 1 - \frac{\lambda}{(1 + x/\beta)^\alpha - \bar{\lambda}} \quad (2.5)$$

e função densidade

$$f(x) = \frac{\lambda \alpha (1 + x/\beta)^{\alpha-1}}{\beta[(1 + x/\beta)^\alpha - \bar{\lambda}]^2}. \quad (2.6)$$

A Figura 2.5 apresenta a densidade da distribuição MOEL para alguns valores dos parâmetros.

Ghitany et al. (2007) demonstraram que, para $\lambda + \alpha(2 - \lambda) < 0$, a equação (2.6) é unimodal, com moda igual a

$$\text{Moda} = \beta \left[\left(\frac{\bar{\lambda} \bar{\alpha}}{1 + \alpha} \right)^{1/\alpha} - 1 \right],$$

em que $\bar{\alpha} = 1 - \alpha$. Para $\lambda + \alpha(2 - \lambda) \geq 0$, (2.6) é decrescente.

A função taxa de falha é dada por

$$r(x) = \frac{\alpha}{\beta[(1 + x/\beta) - \bar{\lambda}(1 + x/\beta)^{1-\alpha}]}. \quad (2.7)$$

Ghitany et al. (2007) demonstraram que a equação (2.7) é unimodal para $\lambda + \alpha(1 - \lambda) < 0$, caso contrário, ela é decrescente. A Figura 2.6 apresenta a função taxa de falha da distribuição MOEL para alguns valores dos parâmetros.

A função quantílica é dada por

$$Q_F(q) = F^{-1}(q) = \beta \left[\left(1 + \frac{q\lambda}{1-q} \right)^{1/\alpha} - 1 \right], \quad 0 \leq q \leq 1. \quad (2.8)$$

A equação (2.8) pode ser utilizada facilmente para gerar números aleatórios da distribuição MOEL, bastando substituir q por $U \sim \text{Uniforme}(0, 1)$.

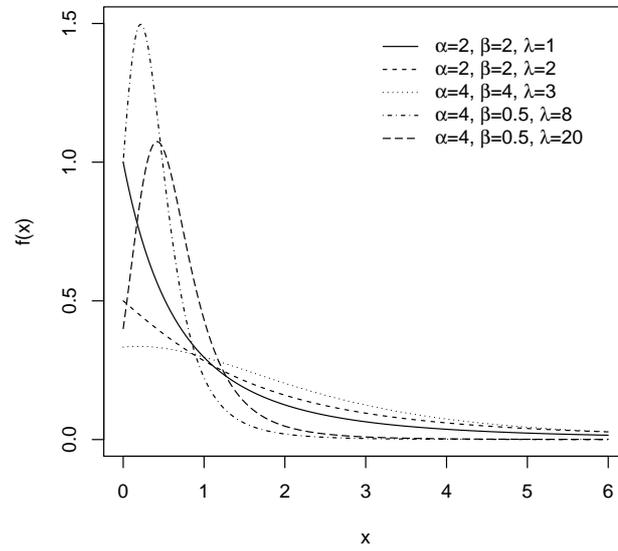


Figura 2.5: Densidade da distribuição MOEL para alguns valores dos parâmetros.

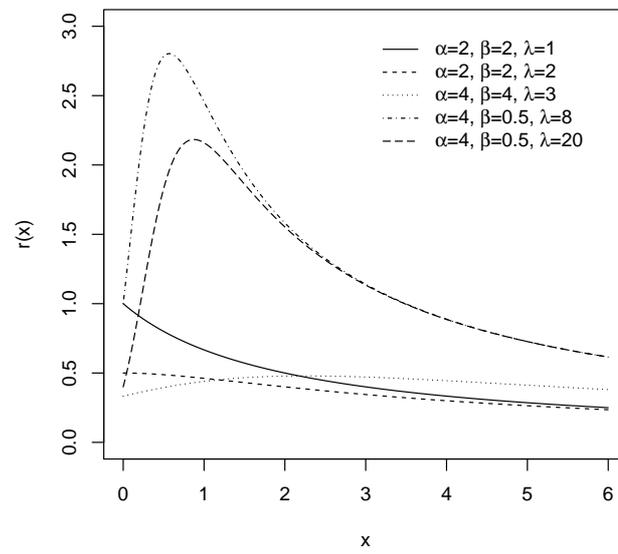


Figura 2.6: Taxa de falha da distribuição MOEL para alguns valores dos parâmetros.

A mediana é obtida da equação (2.8) fazendo $q = 1/2$:

$$\text{Mediana} = M = \beta[(1 + \lambda)^{1/\alpha} - 1].$$

2.3 Algumas propriedades

As propriedades apresentadas na Seção 2.2 foram inicialmente obtidas por Ghitany et al. (2007). Os autores também derivaram outras propriedades como os momentos ordinários e a distribuição dos extremos amostrais, por exemplo.

As propriedades obtidas nesta seção para a distribuição MOEL são inéditas na literatura. Adicionalmente um método alternativo para obtenção dos momentos ordinários é apresentado.

2.3.1 Expansão para a distribuição

Considere a identidade

$$(1 - z)^{-\rho} = \sum_{k=0}^{\infty} \frac{\Gamma(\rho + k)}{\Gamma(\rho) k!} z^k, \quad |z| < 1,$$

em que $\Gamma(a) = \int_{t=0}^{\infty} e^{-t} t^{a-1} dt$ é a função gama. Dessa forma, para $\lambda < 1$, aplicando a identidade em (2.2), obtém-se

$$f(x) = \lambda g(x) \sum_{k=0}^{\infty} \frac{\Gamma(k + 2)}{\Gamma(2) k!} \bar{\lambda}^k \bar{G}(x)^k.$$

Uma vez que $\Gamma(a) = (a - 1)!$, para a inteiro, a densidade pode, então, ser rescrita como

$$f(x) = \sum_{k=0}^{\infty} (k + 1) \lambda \bar{\lambda}^k g(x) \bar{G}(x)^k.$$

Dado o fato que, se $g(x)$ e $\bar{G}(x)$ são a densidade e a função sobrevivência de uma Lomax(α, β), então

$$g(x) \bar{G}(x)^k = (k + 1)^{-1} g_{(k+1)\alpha, \beta}(x), \quad (2.9)$$

em que $g_{(k+1)\alpha, \beta}(x)$ é a densidade da Lomax com parâmetro de forma $(k + 1)\alpha$ e parâmetro de escala β . Assim, é possível reescrever (2.6) como

$$f(x) = \sum_{k=0}^{\infty} v_k(\lambda) g_{(k+1)\alpha, \beta}(x), \quad (2.10)$$

em que

$$v_k(\lambda) = \begin{cases} \lambda \bar{\lambda}^k, & \text{para } 0 < \lambda < 1, \text{ e} \\ [\lambda(k+1)]^{-1} \sum_{j=k}^{\infty} (-1)^k (j+1) \binom{j}{k} (1-\lambda^{-1})^j, & \text{para } \lambda > 1. \end{cases}$$

A obtenção de $v_k(\lambda)$ para $\lambda > 1$ é similar a apresentada. É necessário apenas notar que a densidade dada por (2.2) pode ser reescrita como

$$f(x) = \frac{\lambda^{-1} g(x)}{[1 - (1 - \lambda^{-1})G(x)]^2}.$$

Para aplicar o resultado (2.9) é necessário usar do fato que $G(x) = 1 - \bar{G}(x)$ e utilizar a expansão binomial.

É fácil verificar que $\sum_{k=0}^{\infty} v_k(\lambda) = 1$. Dessa forma, a equação (2.10) revela que a densidade da distribuição MOEL pode ser expressa como uma combinação linear infinita de distribuições Lomax. Assim, diversas propriedades da distribuição gerada podem ser obtidas diretamente das mesmas propriedades da distribuição Lomax. Essas propriedades podem ser encontradas em Lemonte e Cordeiro (2011), por exemplo.

2.3.2 Momentos e Geratriz de Momentos

O p -ésimo momento ordinário de uma variável aleatória X é dado por

$$\mu'_p = E(X^p) = \int_{\mathbb{R}} x^p f(x) dx. \quad (2.11)$$

Tais momentos são importantes, pois são utilizados para obter medidas de tendência, variabilidade e forma. No restante deste capítulo, considera-se que $Y \sim \text{Lomax}(\alpha, \beta)$. O p -ésimo momento ordinário de Y é dado por

$$E(Y^p) = \frac{\beta \Gamma(\alpha - p) \Gamma(p + 1)}{\Gamma(\alpha)}, \quad p < \alpha. \quad (2.12)$$

Utilizando a expansão (2.10), pode-se calcular os momentos ordinários da distribuição gerada por

$$\mu'_p = E(X^p) = \beta \Gamma(p + 1) \sum_{k=0}^{\infty} \frac{v_k(\lambda) \Gamma[\alpha(k+1) - p]}{\Gamma[\alpha(k+1)]}, \quad p < \alpha. \quad (2.13)$$

Os momentos incompletos, importantes para medir desigualdade, são dados por

$$m_p(X; z) = \int_{-\infty}^z x^p f(x) dx. \quad (2.14)$$

Substitutindo (2.10) em (2.14) tem-se

$$m_p(X; z) = \sum_{k=0}^{\infty} v_k m_p(Y_{(k+1)\alpha, \beta}; z), \quad (2.15)$$

em que $m_p(Y_{(k+1)\alpha, \beta}; z)$ é o momento incompleto de $Y_{(k+1)\alpha, \beta} \sim \text{Lomax}((k+1)\alpha, \beta)$. Pode-se utilizar o MAPLE para calcular os momentos incompletos da distribuição Lomax como

$$m_p(Y_{(k+1)\alpha, \beta}; z) = P_p(z) + Q_p(z) \quad (2.16)$$

com

$$P_p(z) = \frac{\Gamma(p+1)\pi \csc[\pi(k+1)\alpha - \pi p]}{\Gamma[(k+1)\alpha + 1]\Gamma[p - (k+1)\alpha + 1]},$$

$$Q_p(z) = \frac{(\beta z)^{p-(k+1)\alpha} {}_2F_1[(k+1)\alpha + 1, (k+1)\alpha - p; (k+1)\alpha - p + 1; \frac{-1}{\beta z}]}{p - (k+1)\alpha}$$

em que \csc é a cossecante e ${}_2F_1(a, b; c; z)$ é a função hipergeométrica definida por

$${}_2F_1(a, b; c; z) = \sum_{j=0}^{\infty} \frac{(a)_j (b)_j}{(c)_j} \frac{z^j}{j!}.$$

Aqui, $(q)_j$ é o símbolo de Pochhammer definido por

$$(q)_j = \begin{cases} 1, & \text{para } j = 0 \text{ e} \\ q(q-1) \times \dots \times (q+j-1), & \text{para } j = 1, 2, \dots \end{cases}$$

Para $p = 1$, Lemonte e Cordeiro (2011) provaram que se $Y \sim \text{Lomax}(\alpha, \beta)$, então seu primeiro momento incompleto é igual a

$$m_1(Y; q) = \frac{\alpha}{\beta} \sum_{j=0}^{\infty} \frac{(-1)^j q^{j+2} \Gamma(\alpha + j + 1)}{(j+2)\beta^j \Gamma(\alpha + 1)j!}. \quad (2.17)$$

Usando (2.15), obtém-se o primeiro momento incompleto da distribuição MOEL

$$m_1(X; z) = \sum_{k,j=0}^{\infty} \frac{(k+1)(-1)^j \alpha z^{j+2} v_k(\lambda) \Gamma[(k+1)\alpha + j + 1]}{(j+2)\beta^{j+1} \Gamma[(k+1)\alpha + 1] j!}. \quad (2.18)$$

O p -ésimo momento fatorial decrescente de X é

$$\mu'_{(p)} = E[X^{(p)}] = E[X(X-1) \times \dots \times (X-p+1)] = \sum_{j=0}^p s(p, j) \mu'_j, \quad (2.19)$$

em que $s(p, j) = (j!)^{-1}[\partial^j x^{(p)}/\partial x^j]_{x=0}$ é o número de Stirling do primeiro tipo. Dessa maneira, o p -ésimo momento fatorial decrescente da nova distribuição é

$$\mu'_{(p)} = \sum_{j=0}^p s(p, j) \beta \Gamma(j+1) \sum_{k=0}^{\infty} \frac{v_k(\lambda) \Gamma[\alpha(k+1)-j]}{\Gamma[\alpha(k+1)]}, \quad p < \alpha.$$

