

**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

**As distribuições Kumaraswamy-log-logística e
Kumaraswamy-logística**

Tiago Viana Flor de Santana

Dissertação apresentada para obtenção do título de Mestre
em Ciências. Área de concentração: Estatística e Experi-
mentação Agronômica

Piracicaba

2010

Tiago Viana Flor de Santana
Matemático

**As distribuições Kumaraswamy-log-logística e
Kumaraswamy-logística**

Orientador:
Profº Drº EDWIN MOISÉS MARCOS ORTEGA

Dissertação apresentada para obtenção do título de Mestre em
Ciências. Área de concentração: Estatística e Experimentação
Agronômica

Piracicaba
2010

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/USP**

Santana, Tiago Viana Flor de

A distribuições Kumaraswamy-log-logística e Kumaraswamy-logística / Tiago Viana Flor de Santana. -- Piracicaba, 2010.
103 p. : il.

Dissertação (Mestrado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2010.
Bibliografia.

1. Análise de regressão e de correlação 2. Análise de sobrevivência 3. Distribuições (Probabilidade) 4. Logística I. Título

CDD 519.532
S232d

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

DEDICATÓRIA

A

DEUS

por me carregar nos momentos mais difíceis.

Aos meus pais,

Levi Flor de Santana e

Valnice Aparecida dos Santos de Santana,

pelas orações, incentivo e esforço que fazem para que eu possa continuar em frente.

À minha irmã

Talita Flor de Santana,

por sempre estar ao meu lado torcendo por mim.

À minha esposa

Ione Viana Flor de Santana,

pelo amor, apoio e compreensão nos momentos ausentes.

AGRADECIMENTO

À Deus, pela força, sustento e por mais esta benção que me tem concedido.

A ao meu pai **Levi Flor de Santana**, a minha mãe **Valnice Aparecida dos Santos de Santana**, a minha irmã **Talita Flor de Santana** e a minha esposa **Ione Viana Flor de Santana** pelas orações, incentivo, amor e apoio incondicional.

Ao meu orientador, Prof. Dr. Edwin Moises Marcos Ortega, pela orientação nesses anos.

Agradecimento especial aos meus tios Livonildo Flor de Santana e Severina Lucia de Souza pelas acomodações, alimento e carinho que contribuiram na evolução desse trabalho.

Agradecimento especial aos meus amigos Diógenes Ferreira Filho, Mateus Augusto Donegá, Epaminondas de Vasconcelos Couto, Tiago Almeida de Oliveira, Marcelino Alves Rosa de Pascoa, Ana Patrícia Bastos Peixoto, Simone Daniela Sartório, Elizabeth Mie Hashimoto, Fabiane de Lima Silva e Gláucia Tatiana Ferrari pelo compartilhamento de conhecimento, dicas e companherismo. E a Eva que tem cuidado de mim por essas estradas.

A coordenadoria da Divisão de Atendimento a Comunidade - Serviço de Promoção Social, em especial a Áurea e a Solange, pelo tratamento e pela Vila Estudantil da Pós Graduação que permitiu minha moradia em Piracicaba e prosseguimento nos estudos enquanto sem bolsa.

Aos professores e funcionários do Departamento de Ciências Exatas da ESALQ/USP, em especial a secretária Solange de Assis Paes Sabadin por estar sempre pronta a ajudar e resolver nossos “*pepinos*”.

Aos amigos dos cursos de mestrado e doutorado do Departamento de Ciências Exatas da ESALQ/USP, em especial ao Vanderlir, Lucimari, Rodrigo Digão, Rafael Kacond, Mariana, Renata e Marina que sempre me auxiliaram em momentos de dúvidas e dificuldades computacionais.

À Coordenadoria para o Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro.

E a todas as pessoas que de forma direta ou indireta, contribuíram para a

“O temor do SENHOR é o princípio da ciência; os loucos desprezam a sabedoria e a instrução”

Provérbios 1:7

“E, de mais disso, filho meu, atenta: não há limite para fazer livros, e o muito estudar enfado é da carne. De tudo o que se tem ouvido, o fim é: Teme a Deus e guarda os seus mandamentos; porque este é o dever de todo homem.”

Eclesiastes 12:12-13

SUMÁRIO

	Página
RESUMO	11
ABSTRACT	13
LISTA DE FIGURAS	15
LISTA DE TABELAS	17
LISTA DE SÍMBOLOS	19
1 INTRODUÇÃO	21
2 DESENVOLVIMENTO	25
2.1 Conceitos básicos em análise de sobrevivência	25
2.2 Classes de distribuições generalizadas	28
2.2.1 Classe de distribuições Marshall e Olkin	29
2.2.2 Classe de distribuições exponenciadas	30
2.2.3 Classe de distribuições estendidas	33
2.2.4 Classe de distribuições betas	35
2.2.5 Classe de distribuições Kumaraswamy	37
3 DISTRIBUIÇÕES LOG-LOGÍSTICA MODIFICADAS	39
3.1 Distribuição log-logística	39
3.2 Distribuição log-logística modificada	42
3.3 Distribuição Kumaraswamy-log-logística	47
3.4 Propriedades da Kumaraswamy-log-logística	53
3.4.1 Expansão da função densidade da Kw-LL	53
3.4.2 Momentos	55
3.4.3 Outras medidas associadas à distribuição Kw-LL	57
3.4.4 Desvio médio	57
3.4.5 Curvas de Bonferroni e Lorenz	58
3.4.6 Estatísticas de ordem	59
3.5 Estimação de máxima verossimilhança com dados censurados	62
4 DISTRIBUIÇÕES LOGÍSTICA GENERALIZADA	65

10	
4.1 Distribuição logística	65
4.2 Distribuição logística modificada	68
4.3 Distribuição Kumaraswamy-logística	69
4.4 Momentos e função geradora de momentos	72
4.5 Modelo de regressão	73
4.5.1 Modelo de regressão Kw-logístico	74
4.6 Estimação de máxima verossimilhança	75
5 APLICAÇÕES	79
5.1 Aplicação 1: Dados de AIDS	79
5.2 Aplicação 2: Dados de níveis de voltagem	84
6 CONCLUSÕES	89
6.1 Pesquisas futuras	89
REFERÊNCIAS	91
APÊNDICE	97

RESUMO

As distribuições Kumaraswamy-log-logística e Kumaraswamy-logística

Neste trabalho apresenta-se duas novas distribuições de probabilidade obtidas de dois métodos de generalização da distribuição log-logística com dois parâmetros ($LL(\alpha, \gamma)$). O primeiro método descrito em Marshall e Olkin (1997) transforma a nova distribuição, agora com três parâmetros e denominada distribuição log-logística modificada ($LLM(v, \alpha, \gamma)$), mais flexível porém, não muda a forma geral da função de taxa de falha e o novo parâmetro v , não influênciaria no cálculo da assimetria e curtoza. O segundo método utiliza a classe de distribuições Kumaraswamy proposta por Cordeiro e Castro (2010), para construir a nova distribuição de probabilidade, denominada distribuição Kumaraswamy log-logística ($Kw-LL(a, b, \alpha, \gamma)$), a qual considera dois novos parâmetros a e b obtendo ganho nas formas da função de taxa de falha, que agora além de modelar dados onde a função de taxa de falha tem forma decrescente e unimodal, modela forma crescente e forma de U. Também foi proposto as distribuições logística modificada ($LM(v, \mu, \sigma)$) e Kumaraswamy logística ($Kw-L(a, b, \mu, \gamma)$) para a variável $Y = \log(T)$, em que $T \sim LLM(v, \alpha, \gamma)$ no caso da distribuição logística modificada e $T \sim Kw-LL(a, b, \alpha, \gamma)$ no caso da distribuição Kw-L. Com reparametrização $\alpha = \exp(\mu)$ e $\gamma = 1/\sigma$. Da mesma forma que a distribuição LLM, não há ganho quanto a forma da função de taxa de falha da distribuição logística modificada e o parâmetro v não contribuiu para o cálculo da assimetria e curtoza desta distribuição. O modelo de regressão locação e escala foi proposto para ambas as distribuições. Por fim, utilizou-se dois conjuntos de dados, para exemplificar o ganho das novas distribuições Kw-LL e Kw-L em relação as distribuições log-logística e logística. O primeiro conjunto refere-se a dados de tempo até a soro-reversão de 143 crianças expostas ao HIV por via vertical, nascidas no Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto no período de 1995 a 2001, onde as mães não foram tratadas. O segundo conjunto de dados refere-se ao tempo até a falha de um tipo de isolante elétrico fluido submetido a sete níveis de voltagem constante.

Palavras-chave: Função de taxa de falha; Distribuição Kumaraswamy log-logística; Distribuição log-logística modificada; Regressão Kumaraswamy logística; Regressão logística modificada

ABSTRACT

Distributions Kumaraswamy-log-logistic and Kumaraswamy-logistic

In this work, are presented two new probability distributions, obtained from two generalization methods of the log-logistic distribution, with two parameters ($LL(\alpha, \gamma)$). The first method described in Marshall e Olkin (1997) turns the new distribution, now with three parameters, called modified log-logistic distribution ($LLM(v, \alpha, \gamma)$). This distribution is more flexible, but, does not change the general shape of the failure rate function, as well as the new parameter v , does not influence the calculus of skewness and kurtosis. The second method, uses the class of distributions Kumaraswamy proposed by Cordeiro and Castro (2010). To build the new probability distribution, called Kumaraswamy log-logistic distribution ($Kw-LL(a, b, \alpha, \gamma)$), which considers two new parameters a and b gaining in the forms of failure rate function, that now, even modeling data where the failure rate function has decreasing and unimodal shape, models the increasing form and the U-shaped. Also, were proposed the distributions modified logistic ($LM(v, \mu, \sigma)$) and Kumaraswamy logistics ($Kw-L(a, b, \mu, \gamma)$) for the variable $Y = \log(T)$, where $T \sim LLM(v, \alpha, \gamma)$ in the case of the modified logistic distribution and $T \sim Kw-LL(a, b, \alpha, \gamma)$ in the case of Kw-L distribution, with reparametrization $\alpha = \exp(\mu)$ and $\gamma = 1/\sigma$. As in the distribution LLM , there is no gain for the shape of the failure rate function of modified logistic distribution and the parameter v does not contribute to the calculation of skewness and kurtosis of the distribution. The location and scale regression models were proposed for both distributions. As illustration, were used two datasets to exemplify the gain of the new distributions $Kw-LL$ and $Kw-L$ compared with the log-logistic and logistic distributions. The first dataset refers to the data of time until soro-reversion of 143 children exposed to HIV through vertical, born in the Hospital of the Medical School of Ribeirão Preto during the period 1995 to 2001, where mothers were not treated. The second dataset refers to the time until the failure of a type of electrical insulating fluid subjected to seven constant voltage levels.

Keywords: Failure rate function; Kumaraswamy log-logistic distribution; Modified log-logistic distribution; Modified logistic regression; Kumaraswamy logistic regression

LISTA DE FIGURAS

	Página
Figura 1 - Exemplo de funções de taxa de falha	27
Figura 2 - Gráficos TTT-plot	28
Figura 3 - Gráficos da fdp, função de sobrevivência e função de taxa de falha da distribuição log-logística para alguns valores γ e $\alpha = 25$	41
Figura 4 - Gráficos da fdp da distribuição LLM para $\gamma = 25$, α assumindo os valores 1, 2 e 4 e v assumindo os valores 1, 5, 10 e 20	44
Figura 5 - Gráficos da função de taxa de falha da distribuição LLM para $\gamma = 25$, α assumindo os valores 1, 2 e 4 e v assumindo os valores 1, 5, 10 e 20	45
Figura 6 - Gráficos da fdp da distribuição Kw-LL. Em (a), (b) e (c) são apresentados os gráficos da fdp fixando $a = 1$, para $\gamma < 1$, $\gamma = 1$ e $\gamma > 1$ respectivamente	48
Figura 7 - Gráficos da fdp da distribuição Kw-LL. Em (a), (b) e (c) são apresentados os gráficos da fdp fixando $b = 1$, para $\gamma < 1$, $\gamma = 1$ e $\gamma > 1$ respectivamente	49
Figura 8 - Gráficos da fdp e da função de sobrevivência da distribuição Kw-LL para alguns valores dos parâmetros a , b , α , γ	50
Figura 9 - Gráficos apresentando as formas da função de taxa de falha da distribuição Kw-LL. (a) Apresenta-se os gráficos da função de taxa de falha crescente e decrescente. (b) Apresenta-se o gráfico da função de taxa de falha unimodal. (c) Apresenta-se o gráfico da função de taxa de falha forma de U ou forma de banheira	51
Figura 10 - Gráficos da assimetria e curtose da distribuição Kw-LL. Em (a) e (b) apresenta-se os gráficos da assimetria e curtose como função do parâmetro a , para alguns valores do parâmetro b . Em (c) e (d) apresenta-se os gráficos da assimetria e curtose como função do parâmetro b , para alguns valores do parâmetro a	57
Figura 11 - Gráficos da fdp, função de sobrevivência e função de taxa de falha da distribuição logística para alguns valores σ e μ	66
Figura 12 - Gráficos da fdp e da função de taxa de falha da distribuição LM, para $\mu = 2$, $\sigma = 3$ e v assumindo valores 0, 5, 1, 5 e 10	69

Figura 13 - Gráficos da fdp da distribuição Kw-L. (a) Apresenta-se os gráficos da fdp da distribuição Kw-L para $\mu = 0$, $\sigma = 1$, $a = 1$ e b assumindo valores 0, 05, 0, 1, 0, 5, 1 e 3. (b) Apresenta-se os gráficos da fdp da distribuição Kw-L para $\mu = 0$, $\sigma = 1$, $b = 1$ e a assumindo valores 0, 05, 0, 01, 0, 5, 1 e 3	71
Figura 14 - Gráficos da fdp e da função de sobrevivência da distribuição Kw-L. (a) Apresenta-se a fdp para alguns valores dos parâmetros a , b , μ e σ . (b) Apresenta-se a função de sobrevivência para alguns valores dos parâmetros a , b , μ e σ	72
Figura 15 - Gráficos da assimetria e curtose da distribuição Kw-L. (a) e (b) apresenta-se os gráficos da assimetria e curtose como função do parâmetro a , para alguns valores do parâmetro b . (c) e (d) apresenta-se os gráficos da assimetria e curtose como função do parâmetro b , para alguns valores do parâmetro a	73
Figura 16 - Gráficos da distribuição dos erros aleatórios para alguns valores dos parâmetros a e b . (a) Apresenta-se os gráficos da fdp da distribuição dos erros aleatórios para $b = 1$ e a assumindo valores 0, 5, 1, 1, 8, 3 e 5. (b) Apresenta-se os gráficos da fdp da distribuição dos erros aleatórios para $a = 1$ e b assumindo valores 0, 5, 1, 1, 8, 3 e 5	75
Figura 17 - Gráfico TTT para os dados de soro reversão de 143 crianças expostas ao HIV por via vertical	80
Figura 18 - Gráfico da função de taxa de falha da distribuição Kw-LL para os valores estimados de a , b , α e γ	82
Figura 19 - Funções de sobrevivência estimada e empírica para os dados de AIDS	82
Figura 20 - Gráfico TTT para os dados de níveis de voltagem	85
Figura 21 - Curvas de sobrevivência estimadas pelos modelos de regressão Kw-L, LM e logístico para dois níveis de voltagem 28 kV e 32 kV. (a) Apresenta-se os ajustes dos modelos Kw-L e logístico. (b) Apresenta-se os ajustes para os modelos Kw-L e LM	87
Figura 22 - Curvas de sobrevivência estimadas pelo modelo de regressão Kw-L para os dados de níveis de voltagem	87

LISTA DE TABELAS

	Página
Tabela 1 - EMV dos parâmetros dos modelos Kw-LL, log-logístico, LLM, Weibull e log-normal para os dados de AIDS, os EP correspondente (em parênteses) e as medidas AIC, BIC e CAIC	81
Tabela 2 - Tempo de falha (em minutos) dos isolantes elétricos em cada um dos sete níveis de voltagens	84
Tabela 3 - EMV dos parâmetros para os modelos de regressão logístico, Kw-logístico e LM ajustados aos dados de níveis de voltagem	86

LISTA DE SÍMBOLOS

T : tempo de falha;

C : tempo de censura ou tempo parcial de falha;

fda: função de distribuição acumulada;

fdp: função de distribuição de probabilidade;

v.a.: variável aleatória;

$\log(\cdot)$: logaritmo natural (base e), $\log(\cdot) = \ln(\cdot)$;

μ : letra grega mi;

β : letra grega beta;

$\boldsymbol{\beta}$: vetor de parâmetros no modelo de regresão locação e escala;

σ : letra grega sigma;

α : letra grega alfa;

γ : letra grega gama;

\mathbb{N} : conjunto dos número naturais;

\mathbb{R} : conjunto dos números reais;

\mathbb{R}^+ : conjunto dos números reais positivos;

\mathbb{R}_*^+ : $\mathbb{R}^+ - \{0\}$;

\mathbb{N} : conjunto dos número naturais;

$\Gamma(\cdot)$: função gama;

$B(\cdot, \cdot)$: função beta;

$B_{G(\cdot)}(\cdot, \cdot)$: função beta incompleta;

$M(\cdot, \cdot, \cdot)$: função ponderada por probabilidade;

1 INTRODUÇÃO

A análise de sobrevivência é uma das áreas da estatística que mais cresceu nas últimas décadas. A razão deste crescimento é o desenvolvimento e aprimoramento de técnicas estatísticas combinadas com computadores cada vez mais velozes, Colosimo e Giolo (2006). Uma evidência quantitativa desse sucesso é o número de aplicações em campos diversos, como medicina, biologia, saúde pública, epidemiologia, engenharia, economia, estudos demográficos, entre outros.

Os dados em análise de sobrevivência referem-se ao tempo até a ocorrência de um evento de interesse. Este tempo é denominado tempo de falha, e apresenta algumas características especiais. A primeira é que a variável resposta, geralmente, apresenta distribuição assimétrica positiva, não sendo portanto, adequado assumir que tenha distribuição normal. A segunda característica é a presença de dados censurados, isto é, para alguns elementos em estudo não se conhece o tempo de interesse exato sabe-se, apenas que o tempo de falha ocorreu à direita ou à esquerda do valor registrado.

Usualmente, dados censurados ocorrem, uma vez que nem sempre é possível que o evento de interesse ocorra para todas as observações em teste. Existem diversos tipos de censuras, e neste trabalho considerou-se censura à direita e mecanismo aleatório, além da censura ser não informativa.

Por outro lado, nos últimos anos tem crescido muito a generalização e/ou a modificação de algumas distribuições utilizadas em análise de sobrevivência. Existem diferentes formas de se modificar uma distribuição de probabilidade, sendo que algumas das classes de generalizações mais trabalhadas são as famílias de distribuições obtidas pelo método desenvolvido por Marshall e Olkin (1997), as distribuições exponenciadas apresentada inicialmente por Mudholkar et al. (1995), as distribuições estendidas discutidas por Barros (2008) e as distribuições betas que receberam maior atenção após o trabalho de Eugene et al. (2002).

Recentemente, Cordeiro e Castro (2010) apresentaram uma classe de generalização baseada na distribuição Kumaraswamy. A classe modela a maioria das formas de risco básicas, ou seja, formas crescente, decrescente, unimodal e forma de U ou de banheira. Cordeiro e Castro (2010) apresentaram ainda duas formas de expandir a função densidade de probabilidade (fdp) de uma distribuição Kumaraswamy generalizada em termos da série

ponderada de potência.

Assim, o objetivo deste trabalho é propor duas generalizações das distribuições log-logística e logística baseadas nos trabalhos de Marshall e Olkin (1997) e Cordeiro e Castro (2010) para modelar dados de sobrevivência. Estas generalizações serão denominadas de distribuições: log-logística modificada (LLM), logística modificada (LM), Kumaraswamy-log-logística (Kw-LL) e Kumaraswamy-logística (Kw-L). Apresenta-se propriedades destas novas distribuições, tais como duas expansões da função densidade de probabilidade (fdp) em série ponderada de potência, os momentos e a função geradora de momentos. São apresentadas também, expressões para algumas medidas, tais como a função do desvio médio da média e da mediana, as curvas de Bonferroni, curvas de Lorenz e as estatísticas de ordem da nova distribuição Kw-LL. Uma estrutura de regressão com dados censurados é proposta utilizando a distribuição Kw-L e LM e duas aplicações são apresentadas demonstrando o melhor ajuste das novas distribuições em relação às distribuições de origens e algumas outras bem conhecidas.

O trabalho está organizado da seguinte maneira. No capítulo 2 apresenta-se uma revisão sobre alguns conceitos fundamentais em análise de sobrevivência, também é apresentado o gráfico TTT proposto por Aarset (1987) que é utilizado para a determinação empírica da forma da função de taxa de falha. Neste capítulo ainda é feita uma revisão sobre os principais trabalhos envolvendo classes de generalizações tais como a classe de distribuições Marshall e Olkin, proposta por Marshall e Olkin (1997), as classes de distribuições exponenciadas apresentada por Mudholkar et al. (1995), as distribuições estendidas discutidas por Barros (2008), as distribuições betas que receberam maior atenção após o trabalho de Eugene et al. (2002) e as distribuições Kumaraswamy desenvolvida por Cordeiro e Castro (2010).

No capítulo 3 duas formas de generalizações da distribuição log-logística são apresentadas, resultando nas novas distribuições log-logística modificada e Kumaraswamy-log-logística e suas propriedades desenvolvidas. Na seção 3.4 as expressões para os momentos, para função geradora de momentos e também duas expansões em séries de potência para a fdp da distribuição Kw-LL é apresentada. E a função do desvio médio para média e mediana, as curvas de Bonferroni e curvas de Lorenz e a densidade das estatísticas de ordem são apresentados na seção 3.4.3. Na seção 3.5 o método de máxima verossimilhança é discutido

para estimar os parâmetros das novas distribuições na presença de dados censurados.

No capítulo 4 outras formas de generalizações da distribuição logística são apresentadas, resultando nas novas distribuições logística modificada e Kw-L e algumas propriedades destas distribuições são discutidas. O modelo de regressão locação e escala utilizando as distribuições logística modificada e Kw-L é proposto no capítulo 4.5. E por fim no capítulo 5 duas aplicações são discutidas utilizando os quatro novos modelos no ajuste dos dados.

2 DESENVOLVIMENTO

Neste capítulo é feito uma revisão sucinta sobre análise de sobrevivência e sobre classes de generalizações de distribuições. Na seção 2.1 os tipos de censuras, os tempos de falha, as funções probabilísticas em análise de sobrevivência e a metodologia de análise gráfica TTT (tempo total em teste) são discutidos.

Na seção 2.2 uma revisão dos principais trabalhos elaborados nos últimos anos envolvendo classes de generalizações é feita. A classe de distribuição Marshall e Olkin é apresentada e discutida na seção 2.2.1. As classes de distribuições exponenciadas, estendidas e betas, são tratadas na seções 2.2.2, 2.2.3 e 2.2.4, respectivamente, e os principais trabalhos comentados. E por fim, a classe de distribuição Kumaraswamy é apresentada e discutida na seção 2.2.5, comparando-a com as demais classes de distribuições.

2.1 Conceitos básicos em análise de sobrevivência

A análise de sobrevivência é caracterizada pelo fato que a variável resposta é composta de dois fatores, o tempo até a ocorrência de um evento de interesse e freqüentemente pelas censuras, Colosimo e Giolo (2006). O evento em estudo é denominado falha e o tempo até a ocorrência da falha é chamado de tempo de falha. A censura é o registro parcial do tempo de falha, devido a perda ou retirada de um elemento do estudo.

A censura é dita ser do tipo I quando ocorre devido ao término do estudo após um período de tempo pré-determinado e, do tipo II quando ocorre devido ao término do estudo após um número de falhas fixado previamente e censura aleatória, sendo a mais comum em situações práticas, quando um elemento deixa o estudo sem que o evento de interesse tenha ocorrido.

As censuras previamente citadas são conhecidas como censura à direita, pois a falha ocorre sempre à direita do tempo registrado. Existem ainda outros mecanismos de censura, como por exemplo, as censuras à esquerda, onde o tempo registrado é maior que o tempo de falha; e censura intervalar, onde não se sabe o tempo exato de falha, e a única informação disponível é que o tempo de falha ocorreu em um certo intervalo de tempo. Neste trabalho, entretanto, será adotado apenas o mecanismo de censura aleatória à direita que

será denominada simplesmente por censura.

Seja T uma variável aleatória não negativa que representa o tempo de falha de um elemento, e seja C uma variável aleatória, independente de T , que representa o tempo de censura associado a este elemento. Assim, os dados observados são representados por $t = \min(T, C)$ e δ o indicador de censura, dado por:

$$\delta = \begin{cases} 1 & , \quad T \leq C \\ 0 & , \quad T > C \end{cases},$$

em que $\delta = 0$ indica censura e $\delta = 1$ indica falha.

A distribuição de probabilidade da variável aleatória T pode ser especificada por meio da fdp, função de sobrevivência ou função de taxa de falha, sendo as três formas equivalentes.

A função de sobrevivência quantifica a probabilidade do tempo de falha de uma observação ser maior que t e é dada por

$$S(t) = P(T \geq t),$$

de onde segue que $S(t) = 1 - F(t)$.

A função de sobrevivência pode ser interpretada como a probabilidade de uma observação não falhar até o tempo t .

A função de taxa de falha pode ser obtida por meio da seguinte função

$$h(t) = \lim_{h \rightarrow 0} \frac{P(t \leq T < t + h | T \geq t)}{h}$$

que pode ser reescrito como

$$h(t) = \frac{f(t)}{S(t)}$$

onde $f(t)$ é a fdp da variável aleatória T .

