

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

**Modelo de regressão log-gama generalizado exponenciado com dados
censurados**

Epaminondas de Vasconcellos Couto

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

**Piracicaba
2010**

Epaminondas de Vasconcellos Couto
Bacharel em Estatística

Modelo de regressão log-gama generalizado exponenciado com dados censurados

Orientador:
Prof. Dr. **EDWIN MOISÉS MARCOS ORTEGA**

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

Piracicaba
2010

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

Couto, Epaminondas de Vasconcellos

Modelo de regressão log-gama generalizado exponenciado com dados censurados /
Epaminondas de Vasconcellos Couto. - - Piracicaba, 2010.
99 p. : il.

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

1. Análise de regressão e de correlação 2. Análise de sobrevivência 3. Dados censurados
4. Distribuições - Probabilidade 5. Processos estocásticos 6. Verossimilhança I. Título

CDD 519.536
C871m

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

Dedicatória

A

DEUS

por sempre ter me abençoado, protegido, confortado e iluminado em todas as horas da minha vida.

Aos meus pais,

Ailton Marques Couto e

Neide Maria de Vasconcellos Couto,

por tudo que hoje sou, que me fizeram acreditar que o mais importante na vida é o crescimento interior.

Aos meus irmãos **Neilton** e **Nívia**, que alegam minha vida dando-me motivos e estímulos para progredir.

As minhas sobrinhas **Nícole Letícia** e **Aléxia Vitória**, amo muito vocês.

A minha avó **Maria de Lourdes Marques Couto** (*in memoriam*), saudades eternas.

AGRADECIMENTOS

Aos meus PAIS pelas noites perdidas após um dia de trabalho exaustivo. Agradeço-os também pelo carinho, compreensão, confiança, paciência, críticas, amor, apoio, conselhos e por todas as outras coisas aqui não citadas.

Aos meus FAMILIARES pelo carinho, amor, amizade, paciência, incentivos e conforto nas horas precisas.

À minha AMIGA Lívia Silva Sousa e FAMILIARES pelo carinho, confiança depositada, conforto e conselhos.

Ao Prof. Dr. Edwin Moisés Marcos Ortega pela orientação, pelo conhecimento compartilhado e apoio, tornando possível a realização deste trabalho.

Aos PROFESSORES do curso de Pós-graduação em Estatística e Experimentação Agronômica, Dr. César Gonçalves de Lima, Dra. Clarice Garcia Borges Demétrio, Dra. Roseli Aparecida Leandro e Dra. Sônia Maria de Stefano Piedade, pela formação.

À PROFESSORA Dra. Giovana de Oliveira Silva pela moradia nos meus primeiros dias na cidade, ótimas conversas, pela ajuda e atenção.

Aos FUNCIONÁRIOS do Departamento de Ciências Exatas da ESALQ/USP pelos auxílios permanentes.

Aos meus COLEGAS, alguns deles, hoje, que se tornaram GRANDES AMIGOS, Alexandre Barbosa (*in memoriam*), Ana Patrícia, Cícero Costa, Eduardo Silva, Elizabeth Hashimoto, Fabiane Silva, Fernanda Rizzato, Fernando Edmar, Jaime Pena, João Gustavo, Juliana Fachini, Kelyny Martins, Luciano Amaral, Ludmila Freitas, Luís Rogério, Marcelino Alves, Márcio Silva, Patrícia Paranaíba, Raphael Dias, Renato Machado, Renato Pereira, Ricardo Olinda, Ricardo Silva, Rodrigo Pescim, Simone Sartorio, Tiago Almeida, Tiago Flor, Vanderson Rabelo e Wecley Prates pela amizade, força, troca de conhecimentos, atenção recebida em todos os momentos e alegrias que sempre me proporcionaram.

À CAPES pelo fundamental suporte financeiro concedido.

Agradeço também a todas as PESSOAS que de alguma forma direta ou indireta contribuíram para o meu crescimento profissional.

Enfim, a todos VOCÊS que iluminam diariamente minha vida, meus sinceros agradecimentos.

SUMÁRIO

RESUMO	7
ABSTRACT	9
LISTA DE FIGURAS	10
LISTA DE TABELAS	11
1 INTRODUÇÃO	13
2 DESENVOLVIMENTO	15
2.1 Análise de sobrevivência	15
2.2 Distribuição gama generalizada	19
2.3 Distribuições modificadas	20
2.4 Distribuição gama generalizada exponenciada	22
2.4.1 Relação com algumas distribuições	24
2.4.2 Formas da função de taxa de falha da GGE	26
3 DISTRIBUIÇÃO GAMA GENERALIZADA EXPONENCIADA ESTENDIDA	27
3.1 Relação com algumas distribuições	30
3.2 Propriedades gerais	32
3.2.1 Momentos	32
3.2.2 Função geratriz dos momentos	34
3.2.3 Desvios médios	37
3.2.4 Confiabilidade	38
3.2.5 Estatísticas de ordem	40
3.3 Estimação de máxima verossimilhança	41
3.4 Teste de hipóteses	42
4 MODELO LOG-GAMA GENERALIZADO EXPONENCIADO	43
4.1 Fórmula geral para os momentos do modelo LGGE	45
4.2 Procedimento inferenciais	47
4.2.1 Características do modelo LGGE sem censuras	47
4.2.2 Características do modelo LGGE com censuras	48
4.2.3 Características do modelo LGGE com censuras e covariáveis	51
4.3 Casos particulares do modelo de regressão LGGE	54

5	MODELO DE LOCAÇÃO E ESCALA PARA DADOS AGRUPADOS	57
5.1	Modelo de regressão com efeito aleatório	57
5.1.1	Verossimilhança aumentada	58
5.2	Modelo de regressão LGGE com efeito aleatório	59
6	APLICAÇÕES	63
6.1	Descrição dos dados	63
6.2	Análise dos dados de residentes dependentes químicos (modelo GGE)	64
6.3	Análise dos dados de níveis de voltagem (modelo de regressão LGGE)	65
6.4	Análise dos dados de animais da raça Nelore (modelo de regressão LGGE com efeito aleatório)	71
7	CONSIDERAÇÕES FINAIS	73
7.1	Conclusões	73
7.2	Pesquisas futuras	73
	REFERÊNCIAS	75
	APÊNDICES	83
	ANEXOS	91

RESUMO

Modelo de regressão log-gama generalizado exponenciado com dados censurados

No presente trabalho, é proposto um modelo de regressão utilizando a distribuição gama generalizada exponenciada (GGE) para dados censurados, esta nova distribuição é uma extensão da distribuição gama generalizada. A distribuição GGE (CORDEIRO et al., 2009) que tem quatro parâmetros pode modelar dados de sobrevivência quando a função de risco tem forma crescente, decrescente, forma de U e unimodal. Neste trabalho apresenta-se uma expansão natural da distribuição GGE para dados censurados, esta distribuição desperta o interesse pelo fato de representar uma família paramétrica que possui como casos particulares outras distribuições amplamente utilizadas na análise de dados de tempo de vida, como as distribuições gama generalizada (STACY, 1962), Weibull, Weibull exponenciada (MUDHOLKAR et al., 1995, 1996), exponencial exponenciada (GUPTA; KUNDU, 1999, 2001), Rayleigh generalizada (KUNDU; RAKAB, 2005), dentre outras, e mostra-se útil na discriminação entre alguns modelos probabilísticos alternativos. Considerando dados censurados, é abordado o método de máxima verossimilhança para estimar os parâmetros do modelo proposto. Outra proposta deste trabalho é introduzir um modelo de regressão log-gama generalizado exponenciado com efeito aleatório. Por fim, são apresentadas três aplicações para ilustrar a distribuição proposta.

Palavras-chave: Dados censurados; Distribuição exponencial exponenciada; Distribuição gama generalizada; Distribuição Weibull; Distribuição Weibull exponenciada; Função de sobrevivência; Máxima verossimilhança

ABSTRACT

The log-exponentiated generalized gamma regression model with censored data

In the present study, we propose a regression model using the exponentiated generalized gamma (EGG) distribution for censored data, this new distribution is an extension of the generalized gamma distribution. The EGG distribution (CORDEIRO et al., 2009) that has four parameters it can model survival data when the risk function is increasing, decreasing, form of U and unimodal-shaped. In this work comes to a natural expansion of the EGG distribution for censored data, is awake distribution the interest for the fact of representing a parametric family that has, as particular cases, other distributions which are broadly used in lifetime data analysis, as the generalized gamma (STACY, 1962), Weibull, exponentiated Weibull (MUDHOLKAR et al., 1995, 1996), exponentiated exponential (GUPTA; KUNDU, 1999, 2001), generalized Rayleigh (KUNDU; RAKAB, 2005), among others, and it is shown useful in the discrimination among some models alternative probabilistics. Considering censored data, the maximum likelihood estimator is considered for the proposed model parameters. Another proposal of this work was to introduce a log-exponentiated generalized gamma regression model with random effect. Finally, three applications were presented to illustrate the proposed distribution.

Keywords: Censored data; Exponentiated exponential distribution; Exponentiated Weibull distribution; Generalized gamma distribution; Maximum likelihood; Survival function; Weibull distribution

LISTA DE FIGURAS

Figura 1 - Gráficos ilustrativos de alguns TTT-plots	18
Figura 2 - Gráfico da função de densidade da distribuição GGE para $\tau = 2$, $k=3$ e $\alpha = 2$	22
Figura 3 - Gráfico da função de sobrevivência da distribuição GGE para $\tau = 2$, $k=3$ e $\alpha = 2$	23
Figura 4 - Função de taxa de falha da distribuição GGE (a) em forma de U e (b) na forma unimodal	27
Figura 5 - Função de taxa de falha da distribuição GGE nas formas crescente e decrescente	27
Figura 6 - Gráfico da função de densidade da distribuição GGE estendida para $\alpha = 1$, $k=1$, $\lambda = 1$ e (a) alguns valores de $\tau > 0$ e (b) alguns valores de $\tau < 0$. . .	28
Figura 7 - Assimetria (a) e curtose (b) da distribuição GGE como função do parâmetro λ , para $\tau = 0.08$, $k = 2$ e alguns valores de α	35
Figura 8 - Assimetria (a) e curtose (b) da distribuição GGE como função do parâmetro λ , para $\alpha = 0.5$, $k = 2$ e alguns valores de τ	35
Figura 9 - Gráfico da função de densidade da distribuição LGGE para (a) $\mu = 2$, $\sigma = 1.2$ e $q = -2$ e (b) $\mu = 2$, $\sigma = 1.2$ e $q = 3$	44
Figura 10 - Assimetria (a) e curtose (b) da distribuição LGGE como função do parâmetro λ , para $\sigma = 0.8$, $q = 2.5$ e alguns valores de μ	46
Figura 11 - Assimetria (a) e curtose (b) da distribuição LGGE como função do parâmetro λ , para $\mu = 0.001$, $q = 0.5$ e alguns valores de σ	46
Figura 12 - TTT-plot para tempo de sobrevivência dos dados de residentes dependentes químicos	64
Figura 13 - Funções de sobrevivência estimadas	66
Figura 14 - Funções de sobrevivência estimadas	66
Figura 15 - TTT-plot para tempo de sobrevivência dos dados de níveis de voltagem . .	67
Figura 16 - Funções de sobrevivência estimadas	68
Figura 17 - Funções de sobrevivência estimadas	69
Figura 18 - Gráfico do método de máxima verossimilhança perfilada	69
Figura 19 - Curvas de sobrevivência estimadas pelo modelo de regressão LGGE considerando três níveis de voltagem: 52.5 kV, 55.0 kV e 57.5 kV	70

LISTA DE TABELAS

Tabela 1 - Estimativas dos parâmetros do modelo para os dados de residentes dependentes químicos, seus correspondentes erros-padrão (entre parênteses) e as estatísticas AIC, AICC e BIC	65
Tabela 2 - Estimativas dos parâmetros do modelo para os dados de níveis de voltagem, seus correspondentes erros-padrão (entre parênteses) e as estatísticas AIC, AICC e BIC	68
Tabela 3 - Estimativas de máxima verossimilhança dos parâmetros do modelo LGGE para os dados de níveis de voltagem	70
Tabela 4 - Estimativas de máxima verossimilhança dos parâmetros do modelo LGGE com efeito aleatório para os dados de animais da raça Nelore	71

1 INTRODUÇÃO

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

Em análise de sobrevivência, a variável resposta é, em geral, o tempo até a ocorrência de um evento de interesse. Este tempo é denominado tempo de falha, podendo ser o tempo até a morte do paciente, bem como até a cura ou recidiva de uma doença.

A principal característica de dados de sobrevivência é a presença de censura, que é a observação parcial da resposta. Isso se refere a situações em que, por alguma razão, o acompanhamento do paciente foi interrompido, seja porque o paciente mudou de cidade, o estudo terminou para a análise dos dados ou, o paciente morreu de causa diferente da estudada. Isto significa que toda informação referente à resposta se resume ao conhecimento de que o tempo de falha é superior àquele observado.

Por outro lado, nos últimos anos tem crescido muito a generalização e/ou a modificação de algumas distribuições utilizadas em análise de sobrevivência. Existem diferentes formas de modificar uma distribuição de probabilidade, em particular, existe as distribuições exponenciadas. As mais conhecidas são a distribuição Weibull exponenciada (WE) proposta por Mudholkar et al. (1995, 1996) e a distribuição exponencial exponenciada (EE) introduzida por Gupta e Kundu (1999, 2001). Segundo Gupta e Kundu (2002), a EE pode ser utilizada com eficácia para análise de tempos de vida, principalmente, na presença de censuras ou dados agrupados. Outras distribuições exponencializadas foram propostas, a distribuição Gumbel exponenciada (NADARAJAH, 2006), a distribuição Fréchet exponenciada (NADARAJAH; KOTZ, 2006) e a distribuição Weibull modificada generalizada (CARRASCO et al., 2008).

Recentemente, Cordeiro et al. (2009) apresenta uma generalização da distribuição gama generalizada, a qual é denominada de gama generalizada exponenciada (GGE). Nela é apresentado como principais resultados os momentos, a função de densidade da i -ésima estatística de ordem, dentre outras propriedades. É utilizado o método da máxima verossimilhança para estimar os parâmetros do modelo proposto.

Assim, no presente trabalho o objetivo é propor uma extensão da distribuição gama generalizada exponenciada (GGE) para modelar dados de sobrevivência. Apresenta-se propriedades desta nova distribuição baseado no trabalho de Cordeiro et al. (2009). A função de densidade da nova distribuição é expressa como forma de mistura das funções de densidade da gama generalizada (GG). Ainda neste trabalho considera-se a nova distribuição denominada de log-gama generalizada exponenciada (LGGE). Esta distribuição LGGE possui como casos particulares várias distribuições como valor-extremo, log-Weibull exponenciado, log-Weibull inverso, dentre outras. Algumas propriedades desta nova distribuição são estudadas. Também, mostra-se que a distribuição LGGE pode ser escrita como um modelo de locação e escala, e isto é importante pois uma estrutura de regressão é considerada, a qual denomina-se modelo de regressão LGGE. Finalmente, em análise de sobrevivência está sendo muito utilizado os modelos de regressão com efeito aleatório. Nesse sentido é estendido o trabalho de Valença (2004) e Silva (2009) e considera-se um modelo de regressão LGGE com efeito aleatório.

O trabalho está organizado da seguinte maneira. Na Seção 2 apresenta-se uma revisão sobre alguns conceitos fundamentais em análise de sobrevivência, também é apresentado o gráfico TTT-plot proposto por Aarset (1987) que é utilizado para a determinação empírica da forma da função de taxa de falha. Nesta Seção ainda é apresentado alguns conceitos relacionados ao modelo gama generalizado proposto por Stacy (1962), e também é discutido algumas formas de como modificar uma distribuição e o modelo gama generalizado exponenciado (GGE) proposto por Cordeiro et al. (2009). A Seção 3 trata de uma extensão da GGE, é demonstrado ainda nesta Seção algumas propriedades deste modelo. Na Seção 4 é proposto uma transformação logarítmica e em seguida uma reparametrização no modelo GGE. Nesta Seção é apresentado os procedimentos inferenciais do modelo proposto. A Seção 5 trata do modelo para dados correlacionados. Na Seção 6 é apresentado a descrição dos dados utilizados, os resultados e as discussões obtidas no presente trabalho. Finalmente, na Seção 7 é apresentado algumas considerações finais.

2 DESENVOLVIMENTO

2.1 Análise de sobrevivência

Análise de sobrevivência é constituída por um conjunto de técnicas estatísticas para analisar dados de tempo até a ocorrência de um determinado evento de interesse, que, em geral, é denominado de tempo de sobrevivência.

O termo análise de sobrevivência refere-se basicamente a situações médicas envolvendo dados censurados (COLOSIMO; GIOLO, 2006). Entretanto, condições similares ocorrem em outras áreas, tais como engenharia, ciências sociais, dentre outras. Em engenharia, são comuns os estudos em que os produtos ou componentes são colocados sob teste para se estimar características relacionadas aos seus tempos de vida, tais como o tempo médio ou a probabilidade de certo produto durar mais do que 5 anos. Os engenheiros denominam esta área de confiabilidade. Criminalistas estudam o tempo entre a liberação de presos e a ocorrência de crimes; demógrafos, usam a análise de sobrevivência onde o interesse pode ser o nascimento ou a morte de um indivíduo, tempo que dura o casamento ou o tempo até o divórcio.

São comumente encontrados conjuntos de dados de tempo de falha em diversas áreas de pesquisa. Por falha, entenda-se como a ocorrência de um evento pré-especificado e, por tempo de falha, o período de tempo decorrido para o evento ocorrer.

Uma característica importante, em análise de sobrevivência, é a presença de observações incompletas ou parciais. Estas observações, denominadas censuras, podem ocorrer por uma variedade de razões, dentre elas, o término do experimento antes da ocorrência do evento; o paciente ter se mudado para uma localidade desconhecida; o paciente ter abandonado o tratamento; a não adaptação do indivíduo ao tratamento e a necessidade de retirá-lo do estudo. Podem ocorrer ainda outros casos, não citados.

Ressalta-se o fato de que, mesmo censurados, todos os resultados provenientes de um estudo devem ser usados na análise estatística. Duas razões justificam tal procedimento: (i) mesmo sendo incompletas, as observações censuradas fornecem informações sobre o tempo de vida de pacientes; (ii) a omissão das censuras no cálculo das estatísticas de interesse pode acarretar em conclusões erradas (COLOSIMO; GIOLO, 2006).

Dentre os tipos de censura destacam-se: censura à direita, à esquerda e intervalar. Censura à direita é a censura em que o tempo até a ocorrência do evento é maior que um tempo

determinado "C", censura à esquerda ocorre quando o tempo registrado é maior que o tempo de falha e censura intervalar é um tipo mais geral de censura que acontece, por exemplo, em estudos em que os pacientes são acompanhados em visitas periódicas e é conhecido somente que o evento de interesse ocorreu em certo intervalo de tempo, ou seja, $(L, U]$, em que L é o limite inferior, U é o limite superior e $L < U$. Lindsey et al. (1998) observam que quando $L = U$ temos tempos exatos de falha, quando $U = \infty$ acontece censura à direita e para $L = 0$ acontece censura à esquerda.

As censuras à direita podem ainda ser caracterizadas como: censura de tipo I, censura de tipo II e censura aleatória. Censura de tipo I é aquela em que o estudo será terminado após um período de tempo pré-estabelecido, Censura de tipo II é aquela em que o estudo será terminado após ter ocorrido o evento de interesse em um número pré-estabelecido de indivíduos e censura aleatória acontece quando um indivíduo em estudo é retirado no decorrer do estudo sem ter ocorrido à falha, ou também, por exemplo, se o indivíduo morrer por uma razão diferente da estudada.

Considere-se, agora, uma variável aleatória T contínua e positiva, que representa o tempo de sobrevivência ou tempo de falha e, que associada a ela, exista uma função de densidade de probabilidade definida como

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P\{T \in [t, t + \Delta t)\}}{\Delta t} = \frac{\partial F(t)}{\partial t},$$

em que $F(t) = P(T \leq t)$ é a função de distribuição acumulada da variável aleatória T .

A probabilidade de uma observação não falhar até um certo tempo t , ou seja, a probabilidade de uma observação sobreviver ao tempo t é representada pela função de sobrevivência dada por

$$S(t) = P(T \geq t) = \int_t^{\infty} f(x)dx = 1 - \int_0^t f(x)dx = 1 - F(t). \quad (1)$$

Na prática, quando não há observações censuradas, a função de sobrevivência é estimada pela proporção de indivíduos que sobreviveram além do instante t com relação ao total de indivíduos no estudo. No entanto, quando há observações censuradas, estimadores mais gerais devem ser considerados. A função de sobrevivência $S(t)$ além de ser não-crescente com o tempo, possui as seguintes propriedades: $\lim_{t \rightarrow 0} S(t) = 1$ e $\lim_{t \rightarrow \infty} S(t) = 0$.

A função de taxa de falha ou de risco é definida como a razão instantânea de morte ou falha no tempo t , dado que o indivíduo sobreviveu até o tempo t , definida como

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P\{T \in [t, t + \Delta t) | T > t\}}{\Delta t} \quad (2)$$

e também pode ser escrita como

$$h(t) = \frac{f(t)}{S(t)} \geq 0.$$

Uma função crescente indica que a taxa de falha do paciente aumenta com o transcorrer do tempo. Uma função constante indica que a taxa de falha não se altera com o passar do tempo. Uma função decrescente mostra que a taxa de falha diminui à medida que o tempo transcorre. Diferentes funções de sobrevivência podem ter formas semelhantes, enquanto as respectivas funções de taxa de falha podem diferir drasticamente.

Em algumas aplicações existem informações qualitativas e, muitas vezes, estruturais a respeito do fenômeno em questão, que pode ser utilizada na determinação empírica da forma da função de taxa de falha. Informações estruturais estão diretamente vinculadas ao conhecimento do pesquisador sobre o fenômeno, enquanto que informações qualitativas podem ser extraídas através de análise gráfica. Neste contexto, um gráfico conhecido como gráfico TTT-plot (tempo total em teste) é de grande utilidade. Este gráfico foi inicialmente proposto por Aarset (1987), o gráfico TTT-plot é obtido a partir de:

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

em que $r = 1, 2, \dots, n$ e $T_{i:n}$, $i = 1, 2, \dots, n$ são estatísticas de ordem da amostra (MUDHOLKAR; SRIVASTAVA; KOLLIA, 1996).

A Figura 1 ilustra as várias formas de uma função de taxa de falha. Se a função de taxa de falha tem forma constante, o gráfico apresenta uma linha diagonal (A), se a função de taxa de falha for decrescente então o gráfico TTT-plot deverá ser convexo (B), caso a função de taxa de falha seja crescente então o gráfico TTT-plot deverá ser côncavo (C). Se o gráfico apresentar uma curvatura primeiramente convexa e depois côncava a função de taxa de falha

tem forma de banheira (D), e se o gráfico apresentar uma curvatura primeiramente côncava e depois convexa a função de taxa de falha possui forma unimodal (E). Várias regiões côncavas e convexas direcionam para o ajuste de uma função de taxa de falha multinomial. Essas curvas provavelmente podem ser ajustadas através da distribuição de múltiplos riscos (LOUZADA NETO, 2000).

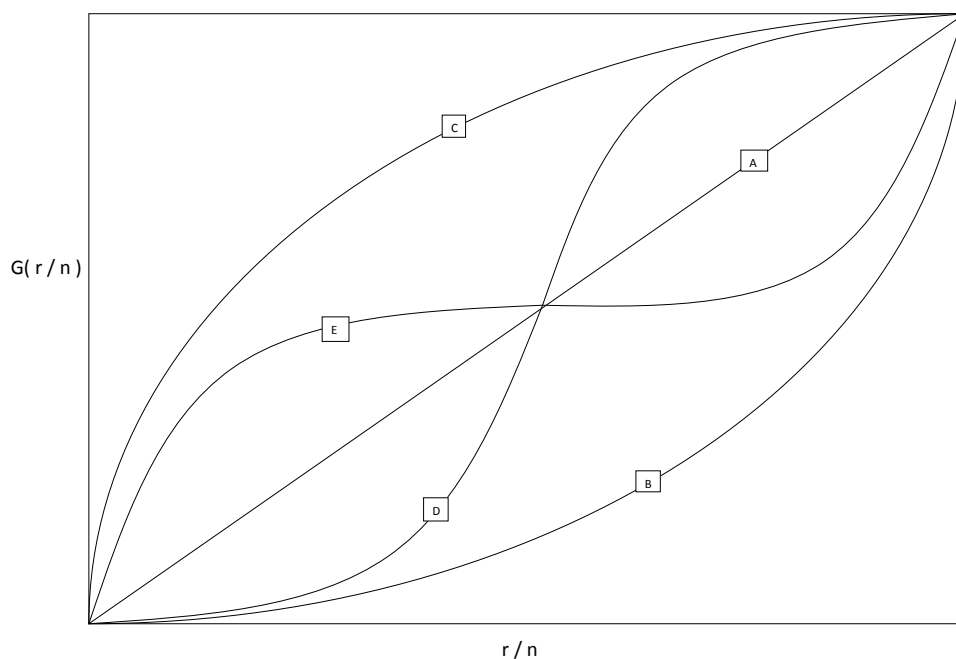


Figura 1 - Gráficos ilustrativos de alguns TTT-plots

Caso se tenha informações sobre covariáveis para cada indivíduo e uma quantidade significativa de indivíduos em cada nível ou combinação destas covariáveis, a curva TTT-plot pode ser construída considerando cada nível de covariável ou combinação das mesmas, separadamente.

2.2 Distribuição gama generalizada

A distribuição gama generalizada (GG) foi introduzida por Stacy (1962) e despertou o interesse de diversos pesquisadores pelo fato de representar uma família paramétrica que possui como casos particulares outras distribuições amplamente utilizadas na análise de dados de tempo de vida. Esta propriedade da distribuição GG faz com que a mesma seja de grande utilidade, como, por exemplo, na discriminação entre modelos probabilísticos alternativos.