Adicionalmente, os momentos centrais e os cumulantes podem ser calculados como

$$\mu_p = \sum_{j=0}^p \binom{p}{j} (-1)^{p-j} \mu_1^{p-j} \mu'_j \quad (2.20)$$

e

$$\kappa_p = \mu'_p - \sum_{j=1}^{p-1} \binom{p-1}{j-1} \kappa_j \mu'_{p-j}. \quad (2.21)$$

Lemonte e Cordeiro (2011) calcularam a função geratriz de momentos (fgm) da distribuição Lomax, $M(t; \alpha, \beta)$, da seguinte forma:

$$M(t; \alpha, \beta) = E(e^{tY}) = e^{-\beta t} + \alpha^{-1} {}_1F_1(1; 1 - \alpha; -\beta t), \quad t < 0, \quad (2.22)$$

em que ${}_1F_1(a; b; c)$ é a função hipergeométrica confluyente, definida por

$${}_1F_1(a; b; c) = \frac{\Gamma(b)}{\Gamma(a)} \sum_{j=0}^{\infty} \frac{\Gamma(a+j) c^j}{\Gamma(b+j) j!}.$$

Assim, a fgm da MOEL pode ser escrita como

$$M(t) = e^{-\beta t} + \alpha^{-1} \sum_{k=0}^{\infty} \frac{v_k(\lambda) {}_1F_1(1; 1 - \alpha(k+1); -\beta t)}{k+1}, \quad t < 0. \quad (2.23)$$

2.3.3 Entropia

Entropia de Shannon

A entropia de Shannon é uma medida relacionada com a quantidade de incerteza de uma variável aleatória. Ela é dada por

$$\vartheta_S(f) = -E\{\log[f(X)]\}. \quad (2.24)$$

Substituindo (2.2) em (2.24) fornece

$$\vartheta_S(f) = -\log(\lambda) - E\{\log[g(X)]\} + 2E\{\log[1 - \bar{\lambda}\bar{G}(X)]\}.$$

Para a MOEL, tem-se

$$\begin{aligned} E\{\log[g(X)]\} &= \int_0^\infty \log[g(x)] f(x) dx \\ &= \sum_{k=0}^{\infty} (k+1)v_x(\lambda) \int_0^\infty \log[g(x)] g(x) \bar{G}(x)^k dx. \end{aligned}$$

Fazendo $\bar{G}(x) = u$, obtém-se

$$E\{\log[g(X)]\} = \sum_{k=0}^{\infty} (k+1)v_x(\lambda) \int_0^1 \log\{g[Q_G(1-u)]\} u^k du,$$

em que $Q_G(\cdot)$ é a função quantílica da Lomax. Assim, tem-se

$$\begin{aligned} E\{\log[g(X)]\} &= \sum_{k=0}^{\infty} (k+1)v_x(\lambda) \int_0^1 \log[\alpha\beta^{-1} u^{\frac{\alpha+1}{\alpha}}] u^k du \\ &= \log\left(\frac{\alpha}{\beta}\right) - \frac{\alpha+1}{\alpha} \sum_{k=0}^{\infty} (k+1)^{-1} v_k(\lambda). \end{aligned}$$

A segunda esperança pode ser calculada da seguinte maneira:

$$\begin{aligned} E\{\log[1 - \bar{\lambda}\bar{G}(X)]\} &= E\left\{-\sum_{j=1}^{\infty} \frac{[\bar{\lambda}\bar{G}(X)]^j}{j}\right\} \\ &= -\sum_{j=1}^{\infty} \frac{\bar{\lambda}^j}{j} \int_0^\infty \bar{G}(x)^j \sum_{k=0}^{\infty} v_k(\lambda) g_{(k+1)\alpha,\beta}(x) dx \\ &= -\sum_{j=1}^{\infty} \frac{\bar{\lambda}^j}{j} \sum_{k=0}^{\infty} \frac{(k+1)v_k(\lambda)}{(k+j+1)} \int_0^\infty (k+j+1)g(x) \bar{G}(x)^{j+k} dx. \end{aligned}$$

A última integral é a densidade de uma Lomax($(k+j+1)\alpha, \beta$). Logo,

$$E\{\log[1 - \bar{\lambda}\bar{G}(X)]\} = -\sum_{j=1}^{\infty} \frac{\bar{\lambda}^j}{j} \sum_{k=0}^{\infty} \frac{(k+1)v_k(\lambda)}{(k+j+1)}.$$

Entropia de Rényi

A entropia de Rényi é dada por

$$\vartheta_R(f; c) = \frac{1}{1-c} \log \left[\int_0^\infty f^c(x) dx \right], \quad c > 0, \quad c \neq 1. \quad (2.25)$$

Se $g(x)$ e $\bar{G}(x)$ são a densidade e a função sobrevivência de uma Lomax(α, β), tem-se

$$g(x)^c \bar{G}(x)^k = \frac{\alpha^c}{\beta^{c-1}[(k+c)\alpha+c-1]} g_{(k+c)\alpha+c-1, \beta}(x). \quad (2.26)$$

Pode-se escrever

$$f(x)^c = \sum_{k=0}^{\infty} w_k(\lambda) g_{\alpha(k+c)+c-1, \beta}(x),$$

em que

$$w_k(\lambda) = \begin{cases} \frac{\lambda^c \bar{\lambda}^k \Gamma(2c+k)}{\Gamma(2c) k!} \frac{\alpha^c}{\beta^{c-1}[\alpha(k+c)+c-1]}, & \text{para } 0 < \lambda < 1, \text{ e} \\ \frac{(-1)^k}{\lambda^c} \sum_{j=k}^{\infty} \frac{\Gamma(2c+j)}{\Gamma(2c) j!} \frac{\binom{j}{k} (1-\lambda^{-1})^j \alpha^c}{\beta^{c-1}[\alpha(k+c)+c-1]}, & \text{para } \lambda > 1. \end{cases}$$

Uma vez que $f(x)^c$ é uma combinação linear de densidades da distribuição Lomax, a entropia de Rényi pode, então, ser escrita como

$$\vartheta_R(f; c) = \frac{1}{1-c} \log \left[\sum_{k=0}^{\infty} w_k(\lambda) \right]. \quad (2.27)$$

2.3.4 Desvios Médios

Os desvios da média e da mediana são uma boa maneira de medir dispersão em uma população. Eles são expressos como

$$\delta_1(X) = \int_0^\infty |x - \mu'_1| f(x) dx \quad \text{e} \quad \delta_2(X) = \int_0^\infty |x - M| f(x) dx,$$

respectivamente. Tem-se

$$\delta_1(X) = 2[\mu'_1 F(\mu'_1) - m_1(X; \mu'_1)] \quad (2.28)$$

e

$$\delta_2(X) = \mu'_1 - 2m_1(X; M), \quad (2.29)$$

em que $F(\cdot)$ é a função distribuição de X e $m_1(X; z)$ é o primeiro momento incompleto dado por (2.18).

Adicionalmente, pode-se utilizar a equação (2.18) para obter as curvas de Lorenz e Bonferroni, índices úteis para medir desigualdade de renda. Eles são expressos como

$$L(\pi) = \frac{m_1(X; q)}{\mu'_1} \quad (2.30)$$

e

$$B(\pi) = \frac{m_1(X; q)}{\pi \mu'_1}, \quad (2.31)$$

em que π é uma probabilidade dada e $q = Q_F(\pi) = \beta[(1 + (\pi \lambda)/(1 - \pi))^{1/\alpha} - 1]$.

2.3.5 Momentos Probabilisticamente Ponderados

Os momentos probabilisticamente ponderados (MPPs) são definidos por

$$M(p, r, s) = E\{X^p [\bar{F}(X)]^r [1 - \bar{F}(X)]^s\}.$$

Usando o teorema binomial generalizado, pode-se escrever

$$M(p, r, s) = \sum_{j=0}^{\infty} (-1)^j \binom{s}{j} \int_0^{\infty} x^p \bar{F}(x)^{r+j} f(x) dx.$$

Após alguma álgebra obtém-se

$$f(x) \bar{F}(x)^a = \sum_{m=0}^{\infty} \omega_m(\lambda) g_{(a+m+1)\alpha, \beta}(x), \quad (2.32)$$

em que

$$\omega_m(\lambda) = \begin{cases} \frac{\lambda^{a+1} \bar{\lambda}^m \Gamma(a+2+m)}{(a+m+1)\Gamma(a+2)m!}, & \text{para } 0 < \lambda < 1, \text{ e} \\ \frac{(-1)^m \lambda^{-1}}{(a+m+1)\Gamma(a+2)} \sum_{k=m}^{\infty} \frac{\binom{k}{m} (1 - \lambda^{-1})^k \Gamma(a+2+k)}{k!}, & \text{para } \lambda > 1. \end{cases}$$

Dessa forma, os MPPs da distribuição MOEL são expressos por

$$M(p, r, s) = \sum_{j,m=0}^{\infty} \frac{(-1)^j \binom{s}{j} \beta \omega_m(\lambda) \Gamma[(r+j+m+1)\alpha - p] \Gamma(p+1)}{\Gamma[(r+j+m+1)\alpha]}, \quad (2.33)$$

para $p < \alpha$.

2.3.6 Estatísticas de Ordem

As estatísticas de ordem têm um importante papel na estatística teórica e prática. Considere X_1, X_2, \dots, X_n como uma amostra aleatória de uma distribuição MOEL. Então, a densidade de $X_{i:n}$, a i -ésima estatística de ordem, é expressa por

$$f_{i:n}(x) = \frac{f(x)}{B(n-i+1, i)} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} \bar{F}(x)^{n+j-i}.$$

Utilizando o resultado (2.32) tem-se

$$f_{i:n}(x) = \sum_{j=0}^{i-1} \frac{(-1)^j \binom{i-1}{j}}{B(n-i+1, i)} \sum_{m=0}^{\infty} \omega_m(\lambda) g_{(m+n+j-i+1)\alpha, \beta}(x). \quad (2.34)$$

A equação (2.34) mostra que a função densidade das estatísticas de ordem de uma amostra da distribuição MOEL pode ser expressa como uma combinação linear infinita de densidades da Lomax. Assim, é possível utilizar (2.34) para obter diversas propriedades das estatísticas de ordem de uma amostra da distribuição MOEL baseadas nas mesmas propriedades da distribuição Lomax.

2.4 Estimadores de Máxima Verossimilhança

Suponha que x_1, x_2, \dots, x_n é uma amostra aleatória realizada da distribuição MOEL. Então, a log-verossimilhança, $\ell = \ell(\alpha, \beta, \lambda; \mathbf{x})$, pode ser expressa como

$$\ell = n \log \left(\frac{\lambda \alpha}{\beta} \right) + (\alpha - 1) \sum_{i=1}^n \log \left(1 + \frac{x_i}{\beta} \right) - 2 \sum_{i=1}^n \log \left[\left(1 + \frac{x_i}{\beta} \right)^\alpha - \bar{\lambda} \right]. \quad (2.35)$$

Ao se maximizar (2.35) em relação a α , β e λ utilizando uma rotina numérica de algum software, por exemplo, a rotina `optim` do software R, obtém-se as estimativas de máxima verossimilhança (EMVs) $\hat{\alpha}$, $\hat{\beta}$ e $\hat{\lambda}$.