A função de taxa de falha é interpretada como a probabilidade de que a falha ocorra no instante t dada que não correu antes de t . O gráfico da função de taxa de falha pode apresentar algumas formas básicas conforme pode ser visualizado na Figura 1, com diferentes

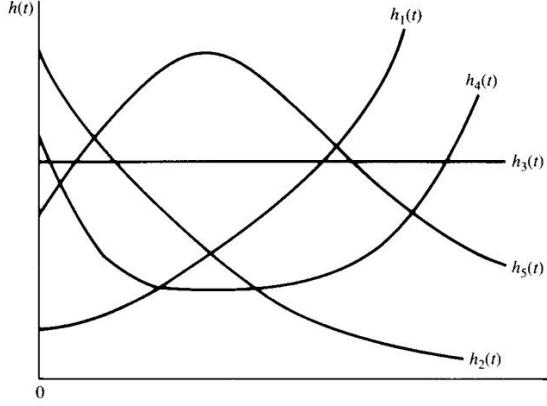


Figura 1 – Exemplo de funções de taxa de falha

interpretações práticas. O gráfico pode ser crescente ou decrescente conforme indicado por $h_1(t)$ e $h_2(t)$ respectivamente, constante representado por $h_3(t)$, ou alternar entre intervalos decrescentes, constantes e crescentes, como por exemplo $h_4(t)$, que é uma forma da função de taxa de falha conhecida por forma de U ou forma de banheira e $h_5(t)$ conhecida por forma unimodal, que inicia com um intervalo crescente até atingir um máximo e então decresce.

A função de taxa de falha é mais informativa do que a função de sobrevivência, pois diferentes funções de sobrevivência podem ter formas parecidas e funções de taxa de falha totalmente diferentes. Portanto, a modelagem da função de taxa de falha é um importante método de análise para dados de tempo de falha, segundo Colosimo e Giolo (2006).

Devido ao fato da existência de diversas distribuições que podem ser associadas ao tempo de falha, é necessário adotar uma metodologia para selecionar um modelo adequado. Os dados amostrais possuem informações qualitativas sobre a função de taxa de falha que podem ser obtidas a partir de uma análise gráfica, como descreve Silva (2008). Assim, o método gráfico proposto por Aarset (1987), denominado gráfico do tempo total em teste (gráfico TTT), é uma metodologia válida de seleção de modelos. O gráfico TTT é obtido plotando os pares $(r/n, G(r/n))$ obtidos a partir da expressão

$$G(r/n) = \frac{(n-r)T_{r:n} + \sum_{i=1}^r T_{i:n}}{\sum_{i=1}^n T_{i:n}}$$

em que $r = 1, \dots, n$ e $T_{i:n}$, $i = 1, \dots, n$ é a estatística de ordem da amostra.

A Figura 2 apresenta as diferentes formas que o gráfico TTT pode assumir.

Quando a curva, obtida do gráfico TTT, é uma reta diagonal (curva A), a função de taxa de falha é constante; quando a curva é convexa (curva B) ou côncava (curva C), a função de taxa de falha é monotonicamente decrescente ou crescente, respectivamente; quando a curva é convexa e então côncava (curva D), a função de taxa de falha tem forma de U, e no caso reverso (curva E) é unimodal.

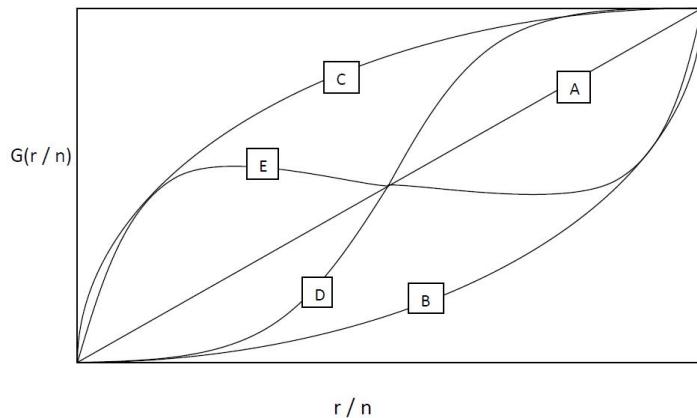


Figura 2 – Gráficos TTT-plot

2.2 Classes de distribuições generalizadas

Diversos autores nos últimos anos tem concentrado seus esforços na generalização de família de distribuições de probabilidade, obtendo, dessa forma, maior flexibilidade e consequentemente, ganho na modelagem de dados e a capacidade de incorporar um grande número de submodelos nas distribuições generalizadas.

Neste trabalho será adotado o termo distribuição base, indicando a família de distribuições a ser generalizada, e será denotado por $G(t)$ a função de distribuição acumulada (fda) da distribuição base, e por $g(t)$ fdp da distribuição base.

Algumas das classes de generalizações mais estudadas nos últimos anos são as famílias de distribuições obtidas pelo método desenvolvido por Marshall e Olkin (1997); as distribuições exponenciadas apresentada inicialmente por Mudholkar et al. (1995); as distribuições estendidas discutidas por Barros (2008); as distribuições betas que receberam maior atenção após o trabalho de Eugene et al. (2002); e recentemente, as distribuições Kumaraswamy, desenvolvida por Cordeiro e Castro (2010). Nas seguintes subseções essas classes

de distribuições serão discutidas e os principais trabalhos apresentados.

2.2.1 Classe de distribuições Marshall e Olkin

Marshall e Olkin (1997) apresentaram um método para obter generalizações de distribuições de probabilidade assumindo um novo parâmetro, $v > 0$, em uma família de distribuições (distribuições bases), dado pela equação

$$S^*(t) = \frac{vS(t)}{1 - (1 - v)S(t)} = \frac{vS(t)}{G(t) + vS(t)}, \quad v > 0 \quad (1)$$

em que $G(t)$ e $S(t)$ é a fda e a função de sobrevivência da distribuição base respectivamente e $S^*(t)$ é a função de sobrevivência da nova família de distribuição, agora com o parâmetro adicional v , obtida da relação (1). Observa-se que se $v = 1$ então $S^*(t) = S(t)$.

Se a distribuição base tem fdp $g(t)$ e função de taxa de falha $h(t)$, então a nova fdp correspondente a $S^*(t)$ é obtida por

$$f^*(t) = \frac{vg(t)}{[G(t) + vS(t)]^2} \quad (2)$$

e a função de taxa de falha por

$$h^*(t) = \frac{h(t)}{1 - (1 - v)S(t)} = \frac{h(t)}{G(t) + vS(t)}.$$

Marshall e Olkin (1997) modificaram a distribuição exponencial baseados na equação (2) e deram o nome a distribuição obtida de distribuição exponencial com dois parâmetros. A distribuição exponencial com dois parâmetros pode servir como concorrente de distribuições conhecidas de dois parâmetros como as distribuições Weibull, gama e lognormal. A mesma modificação é realizada com a distribuição Weibull, a qual é denominada de Weibull com três parâmetros. Marshall e Olkin também desenvolveram as versões bivariadas destas distribuições.

A distribuição Weibull com três parâmetros, obtida por Marshall e Olkin (1997) foi estudada por Zhang e Xie (2007) que destaca que uma das propriedades importantes dessa distribuição é que sua função de taxa de falha modela forma crescente, decrescente ou inici-

almente crescente, e então decrescente, e eventualmente crescente e tem como caso particular a distribuição Weibull para $v = 1$. A caracterização do modelo é estudada, por Zhang e Xie (2007), baseada no gráfico de probabilidade Weibull (WPP). Os autores apresentam um procedimento de estimação de parâmetros com base no WPP e adicionalmente desenvolvem o método de estimação por máxima verossimilhança.

Utilizando o método apresentado por Marshall e Olkin (1997), Thomas e Jose (2004) introduziram a distribuição Marshall-Olkin semi-pareto bivariada (MO-BSP) e Marshall-Olkin pareto bivariada (MO-BP), e estudaram várias características dessas distribuições. O modelo de séries temporais autoregressivo de primeira ordem e de k-ésima ordem foi desenvolvido tendo a distribuição MO-BSP como distribuição marginal estacionária.

Recentemente, Ghitany et al. (2007) investigaram as propriedades da inclusão de um novo parâmetro pelo método de Marshall e Olkin (1997), baseado no modelo lomax, também conhecido como distribuição Pareto de segundo tipo. Os autores mostraram que a distribuição proposta pode ser expressa como uma distribuição de mistura do modelo exponencial. Condições simples e suficientes para o comportamento de forma das funções de densidade e taxa de falha são fornecidas e a estimação de máxima verossimilhança foi utilizada à distribuição proposta para ajustar dados censurados aleatoriamente.

E por fim, Rao et al. (2009) desenvolveram um plano de teste de confiabilidade para não-rejeição ou rejeição de muitos produtos submetidos a inspeção, em que o tempo de vida segue a distribuição exponencial com dois parâmetros de Marshall-Olkin.

2.2.2 Classe de distribuições exponenciadas

A classe de distribuições exponenciadas é obtida elevando-se ao expoente θ ($\theta > 0$) a fda de uma distribuição base, isto é,

$$F(t) = G(t)^\theta , \quad (3)$$

e consequentemente a fdp é dada por $f(t) = \theta g(t)G(t)^{\theta-1}$. A função de sobrevivência e taxa de falha são dados por

$$S(t) = 1 - G(t)^\theta \quad \text{e} \quad h(t) = \frac{\theta g(t)G(t)^{\theta-1}}{1 - G(t)^\theta},$$

respectivamente.

O ganho dessa classe de distribuições é que o novo parâmetro θ flexibiliza a distribuição base.

Dentro dessa classe de distribuições, cita-se a Weibull exponenciada, apresentada por Mudholkar et al. (1995), que é a generalização da distribuição Weibull com parâmetros $\alpha > 0$ e $\gamma > 0$. A distribuição Weibull exponenciada apresenta diferentes formas da função de taxa de falha, como por exemplo, formas crescente, decrescente, de banheira e unimodal, obtidas de quatro regiões do espaço paramétrico determinados pelas equações $\alpha = 1$ e $\theta\alpha = 1$. A distribuição Weibull exponenciada é indicada para modelar dados onde o risco é não-monótono e pode ser utilizada para testar a qualidade do ajuste da distribuição Weibull como submodelo. Cinco conjuntos de dados sobre falha de motor de ônibus, onde Mudholkar et al. (1995) consideraram como falha a quebra abrupta do motor ou quando a máxima potência produzida por este foi menor do que a porcentagem considerada normal para este tipo de motor; e dados de experimento clínico de câncer de cabeça e pescoço são analisados pelos autores, utilizando a distribuição Weibull exponenciada.

Gupta e Kundu (1999) propuseram a distribuição exponencial generalizada (que é a modificação da distribuição exponencial com parâmetros de escala $\lambda > 0$ e locação $-\infty < \mu < \infty$, cuja fda é dada por $G(t) = 1 - e^{-(t-\mu)/\lambda}$). Os autores também estudaram diferentes propriedades desta distribuição, que é um caso particular da distribuição Weibull exponenciada proposta por Mudholkar et al. (1995), quando o parâmetro de locação é igual a zero. Gupta e Kundu observam que esta família possui algumas características interessantes que são similares àquelas das distribuições gama e Weibull. Dessa forma, este modelo pode ser usado como alternativa aos modelos gama e Weibull em diversas situações. Um conjunto de dados de testes sobre a resistência de rolamentos de esfera, citado em Lawless (1982), foi analisado onde o ajuste da distribuição exponencial generalizada foi melhor que o das

distribuições Weibull e gama com três parâmetros.

Algumas das propriedades da distribuição exponencial exponenciada discutida por Gupta et al. (1998) foram estudadas por Gupta e Kundu (2001). A distribuição exponencial exponenciada tem dois parâmetros (escala e forma) similar à distribuição Weibull e gama. Gupta e Kundu (2001) observaram que muitas propriedades desta nova família são similares àquelas das distribuições Weibull e gama, assim como a distribuição exponencial generalizada discutida por Gupta e Kundu (1999). Portanto, esta distribuição também pode ser usada como uma possível alternativa às distribuições Weibull e gama. Dois conjuntos de dados reais de tempo de falha são ajustados ao modelo proposto, o primeiro conjunto de dados refere-se a resistência de rolamentos de esfera, citado em Lawless (1982), e o segundo conjunto refere-se ao tempo de falha de sistema de condicionamento de ar de um avião, citado em Linhart e Zucchini (1986), e comparado o ajuste do modelo com os modelos Weibull e gama. O modelo exponencial exponenciado teve melhor ajuste em um dos conjunto de dados e o modelo Weibull ajustou-se melhor no outro conjunto de dados.

Por outro lado, Nadarajah e Kotz (2006), discutiram algumas distribuições exponenciadas e estendidas (que será tratada na próxima seção) e introduziram a distribuição gama padrão exponenciada, que é a generalização da distribuição gama padrão, cuja fda é $G(t) = \gamma(\alpha, t)/\Gamma(\alpha)$ onde $\gamma(\alpha, t) = \int_0^t t^{\alpha-1} e^{-t} dt$ é a função gama incompleta, elevando-se $G(t)$ a potência θ segundo a equação (3) e desenvolveram matematicamente algumas propriedades estatística de cada distribuição. A fdp e fda foram derivadas na forma de expansão em série para a distribuição gama padrão exponenciada e o comportamento da fdp e da função de taxa de falha, em função dos parâmetros foi discutido. Nadarajah e Kotz (2006) mostraram que a distribuição gama padrão exponenciada modela formas decrescente, unimodal e constante e apresentaram a função geradora de momentos e os momentos dessa distribuição.

Recentemente, Carrasco et al. (2008) definiram e estudaram a distribuição Weibull modificada generalizada que modela formas da função de taxa de falha monótonas e não monótonas, que são comuns em problemas de tempo de vida e tem como sub-modelos as distribuições Weibull, valor extremo, Weibull exponenciada, Rayleigh generalizada e Weibull modificada, entre outras. Os autores também apresentaram duas representações dos momentos em série infinita e a fdp das estatísticas de ordem foi obtida. A estimativa dos parâmetros

foi obtida pelo método de máxima verossimilhança e a matriz de informação observada foi apresentada.

Finalmente, Alkasasbeh e Raqab (2009) estudaram a distribuição logística generalizada, definida por Balakrishnan e Leng (1988) e consideraram a estimativa de máxima verossimilhança dos parâmetros, assim como outros cinco procedimentos de estimativa e compararam a performance destes procedimentos através de extensiva simulação numérica.

2.2.3 Classe de distribuições estendidas

A classe de distribuições estendidas apresentada por Barros (2008) é uma variação das distribuições exponenciadas onde se eleva ao expoente ω , ($\omega > 0$) a função de sobrevivência ($1 - G(t)$), ao invés da fda, ou seja

$$F(t) = 1 - [1 - G(t)]^\omega, \quad (4)$$

e portanto $f(t) = \omega g(t)[1 - G(t)]^{\omega-1}$. As funções de sobrevivência e taxa de falha são dadas respectivamente pelas equações

$$S(t) = [1 - G(t)]^\omega \quad \text{e} \quad h(t) = \frac{\omega g(t)}{1 - G(t)}.$$

A característica principal dessa classe é a maior flexibilidade, devido ao parâmetro adicional ω , e por possuir forma fechada para todas as funções de probabilidade. Porém, observa-se que o parâmetro ω atua apenas de forma multiplicativa na função de taxa de falha, e portanto, não modifica a forma básica da função de taxa de falha da distribuição base.

Barros (2008) estudou a classe de distribuições estendidas, e demonstrou a consistência dessa nova generalização. A fdp, fda e os momentos de várias distribuições estendidas são descritas por Barros (2008), porém a mesma não desenvolveu nenhum tratamento mais detalhado das distribuições apresentadas e não fez nenhuma aplicação mostrando a eficiência dessas novas distribuições.

As distribuições discutidas por Barros (2008) foram as distribuições exponencial estendida, uniforme estendida, Weibull estendida, Pareto estendida, logística padrão

estendida, qui-quadrado estendida, gama estendida, Fréchet estendida e Gumbel estendida.

Entretanto observa-se que algumas distribuições modificadas de acordo com a equação (4) não geram distribuições novas. Como é o caso da distribuição exponencial estendida, obtida pela substituição da fda da distribuição exponencial com parâmetro α na equação (4), e obtendo-se novamente a distribuição exponencial com parâmetro $\alpha\omega$, cuja fda passa a ser dada por $F(t) = 1 - e^{-(\alpha\omega)t}$, $\alpha, \omega > 0$.

A distribuição uniforme estendida é a generalização da distribuição uniforme no intervalo $(0, 1)$, pela equação (4), e resulta na distribuição beta, com parâmetros $\alpha = 1$ e ω , onde sua fda é dada por $G(t) = 1 - (1 - t)^\omega$.

A distribuição Weibull, com parâmetros $\alpha > 0$ e $\lambda > 0$, gera a distribuição Weibull estendida que resumi-se na distribuição Weibull, com pâmetros $\alpha\omega^{-1/\lambda}$ e λ , cuja fda é $F(t) = 1 - \exp[-\omega(t/\alpha)^\lambda]$.

A distribuição Pareto, com parâmetros $k > 0$ e $\alpha > 0$, gera a distribuição Pareto com parâmetros k e $\alpha\omega$ sobre a generalização das distribuições estendidas, cuja fda é $F(t) = 1 - (k/t)^{\alpha\omega}$. A distribuição logística padrão estendida é a generalização da distribuição logística, com parâmetros $\mu = 0$ e $\sigma = 1$, e resulta na distribuição logística generalizada (LG) tipo II de parâmetros ω e 1, com distribuição acumulada dada por $F(y) = 1 - [e^{-\omega y}/(1 + e^{-y})^\omega]$, $y \in \mathbb{R}$.

Diferentemente das distribuições descritas acima as distribuições estendidas qui-quadrado, gama, Fréchet e Gumbel, discutida por Barros (2008), geraram novas distribuições com o parâmetro adicional ω , sendo que as distribuições Fréchet estendida e Gumbel estendida foram apresentadas por Nadarajah e Kotz (2006) com o nome de distribuições Fréchet exponenciada e Gumbel exponenciada.

Nesse mesmo enfoque, Nadarajah e Kotz (2006) introduziram duas novas distribuições. As distribuição Gumbel exponenciada e Fréchet exponenciada, porém essas distribuições apresentam a mesma estrutura das distribuições estendidas. A distribuição Fréchet exponenciada (Fréchet estendida) é a generalização da Fréchet padrão pela classe de distribuições estendidas, Nadarajah e Kotz (2006) apresentaram essa distribuição e discutiram a forma da fdp e da função de taxa de falha. Além disso, representaram a fdp como forma de mistura, desenvolveram a expressão para os momentos e destacaram a importância dessa

distribuição em aplicações abrangendo testes de vida acelerados tais como em dados de terremoto, inundação, corridas de cavalo, precipitação, fila em supermercado, velocidade do vento, entre outros. A distribuição Gumbel exponenciada (Gumbel estendida) é a generalização da Gumbel padrão pela classe de distribuições estendidas. Nadarajah (2006) apresentou um tratamento detalhado das propriedades dessa distribuição, apresentando as formas analíticas da fdp e função de taxa de falha. O autor também calculou o n -ésimo momento e a distribuição assintótica das estatísticas de ordem, investigou a variação das medidas de assimetria e curtose. E por fim, discutiu a estimativa pelo método de máxima verossimilhança e ilustrou sua aplicação com dados de precipitação pluvial em Orlando, Florida, Estados Unidos.

Por ultimo, Barros (2008) desenvolveu os momentos, a função geradora de momentos e a função característica para a distribuição qui-quadrada estendida. A distribuição gama padrão estendida foi introduzida, e apresentada a expressão para a fdp e fda na forma fechada e na forma de expansão em série. Os momentos também foram calculados, assim como outras propriedades.

2.2.4 Classe de distribuições betas

A classe de distribuições betas, que recebeu maior atenção após o trabalho de Eugene et al. (2002), é uma generalização baseada na distribuição Beta, que considera $G(t)$ para definir a fda da distribuição beta generalizada, a qual é expressa como sendo

$$F(t) = \frac{1}{B(a, b)} \int_0^{G(t)} w^{a-1} (1-w)^{b-1} dw , \quad (5)$$

em que $a > 0$ e $b > 0$ são os dois parâmetros novos de forma e $B(a, b)$ é a função beta, definida por $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$. A fdp é dada por

$$f(t) = \frac{1}{B(a, b)} g(t) G(t)^{a-1} (1-G(t))^{b-1} , \quad t > 0 .$$

Observe que a integral na equação (5) é a função beta incompleta, $B_{G(t)}(a, b) = \int_0^{G(t)} w^{a-1} (1-w)^{b-1} dw$. A desvantagem dessa classe é que essa integral muitas vezes não tem solução analítica, devido a fda $G(t)$, ou depende de funções especiais, como a função

$B(a, b)$, necessitando portanto, de abordagens numéricas. Consequentemente, a função de sobrevivência também requer procedimentos numéricos para ser obtida.

Alguns autores trabalharam com essa classe de distribuição, entre eles pode-se citar Ojo e Olapade (2003), que consideraram a distribuição beta logística, denominada por eles de logística generalizada (distribuição esta, desenvolvida inicialmente por George e Ojo (1980)) e provaram alguns teoremas que caracterizam essa distribuição, além de definir uma generalização da distribuição log-logística (beta log-logística), determinar seus momentos e estabelecer algumas relações entre a log-logística generalizada e outras distribuições.

Nadarajah e Kotz (2004) introduziram a distribuição beta Gumbel, apresentaram a forma analítica da fdp e função de taxa de falha, calcularam a expressão para o n -ésimo momento e a distribuição assintótica da estatística de ordem, investigaram a variação das medidas de assimetria e curtose e discutiram o método de estimação de máxima verossimilhança.

Da mesma forma, Nadarajah e Kotz (2005) generalizaram a distribuição exponencial, denominando-a distribuição beta exponencial. As propriedades matemáticas da distribuição beta exponencial foram discutidas e as expressões para a função geradora de momentos e função característica foram derivadas. Os primeiros momentos, a variância, a assimetria e curtose foram apresentados. Também desenvolveram as expressões para o desvio médio sobre a média e sobre a mediana, a entropia de Rényi e Shannon e a distribuição das estatísticas ordem. A estimativa pelo método de máxima verossimilhança e o problema de simulação foram discutidos e uma expressão para a matriz de informação esperada foi apresentada.

Recentemente, Barreto-Souza et al. (2009) propuseram a distribuição beta exponencial generalizada exponenciada, que é a generalização da distribuição exponencial generalizada exponenciada proposta por Gupta e Kundu (1999). Essa distribuição tem como casos particulares as distribuições beta exponencial e exponencial generalizada. A fdp como mistura de densidades da distribuição exponencial generalizada foi expressa, a função geradora de momentos e os momentos foram derivados e a estimativa de parâmetros pelo método de máxima verossimilhança foi discutida e a matriz de informação esperada obtida.

Na mesma abordagem, Pescim et al. (2010) propuseram a chamada distribuição

beta half-normal generalizada, que é a generalização da distribuição half-normal generalizada apresentada por Cooray e Ananda (2008), que contém como casos especiais a half-normal e a half-normal generalizada. As expressões para a fdp e para fda foram derivadas e os momentos foram obtidos, assim como o método de estimação de máxima verossimilhança foi utilizado. A matriz de informação esperada foi calculada, bem como outras propriedades de interesse. A utilidade dessa nova distribuição foi ilustrada através de um conjunto de dados reais mostrando a maior flexibilidade na análise de dados positivos em relação às distribuições half-Normal generalizada, half-Normal, Weibull e beta Weibull.

2.2.5 Classe de distribuições Kumaraswamy

A classe de distribuições Kumaraswamy, recentemente proposta por Cordeiro e Castro (2010), baseia-se na generalização da distribuição Kumaraswamy proposta por Kumaraswamy (1980) para variáveis limitadas inferiormente e superiormente. A fda da distribuição Kumaraswamy é dada por

$$F(x) = 1 - (1 - x^a)^b, \quad 0 < x < 1 \quad \text{e} \quad a, b > 0. \quad (6)$$

Substituindo a variável aleatória x por $G(t)$, na equação (6), obtém-se a classe de distribuições Kumaraswamy, cuja fda é dada por:

$$F(t) = 1 - \{1 - G(t)^a\}^b, \quad 0 < G(t) < 1 \quad \text{e} \quad a, b > 0. \quad (7)$$

em que $a > 0$ e $b > 0$ são os novos parâmetros de forma.

As fdp, função de sobrevivência e função de taxa de falha são dadas por

$$f(t) = abg(t)G(t)^{a-1}[1 - G(t)^a]^{b-1}, \quad (8)$$

$$S(t) = [1 - G(t)^a]^b \quad \text{e} \quad h(t) = \frac{abg(t)G(t)^{a-1}}{1 - G(t)^a},$$

respectivamente.

Os momentos podem ser obtidos pelo cálculo da integral

$$\mu'_k = \int_0^\infty ab[1 - G(t)^a]^{b-1}G(t)^{a-1}g(t)t^k dt \quad (9)$$

$$= b \int_0^1 \{G^{-1}(u^{1/a})\}^k (1-u)^{b-1} du \quad (10)$$

sendo que a transformação usada na integral descrita acima foi $u = G(t)^a$; o que implica em $G^{-1}(u^{1/a}) = t$.

A grande vantagem da classe de distribuições Kumaraswamy é que esta possui forma fechada, ou seja, pode ser escrita de forma analítica, diferentemente da classe de distribuições betas. Essa classe modela formas de risco constante, crescente, decrescente, unimodal e em forma de U. Outra propriedade da classe de distribuição Kumaraswamy é que esta possui como caso particular a classe de distribuições exponenciadas quando $b = 1$, a classe de distribuições estendidas quando $a = 1$ e as distribuições bases $G(t)$ quando $a = 1$ e $b = 1$.

Pode-se citar alguns importantes trabalhos nesta área, como por exemplo, o trabalho de Cordeiro et al. (2010), que propuseram as distribuições Kw-gama, Kw-gumbel, Kw-gaussiana inversa, Kw-normal e Kw-Weibull; e o trabalho de Pascoa et al. (2010) que apresentaram a distribuição Kumaraswamy-gama-generalizada. Tendo como base esses trabalhos, propõem-se nesse trabalho a nova distribuição Kumaraswamy-log-logística e Kumaraswamy-logística, que nos próximos capítulos serão apresentadas e discutidas.