Nos últimos anos diversos trabalhos envolvendo a distribuição GG foram propostos, dentre os quais destacam-se, por exemplo, Nadarajah e Gupta (2007) que usaram a distribuição com aplicações em dados de seca. Ali et al. (2008) derivou a distribuição exata do produto X_1X_2 e do quociente X_1/X_2 onde X_1 e X_2 são variáveis aleatórias independentes GG, provenientes dos resultados de dados de seca. Ortega et al. (2008) comparou três tipos de resíduos baseado no componente de deviance no modelo de regressão GG usando observações censuradas. Nadarajah (2008a) apresentou um estudo sobre o uso da distribuição GG em engenharia elétrica e eletrônica. Cox (2008) discutiu e comparou a família F-generalizada com o modelo GG. Malhotra et al. (2009) modelou o sinal *wireless* por meio da distribuição GG e analisou o desempenho do sinal. Cordeiro et al. (2009) propôs a distribuição gama generalizada exponenciada.

Seja T uma variável aleatória com distribuição GG proposta por Stacy (1962), cuja função de densidade, denotada por $g(t)$, é dada por

$$g(t) = \frac{\tau}{\alpha\Gamma(k)} \left(\frac{t}{\alpha}\right)^{\tau k-1} \exp\left\{-\left(\frac{t}{\alpha}\right)^\tau\right\}, \quad t > 0, \quad (3)$$

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

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

Se T é uma variável aleatória positiva com distribuição GG com parâmetros α , τ e k , então, denota-se que $T \sim \text{GG}(\alpha, \tau, k)$.

As propriedades da distribuição GG foram estudadas por várias pessoas, dentre estas, Stacy e Mihram (1965), Webster (1965), Prentice (1974), Farewell e Prentice (1977) e

Lawless (1980, 2003). A média e a variância da distribuição GG são dadas por

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

A distribuição GG é uma família paramétrica que possui alguns casos particulares. Quando $\tau = k = 1$ na equação (3) temos uma distribuição exponencial, quando $\tau = 1$ temos a distribuição gama. Para $k = 1$ temos a distribuição Weibull, para $\alpha = 1/\lambda$, $\tau = 2$ e $k = 1$ temos a distribuição Rayleigh, outros sub-modelos podem ser visto por exemplo em Lawless (2003).

As funções de distribuição acumulada $G(t)$, de sobrevivência $S(t)$ e de taxa de falha $h(t)$ são expressas, respectivamente, por

$$G(t) = P[T \leq t] = \frac{1}{\Gamma(k)} \int_0^{\left(\frac{t}{\alpha}\right)^\tau} w^{k-1} e^{-w} dw = \left\{ \gamma_1 \left[k, \left(\frac{t}{\alpha}\right)^\tau \right] \right\}, \quad (4)$$

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

$$h(t) = \frac{f(t)}{S(t)} = \frac{t^{\tau k - 1} e^{-\left(\frac{t}{\alpha}\right)^\tau}}{\int_t^\infty x^{\tau k - 1} e^{-\left(\frac{x}{\alpha}\right)^\tau} dx},$$

em que $\gamma_1(k, x) = \frac{1}{\Gamma(k)} \int_0^x u^{k-1} e^{-u} du$ é a função gama incompleta, que é facilmente implementada em vários pacotes estatísticos (R, SAS, Ox).

Algumas outras propriedades da distribuição GG podem ser encontradas em Lawless (2003).

2.3 Distribuições modificadas

Há diferentes formas de modificar uma distribuição de probabilidade. Em particular, existem as distribuições exponenciadas. Considere-se T uma variável aleatória positiva com função de distribuição acumulada (fda) conhecida $G(t)$, então a nova distribuição pode ser obtida exponenciando a função de distribuição acumulada $G(t)$, isto é:

$$F(t) = [G(t)]^\lambda, \quad (5)$$

em que $\lambda > 0$ é um novo parâmetro que caracteriza a forma da nova distribuição.

Nos últimos anos diversos trabalhos envolvendo distribuições modificadas foram propostos, dentre os quais destaca-se, por exemplo, a distribuição Weibull exponenciada (WE) proposta por Mudholkar et al. (1995, 1996), esta distribuição possui várias propriedades, dentre elas, destaca-se que a sua função de taxa de falha pode modelar formas diferentes (constante, crescente, decrescente, unimodal e forma de U). Mudholkar et al. (1996) analisa e discute diferentes aplicações a dados de sobrevivência. A distribuição WE também é utilizada para discriminar modelos, pois possui como casos particulares algumas distribuições conhecidas. A distribuição exponencial exponenciada (EE) proposta por Gupta e Kundu (1999, 2001) e, segundo Gupta e Kundu (2002), uma das vantagens dessa distribuição é que devido a estrutura simples de suas funções de distribuição e sobrevivência, a EE pode ser usada de forma eficaz na análise de dados de tempo de vida, particularmente, na presença de observações censuradas ou dados correlacionados. O método da máxima verossimilhança para obtenção das estimativas dos parâmetros, o método dos momentos, mínimos quadrados e estimadores L-momentos estão descritos em Gupta e Kundu (2007). A distribuição Rayleigh generalizada proposta por Kundu e Rakab (2005), que pode ser utilizada em modelagem de dados de tempo de vida em geral.

Destacam-se ainda, a distribuição Gumbel exponenciada proposta por Nadarajah (2006), esta distribuição é uma generalização da distribuição Gumbel, em que algumas de suas aplicações estão baseadas em modelagem climática (aquecimento global, velocidade do vento e modelagem da chuva). A distribuição Fréchet exponenciada proposta por Nadarajah e Kotz (2006), que é uma generalização da distribuição de Fréchet. A distribuição Weibull modificada generalizada proposta por CARRASCO et al. (2008), essa nova distribuição modela dados onde a função de taxa de falha tem formas crescente, decrescente, forma de U e unimodal, e é um modelo que caracteriza a modelos mais simples como a Weibull, Weibull exponenciada, exponencial exponenciada, entre outros. A distribuição gama exponencializada, definida por Nadarajah e Kotz (2006), que é uma generalização da distribuição gama padrão. A distribuição beta Gumbel proposta por Nadarajah e Kotz (2004), que é uma generalização

da distribuição Gumbel e a mesma pode ser aplicada em diversas áreas da engenharia. A distribuição beta Fréchet proposta por Nadarajah e Gupta (2004). Outras formas de representação da função de distribuição acumulada da beta Fréchet foram obtidas por Souza et al. (2008).

2.4 Distribuição gama generalizada exponenciada

Seja $G(t)$ a função de distribuição acumulada da distribuição GG definida na equação (4). A função de distribuição acumulada da distribuição GGE pode ser definida elevando $G(t)$ a potência λ , isto é, $F(t) = [G(t)]^\lambda$. Conseqüentemente, a função de densidade dessa nova distribuição com quatro parâmetros, $\alpha > 0$, $k > 0$, $\tau > 0$ e $\lambda > 0$ é dada por

$$f(t) = \frac{\lambda\tau}{\alpha\Gamma(k)} \left(\frac{t}{\alpha}\right)^{\tau k-1} \exp\left\{-\left(\frac{t}{\alpha}\right)^\tau\right\} \left\{\gamma_1\left[k, \left(\frac{t}{\alpha}\right)^\tau\right]\right\}^{\lambda-1}, \quad t > 0, \quad (6)$$

esta nova distribuição foi proposta por Cordeiro et al. (2009). A Figura 2 mostra o gráfico da função de densidade da distribuição GGE (CORDEIRO et al., 2009) para $\tau = 2$, $k=3$, $\alpha = 2$ e alguns valores de λ , onde podemos observar a flexibilidade para diferentes valores de λ .

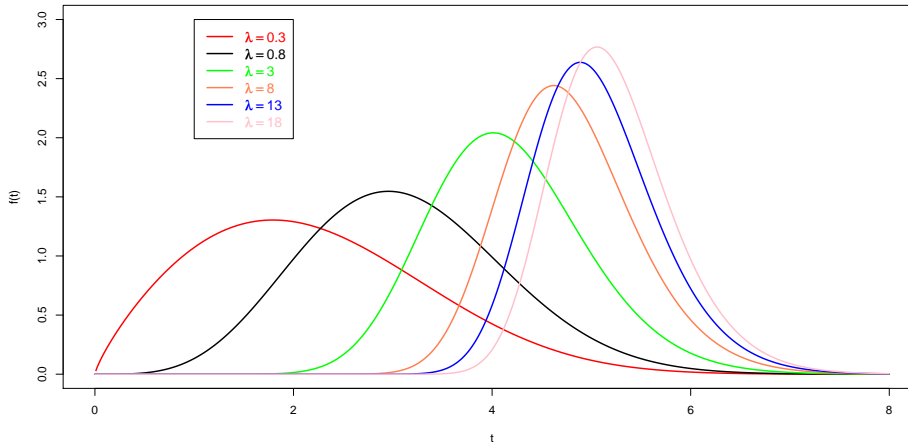


Figura 2 - Gráfico da função de densidade da distribuição GGE para $\tau = 2$, $k=3$ e $\alpha = 2$

Pode-se provar facilmente que a expressão (6) é um função de densidade considerando a substituição $u = \frac{1}{\Gamma(k)} \int_0^x w^{k-1} e^{-w} dw$. Se T é uma variável aleatória positiva com

densidade dada na equação (6), pode-se escrever, então, que $T \sim \text{GGE}(\alpha, \tau, k, \lambda)$.

A função de distribuição acumulada $F(t)$, a função de sobrevivência $S(t)$ e a função de taxa de falha $h(t)$ correspondentes, respectivamente, são:

$$F(t) = \int_0^t f(t)dt = \left\{ \frac{1}{\Gamma(k)} \int_0^{\left(\frac{t}{\alpha}\right)^\beta} w^{k-1} e^{-w} dw \right\}^\lambda = \left\{ \gamma_1 \left[k, \left(\frac{t}{\alpha}\right)^\beta \right] \right\}^\lambda, \quad (7)$$

$$S(t) = 1 - F(t) = 1 - \left\{ \gamma_1 \left[k, \left(\frac{t}{\alpha}\right)^\tau \right] \right\}^\lambda e \quad (8)$$

$$h(t) = \frac{\lambda\tau}{\alpha\Gamma(k)} \left(\frac{t}{\alpha}\right)^{\tau k-1} \exp \left\{ - \left(\frac{t}{\alpha}\right)^\tau \right\} \left\{ \gamma_1 \left[k, \left(\frac{t}{\alpha}\right)^\tau \right] \right\}^{\lambda-1} \left\{ 1 - \left\{ \gamma_1 \left[k, \left(\frac{t}{\alpha}\right)^\tau \right] \right\}^\lambda \right\}^{-1} \quad (9)$$

A Figura 3 mostra o gráfico da função de sobrevivência da distribuição GGE (CORDEIRO et al., 2009) para $\tau = 2$, $k=3$, $\alpha = 2$ e alguns valores de λ .

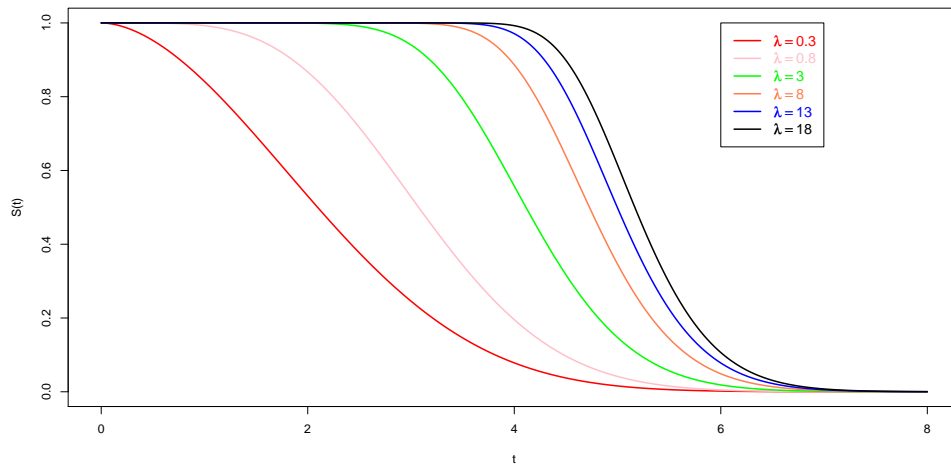


Figura 3 - Gráfico da função de sobrevivência da distribuição GGE para $\tau = 2$, $k=3$ e $\alpha = 2$

Se $T \sim \text{GGE}(\alpha, \tau, k, \lambda)$, Cordeiro et al. (2009) mostra que $S(t)$ pode ser expressa como:

$$S(t) = 1 - \left\{ \frac{\tau}{\alpha^{\tau k} \Gamma(k)} \sum_{j=0}^{\infty} \left[\frac{(-1)^j (\alpha^{-\tau})^j}{j!} \int_0^t x^{\tau k + \tau j - 1} dx \right] \right\}^\lambda. \quad (10)$$

Uma propriedade importante da distribuição GGE (CORDEIRO et al., 2009) é que, se $T \sim \text{GGE}(\alpha, \tau, k, \lambda) \Rightarrow T^m \sim \text{GGE}(\alpha^m, \tau/m, k, \lambda)$, qualquer $m \neq 0$.

2.4.1 Relação com algumas distribuições

A distribuição GGE apresenta algumas distribuições como casos particulares que serão apresentadas a seguir

- Para $\lambda = 1$ na equação (6) a distribuição GGE se reduz a uma distribuição gama generalizada, com função de densidade da forma:

$$f(t) = \frac{\tau}{\alpha\Gamma(k)} \left(\frac{t}{\alpha}\right)^{\tau k-1} \exp\left\{-\left(\frac{t}{\alpha}\right)^\tau\right\}, t > 0.$$

- Para $\tau = 1$ e $k = 1$ na equação (6) a distribuição GGE se reduz a uma distribuição exponencial exponenciada (GUPTA; KUNDU, 1999, 2001), com função de densidade da forma:

$$f(t) = \frac{\lambda}{\alpha} \exp\left\{-\left(\frac{t}{\alpha}\right)\right\} \left\{1 - \exp\left[-\left(\frac{t}{\alpha}\right)\right]\right\}^{\lambda-1}, t > 0.$$

- Para $k = 1$ na equação (6) a distribuição GGE se reduz a uma distribuição Weibull exponenciada (MUDHOLKAR et al., 1995, 1996), com função de densidade da forma:

$$f(t) = \frac{\lambda\tau}{\alpha} \left(\frac{t}{\alpha}\right)^{\tau-1} \exp\left\{-\left(\frac{t}{\alpha}\right)^\tau\right\} \left\{1 - \exp\left[-\left(\frac{t}{\alpha}\right)^\tau\right]\right\}^{\lambda-1}, t > 0.$$

- Para $k = 1$, $\alpha = \frac{1}{\lambda}$ e $\tau = 2$ na equação (6) a distribuição GGE se reduz a uma distribuição Rayleigh generalizada (KUNDU; RAKAB, 2005), com função de densidade da forma:

$$f(t) = 2t\lambda^3 \exp\left\{-\left(t\lambda\right)^2\right\} \left\{1 - \exp\left[-\left(t\lambda\right)^2\right]\right\}^{\lambda-1}, t > 0.$$

- Para $\lambda = 1$ e $k = 1$ na equação (6) a distribuição GGE se reduz a uma distribuição Weibull, com função de densidade da forma:

$$f(t) = \frac{\tau}{\alpha} \left(\frac{t}{\alpha}\right)^{\tau-1} \exp\left\{-\left(\frac{t}{\alpha}\right)^\tau\right\}, t > 0.$$

• Para $\lambda = 1$, $k = 1$ e $\tau = 1$ na equação (6) a distribuição GGE se reduz a uma distribuição exponencial, com função de densidade da forma:

$$f(t) = \frac{1}{\alpha} \exp\left\{-\left(\frac{t}{\alpha}\right)\right\}, t > 0.$$

• Para $\lambda = 1$ e $\tau = 1$ na equação (6) a distribuição GGE se reduz a uma distribuição gama, com função de densidade da forma:

$$f(t) = \frac{1}{\alpha\Gamma(k)} \left(\frac{t}{\alpha}\right)^{k-1} \exp\left\{-\left(\frac{t}{\alpha}\right)\right\}, t > 0.$$

• Para $\lambda = 1$, $\alpha = 2$, $\tau = 1$ e $k = \frac{v}{2}$ na equação (6) a distribuição GGE se reduz a uma distribuição qui-quadrada com v g.l., com função de densidade da forma:

$$f(t) = \frac{1}{\Gamma(\frac{v}{2})} \frac{1}{2^{\frac{v}{2}}} t^{\frac{v}{2}-1} \exp\left\{-\left(\frac{t}{2}\right)\right\}, t > 0.$$

• Para $\lambda = 1$, $\alpha = \sqrt{2}$, $\tau = \frac{1}{2}$ e $k = 2$ na equação (6) a distribuição GGE se reduz a uma distribuição meio normal, com função de densidade da forma:

$$f(t) = \frac{1}{2\sqrt{2}} \exp\left\{-\left(\frac{t}{\sqrt{2}}\right)^{\frac{1}{2}}\right\}, t > 0.$$

• Para $\lambda = 1$, $\alpha = \sqrt{2}$, $\tau = 1$ e $k = 2$ na equação (6) a distribuição GGE se reduz a uma distribuição normal circular, com função de densidade da forma:

$$f(t) = \frac{t}{2} \exp\left\{-\left(\frac{t}{\sqrt{2}}\right)\right\}, t > 0.$$

• Para $\lambda = 1$, $\alpha = \sqrt{2}$, $\tau = \frac{3}{2}$ e $k = 2$ na equação (6) a distribuição GGE se reduz a uma distribuição normal esférica, com função de densidade da forma:

$$f(t) = \frac{3t^2}{4\sqrt{2}} \exp\left\{-\left(\frac{t}{\sqrt{2}}\right)^{\frac{3}{2}}\right\}, t > 0.$$

• Para $k \rightarrow \infty$ na equação (6) a distribuição GGE se reduz a uma distribuição log-normal, com função de densidade da forma:

$$f(t) = \frac{t}{\sqrt{2}} \exp \left\{ - \left(\frac{t}{\sqrt{2}} \right) \right\}, \quad t > 0.$$

2.4.2 Formas da função de taxa de falha da GGE

Uma característica da distribuição GGE é que a sua função de taxa de falha acomoda formas de risco crescente, decrescente, forma de U e unimodal. Estas formas dependem basicamente dos valores dos parâmetros τ , k e λ . O estudo da forma da função de taxa de falha envolve uma análise da primeira derivada $h'(t) = \partial h(t)/\partial t$ dada por

$$h'(t) = h(t) \left\{ (\tau k - 1)t^{-1} - \frac{\tau}{\alpha} \left(\frac{t}{\alpha} \right)^{\tau-1} + (\lambda - 1) \frac{g(t)}{\gamma_1 \left[k, \left(\frac{t}{\alpha} \right)^\tau \right]} + h(t) \right\}, \quad (11)$$

onde $g(t)$ denota a função de densidade da distribuição GG com parâmetros (α, τ, k) . Ver mais detalhes em Cordeiro et al. (2009). Pode-se considerar que: quando $\tau k > 1$, $\lambda > 1$ e $\forall t > 0$, $h'(t) > 0$ e a função de risco tem forma crescente; quando $0 < \tau k < 1$, $0 < \lambda < 1$ e $\forall t > 0$, $h'(t) < 0$ e a função de risco tem forma decrescente; quando $\tau > 1$ e $\tau \lambda < 1$, a função de risco tem forma de U (ou banheira); quando $\tau < 1$ e $\tau \lambda \geq 1$, a função de risco tem forma unimodal.

Algumas formas da função de taxa de falha para alguns valores dos parâmetros da distribuição GGE são dadas nas Figuras 4 e 5. Na Figura 4 encontram-se a função de taxa de falha em forma de U (a) e em forma unimodal (b). Na Figura 5 encontram-se as formas crescente e decrescente.

Outras propriedades da distribuição GGE, assim como aplicações, podem ser encontradas em Cordeiro et al. (2009).

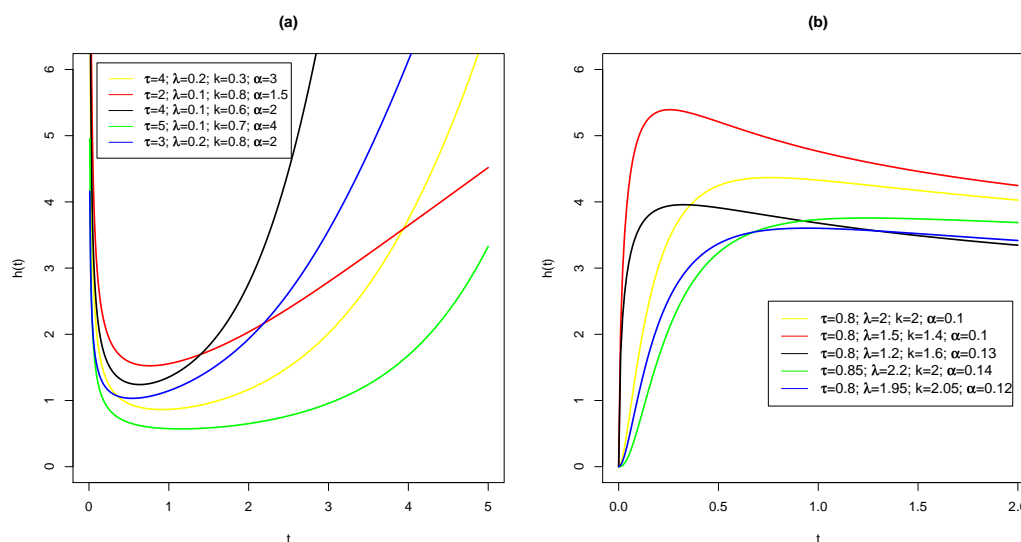


Figura 4 - Função de taxa de falha da distribuição GGE (a) em forma de U e (b) na forma unimodal

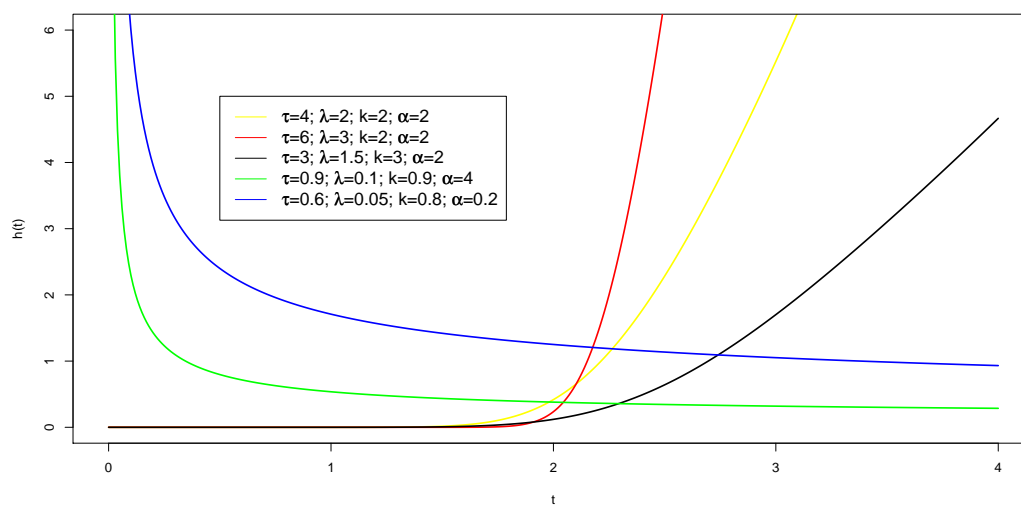


Figura 5 - Função de taxa de falha da distribuição GGE nas formas crescente e decrescente

3 DISTRIBUIÇÃO GAMA GENERALIZADA EXPONENCIADA ESTENDIDA

Seguindo a idéia de Stacy e Mihram (1965) onde apresentaram uma reparametrização incluindo casos onde o parâmetro τ pode ser considerado negativo, neste

trabalho é considerado uma expansão natural de (6), considerando que τ difere de zero, e os demais parâmetros positivos. A função de densidade neste caso é dada por

$$f(t) = \frac{\lambda|\tau|}{\alpha\Gamma(k)} \left(\frac{t}{\alpha}\right)^{\tau k-1} \exp\left\{-\left(\frac{t}{\alpha}\right)^\tau\right\} \left\{\gamma_1\left[k, \left(\frac{t}{\alpha}\right)^\tau\right]\right\}^{\lambda-1}, \quad t > 0, \quad (12)$$

em que $\tau \neq 0$, $\alpha > 0$, $k > 0$ e $\lambda > 0$.

A Figura 6 mostra o gráfico da função de densidade da distribuição GGE estendida para $\alpha = 1$, $k = 1$, $\lambda = 1$ e alguns valores de $\tau > 0$ (a) e $\tau < 0$ (b), observando-se a flexibilidade da função de densidade para valores de τ .

