JULIANO BORTOLINI

DISTRIBUIÇÃO GAMA GENERALIZADA GEOMÉTRICA ESTENDIDA

LAVRAS - MG
2015

JULIANO BORTOLINI

DISTRIBUIÇÃO GAMA GENERALIZADA GEOMÉTRICA ESTENDIDA

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, Área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

Orientador
Prof. Dr. Renato Ribeiro de Lima

Coorientador
Prof. Dr. Marcelino Alves Rosa de Pascoa

LAVRAS - MG
2015

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Bortolini, Juliano.
Distribuição gama generalizada geométrica estendida / Juliano Bortolini. – Lavras: UFLA, 2015.
133 p. : il.

Tese(doutorado)–Universidade Federal de Lavras, 2015.
Orientador: Renato Ribeiro de Lima.
Bibliografia.

1. Análise de Sobrevivência. 2. Distribuição de Probabilidade. 3. Regressão Locação Escala. 4. Bimodal. I. Universidade Federal de Lavras. II. Título.

JULIANO BORTOLINI

DISTRIBUIÇÃO GAMA GENERALIZADA GEOMÉTRICA ESTENDIDA

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, Área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

APROVADA em 3 de dezembro de 2015.

Prof. Dr. Anderson Castro Soares de Oliveira	UFMT
Profa. Dra. Graziela Dutra Rocha Gouvêa	UFOP
Prof. Dr. Lucas Monteiro Chaves	UFLA
Prof. Dr. Mário Javier Ferrua Vivanco	UFLA

Prof. Dr. Renato Ribeiro de Lima
Orientador

Prof. Dr. Marcelino Alves Rosa de Pascoa
Coorientador

LAVRAS - MG
2015

A meus pais Leda e Jesus,

pelo amor, carinho e educação.

A minha irmã Rafaela,

pelo amor, carinho, educação e travessuras.

DEDICO

AGRADECIMENTOS

A Deus, pelo sol que ilumina os meus dias, pelos corpos celestes que clareiam as minhas noites mais escuras, pelas chuvas que lavam a minha alma, pelas primaveras que colorem a minha vida com as suas flores, pelos pássaros que alegram o meu espírito com os seus cantos, pelos alimentos que fortificam o meu corpo, pela verdade que aos poucos me liberta, pelo trabalho que me faz sentir útil, pela vida que me proporciona novas experiências e oportunidades de crescimento, e por tudo que ainda não tenho consciência que fazes por mim.

À minha querida mãe Marileda e ao meu querido pai José Jesus, pelo amor e dedicação na boa educação, em especial por me ensinarem a como deitar e conseguir dormir em paz.

À minha querida irmã Rafaela, pela companhia fraterna, desde sempre, me alegrando e educando.

Aos meus familiares, amigos e minha namorada Thayza, pelo apoio, carinho e compreensão.

Ao meu orientador Prof. Dr. Renato Ribeito de Lima e coorientador Prof. Dr. Marcelino Alves Rosa de Pascoa, pelos conhecimentos e esclarecimentos confiados a mim, pela paciência e compreensão das minhas dificuldades e por me aceitarem como orientando.

Aos Profs. Drs. Anderson Castro Soares de Oliveira, Lucas Monteiro Chaves, Mário Javier Ferrua Vivanco e à Profa. Dra. Graziela Dutra Rocha Gouvêa pelas importantes contribuições para esta tese.

À Universidade Federal de Lavras (UFLA) e ao Departamento de Ciências Exatas (DEX), pela oportunidade de cursar o doutorado.

Aos Profs. Drs. do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária da UFLA, em especial aos Profs. Drs. Augusto Ramalho de Morais, Daniel Furtado Ferreira, João Domingos Scalon, Joel Augusto Muniz, Júlio Silvio de Sousa Bueno Filho, Lucas Monteiro Chaves, Marcelo Ângelo Cirillo, Marcelo Silva de Oliveira, Mário Javier Ferrua Vivanco, Paulo César Lima, Renato Ribeiro de Lima, Solange Gomes Faria Martins e Thelma Sáfadi, pelas relevantes contribuições em suas disciplinas.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão de bolsa de estudos no primeiro ano de doutorado.

Aos colegas de turma Adriana, Cristiane, Danielle, Elayne, Gilberto, Guido Gustavo, Juscelia, Larissa, Naje, Tábata, Tales, Thalita e Wederson, pela companhia nos estudos e nos momentos de descontração.

A todos os demais amigos da estatística, em especial à Adriele, Andrezza, Carolina, Cristina, Flávia, Isabel, Leila, Maíra, Marcília, Mariele, Suzana, Tatiane, Vanessa, ao Allan, Ben Deivide, aos Carlos, Danilo, Diogo, Edmundo, Felipe, Fernando, aos Jair, Joel, Juracy, Leandro, Luciano, Manoel, Rossicley e Walter, pela amizade e pelos momentos de lazer.

Aos companheiros de república André, Carlos, Gilberto e Guido Gustavo, pela companhia e por sempre me receberem tão bem.

A todos os colegas professores do Departamento de Estatística da Universidade Federal de Mato Grosso, Ana Paula, Anderson, Édila, Eveliny, Fabiane, Gilmar, Hudson, José Nilton, José Roberto, Juscelia, Kuang Hongyu, Lia, Luiz André (Zardo), Marcelino, Márcia, Mariano, Neuber, Sílvia, Tábata e Vera, e técnicos Roseane e Sandro, por ajudarem (e muito!) no desenvolvimento deste trabalho e pelos agradáveis momentos.

Aos amigos da Campanha do Quilo, dos Centros Espíritas Allan Kardec e Augusto Silva, por confiarem o trabalho a mim.

A todos que de alguma forma contribuíram, mesmo sem saberem, para a realização deste trabalho.

“A ciência incha,
mas o AMOR edifica.”

Paulo de Tarso

“A ciência pode estar cheia de poder, mas só o amor beneficia. [...] A ciência pode concretizar muitas obras úteis, mas só o amor institui as obras mais altas. Não duvidamos de que a primeira, bem interpretada, possa dotar o homem de um coração corajoso; entretanto, somente o segundo pode dar um coração iluminado.”

Emmanuel

RESUMO GERAL

Novas distribuições de probabilidade são propostas com o objetivo de obter melhores ajustes a dados que apresentem comportamentos mais complexos, tais como os que estão suscetíveis a censuras. Nesta perspectiva, este trabalho propôs novos modelos mais flexíveis para a análise de sobrevivência. O primeiro exposto é a distribuição gama generalizada geométrica estendida de cinco parâmetros, que inclui importantes distribuições como casos particulares, tal como a gama generalizada. Para essa nova distribuição, obteve-se uma expressão para os momentos, função geradora de momentos, função densidade da distribuição de estatística de ordem, desvios médios e confiabilidade. Examinaram-se os estimadores de máxima verossimilhança dos parâmetros e calculou-se a matriz de informação observada. Em sequência, realizou-se uma sutil generalização da distribuição já proposta e a transformação logarítmica, proporcionando o desenvolvimento de um modelo de regressão paramétrico. A utilidade dos novos modelos propostos são ilustrados com uma aplicação a um conjunto de dados de tempo de permanência de imigrantes brasileiros no Japão. Para o conjunto de dados analisado, as estatísticas AIC, BIC e CAIC mostraram que os novos modelos são mais adequados do que outros disponíveis na literatura.

Palavras-chave: Análise de sobrevivência. Regressão locação-escala. Estimação de máxima verossimilhança. Inferência Bayesiana. Bimodal.

GENERAL ABSTRACT

New probability distributions are proposed in order to get better fit to the complex data such as censored, skewed and bimodal. In this perspective, this work proposed new more flexible models for survival analysis. The first model proposed is the extended generalized gamma geometric distribution of five parameters, which includes well-known lifetime special sub-models such as the generalized gamma. We provided a mathematical treatment of the new distribution including explicit expressions for moments, moment generating function, mean deviations, reliability and order statistics. Further, we developed an extension of this distribution by assuming that a shape parameter can take negative values. Additionally, we derived the log-transformed distribution and its regression model. The new regression model represents a parametric family of models that includes as sub-models some widely known regression models that can be applied to censored survival data. Finally, an application of the new models to real data showed that they could provide a better fit than other statistical models frequently used in lifetime data analysis.

Keywords: Survival analysis. Location-scale regression model. Maximum likelihood estimation. Bayesian inference. Bimodal.

SUMÁRIO

PRIMEIRA PARTE

1 INTRODUÇÃO

De um modo geral, qualquer distribuição de probabilidade definida em um conjunto real positivo pode ser considerada como um modelo para o tempo até a ocorrência de um evento de interesse. No entanto, nem todas as distribuições são adequadas para descrever um fenômeno específico de envelhecimento, principalmente quando há a presença de dados censurados, isto é, a informação parcial da resposta.

Em muitas situações práticas, as distribuições usuais, bem como a normal e t-Student, não são adequadas. Dessa forma, a falta de modelos mais flexíveis para a análise de dados de sobrevivência, tais como assimétricos, platicúrticos, leptocúrticos e bimodais, estimulou o desenvolvimento de novas distribuições de probabilidade. Esses novos modelos são úteis em aplicações em diversas áreas, tais como a medicina, biologia, saúde pública, epidemiologia, engenharia, economia, estudos demográficos entre outras.

Na obtenção de novas distribuições, diversos métodos podem ser considerados. Entre os métodos existentes, destacam-se os que envolvem transformações na função de distribuição, resultando em modelos mais gerais, a adição de parâmetros de locação e escala, proporcionando um modelo mais flexível e a mistura de distribuições.

Entre as distribuições já conhecidas, utilizadas em análise de sobrevivência, evidenciam-se a exponencial, gama, gama generalizada, Weibull, log-normal e log-logística. Por isso, novas generalizações são obtidas preferencialmente a partir destas, sob a perspectiva de obter novas distribuições que modelam melhor os comportamentos mais complexos, tais como os que envolvem dados censurados. Desta forma, definiu-se como objetivo geral desta pesquisa, a proposição de novas distribuições de probabilidade mais flexíveis, que se ajustam a dados bimodais, assimétricos e para valores de curtose diferente de zero. Para o seu desenvolvimento foram fixados três objetivos específicos.

O primeiro refere-se à caracterização da distribuição gama generalizada geométrica estendida, ou seja, a sua definição e o cálculo de algumas propriedades, tais como a expansão da função densidade, momentos, função geradora de momentos, desvios médios, confiabilidade, estatísticas de ordem, função de verossimilhança e densidades *a posteriori* e, em seguida, uma aplicação real ilustrando a vantagem do novo modelo para dados bimodais em relação aos seus sub-modelos.

Na sequência, propôs-se apresentar uma generalização da distribuição gama generalizada geométrica estendida e algumas de suas propriedades, quais sejam: a expansão da função densidade, momentos, função geradora de momentos, desvios médios, confiabilidade e estatísticas de ordem.

O terceiro objetivo destina-se a obter a distribuição log-gama generalizada geométrica estendida, o cálculo de seus momentos e o seu modelo de regressão, também com uma aplicação.

Para o desenvolvimento do tema suscitado, o plano de trabalho foi estruturado em duas partes. A primeira é composta por dois capítulos; e a segunda por dois artigos referentes aos resultados da pesquisa, que correspondem aos três objetivos específicos.

Desse modo, o capítulo primeiro destina-se à introdução desta tese, no qual é delimitado o assunto de estudo e são especificados os objetivos da pesquisa.

O segundo capítulo dedica-se ao referencial teórico, no qual são expostos conceitos básicos em análise de sobrevivência, as equivalências entre as funções do tempo de sobrevivência, a função de verossimilhança em análise de sobrevivência, as distribuições gama generalizada e gama generalizada geométrica, classes de generalizações de distribuições, o modelo de regressão paramétrico locação-escala e procedimentos de inferência estatística.

Na sequência, tem-se o início da segunda parte, a qual está reservada a dois artigos e às considerações gerais do trabalho.

O primeiro artigo, referente ao desenvolvimento da distribuição gama generalizada geométrica estendida, está submetido à revista *Hacettepe Journal of Mathematics and Statistics* e está na fase de avaliação pelos revisores.

O segundo artigo, que reporta-se ao cumprimento do segundo e terceiro objetivos específicos, será submetido à mesma revista, após a publicação do primeiro artigo.

Por derradeiro, nas considerações gerais são apresentadas as principais conclusões do trabalho e algumas sugestões para pesquisas futuras.

2 REFERENCIAL TEÓRICO

Nesta seção serão abordados os conceitos básicos e essenciais sobre análise de sobrevivência, a distribuição gama generalizada geométrica, a classe de distribuições estendidas, modelos de regressão locação-escala e métodos inferenciais.

2.1 Conceitos básicos em análise de sobrevivência

A análise de sobrevivência é caracterizada pelo fato de que a variável resposta é, geralmente, o tempo até a ocorrência de um evento de interesse. O evento em estudo é denominado falha e o tempo até a ocorrência da falha é definido por tempo de falha, ou de sobrevivência, ou de sobrevida. Outra característica da análise de sobrevivência é a presença de dados censurados, que é a informação parcial do tempo de falha (COLOSIMO e GIOLO, 2006).

Destacam-se três tipos de censuras: tipo I, quando finaliza-se um tempo pré-estabelecido para o estudo e nem todos os elementos falharam. Neste caso, o tempo até a ocorrência da falha só é conhecido se ela ocorrer antes do final do estudo; tipo II, se um número pré-estabelecido de falhas ocorrem e encerra-se o estudo; tipo III, ou censura aleatória, caso o elemento é retirado do estudo por uma causa alheia ao próprio estudo. Os tipos I, II e III de censura estão ilustrados na Figura 1.

As censuras do tipo I, II e III são conhecidas também como censuras à direita, pois a falha está à direita do tempo registrado. Além dessa, pode ocorrer a censura à esquerda, se o tempo registrado é maior que o tempo de falha, ou seja, o evento de interesse ocorreu antes do elemento ser observado, e a censura intervalar, quando não se pode afirmar o tempo exato de falha, sabe-se apenas que ela ocorreu em um determinado intervalo.

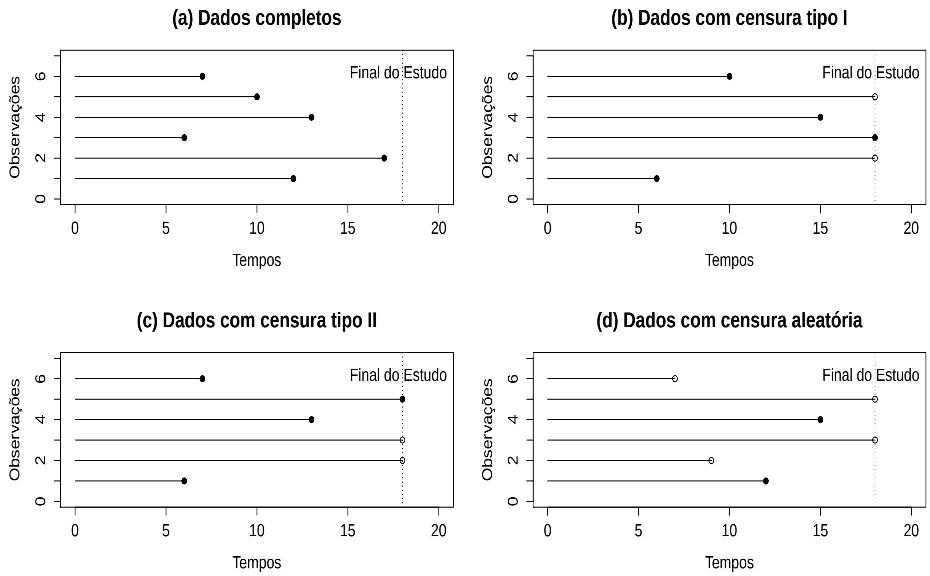

Figura 1 Ilustração de alguns mecanismos de censura em que • representa que o elemento do estudo falhou e ○ que o dado foi censurado. (a) todos os elementos falharam antes do final do estudo, (b) alguns elementos não falharam até o final do estudo, (c) o estudo foi finalizado após a ocorrência de um número pré-estabelecido de falhas e (d) o acompanhamento de alguns elementos foi interrompido por alguma razão e alguns elementos não experimentaram o evento de interesse até o final do estudo. Fonte: adaptado de Colosimo e Giolo (2006).

Considerando o mecanismo de censura aleatória e as variáveis aleatórias não negativas e independentes T e C, que representam, respectivamente, o tempo de falha e o tempo de censura para um elemento em estudo, os dados observados podem ser representados por $t = \min\{T,C\}$ e δ, o indicador de falha ou censura, definido por

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

sendo $\delta = 1$ representando a ocorrência de falha, e $\delta = 0$ de censura.

Em geral, para cada indivíduo i, $i = 1, \cdots, n$, as observações são representadas pelo par (t_i, δ_i), sendo t_i o tempo observado de falha ou censura e δ_i uma variável indicadora assumindo $\delta_i = 1$ se o tempo corresponde a falha ou $\delta_i = 0$ para censura. Quando para cada indivíduo estiver associado um vetor de covariáveis $\mathbf{x}_i = (x_{i1}, \cdots, x_{il})$, os dados podem ser representados por $(t_i, \delta_i, \mathbf{x}_i)$.

A variável aleatória não negativa T, que representa o tempo de falha, pode ser especificada pela função densidade de probabilidade, ou função de distribuição acumulada, ou função de sobrevivência, ou função risco, sendo essas quatro funções matematicamente equivalentes. Assim, especificando qualquer uma dessas funções obtém-se as outras.

2.2 Funções do tempo de sobrevivência

Seja T uma variável aleatória contínua não negativa, de função densidade de probabilidade $f(t)$ e função de distribuição acumulada

$$F(t) = P(T \leq t) = \int_0^t f(x)dx.$$

A probabilidade de um indivíduo sobreviver até o tempo t é dada pela função de sobrevivência

$$S(t) = 1 - F(t) = P(T > t).$$

Note que a função de sobrevivência $S(t)$ é uma função monótona decrescente com $S(0) = 1$ e $\lim_{t \to \infty} S(t) = 0$.

A probabilidade da falha ocorrer em um intervalo específico $[t_1, t_2)$ pode ser expressa em termos da função de sobrevivência como $S(t_1) - S(t_2)$. E a taxa de falha no intervalo $[t_1, t_2)$ é definida como a razão da probabilidade que a falha ocorra neste intervalo, dado que não ocorreu antes de t_1, pelo comprimento do intervalo. Ou seja, a taxa de falha no intervalo $[t_1, t_2)$ é calculada por (COLOSIMO; GIOLO, 2006)

$$\frac{P(t_1 \leq T < t_2|T \geq t_1)}{(t_2 - t_1)} = \frac{S(t_1) - S(t_2)}{(t_2 - t_1)S(t_1)}.$$

Redefinindo o intervalo $[t_1,t_2)$ como $[t, t + \Delta t)$, e considerando Δt bem pequeno, define-se a função taxa de falha, ou função risco, por

$$h\left(t\right) = \lim_{\Delta t \to 0} \frac{P\left(t \leq T < t + \Delta t|T \geq t\right)}{\Delta t} = \frac{f\left(t\right)}{S\left(t\right)}.$$

Esta função é bastante útil para descrever a distribuição do tempo de vida, ela descreve a forma em que a taxa instantânea de falha muda com o tempo.

Segundo Colosimo e Giolo (2006), a função risco é mais informativa que a função de sobrevivência, pois diferentes funções de sobrevivência podem ter formas semelhantes, entretanto suas respectivas funções risco podem diferir drasticamente.

As funções $f(t)$, $F(t)$, $S(t)$ e $h(t)$ são associadas matematicamente, ou seja, especificada qualquer uma delas as outras podem ser obtidas, conforme já citado.

Expressões para $f(t)$ e $S(t)$ podem ser obtidas por $h(t)$. Do fato que $f(t) = -\frac{d}{dt}S(t)$, tem-se, então

$$h\left(t\right) = -\frac{d}{dt}\log\left(S(t)\right).$$

Agora, integrando ambos os lados, e então exponencializando, obtém-se

$$S(t) = \exp\left(-\int_0^t h(x)dx\right).$$

Finalmente, a partir das duas últimas expressões, a função densidade pode ser obtida por

$$f(t) = h(t)\exp\left(-\int_0^t h(x)dx\right).$$

2.3 A função de verossimilhança em análise de sobrevivência

A inferência estatística para modelos paramétricos em análise de sobrevivência pode ser feita pelo método da máxima verossimilhança.

Na obtenção da função de verossimilhança será considerado o mecanismo de censura aleatória, definido na seção 2.1, e não informativa. Conforme Allison (2010), censura não informativa pode ser entendida como: se a observação de um indivíduo é censurada em um tempo c, então esse indivíduo pode ser representado por todos os outros que sobreviveram no tempo c e que apresentam os mesmos valores para as covariáveis. Kalbfleish e Prentice (2002) definem matematicamente censura não informativa como a situação em que a distribuição do tempo de censura não depende dos parâmetros da distribuição do tempo de falha.

Considerando T o tempo de falha e C o de censura, com T e C variáveis aleatórias contínuas independentes. Para $i = 1, \cdots, n$, os dados observados consistem dos pares (t_i, δ_i), em que t_i é uma realização de $\mathcal{T}_i = \min\{T, C\}$ e $\delta_i = 1$ se $T_i \leq C_i$ ou $\delta_i = 0$ se $T_i > C_i$ (LAWLESS, 2003). Todos os tempos de falhas e censuras são independentes e a distribuição de C não depende dos parâmetros $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)$ da distribuição de T. Sejam $f(t;\boldsymbol{\theta})$, $S(t;\boldsymbol{\theta})$, $f_C(t)$ e $S_C(t)$ as funções densidade e sobrevivência de T e C, respectivamente. Então, para a

$i-$ésima observação, tem-se que

$$\begin{aligned} P(\mathcal{T}_i = t_i, \delta_i = 0) &= P(C_i = t_i, T_i > C_i) = P(C_i = t_i, T_i > t_i) \\ &= f_{C_i}(t_i) S(t_i; \boldsymbol{\theta}) \end{aligned}$$

ou

$$\begin{aligned} P(\mathcal{T}_i = t_i, \delta_i = 1) &= P(T_i = t_i, T_i \leq C_i) = P(T_i = t_i, C_i > t_i) \\ &= f(t_i; \boldsymbol{\theta}) S_{C_i}(t_i). \end{aligned}$$

Essas probabilidades podem ser expressas em uma úncia expressão como

$$P(\mathcal{T}_i = t_i, \delta_i) = \left[f(t_i; \boldsymbol{\theta}) S_{C_i}(t_i)\right]^{\delta_i} \left[f_{C_i}(t_i) S(t_i; \boldsymbol{\theta})\right]^{1-\delta_i},$$

e a função densidade conjunta de $(\mathcal{T}_i, \delta_i)$, $i = 1, \cdots, n$, é dada por

$$\prod_{i=1}^{n} \left[f(t_i; \boldsymbol{\theta}) S_{C_i}(t_i)\right]^{\delta_i} \left[f_{C_i}(t_i) S(t_i; \boldsymbol{\theta})\right]^{1-\delta_i}.$$

Sendo assim, a função de verossimilhança de $\boldsymbol{\theta}$ é dada por

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f_{C_i}(t_i)^{1-\delta_i} S_{C_i}(t_i)^{\delta_i} \prod_{i=1}^{n} f(t_i; \boldsymbol{\theta})^{\delta_i} S(t_i; \boldsymbol{\theta})^{1-\delta_i}.$$

Como $f_{C_i}(t_i)$ e $S_{C_i}(t_i)$, $i = 1, \cdots, n$, não dependem de $\boldsymbol{\theta}$, a função de verossimilhança de $\boldsymbol{\theta}$ pode ser reescrita como (KALBFLEISH e PRENTICE, 2002)

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

sendo $\boldsymbol{\theta}$ o vetor de dimensões $p \times 1$ de parâmetros desconhecidos, $f(t_i; \boldsymbol{\theta})$ e $S(t_i; \boldsymbol{\theta})$ as funções densidade de probabilidade e de sobrevivência da variável aleatória T_i, respectivamente.

O logarítmo da função de verossimilhança, ou simplesmente a função log-verossimilhança, é dado por $\ell\left(\boldsymbol{\theta}\right) = \log L\left(\boldsymbol{\theta}\right)$. O vetor de primeiras derivadas $\mathbf{U}\left(\boldsymbol{\theta}\right) = \partial \ell\left(\boldsymbol{\theta}\right) / \partial \boldsymbol{\theta}$, de dimensões $p \times 1$, é denominado de vetor escore.

2.4 Distribuição Gama Generalizada

A distribuição gama generalizada (GG), introduzida por Stacy (1962), é uma generalização da distribuição gama e tem sido bastante discutida na literatura. Como exemplos podem ser citados os trabalhos de Nadarajah e Gupta (2007) que usaram a distribuição com aplicações em dados de seca, Ali, Woo e Nadarajah (2008) que derivaram a distribuição exata do produto de duas variáveis aleatórias independentes GG, Cordeiro, Ortega e Silva (2011) que propuseram a distribuição gama generalizada exponenciada e Ortega, Cordeiro e Pascoa (2011) que desenvolveram a distribuição gama generalizada geométrica, apresentada em detalhes na subseção 2.5.

A função densidade de probabilidade de uma variável aleatória T com distribuição gama generalizada, de parâmetros α, τ e k, é dada por

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

sendo $\alpha > 0$ parâmetro de escala, $\tau > 0$ e $k > 0$ parâmetros de forma e $\Gamma(k)$ a função matemática gama, definida por $\Gamma(k) = \int_0^\infty w^{k-1} \exp(-w) dw$.

A distribuição GG contém alguns modelos bastante conhecidos como casos particulares, tais como a distribuição exponencial ($\tau = k = 1$), gama ($\tau = 1$), Weibull ($k = 1$), Rayleigh ($\alpha = \sigma\sqrt{2}$, $\tau = 2$ e $k = 1$), semi-normal ($\alpha =$

$\sigma\sqrt{\pi}, \tau = 2, k = 1/2)$, entre outros.

Vários trabalhos estudaram as propriedades da distribuição GG, tais como os de Stacy e Mihram (1965), Prentice (1974), Farewal e Prentice (1977), Lawless (2003) e Dadpay, Soofi e Soyer (2007).

Uma propriedade importante da distribuição GG é de ser fechada para transformações potências (STACY; MIHRAM, 1965; DADPAY; SOOFI; SOYER, 2007). Isto é,

$$Y = T^s \sim GG(\alpha^s, \tau/s, k), \quad s > 0.$$

Em particular,

$$X = T^\tau \sim G(\alpha^\tau, k),$$

sendo $G(\alpha^\tau, k)$ denotando a distribuição gama com parâmetro α^τ de escala e k de forma.

A média e a variância da distribuição GG são dadas por:

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

Em geral, o $r-$ésimo momento ordinário da distribuição GG, conforme Stacy e Mihram (1965), pode ser obtido por

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

A função de distribuição acumulada $G(t)$, função de sobrevivência $S(t)$ e função risco $h(t)$ são expressas, respectivamente, por:

$$\begin{aligned} G\,(t) &= \frac{1}{\Gamma(k)}\int_0^{\left(\frac{t}{\alpha}\right)^\tau} w^{k-1}\exp(-w)dw = \frac{\gamma\left[k,\left(\frac{t}{\alpha}\right)^\tau\right]}{\Gamma(k)} \\ &= \gamma_1\left[k,\left(\frac{t}{\alpha}\right)^\tau\right], \end{aligned} \qquad (2)$$

$$S(t) = 1 - G(t) = 1 - \gamma_1\left[k,(t/\alpha)^\tau\right]$$

e

$$h(t) = \frac{t^{\tau k-1}\exp\left[-\left(\frac{t}{\alpha}\right)^\tau\right]}{\int_t^\infty x^{\tau k-1}\exp\left[-\left(\frac{x}{\alpha}\right)^\tau\right]dx},$$

sendo $\gamma(k,x) = \int_0^x w^{k-1}\exp(-w)dw$ a função gama incompleta e $\gamma_1(k,x)$ a função razão gama incompleta, definida por $\gamma_1(k,x) = \gamma(k,x)/\Gamma(k)$.

A função risco $h(t)$ da distribuição gama generalizada pode assumir as formas unimodal, banheira, crescente, decrescente e constante (Figura 2), por isso essa distribuição é bastante aplicada a dados de sobrevivência e considerada na obtenção de novas distribuições, tal como a gama generalizada geométrica.

2.5 Distribuição Gama Generalizada Geométrica

Seguindo a ideia de Adamidis e Loukas (1998), para um processo de mistura de distribuições, Ortega, Cordeiro e Pascoa (2011) propuseram a distribuição gama generalizada geométrica (GGG) com quatro parâmetros.

Suponha que $\{Y_i\}_{i=1}^Z$ sejam variáveis aleatórias independentes e identicamente distribuídas (iid) que possuem função de distribuição acumulada gama generalizada (GG) definida em (2).

Seja Z uma variável aleatória geométrica com função de probabilidade dada por $P(z;p) = (1-p)p^{z-1}$ para $Z \in \mathbb{N}$ e $p \in (0{,}1)$. Seja $X|Z = \min(\{Y_i\}_{i=1}^Z)$. Então, fazendo-se uso da distribuição do mínimo, encontra-se a

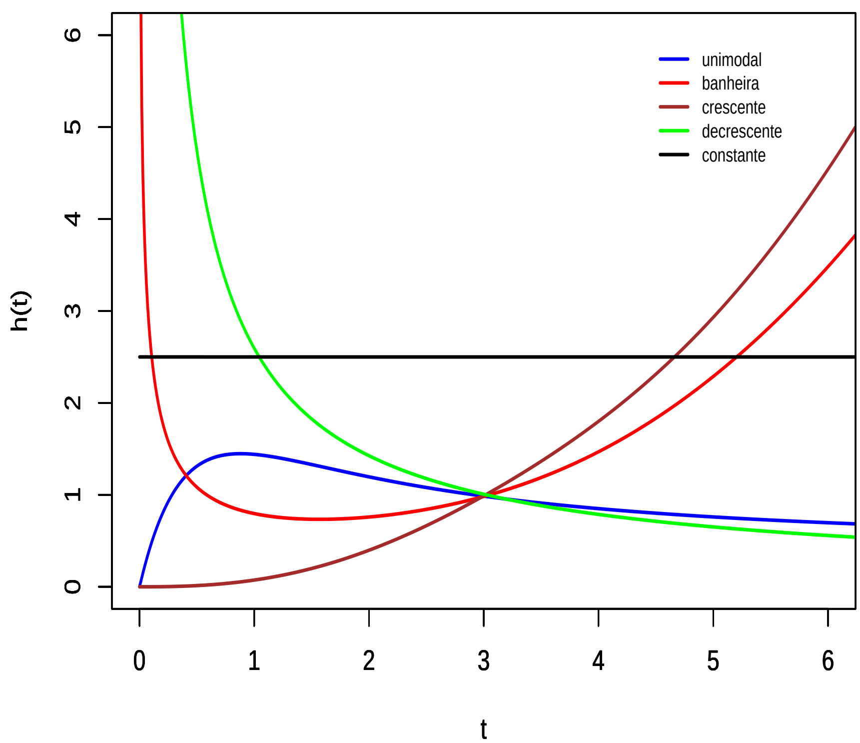

Figura 2 Formas unimodal, banheira, crescente, decrescente e constante da função risco $h(t)$.

função de distribuição acumulada de $X|Z$, representada por

$$F_{X|Z}(x|z) = 1 - \prod_{i=1}^{z}[1 - F_{Y_i}(x)] = 1 - \left\{1 - \gamma_1\left[k, \left(\frac{x}{\alpha}\right)^{\tau}\right]\right\}^{z}. \tag{3}$$

Derivando a expressão (3) em relação a x, encontra-se a função densidade de probabilidade de X condicionada a $Z = z$. Assim,

$$f(x|z;\alpha,\tau,k) = \frac{z\tau}{\alpha\Gamma(k)}\Big(\frac{x}{\alpha}\Big)^{\tau k-1}\exp\Big[-\Big(\frac{x}{\alpha}\Big)^{\tau}\Big]\Big\{1-\gamma_1\Big[k,\Big(\frac{x}{\alpha}\Big)^{\tau}\Big]\Big\}^{z-1}.$$

Sabendo que

$$f(x|z;\alpha,\tau,k) = \frac{f(x,z;\alpha,\tau,k,p)}{P(z;p)},$$

a função densidade de probabilidade conjunta de X e Z é dada por

$$\begin{aligned} f(x,z;\alpha,\tau,k,p) &= f(x|z;\alpha,\tau,k)P(z;p) \\ &= \frac{z\tau}{\alpha\Gamma(k)}\Big(\frac{x}{\alpha}\Big)^{\tau k-1}\exp\Big[-\Big(\frac{x}{\alpha}\Big)^{\tau}\Big] \\ &\times \Big\{1-\gamma_1\Big[k,\Big(\frac{x}{\alpha}\Big)^{\tau}\Big]\Big\}^{z-1}(1-p)p^{z-1}. \end{aligned}$$

Logo, a função densidade de probabilidade marginal de X, $f(x;\alpha,\tau,k,p)$, é dada por

$$\begin{aligned} f(x;\alpha,\tau,k,p) &= \sum_{z=1}^{\infty} f(x,z;\alpha,\tau,k,p) \\ &= \frac{(1-p)\tau}{\alpha\Gamma(k)}\Big(\frac{x}{\alpha}\Big)^{\tau k-1}\exp\Big[-\Big(\frac{x}{\alpha}\Big)^{\tau}\Big] \\ &\times \sum_{z=1}^{\infty} p^{z-1}z\Big\{1-\gamma_1\Big[k,\Big(\frac{x}{\alpha}\Big)^{\tau}\Big]\Big\}^{z-1}. \end{aligned} \tag{4}$$

Mas, note que, como $0 < \gamma_1[k,(x/\alpha)^{\tau}] < 1$, ver (2), e $p \in (0,1)$, então $0 < p\{1-\gamma_1[k,(x/\alpha)^{\tau}]\} < 1$. Assim, da teoria de séries geométricas,

$$\sum_{z=0}^{\infty}\Big(p\Big\{1-\gamma_1\Big[k,\Big(\frac{x}{\alpha}\Big)^{\tau}\Big]\Big\}\Big)^{z} = \frac{1}{1-p\Big\{1-\gamma_1\Big[k,\Big(\frac{x}{\alpha}\Big)^{\tau}\Big]\Big\}}, \tag{5}$$

e derivando ambos os membros de (5) em relação a p, tém-se

$$\sum_{z=1}^{\infty} p^{z-1} z\Big\{1-\gamma_1\Big[k,\Big(\frac{x}{\alpha}\Big)^{\tau}\Big]\Big\}^{z-1} = \Big(1-p\Big\{1-\gamma_1\Big[k,\Big(\frac{x}{\alpha}\Big)^{\tau}\Big]\Big\}\Big)^{-2}. \qquad (6)$$

Portanto, substituindo (6) em (4) obtém-se a função densidade da distribuição GGG com quatro parâmetros $(x > 0)$

$$\begin{aligned} f(x;\alpha,\tau,k,p) = {} & \frac{\tau(1-p)}{\alpha\Gamma(k)}\Big(\frac{x}{\alpha}\Big)^{\tau k-1}\exp\Big[-\Big(\frac{x}{\alpha}\Big)^{\tau}\Big] \\ & \times\Big(1-p\Big\{1-\gamma_1\Big[k,\Big(\frac{x}{\alpha}\Big)^{\tau}\Big]\Big\}\Big)^{-2}. \end{aligned} \qquad (7)$$

A função de distribuição acumulada correspondente a (7) é:

$$F(x;\alpha,\tau,k,p) = \frac{\gamma_1\Big[k,\Big(\frac{x}{\alpha}\Big)^{\tau}\Big]}{1-p\Big\{1-\gamma_1\Big[k,\Big(\frac{x}{\alpha}\Big)^{\tau}\Big]\Big\}}.$$

A variável aleatória X com função densidade (7) é denotada por $X \sim GGG(\alpha,\tau,k,p)$.