3 DISTRIBUIÇÕES LOG-LOGÍSTICA MODIFICADAS

Nesse capítulo são apresentadas duas formas de generalizar a distribuição log-logística. A primeira forma é baseada nos trabalhos de Marshall e Olkin (1997), e a segunda tem por base a classe de distribuições Kumaraswamy introduzida por Cordeiro et al. (2010). Na seção 3.1 é feita uma revisão sobre a distribuição log-logística, sendo apresentadas as principais características e propriedades dessa distribuição, e também é feita uma revisão de trabalhos em várias áreas aplicadas e teórica, utilizando a distribuição log-logística nos últimos anos, evidenciando sua importância nos diversos campos do conhecimento. Na seção 3.2, a nova distribuição log-logística modificada é apresentada. A distribuição Kumaraswamy-log-logística é proposta na seção 3.3.

3.1 Distribuição log-logística

A distribuição log-logística, com parâmetros α e γ , considera que o tempo de falha T tem fdp dado por

$$g(t) = \frac{\gamma}{\alpha} \left(\frac{t}{\alpha} \right)^{\gamma-1} \left[1 + \left(\frac{t}{\alpha} \right)^\gamma \right]^{-2}, \quad t > 0$$

em que $\alpha > 0$ é o parâmetro de escala e $\gamma > 0$ é o parâmetro de forma.

A variável aleatória T , segundo a distribuição log-logística será denotada neste trabalho por $T \sim LL(\alpha, \gamma)$. A fda para a variável aleatória T tem a seguinte forma

$$G(t) = \left(\frac{t}{\alpha} \right)^\gamma \left[1 + \left(\frac{t}{\alpha} \right)^\gamma \right]^{-1}.$$

Consequentemente, a função de sobrevivência é dada por

$$S(t) = \left[1 + \left(\frac{t}{\alpha} \right)^\gamma \right]^{-1}.$$

A função de taxa de falha é dada por

$$h(t) = \frac{\gamma}{\alpha} \left(\frac{t}{\alpha} \right)^{\gamma-1} \left[1 + \left(\frac{t}{\alpha} \right)^\gamma \right]^{-1}.$$

A distribuição LL é caracterizada por modelar funções de taxa de falha decrescentes e unimodal. Possui forma fechada, ou seja, todos os objetos relacionados a distribuição LL, como por exemplo a fda, a fdp, os momentos, a assimetria e a curtose, podem ser obtidos de forma analítica, o que é de grande importância em aplicações práticas, principalmente na área de hidrologia, segundo Ashkar e Mahdi (2006).

A distribuição LL, em muitas situações práticas, tem-se apresentado como alternativa às distribuições de Weibull e a log-normal, conforme discuti Colosimo e Giolo (2006).

E em economia, onde é conhecida como distribuição Fisk, a distribuição LL modela a distribuição de riqueza e renda e na área de hidrologia é utilizada para modelar escoamento fluvial.

A Figura 3a, 3b e 3c mostra os gráficos das funções fdp, de sobrevivência e taxa de falha para $\alpha = 25$ e para alguns valores de γ , respectivamente.

A Figura 3c apresenta as formas unimodal e decrescente da função de taxa de falha, para $\gamma = 4$, $\gamma = 0,67$ e $\gamma = 1$ respectivamente. Note que a função de taxa de falha tem forma unimodal para $\gamma > 1$, decrescente para $\gamma < 1$ e decrescente iniciando em $1/\alpha$ para $\gamma = 1$.

O r -ésimo momento para o tempo de falha é dado por

$$E(T^r) = \alpha^r \frac{r\pi/\gamma}{\operatorname{sen}(r\pi/\gamma)}, \quad r < \gamma .$$

Dessa forma, a esperança e a variância do tempo de falha é dado pelas seguintes expressões

$$E(T) = \alpha \frac{\pi/\gamma}{\operatorname{sen}(\pi/\gamma)}, \quad \gamma > 1 \quad \text{e} \quad \operatorname{Var}(T) = \alpha^2 \frac{2\pi/\gamma}{\operatorname{sen}(2\pi/\gamma)} - \left[\alpha \frac{\pi/\gamma}{\operatorname{sen}(\pi/\gamma)} \right]^2, \quad \gamma > 2 .$$

A função quantílica é dada por

$$G^{-1}(p) = \alpha \left(\frac{p}{1-p} \right)^{1/\gamma}, \quad 0 < p < 1$$

em que $G^{-1}(p)$ é a inversa da fda.

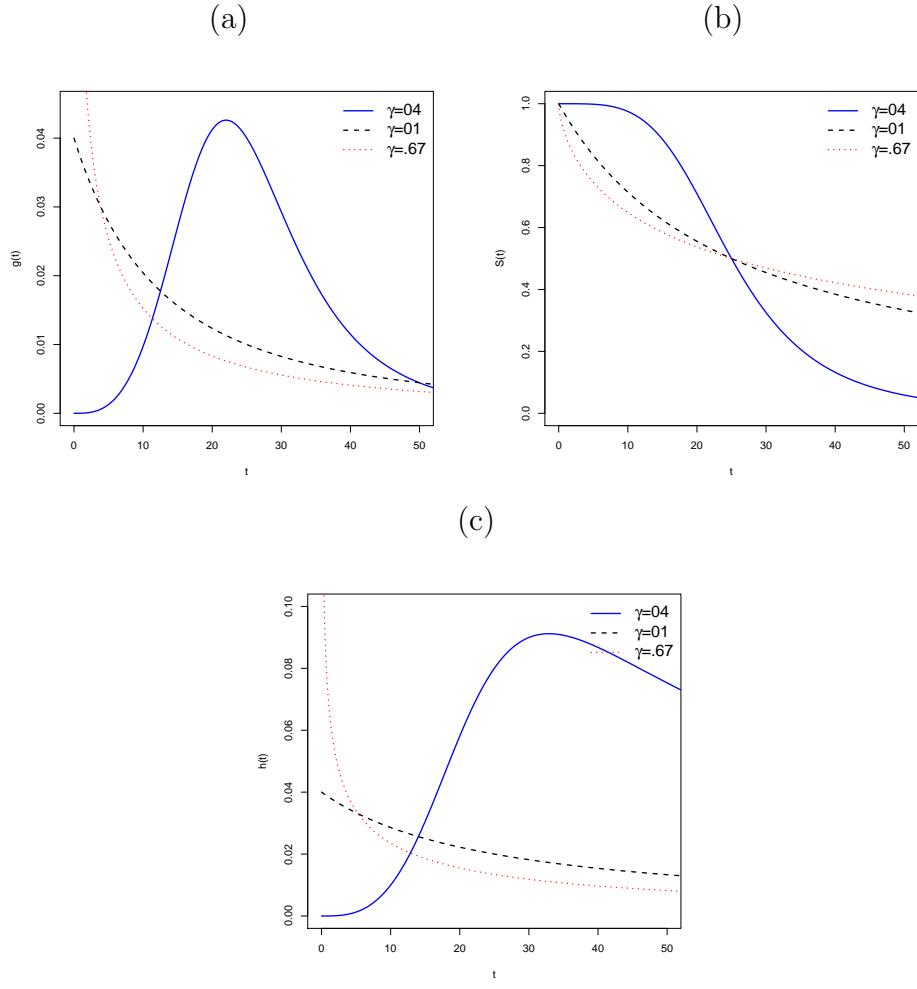


Figura 3 – Gráficos da fdp, função de sobrevivência e função de taxa de falha da distribuição log-logística para alguns valores γ e $\alpha = 25$

A distribuição log-logística possui a mesma relação que a distribuição log-normal tem com a normal, ou seja, se T segue distribuição log-logística, então $Y = \log(T)$ segue distribuição logística.

Nos últimos anos diversos trabalhos foram publicados utilizando a distribuição log-logística ou a distribuição logística, o que demonstra a importância dessas distribuições em diferentes áreas do conhecimento. A seguir é apresentado alguns desses trabalhos.

Ojo e Olapade (2003) definiram uma generalização da distribuição log-logística (beta log-logística), determinaram seus momentos e estabeleceram algumas relações entre a log-logística generalizada e outras distribuições.

Nesse mesmo ano, Leroy et al. (2003) investigaram o tempo de aparecimento de

dentes permanentes de 4468 crianças em Flanders, Bélgica, por meio da análise de sobrevivência, considerando a distribuição log-logística para o cálculo da mediana, média e intervalo de confiança. Indicando o aparecimento de dentes permanentes significativamente mais cedo em meninas do que em meninos.

Por outro lado, Ashkar e Mahdi (2006) investigaram o método dos momentos generalizados para estimação dos parâmetros do modelo log-logístico com dois parâmetros. Os autores calcularam a matriz de covariância assintótica para os estimadores dos parâmetros e apresentaram resultados de simulação sobre a performance do método dos momentos generalizados comparado aos métodos dos momentos ponderados por probabilidade generalizados, de máxima verossimilhança e dos momentos clássicos. Ashkar e Mahdi, ainda apresentaram uma aplicação na área de hidrologia, em estudos envolvendo baixo fluxo fluvial.

Recentemente, Kirshenbaum et al. (2009) estudaram a recaída de dependentes químicos, após um período inicial de abstinência. O objetivo principal do estudo foi verificar se a reacaída de dependentes químicos exibi um padrão típico que pode ser caracterizado por uma função quantitativa comum. Os autores realizaram análises (utilizando as técnicas da análise de sobrevivência) em 20 estudos sobre indivíduos que permaneceram abstinentes depois de um período de abstinência inicial, onde observou-se que a distribuição log-logística foi a mais adequada para modelar os dados com maior precisão.

3.2 Distribuição log-logística modificada

Nesta seção foi usado o método descrito na seção 2.2.1, do capítulo anterior, para introduzir um novo parâmetro na distribuição de probabilidade log-logística e, portanto, obter a distribuição log-logística modificada.

Seja v o novo parâmetro e $S(t)$ a função de sobrevivência da distribuição LL(α, γ). Ao Substituir $S(t)$ na equação (1), obtém-se a distribuição log-logística modificada expressa por

$$S^*(t) = \frac{v}{v + (t/\alpha)^\gamma}, \quad t > 0.$$

A partir da função de sobrevivência pode-se construir a fda e a fdp, que são dadas por

$$F(t) = 1 - \frac{v}{v + (\frac{t}{\alpha})^\gamma} \quad \text{e} \quad f(t) = \frac{v\gamma}{\alpha^\gamma} \frac{t^{\gamma-1}}{\left[v + (\frac{t}{\alpha})^\gamma\right]^2},$$

respectivamente. Observe que integrando $f(t)$ em relação a t obtém-se

$$\int_0^\infty \frac{v\gamma}{\alpha^\gamma} \frac{t^{\gamma-1}}{\left[v + (\frac{t}{\alpha})^\gamma\right]^2} dt = \int_v^\infty \frac{v}{u^2} du = \left[-v \frac{1}{u}\right]_v^\infty = 1,$$

logo, $f(t)$ é fdp da distribuição log-logística modificada, com parâmetros v , α e γ que será denotada por $LLM(v, \alpha, \gamma)$. Observe que se $v = 1$, então tem-se a distribuição $LL(\alpha, \gamma)$. A Figura 4 apresenta os gráficos da fdp da distribuição LLM para alguns valores de α e γ fixados e variando os valores do parâmetro v . A curva contínua representa a densidade da distribuição LL ($v = 1$), note que à medida que v aumenta, a curva vai se achatando, como pode ser observado pelas curvas segmentadas.

Se T segue uma distribuição $LLM(v, \alpha, \gamma)$ ($T \sim LLM(v, \alpha, \gamma)$), pela relação $h(t) = f(t)/S(t)$ obtém-se a função de taxa de falha da distribuição log-logística modificada, dada por

$$h(t) = \frac{\gamma}{\alpha} \frac{(\frac{t}{\alpha})^{\gamma-1}}{v + (\frac{t}{\alpha})^\gamma}.$$

Os gráficos da função de taxa de falha são apresentados na Figura 5 para alguns valores de v , α e γ . Observe que o parâmetro v torna a distribuição mais flexível em relação à distribuição LL , e à medida que v aumenta, mais achatada se torna a curva da função de taxa de falha.

Os momentos são obtidos pelo teorema a seguir.

Teorema 3.1: Se $T \sim LLM(v, \alpha, \gamma)$, então o k -ésimo momento é dado por

$$\mu'_k = v^{k/\gamma} \frac{\alpha^k k \pi / \gamma}{\operatorname{sen}(k \pi / \gamma)}, \quad r < \gamma.$$

Prova: Os momentos da distribuição $LLM(v, \alpha, \gamma)$ são obtidos apartir de

$$\mu'_k = \int_0^\infty t^k f(t) dt = \int_0^\infty t^k \frac{v\gamma}{\alpha^\gamma} \left[v + \left(\frac{t}{\alpha}\right)^\gamma\right]^{-2} dt.$$

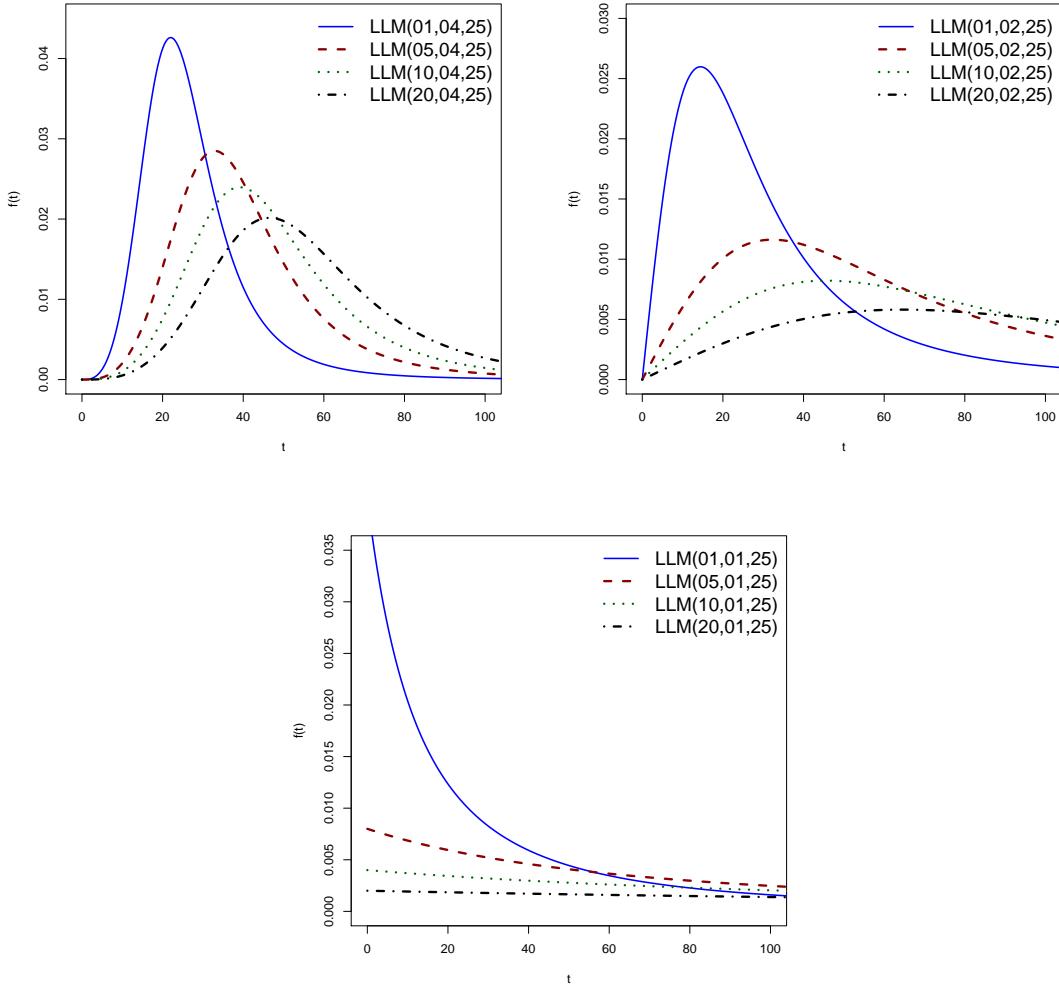


Figura 4 – Gráficos da fdp da distribuição LLM para $\gamma = 25$, α assumindo os valores 1, 2 e 4 e v assumindo os valores 1, 5, 10 e 20

Substituindo-se $w = \left[v + \left(\frac{t}{\alpha}\right)^\gamma\right]^{-1}$, obtém-se

$$\mu'_k = v\alpha^k \int_0^{1/v} w^{-k/\gamma} (1-wv)^{k/\gamma} dw ,$$

reparametrizando novamente $\theta = wv$ obtém-se

$$\mu'_k = v^{k/\gamma} \alpha^k \int_0^1 \theta^{-k/\gamma} (1-\theta)^{k/\gamma} d\theta = v^{k/\gamma} \alpha^k B(1+k/\gamma, 1-k/\gamma) ,$$

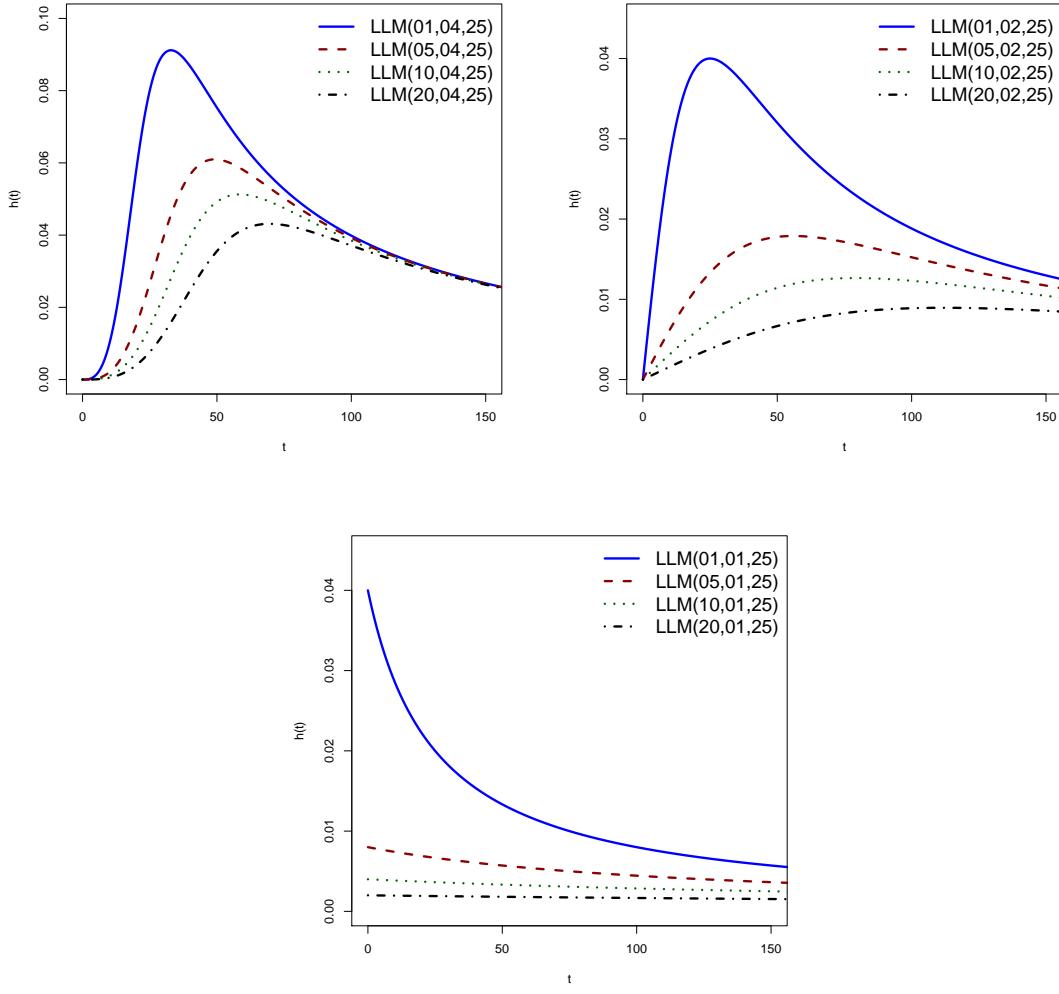


Figura 5 – Gráficos da função de taxa de falha da distribuição LLM para $\gamma = 25$, α assumindo os valores 1, 2 e 4 e v assumindo os valores 1, 5, 10 e 20

em que $B(\cdot, \cdot)$ é a função beta. Da relação, descrita em Lawless (1982),

$$B(1 + k/\gamma, 1 - k/\gamma) = \frac{r\pi/\gamma}{\sin(r\pi/\gamma)},$$

tem-se que os momentos da distribuição log-logística modificada são obtidos da relação

$$\mu'_k = v^{k/\gamma} \frac{\alpha^k k \pi / \gamma}{\sin(k\pi/\gamma)}.$$

□

Do teorema acima, segue que a média e a variância são dadas por

$$E(T) = v^{1/\gamma} \frac{\alpha\pi/\gamma}{\operatorname{sen}(\pi/\gamma)}, \quad \gamma > 1 \quad \text{e} \quad \operatorname{Var}(T) = v^{2/\gamma} \left[\frac{\alpha^2 2\pi/\gamma}{\operatorname{sen}(2\pi/\gamma)} - \frac{\alpha^2 (\pi/\gamma)^2}{\operatorname{sen}^2(\pi/\gamma)} \right] \quad \gamma > 2.$$

A variância, assimetria e curtose podem ser obtidas por meio das relações descritas em (11), segundo Nadarajah e Kotz (2004).

$$\begin{aligned} \operatorname{Var}(T) &= E(T^2) - E^2(T) \\ A(T) &= \frac{E(T^3) - 3E(T)E(T^2) + 2E^3(T)}{\operatorname{Var}^{3/2}(T)} \\ K(T) &= \frac{E(T^4) - 4E(T)E(T^3) + 6E(T^2)E^2(T) - 3E^4(T)}{\operatorname{Var}^2(T)} \end{aligned} \quad (11)$$

logo a assimetria, $A(t)$ e curtose, $K(t)$, são dadas respectivamente por

$$\begin{aligned} A(T) &= \left\{ \frac{3\pi/\gamma}{\operatorname{sen}(3\pi/\gamma)} - \frac{6\pi/\gamma}{\operatorname{sen}(\pi/\gamma)\operatorname{sen}(2\pi/\gamma)} \right\} \left\{ \frac{2\pi/\gamma}{\operatorname{sen}(2\pi/\gamma)} - \frac{(\pi/\gamma)^2}{\operatorname{sen}^2(\pi/\gamma)} \right\}^{-\frac{3}{2}} \\ K(T) &= \left\{ \frac{12(\pi/\gamma)^3}{\operatorname{sen}(2\pi/\gamma)\operatorname{sen}^2(\pi/\gamma)} - \frac{12\pi/\gamma}{\operatorname{sen}(4\pi/\gamma)} \right\} \left\{ \frac{2\pi/\gamma}{\operatorname{sen}(2\pi/\gamma)} - \frac{(\pi/\gamma)^2}{\operatorname{sen}^2(\pi/\gamma)} \right\}^{-2} \end{aligned}$$

Observa-se que tanto a assimetria quanto a curtose não dependem do novo parâmetro v , somente de γ . Dessa forma, o novo parâmetro v introduzido na distribuição log-logística pelo método apresentado por Marshall e Olkin (1997) não influencia no cálculo da assimetria e curtose, que são estatísticas que medem o grau de simetria e achatamento da distribuição.

Assim como as distribuições log-logística e logística estão relacionadas através da relação $Y = \log(T)$, a distribuição LLM está relacionada com a logística modificada, através da relação $Y = \log(T)$. Isto é, se a variável aleatória T tem distribuição LLM(v, α, γ), então a variável $Y = \log(T)$ tem distribuição logística modificada.

Pelo método do Jacobiano tem-se que se $T \sim \text{LLM}(v, \alpha, \gamma)$ e $Y = \log(T)$ então a fdp da variável aleatória Y é dada pela expressão

$$f(y) = \frac{v}{\alpha} \frac{\exp\left(\frac{y-\mu}{\sigma}\right)}{\left[v + \exp\left(\frac{y-\mu}{\sigma}\right)\right]^2} \quad (12)$$

com reparametrização $\gamma = 1/\sigma$ e $\alpha = \exp(\mu)$, em que $\sigma > 0$ e $-\infty < \mu < \infty$.

3.3 Distribuição Kumaraswamy-log-logística

Outra forma de generalizar a distribuição log-logística é obtida utilizando a classe de distribuições Kumaraswamy proposta por Cordeiro e Castro (2010), apresentada na seção 2.2.5, onde substituindo, na equação (7), $G(t)$ pela fda da distribuição log-logística obtém-se, então, a distribuição Kumaraswamy-log-logística com quatro parâmetros, denotada por Kw-LL(a, b, α, γ), cujas funções fda e fdp são definidas pelas equações (13) e (14), respectivamente.

$$F(t) = 1 - \left\{ 1 - \left[1 - \frac{1}{1 + (t/\alpha)^\gamma} \right]^a \right\}^b, \quad (13)$$

$$f(t) = ab\frac{\gamma}{\alpha} \left(\frac{t}{\alpha} \right)^{a\gamma-1} \left[1 + \left(\frac{t}{\alpha} \right)^\gamma \right]^{-(a+1)} \left\{ 1 - \left[1 - \frac{1}{1 + (t/\alpha)^\gamma} \right]^a \right\}^{b-1}. \quad (14)$$

As Figuras 6 e 7 apresentam os gráficos da fdp da distribuição Kw-LL, para alguns valores de a , b e γ . Na Figura 6 (a), (b) e (c) observa-se o gráfico da densidade, sem levar em consideração a influência de a , ou seja, fazendo $a = 1$. Na Figura 7 (a), (b) e (c) observa-se o gráfico da fdp considerando $b = 1$ e variando o parâmetro a . Nas Figura 6(a) e Figura 7(a) são apresentados as formas gráficas da fdp para $\gamma < 1$. Da mesma forma, nas Figuras 6(b) e 7(b) e Figuras 6(c) e 7(c) apresenta-se as formas gráficas da fdp para $\gamma = 1$ e $\gamma > 1$, respectivamente.