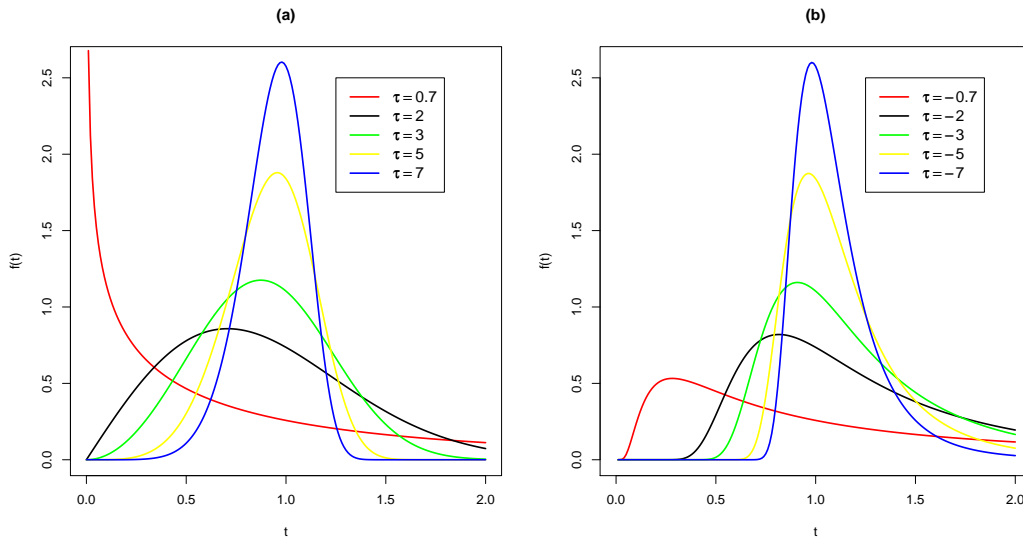


Figura 6 - Gráfico da função de densidade da distribuição GGE estendida para $\alpha = 1$, $k=1$, $\lambda = 1$ e (a) alguns valores de $\tau > 0$ e (b) alguns valores de $\tau < 0$

Note que se $\tau > 0$ temos a densidade dada na equação (6). Neste trabalho chamaremos esta nova distribuição como sendo gama generalizada exponenciada (GGE), pois é uma generalização natural da distribuição GGE (CORDEIRO et al., 2009). Portanto, a função de distribuição acumulada $F(t)$ é dada por

$$F(t) = \begin{cases} \left\{\gamma_1\left[k, \left(\frac{t}{\alpha}\right)^\tau\right]\right\}^\lambda & \text{se } \tau > 0, \\ 1 - \left\{\gamma_1\left[k, \left(\frac{t}{\alpha}\right)^\tau\right]\right\}^\lambda & \text{se } \tau < 0. \end{cases} \quad (13)$$

Conseqüentemente, a função de sobrevivência $S(t)$ é dada por

$$S(t) = \begin{cases} 1 - \left\{ \gamma_1 \left[k, \left(\frac{t}{\alpha} \right)^\tau \right] \right\}^\lambda & \text{se } \tau > 0, \\ \left\{ \gamma_1 \left[k, \left(\frac{t}{\alpha} \right)^\tau \right] \right\}^\lambda & \text{se } \tau < 0. \end{cases} \quad (14)$$

Teorema 1. Se $T \sim \text{GGE}(\alpha, \tau, k, \lambda)$ com função de densidade dada na equação (12), então, a função de densidade pode ser escrita como uma mistura de densidades da distribuição GG, isto é,

$$f(t) = \sum_{m,i=0}^{\infty} w(m, i, k, \lambda) g_{\alpha, \tau, k(m+1)+i}(t), \quad t > 0. \quad (15)$$

em que $g_{\alpha, \tau, k(m+1)+i}(t)$ tem distribuição $\text{GG}(\alpha, \tau, k(m+1)+i)$ e os coeficientes de ponderação $w(m, i, k, \lambda)$ são dados por

$$w(m, i, k, \lambda) = \sum_{m,i=0}^{\infty} \frac{\lambda s_m(\lambda) c_{m,i} \Gamma[k(m+1)+i]}{\Gamma(k)^{m+1}}.$$

onde $s_m(\lambda)$ é dada por

$$s_m(\lambda) = \sum_{j=m}^{\infty} (-1)^{j+m} \binom{\lambda-1}{j} \binom{j}{m}, \quad (16)$$

e os coeficientes $c_{m,i}$ (para $i = 1, 2, \dots$) satisfazem a relação de recorrência

$$c_{m,i} = (i a_0)^{-1} \sum_{p=1}^i (m p - i + p) a_p c_{m,i-p} \quad (17)$$

onde $c_{m,0} = a_0^m$ e $a_p = (-1)^p / (k+p)p!$. O coeficiente $c_{m,i}$ pode ser calculado a partir de $c_{m,0}, \dots, c_{m,i-1}$, e pode ser escrito explicitamente em função das quantidades a_0, \dots, a_i .

Prova:

Seja $g(t)$ a função de densidade estendida proposta por Stacy e Mihram (1965) da distribuição $\text{GG}(\alpha, \tau, k)$, isto é

$$g(t) = \frac{|\tau|}{\alpha \Gamma(k)} \left(\frac{t}{\alpha} \right)^{\tau k - 1} \exp \left[- \left(\frac{t}{\alpha} \right)^\tau \right], \quad t > 0. \quad (18)$$

A partir daqui iremos trabalhar com esta distribuição que é uma extensão natural da distribuição GG proposta por Stacy (1962) definida na equação (3).

A função $[\gamma_1(k, x)]^{\lambda-1}$ pode ser escrita como

$$[\gamma_1(k, x)]^{\lambda-1} = \sum_{m,i=0}^{\infty} \frac{s_m(\lambda)c_{m,i}}{\Gamma(k)^m} x^{km+i}, \quad (19)$$

em que $s_m(\lambda)$ e $c_{m,i}$ são definidos em (16) e (17), respectivamente. Utilizando esta aproximação em (12), a função de densidade da distribuição $GGE(\alpha, \tau, k, \lambda)$ pode ser escrita como

$$f(t) = \frac{\lambda|\tau|}{\alpha\Gamma(k)} \exp\left[-\left(\frac{t}{\alpha}\right)^\tau\right] \sum_{m,i=0}^{\infty} \frac{s_m(\lambda)c_{m,i}}{\Gamma(k)^m} \left(\frac{t}{\alpha}\right)^{\tau[k(m+1)+i]-1}, \quad t > 0.$$

Imediatamente obtemos uma densidade de mistura

$$f(t) = \sum_{m,i=0}^{\infty} w(m, i, k, \lambda) g_{\alpha, \tau, k(m+1)+i}(t), \quad t > 0, \quad (20)$$

em que $g_{\alpha, \tau, k(m+1)+i}(t)$ tem distribuição $GG(\alpha, \tau, k(m+1)+i)$ e os coeficientes de ponderação $w(m, i, k, \lambda)$ são dados por

$$w(m, i, k, \lambda) = \sum_{m,i=0}^{\infty} \frac{\lambda s_m(\lambda) c_{m,i} \Gamma[k(m+1)+i]}{\Gamma(k)^{m+1}}.$$

em que $s_m(\lambda)$ é dada por

$$s_m(\lambda) = \sum_{j=m}^{\infty} (-1)^{j+m} \binom{\lambda-1}{j} \binom{j}{m}, \quad (21)$$

e os coeficientes $c_{m,i}$ (para $i = 1, 2, \dots$) satisfazem a relação de recorrência

$$c_{m,i} = (ia_0)^{-1} \sum_{p=1}^i (mp - i + p) a_p c_{m,i-p} \quad (22)$$

onde $c_{m,0} = a_0^m$ e $a_p = (-1)^p / (k+p)p!$. O coeficiente $c_{m,i}$ pode ser calculado a partir de $c_{m,0}, \dots, c_{m,i-1}$, e pode ser escrito explicitamente em função das quantidades a_0, \dots, a_i .

3.1 Relação com algumas distribuições

Além de ter casos particulares descritos na Seção 2.4.1, a distribuição GGE apresenta algumas distribuições como casos particulares quando $\tau < 0$ que serão apresentadas a seguir

• Para $\lambda = 1$ e $\tau = -1$ na equação (12) a distribuição GGE se reduz a uma distribuição gama inversa, com função de densidade da forma:

$$f(t) = \frac{1}{\alpha\Gamma(k)} \left(\frac{\alpha}{t}\right)^{k+1} \exp\left\{-\left(\frac{\alpha}{t}\right)\right\}, t > 0.$$

• Para $\lambda = 1$, $k = 1$ e $\tau = -1$ na equação (12) a distribuição GGE se reduz a uma distribuição exponencial inversa, com função de densidade da forma:

$$f(t) = \left(\frac{\alpha}{t^2}\right) \exp\left\{-\left(\frac{\alpha}{t}\right)\right\}, t > 0.$$

• Para $\lambda = 1$, $\alpha = \frac{1}{\sqrt{2s^2}}$, $k = 1$ e $\tau = -2$ na equação (12) a distribuição GGE se reduz a uma distribuição Rayleigh inversa, com função de densidade da forma:

$$f(t) = \left(\frac{1}{s^2}\right) \left(\frac{1}{t}\right)^3 \exp\left\{-\left(\frac{1}{2s^2t^2}\right)\right\}, t > 0.$$

• Para $\lambda = 1$, $\alpha = \frac{1}{2}$, $\tau = -1$ e $k = \frac{v}{2}$ na equação (12) a distribuição GGE se reduz a uma distribuição qui-quadrada inversa com v g.l., com função de densidade da forma:

$$f(t) = \frac{2}{\Gamma(\frac{v}{2})} \left(\frac{1}{2t}\right)^{\frac{v}{2}+1} \exp\left\{-\left(\frac{1}{2t}\right)\right\}, t > 0.$$

• Para $\lambda = 1$, $\alpha = \frac{1}{\sqrt{2}}$, $\tau = -2$ e $k = \frac{v}{2}$ na equação (12) a distribuição GGE se reduz a uma distribuição qui inversa com v g.l., com função de densidade da forma:

$$f(t) = \frac{2\sqrt{2}}{\Gamma(\frac{v}{2})} \left(\frac{1}{t\sqrt{2}}\right)^{v+1} \exp\left\{-\left(\frac{1}{2t^2}\right)\right\}, t > 0.$$

• Para $\lambda = 1$, $\alpha = \frac{c}{2}$, $\tau = -1$ e $k = \frac{1}{2}$ na equação (12) a distribuição GGE se reduz a uma distribuição Lévy, com função de densidade da forma:

$$f(t) = \left(\frac{c}{2\pi}\right)^{\frac{1}{2}} \left(\frac{1}{t^{3/2}}\right) \exp\left\{-\left(\frac{c}{2t}\right)\right\}, t > 0.$$

3.2 Propriedades gerais

3.2.1 Momentos

Precisa-se enfatizar a necessidade e a importância dos momentos em qualquer análise estatística. Algumas das características mais importantes de uma distribuição podem ser estudadas pelos momentos (por exemplo, tendência, dispersão, assimetria e curtose).

Teorema 2. Seja $T \sim \text{GGE}(\alpha, \tau, k, \lambda)$ com função de densidade dada na equação (12), então o r -ésimo momento da distribuição GGE é dado por

$$\begin{aligned} \mu'_r &= \frac{\lambda \alpha^r \text{sinal}(\tau)}{\Gamma(k)} \sum_{j=0}^{\infty} \sum_{m=0}^j (-1)^{j+m} \binom{\lambda-1}{j} \binom{j}{m} k^{-m} \Gamma(r/\tau + k(m+1)) \times \\ &F_A^{(m)}(r/\tau + k(m+1); k, \dots, k; k+1, \dots, k+1; -1, \dots, -1). \end{aligned} \quad (23)$$

em que $F_A^{(m)}$ é a função de Lauricella do tipo A (EXTON, 1978; AARTS, 2000) e é definida por

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

onde $(a)_i$ é um fatorial ascendente definido por

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

em que $(a)_0 = 1$.

Prova:

Pela definição de momentos temos que o r -ésimo momento $E(T^r)$ pode ser escrito como

$$\mu'_r = \int_0^{+\infty} t^r \frac{\lambda |\tau|}{\alpha \Gamma(k)} \left(\frac{t}{\alpha}\right)^{\tau k-1} \exp\left\{-\left(\frac{t}{\alpha}\right)^{\tau}\right\} \left\{\gamma_1\left[k, \left(\frac{t}{\alpha}\right)^{\tau}\right]\right\}^{\lambda-1} dt, \quad (24)$$

a qual pode ser escrita como

$$\mu'_r = \frac{\lambda |\tau| \alpha^{r-1}}{\Gamma(k)} \int_0^{+\infty} \left(\frac{t}{\alpha}\right)^{\tau k+r-1} \exp\left\{-\left(\frac{t}{\alpha}\right)^{\tau}\right\} \left\{\gamma_1\left[k, \left(\frac{t}{\alpha}\right)^{\tau}\right]\right\}^{\lambda-1} dt. \quad (25)$$

Fazendo a mudança de variável, $x = \left(\frac{t}{\alpha}\right)^\tau$ tem-se que

$$\mu'_r = \frac{\lambda \alpha^r \text{sinal}(\tau)}{\Gamma(k)} \int_0^{+\infty} x^{k+\frac{r}{\tau}-1} \exp(-x) [\gamma_1(k, x)]^{\lambda-1} dx, \quad (26)$$

em que $\text{sinal}(\cdot)$ é a função sinal. A função sinal é a função que retorna o sinal do número real, ou seja, $\text{sinal}(\tau) = -1$, se $\tau < 0$ ou $\text{sinal}(\tau) = 1$, se $\tau > 0$.

Uma expansão para $[\gamma_1(k, x)]^{\lambda-1}$ para todo $\lambda > 0$, não-inteiro, pode ser escrita como

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

que sempre converge quando $0 < \gamma_1(k, x) < 1$. Então,

$$[\gamma_1(k, x)]^{\lambda-1} = \sum_{j=0}^{\infty} \sum_{m=0}^j (-1)^{j+m} \binom{\lambda-1}{j} \binom{j}{m} \gamma_1(k, x)^m. \quad (27)$$

Substituindo está última aproximação na equação (26) temos que

$$\mu'_r = \frac{\lambda \alpha^r \text{sinal}(\tau)}{\Gamma(k)} \sum_{j=0}^{\infty} \sum_{m=0}^j (-1)^{j+m} \binom{\lambda-1}{j} \binom{j}{m} I\left(k, \frac{r}{\tau}, m\right) \quad (28)$$

em que

$$I\left(k, \frac{r}{\tau}, m\right) = \int_0^{+\infty} x^{k+\frac{r}{\tau}-1} \exp\{-x\} \gamma_1(k, x)^m dx. \quad (29)$$

Utilizando uma expansão em série para a função gama incompleta $\gamma_1(k, x)$ dada por

$$\gamma_1(k, x) = \frac{x^k}{\Gamma(k)} \sum_{i=0}^{\infty} \frac{(-x)^i}{(k+i)i!}, \quad (30)$$

temos que a equação (29) pode ser escrita como

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

Esta integral é determinada para as equações (24) e (25) de Nadarajah (2008b) em termos da função de Lauricella do tipo A (EXTON, 1978; AARTS, 2000) definida por

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

onde $(a)_i$ é um fatorial ascendente definido por

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

em que $(a)_0 = 1$.

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

$$I \left[k, \frac{r}{\tau}, m \right] = k^{-m} \Gamma(r/\tau + k(m+1)) \times F_A^{(m)}(r/\tau + k(m+1); k, \dots, k; k+1, \dots, k+1; -1, \dots, -1). \quad (31)$$

Representações gráficas da assimetria e curtose da distribuição GGE como função do parâmetro λ , para alguns valores escolhidos do parâmetro α e para alguns valores escolhidos do parâmetro τ , são dadas nas Figuras 7 e 8, respectivamente. Na Figura 7, tem-se o gráfico da assimetria (a) e curtose (b) para $\tau = 0.08$, $k = 2$ e alguns valores de α . Na Figura 8, tem-se o gráfico da assimetria (a) e curtose (b) para $\alpha = 0.5$, $k = 2$ e alguns valores de τ . A Figura 7 mostra que a assimetria e a curtose aumentam quando λ aumenta e α diminui. A Figura 8 mostra que a assimetria e a curtose aumentam quando λ aumenta e τ diminui.

3.2.2 Função geratriz dos momentos

Suponha que T é uma variável aleatória tendo função de densidade $GG(\alpha, \tau, k)$, fornecemos uma expressão na forma fechada para a função geratriz dos momentos (fgm), onde $M_{\alpha, \tau, K}(s) = E[\exp(sT)]$, usando a função Wright (WRIGHT, 1935).

Teorema 3. Se $T \sim GGE(\alpha, \tau, k, \lambda)$ com função de densidade dada na equação (12), então a função geratriz dos momentos para a distribuição GGE pode ser escrita na forma de mistura (20) como

$$M(t) = \sum_{m, i=0}^{\infty} w(m, i, k, \lambda) M_{\alpha, \tau, k(m+1)+i}(s), \quad (32)$$

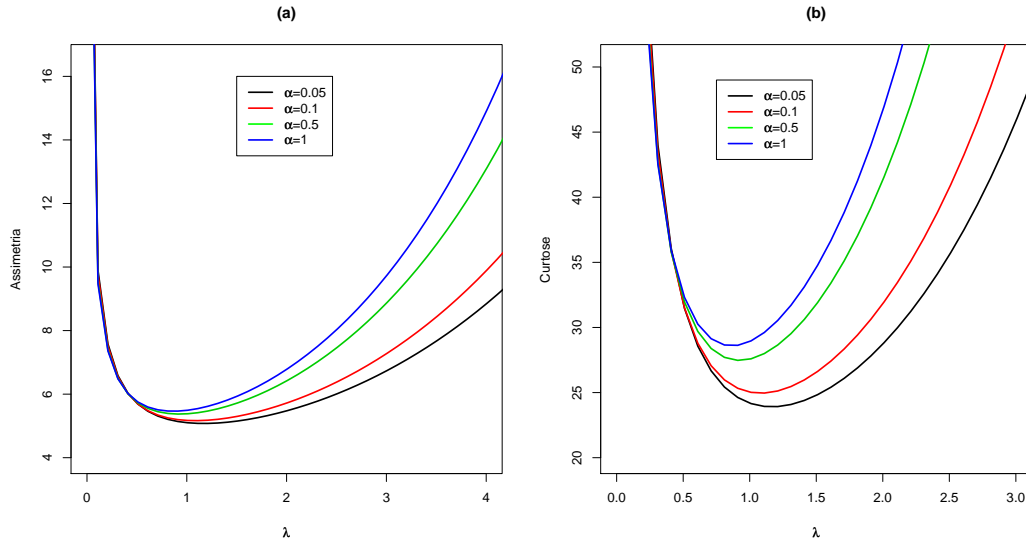


Figura 7 - Assimetria (a) e curtose (b) da distribuição GGE como função do parâmetro λ , para $\tau = 0.08$, $k = 2$ e alguns valores de α

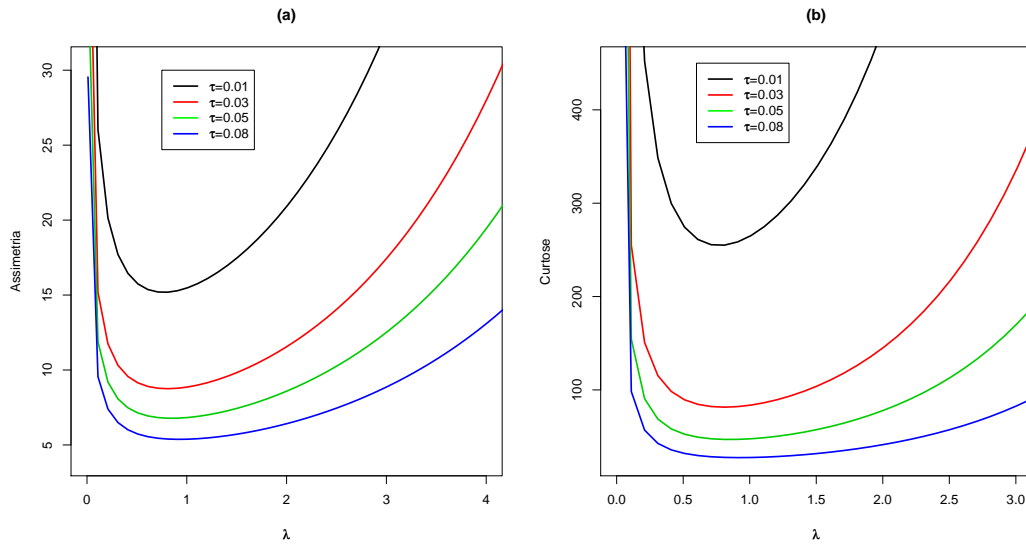


Figura 8 - Assimetria (a) e curtose (b) da distribuição GGE como função do parâmetro λ , para $\alpha = 0.5$, $k = 2$ e alguns valores de τ

em que $M_{\alpha,\tau,k(m+1)+i}(s)$ é dada por

$$M_{\alpha,\tau,k}(s) = \frac{\text{sinal}(\tau)}{\Gamma(k)} \sum_{m=0}^{\infty} \Gamma\left(\frac{m}{\tau} + k\right) \frac{(s\alpha)^m}{m!}. \quad (33)$$

Prova:

Tem-se que, por definição

$$M_{\alpha,\tau,k}(s) = \frac{|\tau|}{\alpha^{\tau k} \Gamma(k)} \int_0^\infty t^{\tau k-1} \exp \left[st - \left(\frac{t}{\alpha} \right)^\tau \right] dt.$$

Seja a substituição $u = t/\alpha$, então

$$M_{\alpha,\tau,k}(s) = \frac{|\tau|}{\Gamma(k)} \int_0^\infty \exp(s\alpha u) u^{\tau k-1} \exp(-u^\tau) du.$$

Expandindo a exponencial em série de Taylor temos

$$M_{\alpha,\tau,k}(s) = \frac{|\tau|}{\Gamma(k)} \sum_{m=0}^{\infty} \frac{(s\alpha)^m}{m!} \int_0^\infty u^{\tau k+m-1} \exp(-u^\tau) du.$$

A integral obtida como $\int_0^\infty u^{\tau k+m-1} \exp(-u^\tau) du = \tau^{-1} \Gamma(k + m/\tau)$, logo

$$M_{\alpha,\tau,k}(s) = \frac{\text{sinal}(\tau)}{\Gamma(k)} \sum_{m=0}^{\infty} \Gamma \left(\frac{m}{\tau} + k \right) \frac{(s\alpha)^m}{m!}. \quad (34)$$

A equação (34) vale para qualquer τ diferente de zero. Entretanto, se $\tau > 1$, podemos simplificá-la considerando a função Wright hipergeométrica generalizada definida por

$${}_p\Psi_q \left[\begin{matrix} (\alpha_1, A_1), \dots, (\alpha_p, A_p) \\ (\beta_1, B_1), \dots, (\beta_q, B_q) \end{matrix} ; x \right] = \sum_{m=0}^{\infty} \frac{\prod_{j=1}^p \Gamma(\alpha_j + A_j m)}{\prod_{j=1}^q \Gamma(\beta_j + B_j m)} \frac{x^m}{m!}. \quad (35)$$

Está função existe se $1 + \sum_{j=1}^q B_j - \sum_{j=1}^p A_j > 0$. Combinando as equações (34) e (35), nós temos

$$M_{\alpha,\tau,k}(s) = \frac{1}{\Gamma(k)} {}_1\Psi_0 \left[\begin{matrix} (1, 1/\tau) \\ - \end{matrix} ; s\alpha \right]. \quad (36)$$

Finalmente, a fgm para a distribuição GGE pode ser escrita na forma de mistura (20) como

$$M(t) = \sum_{m,i=0}^{\infty} w(m, i, k, \lambda) M_{\alpha,\tau,k(m+1)+i}(s), \quad (37)$$

onde $M_{\alpha,\tau,k(m+1)+i}(s)$ é obtida a partir da equação (34) (ou (36)).

3.2.3 Desvios médios

A quantidade da dispersão em uma população é medida até certo ponto, pela totalidade dos desvios da média e mediana. Se T tem distribuição GGE, nós podemos derivar os desvios médios em torno da média $\mu'_1 = E(T)$ e sobre a mediana m das relações

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

respectivamente. Definindo a integral $I(s) = \int_0^s t f(t) dt$, estas medidas podem ser calculadas a partir de

$$\delta_1 = 2\mu'_1 F(\mu'_1) - 2I(\mu'_1) \text{ e } \delta_2 = \mu'_1 - 2I(m), \quad (38)$$

onde $F(\mu'_1)$ e $F(m)$ são facilmente obtidas a partir da equação (13). A integral $I(s)$ pode ser calculada a partir de (20) como

$$I(s) = \sum_{m,i=0}^{\infty} w(m, i, k, \lambda) J(\alpha, \tau, [k(m+1) + i], s), \quad (39)$$

em que

$$J(\alpha, \tau, k, s) = \int_0^s t g_{\alpha, \tau, k}(t) dt.$$

Nós podemos obter a densidade da distribuição $GG(\alpha, \tau, k)$ definindo $x = t/\alpha$

$$J(\alpha, \tau, k, s) = \frac{\alpha|\tau|}{\Gamma(k)} \int_0^{s/\alpha} x^{\tau k} \exp(-x^\tau) dx.$$

Substituindo $w = x^\tau$, podemos escrever $J(\alpha, \tau, k, s)$ em termos da função gama incompleta da seguinte forma

$$J(\alpha, \tau, k, s) = \frac{\alpha \operatorname{ sinal}(\tau)}{\Gamma(k)} \int_0^{(s/\alpha)^\tau} w^{k+\tau^{-1}-1} \exp(-w) dw = \gamma(k + \tau^{-1}, (s/\alpha)^\tau).$$

Assim, inserindo o último resultado em (39) temos

$$I(s) = \sum_{m,i=0}^{\infty} w(m, i, k, \lambda) \gamma(k(m+1) + i + \tau^{-1}, (s/\alpha)^\tau). \quad (40)$$

Os desvios médios para a distribuição GGE podem ser obtidos através das equações (38) e (45).