As funções de sobrevivência e risco correspondentes a (7) são

$$S(x;\alpha,\tau,k,p) = 1-\Big\{\gamma_1\Big[k,\Big(\frac{x}{\alpha}\Big)^{\tau}\Big]\Big(1-p\Big\{1-\gamma_1\Big[k,\Big(\frac{x}{\alpha}\Big)^{\tau}\Big]\Big\}\Big)^{-1}\Big\}$$

e

$$h(x;\alpha,\tau,k,p) = \frac{\frac{\tau(1-p)}{\alpha\Gamma(k)}\Big(\frac{x}{\alpha}\Big)^{\tau k-1}\exp\Big[-\Big(\frac{x}{\alpha}\Big)^{\tau}\Big]\Big(1-p\Big\{1-\gamma_1\Big[k,\Big(\frac{x}{\alpha}\Big)^{\tau}\Big]\Big\}\Big)^{-2}}{1-\Big\{\gamma_1\Big[k,\Big(\frac{x}{\alpha}\Big)^{\tau}\Big]\Big(1-p\Big\{1-\gamma_1\Big[k,\Big(\frac{x}{\alpha}\Big)^{\tau}\Big]\Big\}\Big)^{-1}\Big\}},$$

respectivamente.

Vale lembrar que a função risco da distribuição GGG acomoda as formas crescente, descrecente, unimodal e de banheira, demonstrando-se bastante versátil na modelagem de dados de sobrevivência (ORTEGA; CORDEIRO; PASCOA, 2011).

A distribuição GGG possui diversos modelos como casos particulares, tais como a exponencial geométrica ($\tau = k = 1$), proposta por Adamidis e Loukas (1998), Weibull geométrica ($k = 1$), gama geométrica ($\tau = 1$), gama generalizada ($p \to 0^+$), Weibull ($k = 1$ e $p \to 0^+$), exponencial ($\tau = k = 1$ e $p \to 0^+$), entre outras.

2.6 Generalização de distribuições

Os dados de sobrevivência possuem características especiais, como serem positivos e poderem ter algumas observações censuradas. Desta forma, a distribuição normal, bastante utilizada na análise estatística, não é aconselhada. Sugere-se então, para a análise de dados de sobrevivência, distribuições assimétricas com suporte nos reais positivos.

Entre as distribuições utilizadas na análise de sobrevivência, destacam-se a gama, gama generalizada, Weibull, exponencial, semi-normal, burr XII, entre outras. Mas, essas distribuições nem sempre apresentam ajustes satisfatórios, principalmente quando há necessidade da função risco assumir formas não monótonas (ORTEGA; CORDEIRO; PASCOA, 2011). Dessa forma, pesquisas na área de proposição de distribuições são intensificadas com o objetivo de obter novos modelos que melhor explicam situações mais complexas. Pode-se entender como comportamento complexo os que não são facilmente modelados pelas distribuições já existentes.

Diversas técnicas podem ser utilizadas no desenvolvimento de novas distribuições a partir de uma já existente. A transformação de variáveis aleatórias é utilizada naturalmente, por exemplo, se X é uma variável aleatória de distribuição lognormal, a variável aleatória $Y = \log(X)$ tem distribuição normal.

Outra forma de obter distribuições é diretamente pela caracterização de certos fenômenos ou experimentos, tal como ocorre com a distribuição binomial, referente ao número de sucessos em n tentativas independentes de um experimento, sendo que cada tentativa pode resultar em apenas sucesso ou fracasso, e que a probabilidade p de ocorrer sucesso (ou fracasso) em cada tentativa permanece constante.

Convoluções também podem definir distribuições de probabilidade, como é o caso da distribuição Erlang definida como a soma de exponenciais independentes e identicamente distribuídas.

Lai (2013) discute alguns métodos utilizados no desenvolvimento de novas distribuições de probabilidade, entre os quais destaca-se a transformação de funções de distribuições.

No método transformação de funções de distribuições, denomina-se distribuição base a família de distribuições a ser generalizada e denotar-se-á suas funções de distribuição acumulada e densidade respectivamente por $G(x)$ e $g(x)$. As funções de distribuição acumulada e densidade da distribuição generalizada serão representadas por $F(x)$ e $f(x)$ respectivamente. A transformação utilizada na generalização define a classe de distribuições. É oportuno esclarecer que todas as transformações são bem definidas, no sentido que todas as funções $F(x)$, obtidas por transformações de $G(x)$, são funções de distribuição acumulada.

Algumas classes de generalizações de distribuições são a exponencializadas, proposta por Gupta, Gupta e Gupta (1998), a beta, discutida por Eugene, Lee e Famoye (2002), a gama, exposta por Zografos e Balakrishnan (2009), a Kumaraswamy, desenvolvida por Cordeiro e de Castro (2011), a exponencializada generalizada, apresentada por Cordeiro, Ortega e da Cunha (2013), entre outras.

Nas próximas seções, as principais classes de distribuições serão apresentadas.

2.6.1 Classe de distribuições exponencializadas

As distribuições exponencializadas, propostas por Gupta, Gupta e Gupta (1998), são obtidas elevando-se ao expoente $\alpha > 0$ a função de distribuição acumulada $G(x)$ de uma distribuição base, isto é,

$$F(x) = G(x)^{\alpha}, \quad 0 < G(x) < 1,$$

sendo $\alpha > 0$ o novo parâmetro de forma.

As funções densidade, sobrevivência e risco são dadas respectivamente por

$$f(x) = \alpha g(x) G(x)^{\alpha-1},$$

$$S(x) = 1 - G(x)^{\alpha}$$

e

$$h(x) = \frac{\alpha g(x) G(x)^{\alpha-1}}{1 - G(x)^{\alpha}}.$$

A classe de distribuições exponencializadas possui uma interpretação física quando o parâmetro α é um número inteiro positivo. Considere um equipamento composto por α componentes independentes em um sistema em paralelo. O equipamento falhará se todos os componentes falharem. Sendo assim, seja $X_1, \cdots, X_{\alpha}$ variáveis aleatórias independentes e identicamente distribuídas que denotam o tempo de vida dos componentes, com comum função de distribuição acumulada $G(x)$. Seja X uma variável aleatória que denota o tempo de vida do equipamento. A função de distribuição acumulada $F(x)$ de X é obtida por

$$
\begin{aligned}
F(x) &= P(X \leq x) = P(X_1 \leq x, \cdots, X_\alpha \leq x) \\
&= P(X_1 \leq x)^\alpha = G(x)^\alpha.
\end{aligned}
$$

Portanto, o tempo de vida do equipamento segue a classe de distribuições exponencializadas.

A classe de distribuições exponencializadas é também denominada de Lehmann tipo I por Alexander et al. (2012) e Cordeiro, Ortega e da Cunha (2013), devido às ideias apresentadas por Lehmann (1953).

A apresentação formal da classe exponencializada foi antecedida pela proposição da Weibull exponencializada por Mudholkar, Srivastava e Freimer (1995), que possui diferentes formas para a função risco, tal como crescente, decrescente, unimodal e de banheira.

Dentro dessa classe, destacam-se as distribuições exponencial exponencializada, proposta por Gupta, Gupta e Gupta (1998), a gama exponencializada estudada por Nadarajah e Kotz (2006b) e a Weibull modificada generalizada desenvolvida por Carrasco, Ortega e Cordeiro (2008).

A distribuição exponencial exponencializada possui dois parâmetros, sendo um de forma e o outro de escala, tal como as distribuições Weibull e gama. Por isso, essa distribuição é aconselhada como uma possível alternativa às distribuições Weibull e gama (GUPTA e KUNDU, 2001).

Gupta e Kundu (1999) desenvolveram a distribuição exponencial generalizada, obtida pela transformação da distribuição exponencial de função de distribuição acumulada $G(x) = 1 - e^{-\frac{x-\mu}{\lambda}}$.

2.6.2 Classe de distribuições betas

A classe de distribuições betas introduzida por Eugene, Lee e Famoye (2002) é caracterizada pela distribuição beta e definida por

$$F(x) = \frac{1}{B(a,b)} \int_0^{G(x)} w^{a-1}(1-w)^{b-1} dw, \quad 0 < G(x) < 1, \tag{8}$$

sendo $a > 0$ e $b > 0$ os novos parâmetros de forma e $B(a,b) = \int_0^1 w^{a-1}(1-w)^{b-1}dw$ a função beta.

A função densidade de probabilidade da classe de distribuições betas correspondente à expressão (8) é dada por

$$f(x) = \frac{1}{B(a,b)} g(x)G(x)^{a-1} \left[1 - G(x)\right]^{b-1},$$

sendo $g(x) = dG(x)/dx$ a função densidade da distribuição base.

Essa classe possui a desvantagem de que a integral em (8) pode não ter solução analítica, dependendo da escolha da função de distribuição acumulada $G(x)$. Nestes casos, a obtenção de $F(x)$ é feita por métodos numéricos. Por isso, a classe beta é melhor empregada quando as funções de distribuição acumulada $G(x)$ e densidade $g(x)$ apresentam uma expressão analítica simples, tal como a distribuição exponencial, Weibull e Burr XII.

Nadarajah e Kotz (2004) introduziram a distribuição beta Gumbel, em que foi possível obter analiticamente as funções densidade de probabilidade e risco. Além da proposição da beta Gumbel, os autores calcularam a expressão para o $n-$ésimo momento e a distribuição assintótica da estatística de ordem, investigaram a variação das medidas de assimetria e curtose e discutiram o método de estimação de máxima verossimilhança.

Da mesma forma, Nadarajah e Kotz (2006a) generalizaram a distribuição exponencial, denominando-a distribuição beta exponencial. As propriedades matemáticas da distribuição beta exponencial foram discutidas e as expressões para a função geradora de momentos e função característica foram derivadas. Os primeiros momentos, a variância, a assimetria e curtose foram apresentados. E mais, desenvolveram as expressões para o desvio médio sobre a média, sobre a mediana e a distribuição das estatísticas de ordem. A estimação dos parâmetros pelo método de máxima verossimilhança foi discutido e uma expressão para a matriz de informação esperada foi apresentada.

Barreto-Souza, Santos e Cordeiro (2010) propuseram a distribuição beta exponencial generalizada, que é a generalização da distribuição exponencial generalizada (GUPTA; KUNDU, 1999) pela classe beta. Essa distribuição tem como casos particulares as distribuições beta exponencial e exponencial generalizada. Além do desenvolvimento da nova distribuição, os autores derivaram algumas propriedades a partir da distribuição exponencial generalizada, tais como a expansão da função densidade de probabilidade, explicitação da função geradora de momentos e a estimação de parâmetros pelo método de máxima verossimilhança.

Outra distribuição gerada pela classe em destaque é a beta semi-normal generalizada proposta por Pescim et al. (2010), que generaliza a distribuição semi-normal generalizada apresentada por Cooray e Ananda (2008). A utilidade dessa nova distribuição foi ilustrada na análise de um conjunto de dados reais mostrando-se mais flexível na modelagem de dados positivos em relação às distribuições semi-Normal generalizada, semi-Normal, Weibull e beta Weibull.

2.6.3 Classe de distribuições gama

Zografos e Balakrishnan (2009) propuseram a classe de distribuições gama definida por

$$F(x) = \frac{1}{\Gamma(\phi)}\gamma\left\{-log\left[1 - G(x)\right], \phi\right\}, \tag{9}$$

em que $\phi > 0$ é o novo parâmetro de forma e $\gamma\{w, \phi\} = \int_0^w e^{-u}u^{\phi-1}du$ é a função gama incompleta.

Dessa forma, derivando a função $F(x)$ em (9), obtém-se a sua função densidade de probabildiade dada por

$$f(x) = \frac{g(x)}{\Gamma(\phi)}\left\{-log\left[1 - G(x)\right]\right\}^{\phi-1},$$

em que $g(x) = dG(x)/dx$.

Algumas distribuições baseadas nessa classe são a Weibull exponencializada gama proposta por Pinho, Cordeiro e Nobre (2012), gama-Weibull e gama-log-logística apresentadas pela Hashimoto (2013) e a logística gama desenvolvida por Castellares et al. (2015).

2.6.4 Classe de distribuições Kumaraswamy

A classe de distribuições Kumaraswamy, desenvolvida por Cordeiro e de Castro (2011), é baseada na distribuição Kumaraswamy (1980) para variáveis aleatórias definidas no intervalo $(0{,}1)$, cuja função de distribuição acumulada é dada por

$$F(x) = 1 - (1 - x^a)^b, \quad 0 < x < 1, \quad a{,}b > 0. \tag{10}$$

Substituindo a variável x em (10) por $G(x)$, obtém-se a função de distribuição acumulada de distribuições da classe Kumaraswamy, ou seja,

$$F(x) = 1 - \left[1 - G(x)^a\right]^b, \quad 0 < G(x) < 1,$$

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

As funções densidade de probabilidade, sobrevivência e risco são dadas respectivamente por

$$f(x) = abg(x)G(x)^{a-1}\left[1 - G(x)^a\right]^{b-1},$$

$$S(x) = \left[1 - G(x)^a\right]^b \quad \text{e} \quad h(x) = \frac{abg(x)G(x)^{a-1}}{1 - G(x)^a}.$$

Essa classe possui a vantagem de possuir uma forma fechada para a função de distribuição acumulada, tal como ocorre com as classes estendida e exponencializada.

Uma propriedade interessante da classe Kumaraswamy é a de possuir como caso particular a classe de distribuições exponencializadas quando $b = 1$, a classe estendida quando $a = 1$ e as distribuições base $G(x)$ quando $a = b = 1$.

Entre as distribuições obtidas por essa classe, destacam-se a Kumaraswamy gama, Kumaraswamy gumbel, Kumaraswamy gaussiana inversa, Kumaraswamy normal e Kumaraswamy Weibull todas apresentadas por Cordeiro e de Castro (2011), Kumaraswamy Log-logística e a Kumaraswamy logística discutidas por Santana (2012), a Kumaraswamy gama generalizada proposta por Pascoa, Ortega e Cordeiro (2011) e a Kumaraswamy Burr XII desenvolvida por Paranaíba et al. (2013). Sendo a Kumaraswamy gama generalizada bastante recomendada na modelagem de dados de sobrevivência por possuir como sub-modelo a gama generalizada.

2.6.5 Classe de distribuições estendidas

A classe de distribuições estendidas apresentada por Cordeiro, Ortega e da Cunha (2013) é definida como uma variação da classe exponencializada introduzida por Gupta, Gupta e Gupta (1998). A função de sobrevivência da classe estendida é obtida elevando-se ao expoente λ a função de sobrevivência $1 - G(x)$ de uma distribuição base, ou seja,

$$1 - F(x) = [1 - G(x)]^{\lambda}, \quad 0 < G(x) < 1,$$

sendo $\lambda > 0$ o novo parâmetro de forma.

Consequentemente, a função de distribuição acumulada da classe estendida é dada por

$$F(x) = 1 - [1 - G(x)]^{\lambda}.$$

As funções densidade e risco são dadas respectivamente por

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

e

$$h(x) = \frac{\lambda g(x)}{1 - G(x)}.$$

Alexander et al. (2012) e Cordeiro, Ortega e da Cunha (2013) denominam alternativamente a classe de distribuições estendidas por Lehmann tipo II, devido às ideias apresentadas por Lehmann (1953).

A classe de distribuições estendidas também possui uma interpretação física quando o parâmetro λ é um número inteiro positivo. Considere um equipamento composto por λ componentes independentes em um sistema em série. O equipamento falhará se qualquer componente apresentar falha. Sendo assim, seja $X_1, \cdots, X_\lambda$ variáveis aleatórias independentes e identicamente distribuidas que denotam o tempo de vida dos componentes, com comum função de distribuição acumulada $G(x)$. Seja X uma variável aleatória que denota o tempo de vida do equipamento. A função de distribuição acumulada $F(x)$ de X é obtida por

$$
\begin{aligned}
F(x) = P(X \leq x) &= 1 - P(X > x) = 1 - P(X_1 > x, \cdots, X_\lambda > x) \\
&= 1 - P(X_1 > x)^\lambda = 1 - [1 - P(X_1 \leq x)]^\lambda \\
&= 1 - [1 - G(x)]^\lambda .
\end{aligned}
$$

Portanto, o tempo de vida do equipamento segue a classe de distribuições estendidas.

Outra característica interessante da classe estendida é a de possuir uma forma fechada para a função de distribuição acumulada, o que não ocorre, por exemplo, com a classe beta proposta por Eugene, Lee e Famoye (2002) e apresentada na seção 2.6.2. Além disso, a distribuição base $G(x)$ é um caso particular quando $\lambda = 1$.

O parâmetro de forma adicional λ proporciona maior flexibilidade na nova distribuição. No entanto, nota-se que o parâmetro λ atua apenas de forma multiplicativa na função risco, não modificando a forma da função risco da distribuição base.

Algumas distribuições modificadas pela classe estendida geram as mesmas distribuições mas com parâmetros diferentes, tal como ocorre com as distribuições exponencial, Weibull e Pareto (BARROS, 2008). Para a distribuição exponencial de parâmetro $\theta > 0$, obtém-se a distribuição exponencial de parâmetro $\theta\lambda$, para a Weibull de parâmetros $\alpha > 0$ e $\beta > 0$, obtém-se a distribuição Weibull de parâmetros $\alpha\lambda^{-1/\beta}$ e β, para a Pareto de parâmetros $k > 0$ e $\alpha > 0$, obtém-se a distribuição Pareto de parâmetros k e $\alpha\lambda$.

Por outro lado, algumas distribuições já conhecidas são obtidas pela classe estendida. A distribuição uniforme estendida, que é a generalização da distribuição uniforme no intervalo (0,1) pela classe estendida, resulta na distribuição beta com parâmetros $\alpha = 1$ e λ, a distribuição logística padrão estendida é a generalização da distribuição logística de parâmetros $\mu = 0$ e $\sigma = 1$, resultando na logística generalizada tipo II de parâmetro λ, cuja função de distribuição acumulada é dada por (JOHNSON; KOTZ; BALAKRISHNAN, 1995)

$$F(x) = 1 - [e^{-\lambda x}/(1 + e^{x})^{\lambda}], \quad x \in \mathbb{R}, \quad \lambda > 0.$$

Nadarajah e Kotz (2006b) propuseram duas novas distribuições, a Gumbel exponencializada e Fréchet exponencializada, porém, essas distribuições apresentam a mesma estrutura das distribuições estendidas. A distribuição Fréchet exponencializada (Fréchet estendida) é a generalização da Fréchet padrão, e a Gumbel exponencializada (Gumbel estendida) é a generalização da Gumbel padrão.

2.6.6 Classe de distribuições exponencializadas generalizadas

Cordeiro, Ortega e da Cunha (2013) propuseram a classe de distribuições exponencializadas generalizadas adicionando dois novos parâmetros de forma. A função de distribuição acumulada dessa nova classe é caracterizada pela transformação

$$F(x) = \{1 - [1 - G(x)]^{\alpha}\}^{\beta}, \quad 0 < G(x) < 1, \tag{11}$$

sendo $\alpha > 0$ e $\beta > 0$ os novos parâmetros de forma.

Nota-se que, a função de distribuição acumulada (11) possui expressão mais simples do que a classe beta, que também adiciona dois parâmetros.

A função densidade de probabilidade correspondente a (11) é calculada por

$$f(x) = \alpha\beta g(x) \left[1 - G(x)\right]^{\alpha-1} \left\{1 - \left[1 - G(x)\right]^{\alpha}\right\}^{\beta-1}.$$

Verifica-se que a distribuição base $G(x)$ é caso especial de (11) quando $\alpha = \beta = 1$. E mais, para $\alpha = 1$ tem-se a classe de distribuições exponencializadas e para $\beta = 1$ tem-se a classe de distribuições estendidas. Assim, essa nova classe generaliza as ideias iniciais de Gupta, Gupta e Gupta (1998), tornando as classes exponencializada e estendida como casos particulares.

A classe de distribuições exponencializadas generalizadas possui interpretação física quando seus parâmetros assumem valores inteiros positivos, tal como ocorre com as classes exponencializadas e estendidas. Considere um equipamento composto por β componentes independentes em um sistema em paralelo. Além disso, cada componente é composto por α subcomponentes independentes em um sistema em série. Assim, o equipamento falhará se todos os β componentes falharem, e qualquer componente falhará se pelo menos um subcomponente falhar. Sendo assim, sejam $X_{j,1}, \cdots, X_{j,\alpha}$ variáveis aleatórias independentes e identicamente distribuídas que denotam o tempo de vida dos α subcomponentes dentro do $j-$ésimo componente, $j = 1, \cdots, \beta$, com comum função de distribuição acumulada $G(x)$, $X_1, \cdots, X_\beta$ variáveis aleatórias independentes e identicamente distribuidas que denotam o tempo de vida dos β componentes e X a variável aleatória que denota o tempo de vida do equipamento. A função de distribuição acumulada

$F(x)$ de X é obtida por

$$\begin{aligned}F(x) &= P(X \leq x) = P(X_1 \leq x, \cdots X_\beta \leq x) = P(X_1 \leq x)^\beta \\ &= [1 - P(X_1 > x)]^\beta = [1 - P(X_{11} > x, \cdots, X_{1\alpha} > x)]^\beta \\ &= [1 - P(X_{11} > x)^\alpha]^\beta = \{1 - [1 - P(X_{11} \leq x)]^\alpha\}^\beta \\ &= \{1 - [1 - G(x)]^\alpha\}^\beta .\end{aligned}$$

Portanto, o tempo de vida do equipamento segue a classe de distribuições exponencializadas generalizadas.

Andrade et al. (2015) apresentam a distribuição Gumbel exponencializada generalizada, proposta por Cordeiro, Ortega e da Cunha (2013), e desenvolvem algumas propriedades, tais como expressões para os momentos, função geradora de momentos, estatísticas de ordem e estimação.

2.7 Modelo de regressão locação-escala

Em muitos estudos, tem-se situações em que covariáveis, como tratamentos, indicadores de grupo, características individuais, ou condições ambientais, podem estar relacionadas com os tempos de sobrevivência. Sem dúvidas, essas informações devem ser incluídas na análise estatística dos dados. Um método eficiente e elegante em considerar covariáveis na análise estatística é por meio de modelos de regressão.

Krall, Uthoff e Harley (1975), por exemplo, investigaram se o tempo de sobrevivência de pacientes com mieloma múltiplo* estava relacionado com algumas medidas fisiológicas, tais como a contagem de glóbulos brancos do indivíduo

*O mieloma múltiplo é uma doença hematológica maligna que afeta originalmente a medula óssea e se caracteriza pelo aumento do plasmócito, um tipo de célula que produz imunoglobulina, uma proteína que participa de nosso sistema de defesa. http://www.abrale.org.br/pagina/mieloma-multiplo-mm e a presença ou ausência de infecção no momento do diagnóstico, além de características pessoais como sexo e idade.

Em análise de sobrevivência, o modelo de regressão mais simples, que estabelece a relação entre a variável resposta tempo T e uma única covariável x, é dado por

$$T = \beta_0 + \beta_1 x + \epsilon, \tag{12}$$

em que β_0 e β_1 são os parâmetros a serem estimados e ϵ é o erro aleatório com distribuição normal.

No entanto, conforme Colosimo e Giolo (2006), o modelo (12) nem sempre é adequado para dados de sobrevivência, pois, a distribuição da variável resposta T tende a ser assimétrica, desencorajando o uso da distribuição normal para ϵ.

Alternativamente ao modelo (12), podem ser considerados modelos log-lineares para T, resultando em $Y = \log(T) = \beta_0 + \beta_1 x + \epsilon$ ou, na presença de l covariáveis,

$$Y = \log(T) = \beta_0 + \beta_1 x_1 + \cdots + \beta_l x_l + \epsilon = \mathbf{x}'\boldsymbol{\beta} + \epsilon, \tag{13}$$

em que $\mathbf{x}' = (1, x_1, \cdots, x_l)$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_l)$ e ϵ é o erro aleatório de média zero e parâmetro de escala $\sigma > 0$. Distribuições adequadas para T são, por exemplo, a log-normal, gama, log-logística, Weibull, entre outras. De forma correspondente para ϵ, as distribuições apropriadas são a normal, log-gama, logística, valor extremo, entre outras.

O modelo de regressão (13) é conhecido como modelo de locação e escala, pois a distribuição de Y pertence à família de distribuições que se caracteriza pelo fato de ter um parâmetro de locação $\mu = \mathbf{x}'\boldsymbol{\beta}$ e um parâmetro de escala $\sigma > 0$.

A inferência estatística nos modelos de regressão locação e escala é realizada por meio das propriedades assintóticas dos estimadores de máxima verossimilhança (LAWLESS, 2003).

Diversos autores detalham os modelos de locação e escala, ou modelo de tempo de vida acelerado, tais como Cox e Oakes (1984) e Kalbfleisch e Prentice (2002). Lawless (2003) apresenta mais detalhes dessa classe de modelos, além de vários modelos usando distribuições de probabilidade para T, comumente usadas em análise de sobrevivência.

2.8 Inferência estatística

Assumindo um modelo paramétrico adequado para a análise dos dados, deseja-se fazer inferências com base nesse modelo. Em geral, essa análise torna-se mais complicada quando se necessita incorporar dados censurados.

Nesta seção apresentar-se-á o método da máxima verossimilhança e a abordagem bayesiana para estimação e testes de hipóteses para modelos em análise de sobrevivência.

2.8.1 Método de máxima verossimilhança

Sejam $(y_1,\delta_1,\mathbf{x}_1), \cdots, (y_n,\delta_n,\mathbf{x}_n)$, n observações independentes, com $y_i = \log(t_i)$ o logaritmo do tempo de falha ou censura, δ_i o indicador de censura e $\mathbf{x}_i = (x_{i1}, \cdots, x_{il})'$ o vetor de covariáveis, para todo $i = 1,2, \cdots, n$. O logaritmo da função de verossimilhança definida em (1), considerando uma amostra completa, é dado por

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

em que $f(y)$ e $S(y)$ são as funções densidade e de sobrevivência da variável aleatória Y, $\boldsymbol{\theta}$ é o vetor de parâmetros e F e C denotam os conjuntos de observações não censuradas e censuradas, respectivamente.

Os estimadores de máxima verossimilhança são obtidos derivando $\ell(\boldsymbol{\theta})$ em relação ao vetor de parâmetros desconhecido $\boldsymbol{\theta}$ e resolvendo o sistema de equações

$$U(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0. \tag{14}$$

Quando o sistema de equações (14) é não-linear, as estimativas de máxima verossimilhança são obtidas por procedimentos iterativos, tais como os métodos de otimização Newton-Raphson ou *quasi*-Newton.

As propriedades assintóticas dos estimadores de máxima verossimilhança são usadas na construção de intervalos de confiança e testes de hipóteses sobre os parâmetros do modelo. Considerando o fato que, sob certas condições de regularidade, $\widehat{\boldsymbol{\theta}}$ tem assintoticamente distribuição normal multivariada de média $\boldsymbol{\theta}$ e matriz de variâncias e covariâncias dada pelo inverso da matriz de informação de Fisher $\mathbf{I}(\boldsymbol{\theta})$, em que $\mathbf{I}(\boldsymbol{\theta}) = E[-\ddot{\mathbf{L}}(\boldsymbol{\theta})]$ e $\ddot{\mathbf{L}}(\boldsymbol{\theta}) = \partial^2 \ell(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{'}$. Ou seja, a matriz de variância e covariâncias dos estimadores de máxima verossimilhança é aproximadamente o negativo da inversa da esperança da matriz de derivadas segundas do logaritmo de $L(\boldsymbol{\theta})$.

Como o cálculo da matriz de informação de Fisher $\mathbf{I}(\boldsymbol{\theta})$ é complicado, devido às observações censuradas, pode-se utilizar o negativo da matriz Hessiana, $-\ddot{\mathbf{L}}(\boldsymbol{\theta})$, avaliada em $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, que é um estimador consistente para $\mathbf{I}(\boldsymbol{\theta})$. Portanto, a distribuição assintótica para $\widehat{\boldsymbol{\theta}}$ é especificada por

$$\widehat{\boldsymbol{\theta}}^{'} \overset{a}{\sim} N_{(d)} \left[\boldsymbol{\theta}^{'}; -\ddot{\mathbf{L}}(\widehat{\boldsymbol{\theta}})^{-1} \right],$$

em que $-\ddot{\mathbf{L}}(\widehat{\boldsymbol{\theta}})$ é a matriz de informação observada e d é o número de parâmetros do modelo.

Para avaliar se duas distribuições encaixadas são equivalentes, pode-se utilizar o teste de razão de verossimilhanças (LR, do inglês *likelihood ratio*). Considerando a partição $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$, em que $\boldsymbol{\theta}_1$ é um subconjunto de parâmetros de interesse e $\boldsymbol{\theta}_2$ é um subconjunto de parâmetros remanescentes. A estatística LR para testar a hipótese nula $H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^{(0)}$ *versus* a hipótese alternativa $H_1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_1^{(0)}$ é dada por $LR = 2\{\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widetilde{\boldsymbol{\theta}})\}$, em que $\widetilde{\boldsymbol{\theta}}$ e $\widehat{\boldsymbol{\theta}}$ são os estimadores de máxima verossimilhança sob as hipóteses nulas e alternativa, respectivamente. A estatística LR é assintoticamente distribuída como χ^2_ν, com ν a dimensão do subconjunto de parâmetros de interesse $\boldsymbol{\theta}_1$.

O método de máxima verossimilhança, no contexto da análise de sobrevivência, foi usado por diversos autores, destaca-se Ortega et al. (2012) para o modelo de regressão log-gama generalizada exponencializada e Gomes et al. (2014) na estimação dos parâmetros da distribuição Kumaraswamy Rayleigh generalizada. No âmbito das famílias de distribuições, Cordeiro e de Castro (2011) discutem a estimação de parâmetros na família de distribuições Kumaraswamy, Cordeiro, Alizadeh e Ortega (2014) na classe de distribuições semi-logística exponencializada e Bourguignon, Silva e Cordeiro (2014) na família de distribuições Weibull-G.

2.8.2 Análise bayesiana

Na abordagem clássica paramétrica, θ é considerado como uma quantidade desconhecida fixa. Contudo, na bayesiana, θ é uma quantidade que pode ser descrita pela distribuição de probabilidade *a priori* $\pi(\theta)$.

Combinando a informação *a priori* com a função amostral $f(x|\theta)$ tem-se, pelo teorema de Bayes, a distribuição *a posteriori* $\pi(\theta|x)$, que é obtida por

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int_{\Theta} f(x|\theta)\pi(\theta)d\theta}, \quad \theta \in \Theta.$$

Para uma amostra aleatória observada $\mathbf{x} = x_1, \cdots, x_n$ de $f(x|\theta)$, tem-se que

$$\pi(\theta|\mathbf{x}) = \frac{L(\theta|\mathbf{x})\pi(\theta)}{\int_{\Theta} L(\theta|\mathbf{x})\pi(\theta)d\theta}, \quad \theta \in \Theta, \tag{15}$$

em que $L(\theta|\mathbf{x}) = \prod_{i=1}^{n} f(x_i|\theta)$ é a função de verossimilhança de θ correspondente à amostra observada $\mathbf{x}$.

Note que o denominador em (15) é a distribuição marginal de $\mathbf{X}$, que funciona como uma constante normalizadora de $\pi(\theta|\mathbf{x})$, pois não depende de θ. Então, pode-se omitir o termo $\int_{\Theta} f(\mathbf{x}|\theta)\pi(\theta)d\theta$ em (15), e a igualdade é substituída por uma proporcionalidade, ou seja,

$$\pi(\theta|\mathbf{x}) \propto L(\theta|\mathbf{x})\pi(\theta).$$

Pelo teorema de Bayes é possível atualizar a informação a respeito do parâmetro θ com base na amostra observada $\mathbf{x}$, e utilizá-lo para quantificar esse aumento de informação. Sendo assim, na inferência bayesiana considera-se a maior quantidade de informação disponível para θ, que é a distribuição *a posteriori* $\pi(\theta|\mathbf{x})$.

Considerando o vetor de parâmetros $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_d)$, é necessário obter a distribuição marginal *a posteriori* $\pi(\theta_i|\mathbf{x})$ para cada parâmetro θ_i, $i = 1, \cdots, d$, pela integração da função densidade conjunta *a posteriori* $\pi(\boldsymbol{\theta}|\mathbf{x})$, ou seja,

$$\pi(\theta_i|\mathbf{x}) = \int \cdots \int \pi(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta}_{-i}, \tag{16}$$

em que $\boldsymbol{\theta}_{-i}$ refere-se a todos os parâmetros exceto θ_i.

A resolução analítica da integral (16) é, em geral, impraticável. Uma das alternativas é a utilização de métodos numéricos, dentre os quais destacam-se os algoritmos dos métodos de Monte Carlo via Cadeias de Markov (MCMC), que tem como objetivo simular um passeio aleatório no espaço do parâmetro $\boldsymbol{\theta}$, o qual converge para uma distribuição estacionária, que é a distribuição marginal *a posteriori* de $\boldsymbol{\theta}$ (PAULINO; TURKMAN; MURTEIRA, 2003).

Os métodos MCMC, utilizam a simulação estocástica considerando as distribuições condicionais completas *a posteriori* de cada parâmetro para gerar amostras que convergem para a densidade marginal, com o aumento do tamanho dessa amostra.

A distribuição condicional completa *a posteriori* do parâmetro θ_i, denotada por $\pi(\theta_i|\boldsymbol{\theta}_{-i},\mathbf{x})$, é obtida considerando na densidade conjunta *a posteriori* $\pi(\boldsymbol{\theta}|\mathbf{x})$ todos os parâmetros $\boldsymbol{\theta}_{-i}$ conhecidos e, assim, a expressão se torna menos complexa, já que as constantes podem ser desconsideradas.

Na obtenção da distribuição *a posteriori* de $\boldsymbol{\theta}$ pode-se considerar dois algoritmos: o amostrador de Gibbs quando as condicionais completas possuem formas conhecidas (GELFAND; SMITH, 1990), ou o algoritmo Metropolis-Hastings para o caso em que as condicionais completas possuem formas desconhecidas (HASTINGS, 1970).

A distribuição marginal *a posteriori* de um parâmetro θ_i contém toda a informação probabilística a respeito deste parâmetro. No entanto, algumas vezes, é necessário resumir a informação contida nesta distribuição por meio de alguns poucos valores numéricos. Um caso simples é a estimação pontual de θ_i, em que se resume a distribuição marginal *a posteriori* por meio de um único número $\hat{\theta}_i$. É importante, também, associar alguma informação sobre o quão precisa é a especificação deste número. As medidas de incerteza mais usuais são: a variância e o coeficiente de variação para a média *a posteriori*, a medida de informação observada de Fisher para a moda *a posteriori* e a distância entre quartis para a mediana *a posteriori*.

O próprio conceito de estimação pontual de parâmetros, conduz no cenário bayesiano, a considerar como estimativas os pontos críticos da distribuição *a posteriori*.

Um resumo de $\pi(\boldsymbol{\theta}|\mathbf{x})$ mais informativo do que qualquer estimativa pontual é obtido de uma região do espaço paramétrico de $\boldsymbol{\theta}$, que contenha uma parte substancial da massa probabilística *a posteriori*, chamada de intervalo de credibilidade (PAULINO; TURKMAN; MURTEIRA, 2003).

Um intervalo C, contido no conjunto suporte de θ_i, é definido como intervalo de credibilidade com nível de credibilidade $100(1-\alpha)\%$ de θ_i se $P(\theta_i \in C) \geq 1-\alpha$.

Dada a infinidade de intervalos de credibilidade de probabilidade $100(1-\alpha)\%$, interessa selecionar aquele de menor comprimento possível. O intervalo de credibilidade de comprimento mínimo é obtido tomando os valores de θ_i com maior densidade *a posteriori*, tal intervalo de credibilidade é denominado de máxima densidade *a posteriori*, ou simplesmente HPD (do inglês *hightest posterior density*). Em um intervalo HPD, todos os seus valores possuem densidade maior que qualquer outro ponto fora dele.