As primeiras derivadas de $\ell(\alpha, \beta, \lambda; \mathbf{x})$ em relação a α , β e λ são

$$\begin{aligned} \frac{\partial \ell}{\partial \alpha} &= \frac{n}{\alpha} + \sum_{i=1}^n \log \left(1 + \frac{x_i}{\beta} \right) - 2 \sum_{i=1}^n \left[\left(1 + \frac{x_i}{\beta} \right)^\alpha - \bar{\lambda} \right]^{-1} \left(1 + \frac{x_i}{\beta} \right)^\alpha \\ &\quad \times \log \left(1 + \frac{x_i}{\beta} \right), \end{aligned}$$

$$\frac{\partial \ell}{\partial \beta} = -\frac{n}{\beta} + \sum_{i=1}^n \frac{x_i}{\beta^2} \left\{ 2\alpha \left(1 + \frac{x_i}{\beta}\right)^{\alpha-1} \left[\left(1 + \frac{x_i}{\beta}\right)^{\alpha} - \bar{\lambda} \right]^{-1} + \bar{\alpha} \left(1 + \frac{x_i}{\beta}\right)^{-1} \right\},$$

$$\frac{\partial \ell}{\partial \lambda} = \frac{n}{\lambda} - 2 \sum_{i=1}^n \left[\left(1 + \frac{x_i}{\beta}\right)^{\alpha} - \bar{\lambda} \right]^{-1}.$$

Tamém é possível obter as EMVs ao se solucionar as equações não lineares $\partial \ell / \partial \alpha = 0$, $\partial \ell / \partial \beta = 0$ e $\partial \ell / \partial \lambda = 0$ simultaneamente.

Uma abordagem mais abrangente da estimação de máxima verossimilhança para a distribuição MOEL, como a aplicação em caso de censura na amostra, por exemplo, é dada em Ghitany et al. (2007).

Neste capítulo propõe-se um novo método para compor distribuições e algumas de suas propriedades são apresentadas. A classe estudada possui como casos especiais as distribuições estudadas por Kuş (2007) e Lu e Shi (2012), além de diversos outros modelos.

A classe aqui estudada é correspondente à classe $\exp-G$, proposta por Barreto-Souza e Simas (2013). No trabalho citado, os autores desenvolvem diversas propriedades matemáticas da classe, além de proporem os modelos \exp -Weibull e \exp -Beta. Neste trabalho, diversas novas propriedades são obtidas, bem como novos modelos, pertencentes a classe, são propostos.

3.1 A classe G-Poisson

Considere $G(x)$ como a função distribuição de qualquer distribuição *baseline* contínua. Dessa forma, a função distribuição da classe G-Poisson é expressa por

$$F(x) = \frac{e^{\theta G(x)} - 1}{e^{\theta} - 1}, \quad \theta \in \mathbb{R}, \theta \neq 0. \quad (3.1)$$

A função densidade correspondente a (3.1) é

$$f(x) = \frac{\theta e^{\theta G(x)}}{e^{\theta} - 1} g(x). \quad (3.2)$$

Pode-se motivar o parâmetro adicional θ da seguinte maneira:

- $\theta > 0$: Usando o modelo (1.6), com $Z \sim \text{Poisson}(\theta)$, em que $P(Z = z)$ é dado por (1.8), então X terá distribuição dada por (3.1); e
- $\theta < 0$: Usando o modelo (1.5), com $Z \sim \text{Poisson}(|\theta|)$, em que $P(Z = z)$ é dado por (1.8), então X terá distribuição dada por (3.1).

A distribuição *baseline* é um caso especial de (3.2), uma vez que para $\theta \rightarrow 0$ tem-se que $F(x)$, dado por (3.1), converge para $G(x)$.

3.2 Alguns casos especiais

Nesta seção apresentam-se alguns modelos especiais da distribuição G-Poisson. São apresentados os modelos exponencial-Poisson e Weibull-Poisson, anteriormente considerados em Barreto-Souza e Simas (2013), bem como os novos modelos gama-Poisson e normal-Poisson. A função densidade (3.2) é mais fácil de ser trabalhada quando a função distribuição $G(x)$ e a função densidade $g(x)$ têm expressões analíticas simples.

3.2.1 Exponencial-Poisson

A distribuição exponencial tem distribuição e densidade dadas por (1.4). A densidade da distribuição exponencial-Poisson é

$$f(x; \theta) = \frac{\theta e^{\theta(1 - e^{-x/\beta}) - x/\beta}}{\beta(e^\theta - 1)}. \quad (3.3)$$

A Figura 3.1 apresenta a densidade da exponencial-Poisson para alguns valores dos parâmetros.

3.2.2 Weibull-Poisson

A distribuição Weibull tem distribuição e densidade dadas por (1.3). A densidade da Weibull-Poisson é

$$f(x; \theta) = \frac{\theta \alpha x^{\alpha-1} \exp[\theta(1 - e^{-(x/\beta)^\alpha}) - (x/\beta)^\alpha]}{\beta^\alpha(e^\theta - 1)}. \quad (3.4)$$

Para $\alpha = 1$, a distribuição Weibull-Poisson se reduz à distribuição exponencial-Poisson. A Figura 3.2 apresenta a densidade da distribuição Weibull-Poisson para alguns valores dos parâmetros.

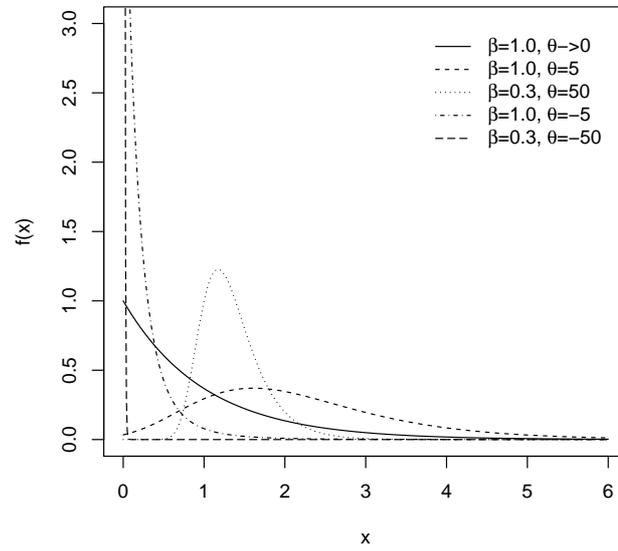


Figura 3.1: Densidade da exponencial-Poisson para alguns valores dos parâmetros.

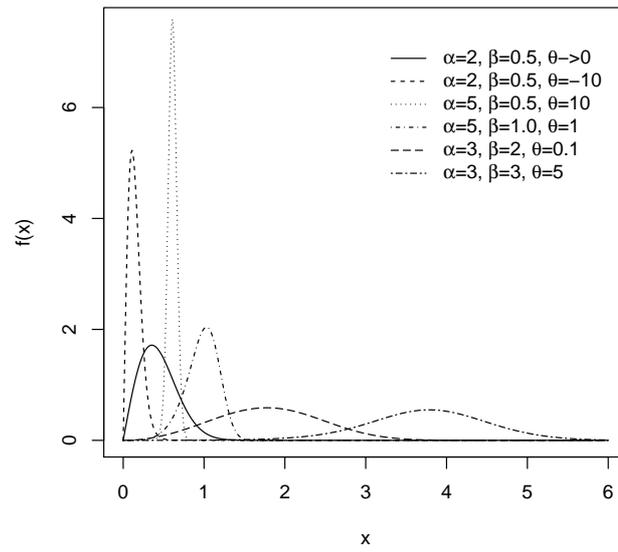


Figura 3.2: Densidade da Weibull-Poisson para alguns valores dos parâmetros.

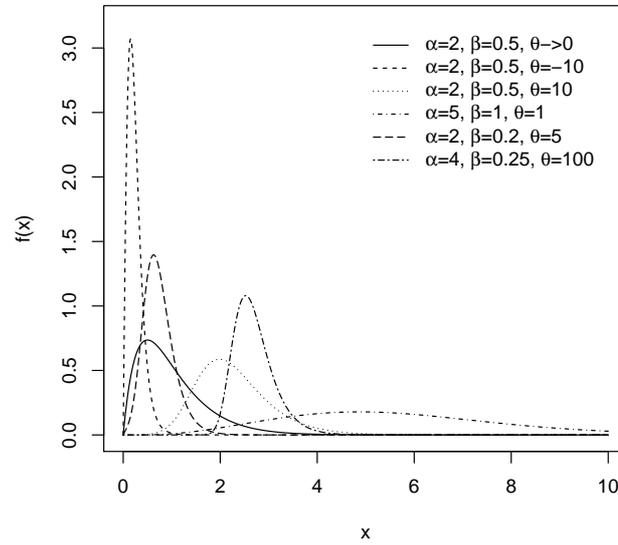


Figura 3.3: Densidade da gama-Poisson para alguns valores dos parâmetros.

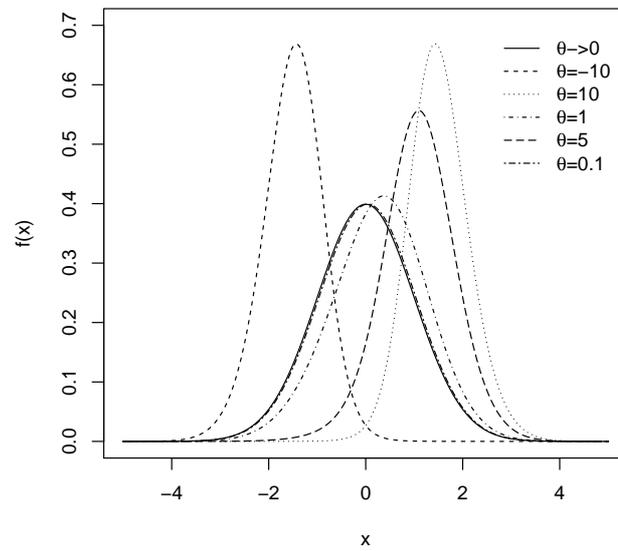


Figura 3.4: Densidade da normal-Poisson padrão para alguns valores de θ .

3.2.3 Gama-Poisson

A distribuição gama possui função distribuição e função densidade iguais a

$$G(x) = \gamma_1(\alpha, x/\beta) = \frac{\gamma(\alpha, x/\beta)}{\Gamma(\alpha)} \quad \text{e} \quad g(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, \quad x > 0,$$

respectivamente, em que $\gamma(\alpha, z) = \int_0^z t^{\alpha-1} e^{-t} dt$ é a função gama incompleta, $\alpha > 0$ é o parâmetro de forma e $\beta > 0$ é o parâmetro de escala. A densidade da distribuição gama-Poisson é, então, dada por

$$f(x; \theta) = \frac{\theta x^{\alpha-1} \exp[\theta \gamma_1(\alpha, x/\beta) - x/\beta]}{\beta^\alpha (e^\theta - 1) \Gamma(\alpha)}. \quad (3.5)$$

Para $\alpha = 1$, a distribuição gama-Poisson se reduz à exponencial-Poisson. A Figura 3.3 apresenta a densidade da gama-Poisson para alguns valores dos parâmetros.

3.2.4 Normal-Poisson

A função distribuição da normal padrão é representada por

$$G(x) = \Phi(x), \quad x \in \mathbb{R}.$$

Logo, a densidade da normal-Poisson é dada por

$$f(x; \theta) = \frac{\theta e^{\theta \Phi(x)}}{e^\theta - 1} \phi(x),$$

em que $\phi(\cdot)$ é a densidade da normal padrão, isto é, $\phi(x) = [\sqrt{2\pi}]^{-1} e^{-x^2/2}$. A Figura 3.4 apresenta a densidade da normal-Poisson para alguns valores dos parâmetros. Uma propriedade interessante é: se $X \sim \text{Normal-Poisson}(0, 1, \theta)$, então, $\mu + \sigma X \sim \text{Normal-Poisson}(\mu, \sigma^2, \theta)$. μ e σ^2 são os parâmetros de locação e escala da distribuição normal. Tal fato é facilmente provado pela própria motivação da classe G-Poisson, considerando que os operadores máximo e mínimo são invariantes quanto as operações de soma e multiplicação por números positivos.

3.3 Algumas Propriedades

3.3.1 Expansão para a distribuição G-Poisson

Expandindo (3.2) em série de Taylor, tem-se

$$f(x) = \sum_{k=0}^{\infty} w_k h_{k+1}(x), \quad (3.6)$$

em que

$$w_k = \frac{\theta^{k+1}}{(k+1)!(e^\theta - 1)}$$

e $h_{k+1}(x)$ é dado por (1.2). Uma equação similar para (3.1) é dada por

$$F(x) = \sum_{k=0}^{\infty} w_k H_{k+1}(x), \quad (3.7)$$

em que $H_{k+1}(x)$ é dado por (1.1). Dessa forma, a expansão (3.6) pode ser utilizada para obter diversas propriedades da distribuição G-Poisson baseando-se nas mesmas propriedades da distribuição G-exponencializada.