A função de sobrevivência, para a variável aleatória T , com distribuição de probabilidade Kw-LL(a, b, α, γ), é dada por

$$S(t) = \left\{ 1 - \left[1 - \frac{1}{1 + (t/\alpha)^\gamma} \right]^a \right\}^b. \quad (15)$$

Os gráficos da fdp e da função de sobrevivência da distribuição Kw-LL são apresentados na Figura 8, para alguns valores dos parâmetros, onde observa-se a maior flexibilidade da fdp e da função de sobrevivência, devido aos parâmetros a e b .

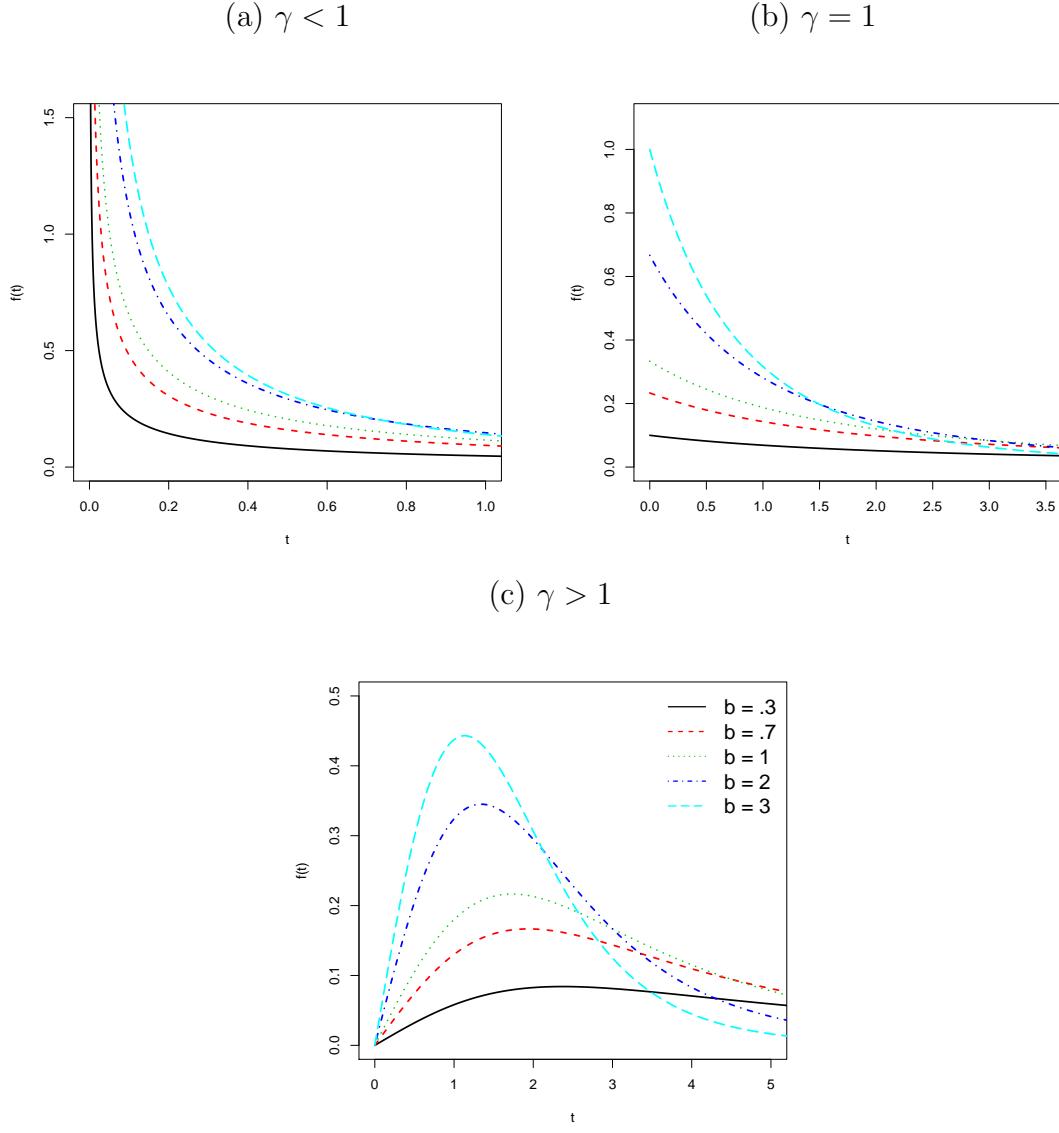


Figura 6 – Gráficos da fdp da distribuição Kw-LL. Em (a), (b) e (c) são apresentados os gráficos da fdp fixando $a = 1$, para $\gamma < 1$, $\gamma = 1$ e $\gamma > 1$ respectivamente

A função de taxa de falha é dada por

$$h(t) = ab \frac{\gamma}{\alpha} \left(\frac{t}{\alpha}\right)^{a\gamma-1} \left[1 + \left(\frac{t}{\alpha}\right)^\gamma\right]^{-(a+1)} \left\{1 - \left[1 - \frac{1}{1 + (t/\alpha)^\gamma}\right]^a\right\}^{-1}. \quad (16)$$

A característica da distribuição Kw-LL é que sua função de taxa de falha acomoda as formas crescente, decrescente, unimodal e forma de banheira, que dependem basicamente dos valores dos parâmetros a e γ .

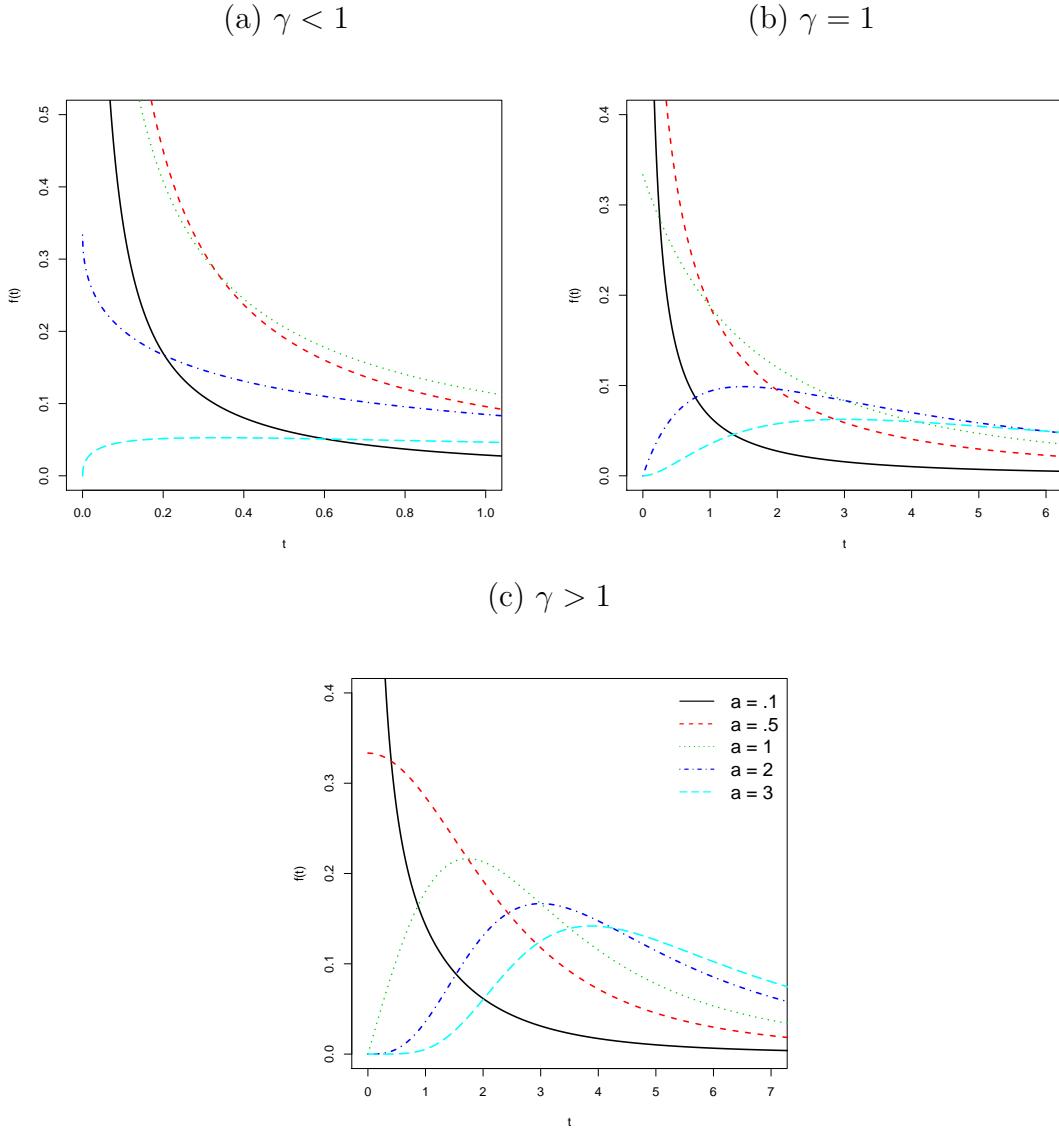


Figura 7 – Gráficos da fdp da distribuição Kw-LL. Em (a), (b) e (c) são apresentados os gráficos da fdp fixando $b = 1$, para $\gamma < 1$, $\gamma = 1$ e $\gamma > 1$ respectivamente

Para estimar as possíveis formas que possa assumir $h(t)$, pode-se fazer o estudo da derivada primeira, cuja expressão é dada por

$$h'(t) = \left\{ -\frac{\gamma+1}{\alpha} + (a+1) \frac{\frac{\gamma}{\alpha} \left(\frac{t}{\alpha}\right)^{-\gamma}}{\left[1 + \left(\frac{t}{\alpha}\right)^{-\gamma}\right]} + a \frac{\frac{\gamma}{\alpha} \left(\frac{t}{\alpha}\right)^{-\gamma}}{\left[1 + \left(\frac{t}{\alpha}\right)^{-\gamma}\right]} \frac{\left[1 + \left(\frac{t}{\alpha}\right)^{-\gamma}\right]^{-a}}{1 - \left[1 + \left(\frac{t}{\alpha}\right)^{-\gamma}\right]^{-a}} \right\} \frac{\alpha}{t} h(t) \quad (17)$$

em que $h(t)$ é a função de taxa de falha da distribuição Kw-LL. Da equação (17), observa-se que o gráfico da função de taxa de falha para:

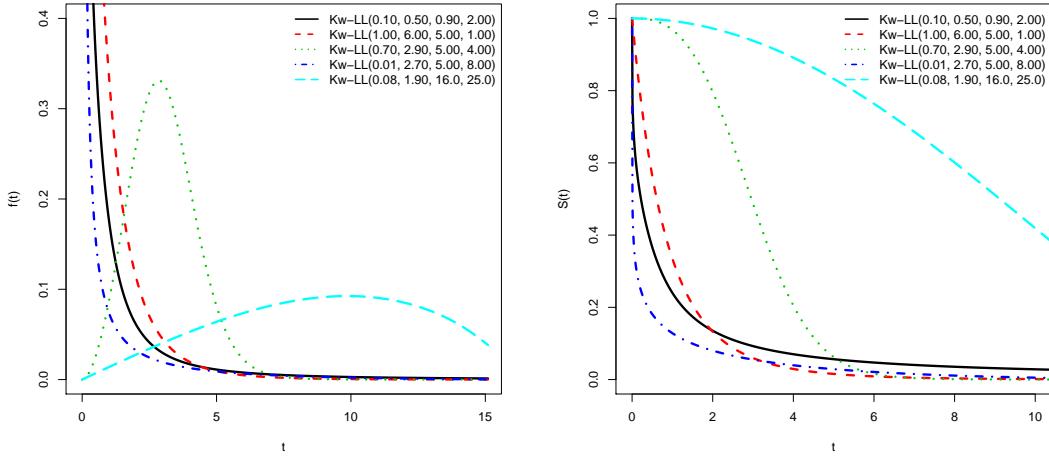


Figura 8 – Gráficos da fdp e da função de sobrevivência da distribuição Kw-LL para alguns valores dos parâmetros a, b, α, γ

- i. $a = 1$ e $\gamma > 1$, inicia-se na origem e aumenta até atingir seu máximo em $t = \alpha(\gamma - 1)^{1/\gamma}$, decrescendo em seguida e tendendo a zero conforme t aumenta;
- ii. $a = 1$ e $\gamma < 1$, inicia-se no infinito e decresce tendendo a zero;
- iii. $a = 1$ e $\gamma = 1$, é decrescente iniciando em b/α e tendendo a zero conforme t cresce;
- iv. $a = 1, b = 1$ e $\gamma > 1$, inicia-se na origem, cresce até um ponto máximo em $t = \alpha(\gamma - 1)^{1/\gamma}$ e em seguida decresce, tendendo a assintoticamente a reta $y = 0$ à medida que t aumenta;
- v. $a = 1, b = 1$ e $\gamma = 1$, inicia em $1/\alpha$, e então decresce monotonicamente;
- vi. $a = 1, b = 1$ e $\gamma < 1$, inicia no infinito e decresce a medida que t aumenta.

A Figura 9 apresenta algumas formas para a função de taxa de falha que a distribuição Kw-LL abrange. Observe que além das formas decrescente e unimodal, a distribuição Kw-LL abrange as formas de U (ou de banheira) e crescente.

Para gerar valores de uma distribuição Kw-LL, utilizamos a seguinte equação

$$t = \alpha \left\{ \left[1 - (1 - u)^{1/b} \right]^{-1/a} - 1 \right\}^{-1/\gamma}, \quad (18)$$

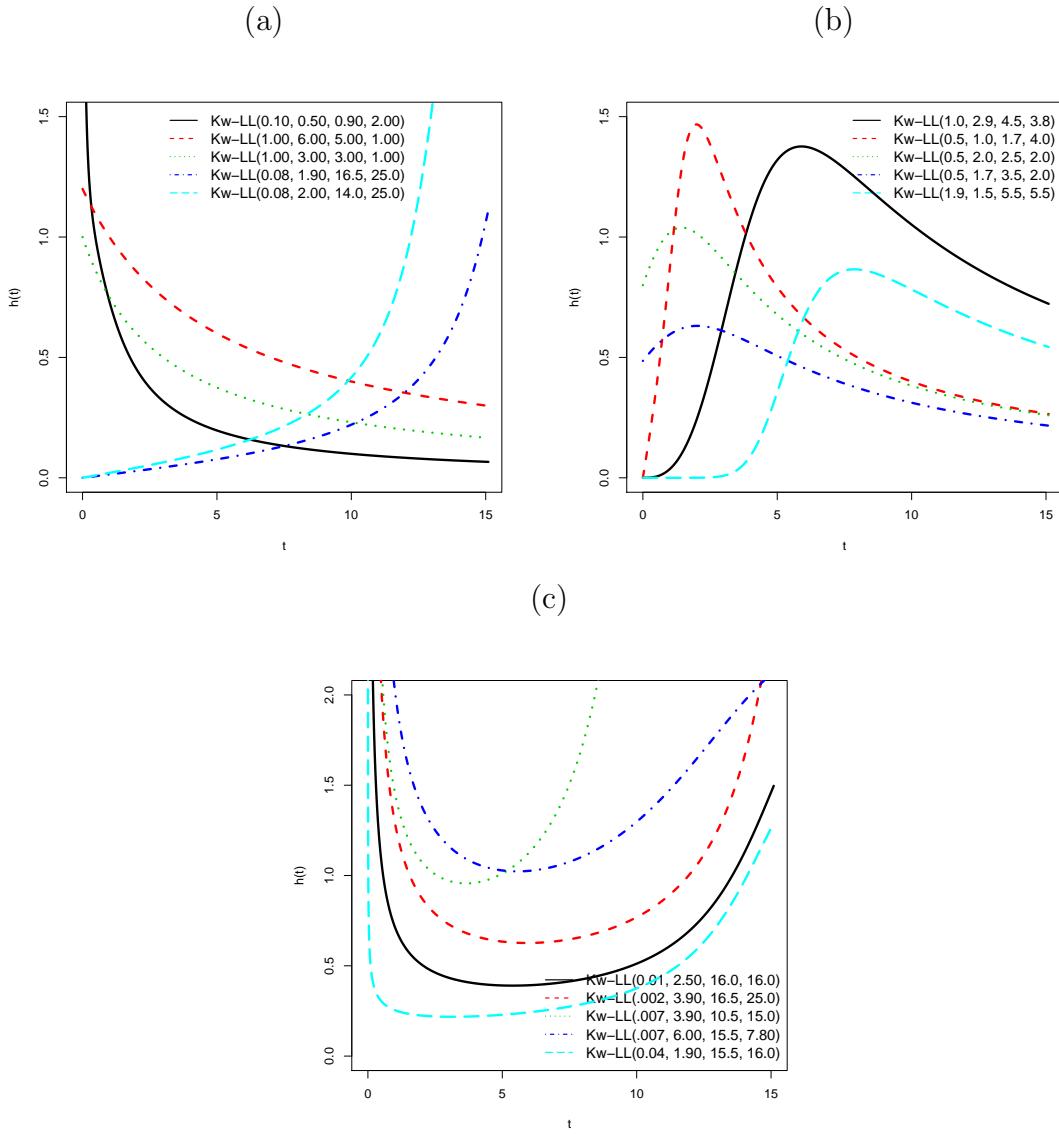


Figura 9 – Gráficos apresentando as formas da função de taxa de falha da distribuição Kw-LL. (a) Apresenta-se os gráficos da função de taxa de falha crescente e decrescente. (b) Apresenta-se o gráfico da função de taxa de falha unimodal. (c) Apresenta-se o gráfico da função de taxa de falha forma de U ou forma de banheira

onde u tem distribuição Uniforme, $U(0, 1)$.

A distribuição Kw-LL possui a mesma propriedade que as distribuições Weibull e log-normal, ou seja, dada uma variável aleatória T com distribuição de probabilidade Kw-LL, a variável aleatória $Y = \log(T)$ tem distribuição Kumaraswamy logística. Pelo método

do Jacobiano tem-se que

$$f(y) = \frac{ab}{\sigma} \exp\left(-\frac{y-\mu}{\sigma}\right) \left\{1 + \exp\left(-\frac{y-\mu}{\sigma}\right)\right\}^{-(a+1)} \left\{1 - \left[1 + \exp\left(-\frac{y-\mu}{\sigma}\right)\right]^{-a}\right\}^{b-1} \quad (19)$$

cuja reparametrização utilizada é $\gamma = 1/\sigma$ e $\alpha = \exp(\mu)$. Dessa forma tem-se que se $T \sim \text{Kw-LL}(a, b, \alpha, \gamma)$ então $Y = \log(T) \sim \text{Kw-L}(a, b, \mu, \sigma)$.

Uma importante propriedade da distribuição Kw-LL é o fato de possuir como submodelos várias distribuições de probabilidade, fixando valores para alguns parâmetros, como por exemplo:

- i. Para $a = 1$ e $b = 1$, obtém-se a distribuição log-logística como caso particular;
- ii. Para $a = 1$, $b = 1$, $\alpha = 1$ e $\gamma = 1$, obtém-se a distribuição log-logística padrão;
- iii. A distribuição Burr XII é um caso particular da Kw-LL, para $a = 1$, $\alpha = 1$ e $\lambda < 0$;
- iv. A distribuição log-logística exponenciada é um caso particular da distribuição Kw-LL, com $b = 1$;
- v. Fazendo $a = 1$, têm-se a distribuição log-logística estendida.

3.4 Propriedades da Kumaraswamy-log-logística

Em qualquer análise estatística é de grande importância o estudo dos momentos e da função geradora de momentos, especialmente em trabalhos aplicados. Algumas das características mais importantes e os aspectos de uma distribuição podem ser estudados por meio dos momentos (por exemplo, tendência, dispersão, assimetria e curtose).

Nas próximas seções apresentam-se algumas propriedades da distribuição Kw-LL. Na seção 3.4.1 será apresentada duas representações para a fdp em expansão de série. Os momentos e a função geradora de momentos serão apresentados na seção 3.4.2.

3.4.1 Expansão da função densidade da Kw-LL

Para $b > 0$ real não inteiro, tem-se que

$$[1 - G(t)^a]^{b-1} = \sum_{i=0}^{\infty} (-1)^i \binom{b-1}{i} G(t)^{ai}, \quad (20)$$

veja por exemplo Cordeiro e Castro (2010). Onde o coeficiente binomial está definido para qualquer real. Substituindo a expansão acima na fórmula (8), pode-se reescrever a fdp da Kw-LL como

$$f(t) = g(t) \sum_{i=0}^{\infty} w_i G(t)^{a(i+1)-1} \quad (21)$$

onde os coeficientes w_i são dados por

$$w_i = ab(-1)^i \binom{b-1}{i}, \quad (22)$$

note que, os coeficientes satisfazem $\sum_{i=1}^{\infty} w_i = 1$.

Se b é inteiro, então o índice i da soma para em $b-1$. Se a é um número inteiro, então a equação (21) mostra que a fdp da distribuição Kw-LL é igual à densidade da distribuição base, multiplicado por uma série de potência ponderada infinita de fda da distribuição base. Por outro lado, se a é um real não inteiro, pode-se expandir $G(t)^{a(i+1)-1}$,

como segue

$$G(t)^{a(i+1)-1} = \{1 - [1 - G(t)]\}^{a(i+1)-1} = \sum_{j=0}^{\infty} (-1)^j \binom{a(i+1)-1}{j} [1 - G(t)]^j$$

e então

$$G(t)^{a(i+1)-1} = \{1 - [1 - G(t)]\}^{a(i+1)-1} = \sum_{j=0}^{\infty} \sum_{r=0}^j (-1)^{j+r} \binom{a(i+1)-1}{j} \binom{j}{r} G(t)^r$$

assim

$$f(t) = g(t) \sum_{i,j=0}^{\infty} \sum_{r=0}^j w_{i,j,r} G(t)^r \quad (23)$$

onde os coeficientes $w_{i,j,r}$ são da forma

$$w_{i,j,r} = ab(-1)^{i+j+r} \binom{a(i+1)-1}{j} \binom{b-1}{i} \binom{j}{r}, \quad (24)$$

novamente note que os coeficientes satisfazem $\sum_{i,j}^{\infty} \sum_{r=0}^j w_{i,j,r} = 1$.

Portanto, a fdp da Kw-LL pode ser reescrita como

1. Se $a > 0$ é inteiro

$$f(t) = \frac{\gamma}{\alpha} \left(\frac{t}{\alpha} \right)^{-\gamma-1} \left[1 + \left(\frac{t}{\alpha} \right)^{-\gamma} \right]^{-a-1} \sum_{i=0}^{\infty} w_i \left[1 + \left(\frac{t}{\alpha} \right)^{-\gamma} \right]^{-ai}.$$

2. Se $a > 0$ é real não inteiro.

$$f(t) = \frac{\gamma}{\alpha} \left(\frac{t}{\alpha} \right)^{-\gamma-1} \left[1 + \left(\frac{t}{\alpha} \right)^{-\gamma} \right]^{-2} \sum_{i,j=0}^{\infty} \sum_{r=0}^j w_{i,j,r} \left[1 + \left(\frac{t}{\alpha} \right)^{-\gamma} \right]^{-r}.$$

3.4.2 Momentos

Substituindo em (10) a função inversa aplicada em $u^{1/a}$ pela inversa da fda da distribuição log-logística, obtém-se os momentos da Kw-LL, dado pela integral

$$\mu'_k = \alpha^k b \int_0^1 u^{k/\gamma a} \left(1 - u^{1/a}\right)^{-k/\gamma} (1-u)^{b-1} du \quad (25)$$

em que $0 < u < 1$.

Alternativamente pode-se escrever os momentos da Kw-LL usando a representação em série conforme descrito no teorema seguinte.

Teorema 1: Se $T \sim \text{Kw-LL}(a, b, \alpha, \gamma)$, então o k -ésimo momento é dado por

i. Para $a > 0$ inteiro

$$\mu'_k = ab\alpha^k \sum_{i=0}^{\infty} w_i B\left[a(i+1) + \frac{k}{\gamma}, 1 - \frac{k}{\gamma}\right]. \quad (26)$$

ii. Para $a > 0$ real não inteiro

$$\mu'_k = ab\alpha^k \sum_{i,j=0}^{\infty} \sum_{r=0}^j w_{i,j,r} B\left[1 + \frac{k}{\gamma} + r, 1 - \frac{k}{\gamma}\right], \quad (27)$$

em que $B(\cdot, \cdot)$ é a função beta e w_i e $w_{i,j,r}$ são definidos em (22) e (24), respectivamente.

Prova: Para a distribuição Kw-LL, os momentos são obtidos solucionando a equação

$$\mu'_k = \int_0^\infty t^k f(t) dt, \quad t > 0. \quad (28)$$

Usando os resultados (21) e (23) obtém-se

$$\mu'_k = \begin{cases} \sum_{i=0}^{\infty} w_i \int_0^\infty t^k g(t) G(t)^{a(i+1)-1} dt \\ \sum_{i,j=0}^{\infty} \sum_{r=0}^j w_{i,j,r} \int_0^\infty t^k g(t) G(t)^r dt \end{cases} \quad (29)$$

A integral em (29) pode ser escrita em termo dos momentos ponderados por probabilidade da variável aleatória, com fda $G(t)$, dada por

$$M(k, r, 0) = E(T^k G(t)^r) = \int_0^\infty t^k g(t) G(t)^r dt.$$

Seja $G(t)$ a fda da distribuição log-logística, então

$$\begin{aligned} M(k, r, 0) &= \int_0^\infty t^k \frac{\gamma}{\alpha} \left(\frac{t}{\alpha}\right)^{-\gamma-1} \left[1 + \left(\frac{t}{\alpha}\right)^{-\gamma}\right]^{-r-2} dt \\ &= \alpha^k \int_0^1 (1-w)^{-k/\gamma} w^{r+k/\gamma} dt \\ &= \alpha^k B\left(r+1+\frac{k}{\gamma}, 1+\frac{k}{\gamma}\right). \end{aligned} \quad (30)$$

Logo,

$$\mu'_k = \begin{cases} \sum_{i=0}^{\infty} w_i B\left[a(i+1) + \frac{k}{\gamma}, 1 - \frac{k}{\gamma}\right], & \text{para } a \text{ inteiro,} \\ \sum_{i,j=0}^{\infty} \sum_{r=0}^j w_{i,j,r} B\left[1 + \frac{k}{\gamma} + r, 1 - \frac{k}{\gamma}\right], & \text{para } a \text{ real não inteiro.} \end{cases}$$

□

Conseqüentemente, se $T \sim \text{Kw-LL}(a, b, \alpha, \gamma)$, então a função geradora de momentos (fgm) é dada por

$$\begin{aligned} M(s) &= \int_0^\infty \exp(st) f(t) dt = \int_0^\infty \sum_{k=0}^{\infty} \frac{(st)^k}{k!} f(t) dt \\ &= \sum_{k=0}^{\infty} \frac{s^k}{k!} \int_0^\infty t^k f(t) dt = \sum_{k=0}^{\infty} \frac{\mu'_k s^k}{k!}. \end{aligned}$$

A assimetria e curtose agora podem ser calculadas a partir dos momentos ordinários usando as relações apresentadas. Os momentos foram calculados utilizando o software estatístico R Development Core Team (2009) e a variância, assimetria e curtose foram obtidas das relações descritas em (11).