3.2.4 Confiabilidade

Existem muitos trabalhos interessados na estimação da confiabilidade $R = Pr(T_2 < T_1)$ quando T_1 e T_2 são variáveis aleatórias independentes e identicamente distribuídas. A forma algébrica para R tem sido obtida para a maioria das distribuições padrões bem conhecidas. Mas, existem ainda muitas outras distribuições para as quais a forma de R não foi obtida (SILVA, 2008). Sabe-se que a forma de R , quando T_1 e T_2 são independentes e têm a mesma distribuição, pode ser expressa como

$$R = \int_0^{\infty} f(t)F(t)dt, \quad (41)$$

onde $f(t)$ e $F(t)$ são dados a partir das equações (12) e (13). Utilizando o Teorema 1 temos que:

- Para $\tau > 0$,

$$R = \int_0^{\infty} \sum_{m,i=0}^{\infty} w(m,i,k,\lambda) \frac{\tau}{\alpha \Gamma[k(m+1)+i]} \exp\left[-\left(\frac{t}{\alpha}\right)^{\tau}\right] \left(\frac{t}{\alpha}\right)^{\tau[k(m+1)+i]-1} \times \left\{ \gamma_1\left[k, \left(\frac{t}{\alpha}\right)^{\tau}\right] \right\}^{\lambda} dt. \quad (42)$$

Fazendo $x = \left(\frac{t}{\alpha}\right)^{\tau}$ em (42) têm

$$R = \sum_{m,i=0}^{\infty} \frac{w(m,i,k,\lambda) \text{sin}(\tau)}{\Gamma[k(m+1)+i]} \int_0^{\infty} x^{k(m+1)+i-1} \exp(-x) \gamma_1(k,x)^{\lambda} dx. \quad (43)$$

Usando uma expansão para $[\gamma_1(k,x)]^{\lambda}$, dada por

$$\begin{aligned} \{1 - [1 - \gamma_1(k,x)]\}^{\lambda} &= \sum_{j=0}^{\infty} \binom{\lambda}{j} (-1)^j [1 - \gamma_1(k,x)]^j \\ &= \sum_{j=0}^{\infty} \sum_{m_1=0}^j (-1)^{j+m_1} \binom{\lambda}{j} \binom{j}{m_1} \gamma_1(k,x)^{m_1}, \end{aligned}$$

e reescrevendo a equação (43) usando a equação anterior, tem-se

$$R = \sum_{m,i,j=0}^{\infty} \sum_{m_1=0}^j v_1(m,m_1,i,j,k,\lambda) I[k(m+1),i,m_1], \quad (44)$$

em que

$$v_1(m, m_1, i, j, k, \lambda) = \frac{(-1)^{j+m_1} \lambda \operatorname{ sinal}(\tau) s_m(\lambda) c_{m,i}}{\Gamma(k)^{m+1}} \binom{\lambda}{j} \binom{j}{m_1},$$

e

$$I[k(m+1), i, m_1] = \int_0^\infty x^{k(m+1)+i-1} \exp(-x) \gamma_1(k, x)^{m_1} dx.$$

Usando a função de Lauricella do tipo A definida na Seção 3.2.1, a última integral pode ser expressa como (veja NADARAJAH, 2008b, equação (23))

$$\begin{aligned} I[k(m+1), i, m_1] &= [k(m+1)]^{-m_1} \Gamma(i + k(m + m_1 + 1)) \times \\ &F_A^{(m_1)}(i + k(m + m_1 + 1); k(m+1), \dots, k(m+1); \\ &k(m+1) + 1, \dots, k(m+1) + 1; -1, \dots, -1). \end{aligned} \quad (45)$$

- Para $\tau < 0$,

$$\begin{aligned} R &= \int_0^\infty \sum_{m,i=0}^\infty w(m, i, k, \lambda) \frac{-\tau}{\alpha \Gamma[k(m+1) + i]} \exp\left[-\left(\frac{t}{\alpha}\right)^\tau\right] \left(\frac{t}{\alpha}\right)^{\tau[k(m+1)+i]-1} \times \\ &\left(1 - \left\{\gamma_1\left[k, \left(\frac{t}{\alpha}\right)^\tau\right]\right\}^\lambda\right) dt. \end{aligned} \quad (46)$$

Fazendo $x = \left(\frac{t}{\alpha}\right)^\tau$ na equação (46) têm-se

$$R = \sum_{m,i=0}^\infty w(m, i, k, \lambda) \left\{1 - \frac{\operatorname{ sinal}(\tau)}{\Gamma[k(m+1)] + i} \int_0^\infty x^{k(m+1)+i-1} \exp(-x) \gamma_1(k, x)^\lambda\right\}. \quad (47)$$

Utilizando novamente uma expansão para $[\gamma_1(k, x)]^\lambda$, dada por

$$\begin{aligned} \{1 - [1 - \gamma_1(k, x)]\}^\lambda &= \sum_{j=0}^\infty \binom{\lambda}{j} (-1)^j [1 - \gamma_1(k, x)]^j \\ &= \sum_{j=0}^\infty \sum_{m_1=0}^j (-1)^{j+m_1} \binom{\lambda}{j} \binom{j}{m_1} \gamma_1(k, x)^{m_1}, \end{aligned}$$

obtem-se então

$$R = \sum_{m,i=0}^{\infty} w(m, i, k, \lambda) \left\{ 1 - \sum_{j=0}^{\infty} \sum_{m_1=0}^j v_2(m, m_1, i, j, k, \tau, \lambda) I[k(m+1), i, m_1] \right\}, \quad (48)$$

onde

$$v_2(m, m_1, i, j, k, \tau, \lambda) = \frac{(-1)^{j+m_1} \text{sinal}(\tau)}{\Gamma[k(m+1) + i]} \binom{\lambda}{j} \binom{j}{m_1}$$

e $w(m, i, k, \lambda)$ e $I[k(m+1), i, m_1]$ são dados por (20) e (45), respectivamente.

3.2.5 Estatísticas de ordem

A função de densidade $f_{i:n}(t)$ da i -ésima estatística de ordem, para $i = 1, \dots, n$, das variáveis aleatórias T_1, \dots, T_n tendo densidade dada na equação (12), pode ser escrita como

$$f_{i:n}(t) = \frac{1}{B(i, n-i+1)} f(t) F(t)^{i-1} [1 - F(t)]^{n-i},$$

onde $f(t)$ e $F(t)$ são as função de densidade de probabilidade e função de densidade acumulada da distribuição GGE, respectivamente, e $B(\cdot, \cdot)$ denota a função beta. Usando uma expansão binomial na equação acima, podemos facilmente obter

$$f_{i:n}(t) = \frac{1}{B(i, n-i+1)} f(t) \sum_{l=0}^{n-i} \binom{n-i}{l} (-1)^l F(t)^{i+l-1}. \quad (49)$$

Seja $f_{\alpha, \tau, k, \lambda}(t)$ a função de densidade da distribuição GGE(α, τ, k, λ). Juntando as expressões (12) e (13) na última equação, podemos escrever

- Para $\tau > 0$,

$$f_{i:n}(t) = \frac{1}{B(i, n-i+1)} \sum_{l=0}^{n-i} \binom{n-i}{l} \frac{(-1)^l}{(i+l)} f_{\alpha, \tau, k, \lambda(i+l)}(t). \quad (50)$$

A equação (50) mostra que a função de densidade da i -ésima estatística de ordem é uma mistura finita de densidades GGE. Assim, os momentos das estatísticas de ordem podem ser calculados diretamente a partir de (50) usando as expressões gerais (28) e (31). O r -ésimo momento para a i -ésima estatística de ordem, $\mu'_{r(i:n)}$, reduz-se a

$$\mu'_{r(i:n)} = \frac{\lambda \alpha^r \text{sinal}(\tau)}{B(i, n-i+1)} \sum_{l=0}^{n-i} \sum_{j=0}^{\infty} \sum_{m=0}^j \frac{(-1)^{l+j+m}}{\Gamma(k)} \binom{n-i}{l} \binom{\lambda(i+l)-1}{j} \binom{j}{m} I \left[k, \frac{r}{\tau}, m \right].$$

- Para $\tau < 0$,

$$f_{i:n}(t) = \frac{-\lambda\tau}{\alpha\Gamma(k)B(i, n-i+1)} \left(\frac{t}{\alpha}\right)^{\tau k-1} \exp\left[-\left(\frac{t}{\alpha}\right)^\tau\right] \left\{\gamma_1\left[k, \left(\frac{t}{\alpha}\right)^\tau\right]\right\}^{\lambda-1} \times \sum_{l=0}^{n-i} \binom{n-i}{l} (-1)^l \left(1 - \left\{\gamma_1\left[k, \left(\frac{t}{\alpha}\right)^\tau\right]\right\}^\lambda\right)^{i+l-1}. \quad (51)$$

Usando um expansão binomial na expressão acima, reescrevemos (51) como

$$f_{i:n}(t) = \frac{1}{B(i, n-i+1)} \sum_{l=0}^{n-i} \sum_{j_1=0}^{i+l-1} \frac{(-1)^{l+j_1}}{(1+j_1)} \binom{n-i}{l} \binom{i+l-1}{j_1} f_{\alpha, \tau, k, \lambda(1+j_1)}(t). \quad (52)$$

A equação (52) mostra que a função de densidade da i -ésima estatística de ordem é uma mistura finita de densidades GGE. Assim, os momentos das estatísticas podem ser calculados diretamente de (52) usando as expressões gerais (28) e (31). Finalmente, o r -ésimo momento da i -ésima estatística de ordem se reduz a

$$\begin{aligned} \mu'_{r(i:n)} &= \frac{\lambda \alpha^r \text{sinal}(\tau)}{B(i, n-i+1)} \sum_{l=0}^{n-i} \sum_{j_1=0}^{i+l-1} \sum_{j=0}^{\infty} \sum_{m=0}^j \frac{(-1)^{l+j+j_1+m}}{\Gamma(k)} \binom{n-i}{l} \binom{i+l-1}{j_1} \\ &\times \binom{\lambda(1+j_1)-1}{j} \binom{j}{m} I\left[k, \frac{r}{\tau}, m\right], \end{aligned} \quad (53)$$

onde $I[k, r/\tau, m]$ é dado em (31).

3.3 Estimação de máxima verossimilhança

Seja t_1, t_2, \dots, t_n uma amostra aleatória de uma distribuição GGE com vetor de parâmetros $\boldsymbol{\theta} = (\alpha, \tau, k, \lambda)^T$, a estimação dos parâmetros pode ser feita pelo método de máxima verossimilhança, sendo o logaritmo da função de verossimilhança dado por

$$\begin{aligned} l(\boldsymbol{\theta}) &= n\{\log(\lambda) + \log(\tau) - \log(\alpha) - \log[\Gamma(k)]\} + (\tau k - 1) \sum_{i=1}^n \log\left(\frac{t_i}{\alpha}\right) \\ &- \sum_{i=1}^n \left(\frac{t_i}{\alpha}\right)^\tau + (\lambda - 1) \sum_{i=1}^n \log\left\{\gamma_1\left[k, \left(\frac{t_i}{\alpha}\right)^\tau\right]\right\}. \end{aligned}$$

Derivando-se $l(\boldsymbol{\theta})$ em relação a α , τ , k e λ , os elementos da função escore são dados por

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \alpha} = \frac{-n\hat{\tau}\hat{k}}{\hat{\alpha}} + \frac{\hat{\tau}}{\hat{\alpha}} \sum_{i=1}^n v_i - \frac{(\hat{\lambda} - 1)\hat{\tau}}{\hat{\alpha}\Gamma(k)} \sum_{i=1}^n \frac{v_i^k \exp(-v_i)}{\gamma_1(k, v_i)},$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \tau} = \frac{n}{\hat{\tau}} + \frac{\hat{k}}{\hat{\tau}} \sum_{i=1}^n \log(v_i) - \frac{1}{\hat{\tau}} \sum_{i=1}^n v_i \log(v_i) + \frac{(\hat{\lambda} - 1)}{\hat{\tau}\Gamma(k)} \sum_{i=1}^n \frac{v_i^k \exp(-v_i) \log(v_i)}{\gamma_1(k, v_i)},$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial k} = -n\Psi(\hat{k})\hat{\lambda} + \sum_{i=1}^n \log(v_i) + \frac{(\hat{\lambda} - 1)}{\Gamma(k)} \sum_{i=1}^n \frac{\gamma_1(k, v_i)}{\gamma_1(k, v_i)} e$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \lambda} = \frac{n}{\hat{\lambda}} + \sum_{i=1}^n \log[\gamma_1(k, v_i)].$$

em que $v_i = \left(\frac{t_i}{\alpha}\right)^\tau$, $\gamma_1(k, v_i) = \int_0^{v_i} w^{k-1} \exp(-w) \log(w) dw$ e $\psi(\hat{k})$ é a função digama que é definida como $\psi(k) = \frac{\partial(\log[\Gamma(k)])}{\partial k} = \frac{\Gamma'(k)}{\Gamma(k)}$, com $k > 0$. Os estimadores de máxima verossimilhança (EMV) são obtidos da solução $\frac{\partial l(\boldsymbol{\theta})}{\partial \alpha} = \frac{\partial l(\boldsymbol{\theta})}{\partial \tau} = \frac{\partial l(\boldsymbol{\theta})}{\partial k} = \frac{\partial l(\boldsymbol{\theta})}{\partial \lambda} = 0$. Para estimação dos parâmetros será utilizado o PROC NLMIXED do SAS.

Neste caso, a matriz de informação de observada $\ddot{L}(\boldsymbol{\theta})$ é dada por

$$\ddot{L}(\boldsymbol{\theta}) = \begin{bmatrix} \ddot{L}_{\alpha,\alpha} & \ddot{L}_{\alpha,\tau} & \ddot{L}_{\alpha,k} & \ddot{L}_{\alpha,\lambda} \\ \ddot{L}_{\tau,\alpha} & \ddot{L}_{\tau,\tau} & \ddot{L}_{\tau,k} & \ddot{L}_{\tau,\lambda} \\ \ddot{L}_{k,\alpha} & \ddot{L}_{k,\tau} & \ddot{L}_{k,k} & \ddot{L}_{k,\lambda} \\ \ddot{L}_{\lambda,\alpha} & \ddot{L}_{\lambda,\tau} & \ddot{L}_{\lambda,k} & \ddot{L}_{\lambda,\lambda} \end{bmatrix},$$

onde os elementos da matriz $\ddot{L}(\boldsymbol{\theta})$ encontram-se no Apêndice A.

3.4 Teste de hipóteses

Suponha que se deseja testar hipóteses do tipo $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, para isso, pode-se utilizar as estatísticas da razão de verossimilhança, Wald ou escore para testar essas hipóteses. Neste sentido, percebe-se a importância da distribuição gama generalizada exponenciada, já que ela pode ser usada para discriminar modelos. Assim sendo, supor que deseja-se em testar as hipóteses $H_0 : \lambda = 1$ versus $H_1 : \lambda \neq 1$, isto é, testar a adequabilidade do modelo gama generalizado. Nesse caso, a estatística da razão de verossimilhança é dada por

$$\Lambda = 2[l(\hat{\alpha}, \hat{\tau}, \hat{k}, \hat{\lambda}) - l(\tilde{\alpha}, \tilde{\tau}, \tilde{k}, 1)]$$

em que $\tilde{\alpha}$, $\tilde{\tau}$ e \tilde{k} são os estimadores de máxima verossimilhança sob H_0 e $\hat{\alpha}$, $\hat{\tau}$, \hat{k} e $\hat{\lambda}$ são os estimadores de máxima verossimilhança sob o espaço paramétrico completo e rejeita-se a hipótese nula quando $\Lambda > \chi_{1-\alpha}^2(1)$ que é o quantil $(1 - \alpha)$ da distribuição qui-quadrada com 1 grau de liberdade.

4 MODELO LOG-GAMA GENERALIZADO EXPONENCIADO

Seja T uma variável aleatória com distribuição GGE com função de densidade dada na equação (12). Considere a nova variável aleatória $Y = \log(T)$. Usando o método do Jacobiano pode-se obter a distribuição da variável aleatória Y que fica expressa da seguinte forma

$$f(y) = f(e^y)|J|,$$

em que $|J| = \frac{\partial t}{\partial y} = |e^y|$ é o Jacobiano da transformação. Portanto, o modelo log-gama generalizado exponenciado (LGGE) é dado da seguinte forma

$$f(y) = \frac{\lambda|\tau|}{\Gamma(k)} \left(\frac{e^y}{\alpha}\right)^{\tau k} \exp\left\{-\left(\frac{e^y}{\alpha}\right)^\tau\right\} \left\{\gamma_1\left[k, \left(\frac{e^y}{\alpha}\right)^\tau\right]\right\}^{\lambda-1}, \quad (54)$$

em que $y \in \mathbb{R}$, $\alpha > 0$, $\tau \neq 0$, $k > 0$ e $\lambda > 0$.

Neste trabalho é proposto uma reparametrização no modelo LGGE com os parâmetros μ , σ e q , da seguinte forma

$$\begin{cases} \mu = \ln(\alpha) + \frac{1}{\tau}\ln(k), \\ \sigma = \frac{1}{\tau\sqrt{k}} \text{ e} \\ q = \frac{1}{\sqrt{k}}. \end{cases}$$

Portanto, o modelo LGGE pode ser reescrito como

$$f(y) = \begin{cases} \frac{\lambda|q|(q^{-2})^{q^{-2}}}{\sigma\Gamma(q^{-2})} \exp\left\{q^{-1}\left(\frac{y-\mu}{\sigma}\right) - q^{-2}\exp\left[q\left(\frac{y-\mu}{\sigma}\right)\right]\right\} \\ \left\{\gamma_1\left[q^{-2}, q^{-2}\exp\left(q\left(\frac{y-\mu}{\sigma}\right)\right)\right]\right\}^{\lambda-1}, & q \neq 0. \end{cases} \quad (55)$$

Usando uma extensão proposta por Prentice (1974) e depois Lawless (2003) no ponto $q = 0$, $f(y)$ pode ser definida como

$$f(y) = \begin{cases} \frac{\lambda|q|(q^{-2})^{q^{-2}}}{\sigma\Gamma(q^{-2})} \exp\left\{q^{-1}\left(\frac{y-\mu}{\sigma}\right) - q^{-2}\exp\left[q\left(\frac{y-\mu}{\sigma}\right)\right]\right\} \\ \left\{\gamma_1\left[q^{-2}, q^{-2}\exp\left(q\left(\frac{y-\mu}{\sigma}\right)\right)\right]\right\}^{\lambda-1}, & q \neq 0, \\ \frac{\lambda}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right] \left[\Phi\left(\frac{y-\mu}{\sigma}\right)\right]^{\lambda-1}, & q = 0, \end{cases} \quad (56)$$

em que $\Phi(\cdot)$ é a função de distribuição acumulada da normal padrão.

Se Y tem distribuição LGGE, denotaremos por $Y \sim \text{LGGE}(\mu, \mu, q, \lambda)$.

A Figura 9 mostra a função de densidade para a variável aleatória Y que é chamada de distribuição LGGE, onde podemos observar a flexibilidade para diferentes valores de λ .

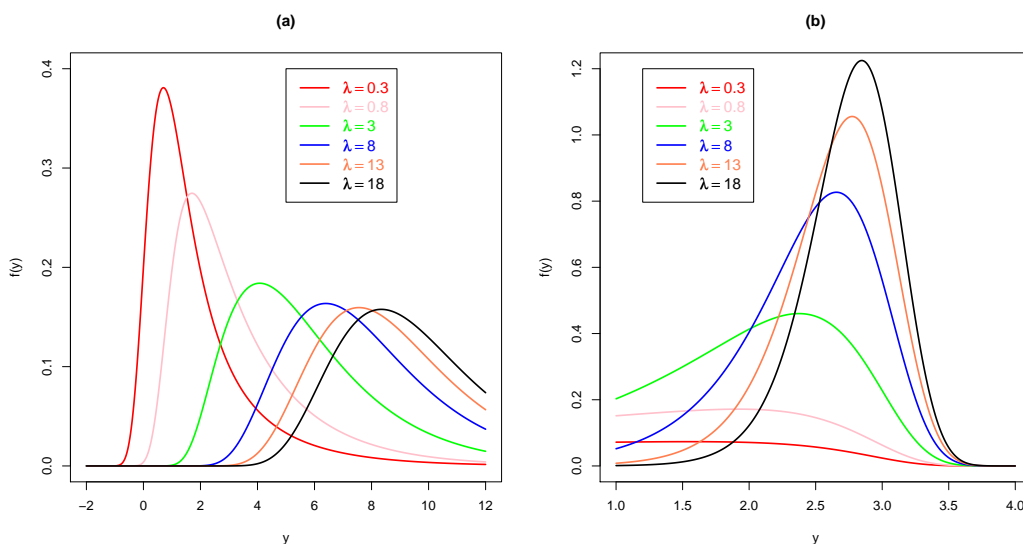


Figura 9 - Gráfico da função de densidade da distribuição LGGE para (a) $\mu = 2$, $\sigma = 1.2$ e $q = -2$ e (b) $\mu = 2$, $\sigma = 1.2$ e $q = 3$

4.1 Fórmula geral para os momentos do modelo LGGE

Teorema 4. Se $Y \sim LGGE(\mu, \sigma, q, \lambda)$, então os momentos da distribuição LGGE podem ser escritos como

$$\mu'_r = \sum_{m,i=0}^{\infty} \sum_{l=0}^r \frac{\lambda \text{sinal}(q) s_m(\lambda) c_{m,i}}{\Gamma(q^{-2})^{m+1}} \binom{r}{l} \left[\frac{2\sigma}{q} \log(|q|) + \mu \right]^{r-l} \left(\frac{\partial \Gamma[q^{-2}(m+1)+i]}{\partial [q^{-2}(m+1)+i]} \right)^{\frac{\sigma l}{q}}.$$

Prova:

Seja o r -ésimo momento $\mu'_r = E(Y^r)$ da distribuição LGGE é dado por

$$\begin{aligned} \mu'_r &= \int_{-\infty}^{\infty} y^r \frac{\lambda |q| (q^{-2})^{q^{-2}}}{\sigma \Gamma(q^{-2})} \exp \left\{ q^{-1} \left(\frac{y-\mu}{\sigma} \right) - q^{-2} \exp \left[q \left(\frac{y-\mu}{\sigma} \right) \right] \right\} \\ &\times \left(\gamma_1 \left\{ q^{-2}, q^{-2} \exp \left[q \left(\frac{y-\mu}{\sigma} \right) \right] \right\} \right)^{\lambda-1} dx. \end{aligned}$$

Fazendo $x = q^{-2} \exp \left[q \left(\frac{y-\mu}{\sigma} \right) \right]$ na equação anterior temos

$$\mu'_r = \int_0^{\infty} \frac{\lambda \text{sinal}(q)}{\Gamma(q^{-2})} \left\{ \frac{\sigma}{q} [\log(x) + 2 \log(|q|)] + \mu \right\}^r x^{q^{-2}-1} \exp(-x) \gamma_1(q^{-2}, x)^{\lambda-1} dx.$$

Uma expansão para $[\gamma_1(q^{-2}, x)]^{\lambda-1}$ expressa na equação (19) é dada por

$$[\gamma_1(q^{-2}, x)]^{\lambda-1} = \sum_{m,i=0}^{\infty} \frac{s_m(\lambda) c_{m,i}}{\Gamma(q^{-2})^m} x^{q^{-2}m+i}.$$

Inserindo a última equação na expressão para μ'_r e intercalando os termos, temos

$$\mu'_r = \int_0^{\infty} \frac{\lambda \text{sinal}(q)}{\Gamma(q^{-2})} \left\{ \frac{\sigma}{q} [\log(x) + 2 \log(|q|)] + \mu \right\}^r x^{q^{-2}-1} \exp(-x) \sum_{m,i=0}^{\infty} \frac{s_m(\lambda) c_{m,i}}{\Gamma(q^{-2})^m} x^{q^{-2}m+i}. \quad (57)$$

Usando binômio de Newton em $\left\{ \frac{\sigma}{q} [\log(x) + 2 \log(|q|)] + \mu \right\}^r$ na expressão (57), temos que

$$\mu'_r = \sum_{m,i=0}^{\infty} \sum_{l=0}^r \frac{\lambda \text{sinal}(q) s_m(\lambda) c_{m,i}}{\Gamma(q^{-2})^{m+1}} \binom{r}{l} \left[\frac{2\sigma}{q} \log(|q|) + \mu \right]^{r-l} \int_0^{\infty} \frac{\sigma l}{q} \log(x) \exp(-x) x^{q^{-2}(m+1)+i-1} dx.$$

O cálculo da última integral é dado em Prudnikov et al. (1986, vol. 1, Seção 2.6.21, integral 1). Portanto, os momentos da distribuição LGGE podem ser escritos como

$$\mu'_r = \sum_{m,i=0}^{\infty} \sum_{l=0}^r \frac{\lambda \text{sinal}(q) s_m(\lambda) c_{m,i}}{\Gamma(q^{-2})^{m+1}} \binom{r}{l} \left[\frac{2\sigma}{q} \log(|q|) + \mu \right]^{r-l} \left(\frac{\partial \Gamma[q^{-2}(m+1)+i]}{\partial [q^{-2}(m+1)+i]} \right)^{\frac{\sigma l}{q}}.$$

Representações gráficas da assimetria e curtose da distribuição LGGE como função do parâmetro λ , para alguns valores escolhidos do parâmetro μ e para alguns valores escolhidos do parâmetro σ , são dadas nas Figuras 10 e 11, respectivamente. Na Figura 10, tem-se o gráfico da assimetria (a) e curtose (b) para $\sigma = 0.8$, $q = 2.5$ e alguns valores de μ . Na Figura 11, tem-se o gráfico da assimetria (a) e curtose (b) para $\mu = 0.001$, $q = 0.5$ e alguns valores de σ .

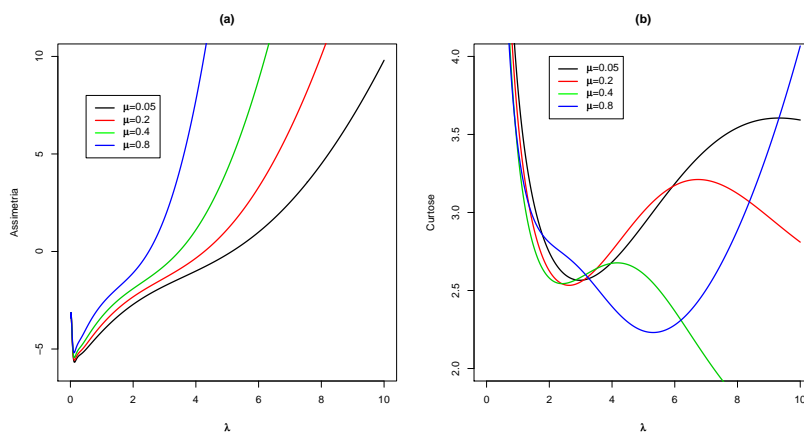


Figura 10 - Assimetria (a) e curtose (b) da distribuição LGGE como função do parâmetro λ , para $\sigma = 0.8$, $q = 2.5$ e alguns valores de μ

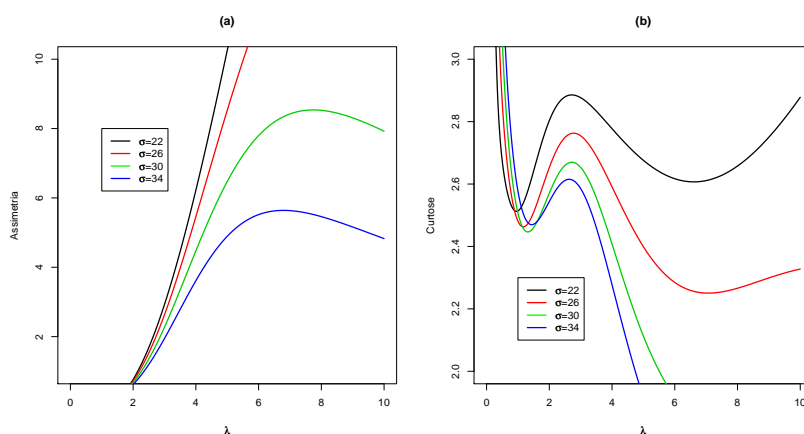


Figura 11 - Assimetria (a) e curtose (b) da distribuição LGGE como função do parâmetro λ , para $\mu = 0.001$, $q = 0.5$ e alguns valores de σ

4.2 Procedimento inferenciais

4.2.1 Características do modelo LGGE sem censuras

As dificuldades básicas na análise estatística do modelo GGE, ocorrem pela presença de quatro parâmetros desconhecidos, e pela complexidade computacional dos procedimentos inferenciais. No modelo original foram encontradas dificuldades pela própria estrutura da função de verossimilhança. Devido as simplificações obtidas pelo uso do modelo LGGE, o mesmo vem sendo utilizado no lugar do modelo GGE original.