3.3.2 Função taxa de falha

A função taxa de falha para a distribuição G-Poisson é dada por

$$r(x) = \frac{\theta e^{-\theta \bar{G}(x)} \bar{G}(x)}{1 - e^{-\theta \bar{G}(x)}} r_G(x), \quad (3.8)$$

em que $r_G(x) = g(x)/\bar{G}(x)$ é a função de risco da *baseline*. Da equação (3.8), segue-se que

$$\lim_{x \rightarrow -\infty} r(x) = \frac{\theta e^{-\theta}}{1 - e^{-\theta}} \lim_{x \rightarrow -\infty} r_G(x) \quad \text{e} \quad \lim_{x \rightarrow +\infty} r(x) = \lim_{x \rightarrow +\infty} r_G(x).$$

Uma vez que $\theta e^{-\theta}(1 - e^{-\theta})^{-1}$ é uma função decrescente em θ com $\theta e^{-\theta}(1 - e^{-\theta})^{-1} \rightarrow 1$ para $\theta \rightarrow 0$, tem-se

$$\frac{\theta e^{-\theta}}{1 - e^{-\theta}} > 1 \text{ para } \theta < 0 \quad \text{e} \quad \frac{\theta e^{-\theta}}{1 - e^{-\theta}} < 1 \text{ para } \theta > 0$$

que conduz a

$$r_G(x) < r(x) < \frac{\theta e^{-\theta}}{1 - e^{-\theta}} r_G(x), \quad \theta < 0; \text{ e}$$

$$\frac{\theta e^{-\theta}}{1 - e^{-\theta}} r_G(x) < r(x) < r_G(x), \quad \theta > 0.$$

A Figura 3.5 apresenta a função de risco da exponencial-Poisson para $\lambda = 1$ e alguns valores de θ .

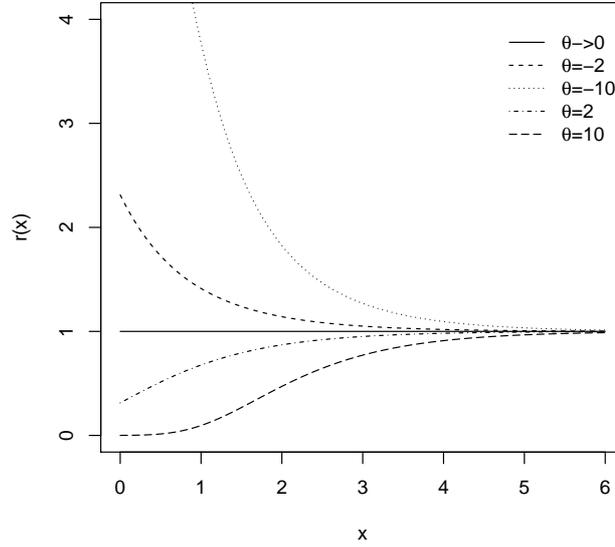


Figura 3.5: A função de risco da exponencial-Poisson para alguns valores de θ com $\lambda = 1$.

3.3.3 Moda

A primeira derivada de (3.2) é dada por

$$f'(x) = \frac{\theta}{e^\theta - 1} e^{\theta G(x)} [g'(x) + \theta g^2(x)].$$

Uma vez que a expressão fora dos colchetes é sempre positiva, o seu comportamento depende se

$$g'(x) + \theta g^2(x) > 0 \quad \text{ou} \quad g'(x) + \theta g^2(x) < 0.$$

No primeiro caso, $f(x)$ é crescente e no segundo decrescente. Para a distribuição exponencial-Poisson, tem-se:

$$g'(x) + \theta g^2(x) = \beta^{-2} e^{-x/\beta} (\theta e^{-x/\beta} - 1).$$

Novamente, o termo fora dos paranteses é sempre maior que zero, deixando a função central

$$(\theta e^{-x/\beta} - 1).$$

Uma vez que $e^{-x/\beta}$ é sempre positivo, para qualquer $\theta < 0$, a moda será em 0 com valor modal dado por $\theta[\beta(e^\theta - 1)]^{-1}$. Este resultado está de acordo com o apresentado por Kuş (2007).

Para $\theta > 0$, a moda é dada por $x = \beta \log(\theta)$ e o valor modal é igual a $e^{\theta-1}[\beta(e^\theta - 1)]^{-1}$.

3.3.4 Simulação

A função quantílica correspondente a (3.1) é dada por

$$F^{-1}(u) = G^{-1} \{ \theta^{-1} \log[(e^\theta - 1)u + 1] \}, \quad (3.9)$$

em que $G^{-1}(\cdot)$ é a função quantílica da distribuição *baseline*. Um número aleatório X para qualquer distribuição dentro da classe G-Poisson pode ser gerado por

$$X = G^{-1} \{ \theta^{-1} \log[(e^\theta - 1)U + 1] \},$$

em que $U \sim \text{Uniforme}(0, 1)$.

3.3.5 Momentos

Substituindo (3.2) em (2.11), tem-se que

$$E(X^p) = \frac{\theta}{e^\theta - 1} \int_{\mathbb{R}} x^p e^{\theta G(x)} g(x) dx.$$

Fazendo uma mudança de variável, $G(x) = u$, na integral acima original

$$E(X^p) = \frac{\theta}{e^\theta - 1} \int_0^1 Q_G(u)^p e^{\theta u} du, \quad (3.10)$$

em que $Q_G(u) = G^{-1}(u)$ é a função quantílica da distribuição *baseline*. A integral pode ser resolvida analiticamente ou numericamente para a maioria dos modelos G-Poisson especiais. Um método alternativo para obter os momentos ordinários pode ser obtido partindo de (3.6). Pelo restante deste capítulo, considera-se $Y_k \sim \text{G-exponencializada}(k + 1)$. Então, é fácil obter

$$E(X^p) = \sum_{k=0}^{\infty} w_k E(Y_k^p). \quad (3.11)$$

Os momentos ordinários de diversas distribuições exponencializadas são dados por Nadarajah e Kotz (2006). Dessa forma, é possível computar (3.11).

Da equação (3.11), é possível obter uma terceira fórmula. Tem-se

$$E(Y_k^p) = (k + 1) \int_{\mathbb{R}} x^p G(x)^k g(x) dx,$$

como $E(Y_k^p)$ pode ser expresso por

$$E(Y_k^p) = (k+1) \int_0^1 Q_G(u)^p u^k du.$$

Os momentos ordinários de X se reduzem a

$$E(X^p) = \sum_{k=0}^{\infty} (k+1) w_k \tau(p, k), \quad (3.12)$$

em que $\tau(p, k) = \int_0^1 Q_G(u)^p u^k du$. Utilizando (3.12), apresentam-se os momentos ordinários das distribuições exponencial-Poisson, logística padrão-Poisson e Pareto-Poisson. Os momentos da exponencial-Poisson são

$$E(X^p) = p! \lambda^p \sum_{k=0}^{\infty} \sum_{j=0}^k \frac{(-1)^{2p+j} (k+1) \binom{k}{j} w_k}{(j+1)^{p+1}}.$$

Para a distribuição logística padrão-Poisson, em que $G(x) = (1 + e^{-x})^{-1}$, tem-se

$$E(X^p) = \sum_{k=0}^{\infty} (k+1) w_k \frac{\partial^p}{\partial t^p} B(t+k+1, 1-t) \Big|_{t=0},$$

em que $B(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt$ é a função beta. Para o modelo Pareto-Poisson, em que $G(x) = 1 - (1+x)^{-\nu}$, $\nu > 0$, obtém-se

$$E(X^p) = \sum_{k,j=0}^{\infty} (-1)^{p+j} (k+1) \binom{p}{j} B(k+1, 1-j\nu^{-1}) w_k.$$

Substituindo (3.6) em (2.14), obtém-se os momentos incompletos como

$$m_p(X; z) = \sum_{k=0}^{\infty} w_k m_p(Y_k; z), \quad (3.13)$$

em que $m_p(Y_k; z)$ é o momento incompleto de $Y_k \sim G$ -exponencializada($k+1$). Desta maneira, é fácil obter os momentos incompletos da distribuição G -Poisson usando (3.13) e os momentos incompletos da distribuição G -exponencializada associada. Mudando variáveis, obtém-se uma segunda fórmula para calcular os momentos incompletos:

$$m_p(X; z) = \sum_{k=0}^{\infty} (k+1) w_k I_k(z), \quad (3.14)$$

em que $I_k(z) = \int_0^{G(z)} Q_G(u)^p u^k du$ pode ser determinado para a maioria das distribuições *baseline* G .

O p -ésimo momento fatorial descendente, bem como os momentos centrais e os cumulantes podem ser obtidos pelas equações (2.19), (2.20) e (2.21), respectivamente.

3.3.6 Função geratriz de momentos

A função geratriz de momentos, $M(t) = E(e^{tX})$, pode ser determinada através da expansão em série de Taylor

$$M(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} \mu'_k, \quad (3.15)$$

em que μ'_k é obtido de (3.10), (3.11) ou (3.12). Substituindo (3.2) na definição acima tem-se

$$M(t) = \frac{\theta}{e^\theta - 1} \int_0^1 e^{tQ_G(u) + \theta u} du, \quad (3.16)$$

em que a integral pode ser resolvida analiticamente ou numericamente. Partindo da expansão (3.6), segue que

$$M(t) = \sum_{k=0}^{\infty} w_k M_k(t), \quad (3.17)$$

em que $M_k(t)$ é a função geratriz de momentos de Y_k . Uma quarta fórmula para $M(t)$ decorre de

$$M(t) = \sum_{k=0}^{\infty} (k+1) w_k \int_{\mathbb{R}} e^{tx} G(x)^k g(x) dx,$$

que pode ser expressa como

$$M(t) = \sum_{k=0}^{\infty} (k+1) w_k \int_0^1 e^{tQ_G(u)} u^k du. \quad (3.18)$$

A função geratriz de momentos para as distribuições exponencial-Poisson, logística padrão-Poisson e Pareto-Poisson são

$$M(t) = \sum_{k=0}^{\infty} (k+1) w_k B(k+1, 1 - \lambda t),$$

$$M(t) = \sum_{k=0}^{\infty} (k+1) w_k B(t+k+1, 1-t)$$

e

$$M(t) = e^{-t} \sum_{k,r=0}^{\infty} \frac{(k+1) w_k B(k+1, 1 - r\nu^{-1})}{r!} t^r,$$

respectivamente.

3.3.7 Entropia

Entropia de Shannon

Substituindo (3.2) em (2.24) fornece

$$\vartheta_S(f) = -\log\left(\frac{\theta}{e^\theta - 1}\right) - \theta E[G(X)] - E\{\log[g(X)]\}$$

em que

$$E[G(X)] = \frac{\theta}{e^\theta - 1} \int_{\mathbb{R}} g(x) G(x) e^{\theta G(x)} dx = \frac{\theta}{e^\theta - 1} \int_0^1 u e^{\theta u} du = \frac{e^\theta(\theta - 1) + 1}{\theta(e^\theta - 1)},$$

$$E\{\log[g(X)]\} = \int_{\mathbb{R}} \log[g(x)] f(x) dx = \sum_{k=0}^{\infty} (k+1) w_k J_k,$$

w_k é dado por (3.6) e $J_k = \int_0^1 u^k \log\{g[Q_G(u)]\} du$. A integral J_k pode ser computada para um bom número de distribuições *baseline*. Para as distribuições exponencial-Poisson, logística padrão-Poisson e Pareto-Poisson, obtém-se

$$J_k = \sum_{j=0}^k \frac{(-1)^{j+1} \binom{k}{j}}{(j+1)^2} - \frac{\log(\lambda)}{k+1},$$

$$J_k = \sum_{j=0}^k \frac{(-1)^{j+1} \binom{k}{j}}{(j+1)^2} - \frac{1}{(k+1)^2},$$

e

$$J_k = (1 + \nu^{-1}) \sum_{j=0}^k \frac{(-1)^{j+1} \binom{k}{j}}{(j+1)^2} + \frac{\log(\nu)}{(k+1)^2},$$

respectivamente.