Em análise de sobrevivência, há diversos trabalhos utilizando a abordagem bayesiana, tais como os de Pascoa et al. (2011) para a distribuição Kumaraswamy gama generalizada, Paranaíba et al. (2013) na distribuição Kumaraswamy Burr XII, Pascoa et al. (2013) para o modelo de regressão log-Kumaraswamy gama generalizada e Hashimoto et al. (2015) no modelo McDonald Weibull estendida.

2.9 Critérios de informação AIC, BIC e CAIC

A seleção de um modelo mais adequado, entre os disponíveis, pode ser realizada com base em algum critério. Em geral, considera-se uma medida de qualidade de ajuste em relação aos dados. Alguns dos critérios conhecidos consideram, também, a complexidade do modelo avaliado.

Akaike (1974) propôs o critério de informação de Akaike (AIC, do inglês *Akaike information criterion*), estabelecendo uma relação entre a máxima verossimilhança e a informação de Kullback-Leibler.

O critério de informação bayesiano (BIC, do inglês *bayesian information criterion*), proposto por Schwarz (1978), é um critério de avaliação de modelos definido em termos da probabilidade *a posteriori*.

Emiliano et al. (2009) apresentam os fundamentos dos critérios AIC e BIC e um estudo de comparação entre eles em modelos normais. Os autores ressaltam que tanto o BIC quanto o AIC tiveram resultados semelhantes para grandes amostras. Entretanto, para amostras de tamanho pequeno, ambos os critérios não apresentaram bons resultados.

Hurvich e Tsai (1989) estabeleceram o critério de informação Akaike corrigido (CAIC, do inglês *corrected Akaike information criterion*), melhorando o tradicional AIC em pequenas amostras. No CAIC é considerada uma ponderação entre o número de parâmetros e o tamanho amostral.

Considerando esses esses critérios, estabelecerá como o melhor modelo aquele que apresentar menor valor para AIC, BIC e CAIC.

Para uma amostra aleatória $X_1, \cdots, X_n$ de tamanho n e vetor de parâmetros $\boldsymbol{\theta}$, as estatísticas AIC, BIC e CAIC podem ser calculadas por

$$\begin{aligned}
\text{AIC} &= -2\ell(\boldsymbol{\theta}) + 2d, \\
\text{BIC} &= -2\ell(\boldsymbol{\theta}) + d\log(n), \\
\text{CAIC} &= \text{AIC} + 2\frac{(d+1)(d+2)}{n-d-2},
\end{aligned}$$

sendo $\ell(\boldsymbol{\theta})$ o logaritmo da função de verossimilhança maximizada e d o número de parâmetros estimados pelo modelo.

Nota-se que esses critérios, essencialmente, penalizam a verossimilhança do modelo pelo número de variáveis estimadas e, eventualmente, pelo tamanho da amostra.

Os critérios AIC, BIC e CAIC foram aplicados por Pascoa et al. (2013) na comparação dos modelos de regressão log-Kumaraswamy gama generalizada e o de Cox (1972). Hashimoto et al. (2015) utilizaram as estatísticas AIC e BIC para comparar o modelo McDonald Weibull estendida com seus submodelos Weibull exponencializada, beta exponencial e Weibull.

REFERÊNCIAS

ADAMIDIS, K.; LOUKAS, S. A Lifetime distribution with decreasing failure rate. **Statistics & Probability Letters**, v. 39, p. 35-42, 1998.

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v. 19, n. 6, p. 716-723, 1974.

ALEXANDER, C.; CORDEIRO, G. M.; ORTEGA, E. M. M.; SARABIA, J. M. Generalized beta-generated distributions. **Computational Statistics & Data Analysis**, v. 56, n. 6, p. 1880-1897, 2012.

ALI, M. M.; WOO, J.; NADARAJAH, S. Generalized gamma variables with drought application. **Journal of the Korean Statistical Society**, v. 37, p. 37-45, 2008.

ALLISON, P. D. **Survival analysis using SAS:** a practical guide. 2. ed. Cary: SAS Institute, 2010. 336p.

ANDRADE, T.; RODRIGUES, H.; BOURGUIGNON, M.; CORDEIRO, G. M. The exponentiated generalized Gumbel distribution. **Revista Colombiana de Estadística**, v. 38, n. 1, p. 123-143, 2015.

BARRETO-SOUZA, W., SANTOS, A. H. S. ; CORDEIRO, G. M. The beta generalized exponential distribution. **Journal of Statistical Computation and Simulation**, v. 80, n. 2, p. 159-172, 2010.

BARROS, A. A. A. **Distribuições exponencializadas e estendidas**. 2008. 75p. Dissertação (Mestrado em Biometria e Estatística Aplicada) - Universidade Federal Rural de Pernambuco, Pernambuco, 2008.

BOURGUIGNON, M.; SILVA, R. B.; CORDEIRO, G. M. The Weibull-G family of probability distributions. **Journal of Data Science**, v. 12, p. 53-68, 2014.

CARRASCO, J. M. F.; ORTEGA, E. M. M.; CORDEIRO, G. M. A generalized modified Weibull distribution for lifetime modeling. **Computational Statistics**

and Data Analysis, v. 53, p. 450-462, 2008.

CASTELLARES, F.; SANTOS, M. A. C.; MONTENEGRO, L. C.; CORDEIRO, G. M. A gamma-generated logistic distribution: properties and inference. **American Journal of Mathematical and Management Sciences**, v. 34, p. 14-39, 2015.

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

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

CORDEIRO, G. M.; ALIZADEH, M.; ORTEGA, E. M. M. The exponentiated half-logistic family of distributions: properties and applications. **Journal of Probability and Statistics**, v. 2014, p. 1-21, 2014.

CORDEIRO, G. M.; de CASTRO, M. A new family of generalized distributions. **Journal of Statistical Computation and Simulation**, v. 81, n. 7, p. 883-898, 2011.

CORDEIRO, G. M.; ORTEGA, E. M. M.; da CUNHA, D. C. C. The exponentiated generalized class of distributions. **Journal of Data Science**, v. 11, p. 1-27, 2013.

CORDEIRO, G. M.; ORTEGA, E. M. M.; SILVA, G. O. The exponentiated generalized gamma distribution with application to lifetime data. **Journal of Statistical Computation and Simulation**, v. 81, n.7, p. 827-842, 2011.

COX, D. R. Regression models and life-tables. **Journal of the Royal Statistical Society, Series B**, v. 34, n. 2, p. 187-220, 1972.

COX, D.R.; OAKES, D. **Analysis of survival data**. New York: Chapman and Hall, 1984. 208p.

DADPAY, A.; SOOFI, E. S.; SOYER, R. Information measures for generalized gamma family. **Journal of Econometrics**, v. 138, p. 568-585, 2007.

EMILIANO, P. C.; VIVANCO, M. J. F.; de MENEZES, F. S.; AVELAR, F. G. Foundations and comparison of information criteria: Akaike and Bayesian. **Revista Brasileira de Biometria**, v. 27, n. 3, p. 394-411, 2009.

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

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

GELFAND, A. E.; SMITH, A. F. M. Sampling-based approaches to calculating marginal densities. **Journal of the American Statistical Association**, v. 85, n. 410, p. 398-409, 1990.

GOMES, A. E.; da-SILVA, C. Q.; CORDEIRO, G. M.; ORTEGA, E. M. M. A new lifetime model: the Kumaraswamy generalized Rayleigh distribution. **Journal of Statistical Computation and Simulation**, v. 84, n. 2, p. 290-309, 2014.

GUPTA, R. C.; GUPTA, P. L.; GUPTA, R. D. Modeling failure time data by lehman alternatives. **Communications in Statistics - Theory and Methods**, v. 27, n. 4, p. 887-904, 1998.

GUPTA, R. D.; KUNDU, D. Generalized exponential distributions. **Australian & New Zealand Journal of Statistics**, v. 41, n. 2, p. 173-188, 1999.

GUPTA, R. D.; KUNDU, D. Exponentiated exponential family: an alternative to gamma and Weibull distributions. **Biometrical Journal**, v. 43, n. 1, p. 117-130, 2001.

HASHIMOTO, E. M. **Modelo de regressão gama-G em análise de sobrevivência**. 2013. 177p. Tese (Doutorado em Estatística e Experimentação Agronômica) - Universidade de São Paulo, Piracicaba, 2013.

HASHIMOTO, E. M.; ORTEGA, E. M. M.; CORDEIRO, G. M.; PASCOA, M. A. R. The McDonald extended Weibull distribution. **Journal of Statistical Theory and Practice**, v. 0, n. 0, p. 1-25, 2015.

HASTINGS, W. K. Monte carlo sampling methods using Markov Chains and theirs applications. **Biometrika**, v. 57, n.1, p.97-109, 1970.

HURVICH, C. M.; TSAI, C. L. Regression and time series model selection in small samples. **Biometrika**, v. 76, n. 2, p.297-307, 1989.

JOHNSON, N.L.; KOTZ, S.; BALAKRISHNAN, N. **Continuous univariate distributions.** 2. ed. New York: John Wiley, 1995. v. 2, 752p.

KALBFLEISH, J.D.; PRENTICE, R.L. **The statistical analysis of failure time data**. New York: John Wiley, 2002. 462p.

KRALL, J.; UTHOFF, V.; HARLEY, J. A step-up procedure for selectiong variables associated with survival. **Biometrics**, v. 31, p. 49-57, 1975.

LAI, C. D. Constructions and applications of lifetime distributions. **Applied Stochastic Models in Business and Industry**, v. 29, p. 127-140, 2013.

LAWLESS, J. F. **Statistical models and methods for lifetime data**. 2. ed. New York: Wiley, 2003. 630p.

LEHMANN, E. L. The power of rank tests. **Annals of Mathematical Statistics**, v. 24, p. 23-43, 1953.

MUDHOLKAR, G. S.; SRIVASTAVA, D. K.; FREIMER, M. The exponentiated Weibull family: a reanalysis of the bus-motor-failure data. **Technometrics**, v. 37, p. 436-445, 1995.

NADARAJAH, S.; GUPTA, A. K. A generalized gamma distribution with application to drought data. **Mathematics and Computers in Simulation**, v. 74, p.1-7, 2007.

NADARAJAH, S.; KOTZ, S. The beta Gumbel distribution. **Mathematical Problems in Engineering**, v. 4, p. 323-332, 2004.

NADARAJAH, S.; KOTZ, S. The beta exponential distribution. **Reliability Engineering & System Safety**, v. 91, n. 6, p. 689-697, 2006a.

NADARAJAH, S.; KOTZ, S. The exponential type distributions. **Acta**

Applicandae Mathematicae, v. 92, n. 2, p. 97-111, 2006b.

ORTEGA, E. M. M.; CORDEIRO, G. M.; PASCOA, M. A. R. The generalized gamma geometric distribution. **Journal of Statistical Theory and Applications**, v. 10, p. 433-454, 2011.

ORTEGA, E. M. M.; CORDEIRO, G. M.; PASCOA, M. A. R.; COUTO, E. V. The log-exponentiated generalized gamma regression model for censored data. **Journal of Statistical Computation and Simulation**, v. 82, n. 8, p. 1169-1189, 2012.

PARANAÍBA, P. F.; ORTEGA, E. M. M.; CORDEIRO, G. M.; PASCOA, M. A. R. The Kumaraswamy Burr XII distribution: theory and practice. **Journal of Statistical Computation and Simulation**, v. 83, n. 11, p. 2117-2143, 2013.

PASCOA, M. A. R.; ORTEGA, E. M. M.; CORDEIRO, G. M. The Kumaraswamy generalized gamma distribution with application in survival analysis. **Statistical Methodology**, v. 8, n. 5, p. 411-433, 2011.

PASCOA, M. A. R.; PAIVA, C. M. M.; CORDEIRO, G. M.; ORTEGA, E. M. M. The Log-Kumaraswamy generalized gamma regression model with application to chemical dependency data. **Journal of Data Science**, v. 11, n. 4, p. 781-818, 2013.

PAULINO, C. D.; TURKMAN, A. A.; MURTEIRA, B. **Estatística bayesiana.** Lisboa: Fundação Calouste Gulbenkian, 2003. 446p.

PESCIM, R. R.; DEMÉTRIO, C. G. B.; CORDEIRO, G. M.; ORTEGA, E. M. M.; URBANO, M. R. The beta generalized half-normal distribution. **Computational statistics & data analysis**, v. 54, n. 4, p. 945-957, 2010.

PINHO, L. G. B.; CORDEIRO, G. M; NOBRE, J. S. The gamma-exponentiated Weibull distribution. **Journal of Statistical Theory and Applications**, v. 11, p. 379-395, 2012.

PRENTICE, R. L. A log-gamma model and its maximum likelihood estimation. **Biometrica**, v. 61, n. 3, p. 539-544, 1974.

SANTANA, T. V. F.; ORTEGA, E. M. M.; CORDEIRO, G. M.; SILVA, G. O.

The Kumaraswamy-log-logistic distribution. **Journal of Statistical Theory and Applications**, v. 11, n. 3, p. 265-291, 2012.

SCHWARZ, G. Estimating the dimensional of a model. **Annals of Statistics** v. 6, n. 2, p. 461-464, 1978.

STACY E. W. A generalization of the gamma distribution. **Annals of Mathematical Statistics** v. 33, p. 1187-1192, 1962.

STACY, E. W.; MIHRAM, G. A. Parameter estimation for a generalized gamma distribution. **Technometrics**, v. 7, n. 3, p. 349-358, 1965.

ZOGRAFOS, K.; BALAKRISHNAN, N. On families of beta- and generalized gamma-generated distributions and associated inference. **Statistical Methodology**, v. 6, p. 344-362, 2009.

SEGUNDA PARTE
Artigos

ARTIGO 1: The Extended Generalized Gamma Geometric Distribution

Redigido conforme as normas da revista Hacettepe Journal of Mathematics and Statistics - versão preliminar.

The Extended Generalized Gamma Geometric Distribution

Juliano Bortolini[*], Marcelino A. R. Pascoa[†‡], Renato R. de Lima[§] and Anderson C. S. de Oliveira[¶]

Abstract

We propose and study the so-called extended generalized gamma geometric distribution. The proposed distribution has five parameters and it can be accommodate increasing, decreasing, bathtub and unimodal shaped hazard functions. The new distribution has a large number of well-known lifetime special sub-models such as the generalized gamma geometric, Weibull geometric, gamma geometric, exponential geometric, Rayleigh geometric, half-normal geometric among others. We provide a mathematical treatment of the new distribution including explicit expressions for moments, moment generating function, mean deviations, reliability and order statistics. The method of maximum likelihood and a Bayesian procedure are adopted for estimating the model parameters. Finally, an application of the new distribution is illustrated in a real data set.

Keywords: Generalized gamma distribution, Weibull geometric distribution, lifetime distribution, maximum likelihood estimation, bimodality.

2000 AMS Classification: 60E05, 62E15, 62P25

1. Introduction

The generalized gamma (GG) distribution, introduced by Stacy [31], is an extensive family that contains a variety of special sub-models, including the exponential, Weibull, log normal, gamma and Rayleigh distributions, among others. This distribution is suitable for modeling lifetime data and for modeling phenomenon with different types of hazard rate function as well as monotonically increasing and decreasing, in the form of bathtub and unimodal [13].

Several distributions based on extensions or mixtures of the distributions were developed in last years providing more flexibility for modeling survival data. Adamidis and Loukas [2] introduced a two-parameter distribution with decreasing hazard rate so-called exponential geometric (EG). Silva *et al.* [29] proposed the

[*]Departamento de Estatística, Universidade Federal de Mato Grosso, Cuiabá 78060-900, Brazil, Email: julianobortolini@ufmt.br

[†]Departamento de Estatística, Universidade Federal de Mato Grosso, Cuiabá 78060-900, Brazil, Email: marcelino.pascoa@gmail.com

[‡]Corresponding Author.

[§]Departamento de Ciências Exatas, Universidade Federal de Lavras, Lavras 37200-000, Brazil, Email: rrlima@dex.ufla.br

[¶]Departamento de Estatística, Universidade Federal de Mato Grosso, Cuiabá 78060-900, Brazil, Email: andersoncso@gmail.com

generalized exponential geometric with decreasing, increasing and unimodal hazard rate depending on their parameters. Gupta and Kundu [17, 18] developed the generalized exponential (GE) and exponentiated exponential distributions which has increasing or decreasing hazard rate depending on the shape parameter. These authors provided more results of the GE distribution in [19, 20]. Mudholkar *et al.* [22] introduced the exponentiated Weibull (EW) which has unimodal hazard rate function, Lai *et al.* [21] proposed the modified Weibull, Carrasco *et al.* [6] presented the generalized modified Weibull (GMW), Silva *et al.* [30] studied the beta modified Weibull distribution which admits only increasing and decreasing hazard rate functions and Barreto-Souza *et al.* [5] defined the Weibull geometric (WG) which is an extension of the EG distribution and considered for modeling monotone or unimodal hazard rates. Cordeiro *et al.* [7] introduced the exponentiated generalized gamma distribution, Pascoa *et al.* [26] proposed the Kumaraswamy generalized gamma, Paranaíba *et al.* [25] exposed the Kumaraswamy Burr XII. Cordeiro *et al.* [9] defined the beta-Weibull geometric, Cordeiro *et al.* [10] presented the Kumaraswamy modified Weibull that contains as special sub-models the GMW, EW among others distributions. Al-Zahrani *et al.* [3] defined the (P-A-L) extended Weibull distribution and Cordeiro *et al.* [11] proposed the Kumaraswamy exponential-Weibull distribution that generalizes a number of well-known special lifetime models such as the Weibull, exponential, Rayleigh, modified Rayleigh, modified exponential and exponentiated Weibull distributions, among others.

Ortega *et al.* [24], following the idea of Adamidis and Loukas [2] for a process of mixing distributions, introduced the generalized gamma geometric (GGG) distribution with four parameters that generalizes the GG, EG and WG distributions.

The GGG distribution has monotonically increasing and decreasing, in the form of bathtub and unimodal hazard rate. However, this distribution and their submodels does not provide a reasonable parametric fit for some practical applications which data may be bimodal shape.

In this work we propose so-called the extended generalized gamma geometric (denoted with the prefix "ExGGG" for short) distribution with five parameters and derive some of their properties with the hope that it will attract wider applications in reliability, engineering and in other areas of research. We are motivated to study the ExGGG distribution because of the wide usage of the GGG distribution and their sub-models in survival analysis. Furthermore, the current extension provides density and hazard rate functions with great flexibility to model complex data in a great variety of applications including the bimodal cases.

The paper is outlined as follows. In Section 2, we define the ExGGG distribution and some of their submodels. Further, we derive useful expansions for its density function. In Section 4 we obtain two alternative expansions for the moments. In Section 5 we provide an explicit expression for the moment generating function. The mean deviations are determined in Section 6. The reliability is derived in Section 7. In Section 8 we derive the density function of the *ith* order statistic. Maximum likelihood method and Bayesian approach for the parameter model are discussed in Section 9. The usefulness of the new model is illustrated by means of an application to real data in Section 10. Some conclusions are offered in Section 11.

2. The ExGGG distribution

The distribution GGG with four parameters $\alpha > 0$, $\tau > 0$, $k > 0$ and $p \in (0,1)$, defined by Ortega *et al.* [24], has the probability density function (pdf) given by

$$g(x) = \frac{\tau(1-p)}{\alpha\Gamma(k)}\left(\frac{x}{\alpha}\right)^{\tau k-1}\exp\left[-\left(\frac{x}{\alpha}\right)^{\tau}\right]\left(1-p\left\{1-\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^{\tau}\right]\right\}\right)^{-2},$$

where $x > 0$, $\gamma(k,x) = \int_0^x w^{k-1}e^{-w}dw$ is the incomplete gamma function, $\Gamma(k) = \int_0^\infty w^{k-1}e^{-w}dw$ is the gamma function and $\gamma_1(k,x) = \gamma(k,x)/\Gamma(k)$ is the incomplete gamma function ratio.

The cumulative density function (cdf) of the GGG distribution is

$$G(x) = \frac{\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^{\tau}\right]}{1-p\left\{1-\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^{\tau}\right]\right\}}, \quad x > 0. \tag{2.1}$$

Let $G(x)$ be a cdf, the extended class of distributions (also referred as the Lehmann type II class of distributions) presented by Cordeiro *et al.* [8] corresponding to $G(x)$ is defined by $F(x) = 1 - [1-G(x)]^{\lambda}$, where λ is a positive real number. Hence, the cdf of the ExGGG with five parameters $\alpha > 0$, $\tau > 0$, $k > 0$, $\lambda > 0$ and $p \in (0,1)$ has the form

$$F(x) = 1 - \left\{1 - \frac{\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^{\tau}\right]}{1-p\left\{1-\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^{\tau}\right]\right\}}\right\}^{\lambda}, \quad x > 0 \tag{2.2}$$

and the pdf is given by

$$\begin{aligned} f(x) &= \frac{\lambda\tau(1-p)}{\alpha\Gamma(k)}\left(\frac{x}{\alpha}\right)^{\tau k-1}\exp\left[-\left(\frac{x}{\alpha}\right)^{\tau}\right]\left(1-p\left\{1-\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^{\tau}\right]\right\}\right)^{-2} \\ &\times \left\{1-\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^{\tau}\right]\left(1-p\left\{1-\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^{\tau}\right]\right\}\right)^{-1}\right\}^{\lambda-1}. \end{aligned} \tag{2.3}$$

A random variable X having pdf (2.3) is denoted by ExGGG$(\alpha,\tau,k,p,\lambda)$. Clearly, when $\lambda = 1$ we have GGG distribution. Some distributions are obtained from (2.3) as particular cases, for example, when $k = 1$ we have the extended Weibull geometric (ExWG), which is a new distribution, for $k = \lambda = 1$ we have the WG distribution, for $\tau = k = \lambda = 1$, we obtain the EG distribution. The GG distribution is the limiting distribution (the limit is defined in terms of the convergence in distribution) of the ExGGG distribution when $p \to 0^+$ and $\lambda = 1$. On the other hand, if $p \to 1^-$, we obtain the distribution of a random variable Y such that $P(Y = 0) = 1$. Hence, the parameter p can be interpreted as a degeneration parameter. Some important ExGGG sub-models are listed in Table 1.

The survival and hazard rate functions corresponding to (2.2) are

$$S(x) = \left\{1 - \frac{\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^{\tau}\right]}{1-p\left\{1-\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^{\tau}\right]\right\}}\right\}^{\lambda}$$

Table 1. Some ExGGG Distributions.

Distribution	τ	α	k	p	λ
Extended Gamma geometric	1	α	k	p	λ
Extended Chi-square geometric	1	2	$\frac{n}{2}$	p	λ
Extended Exponential geometric	1	α	1	p	λ
Extended Weibull geometric	c	α	1	p	λ
Extended Rayleigh geometric	2	$\sigma\sqrt{2}$	1	p	λ
Extended Maxwell geometric	2	$\sigma\sqrt{2}$	$\frac{3}{2}$	p	λ
Extended Half normal geometric	2	$\sigma\sqrt{\pi}$	$\frac{1}{2}$	p	λ
Extended Generalized gamma	τ	α	k	0^+	λ
Extended Gamma	1	α	k	0^+	λ
Extended Chi-square	1	2	$\frac{n}{2}$	0^+	λ
Extended Exponential	1	α	1	0^+	λ
Extended Weibull	c	α	1	0^+	λ
Extended Rayleigh	2	$\sigma\sqrt{2}$	1	0^+	λ
Extended Maxwell	2	$\sigma\sqrt{2}$	$\frac{3}{2}$	0^+	λ
Extended Half normal	2	$\sigma\sqrt{\pi}$	$\frac{1}{2}$	0^+	λ
Generalized gamma geometric	τ	α	k	p	1
Gamma geometric	1	α	k	p	1
Chi-square geometric	1	2	$\frac{n}{2}$	p	1
Exponential geometric	1	α	1	p	1
Weibull geometric	c	α	1	p	1
Rayleigh geometric	2	$\sigma\sqrt{2}$	1	p	1
Maxwell geometric	2	$\sigma\sqrt{2}$	$\frac{3}{2}$	p	1
Half normal geometric	2	$\sigma\sqrt{\pi}$	$\frac{1}{2}$	p	1
Generalized gamma	τ	α	k	0^+	1
Gamma	1	α	k	0^+	1
Chi-square	1	2	$\frac{n}{2}$	0^+	1
Exponential	1	α	1	0^+	1
Weibull	c	α	1	0^+	1
Rayleigh	2	$\sigma\sqrt{2}$	1	0^+	1
Maxwell	2	$\sigma\sqrt{2}$	$\frac{3}{2}$	0^+	1
Half normal	2	$\sigma\sqrt{\pi}$	$\frac{1}{2}$	0^+	1

and

$$h(x) = \frac{\frac{\lambda\tau(1-p)}{\alpha\Gamma(k)}\left(\frac{x}{\alpha}\right)^{\tau k-1}\exp\left[-\left(\frac{x}{\alpha}\right)^{\tau}\right]\left(1-p\left\{1-\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^{\tau}\right]\right\}\right)^{-2}}{1-\frac{\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^{\tau}\right]}{1-p\left\{1-\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^{\tau}\right]\right\}}},$$

respectively. Plots of the ExGGG density and hazard rate function for selected parameter values are given in Figures 1 and 2, respectively. Note in Figure 1(a) that the density function of the ExGGG distribution has very flexible shapes, especially bimodal. This is a great advantage of the distribution proposed in relation to their sub-models because none has bimodal density. The hazard rate function also presents some peculiar shapes. For instance, the blue hazard rate function in Figure 2(a) is initially increasing and then decreasing and finally increasing again.

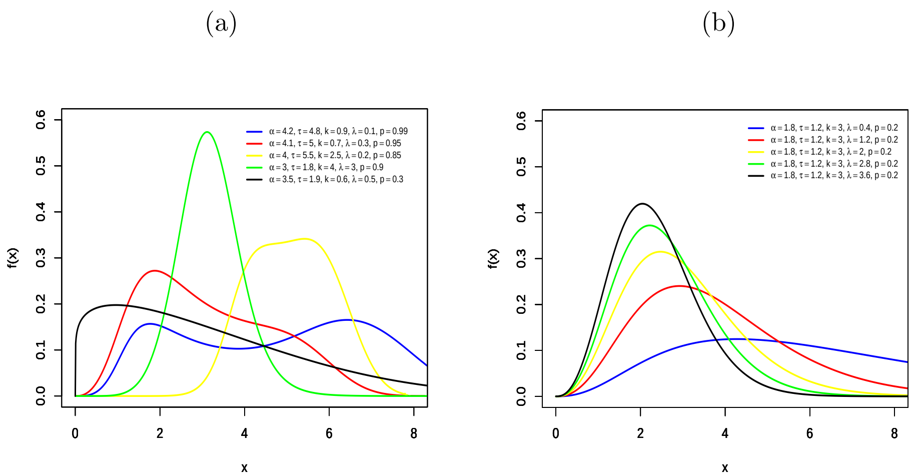

Figure 1. Plots of the ExGGG density for some parameter values.

The ExGGG distribution has an attractive physical interpretation whenever λ is a positive integer. Consider a device made of λ components independent and identically distributed according to $G(x)$ (2.1) in a series system. The device fails if any component fails. Let $X_1, \cdots, X_\lambda$ denote the lifetimes of the components, with common cdf $G(x)$. Let X denote the lifetime of the device. Thus, the cdf $F(x)$ of X is

$$\begin{aligned} F(x) &= P(X \leq x) = 1 - P(X > x) = 1 - P(X_1 > x, \cdots, X_\lambda > x) \\ &= 1 - P(X_1 > x)^\lambda = 1 - [1 - P(X_1 \leq x)]^\lambda \\ &= 1 - [1 - G(x)]^\lambda . \end{aligned}$$

So, the lifetime of the device obeys the ExGGG distribution.

3. Expansion of the density function

Now, we demonstrate that the density function (2.3) can be expressed as a linear combination of GG density functions. This result is important to provide mathematical properties of the ExGGG distribution directly from properties of the GG distribution.

Let $g_{\alpha,\tau,k}(x)$ be the density function of the GG(α, τ, k) distribution given by

$$g_{\alpha,\tau,k}\,(x) = \frac{\tau}{\alpha\Gamma(k)} \left(\frac{x}{\alpha}\right)^{\tau k - 1} \exp\Big[- \left(\frac{x}{\alpha}\right)^{\tau} \Big], \quad x > 0.$$

For $|z| < 1$ and $\rho \in \mathbb{R}$, we consider the power series

$$(1 - z)^\rho = \sum_{j=0}^{\infty} (-1)^j \binom{\rho}{j} z^j, \tag{3.1}$$

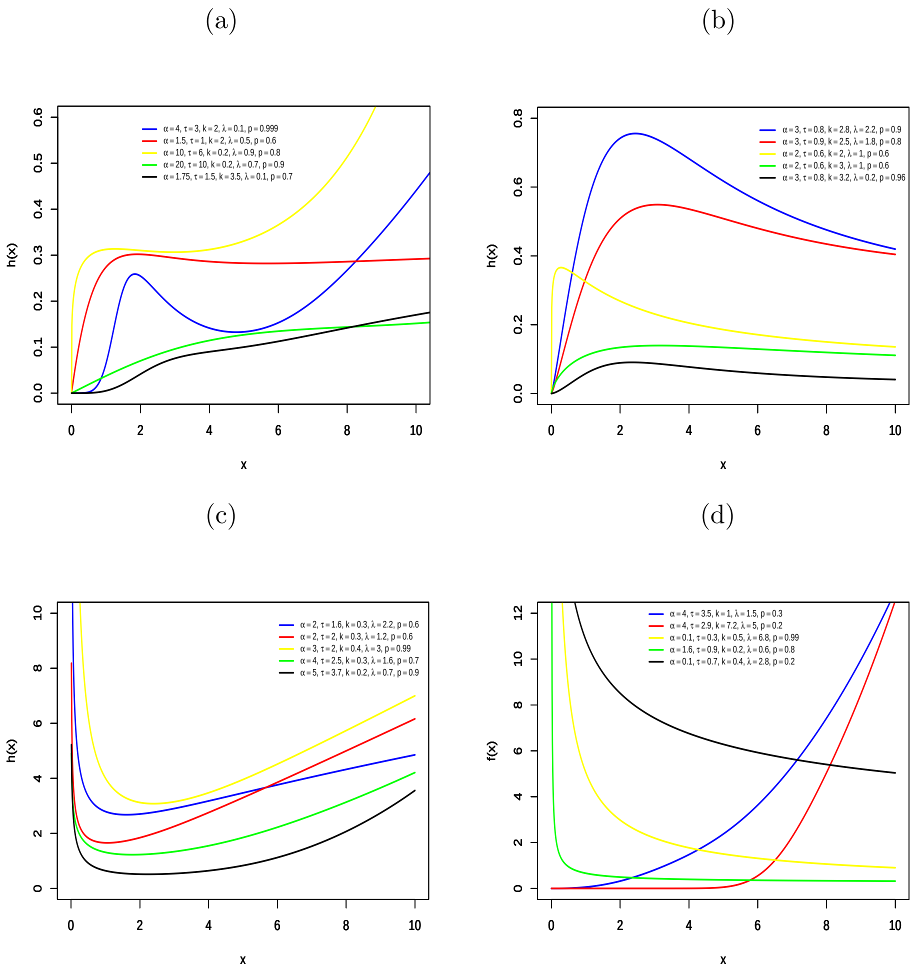

Figure 2. The ExGGG hazard rate function. (a) Plots of the hazard rate function for some parameter values. (b) Unimodal hazard rate function. (c) Bathtub hazard rate function. (d) Increasing and decreasing hazard rate function.

where $\binom{\rho}{j} = \Gamma(\rho+1)/\left[\Gamma(\rho-j+1)j!\right]$.

Considering (3.1) in (2.3), the pdf of the ExGGG$(\alpha,\tau,k,p,\lambda)$ can be written as

$$\begin{aligned} f(x) &= \frac{\lambda\tau(1-p)}{\alpha\Gamma(k)}\left(\frac{x}{\alpha}\right)^{\tau k-1}\exp\left[-\left(\frac{x}{\alpha}\right)^{\tau}\right]\left(1-p\left\{1-\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^{\tau}\right]\right\}\right)^{-2} \\ &\times \sum_{j=0}^{\infty}(-1)^j\binom{\lambda-1}{j}\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^{\tau}\right]^j\left(1-p\left\{1-\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^{\tau}\right]\right\}\right)^{-j}. \end{aligned}$$

Grouping common terms, using (3.1) and binomial expansion, we have that

$$\begin{aligned} f(x) &= \frac{\lambda\tau(1-p)}{\alpha\Gamma(k)}\left(\frac{x}{\alpha}\right)^{\tau k-1}\exp\left[-\left(\frac{x}{\alpha}\right)^{\tau}\right] \\ &\times \sum_{j,l=0}^{\infty}\sum_{m=0}^{l}(-1)^{j+l+m}\binom{\lambda-1}{j}\binom{-(j+2)}{l}\binom{l}{m}p^{l}\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^{\tau}\right]^{j+m} \end{aligned}$$

We can substitute $\sum_{j,l=0}^{\infty}\sum_{m=0}^{l}$ for $\sum_{j,m=0}^{\infty}\sum_{l=m}^{\infty}$ to obtain

$$\begin{aligned} f(x) &= \frac{\lambda\tau(1-p)}{\alpha\Gamma(k)}\left(\frac{x}{\alpha}\right)^{\tau k-1}\exp\left[-\left(\frac{x}{\alpha}\right)^{\tau}\right] \\ &\times \sum_{j,m=0}^{\infty}s_{j,m}(\lambda,p)\,\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^{\tau}\right]^{j+m}, \end{aligned} \tag{3.2}$$

where

$$s_{j,m}(\lambda,p)=\sum_{l=m}^{\infty}(-1)^{j+l+m}\binom{\lambda-1}{j}\binom{-(j+2)}{l}\binom{l}{m}p^{l} \tag{3.3}$$

Therefore, using the result (A.3) (givin in Appendix A) in the expression (3.2), the pdf $f(x)$ can be written as a linear combination of the distribution GG, in the form:

$$f(x)=\sum_{j,m,q=0}^{\infty}w_{j,m,q}(k,p,\lambda)g_{\alpha,\tau,k^{\bullet}}(x), \quad x>0 \tag{3.4}$$

where $k^{\bullet}=k(j+m+1)+q$, $g_{\alpha,\tau,k^{\bullet}}(x)$ has distribution GG$(\alpha,\tau,k^{\bullet})$, the weightings $w_{j,m,q}(k,p,\lambda)$ are given by

$$w_{j,m,q}(k,p,\lambda)=\lambda(1-p)\,s_{j,m}(\lambda,p)\,c_{j+m,q}\frac{\Gamma(k^{\bullet})}{\Gamma(k)^{j+m+1}}, \tag{3.5}$$

and the coefficients $c_{j+m,q}$ are determined from the recurrence relation (A.2) (Appendix A).

Expression (3.4) shows that the density function ExGGG distribution can be written in terms of a linear combination of densities GG.

4. Moments

Some important features of a distribution such as dispersion, skewness and kurtosis can be studied through their moments. In this section we obtain two alternative expansions for the moments of the ExGGG distribution. Initially, we know that the rth ordinary moment of the GG(α,τ,k) distribution, denoted by $\mu'_{r,GG}$, is

$$\mu'_{r,GG}=\frac{\alpha^{r}\Gamma(k+r/\tau)}{\Gamma(k)}. \tag{4.1}$$

Now, follows from expressions (3.4) and (4.1), the rth moment ordinary of the ExGGG$(\alpha,\tau,k,p,\lambda)$ is given by

$$\mu_r' = \alpha^r \sum_{j,m,q=0}^{\infty} w_{j,m,q}(k,p,\lambda)\frac{\Gamma\left(k^{\bullet}+r/\tau\right)}{\Gamma\left(k^{\bullet}\right)}. \tag{4.2}$$

The expression (4.2) depends on the quantities $c_{j+m,q}$ which are obtained recursively by (A.2).

