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

**O modelo de regressão odd log-logística gama generalizada com
aplicações em análise de sobrevivência**

Fábio Prata

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

**Piracicaba
2017**

Fábio PrataViera
Bacharel em Estatística

**O modelo de regressão odd log-logística gama generalizada com
aplicações em análise de sobrevivência**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:

Prof. Dr. **EDWIN MOISES MARCOS ORTEGA**

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

Piracicaba
2017

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

Prataviera, Fábio

O modelo de regressão odd log-logística gama generalizada com aplicações em análise de sobrevivência / Fábio Prataviera. -- versão revisada de acordo com a resolução CoPGr 6018 de 2011. -- Piracicaba, 2017 .

90 p.

Dissertação (Mestrado) -- USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. Análise de sobrevivência
2. Censura aleatória
3. Distribuição odd log-logística gama generalizada
4. Distribuição gama generalizada
5. Simulação
6. Modelo de regressão log-odd log-logística gama generalizada I. Título.

DEDICATÓRIA

Aos meus pais,

João Batista Prativiera e Zilda Aparecida Tolon Prativiera, por todo o amor, o carinho, a paciência e o apoio dedicados a mim.

Ao meu irmão,

Marcelo Henrique Prativiera, pela amizade, risadas, paciência e pela confiança em mim.

Ao meu tio,

Gilberto Aparecido Prativiera pela amizade, pelos ensinamentos e por todo apoio dedicados a mim.

A eles,

dedico este trabalho.

AGRADECIMENTOS

À Deus, pela força, para que eu pudesse prosseguir os meus estudos com perseverança e sabedoria.

Ao meu orientador Prof. Dr. Edwin Moises Marcos Ortega, pela compreensão e pela orientação no desenvolvimento deste trabalho.

À minha Professora de graduação da UFSCar, Teresa Cristina Martins Dias, pela orientação, pelos conselhos e incentivo.

Aos Professores Prof. Dr. Gauss Cordeiro, Prof. Dr. Adriano Kamimura Suzuki e Prof. Dr. Altemir da Silva Braga por toda contribuição no desenvolvimento deste trabalho.

À CAPES - pela bolsa de mestrado concedida.

À todos os professores do curso de Pós-graduação em Estatística e Experimentação Agronômica, pelos ensinamentos e possibilidades que contribuíram para minha formação acadêmica e em especial, aos professores, Profa. Sônia Maria De Stefano Piedade, Profa. Roseli Aparecida Leandro e Prof. Carlos Tadeu Dos Santos Dias.

Às secretárias do Departamento de Ciências Exatas, Luciane Brajão e Solange de Assis Paes Sabadin e aos técnicos de informática, Eduardo Bonilha e Jorge Alexandre Wiendl, pela ajuda e boa vontade de vocês.

À todos os amigos dos cursos de mestrado e doutorado do Programa de Pós-graduação em Estatística e Experimentação Agronômica, pelos momentos de estudos, descontração, a atenção e amizade.

Enfim, a todos que me ajudaram de forma direta ou indireta para o desenvolvimento deste trabalho.

SUMÁRIO

RESUMO	7
ABSTRACT	8
LISTA DE FIGURAS	10
LISTA DE TABELAS	11
1 INTRODUÇÃO	13
2 REVISÃO BIBLIOGRÁFICA	17
2.1 Conceitos básicos em análise de sobrevivência	17
2.1.1 Censura	18
2.1.2 Gráfico TTT-plot	19
2.1.3 A função de verossimilhança em análise de sobrevivência	20
2.2 Distribuição gama generalizada	20
2.2.1 Propriedades e medidas de interesse	21
2.3 Modelo de regressão	22
2.3.1 Modelo de regressão locação-escala	23
2.4 Inferência estatística	23
2.4.1 Método de máxima verossimilhança	24
2.5 Estatísticas AIC, BIC e CAIC	24
3 DISTRIBUIÇÃO ODD LOG-LOGÍSTICA GAMA GENERALIZADA	27
3.1 Introdução	27
3.2 Distribuição odd log-logística gama generalizada	29
3.3 Representação linear para a distribuição OLLGG	30
3.4 Propriedades matemáticas	33
3.5 Estimação via método de máxima verossimilhança com dados censurados	36
3.6 Estudo via simulação	37
3.6.1 Estudo via simulação sem censura	37
3.6.2 Estudo via simulação considerando censura aleatória	38
3.7 Aplicações	39
3.7.1 Aplicação 1	40
3.7.2 Aplicação 2	43
3.8 Conclusões	46
4 DISTRIBUIÇÃO LOG-ODD LOG-LOGÍSTICA GAMA GENERALIZADA	49
4.1 Introdução	49
4.2 Distribuição odd log-logística gama generalizada estendida	50
4.3 Extensões úteis e propriedades estruturais	53
4.3.1 Momentos ordinários	56
4.3.2 Momentos incompletos	58
4.3.3 Função geradora de momentos	59

4.3.4	Desvios médios	60
4.4	Distribuição log-odd log-logística gama generalizada	60
4.5	Modelo de regressão log-odd log-logística gama generalizada	62
4.5.1	Estimação via método de máxima verossimilhança com dados cen- surados	63
4.5.2	Estudo via simulação	65
4.6	Análise de resíduos	66
4.6.1	Estudo via simulação	68
4.7	Aplicação	68
4.8	Conclusões	74
5	Considerações finais	75
5.1	Pesquisas futuras	75
	REFERÊNCIAS	77
	APÊNDICES	83

RESUMO

O modelo de regressão odd log-logística gama generalizada com aplicações em análise de sobrevivência

Propor uma família de distribuição de probabilidade mais ampla e flexível é de grande importância em estudos estatísticos. Neste trabalho é utilizado um novo método de adicionar um parâmetro para uma distribuição contínua. A distribuição gama generalizada, que tem como casos especiais a distribuição Weibull, exponencial, gama, qui-quadrado, é usada como distribuição base. O novo modelo obtido tem quatro parâmetros e é chamado odd log-logística gama generalizada (OLLGG). Uma das características interessante do modelo OLLGG é o fato de apresentar bimodalidade. Outra proposta deste trabalho é introduzir um modelo de regressão chamado log-odd log-logística gama generalizada (LOLLGG) com base na GG (STACY e MIHRAM, 1965). Este modelo pode ser muito útil, quando por exemplo, os dados amostrados possuem uma mistura de duas populações estatísticas. Outra vantagem da distribuição OLLGG consiste na capacidade de apresentar várias formas para a função de risco, crescente, decrescente, na forma de U e bimodal entre outras. Desta forma, são apresentadas em ambos os casos as expressões explícitas para os momentos, função geradora e desvios médios. Considerando dados não-censurados e censurados de forma aleatória, as estimativas para os parâmetros de interesse, foram obtidas via método da máxima verossimilhança. Estudos de simulação, considerando diferentes valores para os parâmetros, porcentagens de censura e tamanhos amostrais foram conduzidos com o objetivo de verificar a flexibilidade da distribuição e a adequabilidade dos resíduos no modelo de regressão. Para ilustrar, são realizadas aplicações em conjuntos de dados reais.

Palavras-chave: Análise de sobrevivência; Censura aleatória; Distribuição odd log-logística gama generalizada; Distribuição gama generalizada; Simulação; Modelo de regressão log-odd log-logística gama generalizada

ABSTRACT

The regression model odd log-logistics generalized gamma with applications in survival analysis

Providing a wider and more flexible probability distribution family is of great importance in statistical studies. In this work a new method of adding a parameter to a continuous distribution is used. In this study the generalized gamma distribution (GG) is used as base distribution. The GG distribution has, as especial cases, Weibull distribution, exponential, gamma, chi-square, among others. For this motive, it is considered a flexible distribution in data modeling procedures. The new model obtained with four parameters is called log-odd log-logistic generalized gamma (OLLGG). One of the interesting characteristics of the OLLGG model is the fact that it presents bimodality. In addition, a regression model regression model called log-odd log-logistic generalized gamma (LOLLGG) based by GG (STACY e MIHRAM, 1965) is introduced. This model can be very useful when, the sampled data has a mixture of two statistical populations. Another advantage of the OLLGG distribution is the ability to present various forms for the failing rate, as increasing, as decreasing, and the shapes of bathtub or U. Explicit expressions for the moments, generating functions, mean deviations are obtained. Considering non-censored and randomly censored data, the estimates for the parameters of interest were obtained using the maximum likelihood method. Simulation studies, considering different values for the parameters, percentages of censoring and sample sizes were done in order to verify the distribution flexibility, and the residues distribution in the regression model. To illustrate, some applications using real data sets are carried out.

Keywords: Survival analysis; Random censoring; Odd log-logistic generalized gamma distribution; Generalized gamma distribution; Simulation; Log-odd log-logistic generalized gamma regression model

LISTA DE FIGURAS

2.1	Gráfico ilustrativo de algumas curvas TTT-plot.	19
2.2	Gráficos da distribuição GG para diferentes valores dos parâmetros. (a) Função de densidade de probabilidade. (b) Função de sobrevivência. (c) Função de risco.	22
3.1	Gráficos da fdp OLLGG para diferentes valores dos parâmetros. (a) Fixando $\lambda = 1$. (b) Fixando $\alpha = 2$, $\tau = 3$ e $k = 10$. (c) Fixando $\alpha = 2$, $\tau = 5$ e $\lambda = 0, 15$	30
3.2	Gráficos da função de risco OLLGG para diferentes valores dos parâmetros. (a) Função de risco com forma de banheira (ou U). (b) Função de risco unimodal. (c) Função de risco crescente, decrescente, constante e outras. .	30
3.3	Assimetria e curtose para a distribuição OLLGG em função de λ para alguns valores de k com $\alpha = 2$ e $\tau = 1$	35
3.4	Assimetria e curtose para a distribuição OLLGG em função de τ para alguns valores de λ com $\alpha = 2$ e $k = 1$	35
3.5	Comportamento da fdp OLLGG real e estimada. (a) $n = 50$. (b) $n = 150$. (c) $n = 350$	39
3.6	Comportamento da fdp OLLGG real e estimada. (a) $n = 50$. (b) $n = 150$. (c) $n = 350$	39
3.7	Comportamento da fdp OLLGG real e estimada considerando 0%, 10% e 30% de censura para os tamanhos de amostras $n = 50, 150$ e 350	41
3.8	(a) Funções de densidade estimadas para os modelos OLLGG, GG e Weibull para os dados de temperatura. (b) Funções de distribuição estimadas para os modelos OLLGG, GG e Weibull e distribuição empírica para os dados de temperatura.	43
3.9	Curva TTT-plot para os dados de Aids.	45
3.10	Função de sobrevivência estimada com ajuste da distribuição OLLGG e alguns outros modelos e sobrevivência empírica para os dados de Aids. (a) OLLGG vs KGG e GG. (b) OLLGG vs BW, EGG e Weibull.	46
3.11	Função de risco estimada com ajuste da distribuição OLLGG e alguns outros modelos e risco empírico para os dados de Aids. (a) OLLGG vs KGG e GG. (b) OLLGG vs BW, EGG e Weibull.	46
4.1	Gráficos da pdf OLLGG considerando diferentes valores dos parâmetros. (a) Para alguns valores de $\tau < 0$ e $\alpha = 1$ fixo. (b) Para alguns valores de $\tau > 0$ e $\alpha = 1$ fixo. (c) Para alguns valores de $\tau < 0$ e $\tau > 0$, $\alpha = 2$ e $\lambda = 0, 15$ fixos.	52

4.2	Gráficos da função de risco para a distribuição OLLGG considerando diferentes valores dos parâmetros. (a) Para alguns valores de $\tau > 0$, fixando $\alpha = 1$. (b) Para alguns valores de $\tau > 0$. (c) Para alguns valores de $\tau < 0$ e $\tau > 0$. (d) Para alguns valores de $\tau < 0$ fixando $\alpha = 1$ e $\lambda = 0,15$. (e) Para alguns valores de $\tau < 0$ e $\tau > 0$ fixando $\alpha = 2$ e $\lambda = 0,15$	54
4.3	Gráficos da fdp LOLLGG para alguns valores dos parâmetros. (a) Para alguns valores de $q < 0$ e $\mu = 2$, $\sigma = 1$ e $\lambda = 0,25$ fixos. (b) Para $q = 0$ fixo $\mu = 0$, $\sigma = 1,5$ e alguns valores de λ . (c) Para alguns valores de $q > 0$ e $\mu = 2$, $\sigma = 1$ e $\lambda = 0,25$ fixos.	61
4.4	Gráficos normal probabilístico para os resíduos r_{D_i} para o cenário 1. (a) Amostra de tamanho $n=50$. (b) Amostra de tamanho $n=150$. (c) Amostra de tamanho $n=350$. Para porcentagens de censuras de 0%, 10% e 30%.	69
4.5	Gráficos normal probabilístico para os resíduos r_{D_i} para o cenário 2. (a) Amostra de tamanho $n=50$. (b) Amostra de tamanho $n=150$. (c) Amostra de tamanho $n=350$. Para porcentagens de censuras de 0%, 10% e 30%.	70
4.6	Curva TTT-plot para os dados de níveis de tensão.	71
4.7	Gráfico da função de sobrevivência estimada com ajuste da distribuição OLLGG e alguns outros modelos e a sobrevivência empírica para os dados de níveis de tensão. (a) OLLGG, OLLW e OLLE. (b) OLLGG, GG e Weibull.	72
4.8	Gráfico da função de risco estimada com ajuste da distribuição OLLGG e alguns outros modelos e o risco empírico para os dados de níveis de tensão. (a) OLLGG, OLLW e OLLE. (b) OLLGG, GG e Weibull.	73
4.9	(a) Função de sobrevivência estimada considerando o modelo de regressão LOLLGG para os dados de níveis de tensão. (b) Gráfico normal probabilístico para os resíduos deviance com envelope.	74

LISTA DE TABELAS

2.1	Casos particulares da distribuição GG.	22
3.1	EMs, desvios padrão e erros quadráticos médio (EQM) para os parâmetros da distribuição OLLGG.	38
3.2	EMs e erros quadráticos médio (EQM) para os parâmetros da distribuição OLLGG.	40
3.3	EMVs para os parâmetros dos modelos OLLGG, KumGG, EGG, GG, Weibull e BW para os dados de temperatura, correspondentes erros padrão (entre parenteses) e os valores das estatísticas AIC, CAIC e BIC.	42
3.4	<i>TRV</i> para os dados de temperatura.	43
3.5	EMVs dos parâmetros dos modelos OLLGG, KumGG, EGG, GG, Weibull e BW para os dados de Aids, correspondentes erros padrão (entre parenteses) e valores das estatísticas AIC, BIC e CAIC.	45
3.6	<i>TRV</i> para os dados de Aids.	45
4.1	Alguns submodelos para $\tau > 0$	51
4.2	Alguns submodelos para $\tau < 0$	52
4.3	EMs e (EQMs) para os parâmetros da distribuição LOLLGG.	66
4.4	EMVs para os parâmetros do modelo para os dados de níveis de tensão, os correnpondentes EPs (dados em parênteses) e as estatísticas AIC, CAIC e BIC.	72
4.5	<i>TRV</i> para os dados de níveis de tensão.	72
4.6	EMVs, EPs e p -valor para o modelo de regressão LOLLGG ajustado para os dados de níveis de tensão.	73

1 INTRODUÇÃO

Estatística é uma ciência que estuda problemas que envolvem conhecimento de diversas áreas, metodologias e técnicas como: probabilidade, análise multivariada, técnicas de amostragem, análise de sobrevivência e outras. Entre estas, a análise de sobrevivência é uma das áreas de grande interesse de estudo e pesquisa. Como exemplo, a aplicação de tal metodologia, pode ser o estudo do tempo de vida de pacientes com determinada doença e o estudo do tempo de falha de um componente eletrônico. O estudo de tempos é denominado análise de sobrevivência na área médica e análise de confiabilidade na área industrial.

Os dados de sobrevivência apresentam características que os diferenciam de outros procedimentos de análise. A primeira é que essa variável possui suporte $\mathbb{R}_+$, portanto, não é razoável assumir que tem distribuição normal e geralmente, apresenta distribuição assimétrica positiva. A segunda característica muito frequente em dados de sobrevivência é a presença de observações censuradas, isto é, para alguns elementos em estudo não se conhece o tempo de interesse exato, mas apenas que ocorre à direita ou à esquerda de um certo valor. Usualmente, dados censurados ocorrem, pois nem sempre é possível esperar que o evento de interesse ocorra para todos os elementos em estudo.

As observações censuradas devem ser introduzidas na análise estatística por duas razões (COLOSIMO e GIOLO, 2006): (i) mesmo sendo incompletas, elas fornecem informações sobre o tempo de vida de pacientes e (ii) a omissão das censuras no cálculo das estatísticas de interesse pode acarretar conclusões viciadas. Existem diversos tipos de censura, neste trabalho considerou-se censura à direita sob o mecanismo aleatório e não informativa.

Além disso, nos estudos de dados relacionados a sobrevivência, na maioria das vezes a variável de interesse, o tempo, é influenciado por uma ou várias covariáveis que apresentam características dos indivíduos em estudo. Desta forma, a inclusão de covariáveis em um modelo de regressão é uma maneira de explicar a heterogeneidade ou a variabilidade presente nos tempos de falha.

Desta forma, sob o enfoque de modelos de regressão paramétricos em análise de sobrevivência algumas distribuições clássicas associadas à variável resposta podem ser consideradas: distribuição Weibull, distribuição exponencial e distribuição log-normal (LAWLESS, 2003). Porém, em algumas situações práticas tais distribuições não possuem flexibilidade para modelar algumas formas que a função de risco pode assumir, como por exemplo, a forma U e bimodal. Isso tem levado a generalizações e modificações de distribuições contínuas com a finalidade de obter maior flexibilidade nas formas da função de risco.

No contexto de novos modelo de probabilidade, MUDHOLKAR *et al.* (1995) apresentaram uma extensão da distribuição Weibull, a chamada distribuição Weibull expo-

nenciada, que possui uma classe mais ampla de funções de risco nas formas unimodal e U. A flexibilidade desta distribuição foi ilustrada por meio de aplicações em cinco conjuntos de dados clássicos com censura. EUGENE *et al.* (2002) apresentaram a classe beta generalizada (Beta-G) de distribuições utilizando a distribuição beta-normal, baseada na composição da distribuição beta e da distribuição normal. Outras generalizações podem ser vistas em MUDHOLKAR *et al.* (1996) e MARSHALL e OLKIN (1997).

CARRASCO *et al.* (2008) propuseram a distribuição Weibull modificada generalizada, uma distribuição com quatro parâmetros, capaz de modelar as formas unimodais e de U da função de risco, e possui como casos particulares as distribuições exponencial exponenciada, a Weibull exponenciada e a Rayleigh generalizada. PASCOA *et al.* (2011) introduzem e estudam a distribuição Kumaraswamy gama generalizada (KumGG) com cinco parâmetros. Sua importância reside na sua capacidade de modelar funções de risco monótona e não-monótona que são bastante comuns em dados de análise de sobrevivência e confiabilidade, e possui alguns submodelos especiais como, a gama generalizada exponenciada, a Weibull exponenciada, a semi normal exponenciada e a semi normal generalizada, entre outros. Outros trabalhos realizados recentemente nestes temas são, CORDEIRO *et al.* (2011), CORDEIRO *et al.* (2013a), CORDEIRO *et al.* (2013b) e CRUZ *et al.* (2016).

O objetivo deste trabalho é propor o estudo de uma nova distribuição. Esta nova distribuição é obtida, modificando-se uma distribuição de base contínua pela adição de um parâmetro de forma por meio da técnica de GLEATON e LYNCH (2006). Desta maneira, é apresentado a modificação da distribuição gama generalizada (STACY, 1962). Esta nova distribuição denominada por distribuição odd log-logística gama generalizada (OLLGG) tem como vantagens a modelagem de dados com bimodalidade, assim como, a obtenção de funções de riscos crescente, decrescente, unimodais, em forma de U entre outras. Também foi considerada outra nova distribuição denominada log-odd log-logística gama generalizada (LOLLGG) que é baseada na distribuição gama generalizada (STACY e MIHRAM, 1965). Algumas propriedades das distribuições propostas são apresentadas. Em particular para o modelo LOLLGG considerou-se o modelo de regressão na forma locação-escala.

O trabalho está organizado da seguinte maneira. No Capítulo 2 são apresentados alguns conceitos básicos em análise de sobrevivência, complementado por um breve estudo da distribuição gama generalizada. No Capítulo 3 a distribuição odd log-logística gama generalizada é obtida a partir da distribuição gama generalizada. Algumas características e propriedades desta distribuição são apresentadas. Um estudo de simulação é realizado para verificar as propriedades assintóticas do novo modelo. Duas aplicações são realizadas para ilustrar a flexibilidade do modelo proposto. No Capítulo 4 o novo modelo chamado modelo de regressão log-odd log-logística gama generalizada é obtido a partir da transformação logarítmica. Sob a abordagem de modelos de regressão locação-escala

um estudo de simulação é realizado para avaliar os resíduos do novo modelo. Uma aplicação com dados censurados é realizada para ilustrar a aplicabilidade do modelo. Por fim, no Capítulo 5, são apresentadas as considerações finais e as perspectivas para trabalhos futuros.

2 REVISÃO BIBLIOGRÁFICA

Neste capítulo são apresentados alguns conceitos básicos para o desenvolvimento de estudos em análise de sobrevivência.

2.1 Conceitos básicos em análise de sobrevivência

Em análise de sobrevivência as observações (ou respostas) são caracterizadas pelos tempos de falhas, ou também pela presença de censura. Tal metodologia é utilizada quando se deseja analisar dados correspondentes ao tempo de ocorrência de um evento de interesse. Desta forma, seja T uma variável aleatória (v.a) não-negativa e contínua que representa o tempo de sobrevivência de um indivíduo (ou um item) e pode ser expressa por funções matemáticas. Na sequência, são definidas as funções de densidade de probabilidade (fdp), $f(t)$, a de distribuição acumulada (fda), $F(t)$, a de sobrevivência, $S(t)$ e de risco ou taxa de falha, $h(t)$.

A fdp de uma v.a T positiva, com probabilidade de um indivíduo sofrer o evento de interesse no tempo t , é dada por

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t}.$$

A função de sobrevivência é definida como a probabilidade de um indivíduo não falhar (ou sobreviver) a partir de um tempo de origem para algum tempo além de t , é dada por

$$\begin{aligned} S(t) &= P(T > t) = 1 - P(T \leq t) \\ &= 1 - \int_0^t f(u) du = 1 - F(t), \end{aligned}$$

em que $F(t)$ é a probabilidade de um indivíduo sofrer o evento de interesse até o tempo t .

A função de risco é a probabilidade de falha instantânea de um indivíduo dado que o mesmo sobreviveu até o tempo t . A função de risco é definida por

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

A partir das funções definidas anteriormente, podemos obter algumas relações importantes, como

$$h(t) = \frac{f(t)}{S(t)} \quad \text{ou} \quad h(t) = -\frac{d}{dt} \{\log S(t)\},$$

e

$$S(t) = \exp [-H(t)],$$

no qual

$$H(t) = \int_0^t h(u) du$$

é a função de risco acumulado. Além disso, o conhecimento de uma função, por exemplo $f(t)$, implica no conhecimento das outras funções, $F(t)$, $S(t)$, $h(t)$ e $H(t)$.

2.1.1 Censura

As razões da ocorrência de dados com censura são várias, como por exemplo, a falha dos indivíduos por razões diferentes daquelas de interesse; planejamento do estudo; perda de registros, dentre outras. Entretanto, em uma análise, é importante considerar todas as informações a respeito da variável de interesse, inclusive os registros de informações censuradas. A omissão de tais observações podem nos levar à conclusões tendenciosas ou viciadas ao calcular estatísticas de interesse.

Para o caso em que todos os indivíduos sob estudo falharam, o tempo até a ocorrência da falha de cada indivíduo é conhecido. Este caso é chamado de dados completos.

Os dados de sobrevivência podem ser representados pelo par (t_i, δ_i) , em que $i = 1, \dots, n$, sendo t_i o tempo de falha ou de censura do i -ésimo indivíduo e δ_i a função indicadora de censura, dada por

$$\delta_i = \begin{cases} 1, & \text{se } t_i \text{ é uma observação completa,} \\ 0, & \text{se } t_i \text{ é uma observação censurada.} \end{cases}$$

No âmbito de estudos que envolvem dados censurados existem diversos mecanismos de censuras para realizar as análises estatísticas:

- **Censura tipo I:** o estudo é terminado após um período pré-estabelecido de tempo, sendo que alguns indivíduos não sofreram o evento de interesse, ou seja, não falharam. Desta maneira, o tempo de falha de todos os indivíduos que não puderam ser observados completamente, passam a ser considerados como censura.
- **Censura tipo II:** antes do início do experimento é pré-estabelecida uma quantidade r de indivíduos (ou itens) e o estudo é terminado após r ocorrência do evento de interesse. Assim, para um estudo com n indivíduos ao final teremos $n-r$ observações censuradas.
- **Censura aleatória:** caracterizada pela perda de acompanhamento do indivíduo no decorrer do estudo sem este ter apresentado o evento de interesse (falha). Na prática, este é o tipo de censura que mais aparece, sendo representado por meio de duas v.a independentes T_i e C_i , em que a v.a T_i representa o tempo de falha do i -ésimo indivíduo e a v.a C_i representa o tempo de censura associado a este indivíduo. Neste caso, o tempo da i -ésima observação é dado por $t_i = \min(T_i, C_i)$.
- **Censura intervalar:** é conhecido que o evento de interesse ocorreu em um certo intervalo de tempo. Ou seja, o tempo de falha não é conhecido exatamente, mas pertence ao intervalo de tempo $(L, U]$.

Outros mecanismos de censura, como por exemplo, à esquerda e à direita podem ser encontrados em MEEKER e ESCOBAR (1998) e COLOSIMO e GIOLO (2006).

2.1.2 Gráfico TTT-plot

BARLOW e CAMPO (1975) propuseram uma técnica gráfica para verificar o comportamento da função de risco, chamado de TTT-plot (tempo total em teste). A curva TTT é obtida construindo um gráfico de

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

em função de $\frac{r}{n}$, em que n é o tamanho da amostra, $r = 1, \dots, n$ e $T_{i:n}$, $i = 1, \dots, n$ são estatísticas de ordem da amostra.

Utilizando-se a curva TTT pode-se detectar o tipo de função de risco que os dados de tempos de vida em estudo possuem. Na Figura 2.1 é apresentado o comportamento da função de risco utilizando-se o TTT-plot.

Figura 2.1. Gráfico ilustrativo de algumas curvas TTT-plot.

Da Figura 2.1 podemos observar os seguintes padrões de curvas:

- função de risco constante (reta A);
- curva convexa se a função de risco é decrescente (curva B);
- curva côncava se a função de risco é crescente (curva C);
- curvatura convexa e depois côncava função de risco em forma de U (curva D);
- curvatura côncava e depois convexa função de risco unimodal (curva E).

Em situações em que a curva apresenta várias regiões côncavas e depois convexas este comportamento direciona para o ajuste de uma função de risco multinomial. Essas curvas podem ser ajustadas através da distribuição de múltiplos riscos (LOUZADA-NETO e DE BRAGANÇA PEREIRA, 2000).