Considerando y_1, y_2, \dots, y_n uma amostra aleatória com distribuição LGGE, o logaritmo da função de verossimilhança é dado por:

$$l(\boldsymbol{\theta}) = \sum_{i=1}^n \log f(y_i; \boldsymbol{\theta}), \quad (58)$$

em que $\boldsymbol{\theta} = (\mu, \sigma, q, \lambda)^T$ é o vetor de parâmetros.

Portanto, substituindo a função de densidade dada na expressão (54) na expressão (58), o logaritmo da função de verossimilhança é dado por

- para $q \neq 0$

$$\begin{aligned} l(\boldsymbol{\theta}) = & n \left[\log(\lambda) + \log(|q|) + q^{-2} \log(q^{-2}) - \log[\sigma \Gamma(q^{-2})] \right] + \\ & \sum_{i=1}^n \left\{ q^{-1} \left(\frac{y_i - \mu}{\sigma} \right) - q^{-2} \exp \left[q \left(\frac{y_i - \mu}{\sigma} \right) \right] \right\} + \\ & \sum_{i=1}^n \left\{ (\lambda - 1) \log \left\{ \gamma_1 \left[q^{-2}, q^{-2} \exp \left[q \left(\frac{y_i - \mu}{\sigma} \right) \right] \right] \right\} \right\}; \end{aligned}$$

- para $q = 0$

$$\begin{aligned} l(\boldsymbol{\theta}) = & n \left[\log(\lambda) - \frac{1}{2} \log(2\Pi) - \log(\sigma) \right] + \sum_{i=1}^n \left\{ -\frac{1}{2} \left(\frac{y_i - \mu}{\sigma} \right)^2 \right\} + \\ & \sum_{i=1}^n \left\{ (\lambda - 1) \log \left[\Phi \left(\frac{y_i - \mu}{\sigma} \right) \right] \right\}. \end{aligned}$$

Para se obterem os estimadores de máxima verossimilhança, basta resolver, simultaneamente, as equações:

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \mu} = 0, \quad \frac{\partial l(\boldsymbol{\theta})}{\partial \sigma} = 0, \quad \frac{\partial l(\boldsymbol{\theta})}{\partial q} = 0 \quad \text{e} \quad \frac{\partial l(\boldsymbol{\theta})}{\partial \lambda} = 0.$$

Entretanto, o método da máxima verossimilhança pode ter problemas de convergência e tem sido sugerido (LAWLESS, 2003) em seu lugar, o uso do método da máxima verossimilhança perfilada. Esse método consiste em atribuir uma grade de valores para q dentro de uma amplitude fixada (-3,3). Para cada valor fixado de q , obtêm-se os valores $\tilde{\mu}(q)$, $\tilde{\sigma}(q)$ e $\tilde{\lambda}(q)$ que maximizam $l(\mu(q), \sigma(q), q, \lambda(q))$, fazendo-se $\frac{\partial l(\mu(q), \sigma(q), q, \lambda(q))}{\partial \mu} = \frac{\partial l(\mu(q), \sigma(q), q, \lambda(q))}{\partial \sigma} = \frac{\partial l(\mu(q), \sigma(q), q, \lambda(q))}{\partial \lambda} = 0$, sendo que, para $q > 0$, por exemplo, têm-se

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \mu} = -\frac{n}{q\sigma} + \frac{1}{q\sigma} \sum_{i=1}^n u_i - \frac{q(\lambda-1)}{\sigma\Gamma(q^{-2})} \sum_{i=1}^n \left\{ \frac{1}{\gamma_1[q^{-2}, q^{-2}u_i]} (q^{-2}u_i)^{q^{-2}} \exp\{-q^{-2}u_i\} \right\},$$

$$\begin{aligned} \frac{\partial l(\boldsymbol{\theta})}{\partial \sigma} &= -\frac{n}{\sigma} - \sum_{i=1}^n \left(\frac{y_i - \mu}{q\sigma^2} \right) + \sum_{i=1}^n \left(\frac{y_i - \mu}{q\sigma^2} \right) u_i \\ &\quad - \frac{q(\lambda-1)}{\Gamma(q^{-2})} \sum_{i=1}^n \left\{ \frac{1}{\gamma_1[q^{-2}, q^{-2}u_i]} (q^{-2}u_i)^{q^{-2}} \exp\{-q^{-2}u_i\} \left(\frac{y_i - \mu}{\sigma^2} \right) \right\} \end{aligned}$$

e

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \lambda} = \frac{n}{\lambda} + \sum_{i=1}^n \left\{ \log[\gamma_1(q^{-2}, q^{-2}u_i)] \right\},$$

em que $u_i = \exp\left[q\left(\frac{y_i - \mu}{\sigma}\right)\right]$.

A seguir faz-se o gráfico de $l(\tilde{\mu}(q), \tilde{\sigma}(q), q, \tilde{\lambda}(q))$ e as estimativas de μ , σ , λ e q serão dadas pelos valores em que $l(\hat{\mu}(q), \hat{\sigma}(q), q, \hat{\lambda}(q))$ é máximo.

4.2.2 Características do modelo LGGE com censuras

Seja y_1, y_2, \dots, y_n uma amostra aleatória com distribuição LGGE com vetor de parâmetros $\boldsymbol{\theta} = (\mu, \sigma, q, \lambda)^T$, em que $y_i = \log(t_i)$, $i = 1, 2, \dots, n$. A cada observação associa-se um indicador de censura δ_i , isto é,

$$\delta_i = \begin{cases} 1, & \text{se for um tempo de falha} \\ 0, & \text{se for um tempo de censura.} \end{cases}$$

Então, a função de verossimilhança considerando censura aleatória é dada por:

$$L(\boldsymbol{\theta}) = \prod_{i:\delta_i=1} f(y_i) \prod_{i:\delta_i=0} S(y_i),$$

ou equivalentemente por

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

em que $f(y_i)$, $S(y_i)$ e $h(y_i)$ são as funções de densidade, sobrevivência e taxa de falha da variável aleatória Y , e o logaritmo da função de verossimilhança pode ser escrito da seguinte forma:

$$l(\boldsymbol{\theta}) = \log[L(\boldsymbol{\theta})] = \sum_{i:\delta_i=1} \log[f(y_i)] + \sum_{i:\delta_i=0} \log[S(y_i)]. \quad (59)$$

A função de sobrevivência para a distribuição LGGE é dada por

- para $q > 0$

$$\begin{aligned} S(y) &= 1 - F(y) = P(Y > y) \\ &= P(\mu + \sigma W > y) \\ &= P\left(W > \frac{y - \mu}{\sigma}\right) = P(W > w) \\ &= \int_w^\infty \frac{\lambda q (q^{-2})^{q^{-2}}}{\Gamma(q^{-2})} \exp\left\{q^{-1}u - q^{-2}\exp(qu)\right\} \left\{\gamma_1\left[q^{-2}, q^{-2}\exp(qu)\right]\right\}^{\lambda-1} du. \end{aligned}$$

- para $q < 0$

$$\begin{aligned} S(y) &= 1 - F(y) = P(Y > y) \\ &= P(\mu + \sigma W > y) \\ &= P\left(W > \frac{y - \mu}{\sigma}\right) = P(W > w) \\ &= \int_w^\infty \frac{-\lambda q (q^{-2})^{q^{-2}}}{\Gamma(q^{-2})} \exp\left\{q^{-1}u - q^{-2}\exp(qu)\right\} \left\{\gamma_1\left[q^{-2}, q^{-2}\exp(qu)\right]\right\}^{\lambda-1} du. \end{aligned}$$

Portanto, a função de sobrevivência $S(y)$ é dada por

$$S(y) = \begin{cases} \int_w^\infty \frac{\lambda q (q^{-2})^{q^{-2}}}{\Gamma(q^{-2})} \exp\left\{q^{-1}u - q^{-2}\exp(qu)\right\} \left\{\gamma_1\left[q^{-2}, q^{-2}\exp(qu)\right]\right\}^{\lambda-1} du, & q > 0; \\ \int_w^\infty \frac{-\lambda q (q^{-2})^{q^{-2}}}{\Gamma(q^{-2})} \exp\left\{q^{-1}u - q^{-2}\exp(qu)\right\} \left\{\gamma_1\left[q^{-2}, q^{-2}\exp(qu)\right]\right\}^{\lambda-1} du, & q < 0; \\ 1 - [\Phi(w)]^\lambda, & q = 0, \end{cases} \quad (60)$$

em que $w = \frac{y-\mu}{\sigma}$ e $\Phi(w)$ é a função de distribuição acumulada da normal padrão.

Resolvendo as integrais dada na expressão (60) e fazendo $w = \frac{y-\mu}{\sigma}$, tem-se:

$$S(y) = \begin{cases} 1 - \left\{ \gamma_1 \left[q^{-2}, q^{-2} \exp \left(q \left(\frac{y-\mu}{\sigma} \right) \right) \right] \right\}^\lambda, & \text{se } q > 0; \\ \left\{ \gamma_1 \left[q^{-2}, q^{-2} \exp \left(q \left(\frac{y-\mu}{\sigma} \right) \right) \right] \right\}^\lambda, & \text{se } q < 0; \\ 1 - \left[\Phi \left(\frac{y-\mu}{\sigma} \right) \right]^\lambda, & \text{se } q = 0. \end{cases} \quad (61)$$

Desta maneira, substituindo a função de densidade (54) e a função de sobrevivência (61) na expressão (59) o logaritmo da função de verossimilhança pode ser escrito como

$$l(\boldsymbol{\theta}) = \begin{cases} \sum_{i:\delta_i=1} \log \left\{ \frac{\lambda q}{\sigma \Gamma(q^{-2})} (q^{-2})^{q^{-2}} \exp \left\{ q^{-2} \left[\left(\frac{y_i-\mu}{\sigma} q \right) - \exp \left(\frac{y_i-\mu}{\sigma} q \right) \right] \right\} \right. \\ \left. \left\{ \gamma_1 \left[q^{-2}, q^{-2} \exp \left(\frac{y_i-\mu}{\sigma} q \right) \right] \right\}^{\lambda-1} \right\} \\ + \sum_{i:\delta_i=0} \log \left\{ 1 - \left\{ \gamma_1 \left[q^{-2}, q^{-2} \exp \left(\frac{y_i-\mu}{\sigma} q \right) \right] \right\}^\lambda \right\}, & \text{se } q > 0; \\ \sum_{i:\delta_i=1} \log \left\{ \frac{-\lambda q}{\sigma \Gamma(q^{-2})} (q^{-2})^{q^{-2}} \exp \left\{ q^{-2} \left[\left(\frac{y_i-\mu}{\sigma} q \right) - \exp \left(\frac{y_i-\mu}{\sigma} q \right) \right] \right\} \right. \\ \left. \left\{ \gamma_1 \left[q^{-2}, q^{-2} \exp \left(\frac{y_i-\mu}{\sigma} q \right) \right] \right\}^{\lambda-1} \right\} \\ + \sum_{i:\delta_i=0} \log \left\{ \left\{ \gamma_1 \left[q^{-2}, q^{-2} \exp \left(\frac{y_i-\mu}{\sigma} q \right) \right] \right\}^\lambda \right\}, & \text{se } q < 0; \\ \sum_{i:\delta_i=1} \log \left\{ \frac{\lambda}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2\sigma^2} (y_i - \mu)^2 \right] \left[\Phi \left(\frac{y_i-\mu}{\sigma} \right) \right]^{\lambda-1} \right\} \\ + \sum_{i:\delta_i=0} \log \left\{ 1 - \left[\Phi \left(\frac{y_i-\mu}{\sigma} \right) \right]^\lambda \right\}, & \text{se } q = 0. \end{cases} \quad (62)$$

Assim, como no modelo em que se consideram apenas falhas, também será utilizado o método da máxima verossimilhança perfilada para a obtenção das estimativas dos parâmetros.

4.2.3 Características do modelo LGGE com censuras e covariáveis

Sabe-se que o tempo de vida pode ser influenciado por variáveis regressoras ou covariáveis, que podem ser: idade, sexo, estágio da doença no diagnóstico etc, sendo de grande interesse prático o modelo de regressão LGGE.

A classe de modelos de locação e escala se caracteriza pelo fato de $Y = \log(T)$ ter uma distribuição com parâmetros de locação $\mu(x)$ dependendo das variáveis regressoras, e um parâmetro de escala σ constante. Pode-se então escrever

$$Y = \mu(x) + \sigma W, \quad (63)$$

em que $\sigma > 0$ e W tem uma distribuição que não depende de x . Agora, por outro lado, considere-se um modelo de regressão em que se assume uma relação linear entre o logaritmo do tempo de vida, denotado por Y e um vetor de covariáveis $\mathbf{x} = (x_1, x_2, \dots, x_p)^T$. Além disso, suponha que Y dado \mathbf{x} tem distribuição LGGE e pode se representada na forma

$$Y = \mathbf{x}^T \boldsymbol{\beta} + \sigma Z, \quad (64)$$

em que $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$ e $\sigma > 0$ são os parâmetros desconhecidos.

A distribuição da variável aleatória Z pode ser obtida utilizando a transformação de variáveis que neste caso é dada por

$$f(z) = \begin{cases} \frac{\lambda|q|}{\Gamma(q^{-2})} (q^{-2})^{q^{-2}} \exp\{q^{-1}z - q^{-2}\exp(qz)\} \{\gamma_1[q^{-2}, q^{-2}\exp(qz)]\}^{\lambda-1}, & \text{se } q \neq 0 \\ \frac{\lambda}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) [\Phi(z)]^{\lambda-1}, & \text{se } q = 0, \end{cases} \quad (65)$$

em que $z = \frac{y - \mathbf{x}^T \boldsymbol{\beta}}{\sigma}$ e $\Phi(\cdot)$ é a função de distribuição acumulada da normal padrão..

Assim, a função de sobrevivência considerando tempos de falha, censuras e covariáveis pode ser escrita como

- para $q > 0$

$$\begin{aligned} S(y) &= 1 - F(y) = P(Y > y) \\ &= P(\mathbf{x}^T \boldsymbol{\beta} + \sigma Z > y) \\ &= P\left(Z > \frac{y - \mathbf{x}^T \boldsymbol{\beta}}{\sigma}\right) \\ &= \int_{\frac{y - \mathbf{x}^T \boldsymbol{\beta}}{\sigma}}^{\infty} \frac{\lambda q (q^{-2})^{q^{-2}}}{\Gamma(q^{-2})} \exp\left\{q^{-1}z - q^{-2}\exp(qz)\right\} \{\gamma_1[q^{-2}, q^{-2}\exp(qz)]\}^{\lambda-1} dz. \end{aligned}$$

- para $q < 0$

$$\begin{aligned}
S(y) &= 1 - F(y) = P(Y > y) \\
&= P(\mathbf{x}^T \boldsymbol{\beta} + \sigma Z > y) \\
&= P\left(Z > \frac{y - \mathbf{x}^T \boldsymbol{\beta}}{\sigma}\right) \\
&= \int_{\frac{y - \mathbf{x}^T \boldsymbol{\beta}}{\sigma}}^{\infty} \frac{-\lambda q (q^{-2})^{q^{-2}}}{\Gamma(q^{-2})} \exp\left\{q^{-1}z - q^{-2} \exp(qz)\right\} \left\{\gamma_1[q^{-2}, q^{-2} \exp(qz)]\right\}^{\lambda-1} dz.
\end{aligned}$$

em que $z = \frac{y - \mathbf{x}^T \boldsymbol{\beta}}{\sigma}$.

Resolvendo as integrais dadas anteriormente para $q > 0$ e $q < 0$, a função de sobrevivência, considerando um modelo de regressão, pode ser escrita como

$$S(y) = \begin{cases} 1 - \left\{ \gamma_1 \left[q^{-2}, q^{-2} \exp\left(q \left(\frac{y - \mathbf{x}^T \boldsymbol{\beta}}{\sigma} \right) \right) \right] \right\}^{\lambda}, & \text{se } q > 0; \\ \left\{ \gamma_1 \left[q^{-2}, q^{-2} \exp\left(q \left(\frac{y - \mathbf{x}^T \boldsymbol{\beta}}{\sigma} \right) \right) \right] \right\}^{\lambda}, & \text{se } q < 0; \\ 1 - \left[\Phi \left(\frac{y - \mathbf{x}^T \boldsymbol{\beta}}{\sigma} \right) \right]^{\lambda}, & \text{se } q = 0, \end{cases} \quad (66)$$

em que $\Phi(\cdot)$ é a função de distribuição acumulada da normal padrão.

Sejam $(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)$ n observações provenientes de uma distribuição LGGE, em que y_i representa o logaritmo do tempo de vida ou tempo até a censura e $\mathbf{x}_i = (1, x_{i1}, x_{i2}, \dots, x_{ip})^T$ o vetor de covariáveis associadas.

Utilizando a expressão (64) o modelo pode ser escrito como

$$y_i = x_i^T \boldsymbol{\beta} + \sigma z_i, \quad i = 1, 2, \dots, n, \quad (67)$$

em que o vetor de parâmetros é $\boldsymbol{\theta} = (\lambda, \sigma, q, \boldsymbol{\beta}^T)^T$ e $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$.

Nesse caso, o logaritmo da função de verossimilhança considerando tempos de

falha, censuras e covariáveis pode ser escrito da seguinte forma:

$$l(\boldsymbol{\theta}) = \begin{cases} \sum_{i:\delta_i=1} \log \left\{ \frac{\lambda q (q^{-2})^{q-2}}{\Gamma(q^{-2})} \exp \left[q^{-1} z_i - q^{-2} \exp(q z_i) \right] \left\{ \gamma_1 \left[q^{-2}, q^{-2} \exp(q z_i) \right] \right\}^{\lambda-1} \right\} \\ + \sum_{i:\delta_i=0} \log \left\{ 1 - \left[\gamma_1 \left(q^{-2}, q^{-2} \exp(q z_i) \right) \right]^\lambda \right\}, & q > 0; \\ \\ \sum_{i:\delta_i=1} \log \left\{ \frac{-\lambda q (q^{-2})^{q-2}}{\Gamma(q^{-2})} \exp \left[q^{-1} z_i - q^{-2} \exp(q z_i) \right] \left\{ \gamma_1 \left[q^{-2}, q^{-2} \exp(q z_i) \right] \right\}^{\lambda-1} \right\} \\ + \sum_{i:\delta_i=0} \log \left\{ \left[\gamma_1 \left(q^{-2}, q^{-2} \exp(q z_i) \right) \right]^\lambda \right\}, & q < 0; \\ \\ \sum_{i:\delta_i=1} \log \left\{ \frac{\lambda}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} z_i^2 \right\} \left[\Phi(z_i) \right]^{\lambda-1} \right\} + \sum_{i:\delta_i=0} \log \left\{ 1 - \left[\Phi(z_i) \right]^\lambda \right\}, & q = 0 \end{cases} \quad (68)$$

em que $z_i = \frac{y_i - x_i^T \boldsymbol{\beta}}{\sigma}$. O vetor de parâmetros para esse modelo, mais geral, é dado por $\boldsymbol{\theta}^T = (\lambda, \sigma, q, \boldsymbol{\beta}^T)$ e $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$.

A dificuldade na estimação dos parâmetros é ainda maior, pois têm-se $(p + 3)$ parâmetros a serem estimados e, então, usa-se o método da máxima verossimilhança.

Para se obterem os estimadores de máxima verossimilhança precisa-se derivar $l(\boldsymbol{\theta})$ em relação a $(\beta_1, \beta_2, \dots, \beta_p)$, λ , σ e q . Entretanto, dificuldades existentes na derivação de $l(\boldsymbol{\theta})$ em relação ao parâmetro q é ainda maior neste caso em que se admite censuras e covariáveis. Usa-se, então, o método descrito por Lawless (2003). É considerado 2 etapas neste método, a primeira etapa considera o parâmetro de forma q fixo, de modo que para um dado valor de q encontra-se os valores $\tilde{\beta}(q)$, $\tilde{\lambda}(q)$ e $\tilde{\sigma}(q)$ que maximizam $l(\boldsymbol{\theta})$. Feito isso, para alguns valores de q pode-se calcular o valor da função de verossimilhança e, então, localizar \hat{q} que leva a verossimilhança máxima. Para estimação dos parâmetros será utilizado o PROC NLMIXED do SAS.

Pode-se ainda calcular estimativas para a variância de $\hat{\boldsymbol{\theta}}$ e construir testes de hipóteses para os parâmetros utilizando o fato de que $\hat{\boldsymbol{\theta}}$ tem distribuição assintótica normal multivariada sob certas condições de regularidade com média $\boldsymbol{\theta}$ e matriz de covariância $I^{-1}(\boldsymbol{\theta})$ com $I(\boldsymbol{\theta}) = -E[\ddot{L}(\boldsymbol{\theta})]$ tal que $\ddot{L}(\boldsymbol{\theta}) = \left\{ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right\}$, ou seja,

$$\hat{\boldsymbol{\theta}} \sim N_{p+3}(\boldsymbol{\theta}, I^{-1}(\boldsymbol{\theta})).$$

Visto que o cálculo de $I(\boldsymbol{\theta})$, denominada matriz de informação de Fischer, não

é possível devido a presença de censuras, pode-se utilizar alternativamente a matriz $-\ddot{L}(\boldsymbol{\theta})$ avaliada em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, denominada matriz de informação observada, a qual é uma estimativa consistente de $I(\boldsymbol{\theta})$. Neste caso, a matriz $\ddot{L}(\boldsymbol{\theta})$ é dada por

$$\ddot{L}(\boldsymbol{\theta}) = \begin{bmatrix} \ddot{L}_{\sigma,\sigma} & \ddot{L}_{\sigma,\lambda} & \ddot{L}_{\sigma,\beta_j} \\ \ddot{L}_{\lambda,\sigma} & \ddot{L}_{\lambda,\lambda} & \ddot{L}_{\lambda,\beta_j} \\ \ddot{L}_{\beta_j,\sigma} & \ddot{L}_{\beta_j,\lambda} & \ddot{L}_{\beta_j,\beta_s} \end{bmatrix},$$

onde os elementos da matriz $\ddot{L}(\boldsymbol{\theta})$ encontram-se no Apêndice B.

Portanto, o intervalo de confiança assintótico para β_i , $i = 1, \dots, p$ considerando $(1 - \alpha)100\%$ de confiança é dado por

$$\hat{\beta}_i \pm z_{\alpha/2} \sqrt{\hat{V}(\hat{\beta}_i)}.$$

Para construção de um teste de hipótese para os parâmetros é utilizado a diagonal principal da matriz $\ddot{L}(\boldsymbol{\theta})$ como estimativa para a variância dos parâmetros. Sendo assim, a estatística para testar as hipóteses

$$\begin{cases} H_0 : \beta_i = \beta_{i0} \\ H_1 : \beta_i \neq \beta_{i0} \end{cases} \quad (69)$$

é dada por $z = \frac{\hat{\beta}_i - \beta_{i0}}{\sqrt{\hat{V}(\hat{\beta}_i)}} \sim N(0, 1)$.

4.3 Casos particulares do modelo de regressão LGGE

O modelo de regressão LGGE apresenta alguns modelos de regressão particulares que serão apresentados a seguir

- Para $\lambda = 1$ e $q = 1$ na equação (66) o modelo de regressão LGGE se reduz ao modelo de regressão log-Weibull ou valor extremo, com função de sobrevivência da forma:

$$S(y) = \exp \left[- \exp \left(\frac{y - x^T \beta}{\sigma} \right) \right], \quad (70)$$

que é o modelo de regressão Weibull clássico (ver, LAWLESS, 2003). Se $\sigma = 1$ e $\sigma = 1/2$

em adição a $\lambda = 1$ e $q = 1$, coincide com os modelos de regressão exponencial e Rayleigh, respectivamente.

- Para $q = 1$ na equação (66) o modelo de regressão LGGE se reduz ao modelo de regressão log-Weibull exponenciado, com função de sobrevivência da forma:

$$S(y) = 1 - \left\{ 1 - \exp \left[- \exp \left(\frac{y - x^T \beta}{\sigma} \right) \right] \right\}^\lambda, \quad (71)$$

que é o modelo de regressão introduzido por Mudholkar et al. (1995), Cancho et al. (1999), Ortega et al. (2006) e Cancho et al. (2008). Se $\sigma = 1$ em adição a $q = 1$, coincide com o modelo de regressão log-exponencial exponenciado. Se $\sigma = 1/2$ em adição a $q = 1$, coincide com o modelo de regressão log-Rayleigh generalizado.

- Para $\lambda = 1$ na equação (66) o modelo de regressão LGGE se reduz ao modelo de regressão log-gama generalizado, com função de sobrevivência da forma:

$$S(y) = \begin{cases} 1 - \gamma_1 \left(q^{-2}, q^{-2} \exp \left[q \left(\frac{y - x^T \beta}{\sigma} \right) \right] \right), & \text{se } q > 0; \\ \gamma_1 \left(q^{-2}, q^{-2} \exp \left[q \left(\frac{y - x^T \beta}{\sigma} \right) \right] \right), & \text{se } q < 0; \\ 1 - \Phi \left(\frac{y - x^T \beta}{\sigma} \right), & \text{se } q = 0, \end{cases} \quad (72)$$

que é o modelo de regressão introduzido por Ortega et al. (2003), Lawless (2003), Ortega et al. (2008) e Ortega et al. (2009). E $\Phi(\cdot)$ é a função de distribuição acumulada da normal padrão.