Another infinite sum representation for μ_r' is obtained computing the moment directly, that is

$$\begin{aligned}
\mu_r' &= \frac{\lambda\tau\left(1-p\right)}{\alpha\Gamma\left(k\right)}\int_0^{\infty} x^r\left(\frac{x}{\alpha}\right)^{\tau k-1}\exp\left[-\left(\frac{x}{\alpha}\right)^{\tau}\right]\\
&\times \left(1-p\left\{1-\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^{\tau}\right]\right\}\right)^{-2}\\
&\times \left\{1-\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^{\tau}\right]\left(1-p\left\{1-\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^{\tau}\right]\right\}\right)^{-1}\right\}^{\lambda-1}dx.
\end{aligned}$$

Setting $y=\left(\frac{x}{\alpha}\right)^{\tau}$ in the last expression,

$$\begin{aligned}
\mu_r' &= \frac{\lambda\left(1-p\right)\alpha^r}{\Gamma\left(k\right)}\int_0^{\infty} y^{k+\frac{r}{\tau}-1}\exp\left(-y\right)\left\{1-p\left[1-\gamma_1\left(k,y\right)\right]\right\}^{-2}\\
&\times \left\{1-\gamma_1\left(k,y\right)\left\{1-p\left[1-\gamma_1\left(k,y\right)\right]\right\}^{-1}\right\}^{\lambda-1}dy.
\end{aligned} \tag{4.3}$$

Considering (3.1) in (4.3) twice conveniently, we have that

$$\begin{aligned}
\mu_r' &= \frac{\lambda\left(1-p\right)\alpha^r}{\Gamma\left(k\right)}\int_0^{\infty} y^{k+\frac{r}{\tau}-1}\exp\left(-y\right)\\
&\times \sum_{j,l=0}^{\infty}\left(-1\right)^{j+l}\binom{\lambda-1}{j}\binom{-\left(j+2\right)}{l}\gamma_1\left(k,y\right)^j p^l\left[1-\gamma_1\left(k,y\right)\right]^l dy.
\end{aligned}$$

Using the binomial expansion in the term $\left[1-\gamma_1\left(k,y\right)\right]^l$, the last expression is rewritten as

$$\begin{aligned}
\mu_r' &= \frac{\lambda\left(1-p\right)\alpha^r}{\Gamma\left(k\right)}\sum_{j,l=0}^{\infty}\sum_{m=0}^{l}\left(-1\right)^{j+l+m}\binom{\lambda-1}{j}\binom{-\left(j+2\right)}{l}\binom{l}{m}p^l\\
&\times \int_0^{\infty} y^{k+\frac{r}{\tau}-1}\exp\left(-y\right)\gamma_1\left(k,y\right)^{j+m}dy.
\end{aligned} \tag{4.4}$$

Replacing $\sum_{j,l=0}^{\infty}\sum_{m=0}^{l}$ for $\sum_{j,m=0}^{\infty}\sum_{l=m}^{\infty}$ in the expression (4.5), we have

$$\begin{aligned}
\mu_r' &= \frac{\lambda\left(1-p\right)\alpha^r}{\Gamma\left(k\right)}\sum_{j,m=0}^{\infty}\sum_{l=m}^{\infty}\left(-1\right)^{j+l+m}\binom{\lambda-1}{j}\binom{-\left(j+2\right)}{l}\binom{l}{m}p^l\\
&\times \int_0^{\infty} y^{k+\frac{r}{\tau}-1}\exp\left(-y\right)\gamma_1\left(k,y\right)^{j+m}dy.
\end{aligned}$$

Therefore μ_r' can be rewritten as

$$\mu_r' = \frac{\lambda (1-p)\, \alpha^r}{\Gamma(k)} s_{j,m}(\lambda, p)\, I\left(k + \frac{r}{\tau}, j+m\right),$$

where $s_{j,m}(\lambda, p)$ is defined by expression (3.3), and

$$I\left(k + \frac{r}{\tau}, j+m\right) = \int_0^\infty y^{k+\frac{r}{\tau}-1} \exp(-y)\, \gamma_1(k, y)^{j+m}\, dy.$$

This integral can be determined from expressions (24) and (25) of Nadarajah [23] in terms of the Lauricella function of type A [14, 1]defined by

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

where $(a)_i$ is the ascending factorial defined by $(a)_i = a(a+1)\cdots(a+i-1)$ assuming $(a)_0 = 1$. Numerical routines for the direct computation of the Lauricella function of type A are available [14, 32]. We obtain

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

The graphic representations of the skewness and kurtosis measures in terms of λ for selected values of α, τ, k and p, are shown in Figure 3.

(a) (b)

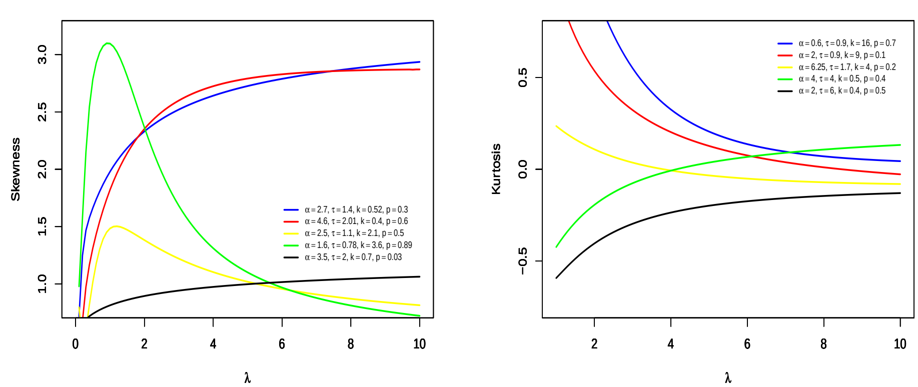

Figure 3. Skewness and kurtosis of the ExGGG distribution as a function of the parameter λ.

5. Moment Generating Function

Here, we provide two expressions for the mgf of ExGGG distribution based on mgf of GG distribution.

Let $M_{\alpha,\tau,k}\left(s\right)=E\left[\exp\left(sX\right)\right]$ be mgf of $X\sim GG\left(\alpha,\tau,k\right)$. We can write

$$M_{\alpha,\tau,k}(s)=\frac{\tau}{\alpha^{\tau k}\Gamma(k)}\int_0^\infty \exp(sx)x^{\tau k-1}\exp\{-(x/\alpha)^\tau\}dx.$$

Setting $u=x/\alpha$, we have

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

Expanding the first exponential in Taylor series and using $\int_0^\infty u^{k\tau+d-1}\exp\left(-u^\tau\right)du=\tau^{-1}\Gamma(k+d/\tau)$, we obtain

$$M_{\alpha,\tau,k}(s)=\frac{1}{\Gamma(k)}\sum_{d=0}^{\infty}\Gamma\left(\frac{d}{\tau}+k\right)\frac{(\alpha s)^d}{d!}. \tag{5.1}$$

Using the result in (3.4), the mgf of ExGGG$(\alpha,\tau,k,p,\lambda)$ is given by

$$M(s)=\sum_{j,m,q,d=0}^{\infty} w_{j,m,q}(k,p,\lambda)M_{\alpha,\tau,k^\bullet}(s),$$

where $k^\bullet=k(j+m+1)+q$ and $w_{j,m,q}(k,p,\lambda)$ is given by (3.5).

Therefore,

$$M(s)=\sum_{j,m,q,d=0}^{\infty}\frac{w_{j,m,q}(k,p,\lambda)}{\Gamma(k^\bullet)}\Gamma\Big(\frac{d}{\tau}+k^\bullet\Big)\frac{(\alpha s)^d}{d!}.$$

However, for $\tau>1$, it can be simplified by considering the Wright generalized hypergeometric function [34] defined by

$${}_p\Psi_q\left[\begin{array}{c}\big(\alpha_1,A_1\big),\cdots,\big(\alpha_p,A_p\big)\\ \big(\beta_1,B_1\big),\cdots,\big(\beta_q,B_q\big)\end{array};x\right]=\sum_{m=0}^{\infty}\frac{\prod_{j=1}^{p}\Gamma(\alpha_j+A_j\,d)}{\prod_{j=1}^{q}\Gamma(\beta_j+B_j\,d)}\frac{x^d}{d!}. \tag{5.2}$$

This function exists if $1+\sum_{j=1}^{q}B_j-\sum_{j=1}^{p}A_j>0$. Combining the results in (5.2) to rewrite (5.1), we have

$$M_{\alpha,\tau,k}(s)=\frac{1}{\Gamma(k)}{}_1\Psi_0\left[\begin{array}{c}(k,1/\tau)\\ -\end{array};\alpha s\right]. \tag{5.3}$$

Finally, the mgf of ExGGG can be written from expressions (3.4) and (5.3) as

$$M(s)=\sum_{j,m,q,d=0}^{\infty}\frac{w_{j,m,q}(k,p,\lambda)}{\Gamma(k^\bullet)}{}_1\Psi_0\left[\begin{array}{c}(k^\bullet,1/\tau)\\ -\end{array};\alpha s\right].$$

6. Mean Deviations

The amount of scatter in a population is evidently measured to some extent by the totality of deviations from the mean and median. If X has the ExGGG distribution with density function $f(x)$, we can derive the mean deviations about the mean $\mu_1' = E(X)$ and about the median m_1 from the relations

$$\delta_1 = \int_0^\infty |x - \mu_1'| f(x) dx \quad \text{and} \quad \delta_2 = \int_0^\infty |x - m_1| f(x) dx.$$

The measures δ_1 and δ_2 can be expressed as

$$\delta_1 = 2\mu_1' F(\mu_1') - 2I(\mu_1') \quad \text{and} \quad \delta_2 = \mu_1' - 2I(m_1), \tag{6.1}$$

where $I(s) = \int_0^s x f(x) dx$ and $F(\mu_1')$ is calculated from (2.2). The integral $I(m_1)$ can be obtained from the expression (3.4) as

$$\begin{aligned} I(m_1) &= \sum_{j,m,q=0}^{\infty} w_{j,m,q}(k,p,\lambda) \int_0^{m_1} x g_{\alpha,\tau,k^\bullet}(x) dx \\ &= \sum_{j,m,q=0}^{\infty} w_{j,m,q}(k,p,\lambda) J(\alpha,\tau,k^\bullet,m_1). \end{aligned} \tag{6.2}$$

By setting $u = x/\alpha$ in the expression (6.2), we obtain

$$J(\alpha,\tau,k^\bullet,m_1) = \frac{\alpha\tau}{\Gamma(k^\bullet)} \int_0^{m_1/\alpha} u^{\tau k^\bullet} \exp(-u^\tau) du.$$

The substitution $w = u^\tau$ yields $J(\alpha,\tau,k^\bullet,m_1)$ in terms of the incomplete gamma function

$$\begin{aligned} J(\alpha,\tau,k^\bullet,m_1) &= \frac{\alpha}{\Gamma(k^\bullet)} \int_0^{(m_1/\alpha)^\tau} w^{k^\bullet+\tau^{-1}-1} \exp(-w) dw \\ &= \frac{\alpha}{\Gamma(k^\bullet)} \gamma[k^\bullet + \tau^{-1}, (m_1/\alpha)^\tau] \end{aligned}$$

Hence, inserting the last result into (6.2) gives

$$I(m_1) = \sum_{j,m,q=0}^{\infty} \frac{\alpha w_{j,m,q}(k,p,\lambda)}{\Gamma(k^\bullet)} \gamma[k^\bullet + \tau^{-1}, (m_1/\alpha)^\tau].$$

The ExGGG mean deviations follow from (6.1) and the last expression. The result is analogous to $I(\mu_1')$.

7. Reliability

In the context of reliability, the stress-strength model describes the life of a component which has a random strength X_1 that is subjected to a random stress X_2. The component fails at the instant that the stress applied to it exceeds the strength, and the component will function satisfactorily whenever $X_1 > X_2$. Hence, $R = P(X_2 < X_1)$ is a measure of component reliability. It has many applications especially in engineering concepts such as structures, deterioration of rocket motors, static fatigue of ceramic components, fatigue failure of aircraft structures, and the aging of concrete pressure vessels. We now derive the corresponding form for the reliability R when X_1 and X_2 are independent and have identical ExGGG distribution. The reliability of the ExGGG distribution is

$$R = \int_0^\infty f(x)F(x)dx,$$

where $f(x)$ and $F(x)$ are calculated from (3.4) and (2.2), respectively. The reliability can be written explicitly as follows

$$\begin{aligned} R &= \int_0^\infty \sum_{j,m,q=0}^\infty w_{j,m,q}(k,p,\lambda) g_{\alpha,\tau,k^\bullet}(x) \\ &\times \left(1 - \left\{1 - \frac{\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^\tau\right]}{1-p\left\{1-\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^\tau\right]\right\}}\right\}^\lambda\right) dx \\ &= \sum_{j,m,q=0}^\infty w_{j,m,q}(k,p,\lambda) \\ &\times \left(1 - \int_0^\infty g_{\alpha,\tau,k^\bullet}(x) \left\{1 - \frac{\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^\tau\right]}{1-p\left\{1-\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^\tau\right]\right\}}\right\}^\lambda dx\right). \end{aligned}$$

Using the expansion (3.1) twice conveniently in the expression above, we have

$$\begin{aligned} R &= \sum_{j,m,q=0}^\infty w_{j,m,q}(k,p,\lambda)\left(1 - \int_0^\infty g_{\alpha,\tau,k^\bullet}(x)\right. \\ &\left. \times \sum_{r,u=0}^\infty (-1)^{r+u}\binom{\lambda}{r}\binom{-r}{u}p^u\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^\tau\right]^r \left\{1-\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^\tau\right]\right\}^u dx\right). \end{aligned} \tag{7.1}$$

Using the binomial expansion in expression (7.1), we have

$$\begin{aligned} R &= \sum_{j,m,q=0}^\infty w_{j,m,q}(k,p,\lambda)\left\{1 - \int_0^\infty g_{\alpha,\tau,k^\bullet}(x)\right. \\ &\left. \times \sum_{r,u=0}^\infty \sum_{t=0}^u (-1)^{r+u+t}\binom{\lambda}{r}\binom{-r}{u}\binom{u}{t}p^u\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^\tau\right]^{r+t} dx\right\}. \end{aligned}$$

We can substitute $\sum_{r,u=0}^\infty \sum_{t=0}^u$ for $\sum_{r,t=0}^\infty \sum_{u=t}^\infty$ to obtain

$$\begin{aligned} R &= \sum_{j,m,q=0}^\infty w_{j,m,q}(k,p,\lambda) \\ &\times \left\{1 - \int_0^\infty g_{\alpha,\tau,k^\bullet}(x) \sum_{r,t=0}^\infty s_{r,t}(p,\lambda)\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^\tau\right]^{r+t} dx\right\}, \end{aligned}$$

where

$$s_{r,t}(p,\lambda) = \sum_{u=t}^\infty (-1)^{r+u+t}\binom{\lambda}{r}\binom{-r}{u}\binom{u}{t}p^u.$$

How $g_{\alpha,\tau,k^\bullet}(x)$ has distribution $\mathrm{GG}(\alpha,\tau,k^\bullet)$, we obtain

$$R=\sum_{j,m,q=0}^{\infty} w_{j,m,q}(k,p,\lambda)\Bigg\{1-\int_0^\infty \frac{\tau}{\alpha\Gamma\left(k^\bullet\right)}\left(\frac{x}{\alpha}\right)^{\tau k^\bullet-1}$$

$$\times\exp\left[-\left(\frac{x}{\alpha}\right)^{\tau}\right]\sum_{r,t=0}^{\infty} s_{r,t}(p,\lambda)\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^{\tau}\right]^{r+t}dx\Bigg\}. \tag{7.2}$$

Finally, setting $w=\left(\frac{x}{\alpha}\right)^{\tau}$ in (7.2), we have

$$R=\sum_{j,m,q=0}^{\infty} w_{j,m,q}(k,p,\lambda)\left\{1-\sum_{r,t=0}^{\infty}\frac{s_{r,t}(p,\lambda)}{\Gamma\left(k^\circ+q\right)}I\left(k^\circ+q,r+t\right)\right\},$$

where

$$I\left(k^\circ+q,r+t\right)=\int_0^\infty w^{k^\circ+q-1}\exp(-w)\gamma_1(k,w)^{r+t}dw,$$

$k^\circ=k(j+m+1)$ and $w_{j,m,q}(k,p,\lambda)$ is defined by expression (3.5). Using the Lauricella function of type A (defined in Section 4), the last integral can be written as

$$I\left(k^\circ+q,r+t\right)=\left(k^\circ\right)^{-r-t}\Gamma\left[q+k^\circ(r+t+1)\right]$$
$$\times F_A^{(r+t)}(q+k^\circ(r+t+1);k^\circ,\cdots,k^\circ;k^\circ+1,\cdots,k^\circ+1;-1,\cdots,-1).$$

8. Order Statistics

The density function $f_{i:n}(x)$ of the ith order statistic, say $X_{i:n}$, for $i=1,\ldots,n$, from random variables $X_1,\ldots,X_n$ having density (2.3), is given by

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

where $f(x)$ and $F(x)$ are the pdf and cdf of the ExGGG distribution, respectively and $B(\cdot,\cdot)$ denotes the beta function. We readily obtain using the binomial expansion

$$f_{i:n}(x)=\frac{1}{B(i,n-i+1)}f(x)\sum_{j_1=0}^{n-i}(-1)^{j_1}\binom{n-i}{j_1}F(x)^{i+j_1-1}.$$

However, if $F(x)$ is the cdf of ExGGG distribution defined in (2.2) has

$$F(x)^{i+j_1-1}=\left(1-\left\{1-\frac{\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^{\tau}\right]}{1-p\left\{1-\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^{\tau}\right]\right\}}\right\}^{\lambda}\right)^{i+j_1-1}.$$

Using the binomial expansion in the last expression, we obtain

$$
\begin{aligned}
F(x)^{i+j_1-1} &= \sum_{l_1=0}^{i+j_1-1} (-1)^{l_1} \binom{i+j_1-1}{l_1} \\
&\times \left\{ 1 - \frac{\gamma_1\left[k, \left(\frac{x}{\alpha}\right)^{\tau}\right]}{1 - p\left\{1-\gamma_1\left[k, \left(\frac{x}{\alpha}\right)^{\tau}\right]\right\}} \right\}^{l_1\lambda}. \qquad (8.1)
\end{aligned}
$$

Considering (3.1) in (8.1) twice conveniently, we have that

$$
\begin{aligned}
F(x)^{i+j_1-1} &= \sum_{l_1=0}^{i+j_1-1} \sum_{s,a=0}^{\infty} (-1)^{l_1+s+a} \binom{i+j_1-1}{l_1} \binom{l_1\lambda}{s} \binom{-s}{a} p^a \\
&\times \gamma_1\left[k, \left(\frac{x}{\alpha}\right)^{\tau}\right]^s \left\{1-\gamma_1\left[k, \left(\frac{x}{\alpha}\right)^{\tau}\right]\right\}^a.
\end{aligned}
$$

Now, using the binomial expansion in the expression $\left\{1-\gamma_1\left[k, \left(\frac{x}{\alpha}\right)^{\tau}\right]\right\}^a$, we have

$$
\begin{aligned}
F(x)^{i+j_1-1} &= \sum_{l_1=0}^{i+j_1-1} \sum_{s,a=0}^{\infty} \sum_{b=0}^{a} (-1)^{l_1+s+a+b} \\
&\times \binom{i+j_1-1}{l_1} \binom{l_1\lambda}{s} \binom{-s}{a} \binom{a}{b} p^a \gamma_1\left[k, \left(\frac{x}{\alpha}\right)^{\tau}\right]^{s+b}.
\end{aligned}
$$

We can substitute $\sum_{a=0}^{\infty}\sum_{b=0}^{a}$ for $\sum_{b=0}^{\infty}\sum_{a=b}^{\infty}$ to obtain

$$
F(x)^{i+j_1-1} = \sum_{l_1=0}^{i+j_1-1} \sum_{s,b=0}^{\infty} s_{l_1,s,b}(\lambda,p) \binom{i+j_1-1}{l_1} \gamma_1\left[k, \left(\frac{x}{\alpha}\right)^{\tau}\right]^{s+b}, \qquad (8.2)
$$

where

$$
s_{l_1,s,b}(\lambda,p) = \sum_{a=b}^{\infty} (-1)^{l_1+s+a+b} \binom{l_1\lambda}{s} \binom{-s}{a} \binom{a}{b} p^n.
$$

The expression (8.2) can be rewritten as

$$
F(x)^{i+j_1-1} = \sum_{s,b=0}^{\infty} \rho_{s,b,i+j_1-1}(\lambda,p) \gamma_1\left[k, \left(\frac{x}{\alpha}\right)^{\tau}\right]^{s+b},
$$

where

$$
\rho_{s,b,u}(\lambda,p) = \sum_{l_1=0}^{u} s_{l_1,s,b}(\lambda,p) \binom{u}{l_1}. \qquad (8.3)
$$

For $s, b, u = 0, 1, \ldots$, the density function of the *ith* ExGGG order statistic becomes

$$
\begin{aligned}
f_{i:n}(x) &= \frac{1}{B(i, n-i+1)} f(x) \sum_{j_1=0}^{n-i} (-1)^{j_1} \binom{n-i}{j_1} \\
&\times \sum_{s,b=0}^{\infty} \rho_{s,b,i+j_1-1}(\lambda,p) \gamma_1\left[k, \left(\frac{x}{\alpha}\right)^{\tau}\right]^{s+b}.
\end{aligned}
$$

Using the density (3.4), $f_{i:n}(x)$ can be written as

$$\begin{aligned} f_{i:n}(x) \quad &= \quad \frac{1}{B(i,n-i+1)} \sum_{j,m,q,s,b=0}^{\infty} \sum_{j_1=0}^{n-i} (-1)^{j_1} \binom{n-i}{j_1} w_{j,m,q}(k,p,\lambda) \\ &\times \quad \rho_{s,b,i+j_1-1}(\lambda,p)\gamma_1 \left[k, \left(\frac{x}{\alpha}\right)^{\tau}\right]^{s+b} g_{\alpha,\tau,k^{\bullet}}(x). \end{aligned} \tag{8.4}$$

By applying expansion (A.3) in the expression (8.4), the density $f_{i:n}(x)$ is expressed by

$$f_{i:n}(x) = \frac{1}{B(i,n-i+1)} \sum_{j,m,q,s,b,v=0}^{\infty} \sum_{j_1=0}^{n-i} \frac{(-1)^{j_1}\binom{n-i}{j_1}}{\Gamma(k)^{s+b}\alpha^{\tau[k(s+b)+v]}} c_{s+b,v} w_{j,m,q}(k,p,\lambda)$$
$$\times \rho_{s,b,i+j_1-1}(\lambda,p) x^{\tau[k(s+b)+v]} g_{\alpha,\tau,k^{\bullet}}(x),$$

on which the quantities $w_{j,m,q}(k,p,\lambda)$ and $\rho_{s,b,i+j_1-1}(\lambda,p)$ are defined in (3.5) and (8.3), $c_{s+b,v}$ is calculated recursively as (A.2).

9. Inference and estimation

In this Section, we discuss the maximum likelihood method and Bayesian approach for the inference and estimation of the ExGGG parameter model. We also assess the performance of the maximum likelihood method for estimating the ExGGG parameters using Monte Carlo simulation.

9.1. Maximum likelihood estimation. Here, we consider the estimation of the parameters of the ExGGG distribution by maximum likelihood method.

Let X_i be a random variable following (2.3) with the vector of parameters $\boldsymbol{\theta} = (\alpha, \tau, k, p, \lambda)^T$. Suppose that the data consist of n independent observations x_i of X_i for $i = 1, \dots, n$. Parametric inference for such data are usually based on likelihood methods and their asymptotic theory. The log-likelihood $\ell(\boldsymbol{\theta})$ for the model parameters can be expressed as

$$\begin{aligned} \ell(\boldsymbol{\theta}) = n\log\left[\frac{\lambda\tau\,(1-p)}{\alpha\Gamma\,(k)}\right] &+ (\tau k - 1)\sum_{i=1}^{n}\log\left(\frac{x_i}{\alpha}\right) - \sum_{i=1}^{n}\left(\frac{x_i}{\alpha}\right)^{\tau} \\ &- 2\sum_{i=1}^{n}\log\left(1 - p\left\{1 - \gamma_1\left[k, \left(\frac{x_i}{\alpha}\right)^{\tau}\right]\right\}\right) \\ &+ (\lambda - 1)\sum_{i=1}^{n}\log\left\{1 - \frac{\gamma_1\left[k, \left(\frac{x_i}{\alpha}\right)^{\tau}\right]}{1 - p\left\{1 - \gamma_1\left[k, \left(\frac{x_i}{\alpha}\right)^{\tau}\right]\right\}}\right\}. \end{aligned} \tag{9.1}$$

The score components corresponding to the parameters in $\boldsymbol{\theta}$ are

$$U_{\alpha}(\boldsymbol{\theta}) = -\frac{n\tau k}{\alpha} + \frac{\tau}{\alpha}\sum_{i=1}^{n} u_i + \frac{\tau}{\alpha\Gamma\,(k)}\sum_{i=1}^{n}(2pq_i + \lambda - 1)\,\frac{g_i}{q_i} u_i^k \exp(-u_i),$$

$$U_\tau(\boldsymbol{\theta}) = \frac{n}{\tau} + \frac{1}{\tau}\sum_{i=1}^{n}(k-u_i)\log u_i - \frac{2p}{\tau\Gamma(k)}\sum_{i=1}^{n} g_i u_i^k \exp(-u_i)\log u_i$$
$$-\frac{(\lambda-1)}{\tau\Gamma(k)}\sum_{i=1}^{n}\frac{g_i}{q_i}u_i^k \exp(-u_i)\log u_i,$$

$$U_k(\boldsymbol{\theta}) = -n\psi(k) + \sum_{i=1}^{n}\log u_i - 2p\sum_{i=1}^{n} g_i\left[w_i - \psi(k)\gamma_1(k,u_i)\right]$$
$$-(\lambda-1)\sum_{i=1}^{n}\frac{g_i}{q_i}\left[w_i - \psi(k)\gamma_1(k,u_i)\right],$$

$$U_p(\boldsymbol{\theta}) = -\frac{n}{1-p} + 2\sum_{i=1}^{n} g_i q_i - \frac{(\lambda-1)}{1-p}\sum_{i=1}^{n} g_i\gamma_1(k,u_i)$$

and

$$U_\lambda(\boldsymbol{\theta}) = \frac{n}{\lambda} + \sum_{i=1}^{n}\log\left[1-g_i\gamma_1(k,u_i)\right],$$

where,

$$u_i = \left(\frac{x_i}{\alpha}\right)^\tau, \quad q_i = 1-\gamma_1(k,u_i), \quad g_i = \frac{1}{1-pq_i}, \quad w_i = \frac{\dot\gamma(k,u_i)_k}{\Gamma(k)},$$

$$\dot\gamma(k,u_i)_k = \sum_{n=0}^{\infty}\frac{(-1)^n}{n!}J(u_i,k+n-1,1),$$

$\psi(.)$ is the digamma function and $J(u_i,k+n-1,1)$ is defined in Appendix B.

The maximum likelihood estimates (MLEs) $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is obtained numerically from the nonlinear equations

$$U_\alpha(\boldsymbol{\theta}) = U_\tau(\boldsymbol{\theta}) = U_k(\boldsymbol{\theta}) = U_p(\boldsymbol{\theta}) = U_\lambda(\boldsymbol{\theta}) = 0.$$

For interval estimation and hypothesis tests on the model parameters, we require the 5×5 unit observed information matrix, say

$$J = J(\boldsymbol{\theta}) = \begin{bmatrix} J_{\alpha\alpha} & J_{\alpha\tau} & J_{\alpha k} & J_{\alpha p} & J_{\alpha\lambda} \\ J_{\tau\alpha} & J_{\tau\tau} & J_{\tau k} & J_{\tau p} & J_{\tau\lambda} \\ J_{k\alpha} & J_{k\tau} & J_{kk} & J_{kp} & J_{k\lambda} \\ J_{p\alpha} & J_{p\tau} & J_{pk} & J_{pp} & J_{p\lambda} \\ J_{\lambda\alpha} & J_{\lambda\tau} & J_{\lambda k} & J_{\lambda p} & J_{\lambda\lambda} \end{bmatrix},$$

whose elements are given in Appendix B.

Under conditions that are fulfilled for parameters in the interior of the parameter space but not on the boundary, the asymptotic distribution of $\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})$ is $N_5(0, I(\boldsymbol{\theta})^{-1})$, where $I(\boldsymbol{\theta})$ is the expected information matrix. This matrix can be replaced by $J(\widehat{\boldsymbol{\theta}})$, i.e., the observed information matrix evaluated $\widehat{\boldsymbol{\theta}}$. The multivariate normal $N_5(0, J(\widehat{\boldsymbol{\theta}})^{-1})$ distribution can be used to construct approximate confidence intervals for the individual parameters.

We can compute the maximum values of the unrestricted and restricted log-likelihoods to construct likelihood ratio (LR) statistics for testing some submodels (see Section 2) of the ExGGG distribution. For example, we may use LR statistics to check if the fit using the ExGGG distribution is statistically "superior" to the fit using the GGG distribution for a given data set. That is, to test $H_0 : \lambda = 1$ versus $H_1 : \lambda \neq 1$ the LR statistics is

$$LR = 2\left\{\ell(\hat{\alpha}, \hat{\tau}, \hat{k}, \hat{p}, \hat{\lambda}) - \ell(\tilde{\alpha}, \tilde{\tau}, \tilde{k}, \tilde{p}, 1)\right\},$$

where $\hat{\alpha}, \hat{\tau}, \hat{k}, \hat{p}$ and $\hat{\lambda}$ are the MLEs under H_1 and $\tilde{\alpha}, \tilde{\tau}, \tilde{k}$ and $\tilde{p}$ are the estimates under H_0.

9.2. Bayesian approach. In the Bayesian approach, the information referring to the model parameters is obtained through a posterior marginal distribution. Here, we use the simulation method of Markov Chain Monte Carlo (MCMC), such as the Metropolis-Hastings algorithm. Since we have no prior information from historical data or from previous experiment, we assign conjugate but weakly informative prior distributions to the parameters. Since we assumed informative (but weakly) prior distribution, the posterior distribution is a well-defined proper distribution. Here, we assume that the elements of the parameter vector to be independent and consider that the joint prior distribution of all unknown parameters has a density function given by

$$\pi(\alpha, \tau, k, p, \lambda) \propto \pi(\alpha) \times \pi(\tau) \times \pi(k) \times \pi(p) \times \pi(\lambda). \tag{9.2}$$

Here, $\alpha \sim \Gamma(c_1, b_1)$, $\tau \sim \Gamma(c_2, b_2)$, $k \sim \Gamma(c_3, b_3)$, $\lambda \sim \Gamma(c_4, b_4)$ and $p \sim \mathrm{Be}(a, b)$, where $\mathrm{Be}(a, b)$ denotes a beta distribution with mean $\frac{a}{a+b}$, variance $\frac{ab}{(a+b)^2(a+b+1)}$ and density function given by

$$f(\upsilon; a, b) = \frac{1}{\mathrm{B}(a, b)}\upsilon^{a-1}(1-\upsilon)^{b-1},$$

where $\upsilon \in (0, 1)$, $a > 0$ and $b > 0$, $\Gamma(c_i, b_i)$ denotes a gamma distribution with mean c_i/b_i, variance c_i/b_i^2 and density function given by

$$f(\upsilon; c_i, b_i) = \frac{b_i^{c_i}\upsilon^{c_i-1}\mathrm{e}^{-\upsilon \mathrm{b_i}}}{\Gamma(c_i)},$$

where $\upsilon > 0$, $c_i > 0$ and $b_i > 0$. All hyper-parameters are specified. Combining the likelihood function (9.1) and the prior distribution (9.2), the joint posterior distribution for α, τ, k, p and λ reduces to

$$\begin{aligned}\pi(\alpha, \tau, k, p, \lambda|x) \propto n\log\left[\frac{\lambda\tau\,(1-p)}{\alpha\Gamma\,(k)}\right] &+ (\tau k - 1)\sum_{i=1}^{n}\log\left(\frac{x_i}{\alpha}\right) - \sum_{i=1}^{n}\left(\frac{x_i}{\alpha}\right)^{\tau} \\ &- 2\sum_{i=1}^{n}\log\left(1 - p\left\{1 - \gamma_1\left[k, \left(\frac{x_i}{\alpha}\right)^{\tau}\right]\right\}\right) \\ &+ (\lambda - 1)\sum_{i=1}^{n}\log\left\{1 - \frac{\gamma_1\left[k, \left(\frac{x_i}{\alpha}\right)^{\tau}\right]}{1 - p\left\{1 - \gamma_1\left[k, \left(\frac{x_i}{\alpha}\right)^{\tau}\right]\right\}}\right\} \\ &+ \log\left[\pi(\alpha, \tau, k, p, \lambda)\right]. \end{aligned} \tag{9.3}$$

The joint posterior density (9.3) is analytically intractable because the integration of the joint posterior density is not easy to perform. So, the inference can be based on MCMC simulation methods such as the Gibbs sampler and Metropolis-Hastings algorithm, which can be used to draw samples, from which features of the marginal distributions of interest can be inferred. In this direction, we first obtain the full conditional distributions of each unknown quantity, which are given by

$$\begin{aligned}\pi(\alpha|x,\tau,k,p,\lambda) \propto & -n\log(\alpha) + (\tau k - 1)\sum_{i=1}^{n}\log\left(\frac{x_i}{\alpha}\right) - \sum_{i=1}^{n}\left(\frac{x_i}{\alpha}\right)^{\tau} \\ & - 2\sum_{i=1}^{n}\log\left(1 - p\left\{1 - \gamma_1\left[k, \left(\frac{x_i}{\alpha}\right)^{\tau}\right]\right\}\right) \\ & + (\lambda - 1)\sum_{i=1}^{n}\log\left\{1 - \frac{\gamma_1\left[k, \left(\frac{x_i}{\alpha}\right)^{\tau}\right]}{1 - p\left\{1 - \gamma_1\left[k, \left(\frac{x_i}{\alpha}\right)^{\tau}\right]\right\}}\right\} \\ & + \log\left[\pi(\alpha)\right],\end{aligned}$$

$$\begin{aligned}\pi(\tau|x,\alpha,k,p,\lambda) \propto & \; n\log(\tau) + (\tau k)\sum_{i=1}^{n}\log\left(\frac{x_i}{\alpha}\right) - \sum_{i=1}^{n}\left(\frac{x_i}{\alpha}\right)^{\tau} \\ & - 2\sum_{i=1}^{n}\log\left(1 - p\left\{1 - \gamma_1\left[k, \left(\frac{x_i}{\alpha}\right)^{\tau}\right]\right\}\right) \\ & + (\lambda - 1)\sum_{i=1}^{n}\log\left\{1 - \frac{\gamma_1\left[k, \left(\frac{x_i}{\alpha}\right)^{\tau}\right]}{1 - p\left\{1 - \gamma_1\left[k, \left(\frac{x_i}{\alpha}\right)^{\tau}\right]\right\}}\right\} \\ & + \log\left[\pi(\tau)\right],\end{aligned}$$

$$\begin{aligned}\pi(k|x,\alpha,\tau,p,\lambda) \propto & -n\log\left[\Gamma(k)\right] + (\tau k)\sum_{i=1}^{n}\log\left(\frac{x_i}{\alpha}\right) \\ & - 2\sum_{i=1}^{n}\log\left(1 - p\left\{1 - \gamma_1\left[k, \left(\frac{x_i}{\alpha}\right)^{\tau}\right]\right\}\right) \\ & + (\lambda - 1)\sum_{i=1}^{n}\log\left\{1 - \frac{\gamma_1\left[k, \left(\frac{x_i}{\alpha}\right)^{\tau}\right]}{1 - p\left\{1 - \gamma_1\left[k, \left(\frac{x_i}{\alpha}\right)^{\tau}\right]\right\}}\right\} \\ & + \log\left[\pi(k)\right],\end{aligned}$$

$$\begin{aligned}\pi(p|x,\alpha,\tau,k,\lambda) \propto & \; n\log(1 - p) - 2\sum_{i=1}^{n}\log\left(1 - p\left\{1 - \gamma_1\left[k, \left(\frac{x_i}{\alpha}\right)^{\tau}\right]\right\}\right) \\ & + (\lambda - 1)\sum_{i=1}^{n}\log\left\{1 - \frac{\gamma_1\left[k, \left(\frac{x_i}{\alpha}\right)^{\tau}\right]}{1 - p\left\{1 - \gamma_1\left[k, \left(\frac{x_i}{\alpha}\right)^{\tau}\right]\right\}}\right\} \\ & + \log\left[\pi(p)\right]\end{aligned}$$

and

$$\pi(\lambda|x,\alpha,\tau,k,p) \propto n\log(\lambda) + \lambda\sum_{i=1}^{n}\log\left\{1 - \frac{\gamma_1\left[k,\left(\frac{x_i}{\alpha}\right)^{\tau}\right]}{1-p\left\{1-\gamma_1\left[k,\left(\frac{x_i}{\alpha}\right)^{\tau}\right]\right\}}\right\} + \log\left[\pi(\lambda)\right].$$

Since the full conditional distributions for α, τ, k, p and λ do not have explicit expressions, we require the use of the Metropolis-Hastings algorithm.