2.1.3 A função de verossimilhança em análise de sobrevivência

Através do método de máxima verossimilhança é possível incorporar as censuras presentes em muitos dados de tempo de vida.

Seja uma amostra de v.a's independentes $T_1, T_2, \dots, T_n$ de tempos de sobrevivência. Supondo que os dados consistem em n pares observados $(t_1, \delta_1), (t_2, \delta_2), \dots, (t_n, \delta_n)$ em que t_i é o tempo de sobrevivência ou censura, δ_i é o indicador de censura. Neste caso, a função de verossimilhança considerando censura não informativa é dada por

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n [f(t_i; \boldsymbol{\theta})]^{\delta_i} [S(t_i; \boldsymbol{\theta})]^{1-\delta_i},$$

em que $\boldsymbol{\theta}$ é o vetor de parâmetros desconhecidos, $f(t_i)$ e $S(t_i)$ são as funções de densidade de probabilidade e de sobrevivência para cada v.a T_i , respectivamente. Observa-se que a contribuição de cada observação não censurada é a sua função de densidade e que cada observação censurada contribui com a função de sobrevivência.

2.2 Distribuição gama generalizada

Em uma forma mais geral, a distribuição gama possui três parâmetros, e é conhecida como gama generalizada (GG). A distribuição GG introduzida por STACY (1962) é um modelo bastante utilizado para analisar dados assimétricos. Esta distribuição, possui como casos particulares, além da distribuição gama, outras distribuições importantes, como a exponencial, Weibull, Rayleigh e qui-quadrado. A distribuição GG acomoda diferentes formas para a função de risco, como crescente, decrescente, unimodal e de U.

Seja T uma variável aleatória contínua, que assume valores não-negativos. Tem-se que T segue uma distribuição de probabilidade gama generalizada, se sua fdp é dada por

$$g(t; \alpha, \tau, k) = \frac{\tau}{\alpha \Gamma(k)} \left(\frac{t}{\alpha}\right)^{\tau k - 1} \exp \left\{ - \left(\frac{t}{\alpha}\right)^{\tau} \right\}, \quad (2.1)$$

com $t > 0$, $\alpha > 0$ é o parâmetro de escala, $\tau > 0$ e $k > 0$ são os parâmetros de forma e $\Gamma(k)$ é a função gama definida por

$$\Gamma(k) = \int_0^{\infty} w^{k-1} e^{-w} dw.$$

A função GG definida pela equação (2.1) está implementada no *software* R (R DEVELOPMENT CORE TEAM, 2014) no pacote *flexsurv* utilizando a função *dgen-gamma.orig*, que representa a original parametrização de STACY (1962). Desta forma, é possível utilizá-la em procedimentos computacionais como, simulações de valores e construção de gráficos como os apresentados na Subseção 2.2.1.

2.2.1 Propriedades e medidas de interesse

Alguns autores estudaram as propriedades da distribuição GG. Dentre eles STACY e MIHRAM (1965), PRENTICE (1974) e LAWLESS (2003). A esperança e a variância da distribuição GG são dadas por

$$E[T] = \frac{\alpha \Gamma\left(\frac{\tau k + 1}{\tau}\right)}{\Gamma(k)}$$

e

$$V[T] = \frac{\alpha^2}{\Gamma(k)} \left[\Gamma\left(\frac{\tau k + 2}{\tau}\right) - \frac{[\Gamma\left(\frac{\tau k + 1}{\tau}\right)]^2}{\Gamma(k)} \right].$$

A fda $G(t; \alpha, \tau, k)$, a função de sobrevivência $S(t; \alpha, \tau, k)$ e a função de risco $h(t; \alpha, \tau, k)$ são expressas, respectivamente, por

$$G(t; \alpha, \tau, k) = \gamma_1\left(k, \left(\frac{t}{\alpha}\right)^\tau\right) = \frac{\gamma(k, (t/\alpha)^\tau)}{\Gamma(k)} = \gamma_1\left(k, \left(\frac{t}{\alpha}\right)^\tau\right), \quad t > 0, \quad (2.2)$$

$$S(t; \alpha, \tau, k) = 1 - G(t; \alpha, \tau, k) = 1 - \gamma_1\left(k, \left(\frac{t}{\alpha}\right)^\tau\right) \quad \text{e}$$

$$h(t; \alpha, \tau, k) = \frac{f(t; \alpha, \tau, k)}{S(t; \alpha, \tau, k)} = \frac{t^{\tau k - 1} \exp\left\{-\left(\frac{t}{\alpha}\right)^\tau\right\}}{\int_t^\infty w^{\tau k - 1} \exp\left\{-\left(\frac{w}{\alpha}\right)^\tau\right\} dw},$$

em que $\gamma(k, x) = \int_0^x w^{k-1} e^{-w} dw$ é a função gama incompleta. Outras propriedades de interesse são os momentos e a função geradora de momentos. O j -ésimo momento da variável aleatória T com distribuição GG é dada por

$$\mu_j = \frac{\alpha^j}{\Gamma(k)} \Gamma(k + j/\tau), \quad j/\tau > -k.$$

A função geradora de momentos da distribuição GG é dado por

$$M_T(r) = \frac{1}{\Gamma(k)} \sum_{m=0}^{\infty} \Gamma(k + m/\tau) \frac{(\alpha r)^m}{m!}.$$

Um fato interessante da distribuição GG é que esta possui como casos particulares distribuições conhecidas na literatura. Na Tabela 2.1 são listadas algumas destas distribuições. O comportamento da distribuição GG é apresentado na Figura 2.2 considerando diferentes valores dos parâmetros para as funções fdp (a), sobrevivência (b) e de risco (c). Esta figura mostra o quanto a distribuição GG é flexível como, por exemplo, Figura 2.2c as formas que a função de risco pode assumir, como unimodal e em forma de U.

Tabela 2.1. Casos particulares da distribuição GG.

Distribuição	α	τ	k
Gama	α	1	k
Weibull	α	τ	1
Exponencial	α	1	1
Qui-Quadrado	2	1	$\frac{n}{2}$
Qui	$\sqrt{2}$	2	$\frac{n}{2}$
Qui-escala	$\sqrt{2}\sigma$	2	$\frac{n}{2}$
Rayleigh	α	2	1
Maxwell	α	2	$\frac{3}{2}$
Normal dupla	$\sqrt{2}$	2	$\frac{1}{2}$
Normal circular	$\sqrt{2}$	2	1
Normal esférica	$\sqrt{2}$	2	$\frac{3}{2}$
Semi Normal Generalizada	$2^{\frac{1}{2\gamma}}\theta$	2γ	$\frac{1}{2}$
Semi Normal	$2^{\frac{1}{2}}\theta$	2	$\frac{1}{2}$

Figura 2.2. Gráficos da distribuição GG para diferentes valores dos parâmetros. (a) Função de densidade de probabilidade. (b) Função de sobrevivência. (c) Função de risco.

2.3 Modelo de regressão

Na prática, é comum a ocorrência de situações em que uma ou mais covariáveis estão relacionadas aos tempos de sobrevivência, isto é, os tempos de falha são influenciados por covariáveis. Na área biomédicas, por exemplo, podem ser a idade, a altura, um tipo de tumor cancerígeno, a quantidade de hemoglobina no sangue, etc. Tais covariáveis explicam parte da heterogeneidade do tempo até a ocorrência do evento de interesse.

Considere T uma v.a e seja $\mathbf{x} = (x_1, x_2, \dots, x_p)^\top$ um vetor com p covariáveis. Uma maneira de estabelecer relação entre T e $\mathbf{x}$ é por meio da utilização de modelos de regressão.

O modelo de regressão paramétrico é uma maneira de estudar o efeito que as covariáveis causam no tempo de resposta. Na literatura esses modelos são conhecidos como modelos de tempo de vida acelerados ou também locação e escala.

Muitos trabalhos sobre diversos tipos de modelos de regressão podem ser encontrados em COX (1972), COX e OAKES (1984), KALBFLEISCH e PRENTICE (2002), LAWLESS (2003) e NELSON (2009), entre outros.

2.3.1 Modelo de regressão locação-escala

A classe de modelos de locação e escala consiste em utilizar a transformação logarítmica dos tempos, $Y = \log(T)$, de tal forma que para um vetor de covariáveis, o logaritmo do tempo possui uma distribuição com um parâmetro de locação, μ ($-\infty < \mu < \infty$) que depende das variáveis regressoras (covariáveis), e um parâmetro de escala, σ ($\sigma > 0$).

As distribuições que pertencem a essa família têm fdp dada por

$$f(y; \mu, \sigma) = \frac{1}{\sigma} g\left(\frac{y - \mu}{\sigma}\right), \quad -\infty < y < \infty,$$

e pode-se escrever o modelo log-linear da seguinte forma

$$\mathbf{Y} = \mu(\mathbf{x}) + \sigma \mathbf{Z},$$

em que $\mathbf{Y}$ pertence a família de distribuições locação e escala, $\mathbf{Z}$ é o erro aleatório e $\mathbf{x}$ é o vetor de covariáveis.

Uma importante característica deste modelo é que as variáveis regressoras possuem efeito multiplicativo sobre $\mathbf{T}$, isto é, $\mathbf{T} = \exp[\mu(\mathbf{x})] \exp[\sigma \mathbf{Z}]$. Assim, o parâmetro de locação μ escrito em termos das covariáveis é dado por $\mu(\mathbf{x}) = \mathbf{x}^\top \boldsymbol{\beta}$, em que $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^\top$ é o vetor de parâmetros desconhecidos. Este modelo possui efeito linear em $\mathbf{Y}$ e a função de sobrevivência de $\mathbf{Y}$ dado $\mathbf{x}$ é da forma $G\left(\frac{y - \mu(\mathbf{x})}{\sigma}\right)$, em que $G(z)$ é a função de sobrevivência de $\mathbf{Z}$ e $z = \left(\frac{y - \mu(\mathbf{x})}{\sigma}\right)$.

2.4 Inferência estatística

Assumindo que o modelo de regressão utilizado é adequado para análise dos dados, a próxima etapa é utilizar um método de estimação para os parâmetros e realizar o processo de inferência. Para estimação dos parâmetros do modelo de regressão, várias abordagens podem ser utilizadas, tais como o método de máxima verossimilhança, o método de Jackknife, a análise bayesiana, entre outros. Neste trabalho, é considerado o processo de estimação via método de máxima verossimilhança.

2.4.1 Método de máxima verossimilhança

Sejam $(y_1, x_{i1}, \delta_1), \dots, (y_n, x_{in}, \delta_n)$, n observações independentes em que $y_i = \log(t_i)$, representa o logaritmo do tempo de falha ou censura, $x_i = (x_{i1}, \dots, x_{ip})^T$ é o vetor de covariáveis e δ_i é o indicador de censura, para $i = 1, \dots, n$. Desta maneira, o logaritmo da função de verossimilhança para o vetor de parâmetros $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^T$ é dado por

$$l(\boldsymbol{\theta}) = \sum_{i \in F} \log [f(y_i)] \sum_{i \in C} \log [S(y_i)],$$

em que $f(y)$ e $S(y)$ são as funções densidade de probabilidade e sobrevivência da v.a Y e F e C denotam o conjunto de observações não censuradas e censuradas, respectivamente.

As propriedades assintóticas dos estimadores de máxima verossimilhança são necessárias para construção de intervalos de confiança e testes de hipóteses sobre os parâmetros do modelo. Sob certas condições de regularidade $\hat{\boldsymbol{\theta}}$ tem assintoticamente uma distribuição normal multivariada com média $\boldsymbol{\theta}$ e matriz de variância e covariância dada pelo inverso da matriz de informação de Fisher ($I(\boldsymbol{\theta})^{-1}$), em que $I(\boldsymbol{\theta}) = E [\ddot{L}(\boldsymbol{\theta})]$, tal que, $\ddot{L}(\boldsymbol{\theta}) = - \left\{ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right\}$.

Como o cálculo da matriz de informação $I(\boldsymbol{\theta})$ é complicado, devido às observações censuradas, pode-se usar a matriz Hessiana, $-\ddot{L}(\boldsymbol{\theta})$, avaliada em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, que é um estimador consistente para $I(\boldsymbol{\theta})$. Portanto, a distribuição para $\hat{\boldsymbol{\theta}}$ é especificada por $\hat{\boldsymbol{\theta}}^T \sim N_{(d)} \left(\boldsymbol{\theta}^T, \ddot{L}(\boldsymbol{\theta})^{-1} \right)$, em que $\ddot{L}(\boldsymbol{\theta})$ é a matriz de informação observada, denotada neste trabalho por $J(\boldsymbol{\theta})$ ou $K(\boldsymbol{\theta})$ e d é o número de parâmetros do modelo. Além disso, pode-se obter os intervalos de confiança aproximados para os parâmetros individualmente e também a realização de testes de hipóteses na discriminação do modelo proposto em relação aos modelos encaixados.

2.5 Estatísticas AIC, BIC e CAIC

Alguns critérios comuns na literatura podem ser utilizados para seleção de modelos. Esses critérios levam em consideração a complexidade do modelo no critério de seleção. São critérios que penalizam a verossimilhança, utilizando o número de parâmetros a serem estimados e o tamanho da amostra.

Entre os critérios de seleção de modelos, os utilizados neste estudo são:

- AIC - Critério de informação de Akaike (AKAIKE, 1974), dado por

$$\text{AIC} = -2l(\boldsymbol{\theta}) + 2d.$$

- CAIC - Critério de informação de Akaike corrigido (BOZDOGAN, 1987), dado por

$$\text{CAIC} = \text{AIC} + \frac{2d(d+1)}{n-d-1},$$

- BIC - Critério de informação bayesiano (SCHWARZ *et al.*, 1978), dado por

$$\text{BIC} = -2l(\boldsymbol{\theta}) + d \log(n).$$

em que $l(\boldsymbol{\theta})$ é o logaritmo da função de verossimilhança, d o número de parâmetros do modelo e n o tamanho da amostra em estudo.

3 DISTRIBUIÇÃO ODD LOG-LOGÍSTICA GAMA GENERALIZADA

3.1 Introdução

A literatura estatística é preenchida com centenas de distribuições univariadas contínuas. Desenvolvimentos recentes concentram-se em novas técnicas para a construção de modelos significativos e muitos métodos tem sido propostos para introduzir-se um ou mais parâmetros para gerar novas distribuições. Entre estes métodos, a composição de algumas distintas e importantes distribuições para o tempo de vida tem sido a vanguarda da modelagem em estudos de análise de sobrevivência e várias famílias de distribuições foram investigadas.

A distribuição log-logística (LL) com um parâmetro de forma $\lambda > 0$ é um modelo útil para a análise de sobrevivência e é uma alternativa à distribuição log-normal. Ao contrário da distribuição de Weibull mais comumente utilizada, a distribuição LL tem uma função de risco não-monotônica, que a torna adequada para a modelagem em sobrevivência para alguns tipos de dados relacionados ao estudo de câncer. Para $\lambda > 1$, a função taxa de risco é unimodal e quando $\lambda = 1$, o risco diminui monotonicamente. O fato de sua fda ter uma expressão analítica em termos de funções conhecidas é particularmente útil para a análise de dados de sobrevivência com censura.

Uma generalização da LL é a família de distribuições odd log-logística (OLL) proposta por GLEATON e LYNCH (2006). Os autores chamam esta família de família log-logística generalizada (GLL). Recentemente, CRUZ *et al.* (2016) propuseram a odd log-logística Weibull; DA SILVA BRAGA *et al.* (2016) estudaram a distribuição odd log-logística normal e CORDEIRO *et al.* (2016) propuseram a família beta odd log-logística generalizada. Neste sentido, é desenvolvido uma metodologia semelhante para propor um novo modelo baseado na distribuição gama generalizada (GG). A distribuição GG desempenha um papel muito importante em problemas de inferência estatística pela flexibilidade e por ter modelos conhecidos como caso particular o que permite testar a qualidade de ajuste com tais modelos. Ao modelar a funções de risco monótonas, a distribuição Weibull pode ser uma alternativa inicial devido suas formas de densidade negativa e positivamente inclinada. No entanto, a distribuição não apresenta um ajuste paramétrico razoável para fenômeno de modelagem com riscos não monótonos, como a forma de banheira e unimodal, que são comuns em estudos biológicos e de confiabilidade.

Outras extensões da distribuição GG foram desenvolvidos para modelagem de dados de sobrevivência. Por exemplo, CORDEIRO *et al.* (2011) definiram a gama generalizada exponenciada, PASCOA *et al.* (2011) introduziram a distribuição Kumaraswamy gama generalizada, ORTEGA *et al.* (2011) propuseram a gama generalizada geométrica, CORDEIRO *et al.* (2013b) estudaram a distribuição beta gama generalizada, LUCENA *et al.*

(2015) definiram a distribuição gama generalizada transmutada e SILVA *et al.* (2016) propuseram a distribuição série de potência gama generalizada.

Dada uma fda contínua de base $G(t; \xi)$, com um vetor de parâmetros ξ , a fda da distribuição *odd log-logistic-G* (“OLL-G”), com um parâmetro de forma adicional $\lambda > 0$, é definida por

$$F(t) = \int_0^{G(t; \xi)/\bar{G}(t; \xi)} \frac{\lambda x^{\lambda-1}}{(1+x^\lambda)^2} dx = \frac{G(t; \xi)^\lambda}{G(t; \xi)^\lambda + \bar{G}(t; \xi)^\lambda}. \quad (3.1)$$

O parâmetro λ pode ser escrito como

$$\lambda = \frac{\log [F(t)/\bar{F}(t)]}{\log [G(t)/\bar{G}(t)]} \quad \text{e} \quad \bar{G}(t; \xi) = 1 - G(t; \xi)$$

e λ representa o quociente do logaritmo da razão entre as distribuições gerada, $F(t)$ e de base, $G(t; \xi)$. Nota-se que não há uma função complicada na equação (3.1) em comparação com a família beta generalizada (EUGENE *et al.*, 2002), a qual inclui dois parâmetros extras e também envolve a função beta incompleta. A distribuição de base $G(t; \xi)$ é um caso especial de (3.1) quando $\lambda = 1$. Se $G(t; \xi) = t/(1+t)$, esta reduz-se a distribuição LL. A partir da equação (3.1) é possível gerar várias distribuições. Por exemplo, as distribuições odd log-logística Fréchet e a odd log-logística gama são obtidas tomando $G(t; \xi)$ como sendo a fda Fréchet e gama, respectivamente.

A fdp da nova distribuição é dada por

$$f(t) = \frac{\lambda g(t; \xi) \{G(t; \xi)[1 - G(t; \xi)]\}^{\lambda-1}}{\left\{G(t; \xi)^\lambda + [1 - G(t; \xi)]^\lambda\right\}^2}. \quad (3.2)$$

A família OLL-G de densidades (3.2) permite uma maior flexibilidade de suas caudas e pode ser amplamente aplicada em muitas áreas da engenharia e biologia. Algumas de suas propriedades matemáticas podem ser estudadas, pois é possível a extensão para várias distribuições bem conhecidas. A parte inferencial deste modelo é realizada utilizando a distribuição assintótica dos estimadores de máxima verossimilhança. Porém, nas situações em que o tamanho da amostra é pequeno ou moderado, os resultados da inferência em relação aos parâmetros do modelo podem ser ruins via estudos de simulações.

A sequência do capítulo está organizado da seguinte forma: na Seção 3.2, é apresentado a definição da distribuição odd log-logística gama generalizada (OLLGG). Na Seção 3.3, uma representação linear para a distribuição OLLGG é apresentado. Algumas propriedades matemática são desenvolvidas na Seção 3.4. Assumindo dados censurados, é realizado uma análise clássica via método de máxima verossimilhança para os parâmetros do modelo na Seção 3.5. Na Seção 3.6 são apresentados os resultados via estudo de simulação. Duas aplicações para conjuntos de dados reais são realizadas na Seção 3.7. Conclusões são apresentadas na Seção 3.8.

3.2 Distribuição odd log-logística gama generalizada

A fdp e a fda da distribuição OLLGG (para $t > 0$) são definidas substituindo (2.2) e (2.1) nas equações (3.1) e (3.2), respectivamente. Assim, a fdp da OLLGG é dada por

$$f(t) = \frac{\lambda \tau (t/\alpha)^{\tau k-1} \exp[-(t/\alpha)^\tau] \{\gamma_1(k, (t/\alpha)^\tau) [1 - \gamma_1(k, (t/\alpha)^\tau)]\}^{\lambda-1}}{\alpha \Gamma(k) \{\gamma_1^\lambda(k, (t/\alpha)^\tau) + [1 - \gamma_1(k, (t/\alpha)^\tau)]^\lambda\}^2}, \quad t > 0, \quad (3.3)$$

em que $\alpha > 0$ é o parâmetro de escala e os outros parâmetros positivos τ , k e λ são de forma. A vantagem de (3.3) é a sua capacidade de ajustar dados assimétricos e bimodais devido a inclusão do parâmetro λ .

As distribuições Weibull e GG são os mais importantes submodelos de (3.3) para $\lambda = k = 1$ e $\lambda = 1$, respectivamente. A distribuição OLLGG se aproxima da distribuição log-normal (LN), quando $\lambda = 1$ e $k \rightarrow \infty$. Outros submodelos podem ser imediatamente obtidos a partir da Tabela 2.1: OLL - Gamma, OLL - Qui - Quadrado, OLL - Qui, OLL - Qui-Escalar, OLL - exponencial, OLL - Weibull, OLL - Rayleigh, OLL - Maxwell e OLL - Normal dupla, OLL - Normal Circular, OLL - Normal esférica, OLL - Semi normal generalizada e OLL - Semi normal.

Se T é uma v.a com fdp (3.3), pode-se denotar $T \sim \text{OLLGG}(\alpha, \tau, k, \lambda)$. As funções de sobrevivência e risco correspondentes a (3.3) são

$$S(t) = 1 - F(t) = \frac{[1 - \gamma_1(k, (t/\alpha)^\tau)]^\lambda}{\gamma_1^\lambda(k, (t/\alpha)^\tau) + [1 - \gamma_1(k, (t/\alpha)^\tau)]^\lambda} \quad (3.4)$$

e

$$h(t) = \frac{\lambda \tau (t/\alpha)^{\tau k-1} \exp[-(t/\alpha)^\tau] \gamma_1^{\lambda-1}(k, (t/\alpha)^\tau) \{\gamma_1^\lambda(k, (t/\alpha)^\tau) + [1 - \gamma_1(k, (t/\alpha)^\tau)]^\lambda\}}{\alpha \Gamma(k) \{\gamma_1^\lambda(k, (t/\alpha)^\tau) + [1 - \gamma_1(k, (t/\alpha)^\tau)]^\lambda\}^2 [1 - \gamma_1(k, (t/\alpha)^\tau)]}, \quad (3.5)$$

respectivamente. A representação gráfica da fdp OLLGG para alguns valores dos parâmetros selecionados são apresentados na Figura 3.1. Nesta figura, é verificado que a forma bimodal é obtida para valores de $\lambda < 1$. A função de risco (3.5) é bastante flexível para modelar os dados de sobrevivência como apresentadas na Figura 3.2 para diferentes valores dos parâmetros. A partir de estudos numéricos para vários valores de parâmetros é mostrado na Figura 3.2c uma nova forma para a função de risco.

A OLLGG pode ser simulada invertendo a equação (3.1) e assumindo como função de base a GG, assim

$$t = Q_{GG} \left(\frac{u^{1/\lambda}}{(1-u)^{1/\lambda} + u^{1/\lambda}}, \alpha, \tau, k \right), \quad (3.6)$$

em que u tem distribuição uniforme $U(0, 1)$ e $Q_{GG}(\cdot) = G^{-1}(\cdot)$ é a função quantílica da GG.

Algumas propriedades da distribuição OLLGG são:

- Se $T \sim \text{OLLGG}(\alpha, \tau, k, \lambda)$ então $bT \sim \text{OLLGG}(b\alpha, \tau, k, \lambda)$, $\forall b > 0$.
- Se $T \sim \text{OLLGG}(\alpha, \tau, k, \lambda)$ então $T^m \sim \text{OLLGG}(\alpha^m, \tau/m, k, \lambda)$, $\forall m \neq 0$.

Figura 3.1. Gráficos da fdp OLLGG para diferentes valores dos parâmetros. (a) Fixando $\lambda = 1$. (b) Fixando $\alpha = 2$, $\tau = 3$ e $k = 10$. (c) Fixando $\alpha = 2$, $\tau = 5$ e $\lambda = 0, 15$.

Figura 3.2. Gráficos da função de risco OLLGG para diferentes valores dos parâmetros. (a) Função de risco com forma de banheira (ou U). (b) Função de risco unimodal. (c) Função de risco crescente, decrescente, constante e outras.

3.3 Representação linear para a distribuição OLLGG

Primeiro, defini-se a distribuição gama generalizada exponenciada (“Exp-GG”), denotada por $W \sim \text{Exp}^c(GG)$ com parâmetro de potência $c > 0$, se W tem fda e fdp dadas por

$$H_c(t) = G(t; \alpha, \tau, k)^c \quad \text{e} \quad h_c(t) = \frac{c\tau}{\alpha\Gamma(k)} \left(\frac{t}{\alpha}\right)^{\tau k - 1} \exp\left[-\left(\frac{t}{\alpha}\right)^\tau\right] G(t; \alpha, \tau, k)^{c-1},$$

respectivamente. Em um contexto geral, as propriedades das distribuições exponenciada-G (Exp-G) foram estudadas por vários autores para alguns modelos de base G, ver

MUDHOLKAR *et al.* (1995) e MUDHOLKAR *et al.* (1996) para Weibull exponenciada, NADARAJAH (2006) para Gumbel exponenciada, KAKDE e SHIRKE (2006) para log-normal exponenciada e NADARAJAH e GUPTA (2007) para distribuições gama exponenciada. Veja, também, NADARAJAH e KOTZ (2006), entre outros.

A seguir, é obtido uma expansão para $F(t; \alpha, \tau, k, \lambda)$ usando a série de potência para $G(t; \alpha, \tau, k)^\lambda$ ($\lambda > 0$ real)

$$G(t; \alpha, \tau, k)^\lambda = \sum_{j=0}^{\infty} a_j G(t; \alpha, \tau, k)^j, \quad (3.7)$$

em que

$$a_j = a_j(\lambda) = \sum_{u=j}^{\infty} (-1)^{j+u} \binom{\lambda}{u} \binom{u}{j}.$$

Para qualquer real $\lambda > 0$, é considerado a expansão binomial generalizada dada por

$$[1 - G(t; \alpha, \tau, k)]^\lambda = \sum_{j=0}^{\infty} (-1)^j \binom{\lambda}{j} G(t; \alpha, \tau, k)^j. \quad (3.8)$$

Substituindo (3.7) e (3.8) na equação (3.1), obtém-se

$$F(t; \alpha, \tau, k, \lambda) = \frac{\sum_{j=0}^{\infty} a_j G(t; \alpha, \tau, k)^j}{\sum_{j=0}^{\infty} b_j G(t; \alpha, \tau, k)^j},$$

em que $b_j = a_j + (-1)^j \binom{\lambda}{j}$ para $j \geq 0$.