- Para $q = -1$ na equação (66) o modelo de regressão LGGE se reduz ao modelo de regressão log-Weibull inversa generalizado, com função de sobrevivência da forma:

$$S(y) = \left\{ 1 - \exp \left(- \exp \left[- \left(\frac{y - x^T \beta}{\sigma} \right) \right] \right) \right\}^\lambda, \quad (73)$$

que é o modelo de regressão introduzido por Gusmão et al. (2009). Se $\lambda = 1$ em adição a $q = -1$, coincide com o modelo de regressão log-Weibull inversa.

- Para $q = 0$ na equação (66) o modelo de regressão LGGE se reduz ao modelo de regressão log-normal exponenciado, com função de sobrevivência da forma:

$$S(y) = 1 - \left[\Phi\left(\frac{y - x^T \beta}{\sigma}\right) \right]^\lambda, \quad (74)$$

que é o modelo de regressão introduzido por Gusmão et al. (2009). Se $\lambda = 2$, coincide com o modelo de regressão log-normal skew.

5 MODELO DE LOCAÇÃO E ESCALA PARA DADOS AGRUPADOS

Foram apresentados anteriormente modelos de regressão que assumem independência entre os tempos de sobrevivência. Apesar dessa suposição ser válida para algumas aplicações, ela pode ser inadequada para outras. Assim, é necessário construir modelos de regressão que considerem a existência de uma possível associação entre os tempos e nesta situação um modelo frequentemente utilizado é o modelo de regressão com efeito aleatório. Esta parte do trabalho é baseada nos trabalhos de Valença (2003) e Silva (2008).

5.1 Modelo de regressão com efeito aleatório

Admitindo-se que os tempos de sobrevivência estão sujeitos a censura à direita, que o mecanismo de censura é aleatório e que a censura é não informativa. Considere uma amostra dividida em m grupos e seja T_{ji} o tempo de sobrevivência do i -ésimo indivíduo no j -ésimo grupo, com $i = 1, \dots, n_j$ e $j = 1, \dots, m$. Sejam x_{ji} o vetor de covariáveis e E_{ji} uma variável aleatória, independente de T_{ji} , que representa o tempo de censura associado a este indivíduo. Assim, os dados observados são $t_{ji} = \min\{T_{ji}, E_{ji}\}$ e

$$\delta_{ji} = \begin{cases} 1, & \text{se } T_{ji} \leq E_{ji} \\ 0, & \text{se } T_{ji} > E_{ji}, \end{cases}$$

em que δ_{ji} é o indicador de censura.

No modelo de locação e escala com efeito aleatório, todos os indivíduos do mesmo grupo tem um efeito aleatório comum, denotado por W_j , que é uma variável aleatória não observável. Nesse contexto, o modelo de locação e escala com efeito aleatório é representado da seguinte forma

$$Y_{ji} = W_j + x_{ji}^T \beta + \sigma Z_{ji}, \quad (75)$$

em que $Y_{ji} = \log(T_{ji})$, Z_{ji} 's são os erros aleatórios independentes e identicamente distribuídos e os elementos do vetor $\theta = (\beta^T, \sigma)^T$ são parâmetros desconhecidos. Cada grupo tem um efeito aleatório W_j , que é representado por variáveis aleatórias independentes e identicamente distribuídas com densidade $g(w_j, \mathbf{D})$ com variância ξ e \mathbf{D} é o vetor de parâmetros.

Assuma que $Cov(W_j, Z_{ji}) = 0$, e que, condicionado ao efeito aleatório W_j , as respostas dentro do grupo j são independentes. Assuma também que os efeitos aleatórios são independentes dos tempos de censura. A covariância entre o logaritmo dos tempos de sobrevivência dentre um grupo é dada por ξ , portanto o modelo reduz-se ao modelo de locação e escala usual quando $\xi = 0$.

Para estimar o vetor de parâmetros $\Psi = (\boldsymbol{\theta}^T, D^T)^T$ pode-se utilizar o método de máxima verossimilhança descrita a seguir.

5.1.1 Verossimilhança aumentada

Considere o modelo dado na equação (75), a função de verossimilhança condicional ao efeito aleatório para o i -ésimo indivíduo no j -ésimo grupo é dada por

$$f(y_{ji}|w_j, x_{ji})^{\delta_{ji}} S(y_{ji}|w_j, x_{ji})^{1-\delta_{ji}}$$

em que $f(\cdot)$ e $S(\cdot)$ denotam, respectivamente, as funções de densidade e sobrevivência de \mathbf{Y}_{ji} condicional ao efeito aleatório W_j .

Para o grupo j , o vetor da variável resposta é representado por $Y_{ji} = (Y_{j1}, Y_{j2}, \dots, Y_{jn_j})^T$ e de maneira análoga defina δ_j o vetor contendo o indicador de censuras. A função de verossimilhança condicionada ao efeito aleatório (independência dentre o grupo) para os indivíduos do grupo j é dada por

$$L_j(\boldsymbol{\theta}, W_j) = \prod_{i=1}^{n_j} f(y_{ji}|w_j, x_{ji})^{\delta_{ji}} S(y_{ji}|w_j, x_{ji})^{1-\delta_{ji}}. \quad (76)$$

Então, a verossimilhança aumentada para Ψ no grupo j é

$$L_j(\Psi; y_j, \delta_j, w_j) = L_j(\boldsymbol{\theta}, W_j)g(w_j; D), \quad (77)$$

em que $L_j(\boldsymbol{\theta}, W_j)$ está definida na equação (76).

Assumindo que existe independência entre os vetores (Y_1, Y_2, \dots, Y_m) a função de verossimilhança aumentada para toda a amostra é dada por

$$L_j(\Psi; y, \delta, w) = \prod_{j=1}^m L_j(\Psi; y_j, \delta_j, w_j), \quad (78)$$

em que $y = (y_1^T, y_2^T, \dots, y_m^T)^T$, $\delta = (\delta_1^T, \delta_2^T, \dots, \delta_m^T)^T$, $w = (w_1, w_2, \dots, w_m)^T$ e $L_j(\Psi; y_j, \delta_j, w_j)$ está definida na equação (77).

O logaritmo da função de verossimilhança correspondente é

$$l(\Psi; y, \delta, w) = \sum_{j=1}^m \log[L_j(\boldsymbol{\theta}, w_j)] + \sum_{j=1}^m \log[g(w_j, D)], \quad (79)$$

esta expressão pode ser dividida em duas partes, uma que depende somente do vetor de parâmetros \mathbf{D} e outra que depende do vetor de parâmetros $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma)^T$.

5.2 Modelo de regressão LGGE com efeito aleatório

Considere que $Y = \log(T)$ seja uma variável aleatória com distribuição LGGE, com densidade dada na equação (54), e função de sobrevivência dada na equação (61). Neste caso, para o modelo de locação e escala com efeito aleatório dado na equação (75), a densidade de Y para o i -ésimo indivíduo no j -ésimo grupo, denotada por $f(y_{ji}|w_j; \lambda, q, \sigma, \beta, x_{ji})$, dado x_{ji} e w_j é definida como

$$f(y_{ji}|w_j; \lambda, q, \sigma, \beta, x_{ji}) = \begin{cases} \frac{\lambda q (q^{-2})^{q-2}}{\sigma \Gamma(q^{-2})} \exp \left\{ q^{-1} \left(\frac{y_{ji} - x_{ji}^T \beta - w_j}{\sigma} \right) - q^{-2} \exp \left[q \left(\frac{y_{ji} - x_{ji}^T \beta - w_j}{\sigma} \right) \right] \right\} \\ \left\{ \gamma_1 \left[q^{-2}, q^{-2} \exp \left(q \left(\frac{y_{ji} - x_{ji}^T \beta - w_j}{\sigma} \right) \right) \right] \right\}^{\lambda-1}, & q \neq 0 \\ \frac{\lambda}{\sqrt{2\pi}\sigma} \exp \left[-\frac{1}{2} \left(\frac{y_{ji} - x_{ji}^T \beta - w_j}{\sigma} \right)^2 \right] \left[\Phi \left(\frac{y_{ji} - x_{ji}^T \beta - w_j}{\sigma} \right) \right]^{\lambda-1}, & q = 0. \end{cases} \quad (80)$$

A função de sobrevivência correspondente, denotada por $S(y_{ji}|w_j; \lambda, q, \sigma, \beta, x_{ji})$, é dada por

$$S(y_{ji}|w_j; \lambda, q, \sigma, \beta, x_{ji}) = \begin{cases} 1 - \left\{ \gamma_1 \left[q^{-2}, q^{-2} \exp \left(q \left(\frac{y_{ji} - x_{ji}^T \beta - w_j}{\sigma} \right) \right) \right] \right\}^\lambda, & \text{se } q > 0; \\ \left\{ \gamma_1 \left[q^{-2}, q^{-2} \exp \left(q \left(\frac{y_{ji} - x_{ji}^T \beta - w_j}{\sigma} \right) \right) \right] \right\}^\lambda, & \text{se } q < 0; \\ 1 - \left[\Phi \left(\frac{y_{ji} - x_{ji}^T \beta - w_j}{\sigma} \right) \right]^\lambda, & \text{se } q = 0. \end{cases} \quad (81)$$

Assuma para as variáveis aleatórias W_j , $j=1, \dots, m$, a distribuição $\Gamma(\frac{1}{\xi}, \frac{1}{\xi})$ que apresenta função de densidade dada por

$$g\left(w; \frac{1}{\xi}, \frac{1}{\xi}\right) = \frac{1}{\xi} w^{\frac{1}{\xi}-1} \exp\left(-\frac{w}{\xi}\right), \quad (82)$$

em que $w \geq 0$. Segue, então, que $E(W_j) = 1$ e $V(W_j) = \xi$. Se $\xi = 0$ então, obtém-se o modelo usual de locação e escala. Para mais detalhes, veja, Colosimo e Giolo (2006) ou LI (2006).

O procedimento de estimação é baseado na construção do logaritmo da função de verossimilhança. Para o modelo de regressão LGGE com efeito aleatório seguindo distribuição dada em (82), o logaritmo da função de verossimilhança, denotado por $l(\boldsymbol{\theta}; \xi)$ é dado por

- Para $q > 0$

$$l(\boldsymbol{\theta}; \xi) = \begin{cases} \sum_{i:\delta_i=1} \log \left\{ \frac{\lambda q}{\sigma \Gamma(q^{-2})} (q^{-2})^{q^{-2}} \exp \left\{ q^{-2} \left[\left(\frac{y_{ji} - x_{ji}^T \beta - w_j}{\sigma} \right) q \right] - \exp \left(\left(\frac{y_{ji} - x_{ji}^T \beta - w_j}{\sigma} \right) q \right) \right\} \right. \\ \left. \left\{ \gamma_1 \left[q^{-2}, q^{-2} \exp \left(\frac{y_{ji} - x_{ji}^T \beta - w_j}{\sigma} q \right) \right] \right\}^{\lambda-1} \right\} \\ + \sum_{i:\delta_i=0} \log \left\{ 1 - \left\{ \gamma_1 \left[q^{-2}, q^{-2} \exp \left(\frac{y_{ji} - x_{ji}^T \beta - w_j}{\sigma} q \right) \right] \right\}^\lambda \right\} \\ \left(\frac{1}{\xi} - 1 \right) \sum_{j=1}^m \log(w_j) - \frac{1}{\xi} \sum_{j=1}^m (w_j) - m \left\{ \frac{1}{\xi} \log(\xi) + \log \left[\Gamma \left(\frac{1}{\xi} \right) \right] \right\}; \end{cases} \quad (83)$$

- Para $q < 0$

$$l(\boldsymbol{\theta}; \xi) = \begin{cases} \sum_{i:\delta_i=1} \log \left\{ \frac{-\lambda q}{\sigma \Gamma(q^{-2})} (q^{-2})^{q^{-2}} \exp \left\{ q^{-2} \left[\left(\frac{y_{ji} - x_{ji}^T \beta - w_j}{\sigma} \right) q \right] - \exp \left(\frac{(y_{ji} - x_{ji}^T \beta - w_j)}{\sigma} q \right) \right\} \right\} \\ \left\{ \gamma_1 \left[q^{-2}, q^{-2} \exp \left(\frac{(y_{ji} - x_{ji}^T \beta - w_j)}{\sigma} q \right) \right] \right\}^{\lambda-1} \\ + \sum_{i:\delta_i=0} \log \left\{ \left\{ \gamma_1 \left[q^{-2}, q^{-2} \exp \left(\frac{(y_{ji} - x_{ji}^T \beta - w_j)}{\sigma} q \right) \right] \right\}^\lambda \right\} \\ \left(\frac{1}{\xi} - 1 \right) \sum_{j=1}^m \log(w_j) - \frac{1}{\xi} \sum_{j=1}^m (w_j) - m \left\{ \frac{1}{\xi} \log(\xi) + \log \left[\Gamma \left(\frac{1}{\xi} \right) \right] \right\}; \end{cases} \quad (84)$$

- Para $q = 0$

$$l(\boldsymbol{\theta}; \xi) = \begin{cases} \sum_{i:\delta_i=1} \log \left\{ \frac{\lambda}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2\sigma^2} (y_{ji} - x_{ji}^T \beta - w_j)^2 \right] \left[\Phi \left(\frac{y_{ji} - x_{ji}^T \beta - w_j}{\sigma} \right) \right]^{\lambda-1} \right\} \\ + \sum_{i:\delta_i=0} \log \left\{ 1 - \left[\Phi \left(\frac{y_{ji} - x_{ji}^T \beta - w_j}{\sigma} \right) \right]^\lambda \right\} \\ \left(\frac{1}{\xi} - 1 \right) \sum_{j=1}^m \log(w_j) - \frac{1}{\xi} \sum_{j=1}^m (w_j) - m \left\{ \frac{1}{\xi} \log(\xi) + \log \left[\Gamma \left(\frac{1}{\xi} \right) \right] \right\}. \end{cases} \quad (85)$$

Esta expressão pode ser separada em duas partes, uma que depende somente de ξ e outra que depende dos parâmetros λ , q , $\boldsymbol{\beta}^T$, σ . Dessa forma, tem-se que $l(\boldsymbol{\theta}; \xi) = l_1(\boldsymbol{\theta}) + l_2(\xi)$.

Para estimação dos parâmetros será utilizado o PROC NLMIXED do SAS.

6 APLICAÇÕES

6.1 Descrição dos dados

Para a realização deste trabalho foram avaliados três conjuntos de dados distintos.

O primeiro conjunto de dados foi fornecido pela Associação Mãe Admirável, situada na cidade de Caratinga, MG. Foram avaliados 141 residentes, dependentes químicos, no período de 2000 a 2005. A variável resposta foi o tempo de permanência na comunidade até a desistência do tratamento, considerando que cada residente permanece na Comunidade por um período máximo de 270 dias, sem qualquer contato com as drogas e quem alcança esta meta, foi considerado, neste trabalho, como um dado censurado.

O segundo conjunto de dados foi retirado do livro de Lawless (1944), que apresenta um estudo de isolamento elétrico (STONE, 1978) considerando três níveis de voltagem: 52.5 Kv, 55.0 kV e 57.5 kV, onde foram testadas 20 amostras para cada um dos três níveis. Os dados utilizados encontram-se no Apêndice D.

O terceiro conjunto de dados utilizado foi retirado do livro de Colosimo e Giolo (2006), que apresenta um estudo com animais da raça Nelore. O gado da raça Nelore é comumente usado no Brasil para a produção comercial de carne. Tempos não muito longos para um ganho específico de peso no período do nascimento até a desmama, bem como da desmama ao abate são, portanto, economicamente desejáveis. Identificar touros que produzam animais com um ganho rápido e específico de peso em um desses períodos é, portanto, um dos interesses dos produtores dessa raça de gado.

Estudos dessa natureza envolvem, em geral, uma grande quantidade de animais e, desse modo, apenas um subconjunto dos dados analisados por Giolo et al. (2003) foi considerado para análise neste trabalho. Esse subconjunto é composto de 4 touros Nelore que produziram um total de 155 animais. Todos os animais produzidos por estes touros nasceram na primavera entre os anos de 1993 e 1998, sendo 68% fêmeas. O número de animais por touro variou de 31 a 56. A variável resposta de interesse foi estabelecida como sendo o tempo, em dias, que um animal levou para ganhar 160 Kg no período do nascimento até o desmama. De acordo com Albuquerque e Fries (1998), em estudo com gado de corte da raça Nelore, 160 Kg é um ganho de peso realístico para esse período. Sexo dos animais (1 se macho e 0 se fêmea)

e idade da vaca no parto, que variou de 3 a 16 anos, foram as covariáveis fixas consideradas nesta análise.

6.2 Análise dos dados de residentes dependentes químicos (modelo GGE)

Para o conjunto de dados de residentes dependentes químicos foi construída a curva de risco empírico TTT-plot (Figura 12), indicando que a função de taxa de falha, associada ao conjunto de dados tem forma unimodal, pois verifica-se que o gráfico obtido apresenta uma curvatura primeiramente côncava e depois convexa. Portanto, a distribuição GGE será utilizada para modelar esse conjunto de dados.

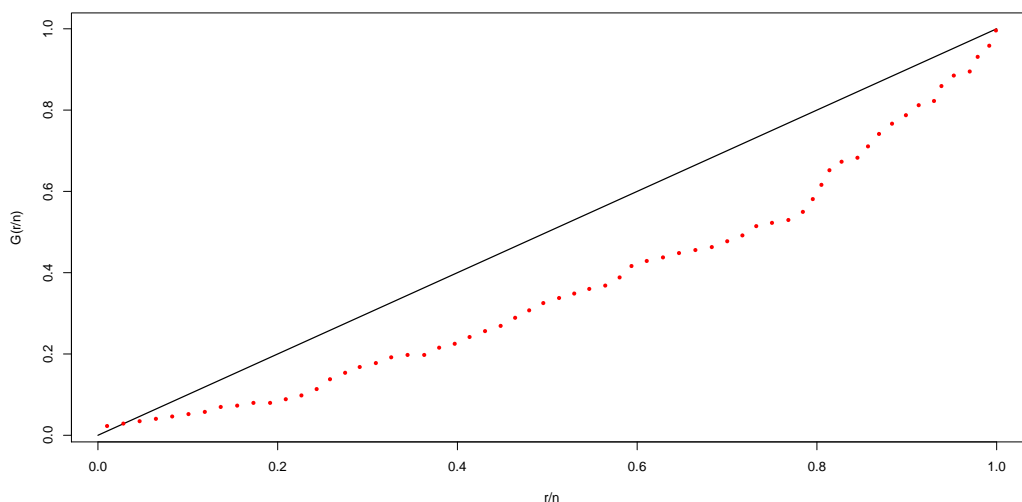


Figura 12 - TTT-plot para tempo de sobrevivência dos dados de residentes dependentes químicos

Verifica-se através da Tabela 1 que o modelo GGE ajusta melhor os dados de residentes dependentes químicos, quando comparado a outros modelos, pois observando os valores das estatísticas AIC e AICC dadas na Tabela 1, o modelo GGE possui os menores valores, ou seja, esse modelo pode ser escolhido como o melhor modelo para analisar os dados de residentes dependentes químicos.

Para confirmação do resultado obtido na Tabela 1, foram feitos também alguns gráficos da função de sobrevivência estimada versus os tempos de sobrevivência, verificando realmente que a distribuição GGE se ajusta melhor aos dados quando comparada a outras

Tabela 1 - Estimativas dos parâmetros do modelo para os dados de residentes dependentes químicos, seus correspondentes erros-padrão (entre parênteses) e as estatísticas AIC, AICC e BIC

Modelo	α	τ	κ	λ	AIC	AICC	BIC
GGE	220.81 (40.3460)	0.5504 (0.09051)	0.08638 (0.06074)	19.9877 (19.7494)	1224.9	1225.1	1236.6
GG	220.84 (109.68)	0.8005 (0.2096)	0.9112 (0.3150)	1 (-)	1228.2	1228.4	1237.0
EE	270.58 (47.9597)	1 (-)	1 (-)	0.6906 (0.08264)	1228.8	1228.9	1234.7
WE	202.33 (134.65)	0.6786 (0.1949)	1 (-)	1.0434 (0.5058)	1228.1	1228.2	1236.8
E	188.31 (18.9261)	1 (-)	1 (-)	1 (-)	1237.1	1237.2	1240.1
W	188.32 (25.2426)	0.7393 (0.06419)	1 (-)	1 (-)	1225.5	1225.6	1231.4

distribuições amplamente utilizadas na análise de tempos de vida. Este resultado pode ser observado na Figura 13, que compara a distribuição GGE versus as distribuições Weibull exponenciada e exponencial exponenciada, respectivamente, e na Figura 14, que compara a distribuição GGE versus as distribuições GG e Weibull, respectivamente. Estes resultados foram obtidos utilizando o software R.

6.3 Análise dos dados de níveis de voltagem (modelo de regressão LGGE)

Para o conjunto de dados de níveis de voltagem foi construída a curva de risco empírico TTT-plot (Figura 15), indicando que a função de taxa de falha, associada ao conjunto de dados tem forma unimodal, pois verifica-se que o gráfico obtido apresenta uma curvatura primeiramente côncava e depois convexa. Portanto, a distribuição GGE será utilizada para modelar esse conjunto de dados.

Verifica-se através da Tabela 2 que o modelo GGE ajusta melhor os dados de níveis de voltagem, quando comparado a outros modelos, pois observando os valores das estatísticas AIC e AICC dadas na Tabela 2, o modelo GGE possui os menores valores, ou seja,

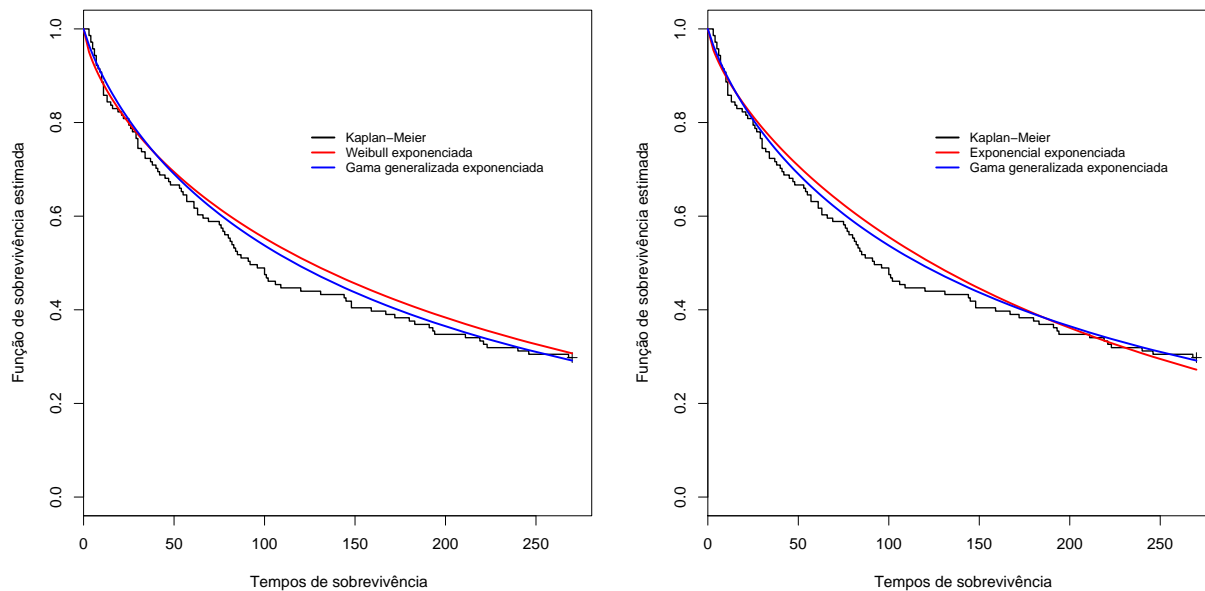


Figura 13 - Funções de sobrevivência estimadas

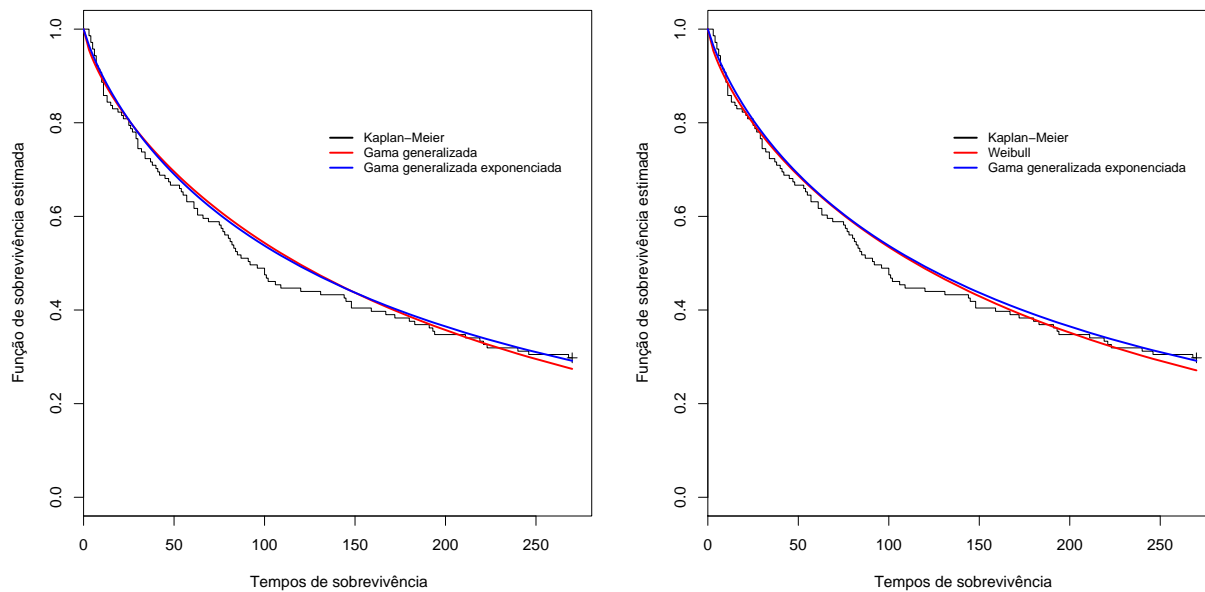


Figura 14 - Funções de sobrevivência estimadas

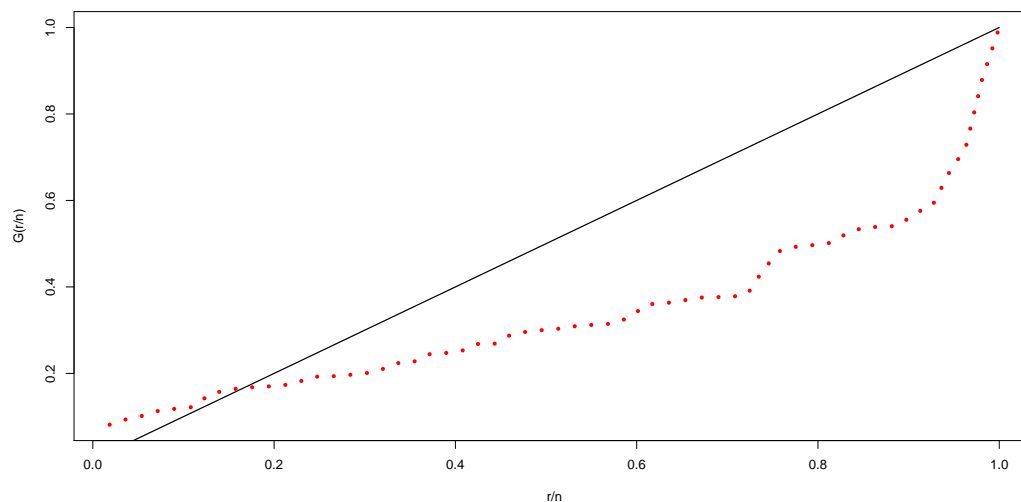


Figura 15 - TTT-plot para tempo de sobrevivência dos dados de níveis de voltagem

esse modelo pode ser escolhido como o melhor modelo para analisar os dados de níveis de voltagem.