9.3. Simulation study. Here, we assess the performance of the MLEs by means Monte Carlo simulation experiment with $1,000$ replications. We considered no censoring case for simplicity. All results were carried out using the statistical software package `R`. For maximizing the log-likelihood function, we used the subroutine `optim`.

The evaluation of point estimation was performed based on the following quantities for each sample size: the empirical mean and the mean squared error (MSE). We set the sample size at $n = 100$, 200, 400 and 800 and considered different values for the parameters α, τ, k, p and λ. The empirical results are given in Table 2.

The values in Table 2 indicate that the estimates are close to the true values of the parameters for these sample sizes. Additionally, as the sample size increases, the MSEs decrease as expected, which means that the maximum likelihood method can be used effectively for estimating the parameters of the ExGGG distribution.

10. Application: permanence time in Japan

In this section, we provide one application to real data set in order to illustrate the importance and flexibility of the proposed distribution using both MLEs and Bayesian approaches. The data come from a study on the permanence time in Japan of the Brazilian immigrants.

The data (Table 3) were obtained from an electronic survey (e-survey) according to the method developed by Babbie [4], which serves to obtain data on the characteristics, actions or opinions of groups using the Internet as a research tool. The survey was carried out in the first half of the year 2010, by means of a reserved site with limited access, by 246 respondents which filled out the questionnaires. Nevertheless only 147 were used in the analysis because some immigrants were from other nationalities. We considered the main variable of interest as permanence time in Japan in years, counted from the first arrival date until the research date. Table 4 gives some descriptive statistics of these data, which include center tendency and dispersion. Figure 4 display the histogram for these data that has bimodal shape.

By using MLEs method, we fit the ExGGG, ExWG, GGG [24], WG [5], GG [31], Weibull [33], gamma and exponential distributions to these data. To obtain the MLEs of the model parameters, we use the NLMixed procedure in SAS. Table 5 lists the MLEs (and the corresponding standard errors in parentheses) of the model parameters.

Table 2. Empirical means and the MSEs in parentheses.

	$\alpha = 1.00$	$\tau = 1.50$	$k = 3.00$	$p = 0.90$	$\lambda = 0.50$
n	$\hat{\alpha}$	$\hat{\tau}$	$\hat{k}$	$\hat{p}$	$\hat{\lambda}$
100	0.9242	1.5722	3.4094	0.9168	0.4118
	(0.0490)	(0.0437)	(0.6373)	(0.0058)	(0.0312)
200	0.9456	1.5356	3.2579	0.9052	0.4550
	(0.0504)	(0.0353)	(0.3920)	(0.0033)	(0.0193)
400	0.9660	1.5178	3.1565	0.9002	0.4781
	(0.0487)	(0.0323)	(0.2966)	(0.0021)	(0.0115)
800	0.9909	1.5238	3.0790	0.8992	0.4849
	(0.0469)	(0.0299)	(0.2269)	(0.0013)	(0.0065)
	$\alpha = 5.00$	$\tau = 6.00$	$k = 0.20$	$p = 0.50$	$\lambda = 0.30$
n	$\hat{\alpha}$	$\hat{\tau}$	$\hat{k}$	$\hat{p}$	$\hat{\lambda}$
100	4.7765	6.0793	0.2281	0.6413	0.2414
	(0.1142)	(0.0858)	(0.0046)	(0.0730)	(0.0078)
200	4.8194	6.0500	0.2213	0.6192	0.2570
	(0.1043)	(0.0912)	(0.0024)	(0.0526)	(0.0065)
400	4.8552	6.0425	0.2144	0.5953	0.2666
	(0.0896)	(0.0863)	(0.0012)	(0.0424)	(0.0051)
800	4.8965	6.0182	0.2086	0.5554	0.2791
	(0.0797)	(0.0854)	(0.0007)	(0.0313)	(0.0042)
	$\alpha = 5.00$	$\tau = 0.75$	$k = 3.00$	$p = 0.90$	$\lambda = 1.50$
n	$\hat{\alpha}$	$\hat{\tau}$	$\hat{k}$	$\hat{p}$	$\hat{\lambda}$
100	5.0525	0.7103	3.4942	0.9395	1.4870
	(0.0939)	(0.0253)	(0.5615)	(0.0071)	(0.3453)
200	5.0443	0.7188	3.3180	0.9249	1.5328
	(0.1044)	(0.0206)	(0.3953)	(0.0043)	(0.1972)
400	5.0267	0.7337	3.1780	0.9143	1.5237
	(0.1026)	(0.0146)	(0.2455)	(0.0026)	(0.1243)
800	5.0541	0.7580	3.0419	0.9025	1.5041
	(0.1130)	(0.0128)	(0.1684)	(0.0018)	(0.0934)

The Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Corrected Akaike Information Criterion (CAIC), Anderson-Darling (A^*), Cramér-von Mises (W^*) and Kolmogorov-Smirnov (KS) statistics for the fitted distributions are reported in Table 6. In general, the lower values of these statistics indicate the best fit for the data. These results indicate that the ExGGG distribution has the lowest AIC, BIC, CAIC, A^*, W^* and KS values, and therefore our new model can be chosen as the best one.

Comparisons of the proposed distribution with some of their sub-models using LR statistics are performed in Table 7. The numbers in this table, specially the p-values, suggest that the ExGGG model yields a better fit to these data than the other distributions.

Table 3. Permanence time (years) in Japan of the Brazilian immigrants ($n = 147$).

1	1	1	2	2	2	3	3	3	3	3	3	3	4	4
4	4	4	4	5	5	5	5	5	5	5	5	5	6	6
6	6	6	6	6	6	6	7	7	7	8	8	9	9	9
9	10	10	10	10	10	10	10	10	11	12	12	12	12	12
13	13	13	13	13	13	13	13	13	13	14	14	14	14	14
14	14	14	14	15	15	15	15	15	15	15	16	16	16	16
16	16	17	17	17	17	17	18	18	18	18	18	18	18	18
18	18	18	18	18	18	18	18	18	18	19	19	19	19	19
19	19	19	19	19	19	19	20	20	20	20	20	20	20	20
20	20	20	20	21	21	21	21	22	22	22	22			

Table 4. Descriptive statistics of the permanence time in Japan.

Minimum	Mean	Median	Maximum	Variance
1.00	12.81	14.00	22.00	37.78

Table 5. MLEs of the model parameters for the permanence time data in Japan and the corresponding SEs in parentheses.

Model	$\hat{\alpha}$	$\hat{\tau}$	$\hat{k}$	$\hat{p}$	$\hat{\lambda}$
ExGGG	17.6851	12.0731	0.1610	0.9044	0.2283
	(0.0219)	(0.0683)	(0.0014)	(0.0305)	(0.0232)
ExWG	10.2767	3.8861	1	0.9851	0.1169
	(0.0030)	(0.0001)		(0.0066)	(0.0111)
GGG	345.3600	3.8158	0.6843	0.9998	1
	(0.0028)	(0.7310)	(0.1382)	(0.0001)	
WG	343.5700	2.6131	1	0.9998	1
	(0.0001)	(0.1732)		(0.0001)	
GG	21.9112	33.2664	0.04257	0	1
	(0.3329)	(18.6065)	(0.02505)		
Weibull	14.3931	2.1801	1	0	1
	(0.5679)	(0.1553)			
Gamma	4.4728	1	2.8639	0	1
	(0.5403)		(0.3165)		
Exponential	12.8095	1	1	0	1
	(1.0565)				

More information is provided by a visual comparison of the histogram and empirical cumulative distribution function of the data with the fitted ExGGG, ExWG, GGG, WG, GG, Weibull, gamma and exponential distributions. The plots of the estimated density and cumulative distribution functions are displayed

Table 6. Goodness-of-fit statistics for the permanence time data in Japan and the corresponding p-values in parentheses.

Model	AIC	BIC	CAIC	A*	W*	KS
ExGGG	**899.8**	**914.7**	**900.2**	**0.59**	**0.09**	**0.08**
				(0.66)	(0.62)	(0.32)
ExWG	930.3	942.3	930.6	3.17	0.56	0.13
				(0.02)	(0.03)	(0.01)
GGG	1,011.1	1,023.0	1,011.3	7.72	0.97	0.19
				(< 0.01)	(< 0.01)	(< 0.01)
WG	1,009.1	1,018.0	1,009.2	8.80	1.09	0.21
				(< 0.01)	(< 0.01)	(< 0.01)
GG	905.3	914.2	905.4	1.50	0.26	0.11
				(0.18)	(0.18)	(0.04)
Weibull	957.4	963.4	957.5	5.38	0.86	0.14
				(< 0.01)	(< 0.01)	(< 0.01)
Gamma	978.9	984.9	979.0	6.21	1.10	0.18
				(< 0.01)	(< 0.01)	(< 0.01)
Exponential	1,045.8	1,048.7	1,045.8	16.14	3.02	0.23
				(< 0.01)	(< 0.01)	(< 0.01)

in Figures 4 and 5. Clearly, the ExGGG distribution provides a closer fit to the histogram and empirical cumulative distribution function.

Table 7. LR tests.

Model	Hypotheses	LR statistics	p-value
ExGGG vs ExWG	$H_0 : k = 1$ vs $H_1 : H_0$ is false	32.5	< 0.01
ExGGG vs GGG	$H_0 : \lambda = 1$ vs $H_1 : H_0$ is false	113.3	< 0.01
ExGGG vs WG	$H_0 : \lambda = k = 1$ vs $H_1 : H_0$ is false	113.3	< 0.01
ExGGG vs GG	$H_0 : p = 0$ and $\lambda = 1$ vs $H_1 : H_0$ is false	9.5	0.01
ExWG vs WG	$H_0 : \lambda = 1$ vs $H_1 : H_0$ is false	80.8	< 0.01
GGG vs WG	$H_0 : k = 1$ vs $H_1 : H_0$ is false	< 0.01	0.99

Regarding bayesian analysis, we fit the ExGGG distribution. The following independent priors were considered to perform the Metropolis-Hastings algorithm: $\alpha \sim \Gamma(0.01, 0.01)$, $\tau \sim \Gamma(0.01, 0.01)$, $k \sim \Gamma(0.01, 0.01)$, $p \sim \mathrm{Be}(0,5; 0,5)$ and $\lambda \sim \Gamma(0.01, 0.01)$, so that we have a vague prior distribution. Considering these prior density functions, we generate two parallel independent runs of the Metropolis-Hastings with size $100,000$ for each parameter, disregarding the first $10,000$ iterations to eliminate the effect of the initial values and, to avoid correlation problems, we consider a spacing of size 10, obtaining a sample of size $9,000$ from each chain. To monitor the convergence of the Metropolis-Hastings, we perform the methods suggested by Cowles and Carlin [12]. To monitor the convergence of the Metropolis-Hastings, we use the between and within sequence information, following the approach developed in Gelman and Rubin [15], to obtain the potential

(a) (b)

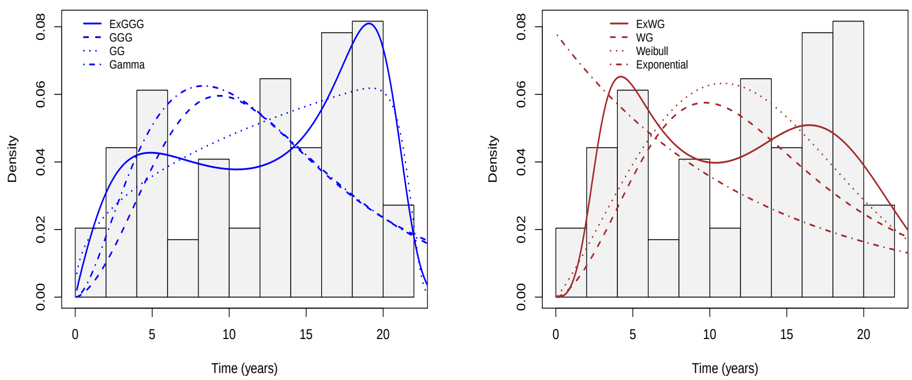

Figure 4. Histogram and fitted density functions for the permanence time data in Japan. (a) Fitted ExGGG, GGG, GG and gamma distributions. (b) Fitted ExWG, WG, Weibull and exponential distributions.

(a) (b)

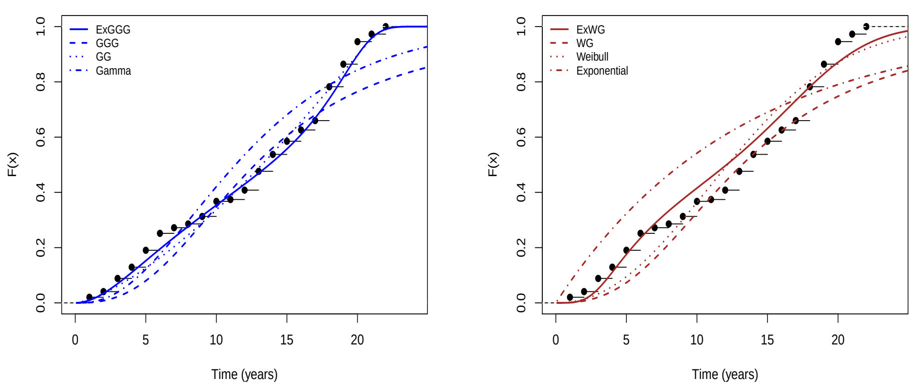

Figure 5. Estimated and empirical cumulative distribution functions for the permanence time in Japan. (a) Fitted ExGGG, GGG, GG and gamma distributions. (b) Fitted ExWG, WG, Weibull and exponential distributions.

scale reduction, $\widehat{R}$. In all cases, these values were close to one, indicating the convergence of the chain. The approximate posterior marginal density functions for the parameters are presented in Figure 6. In Table 8, we report posterior summaries for the parameters of the ExGGG model, where SD represents the standard deviation from the posterior distributions of the parameters and HPD represents the 95% highest posterior density (HPD) intervals. We note that the values for the means a posteriori (Table 8) are quite close (as expected) to the MLEs given in Table 5.

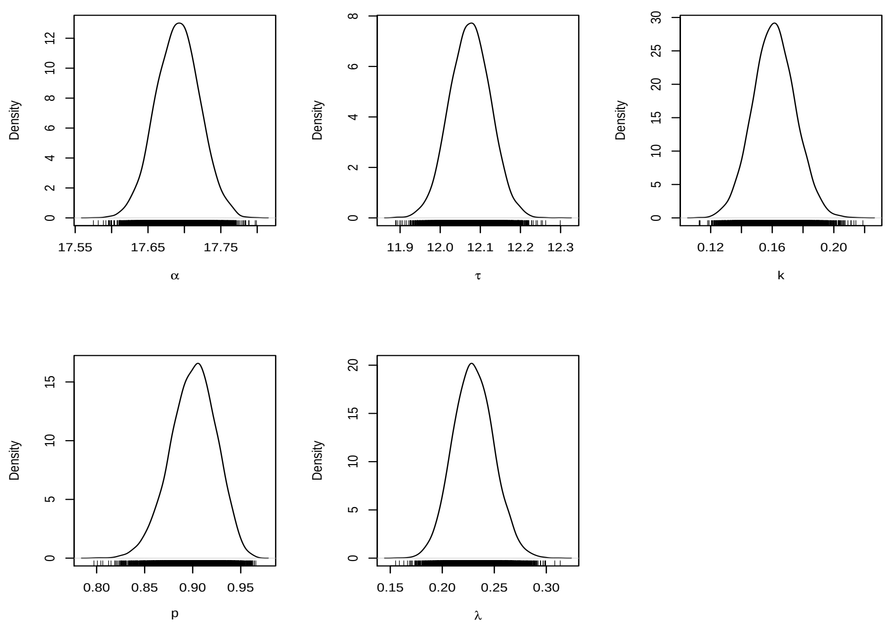

Figure 6. Approximate posterior marginal densities for the parameters from the ExGGG model for the permanence time in Japan.

Table 8. Posterior summaries for the parameters from the ExGGG model for the permanence time in Japan.

Parameter	Mean	SD	HPD (95%)	$\hat{R}$
α	17.6910	0.0298	(17.6298; 17.7469)	1.0016
τ	12.0737	0.0301	(12.0154; 12.1331)	1.0019
k	0.1616	0.0129	(0.1363; 0.1864)	1.0011
p	0.9029	0.0177	(0.8676; 0.9364)	1.0007
λ	0.2292	0.0171	(0.1974; 0.2640)	1.0001

Considering the MLEs of ExGGG and ExWG distributions (Table 5) and their estimated survival function (Figure 5), we obtain some useful results. The estimate for the median permanence time of Brazilian immigrants in Japan is approximately equal to thirteen years and nine months for ExGGG distribution and twelve years and one months for ExWG distribution. The probability of a Brazilian to stay less than five years in Japan is 15.25% for ExGGG distribution and 17.55% for ExWG distribution. The probability of a immigrant to remain less than twenty years is 89.60% for ExGGG distribution and 87.08% for ExWG distribution.

11. Conclusions

We introduced a new five parameter distribution called the extended generalized gamma geometric (ExGGG) distribution that provides a rather general and flexible framework for statistical analysis of positive data. The new distribution has positive and negative skewness (kurtosis), and further bimodal density, depending on the values of their parameters. Its hazard rate function has the forms: increasing, decreasing, in the form of bathtub and unimodal. Another important characteristic of this distribution is that it contains special sub-models such as generalized gamma geometric, Weibull geometric, exponential geometric, Rayleigh geometric, among some other distributions. Therefore the ExGGG distribution is suggested in a variety of problems for modeling lifetime data, such as bimodal and skewed. We demonstrated that the ExGGG density function can be expressed as a mixture of GG density functions. We derived explicit expressions for moments, moment generating function, mean deviations, reliability and order statistics. The estimation of parameters was approached by the method of maximum likelihood and by Bayesian method. Additionally the observed information matrix was determined. Furthermore, an application of the ExGGG distribution to real data showed that it could provide a better fit than other statistical models frequently used in lifetime data analysis.

Appendix A. Series expansion for the incomplete gamma ratio function

Pascoa *et al.* [27] developed a series expansion for the incomplete gamma ratio function given by

$$\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^{\tau}\right]=\frac{1}{\Gamma(k)}\left(\frac{x}{\alpha}\right)^{\tau k}\sum_{q=0}^{\infty}\left[-\left(\frac{x}{\alpha}\right)^{\tau}\right]^{q}\frac{1}{(k+q)q!}.$$

They used an equation in Section 0.314 of Gradshteyn and Ryzhik [16] for a power series raised to a positive integer m

$$\left[\sum_{q=0}^{\infty}a_q\left(\frac{x}{\alpha}\right)^{\tau q}\right]^{m}=\sum_{q=0}^{\infty}c_{m,q}\left(\frac{x}{\alpha}\right)^{\tau q}, \tag{A.1}$$

whose coefficients $c_{m,q}$ (for $q=1,2,\cdots$) are calculated from the recurrence equation

$$c_{m,q}=(qa_0)^{-1}\sum_{r=1}^{q}(mr-q+r)\,a_r\,c_{m,q-r} \tag{A.2}$$

and $c_{m,0}=a_0^m$, where $a_q=(-1)^q\,[(k+q)q!]^{-1}$. The coefficient $c_{m,q}$ can be obtained from $c_{m,0},\ldots,c_{m,q-1}$. It can also be written explicitly as functions of the quantities $a_0,\ldots,a_q$ using algebraic software such as Maple and Mathematica, although it is not necessary for programming numerically our expansions. Here, $c_{m,0}=k^{-m}$,

$c_{m,1} = -m[(k+1)k^{m-1}]^{-1}$, $c_{m,2} = m[2(k+2)k^{m-1}]^{-1} + m(m-1)[2(k+1)^2 k^{m-2}]^{-1}$, etc. Equation (A.1) yields

$$\gamma_1 \left[k, \left(\frac{x}{\alpha}\right)^{\tau}\right]^m = \frac{1}{\Gamma(k)^m} \sum_{q=0}^{\infty} c_{m,q} \left(\frac{x}{\alpha}\right)^{\tau(km+q)}, \tag{A.3}$$

where the coefficients $c_{m,q}$ are calculated from (A.2).

Appendix B. Observed information matrix $J(\boldsymbol{\theta})$

By differentiating (9.1), the elements of the observed information matrix $J(\boldsymbol{\theta})$ for the parameters $(\alpha, \tau, k, p, \lambda)$ are

$$\begin{aligned} J_{\alpha\alpha}(\boldsymbol{\theta}) = & -\frac{n\tau k}{\alpha^2} + (\tau+1)\frac{\tau}{\alpha^2}\sum_{i=1}^{n} u_i \\ & + \frac{2p\tau}{\alpha^2 \Gamma(k)} \sum_{i=1}^{n} g_i^2 u_i^k \exp(u_i) \\ & \times \left\{ (\tau k + 1 - \tau u_i)(1 - pq_i) - p\left[\frac{\tau}{\Gamma(k)} u_i^k \exp(u_i)\right]\right\} \\ & + \frac{(\lambda-1)\tau}{\alpha^2 \Gamma(k)} \sum_{i=1}^{n} \left(\frac{g_i}{q_i}\right)^2 u_i^k \exp(u_i) \\ & \times \left\{ q_i (1 - pq_i)(\tau k + 1 - \tau u_i) - \left[\frac{\tau}{\Gamma(k)} u_i^k \exp(u_i)\right](-1 + 2pq_i)\right\}, \end{aligned}$$

$$\begin{aligned} J_{\alpha\tau}(\boldsymbol{\theta}) = & \frac{nk}{\alpha} - \frac{1}{\alpha}\sum_{i=1}^{n} u_i (1 + \log u_i) \\ & - \frac{2p}{\alpha \Gamma(k)} \sum_{i=1}^{n} g_i^2 u_i^k \exp(u_i) \\ & \times \left\{ [1 + (k - u_i)\log u_i](1 - pq_i) - \frac{p}{\Gamma(k)} u_i^k \exp(u_i) \log u_i \right\} \\ & - \frac{\lambda - 1}{\alpha \Gamma(k)} \sum_{i=1}^{n} \left(\frac{g_i}{q_i}\right)^2 u_i^k \exp(u_i) \\ & \times \left\{ [1 + (k - u_i)\log u_i](1 - pq_i) q_i + \frac{(1 - 2pq_i)}{\Gamma(k)} u_i^k \exp(u_i) \log u_i \right\}, \end{aligned}$$

$$
\begin{aligned}
J_{\alpha k}(\boldsymbol{\theta}) &= \frac{n\tau}{\alpha} + \frac{2p\tau}{\alpha\Gamma(k)} \sum_{i=1}^{n} g_i^2 u_i^k \exp(u_i) \\
&\times \left\{ \left(\psi(k) - \frac{k}{\tau} \log u_i \right) (1 - pq_i) + p \left[w_i - \psi(k)\gamma_1(k, u_i) \right] \right\} \\
&+ \frac{(\lambda - 1)\tau}{\alpha\Gamma(k)} \sum_{i=1}^{n} u_i^k \exp(u_i) \left(\frac{g_i}{q_i} \right)^2 \\
&\times \left\{ \left(\psi(k) - \frac{k}{\tau} \log u_i \right) q_i (1 - pq_i) - \left[w_i - \psi(k)\gamma_1(k, u_i) \right] (1 - 2pq_i) \right\},
\end{aligned}
$$

$$
\begin{aligned}
J_{\tau\tau}(\boldsymbol{\theta}) &= \frac{n}{\tau^2} + \frac{1}{\tau^2} \sum_{i=1}^{n} u_i \log^2 u_i \\
&+ \frac{2p}{\tau^2\Gamma(k)} \sum_{i=1}^{n} g_i^2 u_i^k \exp(u_i) \log^2 u_i \\
&\times \left[(\tau - u_i)(1 - pq_i) - \frac{p}{\Gamma(k)} u_i^k \exp(u_i) \right] \\
&+ \frac{(\lambda - 1)}{\tau^2\Gamma(k)} \sum_{i=1}^{n} \left(\frac{g_i}{q_i} \right)^2 u_i^k \exp(u_i) \log^2 u_i \\
&\times \left[(\tau - u_i) q_i (1 - pq_i) + \frac{(1 - 2pq_i)}{\Gamma(k)} u_i^k \exp(u_i) \right],
\end{aligned}
$$

$$
\begin{aligned}
J_{\tau k}(\boldsymbol{\theta}) &= -\frac{1}{\tau} \sum_{i=1}^{n} \log u_i \\
&+ \frac{2p}{\tau\Gamma(k)} \sum_{i=1}^{n} g_i^2 u_i^k \exp(u_i) \log u_i \\
&\times \left\{ \left[\log u_i - \psi(k) \right] (1 - pq_i) - \frac{p}{\tau\Gamma(k)} u_i^k \exp(u_i) \log u_i \right\} \\
&+ \frac{(\lambda - 1)}{\tau\Gamma(k)} \sum_{i=1}^{n} \left(\frac{g_i}{q_i} \right)^2 u_i^k \exp(u_i) \log u_i \\
&\times \left\{ \left[\log u_i - \psi(k) \right] q_i (1 - pq_i) + \frac{1}{\tau\Gamma(k)} (1 - 2pq_i) u_i^k \exp(u_i) \log u_i \right\},
\end{aligned}
$$

$$
J_{\tau p}(\boldsymbol{\theta}) = \frac{(\lambda + 1)}{\tau\Gamma(k)} \sum_{i=1}^{n} g_i^2 u_i^k \exp(u_i) \log u_i,
$$

$$
\begin{aligned}
J_{kk}(\boldsymbol{\theta}) &= n\psi'(k) \\
&+ 2p\sum_{i=1}^{n} g_i^2 \Bigg\{ \left[\frac{\ddot{\gamma}\left(k,u_i\right)_{kk}}{\Gamma\left(k\right)} - \gamma_1\left(k,u_i\right)\left[\psi'(k)+\psi^2(k)\right] \right] \\
&\times\left(1-pq_i\right) - p\left[w_i - \psi(k)\gamma_1\left(k,u_i\right)\right]^2 \Bigg\} \\
&+ (\lambda-1)\sum_{i=1}^{n}\left(\frac{g_i}{q_i}\right)^2 \Bigg(\left\{ \frac{\ddot{\gamma}\left(k,u_i\right)_{kk}}{\Gamma\left(k\right)} - \gamma_1\left(k,u_i\right)\left[\psi'(k)+\psi^2(k)\right] \right\} \\
&\times q_i\left(1-pq_i\right) + \left[w_i - \psi(k)\gamma_1\left(k,u_i\right)\right]^2\left(1-2pq_i\right) \Bigg),
\end{aligned}
$$

$$
J_{kp}(\boldsymbol{\theta}) = (\lambda+1)\sum_{i=1}^{n} g_i^2\left[w_i - \psi(k)\gamma_1\left(k,u_i\right)\right],
$$

$$
J_{pp}(\boldsymbol{\theta}) = \frac{n}{(1-p)^2} - 2\sum_{i=1}^{n} q_i^2 g_i^2 + \frac{(\lambda-1)}{1-p}\sum_{i=1}^{n} g_i\gamma_1\left(k,u_i\right)\left(1-2pq_i+q_i\right),
$$

$$
J_{\lambda\lambda}(\boldsymbol{\theta}) = \frac{n}{\lambda^2},
$$

$$
J_{\lambda p}(\boldsymbol{\theta}) = \frac{1}{1-p}\sum_{i=1}^{n} g_i\gamma_1\left(k,u_i\right),
$$

$$
J_{\lambda k}(\boldsymbol{\theta}) = \sum_{i=1}^{n}\frac{g_i}{q_i}\left[w_i - \psi(k)\gamma_1\left(k,u_i\right)\right],
$$

$$
J_{\lambda\tau}(\boldsymbol{\theta}) = \frac{1}{\tau\Gamma\left(k\right)}\sum_{i=1}^{n}\frac{g_i}{q_i}u_i^k\exp\left(u_i\right)\log u_i
$$

and

$$
J_{\lambda\alpha}(\boldsymbol{\theta}) = -\frac{\tau}{\alpha\Gamma\left(k\right)}\sum_{i=1}^{n}\frac{g_i}{q_i}u_i^k\exp\left(u_i\right),
$$

where

$$
\dot{\gamma}(k,u_i)_k = \sum_{n=0}^{\infty}\frac{(-1)^n}{n!}J(u_i,k+n-1,1),
$$

$$
\ddot{\gamma}(k,u_i)_{kk} = \sum_{n=0}^{\infty}\frac{(-1)^n}{n!}J(u_i,k+n-1,2),
$$

$\psi'(\cdot)$ is the derivative of the digamma function, u_i, g_i, q_i and w_i are defined in Section 9. The $J(\cdot,\cdot,\cdot)$ function can be easily calculated from the integral given by Prudnikov *et al.* [28] in Section 2.6.3.

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

and

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

References

[1] Aarts, R.M. *Lauricella functions.* From MathWorld - A Wolfram Web Resource, created by E. W. Weisstein. Available at www.mathworld.wolfram.com/LauricellaFunctions.html, 2000.

[2] Adamidis, K. and Loukas, S. *A Lifetime distribution with decreasing failure rate*, Statistics & Probability Letters **39** (1), 35-42, 1998.

[3] Al-Zahrani, B., Marinho, P.R.D., Fattah, A.A., Ahmed, A.H.N., and Cordeiro, G.M. *The (P-A-L) extended Weibull distribution: a new generalization of the Weibull distribution*, Hacettepe Journal of Mathematics and Statistics **45**, 1-19, 2015.

[4] Babbie, E. *Métodos de pesquisas de Survey.* (Ed. UFMG, Belo Horizonte, 1999).

[5] Barreto-Souza, W., de Morais, A.L. and Cordeiro, G.M. *The Weibull-geometric distribution*, Journal of Statistical Computation and Simulation **81** (5), 645-657, 2011.

[6] Carrasco, J.M.F., Ortega, E.M.M. and Cordeiro, G.M. *A generalized modified Weibull distribution for lifetime modeling*, Computational Statistics and Data Analysis **53** (2), 450-462, 2008.

[7] Cordeiro, G.M., Ortega, E.M.M. and Silva, G.O. *The exponentiated generalized gamma distribution with application to lifetime data*, Journal of Statistical Computation and Simulation **81** (7), 827-842, 2011.

[8] Cordeiro, G.M., Ortega, E.M.M. and da Cunha, D.C.C. *The exponentiated generalized class of distributions*, Journal of Data Science **11**, 1-27, 2013.

[9] Cordeiro, G.M., Silva, G.O. and Ortega, E.M.M. *The beta-Weibull geometric distribution*, Statistics **47** (4), 817-834, 2013.

[10] Cordeiro, G.M., Ortega, E.M.M. and Silva, G.O. *The Kumaraswamy modified Weibull distribution: theory and applications*, Journal of Statistical Computation and Simulation **84** (7), 1387-1411, 2014.

[11] Cordeiro, G.M., Saboor, A., Khan, M.N., Ozel, G. and Pascoa, M.A.R. *The Kumaraswamy exponential-Weibull distribution: theory and applications*, Hacettepe Journal of Mathematics and Statistics **45**, 1-30, 2015.

[12] Cowles, M.K. and Carlin, B.P. *Markov chain Monte Carlo convergence diagnostics: a comparative review*, Journal of the American Statistical Association **91**, 883-904, 1996.

[13] Cox, C., Chu, M.F and Muñoz, A. *Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution* Statistics in Medicine **26**, 4352-4374, 2007.

[14] Exton, H. *Handbook of hypergeometric integrals: theory, applications, tables, computer programs.* (Halsted Press, New York, 1978).

[15] Gelman, A. and Rubin, D.B. *Inference from iterative simulation using multiple sequences (with discussion).* Statistical Science **7**, 457-472, 1992.

[16] Gradshteyn, I.S. and Ryzhik, I.M. *Table of integrals, series, and products.* (Academic Press, New York, 2000).

[17] Gupta, R.D. and Kundu, D. *Generalized exponential distribution*, Australian and New Zealand Journal of Statistics **41** (2), 173–188, 1999.

[18] Gupta, R.D. and Kundu, D. *Exponetiated exponential family: an alternative to gamma and Weibull distributions*, Biometrical Journal **43**, 117-130, 2001a.

[19] Gupta, R. D. and Kundu, D. *Generalized exponential distribution: different method of estimations*, Journal of Statistical Computation and Simulation, **69**, 315-337, 2001b.

[20] Gupta, R.D. and Kundu, D. *Generalized exponential distribution: existing results and some recent developments*, Journal of Statistical Planning and Inference **137** (11), 3537-3547, 2007.

[21] Lai, C.D., Xie, M. and Murthy, D.N.P. *A modified Weibull distribution.* Trans Reliab **52** (1), 33-37, 2003.

[22] Mudholkar, G.S., Srivastava, D.K. and Freimer, M. *The exponentiated Weibull family: a reanalysis of the bus-motor-failure data*, Technometrics **37** (4), 436-445, 1995.

[23] Nadarajah, S. *Explicit expressions for moments of order statistics*, Statistics and Probability Letters **78** (2), 196-205, 2008.

[24] Ortega, E.M.M., Cordeiro, G.M. and Pascoa, M.A.R. *The generalized gamma geometric distribution*, Journal of Statistical Theory and Applications **10**, 433-454, 2011.

[25] Paranaíba, P.F., Ortega, E.M.M., Cordeiro, G.M. and Pascoa, M.A.R. *The Kumaraswamy Burr XII distribution: theory and practice*, Journal of Statistical Computation and Simulation **83** (11), 2117-2143, 2013.

[26] Pascoa, M.A.R., Ortega, E.M.M. and Cordeiro, G.M. *The Kumaraswamy generalized gamma distribution with application in survival analysis*. Statistical Methodology **8** (5), 411-433, 2011.

[27] Pascoa, M.A.R., de Paiva, C.M.M., Cordeiro, G.M. and Ortega, E.M.M. *The Log-Kumaraswamy generalized gamma regression model with application to chemical dependency data*. Journal of Data Science **11**, 781-818, 2013.

[28] Prudnikov, A.P., Brychkov, Y.A. and Marichev, O.I. *Integrals and Series.* (Gordon and Breach Science Publishers, Amsterdam, 1986).

[29] Silva, R.B., Barreto-Souza, W. and Cordeiro, G.M. *A new distribution with decreasing, increasing and upside-down bathtub failure rate*, Computational Statistics and Data Analysis **54** (4), 935-944, 2010.

[30] Silva, G.O., Ortega, E.M.M. and Cordeiro, G.M. *The beta modified Weibull distribution*, Lifetime Data Analysis **16**, 409-430, 2010.

[31] Stacy, E.W. *A generalization of the gamma distribution*, Annals of Mathematical Statistics **33** (3), 1187-1192, 1962.

[32] Trott, M. *The mathematica guidebook for symbolics.* (Springer, New York, 2006).

[33] Weibull, W. *A statistical distribution function of wide applicability*, Journal of Applied Mechanics **18** (3), 293-297, 1951.

[34] Wright, E.M. *The asymptotic expansion of the generalized Bessel function.* Proceedings of the London Mathematical Society **38**, 257-270, 1935.

ARTIGO 2: A New Extended Generalized Gamma Geometric Distribution And Its Regression Model

Redigido conforme as normas da revista Hacettepe Journal of Mathematics and Statistics - versão preliminar.