A razão entre duas séries de potência pode ser expressa como

$$F(t; \alpha, \tau, k, \lambda) = \sum_{j=0}^{\infty} c_j G(t; \alpha, \tau, k)^j, \quad (3.9)$$

em que $c_0 = a_0/b_0$ e os coeficientes c_j 's (para $j \geq 1$) são determinados pela equação de recorrência

$$c_j = b_0^{-1} \left(a_j - \sum_{r=1}^j b_r c_{j-r} \right).$$

A fdp para T é obtida derivando (3.9) como

$$f(t; \alpha, \tau, k, \lambda) = \sum_{j=0}^{\infty} c_{j+1} h_{j+1}(t), \quad (3.10)$$

em que

$$h_{j+1}(t) = \frac{(j+1)\tau}{\alpha \Gamma(k)} \left(\frac{t}{\alpha} \right)^{\tau k - 1} \exp \left[- \left(\frac{t}{\alpha} \right)^\tau \right] G(t; \alpha, \tau, k)^j$$

é a função densidade Exp-GG com parâmetro de potência $j + 1$.

Para $j \geq 0$, pode-se escrever

$$h_{j+1}(t) = \frac{(j+1)\tau}{\alpha \Gamma(k)} (t/\alpha)^{\tau k-1} \exp \left[- \left(\frac{t}{\alpha} \right)^\tau \right] \gamma_1(k, (t/\alpha)^\tau)^j, \quad (3.11)$$

em que $\gamma_1(k, (t/\alpha)^\tau) = \gamma(k, (t/\alpha)^\tau) / \Gamma(k)$.

A potência de uma série de potência pode ser escrita para qualquer j positivo (GRADSHTEYN e RYZHIK, 2000) como

$$\left(\sum_{i=0}^{\infty} a_i x^i \right)^j = \sum_{i=0}^{\infty} d_{j,i} x^i, \quad (3.12)$$

em que o coeficiente $d_{j,i}$ (para $i = 1, 2, \dots$) satisfaz a relação de recorrência

$$d_{j,i} = (ia_0)^{-1} \sum_{p=1}^i [j(p+1) - i] a_p d_{j,i-p} \quad (3.13)$$

e $d_{j,0} = a_0^j$. O coeficiente $d_{j,i}$ pode ser expresso explicitamente a partir de $d_{j,0}, \dots, d_{j,i-1}$ e para $a_0, \dots, a_i$.

Além disso, usando a equação (3.12), pode-se escrever (para $j \geq 1$)

$$\gamma_1(k, (t/\alpha)^\tau)^j = \frac{(t/\alpha)^{jk\tau}}{\Gamma(k)^j} \sum_{i=0}^{\infty} d_{j,i} (t/\alpha)^{i\tau}, \quad (3.14)$$

em que os coeficientes $d_{j,i}$ (para $i \geq 1$) são determinados por (3.13) com $a_p = (-1)^p / [(k+p)p!]$.

Com base na equação (3.14) pode-se escrever a função de densidade Exp-GG (para $j \geq 1$) (3.11) como

$$h_{j+1}(t) = \frac{(j+1)\tau}{\alpha \Gamma(k)^{j+1}} \exp \left[- \left(\frac{t}{\alpha} \right)^\tau \right] \sum_{i=0}^{\infty} d_{j,i} \left(\frac{t}{\alpha} \right)^{[i+(j+1)k]\tau-1}.$$

A última densidade pode ser expressa em termos da função de densidade GG. Pela forma de (3.11), é escrita (para $j \geq 1$) como

$$h_{j+1}(t) = \sum_{i=0}^{\infty} e_{j,i} g(t; \alpha, \tau, i + (j+1)k), \quad (3.15)$$

em que $g(t; \alpha, \tau, i + (j+1)k)$ é a fdp GG com parâmetros α , τ e $i + (j+1)k$ e

$$e_{j,i} = \frac{(j+1) \Gamma(i + (j+1)k)}{\Gamma(k)^{j+1}} d_{j,i}. \quad (3.16)$$

Para $j = 0$, é obtido de (3.11) que $h_1(t) = \frac{\tau}{\alpha \Gamma(k)} (t/\alpha)^{\tau k-1} \exp \left[- \left(\frac{t}{\alpha} \right)^\tau \right] = g(t; \alpha, \tau, k)$. Combinando o resultado (3.15) (para $j \geq 1$) com o de $j = 0$, pode-se escrever $f(t) = f(t; \alpha, \tau, k, \lambda)$ em (3.10) como

$$f(t) = c_1 g(t; \alpha, \tau, k) + \sum_{j=1}^{\infty} \sum_{i=0}^{\infty} e_{j,i} c_{j+1} g(t; \alpha, \tau, i + (j+1)k). \quad (3.17)$$

A equação (3.17) é o resultado principal dessa seção e revela que a função de densidade OLLGG é uma combinação linear de densidades Exp-GG. Assim, algumas propriedades matemáticas da distribuição OLLGG podem seguir diretamente das propriedades da distribuição GG. Por exemplo, os momentos e a função geradora de momentos (fmg) da distribuição proposta podem ser obtidos a partir da mesma combinação linear infinita ponderada das quantidades correspondentes para a distribuição GG.

3.4 Propriedades matemáticas

Algumas das características mais importantes de uma distribuição podem ser estudadas através de momentos (por exemplo, tendência, dispersão, assimetria e curtose). Nesta seção, são apresentadas duas expansões diferentes para calcular os momentos da distribuição EGG.

Primeiro, é obtido uma representação de soma infinita do r -ésimo momento ordinário μ'_r para a distribuição EGG com base na equação (3.17). O r -ésimo momento para a distribuição $GG(\alpha, \tau, k)$ é dado por

$$\mu'_{r,GG} = \frac{\alpha^r \Gamma(k + r/\tau)}{\Gamma(k)}, \quad r/\tau > -k.$$

Desta forma, pela equação (3.17) tem-se

$$\mu'_r = \frac{c_1 \alpha^r \Gamma(k + r/\tau)}{\Gamma(k)} + \frac{\alpha^r}{\Gamma(k)} \sum_{j=1}^{\infty} \sum_{i=0}^{\infty} e_{j,i} c_{j+1} \Gamma(i + (j+1)k + r/\tau). \quad (3.18)$$

A equação (3.18) revela que o momento μ'_r depende das quantidades $e_{j,i}$ dadas por (3.16).

Outra representação de soma infinita para μ'_r calculando o r -ésimo momento diretamente sem requerer as quantidades $e_{j,i}$. É obtido

$$\mu'_r = \frac{\lambda \tau \alpha^{r-1}}{\Gamma(k)} \int_0^{\infty} \left(\frac{t}{\alpha}\right)^{\tau k + r - 1} \exp\left[-\left(\frac{t}{\alpha}\right)^{\tau}\right] \left\{\gamma_1\left[k, \left(\frac{t}{\alpha}\right)^{\tau}\right]\right\}^{\lambda-1} dt$$

e com a mudança de variável $x = (t/\alpha)^{\tau}$ pode ser escrita como

$$\mu'_r = \frac{\lambda \alpha^r}{\Gamma(k)^{\lambda}} \int_0^{\infty} x^{k+r/\tau-1} e^{-x} \gamma(k, x)^{\lambda-1} dx.$$

Usando a expressão (3.14) para $\gamma(k, x)$ tem-se

$$\gamma(k, x)^{\lambda-1} = \sum_{j=0}^{\infty} \sum_{m=0}^j (-1)^{j+m} \binom{\lambda-1}{j} \binom{j}{m} \gamma(k, x)^m.$$

Inserindo a última equação na expressão para μ'_r e trocando os termos, é obtido

$$\mu'_r = \frac{\lambda \alpha^r}{\Gamma(k)^{\lambda}} \sum_{j=0}^{\infty} \sum_{m=0}^j (-1)^{j+m} \binom{\lambda-1}{j} \binom{j}{m} I(k, r/\tau, m), \quad (3.19)$$

em que

$$I(k, r/\tau, m) = \int_0^\infty x^{k+r/\tau-1} e^{-x} \gamma(k, x)^m dx.$$

Com a expansão em série dada pela equação (3.14) a integral pode ser escrita como

$$I(k, r/\tau, m) = \int_0^\infty x^{k+r/\tau-1} e^{-x} \left\{ x^k \sum_{p=0}^\infty \frac{(-x)^p}{(k+p)p!} \right\}^m dx.$$

Esta integral pode ser obtida a partir das equações (24) e (25) de NADARAJAH (2008) em termos da função de Lauricella do tipo A (AARTS, 2000) definido por

$$F_A^{(n)}(a; b_1, \dots, b_n; c_1, \dots, c_n; x_1, \dots, x_n) = \sum_{m_1=0}^\infty \dots \sum_{m_n=0}^\infty \frac{(a)_{m_1+\dots+m_n} (b_1)_{m_1} \dots (b_n)_{m_n}}{(c_1)_{m_1} \dots (c_n)_{m_n}} \frac{x_1^{m_1} \dots x_n^{m_n}}{m_1! \dots m_n!},$$

em que $(a)_i$ é um fatorial ascendente definido por (com a convenção que $(a)_0 = 1$)

$$(a)_i = a(a+1) \dots (a+i-1).$$

Rotinas numéricas para o cálculo direto da função Lauricella do tipo A estão disponíveis, veja EXTON (1978) e Mathematica (TROTT, 2006). Assim

$$I(k, r/\tau, m) = k^{-m} \Gamma(r/\tau + k(m+1)) \times F_A^{(m)}(r/\tau + k(m+1); k, \dots, k; k+1, \dots, k+1; -1, \dots, -1). \quad (3.20)$$

Portanto, como uma forma alternativa para a equação (3.18), o r -ésimo momento da distribuição EGG segue de ambas as fórmulas (3.19) e (3.20) como uma soma ponderada infinita das funções de Lauricella do tipo A. Nas Figuras 3.3 e 3.4, são apresentados os gráficos da assimetria e curtose da distribuição OLLGG para alguns valores dos parâmetros.

Como os momentos não existem explicitamente são utilizadas as medidas de assimetria e curtose tratáveis computacionalmente utilizando a função quantil (3.6). Assim, podemos estudar os efeitos dos parâmetros de forma e bimodalidade na assimetria e curtose. A assimetria de Bowley apresentada por KENNEY e KEEPING (1962) é dada por

$$\mathcal{B} = \frac{Q(3/4) + Q(1/4) - 2Q(2/4)}{Q(3/4) - Q(1/4)}.$$

A curtose de Moors apresentada por MOORS (1988) é dada por

$$\mathcal{M} = \frac{Q(7/8) - Q(5/8) + Q(3/8) - Q(1/8)}{Q(6/8) - Q(2/8)},$$

em que $Q(\cdot)$ neste caso é a função quantil definida em (3.6).

As medidas apresentadas são menos sensíveis na presença de observações discrepantes e existem para distribuições na qual não é possível obter os momentos. Para a

Figura 3.3. Assimetria e curtose para a distribuição OLLGG em função de λ para alguns valores de k com $\alpha = 2$ e $\tau = 1$.

Figura 3.4. Assimetria e curtose para a distribuição OLLGG em função de τ para alguns valores de λ com $\alpha = 2$ e $k = 1$.

distribuição normal padrão, essas medidas são zero (Bowley) e 1,2331 (Moors) (ALEXANDER *et al.*, 2012).

3.5 Estimação via método de máxima verossimilhança com dados censurados

Seja T uma v.a com distribuição OLLGG (3.3) com vetor de parâmetros $\boldsymbol{\theta} = (\alpha, \tau, k, \lambda)^T$. Considerando que os dados consistem de n observações independentes $t_i = \min(T_i, C_i)$ para $i = 1, \dots, n$. Assim, o logaritmo da função de verossimilhança, considerando a distribuição OLLGG é dado por

$$\begin{aligned} l(\boldsymbol{\theta}) = & r \log \left[\frac{\lambda \tau}{\alpha \Gamma(k)} \right] - \sum_{i \in F} \left(\frac{t_i}{\alpha} \right)^\tau + (\tau k - 1) \sum_{i \in F} \log \left(\frac{t_i}{\alpha} \right) + \\ & (\lambda - 1) \sum_{i \in F} \log(u_i) + (\lambda - 1) \sum_{i \in F} \log(\bar{u}_i) - 2 \sum_{i \in F} \log[u_i^\lambda + \bar{u}_i^\lambda] + \\ & \lambda \sum_{i \in C} \log(\bar{u}_i) - \sum_{i \in C} \log[u_i^\lambda + \bar{u}_i^\lambda], \end{aligned} \quad (3.21)$$

em que $u_i = \gamma_1 \left(k, \left(\frac{t_i}{\alpha} \right)^\tau \right)$, $\bar{u}_i = 1 - u_i$, r é o número de falhas, F e C denotam os dados não censurados e censurados, respectivamente.

Os componentes do vetor score (cujos elementos são as derivadas de primeira ordem) correspondentes aos parâmetros em $\boldsymbol{\theta}$ são

$$\begin{aligned} U_\alpha(\boldsymbol{\theta}) = & -\frac{r \tau k}{\alpha} + \frac{\tau}{\alpha} \sum_{i \in F} \left(\frac{t_i}{\alpha} \right)^\tau + (\lambda - 1) \sum_{i \in F} \frac{[\dot{u}_i]_\alpha}{u_i} - (\lambda - 1) \sum_{i \in F} \frac{[\dot{u}_i]_\alpha}{\bar{u}_i} \\ & - 2\lambda \sum_{i \in F} \frac{[\dot{u}_i]_\alpha [u_i^{\lambda-1} - \bar{u}_i^{\lambda-1}]}{[u_i^\lambda + \bar{u}_i^\lambda]} - \lambda \sum_{i \in C} \frac{[\dot{u}_i]_\alpha}{\bar{u}_i} - \lambda \sum_{i \in C} \frac{[\dot{u}_i]_\alpha [u_i^{\lambda-1} - \bar{u}_i^{\lambda-1}]}{[u_i^\lambda + \bar{u}_i^\lambda]}, \end{aligned}$$

$$\begin{aligned} U_\tau(\boldsymbol{\theta}) = & \frac{r}{\tau} - \sum_{i \in F} \left(\frac{t_i}{\alpha} \right)^\tau \log \left(\frac{t_i}{\alpha} \right) + k \sum_{i \in F} \log \left(\frac{t_i}{\alpha} \right) + (\lambda - 1) \left\{ \sum_{i \in F} \frac{[\dot{u}_i]_\tau}{u_i} - \sum_{i \in F} \frac{[\dot{u}_i]_\tau}{\bar{u}_i} \right\} \\ & - 2\lambda \sum_{i \in F} \frac{[\dot{u}_i]_\tau [u_i^{\lambda-1} - \bar{u}_i^{\lambda-1}]}{[u_i^\lambda + \bar{u}_i^\lambda]} - \lambda \sum_{i \in C} \frac{[\dot{u}_i]_\tau}{\bar{u}_i} - \lambda \sum_{i \in C} \frac{[\dot{u}_i]_\tau [u_i^{\lambda-1} - \bar{u}_i^{\lambda-1}]}{[u_i^\lambda + \bar{u}_i^\lambda]}, \end{aligned}$$

$$\begin{aligned} U_k(\boldsymbol{\theta}) = & -r\psi(k) + \tau \sum_{i \in F} \log \left(\frac{t_i}{\alpha} \right) + (\lambda - 1) \sum_{i \in F} \frac{[\dot{u}_i]_k}{u_i} - (\lambda - 1) \sum_{i \in F} \frac{[\dot{u}_i]_k}{\bar{u}_i} \\ & - 2\lambda \sum_{i \in F} \frac{[\dot{u}_i]_k [u_i^{\lambda-1} - \bar{u}_i^{\lambda-1}]}{[u_i^\lambda + \bar{u}_i^\lambda]} - \lambda \sum_{i \in C} \frac{[\dot{u}_i]_k}{\bar{u}_i} - \lambda \sum_{i \in C} \frac{[\dot{u}_i]_k [u_i^{\lambda-1} - \bar{u}_i^{\lambda-1}]}{[u_i^\lambda + \bar{u}_i^\lambda]}, \end{aligned}$$

e

$$\begin{aligned} U_\lambda(\boldsymbol{\theta}) = & \frac{r}{\lambda} + \sum_{i \in F} \log(u_i) + \sum_{i \in F} \log(\bar{u}_i) - 2 \sum_{i \in F} \frac{u_i^\lambda \log(u_i) + \bar{u}_i^\lambda \log(\bar{u}_i)}{[u_i^\lambda + \bar{u}_i^\lambda]} \\ & + \sum_{i \in C} \log(\bar{u}_i) - \sum_{i \in C} \frac{u_i^\lambda \log(u_i) + \bar{u}_i^\lambda \log(\bar{u}_i)}{[u_i^\lambda + \bar{u}_i^\lambda]}, \end{aligned}$$

em que

$$[\dot{u}_i]_\alpha = \frac{\partial \gamma_1 \left(k, \left(\frac{t_i}{\alpha} \right)^\tau \right)}{\partial \alpha}, \quad [\dot{u}_i]_\tau = \frac{\partial \gamma_1 \left(k, \left(\frac{t_i}{\alpha} \right)^\tau \right)}{\partial \tau}, \quad [\dot{u}_i]_k = \frac{\partial \gamma_1 \left(k, \left(\frac{t_i}{\alpha} \right)^\tau \right)}{\partial k},$$

$\psi(\cdot)$ é a função digama com $i = 1, \dots, n$.

As estimativas de máxima verossimilhança (EMVs) $\hat{\theta}$ de θ são obtidas numericamente a partir das equações não lineares $U_\alpha(\theta) = U_\tau(\theta) = U_k(\theta) = U_\lambda(\theta) = 0$. A construção de intervalos de confiança e realização de testes de hipóteses sobre os parâmetros do modelo são realizados baseando-se na teoria assintótica dos EMV. Assim, sob certas condições de regularidade, a distribuição assintótica das EMV é dada por $(\hat{\theta} - \theta) \sim N_4(0, I(\theta)^{-1})$, em que $I(\theta)$ é a matriz de informação esperada. Esta matriz pode ser substituída por $J(\hat{\theta})$ que é a matriz de informação observada avaliada em $\hat{\theta}$. Além da construção de intervalos de confiança aproximados para os parâmetros individuais pode-se também usar o teste da razão de verossimilhança (*TRV*) para comparar esta distribuição com alguns de seus submodelos especiais. Por exemplo, o teste de $H_0 : \lambda = 1$ versus $H_1 : \lambda \neq 1$, o que é equivalente a comparar a distribuição OLLGG com a distribuição GG e o *TRV* é dado por

$$w = 2\{\ell(\hat{\alpha}, \hat{\tau}, \hat{k}, \hat{\lambda}) - \ell(\tilde{\alpha}, \tilde{\tau}, \tilde{k}, 1)\},$$

em que $\hat{\alpha}$, $\hat{\tau}$, $\hat{k}$, e $\hat{\lambda}$ são as EMVs sob H_1 e $\tilde{\alpha}$, $\tilde{\tau}$ e $\tilde{k}$ são as EMVs sob H_0 .

3.6 Estudo via simulação

Nesta seção, um estudo de simulação é apresentado para avaliar algumas propriedades das EMVs, usando a sub-rotina *optim* pelo método BFGS do *software* R (R DEVELOPMENT CORE TEAM, 2014). Desta forma, a distribuição OLLGG é simulada considerando a forma bimodal, a partir da equação (3.6) usando uma v.a U com distribuição uniforme (0,1). Assim, usando simulação de Monte Carlo considerando três diferentes tamanhos de amostras $n = 50$, $n = 150$ e $n = 350$. O processo é repetido 1000 vezes e as estimativas médias (EMs), desvios padrão e os erros quadráticos médio (EQM) são calculados para os parâmetros α , τ , k e λ .

3.6.1 Estudo via simulação sem censura

Neste estudo, são considerados dois cenários. No primeiro cenário é considerado que os valores dos parâmetros são $\alpha = 2$, $\tau = 5$, $k = 10$ e $\lambda = 0, 15$. No segundo cenário, são usados os valores estimados no ajuste aos dados reais de temperatura, Subseção 3.7.1 ($\alpha = 21, 2911$, $\tau = 13, 0661$, $k = 2, 8755$, $\lambda = 0, 2882$). As estimativas para α , τ , k e λ são determinadas resolvendo o sistema de equações não lineares $U_\alpha(\theta) = U_\tau(\theta) = U_k(\theta) = U_\lambda(\theta) = 0$. Os resultados do estudo pelo processo de Monte Carlo são dados na Tabela 3.1. É verificado que os EQMs e os desvios padrão das EMVs de α , τ , k e λ diminuem a medida que o tamanho das amostras aumentam, o que é esperado sob a teoria assintótica padrão. Nas Figuras 3.5 e 3.6, são apresentadas as fdp estimadas com base em 1000 amostras das EMs dos parâmetros e a fdp com os verdadeiros valores dos parâmetros. Os

comportamentos observados nos gráficos estão de acordo com a teoria assintótica padrão para EMVs.

Tabela 3.1. EMs, desvios padrão e erros quadráticos médio (EQM) para os parâmetros da distribuição OLLGG.

cenário 1					
n	Parâmetros	Valor Real	EMs	Desvios Padrão	EQM
50	α	2,00	2,0404	0,4436	0,1984
	τ	5,00	5,3257	1,3214	1,8523
	k	10,00	10,7653	5,4143	29,9000
	λ	0,15	0,1708	0,1053	0,0115
150	α	2,00	2,0393	0,1504	0,0242
	τ	5,00	5,1585	0,4265	0,2070
	k	10,00	9,8491	1,4045	1,9955
	λ	0,15	0,1528	0,0330	0,0011
350	α	2,00	2,0065	0,0483	0,0024
	τ	5,00	5,0417	0,1608	0,0276
	k	10,00	10,012	0,4711	0,2220
	λ	0,15	0,1511	0,0113	0,0001
cenário 2					
n	Parâmetros	Valor Real	EMs	Desvios Padrão	EQM
50	α	21,2911	21,1422	2,7809	7,7557
	τ	13,0661	15,5491	7,6516	64,7128
	k	2,8755	4,5288	4,4186	22,2571
	λ	0,2882	0,3400	0,2565	0,0685
150	α	21,2911	21,3407	1,4791	2,1903
	τ	13,0661	13,8973	3,0415	9,9415
	k	2,8755	3,2666	1,7958	3,3779
	λ	0,2882	0,3060	0,1279	0,0167
350	α	21,2911	21,2908	0,9161	0,8393
	τ	13,0661	13,3138	1,7378	3,0814
	k	2,8755	3,0593	1,0807	1,2018
	λ	0,2882	0,2956	0,0757	0,0058

3.6.2 Estudo via simulação considerando censura aleatória

Um estudo de simulação é considerado para dados com a presença de censura. Os tempos de censura C_i são gerados a partir de uma distribuição uniforme no intervalo $(0, \nu)$, em que ν denota a proporção de observações censuradas. Neste estudo, as proporções de observações censuradas são aproximadamente iguais a 10% e 30%. Neste cenário os valores considerados para os parâmetros são $\alpha = 2$, $\tau = 5$, $k = 10$, $\lambda = 0,15$. Na Tabela 3.2 são listadas as EMs e os EQMs. Os valores apresentados nessa tabela indicam que o EQM aumenta quando a porcentagem de observações censuradas aumenta. Além disso, os EQMs das EMVs de α , τ , k e λ tende a zero quando o tamanho da amostra aumenta, como esperado sob a teoria assintótica padrão. Na Figura 3.7 são apresentadas as fdp real

Figura 3.5. Comportamento da fdp OLLGG real e estimada. (a) $n = 50$. (b) $n = 150$. (c) $n = 350$.

Figura 3.6. Comportamento da fdp OLLGG real e estimada. (a) $n = 50$. (b) $n = 150$. (c) $n = 350$.

e estimadas considerando as diferentes proporções de censura e tamanhos de amostras. Desta forma, estes gráficos auxiliam na representação do comportamento da distribuição, estando de acordo com a teoria assintótica padrão para as EMVs.

3.7 Aplicações

Nesta seção, são apresentadas duas aplicações para dois conjuntos de dados reais como forma de mostrar empiricamente a flexibilidade do modelo OLLGG. Os resultados foram obtidos utilizando os *softwares* R (R DEVELOPMENT CORE TEAM, 2014) e SAS

Tabela 3.2. EMs e erros quadráticos médio (EQM) para os parâmetros da distribuição OLLGG.

n	Parâmetros	Valor Real	0%	10%	30%
50	α	2,00	2,0404 (0,1984)	2,0366 (0,2257)	2,0441 (0,2836)
	τ	5,00	5,3257 (1,8523)	5,3950 (3,3121)	5,5626 (4,3955)
	k	10,00	10,7653 (29,900)	10,9566 (32,046)	11,2739 (36,305)
	λ	0,15	0,1708 (0,0115)	0,1708 (0,0149)	0,1736 (0,0201)
150	α	2,00	2,0393 (0,0242)	2,0382 (0,0322)	2,0427 (0,0621)
	τ	5,00	5,1585 (0,2070)	5,1763 (0,2784)	5,2257 (0,5882)
	k	10,00	9,8491 (1,9955)	9,9663 (3,2201)	10,0686 (7,2771)
	λ	0,15	0,1528 (0,0011)	0,1521 (0,0015)	0,1539 (0,0022)
350	α	2,00	2,0065 (0,0024)	2,0089 (0,0033)	2,0181 (0,0115)
	τ	5,00	5,0417 (0,0276)	5,0483 (0,0315)	5,0823 (0,0969)
	k	10,00	10,012 (0,2220)	9,9941 (0,3281)	9,9645 (1,3263)
	λ	0,15	0,1511 (0,0001)	0,1506 (0,0002)	0,1510 (0,0005)

(SAS ONDEMAND FOR ACADEMICS, 2015) no procedimento NLMIXED. Na primeira aplicação, é utilizado um conjunto de dados com bimodalidade. Na segunda aplicação, é analisado a flexibilidade da distribuição OLLGG para um conjunto de dados com censura. Em ambas as aplicações o modelo OLLGG é comparado com os modelos GG, Weibull, gama generalizada exponenciada (EGG), beta Weibull (BW) e Kumaraswamy gama generalizada (KumGG).

3.7.1 Aplicação 1

Os dados usados nesta aplicação são as temperaturas em ($^{\circ}C$) diárias correspondente ao período de 01 de Janeiro a 31 de Dezembro de 2011 da cidade de Piracicaba, obtidos no Departamento de Engenharia de Biossistemas da Escola Superior de Agricultura Luiz de Queiroz (ESALQ), da Universidade de São Paulo (USP).

Nesta aplicação é verificado a flexibilidade da distribuição OLLGG em comparação com alguns dos seus submodelos e também para os seguintes modelos não aninhados: as distribuições gama generalizada exponenciada (EGG) proposta por CORDEIRO *et al.* (2011) e beta Weibull (BW). A função de distribuição BW foi introduzida por FAMOYE *et al.* (2005) é dada por

$$F(t) = \frac{1}{B(a, b)} \int_0^{\{1 - \exp[-(t/\alpha)^\gamma]\}} w^{a-1} (1 - w)^{b-1} dw.$$

A distribuição Kumaraswamy gama generalizada (KumGG) (para $t > 0$) é definida por PASCOA *et al.* (2011). Assim, a fdp associada com cinco parâmetros positivos

Figura 3.7. Comportamento da fdp OLLGG real e estimada considerando 0%, 10% e 30% de censura para os tamanhos de amostras $n = 50$, 150 e 350.