Os gráficos da assimetria e curtose são apresentados na Figura 10 para $\gamma = 1$,

$\alpha = 2$, e variando os parâmetros a e b . Os gráficos mostram imediatamente que as curvas de assimetria e curtose aumentam ou diminuem conforme os parâmetros a ou b variam.

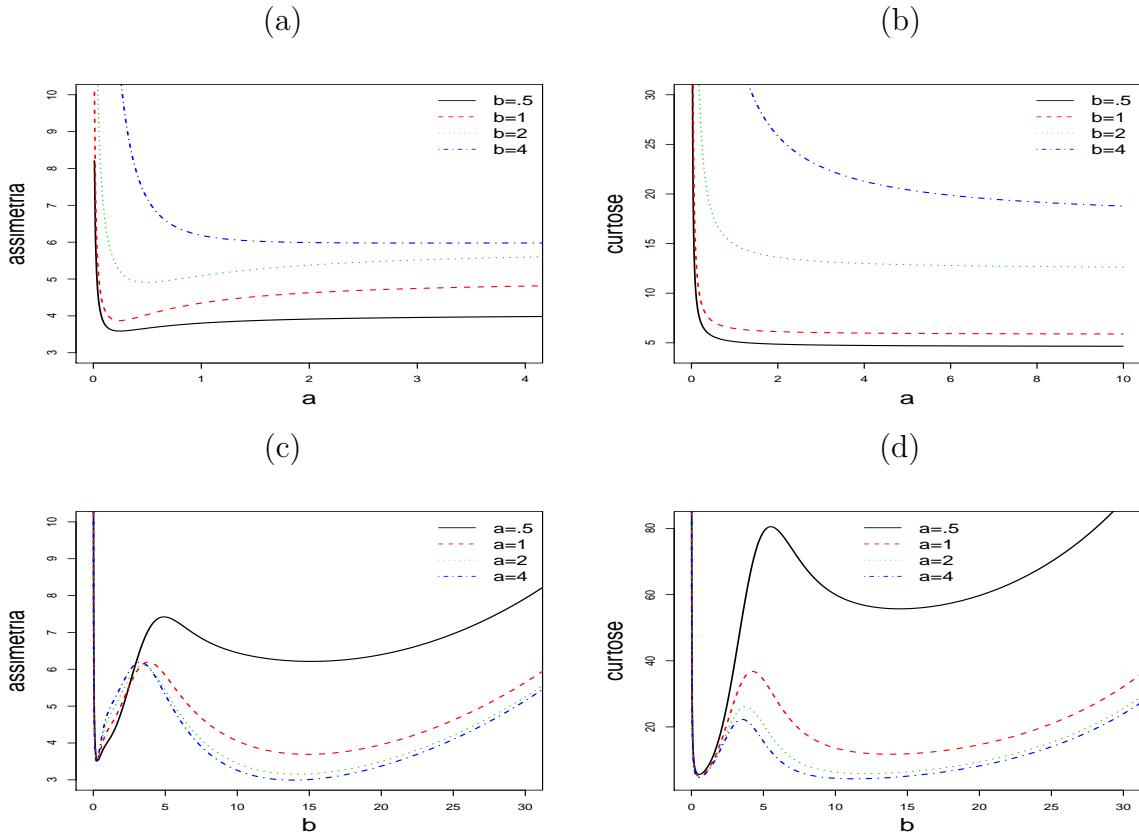


Figura 10 – Gráficos da assimetria e curtose da distribuição Kw-LL. Em (a) e (b) apresenta-se os gráficos da assimetria e curtose como função do parâmetro a , para alguns valores do parâmetro b . Em (c) e (d) apresenta-se os gráficos da assimetria e curtose como função do parâmetro b , para alguns valores do parâmetro a

3.4.3 Outras medidas associadas à distribuição Kw-LL

3.4.4 Desvio médio

O total de dispersão em uma população é evidentemente medido até certo ponto pelo total de desvios da média e mediana. Se T tem distribuição Kw-LL com fdp $f(t)$, então deriva-se os desvios médios da média $\mu'_1 = E(T)$ e da mediana m das relações

$$\delta_1 = \int_0^\infty |t - \mu'_1| f(t) dt \quad \text{e} \quad \delta_2 = \int_0^\infty |t - m| f(t) dt,$$

respectivamente. As medidas δ_1 and δ_2 podem ser calculadas usando as relações

$$\delta_1 = 2\mu'_1 F(\mu'_1) - 2 \int_0^{\mu'_1} t f(t) dt \quad \text{e} \quad \delta_2 = \mu'_1 - 2 \int_0^m t f(t) dt.$$

Usando a equação (20), e integrando por parte, pode-se mostrar que

$$\begin{aligned} \int_0^z t f(t) dt &= ab \int_0^z t g(t) G(t)^{a-1} [1 - G(t)^a]^{b-1} dt \\ &= ab \sum_{k=0}^{\infty} \binom{b-1}{k} (-1)^k \int_0^z t g(t) G(t)^{a(k+1)-1} dt \\ &= ab \sum_{k=0}^{\infty} \binom{b-1}{k} (-1)^k \int_0^z t \frac{\gamma}{\alpha} \left(\frac{t}{\alpha} \right)^{-\gamma-1} \left[1 + \left(\frac{t}{\alpha} \right)^{-\gamma} \right]^{-1-a(k+1)} dt \\ &= ab\alpha \sum_{k=0}^{\infty} \binom{b-1}{k} (-1)^k B_{G(z)} \left(a(k+1) - \frac{1}{\gamma}, 1 + \frac{1}{\gamma} \right) \end{aligned} \quad (31)$$

em que $B_{G(\cdot)}(\cdot, \cdot)$ é a função beta incompleta e $G(z)$ é a fda da distribuição log-logística.

Assim,

$$\delta_1 = 2\mu'_1 F(\mu'_1) - 2ab\alpha \sum_{k=0}^{\infty} \binom{b-1}{k} (-1)^k B_{G(\mu'_1)} \left(a(k+1) - \frac{1}{\gamma}, 1 + \frac{1}{\gamma} \right)$$

$$\delta_2 = \mu'_1 - 2ab\alpha \sum_{k=0}^{\infty} \binom{b-1}{k} (-1)^k B_{G(m)} \left(a(k+1) - \frac{1}{\gamma}, 1 + \frac{1}{\gamma} \right)$$

como $m = \text{mediana}(T)$, então $G(m) = 0,5$.

3.4.5 Curvas de Bonferroni e Lorenz

As curvas de Bonferroni e Lorenz tem aplicações não somente na área de economia em estudos de renda e pobreza, mas também em outras áreas como confiabilidade, estudos demográficos, seguro e medicina. Elas são definidas por

$$B(p) = \frac{1}{p\mu'_1} \int_0^q t f(t) dx \quad \text{e} \quad L(p) = \frac{1}{\mu'_1} \int_0^q t f(t) dx,$$

respectivamente, onde $\mu'_1 = E(T)$ e $q = F^{-1}(p)$. Da equação (18) obtém-se q para um dado p e de (31) têm-se

$$\begin{aligned} B(p) &= \frac{1}{p\mu'_1} ab\alpha \sum_{k=0}^{\infty} \binom{b-1}{k} (-1)^k B_{G(q)} \left(a(k+1) - \frac{1}{\gamma}, 1 + \frac{1}{\gamma} \right) \\ L(p) &= \frac{1}{\mu'_1} ab\alpha \sum_{k=0}^{\infty} \binom{b-1}{k} (-1)^k B_{G(q)} \left(a(k+1) - \frac{1}{\gamma}, 1 + \frac{1}{\gamma} \right). \end{aligned}$$

onde $G(q) = [1 - (1-p)^{1/b}]^{1/a}$.

3.4.6 Estatísticas de ordem

A função de densidade $f_{l:n}(t)$ da l -ésima estatística de ordem, para $l = 1, \dots, n$, das variáveis aleatórias T_1, \dots, T_n com fdp Kw-LL, apresentada em (14), é dada por

$$f_{l:n}(t) = \frac{f(t)}{B(l, n-l+1)} F(t)^{l-1} \{1 - F(t)\}^{n-l}, \quad (32)$$

onde $f(\cdot)$ e $F(\cdot)$ são a fdp e fda da distribuição Kw-LL, respectivamente e $B(\cdot, \cdot)$ denota a função beta. Da relação descrita em (20), obtém-se a representação em série para $[1 - F(t)]^{n-l}$, para n inteiro. Portanto pode-se reescrever (32) como

$$f_{l:n}(t) = \frac{f(t)}{B(l, n-l+1)} \sum_{c=0}^{n-l} \binom{n-l}{c} (-1)^c F(t)^{l+c-1}. \quad (33)$$

A equação (33) é uma expressão para a densidade da estatística de ordem da distribuição Kw-LL. Da equação (7) e (20), obtém-se a expansão para $F(t)^{l+c-1}$

$$F(t)^{l+c-1} = \{1 - [1 - G(t)^a]^b\}^{l+c-1} = \sum_{k=0}^{l+c-1} (-1)^k \binom{l+c-1}{k} [1 - G(t)^a]^{kb}. \quad (34)$$

Usando a expansão em série para $[1 - G(t)^a]^{kb}$

$$[1 - G(t)^a]^{kb} = \sum_{m=0}^{\infty} (-1)^m \binom{kb}{m} G(t)^{ma}. \quad (35)$$

Expandindo em série de potência $G(t)^{ma}$, em (35) tem-se que

$$\begin{aligned}
 G(t)^{ma} &= \{1 - [1 - G(t)]\}^{ma} = \\
 &= \sum_{h=0}^{\infty} (-1)^h \binom{ma}{h} \{1 - G(t)\}^h = \\
 &= \sum_{h=0}^{\infty} (-1)^h \binom{ma}{h} \sum_{d=0}^h (-1)^d \binom{h}{d} G(t)^d.
 \end{aligned} \tag{36}$$

Logo, substituindo (36) em (35) e (35) em (34), tem-se que

$$F(t)^{l+c-1} = \sum_{k=0}^{l+c-1} (-1)^k \binom{l+c-1}{k} \sum_{m=0}^{\infty} (-1)^m \binom{kb}{m} \sum_{h=0}^{\infty} (-1)^h \binom{ma}{h} \sum_{d=0}^h (-1)^d \binom{h}{d} G(t)^d.$$

Substituindo $\sum_{h=0}^{\infty} \sum_{d=0}^h$ por $\sum_{d=0}^{\infty} \sum_{h=d}^{\infty}$, obtém-se

$$F(t)^{l+c-1} = \sum_{k=0}^{l+c-1} \sum_{m=0}^{\infty} \sum_{d=0}^{\infty} \sum_{h=d}^{\infty} (-1)^{k+m+h+d} \binom{l+c-1}{k} \binom{kb}{m} \binom{ma}{h} \binom{h}{d} G(t)^d. \tag{37}$$

para $d, l + c - 1 = 0, 1, \dots$

Se a é real não inteiro, substituindo (23) e (37) em (33), pode-se escrever a densidade $f_{l:n}(t)$ na forma

$$\begin{aligned}
 f_{l:n}(t) &= \frac{abg(t)}{B(l, n-l+1)} \sum_{c=0}^{n-l} (-1)^c \binom{n-l}{c} \times \\
 &\quad \times \sum_{i,j=0}^{\infty} \sum_{r=0}^j (-1)^{i+j+r} \binom{a(i+1)-1}{j} \binom{b-1}{i} \binom{j}{r} \times \\
 &\quad \times \sum_{k=0}^{l+c-1} \sum_{m=0}^{\infty} \sum_{d=0}^{\infty} \sum_{h=d}^{\infty} (-1)^{k+m+h+d} \binom{l+c-1}{k} \binom{kb}{m} \binom{ma}{h} \binom{d}{h} G(t)^{d+r}.
 \end{aligned} \tag{38}$$

Se a é inteiro, obtém-se das equações (21), (33) e (37)

$$\begin{aligned}
 f_{l:n}(t) &= \frac{abg(t)}{B(l, n-l+1)} \sum_{c=0}^{n-l} (-1)^c \binom{n-l}{c} \sum_{i=0}^{\infty} (-1)^i \binom{b-1}{i} \times \\
 &\quad \times \sum_{k=0}^{l+c-1} \sum_{m,d=0}^{\infty} \sum_{h=d}^{\infty} (-1)^{k+m+h+d} \binom{l+c-1}{k} \binom{kb}{m} \binom{ma}{h} \binom{h}{d} G(t)^{a(i+1)+d-1}.
 \end{aligned} \tag{39}$$

As equações (38) e (39) dão imediatamente a densidade das estatísticas de ordem da distribuição Kw-LL como função da densidade da distribuição base multiplicado pela soma ponderada infinita de potências de $G(t)$.

3.5 Estimação de máxima verossimilhança com dados censurados

Seja T uma variável aleatória com distribuição Kw-LL com vetor de parâmetros $\boldsymbol{\theta} = (a, b, \alpha, \gamma)^T$. Os dados encontrados em análises de sobrevivência e estudos de confiabilidade são frequentemente censurados. Um caso muito simples de mecanismo de censura aleatória e que é frequente em situações reais, é aquele em que cada indivíduo i assume ter tempo de vida T_i e tempo de censura C_i , onde T_i e C_i são variáveis independentes. Supondo que os dados consistam de n observações independentes $t_i = \min(T_i, C_i)$ para $i = 1, \dots, n$. A distribuição de C_i não depende de qualquer parâmetro desconhecido de T_i . O logarítmo da função de verossimilhança considerando a distribuição Kw-LL é dado por

$$\begin{aligned} l(\boldsymbol{\theta}) &= r \log \left(\frac{ab\gamma}{\alpha} \right) + (a\gamma - 1) \sum_{i \in F} \log \left(\frac{t_i}{\alpha} \right) - (a+1) \sum_{i \in F} \log(u_i) + \\ &= +(b-1) \sum_{i \in F} \log \left[1 - u_i^{-a} \left(\frac{t_i}{\alpha} \right)^{a\gamma} \right] + b \sum_{i \in C} \log \left[1 - u_i^{-a} \left(\frac{t_i}{\alpha} \right)^{a\gamma} \right], \end{aligned} \quad (40)$$

em que $u_i = 1 + \left(\frac{t_i}{\alpha} \right)^\gamma$, r é o número de falhas e F e C denotam o conjunto de observações não censuradas e censuradas, respectivamente.

As derivadas de primeira ordem de (40) em relação aos parâmetros em $\boldsymbol{\theta}$ são dadas por

$$U_a(\boldsymbol{\theta}) = \frac{r}{a} - \sum_{i \in F} \log(u_i) + (b-1) \sum_{i \in F} q_i \log(u_i) + b \sum_{i \in C} q_i \log(u_i)$$

$$U_b(\boldsymbol{\theta}) = \frac{r}{b} + \sum_{i \in F} \log(1 - u_i^{-a}) + \sum_{i \in C} \log(1 - u_i^{-a}),$$

$$\begin{aligned} U_\alpha(\boldsymbol{\theta}) &= -\frac{r}{\alpha} + \frac{\gamma-1}{\alpha^r} - \frac{(a+1)\gamma}{\alpha} \sum_{i \in F} \frac{1}{u_i} \left(\frac{t_i}{\alpha} \right)^{-\gamma} + \frac{(b-1)a\gamma}{\alpha} \sum_{i \in F} q_i \left(\frac{t_i}{\alpha} \right)^{-\gamma} \\ &+ \frac{ab\gamma}{\alpha} \sum_{i \in C} q_i \left(\frac{t_i}{\alpha} \right)^{-\gamma}, \end{aligned}$$

e

$$\begin{aligned} U_\gamma(\boldsymbol{\theta}) &= \frac{r}{\gamma} - \sum_{i \in F} \log \left(\frac{t_i}{\alpha} \right) + (a+1) \sum_{i \in F} \frac{1}{u_i} \left(\frac{t_i}{\alpha} \right)^{-\gamma} \log \left(\frac{t_i}{\alpha} \right) \\ &\quad - (b-1)a \sum_{i \in F} q_i \left(\frac{t_i}{\alpha} \right)^{-\gamma} \log \left(\frac{t_i}{\alpha} \right) - ab \sum_{i \in C} q_i \left(\frac{t_i}{\alpha} \right)^{-\gamma} \log \left(\frac{t_i}{\alpha} \right), \end{aligned}$$

onde

$$q_i = \frac{u_i^{-a}}{1 - u_i^{-a}}.$$

Consequentemente, o estimador de máxima verossimilhança (EMV) $\hat{\boldsymbol{\theta}}$ de $\boldsymbol{\theta}$ é obtido numéricamente a partir das equações não lineares

$$U_a(\boldsymbol{\theta}) = U_b(\boldsymbol{\theta}) = U_k(\alpha) = U_\gamma(\boldsymbol{\theta}) = 0.$$

Para a estimação de intervalo e testes de hipóteses sobre os parâmetros do modelo é necessário a matriz 4×4 de informação observada

$$J = J(\boldsymbol{\theta}) = \begin{pmatrix} j_{aa} & j_{ab} & j_{a\alpha} & j_{a\gamma} \\ j_{bb} & j_{b\alpha} & j_{b\gamma} & \\ j_{\alpha a} & j_{\alpha b} & j_{\alpha\gamma} & \\ & & j_{\gamma\gamma} & \end{pmatrix},$$

cujos elementos são apresentados no Apêndice A.

Sobre as condições de regularidade, a distribuição assintótica dos EMV é dada por

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \text{ é } N_4(0, I(\boldsymbol{\theta})^{-1}),$$

onde $I(\boldsymbol{\theta})$ é a matriz de informação esperada. Esta distribuição aproximada permanece quando $I(\boldsymbol{\theta})$ é substituído por $J(\hat{\boldsymbol{\theta}})$, isto é, a matriz de informação observada estimada em $\hat{\boldsymbol{\theta}}$. A distribuição normal multivariada $N_4(0, J(\hat{\boldsymbol{\theta}})^{-1})$ pode ser utilizada na construção de intervalos de confiabilidade aproximados para os parâmetros individuais e para as funções de sobrevivência e de taxa de falha. De fato, o intervalo de confiança assintótico com nível de

significância γ , para cada parâmetro θ_r é obtido por meio da expressão

$$ACI(\theta_r, 100(1 - \gamma)\%) = \left(\hat{\theta}_r - z_{\gamma/2} \sqrt{\frac{\hat{j}^{\theta_r, \theta_r}}{n}}, \hat{\theta}_r + z_{\gamma/2} \sqrt{\frac{\hat{j}^{\theta_r, \theta_r}}{n}} \right),$$

em que, $\hat{j}^{\theta_r, \theta_r}$ é o r -ésimo elementos diagonal estimado de $J(\hat{\boldsymbol{\theta}})^{-1}$ para $r = 1, \dots, 4$ e $z_{\gamma/2}$ é o quantil $1 - \gamma/2$ da distribuição Normal padrão.

A estatística razão de verossimilhança (LR) é útil para testar o ajuste da distribuição Kw-LL e para compará-la com alguns de seus submodelos. Por exemplo, pode-se usar a estatística LR para checar se o ajuste usando a distribuição Kw-LL é estatisticamente “superior” ao ajuste usando as distribuições log-logística exponenciada e log-logística para um determinado banco de dados. Em qualquer caso, teste de hipóteses do tipo $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, onde $\boldsymbol{\theta}_0$ é um vetor especificado, pode ser realizado usando a estatística LR . Por exemplo, o teste de $H_0 : b = 1$ versus $H_1 : H_0$ não é verdade, é equivalente a comparar a distribuição log-logística exponenciada com a distribuição Kw-LL para o qual a estatística LR reduz a

$$w = 2[\ell(\hat{a}, \hat{b}, \hat{\alpha}, \hat{\gamma}) - \ell(\tilde{a}, 1, \tilde{\alpha}, \tilde{\gamma})],$$

em que $\hat{a}, \hat{b}, \hat{\alpha}$ e $\hat{\gamma}$ são os estimadores de máxima verossimilhança sobre H e $\tilde{a}, \tilde{\alpha}$ e $\tilde{\gamma}$ são os estimadores sobre H_0 .

4 DISTRIBUIÇÕES LOGÍSTICA GENERALIZADA

Nesse capítulo são apresentadas duas formas de generalizações da distribuição logística, resultando nas novas distribuições logística modificada (LM) e Kumaraswamy-logística (Kw-L). Na seção 4.1 faz-se uma revisão sobre a distribuição logística, apresentando as principais características e propriedades dessa distribuição. Também é feita uma revisão de trabalhos em diversas áreas aplicadas e teóricas que utilizaram a distribuição logística nos últimos anos, evidenciando sua importância nos diversos campos do conhecimento. Na seção 4.2, a nova distribuição LM é introduzida. A distribuição Kw-L é discutida na seção 4.3.

4.1 Distribuição logística

A distribuição logística, com parâmetros μ e σ , considera que $Y = \log(T)$, em que $T \sim LL(\alpha, \gamma)$, tem fdp dada por

$$g(y) = \frac{\exp\left(\frac{y-\mu}{\sigma}\right)}{\sigma \left[1 + \exp\left(\frac{y-\mu}{\sigma}\right)\right]^2} = \frac{1}{4\sigma} \operatorname{sech}^2\left(\frac{y-\mu}{2\sigma}\right), \quad -\infty < y < \infty,$$

em que $-\infty < \mu < \infty$ é o parâmetro de locação e $\sigma > 0$ é o parâmetro de escala, com reparametrização $\alpha = \exp(\mu)$ e $\gamma = 1/\sigma$.

A variável aleatória Y seguindo a distribuição logística será denotada, neste trabalho, por $Y \sim \text{Logística}(\mu, \sigma)$. A fda para a variável aleatória Y tem a seguinte forma

$$G(y) = \left[1 + \exp\left(\frac{-(y-\mu)}{\sigma}\right)\right]^{-1} = \frac{1}{2} + \frac{1}{2} \tanh\left(\frac{y-\mu}{2\sigma}\right),$$

consequentemente, a função de sobrevivência é dada por

$$S(y) = \left[1 + \exp\left(\frac{y-\mu}{\sigma}\right)\right]^{-1}.$$

A função de taxa de falha é dada por

$$h(y) = \frac{1}{\sigma} \exp\left(\frac{y-\mu}{\sigma}\right) \left[1 + \exp\left(\frac{y-\mu}{\sigma}\right)\right]^{-1}.$$

A distribuição logística modela dados de crescimento em populações humanas e

em alguns organismos biológicos e é aplicada na área de saúde pública, Balakrishnan (1992). Devido à fdp da distribuição logística poder ser expressa, alternativamente, em termos da secante hiperbólica ao quadrado, ela também é conhecida como distribuição secante ao quadrado, como ressalta Balakrishnan (1992).

Os gráficos da fdp, função de sobrevivência e função de taxa de falha são apresentados na Figura 11 para alguns valores de μ e σ . O gráfico fdp da distribuição logística assume forma unimodal, como pode ser visto na Figura 11, onde o valor máximo da fdp é determinado pelo parâmetro de locação μ , ou seja $g(y)$ é máximo quando $y = \mu$.

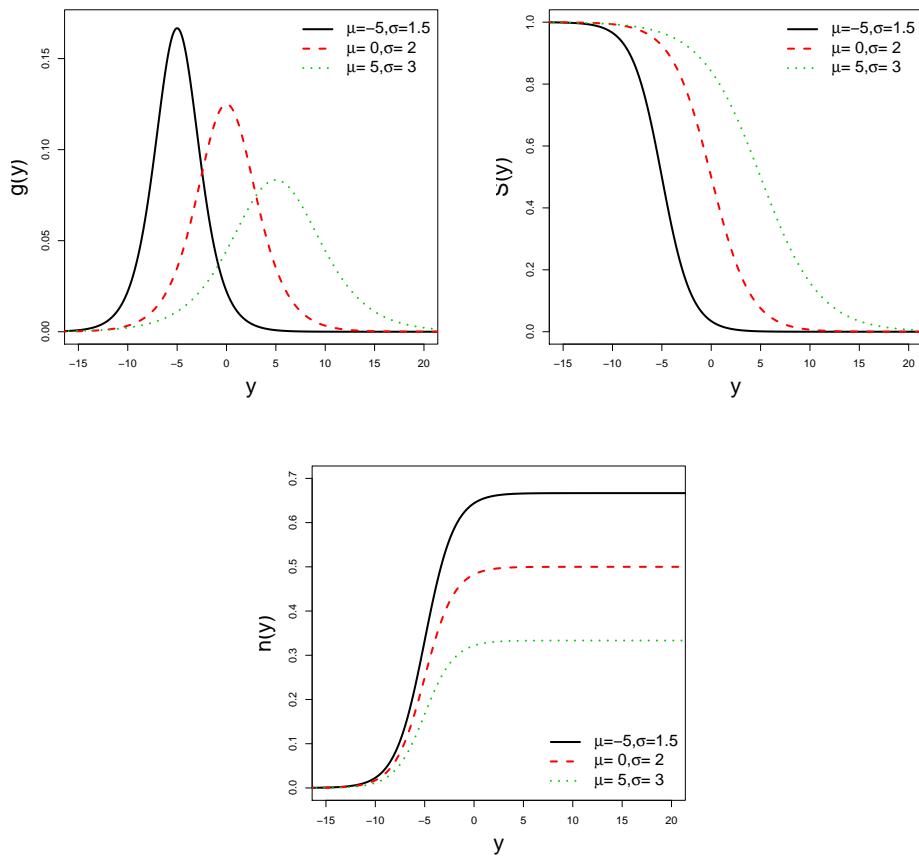


Figura 11 – Gráficos da fdp, função de sobrevivência e função de taxa de falha da distribuição logística para alguns valores σ e μ

A seguir é apresentada uma revisão de alguns trabalhos publicados nos últimos anos utilizando a distribuição logística exemplificando a importância dessas distribuições nas diversas áreas do conhecimento.