A New Extended Generalized Gamma Geometric Distribution And Its Regression Model

Juliano Bortolini*, Marcelino A. R. Pascoa†‡, Renato R. de Lima§ and Anderson C. S. de Oliveira¶

Abstract

The five parameter extended generalized gamma geometric model proposed by Bortolini *et al.* [8] includes some important distributions as special cases and it is very useful for modeling lifetime data. We propose an extension of this distribution by assuming that a shape parameter can take negative values. The new distribution can accommodate increasing, decreasing, bathtub and unimodal shaped hazard functions. A second advantage is that it also includes as special models reciprocal distributions like the reciprocal gamma geometric and inverse Weibull distributions. We provide a mathematical treatment of the new distribution including explicit expressions for moments, generating function, mean deviations, reliability and order statistics. Further, we derive the log-transformed distribution and its regression model. The new regression model represents a parametric family of models that includes as sub-models some widely known regression models that can be applied to censored survival data. The method of maximum likelihood and a Bayesian procedure are adopted for estimating the regression model parameters. Finally, we analyze a real data set using the log-extended generalized gamma geometric regression model to illustrate its usefulness.

Keywords: Censored data, Lifetime data, Regression model, Survival analysis, Log-gamma distribution, Location-scale regression model.

2000 AMS Classification: 60E05, 62E15, 62P25

---

*Departamento de Estatística, Universidade Federal de Mato Grosso, Cuiabá 78060-900, Brazil, Email: `julianobortolini@ufmt.br`

†Departamento de Estatística, Universidade Federal de Mato Grosso, Cuiabá 78060-900, Brazil, Email: `marcelino.pascoa@gmail.com`

‡Corresponding Author.

§Departamento de Ciências Exatas, Universidade Federal de Lavras, Lavras 37200-000, Brazil, Email: `rrlima@dex.ufla.br`

¶Departamento de Estatística, Universidade Federal de Mato Grosso, Cuiabá 78060-900, Brazil, Email: `andersoncso@gmail.com`

1. Introduction

The gamma distribution is the most popular model for analyzing skewed data. Stacy [50] introduced the generalized gamma (GG) distribution that includes as their special sub-models the exponential, Weibull, gamma and Rayleigh distributions, among others. It is suitable for modeling data with hazard rate function of different forms (increasing, decreasing, bathtub and unimodal). The GG distribution has been used in several research areas such as engineering, hydrology and survival analysis. Nadarajah and Gupta [36] used this distribution with applications to drought data and Cox *et al.* [17] presented a parametric survival analysis and taxonomy of the GG hazard rate functions. Further, Gomes *et al.* [25] focused on parameter estimation. Barkauskas *et al.* [6] modeled the noise part of a spectrum as an autoregressive moving average model with innovations having the GG distribution and Malhotra *et al.* [34] provided a unified analysis for wireless system over generalized fading channels that is modeled by a two parameter GG model. Cox and Matheson [19] discussed and compared the exponentiated Weibull (EW) distribution, proposed by Mudholkar *et al.* [35], with the GG model and showed that for each EW distribution, there is a corresponding GG distribution which is nearly identical.

New more flexible distributions have been developed for modeling complex survival data based on extensions of the GG distribution or their sub-models. These distributions allow us to obtain density and failure rate functions with great flexibility and are useful to develop more realistic statistical models in a great variety of applications.

Adamidis and Loukas [2] introduced the exponential geometric (EG) distribution to model lifetime data with decreasing failure rate function. Gupta and Kundu [27, 28, 29, 30] provided a comprehensive mathematical treatment of the so-called generalized exponential (GE) (also refereed as the exponentiated exponential) distribution, Carrasco *et al.* [9] proposed a generalized modified Weibull (GMW) with applications in survival analysis and Silva *et al.* [49] studied in details the beta modified Weibull distribution which admits only increasing and decreasing failure rate functions. Following the same idea of GE distribution, Silva *et al.* [48] defined the generalized exponential geometric distribution and demonstrated that its failure rate function can be increasing, decreasing or unimodal. The Weibull geometric (WG) extension of the EG distribution was proposed by Barreto-Souza *et al.* [7] for modeling monotone or unimodal failure rates. Ortega *et al.* [42], following the idea of Adamidis and Loukas [2] for a process of mixing distributions, introduced the generalized gamma geometric (GGG) distribution with four parameters and four standard types of the failure rate function (i.e. increasing, decreasing, unimodal and bathtub). Pascoa *et al.* [45] introduced and studied the Kumaraswamy generalized gamma distribution that is capable to model monotone and non-monotone hazard rate functions which are quite common in lifetime data analysis and reliability. Cordeiro *et al.* [13] defined the beta-Weibull geometric and Cordeiro *et al.* [14] presented the Kumaraswamy modified Weibull that contains as special sub-models the GMW, EW among others distributions. Al-Zahrani *et al.* [3] defined the (P-A-L) extended Weibull distribution and Cordeiro *et al.* [15] proposed the Kumaraswamy exponential-Weibull distribution that generalizes a number of well-known special lifetime models such as the Weibull, exponential, Rayleigh, modified Rayleigh, modified exponential and exponentiated Weibull distributions, among others.

Bortolini *et al.* [8] proposed the extended generalized gamma geometric (ExGGG) distribution that can model four standard types of the failure rate function depending on the values of their parameters. It is also suitable for testing goodness-of-fit of some special sub-models, such as the GGG, GG, WG, EG, Weibull and exponential distributions and suggested in a variety of problems for modeling lifetime data, such as bimodal, skewed and heavy-tailed. However, the sub-models of this distribution does not include reciprocal type distributions. In this article, we define a new ExGGG distribution to cope with several reciprocal type distributions, such as inverse Weibull [22], and study some of their structural properties, such as moments, generating function, mean deviations, reliability and order statistics.

In many studies, the lifetimes are affected by explanatory variables such as treatment, group indicator, individual characteristics, environmental conditions and many others. In these circumstances the regression models can be used. Among them, the location-scale GG regression model [20, 32, 33] is frequently used in survival analysis. Furthermore, new regression models for modeling survival data based on extensions of the GG distribution or their sub-models were developed. Pascoa *et al.* [46] proposed the log-Kumaraswamy generalized gamma regression model which is a alternative for modeling the four standard shapes of failure rate functions, Ortega *et al.* [44] defined the negative binomial-generalized gamma regression model with a cure fraction and da Cruz *et al.* [21] presented the log-odd log-logistic Weibull regression model. Here, we propose the log-extended generalized gamma geometric (LExGGG) distribution and its location-scale regression model which is great flexibility and suggested a variety of problems for modeling lifetime data, such as bimodal, skewed and heavy-tailed.

The paper is organized as follows. In Section 2, we define a new ExGGG distribution. Their general properties are presented in section 3. In Section 4, we define the LExGGG distribution and derive an expansion for their moments. In Section 5, we propose a LExGGG regression model for censored data. We consider the methods of the maximum likelihood and Bayesian to estimate the model parameters. In Section 6, a real data set is analyzed which illustrate the usefulness of the proposed model. Finally, concluding remarks are given in last section.

Unless otherwise stated, all of the results presented in this article are new and original. It is expected that they encourage further research on the new distribution and regression models.

2. The new ExGGG Distribution

Bortolini *et al.* [8] defined the ExGGG distribution with five parameters α, τ, k, λ positive, and $p \in (0, 1)$ to extend the GGG distributions introduced by Ortega *et al.* [42]. The probability density function (pdf) of the ExGGG distribution has the form (for $t > 0$)

$$
\begin{aligned}
f(t) \quad &= \quad \frac{\lambda\tau(1-p)}{\alpha\Gamma(k)}\left(\frac{t}{\alpha}\right)^{\tau k-1}\exp\left[-\left(\frac{t}{\alpha}\right)^{\tau}\right]\left(1-p\left\{1-\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]\right\}\right)^{-2}\\
&\times \quad \left\{1-\frac{\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]}{1-p\left\{1-\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]\right\}}\right\}^{\lambda-1},
\end{aligned}
\tag{2.1}
$$

where $\Gamma(\cdot)$ is the gamma function, $\gamma_1(k,x) = \gamma(k,x)/\Gamma(k)$ is the incomplete gamma function ratio and $\gamma(k,x) = \int_0^x w^{k-1}\,\mathrm{e}^{-\mathrm{w}}\mathrm{dw}$ is the incomplete gamma function. In the density function (2.1), α is a scale parameter and τ, k, p and λ are shape parameters.

Clearly, the GGG distribution is obtained from (2.1) when $\lambda = 1$. Some distributions are obtained from (2.1) as particular cases, for example, when $k = \lambda = 1$ we have the WG distribution, for $\tau = k = \lambda = 1$, we obtain the EG distribution. The GG distribution is the limiting distribution (the limit is defined in terms of the convergence in distribution) when $p \to 0^+$ and $\lambda = 1$. On the other hand, if $p \to 1^-$, we obtain the distribution of a random variable Y such that $P(Y = 0) = 1$. Hence, the parameter p can be interpreted as a degeneration parameter. When p approaches to zero (and $k = \lambda = 1$), it leads to the Weibull distribution.

Now, following the same idea of Stacy and Mihram [51], Ortega *et al.* [43] and Pascoa *et al.* [46], we define an extended form of the density function (2.1) (for $t > 0$) given by

$$
\begin{aligned}
f(t) \quad &= \quad \frac{\lambda|\tau|(1-p)}{\alpha\Gamma(k)}\left(\frac{t}{\alpha}\right)^{\tau k-1}\exp\left[-\left(\frac{t}{\alpha}\right)^{\tau}\right]\left(1-p\left\{1-\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]\right\}\right)^{-2}\\
&\times \quad \left\{1-\frac{\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]}{1-p\left\{1-\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]\right\}}\right\}^{\lambda-1},
\end{aligned}
\tag{2.2}
$$

where τ is not zero, $\alpha, k, \lambda > 0$ and $p \in (0,1)$. The terminology extended generalized gamma geometric (ExGGG) distribution is maintained for (2.2) and then if T is a random variable having this density, we write $T \sim \mathrm{ExGGG}(\alpha,\tau,k,p,\lambda)$. Several special models of (2.2) when $\tau > 0$ can be seen in Bortolini *et al.* [8]. For $\tau = -1$, we obtain the reciprocal extended gamma geometric distribution (new). The case $\tau < 0$ and $\lambda = 1$ yields the reciprocal generalized gamma geometric distribution (new). For $\tau = -1$ and $k = \lambda = 1$, we obtain the reciprocal exponential geometric distribution (new). If $\tau = -2$ and $k = \lambda = 1$, we obtain the reciprocal Rayleigh geometric distribution (new). For $\tau = -1$, $\alpha = 2$, $k = n/2$ and $\lambda = 1$, we obtain reciprocal chi-square geometric distribution (new). If $\tau = -2$, $k = 3/2$ and $\lambda = 1$, we obtain the reciprocal Maxwell geometric distribution (new). If $\tau = -2$, $\alpha = \sqrt{2}$, $k = 1/2$ and $\lambda = 1$, we obtain the reciprocal half normal geometric distribution (new). Pascoa *et al.* [46] presented some reciprocal distributions such as generalized reciprocal gamma ($\tau < 0$, $p \to 0^+$ and $\lambda = 1$), reciprocal Weibull or FrÚchet distribution ($\tau < 0$, $p \to 0^+$ and $k = \lambda = 1$), reciprocal exponential ($\tau = -1$, $p \to 0^+$ and $k = \lambda = 1$), among others. Plots of the ExGGG density function for selected values of $\tau > 0$ and $\tau < 0$ are shown in Figure 1.

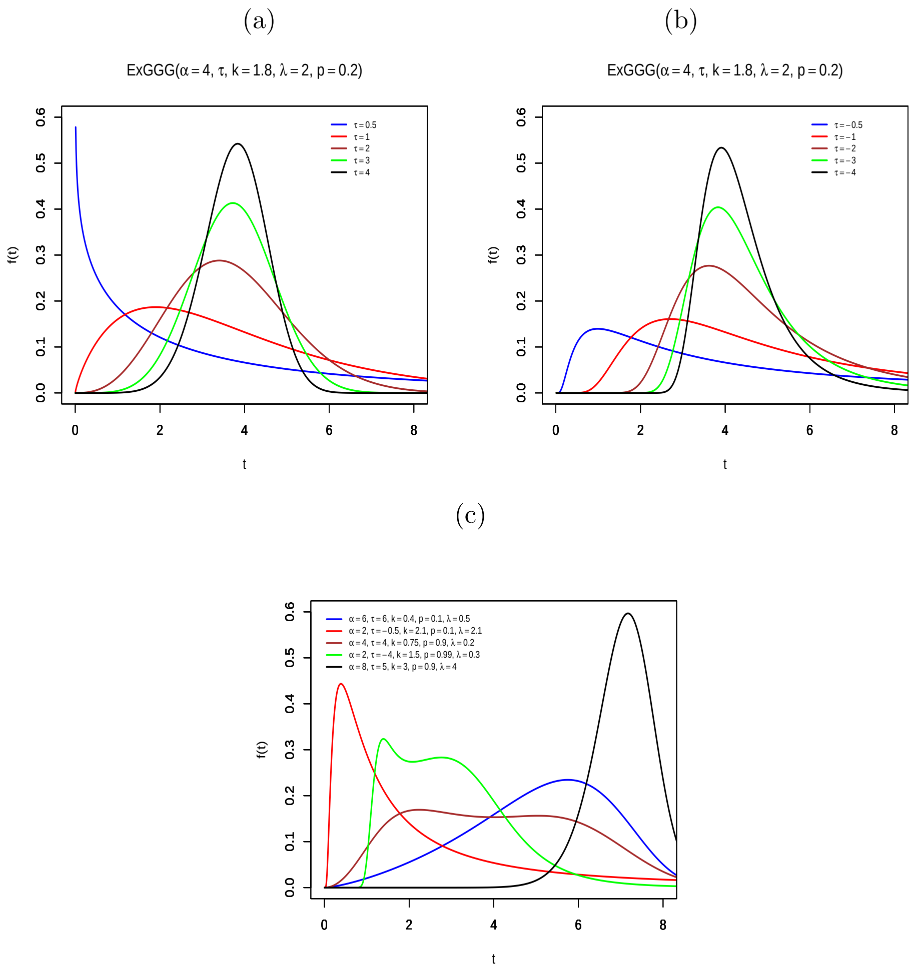

Figure 1. The ExGGG density curves: (a) For some values of $\tau > 0$. (b) For some values of $\tau < 0$. (c) For some values of $\tau > 0$ and $\tau < 0$.

The cumulative distribution function (cdf) corresponding to (2.2) is given by

$$F(t) = \begin{cases} 1 - \left\{ 1 - \frac{\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]}{1-p\left\{1-\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]\right\}} \right\}^{\lambda} & \text{if} \quad \tau > 0, \\ \left\{ 1 - \frac{\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]}{1-p\left\{1-\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]\right\}} \right\}^{\lambda} & \text{if} \quad \tau < 0. \end{cases} \tag{2.3}$$

The evidence that the function $F(t)$ in (2.3) is a cumulative distribution follows that, for $\alpha, k > 0$ and $\tau \neq 0$ fixed, the function $\gamma_1\left[k, (t/\alpha)^{\tau}\right]$ is right-continuous at $t > 0$, $\lim_{t\to\infty} \gamma_1\left[k, (t/\alpha)^{\tau}\right] = 1$ and $\lim_{t\to 0^+} \gamma_1\left[k, (t/\alpha)^{\tau}\right] = 0$ for $\tau > 0$, and $\lim_{t\to\infty} \gamma_1\left[k, (t/\alpha)^{\tau}\right] = 0$ and $\lim_{t\to 0^+} \gamma_1\left[k, (t/\alpha)^{\tau}\right] = 1$ for $\tau < 0$.

Now, we demonstrate that the density function (2.2) can be expressed as a linear combination of GG density functions. This result is important to provide mathematical properties of the ExGGG distribution directly from properties of the GG distribution.

Let $g_{\alpha,\tau,k}(t)$ be the density function of the GG(α, τ, k) distribution [51] given by (for $t > 0$)

$$g_{\alpha,\tau,k}(t) = \frac{|\tau|}{\alpha\,\Gamma(k)} \left(\frac{t}{\alpha}\right)^{\tau k-1} \exp\left[-\left(\frac{t}{\alpha}\right)^{\tau}\right]. \tag{2.4}$$

For $|z| < 1$ and $\rho \in \mathbb{R}$, we consider the power series

$$(1-z)^{\rho} = \sum_{j=0}^{\infty} (-1)^j \binom{\rho}{j} z^j, \tag{2.5}$$

where $\binom{\rho}{j} = \Gamma(\rho+1)/\left[\Gamma(\rho-j+1)j!\right]$.

Using (2.5) in equation (2.2), the ExGGG$(\alpha, \tau, k, p, \lambda)$ density function can be written as

$$\begin{aligned} f(t) &= \frac{\lambda|\tau|(1-p)}{\alpha\Gamma(k)} \left(\frac{t}{\alpha}\right)^{\tau k-1} \exp\left[-\left(\frac{t}{\alpha}\right)^{\tau}\right] \left(1-p\left\{1-\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]\right\}\right)^{-2} \\ &\times \sum_{j=0}^{\infty} (-1)^j \binom{\lambda-1}{j} \gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]^j \left(1-p\left\{1-\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]\right\}\right)^{-j}. \end{aligned}$$

Grouping common terms, using (2.5) and binomial expansion, we have that

$$\begin{aligned} f(t) &= \frac{\lambda|\tau|(1-p)}{\alpha\Gamma(k)} \left(\frac{t}{\alpha}\right)^{\tau k-1} \exp\left[-\left(\frac{t}{\alpha}\right)^{\tau}\right] \\ &\times \sum_{j,l=0}^{\infty} \sum_{m=0}^{l} (-1)^{j+l+m} \binom{\lambda-1}{j} \binom{-(j+2)}{l} \binom{l}{m} p^l \gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]^{j+m} \end{aligned}$$

We can substitute $\sum_{j,l=0}^{\infty}\sum_{m=0}^{l}$ for $\sum_{j,m=0}^{\infty}\sum_{l=m}^{\infty}$ to obtain

$$\begin{aligned} f(t) &= \frac{\lambda|\tau|(1-p)}{\alpha\Gamma(k)} \left(\frac{t}{\alpha}\right)^{\tau k-1} \exp\left[-\left(\frac{t}{\alpha}\right)^{\tau}\right] \\ &\times \sum_{j,m=0}^{\infty} s_{j,m}(\lambda,p)\, \gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]^{j+m}, \end{aligned} \tag{2.6}$$

where

$$s_{j,m}(\lambda,p) = \sum_{l=m}^{\infty} (-1)^{j+l+m} \binom{\lambda-1}{j} \binom{-(j+2)}{l} \binom{l}{m} p^l. \tag{2.7}$$

Therefore, using the result (A.3) (given in Appendix A) in the expression (2.6), the pdf $f(t)$ can be written as a linear combination of the distribution GG, in the form:

$$f(t) = \sum_{j,m,d=0}^{\infty} w_{j,m,d}(k,p,\lambda) g_{\alpha,\tau,k^{\bullet}}(t), \quad t > 0 \tag{2.8}$$

where $k^{\bullet} = k\,(j+m+1)+d$, $g_{\alpha,\tau,k^{\bullet}}\,(t)$ has distribution GG$(\alpha,\tau,k^{\bullet})$, the weightings $w_{j,m,d}(k,p,\lambda)$ are given by

$$w_{j,m,d}(k,p,\lambda) = \lambda\,(1-p)\,s_{j,m}\,(\lambda,p)\,c_{j+m,d}\frac{\Gamma\,(k^{\bullet})}{\Gamma\,(k)^{j+m+1}}, \tag{2.9}$$

and the coefficients $c_{j+m,d}$ are determined from the recurrence relation (A.2) (Appendix A).

Equation (2.8) shows that the ExGGG density function can be written in terms of a linear combination of GG densities. Hence, some mathematical properties of the new distribution (for example, ordinary moments and moment generating function) can be easily obtained from those GG properties.

3. General Properties

In this section, we present some properties of the ExGGG distribution. Here and henceforth, let T and Y be random variables having the distributions (2.2) and (2.4), respectively.

3.1. Moments. Some important features of a distribution such as dispersion, skewness and kurtosis can be studied through their moments. Here, we give two alternative expansions for the moments of the ExGGG distribution. First, the rth moment of the GG(α,τ,k) distribution can be expressed as

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

Let $\mu'_r = E(T^r)$ be the rth ordinary moment of ExGGG$(\alpha,\tau,k,p,\lambda)$ distribution. Hence, μ'_r follows from equation (2.8) as

$$\mu'_r = \alpha^r \sum_{j,m,d=0}^{\infty} \lambda\,(1-p)\,s_{j,m}\,(\lambda,p)\,c_{j+m,d}\frac{\Gamma\,(k^{\bullet}+r/\tau)}{\Gamma\,(k)^{j+m+1}}, \tag{3.1}$$

where $k^{\bullet} = k(j+m+1)+d$ and the quantity $s_{j,m}\,(\lambda,p)$ comes from equation (2.7). Equation (3.1) depends on the quantities $c_{j+m,d}$ that can only be calculated recursively from (A.2). This equation is readily computed numerically using standard statistical software. It (and other expansions in this article) can also be evaluated in symbolic computation software such as Mathematica and Maple. In numerical applications, a large natural number N can be used in the sums instead of infinity.

Now, we derive another infinite sum representation for μ'_r by computing it directly. The rth moment of the ExGGG distribution is obtained from (2.2) as

$$\begin{aligned}
\mu'_r &= \frac{\lambda|\tau|\,(1-p)}{\alpha\Gamma\,(k)}\int_0^\infty t^r\left(\frac{t}{\alpha}\right)^{\tau k-1}\exp\left[-\left(\frac{t}{\alpha}\right)^{\tau}\right]\\
&\times \left(1-p\left\{1-\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]\right\}\right)^{-2}\\
&\times \left\{1-\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]\left(1-p\left\{1-\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]\right\}\right)^{-1}\right\}^{\lambda-1} dt.
\end{aligned}$$

Setting $w = \left(\frac{t}{\alpha}\right)^{\tau}$ in the last equation, we have

$$\mu_r' = \frac{\lambda\left(1-p\right)\alpha^r \operatorname{sgn}(\tau)}{\Gamma\left(k\right)} \int_0^{\infty} w^{k+\frac{r}{\tau}-1} \exp\left(-w\right)\left\{1-p\left[1-\gamma_1\left(k,w\right)\right]\right\}^{-2}$$

$$\times\left\{1-\gamma_1\left(k,w\right)\left\{1-p\left[1-\gamma_1\left(k,w\right)\right]\right\}^{-1}\right\}^{\lambda-1} dw. \tag{3.2}$$

Considering (2.5) in (3.2) twice conveniently, we have that

$$\mu_r' = \frac{\lambda\left(1-p\right)\alpha^r \operatorname{sgn}(\tau)}{\Gamma\left(k\right)} \int_0^{\infty} w^{k+\frac{r}{\tau}-1} \exp\left(-w\right)$$

$$\times \sum_{j,l=0}^{\infty} \left(-1\right)^{j+l} \binom{\lambda-1}{j} \binom{-\left(j+2\right)}{l} \gamma_1\left(k,w\right)^j p^l \left[1-\gamma_1\left(k,w\right)\right]^l dw. \tag{3.3}$$

Using the binomial expansion in the term $\left[1-\gamma_1\left(k,w\right)\right]^l$ the expression (3.3) is rewritten as

$$\mu_r' = \frac{\lambda\left(1-p\right)\alpha^r \operatorname{sgn}(\tau)}{\Gamma\left(k\right)} \sum_{j,l=0}^{\infty} \sum_{m=0}^{l} \left(-1\right)^{j+l+m} \binom{\lambda-1}{j} \binom{-\left(j+2\right)}{l} \binom{l}{m} p^l$$

$$\times \int_0^{\infty} w^{k+\frac{r}{\tau}-1} \exp\left(-w\right) \gamma_1\left(k,w\right)^{j+m} dw.$$

Replacing $\sum_{j,l=0}^{\infty}\sum_{m=0}^{l}$ for $\sum_{j,m=0}^{\infty}\sum_{l=m}^{\infty}$, we have

$$\mu_r' = \frac{\lambda\left(1-p\right)\alpha^r \operatorname{sgn}(\tau)}{\Gamma\left(k\right)} \sum_{j,m=0}^{\infty} \sum_{l=m}^{\infty} \left(-1\right)^{j+l+m} \binom{\lambda-1}{j} \binom{-\left(j+2\right)}{l} \binom{l}{m} p^l$$

$$\times \int_0^{\infty} w^{k+\frac{r}{\tau}-1} \exp\left(-w\right) \gamma_1\left(k,w\right)^{j+m} dw.$$

Therefore μ_r' can be rewritten as

$$\mu_r' = \frac{\lambda\left(1-p\right)\alpha^r \operatorname{sgn}(\tau)}{\Gamma\left(k\right)} s_{j,m}\left(\lambda,p\right) I\left(k+\frac{r}{\tau}, j+m\right), \tag{3.4}$$

where $s_{j,m}\left(\lambda,p\right)$ is defined by expression (2.7) and

$$I\left(k+\frac{r}{\tau}, j+m\right) = \int_0^{\infty} w^{k+\frac{r}{\tau}-1} \exp\left(-w\right) \gamma_1\left(k,w\right)^{j+m} dw.$$

This integral can be determined from equations (24) and (25) of Nadarajah [38] in terms of the Lauricella function of type A [23, 1] defined by

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

where $(a)_i$ is the ascending factorial defined by $(a)_i = a(a+1)\cdots(a+i-1)$ assuming $(a)_0 = 1$. Numerical routines for the direct computation of the Lauricella function of type A are available, in Exton [23] and Trott [52]. We obtain

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

The moments of the ExGGG distribution given by (3.1) and (3.4) are the main results of this section.

The plots of skewness and kurtosis when $k=0.5$ and $p=0.6$, as function of λ for selected values of α and τ, are shown in Figures 2 (a) and (b), respectively. These plots indicate that the skewness and kurtosis of ExGGG distribution have great flexibility.

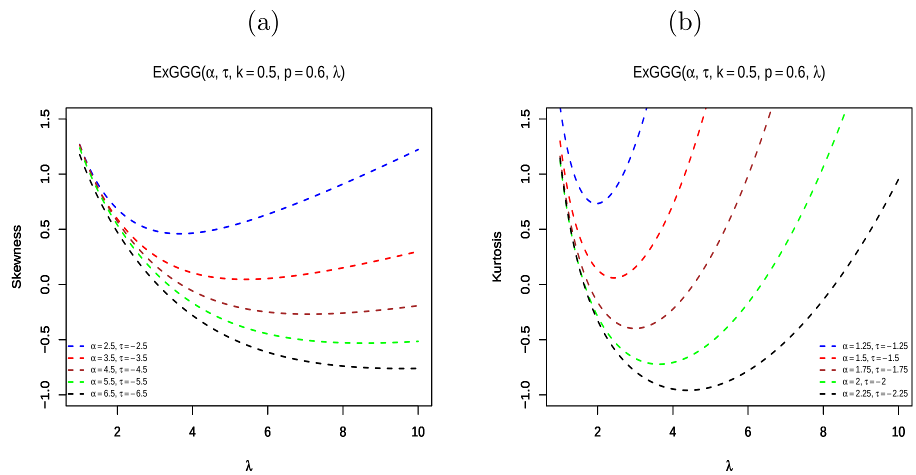

Figure 2. (a) Skewness and (b) kurtosis of the ExGGG distribution when $k=0.5$ and $p=0.6$, as function of λ for selected values of α and τ.

3.2. Moment Generating Function. This section we provide an expressions for the moment generating function (mgf) of Y, say $M_{\alpha,\tau,k}(s)=E[\exp(sY)]$, using the Wright function [53] and further the mgf for ExGGG distribution. We can write

$$M_{\alpha,\tau,k}(s)=\frac{|\tau|}{\alpha^{\tau k}\Gamma(k)}\int_0^\infty \exp(sy)y^{\tau k-1}\exp\{-(y/\alpha)^\tau\}dy.$$

Setting $u=y/\alpha$, we obtain

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

Expanding the first exponential in Taylor series and using $\int_0^\infty u^{k\tau+h-1}\exp(-u^\tau)du = \tau^{-1}\Gamma(k+h/\tau)$, we have

$$M_{\alpha,\tau,k}(s) = \frac{\operatorname{sgn}(\tau)}{\Gamma(k)}\sum_{h=0}^{\infty}\Gamma\left(\frac{h}{\tau}+k\right)\frac{(\alpha s)^h}{h!}. \tag{3.5}$$

Equation (3.5) holds for any τ different from zero. However, for $\tau > 1$, it can be simplified by considering the Wright generalized hypergeometric function [53] defined by

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

This function exists if $1+\sum_{j=1}^{q}B_j-\sum_{j=1}^{p}A_j > 0$. By combining the last two equations, we have

$$M_{\alpha,\tau,k}(s) = \frac{1}{\Gamma(k)}\,{}_1\Psi_0\left[\begin{array}{c}(k,\tau^{-1})\\-\end{array};\alpha s\right]. \tag{3.6}$$

Using the results (3.5) and (2.8), the mgf of ExGGG distribution for any τ different from zero is given by

$$M(s) = \sum_{j,m,d=0}^{\infty} w_{j,m,d}(k,p,\lambda)M_{\alpha,\tau,k^\bullet}(s),$$

where $k^\bullet = k(j+m+1)+d$ and $w_{j,m,d}(k,p,\lambda)$ is given by (2.9).

Therefore,

$$M(s) = \sum_{j,m,d,h=0}^{\infty}\frac{w_{j,m,d}(k,p,\lambda)}{\Gamma(k^\bullet)}\Gamma\Big(\frac{h}{\tau}+k^\bullet\Big)\frac{(\alpha s)^h}{h!}.$$

Finally, using the Wright function, the mgf of T for $\tau > 1$ can be written from equations (2.8) and (3.6) as

$$M(s) = \sum_{j,m,d=0}^{\infty}\frac{w_{j,m,d}(k,p,\lambda)}{\Gamma(k^\bullet)}\,{}_1\Psi_0\left[\begin{array}{c}(k^\bullet,1/\tau)\\-\end{array};\alpha s\right].$$

3.3. Mean Deviations. The amount of scatter in a population is evidently measured to some extent by the totality of deviations from the mean and median. If T has the ExGGG distribution with density function $f(t)$, we can derive the mean deviations about the mean $\mu_1' = E(T)$ and about the median m_1 from the relations

$$\delta_1 = \int_0^\infty |t-\mu_1'|f(t)dt \quad\text{and}\quad \delta_2 = \int_0^\infty |t-m_1|f(t)dt.$$

These measures can be expressed, respectively, as

$$\delta_1 = 2\mu_1'F(\mu_1') - 2I(\mu_1') \quad\text{and}\quad \delta_2 = \mu_1' - 2I(m_1), \tag{3.7}$$

where $F(\mu_1')$ is easily calculated from (2.3) and $I(s) = \int_0^s t\,f(t)dt$ is the first incomplete moment of T.

The integral $I(s)$ can be obtained from equation (2.8) as

$$
\begin{aligned}
I(s) &= \sum_{j,m,d=0}^{\infty} w_{j,m,d}(k,p,\lambda) \int_0^s t g_{\alpha,\tau,k^\bullet}(t)dt \\
&= \sum_{j,m,d=0}^{\infty} w_{j,m,d}(k,p,\lambda) J(\alpha,\tau,k^\bullet,s), \qquad (3.8)
\end{aligned}
$$

where

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

By setting $w = t/\alpha$, we can write from the $\mathrm{GG}(\alpha,\tau,k^\bullet)$ density function

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

The substitution $z = w^\tau$ yields $J(\alpha,\tau,k^\bullet,s)$ in terms of the incomplete gamma function

$$
\begin{aligned}
J(\alpha,\tau,k^\bullet,s) &= \frac{\alpha\,\mathrm{sgn}(\tau)}{\Gamma(k^\bullet)} \int_0^{(s/\alpha)^\tau} z^{k^\bullet+\tau^{-1}-1} \exp(-z)dz \\
&= \frac{\alpha\,\mathrm{sgn}(\tau)}{\Gamma(k^\bullet)} \gamma(k^\bullet+\tau^{-1},(s/\alpha)^\tau).
\end{aligned}
$$

Hence, inserting the last result into (3.8) gives

$$I(s) = \sum_{j,m,d=0}^{\infty} \frac{\alpha \mathrm{sgn}(\tau) w_{j,m,d}(k,p,\lambda)}{\Gamma(k^\bullet)} \gamma[k^\bullet+\tau^{-1},(s/\alpha)^\tau]. \qquad (3.9)$$

Equation (3.9) is the main result of this section from which δ_1 and δ_2, defined in (3.7), are immediately determined.

Important applications of (3.9) refer to the Lorenz and Bonferroni curves in fields like economics, reliability, demography, insurance and medicine. For a given probability π, they are defined by $L(\pi) = I(q)/\mu_1'$ and $B(\pi) = I(q)/(\pi\mu_1')$, respectively, where $q = F^{-1}(\pi)$.

3.4. Reliability. In the context of reliability, the stress-strength model describes the life of a component which has a random strength X_1 that is subjected to a random stress X_2. The component fails at the instant that the stress applied to it exceeds the strength, and the component will function satisfactorily whenever $X_1 > X_2$. Hence, $R = P(X_2 < X_1)$ is a measure of component reliability. It has many applications especially in engineering concepts such as structures, deterioration of rocket motors, static fatigue of ceramic components, fatigue failure of aircraft structures, and the aging of concrete pressure vessels. We now derive the corresponding form for the reliability R when X_1 and X_2 are independent and have identical ExGGG distribution. The reliability of the ExGGG distribution is

$$R = \int_0^\infty f(x)F(x)dx,$$

where $f(x)$ and $F(x)$ are calculated from (2.2) and (2.3), respectively.