α , τ , k , λ e φ tem a forma

$$f(t) = \frac{\lambda \varphi \tau}{\alpha \Gamma(k)} \left(\frac{t}{\alpha}\right)^{\tau k-1} \exp\left[-\left(\frac{t}{\alpha}\right)^\tau\right] \left\{\gamma_1\left[k, \left(\frac{t}{\alpha}\right)^\tau\right]\right\}^{\lambda-1} \times \\ \left(1 - \left\{\gamma_1\left[k, \left(\frac{t}{\alpha}\right)^\tau\right]\right\}^\lambda\right)^{\varphi-1}, \quad (3.22)$$

em que $\gamma_1(k, x) = \gamma(k, x)/\Gamma(k)$ é a função gama incompleta, α é parâmetro de escala e os outros parâmetros positivos τ , k , φ e λ são parâmetros de forma.

Desta forma, obtemos os EMVs (e os correspondentes erros padrão entre parênteses) dos parâmetros e os valores do Critério de Informação de Akaike (AIC), Critério de Informações Consistentes (CAIC) e Critério de Informação Bayesiano (BIC). Quanto menor estes critérios, melhor o ajuste. Em cada caso, os parâmetros são estimados via método de máxima verossimilhança, utilizando o procedimento NLMIXED no SAS.

Os resultados são apresentados na Tabela 3.3. Os três critérios de informação concordam com a classificação do modelo. Ou seja, os três critérios utilizados apresentam menores valores para o modelo OLLGG comparado com os outros modelos ajustados. Portanto, o modelo OLLGG é o mais adequado para os dados de temperatura. Para

Tabela 3.3. EMVs para os parâmetros dos modelos OLLGG, KumGG, EGG, GG, Weibull e BW para os dados de temperatura, correspondentes erros padrão (entre parenteses) e os valores das estatísticas AIC, CAIC e BIC.

Modelos	α	τ	k	λ	φ	AIC	CAIC	BIC
OLLGG	21,2911 (0,0012)	13,0661 (0,0234)	2,8755 (0,1095)	0,2882 (0,0127)		1752,1	1752,2	1767,7
KumGG	25,3965 (1,6147)	25,2759 (3,0850)	12,8897 (0,6885)	0,0243 (0,0079)	2,3730 (2,4887)	1780,6	1780,7	1800,1
EGG	23,8850 (2,8175)	22,9475 (7,7331)	12,8766 (9,1805)	0,0215 (0,0019)	1 (-)	1777,6	1777,7	1793,2
GG	26,1868 (0,1877)	33,1789 (7,5737)	0,1888 (0,0514)	1 (-)		1776,6	1776,7	1788,3
Weibull	23,5808 (0,1376)	9,4296 (0,4038)	1 (-)	1 (-)		1796,4	1796,5	1804,2
	α	γ	a	b				
BW	25,0516 (1,3335)	25,7636 (8,2195)	0,2460 (0,0858)	0,6159 (0,4512)		1778,1	1778,3	1793,7

ilustrar a utilização da distribuição OLLGG para discriminação de modelos, é realizado por meio do *TRV* a comparação do modelo proposto com os modelos GG e Weibull. Os resultados do *TRV* são apresentados na Tabela 3.4. As fdp ajustadas ao histograma e a fda dos modelos são apresentadas na Figura 3.8. Os resultados da Tabela 3.4 sugerem que o modelo OLLGG possui o ajuste mais adequado aos dados de temperatura quando comparado com os outros modelos. A rejeição é significativa e fornece evidência da ne-

Tabela 3.4. *TRV* para os dados de temperatura.

Modelos	Hipóteses	Estatísticas w	p -valor
OLLGG vs GG	$H_0 : \lambda = 1$ vs $H_1 : H_0$ é falso	26,5	<0,0001
OLLGG vs Weibull	$H_0 : \lambda = k = 1$ vs $H_1 : H_0$ é falso	48,3	<0,0001

cessidade de incluir o parâmetro de forma λ na modelagem destes dados reais. Mais uma informação é obtida por uma comparação visual do histograma dos dados e das fdp e das fda Figuras 3.8a e 3.8b, respectivamente. Assim, é possível verificar que de fato a distribuição OLLGG apresenta bom ajuste quando os dados apresentam bimodalidade.

Figura 3.8. (a) Funções de densidade estimadas para os modelos OLLGG, GG e Weibull para os dados de temperatura. (b) Funções de distribuição estimadas para os modelos OLLGG, GG e Weibull e distribuição empírica para os dados de temperatura.

3.7.2 Aplicação 2

A Aids é uma patologia que mobiliza seus portadores por causa das implicações em suas relações interpessoais e de reprodução. Avanços terapêuticos tem possibilitado à mulheres soropositivas terem filhos com segurança. A este respeito, o serviço de imunologia pediátrica e serviço social do Hospital das Clínicas tem um programa especial para cuidar de recém-nascidos de mães soropositivas, para fornecer orientação e apoio para a terapia antirretroviral, para permitir que essas mulheres e seus bebês possam viver o mais normalmente possível. Neste estudo, é considerado o conjunto de dados de tempo até a soro-reversão de 148 crianças expostas ao HIV por via vertical, dos quais 56% são censurados, nascidas no Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto entre 1986 e 2001, e cujas mães não foram tratadas ((SILVA, 2004); (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, em que mais ocorre, 65% dos casos; ou na amamentação, variando de 7% a 22%.

A soro-reversão ou reversão sorológica pode ocorrer em crianças nascidas de mães contaminadas com o vírus HIV. Este é o processo de desaparecimento dos anticorpos anti-HIV do sangue em indivíduos que apresenta 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, a qual foi administrada durante 6 semanas. Assume-se que os tempos de vida são distribuídos de forma independente, e também independente do mecanismo de censura.

A curva TTT-plot, definida na Subseção 2.1.2, para o conjunto de dados de Aids é apresentada na Figura 3.9 e indica uma função de risco crescente. Desta forma, foi considerada a distribuição OLLGG e algumas outras distribuições conhecidas para analisar esse conjunto de dados.

Na Tabela 3.5, é apresentado as EMVs (e os correspondentes erros padrão que estão entre parenteses) dos parâmetros e os valores das estatísticas AIC, BIC e CAIC. Estes resultados indicam que o modelo OLLGG tem os menores valores AIC, BIC e CAIC entre os submodelos GG e Weibull, e, portanto, pode ser escolhido como modelo mais adequado.

Nota-se que o modelo KumGG é competitivo com o modelo proposto. No entanto, o modelo KumGG tem duas desvantagens:

- O modelo KumGG não modela dados com comportamento bimodal.
- O modelo KumGG tem cinco parâmetros, ou seja, é menos parcimonioso.

Uma comparação da distribuição proposta com alguns dos seus submodelos usando estatísticas TRV é apresentada na Tabela 3.6. Os valores nesta tabela, especialmente os p -valores, sugerem que o modelo OLLGG produz um melhor ajuste para estes dados do que as outras três distribuições. A fim de avaliar se o modelo está adequado, são apresentadas na Figura 3.10 as funções de sobrevivência estimadas das distribuições OLLGG, KumGG, EGG, GG, Weibull e BW e a função de sobrevivência empírica. E como uma outra alternativa para verificar a qualidade do ajuste, na Figura 3.11 apresentamos as funções de risco estimadas das distribuições OLLGG, KumGG, EGG, GG, Weibull e BW e a função de risco empírica. Conclui-se que a distribuição OLLGG fornece um bom ajuste para estes dados.

Tabela 3.5. EMVs dos parâmetros dos modelos OLLGG, KumGG, EGG, GG, Weibull e BW para os dados de Aids, correspondentes erros padrão (entre parenteses) e valores das estatísticas AIC, BIC e CAIC.

Modelos	α	τ	k	λ	φ	AIC	BIC	CAIC
OLLGG	352,0 (1,0590)	46,9706 (1,4847)	0,1043 (0,0324)	0,4468 (0,0881)		771,1	783,6	771,9
KumGG	350,05 (1,5707)	49,8303 (5,8895)	0,2176 (0,0073)	0,1282 (0,0236)	0,3424 (0,0522)	770,7	785,7	771,1
EGG	350,45 (2,4187)	22,2991 (0,0375)	1,0741 (0,0004)	0,1072 (0,0113)	1 (-)	798,1	810,1	798,3
GG	379,40 (8,8211)	24,5312 (10,3258)	0,0974 (0,0402)	1 (-)	1 (-)	783,7	792,7	783,9
Weibull	307,62 (12,3523)	3,1132 (0,3250)	1 (-)	1 (-)	1 (-)	808,0	814,0	808,1
	α	γ	a	b				
BW	349,99 (23,0923)	6,3895 (0,7657)	0,3944 (0,0468)	0,9273 (0,3361)		797,9	809,9	798,2

Tabela 3.6. TRV para os dados de Aids.

Modelos	Hipóteses	Estatísticas w	p -valor
OLLGG vs GG	$H_0 : \lambda = 1$ vs $H_1 : H_0$ é falso	13,0	0,00031
OLLGG vs Weibull	$H_0 : \varphi = \lambda = k = 1$ vs $H_1 : H_0$ é falso	40,3	<0,0001

Figura 3.9. Curva TTT-plot para os dados de Aids.

Figura 3.10. Função de sobrevivência estimada com ajuste da distribuição OLLGG e alguns outros modelos e sobrevivência empírica para os dados de Aids. (a) OLLGG vs KGG e GG. (b) OLLGG vs BW, EGG e Weibull.

Figura 3.11. Função de risco estimada com ajuste da distribuição OLLGG e alguns outros modelos e risco empírico para os dados de Aids. (a) OLLGG vs KGG e GG. (b) OLLGG vs BW, EGG e Weibull.

3.8 Conclusões

Neste capítulo, foi proposto uma nova distribuição de probabilidade que possui quatro parâmetros e é chamada odd log-logística gama generalizada (OLLGG). Por meio dos estudos realizados verifica-se que a distribuição OLLGG, além de possibilitar a mode-

lagem de dados com bimodalidade, a sua função de riscos possui formas bastante flexíveis, como forma de U, crescente, decrescente e constante e uma nova forma denominada, forma de N. O fato de ter como casos particulares distribuições conhecidas - Weibull, exponencial, gama, GG, etc - permite testar a qualidade do ajuste com tais modelos. O ajuste e o processo de estimação foi realizado utilizando o método de máxima verossimilhança por meio de rotinas computacionais implementadas no *software* R e considerando os conjuntos de dados reais o *software* SAS. A adequabilidade do modelo é apresentada em duas situações práticas. Nestas aplicações, são verificados que o modelo pode ser uma alternativa quando se tem dados com bimodalidade e na presença de censura.

4 DISTRIBUIÇÃO LOG-ODD LOG-LOGÍSTICA GAMA GENERALIZADA

4.1 Introdução

Recentemente, diferentes generalizações e modificações para a distribuição GG para modelar tempo de vida foram propostas na literatura. Por exemplo, BARKAUSKAS *et al.* (2009) estudaram a parte de ruído de um espectro como modelo de média móvel autorregressiva (ARMA) com as inovações da distribuição GG, MALHOTRA *et al.* (2009) apresentaram uma análise unificada para o sistema sem fio sobre canais de desvanecimento generalizado que é modelado pela GG e CORDEIRO *et al.* (2011) definiram a distribuição gama generalizada exponenciada (EGG). Além disso, PASCOA *et al.* (2011) propuseram a distribuição Kumaraswamy-GG com aplicações em análise de sobrevivência, ORTEGA *et al.* (2011) definiram o modelo GG geométrica, CORDEIRO *et al.* (2013b) introduziram a distribuição beta GG, CORDEIRO *et al.* (2014) estudaram a distribuição Kummer beta GG, LUCENA *et al.* (2015) definiram a distribuição GG transmutada com propriedades e aplicações, BOURGUIGNON *et al.* (2015) investigaram uma distribuição GG alternativa e SILVA *et al.* (2016) definiram a distribuição série de potência GG. Ressalta-se que em todos esses artigos não é discutida uma estrutura de regressão.

Neste capítulo, é proposto uma extensão da distribuição OLLGG em que o parâmetro τ pode ser positivo ou negativo, que tem flexibilidade para acomodar várias formas da função de risco, como crescente, decrescente, em forma de U e unimodal. Por facilidade de notação esta nova extensão é denominada também de OLLGG. Este modelo também é apropriado para discriminar entre alguns modelos contínuos alternativos, como as distribuições odd log-logística Weibull (OLLW) proposto por da CRUZ *et al.* (2016), GG, Weibull e exponencial, comumente usados na análise do tempo de sobrevivência. Portanto, é esperado que a nova distribuição seja útil em uma variedade de aplicações na análise de dados de sobrevivência e confiabilidade.

Além disso, é comum em situações práticas que o tempo de sobrevida seja relacionado às covariáveis que ajudam a explicar a sua variabilidade deste tempo. Para estudar os efeitos dessas covariáveis sobre o tempo de sobrevivência, é necessário usar um modelo de regressão apropriado para dados censurados e para os tempos de falha. Em outras palavras, para formular modelos de regressão paramétrico, é necessário definir um modelo probabilístico para o tempo de sobrevivência. Portanto, outro objetivo deste trabalho é propor modelos de regressão, na forma de locação e escala, chamado modelo de regressão *log-odd log-logística gama generalizada* (LOLLGG), para dados censurados que têm como casos particulares os modelos de regressão log-gama generalizada (LGG), log-odd log-logística Weibull (LOLLW) e log-Weibull (LW), entre outros.

Neste estudo, são realizados os estudos dos resíduos para avaliar os pressupostos

em relação ao erro e para detectar observações extremas no modelo de regressão LOLLGG com observações censuradas. São realizados estudos de simulação para diferentes configurações de parâmetros, tamanhos de amostra e porcentagens de censura e a distribuição empírica dos resíduos é comparada com a distribuição normal padrão.

O restante do capítulo está organizado da seguinte forma. Na Seção 4.2, é definido uma versão estendida da distribuição GG. Na Seção 4.3, algumas propriedades estruturais para a fdp da OLLGG e expressões explícitas para os momentos ordinários, momentos incompletos, a função geradora de momentos (fmg) e os desvios médios são apresentados. Na Seção 4.4, é definida a distribuição LOLLGG. Na Seção 4.5, é proposto o modelo de regressão LOLLGG para análise de dados censurados e estimação dos parâmetros do modelo por máxima verossimilhança. Na Seção 4.6, são examinados os resíduos e apresentados os resultados de alguns estudos via simulação. Na Seção 4.7, a potencialidade do novo modelo de regressão é ilustrado por meio de um conjunto de dados reais de níveis de voltagem. Seção 4.8, termina com algumas observações e comentários.

4.2 Distribuição odd log-logística gama generalizada estendida

Desenvolvimentos recentes visam definir novas famílias que estendem distribuições bem conhecidas e ao mesmo tempo, oferecem grande flexibilidade na modelagem de dados na prática. Por exemplo, a distribuição Weibull de dois parâmetros tem sido o modelo mais popular para dados de vida útil. No entanto, a sua principal fraqueza é a sua incapacidade de acomodar as formas da função taxa de falha não-monótona (em particular, banheira e formas bimodais). Esse fato, leva a buscar generalizações dessa distribuição. O modelo GG (STACY (1962), STACY e MIHRAM (1965)) tem fdp e fda dadas por

$$g(t; \alpha, \tau, k) = \frac{|\tau|}{\alpha \Gamma(k)} \left(\frac{t}{\alpha}\right)^{\tau k - 1} \exp \left[- \left(\frac{t}{\alpha}\right)^{\tau} \right] \quad (4.1)$$

e

$$G(t; \alpha, \tau, k) = \begin{cases} \gamma_1 \left(k, \left(\frac{t}{\alpha}\right)^{\tau}\right) = \frac{\gamma(k, \left(\frac{t}{\alpha}\right)^{\tau})}{\Gamma(k)}, & \text{se } \tau > 0, \\ 1 - \gamma_1 \left(k, \left(\frac{t}{\alpha}\right)^{\tau}\right) = 1 - \frac{\gamma(k, \left(\frac{t}{\alpha}\right)^{\tau})}{\Gamma(k)}, & \text{se } \tau < 0, \end{cases} \quad (4.2)$$

respectivamente, em que $\tau \neq 0$ e $k > 0$ são os parâmetros de forma, $\alpha > 0$ é o parâmetro de escala, $\Gamma(k) = \int_0^{\infty} w^{k-1} e^{-w} dw$ (para $k > 0$) é a função gama, $\gamma_1(k, x) = \gamma(k, x)/\Gamma(k)$ é a razão de uma função gama incompleta e $\gamma(k, x) = \int_0^x w^{k-1} e^{-w} dw$ é a função gama incompleta. A função gama incompleta é definida aqui para os reais positivos k e x .

Seguindo a mesma ideia de GLEATON e LYNCH (2006), CRUZ *et al.* (2016) e CORDEIRO *et al.* (2017), definimos a fda OLLGG com um parâmetro de forma adicional

$\lambda > 0$ por

$$F(t) = \begin{cases} \int_0^{\frac{\gamma_1(k, (t/\alpha)^\tau)}{1-\gamma_1(k, (t/\alpha)^\tau)}} \frac{\lambda x^{\lambda-1}}{(1+x^\lambda)^2} dx, & \text{se } \tau > 0, \\ \int_0^{\frac{1-\gamma_1(k, (t/\alpha)^\tau)}{\gamma_1(k, (t/\alpha)^\tau)}} \frac{\lambda x^{\lambda-1}}{(1+x^\lambda)^2} dx, & \text{se } \tau < 0, \end{cases} \quad (4.3)$$

com

$$\lambda = \frac{\log \left[\frac{F(t; \alpha, \tau, k, \lambda)}{1-F(t; \alpha, \tau, k, \lambda)} \right]}{\log \left[\frac{\gamma_1(k, (t/\alpha)^\tau)}{1-\gamma_1(k, (t/\alpha)^\tau)} \right]}.$$

Assim, o parâmetro λ representa o quociente do logaritmo da razão entre as distribuições gerada e de base. A fda GG (4.2) é claramente um caso especial de (4.3) no caso em que $\lambda = 1$.

Neste estudo, definimos a fdp OLLGG (para $t > 0$) como

$$f(t) = \frac{\lambda |\tau| (t/\alpha)^{\tau k-1} \exp[-(t/\alpha)^\tau] \{\gamma_1(k, (t/\alpha)^\tau) [1 - \gamma_1(k, (t/\alpha)^\tau)]\}^{\lambda-1}}{\alpha \Gamma(k) \{\gamma_1^\lambda(k, (t/\alpha)^\tau) + [1 - \gamma_1(k, (t/\alpha)^\tau)]^\lambda\}^2}, \quad (4.4)$$

em que τ é diferente de zero e os outros parâmetros são positivos. A terminologia distribuição OLLGG é adotada para (4.4). Doravante, para uma v.a T tendo esta fdp, denota-se $T \sim \text{OLLGG}(\alpha, \tau, k, \lambda)$. Para $\tau > 0$ e $\lambda = 1$, é idêntica à distribuição GG amplamente utilizada por muitos pesquisadores. Para $\tau > 0$ e $\lambda = k = 1$, obtemos a distribuição *Weibull*. Na Tabela 4.1 são listados treze submodelos para $\tau > 0$. A correspondente *odd log-logística recíproca gama generalizada* corresponde para $\tau < 0$. Os casos especiais para $\tau < 0$ são listados na Tabela 4.2. Os gráficos da fdp OLLGG para valores $\tau > 0$ e $\tau < 0$

Tabela 4.1. Alguns submodelos para $\tau > 0$.

Distribuição	α	τ	k	λ
OLL-Gama	α	1	k	λ
OLL-Weibull	α	τ	1	λ
OLL-Exponencial	α	1	1	λ
OLL-Qui-quadrado	2	1	$\frac{n}{2}$	λ
OLL-Qui	$\sqrt{2}$	2	$\frac{n}{2}$	λ
OLL-Qui escala	$\sqrt{2}\sigma$	2	$\frac{n}{2}$	λ
OLL-Rayleigh	α	2	1	λ
OLL-Maxwell	α	2	$\frac{3}{2}$	λ
OLL-Folded normal	$\sqrt{2}$	2	$\frac{1}{2}$	λ
OLL-Normal circular	$\sqrt{2}$	2	1	λ
OLL-Normal esférica	$\sqrt{2}$	2	$\frac{3}{2}$	λ
OLL-Semi-normal generalizada	$2^{\frac{1}{2\gamma}}\theta$	2γ	$\frac{1}{2}$	λ
OLL-Semi-normal	$2^{\frac{1}{2}}\theta$	2	$\frac{1}{2}$	λ

são apresentados na Figura 4.1. Para $\lambda = 1$, é obtido a fdp (STACY e MIHRAM, 1965)

Tabela 4.2. Alguns submodelos para $\tau < 0$.

Distribuição	α	τ	k	λ
OLL-Recíproca Gamma	α	-1	k	λ
OLL-Recíproca Weibull	α	$-\tau$	1	λ
OLL-Recíproca Exponencial	α	-1	1	λ
OLL-Recíproca Qui-quadrado	2	-1	$\frac{n}{2}$	λ
OLL-Recíproca Qui	$\sqrt{2}$	-2	$\frac{n}{2}$	λ
OLL-Recíproca Qui escala	$\sqrt{2}\sigma$	-2	$\frac{n}{2}$	λ
OLL-Recíproca Rayleigh	α	-2	1	λ
OLL-Recíproca Maxwell	α	-2	$\frac{3}{2}$	λ
OLL-Recíproca Folded normal	$\sqrt{2}$	-2	$\frac{1}{2}$	λ
OLL-Recíproca Normal circular	$\sqrt{2}$	-2	1	λ
OLL-Recíproca Normal esférica	$\sqrt{2}$	-2	$\frac{3}{2}$	λ
OLL-Recíproca Semi-normal generalizada	$2^{\frac{1}{2\gamma}}\theta$	-2γ	$\frac{1}{2}$	λ
OLL-Recíproca Semi-normal	$2^{\frac{1}{2}}\theta$	-2	$\frac{1}{2}$	λ

Figura 4.1. Gráficos da pdf OLLGG considerando diferentes valores dos parâmetros. (a) Para alguns valores de $\tau < 0$ e $\alpha = 1$ fixo. (b) Para alguns valores de $\tau > 0$ e $\alpha = 1$ fixo. (c) Para alguns valores de $\tau < 0$ e $\tau > 0$, $\alpha = 2$ e $\lambda = 0, 15$ fixos.

dada pela equação (4.1).

A fda correspondente para (4.4) é dada por

$$F(t) = \begin{cases} \frac{\gamma_1^\lambda(k, (t/\alpha)^\tau)}{\gamma_1^\lambda(k, (t/\alpha)^\tau) + [1 - \gamma_1(k, (t/\alpha)^\tau)]^\lambda}, & \text{se } \tau > 0, \\ \frac{[1 - \gamma_1(k, (t/\alpha)^\tau)]^\lambda}{\gamma_1^\lambda(k, (t/\alpha)^\tau) + [1 - \gamma_1(k, (t/\alpha)^\tau)]^\lambda}, & \text{se } \tau < 0. \end{cases} \quad (4.5)$$

A função de risco pode ser constante, decrescente, crescente, forma de banheira para baixo (unimodal), forma de U e forma bimodal. A função de risco para T é dada

por

$$h(t) = \begin{cases} \frac{\lambda \tau (t/\alpha)^{\tau k-1} \exp[-(t/\alpha)^\tau] \gamma_1^{\lambda-1}(k, (\frac{t}{\alpha})^\tau)}{\alpha \Gamma(k) \left\{ \gamma_1^\lambda(k, (\frac{t}{\alpha})^\tau) + [1 - \gamma_1(k, (\frac{t}{\alpha})^\tau)]^\lambda \right\} [1 - \gamma_1(k, (\frac{t}{\alpha})^\tau)]}, & \text{se } \tau > 0, \\ \frac{\lambda (-\tau) (t/\alpha)^{\tau k-1} \exp[-(t/\alpha)^\tau] [1 - \gamma_1(k, (\frac{t}{\alpha})^\tau)]^{\lambda-1}}{\alpha \Gamma(k) \left\{ \gamma_1^\lambda(k, (\frac{t}{\alpha})^\tau) + [1 - \gamma_1(k, (\frac{t}{\alpha})^\tau)]^\lambda \right\} \gamma_1(k, (\frac{t}{\alpha})^\tau)}, & \text{se } \tau < 0. \end{cases} \quad (4.6)$$

A Figura 4.2 mostra o comportamento da função de risco dada pela equação (4.6) para diferentes valores de parâmetros. Podemos observar varias formas nos gráficos de $h(t)$ como o crescente, decrescente, unimodal, forma de U entre outras. Essas formas indicam que a distribuição pode ser um modelo interessante na modelagem de dados de sobrevivência em diversas condições.

4.3 Extensões úteis e propriedades estruturais

Em um contexto geral, as propriedades das distribuições exponenciada-G (Exp-G) foram estudadas por vários autores para alguns modelos de base G, ver por exemplo, CORDEIRO *et al.* (2017). Nesta seção, é demonstrado que a fdp OLLGG é uma combinação linear de densidades Exp-GG.

Inicialmente, é definido a distribuição gama generalizada exponenciada (“Exp-GG”), denotada por $W \sim \text{Exp}^c[G(\alpha, \tau, k)]$ com parâmetro de potência $c > 0$, se W têm fda e fdp dadas por

$$H_c(t) = G(t; \alpha, \tau, k)^c \quad \text{e} \quad h_c(t) = \frac{c|\tau|}{\alpha \Gamma(k)} \left(\frac{t}{\alpha}\right)^{\tau k-1} \exp\left[-\left(\frac{t}{\alpha}\right)^\tau\right] G(t; \alpha, \tau, k)^{c-1},$$

respectivamente. A seguir, é obtido uma expansão para $F(t)$ em (4.2) usando uma série de potência $G(t; \alpha, \tau, k)^\lambda$ ($\lambda > 0$ real)

$$G(t; \alpha, \tau, k)^\lambda = \sum_{j=0}^{\infty} a_j G(t; \alpha, \tau, k)^j, \quad (4.7)$$

em que

$$a_j = a_j(\lambda) = \sum_{j=u}^{\infty} (-1)^{u+j} \binom{\lambda}{u} \binom{u}{j}.$$

A expansão binomial generalizada converge para $|z| < 1$ e um real qualquer $b > 0$

$$(1 - z)^b = \sum_{j=0}^{\infty} (-1)^j \binom{b}{j} z^j, \quad (4.8)$$

em que $\binom{b}{j} = \Gamma(b+1)/[\Gamma(b-j+1)j!]$.

Para qualquer $\lambda > 0$ real, usando a equação (4.8) para obter

$$[1 - G(t; \alpha, \tau, k)]^\lambda = \sum_{j=0}^{\infty} (-1)^j \binom{\lambda}{j} G(t; \alpha, \tau, k)^j. \quad (4.9)$$

Figura 4.2. Gráficos da função de risco para a distribuição OLLGG considerando diferentes valores dos parâmetros. (a) Para alguns valores de $\tau > 0$, fixando $\alpha = 1$. (b) Para alguns valores de $\tau > 0$. (c) Para alguns valores de $\tau < 0$ e $\tau > 0$. (d) Para alguns valores de $\tau < 0$ fixando $\alpha = 1$ e $\lambda = 0, 15$. (e) Para alguns valores de $\tau < 0$ e $\tau > 0$ fixando $\alpha = 2$ e $\lambda = 0, 15$

Substituindo (4.7) e (4.9) na equação (4.5) obtêm-se as seguintes expressões

$$F(t) = \begin{cases} \frac{\sum_{j=0}^{\infty} a_j G(t; \alpha, \tau, k)^j}{\sum_{j=0}^{\infty} b_j G(t; \alpha, \tau, k)^j}, & \text{if } \tau > 0, \\ \frac{\sum_{j=0}^{\infty} a_j^* G(t; \alpha, \tau, k)^j}{\sum_{j=0}^{\infty} b_j G(t; \alpha, \tau, k)^j}, & \text{if } \tau < 0, \end{cases}$$

tais que $a_j^* = (-1)^j \binom{\lambda}{j}$ e $b_j = a_j + (-1)^j \binom{\lambda}{j}$ para $j \geq 0$.