Entropia de Rényi

Substituindo (3.2) em (2.25), obtém-se

$$\int_{\mathbb{R}} f^c(x) dx = \frac{\theta^c}{(e^\theta - 1)^c} \int_{\mathbb{R}} e^{c\theta G(x)} g^c(x) dx.$$

Expandido-se em série de Taylor

$$\int_{\mathbb{R}} f^c(x) dx = (e^\theta - 1)^{-c} \sum_{k=0}^{\infty} \frac{c^k \theta^{c+k}}{k!} I_k,$$

em que $I_k = \int_{\mathbb{R}} g^c(x) G^k(x) dx$ pode ser calculado para cada distribuição *baseline*. Para as distribuições exponencial, logística padrão e Pareto, tem-se $I_k = \lambda^{1-c} B(k+1, c)$, $I_k = B(k+c, c)$ e $I_k = \nu^{c-1} B(k+1, \nu^{-1}(c-1) + c)$, respectivamente.

3.3.8 Desvios Médios

Para obter δ_1 , o desvio com relação a média, e δ_2 , o desvio com a relação a mediana, a partir das equações (2.28) e (2.29), respectivamente, é necessário se conhecer a mediana e o primeiro momento incompleto da distribuição *baseline*.

A mediana M é obtida por

$$M = G^{-1} \{ \theta^{-1} \log[(e^\theta - 1)/2 + 1] \}.$$

O primeiro momento incompleto, $m_1(X; \cdot)$, é dado por (3.13) ou por (3.14). Usando (3.14), o primeiro momento incompleto para as distribuições exponencial-Poisson, logística padrão-Poisson e Pareto-Poisson são dados por

$$m_1(X; z) = \sum_{k=0}^{\infty} \sum_{j=1}^k (k+1) w_k \lambda \frac{(1 - e^{-z/\lambda})^{j+k+1}}{j(j+k+1)},$$

$$m_1(X; z) = \sum_{k=0}^{\infty} (k+1) w_k \left\{ \sum_{r=1}^{\infty} \frac{(1 + e^{-z})^{-(r+k+1)}}{r(r+k+1)} - \sum_{j=1}^{\infty} \frac{B[(1 + e^{-z})^{-1}; j, k]}{j} \right\}$$

e

$$m_1(X; z) = \sum_{k=0}^{\infty} (k+1) w_k \left\{ B[1 - (1+z)^{-v}; 1 - v^{-1}, k+1] - \frac{[1 - (1+z)^{-v}]^{k+1}}{k+1} \right\},$$

em que $B(z; a, b) = \int_0^z t^{a-1} (1-t)^{b-1} dt$ é a função beta incompleta.

As curvas de Lorenz e Bonferroni, dadas por (2.30) e (2.31), respectivamente, também podem ser calculadas, utilizando-se dos primeiros momentos incompletos fornecidos anteriormente.

3.3.9 Momentos Probabilisticamente Ponderados

Os momentos probabilisticamente ponderados (MPPs) são úteis para obter estimadores para parâmetros desconhecidos. Nesta seção eles serão calculados como

$$M(p, r, s) = E\{X^p [F(X)]^r [1 - F(X)]^s\}.$$

Utilizando o teorema binomial generalizado, pode-se escrever

$$M(p, r, s) = \sum_{j=0}^{\infty} (-1)^j \binom{s}{j} \int_{\mathbb{R}} x^p F(x)^{r+j} f(x) dx.$$

Substituindo (3.1) na equação acima e utilizando o teorema binomial novamente fornece

$$M(p, r, s) = \sum_{j=0}^{\infty} \frac{(-1)^j \binom{s}{j}}{(1 - e^{\theta})^{r+j}} \sum_{k=0}^{\infty} (-1)^k \binom{r+j}{k} \int_{\mathbb{R}} x^p e^{k\theta G(x)} f(x) dx.$$

Substituindo $f(x)$ por (3.2) vem

$$M(p, r, s) = \sum_{j=0}^{\infty} \frac{(-1)^{j+1} \binom{s}{j}}{(1 - e^{\theta})^{r+j+1}} \sum_{k=0}^{\infty} (-1)^k \binom{r+j}{k} \theta \int_{\mathbb{R}} x^p e^{(k+1)\theta G(x)} g(x) dx.$$

Os MPPs podem, então, ser expressos como

$$M(p, r, s) = \sum_{j,k=0}^{\infty} \frac{(-1)^{j+k+1} \binom{s}{j} \binom{r+j}{k} \theta}{(1 - e^{\theta})^{r+j+1}} J(p, k, \theta), \quad (3.19)$$

em que

$$J(p, k, \theta) = \int_{\mathbb{R}} x^p e^{(k+1)\theta G(x)} g(x) dx = \int_0^1 Q_G(u)^p e^{(k+1)\theta u} du$$

pode ser obtido analiticamente ou numericamente.

3.3.10 Estatísticas de ordem

Considere X_1, X_2, \dots, X_n uma amostra aleatória de uma distribuição G-Poisson especificada. Então, a densidade de $X_{i:n}$, a i -ésima estatística de ordem, equivale a

$$\begin{aligned} f_{i:n}(x) &= \frac{f(x)}{B(n-i+1, i)} F(x)^{i-1} [1 - F(x)]^{n-i} \\ &= \frac{f(x)}{B(n-i+1, i)} \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} F(x)^{i+j-1}. \end{aligned}$$

Partindo de (1.1), (3.6) e (3.7), pode-se escrever

$$f_{i:n}(x) = \sum_{k=0}^{\infty} \frac{w_k (k+1) G(x)^k g(x)}{B(n-i+1, i)} \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} \left(\sum_{r=0}^{\infty} w_r G(x)^{r+1} \right)^{i+j-1}.$$

Utilizando o resultado 0.314 de Gradshteyn e Ryzhik (2007), obtém-se

$$\left(\sum_{r=0}^{\infty} w_r G(x)^{r+1} \right)^{i+j-1} = \sum_{r=0}^{\infty} b_{i+j-1,r} G(x)^{r+i+j-1},$$

em que $b_{i+j-1,0} = w_0^{i+j-1}$ e

$$b_{i+j-1,m} = (m w_0)^{-1} \sum_{l=1}^m [l(i+j) - m] w_l b_{i+j-1,m-l}.$$

Finalmente,

$$f_{i:n}(x) = \sum_{j=0}^{n-i} \sum_{k,r=0}^{\infty} t_{j,k,r} h_{k+r+i+j}(x), \quad (3.20)$$

em que

$$t_{j,k,r} = \frac{(-1)^j \binom{n-i}{j}}{B(n-i+1, i)} \frac{(k+1) w_k b_{i+j-1,r}}{(k+r+i+j)}.$$

Dessa forma, diversas propriedades matemáticas da distribuição das estatísticas de ordem da classe G-Poisson, como momentos e função geratriz de momentos, podem ser obtidas das mesmas propriedades da classe G-exponencializada, utilizando-se a equação (3.20).

3.4 Estimação

O método de máxima verossimilhança será considerado para obter os parâmetros desconhecidos da distribuição G-Poisson. Suponha que x_1, x_2, \dots, x_n é uma amostra aleatória de $X \sim G\text{-Poisson}(\theta, \boldsymbol{\tau})$, em que $\boldsymbol{\tau}$ é um vetor $p \times 1$ de parâmetros desconhecidos indexando a distribuição *baseline* $G(x)$. Então, a função de verossimilhança pode ser expressa como

$$\ell(\theta, \boldsymbol{\tau}; \mathbf{x}) = n \log \left(\frac{\theta}{e^\theta - 1} \right) + \theta \sum_{k=1}^n G(x_k) + \ell(\boldsymbol{\tau}; \mathbf{x}), \quad (3.21)$$

em que $\ell(\boldsymbol{\tau}; \mathbf{x})$ denota a log-verossimilhança da distribuição original G. Ao maximizar (3.21) em relação a θ e $\boldsymbol{\tau}$ usando alguma rotina numérica, como a função `optim` do software R, por exemplo, obtém-se as estimativas de máxima verossimilhança (EMVs) $\hat{\theta}$ e $\hat{\boldsymbol{\tau}}$.

As primeiras derivadas de $\ell = \ell(\theta, \boldsymbol{\tau}; \mathbf{x})$ em relação a θ e $\boldsymbol{\tau}$ são

$$\begin{aligned}\frac{\partial \ell}{\partial \theta} &= \sum_{k=1}^n G(x_k) + n \frac{e^\theta(1-\theta) - 1}{\theta(e^\theta - 1)}, \\ \frac{\partial \ell}{\partial \tau_j} &= \theta \sum_{k=1}^n \frac{\partial G(x_k)}{\partial \tau_j} + \frac{\partial \ell(\boldsymbol{\tau}; \mathbf{x})}{\partial \tau_j}, \quad j = 1, 2, \dots, p.\end{aligned}$$

Também é possível obter as EMVs ao se solucionar as equações não lineares $\partial \ell / \partial \theta = 0$ e $\partial \ell / \partial \tau_j = 0$, $j = 1, 2, \dots, p$, simultaneamente.

A matriz de informação observada para $(\theta, \boldsymbol{\tau})$ é

$$\mathbf{K} = - \begin{pmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} \\ \mathbf{K}_{12} & \mathbf{K}_{22} \end{pmatrix},$$

em que

$$\begin{aligned}\mathbf{K}_{11} &= \frac{\partial^2 \ell}{\partial \theta^2} = n \frac{\theta^2 e^\theta - (e^\theta - 1)^2}{[\theta(e^\theta - 1)]^2}, \\ \mathbf{K}_{12} &= \frac{\partial^2 \ell}{\partial \theta \partial \tau_j} = \sum_{k=1}^n \frac{\partial G(x_k)}{\partial \tau_j}, \quad j = 1, 2, \dots, p, \\ \mathbf{K}_{22} &= \frac{\partial^2 \ell}{\partial \tau_j \partial \tau_m} = \theta \sum_{k=1}^n \frac{\partial^2 G(x_k)}{\partial \tau_j \partial \tau_m} + \frac{\partial \ell(\boldsymbol{\tau}; \mathbf{x})}{\partial \tau_j \partial \tau_m}, \quad j, m = 1, 2, \dots, p.\end{aligned}$$

A matriz \mathbf{K} é necessária para construir estimativas intervalares e realizar testes de hipóteses. Seja $\boldsymbol{\Theta} = (\theta, \boldsymbol{\tau})$ o conjunto de parâmetros desconhecidos e $\widehat{\boldsymbol{\Theta}}$ $(p+1) \times 1$ o EMV de $\boldsymbol{\Theta} = (\theta, \boldsymbol{\tau})$. Sob condições padrões de regularidade, $\sqrt{n}(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta})$ é assintoticamente normal, $N_{p+1}(0, \mathbf{I}(\boldsymbol{\Theta})^{-1})$, em que $\mathbf{I}(\boldsymbol{\Theta})$ é a matriz de informação esperada. É possível substituir $\mathbf{I}(\boldsymbol{\Theta})$ por $\mathbf{K}(\widehat{\boldsymbol{\Theta}})$, a matriz de informação observada avaliada em $\widehat{\boldsymbol{\Theta}}$, uma vez que esta é um estimador consistente de $\mathbf{I}(\boldsymbol{\Theta})$. Intervalos de confiança assintóticos para os parâmetros do modelo G-Poisson podem ser construídos a partir da distribuição normal multivariada $N_{p+1}(0, \mathbf{K}(\widehat{\boldsymbol{\Theta}})^{-1})$.

As estatísticas de razão de verossimilhanças (RV) podem ser construídas de forma a verificar se a distribuição G-Poisson apresenta um melhor ajuste que qualquer submodelo, por exemplo, a distribuição *baseline*. Nesse caso, a hipótese $H_0 : \theta = 0$ seria testada contra a hipótese $H_1 : \theta \neq 0$. Para qualquer par de hipóteses, a estatística RV é igual a

$$RV = 2[\ell(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\tau}}) - \ell(\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\tau}})],$$

em que $\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\tau}}$ são as EMVs sob H_1 e $\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\tau}}$ são as estimativas sob H_0 .