For $\tau > 0$ the reliability can be written as

$$R = \int_0^\infty \sum_{j,m,d=0}^{\infty} w_{j,m,d}(k,p,\lambda) g_{\alpha,\tau,k^\bullet}(t)$$
$$\times \left(1 - \left\{1 - \frac{\gamma_1\left[k, \left(\frac{t}{\alpha}\right)^\tau\right]}{1 - p\left\{1 - \gamma_1\left[k, \left(\frac{t}{\alpha}\right)^\tau\right]\right\}}\right\}^\lambda\right) dt$$
$$= \sum_{j,m,d=0}^{\infty} w_{j,m,d}(k,p,\lambda)$$
$$\times \left(1 - \int_0^\infty g_{\alpha,\tau,k^\bullet}(t)\left\{1 - \frac{\gamma_1\left[k, \left(\frac{t}{\alpha}\right)^\tau\right]}{1 - p\left\{1 - \gamma_1\left[k, \left(\frac{t}{\alpha}\right)^\tau\right]\right\}}\right\}^\lambda dt\right).$$

Using the expansion (2.5) twice conveniently, we have

$$R = \sum_{j,m,d=0}^{\infty} w_{j,m,d}(k,p,\lambda)\Bigg(1 - \int_0^\infty g_{\alpha,\tau,k^\bullet}(t)$$
$$\times \sum_{r,u=0}^{\infty} (-1)^{r+u}\binom{\lambda}{r}\binom{-r}{u} p^u \gamma_1\left[k, \left(\frac{t}{\alpha}\right)^\tau\right]^r \left\{1 - \gamma_1\left[k, \left(\frac{t}{\alpha}\right)^\tau\right]\right\}^u dt\Bigg).$$

Using the binomial expansion, we obtain

$$R = \sum_{j,m,d=0}^{\infty} w_{j,m,d}(k,p,\lambda)\Bigg\{1 - \int_0^\infty g_{\alpha,\tau,k^\bullet}(t)$$
$$\times \sum_{r,u=0}^{\infty}\sum_{t_1=0}^{u} (-1)^{r+u+t_1}\binom{\lambda}{r}\binom{-r}{u}\binom{u}{t_1} p^u \gamma_1\left[k, \left(\frac{t}{\alpha}\right)^\tau\right]^{r+t_1} dt\Bigg\}.$$

We can substitute $\sum_{r,u=0}^{\infty}\sum_{t_1=0}^{u}$ for $\sum_{r,t_1=0}^{\infty}\sum_{u=t_1}^{\infty}$ to obtain

$$R = \sum_{j,m,d=0}^{\infty} w_{j,m,d}(k,p,\lambda)$$
$$\times \left\{1 - \int_0^\infty g_{\alpha,\tau,k^\bullet}(t) \sum_{r,t_1=0}^{\infty} s_{r,t_1}(p,\lambda)\gamma_1\left[k, \left(\frac{t}{\alpha}\right)^\tau\right]^{r+t_1} dt\right\}, \tag{3.10}$$

where

$$s_{r,t_1}(p,\lambda) = \sum_{u=t_1}^{\infty} (-1)^{r+u+t_1}\binom{\lambda}{r}\binom{-r}{u}\binom{u}{t_1} p^u. \tag{3.11}$$

As $g_{\alpha,\tau,k^\bullet}(t)$ has distribution $\mathrm{GG}(\alpha,\tau,k^\bullet)$, the expression (3.10) can be rewritten as

$$R=\sum_{j,m,d=0}^{\infty} w_{j,m,d}(k,p,\lambda)\left\{1-\int_0^\infty \frac{|\tau|}{\alpha\Gamma\left(k^\bullet\right)}\left(\frac{t}{\alpha}\right)^{\tau k^\bullet-1}\right.$$

$$\left.\times\exp\left[-\left(\frac{t}{\alpha}\right)^{\tau}\right]\sum_{r,t_1=0}^{\infty} s_{r,t_1}(p,\lambda)\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]^{r+t_1}dt\right\}. \tag{3.12}$$

Setting $w=\left(\frac{t}{\alpha}\right)^{\tau}$ in (3.12), we have

$$R=\sum_{j,m,d=0}^{\infty} w_{j,m,d}(k,p,\lambda)\left\{1-\sum_{r,t_1=0}^{\infty} v_{r,t_1}\left(k^\bullet,p,\lambda\right)I\left(k^\circ+d,r+t_1\right)\right\},$$

where

$$v_{r,t_1}\left(k^\bullet,p,\lambda\right)=\frac{s_{r,t_1}(p,\lambda)\operatorname{sgn}(\tau)}{\Gamma\left(k^\bullet\right)}, \tag{3.13}$$

$$I\left(k^\circ+d,r+t_1\right)=\int_0^\infty w^{k^\circ+d-1}\exp(-w)\gamma_1(k,w)^{r+t_1}dw, \tag{3.14}$$

$k^\circ=k(j+m+1)$ and $w_{j,m,d}(k,p,\lambda)$ is defined by the expression (2.9). Using the Lauricella function of type A (defined in Section 4), the last integral can be written as

$$\begin{aligned}I\left(k^\circ+d,r+t_1\right)&=\left(k^\circ\right)^{-r-t_1}\Gamma\left[d+k^\circ(r+t_1+1)\right]\\&\times F_A^{(r+t_1)}(d+k^\circ(r+t_1+1);k^\circ,\cdots,k^\circ;k^\circ+1,\cdots,k^\circ+1;-1,\cdots,-1).\end{aligned} \tag{3.15}$$

Now, for $\tau<0$ the reliability can be written as

$$R=\int_0^\infty\sum_{j,m,d=0}^{\infty} w_{j,m,d}(k,p,\lambda)g_{\alpha,\tau,k^\bullet}(t)\left\{1-\frac{\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]}{1-p\left\{1-\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]\right\}}\right\}^{\lambda}dt.$$

Exchanging the addition and integration operations and using the expansion (2.5) twice conveniently in the last expression, we have

$$\begin{aligned}R=&\sum_{j,m,d=0}^{\infty} w_{j,m,d}(k,p,\lambda)\int_0^\infty g_{\alpha,\tau,k^\bullet}(t)\\&\times\sum_{r,u=0}^{\infty}(-1)^{r+u}\binom{\lambda}{r}\binom{-r}{u}p^u\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]^{r}\left\{1-\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]\right\}^{u}dt.\end{aligned}$$

Using the binomial expansion gets

$$\begin{aligned}R=&\sum_{j,m,d=0}^{\infty} w_{j,m,d}(k,p,\lambda)\int_0^\infty g_{\alpha,\tau,k^\bullet}(t)\\&\times\sum_{r,u=0}^{\infty}\sum_{t_1=0}^{u}(-1)^{r+u+t_1}\binom{\lambda}{r}\binom{-r}{u}\binom{u}{t_1}p^u\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]^{r+t_1}dt.\end{aligned}$$

We can substitute $\sum_{r,u=0}^{\infty}\sum_{t_1=0}^{u}$ for $\sum_{r,t_1=0}^{\infty}\sum_{u=t_1}^{\infty}$ to obtain

$$R=\sum_{j,m,d,r,t_1=0}^{\infty} w_{j,m,d}(k,p,\lambda)\int_0^\infty g_{\alpha,\tau,k^\bullet}(t)s_{r,t_1}(p,\lambda)\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^\tau\right]^{r+t_1}dt, \tag{3.16}$$

where $s_{r,t_1}(p,\lambda)$ is defined in (3.11).

How $g_{\alpha,\tau,k^\bullet}(t)$ has distribution $\mathrm{GG}(\alpha,\tau,k^\bullet)$, the expression (3.16) can be rewritten as

$$\begin{aligned}R=&\sum_{j,m,d,r,t_1=0}^{\infty} w_{j,m,d}(k,p,\lambda)\int_0^\infty \frac{|\tau|}{\alpha\Gamma\left(k^\bullet\right)}\left(\frac{t}{\alpha}\right)^{\tau k^\bullet-1}\\ &\times\exp\left[-\left(\frac{t}{\alpha}\right)^\tau\right]s_{r,t_1}(p,\lambda)\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^\tau\right]^{r+t_1}dt.\end{aligned} \tag{3.17}$$

Setting $w=\left(\frac{t}{\alpha}\right)^\tau$ in (3.17), we have

$$R=\sum_{j,m,d,r,t_1=0}^{\infty} w_{j,m,d}(k,p,\lambda)v_{r,t_1}\left(k^\bullet,p,\lambda\right)I\left(k^\circ+d,r+t_1\right),$$

where $v_{r,t_1}\left(k^\bullet,p,\lambda\right)$, $I\left(k^\circ+d,r+t_1\right)$ and $w_{j,m,d}(k,p,\lambda)$ are given by (3.13), (3.14) and (2.9), respectively. The integral $I\left(k^\circ+d,r+t_1\right)$ also can be written by (3.15).

3.5. Order Statistics. Order statistics make their appearance in many areas of statistical theory and practice. The density function $f_{i:n}(t)$ of the ith order statistic, say $T_{i:n}$, for $i=1,\ldots,n$, from random variables $T_1,\ldots,T_n$ having density (2.2), is given by

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

where $f(t)$ and $F(t)$ are the pdf and cdf of the ExGGG distribution, respectively and $B(\cdot,\cdot)$ denotes the beta function. We readily obtain using the binomial expansion

$$f_{i:n}(t)=\sum_{j_1=0}^{n-i}\frac{(-1)^{j_1}\binom{n-i}{j_1}}{B(i,n-i+1)}f(t)F(t)^{i+j_1-1}. \tag{3.18}$$

We now derive an expression for the density of the ith order statistics as a linear combination of GG densities. We shall consider two cases: $\tau>0$ and $\tau<0$.

For $\tau > 0$ and u positive integer, we have

$$\begin{aligned}
f(t)F(t)^u &= \frac{\lambda|\tau|\,(1-p)}{\alpha\Gamma\,(k)}\left(\frac{t}{\alpha}\right)^{\tau k-1}\exp\left[-\left(\frac{t}{\alpha}\right)^{\tau}\right] \\
&\times \left(1-p\left\{1-\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]\right\}\right)^{-2} \\
&\times \left\{1-\frac{\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]}{\left(1-p\left\{1-\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]\right\}\right)}\right\}^{\lambda-1} \\
&\times \left(1-\left\{1-\frac{\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]}{1-p\left\{1-\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]\right\}}\right\}^{\lambda}\right)^{u}.
\end{aligned}$$

Using the binomial expansion and (2.5) twice conveniently, we have

$$\begin{aligned}
f(t)F(t)^u = &\frac{\lambda|\tau|\,(1-p)}{\alpha\Gamma\,(k)}\left(\frac{t}{\alpha}\right)^{\tau k-1}\exp\left[-\left(\frac{t}{\alpha}\right)^{\tau}\right] \\
&\times \sum_{l_1=0}^{u}\sum_{j,a=0}^{\infty}(-1)^{l_1+j+a}\binom{u}{l_1}\binom{\lambda(l_1+1)-1}{j}\binom{-(j+2)}{a}p^a \\
&\times \gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]^{j}\left\{1-\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]\right\}^{a}.
\end{aligned}$$

Now, using the binomial expansion in the expression $\left\{1-\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]\right\}^{a}$, we obtain

$$\begin{aligned}
f(t)F(t)^u &= \frac{\lambda|\tau|\,(1-p)}{\alpha\Gamma\,(k)}\left(\frac{t}{\alpha}\right)^{\tau k-1}\exp\left[-\left(\frac{t}{\alpha}\right)^{\tau}\right] \\
&\times \sum_{l_1=0}^{u}\sum_{j,a=0}^{\infty}\sum_{m=0}^{a}(-1)^{l_1+j+a+m}\binom{u}{l_1}\binom{\lambda(l_1+1)-1}{j}\binom{-(j+2)}{a}\binom{a}{m}p^a \\
&\times \gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]^{j+m}.
\end{aligned}$$

We can substitute $\sum_{a=0}^{\infty}\sum_{m=0}^{a}$ for $\sum_{m=0}^{\infty}\sum_{a=m}^{\infty}$ to obtain

$$\begin{aligned}
f(t)F(t)^u = &\frac{\lambda|\tau|\,(1-p)}{\alpha\Gamma\,(k)}\left(\frac{t}{\alpha}\right)^{\tau k-1}\exp\left[-\left(\frac{t}{\alpha}\right)^{\tau}\right] \\
&\times \sum_{j,m=0}^{\infty}\rho_{j,m,u}(p,\lambda)\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]^{j+m}, \qquad (3.19)
\end{aligned}$$

where

$$\rho_{j,m,u}(p,\lambda) = \sum_{l_1=0}^{u}\sum_{a=m}^{\infty}(-1)^{l_1+j+a+m}\binom{u}{l_1}\binom{\lambda(l_1+1)-1}{j}\binom{-(j+2)}{a}\binom{a}{m}p^a.$$

Inserting $f(t)F(t)^u$ given by (3.19) in expression (3.18), applying expansion (A.3) and rearranging terms, the density function of the ith ExGGG order statistic for $\tau > 0$ is expressed by

$$(3.20)\quad f_{i:n}(t) = \sum_{j,m,d=0}^{\infty} v_{i:n,j,m,d}(k,p,\lambda) g_{\alpha,\tau,k^{\bullet}}(t),$$

where

$$v_{i:n,j,m,d} = \frac{\lambda\,(1-p)\,c_{j+m,d}\Gamma\left(k^{\bullet}\right)}{B(i,n-i+1)\Gamma\left(k\right)^{j+m+1}} \sum_{j_1=0}^{n-i} (-1)^{j_1} \binom{n-i}{j_1} \rho_{j,m,i+j_1-1}(p,\lambda),$$

$k^{\bullet} = k(j+m+1)+d$, $\rho_{j,m,i+j_1-1}(p,\lambda)$ is defined above and $c_{j+m,d}$ is calculated recursively by (A.2).

For $\tau < 0$ and u positive integer, we have that $f(t)F(t)^u$ can be written as

$$\begin{aligned} f(t)F(t)^u &= \frac{\lambda|\tau|\,(1-p)}{\alpha\Gamma\left(k\right)} \left(\frac{t}{\alpha}\right)^{\tau k-1} \exp\left[-\left(\frac{t}{\alpha}\right)^{\tau}\right] \\ &\times \left(1-p\left\{1-\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]\right\}\right)^{-2} \\ &\times \left\{1-\frac{\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]}{\left(1-p\left\{1-\gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]\right\}\right)}\right\}^{\lambda(u+1)-1}. \end{aligned}$$

Using the power series (2.5) twice and binomial expansion, we have

$$\begin{aligned} f(t)F(t)^u &= \frac{\lambda|\tau|\,(1-p)}{\alpha\Gamma\left(k\right)} \left(\frac{t}{\alpha}\right)^{\tau k-1} \exp\left[-\left(\frac{t}{\alpha}\right)^{\tau}\right] \\ &\times \sum_{j,a=0}^{\infty} \sum_{m=0}^{a} (-1)^{j+a+m} \binom{\lambda(u+1)-1}{j} \binom{-(j+2)}{a} \binom{a}{m} p^a \\ &\times \gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]^{j+m}. \end{aligned}$$

We can substitute $\sum_{a=0}^{\infty}\sum_{m=0}^{a}$ for $\sum_{m=0}^{\infty}\sum_{a=m}^{\infty}$ to obtain

$$\begin{aligned} f(t)F(t)^u = {} & \frac{\lambda|\tau|\,(1-p)}{\alpha\Gamma\left(k\right)} \left(\frac{t}{\alpha}\right)^{\tau k-1} \exp\left[-\left(\frac{t}{\alpha}\right)^{\tau}\right] \\ (3.21)\qquad & \times \sum_{j,m=0}^{\infty} \rho_{j,m,u}(p,\lambda) \gamma_1\left[k,\left(\frac{t}{\alpha}\right)^{\tau}\right]^{j+m}, \end{aligned}$$

where

$$\rho_{j,m,u}(p,\lambda) = \sum_{a=m}^{\infty} (-1)^{j+a+m} \binom{\lambda(u+1)-1}{j} \binom{-(j+2)}{a} \binom{a}{m} p^a.$$

Inserting $f(t)F(t)^u$ given by (3.21) in expression (3.18), applying expansion (A.3) and rearranging terms, the density function of the ith ExGGG order statistic for $\tau < 0$ is expressed by

$$(3.22)\quad f_{i:n}(t) = \sum_{j,m,d=0}^{\infty} v_{i:n,j,m,d}(k,p,\lambda) g_{\alpha,\tau,k^{\bullet}}(t),$$

where

$$v_{i:n,j,m,d} = \frac{\lambda\,(1-p)\,c_{j+m,d}\Gamma\left(k^{\bullet}\right)}{B(i,n-i+1)\Gamma\left(k\right)^{j+m+1}} \sum_{j_1=0}^{n-i} (-1)^{j_1} \binom{n-i}{j_1} \rho_{j,m,i+j_1-1}(p,\lambda),$$

$k^{\bullet} = k(j+m+1)+d$, $\rho_{j,m,i+j_1-1}(p,\lambda)$ is defined above and $c_{j+m,d}$ is calculated recursively by (A.2).

Density functions (3.20) and (3.22) gives the density function of the ith order statistic as a linear combination of GG densities. Hence, some of the mathematical quantities of the ExGGG order statistics can be derived by knowing those of the GG distribution. For example, the rth ordinary moment and the mgf $M_{i:n}(s)$ of $T_{i:n}$ are

$$E(T_{i:n}^{r}) = \alpha^{r} \sum_{j,m,d=0}^{\infty} v_{i:n,j,m,d}(k,p,\lambda) \frac{\Gamma\left(k^{\bullet}+r/\tau\right)}{\Gamma\left(k^{\bullet}\right)}$$

and

$$M_{i:n}(s) = \sum_{j,m,d=0}^{\infty} v_{i:n,j,m,d}(k,p,\lambda) M_{\alpha,\tau,k^{\bullet}}(s),$$

where $M_{\alpha,\tau,k^{\bullet}}(s)$ can be calculated from expression (3.5) or (3.6).

4. The Log-Extended Generalized Gamma Geometric Distribution

Henceforth, T is a random variable following the ExGGG density function (2.2) and Y is now defined by $Y = \log(T)$. By setting $k = q^{-2}$, $\tau = (\sigma\sqrt{k})^{-1}$ and $\alpha = \exp\left\{\mu - \log(k)\tau^{-1}\right\}$, the density function of Y can be expressed as

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

where $-\infty < y,\ \mu < \infty$, $\sigma > 0$, $\lambda > 0$ and q is different from zero. We consider an extended form including the case $q = 0$ [33]. Thus, the density of Y can be expressed as

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

where $\Phi(\cdot)$ is the standard normal cumulative distribution. We refer to equation (4.1) as the LExGGG distribution, say $Y \sim \text{LExGGG}(\mu,\sigma,q,p,\lambda)$, where $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ is the scale parameter and q and p are shape parameters. For $q = 0$ and $\lambda = 1$ we obtain the normal geometric distribution (new), for $q = 0$ and $p \to 0^+$ we have extended normal [12] and for $q > 0$ and $\lambda = 1$ we have generalized gamma geometric distribution [42]. Thus,

$$
\text{if} \quad T \sim \text{ExGGG}(\alpha,\tau,k,p,\lambda) \quad \text{then, we have} \quad Y=\log(T) \sim \text{LExGGG}(\mu,\sigma,q,p,\lambda).
$$

Letting $\mu = 0$ and $\sigma = 1$, the plots of the density function (4.1) for selected values of p, when $q < 0$, $q > 0$ and $q = 0$, are given in Figure 3. These plots clearly show that the LExGGG distribution could be very flexible for modeling its kurtosis.

The survival function of Y becomes

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

4.1. Moments. Now, we derive the rth moment of $Y \sim \text{LExGGG}(\mu,\sigma,q,p,\lambda)$, say μ'_r.

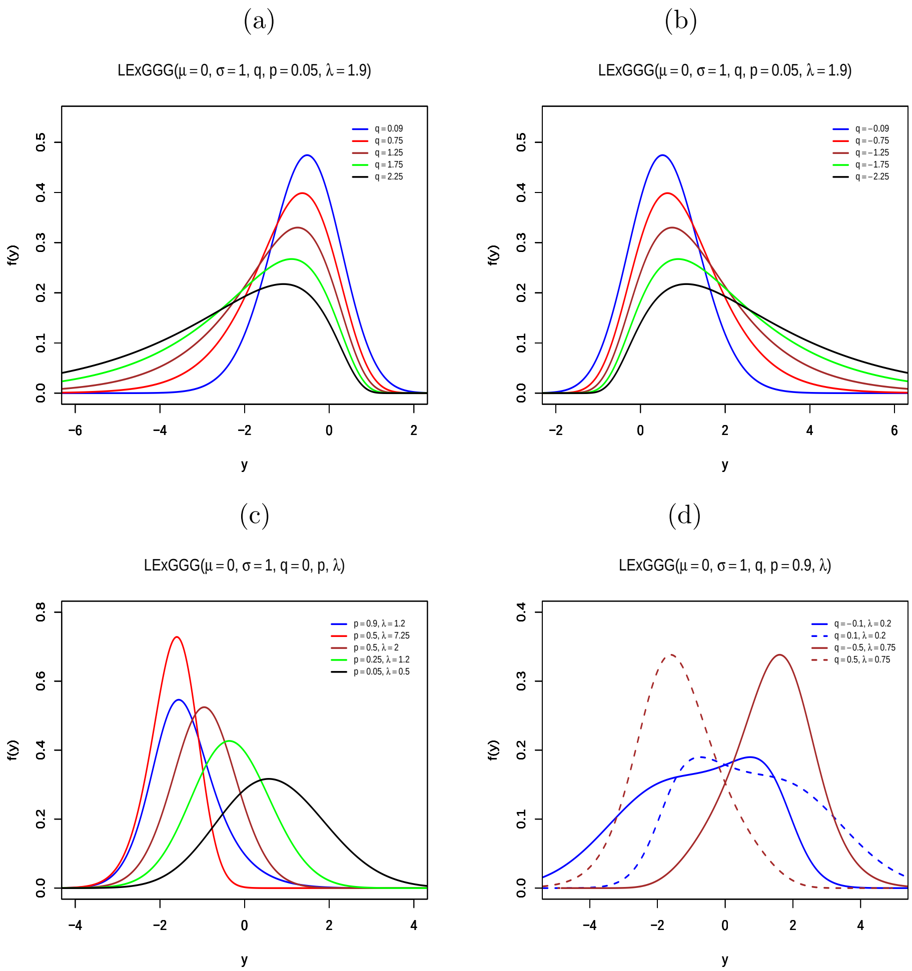

Figure 3. The LExGGG density curves: (a) For some values of $q > 0$. (b) For some values of $q < 0$. (c) For some values of $q = 0$. (c) For some values of q and λ.

For $q \neq 0$, we obtain

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

Setting $x = q^{-2}\exp\left[q\left(\frac{y-\mu}{\sigma}\right)\right]$, using (2.5) twice and binomial expansion, μ'_r reduces to

$$\mu'_r = \frac{\lambda\,\mathrm{sgn}(q)(1-p)}{\Gamma(q^{-2})}\sum_{j,l=0}^{\infty}\sum_{m=0}^{l}(-1)^{j+l+m}\binom{\lambda-1}{j}\binom{-(j+2)}{l}\binom{l}{m}p^l$$
$$\times\int_0^\infty\left\{\frac{\sigma}{q}\left[2\log\left(|q|\right)+\log\left(x\right)\right]+\mu\right\}^r x^{q^{-2}-1}\exp\left(-x\right)\gamma_1\big(q^{-2},x\big)^{j+m}dx.$$

Substituting $\sum_{j,l=0}^{\infty}\sum_{m=0}^{l}$ by $\sum_{j,m=0}^{\infty}\sum_{l=m}^{\infty}$, we obtain

$$\mu'_r = \frac{\lambda\,\mathrm{sgn}(q)(1-p)}{\Gamma(q^{-2})}\sum_{j,m=0}^{\infty}s_{j,m}(\lambda,p)$$
$$\times\int_0^\infty\left\{\frac{\sigma}{q}\left[2\log\left(|q|\right)+\log\left(x\right)\right]+\mu\right\}^r x^{q^{-2}-1}\exp\left(-x\right)\gamma_1\big(q^{-2},x\big)^{j+m}dx \tag{4.3}$$

where the quantity $s_{j,m}(\lambda,p)$ is given by equation (2.7). Using the equation (A.3) in the expression (4.3), we obtain

$$\mu'_r = \sum_{j,m,d=0}^{\infty}\frac{\lambda\,\mathrm{sgn}(q)(1-p)s_{j,m}(\lambda,p)c_{j+m,d}}{\Gamma(q^{-2})^{j+m+1}}$$
$$\times\int_0^\infty\left\{\frac{\sigma}{q}\left[2\log\left(|q|\right)+\log\left(x\right)\right]+\mu\right\}^r x^{q^{-2}(j+m+1)+d-1}\exp\left(-x\right)dx,$$

where the coefficients $c_{j+m,d}$ are calculated from (A.2).

Note that,

$$\left\{\frac{2\sigma}{q}\log(|q|)+\mu+\frac{\sigma}{q}\log(x)\right\}^r = \sum_{l=0}^{r}\binom{r}{l}\left[\frac{2\sigma}{q}\log(|q|)+\mu\right]^{r-l}\left(\frac{\sigma}{q}\right)^l\log^l(x),$$

and then

$$\mu'_r = \sum_{j,m,d=0}^{\infty}\sum_{l=0}^{r}\frac{\lambda\,\mathrm{sgn}(q)(1-p)s_{j,m}(\lambda,p)c_{j+m,d}}{\Gamma(q^{-2})^{j+m+1}}\binom{r}{l}$$
$$\times\left[\frac{2\sigma}{q}\log\left(|q|\right)+\mu\right]^{r-l}\left(\frac{\sigma}{q}\right)^l\int_0^\infty\log^l\left(x\right)x^{q^{-2}(j+m+1)+d-1}\exp\left(-x\right)dx.$$

The last integral is given in Prudnikov *et al.* [47] (vol 1, Section 2.6.21) and then

$$\mu'_r = \sum_{j,m,d=0}^{\infty}\sum_{l=0}^{r}\frac{\lambda\,\mathrm{sgn}(q)(1-p)s_{j,m}(\lambda,p)c_{j+m,d}}{\Gamma(q^{-2})^{j+m+1}}\binom{r}{l}$$
$$\times\left[\frac{2\sigma}{q}\log\left(|q|\right)+\mu\right]^{r-l}\left(\frac{\sigma}{q}\right)^l\Gamma^{(l)}\left[q^{-2}(j+m+1)+d\right],$$

where $\Gamma^{(n)}(p) = \frac{\partial^n\Gamma(p)}{\partial p^n}$.

For $q=0$, we obtain

$$\mu_r' = \frac{\lambda(1-p)}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} y^r \exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right]$$
$$\times \left\{1-p\left[1-\Phi\left(\frac{y-\mu}{\sigma}\right)\right]\right\}^{-2}\left\{1-\frac{\Phi\left(\frac{y-\mu}{\sigma}\right)}{1-p\left[1-\Phi\left(\frac{y-\mu}{\sigma}\right)\right]}\right\}^{\lambda-1} dy.$$

Using (2.5) twice and binomial expansion, we have

$$\mu_r' = \frac{\lambda(1-p)}{\sqrt{2\pi}\sigma}\sum_{j,l=0}^{\infty}\sum_{m=0}^{l}(-1)^{j+l+m}\binom{\lambda-1}{j}\binom{-(j+2)}{l}\binom{l}{m}p^l$$
$$\times \int_{-\infty}^{\infty} y^r \exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right]\Phi^{j+m}\left(\frac{y-\mu}{\sigma}\right)dy.$$

Substituting $\sum_{j,l=0}^{\infty}\sum_{m=0}^{l}$ by $\sum_{j,m=0}^{\infty}\sum_{l=m}^{\infty}$, we obtain

$$\mu_r' = \frac{\lambda(1-p)}{\sqrt{2\pi}\sigma}\sum_{j,m=0}^{\infty} s_{j,m}(\lambda,p)\int_{-\infty}^{\infty} y^r \exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right]\Phi^{j+m}\left(\frac{y-\mu}{\sigma}\right)dy,$$

where $s_{j,m}(\lambda,p)$ is given by equation (2.7). Setting $y=\mu+\sigma z$, we obtain

$$\mu_r' = \frac{\lambda(1-p)}{\sqrt{2\pi}}\sum_{j,m=0}^{\infty} s_{j,m}(\lambda,p)\int_{-\infty}^{\infty}(\mu+\sigma z)^r \exp\left(-\frac{z^2}{2}\right)\Phi^{j+m}(z)\,dz \tag{4.4}$$

Expanding the binomial term $(\mu+\sigma z)^r$ gives

$$\mu_r' = \lambda(1-p)\sum_{j,m=0}^{\infty}\sum_{l=0}^{r} s_{j,m}(\lambda,p)\mu^{r-l}\sigma^l\binom{r}{l}\nu_{l,j+m}$$

where $\nu_{j,p}$ is the probability weighted moments (PWM) for j and p non-negative integers of the standardized normal distribution, defined by

$$\nu_{j,p} = \int_{-\infty}^{\infty} z^j\phi(z)\Phi^p(z)dz,$$

where $\phi(z)$ is the standard normal density.

The standard cumulative normal can be expressed as

$$\Phi(x) = \frac{1}{2}\left\{1+\operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right\}, \quad x\in\mathbb{R},$$

where $\operatorname{erf}(\cdot)$ denotes the error function defined by

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x \exp(-t^2)dt.$$

Using the binomial expansion and interchanging terms, we obtain

$$\nu_{j,p} = \frac{1}{2^p\sqrt{2\pi}}\sum_{i=0}^{p}\binom{p}{i}\int_{-\infty}^{\infty} z^j\exp\left(-\frac{z^2}{2}\right)\left[\operatorname{erf}\left(\frac{z}{\sqrt{2}}\right)\right]^i dz.$$

The last integral follows from equations (9)-(11) of Nadarajah [38]. When $i+j$ is even, we have

$$\begin{aligned} \nu_{j,p} = & \sum_{i=0}^{p} \binom{p}{i} 2^{j/2+i-p} \pi^{-(i+1)/2} \Gamma\left(\frac{i+j+1}{2}\right) \times \\ & F_A^{(i)}\left(\frac{i+j+1}{2}; \frac{1}{2}, \cdots, \frac{1}{2}; \frac{3}{2}, \cdots, \frac{3}{2}; -1, \cdots, -1\right). \end{aligned} \tag{4.5}$$

Inserting (4.5) in (4.4) yields the sth moment of the LExGGG distribution when $q = 0$. The skewness and kurtosis measures can be calculated from the ordinary moments using well-known relationships. Plots of the skewness and kurtosis for selected values of μ, σ, q, p and λ are given in Figures 4 (a) and (b), respectively.

(a) (b)

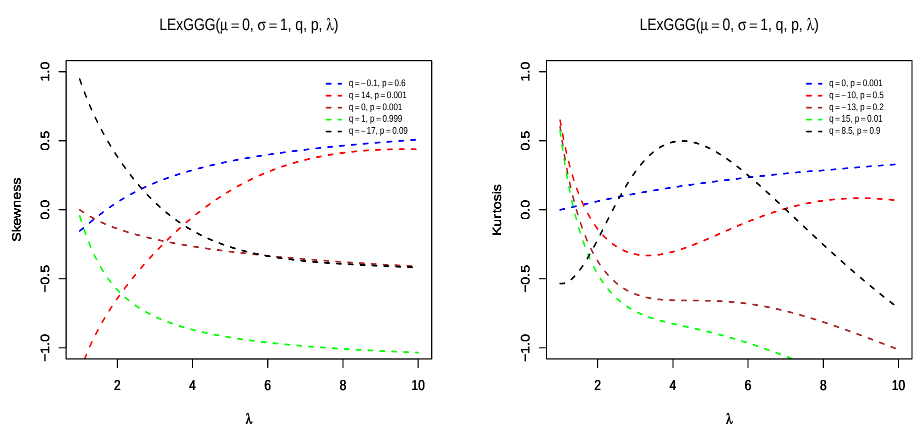

Figure 4. Skewness and kurtosis of the LExGGG distribution as a function of λ for some values of q and p with $\mu = 0$ and $\sigma = 1$.

5. The Log-Extended Generalized Gamma Geometric Regression Model

When the lifetime is affected by explanatory variables, such as treatment, group indicator, individual characteristics, environmental conditions and many others, parametric regression models can be used. A parametric model that provides a good fit to lifetime data tends to yield more precise estimates of the quantities of interest.

If $Y \sim \text{LExGGG}(\mu, \sigma, q, p, \lambda)$, we define the standardized random variable $Z = (Y - \mu)/\sigma$ with a density function given by

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

and we write $Z \sim \text{LExGGG}(0, 1, q, p, \lambda)$.

Now, based on the LExGGG density function, we propose a linear location-scale regression model linking the response variable y_i and the explanatory variable vector $\mathbf{x}_i^T = (x_{i1}, \ldots, x_{il})$ by

$$
(5.2)\qquad y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \sigma z_i, \; i = 1, \cdots, n,
$$

where the random error z_i has density function (5.1), $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_l)^T$, $\sigma > 0$, $-\infty < q < \infty$, $p \in (0, 1)$ and $\lambda > 0$ are unknown parameters. The parameter $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$ is the location of y_i. The location parameter vector $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_n)^T$ is represented by a linear model $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, where $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_n)^T$ is a known model matrix. The LExGGG model (5.2) opens new possibilities for fitting many different types of data. It contains as special sub-models some well-known regression models. For $\lambda = 1$, we obtain the log-generalized gamma geometric (LGGG) regression model (new). If $\lambda = 1$ and $p \to 0^+$, the LExGGG reduces to the log-generalized gamma (LGG) regression model [33, 39]. The LGG regression model has been studied in more detail by Ortega *et al.* [40, 41]. For $q = 1$, we obtain the log-extended Weibull geometric (LExWG) regression model (new). If $\sigma = 1$ and $\sigma = 0.5$, in addition to $q = \lambda = 1$ and $p \to 0^+$, we have the exponential and Rayleigh regression model, respectively. If $\lambda = 1$ and $p \to 0^+$, in addition to $q = -1$, it follows the log-inverse Weibull regression model [31]. The case $q = \lambda = 1$ and $p \to 0^+$ refers to the classical log-Weibull (or extreme value) regression model [33]. For $q = 0$, the LExGGG regression model reduces to the log-extended normal geometric regression model (new). Finally, for $q = 0$, $\lambda = 1$ and $p \to 0^+$, we obtain the normal regression model.

5.1. Maximum Likelihood Estimation. Consider a sample $(y_1, \mathbf{x}_1), \cdots, (y_n, \mathbf{x}_n)$ of size n from the random variables $Y_i \sim \text{LExGGG}(\mu_i, \sigma, q, p, \lambda)$, where $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$ is the location parameter of Y_i and $\mathbf{x}_i$ is the explanatory variable vector associated with the ith individual, $i = 1, \cdots, n$. The log-likelihood function for the vector of parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma, q, p, \lambda)^T$ from model (5.2) can be expressed as

$$(5.3) \qquad \ell(\boldsymbol{\theta}) = \begin{cases} n \log\left[\frac{\lambda q(1-p)(q^{-2})^{q^{-2}}}{\Gamma(q^{-2})}\right] + q^{-1}\sum_{i=1}^{n} z_i \\ -q^{-2}\sum_{i=1}^{n}\exp(qz_i) \\ -2\sum_{i=1}^{n}\log\left(1-p\left\{1-\gamma_1\left[q^{-2}, q^{-2}\exp(qz_i)\right]\right\}\right) \\ +(\lambda-1)\sum_{i=1}^{n}\log\left(1-\frac{\gamma_1\left[q^{-2},q^{-2}\exp(qz_i)\right]}{1-p\{1-\gamma_1[q^{-2},q^{-2}\exp(qz_i)]\}}\right), & \text{if} \quad q>0, \\ n \log\left[\frac{\lambda(-q)(1-p)(q^{-2})^{q^{-2}}}{\Gamma(q^{-2})}\right] + q^{-1}\sum_{i=1}^{n} z_i \\ -q^{-2}\sum_{i=1}^{n}\exp(qz_i) \\ -2\sum_{i=1}^{n}\log\left(1-p\left\{1-\gamma_1\left[q^{-2}, q^{-2}\exp(qz_i)\right]\right\}\right) \\ +(\lambda-1)\sum_{i=1}^{n}\log\left(1-\frac{\gamma_1\left[q^{-2},q^{-2}\exp(qz_i)\right]}{1-p\{1-\gamma_1[q^{-2},q^{-2}\exp(qz_i)]\}}\right), & \text{if} \quad q<0, \\ n\log\left[\frac{\lambda(1-p)}{\sqrt{2\pi}}\right] - \frac{1}{2}\sum_{i=1}^{n} z_i^2 \\ -2\sum_{i=1}^{n}\log\left\{1-p\left[1-\Phi(z_i)\right]\right\} \\ +(\lambda-1)\sum_{i=1}^{n}\log\left\{1-\frac{\Phi(z_i)}{1-p[1-\Phi(z_i)]}\right\}, & \text{if} \quad q=0, \end{cases}$$

where $z_i = (y_i - \mathbf{x}_i^T\boldsymbol{\beta})/\sigma$.