A razão de duas séries de potência, para $\tau < 0$, pode ser expressa como

$$F(t) = \sum_{j=0}^{\infty} c_j G(t; \alpha, \tau, k)^j, \quad (4.10)$$

em que $c_0 = a_0^*/b_0$ e os coeficientes c_j 's (para $j \geq 1$) são determinados pela equação de recorrência

$$c_j = b_0^{-1} \left(a_j^* - \sum_{r=1}^j b_r a_{j-r}^* \right).$$

A fdp para T é obtida a partir da derivada de primeira ordem de (4.10) e é dada por

$$f(t) = \sum_{j=0}^{\infty} c_{j+1} h_{j+1}(t), \quad (4.11)$$

tal que $h_{j+1}(t)$ (para $j \geq 0$) é a função de densidade Exp-GG com parâmetro de potência $j+1$ dada por

$$h_{j+1}(t) = \frac{(j+1)|\tau|}{\alpha \Gamma(k)} (t/\alpha)^{\tau k-1} \exp \left[- \left(\frac{t}{\alpha} \right)^{\tau} \right] \gamma_1(k, (t/\alpha)^{\tau})^j, \quad (4.12)$$

em que $\gamma_1(k, (t/\alpha)^{\tau}) = \gamma(k, (t/\alpha)^{\tau})/\Gamma(k)$.

A equação (4.11) mostra que a fdp OLLGG é uma combinação linear da densidades Exp-GG. Assim, algumas das suas propriedades, tais como os momentos ordinários e incompletos e a função geradora podem ser determinados a partir de propriedades da distribuição Exp-GG. Esta equação é o principal resultado desta seção.

Considere uma equação da Seção 0.314 de GRADSHTEYN e RYZHIK (2000) para uma série de potência aumentada para um número inteiro positivo j

$$\left(\sum_{i=0}^{\infty} a_i x^i \right)^j = \sum_{i=0}^{\infty} d_{j,i} x^i, \quad (4.13)$$

em que os coeficientes $d_{j,i}$ (para $i = 1, 2, \dots$) são obtidos pela relação de recorrência

$$d_{j,i} = (ia_0)^{-1} \sum_{p=1}^i [j(p+1) - i] a_p d_{j,i-p} \quad (4.14)$$

e $d_{j,0} = a_0^j$.

Além disso, utilizando a equação (4.13), pode-se escrever (para $j \geq 1$)

$$\gamma_1(k, (t/\alpha)^{\tau})^j = \frac{(t/\alpha)^{jk\tau}}{\Gamma(k)^j} \sum_{i=0}^{\infty} d_{j,i} (t/\alpha)^{i\tau}, \quad (4.15)$$

em que os coeficientes $d_{j,i}$ (para $i \geq 1$) são determinados de (4.14) com $a_p = (-1)^p / [(k+p)p!]$.

Com base na equação (4.15), pode-se escrever a densidade da Exp-GG (para $j \geq 1$) de (4.12) como

$$h_{j+1}(t) = \frac{(j+1)|\tau|}{\alpha \Gamma(k)^{j+1}} \exp \left[- \left(\frac{t}{\alpha} \right)^{\tau} \right] \sum_{i=0}^{\infty} d_{j,i} \left(\frac{t}{\alpha} \right)^{[i+(j+1)k]\tau-1}.$$

A última fdp pode ser expressa em termos das funções de densidade GG. Observando a forma de (4.1), $h_{j+1}(t)$ (para $j \geq 1$) é escrita como

$$h_{j+1}(t) = \sum_{i=0}^{\infty} e_{j,i} g(t; \alpha, \tau, i + (j+1)k), \quad (4.16)$$

em que $g(t; \alpha, \tau, i + (j+1)k)$ é a fdp GG (4.1) com parâmetros α , τ e $i + (j+1)k$ e

$$e_{j,i} = \frac{(j+1)\Gamma(i + (j+1)k)}{\Gamma(k)^{j+1}} d_{j,i}. \quad (4.17)$$

Para $j = 0$, é obtido de (4.12) que $h_1(t) = \frac{|\tau|}{\alpha\Gamma(k)} (t/\alpha)^{\tau k-1} \exp\left[\left(\frac{t}{\alpha}\right)^\tau\right] = g(t; \alpha, \tau, k)$. Combinando o resultado (4.16) (para $j \geq 1$) e aquele para $j = 0$, tem-se $f(t) = f(t; \alpha, \tau, k, \lambda)$ em (4.11) como

$$f(t) = c_1 g(t; \alpha, \tau, k) + \sum_{j=1}^{\infty} \sum_{i=0}^{\infty} e_{j,i} c_{j+1} g(t; \alpha, \tau, i + (j+1)k). \quad (4.18)$$

A equação (4.18) mostra que algumas propriedades matemáticas da distribuição OLLGG podem ser obtidas diretamente das propriedades da distribuição GG. Por exemplo, os momentos ordinário, central e fatorial e a fmg da distribuição proposta seguem a mesma combinação linear infinita ponderada das quantidades correspondentes para a distribuição GG. Esta equação é o resultado principal desta seção.

4.3.1 Momentos ordinários

Nesta seção são obtidas duas expressões para o r -ésimo momento para T , denotado por $\mu'_r = E(T^r)$.

Primeiro, é obtido uma soma infinita para μ'_r da equação (4.18). O r -ésimo momento para a distribuição $GG(\alpha, \tau, k)$ é bem conhecido (STACY e MIHRAM, 1965) e dado por

$$\mu'_{r,GG} = \frac{\alpha^r \Gamma(k + r/\tau)}{\Gamma(k)}, \quad r/\tau > -k, \quad \forall \tau.$$

Então, da equação (4.18) tem-se

$$\mu'_r = \frac{c_1 \alpha^r \Gamma(k + r/\tau)}{\Gamma(k)} + \frac{\alpha^r}{\Gamma(k)} \sum_{j=1}^{\infty} \sum_{i=0}^{\infty} e_{j,i} c_{j+1} \Gamma(i + (j+1)k + r/\tau). \quad (4.19)$$

A equação (4.19) mostra que o momento μ'_r tem o inconveniente de depender das quantidades $e_{j,i}$ dado por (4.17). O seguinte teorema fornece o r -ésimo momento ordinário para T .

Teorema 1. O r -ésimo momento ordinário para T é dado por

$$\mu'_r = \frac{\lambda \alpha^r}{\Gamma(k)^\lambda} \sum_{j=0}^{\infty} \sum_{m=0}^j (-1)^{j+m} \binom{\lambda-1}{j} \binom{j}{m} I(k, r/\tau, m), \quad (4.20)$$

tal que $I(k, r/\tau, m) = \int_0^\infty x^{k+r/\tau-1} e^{-x} \gamma(k, x)^m dx$ pode ser dado em termos da função Lauricella do tipo A (AARTS, 2000).

Demonstração. Outra representação de soma infinita para μ'_r sem a necessidade das quantidades $e_{j,i}$. Obtendo assim

$$\mu'_r = \frac{\lambda |\tau| \alpha^{r-1}}{\Gamma(k)} \int_0^\infty \left(\frac{t}{\alpha}\right)^{\tau k+r-1} \exp\left\{-\left(\frac{t}{\alpha}\right)^\tau\right\} \left\{\gamma_1\left[k, \left(\frac{t}{\alpha}\right)^\tau\right]\right\}^{\lambda-1} dt$$

e então $x = (t/\alpha)^\tau$, para $\tau > 0$, dado

$$\mu'_r = \frac{\lambda \alpha^r}{\Gamma(k)^\lambda} \int_0^\infty x^{k+r/\tau-1} e^{-x} \gamma(k, x)^{\lambda-1} dx.$$

Pode-se escrever para qualquer $\lambda > 0$

$$\gamma_1(k, x)^{\lambda-1} = [1 - \{1 - \gamma_1(k, x)\}]^{\lambda-1} = \sum_{j=0}^{\infty} (-1)^j \binom{\lambda-1}{j} \{1 - \gamma_1(k, x)\}^j,$$

que sempre converge desde que $0 < \gamma_1(k, x) < 1$. Consequentemente,

$$\gamma_1(k, x)^{\lambda-1} = \sum_{j=0}^{\infty} \sum_{m=0}^j (-1)^{j+m} \binom{\lambda-1}{j} \binom{j}{m} \gamma_1(k, x)^m.$$

Inserindo a última equação na expressão para μ'_r e trocando os termos, é obtido

$$\mu'_r = \frac{\lambda \alpha^r}{\Gamma(k)^\lambda} \sum_{j=0}^{\infty} \sum_{m=0}^j (-1)^{j+m} \binom{\lambda-1}{j} \binom{j}{m} I(k, r/\tau, m), \quad (4.21)$$

em que

$$I(k, r/\tau, m) = \int_0^\infty x^{k+r/\tau-1} e^{-x} \gamma(k, x)^m dx.$$

Reescrevendo a última integral usando a série de potência para a função gama incompleta

$$I(k, r/\tau, m) = \int_0^\infty x^{k+r/\tau-1} e^{-x} \left[x^k \sum_{p=0}^{\infty} \frac{(-x)^p}{(k+p)p!} \right]^m dx.$$

Em seguida, esta integral pode ser obtida a partir das equações (24) e (25) de NADARAJAH (2008) em termos da função Lauricella do tipo A (AARTS, 2000) definida por

$$F_A^{(n)}(a; b_1, \dots, b_n; c_1, \dots, c_n; x_1, \dots, x_n) = \sum_{m_1=0}^{\infty} \dots \sum_{m_n=0}^{\infty} \frac{(a)_{m_1+\dots+m_n} (b_1)_{m_1} \dots (b_n)_{m_n}}{(c_1)_{m_1} \dots (c_n)_{m_n}} \frac{x_1^{m_1} \dots x_n^{m_n}}{m_1! \dots m_n!},$$

tal que $(a)_i$ é o fatorial ascendente definido por (com a convenção de que $(a)_0 = 1$)

$$(a)_i = a(a+1) \cdots (a+i-1).$$

Assim, é obtido

$$I(k, r/\tau, m) = k^{-m} \Gamma(r/\tau + k(m+1)) \times \\ F_A^{(m)}(r/\tau + k(m+1); k, \dots, k; k+1, \dots, k+1; -1, \dots, -1). \quad (4.22)$$

Então, como uma forma alternativa à equação (4.19), o r -ésimo momento para T segue de ambas as fórmulas (4.21) e (4.22) como uma soma ponderada infinita das funções de Lauricella do tipo A. As rotinas numéricas para a computação direta da função Lauricella do tipo A estão disponíveis, em (EXTON, 1978) e (TROT, 2006).

4.3.2 Momentos incompletos

Nesta seção, é obtido o n -ésimo momento incompleto para T dado por $m_n(y) = \int_0^y t^n f(t) dt$. Com base na combinação linear dada pela equação (4.18) tem-se o seguinte teorema para $m_n(y)$.

Teorema 2. O n -ésimo momento incompleto para T é dado por

$$m_n(y) = \sum_{j=0}^{\infty} f_j H(k, j, n, \tau, i, \alpha, u), \quad (4.23)$$

em que $f_j = \sum_{i=0}^{\infty} \frac{(j+1)|\tau|}{\Gamma(k)^{j+1}} d_{j,i}$ e $H(k, j, n, \tau, i, \alpha, u)$ é a função Whittaker definida abaixo.

Demonstração. Uma expressão para $m_n(y)$ segue das equações (4.11) e (4.12) como

$$m_n(y) = \sum_{j=0}^{\infty} \frac{(j+1)|\tau| c_{j+1}}{\alpha \Gamma(k)} \int_0^y t^n (t/\alpha)^{\tau k-1} \exp \left[- \left(\frac{t}{\alpha} \right)^{\tau} \right] \gamma_1(k, (t/\alpha)^{\tau})^j dt. \quad (4.24)$$

A equação (4.15) dada como uma série de potência $\gamma_1(k, (t/\alpha)^{\tau})^j$. Inserindo (4.15) e com $x = (t/\alpha)^{\tau}$ na equação (4.24) obtém-se

$$m_n(y) = \sum_{i,j=0}^{\infty} \frac{(j+1)|\tau| \alpha^n c_{j+1} d_{j,i}}{\Gamma(k)^{j+1}} \int_0^{(y/\alpha)^{\tau}} x^{(j+1)k+n/\tau+i-1} e^{-x} dx. \quad (4.25)$$

Consideramos a função hipergeométrica confluyente ${}_1F_1(a; b; z) = \sum_{k=0}^{\infty} \frac{(a)_k}{(b)_k} \frac{z^k}{k!}$, a função hipergeométrica ${}_2F_1(a, b; c; z) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k} \frac{z^k}{k!}$ e a função Kummer's dada por $U(a, b; z) = z^{-a} {}_2F_0(a, 1+a-b; -z^{-1})$. Todas estas funções podem ser calculadas usando o programa Mathematica. (Ver <http://mathworld.wolfram.com/ConfluentHypergeometricFunctionoftheFirstKind.html>.)

Pode-se mostrar que (Whittaker e Watson, 1990),

$H(k, j, n, \tau, i, \alpha, y) = \int_0^{(y/\alpha)^\tau} x^{(j+1)k+n/\tau+i-1} e^{-x} dx$ para $y > 0$, é dado por

$$H(k, j, n, \tau, i, \alpha, u) = \frac{u^{\left(\frac{\tau[(j+1)k+1-3]+n}{2\tau}\right)} (\tau^2 + \tau \omega_1) e^{-\frac{u}{2}}}{\omega_0 \omega_1 \omega_2} N_{\omega_0 - \frac{\omega_1}{2\tau}, \frac{\omega_1}{2\tau} + \frac{1}{2}}(u) \\ + \frac{u^{\left(\frac{\tau[(j+1)k+1-3]+n}{2\tau}\right)} \omega_1 e^{-\frac{u}{2}}}{\tau \omega_0 \omega_1} N_{\omega_0 - \frac{\omega_1}{2\tau} + 1, \frac{\omega_1}{2\tau} + \frac{1}{2}}(u),$$

em que $\omega_0 = (j+1)k + \frac{n}{\tau} + i$, $\omega_1 = \tau + (j+1)k\tau + n + i\tau$, $\omega_2 = 2\tau + (j+1)k\tau + n + i\tau$ com $u = (y/\alpha)^\tau$ e $N_{k,m}(x)$ é a função Whittaker (Whittaker e Watson 1990, pp. 339-351) dada por

$$N_{k,m}(x) = e^{-x/2} x^{m+1/2} {}_1F_1\left(\frac{1}{2} + m - k, 1 + 2m; x\right)$$

ou

$$N_{k,m}(x) = e^{-x/2} x^{m+1/2} U\left(\frac{1}{2} + m - k, 1 + 2m; x\right).$$

Estas funções são implementadas no Mathematica como $\text{WhittakerM}[k, m, x]$ e $\text{WhittakerW}[k, m, x]$, respectivamente. Deste modo, a equação (4.25) reduz a (4.23).

4.3.3 Função geradora de momentos

Aqui, é apresentado uma expressão explícita para o mgf de T , denotada por $M_T(s) = E(e^{sT})$, para $\tau > 1$, com base na função Wright (WRIGHT, 1935). Considerando a definição para $M_T(s)$ e a representação linear (4.11), pode-se escrever

$$M_T(s) = \frac{(j+1)|\tau|}{\alpha^{\tau k} \Gamma(k)} \sum_{j=0}^{\infty} c_{j+1} \int_0^{\infty} \exp(st) t^{\tau k-1} \exp\{-(t/\alpha)^\tau\} \gamma_1(k, (t/\alpha)^\tau)^j dt. \quad (4.26)$$

Inserindo (4.15) e fazendo $u = (t/\alpha)$ na equação (4.26), é obtido

$$M_T(s) = \frac{(j+1)|\tau|}{\Gamma(k)^{j+1}} \sum_{i,j=0}^{\infty} c_{j+1} d_{j,i} \int_0^{\infty} \exp(su\alpha) u^{(j+1)k\tau+\tau i-1} \exp(-u^\tau) du.$$

Agora, assumindo $\tau > 0$, e expandindo a primeira exponencial em série de Taylor e usando que $\int_0^{\infty} u^{(j+1)k\tau+\tau i+m-1} e^{-u^\tau} du = \tau^{-1} \Gamma(m/\tau + k + j\tau + i)$, obtendo

$$M_T(s) = \sum_{i,j,m=0}^{\infty} \frac{(j+1) c_{j+1} d_{j,i}}{\Gamma(k)} \Gamma\left(\frac{m}{\tau} + k + j\tau + i\right) \frac{(\alpha s)^m}{m!}. \quad (4.27)$$

Considerando $\tau < 0$, analogamente a (4.27) usando $\int_0^{\infty} u^{-(j+1)k\tau-\tau i+m-1} e^{-u^{-\tau}} du = -\tau^{-1} \Gamma(-m/\tau + k + j\tau + i)$, obtém-se

$$M_T(s) = \sum_{i,j,m=0}^{\infty} \frac{(j+1) c_{j+1} d_{j,i}}{\Gamma(k)} \Gamma\left(-\frac{m}{\tau} + k + j\tau + i\right) \frac{(\alpha s)^m}{m!}. \quad (4.28)$$

No entanto, para $\tau > 1$, pode ser simplificado considerando a função hipergeométrica generalizada de (WRIGHT, 1935) definida por

$${}_p\Psi_q \left[\begin{matrix} (\alpha_1, A_1), \dots, (\alpha_p, A_p) \\ (\beta_1, B_1), \dots, (\beta_q, B_q) \end{matrix} ; x \right] = \sum_{m=0}^{\infty} \frac{\prod_{j=1}^p \Gamma(\alpha_j + A_j m)}{\prod_{j=1}^q \Gamma(\beta_j + B_j m)} \frac{x^m}{m!}.$$

Esta função existe se $1 + \sum_{j=1}^q B_j - \sum_{j=1}^p A_j > 0$. Pela combinação das duas últimas equações, temos

$$M_{\alpha, \tau, k}(s) = \frac{1}{\Gamma(k)} {}_1\Psi_0 \left[\begin{matrix} (k, \tau^{-1}) \\ - \end{matrix} ; s \alpha \right]. \quad (4.29)$$

Finalmente, a fmg de T (para $\tau > 1$) pode ser escrita pelas equações (4.18) e (4.29) como

$$M(s) = \sum_{j,i=0}^{\infty} \frac{c_{j+1} e_{j,i}}{\Gamma(i + (j+1)k)} {}_1\Psi_0 \left[\begin{matrix} (i + (j+1)k, \tau^{-1}) \\ - \end{matrix} ; s \alpha \right]. \quad (4.30)$$

As equações (4.27)-(4.30) são os principais resultados desta seção.

4.3.4 Desvios médios

Os desvios médios de T sobre a média μ'_1 e sobre a mediana M em termos de seu primeiro momento incompleto, podem ser obtidos da seguinte forma

$$\delta_1 = 2[\mu'_1 F(\mu'_1) - M_1(\mu'_1)] \quad \text{e} \quad \delta_2 = \mu'_1 - 2m_1(M), \quad (4.31)$$

em que $\mu'_1 = E(T)$ e $m_1(q) = \int_{-\infty}^q t f(t) dt$. A quantidade $m_1(q)$ é determinada de (4.23) com $n = 1$ e as medidas δ_1 e δ_2 são avaliadas em $q = \mu'_1$ e $q = M$, respectivamente.

4.4 Distribuição log-odd log-logística gama generalizada

Com base na v.a T com fdp OLLGG (4.4), seja $Y = \log(T)$. Fazendo $k = q^{-2}$, $\tau = (\sigma\sqrt{k})^{-1}$ e $\alpha = \exp[\mu - \tau^{-1} \log(k)]$, a fdp para Y pode ser expressa como ($y \in \mathbb{R}$)

$$\begin{aligned} f(y) &= \frac{\lambda |q| (q^{-2})^{q^{-2}}}{\sigma \Gamma(q^{-2})} \exp \left\{ q^{-1} \left(\frac{y - \mu}{\sigma} \right) - q^{-2} \exp \left[q \left(\frac{y - \mu}{\sigma} \right) \right] \right\} \times \\ &\quad \left\{ \gamma_1 \left(q^{-2}, q^{-2} \exp \left[q \left(\frac{y - \mu}{\sigma} \right) \right] \right) \left[1 - \gamma_1 \left(q^{-2}, q^{-2} \exp \left[q \left(\frac{y - \mu}{\sigma} \right) \right] \right) \right] \right\}^{\lambda-1} \times \\ &\quad \left\{ \gamma_1^{\lambda} \left(q^{-2}, q^{-2} \exp \left[q \left(\frac{y - \mu}{\sigma} \right) \right] \right) + \left[1 - \gamma_1 \left(q^{-2}, q^{-2} \exp \left[q \left(\frac{y - \mu}{\sigma} \right) \right] \right) \right]^{\lambda} \right\}^{-2}, \end{aligned}$$

tais que $\mu \in \mathbb{R}$, $\sigma > 0$, $\lambda > 0$ e q é diferente de zero. Considerando uma forma estendida incluindo o caso $q = 0$ (LAWLESS, 2003). Assim, a fdp de Y pode ser expressa como

$$f(y) = \begin{cases} \frac{\lambda |q| (q^{-2})^{q-2}}{\sigma \Gamma(q^{-2})} \exp \left\{ q^{-1} \left(\frac{y-\mu}{\sigma} \right) - q^{-2} \exp \left[q \left(\frac{y-\mu}{\sigma} \right) \right] \right\} \times \\ \left\{ \gamma_1 \left(q^{-2}, q^{-2} \exp \left[q \left(\frac{y-\mu}{\sigma} \right) \right] \right) \left[1 - \gamma_1 \left(q^{-2}, q^{-2} \exp \left[q \left(\frac{y-\mu}{\sigma} \right) \right] \right) \right] \right\}^{\lambda-1} \times \\ \left\{ \gamma_1^\lambda \left(q^{-2}, q^{-2} \exp \left[q \left(\frac{y-\mu}{\sigma} \right) \right] \right) + \left[1 - \gamma_1 \left(q^{-2}, q^{-2} \exp \left[q \left(\frac{y-\mu}{\sigma} \right) \right] \right) \right]^\lambda \right\}^{-2}, \text{ se } q \neq 0, \\ \frac{\lambda}{\sigma} \phi \left(\frac{y-\mu}{\sigma} \right) \left\{ \Phi \left(\frac{y-\mu}{\sigma} \right) \left[1 - \Phi \left(\frac{y-\mu}{\sigma} \right) \right] \right\}^{\lambda-1} \left\{ \Phi^\lambda \left(\frac{y-\mu}{\sigma} \right) + \left[1 - \Phi \left(\frac{y-\mu}{\sigma} \right) \right]^\lambda \right\}^{-2}, \text{ se } q = 0, \end{cases} \quad (4.32)$$

em que $\phi(\cdot)$ e $\Phi(\cdot)$ são a função densidade e a função distribuição da normal padrão, respectivamente.

A distribuição dada pela equação (4.32) é denotada como a distribuição LOLGG, $Y \sim \text{LOLLGG}(\mu, \sigma, q, \lambda)$, em que $\mu \in \mathbb{R}$ é o parâmetro de locação, $\sigma > 0$ é o parâmetro de escala, q e λ são os parâmetros de forma. Para $q = 0$ e $\lambda = 1$ e $q = 1$, obtêm-se de (4.32) as distribuições *normal* e *valor extremo*, respectivamente. Para $\lambda = 1$, é obtido a distribuição *log-gama generalizada*. O caso $\lambda = 1$ e $q = -1$ corresponde a distribuição *log-inversa Weibull*. Portanto, se $T \sim \text{OLLGG}(\alpha, \tau, k, \lambda)$ então $Y = \log(T) \sim \text{LOLLGG}(\mu, \sigma, q, \lambda)$.

Os gráficos da fdp (4.32) para valores selecionados de λ , μ , σ e q são apresentados na Figura 4.3. Esses gráficos indicam claramente que a distribuição LOLGG pode ser muito flexível para modelar dados com bimodalidade assimétrica.

Figura 4.3. Gráficos da fdp LOLGG para alguns valores dos parâmetros. (a) Para alguns valores de $q < 0$ e $\mu = 2$, $\sigma = 1$ e $\lambda = 0, 25$ fixos. (b) Para $q = 0$ fixo $\mu = 0$, $\sigma = 1, 5$ e alguns valores de λ . (c) Para alguns valores de $q > 0$ e $\mu = 2$, $\sigma = 1$ e $\lambda = 0, 25$ fixos.

A correspondente função de sobrevivência para Y é

- Para $q > 0$,

$$\begin{aligned} S(y) &= 1 - F(y) = P(Y > y) = P(\mu + \sigma Z > y) = P(Z > z) \\ &= \frac{\lambda q(q^{-2})^{q^{-2}}}{\Gamma(q^{-2})} \int_z^\infty \frac{\exp(q^{-1}u - u^*) \{\gamma_1(q^{-2}, u^*)[1 - \gamma_1(q^{-2}, u^*)]\}^{\lambda-1}}{\{\gamma_1^\lambda(q^{-2}, u^*) + [1 - \gamma_1(q^{-2}, u^*)]^\lambda\}^2} du, \end{aligned}$$

- Para $q < 0$,

$$\begin{aligned} S(y) &= 1 - F(y) = P(Y > y) = P(\mu + \sigma Z > y) = P(Z > z) \\ &= \frac{-\lambda q(q^{-2})^{q^{-2}}}{\Gamma(q^{-2})} \int_z^\infty \frac{\exp(q^{-1}u - u^*) \{\gamma_1(q^{-2}, u^*)[1 - \gamma_1(q^{-2}, u^*)]\}^{\lambda-1}}{\{\gamma_1^\lambda(q^{-2}, u^*) + [1 - \gamma_1(q^{-2}, u^*)]^\lambda\}^2} du, \end{aligned}$$

em que $u^* = q^{-2} \exp[q(\frac{y-\mu}{\sigma})]$ and $u = \frac{y-\mu}{\sigma}$.