Ojo e Olapade (2003) consideraram a distribuição logística generalizada (Beta logística), apresentada por George e Ojo (1980). Os autores apresentaram e provaram alguns teoremas que relacionam a distribuição logística generalizada com outras distribuições.

Nadarajah (2009a) motivado pela importância dos produtos de variáveis aleatórias na área de ciências sociais e áreas relacionadas, derivou a distribuição exata de $|Y_1 Y_2|$ e $|Y_1 Y_2 \dots Y_p|$ quando Y_m , $m = 1, 2, \dots, p$, são variáveis aleatórias independentes com distribuição logística e discutiu possíveis extensões.

Nesse mesmo enfoque, Nadarajah (2009b) derivou as propriedades da distribuição logística assimétrica de Azzalini e apresentou a fda, o n -ésimo momento e expressões para os primeiros dez momentos, o n -ésimo momento central, a função geradora de momentos, função característica, desvio médio sobre a media e mediana, entropia de Rényi e Shannon, estatística de ordem, a distribuição assintótica das estatísticas de ordem extrema, estimativa pelo método dos momentos e máxima verossimilhança e uma aplicação para a regressão logística é discutida.

Por outro lado, Miyazaki e Hoshino (2009) propuseram um novo modelo semi-paramétrico, em teoria da resposta ao item, usando o processo de Dirichlet na distribuição logística de mistura que requer que as curvas de característica de item (ICCs) sejam funções monótonas não-decrescentes. Ou seja, pode-se tratar mais tipos de padrões de resposta ao item que os métodos existentes, como o modelo normal ogive de um parâmetro ou os modelos logístico com dois ou três parâmetros. Também foram conduzidos dois estudos de simulação cujos resultados indicaram que o método proposto pode expressar mais padrões de formas para ICCs e podem estimar os parâmetros de habilidade com mais precisão que os métodos paramétrico e não-paramétricos existentes. O modelo proposto também foi aplicado a dados de reconhecimento de expressão facial com resultados notáveis.

Outro trabalho que utilizada da distribuição logística, é dado por Stamatopoulos et al. (2009) os quais modelaram o processo de formação de coalizão entre primatas como um jogo sequencial. A população consistia de indivíduos que têm graus sociais distintos, determinados pelo recurso do indivíduo que asegura potencial. Cada membro da população estava interessado em ganhar acesso a um recurso de comida, individualmente ou por uma coalizão. Em qualquer fase do jogo, um jogador poderia propor uma coalizão específica ou

poderia ser proposto para unir-se a uma. Conseqüentemente, a estratégia de um jogador consistia em uma sucessão de decisões que considerava propor a formação de uma coalizão e quais propostas aceitar ou rejeitar. As preferências dos jogadores eram derivadas em cima das várias estruturas de coalizão sobre da suposição que a probabilidade de uma coalizão para obter os recursos é determinada pela distribuição logística como uma função de forças relativas dos jogadores e é mostrado que, dado um determinado comportamento estratégico dos primatas, uma variedade de estruturas de coalizão diferente pode emergir em equilíbrio.

Recentemente, Alkasasbeh e Raqab (2009) estudaram a distribuição logística generalizada, definida por Balakrishnan e Leng (1988), e considerando a estimativa de máxima verossimilhança de diferentes parâmetros, assim como outros cinco procedimentos de estimativa e compararam a performance destes procedimentos através de extensiva simulação numérica.

4.2 Distribuição logística modificada

A variável aleatória dada por $Y = \log(T)$, em que $T \sim \text{LLM}(v, \alpha, \gamma)$, segue distribuição logística modificada, com reparametrização $\mu = \log(\alpha)$, $\alpha > 0$ e $\sigma = 1/\gamma$, $\gamma > 0$, cuja função de sobrevivência é dada por

$$S(y) = \frac{v}{v + \exp\left(\frac{y-\mu}{\sigma}\right)}, \quad -\infty < y < \infty.$$

em que $v > 0$, $\sigma > 0$ e $-\infty < \mu < \infty$. A distribuição logística modificada, com parâmetros v , μ e σ , será denotada por $\text{LM}(v, \mu, \sigma)$.

A partir da função de sobrevivência pode-se construir a fda, a fdp e a função de taxa de falha. Que são dadas respectivamente por

$$F(y) = 1 - \frac{v}{v + \exp\left(\frac{y-\mu}{\sigma}\right)}, \quad f(y) = \frac{v}{\sigma} \frac{\exp\left(\frac{y-\mu}{\sigma}\right)}{\left[v + \exp\left(\frac{y-\mu}{\sigma}\right)\right]^2} \quad \text{e} \quad h(y) = \frac{1}{\sigma} \frac{\exp\left(\frac{y-\mu}{\sigma}\right)}{v + \exp\left(\frac{y-\mu}{\sigma}\right)}.$$

Os gráficos da fdp e da função de taxa de falha são apresentados na Figura 12, para alguns valores de v , μ e σ . O gráfico da fdp da distribuição LM assume forma unimodal, como pode ser visto na Figura 12(a), e à medida que o parâmetro v aumenta, o gráfico é

deslocado para a direita, o que também se verifica no gráfico da função de taxa de falha, Figura 12(b).

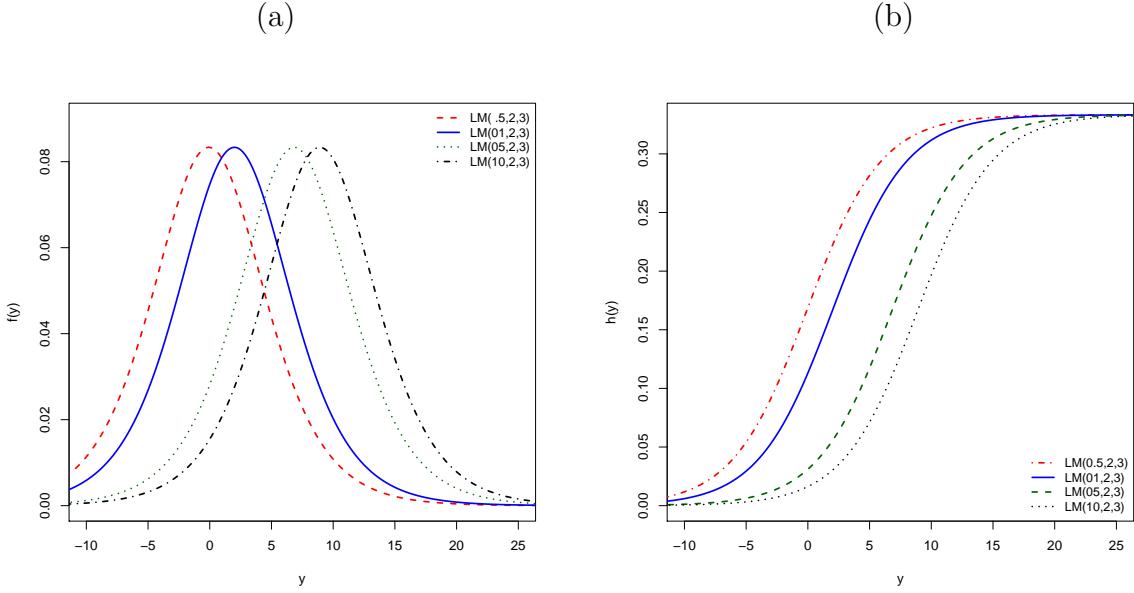


Figura 12 – Gráficos da fdp e da função de taxa de falha da distribuição LM, para $\mu = 2$, $\sigma = 3$ e v assumindo valores 0, 5, 1, 5 e 10

Os momentos da distribuição LM podem ser obtidos pela expressão

$$E(Y^r) = \int_{-\infty}^{\infty} [\sigma(\ln(v) + \theta) + \mu]^r \frac{1}{4} \sec h^2\left(\frac{\theta}{2}\right) d\theta$$

onde $\theta = \frac{y-\mu}{\sigma} - \ln(v)$. Assim, para $r = 1$, obtemos

$$E(Y) = \sigma \ln(v) + \mu$$

4.3 Distribuição Kumaraswamy-logística

A distribuição logística foi generalizada, neste capítulo, tendo por base a classe de distribuições Kumaraswamy proposta por Cordeiro e Castro (2010) e apresentada na seção 2.2.5.

Seja a variável aleatória $Y = \log(T)$, onde T é variável aleatória representando o tempo de falha, com distribuição Kw-LL(a, b, α, γ). Então Y tem distribuição Kumaraswamy-logística com quatro parâmetros, denotada por Kw-L(a, b, μ, σ) e reparametrização $\mu = \log(\alpha)$

e $\sigma = 1/\gamma$, cuja fdp é definida pela expressão

$$\begin{aligned} f(y) &= \frac{ab}{\sigma} \frac{1}{1 + \exp(\frac{y-\mu}{\sigma})} \left[\frac{\exp(\frac{y-\mu}{\sigma})}{1 + \exp(\frac{y-\mu}{\sigma})} \right]^a \times \\ &\quad \times \left\{ 1 - \left[1 - \frac{1}{1 + \exp(\frac{y-\mu}{\sigma})} \right]^a \right\}^{b-1} \quad -\infty < y < \infty , \end{aligned} \quad (41)$$

sendo $a > 0$, $b > 0$, $-\infty < \mu < \infty$ e $\sigma > 0$. A fda é dada por

$$F(y) = 1 - \left\{ 1 - \left[1 - \frac{1}{1 + \exp(\frac{y-\mu}{\sigma})} \right]^a \right\}^b . \quad (42)$$

A Figura 13 apresenta algumas formas da função densidade da Kw-L para $\mu = 0$, $\sigma = 1$, variando um dos parâmetros a ou b e fixando o outro parâmetro igual a 1. Na figura 13(a) o parâmetro a é igual a 1. Os valores atribuídos para b na figura 13(a) são 0, 05; 0, 01; 0, 50; 1, 00 e 3, 00. Na figura 13(b) a pertence ao conjunto $\{0, 05; 0, 01; 0, 50; 1, 00\}$ e b é igual a 1. Observa-se a grande flexibilidade introduzida pelos novos parâmetros a e b . Assim (41) permite uma maior flexibilidade, sendo, portanto, muito útil na maioria das situações práticas.

A função de sobrevivência da distribuição Kw-L é apresentada em (43) e seu gráfico é apresentado na Figura 14 para alguns valores dos parâmetros.

$$S(y) = \left\{ 1 - \left[1 - \frac{1}{1 + \exp(\frac{y-\mu}{\sigma})} \right]^a \right\}^b \quad (43)$$

A função de taxa de falha é dada pela equação.

$$h(y) = \frac{ab}{\sigma} \frac{1}{1 + \exp(\frac{y-\mu}{\sigma})} \left[\frac{\exp(\frac{y-\mu}{\sigma})}{1 + \exp(\frac{y-\mu}{\sigma})} \right]^a \left\{ 1 - \left[1 - \frac{1}{1 + \exp(\frac{y-\mu}{\sigma})} \right]^a \right\}^{-1} \quad (44)$$

Uma importante característica da distribuição Kw-L é que ela contém alguns importantes sub-modelos, como por exemplo:

- i. Para $a = b = 1$, ela é simplificada para a distribuição logística;
- ii. A distribuição logística exponenciada é um caso particular da distribuição Kw-L, com

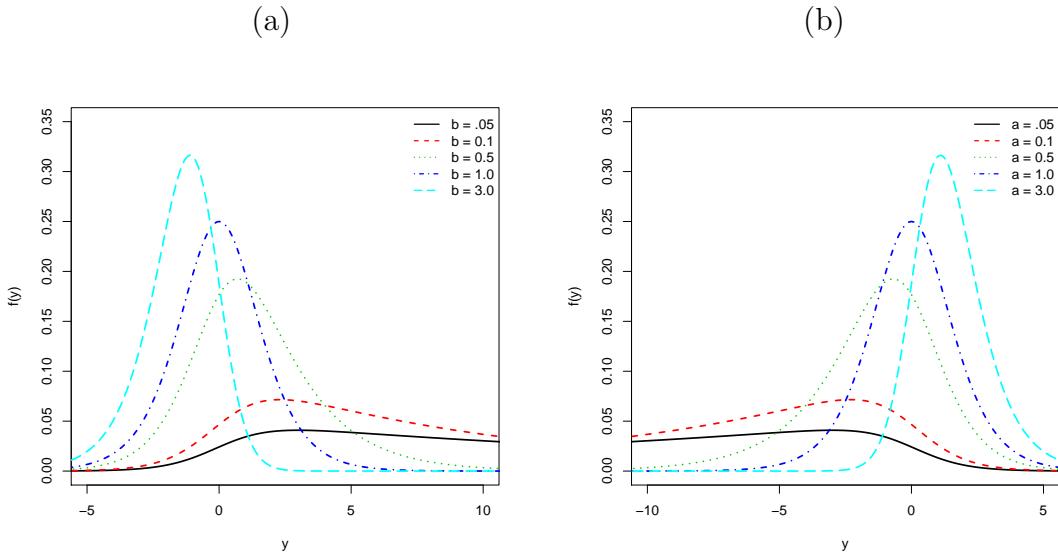


Figura 13 – Gráficos da fdp da distribuição Kw-L. (a) Apresenta-se os gráficos da fdp da distribuição Kw-L para $\mu = 0$, $\sigma = 1$, $a = 1$ e b assumindo valores 0, 0,05, 0, 1, 0, 5, 1 e 3. (b) Apresenta-se os gráficos da fdp da distribuição Kw-L para $\mu = 0$, $\sigma = 1$, $b = 1$ e a assumindo valores 0, 0,05, 0, 01, 0, 5, 1 e 3

$b = 1$;

- Se $a = 1$, obtém-se a distribuição logística estendida;
- Se $a = 2$ e $b = 1$, reduz a distribuição logística skew, veja por, exemplo, Nadarajah (2009b);
- Para $a = 1$, $\mu = 0$ e $\sigma = 1$, obtém-se a distribuição logística generalizada tipo II, com parâmetros b e 1, $\text{LG}(\text{II})(b, 1)$;
- A distribuição logística skew de Azzalini, discutida em Nadarajah (2009b), com três parâmetros $\lambda = 1$, σ e μ , é um caso particular da distribuição Kw-L para $a = 2$, e $b = 1$;
- A distribuição logística generalizada, apresentada por Alkasasbeh e Raqab (2009), é um caso particular da Kw-L, quando $\sigma = 1/\lambda$, $a = \alpha$ e $x = y + \mu$.

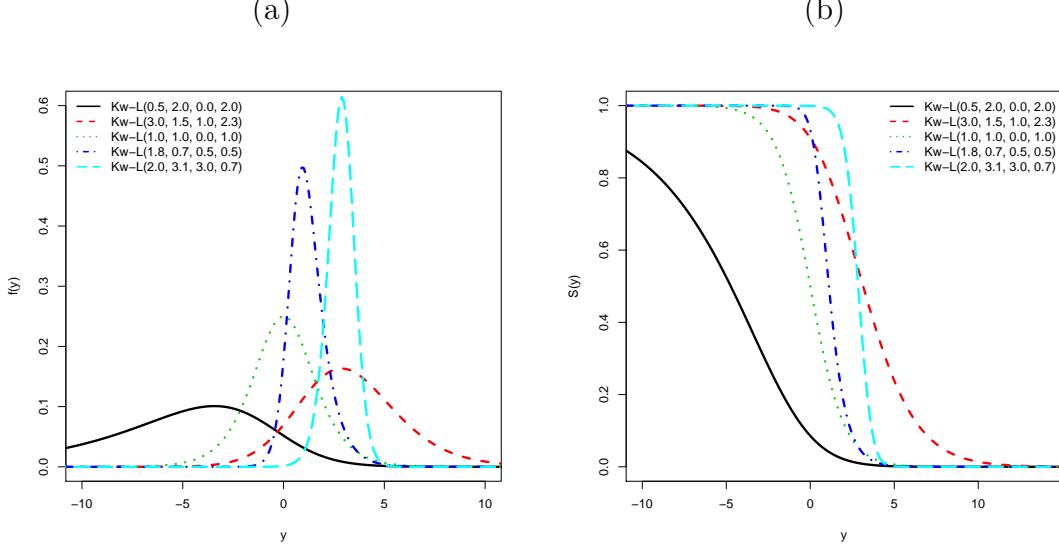


Figura 14 – Gráficos da fdp e da função de sobrevivência da distribuição Kw-L. (a) Apresenta-se a fdp para alguns valores dos parâmetros a , b , μ e σ . (b) Apresenta-se a função de sobrevivência para alguns valores dos parâmetros a , b , μ e σ

4.4 Momentos e função geradora de momentos

Substituindo em (10) a função inversa aplicada em $u^{1/a}$, pela inversa da fda da distribuição logística obtém-se os momentos da Kw-L, dado pela integral

$$\mu'_k = b \int_0^1 \left[\mu + \sigma \ln \left(\frac{w^{1/a}}{1-w^{1/a}} \right) \right]^k (1-w)^{b-1} dw \quad (45)$$

onde $G^{-1}(w^{1/a}) = \mu + \sigma \ln \left(\frac{w^{1/a}}{1-w^{1/a}} \right)$.

Os gráficos da assimetria e curtose são obtidas a partir das relações apresentadas em (11) e apresentados na Figura 10 para $\mu = 1$ e $\sigma = 2$ e variando os parâmetros a e b . Os gráficos mostram imediatamente que a curva de assimetria e curtose aumenta ou diminui conforme os parâmetros a ou b variam.

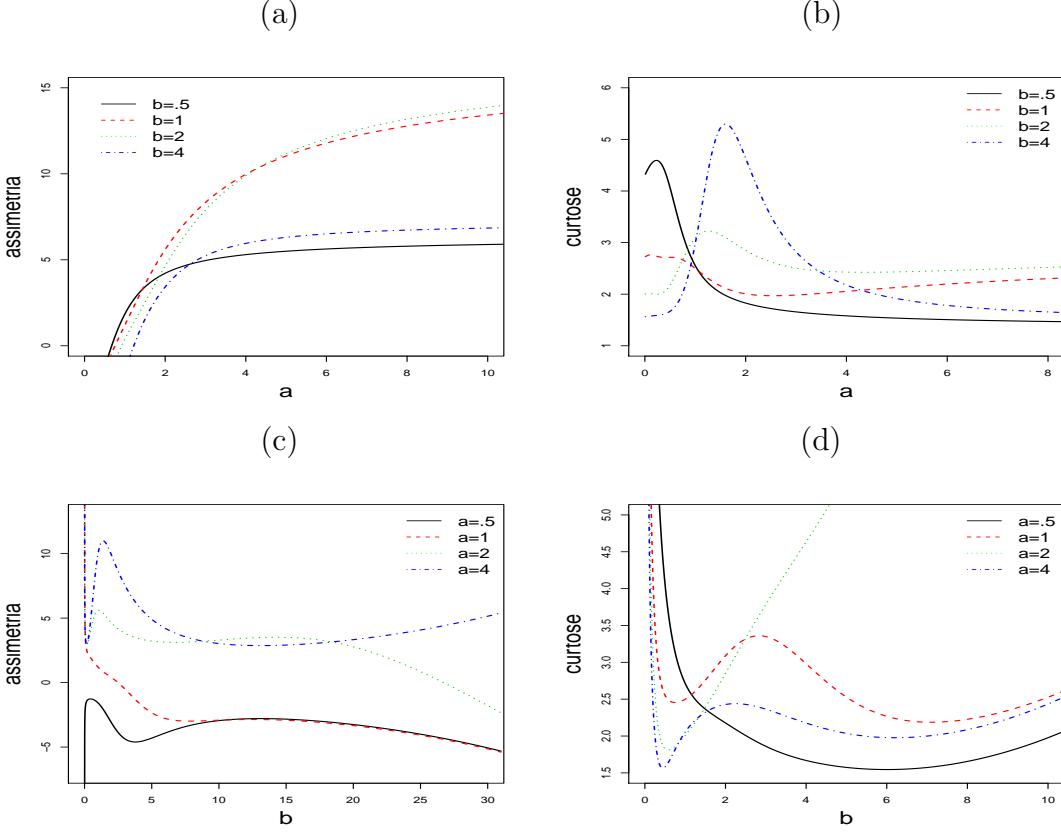


Figura 15 – Gráficos da assimetria e curtose da distribuição Kw-L. (a) e (b) apresenta-se os gráficos da assimetria e curtose como função do parâmetro a , para alguns valores do parâmetro b . (c) e (d) apresenta-se os gráficos da assimetria e curtose como função do parâmetro b , para alguns valores do parâmetro a

A função geradora de momentos (fgm) da Kw-L é dada por

$$\begin{aligned} M(s) &= \int_0^\infty \exp(st) f(t) dt = \int_0^\infty \sum_{k=0}^\infty \frac{(st)^k}{k!} f(t) dt \\ &= \sum_{k=0}^\infty \frac{s^k}{k!} \int_0^\infty t^k f(t) dt = \sum_{k=0}^\infty \frac{\mu'_k s^k}{k!}. \end{aligned}$$

em que μ'_k é dado na equação (45).

4.5 Modelo de regressão

Muitas vezes, no estudo do tempo de falha, podem existir covariáveis relacionadas ao tempo de falha de uma observação, como por exemplo, na indústria, o tempo até a

falha de um determinado equipamento pode ser influenciado pelo nível de voltagem ao qual o equipamento é submetido; e na área médica, o tempo até a morte de um paciente pode estar relacionado com o tipo de tumor ou com a quantidade de hemoglobina presente no sangue do paciente, Silva (2008).

A influência de uma ou mais covariáveis no tempo de sobrevivência pode ser estudada através de modelos de regressão. Existem duas classes de modelos de regressão, os modelos paramétricos, também chamados de modelos de tempo de vida acelerado ou modelos de locação e escala, e os semi-paramétricos, conhecidos como modelos de riscos proporcionais ou modelo de regressão de Cox, Silva (2008). Há uma vasta literatura sobre os modelos de regressão tanto paramétrico quanto semi-paramétrico e podem ser consultadas, por exemplo, em Cox (1972), Lawless (1982), Cox e Oakes (1984), Kalbfleish e Prentice (2002), dentre outros. Neste trabalho será considerado somente o modelo de locação e escala (modelo paramétrico).

4.5.1 Modelo de regressão Kw-logístico

Em muitas aplicações práticas, o tempo de vida t_i é afetado pelas variáveis explanatórias, tais como o nível de colesterol, pressão sanguínea, entre outras. Seja $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ o vetor de variáveis explanatórias associado com a i -ésima variável resposta y_i para $i = 1, \dots, n$.

O modelo de regressão log-linear construído para a variável resposta y_i baseado na fdp da distribuição Kw-L é dado por

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \sigma z_i, \quad i = 1, \dots, n, \quad (46)$$

em que os erros aleatórios z_i seguem distribuição

$$f(z) = ab \frac{1}{1 + \exp(z)} \left[\frac{\exp(z)}{1 + \exp(z)} \right]^a \left\{ 1 - \left[1 - \frac{1}{1 + \exp(z)} \right]^a \right\}^{b-1}, \quad -\infty < z < \infty. \quad (47)$$

e $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$, $\sigma > 0$, $a > 0$ e $b > 0$ são os parâmetros de escala desconhecidos e \mathbf{x}_i é o vetor de variáveis explanatórias modelando o parâmetro de locação $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$. Logo, o vetor de parâmetros de locação $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$ do modelo Kw-L tem uma estrutura

linear $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, onde $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$ é uma matriz de modelo conhecido. O modelo de regressão logístico é definido pela equação (46) com $a = 1$ e $b = 1$.

Na Figura 16 é apresentado os gráficos da distribuição dos erros aleatórios obtidos da equação (47) para alguns valores de a e b . Note que o gráfico de $f(z)$ é unimodal, e para $a = 1$ e $b = 1$ obtém-se a distribuição logística assumindo valor máximo para $f(z)$ em $z = 0$.

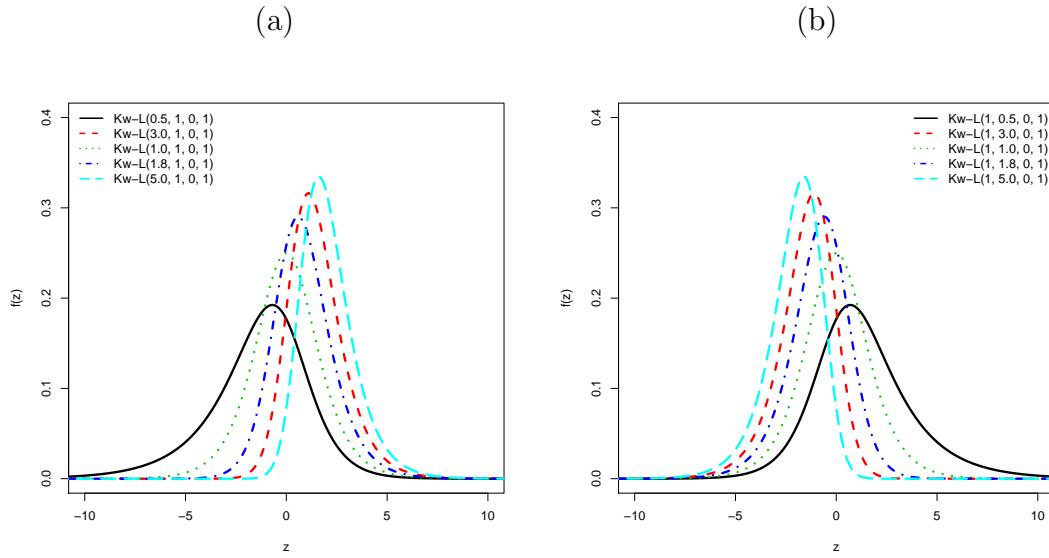


Figura 16 – Gráficos da distribuição dos erros aleatórios para alguns valores dos parâmetros a e b . (a) Apresenta-se os gráficos da fdp da distribuição dos erros aleatórios para $b = 1$ e a assumindo valores 0, 5, 1, 1, 8, 3 e 5. (b) Apresenta-se os gráficos da fdp da distribuição dos erros aleatórios para $a = 1$ e b assumindo valores 0, 5, 1, 1, 8, 3 e 5

4.6 Estimação de máxima verossimilhança

Considere uma amostra $(y_1, \mathbf{x}_1), \dots, (y_n, \mathbf{x}_n)$ de n observações independentes, onde cada resposta aleatória é definida por $y_i = \min\{\log(t_i), \log(c_i)\}$. Assumindo censura não informativa e que os tempo de vida observado e tempo de censura são independentes.