Para confirmação do resultado obtido na Tabela 2, foram feitos também alguns gráficos da função de sobrevivência estimada versus os tempos de sobrevivência, verificando realmente que a distribuição GGE se ajusta melhor aos dados quando comparada a outras distribuições amplamente utilizadas na análise de tempos de vida. Este resultado pode ser observado na Figura 16, que compara a distribuição GGE versus as distribuições Weibull exponenciada e exponencial exponenciada, respectivamente, e na Figura 17, que compara a distribuição GGE versus as distribuições GG e Weibull, respectivamente. Estes resultados foram obtidos utilizando o software R.

O conjunto de dados de níveis de voltagem é analisado sob a perspectiva do modelo de regressão LGGE, o qual pode ser expresso da seguinte maneira

$$y_i = \beta_0 + \beta_1 x_{i1} + \sigma Z_i, \quad i = 1, \dots, 60,$$

em que Z_i é o erro aleatório independente e identicamente distribuído, e a resposta y_i denota o logaritmo do tempo de sobrevivência.

Ajustando o modelo LGGE via método de máxima verossimilhança perfilada

Tabela 2 - Estimativas dos parâmetros do modelo para os dados de níveis de voltagem, seus correspondentes erros-padrão (entre parênteses) e as estatísticas AIC, AICC e BIC

Modelo	α	τ	κ	λ	AIC	AICC	BIC
GGE	1248.33 (3.3370)	0.7767 (0.009786)	0.00096 (0.000125)	1932.78 (0.001011)	869.8	870.5	878.1
GG	1154.19 (656.67)	0.9360 (0.2346)	1.0435 (0.4371)	1 (-)	875.1	875.6	881.4
EE	1154.18 (230.65)	1 (-)	1 (-)	1.0512 (0.2001)	873.4	873.6	877.6
WE	2163.25 (1494.33)	1.4776 (0.6925)	1 (-)	0.4696 (0.3648)	885.6	886.0	891.8
E	1154.91 (157.18)	1 (-)	1 (-)	1 (-)	871.6	871.6	873.7
W	1154.20 (172.41)	0.9512 (0.09471)	1 (-)	1 (-)	873.3	873.5	877.5

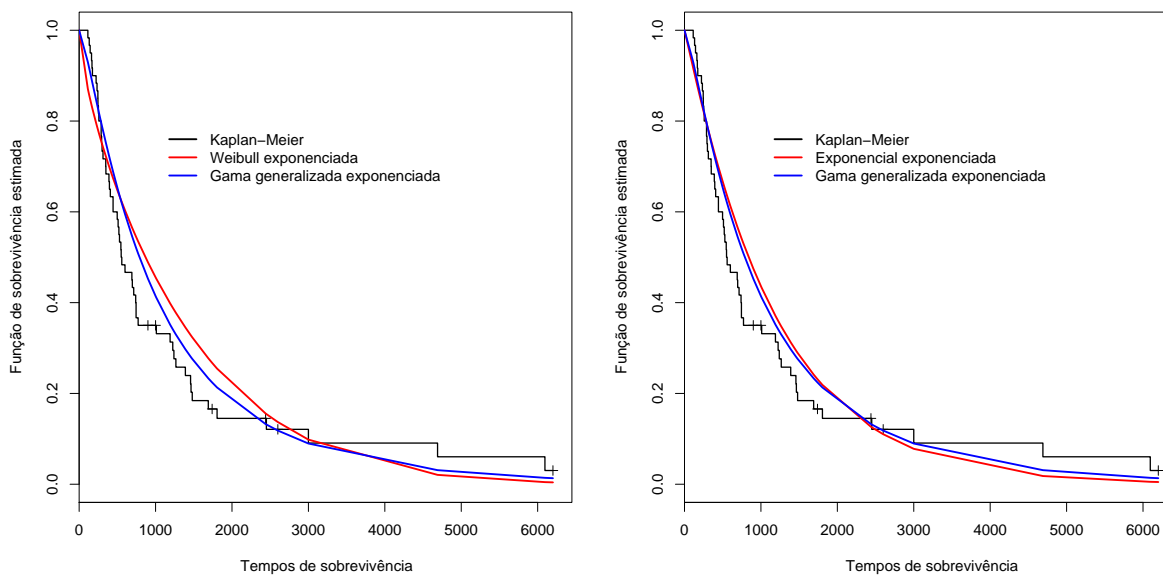


Figura 16 - Funções de sobrevivência estimadas

pode-se observar através da Figura 18 que quando $q = -0.5393$, $l(\hat{\mu}(q), \hat{\sigma}(q), q, \hat{\lambda}(q))$ atinge o valor máximo. Assim, assume-se este valor como a estimativa de máxima verossimilhança

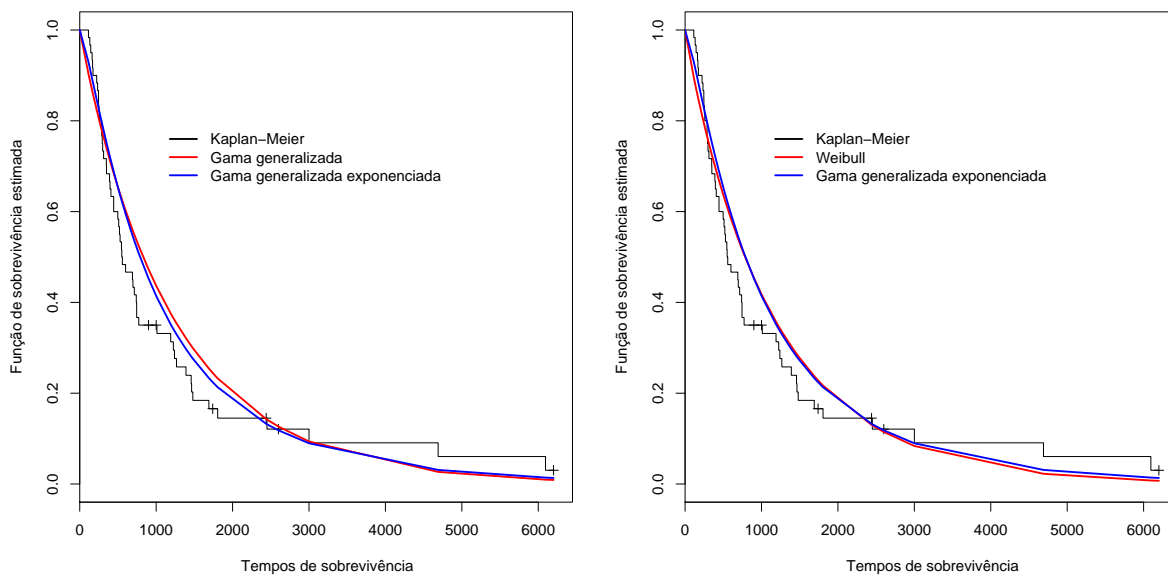


Figura 17 - Funções de sobrevivência estimadas

para o parâmetro q .

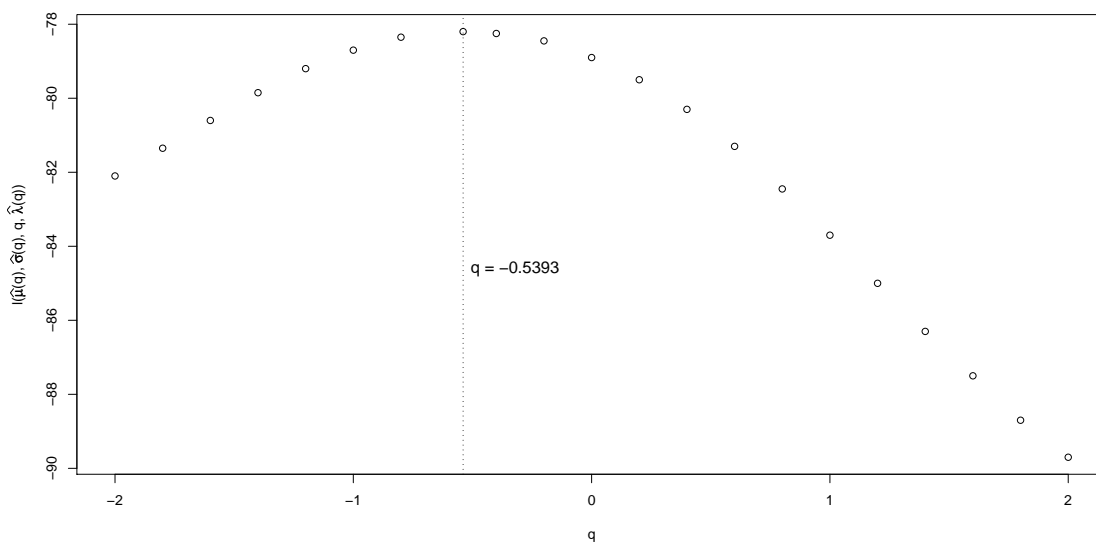


Figura 18 - Gráfico do método de máxima verossimilhança perfilada

As estimativas de máxima verossimilhança dos parâmetros do modelo LGGE encontram-se na Tabela 3. Por meio da Tabela 3, pode-se observar que a variável x_{i1} (nível de

voltagem) é significativa considerando um nível de significância de 5%.

Tabela 3 - Estimativas de máxima verossimilhança dos parâmetros do modelo LGGE para os dados de níveis de voltagem

Parâmetro	Estimativa	Erro-padrão	Valor P
λ	1.0153	1.1729	-
σ	0.9015	0.4549	-
β_0	16.0476	4.1492	0.0003
β_1	-0.1775	0.06653	0.0098

A Figura 19 mostra a função de sobrevivência estimada pelo modelo de regressão LGGE para os níveis de voltagem 52.5 kV, 55.0 kV e 57.5 kV.

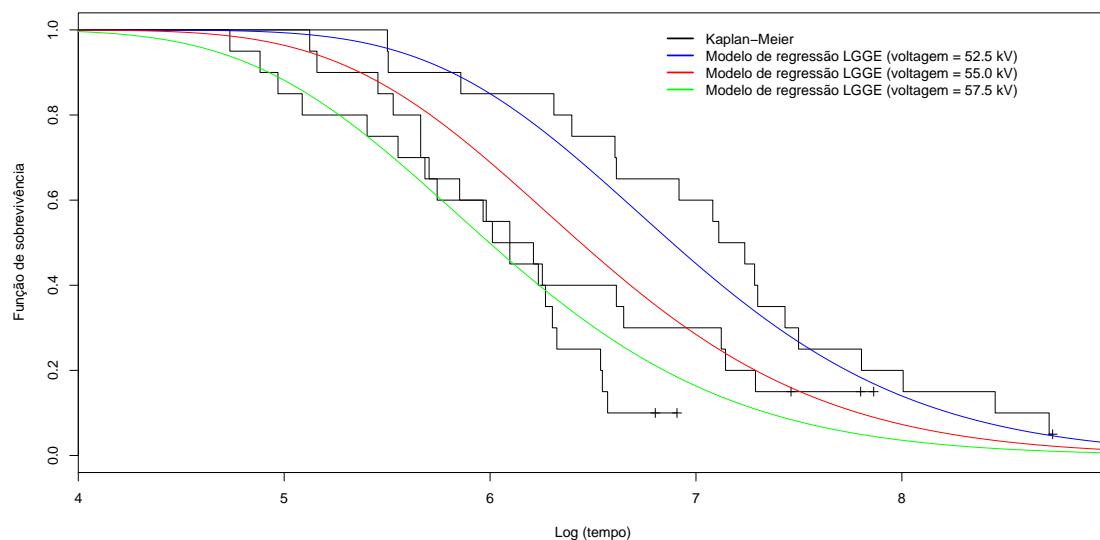


Figura 19 - Curvas de sobrevivência estimadas pelo modelo de regressão LGGE considerando três níveis de voltagem: 52.5 kV, 55.0 kV e 57.5 kV

6.4 Análise dos dados de animais da raça Nelore (modelo de regressão LGGE com efeito aleatório)

O conjunto de dados de animais da raça Nelore é analisado sob a perspectiva do modelo de regressão LGGE com efeito aleatório, o qual pode ser expresso da seguinte maneira

$$y_{ji} = W_j + \beta_0 + \beta_1 x_{ji1} + \beta_2 x_{ji2} + \sigma Z_{ji}, \quad i = 1, \dots, 155; j = 1, \dots, 4,$$

em que W_j é o efeito aleatório e Z_{ji} é o erro aleatório independente e identicamente distribuído.

As estimativas de máxima verossimilhança dos parâmetros do modelo LGGE com efeito aleatório encontram-se na Tabela 4. Verifica-se que tanto a variável x_{ji1} (sexo do animal) quanto a variável x_{ji2} (idade da vaca no parto) são não significativas ao nível de significância de 5%.

Tabela 4 - Estimativas de máxima verossimilhança dos parâmetros do modelo LGGE com efeito aleatório para os dados de animais da raça Nelore

Parâmetro	Estimativa	Erro-padrão	Valor P
β_0	4.9959	0.1253	< 0.0001
β_1	0.006518	0.02430	0.8059
β_2	0.003635	0.005610	0.5632
σ	0.1095	0.1148	0.4107
λ	1.0000	1.1667	0.4544
θ	11.1559	805.87	0.9898

7 CONSIDERAÇÕES FINAIS

7.1 Conclusões

Neste trabalho foram propostas duas distribuições novas baseada na distribuição gama generalizada exponenciada proposta por Cordeiro et al. (2009). A primeira é uma expansão natural da distribuição gama generalizada exponenciada, seguindo a mesma idéia de Stacy e Mihram (1965). A principal vantagem do novo modelo é a flexibilidade em acomodar diferentes formas da função de taxa de falha assim como possuir diversas distribuições muito utilizadas em análise de sobrevivência. Também foram explicitadas expressões para calcular os momentos e a função geratriz de momentos usando soma infinitas. Também foram calculadas diversas propriedades como desvio médio, confiabilidade e estatísticas de ordem desta nova distribuição.

Também foi proposto um modelo de regressão baseado na distribuição log-gama generalizada exponenciada para dados considerando censura à direita e com mecanismo aleatório. Nessa distribuição log-gama generalizada exponenciada foram calculadas os momentos e construídos os gráficos da assimetria e curtose dessa distribuição, mostrando a flexibilidade em relação aos novos parâmetros. Para estimar os parâmetros desses modelos foi utilizado o método de máxima verossimilhança e realizados testes de hipóteses utilizando a distribuição assintótica dos estimadores de máxima verossimilhança que exigem certas condições de regularidade para a função de verossimilhança.

Outro enfoque abordado foi modelo de regressão com efeito aleatório utilizando a distribuição log-gama generalizado exponenciada. Finalmente, aplicações considerando três conjuntos de dados foram desenvolvidas baseadas nos modelos propostos.

7.2 Pesquisas futuras

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

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

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

REFERÊNCIAS

- AARSET, M. V. How to identify bathtub hazard rate. **IEEE Transactions on Reliability**, New York, v. 36, p. 106-108, 1987.
- AARTS, R. M. **Lauricella functions**. Disponível em: www.mathworld.wolfram.com/LauricellaFunctions.html. From MathWorld - A Wolfram Web Resource, created by Eric W. Weisstein. Acesso em: 15 de set. 2009.
- ALBUQUERQUE, L. G.; FRIES, L. A. Precocidad: estrategias de selección. In: SIMPOSIO LATINO-AMERICANO PRODUCTIVIDAD EN GANADO DE CORTE, 1., 1998, Santa Cruz de la Sierra. **Anais...** São Paulo: DBO Sul, 1998. p. 19-24.
- ALI, M. M.; WOO, J.; NADARAJAH, S. Generalized gamma variables with drought application. **Journal of the Korean Statistical Society**, Seoul, v. 37, p. 37-45, 2008.
- BARLOW, R. E.; CAMPO, R. Total time on test processes and applications to failure data analysis. In: _____. **reliability and fault tree analysis**. Philadelphia: Society for Industrial and Applied Mathematics, 1975. p. 451-481.
- BEBBINGTON, M.; LAI, C. D.; ZITIKIS, R. A flexible Weibull extension. **Reliability Engineering and System Safety**, Barking, v. 92, n. 6, p. 719-726, 2007.
- CANCHO, V. G.; BOLFARINE, H.; ACHCAR, J. A. A Bayesian Analysis for the Exponentiated-Weibull Distribution. **Journal of Applied Statistical Science**, New York, v. 8, n. 4, p. 227-242, 1999.
- CANCHO, V. G.; ORTEGA, E. M. M.; BOLFARINE, H. The Exponentiated-Weibull Regression Models with Cure Rate. **Journal of Applied Probability and Statistics**. In press. 2008.

CARRASCO, J. M. F.; ORTEGA, E. M. M.; CORDEIRO, G. M. (2008). A generalized modified Weibull distribution for lifetime modeling. **Computational Statistics and Data Analysis**: doi 10.1016/j.CSDA.2008.08.023.

CHOUDHURY, A. A Simple Derivation of Moments of the Exponentiated Weibull Distribution. **Metrika**, New York, v. 62, p. 17-22, 2005.

COLLET, D. **Modelling survival data in medical research**. London: Chapman and Hall, 1994. 320p.

COLOSIMO, E. A.; GIOLO, S. R. **Análise de sobrevivência aplicada**. São Paulo: Edgard Blücher, 2006. 392 p.

CORDEIRO, G. M.; SILVA, G. O.; ORTEGA, E. M. M. The exponentiated generalized gamma distribution with application to lifetime data. In press. 2009.

COX, C. The generalized F distribution: an umbrella for parametric survival analysis. **Statistics in Medicine**, New York, v. 27, p. 4301-4312, 2008.

EXTON, H. **Handbook of hypergeometric integrals**: theory, applications, tables, computer programs. New York: Halsted Press. 1978.

FAREWELL, V. T.; PRENTICE, R. L. A study of distributional shape in life testing. **Technometrics**, Wisconsin, v. 19, p. 69-75, 1977.

GIOLO, S. R.; HENDERSON, R.; DEMÉTRIO, C. G. B. Um critério para seleção de touros Nelore usando modelos de sobrevivência. **Revista de Matemática e Estatística**, Jaboticabal, v. 21, n. 3, p. 115-123, 2003.

GLASER, R. E. Bathtub and related failure rate characterizations. **Journal of the American Statistical Association**, Alexandria, v. 75, p. 667-672, 1980.

GRADSHTEYN, I. S.; RYZHIK, I. M. **Table of integrals, series and products**. 6th ed. In: by JEFFREY, A.; ZWILLINGER, D. New York, 2000.

GUPTA, R. D.; KUNDU, D. Generalized exponential distributions. **Australian and New Zealand Journal of Statistics**, Oxford, v. 42, p. 173-188, 1999.

_____. Exponentiated exponential distribution: an alternative to gamma and Weibull distributions. **Biometrical Journal**, Berlin, v. 43, p. 117-130, 2001.

_____. Generalized exponential distributions: statistical inferences. **Journal of Statistical Theory and Applications**, Amsterdam, v. 1, n. 1, p. 101-118, 2002.

_____. Generalized exponential distribution: existing results and some recent developments. **Journal of Statistical Planning and Inference**, Amsterdam, v. 137, n. 11, p. 3537-3547, 2007.

GUSMÃO, F. R. S.; ORTEGA, E. M. M.; CORDEIRO, G. M. (2009). The generalized inverse Weibull distribution. **Statistical Papers**. DOI 10.1007/s00362-009-0271-3.

HAGER, H. W.; BRAIN, L. J. Inferential procedures for the generalized gamma distributions. **Journal of the American Statistical Association**, Alexandria, v. 65, p. 1601-1609, 1970.

HAUPT, E.; SCHABE, H. A new model for a lifetime distribution with bathtub shaped failure rate. **Microelectronic and Reliability**, Kidlington, v. 32, p. 633-639, 1992.

HJORTH, U. A reliability distribution with increasing, decreasing, constant and bathtub failure rates. **Technometrics**, Alexandria, v. 22, p. 99-107, 1980.

KUNDU, D.; RAKAB, M. Z. Generalized Rayleigh distribution: different methods of estimation. **Computational Statistics and Data Analysis**, Amsterdam, v. 49, p. 187-200, 2005.

LAI, C. D.; XIE, M.; MURTHY, D. N. P. A Modified Weibull Distribution. **IEEE Transactions on Reliability**, New York, v. 52, p. 33-37, 2003.

LAWLESS, J. F. **Statistical models and methods for lifetime data**. New York: John Wiley e Sons, 1944.

_____. Inference in the generalized gamma and log-gamma distributions. **Technometrics**, Wisconsin, v. 22, n. 3, p. 409-419, 1980.

_____. **Statistical models and methods for lifetime data**. 2nd ed. New York: Wiley, 2003. 630 p.

LI, H. **Survival analysis methods in genetic epidemiology**. Upenn Biostatistics Working Papers. Working Paper 7. Feb. 2006. <Disponível em: <http://biostats.bepress.com/upennbiostat/paper/art7>>. Acesso em: 04 de Sept. 2009.

LINDSEY, J. C.; RYAN, L. M. Tutorial in Biostatistics: methods for interval-censored data. **Statistics in Medicine**, Boston, v. 17, p. 219-238, 1998.

LOUZADA NETO, F. Polyhazard Models for Lifetime Data. **Biometrics**, Washington, v. 55, p. 1281-1285, 2000.

LOUZADA NETO, F.; MAZUCHELI, J.; ACHCAR, J. A. **Uma introdução à análise de sobrevivência e confiabilidade**. In: Jornadas Nacionales del Estadística, 28., 2001. Antofagasta. Anais... Antofagasta: s. ed. 2001.

MALHOTRA, J.; SHARMA, A. K.; KALER, R. S. On the performance analysis of wireless receiver using generalized-gamma fading model. **Annals of Telecommunications**, Paris, v. 64, p. 147-153, 2009.

MUDHOLKAR, G. S.; SRIVASTAVA, D. K.; FRIEMER, M. The exponentiated Weibull family: a reanalysis of the bus-motor-failure data. **American Society for Quality Control and American Statistical Association**, Alexandria, v. 37, n. 4, p. 436-445, 1995.

MUDHOLKAR, G. S.; SRIVASTAVA, D. K.; KOLLIA, G. D. A generalization of the Weibull distribution with application to the analysis of survival data. **Journal of the American Statistical Association**, Boston, v. 91, p. 1575-1583, 1996.

NADARAJAH, S. On the moments of the modified Weibull distribution. **Reliability Engineering and System Safety**, Barking, v. 90, p. 114-117, 2005.

_____. The exponentiated Gumbel distribution with climate application. **Environmetrics**, New York, v. 17, p. 13-23. 2006.

_____. On the use of the generalized gamma distribution. **International Journal of Electronic**, New York, v. 95, p. 1029-1032, 2008a.

_____. Explicit expressions for moments of order statistics. **Statistics and Probability Letters**, Amsterdam, v. 78, p. 196-205, 2008b.

NADARAJAH, S.; GUPTA, A. K. The Beta Fréchet Distribution. **Far East Journal of Theoretical Statistics**, Allahabad, v. 14, n. 1, p. 15-24, 2004.

_____. A generalized gamma distribution with application to drought data. **Mathematics and Computers in Simulation**, Elsevier, v. 74, p. 1-7, 2007.

NADARAJAH, S.; KOTZ, S. The Beta Gumbel Distribution. **Mathematical Problems in Engineering**, New York, v. 4, p. 323-332, 2004.

_____. On some recent modifications of Weibull distribution. **IEEE Transactions Reliability**, New York, v. 54, p. 561-562, 2005.

_____. The exponentiated type distributions. **Acta Applicandae Mathematicae: An International Survey Journal on Applying Mathematics and Mathematical Applications**, Amsterdam, v. 92, n. 2, p. 97-111, 2006.

NELSON, W. **Applied life data analysis**. New York: Wiley, 1982. 634 p.

ORTEGA, E. M. M. **Análise de influência e resíduos em modelos de regressão log-gama generalizados**. Tese (Doutorado em Estatística) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2001.

ORTEGA, E. M. M.; BOLFARINE, H.; PAULA, G. A. Influence diagnostics in generalized log-gamma regression models. **Computational Statistics and Data Analysis**, New York, v. 42, p. 165-186, 2003.