3.5 Aplicação

Para ilustrar a utilidade da classe apresentada neste capítulo, dois conjuntos de dados serão apresentados e ajustes por distribuições usuais em estudos de sobrevivência serão comparados com os ajustes fornecidos pela classe G-Poisson. As distribuições consideradas são exponencial, Weibull, gama, exponencial exponencializada, exponencial-Poisson, Weibull-Poisson e gama-Poisson. Todos os ajustes foram feitos utilizando o software R.

Primeiro conjunto de dados (Jong e Heller, 2008; Paula, 2012): Este conjunto contém os valores de 769 sinistros (em mil dólares australianos) de seguros pessoais, datando de Janeiro de 1989 até Junho de 1999, e pode ser encontrado em <http://www.ime.usp.br/~giapaula/insurance.dat>.

Na Tabela 3.1 são listados as EMVs, as estimativas dos erros-padrões, o critério de informação de Akaike (AIC) e o critério de informação Bayesiana (BIC) para as distribuições ajustadas. Os resultados indicam que a distribuição Weibull-Poisson tem os menores valores AIC e BIC entre todos os modelos ajustados, provendo, portanto, um melhor ajuste. A Figura 3.6 apresenta o histograma dos dados bem como as distribuições Weibull-Poisson e gama-Poisson ajustadas.

Tabela 3.1: EMVs dos parâmetros dos modelos e as estimativas dos erros-padrões para o primeiro conjunto de dados bem como as estatísticas AIC e BIC correspondentes.

Distribuição	α	β	θ	k	a	AIC	BIC
Exponencial	-	7,83 (0,004)	-	-	-	4706,55	4711,19
Weibull	1,08 (0,028)	8,08 (0,282)	-	-	-	4699,97	4709,26
Gama	1,16 (0,052)	6,74 (0,008)	-	-	-	4698,16	4707,45
Exp. Exp.	-	7,06 (0,006)	-	-	1,17 (0,057)	4697,31	4706,60
Exp. Poisson	-	5,76 (0,009)	1,31 (0,231)	-	-	4682,36	4691,65
Wei. Poisson	0,73 (0,038)	2,92 (0,364)	3,56 (0,431)	-	-	4652,10	4666,03
Gam. Poisson	0,46 (0,117)	6,80 (0,011)	4,13 (0,981)	-	-	4665,04	4678,97

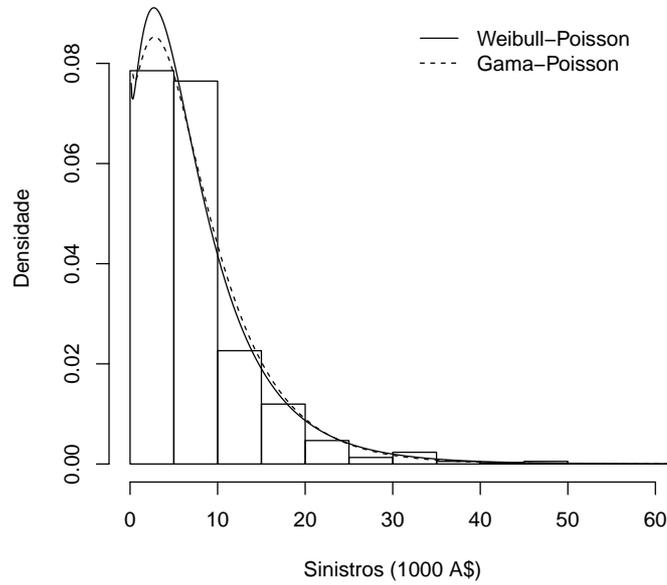


Figura 3.6: Histograma dos sinistros e densidades ajustadas.

Tabela 3.2: EMVs dos parâmetros dos modelos e as estimativas dos erros-padrões para o segundo conjunto de dados bem como as estatísticas AIC e BIC correspondentes.

Distribuição	α	β	θ	k	a	AIC	BIC
Exponencial	-	176,81	-	-	-	891,21	893,49
	-	(0,0006)	-	-	-	-	-
Weibull	1,82	199,85	-	-	-	858,72	863,27
	(0,158)	(13,65)	-	-	-	-	-
Gama	3,08	57,34	-	-	-	855,60	860,15
	(0,485)	(0,002)	-	-	-	-	-
Exp. Exp.	-	88,46	-	-	3,64	855,61	860,17
	-	(0,001)	-	-	(0,721)	-	-
Exp. Poisson	-	74,58	5,81	-	-	853,77	858,33
	-	(0,001)	(1,143)	-	-	-	-
Wei. Poisson	0,87	57,71	7,50	-	-	855,26	862,09
	(0,155)	(21,31)	(2,858)	-	-	-	-
Gam. Poisson	0,77	81,59	7,25	-	-	855,52	862,35
	(0,51)	(0,002)	(4,249)	-	-	-	-

Segundo conjunto de dados (Bjerkedal, 1960; Gupta et al., 1997): Este conjunto de dados contém o tempo de sobrevivência (em dias) de 72 porcos-da-índia infectados com o bacilo da tuberculose. Os animais em uma mesma gaiola pertenciam a um mesmo regime. Considerou-se apenas o regime 4,3.

Na Tabela 3.2 são listados as EMVs, as estimativas dos erros-padrões, o AIC e o BIC para as distribuições ajustadas. Os resultados indicam que a distribuição exponencial-Poisson tem os menores valores AIC e BIC entre todos os modelos ajustados, provendo, portanto, um melhor ajuste. A Figura 3.7, mostra o histograma dos dados bem como as distribuições exponencial-Poisson, Weibull e exponencial exponencializada ajustadas.

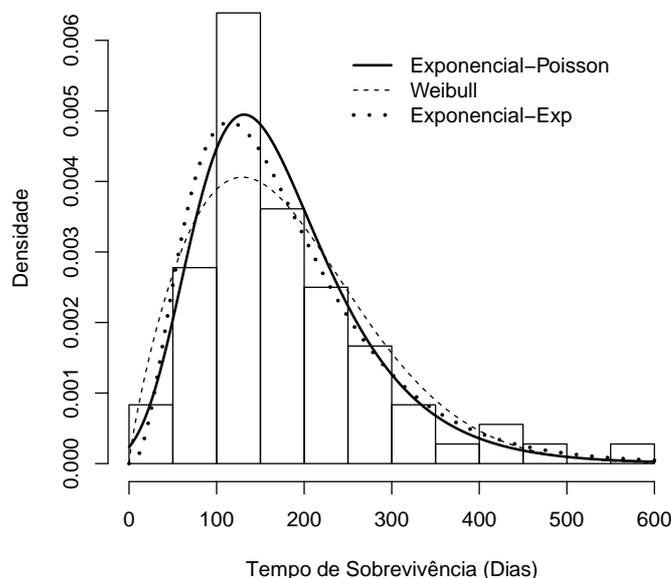


Figura 3.7: Histograma do tempo de sobrevivência e densidades ajustadas.

A distribuição Lomax-Poisson

Neste capítulo, propõe-se, utilizando a metodologia abordada no Capítulo 3, a distribuição Lomax-Poisson. Diversas de suas propriedades matemáticas são apresentadas.

4.1 A distribuição Lomax-Poisson

Substituindo (2.4) em (3.1), obtém-se a função distribuição da Lomax-Poisson (LP) como

$$F(x) = \frac{\exp\{\theta[1 - (1 + x/\beta)^{-\alpha}]\} - 1}{e^\theta - 1}, \quad x > 0. \quad (4.1)$$

Sua densidade correspondente é dada por

$$f(x) = \frac{\alpha \theta \exp\{\theta[1 - (1 + x/\beta)^{-\alpha}]\}}{\beta(e^\theta - 1)} \left(1 + \frac{x}{\beta}\right)^{-(\alpha+1)}, \quad x > 0. \quad (4.2)$$

A Figura 4.1 apresenta a densidade da distribuição LP para alguns valores dos parâmetros.

A função taxa de falha da distribuição LP é dada por

$$r(x) = \frac{\alpha \theta \exp\{\theta[1 - (1 + x/\beta)^{-\alpha}]\}}{\beta(e^\theta - \exp\{\theta[1 - (1 + x/\beta)^{-\alpha}]\})} \left(1 + \frac{x}{\beta}\right)^{-(\alpha+1)}. \quad (4.3)$$

A Figura 4.2 apresenta a função taxa de falha da distribuição LP para alguns valores dos parâmetros.

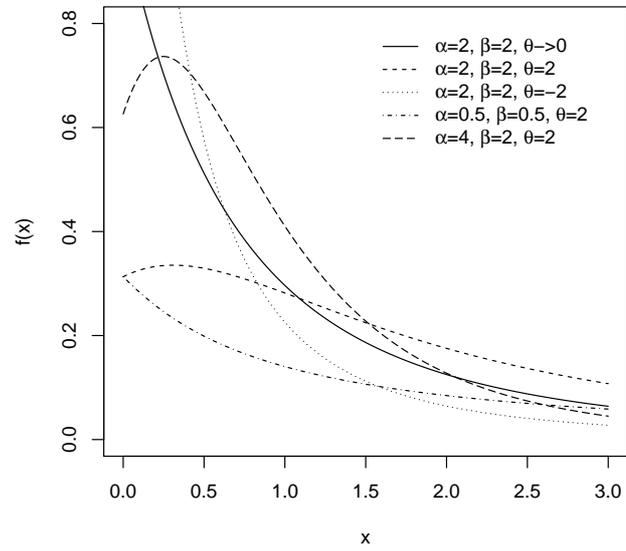


Figura 4.1: Densidade da distribuição LP para alguns valores dos parâmetros.

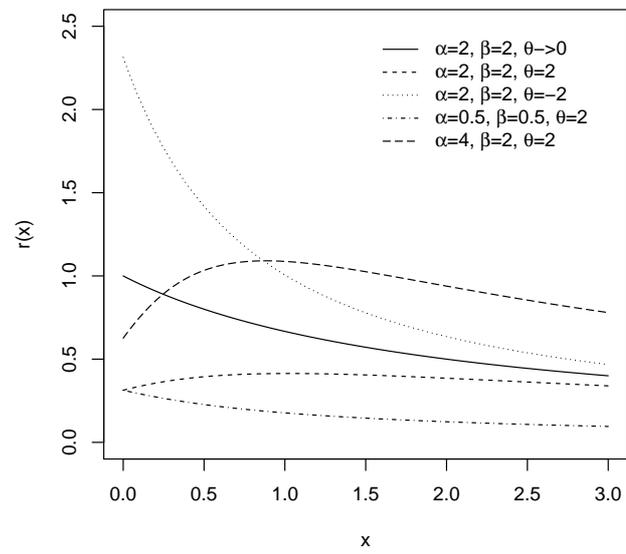


Figura 4.2: Taxa de falha da distribuição LP para alguns valores dos parâmetros.

A função quantílica correspondente a (4.1) é dada por

$$Q(u) = F^{-1}(u) = \beta \left[(1 - \theta^{-1} \log[(e^\theta - 1)u + 1])^{-1/\alpha} - 1 \right], \quad 0 \leq u \leq 1. \quad (4.4)$$

Um número aleatório da distribuição LP pode ser gerado por

$$X = \beta \left[(1 - \theta^{-1} \log[(e^\theta - 1)U + 1])^{-1/\alpha} - 1 \right],$$

em que $U \sim \text{Uniforme}(0, 1)$.

A mediana é dada por

$$M = \beta \left[(1 - \theta^{-1} \log[(e^\theta - 1)/2 + 1])^{-1/\alpha} - 1 \right]. \quad (4.5)$$

4.2 Propriedades

4.2.1 Expansão para a distribuição LP

Pode-se reescrever (3.6) como

$$f(x) = \sum_{j=0}^{\infty} \frac{\theta^{j+1}}{(e^\theta - 1)j!} g(x) \sum_{k=0}^j (-1)^k \binom{j}{k} \overline{G}^k(x).$$

Então, utilizando (2.9) e trocando a ordem dos somatórios obtém-se

$$f(x) = \sum_{k=0}^{\infty} v_k g_{(k+1)\alpha, \beta}(x), \quad (4.6)$$

em que

$$v_k = \frac{(-1)^k \theta^{k+1} e^\theta}{(e^\theta - 1)(k+1)!}$$

É fácil verificar que $\sum_{k=0}^{\infty} v_k = 1$. A equação (4.6) revela que a densidade da distribuição LP pode ser expressa como uma combinação linear de distribuições Lomax. Assim, diversas propriedades da distribuição gerada podem ser obtidas diretamente das mesmas propriedades da distribuição Lomax. Essas propriedades podem ser encontradas em Lemonte e Cordeiro (2011), por exemplo.