Essas integrais podem ser calculadas explicitamente e resultam na seguinte expressão para a função de sobrevivência

$$S(y) = \begin{cases} \frac{\{1 - \gamma_1(q^{-2}, q^{-2} \exp[q(\frac{y-\mu}{\sigma})])\}^\lambda}{\gamma_1^\lambda(q^{-2}, q^{-2} \exp[q(\frac{y-\mu}{\sigma})]) + \{1 - \gamma_1(q^{-2}, q^{-2} \exp[q(\frac{y-\mu}{\sigma})])\}^\lambda}, & \text{se } q > 0, \\ \frac{\gamma_1^\lambda(q^{-2}, q^{-2} \exp[q(\frac{y-\mu}{\sigma})])}{\gamma_1^\lambda(q^{-2}, q^{-2} \exp[q(\frac{y-\mu}{\sigma})]) + \{1 - \gamma_1(q^{-2}, q^{-2} \exp[q(\frac{y-\mu}{\sigma})])\}^\lambda}, & \text{se } q < 0, \\ \frac{[1 - \Phi(\frac{y-\mu}{\sigma})]^\lambda}{\Phi^\lambda(\frac{y-\mu}{\sigma}) + [1 - \Phi(\frac{y-\mu}{\sigma})]^\lambda}, & \text{se } q = 0. \end{cases} \quad (4.33)$$

- Para $q = 0$, DA SILVA BRAGA *et al.* (2016) apresenta várias propriedades estruturais de Y incluindo expressões explícitas para os momentos ordinários e incompletos, função geradora e desvios médios.

4.5 Modelo de regressão log-odd log-logística gama generalizada

Em muitas aplicações práticas, os tempos de vida são afetados por variáveis explicativas ou covariáveis, como o nível de colesterol, a pressão arterial, o peso e muitos outros. A última década possui muito trabalhos sobre classes generalizadas de modelos de regressão, que são sempre de interesse para os estatísticos aplicados. No contexto da análise de sobrevivência, algumas distribuições foram utilizadas para analisar dados censurados. Por exemplo, CRUZ *et al.* (2016) propuseram o modelo de regressão log-odd log-logística Weibull e LANJONI *et al.* (2016) definiram o modelo de regressão Burr XII estendido. Assim, usando a mesma abordagem adotada nestes artigos, uma distribuição obtida a partir da distribuição LOLLGG será expressa na forma da classe de modelos de locação-escala. Um modelo paramétrico que fornece um bom ajuste aos dados de tempo de vida tende a produzir estimativas mais precisas das quantidades de interesse.

Se $Y \sim LOLLGG(\mu, \sigma, q, \lambda)$, a v.a padronizada $Z = (Y - \mu)/\sigma$ tem fdp dada por

$$f(z) = \begin{cases} \frac{\lambda |q|(q^{-2})^{q-2}}{\Gamma(q^{-2})} \exp\{q^{-1}z - q^{-2}\exp(qz)\} \times \\ \{\gamma_1(q^{-2}, q^{-2}\exp(qz)) [1 - \gamma_1(q^{-2}, q^{-2}\exp(qz))]\}^{\lambda-1} \times \\ \left\{ \gamma_1^\lambda(q^{-2}, q^{-2}\exp(qz)) + [1 - \gamma_1(q^{-2}, q^{-2}\exp(qz))]^\lambda \right\}^{-2}, & \text{se } q \neq 0, \\ \lambda \phi(z) \{\Phi(z) [1 - \Phi(z)]\}^{\lambda-1} \left\{ \Phi^\lambda(z) + [1 - \Phi(z)]^\lambda \right\}^{-2}, & \text{se } q = 0. \end{cases} \quad (4.34)$$

Denotando $Z \sim LOLLGG(0, 1, q, \lambda)$. Além disso, é proposto um modelo de regressão linear de locação-escala ligando a variável resposta y_i ao vetor de covariáveis $\mathbf{x}_i^\top = (x_{i1}, \dots, x_{ip})$ de

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \sigma z_i, \quad i = 1, \dots, n, \quad (4.35)$$

em que o erro aleatório z_i tem fdp (4.34), $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$, $\sigma > 0$, $\lambda > 0$ e $q \in \mathbb{R}$ são os parâmetros desconhecidos. O parâmetro $\mu_i = \mathbf{x}_i^\top \boldsymbol{\beta}$ é de locação de y_i . O vetor de parâmetros de locação $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^\top$ é definido de modo linear $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, em que $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top$ é a matriz de covariáveis conhecida.

O modelo LOLLGG (4.35) abre novas possibilidades para vários tipos diferentes de dados. Contém como modelos especiais alguns modelos de regressão bem conhecidos:

- O caso em que $\lambda = 1$ tem-se o modelo de regressão *log-gama generalizado* (LGG).
- O caso em que $q = 1$ refere ao modelo de regressão *log-odd log-logística Weibull* (LOLLW) (CRUZ *et al.*, 2016).
- Para $q = -1$, obtém-se o novo modelo de regressão *log-odd log-logística inversa Weibull* (LOLLIW).
- Para $q = 0$, obtém-se o modelo de regressão *odd log-logística normal* (OLLN) (DA SILVA BRAGA *et al.*, 2016).
- Para $\lambda = q = 1$, tem-se o clássico modelo de regressão *log-Weibull* (ou valor extremo) (ver, LAWLESS (2003)).
- Se $\lambda = 1$, em adição de $q = -1$, segue o modelo de regressão *log-inversa Weibull* (LIW).
- Se $\sigma = 0.5$, em adição de $q = 1$, é obtido o novo modelo de regressão *log-odd log-logística Rayleigh* (LOLLR).

4.5.1 Estimação via método de máxima verossimilhança com dados censurados

Considere uma amostra $(y_1, \mathbf{x}_1), \dots, (y_n, \mathbf{x}_n)$ para n observações independentes, em que cada resposta aleatória é definida por $y_i = \min\{\log(t_i), \log(c_i)\}$. Aqui, os t_i 's são os tempos de falha e os c_i 's são os tempos de censura. Assumindo censura não informativa,

de tal forma que os tempos de vida observados e os tempos de censura são independentes. Seja F e C os conjuntos de indivíduos para os quais y_i é o logaritmo do tempo de vida ou logaritmo da censura, respectivamente. As técnicas convencionais de estimação de verossimilhança podem ser aplicadas aqui. O logaritmo da função de verossimilhança (log-verossimilhança) para o vetor de parâmetros $\boldsymbol{\theta} = (q, \lambda, \sigma, \boldsymbol{\beta}^\top)^\top$ do modelo (4.35) tem a forma $l(\boldsymbol{\theta}) = \sum_{i \in F} l_i(\boldsymbol{\theta}) + \sum_{i \in C} l_i^{(c)}(\boldsymbol{\theta})$, em que $l_i(\boldsymbol{\theta}) = \log[f(y_i)]$, $l_i^{(c)}(\boldsymbol{\theta}) = \log[S(y_i)]$, $f(y_i)$ é a fdp (4.32) e $S(y_i)$ é a função de sobrevivência (4.33) para Y_i . A função log-verossimilhança completa para $\boldsymbol{\theta}$ pode ser particionada como

$$l(\boldsymbol{\theta}) = \begin{cases} r \log \left[\frac{\lambda q (q^{-2})^{q-2}}{\sigma \Gamma(q-2)} \right] + q^{-1} \sum_{i \in F} z_i - q^{-2} \sum_{i \in F} \exp(q z_i) + (\lambda - 1) \sum_{i \in F} \log[u_i(1 - u_i)] - \\ 2 \sum_{i \in F} \log[u_i^\lambda + (1 - u_i)^\lambda] + \lambda \sum_{i \in C} \log(1 - u_i) + \sum_{i \in C} \log[u_i^\lambda + (1 - u_i)^\lambda], & \text{se } q > 0, \\ r \log \left[\frac{-\lambda q (q^{-2})^{q-2}}{\sigma \Gamma(q-2)} \right] + q^{-1} \sum_{i \in F} z_i - q^{-2} \sum_{i \in F} \exp(q z_i) + (\lambda - 1) \sum_{i \in F} \log[u_i(1 - u_i)] - \\ 2 \sum_{i \in F} \log[u_i^\lambda + (1 - u_i)^\lambda] + \lambda \sum_{i \in C} \log(u_i) + \sum_{i \in C} \log[u_i^\lambda + (1 - u_i)^\lambda], & \text{se } q < 0, \\ r \log \left(\frac{\lambda}{\sigma} \right) + \sum_{i \in F} \log[\phi(z_i)] + (\lambda - 1) \sum_{i \in F} \log\{\Phi(z_i)[1 - \Phi(z_i)]\} + \lambda \sum_{i \in C} \log[1 - \Phi(z_i)] - \\ 2 \sum_{i \in F} \log\{\Phi^\lambda(z_i) + [1 - \Phi(z_i)]^\lambda\} + \sum_{i \in C} \log\{\Phi^\lambda(z_i) + [1 - \Phi(z_i)]^\lambda\}, & \text{se } q = 0, \end{cases} \quad (4.36)$$

em que r é o número de observações não censuradas (falhas),

$$u_i = \gamma_1(q^{-2}, q^{-2} \exp(q z_i)) \quad \text{e} \quad z_i = \frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma}.$$

As EMVs $\hat{\boldsymbol{\theta}}$ para o vetor de parâmetros do modelo pode ser calculado maximizando a função log-verossimilhança (4.36). Para valores iniciais $\boldsymbol{\beta}$ e σ obtidos através do ajuste do modelo de regressão log-Weibull $\lambda = 1$ e $q = 1$. Se $\hat{z}_i = (y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})/\hat{\sigma}$, o ajuste do modelo LOLLGG produz a função de sobrevivência estimada para y_i

$$S(y_i; \hat{q}, \hat{\lambda}, \hat{\sigma}, \hat{\boldsymbol{\beta}}^\top) = \begin{cases} \frac{\{1 - \gamma_1(\hat{q}^{-2}, \hat{q}^{-2} \exp(\hat{q} \hat{z}_i))\}^\lambda}{\gamma_1^\lambda(\hat{q}^{-2}, \hat{q}^{-2} \exp(\hat{q} \hat{z}_i)) + \{1 - \gamma_1(\hat{q}^{-2}, \hat{q}^{-2} \exp(\hat{q} \hat{z}_i))\}^\lambda}, & \text{se } q > 0, \\ \frac{\gamma_1^\lambda(\hat{q}^{-2}, \hat{q}^{-2} \exp(\hat{q} \hat{z}_i))}{\gamma_1^\lambda(\hat{q}^{-2}, \hat{q}^{-2} \exp(\hat{q} \hat{z}_i)) + \{1 - \gamma_1(\hat{q}^{-2}, \hat{q}^{-2} \exp(\hat{q} \hat{z}_i))\}^\lambda}, & \text{se } q < 0, \\ \frac{[1 - \Phi(\hat{z}_i)]^\lambda}{\Phi^\lambda(\hat{z}_i) + [1 - \Phi(\hat{z}_i)]^\lambda}, & \text{se } q = 0. \end{cases} \quad (4.37)$$

Em certas condições que são atendidas para o vetor de parâmetro $\boldsymbol{\theta}$, a distribuição assintótica das EMVs é de $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ dada por uma normal multivariada $N_{p+3}(0, K(\boldsymbol{\theta})^{-1})$, em que $K(\boldsymbol{\theta})$ é a matriz de informação total. A matriz de covariância assintótica $K(\boldsymbol{\theta})^{-1}$ para $\hat{\boldsymbol{\theta}}$ pode ser aproximada pela inversa da matriz de informação observada $-\ddot{\mathbf{L}}(\boldsymbol{\theta})$ com dimensão $(p + 3) \times (p + 3)$. Os elementos de $-\ddot{\mathbf{L}}(\boldsymbol{\theta})$, são denotados $-\mathbf{L}_{qq}$, $-\mathbf{L}_{q\lambda}$, $-\mathbf{L}_{q\sigma}$, $-\mathbf{L}_{q\beta_j}$, $-\mathbf{L}_{\lambda\lambda}$, $-\mathbf{L}_{\lambda\sigma}$, $-\mathbf{L}_{\lambda\beta_j}$, $-\mathbf{L}_{\sigma\sigma}$, $-\mathbf{L}_{\sigma\beta_j}$ e $-\mathbf{L}_{\beta_j\beta_s}$ (for $j, s = 1, \dots, p$), são calculados numericamente. A distribuição normal multivariada aproximada $N_{p+3}(0, -\ddot{\mathbf{L}}(\boldsymbol{\theta})^{-1})$ para $\hat{\boldsymbol{\theta}}$ pode ser usado para construir intervalos de confiança para os componentes de $\boldsymbol{\theta}$.

Além disso, é possível utilizar o *TRV* para comparar alguns submodelos com o modelo LOLLGG. Considerando a partição $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$, em que $\boldsymbol{\theta}_1$ é um subconjunto

de parâmetros de interesse e θ_2 é um subconjunto dos parâmetros restantes. O *TRV* para testar a hipótese nula $H_0 : \theta_1 = \theta_1^{(0)}$ versus a hipótese alternativa $H_1 : \theta_1 \neq \theta_1^{(0)}$ é dado por $w = 2\{\ell(\hat{\theta}) - \ell(\tilde{\theta})\}$, em que $\tilde{\theta}$ e $\hat{\theta}$ são as EMVs sob H_1 e sob H_0 , respectivamente. A estatística w segue assintoticamente ($n \rightarrow \infty$) uma distribuição χ_k^2 , tal que k é a dimensão do subconjunto de parâmetros θ_1 de interesse.

4.5.2 Estudo via simulação

Um estudo de simulação de Monte Carlo para avaliar o comportamento amostral das EMVs de q , α , σ e β é realizado. Os resultados são obtidos a partir de 1000 repetições de Monte Carlo realizadas usando a sub-rotina *optim* pelo método BFGS do *software* R (R DEVELOPMENT CORE TEAM, 2014). Em cada repetição, uma amostra aleatória de tamanho n é gerada do modelo LOLLGG($q, \lambda, \sigma, \beta$) e os parâmetros são estimados via método de máxima verossimilhança. O logaritmo dos tempos denotado por $\log(t_1), \dots, \log(t_n)$ são gerados do modelo de regressão LOLLGG (4.35), em que $\mathbf{x}_i^\top \beta = \beta_0 + \beta_1 x_i$ e x_i é gerado de uma distribuição uniforme no intervalo $(0, 1)$. Assim, dois cenários são considerados para as simulações com tamanhos de amostra $n=50$, $n=150$ e $n=350$, e porcentagens de censura de 0%, 10% e 30%. O primeiro cenário refere-se a $q=0,65$, $\lambda=\sigma=0,5$, $\beta_0=2,0$ e $\beta_1=3,0$. No segundo cenário, é considerado o modelo LOLLIW com $q=-1,0$, $\lambda=1,2$, $\sigma=0,5$, $\beta_0=1,0$ e $\beta_1=0,7$. Os tempos de sobrevivência são gerados considerando o mecanismo de censura aleatória do seguinte modo:

- i. Gerar $x_i \sim \text{uniforme}(0,1)$.
- ii. Gerar $z \sim \text{LOLLGG}(0,1,q,\lambda)$, os valores da densidade (4.34).
- iii. Escrever $y = \beta_0 + \beta_1 x_i + \sigma z$.
- iv. Criar um vetor δ de dimensão n contendo 1's.
- v. Obter uma amostra de tamanho np (em que p representa a porcentagem de censura fixada) de índices $1, \dots, n$ e alocar em w .
- vi. Para $j = 1, \dots, \text{length}(w)$, define $y[w[j]] = \text{uniforme}(0, y[w[j]])$ e $\delta[w[j]] = 0$. Isto é, w contém as posições em que y recebe as porcentagens de censuras com porcentagem p fixo.

Desta forma, o modelo de regressão LOLLGG (4.35) é ajustado para cada conjunto de dados gerados, em que $\mu_i = \beta_0 + \beta_1 x_i$. A partir dos resultados das simulações apresentados na Tabela 4.3, são verificados que os EQMs das EMVs de λ , σ , β_0 e β_1 tende a zero quando o tamanho da amostra aumenta, como normalmente é esperado pela teoria assintótica padrão. À medida que o tamanho da amostra n aumenta, as EMs dos parâmetros tendem a estar mais próximas dos verdadeiros valores dos parâmetros. Este fato,

indica que assintoticamente, a distribuição normal fornece uma aproximação adequada à distribuição das estimativas para amostras finitas.

Tabela 4.3. EMs e (EQMs) para os parâmetros da distribuição LOLGG.

cenário 1					
n	Parâmetros	Valor Real	0%	10%	30%
50	σ	0,50	0,4390 (0,0475)	0,4297 (0,0508)	0,3982 (0,0652)
	q	0,65	0,9301 (3,0041)	0,9418 (2,8302)	0,9408 (2,6080)
	λ	0,50	0,5341 (0,2070)	0,5381 (0,2243)	0,5328 (0,2728)
	β_0	2,00	2,0719 (0,1428)	2,1078 (0,1502)	2,1984 (0,1852)
	β_1	3,00	2,9690 (0,1586)	2,9933 (0,1697)	2,9506 (0,2163)
150	σ	0,50	0,4843 (0,0174)	0,4869 (0,0196)	0,4718 (0,0226)
	q	0,65	0,6942 (0,2934)	0,7922 (1,0065)	0,8217 (0,9930)
	λ	0,50	0,5141 (0,0647)	0,5462 (0,0798)	0,5664 (0,2080)
	β_0	2,00	2,0177 (0,0403)	2,0627 (0,0516)	2,1679 (0,0842)
	β_1	3,00	3,0014 (0,0425)	2,9989 (0,0475)	2,9493 (0,0619)
350	σ	0,50	0,4930 (0,0067)	0,4876 (0,0082)	0,4792 (0,0102)
	q	0,65	0,6610 (0,0350)	0,6835 (0,1840)	0,7119 (0,0558)
	λ	0,50	0,5024 (0,0200)	0,5462 (0,0277)	0,5357 (0,0360)
	β_0	2,00	2,0130 (0,0131)	2,0460 (0,0183)	2,1317 (0,0341)
	β_1	3,00	2,9961 (0,0181)	2,9836 (0,0189)	2,9673 (0,0246)
cenário 2					
n	Parâmetros	Valor Real	0%	10%	30%
50	σ	0,50	0,3883 (0,0634)	0,4950 (0,0824)	0,5517 (0,0949)
	q	-1,00	-1,3803 (4,2521)	-1,0528 (1,1017)	-1,0368 (0,9019)
	λ	1,20	1,1455 (0,8730)	1,4018 (1,3438)	1,5648 (1,5713)
	β_0	1,00	0,9690 (0,0937)	0,9736 (0,0983)	0,9986 (0,1226)
	β_1	0,70	0,6888 (0,0631)	0,6777 (0,0623)	0,6649 (0,0801)
150	σ	0,50	0,4430 (0,0258)	0,5058 (0,0246)	0,5435 (0,0251)
	q	-1,00	-1,3850 (3,0003)	-1,0756 (0,4781)	-1,1179 (0,4265)
	λ	1,20	1,2625 (0,4776)	1,3478 (0,5348)	1,4515 (0,5688)
	β_0	1,00	0,9430 (0,0702)	0,9673 (0,0548)	0,9844 (0,0590)
	β_1	0,70	0,7023 (0,0169)	0,7008 (0,0179)	0,6755 (0,0224)
350	σ	0,50	0,4748 (0,0130)	0,5151 (0,0116)	0,5476 (0,0306)
	q	-1,00	-1,2271 (1,1598)	-1,1040 (0,2672)	-1,0912 (0,4626)
	λ	1,20	1,2908 (0,3308)	1,3508 (0,3093)	1,4517 (0,6326)
	β_0	1,00	0,9472 (0,0473)	0,9663 (0,0343)	0,9869 (0,0629)
	β_1	0,70	0,9472 (0,0075)	0,6922 (0,0073)	0,6857 (0,0220)

4.6 Análise de resíduos

O objetivo da análise de resíduos é avaliar a adequação do modelo para um conjunto de dados é verificar características intrínsecas aos dados. Para verificar algumas destas características, por exemplo, a presença de *outliers*, várias abordagens têm sido propostas por COX e SNELL (1968), COOK e WEISBERG (1982), ORTEGA *et al.* (2008) e

SILVA *et al.* (2011). De acordo com COX e SNELL (1968) as vantagens das técnicas mais formais da análise de resíduos são que elas têm testes de significância sensíveis associados a elas e que são diretamente construtivas no sentido de que, se o modelo inicial não se encaixa bem aos dados, um modelo específico mais adequado é obtido imediatamente a partir da análise. Ela indicará a natureza de uma saída do modelo inicial, mas não explicitamente como estender ou substituir o modelo.

No contexto da análise de sobrevivência, os resíduos deviance foram mais amplamente utilizados porque levam em conta a informação dos tempos censurados (SILVA *et al.*, 2011). Assim, o gráfico dos resíduos deviance em relação aos tempos observados fornece uma maneira de verificar a adequação do modelo ajustado e detectar observações atípicas.

Os resíduos deviance para o modelo de regressão LOLLGG com dados censurados pode ser expresso como

$$r_{D_i} = \text{sgn}(\hat{r}_{M_i}) \{-2 [\hat{r}_{M_i} + \delta_i \log(\delta_i - \hat{r}_{M_i})]\}^{1/2}, \quad (4.38)$$

em que

$$\hat{r}_{M_i} = \begin{cases} \delta_i + \log \left\{ \frac{\{1 - \gamma_1(\hat{q}^{-2}, \hat{q}^{-2} \exp(\hat{q} \hat{z}_i))\}^{\hat{\lambda}}}{\gamma_1^{\hat{\lambda}}(\hat{q}^{-2}, \hat{q}^{-2} \exp(\hat{q} \hat{z}_i)) + \{1 - \gamma_1(\hat{q}^{-2}, \hat{q}^{-2} \exp(\hat{q} \hat{z}_i))\}^{\hat{\lambda}}} \right\} & \text{se } q > 0, \\ \delta_i + \log \left\{ \frac{\gamma_1^{\hat{\lambda}}(\hat{q}^{-2}, \hat{q}^{-2} \exp(\hat{q} \hat{z}_i))}{\gamma_1^{\hat{\lambda}}(\hat{q}^{-2}, \hat{q}^{-2} \exp(\hat{q} \hat{z}_i)) + \{1 - \gamma_1(\hat{q}^{-2}, \hat{q}^{-2} \exp(\hat{q} \hat{z}_i))\}^{\hat{\lambda}}} \right\} & \text{se } q < 0, \\ \delta_i + \log \left\{ \frac{[1 - \Phi(\hat{z}_i)]^{\hat{\lambda}}}{\Phi^{\hat{\lambda}}(\hat{z}_i) + [1 - \Phi(\hat{z}_i)]^{\hat{\lambda}}} \right\} & \text{se } q = 0, \end{cases} \quad (4.39)$$

é o resíduo martingale, e δ_i é o indicador de censura, $\text{sign}(\cdot)$ é a função sinal que retorna os valores $+1$ se o argumento é positivo e -1 se o argumento é negativo e $\hat{z}_i = (y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) / \hat{\sigma}$.

Se o modelo for apropriado, os resíduos deviance modificado devem apresentar um comportamento aleatório em torno de zero. Os gráficos dos resíduos, martingale ou deviance modificado, em relação aos valores ajustados fornecem uma maneira simples de verificar a adequação do modelo e detectar *outliers*. ATKINSON (1985) sugeriu a construção de envelopes para permitir uma melhor interpretação do gráfico normal de probabilidades dos resíduos. Estes envelopes possuem faixas de confiança simuladas que contêm os resíduos, de tal forma que se o modelo estiver bem ajustado, a maioria dos pontos estarão dentro destas bandas e distribuídos aleatoriamente. A construção das faixas de confiança segue as seguintes etapas:

- Ajustar o modelo proposto e calcular o resíduo deviance r_{D_i} 's;
- Simular k amostras para a variável resposta usando o modelo ajustado;

- Ajustar o modelo para cada amostra e calcular os resíduos $r_{D_{ij}}$, $j = 1, 2, \dots, k$ e $i = 1, 2, \dots, n$;
- Organizar cada grupo de n resíduos em ordem crescente para obter $r_{D_{(i)j}}$ para $j = 1, 2, \dots, k$ e $i = 1, 2, \dots, n$;
- Para cada i , obter a média, mínimo e máximo de $r_{D_{(i)j}}$, renomeando

$$r_{D_{(i)M}} = \sum_{j=1}^k \frac{r_{D_{(i)j}}}{k}, r_{D_{(i)I}} = \min\{r_{D_{(i)j}} : 1 \leq j \leq k\} \text{ e } r_{D_{(i)S}} = \max\{r_{D_{(i)j}} : 1 \leq j \leq k\};$$

- Incluir as médias, mínimos e máximos juntamente com os valores de r_{D_i} contra os percentis esperados da distribuição normal padrão.

4.6.1 Estudo via simulação

Neste estudo de simulação é avaliado a precisão dos EMVs dos parâmetros no modelo de regressão LOLLGG com dados censurados e também para investigar o comportamento da distribuição empírica dos resíduos deviance. Para o estudo de simulação, são geradas as variáveis $z_1, \dots, z_n$ da distribuição LOLLGG (4.34).

Em seguida, 1000 amostras são geradas para cada cenário apresentado na Seção 4.5.2 bem como o algoritmo para gerar os tempos de sobrevivência e os tempos censurados. Pelo método de simulação de Monte Carlo, os EMVs para os parâmetros do modelo são calculados para cada uma das amostras. Para cada ajuste do modelo, os resíduos r_{D_i} 's dados por (4.38) são determinados.

Nas Figuras 4.4 e 4.5 são apresentados os gráficos dos resíduos em relação aos valores esperados das estatísticas da ordem da distribuição normal padrão. Estes gráficos são conhecidos como gráfico normal probabilístico utilizados para avaliar desvios na suposição de normalidade dos resíduos (WEISBERG, 2005). Consequentemente, a seguinte interpretação é obtida a partir destes gráficos: a distribuição empírica dos resíduos deviance concorda com a distribuição normal padrão quando o tamanho da amostra aumenta como ilustra as Figuras 4.4 e 4.5.

4.7 Aplicação

Como aplicação do modelo de regressão LOLLGG na presença de censura é usado o conjunto de dados reais de LAWLESS (2003). Os dados correspondem a um estudo de isolamento elétrico (STONE, 1978) considerando três níveis de tensão: 52,5 kV, 55 kV e 57,5 kV em que foram testadas 20 amostras para cada um dos três níveis. As variáveis consideradas neste estudo são t_{ij} = tempo de falha em minutos do i -ésimo isolamento elétrico no j -ésimo nível de tensão e x_{ij} = níveis de tensão para o i -ésimo isolamento elétrico sob o j -ésimo nível de tensão, para $i = 1, \dots, 60$ e $j = 1, 2$ e 3.

Figura 4.4. Gráficos normal probabilístico para os resíduos r_{D_i} para o cenário 1. (a) Amostra de tamanho $n=50$. (b) Amostra de tamanho $n=150$. (c) Amostra de tamanho $n=350$. Para porcentagens de censuras de 0%, 10% e 30%.

Figura 4.5. Gráficos normal probabilístico para os resíduos r_{D_i} para o cenário 2. (a) Amostra de tamanho $n=50$. (b) Amostra de tamanho $n=150$. (c) Amostra de tamanho $n=350$. Para porcentagens de censuras de 0%, 10% e 30%.

A porcentagem de observações censuradas para os tempos de sobrevivência é de 10%. Considerando, em uma análise preliminar, apenas os tempos de sobrevivência e censura, o gráfico TTT-plot é exibido na Figura 4.6, que indica que a função de risco associada ao conjunto de dados tem uma forma unimodal, uma vez que o gráfico mostra uma primeira curvatura côncava e, em seguida, convexa. Portanto, a distribuição OLLGG será usada para estes dados. Os parâmetros do modelo são estimados via método de máxima verossimilhança usando a sub-rotina *optim* pelo método BFGS no *software* R. Serão apresentadas as EMVs e seus erros padrão (EPs) entre parênteses e o critério de informação Akaike (AIC), critério de informação Akaike corrigido (CAIC) e Bayesiano (BIC).