Seja F e C os conjuntos de indivíduos para o qual y_i é o logaritmo do tempo de vida ou o logaritmo do tempo de censura, respectivamente. O logaritmo da função de

verossimilhança para os parâmetros do modelo $\boldsymbol{\theta} = (a, b, \sigma, \boldsymbol{\beta}^T)^T$ é dado por

$$\begin{aligned} l(\boldsymbol{\theta}) &= r \log \left(\frac{ab}{\sigma} \right) + a \sum_{i \in F} z_i - (a+1) \sum_{i \in F} \log (1 + e^{z_i}) + \\ &+ (b-1) \sum_{i \in F} \log \left\{ 1 - \left[1 - \frac{1}{1 + e^{z_i}} \right]^a \right\} + b \sum_{i \in C} \log \left\{ 1 - \left[1 - \frac{1}{1 + e^{z_i}} \right]^a \right\}, \quad (48) \end{aligned}$$

onde $z_i = (y_i - \mathbf{x}_i^T \boldsymbol{\beta})/\sigma$ e r é o número de falhas observadas. O estimador de máxima verossimilhança $\hat{\boldsymbol{\theta}}$ de $\boldsymbol{\theta}$ pode ser obtido pela maximização do logarítmico da função de verossimilhança (48). Utilizou-se o procedimento NLMixed do SAS, para o cálculo das estimativas dos parâmetros. Do modelo ajustado (46), a função de sobrevivência para y_i pode ser estimada por

$$S(y_i; \hat{a}, \hat{b}, \hat{\sigma}, \hat{\boldsymbol{\beta}}^T) = \left\{ 1 - \left[1 - \frac{1}{1 + \exp\left(\frac{y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}}{\hat{\sigma}}\right)} \right]^{\hat{a}} \right\}^{\hat{b}},$$

ou equivalentemente

$$S(t_i; \hat{a}, \hat{b}, \hat{\gamma}, \hat{\boldsymbol{\beta}}^T) = \left\{ 1 - \left[1 - \frac{1}{1 + \left(t_i / \exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}) \right)^{\hat{\gamma}}} \right]^{\hat{a}} \right\}^{\hat{b}}.$$

Sobre condições gerais de regularidade, a distribuição assintótica de $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ é a normal multivariada $N_{p+3}(0, K(\boldsymbol{\theta})^{-1})$, onde $K(\boldsymbol{\theta})$ é a matriz de informação esperada. A matriz de covariância assintótica $K(\boldsymbol{\theta})^{-1}$ de $\hat{\boldsymbol{\theta}}$ pode ser aproximada pela inversa da matriz $(p+3) \times (p+3)$ de informação observada $J(\boldsymbol{\theta})$ e então a inferência assintótica para o vetor de parâmetros $\boldsymbol{\theta}$ pode ser baseada na aproximação normal $N_{p+3}(0, J(\boldsymbol{\theta})^{-1})$ para $\hat{\boldsymbol{\theta}}$. A matriz de informação observada é

$$J(\boldsymbol{\theta}) = \begin{pmatrix} J_{aa} & J_{ab} & J_{a\sigma} & J_{a\beta_j} \\ . & J_{bb} & J_{b\sigma} & J_{b\beta_j} \\ . & . & J_{\sigma\sigma} & J_{\sigma\beta_j} \\ . & . & . & J_{\beta_j\beta_s} \end{pmatrix},$$

cujos elementos são apresentados no Apêndice B.

A distribuição Normal multivariada assintótica $N_{p+3}(0, J(\boldsymbol{\theta})^{-1})$ pode ser usada para construir regiões de confiança aproximada para alguns parâmetros em $\boldsymbol{\theta}$ e para a função de sobrevivência e função de taxa de falha. De fato, um intervalo de confiança assintótico de $100(1 - \alpha)\%$ para cada parâmetro θ_r é dado por

$$ICA_r = \left(\hat{\theta}_r - z_{\alpha/2} \sqrt{-\hat{L}^{r,r}}, \hat{\theta}_r + z_{\alpha/2} \sqrt{-\hat{L}^{r,r}} \right),$$

onde $-\hat{L}^{r,r}$ denota o r -ésimo elemento da diagonal da inversa da matriz de informação observada estimada $J(\hat{\boldsymbol{\theta}})^{-1}$ e $z_{\alpha/2}$ é o quantil $1 - \alpha/2$ da distribuição Normal padrão. A normalidade assintótica é também útil para teste de ajuste de alguns submodelos e para comparar alguns sub-modelos especiais usando a estatística LR.

A interpretação dos coeficientes estimados é baseada na razão de tempos medianos segundo Hosmer Jr. e Lemeshow (1999) válida para variáveis contínuas ou categóricas. Quando a variável explanatória é binária (0 ou 1), e considerando a razão dos tempos medianos com $x = 1$ no numerador, se $\hat{\beta}$ é negativo (positivo), isto implica que os indivíduos com $x = 1$ apresentam reduzido (ampliado) tempo de sobrevivência em $[\exp(\hat{\beta}) \times 100\%]$ quando comparados àqueles indivíduos do grupo $x = 0$, assumindo as outras variáveis explanatórias fixada.

É de interesse investigar se o modelo Kw-L é um bom modelo para ajustar dados sobre investigação. Claramente, a estatística LR pode ser usada para discriminar entre os modelos logístico exponenciado e Kw-L, desde que sejam modelos encaixados. As hipóteses a serem testadas neste caso são $H_0 : b = 1$ versus $H_1 : b \neq 1$, e a estatística LR é reduzida a $w = 2\{l(\hat{\boldsymbol{\theta}}) - l(\tilde{\boldsymbol{\theta}})\}$, onde $\tilde{\boldsymbol{\theta}}$ é o estimador de máxima verossimilhança de $\boldsymbol{\theta}$ sobre H_0 . A hipótese nula é rejeitada se $w > \chi^2_{1-\alpha}(1)$, onde $\chi^2_{1-\alpha}(1)$ é o quantil da distribuição Qui-quadrado com um grau de liberdade.

5 APLICAÇÕES

Neste capítulo foram ajustados dois conjuntos de dados às distribuições Kw-LL e Kw-L comparando-as com as distribuições log-logística e logística respectivamente. Evidenciando a maior flexibilidade das distribuições Kw-LL e Kw-L em comparação às distribuições log-logística e logística, propiciando melhor ajuste ao conjunto de dados.

O primeiro conjunto de dados analisado na seção 5.1, refere-se a dados de tempo até a soro-reversão de 143 crianças expostas ao HIV por via vertical, nascidas no Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto no período entre 1995 a 2001, onde as mães não foram tratadas, retirados de Silva (2004). Os modelos ajustados a esse banco de dados, foram Kw-LL, log-logístico, exponencial, Weibull e log-normal.

O segundo conjunto de dados, analisado na seção 5.2, refere-se ao tempo até a falha de um tipo de isolante elétrico fluido submetido a sete níveis de voltagem constata, descrito em Nelson (1972). Foram analisados $n = 76$ isolantes divididos em sete grupos de voltagem de 26 a 38 kilovolts (kV), como apresentado na Tabela 2.

Os softwares utilizados para estimativa dos parâmetros das distribuições foram o R Development Core Team (2009) e o SAS - Statistical Analysis System (Institute, 2004).

5.1 Aplicação 1: Dados de AIDS

A Aids é uma patologia que mobiliza os portadores pelas implicações sócio afetivas de relacionamento e reprodução. Avanços terapêuticos têm possibilitado à mulher soropositiva direito à reprodução. Neste sentido o ambulatório da Imuno Pediatria do Hospital das clínicas e o Serviço Social de Ribeirão Preto atuam na linha do cuidado dos recém nascidos de mães soropositivas no controle e adesão à terapia antiretroviral de forma a potencializar acesso às informações e orientações, garantindo direitos em busca da emancipação. Neste trabalho considera-se o conjunto de dados referente ao tempo até a soro-reversão de 143 crianças expostas ao HIV por via vertical, nascidas no Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto, entre 1995 a 2001, onde as mães não foram tratadas Silva (2004) e Perdoná (2006). A transmissão por via vertical do vírus HIV pode ocorrer durante a gestação em torno de 35% dos casos, durante o trabalho de parto e o parto propriamente dito,

onde mais ocorre, 65% dos casos, ou na amamentação variando de 7% a 22%. A soro-reversão ou reversão sorológico pode ocorrer em crianças nascidas de mães contaminadas com o vírus HIV, ela é o processo de desaparecimento dos anticorpos anti-HIV do sangue em indivíduo que apresentava infecção pelo HIV. Com o passar dos meses os anticorpos maternos vão sendo eliminados e a sorologia anti HIV deixa de ser positiva. Os recém-nascidos expostos foram monitorados até a definição da condição sorológica, fazendo uso da droga Zidovudina (AZT) nas primeiras 24h de vida e sendo administrada durante 6 semanas.

O gráfico TTT (página 27), apresentado na Figura 17, apresenta uma curva predominantemente côncava, porém com uma pequena parte final abaixada da reta diagonal (curva E) indicando que a função de taxa de falha empírica é unimodal. Logo, os modelos LLM, Kw-LL e log-logístico são indicados ao ajuste dos dados, pois modelam forma unimodal. Os modelos Weibull e log-normal também foram ajustados aos dados em questão e comparados entre si através de comparação gráfica, Figura 19, e dos valores das estatísticas AIC (Critério de Informação Akaike), CAIC (Critério Akaike de Informação Consistente) e BIC (Critério de Informação Bayesiano).

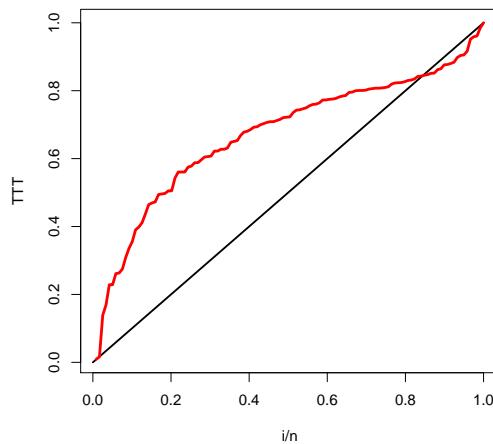


Figura 17 – Gráfico TTT para os dados de soro reversão de 143 crianças expostas ao HIV por via vertical

A Tabela 1 mostra as estimativas de máxima verossimilhança (e os correspondentes erros padrões entre parênteses) dos parâmetros dos modelos e os valores das estatísticas AIC, BIC e CAIC, para os modelos Kw-LL, log-logístico, LLM, Weibull e log-normal. Os cál-

culos foram realizados usando a subrotina NLMIXED do SAS. Estes resultados indicam que o modelo Kw-LL tem os menores valores de AIC, BIC e CAIC entre os valores dos modelos ajustados, e portanto ele é a escolha mais apropriada como melhor modelo.

Tabela 1 – EMV dos parâmetros dos modelos Kw-LL, log-logístico, LLM, Weibull e log-normal para os dados de AIDS, os EP correspondente (em parênteses) e as medidas AIC, BIC e CAIC

Modelo	a	b	α	γ	AIC	CAIC	BIC
Kw-LL	0,08 (0,02)	0,28 (0,07)	499,37 (29,34)	15,46 (3,78)	1616,5	1616,8	1628,4
Log-logístico	1	1	454,87 (20,58)	3,29 (0,26)	1650,4	1650,5	1656,4

Modelo	v	α	γ	AIC	AICC	BIC	
LLM	-	482,17 (0,46)	69,43 (10,82)	3,29 (0,26)	1652,4	1652,6	1661,3

Modelo		α_1	γ_2	AIC	AICC	BIC
Weibull	-	537,46 (21,41)	2,35 (0,16)	1630,6	1630,7	1636,5

Modelo		μ	σ	AIC	AICC	BIC
Log-normal	-	6,03 (0,05)	0,71 (0,05)	1704,1	1704,2	1710,0

Dessa forma, calcula-se se o modelo é apropriado. A Figura 18 apresenta o gráfico da função de taxa de falha estimada onde observa-se forma unimodal da função de taxa de falha da distribuição Kw-LL, confirmando o resultado obtido pelo método gráfico TTT.

A Figura 19 mostra os gráficos da função de sobrevivência empírica e estimada pelas distribuições Kw-LL, LLM, log-logística, Weibull e log-normal. O gráfico à esquerda da Figura 19 apresenta o ajuste pelos modelos Kw-LL, LLM e log-logístico, sendo que o modelo Kw-LL foi o que teve melhor ajuste e o modelo LLM teve ajuste semelhante ao modelo log-logístico. O gráfico a direita da Figura 19 apresenta os ajustes dos modelos Kw-LL, Weibull e log-normal, e novamente o modelo Kw-LL teve melhor ajuste em comparação aos demais, seguido do modelo Weibull.

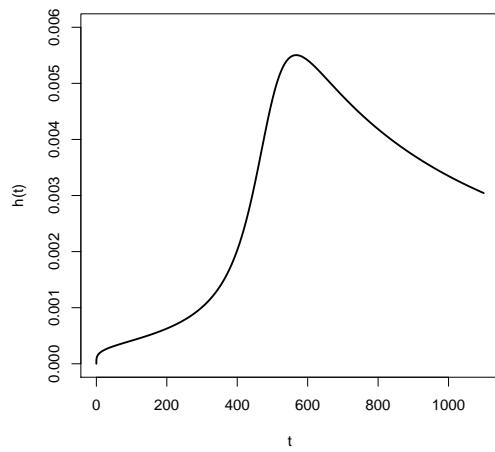


Figura 18 – Gráfico da função de taxa de falha da distribuição Kw-LL para os valores estimados de a , b , α e γ

Os valores numéricos das estatísticas AIC, CAIC e BIC são apresentados na Tabela 1 para os modelos ajustados, os quais comprovam o melhor ajuste do modelo Kw-LL, que obteve os menores valores de AIC, CAIC e BIC. Apesar de visualmente o modelo LLM apresentar ajuste semelhante ao modelo log-logístico, as estatísticas AIC, CAIC e BIC foram maiores em comparação ao modelo log-logístico.

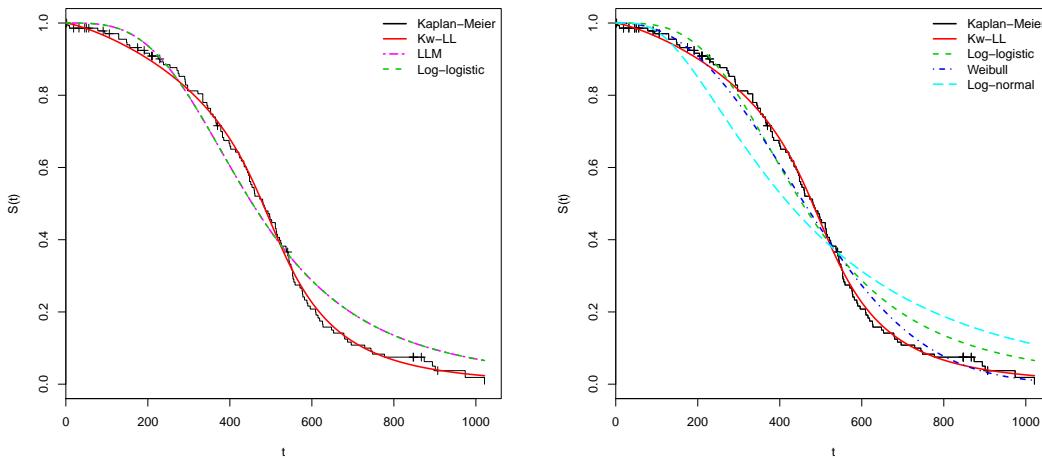


Figura 19 – Funções de sobrevivência estimada e empírica para os dados de AIDS

As expressões das estimativas das funções de sobrevivência para os modelos Kw-LL e log-logístico são respectivamente,

$$83$$

$$\hat{S}_{KLL}(t)=\left\{1-\left[1-\frac{1}{1+(t/499,37)^{15,46}}\right]^{0,08}\right\}^{0,28}$$

$$\hat{S}_{LL}(t) = \frac{1}{1 + (t/3,31)^{456,87}}$$

5.2 Aplicação 2: Dados de níveis de voltagem

Os dados analisados são referentes ao tempo até a falha de um tipo de isolante elétrico fluido submetido a sete níveis de voltagem constante, descrito em Nelson (1972). Foram analisados $n = 76$ isolantes divididos em sete grupos de voltagem de 26 a 38 kilovolts (kV), como apresentado na Tabela 2. O experimento durou tempo suficiente para observar a falha de todos os isolantes elétricos testados.

Tabela 2 – Tempo de falha (em minutos) dos isolantes elétricos em cada um dos sete níveis de voltagens

Nível de Voltagem	Tempo de falha											
26	5,79	1579,52	2323,7									
28	68,85	426,07	110,29	108,29	1067,6							
30	17,05	22,66	21,02	175,88	139,07	144,12	20,46	43,40	194,90			
	47,30	7,74										
32	0,40	82,85	9,88	89,29	215,10	2,75	0,79	15,93	3,91	0,27	0,69	
	100,58	27,80	13,95	53,24								
34	0,96	4,15	0,19	0,78	8,01	31,75	7,35	6,50	8,27	33,91	32,52	3,16
	4,85	2,78	4,67	1,31	12,06	36,71	72,89					
36	1,97	0,59	2,58	1,69	2,71	25,50	0,35	0,99	3,99	3,67	2,07	0,96
	5,35	2,90	13,77									
38	0,47	0,73	1,40	0,74	0,39	1,13	0,09	2,38				

Os dados foram analisados usando os modelos de regressão Kw-L, LM e logístico considerando:

t_i : tempo, em minutos, até a falha do i -ésimo isolante elétrico;

δ_i : indicador de censura da i -ésima observação, $i = 1, 2, \dots, 76$;

x_1 : nível de voltagem em kilovolts (kV);

e a equação

$$y_i = \beta_0 + \beta_1 x_{i1} + \sigma z_i, \quad i = 1, \dots, 76, \quad (49)$$

em que os erros z_1, \dots, z_{76} são variáveis aleatórias independentes com função de densidade (47).

O gráfico TTT, apresentado na Figura 20, foi construído afim de verificar o modelo apropriado para o tempo de sobrevivência desse conjunto de dados. A curva do

gráfico TTT apresenta forma convexa (curva B), indicando que a função de taxa de falha empírica é decrescente. Logo os modelos Kw-L, LM e logístico são indicados ao ajuste dos dados, pois modelam forma decrescente.

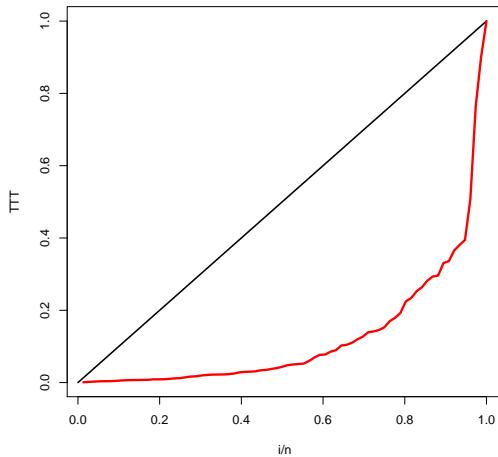


Figura 20 – Gráfico TTT para os dados de níveis de voltagem

A Tabela 3 apresenta as estimativas para os parâmetros de cada modelo com os respectivos erros padrões (E.P.), o valor de p e na base da tabela as estatísticas AIC, CAIC e BIC para ambos os modelos, usando o procedimento NLMIXED do SAS.

Um resumo dos valores AIC, BIC e CAIC para comparar os modelos de regressão logístico, Kw-L e LM é fornecido na Tabela 3. O modelo de regressão Kw-L superou os modelos logístico e LM nos respectivos critérios, apresentando menores valores, e dessa forma pode ser usado na análise desses dados.

A Figura 21 apresenta os gráficos da função de sobrevivência empírica e estimada pelos modelos de regressão Kw-L, LM e logístico, para os níveis de voltagem 28 kV e 34 kV. A Figura 21(a) apresenta o ajuste dos modelos Kw-L e logístico, sendo que o modelo Kw-L obteve melhor ajuste em comparação ao modelo logístico. A Figura 21(b) apresenta o ajuste dos modelos Kw-L e LM e novamente o modelo Kw-L ajustou melhor aos dados.

O modelo final é dado então por $y_i = 23,784 - 0,555x_1 + 1,333z_i$ ou equivalente por $t_i = e^{23,784}e^{-0,555x_{i1}}e^{1,333z_i}$, e $Z \sim \text{Kw-L}(a, b, 0, 1)$.

A função de sobrevivência obtida pelo modelo de regressão Kw-L ajustado para os dados de níveis de voltagem é, portanto, expressa por:

Tabela 3 – EMV dos parâmetros para os modelos de regressão logístico, Kw-logístico e LM ajustados aos dados de níveis de voltagem

Modelo	a	b	σ	β_0	β_1
Kw-L	1,145	10,020	1,333	23,784	-0,555
EP	1,379	25,089	1,347	3,101	0,049
p-valor	-	-	-	<0,001	<0,001
AIC = 284,8	CAIC = 285,6	BIC = 296,4			
Modelo	a	b	σ	β_0	β_1
Logístico	1	1	0,858	19,869	-0,531
EP	-	-	0,082	1,833	0,054
p-valor	-	-	-	<0,001	<0,001
AIC = 286,7	CAIC = 287,0	BIC = 293,7			
Modelo	v	σ	β_0	β_1	
LM	11,712	0,858	17,757	-0,531	
EP	692,95	0,082	50,823	0,054	
p-valor	-	-	0,727	<0,001	
AIC = 288,7	CAIC = 289,3	BIC = 298,0			

$$\hat{S}(t) = \left\{ 1 - \left[1 - \frac{1}{1 + (t/\exp(23,784 - 0,555x_1))^{1/1,333}} \right]^{1,145} \right\}^{10,020}, \quad t \geq 0 \quad (50)$$

Note, a partir de (50), que $\hat{\beta}_1$ é negativo, o que implica que quanto maior o valor de x_1 , menor a probabilidade de sobrevivência estimada. Este fato pode ser claramente observado na Figura 22, em que as curvas de sobrevivência estimadas para dois níveis de voltagem, um com $x_1 = 34$ e outro com $x_1 = 36$, são apresentadas.

A partir da Figura 22 pode-se, ainda, observar que $\hat{S}(20|x_1 = 34) = 0,191$, o que significa que em torno de 20% dos isolantes elétricos submetidos a uma voltagem de 34 Kilovolts, não falham até o tempo $t = 20$ minutos. Por outro lado, estima-se, para isolantes elétricos submetidos a uma voltagem de 36 Kilovolts, que em torno de 2,5% deles não falham até o vigésimo minuto, visto que $\hat{S}(20|x_1 = 36) = 0,025$.

Dos resultados apresentados, verificou-se, portanto, que o modelo de regressão Kw-L ajustou-se satisfatoriamente aos dados de tempos de falha de isolantes elétricos. De modo geral pode-se ainda concluir, que o tempo de falha estimado dos isolantes elétricos diminui à medida que são submetidos ao um nível de voltagem maior.