4.2.2 Momentos e Geratriz de Momentos

Utilizando (4.6) e (2.12), pode-se calcular os momentos ordinários da distribuição LP por

$$\mu'_p = E(X^p) = \beta \Gamma(p+1) \sum_{k=0}^{\infty} \frac{v_k \Gamma[(k+1)\alpha - p]}{\Gamma[(k+1)\alpha]}, \quad p < \alpha. \quad (4.7)$$

Substituindo (4.6) em (2.14) se obtém os momentos incompletos como

$$m_p(X; z) = \sum_{k=0}^{\infty} v_k m_p(Y_{(k+1)\alpha, \beta}; z), \quad (4.8)$$

em que $m_p(Y_{(k+1)\alpha, \beta}; z)$ é o momento incompleto de $Y_{(k+1)\alpha, \beta} \sim \text{Lomax}((k+1)\alpha, \beta)$ dado por (2.16).

Para $p = 1$, substituindo (2.17) em (4.8), tem-se

$$m_1(X; z) = \sum_{k,j=0}^{\infty} \frac{(-1)^j (k+1)\alpha z^{j+2} v_k \Gamma[(k+1)\alpha + j + 1]}{(j+2)\beta^{j+1} \Gamma[(k+1)\alpha + 1] j!}. \quad (4.9)$$

O p -ésimo momento fatorial descendente, bem como os momentos centrais e os cumulantes podem ser obtidos pelas equações (2.19), (2.20) e (2.21), respectivamente.

Partindo de (2.22) e (4.6) obtém-se a fgm da distribuição LP como

$$M(t) = e^{-\beta t} + \alpha^{-1} \sum_{k=0}^{\infty} \frac{v_k {}_1F_1(1; 1 - (k+1)\alpha; -\beta t)}{(k+1)}, \quad t < 0. \quad (4.10)$$

4.2.3 Entropias

Entropia de Shannon

Substituindo (4.6) em (2.24) fornece

$$\vartheta_S(f) = -\log\left(\frac{\theta}{e^\theta - 1}\right) - \theta E[G(X)] - E\{\log[g(X)]\},$$

em que

$$E[G(X)] = \frac{e^\theta(\theta - 1) + 1}{\theta(e^\theta - 1)}, \quad \text{como calculado no capítulo anterior e}$$

$$E\{\log[g(X)]\} = \int_0^\infty \log[g(x)] f(x) dx.$$

Substituindo (3.2), expandindo em serie de Taylor e realizando uma mudança de variável origina

$$= \sum_{k=0}^{\infty} \frac{\theta^{k+1}}{(e^{\theta} - 1)k!} \int_0^1 u^k \log\{g[Q_G(u)]\} du,$$

em que $Q_G(u)$ é a função quantílica da distribuição Lomax. Assim,

$$\begin{aligned} E\{\log[g(X)]\} &= \sum_{k=0}^{\infty} \frac{\theta^{k+1}}{(e^{\theta} - 1)k!} \int_0^1 u^k \log \left[\frac{\alpha}{\beta} (1-u)^{\frac{\alpha+1}{\alpha}} \right] du \\ &= \log \left(\frac{\alpha}{\beta} \right) \sum_{k=0}^{\infty} w_k + \frac{\alpha+1}{\alpha} \sum_{k=0}^{\infty} \frac{\theta^{k+1}}{(e^{\theta} - 1)k!} \int_0^1 u^k \log(1-u) du \\ &= \log(\alpha) - \log(\beta) + \frac{\alpha+1}{\alpha} \sum_{k=0}^{\infty} \frac{\theta^{k+1}}{(e^{\theta} - 1)k!} \sum_{m=0}^k \frac{(-1)^{m+1} \binom{k}{m}}{(m+2)^2}. \end{aligned}$$

Entropia de Rényi

Para a distribuição LP pode-se expandir $f^c(x)$ em série de Taylor como

$$f(x)^c = \frac{\theta^c}{(e^{\theta} - 1)^c} \sum_{m=0}^{\infty} \frac{(\theta c)^m}{m!} \sum_{l=0}^m (-1)^l \binom{m}{l} \overline{G}(x)^l g(x)^c.$$

Então, utilizando (2.26), pode-se escrever a entropia de Rényi, dada por (2.25), da distribuição LP como

$$\vartheta_R(f; c) = \frac{1}{1-c} \log \left\{ \beta \left[\frac{\theta \alpha}{\beta(e^{\theta} - 1)} \right]^c \sum_{m=0}^{\infty} \sum_{l=0}^m \frac{(-1)^l \binom{m}{l} (\theta c)^m}{[(l+c)\alpha + c - 1] m!} \right\}.$$

4.2.4 Desvios Médios

Para obter δ_1 , o desvio com relação a média, e δ_2 , o desvio com a relação a mediana, a partir das equações (2.28) e (2.29) é necessário calcular o primeiro momento incompleto, dado por (4.9), e a mediana, dada por (4.5).

As curvas de Lorenz e Bonferroni, dadas por (2.30) e (2.31), respectivamente, também podem ser calculadas, utilizando-se (4.9).

4.2.5 Momentos Probabilisticamente Ponderados

Os momentos probabilisticamente ponderados (MPPs) são aqui definidos por

$$M(p, r, s) = E\{X^p [1 - \overline{F}(X)]^r [\overline{F}(X)]^s\},$$

em que p , r e s são inteiros positivos. Usando o teorema binomial, pode-se escrever

$$M(p, r, s) = \sum_{j=0}^r (-1)^j \binom{r}{j} E[X^p \bar{F}(X)^{s+j}].$$

Para a distribuição LP o seguinte resultado é verdadeiro:

$$f(x) \bar{F}(x)^n = \sum_{m=0}^n \sum_{k=0}^{\infty} \frac{(-1)^{m+k} \theta^{k+1} (m+1)^k \binom{n}{m}}{(1 - e^{-\theta})^{n+1} (k+1)!} g_{(k+1)\alpha, \beta}(x). \quad (4.11)$$

Dessa forma, utilizando (4.11), é possível reescrever os MPPs, para $p < \alpha$, como

$$M(p, r, s) = \sum_{k=0}^{\infty} \sum_{j=0}^r \sum_{m=0}^{s+j} \frac{(-1)^{j+m+k} \theta^{k+1} (m+1)^k \binom{r}{j} \binom{s+j}{m}}{(1 - e^{-\theta})^{s+j+1} (k+1)!} E[Y_{(k+1)\alpha, \beta}^p],$$

em que $E[Y_{(k+1)\alpha, \beta}^p]$ é dado por (2.12).

4.2.6 Estatísticas de Ordem

Seja X_1, X_2, \dots, X_n uma amostra aleatória da distribuição LP. A densidade de $X_{i:n}$, a i -ésima estatística de ordem, é dada por

$$f_{i:n}(x) = \frac{f(x)}{B(n-i+1, i)} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} \bar{F}(x)^{n+j-i}.$$

De (4.11), têm-se

$$f_{i:n}(x) = \sum_{k=0}^{\infty} \omega_k g_{(k+1)\alpha, \beta}(x), \quad (4.12)$$

em que

$$\omega_k = \sum_{j=0}^{i-1} \frac{(-1)^j \binom{i-1}{j}}{B(n-i+1, i)} \sum_{m=0}^{n+j-i} \frac{(-1)^{m+k} \theta^{k+1} (m+1)^k \binom{n+j-i}{m}}{(1 - e^{-\theta})^{n+j-i+1} (k+1)!}.$$

A equação (4.12) mostra que a função densidade das estatísticas de ordem de uma amostra da distribuição LP pode ser expressa como uma combinação linear de densidades da Lomax. Assim, é possível utilizar (4.12) para obter diversas propriedades das estatísticas de ordem de uma amostra da distribuição LP baseadas nas mesmas propriedades da distribuição Lomax. Por exemplo, o p -ésimo momento ordinário da i -ésima estatística de ordem, para $p < \alpha$, é dado por

$$E(X_{i:n}^p) = \sum_{k=0}^{\infty} \omega_k E[Y_{(k+1)\alpha, \beta}^p].$$

em que $E[Y_{(k+1)\alpha, \beta}^p]$ é dado por (2.12).

4.3 Estimadores de Máxima Verossimilhança

Suponha que x_1, x_2, \dots, x_n é uma amostra aleatória de $X \sim \text{LP}(\alpha, \beta, \theta)$. Então, a função de log-verossimilhança $\ell = \ell(\alpha, \beta, \theta; \mathbf{x})$, pode ser expressa como

$$\ell = n \log \left[\frac{\alpha \theta}{\beta(e^\theta - 1)} \right] + \theta \left[n - \sum_{i=1}^n \left(1 + \frac{x_i}{\beta} \right)^{-\alpha} \right] - (\alpha + 1) \sum_{i=1}^n \log \left(1 + \frac{x_i}{\beta} \right). \quad (4.13)$$

Ao se maximizar (4.13) em relação a α , β e θ utilizando uma rotina numérica de algum software, como a função `optim` do software R, por exemplo, obtém-se as estimativas de máxima verossimilhança (EMVs) $\hat{\alpha}$, $\hat{\beta}$ e $\hat{\theta}$.

As primeiras derivadas de ℓ em relação a α , β e θ são

$$\begin{aligned} \frac{\partial \ell}{\partial \alpha} &= \frac{n}{\alpha} + \sum_{i=1}^n \log \left(1 + \frac{x_i}{\beta} \right) \left[\theta \left(1 + \frac{x_i}{\beta} \right)^{-\alpha} - 1 \right], \\ \frac{\partial \ell}{\partial \beta} &= \sum_{i=1}^n \left(1 + \frac{x_i}{\beta} \right)^{-1} \left(\frac{-x_i}{\beta^2} \right) \left[\alpha \theta \left(1 + \frac{x_i}{\beta} \right)^{-\alpha} - (\alpha + 1) \right] - \frac{n}{\beta}, \\ \frac{\partial \ell}{\partial \theta} &= n \left(\frac{e^\theta(1 - \theta) - 1}{\theta(e^\theta - 1)} + 1 \right) - \sum_{i=1}^n \left(1 + \frac{x_i}{\beta} \right)^{-\alpha}. \end{aligned}$$

Também é possível obter as EMVs ao se solucionar as equações não lineares $\partial \ell / \partial \alpha = 0$, $\partial \ell / \partial \beta = 0$ e $\partial \ell / \partial \theta = 0$ simultaneamente.