Figura 4.6. Curva TTT-plot para os dados de níveis de tensão.

As EMVs para os parâmetros α , τ , k e λ (e o correspondente EPs em parênteses) e os valores das estatísticas AIC, BIC e CAIC são apresentadas na Tabela 4.4. O ajuste do modelo OLLGG é comparado com as distribuições OLL-Weibull (OLLW), OLL-exponential (OLLE), GG e Weibull. Tais distribuições são casos especiais da distribuição OLLGG. Observa-se que o modelo OLLGG apresenta os valores mais baixos das estatísticas AIC, BIC e CAIC, ou seja, o modelo OLLGG ajusta melhor os dados de níveis de tensão em comparação com outros modelos especiais. A estatística *TRV* para comparar os modelos ajustados são apresentadas na Tabela 4.5.

Para avaliar se o modelo é apropriado, é mostrado na Figura 4.7 as funções de sobrevivência estimadas das distribuições OLLGG, OLLW, OLLE, GG e Weibull e a função de sobrevivência empírica. Como uma alternativa para verificar a qualidade do ajuste, a Figura 4.8 apresenta as estimativas da função de risco das distribuições OLLGG,

Tabela 4.4. EMVs para os parâmetros do modelo para os dados de níveis de tensão, os correspondentes EPs (dados em parênteses) e as estatísticas AIC, CAIC e BIC.

Model	α	τ	k	λ	AIC	CAIC	BIC
OLLGG	1154,1730 (4290,9692)	0,0544 (0,0749)	1,2817 (0,2074)	18,8330 (25,5280)	862,9	864,5	871,3
OLLW	1310,4937 (284,0340)	0,4846 (0,1398)	1 (-)	2,3717 (0,7739)	865,9	867,0	872,2
OLLE	1161,1952 (150,8100)	1 (-)	1 (-)	1,07698 (0,1241)	873,1	873,8	877,3
GG	1148,0888 (657,75)	0,9342 (0,2347)	1,0474 (0,4480)	1 (-)	875,1	876,2	881,4
Weibull	1146,9611 (171,1271)	0,9484 (0,0944)	1 (-)	1 (-)	873,2	874,0	877,4

Tabela 4.5. TRV para os dados de níveis de tensão.

Modelos	Hipóteses	Estatística w	p -valor
OLLGG vs OLLW	$H_0 : k = 1$ vs $H_1 : H_0$ é falso	4,99	0,0252
OLLGG vs OLLE	$H_0 : \tau = k = 1$ vs $H_1 : H_0$ é falso	14,19	0,0008
OLLGG vs GG	$H_0 : \lambda = 1$ vs $H_1 : H_0$ é falso	14,15	0,0001
OLLGG vs Weibull	$H_0 : \lambda = k = 1$ vs $H_1 : H_0$ é falso	14,31	0,0007

OLLW, OLLE, GG e Weibull e a função de risco empírica. Conclui-se que a distribuição OLLGG tem um bom ajuste para esses dados de níveis de tensão.

Figura 4.7. Gráfico da função de sobrevivência estimada com ajuste da distribuição OLLGG e alguns outros modelos e a sobrevivência empírica para os dados de níveis de tensão. (a) OLLGG, OLLW e OLLE. (b) OLLGG, GG e Weibull.

O modelo de regressão LOLLGG para os dados de níveis de tensão pode ser

Figura 4.8. Gráfico da função de risco estimada com ajuste da distribuição OLLGG e alguns outros modelos e o risco empírico para os dados de níveis de tensão. (a) OLLGG, OLLW e OLLE. (b) OLLGG, GG e Weibull.

expresso como

$$y_i = \beta_0 + \beta_1 x_i + \sigma z_i,$$

em que $z_1, \dots, z_{60}$ são variáveis aleatórias independentes com fdp (4.34).

As EMVs para o modelo LOLLGG são apresentadas na Tabela 4.6. Assim, ao estabelecer um nível de significância de 5%, observa-se que a variável de nível de tensão x_i é significativa para o modelo ajustado.

Tabela 4.6. EMVs, EPs e p -valor para o modelo de regressão LOLLGG ajustado para os dados de níveis de tensão.

Parâmetros	Estimativas	EP	p -valor
λ	3,3549	2,0714	-
σ	0,7947	1,7048	-
q	-4,7399	9,2365	-
β_0	14,0699	3,3717	0,0001
β_1	-0,1785	0,0607	0,0046

Como alternativa para avaliar a qualidade do ajuste é apresentado na Figura 4.9a o gráfico que compara a função de sobrevivência empírica e a função sobrevivência estimada para o modelo de regressão LOLLGG. Com base neste gráfico, observa-se que o modelo LOLLGG apresenta um bom ajuste para os dados estudados.

Finalmente, verificando a qualidade do ajuste do modelo de regressão LOLLGG, construindo na Figura 4.9b o gráfico normal de probabilidade para o resíduo deviance

com envelope simulado (ATKINSON, 1985). Este gráfico também indica um bom ajuste do modelo de regressão LOLLGG aos dados de níveis de tensão.

Figura 4.9. (a) Função de sobrevivência estimada considerando o modelo de regressão LOLLGG para os dados de níveis de tensão. (b) Gráfico normal probabilístico para os resíduos deviance com envelope.

4.8 Conclusões

Neste capítulo, foi proposto uma forma estendida da distribuição odd log-logística gama generalizada (OLLGG) para a qual a função de risco acomoda os quatro tipos de formas, isto é, crescente, decrescente, forma de U e unimodal. O modelo OLLGG inclui como casos especiais vários modelos úteis para modelagem de tempos de vida. São apresentadas expressões explícitas para seus momentos ordinários, incompletos, desvios médios e função geradora. Além disso, é introduzida a distribuição chamada log-odd log-logística gama generalizada (LOLLGG). Com base nessa nova distribuição, o modelo de regressão LOLLGG, que é muito adequado para a modelagem de dados de tempo de vida censurados e não censurados é definido. O novo modelo de regressão permite testar como modelos especiais a qualidade de ajuste de alguns modelos de regressão amplamente conhecidos. Por isso, representa uma boa alternativa para a análise de dados de tempo da vida. O procedimento de máxima verossimilhança é utilizado para obter as estimativas dos parâmetros do modelo. Finalmente, utilizando um conjunto de dados reais, verifica-se que o modelo LOLLGG pode ter um ajuste melhor do que o de alguns modelos de regressão conhecidos.

5 CONSIDERAÇÕES FINAIS

Neste trabalho foram apresentado duas novas distribuições de probabilidade, denominadas de distribuição odd log-logística gama generalizada e distribuição log-odd log-logística gama generalizada. As distribuições foram obtidas a partir da família de distribuições odd log-logística de distribuições proposta por (GLEATON e LYNCH, 2006). Como distribuição de base são utilizadas as distribuições gama generalizada apresentadas por STACY (1962) e STACY e MIHRAM (1965), respectivamente. Uma das vantagens dessas novas distribuições é a de possuir formas que flexibilizam a função de risco e também servem para discriminação de modelos, pois possuem como casos particulares, distribuições conhecidas como, Weibull, exponencial, gama entre outras.

Podem-se destacar como contribuição deste trabalho:

- Foram obtidas duas novas distribuições a partir do procedimento de (GLEATON e LYNCH, 2006). Estas distribuições apresentam maior flexibilidade na modelagem de dados.
- A apresentação de um modelo de regressão locação-escala com base na distribuição log-odd log-logística gama generalizada.
- Estudos de simulação foram realizados considerando algumas porcentagens de censura, o que permitiu verificar a concordância do modelo em relação às propriedades assintóticas e no caso do modelo de regressão log-odd log-logística gama generalizado verificou-se a adequação dos resíduos, cuja distribuição empírica sugere estar de acordo com a distribuição normal padrão.
- A utilização dos modelos na modelagem de dados reais, na área de climatologia, medicina e engenharia.

5.1 Pesquisas futuras

Como continuidade a esta pesquisa os seguintes temas podem ser abordados:

- A proposta de modelagem com base na distribuição odd log-logística gama generalizada como, a inclusão de efeitos aleatórios, com fração de cura e riscos competitivos. E também o diagnóstico dos resíduos do modelo.
- Desenvolver um pacote no *software* R a partir da distribuição odd log-logística gama generalizada.

REFERÊNCIAS

- AARTS, R. M., 2000 *Lauricella functions*. From MathWorld, A Wolfram Web Resource, created by Eric W. Weisstein.
- AKAIKE, H., 1974 A new look at the statistical model identification. *IEEE transactions on automatic control* **19**: 716–723.
- ALEXANDER, C., G. M. CORDEIRO, E. M. M. ORTEGA, e J. M. SARABIA, 2012 Generalized beta-generated distributions. *Computational Statistics & Data Analysis* **56**: 1880–1897.
- ATKINSON, A. C., 1985 *Plots, transformations, and regression: an introduction to graphical methods of diagnostic regression analysis*. Clarendon Press Oxford.
- BARKAUSKAS, D. A., S. R. KRONEWITTER, C. B. LEBRILLA, e D. M. ROCKE, 2009 Analysis of maldi ft-icr mass spectrometry data: A time series approach. *Analytica Chimica Acta* **648**: 207–214.
- BARLOW, R. E. e R. A. CAMPO, 1975 Total time on test processes and applications to failure data analysis .
- BOURGUIGNON, M., M. D. C. S. LIMA, J. LEÃO, A. D. NASCIMENTO, L. G. B. PINHO, e G. M. CORDEIRO, 2015 A new generalized gamma distribution with applications. *American Journal of Mathematical and Management Sciences* **34**: 309–342.
- BOZDOGAN, H., 1987 Model selection and akaike's information criterion (aic): The general theory and its analytical extensions. *Psychometrika* **52**: 345–370.
- CARRASCO, J., E. ORTEGA, e G. PAULA, 2008 Log-modified weibull regression models with censored data: Sensitivity and residual analysis. *Computational Statistics and Data Analysis* **52**: 4021–4039.
- COLOSIMO, E. e S. GIOLO, 2006 *Análise de Sobrevida Aplicada*. Edigard Blucher, São Paulo.
- COOK, R. D. e S. WEISBERG, 1982 *Residuals and influence in regression*. New York: Chapman and Hall.
- CORDEIRO, G., F. CASTELLARES, G. SILVA, e E. ORTEGA, 2013a The beta-weibull geometric distribution. *Statistics* **47**: 34–78.
- CORDEIRO, G., R. PESCIM, C. DEMÉTRIO, e E. ORTEGA, 2014 The kummer beta generalized gamma distribution. *Journal of Data Science* **12**: 661–698.

- CORDEIRO, G. M., M. ALIZADEH, G. OZEL, B. HOSSEINI, E. M. M. ORTEGA, e E. ALTUN, 2017 The generalized odd log-logistic family of distributions: properties, regression models and applications. *Journal of Statistical Computation and Simulation* **87**: 908–932.
- CORDEIRO, G. M., M. ALIZADEH, M. TAHIR, M. MANSOOR, M. BOURGUIGNON, e G. HAMEDANI, 2016 The beta odd log-logistic generalized family of distributions. *Hacettepe Journal of Mathematics and Statistics* **45**: 1175–1202.
- CORDEIRO, G. M., F. CASTELLARES, L. C. MONTENEGRO, e M. DE CASTRO, 2013b The beta generalized gamma distribution. *Statistics* **47**: 888–900.
- CORDEIRO, G. M., E. M. ORTEGA, e G. O. SILVA, 2011 The exponentiated generalized gamma distribution with application to lifetime data. *Journal of statistical computation and simulation* **81**: 827–842.
- COX, D., 1972 Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society - Series B* **34**: 187–220.
- COX, D. R. e D. OAKES, 1984 *Analysis of survival data*, volume 21. CRC Press.
- COX, D. R. e E. J. SNELL, 1968 A general definition of residuals. *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 248–275.
- CRUZ, J. N. D., E. M. ORTEGA, e G. M. CORDEIRO, 2016 The log-odd log-logistic weibull regression model: modelling, estimation, influence diagnostics and residual analysis. *Journal of Statistical Computation and Simulation* **86**: 1516–1538.
- DA SILVA BRAGA, A., G. CORDEIRO, E. ORTEGA, e J. NILTON DA CRUZ, 2016 The odd log-logistic normal distribution: Theory and applications in analysis of experiments. *Journal of Statistical Theory and Practice* **10**: 311–335.
- EUGENE, N., C. LEE, e F. FAMOYE, 2002 Beta-normal distribution and its applications. *Communication in Statistics - Theory and Methods* **31**: 497–512.
- EXTON, H., 1978 *Handbook of hypergeometric integrals* .
- FAMOYE, F., C. LEE, e O. OLUMOLADE, 2005 The beta-weibull distribution. *Journal of Statistical Theory and Applications* **4**: 121–136.
- GLEATON, J. e J. LYNCH, 2006 On the distribution of the breaking strain of a bundle of brittle elastic fibers. *Advances in Applied Probability* **36**: 98–115.
- GRADSHTEYN, I. S. e I. RYZHIK, 2000 *Table of integrals, series, and products*. translated from the russian. translation edited and with a preface by alan jeffrey and daniel zwillinger.

- KAKDE, C. e D. SHIRKE, 2006 On exponentiated lognormal distribution. *International Journal of Agricultural and Statistical Sciences* **2**: 319–326.
- KALBFLEISCH, J. D. e R. L. PRENTICE, 2002 *Wiley series in probability and statistics. The Statistical Analysis of Failure Time Data*, Second Edition pp. 441–447.
- KENNEY, J. F. e E. S. KEEPING, 1962 *Mathematics of Statistics*. Part 1, 3rd edition., Van Nostrand, New Jersey.
- LANJONI, B. R., E. M. ORTEGA, e G. M. CORDEIRO, 2016 Extended burr xii regression models: theory and applications. *Journal of agricultural, biological, and environmental statistics* **21**: 203–224.
- LAWLESS, J., 2003 *Statistical Models and Methods for Lifetime Data*. Number 2, John Wiley & Sons, New Jersey.
- LOUZADA-NETO, F. e B. DE BRAGANÇA PEREIRA, 2000 Modelos em análise de sobrevivência. *Cadernos Saúde Coletiva*, Rio de Janeiro **8**: 8–26.
- LUCENA, S., A. SILVA, e G. CORDEIRO, 2015 The transmuted generalized gamma distribution: properties and application. *Journal of Data Science* **13**: 409–420.
- MALHOTRA, J., A. K. SHARMA, e R. KALER, 2009 On the performance analysis of wireless receiver using generalized-gamma fading model. *annals of telecommunications-Annales des télécommunications* **64**: 147–153.
- MARSHALL, A. e I. OLKIN, 1997 A new method for adding a parameter to a family of distributions with application to the exponential and weibull families. *Biometrika* **84**: 641–652.
- MEEKER, J. F. e L. A. ESCOBAR, 1998 *Statistical Methods for Reliability Data*. John Wiley & Sons Inc.
- MOORS, J., 1988 A quantile alternative for kurtosis. *The statistician* pp. 25–32.
- MUDHOLKAR, G., D. SRIVASTAVA, e M. FREIMER, 1995 The exponentiated weibull family: A reanalysis of the bus-motor-failure data. *Technometrics* **37**: 436–445.
- MUDHOLKAR, G., D. SRIVASTAVA, e G. KOLLIA, 1996 A generalization of the weibull distribution with application to the analysis of survival data. *Journal of American Statistical Association* **91**: 1575–1583.
- NADARAJAH, S., 2006 The exponentiated gumbel distribution with climate application. *Environmetrics* **17**: 13–23.

- NADARAJAH, S., 2008 A review of results on sums of random variables. *Acta Applicandae Mathematicae* **103**: 131–140.
- NADARAJAH, S. e A. K. GUPTA, 2007 The exponentiated gamma distribution with application to drought data. *Calcutta Statistical Association Bulletin* **59**: 29–54.
- NADARAJAH, S. e S. KOTZ, 2006 The exponentiated type distributions. *Acta applicandae mathematicae* **92**: 97–111.
- NELSON, W. B., 2009 *Accelerated testing: statistical models, test plans, and data analysis*, volume 344. John Wiley & Sons.
- ORTEGA, E. M., G. CORDEIRO, e M. PASCOA, 2011 The generalized gamma geometric distribution. *Journal of Statistical Theory and Applications* **10**: 433–454.
- ORTEGA, E. M., G. A. PAULA, e H. BOLFARINE, 2008 Deviance residuals in generalised log-gamma regression models with censored observations. *Journal of Statistical Computation and Simulation* **78**: 747–764.
- PASCOA, M., E. ORTEGA, e G. CORDEIRO, 2011 The kumaraswamy generalized gamma distribution with application in survival analysis. *Statistical Methodology* **8**: 411–433.
- PERDONÁ, G., 2006 *Modelos de riscos aplicados à análise de sobrevivência*. Tese (Doutorado em Ciências - Ciências da Computação e Matemática Computacional) - Instituto de Ciências Matemática e de Computação, Universidade de São Paulo, São Carlos.
- PRENTICE, R., 1974 A log-gama model and its maximum likelihood estimation. *Biometrika* **61**: 539–544.
- R DEVELOPMENT CORE TEAM, 2014 *R*. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
- SAS ONDEMAND FOR ACADEMICS, 2015 *SAS*. SAS Institute Inc, Cary, NC.
- SCHWARZ, G. *et al.*, 1978 Estimating the dimension of a model. *The annals of statistics* **6**: 461–464.
- SILVA, A., 2004 *Estudo evolutivo das crianças expostas ao HIV e notificadas pelo núcleo de vigilância epidemiológica do HCFMRP - USP*. Dissertação (Mestrado em Saúde na Comunidade) - Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto.
- SILVA, G. O., E. M. ORTEGA, e G. A. PAULA, 2011 Residuals for log-burr xii regression models in survival analysis. *Journal of applied statistics* **38**: 1435–1445.

- SILVA, R. B., M. BOURGUIGNON, e G. M. CORDEIRO, 2016 A new compounding family of distributions: The generalized gamma power series distributions. *Journal of Computational and Applied Mathematics* **303**: 119–139.
- STACY, E., 1962 A generalization of the gamma distribution. *Annals of Mathematical Statistics* **33**: 1187–1192.
- STACY, E. e G. MIHRAM, 1965 Parameter estimation for a generalized gamma distribution. *Tecnometrics* **7**: 349–358.
- STONE, G., 1978 *Statistical analysis of accelerated aging tests on solid electrical insulation*. Unpublished MA Sc. thesis, University of Waterloo, Waterloo, Ontario, Canada.
- TROTT, M., 2006 *The Mathematica guidebook for numerics*. Springer Science & Business Media.
- WEISBERG, S., 2005 *Applied linear regression*, volume 528. John Wiley & Sons.
- WRIGHT, E. M., 1935 The asymptotic expansion of the generalized hypergeometric function. *Journal of the London Mathematical Society* **1**: 286–293.

APÊNDICES

Apêndice A

Derivadas de segunda ordem (Capítulo 3)

Após obter as derivadas de primeira ordem de (3.21), as componentes da matriz de informação observada $J(\boldsymbol{\theta})$ para os parâmetros α , τ , k e λ são:

$$\begin{aligned}
 J_{\alpha\alpha}(\boldsymbol{\theta}) = & \frac{r\tau k}{\alpha^2} - (1 + \tau) \frac{\tau}{\alpha^2} \sum_{i \in F} \left(\frac{t_i}{\alpha} \right)^\tau + (\lambda - 1) \sum_{i \in F} \frac{[\ddot{u}_i]_\alpha \bar{u}_i - 2[\dot{u}_i^2]_\alpha - u_i [\ddot{u}_i]_\alpha}{u_i \bar{u}_i} \\
 & - (\lambda - 1) \sum_{i \in F} \left[\frac{[\dot{u}_i]_\alpha \bar{u}_i - u_i [\dot{u}_i]_\alpha}{u_i \bar{u}_i} \right]^2 - 2\lambda(\lambda - 1) \sum_{i \in F} \frac{u_i^{(\lambda-2)} [\dot{u}_i^2]_\alpha}{u_i^\lambda \bar{u}_i^\lambda} \\
 & - 2\lambda \sum_{i \in F} \frac{u_i^{(\lambda-1)} [\ddot{u}_i]_\alpha}{u_i^\lambda \bar{u}_i^\lambda} - 2\lambda(\lambda - 1) \sum_{i \in F} \frac{\bar{u}_i^{(\lambda-2)} [\dot{u}_i^2]_\alpha}{u_i^\lambda \bar{u}_i^\lambda} + 2\lambda \sum_{i \in F} \frac{\bar{u}_i^{(\lambda-1)} [\ddot{u}_i]_\alpha}{u_i^\lambda \bar{u}_i^\lambda} \\
 & - 2\lambda \sum_{i \in F} \left[\frac{u_i^{(\lambda-1)} [\dot{u}_i]_\alpha - \bar{u}_i^{(\lambda-1)} [\ddot{u}_i]_\alpha}{u_i^\lambda \bar{u}_i^\lambda} \right]^2 + \lambda \sum_{i \in C} \frac{[\ddot{u}_i]_\alpha}{\bar{u}_i} + \lambda \sum_{i \in C} \frac{[\dot{u}_i^2]_\alpha}{\bar{u}_i^2} \\
 & - \lambda \sum_{i \in C} \frac{[\ddot{u}_i]_\alpha [u_i^{(\lambda-1)} - \bar{u}_i^{(\lambda-1)}]}{u_i^\lambda + \bar{u}_i^\lambda} + \lambda(\lambda - 1) \sum_{i \in C} \frac{[\dot{u}_i^2]_\alpha [u_i^{(\lambda-2)} - \bar{u}_i^{(\lambda-2)}]}{u_i^\lambda + \bar{u}_i^\lambda} \\
 & + \lambda^2 \sum_{i \in C} \left[\frac{[\dot{u}_i]_\alpha (u_i^{(\lambda-1)} - \bar{u}_i^{(\lambda-1)})}{u_i^\lambda + \bar{u}_i^\lambda} \right]^2,
 \end{aligned}$$

$$\begin{aligned}
 J_{\alpha\tau}(\boldsymbol{\theta}) = & -\frac{rk}{\alpha} + \frac{1}{\alpha} \sum_{i \in F} \left(\frac{t_i}{\alpha} \right)^\tau + \frac{\tau}{\alpha} \sum_{i \in F} \left[\left(\frac{t_i}{\alpha} \right)^\tau \log \left(\frac{t_i}{\alpha} \right) \right] + (\lambda - 1) \sum_{i \in F} \frac{[\dot{u}_i]_{\alpha\tau} [\bar{u}_i - u_i]}{u_i \bar{u}_i} \\
 & - 2(\lambda - 1) \sum_{i \in F} \frac{[\dot{u}_i]_\alpha [\dot{u}_i]_\tau}{u_i \bar{u}_i} - (\lambda - 1) \sum_{i \in F} \frac{[\dot{u}_i]_\alpha [\dot{u}_i]_\tau [\bar{u}_i - u_i]^2}{[u_i \bar{u}_i]^2} \\
 & - 2\lambda \sum_{i \in F} \frac{[\dot{u}_i]_{\alpha\tau} [u_i^{(\lambda-1)} - \bar{u}_i^{(\lambda-1)}]}{[u_i^\lambda + \bar{u}_i^\lambda]} - 2\lambda(\lambda - 1) \sum_{i \in F} \frac{[\dot{u}_i]_\alpha [\dot{u}_i]_\tau [u_i^{(\lambda-2)} - \bar{u}_i^{(\lambda-2)}]}{[u_i^\lambda + \bar{u}_i^\lambda]} \\
 & + 2\lambda^2 \sum_{i \in F} \frac{[\dot{u}_i]_\alpha [\dot{u}_i]_\tau [u_i^{(\lambda-1)} - \bar{u}_i^{(\lambda-1)}]^2}{[u_i^\lambda + \bar{u}_i^\lambda]^2} - \lambda \sum_{i \in C} \frac{[\dot{u}_i]_{\alpha\tau}}{\bar{u}_i} - \lambda \sum_{i \in C} \frac{[\dot{u}_i]_\alpha [\dot{u}_i]_\tau}{\bar{u}_i^2} \\
 & - \lambda \sum_{i \in C} \frac{[\dot{u}_i]_{\alpha\tau} [u_i^{(\lambda-1)} - \bar{u}_i^{(\lambda-1)}]}{[u_i^\lambda + \bar{u}_i^\lambda]} - \lambda(\lambda - 1) \sum_{i \in C} \frac{[\dot{u}_i]_\alpha [\dot{u}_i]_\tau [u_i^{(\lambda-2)} - \bar{u}_i^{(\lambda-2)}]}{[u_i^\lambda + \bar{u}_i^\lambda]} \\
 & - \lambda^2 \sum_{i \in C} \frac{[\dot{u}_i]_\alpha [\dot{u}_i]_\tau [u_i^{(\lambda-1)} - \bar{u}_i^{(\lambda-1)}]^2}{[u_i^\lambda + \bar{u}_i^\lambda]^2},
 \end{aligned}$$

$$\begin{aligned}
J_{\alpha k}(\boldsymbol{\theta}) = & -\frac{r\tau}{\alpha} + (\lambda - 1) \sum_{i \in F} \frac{[\dot{u}_i]_{\alpha k} [\bar{u}_i - u_i]}{u_i \bar{u}_i} - 2(\lambda - 1) \sum_{i \in F} \frac{[\dot{u}_i]_{\alpha} [\dot{u}_i]_k}{u_i \bar{u}_i} \\
& - (\lambda - 1) \sum_{i \in F} \frac{[\dot{u}_i]_{\alpha} [\dot{u}_i]_k [\bar{u}_i - u_i]^2}{[u_i \bar{u}_i]^2} - 2\lambda \sum_{i \in F} \frac{[\dot{u}_i]_{\alpha k} [u_i^{(\lambda-1)} - \bar{u}_i^{(\lambda-1)}]}{[u_i^{\lambda} + \bar{u}_i^{\lambda}]} \\
& - 2\lambda(\lambda - 1) \sum_{i \in F} \frac{[\dot{u}_i]_{\alpha} [\dot{u}_i]_k [u_i^{(\lambda-2)} - \bar{u}_i^{(\lambda-2)}]}{[u_i^{\lambda} + \bar{u}_i^{\lambda}]} + 2\lambda^2 \sum_{i \in F} \frac{[\dot{u}_i]_{\alpha} [\dot{u}_i]_k [u_i^{(\lambda-1)} - \bar{u}_i^{(\lambda-1)}]^2}{[u_i^{\lambda} + \bar{u}_i^{\lambda}]^2} \\
& - \lambda \sum_{i \in C} \frac{[\dot{u}_i]_{\alpha k}}{\bar{u}_i} - \lambda \sum_{i \in C} \frac{[\dot{u}_i]_{\alpha} [\dot{u}_i]_k}{\bar{u}_i^2} - \lambda \sum_{i \in C} \frac{[\dot{u}_i]_{\alpha k} [u_i^{(\lambda-1)} - \bar{u}_i^{(\lambda-1)}]}{[u_i^{\lambda} + \bar{u}_i^{\lambda}]} \\
& - \lambda(\lambda - 1) \sum_{i \in C} \frac{[\dot{u}_i]_{\alpha} [\dot{u}_i]_k [u_i^{(\lambda-2)} - \bar{u}_i^{(\lambda-2)}]}{[u_i^{\lambda} + \bar{u}_i^{\lambda}]} - \lambda^2 \sum_{i \in C} \frac{[\dot{u}_i]_{\alpha} [\dot{u}_i]_k [u_i^{(\lambda-1)} - \bar{u}_i^{(\lambda-1)}]^2}{[u_i^{\lambda} + \bar{u}_i^{\lambda}]^2},
\end{aligned}$$