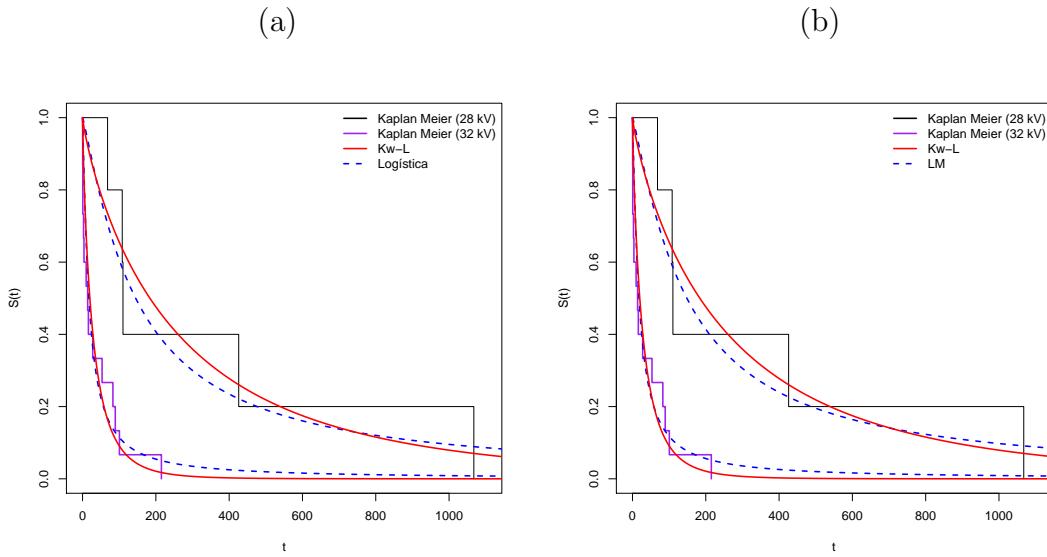


Figura 21 – Curvas de sobrevivência estimadas pelos modelos de regressão Kw-L, LM e logístico para dois níveis de voltagem 28 kV e 32 kV. (a) Apresenta-se os ajustes dos modelos Kw-L e logístico. (b) Apresenta-se os ajustes para os modelos Kw-L e LM

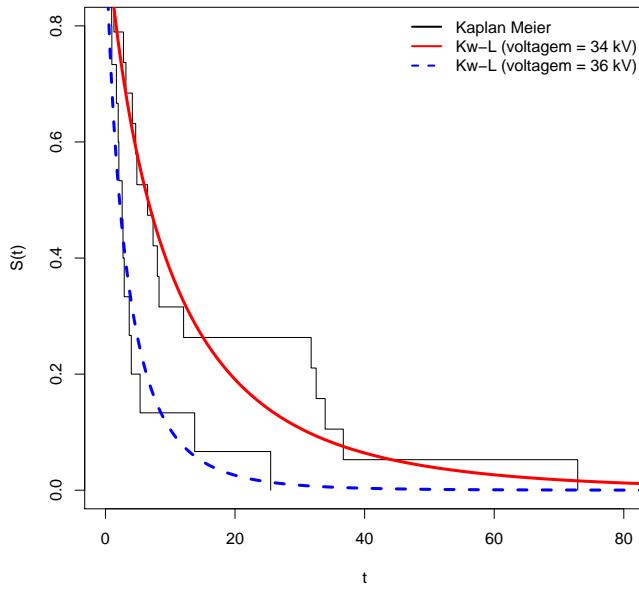


Figura 22 – Curvas de sobrevivência estimadas pelo modelo de regressão Kw-L para os dados de níveis de voltagem

6 CONCLUSÕES

Estudou-se a distribuição de tempo de falha com quatro parâmetros, denominada “distribuição Kumaraswamy-log-logística (Kw-LL)”, que é uma simples extensão da distribuição log-logística. O novo modelo compreende várias distribuições amplamente usadas na literatura de tempo de vida e é mais flexível que as distribuições log-logística exponenciada, Burr XII e log-logística. A distribuição proposta acomoda formas da função de taxa de falha crescente, decrescente, unimodal e forma de banheira. Ela é muito versátil para modelar dados de tempo de vida com função de taxa de falha na forma de banheira e também modela uma variedade de situações incertas.

Baseado nesta nova distribuição, propôs-se o modelo de regressão Kumaraswamy-logístico, que é muito adequado para modelar dados de tempo de vida censurados e não censurados. O novo modelo de regressão permite testar a qualidade do ajuste de alguns modelos conhecidos como sub-modelos especiais. Portanto, o modelo de regressão proposto serve como uma boa alternativa para análise de dados de tempo de vida. O procedimento NLMIXED do SAS foi usado para obter as estimativas de máxima verossimilhança e realizar testes assintóticos para os parâmetros baseado na distribuição assintótica destas estimativas. Foi demonstrado em uma aplicação real que as distribuições Kw-LL e Kw-L podem produzir melhores ajustes que seus sub-modelos.

6.1 Pesquisas futuras

Dando continuidade a esta pesquisa, pretende-se realizar um estudo computacional comparando a eficiência de diferentes métodos de estimação como, por exemplo, bootstrap, jackknife e Bayesiano para o modelo de regressão proposto para obter um melhor desempenho no ajuste do modelo de regressão proposto.

Outra pesquisa será propor medidas de diagnóstico baseados na teoria de influência local e global para os modelos de regressão.

Outra linha de pesquisa ainda seria a proposta de resíduos para esses modelos de regressão e mediante simulações tentar estudar a distribuição empírica desses resíduos.

REFERÊNCIAS

- AARSET, M. V. How to identify bathtub hazard rate. **Transactions on Reliability**, New York, v.36, p.106–108, 1987.
- AHMAD, M. I.; SINCLAIR, C. D.; WERRITTY, A. Log-Logistic Flood Frequency Analysis. **Journal of Hydrology**, New York, v.98, p.205–224, 1988.
- ALKASASBEH, M. R.; RAQAB, M. Z. Estimation of the generalized logistic distribution parameters: Comparative study. **Statistical Methodology**, Oxford, v.6, p.262–279, 2009.
- ASHKAR, F.; MAHDI, S. Fitting the log-logistic distribution by generalized moments. **Journal of Hydrology**, New York, v.328, p.694–703, 2006.
- BALAKRISHNAN, N. **Hand Book of the Logistic Distribution**. New York: Dekker, 1992. 250p.
- BALAKRISHNAN, N.; LENG, M. Y. Order statistics from the type I generalized logistic distribution. **Communications in Statistical - Simulation and Computation**, London, v.17, n.1, p.25–50, 1988.
- BARRETO-SOUZA, W.; SANTOS, A. H. S.; CORDEIRO, G. M. The beta generalized exponential distribution. **Journal of Statistical Computation and Simulation**, London, p.1–14, 2009.
- BARROS, A. A. A. **Distribuições Exponencializadas e Estendidas**, 2008. 75p. Dissertação (Mestrado em Biometria e Estatística Aplicada) – Universidade Federal Rural de Pernambuco, Pernambuco, 2008.
- CARRASCO, J. M. F.; ORTEGA, E. M. M.; CORDEIRO, G. M. A generalized modified Weibull distribution for lifetime modeling. **Computacional Statistics and Data Analysis**, New York, v.53, p.450–462, 2008.

COLOSIMO, E. A.; GIOLO, S. R. **Análise de sobrevivência aplicada.** São Paulo: Edgard Blücher, 2006. 370p.

COORAY, K.; ANANDA, M. M. A. A generalization of the half-normal distribution with applications to lifetime data. **Communication in Statistical Theory and Methods**, v.37, p.1323–1337, 2008.

CORDEIRO, G. M.; CASTRO, M. A new family of generalized distributions. **Journal of Statistical Computation and Simulation**, London, 2010.

CORDEIRO, G. M.; ORTEGA, E. M. M.; SILVA, G. O. The Exponentiated Generalized Gamma Distribution with Application to Lifetime Data. **Jornal of Statistical Computation and Simulation**, London, 2010. In press.

COX, D. R. Regression models and life-tables (with discussion). **Journal of the Royal Statistical Society**, Oxford, v.34, n.2, p.187–220, 1972.

COX, D. R.; OAKES, D. **Analysis of survival data.** , 1984. 200p.

EUGENE, N.; LEE, C.; FAMOYE, F. Beta-Normal distribution and its applications. **Communications in Statistics - Theory and Methods**, New York, v.31, p.497–512, 2002.

GEORGE, E. O.; OJO, M. O. On a generalization of the logistic distribution. **Annals of Statistical Mathematics**, Hayward, v.32, n.2, p.161–169, 1980.

GHITANY, M. E.; AL-AWADHI, F. A.; ALKHALFAN, L. A. Marshall-Olkin Extended Lomax Distribution and Its Application to Censored Data. **Communications in Statistics - Theory and Methods**, New York, v.36, p.1855–1866, 2007.

GUPTA, R. C.; GUPTA, P. L.; GUPTA, R. D. Modeling failure time data by Lehman alternatives. **Communications in Statistics - Theory and Methods**, New York, v.27, p.887–904, 1998.

GUPTA, R. D.; KUNDU, D. Generalized Exponential Distributions. **Australian & New Zealand Journal of Statistics**, Oxford, v.41, n.2, p.173–188, 1999.

GUPTA, R. D.; KUNDU, D. Exponentiated Exponential Family: An Alternative to Gamma and Weibull Distributions. **Biometrical Journal**, Berlin, v.43, n.1, p.117–130, 2001.

HOSMER JR., D. W.; LEMESHOW, S. **Applied survival analysis:** regression modeling of time to event data. New York: John Wiley and Sons, 1999. 386p.

INSTITUTE, S. **SAS/STAT User's Guide:** Version 9. Cary: SAS Institute, 2004.

KALBFLEISH, J. D.; PRENTICE, R. L. **The Statistical analysis of Failure Time Data.** 2nd ed. New York: John Wiley, 2002. 439p.

KIRSHENBAUM, A. P.; OLSEN, D. M.; BICKEL, W. K. A quantitative review of the ubiquitous relapse curve. **Journal of Substance Abuse Treatment**, v.36, p.8–17, 2009.

KLEIBER, C.; KOTZ, S. **Statistical Size Distributions in Economics and Actuarial Sciences.** John Wiley Professio, 2003.

KUMARASWAMY, P. A generalized probability density function for double-bounded random processes. **Journal of Hydrology**, New York, 46, , n.46, p.79–88, 1980.

LAWLESS, J. F. **Statistical models and methods for lifetime data.** New York: Wiley, 1982. 630p.

LEE, E. T.; WANG, J. W. **Statistical methods for survival data analysis.** 3rd ed. Oklahoma: Wiley - Interscience, 2003. 527p.

LEROY, R.; BOGAERTS, K.; LESAFFRE, E.; DECLERCK, D. The emergence of permanent teeth in Flemish children. **Community Dentistry and Oral Epidemiology**, v.31, p.30–39, 2003.

LINHART, H.; ZUCCHINI, W. **Model Selection.** New York: Wiley, 1986. 300p.

MARSHALL, A. W.; OLKIN, I. A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. **Biometrika**, London, v.3, n.84, p.641–652, 1997.

MIYAZAKI, K.; HOSHINO, T. A bayesian semiparametric item response model with Dirichlet process priors. **Psychometrika**, Nagoya, v.74, n.3, p.375–393, 2009.

MUDHOLKAR, G. S.; SRIVASTAVA, D. K.; FREIMER, M. The Exponentiated Weibull Family: A Reanalysis of the Bus-Motor-Failure Data. **Technometrics**, Alexandria, v.37, p.436–445, 1995.

NADARAJAH, S. The exponential Gumbel distribution with climate application. **Environmetrics**, New York, v.17, p.13–23, 2006.

NADARAJAH, S. Exact Distribution of the Product of Two or More Logistic Random Variables. **Methodol Comput Appl Probab**, Manchester, v.11, p.651–660, 2009a.

NADARAJAH, S. The skew logistic distribution. **AStA Advances in Statistical Analysis**, Verlag, v.93, p.187–203, 2009b.

NADARAJAH, S.; KOTZ, S. The beta Gumbel distribution. **Mathematical Problems in Engineering**, New York, v.10, p.323–332, 2004.

NADARAJAH, S.; KOTZ, S. The Beta Exponential distribution. **Reliability Engineering and System Safety**, Barking, v.91, p.689–697, 2005.

NADARAJAH, S.; KOTZ, S. The Exponential Type Distributions. **Acta Applicandae Mathematicae: An International Survey Journal on Applying Mathematics and Mathematical Applications**, Amsterdam, v.92, n.2, p.97–111, 2006.

NELSON, W. B. Graphical analysis of accelerated life test data with the inverse power law model. **IEEE Transactions on Reliability**, New York, p.2–11, 1972.

OJO, M. O.; OLAPADE, A. K. On the generalized Logistic and Log-logistic distributions. **Kragujevac Journal Mathematics**, Ile-Ife, v.25, p.65–73, 2003.

PASCOA, M. A. R.; ORTEGA, E. M. M.; CORDEIRO, G. M. The Kw-Generalized Gamma Distribution with Application in Survival Analysis. **Statistical Methodology**, Oxford, 2010. In press.

PERDONÁ, G. S. C. **Modelos de riscos aplicados à análise de sobrevivência**, 2006. 148p. Tese (Doutorado em Ciências - Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemática e de Computação, Universidade de São Paulo, São Carlos, 2006.

PESCIM, R. R.; DEMÉTRIO, C. G. B.; CORDEIRO, G. M.; ORTEGA, E. M. M.; URBANO, M. R. The Beta Generalized Half-Normal Distribuiton. **Computacional Statistics and Data Analysis**, New York, 2010.

R DEVELOPMENT CORE TEAM. R: a language and environment for statistical computing. R Foundation for Statistical Computing, 2009.

RAO, G. S.; GHITANY, M. E.; KANTAM, R. R. L. Reliability Test Plans for Marshall-Olkin Extended Exponential Distribution. **Applied Mathematical Sciences**, New York, v.3, n.55, p.2745–2755, 2009.

SILVA, A. N. F. **Estudo evolutivo das crianças expostas ao HIV e notificadas pelo núcleo de vigilância epidemiológica do HCFMRP-USP**, 2004. 65p. Dissertação (Mestrado em Saúde na Comunidade) – Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2004.

SILVA, G. O. **Modelos de regressão quando a função de taxa de falha não é monótona e o modelo probabilístico beta Weibull modificada**, 2008. 207p. Tese (Doutorado em Agronomia - Estatística e Experimentação Agronômica) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2008.

STAMATOPOULOS, G.; SENGUPTA, A.; VOGEL, E.; JANSON, C. A game-theoretic model of coalition formation among primates. **Journal of Bioeconomics**, v.11, p.165–183, 2009.

THOMAS, A.; JOSE, K. K. Bivariate semi-Pareto minification processes. **Metrika**, London, v.59, p.305–313, 2004.

ZHANG, T.; XIE, M. Failure Data Analysis with Extended Weibull Distribution. **Communications in Statistical - Simulation and Computation**, London, v.36, n.3, p.579–592, 2007.

APÊNDICE

Apêndice A

As derivadas parciais de segunda ordem de $l(a, b, \alpha, \gamma)$ são dadas a seguir.

$$\begin{aligned}
\frac{\partial^2 l(\theta)}{\partial a^2} &= -\frac{r}{a^2} - (b-1) \sum_{i \in F} h_i \log^2(u_i) - b \sum_{i \in F} h_i \log^2(u_i) \\
\frac{\partial^2 l(\theta)}{\partial b^2} &= -\frac{r}{a^2} \\
\frac{\partial^2 l(\theta)}{\partial \alpha^2} &= \frac{r}{\alpha^2} - \frac{(\gamma-1)r}{\alpha^{r+1}} - \frac{(a+1)\gamma}{\alpha} \sum_{i \in F} \left\{ \frac{(\gamma-1)}{\alpha} \left(\frac{t_i}{\alpha} \right)^{-\gamma} \frac{1}{u_i} - \frac{\gamma}{\alpha} \left(\frac{t_i}{\alpha} \right)^{-2\gamma} \frac{1}{u_i^2} \right\} - \\
&\quad - \frac{(b-1)a\gamma^2}{\alpha} \sum_{i \in F} u_i^{-a} \left(\frac{t_i}{\alpha} \right)^{-\gamma} h_i + \frac{ab\gamma^2}{\alpha} \sum_{i \in C} u_i^{-a} \left(\frac{t_i}{\alpha} \right)^{-\gamma} h_i \\
\frac{\partial^2 l(\theta)}{\partial \gamma^2} &= -\frac{r}{\alpha^2} + (a+1) \sum_{i \in F} \left\{ -\gamma \left(\frac{t_i}{\alpha} \right)^{-\gamma} \log \left(\frac{t_i}{\alpha} \right) \frac{1}{u_i} + \left(\frac{t_i}{\alpha} \right)^{-2\gamma} \log^2 \left(\frac{t_i}{\alpha} \right) \frac{1}{u_i^2} \right\} + \\
&\quad + (b-1)a \sum_{i \in F} \gamma \left(\frac{t_i}{\alpha} \right)^{-\gamma} \log \left(\frac{t_i}{\alpha} \right) h_i (1 - u_i^{-a}) + \\
&\quad + (b-1)a \sum_{i \in F} \left(\frac{t_i}{\alpha} \right)^{-2\gamma} \log^2 \left(\frac{t_i}{\alpha} \right) \frac{h_i}{u_i^2} (u_i^{-a} - a - 1) + \\
&\quad + ab \sum_{i \in C} \gamma \left(\frac{t_i}{\alpha} \right)^{-\gamma} \log \left(\frac{t_i}{\alpha} \right) h_i (1 - u_i^{-a}) + \\
&\quad + ab \sum_{i \in C} \left(\frac{t_i}{\alpha} \right)^{-2\gamma} \log^2 \left(\frac{t_i}{\alpha} \right) \frac{h_i}{u_i^2} (u_i^{-a} - a - 1)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 l(\theta)}{\partial b \partial a} &= \sum_{i \in F} h_i (1 - u_i^{-a}) \log(u_i) + \sum_{i \in C} h_i (1 - u_i^{-a}) \log(u_i) \\
\frac{\partial^2 l(\theta)}{\partial \alpha \partial a} &= \frac{\gamma}{\alpha} \sum_{i \in F} \left(\frac{t_i}{\alpha} \right)^{-\gamma} \frac{1}{u_i} - (b-1) \sum_{i \in F} \frac{\gamma}{\alpha} \left(\frac{t_i}{\alpha} \right)^{-\gamma} (1 - u_i^{-a}) \frac{h_i}{u_i} [1 + a \log(u_i)] - \\
&\quad - b \sum_{i \in C} \frac{\gamma}{\alpha} \left(\frac{t_i}{\alpha} \right)^{-\gamma} (1 - u_i^{-a}) \frac{h_i}{u_i} [1 + a \log(u_i)] \\
\frac{\partial^2 l(\theta)}{\partial \gamma \partial a} &= \sum_{i \in F} \left(\frac{t_i}{\alpha} \right)^{-\gamma} \log \left(\frac{t_i}{\alpha} \right) \frac{1}{u_i} + (b-1) \sum_{i \in F} a \left(\frac{t_i}{\alpha} \right)^{-\gamma} \log \left(\frac{t_i}{\alpha} \right) \frac{h_i}{u_i} \log(u_i) + \\
&\quad + (b-1) \sum_{i \in F} \left(\frac{t_i}{\alpha} \right)^{-\gamma} \log \left(\frac{t_i}{\alpha} \right) (1 - u_i^{-a}) \frac{h_i}{u_i} + \\
&\quad + b \sum_{i \in C} \left(\frac{t_i}{\alpha} \right)^{-\gamma} \log \left(\frac{t_i}{\alpha} \right) \frac{h_i}{u_i} [a \log(u_i) - u_i^{-a} + 1] \\
\frac{\partial^2 l(\theta)}{\partial \alpha \partial b} &= \sum_{i \in F} \frac{a\gamma}{\alpha} \left(\frac{t_i}{\alpha} \right)^{-\gamma} (1 - u_i^{-a}) \frac{h_i}{u_i} + \sum_{i \in C} \frac{a\gamma}{\alpha} \left(\frac{t_i}{\alpha} \right)^{-\gamma} (1 - u_i^{-a}) \frac{h_i}{u_i} \\
\frac{\partial^2 l(\theta)}{\partial \gamma \partial b} &= -a \sum_{i \in F} \left(\frac{t_i}{\alpha} \right)^{-\gamma} \log \left(\frac{t_i}{\alpha} \right) (1 - u_i^{-a}) \frac{h_i}{u_i} - a \sum_{i \in C} \left(\frac{t_i}{\alpha} \right)^{-\gamma} \log \left(\frac{t_i}{\alpha} \right) (1 - u_i^{-a}) \frac{h_i}{u_i} \\
\frac{\partial^2 l(\theta)}{\partial \gamma \partial \alpha} &= -\frac{a+1}{\alpha} \sum_{i \in F} \left\{ \frac{1}{u_i} - \gamma^2 \left(\frac{t_i}{\alpha} \right)^{-1} \frac{1}{u_i} + \gamma \left(\frac{t_i}{\alpha} \right)^{-\gamma} \log \left(\frac{t_i}{\alpha} \right) \frac{1}{u_i^2} \right\} \left(\frac{t_i}{\alpha} \right)^{-\gamma} + \\
&\quad + \frac{(b-1)a}{\alpha} \sum_{i \in F} \left[1 - \gamma^2 \left(\frac{t_i}{\alpha} \right)^{-1} \right] (1 - u_i^{-a}) \frac{h_i}{u_i} \left(\frac{t_i}{\alpha} \right)^{-\gamma} - \\
&\quad - \frac{(b-1)a}{\alpha} \sum_{i \in F} \gamma \left(\frac{t_i}{\alpha} \right)^{-2\gamma} \log \left(\frac{t_i}{\alpha} \right) h_i [u_i^{-a-2} - (a+1)u_i^{-2}] + \\
&\quad + \frac{ba}{\alpha} \sum_{i \in F} \left[1 - \gamma^2 \left(\frac{t_i}{\alpha} \right)^{-1} \right] (1 - u_i^{-a}) \frac{h_i}{u_i} \left(\frac{t_i}{\alpha} \right)^{-\gamma} - \\
&\quad - \frac{(b-1)a}{\alpha} \sum_{i \in F} \gamma \left(\frac{t_i}{\alpha} \right)^{-2\gamma} \log \left(\frac{t_i}{\alpha} \right) h_i [u_i^{-a-2} - (a+1)u_i^{-2}]
\end{aligned}$$

Apêndice B

Os estimadores de máxima verossimilhança para os parâmetros a , b , σ e β são obtidos do sistema das equações

$$\frac{\partial l(\boldsymbol{\theta})}{\partial a} = \frac{r}{a} - \sum_{i \in F} \log(1 + e^{-z_i}) + (b-1) \sum_{i \in F} q_i \log(1 + e^{-z_i}) + \sum_{i \in C} q_i \log(1 + e^{-z_i}) = 0$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial b} = \frac{r}{b} + \sum_{i=1}^n \log \left[1 - (1 + e^{-z_i})^{-a} \right] = 0$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \sigma} = -\frac{r}{\sigma} + \frac{1}{\sigma} \sum_{i \in F} z_i - \frac{(a+1)}{\sigma} \sum_{i \in F} z_i r_i + \frac{a(b-1)}{\sigma} \sum_{i \in F} z_i r_i q_i + \frac{ab}{\sigma} \sum_{i \in C} z_i r_i q_i = 0$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \beta_l} = \frac{1}{\sigma} \sum_{i \in F} x_{il} - \frac{(a+1)}{\sigma} \sum_{i \in F} x_{il} r_i + \frac{a(b-1)}{\sigma} \sum_{i \in F} x_{il} r_i q_i + \frac{ab}{\sigma} \sum_{i \in C} x_{il} r_i + q_i = 0$$

onde $q_i = \frac{(1+e^{-z_i})^{-a}}{1-(1+e^{-z_i})^{-a}}$ e $r_i = \frac{e^{-z_i}}{1+e^{-z_i}}$.

As derivadas de segunda ordem do logarítmico da função de verossimilhança $l(a, b, \sigma, \beta)$ são dadas a seguir:

$$\frac{\partial^2 l}{\partial b^2} = -\frac{r}{b^2}$$

$$\frac{\partial^2 l}{\partial a \partial b} = \sum_{i=1}^n q_i \log [1 + e^{-z_i}]$$

$$\frac{\partial^2 l}{\partial \sigma \partial b} = \frac{a}{\sigma} \sum_{i=1}^n z_i r_i q_i$$

$$\frac{\partial^2 l}{\partial \beta_l \partial b} = \frac{a}{\sigma} \sum_{i=1}^n x_{il} r_i q_i$$

$$\begin{aligned}
\frac{\partial^2 l}{\partial a^2} &= -\frac{r}{a^2} - 2(b-1) \sum_{i \in F} \log(u_i)p_i - 2b \sum_{i \in C} \log(u_i)p_i \\
\frac{\partial^2 l}{\partial \sigma \partial a} &= -\frac{1}{\sigma} \sum_{i \in F} z_i r_i + \frac{b-1}{\sigma} \sum_{i \in F} z_i r_i q_i - \frac{az_i p_i}{u_i} \log(u_i) + \frac{b}{\sigma} \sum_{i \in C} z_i r_i q_i - \frac{az_i p_i}{u_i} \log(u_i) \\
\frac{\partial^2 l}{\partial \beta_l \partial a} &= -\frac{1}{\sigma} \sum_{i \in F} x_{il} r_i + \frac{b-1}{\sigma} \sum_{i \in F} \left\{ x_{il} r_i q_i - \frac{ax_{il} p_i}{u_i} \log(u_i) \right\} + \frac{b}{\sigma} \sum_{i \in F} \left\{ x_{il} r_i q_i - \frac{ax_{il} p_i}{u_i} \log(u_i) \right\} \\
\frac{\partial^2 l}{\partial \sigma^2} &= \frac{r}{\sigma^2} - \frac{1}{\sigma^2} \sum_{i \in F} z_i - \frac{a+1}{\sigma^2} \sum_{i \in F} \left(\frac{z_i}{u_i} - 1 \right) z_i r_i + \frac{a(b-1)}{\sigma^2} \sum_{i \in F} \left[\left(\frac{z_i}{u_i} - 1 \right) q_i - \frac{az_i p_i}{u_i} \right] z_i r_i + \\
&+ \frac{ab}{\sigma^2} \sum_{i \in C} \left[\left(\frac{z_i}{u_i} - 1 \right) q_i - \frac{az_i p_i}{u_i} \right] z_i r_i \\
\frac{\partial^2 l}{\partial \beta_l \partial \sigma} &= -\frac{1}{\sigma^2} \sum_{i \in F} x_{il} - \frac{a+1}{\sigma^2} \sum_{i \in F} \left(\frac{z_i}{u_i} - 1 \right) r_i x_{il} + \frac{a(b-1)}{\sigma^2} \sum_{i \in F} \left[\left(\frac{z_i}{u_i} - 1 \right) q_i - \frac{az_i p_i}{u_i} \right] x_{il} r_i + \\
&+ \frac{ab}{\sigma^2} \sum_{i \in C} \left[\left(\frac{z_i}{u_i} - 1 \right) q_i - \frac{az_i p_i}{u_i} \right] x_{il} r_i \\
\frac{\partial^2 l}{\partial \beta_s \partial \beta_l} &= -\frac{a+1}{\sigma^2} \sum_{i \in F} \frac{x_{il} x_{is} r_i}{u_i} + \frac{a(b-1)}{\sigma^2} \sum_{i \in F} \frac{x_{il} x_{is} r_i}{u_i} \left(1 - \frac{ap_i}{u_i} \right) + \frac{ab}{\sigma^2} \sum_{i \in C} \frac{x_{il} x_{is} r_i}{u_i} \left(1 - \frac{ap_i}{u_i} \right)
\end{aligned}$$

onde $p_i = \frac{(1+e^{-z_i})^{-a}}{[1-(1+e^{-z_i})^{-a}]^2}$ e $u_i = 1 + e^{-z_i}$.