A matriz de informação observada para (α, β, θ) é

$$\mathbf{K} = - \begin{pmatrix} \partial^2 \ell / \partial \alpha^2 & \partial^2 \ell / \partial \alpha \partial \beta & \partial^2 \ell / \partial \alpha \partial \theta \\ \partial^2 \ell / \partial \beta \partial \theta & \partial^2 \ell / \partial \beta^2 & \partial^2 \ell / \partial \beta \partial \theta \\ \partial^2 \ell / \partial \alpha \partial \theta & \partial^2 \ell / \partial \alpha \partial \beta & \partial^2 \ell / \partial \theta^2 \end{pmatrix},$$

em que

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \alpha^2} &= -\frac{n}{\alpha^2} - \theta \sum_{i=1}^n A(x_i)^{-\alpha} \{\log[A(x_i)]\}^2, \\ \frac{\partial^2 \ell}{\partial \alpha \partial \beta} &= \theta \sum_{i=1}^n A(x_i)^{-(\alpha+1)} \left(\frac{-x_i}{\beta^2} \right) \{-\alpha \log[A(x_i)] + 1\} + \sum_{i=1}^n \frac{x_i}{A(x_i)\beta^2}, \\ \frac{\partial^2 \ell}{\partial \alpha \partial \theta} &= \sum_{i=1}^n A(x_i)^{-\alpha} \log[A(x_i)], \\ \frac{\partial^2 \ell}{\partial \beta^2} &= \frac{n}{\beta^2} + \alpha \theta \sum_{i=1}^n \left[-(\alpha + 1) A(x_i)^{-(\alpha+2)} \left(\frac{-x_i}{\beta^2} \right)^2 + A(x_i)^{-(\alpha+1)} \times \right. \end{aligned}$$

$$\left(\frac{2x_i}{\beta^3}\right) - (\alpha + 1) \sum_{i=1}^n \left[-A(x_i)^{-2} \left(\frac{-x_i}{\beta^2}\right)^2 + A(x_i)^{-1} \left(\frac{2x_i}{\beta^3}\right) \right],$$

$$\frac{\partial^2 \ell}{\partial \beta \partial \theta} = \alpha \sum_{i=1}^n A(x_i)^{-(\alpha+1)} \left(\frac{-x_i}{\beta^2}\right) \text{ e}$$

$$\frac{\partial^2 \ell}{\partial \theta^2} = n \left[\frac{\theta^2 e^\theta - (e^\theta - 1)^2}{\theta^2 (e^\theta - 1)^2} \right],$$

com $A(x_i) = 1 + \beta^{-1}x_i$. A matriz \mathbf{K} é necessária para construir estimativas intervalares e realizar testes de hipóteses. Seja $\Theta = (\alpha, \beta, \theta)$ o conjunto de parâmetros desconhecidos e $\hat{\Theta}$ o estimador de máxima verossimilhança de Θ . Sob condições padrões de regularidade, $\sqrt{n}(\hat{\Theta} - \Theta)$ é assintoticamente normal, $N_3(0, \mathbf{I}(\Theta)^{-1})$, em que $\mathbf{I}(\Theta)$ é a matriz de informação esperada. É possível substituir $\mathbf{I}(\Theta)$ por $\mathbf{K}(\hat{\Theta})$, a matriz de informação observada avaliada em $\hat{\Theta}$, uma vez que esta é um estimador consistente de $\mathbf{I}(\Theta)$. Intervalos de confiança assintóticos para os parâmetros do modelo LP podem ser construídos a partir da distribuição normal multivariada $N_3(0, \mathbf{K}(\hat{\Theta})^{-1})$.

As estatísticas de razão de verossimilhanças (RV) podem ser construídas de forma a verificar se a distribuição LP apresenta um melhor ajuste que a distribuição Lomax, por exemplo. Nesse caso, a hipótese $H_0 : \theta = 0$ seria testada contra a hipótese $H_1 : \theta \neq 0$. Para qualquer par de hipóteses, a estatística RV é igual a

$$LR = 2[\ell(\hat{\alpha}, \hat{\beta}, \hat{\theta}) - \ell(\tilde{\alpha}, \tilde{\beta}, \tilde{\theta})],$$

em que $\hat{\alpha}$, $\hat{\beta}$, $\hat{\theta}$ são as EMVs sob H_1 e $\tilde{\alpha}$, $\tilde{\beta}$, $\tilde{\theta}$ são as estimativas sob H_0 .

4.4 Aplicação

Para ilustrar a utilidade da distribuição LP, será utilizado um conjunto de dados reais, ao qual serão ajustadas diversas outras distribuições de sobrevivência, além do modelo LP. O conjunto de dados representa o tempo de remissão (em meses) de uma amostra aleatória de 128 pacientes com câncer de bexiga, apresentado em Lee e Wang (2003). Na Tabela 4.1, apresentam-se as EMVs, as estimativas dos erros-padrões, o AIC e o BIC para as distribuições Lomax, Weibull, MOEE, MOEW, MOEL e LP. Os resultados indicam que a distribuição LP tem os menores valores AIC e BIC, entre todos os modelos ajustados, provendo, portanto, um melhor ajuste. É importante salientar que as distribuições

Tabela 4.1: EMVs dos parâmetros dos modelos e as estimativas dos erros-padrões bem como as estatísticas AIC e BIC correspondentes.

Distribuição	α	β	λ	θ	a	AIC	BIC
Lomax	12,11 (10,74)	104,1 (99,28)	-	-	-	831,68	837,38
Weibull	1,04 (0,06)	9,56 (0,85)	-	-	-	832,17	837,87
WE	0,65 (0,13)	3,34 (1,88)	-	-	2,79 (1,26)	827,36	835,91
MOEE	-	9,10 (1,64)	1,05 (0,32)	-	-	832,65	838,35
MOEW	1,06 (0,16)	34,80 (15,24)	0,06 (0,06)	-	-	826,18	834,74
MOEL	2,31 (0,55)	2,24 (2,62)	21,20 (29,06)	-	-	825,08	833,64
LP	2,87 (0,88)	8,26 (4,87)	-	3,35 (1,03)	-	824,77	833,32

WE, MOEL e LP possuem o mesmo número de parâmetros e que o modelo WE é muito popular em estudos de sobrevivência. A Figura 4.3, mostra o histograma dos dados bem como as distribuições MOEL e LP ajustadas.

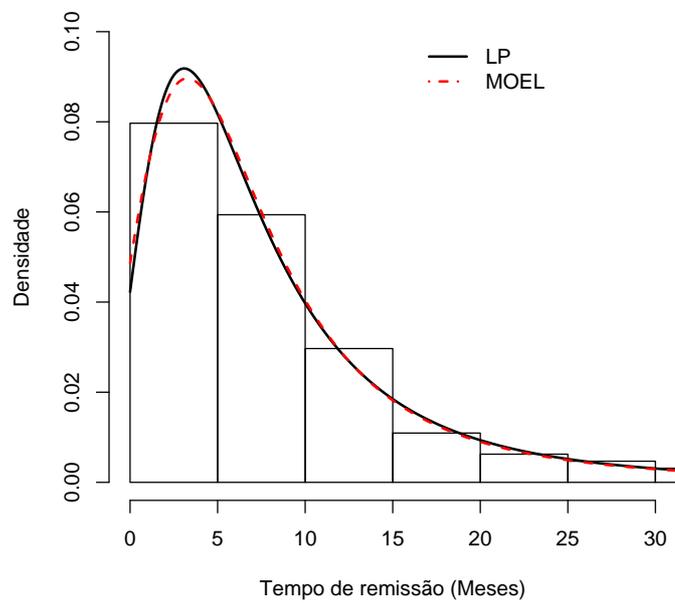


Figura 4.3: Histograma do tempo de remissão (meses) e as densidades ajustadas.

Considerações finais

Nesse trabalho foram abordadas duas classes de distribuições, a classe Marshall-Olkin, proposta por Marshall e Olkin (1997), e a classe G-Poisson, equivalente a classe exp-G , proposta por Barreto-Souza e Simas (2013). Novas propriedades da distribuição Marshall-Olkin extendida Lomax foram calculadas bem como diversas novas propriedades da classe G-Poisson. A distribuição Lomax-Poisson também foi proposta e estudada neste trabalho.

Novas abordagens, contudo, ainda são necessárias. Um possível estudo futuro seria uma análise multivariada da classe G-Poisson. Uma comparação entre as classes G-exponencializada, Marshall-Olkin e G-Poisson, também é de interesse, visto que os três métodos adicionam um novo parâmetro a uma distribuição conhecida.

Referências Bibliográficas

- [1] Adamidis, K., Loukas, S. (1998). A lifetime distribution with decreasing failure rate. *Statistics & Probability Letters*, **39**, 35-42.
- [2] Bakouch, H.S., Ristić, M.M., Asgharzadeh, A., Esmaily, L., Al-Zahrani, B. M. (2012). An exponentiated exponential binomial distribution with application. *Statistics and Probability Letters*, **82**, 1067-1081.
- [3] Barreto-Souza, W., Bakouchb, H.S. (2011). A new lifetime model with decreasing failure rate. *Statistics*, doi:10.1080/02331888.2011.595489.
- [4] Barreto-Souza, W., Morais, A.L., Cordeiro, G.M. (2011). The Weibull-geometric distribution. *Journal of Statistical Computation and Simulation*, **81**, 645-657.
- [5] Barreto-Souza, W., Simas, A.B. (2013). The exp- G family of probability distributions. *Brazilian Journal of Probability and Statistics*, **27**, 84-109.
- [6] Bjerkedal, T. (1960). Acquisition of resistance in guinea pigs infected with different doses of virulent tubercle bacilli. *American Journal of Hygiene*, **72**, 130-148.
- [7] Chahkandi, M., Ganjali, M. (2009). On some lifetime distributions with decreasing failure rate. *Computational Statistics & Data Analysis*, **53**, 4433-4440.
- [8] Ciumara, R., Preda, V. (2009). The Weibull-logarithmic distribution in lifetime analysis and its properties. *The XIII International Conference "Applied Stochastic*

- Models and Data Analysis*”, Institute of Mathematics and Informatics, Vilnius, 395-399.
- [9] Ghitany, M.E., Al-Awadhi, F.A., Alkhalfan, L.A. (2007). Marshall-Olkin extended Lomax distribution and its application to censored data. *Communication in Statistics - Theory and Methods*, **36**, 1855-1866.
- [10] Gradshteyn, I.S., Ryzhik, I.M. (2007). *Table of Integrals, Series and Products*. Academic Press, San Diego.
- [11] Gupta, R.C., Kannan, N., Raychaudhuri, A. (1997). Analysis of Lognormal Survival Data. *Mathematical Biosciences*, **139**, 103-105.
- [12] Gupta, R.D., Kundu, D. (1999). Generalized Exponential Distributions. *Australian & New Zealand Journal of Statistics*, **41**, 173-188.
- [13] Jong, P., Heller, G.Z. (2008). *Generalized Linear Models for Insurance Data*. Cambridge University Press, New York.
- [14] Kus, C. (2007). A new lifetime distribution. *Computational Statistics & Data Analysis*, **51**, 4497-4509.
- [15] Lee, E., Wang, J.W. (2003). *Statistical Methods for Survival Data Analysis*. Wiley, New York.
- [16] Lemonte, A.J., Cordeiro, G.M. (2011). An Extended Lomax Distribution. *Statistics: A Journal of Theoretical and Applied Statistics*, doi:10.1080/02331888.2011.568119.
- [17] Lomax, K.S. (1954). Another Example of the Analysis of Failure Data. *Journal of the American Statistical Association*, **49**, 847-852.
- [18] Lu, W., Shi, D. (2012). A new compounding life distribution: the Weibull-Poisson distribution. *Journal of Applied Statistics*, **39**, 21-38.
- [19] Marshall, A.W., Olkin, I. (1997). A New Method for Adding a Parameter to a Family of Distributions with Application to the Exponential and Weibull Families. *Biometrika*, **84**, 641-652.

- [20] Morais, A.L., Barreto-Souza, W. (2011). A compound class of Weibull and power series distributions. *Computational Statistics & Data Analysis*, **55**, 1410-1425.
- [21] Mudholkar, G.S., Srivastava, D.K. (1993). Exponentiated Weibull Family for Analyzing Bathtub Failure-Rate Data. *IEEE Transactions on Reliability*, **42**, 299-302.
- [22] Nadarajah, S., Kotz, S. (2006). The Exponentiated Type Distributions. *Acta Appl Math*, **92**, 97-111.
- [23] Paula, G.A. (2012). *MODELOS DE REGRESSÃO com apoio computacional*. Instituto de Matemática e Estatística - USP, São Paulo.
- [24] Ristić, M.M., Nadarajah, S. (2010). *A New Lifetime Distribution*. Technical report, University of Manchester.
- [25] Rodrigues, C., Cordeiro, G.M., Demétrio, C.G.B., Ortega, E.M.M. (2011). The Weibull Negative Binomial Distribution *Advances and Applications in Statistics*, **22**, 25-55.
- [26] Tahmasbi, R., Rezaei, S. (2008). A two-parameter lifetime distribution with decreasing failure rate. *Computational Statistics & Data Analysis*, **52**, 3889-3901.
- [27] Weibull, W. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics*, **18**, 293-296.