ORTEGA, E. M. M.; CANCHO, V. G.; BOLFARINE, H. Influence diagnostics in exponentiated-Weibull regression models with censored data. **Statistics and Operations Research Transactions**, Catalunya, v. 30, n. 2, p. 172-192, 2006.

ORTEGA, E. M. M.; PAULA, G. A.; BOLFARINE, H. Deviance residuals in generalized log-gamma regression models with censored observations. **Journal of Statistical Computation and Simulation**, New York, v. 78, p. 747-764, 2008.

ORTEGA, E. M. M.; CANCHO, V. G.; PAULA, G. A. Generalized log-gamma regression models with cure fraction. **Lifetime Data Analysis**, Dordrecht, v. 15, p. 79-106, 2009.

PHAM, H.; LAI, C. D. On Recent Generalizations of the Weibull Distribution. **IEEE Transactions on Reliability**, New York, v. 56, p. 454-458, 2007.

PRENTICE, R. L. Exponential survival with censoring and explanatory variables. **Biometrika**, London, v. 60, p. 279-288, 1973.

_____. A log-gamma model and its maximum likelihood estimation. **Biometrika**, London, v. 61, n. 3, p. 539-544, 1974.

PRUDNIKOV, A. P.; BRYCHKOV, Y. A.; MARICHEV, O. I. **Integrals and series**. Amsterdam: Gordon and Breach Science Publishers, 1986. v. 1.

RAJARSHI, S.; RAJARSHI, M. B. Bathtub distribution: a review. **Communications in Statistics: Part A, Theory and Methods**, New York, v. 17, p. 2597-2621, 1988.

RIZATO, F. B. **Modelos de regressão log-gama generalizado com fração de cura**. 73 p. 2007. Dissertação (Mestrado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2007.

- SILVA, G. O. **Modelos de regressão quando a função de taxa de falha não é monótona e o modelo probabilístico beta Weibull modificado**. Tese (Doutorado em Estatística e Experimentação Agronômica). 2008. 207 p. Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2008.
- SOUZA, W. B.; CORDEIRO, G. M.; SIMAS, A. B. Some new results on the beta Fréchet distribution. No prelo, 2008.
- STACY, E. W. A Generalization of the gamma distribution, **Annals of Mathematical Statistics**, Michigan, v. 33, p. 1187-1192, 1962.
- STACY, E. W.; MIHRAM, G. A. Parameter estimation for a generalized gamma distribution. **Technometrics**, New York, v. 7, n. 3, p. 349-358, 1965.
- STONE, G. C. **Statistical analysis of accelerated aging tests on solid electrical insulation**. 1978. Thesis (Phylosophy Doctor), University of Waterloo. Waterloo, 1978.
- TROTT, M. **The mathematica guidebook for symbolics**. New York: Springer. 2006. 1 CD-ROM.
- VALENÇA, D. M. **Teste de homogeneidade e estimação para dados de sobrevivência agrupados e com erros de medidas**. 2003. 151p. Tese (Doutorado em Estatística) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2003.
- WRIGHT, E. M. The asymptotic expansion of the generalized Bessel function. **Proccedings of the London Mathematical Society**, London, v. 38, p. 257-270, 1935.
- XIE, M.; LAI, C. D. Reliability analysis using an additive Weibull model with bathtub-shaped failure rate function. **Reliability Engineering and System Safety**, Barking, v. 52, p. 87-93, 1996.

XIE, M.; TANG, Y.; GOH, T. N. A modified Weibull extension with bathtub failure rate function. **Reliability Engineering and System Safety**, Barking, v. 76, p. 279-285, 2002.

APÊNDICES

APÊNDICE A - Elementos da matriz de informação observada $\ddot{L}(\theta)$ para a distribuição GGE

(i) para $q > 0$

$$\begin{aligned} \ddot{L}_{\alpha,\alpha} = & \frac{n\tau\kappa}{\alpha^2} - \frac{\tau}{\alpha^2} \sum_{i=1}^n (v_i + \tau v_i) - \frac{(\lambda-1)\tau}{\alpha^2\Gamma(\kappa)} \sum_{i=1}^n \left\{ \frac{(v_i)^\kappa \exp\{-v_i\}}{\gamma_1(\kappa, v_i)} - \frac{\tau\kappa(v_i)^\kappa \exp\{-v_i\}}{\gamma_1(\kappa, v_i)} \right. \\ & \left. + \frac{\tau(v_i)^{\kappa+1} \exp\{-v_i\}}{\gamma_1(\kappa, v_i)} + \frac{\tau}{\Gamma(\kappa)} \frac{[(v_i)^\kappa]^2 [\exp\{-v_i\}]^2}{[\gamma_1(\kappa, v_i)]^2} \right\} \end{aligned}$$

$$\begin{aligned} \ddot{L}_{\alpha,\tau} = & -\frac{n\kappa}{\alpha} + \frac{1}{\alpha} \sum_{i=1}^n [v_i + v_i \ln(v_i)] - \frac{(\lambda-1)}{\alpha\Gamma(\kappa)} \sum_{i=1}^n \left\{ \frac{(v_i)^\kappa \exp\{-v_i\}}{\gamma_1(\kappa, v_i)} + \frac{(v_i)^\kappa \exp\{-v_i\} \ln(v_i)^\kappa}{\gamma_1(\kappa, v_i)} \right. \\ & \left. - \frac{(v_i)^{\kappa+1} \exp\{-v_i\} \ln(v_i)}{\gamma_1(\kappa, v_i)} - \frac{1}{\Gamma(\kappa)} \frac{[(v_i)^\kappa]^2 [\exp\{-v_i\}]^2 \ln(v_i)}{[\gamma_1(\kappa, v_i)]^2} \right\} \end{aligned}$$

$$\ddot{L}_{\alpha,\kappa} = -\frac{n\tau}{\alpha} - \frac{\tau(\lambda-1)}{\alpha\Gamma(\kappa)} \sum_{i=1}^n (v_i)^\kappa \exp\{-v_i\} \left\{ \frac{\ln(v_i)}{\gamma_1(\kappa, v_i)} - \frac{\dot{\gamma}_1(\kappa, v_i)}{\Gamma(\kappa)[\gamma_1(\kappa, v_i)]^2} \right\}$$

$$\ddot{L}_{\alpha,\lambda} = -\frac{\tau}{\alpha\Gamma(\kappa)} \sum_{i=1}^n \frac{(v_i)^\kappa \exp\{-v_i\}}{\gamma_1(\kappa, v_i)}$$

$$\begin{aligned} \ddot{L}_{\tau,\tau} = & -\frac{n}{\tau^2} - \frac{1}{\tau^2} \sum_{i=1}^n v_i [\ln(v_i)]^2 + \frac{(\lambda-1)}{\tau^2\Gamma(\kappa)} \sum_{i=1}^n \left\{ \frac{(v_i)^\kappa \exp\{-v_i\} \ln(v_i)^{\kappa+1}}{\gamma_1(\kappa, v_i)} \right. \\ & \left. - \frac{(v_i)^{\kappa+1} \exp\{-v_i\} [\ln(v_i)]^2}{\gamma_1(\kappa, v_i)} - \frac{[(v_i)^\kappa]^2 [\exp\{-v_i\}]^2 [\ln(v_i)]^2}{\Gamma(\kappa)[\gamma_1(\kappa, v_i)]^2} \right\} \end{aligned}$$

$$\ddot{L}_{\tau,\kappa} = \frac{1}{\tau} \sum_{i=1}^n \ln(v_i) + \frac{(\lambda-1)}{\tau\Gamma(\kappa)} \sum_{i=1}^n (v_i)^\kappa \exp\{-v_i\} \ln(v_i) \left\{ \frac{\ln(v_i)}{\gamma_1(\kappa, v_i)} - \frac{\dot{\gamma}_1(\kappa, v_i)}{\Gamma(\kappa)[\gamma_1(\kappa, v_i)]^2} \right\}$$

$$\ddot{L}_{\tau,\lambda} = \frac{1}{\tau\Gamma(\kappa)} \sum_{i=1}^n \frac{(v_i)^\kappa \exp\{-v_i\} \ln(v_i)}{\gamma_1(\kappa, v_i)}$$

$$\ddot{L}_{\kappa,\kappa} = -n\lambda\Psi'(\kappa) + \frac{(\lambda-1)}{\Gamma(\kappa)} \sum_{i=1}^n \left\{ \frac{\ddot{\gamma}_1(\kappa, v_i)}{\gamma_1(\kappa, v_i)} - \frac{[\dot{\gamma}_1(\kappa, v_i)]^2}{\Gamma(\kappa)[\gamma_1(\kappa, v_i)]^2} \right\}$$

$$\ddot{L}_{\kappa,\lambda} = -n\Psi(\kappa) + \frac{1}{\Gamma(\kappa)} \sum_{i=1}^n \left\{ \frac{\dot{\gamma}_1(\kappa, v_i)}{\gamma_1(\kappa, v_i)} \right\}$$

$$\ddot{L}_{\lambda,\lambda} = \frac{-n}{\lambda^2}$$

em que $v_i = \left(\frac{t_i}{\alpha}\right)^\tau$, $\gamma_1(\kappa, v_i) = \int_0^{v_i} w^{\kappa-1} \exp(-w) \ln(w) dw$, $\dot{\gamma}_1(\kappa, v_i) = \int_0^{v_i} w^{\kappa-1} \exp(-w) [\ln(w)]^2 dw$, $\Psi(\cdot)$ é a função digama e $\Psi'(\cdot)$ é a derivada da função digama.

APÊNDICE B - Elementos da matriz de informação observada $\ddot{L}(\theta)$ para o modelo LGGE com dados censurados e covariáveis

(i) para $q > 0$

$$\begin{aligned} \ddot{L}_{\sigma,\sigma} = & - \sum_{i:F} \left\{ q^{-1} u_i \left[-\frac{2}{\sigma^3} + \left(\frac{v_i}{\sigma^4} \{ q u_i + 2\sigma \} \right) \right] \right\} \\ & - \sum_{i:F} (\lambda - 1) u_i \left\{ \dot{f}_z \left(-\frac{q^{-1} z_i}{\sigma} [1 - v_i] \right) [\gamma_1(q^{-2}, q^{-2} v_i)] \sigma^2 \right. \\ & \left. + \dot{f}_z [u_i \dot{f}_z - 2\sigma \gamma_1(q^{-2}, q^{-2} v_i)] \right\} ([\gamma_1(q^{-2}, q^{-2} v_i)] \sigma^2)^{-2} \\ & + \sum_{i:C} \lambda u_i \left\{ \left[-\frac{(\lambda - 1) [\gamma_1(q^{-2}, q^{-2} v_i)]^{\lambda-2} x_i^T [\dot{f}_z]^2}{\sigma} \right. \right. \\ & \left. \left. - [\gamma_1(q^{-2}, q^{-2} v_i)]^{\lambda-1} \dot{f}_z \left(\frac{q^{-1} z_i}{\sigma} [1 - v_i] \right) \right] (1 - [\gamma_1(q^{-2}, q^{-2} v_i)]^\lambda) \sigma^2 \right. \\ & \left. - [\gamma_1(q^{-2}, q^{-2} v_i)]^{\lambda-1} \dot{f}_z [2\sigma - (2\sigma [\gamma_1(q^{-2}, q^{-2} v_i)]^\lambda - \lambda [\gamma_1(q^{-2}, q^{-2} v_i)]^{\lambda-1} u_i \dot{f}_z)] \right\} \\ & [(1 - [\gamma_1(q^{-2}, q^{-2} v_i)]^\lambda) \sigma^2]^{-2} \end{aligned}$$

$$\begin{aligned} \ddot{L}_{\sigma,\beta_j} = & - \sum_{i:F} \frac{q^{-1}}{\sigma^2} \left[\frac{q x_j}{\sigma} v_i u_i - x_j (1 - v_i) \right] \\ & - \sum_{i:F} \left\{ \frac{(\lambda - 1)}{\sigma^2} \{ (-x_j \dot{f}_z + [\dot{f}_z \{ -q^{-1} x_j (1 - v_i) u_i \}]) \} [\gamma_1(q^{-2}, q^{-2} v_i)] \right. \\ & \left. - u_i \dot{f}_z \left(-\frac{x_j}{\sigma} \dot{f}_z \right) \right\} [\gamma_1(q^{-2}, q^{-2} v_i)]^{-2} \\ & + \sum_{i:C} \frac{\lambda}{\sigma^2} \left\{ -(\lambda - 1) [\gamma_1(q^{-2}, q^{-2} v_i)]^{\lambda-2} x_j z_i [\dot{f}_z]^2 \right. \\ & \left. + [\gamma_1(q^{-2}, q^{-2} v_i)]^{\lambda-1} (-x_j \dot{f}_z + [\dot{f}_z \{ -q^{-1} x_j (1 - v_i) u_i \}]) (1 - [\gamma_1(q^{-2}, q^{-2} v_i)]^\lambda) \right. \\ & \left. - [\gamma_1(q^{-2}, q^{-2} v_i)]^{\lambda-1} u_i \dot{f}_z \left(\frac{\lambda [\gamma_1(q^{-2}, q^{-2} v_i)]^{\lambda-1} x_j \dot{f}_z}{\sigma} \right) \right\} (1 - [\gamma_1(q^{-2}, q^{-2} v_i)]^\lambda)^{-2} \end{aligned}$$

$$\ddot{L}_{\sigma,\lambda} = - \sum_{i:F} \left\{ \frac{z_i \dot{f}_z}{\sigma [\gamma_1(q^{-2}, q^{-2} v_i)]} \right\} + \sum_{i:C} \left\{ \frac{\lambda [\gamma_1(q^{-2}, q^{-2} v_i)]^{\lambda-1} z_i \ln [\gamma_1(q^{-2}, q^{-2} v_i)] \dot{f}_z}{\sigma (1 - [\gamma_1(q^{-2}, q^{-2} v_i)]^\lambda)^2} \right\}$$

$$\ddot{L}_{\lambda,\lambda} = -\frac{r}{\lambda^2} - \sum_{i:C} \left\{ \frac{[\gamma_1(q^{-2}, q^{-2} v_i)]^\lambda \{ \ln [\gamma_1(q^{-2}, q^{-2} v_i)] \}^2}{\{ 1 - [\gamma_1(q^{-2}, q^{-2} v_i)]^\lambda \}^2} \right\}$$

$$\begin{aligned} \ddot{L}_{\lambda, \beta_j} = & - \sum_{i:F} \left\{ \frac{x_j \dot{f}_z}{\sigma [\gamma_1(q^{-2}, q^{-2}v_i)]} \right\} - \sum_{i:C} \left\{ \frac{x_j}{\sigma} \dot{f}_z [\gamma_1(q^{-2}, q^{-2}v_i)]^{2\lambda-1} \right. \\ & \left. - \frac{x_j}{\sigma} \dot{f}_z [\gamma_1(q^{-2}, q^{-2}v_i)]^{\lambda-1} (1 + \lambda \ln[\gamma_1(q^{-2}, q^{-2}v_i)]) \right\} (1 - [\gamma_1(q^{-2}, q^{-2}v_i)]^\lambda)^{-2} \end{aligned}$$

$$\begin{aligned} \ddot{L}_{\beta_j, \beta_s} = & - \sum_{i:F} \left\{ \frac{x_j x_s v_i}{\sigma^2} \right\} + \sum_{i:F} \left\{ \frac{(\lambda-1)x_j}{\sigma} \left[\dot{f}_z [\gamma_1(q^{-2}, q^{-2}v_i)] \left(\frac{q^{-1}x_s}{\sigma} [1-v_i] \right) \right. \right. \\ & \left. \left. + \frac{x_s}{\sigma} [\dot{f}_z]^2 \right] [\gamma_1(q^{-2}, q^{-2}v_i)]^{-2} \right\} - \sum_{i:C} \left\{ \frac{\lambda x_j}{\sigma} \left[\left(\frac{(\lambda-1)x_s [\gamma_1(q^{-2}, q^{-2}v_i)]^{\lambda-2} [\dot{f}_z]^2}{\sigma} \right) \right. \right. \\ & \left. \left. + [\gamma_1(q^{-2}, q^{-2}v_i)]^{\lambda-1} \dot{f}_z \left[\frac{q^{-1}x_s}{\sigma} [1-v_i] \right] \right] (1 - [\gamma_1(q^{-2}, q^{-2}v_i)]^\lambda) \right. \\ & \left. + \frac{\lambda x_s [\gamma_1(q^{-2}, q^{-2}v_i)]^{2(\lambda-1)} [\dot{f}_z]^2}{\sigma} \right\} (1 - [\gamma_1(q^{-2}, q^{-2}v_i)]^\lambda)^{-2} \end{aligned}$$

em que $z_i = \frac{u_i}{\sigma}$, $u_i = y_i - x_i^T \beta_j$, $v_i = \exp(qz_i)$ e $\dot{f}_z = \frac{q(q^{-2})^{q^{-2}}}{\Gamma(q^{-2})} \exp\{q^{-1}z_i - q^{-2}v_i\}$.

APÊNDICE C

Uma expansão para $\gamma_1(k, x)^{\lambda-1}$ para todo $\lambda > 0$, não-inteiro, pode ser escrito como

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

que sempre converge quando $0 < \gamma_1(k, x) < 1$. Então,

$$\gamma_1(k, x)^{\lambda-1} = \sum_{j=0}^{\infty} \sum_{m=0}^j (-1)^{j+m} \binom{\lambda-1}{j} \binom{j}{m} \gamma_1(k, x)^m. \quad (86)$$

Substituindo $\sum_{j=0}^{\infty} \sum_{m=0}^j$ por $\sum_{m=0}^{\infty} \sum_{j=m}^{\infty}$ obtêm

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

Então

$$\gamma_1(k, x)^{\lambda-1} = \sum_{m=0}^{\infty} s_m(\lambda) \gamma_1(k, x)^m, \quad (87)$$

onde

$$s_m(\lambda) = \sum_{j=m}^{\infty} (-1)^{j+m} \binom{\lambda-1}{j} \binom{j}{m}. \quad (88)$$

Usando agora uma expansão em série para a função gama incompleta dada por

$$\gamma_1(k, x) = \frac{x^k}{\Gamma(k)} \sum_{i=0}^{\infty} \frac{(-x)^i}{(k+i)!}. \quad (89)$$

Através da aplicação de uma equação na Seção 0.314 de Gradshteyn e Ryzhik (2000) para séries de potências elevadas a potências, obtêm para qualquer m inteiro positivo

$$\left(\sum_{i=0}^{\infty} a_i x^i \right)^m = \sum_{i=0}^{\infty} c_{m,i} x^i, \quad (90)$$

onde os coeficientes $c_{m,i}$ (para $i = 1, 2, \dots$) satisfazem a relação de recorrência

$$c_{m,i} = (ia_0)^{-1} \sum_{p=1}^i (mp - i + p) a_p c_{m,i-p} \quad (91)$$

onde $c_{m,0} = a_0^m$ e $a_p = (-1)^p/(k+p)p!$. O coeficiente $c_{m,i}$ pode ser calculado a partir de $c_{m,0}, \dots, c_{m,i-1}$, e pode ser escrito explicitamente em função das quantidades a_0, \dots, a_i . Além disso, usando a equação (90), obtêm

$$\gamma_1(k, x)^m = \frac{x^{km}}{\Gamma(k)^m} \sum_{i=0}^{\infty} c_{m,i} x^i, \quad (92)$$

onde os coeficientes $c_{m,i}$ são obtidos pela equação (91). Combinando (87) e (92), pode-se reescrever (87) como

$$\gamma_1(k, x)^{\lambda-1} = \sum_{m,i=0}^{\infty} \frac{s_m(\lambda) c_{m,i}}{\Gamma(k)^m} x^{km+i}. \quad (93)$$

ANEXOS

ANEXO A - Programa no SAS usando proc NLMIXED para estimar os parâmetros do modelo para os dados de níveis de voltagem, seus respectivos erros-padrão (entre parênteses) e as estatísticas AIC, AICC e BIC

```

data dados4;
input t censur;
cards;
4690 1
740 1
1010 1
1190 1
2450 1
:
444 1
390 1
168 1
558 1
288 1
;
proc print data=dados4;
run;

# ESTIMAÇÃO EXPONENCIAL
proc nlmixed cov data=dados4;
parms alfa=6.0193;
bounds alfa>0;
logp=censur*(-log(alfa)-(t/alfa))+(1-censur)*(-(t/alfa));
model t ~ general(logp);
run;

# ESTIMAÇÃO WEIBULL
proc nlmixed cov data=dados4;
parms alfa=1154.91 beta=1;
bounds alfa>0;
bounds beta>0;
logp=censur*(log(beta)-beta*log(alfa)+(beta-1)*log(t)-(t/alfa)**beta)
+(1-censur)*(-(t/alfa)**beta);
model t ~ general(logp);
run;

# ESTIMAÇÃO GAMA GENERALIZADA
proc nlmixed cov data=dados4;
parms alfa=1154.20 beta=0.9512 k=10;
bounds alfa>0;
bounds beta>0;
bounds k>0;
logp=censur*(log(beta)-log(GAMMA(k))-(beta*k)*log(alfa)
+(beta*k-1)*log(t)-(t/alfa)**beta)+(1-censur)*(log(1-probgam((t/alfa)**beta,k)));
model t ~ general(logp);
run;

```

```
# ESTIMAÇÃO GAMA GENERALIZADA EXPONENCIADA
```

```
proc nlmixed cov data=dados4;
parms alfa=1154.19 beta=0.9360 k=1.0435 lambda=35;
bounds alfa>0;
bounds beta>0;
bounds k>0;
bounds lambda>0;
logp=censur*(log(lambda)+log(beta)-log(GAMMA(k))-log(alfa)
+(beta*k-1)*log(t)-(beta*k-1)*log(alfa)-(t/alfa)**beta
+(lambda-1)*log(probgam((t/alfa)**beta,k)))
+(1-censur)*(log(1-(probgam((t/alfa)**beta,k)**lambda)));
model t ~ general(logp);
run;
```

```
# ESTIMAÇÃO EXPONENCIAL EXPONENCIADA
```

```
proc nlmixed cov data=dados4;
parms alfa=1154.19 lambda=5;
bounds alfa>0;
bounds lambda>0;
logp=censur*(log(lambda)-log(alfa)-(t/alfa)+(lambda-1)*log(1-exp(-(t/alfa))))
+(1-censur)*(log(1-(1-exp(-(t/alfa)))**lambda)));
model t ~ general(logp);
run;
```

```
# ESTIMAÇÃO WEIBULL EXPONENCIADA
```

```
proc nlmixed cov data=dados4;
parms alfa=1154.19 beta=0.9360 lambda=75;
bounds alfa>0;
bounds beta>0;
bounds lambda>0;
logp=censur*(log(lambda)+log(beta)-log(alfa)+(beta-1)*log(t/alfa)
-(t/alfa)**(beta)+(lambda-1)*log(1-exp(-(t/alfa)**(beta))))
+(1-censur)*(log(1-(1-exp(-(t/alfa)**(beta)))**lambda)));
model t ~ general(logp);
run;
```

ANEXO B - Programa no SAS usando proc NLMIXED para estimar os parâmetros do modelo LGGE com efeito aleatório para os dados de animais da raça Nelore

```

data touros;
input tempo grupo delta inter sexo idade;
y=log(tempo);
cards;
198 1 0 1 0 13
176 1 1 1 1 12
184 1 0 1 0 9
:
199 4 0 1 0 6
148 4 1 1 0 6
204 4 0 1 0 5
;
proc print data=touros;
run;

# VARIÁVEIS SEXO DO ANIMAL E IDADE DA VACA NO PARTO
proc nlmixed data=touros method=GAUSS NOAD fd qpoints=10;
parms b0=4.9959 b1=0.0065 b2=0.0036 sigma=0.1095 lambda=3 teta=5;
bounds sigma>0;
bounds lambda>0;
bounds teta>0;
q=-4.7888;
pi=CDF("NORMAL",u);
if (pi>0.999999) then pi=0.999999;
gi2=quantile("GAMMA",pi,1/teta);
gi=teta*gi2;
eta=b0+b1*sexo+b2*idade+(gi);
z=(y-eta)/sigma;
A=(-lambda*q)/(sigma*gamma(q**(-2)));
B=(q**(-2))**(q**(-2));
C=exp(q**(-2)*(q*z-exp(q*z)));
D=(probgam((q**(-2))*exp(q*z),q**(-2))**(lambda-1));
E=(probgam((q**(-2))*exp(q*z),q**(-2))**(lambda));
logp=delta*(log(A*B*C*D))+(1-delta)*(log(E));
model y ~ general(logp);
random u ~ normal(0,1) subject=grupo;
run;

```


ANEXO C - Dados utilizados no estudo de níveis de voltagem

(continua)

Tempos de falha (minutos)	Censura	Voltagem (kV)
4690	1	52.5
740	1	52.5
1010	1	52.5
1190	1	52.5
2450	1	52.5
1390	1	52.5
350	1	52.5
6095	1	52.5
3000	1	52.5
1458	1	52.5
6200	0	52.5
550	1	52.5
1690	1	52.5
745	1	52.5
1225	1	52.5
1480	1	52.5
245	1	52.5
600	1	52.5
246	1	52.5
1805	1	52.5
258	1	55.0
114	1	55.0
312	1	55.0
772	1	55.0
498	1	55.0
162	1	55.0
444	1	55.0
1464	1	55.0

ANEXO C - Dados utilizados no estudo de níveis de voltagem

(continuação)

Tempos de falha (minutos)	Censura	Voltagem (kV)
132	1	55.0
1740	0	55.0
1266	1	55.0
300	1	55.0
2440	0	55.0
520	1	55.0
1240	1	55.0
2600	0	55.0
222	1	55.0
144	1	55.0
745	1	55.0
396	1	55.0
510	1	57.5
1000	0	57.5
252	1	57.5
408	1	57.5
528	1	57.5
690	1	57.5
900	0	57.5
714	1	57.5
348	1	57.5
546	1	57.5
174	1	57.5
696	1	57.5
294	1	57.5
234	1	57.5
288	1	57.5
444	1	57.5

ANEXO C - Dados utilizados no estudo de níveis de voltagem

(conclusão)

Tempos de falha (minutos)	Censura	Voltagem (kV)
390	1	57.5
168	1	57.5
558	1	57.5
288	1	57.5