$$\begin{aligned}
J_{\alpha \lambda}(\boldsymbol{\theta}) = & \sum_{i \in F} \frac{[\dot{u}_i]_{\alpha} [\bar{u}_i - u_i]}{u_i \bar{u}_i} + 2\lambda \sum_{i \in F} \frac{[\dot{u}_i]_{\alpha} [u_i^{(\lambda-1)} - \bar{u}_i^{(\lambda-1)}] [u_i^{(\lambda-1)} \log(u) - \bar{u}_i^{(\lambda-1)} \log(\bar{u}_i)]}{[u_i^{\lambda} + \bar{u}_i^{\lambda}]^2} \\
& - 2\lambda \sum_{i \in F} \frac{[\dot{u}_i]_{\alpha} [u_i^{(\lambda-1)} \log(u) - \bar{u}_i^{(\lambda-1)} \log(\bar{u}_i)]}{u_i^{\lambda} + \bar{u}_i^{\lambda}} - 2 \sum_{i \in F} \frac{[\dot{u}_i]_{\alpha} [u_i^{(\lambda-1)} - \bar{u}_i^{(\lambda-1)}]}{u_i^{\lambda} + \bar{u}_i^{\lambda}} \\
& - \sum_{i \in C} \frac{[\dot{u}_i]_{\alpha}}{\bar{u}_i} - \sum_{i \in C} \frac{[\dot{u}_i]_{\alpha} [u_i^{(\lambda-1)} - \bar{u}_i^{(\lambda-1)}]}{u_i^{\lambda} + \bar{u}_i^{\lambda}} - \sum_{i \in C} \frac{[\dot{u}_i]_{\alpha} [u_i^{(\lambda-1)} \log(u) - \bar{u}_i^{(\lambda-1)} \log(\bar{u}_i)]}{u_i^{\lambda} + \bar{u}_i^{\lambda}} \\
& + \sum_{i \in C} \frac{[\dot{u}_i]_{\alpha} [u_i^{(\lambda-1)} - \bar{u}_i^{(\lambda-1)}] [u_i^{(\lambda-1)} \log(u) - \bar{u}_i^{(\lambda-1)} \log(\bar{u}_i)]}{[u_i^{\lambda} + \bar{u}_i^{\lambda}]^2},
\end{aligned}$$

$$\begin{aligned}
J_{\tau\tau}(\boldsymbol{\theta}) = & -\frac{r}{\tau^2} - \sum_{i \in F} \left[\left(\frac{t_i}{\alpha} \right)^{\tau} \log^2 \left(\frac{t_i}{\alpha} \right) \right] + (\lambda - 1) \sum_{i \in F} \frac{[\ddot{u}_i]_{\tau} \bar{u}_i - 2[\dot{u}_i^2]_{\tau} - u_i [\ddot{u}_i]_{\tau}}{u_i \bar{u}_i} \\
& - (\lambda - 1) \sum_{i \in F} \left[\frac{[\dot{u}_i]_{\tau} \bar{u}_i - u_i [\dot{u}_i]_{\tau}}{u_i \bar{u}_i} \right]^2 - 2\lambda(\lambda - 1) \sum_{i \in F} \frac{u_i^{(\lambda-2)} [\dot{u}_i^2]_{\tau}}{u_i^{\lambda} \bar{u}_i^{\lambda}} \\
& - 2\lambda \sum_{i \in F} \frac{u_i^{(\lambda-1)} [\ddot{u}_i]_{\tau}}{u_i^{\lambda} \bar{u}_i^{\lambda}} - 2\lambda(\lambda - 1) \sum_{i \in F} \frac{\bar{u}_i^{(\lambda-2)} [\dot{u}_i^2]_{\tau}}{u_i^{\lambda} \bar{u}_i^{\lambda}} + 2\lambda \sum_{i \in F} \frac{\bar{u}_i^{(\lambda-1)} [\ddot{u}_i]_{\tau}}{u_i^{\lambda} \bar{u}_i^{\lambda}} \\
& - 2\lambda \sum_{i \in F} \left[\frac{u_i^{(\lambda-1)} [\dot{u}_i]_{\tau} - \bar{u}_i^{(\lambda-1)} [\ddot{u}_i]_{\tau}}{u_i^{\lambda} \bar{u}_i^{\lambda}} \right]^2 + \lambda \sum_{i \in C} \frac{[\ddot{u}_i]_{\tau}}{\bar{u}_i} + \lambda \sum_{i \in C} \frac{[\dot{u}_i^2]_{\tau}}{\bar{u}_i^2} \\
& - \lambda \sum_{i \in C} \frac{[\ddot{u}_i]_{\tau} [(u_i^{(\lambda-1)} - \bar{u}_i^{(\lambda-1)})]}{u_i^{\lambda} + \bar{u}_i^{\lambda}} + \lambda(\lambda - 1) \sum_{i \in C} \frac{[\dot{u}_i^2]_{\tau} [u_i^{(\lambda-2)} - \bar{u}_i^{(\lambda-2)}]}{u_i^{\lambda} + \bar{u}_i^{\lambda}} \\
& + \lambda^2 \sum_{i \in C} \left[\frac{[\dot{u}_i]_{\tau} (u_i^{(\lambda-1)} - \bar{u}_i^{(\lambda-1)})}{u_i^{\lambda} + \bar{u}_i^{\lambda}} \right]^2,
\end{aligned}$$

$$\begin{aligned}
J_{\tau k}(\boldsymbol{\theta}) = & \sum_{i \in F} \log \left(\frac{t_i}{\tau} \right) + (\lambda - 1) \sum_{i \in F} \frac{[\dot{u}_i]_{\tau k} [\bar{u}_i - u_i]}{u_i \bar{u}_i} - 2(\lambda - 1) \sum_{i \in F} \frac{[\dot{u}_i]_{\tau} [\dot{u}_i]_k}{u_i \bar{u}_i} \\
& - (\lambda - 1) \sum_{i \in F} \frac{[\dot{u}_i]_{\tau} [\dot{u}_i]_k [\bar{u}_i - u_i]^2}{[u_i \bar{u}_i]^2} - 2\lambda \sum_{i \in F} \frac{[\dot{u}_i]_{\tau k} [u_i^{(\lambda-1)} - \bar{u}_i^{(\lambda-1)}]}{[u_i^\lambda + \bar{u}_i^\lambda]} \\
& - 2\lambda(\lambda - 1) \sum_{i \in F} \frac{[\dot{u}_i]_{\tau} [\dot{u}_i]_k [u_i^{(\lambda-2)} - \bar{u}_i^{(\lambda-2)}]}{[u_i^\lambda + \bar{u}_i^\lambda]} + 2\lambda^2 \sum_{i \in F} \frac{[\dot{u}_i]_{\tau} [\dot{u}_i]_k [u_i^{(\lambda-1)} - \bar{u}_i^{(\lambda-1)}]^2}{[u_i^\lambda + \bar{u}_i^\lambda]^2} \\
& - \lambda \sum_{i \in C} \frac{[\dot{u}_i]_{\tau k}}{\bar{u}_i} - \lambda \sum_{i \in C} \frac{[\dot{u}_i]_{\tau} [\dot{u}_i]_k}{\bar{u}_i^2} - \lambda \sum_{i \in C} \frac{[\dot{u}_i]_{\tau k} [u_i^{(\lambda-1)} - \bar{u}_i^{(\lambda-1)}]}{[u_i^\lambda + \bar{u}_i^\lambda]} \\
& - \lambda(\lambda - 1) \sum_{i \in C} \frac{[\dot{u}_i]_{\tau} [\dot{u}_i]_k [u_i^{(\lambda-2)} - \bar{u}_i^{(\lambda-2)}]}{[u_i^\lambda + \bar{u}_i^\lambda]} - \lambda^2 \sum_{i \in C} \frac{[\dot{u}_i]_{\tau} [\dot{u}_i]_k [u_i^{(\lambda-1)} - \bar{u}_i^{(\lambda-1)}]^2}{[u_i^\lambda + \bar{u}_i^\lambda]^2}
\end{aligned}$$

$$\begin{aligned}
J_{\tau \lambda}(\boldsymbol{\theta}) = & \sum_{i \in F} \frac{[\dot{u}_i]_{\tau} [\bar{u}_i - u_i]}{u_i \bar{u}_i} + 2\lambda \sum_{i \in F} \frac{[\dot{u}_i]_{\tau} [u_i^{(\lambda-1)} - \bar{u}_i^{(\lambda-1)}] [u_i^{(\lambda-1)} \log(u) - \bar{u}_i^{(\lambda-1)} \log(\bar{u}_i)]}{[u_i^\lambda + \bar{u}_i^\lambda]^2} \\
& - 2\lambda \sum_{i \in F} \frac{[\dot{u}_i]_{\tau} [u_i^{(\lambda-1)} \log(u) - \bar{u}_i^{(\lambda-1)} \log(\bar{u}_i)]}{u_i^\lambda + \bar{u}_i^\lambda} - 2 \sum_{i \in F} \frac{[\dot{u}_i]_{\tau} [u_i^{(\lambda-1)} - \bar{u}_i^{(\lambda-1)}]}{u_i^\lambda + \bar{u}_i^\lambda} \\
& - \sum_{i \in C} \frac{[\dot{u}_i]_{\tau} [u_i^{(\lambda-1)} - \bar{u}_i^{(\lambda-1)}]}{u_i^\lambda + \bar{u}_i^\lambda} - \sum_{i \in C} \frac{[\dot{u}_i]_{\tau} [u_i^{(\lambda-1)} \log(u) - \bar{u}_i^{(\lambda-1)} \log(\bar{u}_i)]}{u_i^\lambda + \bar{u}_i^\lambda} \\
& - \sum_{i \in C} \frac{[\dot{u}_i]_{\tau}}{\bar{u}_i} + \sum_{i \in C} \frac{[\dot{u}_i]_{\tau} [u_i^{(\lambda-1)} - \bar{u}_i^{(\lambda-1)}] [u_i^{(\lambda-1)} \log(u) - \bar{u}_i^{(\lambda-1)} \log(\bar{u}_i)]}{[u_i^\lambda + \bar{u}_i^\lambda]^2},
\end{aligned}$$

$$\begin{aligned}
J_{kk}(\boldsymbol{\theta}) = & -\psi'(k) + \tau \sum_{i \in F} \log \left(\frac{t_i}{\alpha} \right) + (\lambda - 1) \sum_{i \in F} \frac{[\ddot{u}_i]_k \bar{u}_i - 2[\dot{u}_i^2]_k - u_i [\ddot{u}_i]_k}{u_i \bar{u}_i} \\
& - (\lambda - 1) \sum_{i \in F} \left[\frac{[\dot{u}_i]_k \bar{u}_i - u_i [\dot{u}_i]_k}{u_i \bar{u}_i} \right]^2 - 2\lambda(\lambda - 1) \sum_{i \in F} \frac{u_i^{(\lambda-2)} [\dot{u}_i^2]_k}{u_i^\lambda \bar{u}_i^\lambda} \\
& - 2\lambda \sum_{i \in F} \frac{u_i^{(\lambda-1)} [\ddot{u}_i]_k}{u_i^\lambda \bar{u}_i^\lambda} - 2\lambda(\lambda - 1) \sum_{i \in F} \frac{\bar{u}_i^{(\lambda-2)} [\dot{u}_i^2]_k}{u_i^\lambda \bar{u}_i^\lambda} + 2\lambda \sum_{i \in F} \frac{\bar{u}_i^{(\lambda-1)} [\ddot{u}_i]_k}{u_i^\lambda \bar{u}_i^\lambda} \\
& - 2\lambda \sum_{i \in F} \left[\frac{u_i^{(\lambda-1)} [\dot{u}_i]_k - \bar{u}_i^{(\lambda-1)} [\ddot{u}_i]_k}{u_i^\lambda \bar{u}_i^\lambda} \right]^2 + \lambda \sum_{i \in C} \frac{[\ddot{u}_i]_k}{\bar{u}_i} + \lambda \sum_{i \in C} \frac{[\dot{u}_i^2]_k}{\bar{u}_i^2} \\
& - \lambda \sum_{i \in C} \frac{[\ddot{u}_i]_k [u_i^{(\lambda-1)} - \bar{u}_i^{(\lambda-1)}]}{u_i^\lambda + \bar{u}_i^\lambda} + \lambda(\lambda - 1) \sum_{i \in C} \frac{[\dot{u}_i^2]_k [u_i^{(\lambda-2)} - \bar{u}_i^{(\lambda-2)}]}{u_i^\lambda + \bar{u}_i^\lambda} \\
& + \lambda^2 \sum_{i \in C} \left[\frac{[\dot{u}_i]_k (u_i^{(\lambda-1)} - \bar{u}_i^{(\lambda-1)})}{u_i^\lambda + \bar{u}_i^\lambda} \right]^2,
\end{aligned}$$

$$\begin{aligned}
J_{k\lambda}(\boldsymbol{\theta}) &= \sum_{i \in F} \frac{[\dot{u}_i]_k [\bar{u}_i - u_i]}{u_i \bar{u}_i} + 2\lambda \sum_{i \in F} \frac{[\dot{u}_i]_k \left[u_i^{(\lambda-1)} - \bar{u}_i^{(\lambda-1)} \right] \left[u_i^{(\lambda-1)} \log(u) - \bar{u}_i^{(\lambda-1)} \log(\bar{u}_i) \right]}{\left[u_i^\lambda + \bar{u}_i^\lambda \right]^2} \\
&\quad - 2\lambda \sum_{i \in F} \frac{[\dot{u}_i]_k \left[u_i^{(\lambda-1)} \log(u) - \bar{u}_i^{(\lambda-1)} \log(\bar{u}_i) \right]}{u_i^\lambda + \bar{u}_i^\lambda} - 2 \sum_{i \in F} \frac{[\dot{u}_i]_k \left[u_i^{(\lambda-1)} - \bar{u}_i^{(\lambda-1)} \right]}{u_i^\lambda + \bar{u}_i^\lambda} \\
&\quad - \sum_{i \in C} \frac{[\dot{u}_i]_k}{\bar{u}_i} - \sum_{i \in C} \frac{[\dot{u}_i]_k \left[u_i^{(\lambda-1)} - \bar{u}_i^{(\lambda-1)} \right]}{u_i^\lambda + \bar{u}_i^\lambda} - \sum_{i \in C} \frac{[\dot{u}_i]_k \left[u_i^{(\lambda-1)} \log(u) - \bar{u}_i^{(\lambda-1)} \log(\bar{u}_i) \right]}{u_i^\lambda + \bar{u}_i^\lambda} \\
&\quad + \sum_{i \in C} \frac{[\dot{u}_i]_k \left[u_i^{(\lambda-1)} - \bar{u}_i^{(\lambda-1)} \right] \left[u_i^{(\lambda-1)} \log(u) - \bar{u}_i^{(\lambda-1)} \log(\bar{u}_i) \right]}{\left[u_i^\lambda + \bar{u}_i^\lambda \right]^2},
\end{aligned}$$

$$\begin{aligned}
J_{\lambda\lambda}(\boldsymbol{\theta}) &= -\frac{r}{\lambda^2} + 2 \sum_{i \in F} \left[\frac{\bar{u}_i^\lambda \log(\bar{u}_i) + u_i^\lambda \log(u_i)}{\bar{u}_i^\lambda + u_i^\lambda} \right]^2 - 2 \sum_{i \in F} \frac{\bar{u}_i^\lambda \log^2(\bar{u}_i) + u_i^\lambda \log^2(u_i)}{\bar{u}_i^\lambda + u_i^\lambda} \\
&\quad + \sum_{i \in C} \left[\frac{\bar{u}_i^\lambda \log(\bar{u}_i) + u_i^\lambda \log(u_i)}{\bar{u}_i^\lambda + u_i^\lambda} \right]^2 - \sum_{i \in C} \frac{\bar{u}_i^\lambda \log^2(\bar{u}_i) + u_i^\lambda \log^2(u_i)}{\bar{u}_i^\lambda + u_i^\lambda},
\end{aligned}$$

em que

$$\begin{aligned}
[\ddot{u}_i]_\alpha &= \frac{\partial^2 \gamma_1 \left(k, \left(\frac{t_i}{\alpha} \right)^\tau \right)}{\partial \alpha \partial \alpha}; & [\ddot{u}_i]_\tau &= \frac{\partial^2 \gamma_1 \left(k, \left(\frac{t_i}{\alpha} \right)^\tau \right)}{\partial \tau \partial \tau}; & [\ddot{u}_i]_k &= \frac{\partial^2 \gamma_1 \left(k, \left(\frac{t_i}{\alpha} \right)^\tau \right)}{\partial k \partial k}, \\
[\dot{u}_i]_{\alpha\tau} &= \frac{\partial^2 \gamma_1 \left(k, \left(\frac{t_i}{\alpha} \right)^\tau \right)}{\partial \alpha \partial \tau}; & [\dot{u}_i]_{\alpha k} &= \frac{\partial^2 \gamma_1 \left(k, \left(\frac{t_i}{\alpha} \right)^\tau \right)}{\partial \alpha \partial k}; & [\dot{u}_i]_{\tau k} &= \frac{\partial^2 \gamma_1 \left(k, \left(\frac{t_i}{\alpha} \right)^\tau \right)}{\partial \tau \partial k}.
\end{aligned}$$

Temos assim

$$[\dot{u}_i]_\alpha = -\frac{\tau}{\alpha \Gamma(k)} \left(\frac{t_i}{\alpha} \right)^{\tau k} \exp \left\{ -\left(\frac{t_i}{\alpha} \right)^\tau \right\},$$

$$[\dot{u}_i]_\tau = \frac{1}{\Gamma(k)} \left(\frac{t_i}{\alpha} \right)^{\tau k} \log \left(\frac{t_i}{\alpha} \right) \exp \left\{ -\left(\frac{t_i}{\alpha} \right)^\tau \right\},$$

$$[\dot{u}_i]_k = \sum_{m=0}^{\infty} \frac{(-1)^m}{n} J(u_i, k+m-1, 1),$$

$$[\ddot{u}_i]_\alpha = \frac{\tau}{\alpha^2 \Gamma(k)} \left(\frac{t_i}{\alpha} \right)^{\tau k} \exp \left\{ -\left(\frac{t_i}{\alpha} \right)^\tau \right\} \left[1 + \tau k - \tau \left(\frac{t_i}{\alpha} \right)^\tau \right],$$

$$[\ddot{u}_i]_\tau = \left(\frac{t_i}{\alpha} \right)^{\tau k} \left[\log \left(\frac{t_i}{\alpha} \right) \right]^2 \exp \left\{ -\left(\frac{t_i}{\alpha} \right)^\tau \right\} \left[k - \left(\frac{t_i}{\alpha} \right)^\tau \right],$$

$$[\ddot{u}_i]_k = \sum_{m=0}^{\infty} \frac{(-1)^m}{n} J(u_i, k+m-1, 2),$$

$$[\dot{u}_i]_{\alpha\tau} = \frac{\tau}{\alpha\Gamma(k)} \left(\frac{t_i}{\alpha}\right)^{\tau k} \exp\left\{-\left(\frac{t_i}{\alpha}\right)^\tau\right\} \left[1 - \tau k \log\left(\frac{t_i}{\alpha}\right) + \tau \left(\frac{t_i}{\alpha}\right)^\tau \log\left(\frac{t_i}{\alpha}\right)\right],$$

$$[\dot{u}_i]_{\alpha k} = \frac{\tau}{\alpha\Gamma(k)} \left(\frac{t_i}{\alpha}\right)^{\tau k} \exp\left\{-\left(\frac{t_i}{\alpha}\right)^\tau\right\} \left[\frac{\psi(k)}{\Gamma(k)} - \tau \log\left(\frac{t_i}{\alpha}\right)\right],$$

$$[\dot{u}_i]_{\tau k} = \frac{1}{\Gamma(k)} \left(\frac{t_i}{\alpha}\right)^{\tau k} \exp\left\{-\left(\frac{t_i}{\alpha}\right)^\tau\right\} \left[\tau \log\left(\frac{t_i}{\alpha}\right) - \frac{\psi(k)}{\Gamma(k)}\right],$$

em que $\psi'(\cdot)$ é a derivada da função digama. A função $J(\cdot, \cdot, \cdot)$ pode ser calculada pela integral dada por

$$J(a, p, 1) = \int_0^a x^p \log(x) dx = \frac{a^{p+1}}{(p+1)^2} [(p+1) \log(a) - 1],$$

e

$$J(a, p, 2) = \int_0^a x^p \log(x)^2 dx = \frac{a^{p+1}}{(p+1)^3} \{2 - (p+1) \log(a) [2 - (p+1) \log(a)]\}.$$

Apêndice B

Programas utilizados no R (Capítulo 3)

Programa desenvolvido no R para construção da distribuição OLLGG e geração de valores.

```
#Pacote que contem a GG (Stacy , 1962)
library("flexsurv")
#Construindo a distribuição OLLGG
OLL.GG <- function(t,lambda,alpha,tau,k){
#função densidade de probabilidade da gama generalizada
g <- dgengamma.orig(t, shape=tau,scale=alpha,k)
#função distribuição de probabilidade da gama generalizada
G <- pgengamma.orig(t,shape=tau,scale=alpha,k)
#construindo a densidade da OLLGG;
f <- (lambda*g*((G*(1-G))^(lambda-1)))/(((G^lambda)+(1-G)^lambda)^2)
#construindo a distribuição da OLLGG;
F <- (G^lambda)/(((G^lambda)+(1-G)^lambda))
#construindo a sobrevivência da OLLGG
S <- 1-F
#construindo o risco da OLLGG
h <-f/S
return(list(f= f,F=F, S=S,h=h) )
}

#Construção do gerador de valores da OLLGG
```



```

geradora <- function(n){
u <- runif(n,0,1)
l <- 1/lambda
u1 <- (u^l)/(((1-u)^l)+u^l)
t <- qgengamma.orig(u1,shape = tau,scale = alpha,k=k)
return(t)
}

x <- geradora(n)

```

Programas utilizados no R (Capítulo 4)

Programa desenvolvido no R para construção da distribuição LOLLGG e geração de valores para o modelo de regressão.

```

library('flexsurv')
#Caso em que q é diferente de zero
LOLL.GG <- function(x,lambda,mu,sigma,q){
#y = x
#função densidade de probabilidade da log-gama generalizada;
w <- (x-mu)/sigma
g1 <- (abs(q)/(sigma*gamma(q^-2)))*
((q^-2)^(q^-2))*exp((q^-2)*(q*w - exp(q*w)))
#função distribuição de probabilidade da log-gama generalizada;
G1 <- pgamma(exp(x*(q/sigma)),scale=exp((q/sigma)*
mu + log(q^2)),q^-2)
#construindo a densidade da LOLLGG;
f <- (lambda*g1*((G1*(1-G1))^(lambda-1)))/(((G1^lambda)+
(1-G1)^lambda)^2)
#construindo a distribuição da LOLLGG;
if(q>0) F1 <- (G1^lambda)/(((G1^lambda)+(1-G1)^lambda))
else F1 <- ((1-G1)^lambda)/(((G1^lambda)+(1-G1)^lambda))
F <- F1
#construindo a sobrevivência da LOLLGG
S <- 1-F
#construindo o risco da LOLLGG
h <-f/S
return(list(f= f,F=F, S=S,h=h) )
}

#Caso em que q é igual a zero
OLL.N <- function(x,lambda,mu,sigma){
z <- (x - mu)/sigma
g1 <- (1/(sigma*sqrt(2*pi)))*exp(-0.5*(z)^2)
G1 <- pnorm(z)
#construindo a densidade da OLLN;
f <- (lambda*g1*((G1*(1-G1))^(lambda-1)))/(((G1^lambda)+
(1-G1)^lambda)^2)

```

```

F <- (G1^lambda)/(((G1^lambda)+(1-G1)^lambda))
#construindo a sobrevivência da OLLN;
S <- 1-F
#construindo o risco da OLLN;
h <-f/S
return(list(f= f,F=F, S=S,h=h) )
}

#Função de densidade log-OLLGG com os casos q <0, q>0 e q=0;
Dens.Log <- function(x,lambda,mu,sigma,q){
f1 <- if(q!=0) LOLG.GG(x,lambda,mu,sigma,q)$f else
  OLL.N(x,lambda,mu,sigma)$f
return (f1)
}

#Função de distribuição log-OLLGG com os casos q <0, q>0 e q=0;
Dist.Log <- function(x,lambda,mu,sigma,q){
F1 <- if(q!=0) LOLG.GG(x,lambda,mu,sigma,q)$F else
  OLL.N(x,lambda,mu,sigma)$F
return (F1)
}

#Função de sobrevivência log-OLLGG com os casos q <0, q>0 e q=0;
Sobre.Log <- function(x,lambda,mu,sigma,q){
S1 <- if(q!=0) LOLG.GG(x,lambda,mu,sigma,q)$S else
  OLL.N(x,lambda,mu,sigma)$S
return (S1)
}

#Função de risco log-OLLGG com os casos q <0, q>0 e q=0;
Risco.Log <- function(x,lambda,mu,sigma,q){
h1 <- if(q!=0) LOLG.GG(x,lambda,mu,sigma,q)$h else
  OLL.N(x,lambda,mu,sigma)$h
return (h1)
}

### Função para geração da distribuição OLLGG, por meio da função
#quantílica da função de base (GG).
rLog.OLLGG <- function(n,lambda,mu,sigma,q){
u <- runif(n,0,1)
l <- 1/lambda
u1 <- (u^l)/(((1-u)^l)+u^l)
t.rep <- qgengamma(u1, mu=mu, sigma=sigma, Q=q)
#entrar com as covariáveis
y <- log(t.rep)
return(y)
}

```

```
##### Gera OLLN
rLog.OLLN <- function(n,lambda,mu,sigma){
  u <- runif(n,0,1)
  l <- 1/lambda
  u1 <- (u^l)/(((1-u)^l)+u^l)
  y <- qnorm(u1, mean=mu, sd=sigma)#entrar com as covariáveis
  return(y)
}

#Função para gerar valores log-OLLGG com os casos q <0, q>0 e q=0;
Gera.LOLLGG <- function(n,lambda,mu,sigma,q){
  y <- if (q!=0) rLog.OLLGG(n,lambda,mu,sigma,q) else
    rLog.OLLN(n,lambda,mu,sigma)
  return (y)
}

## Gerando o modelo de regressão com uma covariável
beta0 <- 2; beta1 <- 3; sigma0 <- 0.50;
lambda0 <- 0.50; q0 <- -0.65

gera.regre <- function(n){
  x1 <- runif(n,0,1)
  mu.x <- beta0 + beta1*x1
  zi <- Gera.LOLLGG(n,lambda=lambda0,mu=0,sigma=1,q=q0)
  #Gera LOLLGG padrão LOLLGG(lambda=?,mu=0,sigma=1,q=?)
  yi <- mu.x + sigma0*zi
  return(list(yi=yi,x1=x1, zi=zi))
}

dados <- gera.regre(n)
```