The maximum likelihood estimate (MLE) $\widehat{\boldsymbol{\theta}}$ of the vector $\boldsymbol{\theta}$ of unknown parameters can be calculated by maximizing the log-likelihood (5.3) that can be either directly using the SAS (Proc NLMixed) or the R (optim). Initial values for σ and q can be taken from the fit of the LGG regression model with $p = 0$ and $\lambda = 1$. If $\hat{z}_i = (y_i - \mathbf{x}_i^T\hat{\boldsymbol{\beta}})/\hat{\sigma}$, the fit of the LExGGG model yields the estimated survival function for y_i

$$\hat{S}(y_i; \widehat{\boldsymbol{\beta}}^T, \hat{\sigma}, \hat{q}, \hat{p}, \hat{\lambda}) = \begin{cases} \left(1-\frac{\gamma_1\left[\hat{q}^{-2},\hat{q}^{-2}\exp(\hat{q}\hat{z}_i)\right]}{1-\hat{p}\{1-\gamma_1[\hat{q}^{-2},\hat{q}^{-2}\exp(\hat{q}\hat{z}_i)]\}}\right)^{\hat{\lambda}}, & \text{if} \quad q>0, \\ 1-\left(1-\frac{\gamma_1\left[\hat{q}^{-2},\hat{q}^{-2}\exp(\hat{q}\hat{z}_i)\right]}{1-\hat{p}\{1-\gamma_1[\hat{q}^{-2},\hat{q}^{-2}\exp(\hat{q}\hat{z}_i)]\}}\right)^{\hat{\lambda}}, & \text{if} \quad q<0, \\ \left\{1-\frac{\Phi(\hat{z}_i)}{1-\hat{p}[1-\Phi(\hat{z}_i)]}\right\}^{\hat{\lambda}}, & \text{if} \quad q=0. \end{cases}$$

For interval estimation and hypothesis tests on the model parameters, we require the $(l+4)\times(l+4)$ total observed information matrix $J(\boldsymbol{\theta})$. Under conditions that are fulfilled for the parameter vector $\boldsymbol{\theta}$ in the interior of the parameter space but not on the boundary, the asymptotic distribution of $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is multivariate normal $N_{l+4}(0, I(\boldsymbol{\theta})^{-1})$, where $I(\boldsymbol{\theta})$ is the expected information matrix. In practice, we can replace $I(\boldsymbol{\theta})$ by $J(\widehat{\boldsymbol{\theta}})$, that is, the observed information matrix evaluated at $\widehat{\boldsymbol{\theta}}$, say

$$J = J(\boldsymbol{\theta}) = \begin{pmatrix} J_{\beta_j\beta_s} & J_{\beta_j\sigma} & J_{\beta_j q} & J_{\beta_j p} & J_{\beta_j\lambda} \\ \cdot & J_{\sigma\sigma} & J_{\sigma q} & J_{\sigma p} & J_{\sigma\lambda} \\ \cdot & \cdot & J_{qq} & J_{qp} & J_{q\lambda} \\ \cdot & \cdot & \cdot & J_{pp} & J_{p\lambda} \\ \cdot & \cdot & \cdot & \cdot & J_{\lambda\lambda} \end{pmatrix},$$

whose elements are given in Appendix B, or it can also be calculated numerically. The approximate multivariate normal distribution $N_{l+4}(0, J(\widehat{\boldsymbol{\theta}})^{-1})$ for $\widehat{\boldsymbol{\theta}}$ can be used in the classical way to construct approximate confidence regions for some components of parameters in $\boldsymbol{\theta}$. Further, the likelihood ratio (LR) statistic can be used for comparing some special sub-models with the LExGGG model. Consider the partition $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$, where $\boldsymbol{\theta}_1$ is a subset of parameters of interest and $\boldsymbol{\theta}_2$ is a subset of remaining parameters. The LR statistic for testing the null hypothesis $H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^{(0)}$ versus the alternative hypothesis $H_1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_1^{(0)}$ is given by $LR = 2\{\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widetilde{\boldsymbol{\theta}})\}$, where $\widetilde{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}$ are the MLEs under the null and alternative hypotheses, respectively. The LR statistic is asymptotically (as $n \to \infty$) distributed as χ^2_ν, where ν is the dimension of the subset of parameters $\boldsymbol{\theta}_1$ of interest.

The interpretation of the estimated coefficients is based on the ratio of the median times. When the explanatory variable x is binary (0 or 1) and considering the ratio of the median times with $x = 1$ in the numerator, if $\hat{\beta}$ is negative (positive), this implies that individuals with $x = 1$ present decreased (increased) median survival time in $e^{\hat{\beta}} \times 100\%$ relative to the individuals in the group with $x = 0$, excluding the effects of the other explanatory variables [21]. This interpretation can be extended to continuous or categorical variables.

5.2. A Bayesian analysis. The Bayesian analysis incorporates previous knowledge of the parameters through informative priori density functions. When this information is not available, we can consider a noninformative prior. In the Bayesian approach, the information referring to the model parameters is obtained through a posterior marginal distribution. Here two difficulties usually arise. The first refers to attaining marginal posterior distribution, and the second to the calculation of the moments of interest. Both cases require numerical integration that, many times, do not present an analytical solution. So we use the simulation method of Markov Chain Monte Carlo (MCMC), such as the Gibbs sampler and Metropolis-Hastings algorithm [46].

Since we have no prior information from historical data or from previous experiment, we assign conjugate but weakly informative prior distributions to the parameters such as Al-Zahrani *et al.* [3] and Pascoa *et al.* [46]. Since we assumed informative (but weakly) prior distribution, the posterior distribution is a well-defined proper distribution. Here, we assume the elements of the parameter vector to be independent and consider that the joint prior distribution of all unknown parameters has a density function given by

$$\pi(q, p, \lambda, \sigma, \boldsymbol{\beta}) \propto \pi(q) \times \pi(p) \times \pi(\lambda) \times \pi(\sigma) \times \pi(\boldsymbol{\beta}) \tag{5.4}$$

Here, $q \sim N(\mu_1, \sigma_1^2)$, $p \sim \text{Beta}(a, b)$, $\lambda \sim \Gamma(a_1, b_1)$, $\sigma \sim \Gamma(a_2, b_2)$ and $\boldsymbol{\beta}_s \sim N(\mu_s, \sigma_s^2)$. All hyper-parameters are specified. Combining the likelihood function (5.3) and the prior distribution (5.4), the joint posterior distribution for β_s^T, σ, q, p and λ reduces to

$$
\begin{aligned}
\pi(\boldsymbol{\beta}, \sigma, q, p, \lambda | y) \quad &\propto \quad \left[\frac{\lambda |q| (1-p) (q^{-2})^{q^{-2}}}{\sigma \Gamma(q^{-2})} \right]^n \exp\left(q^{-1} \sum_{i=1}^{n} \frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma} \right) \\
&\times \quad \exp\left\{ -q^{-2} \sum_{i=1}^{n} \exp\left[q \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma} \right) \right] \right\} \\
&\times \quad \prod_{i=1}^{n} \left\{ 1 - p \left(1 - \gamma_1 \left\{ q^{-2}, q^{-2} \exp\left[q \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma} \right) \right] \right\} \right) \right\}^{-2} \\
&\times \quad \prod_{i=1}^{n} \left\{ 1 - \frac{\gamma_1 \left\{ q^{-2}, q^{-2} \exp\left[q \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma} \right) \right] \right\}}{1 - p \left(1 - \gamma_1 \left\{ q^{-2}, q^{-2} \exp\left[q \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma} \right) \right] \right\} \right)} \right\}^{\lambda - 1} \\
&\times \quad \pi(q, p, \lambda, \sigma, \boldsymbol{\beta}). \qquad (5.5)
\end{aligned}
$$

The joint posterior density (5.5) is analytically intractable because the integration of the joint posterior density is not easy to perform. So, the inference can be based on MCMC simulation methods such as the Gibbs sampler and Metropolis-Hastings algorithm, which can be used to draw samples, from which features of the marginal distributions of interest can be inferred. In this direction, we first obtain the full conditional distributions of each unknown quantity, which are given by

$$
\begin{aligned}
\pi(\boldsymbol{\beta} | y, \sigma, q, p, \lambda) \quad &\propto \quad \exp\left(q^{-1} \sum_{i=1}^{n} \frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma} \right) \\
&\times \quad \exp\left\{ -q^{-2} \sum_{i=1}^{n} \exp\left[q \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma} \right) \right] \right\} \\
&\times \quad \prod_{i=1}^{n} \left\{ 1 - p \left(1 - \gamma_1 \left\{ q^{-2}, q^{-2} \exp\left[q \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma} \right) \right] \right\} \right) \right\}^{-2} \\
&\times \quad \prod_{i=1}^{n} \left\{ 1 - \frac{\gamma_1 \left\{ q^{-2}, q^{-2} \exp\left[q \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma} \right) \right] \right\}}{1 - p \left(1 - \gamma_1 \left\{ q^{-2}, q^{-2} \exp\left[q \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma} \right) \right] \right\} \right)} \right\}^{\lambda - 1} \\
&\times \quad \pi(\boldsymbol{\beta}),
\end{aligned}
$$

$$
\begin{aligned}
\pi(\sigma|y,\boldsymbol{\beta},q,p,\lambda) &\propto \left(\frac{1}{\sigma}\right)^{n} \exp\left(q^{-1}\sum_{i=1}^{n}\frac{y_i-\mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right) \\
&\times \exp\left\{-q^{-2}\sum_{i=1}^{n}\exp\left[q\left(\frac{y_i-\mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right)\right]\right\} \\
&\times \prod_{i=1}^{n}\left\{1-p\left(1-\gamma_1\left\{q^{-2},q^{-2}\exp\left[q\left(\frac{y_i-\mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right)\right]\right\}\right)\right\}^{-2} \\
&\times \prod_{i=1}^{n}\left\{1-\frac{\gamma_1\left\{q^{-2},q^{-2}\exp\left[q\left(\frac{y_i-\mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right)\right]\right\}}{1-p\left(1-\gamma_1\left\{q^{-2},q^{-2}\exp\left[q\left(\frac{y_i-\mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right)\right]\right\}\right)}\right\}^{\lambda-1} \\
&\times \pi(\sigma),
\end{aligned}
$$

$$
\begin{aligned}
\pi(q|y,\boldsymbol{\beta},\sigma,p,\lambda) &\propto \left[\frac{|q|(q^{-2})^{q^{-2}}}{\Gamma(q^{-2})}\right]^{n} \exp\left(q^{-1}\sum_{i=1}^{n}\frac{y_i-\mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right) \\
&\times \exp\left\{-q^{-2}\sum_{i=1}^{n}\exp\left[q\left(\frac{y_i-\mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right)\right]\right\} \\
&\times \prod_{i=1}^{n}\left\{1-p\left(1-\gamma_1\left\{q^{-2},q^{-2}\exp\left[q\left(\frac{y_i-\mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right)\right]\right\}\right)\right\}^{-2} \\
&\times \prod_{i=1}^{n}\left\{1-\frac{\gamma_1\left\{q^{-2},q^{-2}\exp\left[q\left(\frac{y_i-\mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right)\right]\right\}}{1-p\left(1-\gamma_1\left\{q^{-2},q^{-2}\exp\left[q\left(\frac{y_i-\mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right)\right]\right\}\right)}\right\}^{\lambda-1} \\
&\times \pi(q),
\end{aligned}
$$

$$
\begin{aligned}
\pi(p|y,\boldsymbol{\beta},\sigma,q,\lambda) &\propto (1-p)^{n} \\
&\times \prod_{i=1}^{n}\left\{1-p\left(1-\gamma_1\left\{q^{-2},q^{-2}\exp\left[q\left(\frac{y_i-\mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right)\right]\right\}\right)\right\}^{-2} \\
&\times \prod_{i=1}^{n}\left\{1-\frac{\gamma_1\left\{q^{-2},q^{-2}\exp\left[q\left(\frac{y_i-\mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right)\right]\right\}}{1-p\left(1-\gamma_1\left\{q^{-2},q^{-2}\exp\left[q\left(\frac{y_i-\mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right)\right]\right\}\right)}\right\}^{\lambda-1} \\
&\times \pi(p)
\end{aligned}
$$

and

$$
\begin{aligned}
\pi(\lambda|y,\boldsymbol{\beta},\sigma,q,p) &\propto \lambda^{n}\prod_{i=1}^{n}\left\{1-\frac{\gamma_1\left\{q^{-2},q^{-2}\exp\left[q\left(\frac{y_i-\mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right)\right]\right\}}{1-p\left(1-\gamma_1\left\{q^{-2},q^{-2}\exp\left[q\left(\frac{y_i-\mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right)\right]\right\}\right)}\right\}^{\lambda} \\
&\times \pi(\lambda).
\end{aligned}
$$

Since the full conditional distributions for $\boldsymbol{\beta}$, σ, q, p and λ do not have a closed form, we require the use of the Metropolis-Hastings algorithm.

6. Application: permanence time in Japan

In this section, we provide an application of the proposed regression model to a real data set on permanence time in Japan of the Brazilian immigrants.

The data were obtained from an electronic survey (e-survey) according to the method developed by Babbie [5], which serves to obtain data on the characteristics, actions or opinions of groups using the Internet as a research tool. The survey was carried out in the first half of the year 2010, by means of a reserved site with limited access, by 246 respondents which filled out the questionnaires. Nevertheless only 147 were used in the analysis because some immigrants were from other nationalities. We considered the main variable of interest as permanence time in Japan in years, counted from the first arrival date until the research date. The following variables were associated with each immigrant ($i = 1, \cdots, 147$)

i) t_i : permanence time in Japan (years) and
ii) x_i : sex (0 = female, 1 = male).

We fit the LExGGG, LExWG, LGGG and Log-Weibull (or extreme value) regression models with regressor variable

$$y_i = \log(t_i) = \beta_0 + \beta_1 x_i + \sigma z_i,$$

where the errors $z_1, \ldots, z_{147}$ are independent random variables with density function (5.1).

For obtaining the MLEs of the model parameters, we use the NLMixed procedure in SAS. Table 1 lists the MLEs of the parameters for each regression model. The estimate of the parameter of the explanatory variable x is significant only for LExGGG and LExWG models at a significance level of 5%. This shows the advantage of new models in relation to their sub-models. We can note that β_1 is negative, which means that male Brazilian immigrants ($x = 1$) have an estimated probability of remaining in Japan less than women ($x = 0$) as shown in Figure 5. Furthermore, the permanence median time of the male Brazilian immigrants is $e^{-0.1012} \times 100 = 90.38\%$ of the median time of women for LExGGG model and $e^{-0.1881} \times 100 = 82.85\%$ for LExWG model.

Table 1. MLEs of the parameters of regression models for the permanence time in Japan and the corresponding SEs in parentheses.

Model	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\sigma}$	$\hat{q}$	$\hat{p}$	$\hat{\lambda}$
LExGGG	2.6915	−0.1012	0.2152	2.3873	0.9149	0.2004
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0275)	(0.0198)
LExWG	2.1928	−0.1881	0.2973	1	0.9644	0.1053
	(0.0020)	(0.0001)	(0.0007)		(0.0163)	(0.0098)
LGGG	4.0046	−0.0444	1.5766	0.0841	0.6197	1
	(0.3712)	(0.2953)	(0.4360)	(0.0111)	(0.0633)	
Log-Weibull	2.6870	−0.1482	0.4559	1	1	1
	(0.0422)	(0.1055)	(0.0325)			

A summary of the values of the Akaike Information Criterion (AIC), Consistent Akaike Information Criterion (CAIC) and Bayesian Information Criterion (BIC)

(a) (b)

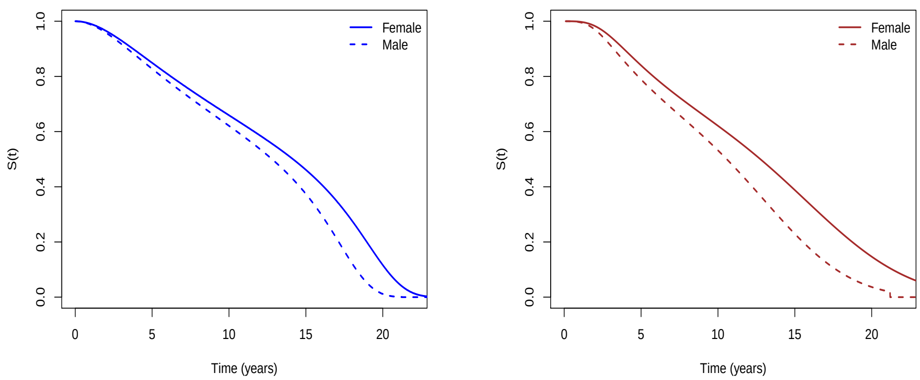

Figure 5. Estimated survival function for the (a) LExGGG and (b) LExWG regression models stratified by covariate sex.

to compare the LExGGG, LExWG, LGGG and Log-Weibull regression models is given in Table 2. These results indicate that the LExGGG regression model has the lowest AIC, BIC and CAIC values among those values of the fitted models, and therefore it could be chosen as the best model.

Table 2. AIC, CAIC and BIC statistics for comparing the LExGGG, LExWG, LGGG and Log-Weibull.

Model	AIC	CAIC	BIC
LExGGG	203.2	203.8	221.2
LExWG	235.0	235.4	249.9
LGGG	458.3	458.7	473.3
Log-Weibull	262.1	262.3	271.1

A comparison of the proposed model with some of their sub-models using LR statistics is performed in Table 3. The figures in this table, specially the p-values, indicate that the LExGGG regression model yields a better fit to these data than their sub-models.

Table 3. LR statistics for the permanence time in Japan. .

Model	Hypotheses	LR Statistics	p-value
LExGGG vs LExWG	$H_0 : q = 1$ vs $H_1 : H_0$ is false	33.8	< 0.01
LExGGG vs LGGG	$H_0 : \lambda = 1$ vs $H_1 : H_0$ is false	257.1	< 0.01
LExGGG vs Log-Weibull	$H_0 : \lambda = 1 - p = q = 1$ vs $H_1 : H_0$ is false	64.9	< 0.01

In the Bayesian analysis following independent priors were considered to perform the Metropolis-Hastings algorithm: $\lambda \sim \Gamma(0.01, 0.01)$, $\sigma \sim \Gamma(0.01, 0.01)$, $p \sim \mathrm{Be}(0.5; 0.5)$, $q \sim N(0, 10)$ and $\boldsymbol{\beta}_s \sim N(0, 10)$, so that we have a vague prior distribution. Considering these prior density functions, we generated two parallel independent runs of the Metropolis-Hastings with size $200,000$ for each parameter, disregarding the first $20,000$ iterations to eliminate the effect of the initial values and, to avoid correlation problems, we considered a spacing of size 10, obtaining a sample of size $18,000$ from each chain. To monitor the convergence of the Metropolis-Hastings, we performed the methods suggested by Cowles and Carlin [16]. To monitor the convergence of the Metropolis-Hastings, we used the between and within sequence information, following the approach developed in Gelman and Rubin [24] to obtain the potential scale reduction, $\widehat{R}$. In all cases, these values were close to the value 1, indicating the convergence of the chain. The approximate posterior marginal density functions for the parameters are presented in Figure 6. In Table 4, we report posterior summaries for the parameters of the LExGGG model. We note that the values for the means a posteriori (Table 4) are quite close (as expected) to the MLEs obtained for the LExGGG model given in Table 1. SD represents the standard deviation from the posterior distributions of the parameters and HPD represents the 95% highest posterior density (HPD) intervals.

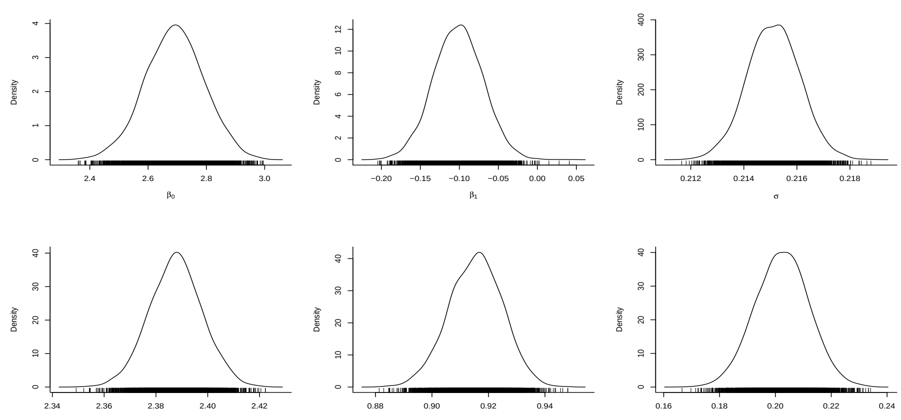

Figure 6. Approximate posterior marginal densities for the parameters from the LExGGG modelfor the permanence time in Japan.

Table 4. Posterior summaries for the parameters of the LExGGG model for the permanence time in Japan.

Parameter	Mean	SD	HPD (95%)	$\hat{R}$
β_0	2.6909	0.1005	$(2.5056; 2.9017)$	0.9999
β_1	-0.1012	0.0315	$(-0.1624; -0.0393)$	1.0002
σ	0.2151	0.0010	$(0.2132; 0.2170)$	1.0007
q	2.3877	0.0100	$(2.3680; 2.4072)$	1.0001
p	0.9156	0.0095	$(0.8969; 0.9342)$	0.9999
λ	0.2024	0.0097	$(0.1832; 0.2213)$	1.0003

7. Concluding remarks:

We studied a five-parameter model called the extended generalized gamma geometric (ExGGG) distribution that can model four standard types of the hazard rate function. We derive expansions for their density function, moments, moment generating function, mean deviations, reliability and order statistics. This extension provides more flexibility to analyze skewed data. Further, we also introduce the log-extended generalized gamma geometric (LExGGG) distribution and obtained expansions for their moments. Based on this new distribution, we define the LExGGG regression model which is very suitable for modeling censored and uncensored lifetime data. We discussed the parameter estimation by maximum likelihood and Bayesian approach. The usefulness of the regression model is showed through the analysis of a real data set. The proposed regression model serves as a good alternative for lifetime data analysis and can be more flexible than the generalized gamma geometric and Weibull models.

Appendix A. Series expansion for the incomplete gamma ratio function

Pascoa *et al.* [46] developed a series expansion for the incomplete gamma ratio function given by

$$\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^{\tau}\right]=\frac{1}{\Gamma(k)}\left(\frac{x}{\alpha}\right)^{\tau k}\sum_{d=0}^{\infty}\left[-\left(\frac{x}{\alpha}\right)^{\tau}\right]^{d}\frac{1}{(k+d)d!}.$$

They used an equation in Section 0.314 of Gradshteyn and Ryzhik [26] for a power series raised to a positive integer m

$$\left[\sum_{d=0}^{\infty}a_d\left(\frac{x}{\alpha}\right)^{\tau d}\right]^{m}=\sum_{d=0}^{\infty}c_{m,d}\left(\frac{x}{\alpha}\right)^{\tau d},\tag{A.1}$$

whose coefficients $c_{m,d}$ (for $d=1,2,\cdots$) are calculated from the recurrence equation

$$c_{m,d}=(da_0)^{-1}\sum_{r=1}^{d}(mr-d+r)\,a_r\,c_{m,d-r}\tag{A.2}$$

and $c_{m,0}=a_0^m$, where $a_d=(-1)^d\,[(k+d)d!]^{-1}$. The coefficient $c_{m,d}$ can be obtained from $c_{m,0},\ldots,c_{m,d-1}$. It can also be written explicitly as functions of the quantities $a_0,\ldots,a_d$ using algebraic software such as Maple and Mathematica, although it is not necessary for programming numerically our expansions. Here, $c_{m,0}=k^{-m}$, $c_{m,1}=-m[(k+1)k^{m-1}]^{-1}$, $c_{m,2}=m[2(k+2)k^{m-1}]^{-1}+m(m-1)[2(k+1)^2k^{m-2}]^{-1}$, etc. Equation (A.1) yields

$$\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^{\tau}\right]^{m}=\frac{1}{\Gamma(k)^m}\sum_{d=0}^{\infty}c_{m,d}\left(\frac{x}{\alpha}\right)^{\tau(km+d)},\tag{A.3}$$

where the coefficients $c_{m,d}$ are calculated from (A.2).

Appendix B. Observed information matrix $J(\boldsymbol{\theta})$

Here, we give the elements of the observed information matrix $J(\boldsymbol{\theta})$ for the parameters $\boldsymbol{\theta} = (\beta_j, \sigma, q, p, \lambda)$. After some algebraic manipulations, we obtain

$$
\begin{aligned}
J_{\beta_j \beta_s} = {} & \frac{1}{\sigma^2} \sum_{i=1}^{n} x_{ij} x_{is} \exp(q z_i) \\
& + \frac{2p}{\sigma^2 \Gamma(q^{-2})} \sum_{i=1}^{n} \frac{x_{ij} x_{is} w_i^{q^{-2}-1} \exp(q z_i - w_i)(q^{-2} + w_i - 2)}{1 - p[1 - \gamma_1(q^{-2}, w_i)]} \\
& - 2\left[\frac{p}{q\sigma\Gamma(q^{-2})}\right]^2 \sum_{i=1}^{n} x_{ij} x_{is} \left\{ \frac{w_i^{q^{-2}-1} \exp(q z_i - w_i)}{1 - p[1 - \gamma_1(q^{-2}, w_i)]} \right\}^2 \\
& - \frac{(\lambda - 1)(1 - p)}{q\sigma\Gamma(q^{-2})} \sum_{i=1}^{n} \frac{x_{ij} w_i^{q^{-2}-1} \exp(q z_i - w_i)}{\{1 - p[1 - \gamma_1(q^{-2}, w_i)]\}^2},
\end{aligned}
$$

$$
\begin{aligned}
J_{\beta_j \sigma} = {} & -\frac{1}{q\sigma^2} \sum_{i=1}^{n} x_{ij} + \frac{1}{q\sigma^2} \sum_{i=1}^{n} x_{ij} \exp(q z_i)(1 + q z_i) \\
& + \frac{2p}{q^2 \sigma^2 \Gamma(q^{-2})} \sum_{i=1}^{n} \frac{x_{ij} z_i w_i^{q^{-2}-1} \exp(q z_i - w_i) \left(1 + w_i^{q^{-2}-1} (q^2 - 1)\right)}{1 - p[1 - \gamma_1(q^{-2}, w_i)]} \\
& + 2\left[\frac{p}{q\sigma\Gamma(q^{-2})}\right]^2 \sum_{i=1}^{n} x_{ij} z_i \left\{ \frac{w_i^{q^{-2}-1} \exp(q z_i - w_i)}{1 - p[1 - \gamma_1(q^{-2}, w_i)]} \right\}^2 \\
& - \frac{(\lambda - 1)(1 - p)}{q\sigma\Gamma(q^{-2})} \sum_{i=1}^{n} \frac{z_i w_i^{q^{-2}-1} \exp(q z_i - w_i)}{\{1 - p[1 - \gamma_1(q^{-2}, w_i)]\}^2},
\end{aligned}
$$

$$
\begin{aligned}
J_{\beta_j p} = {} & -\frac{2}{q\sigma\Gamma(q^{-2})} \sum_{i=1}^{n} \frac{x_{ij} w_i^{q^{-2}-1} \exp(q z_i - w_i)}{\{1 - p[1 - \gamma_1(q^{-2}, w_i)]\}^2} \\
& - \frac{(\lambda - 1)(1 + p^2)}{q\sigma\Gamma(q^{-2})(1 - p)^2} \sum_{i=1}^{n} \frac{x_{ij} w_i^{q^{-2}-1} \exp(q z_i - w_i)}{\{1 - p[1 - \gamma_1(q^{-2}, w_i)]\}^2},
\end{aligned}
$$

$$
J_{\beta_j \lambda} = -\frac{(1 - p)}{q\sigma\Gamma(q^{-2})} \sum_{i=1}^{n} \frac{x_{ij} w_i^{q^{-2}-1} \exp(q z_i - w_i)}{\{1 - p[1 - \gamma_1(q^{-2}, w_i)]\}^2},
$$

$$
\begin{aligned}
J_{\sigma\sigma} = {} & -\frac{2}{q\sigma^2} \sum_{i=1}^{n} z_i + \frac{1}{q\sigma^2} \sum_{i=1}^{n} z_i \exp(q z_i)(2 + q z_i) \\
& - \frac{2p + (\lambda - 1)(1 - p)}{q\sigma^2 \Gamma(q^{-2})} \sum_{i=1}^{n} \frac{z_i w_i^{q^{-2}-1} \exp(q z_i - w_i) \left(q z_i w_i - q^{-1} z_i - 2\right)}{1 - p[1 - \gamma_1(q^{-2}, w_i)]} \\
& - \frac{2p^2 + (\lambda - 1)(1 - p)}{q^2 \sigma^2 \Gamma(q^{-2})^2} \sum_{i=1}^{n} \left\{ \frac{z_i w_i^{q^{-2}-1} \exp(q z_i - w_i)}{1 - p[1 - \gamma_1(q^{-2}, w_i)]} \right\}^2,
\end{aligned}
$$

$$
\begin{aligned}
J_{\sigma p} = & -\frac{2}{q\sigma\Gamma\left(q^{-2}\right)} \sum_{i=1}^{n} \frac{z_i w_i^{q^{-2}-1} \exp(q z_i - w_i)}{\left\{1 - p[1 - \gamma_1(q^{-2}, w_i)]\right\}^2} \\
& - \frac{(\lambda - 1)\left(1 + p^2\right)}{q\sigma\Gamma\left(q^{-2}\right)(1-p)^2} \sum_{i=1}^{n} \frac{z_i w_i^{q^{-2}-1} \exp(q z_i - w_i)}{\left\{1 - p[1 - \gamma_1(q^{-2}, w_i)]\right\}^2},
\end{aligned}
$$

$$
J_{\sigma\lambda} = -\frac{1-p}{q\sigma\Gamma\left(q^{-2}\right)} \sum_{i=1}^{n} \frac{z_i w_i^{q^{-2}-1} \exp(q z_i - w_i)}{\left\{1 - p[1 - \gamma_1(q^{-2}, w_i)]\right\}^2},
$$

$$
\begin{aligned}
J_{pp} = & \frac{n}{(1-p)^2} + 2\sum_{i=1}^{n} \left\{ \frac{1 - \gamma_1(q^{-2}, w_i)}{1 - p[1 - \gamma_1(q^{-2}, w_i)]} \right\}^2 \\
& + \frac{\lambda - 1}{1-p} \sum_{i=1}^{n} \frac{\gamma_1(q^{-2}, w_i)\left\{1 + (1-2p)\left[1 - \gamma_1(q^{-2}, w_i)\right]\right\}}{1 - p[1 - \gamma_1(q^{-2}, w_i)]},
\end{aligned}
$$

$$
J_{p\lambda} = \frac{1}{1-p} \sum_{i=1}^{n} \frac{\gamma_1(q^{-2}, w_i)}{1 - p[1 - \gamma_1(q^{-2}, w_i)]}
$$

and

$$
J_{\lambda\lambda} = \frac{n}{\lambda^2},
$$

where $w_i = q^{-2} \exp(q z_i)$ and $z_i = (y_i - \mathbf{x}_i^T \boldsymbol{\beta})/\sigma$.

References

[1] Aarts, R.M. *Lauricella functions.* From MathWorld - A Wolfram Web Resource, created by E. W. Weisstein. Available at www.mathworld.wolfram.com/LauricellaFunctions.html, 2000.

[2] Adamidis, K. and Loukas, S. *A Lifetime distribution with decreasing failure rate*, Statistics & Probability Letters **39** (1), 35-42, 1998.

[3] Al-Zahrani, B., Marinho, P.R.D., Fattah, A.A., Ahmed, A.H.N., and Cordeiro, G.M. *The (P-A-L) extended Weibull distribution: a new generalization of the Weibull distribution*, Hacettepe Journal of Mathematics and Statistics **45**, 1-19, 2015.

[4] Almpanidis, G. and Kotropoulos, C. *Phonemic segmentation using the generalized gamma distribution and small sample Bayesian information criterion*, Speech Communication **50** (1), 38-55, 2008.

[5] Babbie, E. *Métodos de pesquisas de Survey.* (Ed. UFMG, Belo Horizonte, 1999).

[6] Barkauskas, D.A., Kronewitter, S.R., Lebrilla, C.B. and Rocke, D.M. *Analysis of MALDI FT-ICR mass spectrometry data: A time series approach*, Analytica Chimica Acta **648** (2), 207-214, 2009.

[7] Barreto-Souza, W., de Morais, A.L. and Cordeiro, G.M. *The Weibull-geometric distribution*, Journal of Statistical Computation and Simulation **81** (5), 645-657, 2011.

[8] Bortolini, J., Pascoa, M.A.R., de Lima, R.R. and Oliveira, A.C.S. *The Extended Generalized Gamma Geometric Distribution*, Hacettepe Journal of Mathematics and Statistics **submitted**, 2015.

[9] Carrasco, J.M.F., Ortega, E.M.M. and Cordeiro, G.M. *A generalized modified Weibull distribution for lifetime modeling*, Computational Statistics and Data Analysis **53** (2), 450-462, 2008.

[10] Cordeiro, G.M. and de CASTRO, M. *A new family of generalized distributions*, Journal of Statistical Computation and Simulation **81** (7), 883-898, 2011.

[11] Cordeiro, G.M., Ortega, E.M.M. and Silva, G.O. *The exponentiated generalized gamma distribution with application to lifetime data*, Journal of Statistical Computation and Simulation **81** (7), 827-842, 2011.

[12] Cordeiro, G.M., Ortega, E.M.M. and da Cunha, D.C.C. *The exponentiated generalized class of distributions*, Journal of Data Science **11**, 1-27, 2013.

[13] Cordeiro, G.M., Silva, G.O. and Ortega, E.M.M. *The beta-Weibull geometric distribution*, Statistics **47** (4), 817-834, 2013.

[14] Cordeiro, G.M., Ortega, E.M.M. and Silva, G.O. *The Kumaraswamy modified Weibull distribution: theory and applications*, Journal of Statistical Computation and Simulation **84** (7), 1387-1411, 2014.

[15] Cordeiro, G.M., Saboor, A., Khan, M.N., Ozel, G. and Pascoa, M.A.R. *The Kumaraswamy exponential-Weibull distribution: theory and applications*, Hacettepe Journal of Mathematics and Statistics **45**, 1-30, 2015.

[16] Cowles, M.K. and Carlin, B.P. *Markov chain Monte Carlo convergence diagnostics: a comparative review*, Journal of the American Statistical Association **91**, 883-904, 1996.

[17] Cox, C., Chu, M.F and Muñoz, A. *Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution* Statistics in Medicine **26**, 4352-4374, 2007.

[18] Cox, C. *The generalized F distribution: an umbrella for parametric survival analysis*, Statistics in Medicine **27** (21), 4301-4312, 2008.

[19] Cox, C. and Matheson, M. *A comparison of the generalized gamma and exponentiated Weibull distributions*, Statistics in Medicine **33** (33), 3772-3780, 2014.

[20] Cox, D.R. and Oakes, D. *Analysis of survival data.* (Chapman and Hall, New York, 1984).

[21] da Cruz, J.N., Ortega, E.M.M. and Cordeiro, G.M. *The log-odd log-logistic Weibull regression model: modelling, estimation, influence diagnostics and residual analysis*, Journal of Statistical Computation and Simulation, 2015.

[22] Erto, P. *Genesis, properties and identification of the inverse Weibull lifetime lodel*, Statistica Applicata **1**, 117-128, 1989.

[23] Exton, H. *Handbook of hypergeometric integrals: theory, applications, tables, computer programs.* (Halsted Press, New York, 1978).

[24] Gelman, A. and Rubin, D.B. *Inference from iterative simulation using multiple sequences (with discussion)*. Statistical Science **7**, 457-472, 1992.

[25] Gomes, O., Combes, C and Dussauchoy, A. *Parameter estimation of the generalized gamma distribution*, Mathematics and Computers in Simulation **79** (4), 955-963, 2008.

[26] Gradshteyn, I.S. and Ryzhik, I.M. *Table of integrals, series, and products.* (Academic Press, New York, 2000).

[27] Gupta, R.D. and Kundu, D. *Generalized exponential distribution*, Australian and New Zealand Journal of Statistics **41** (2), 173–188, 1999.

[28] Gupta, R.D. and Kundu, D. *Exponetiated exponential family: an alternative to gamma and Weibull distributions*, Biometrical Journal **43**, 117-130, 2001a.

[29] Gupta, R. D. and Kundu, D. *Generalized exponential distribution: different method of estimations*, Journal of Statistical Computation and Simulation, **69**, 315-337, 2001b.

[30] Gupta, R.D. and Kundu, D. *Generalized exponential distribution: existing results and some recent developments*, Journal of Statistical Planning and Inference **137** (11), 3537-3547, 2007.

[31] Gusmão, F.R.S., Ortega, E.M.M. and Cordeiro, G.M. *The generalized inverse Weibull distribution*, Statistical Papers **52** (3), 591-619, 2009.

[32] Kalbfleisch, J.D. and Prentice, R.L. *The statistical analysis of failure time data.* (John Wiley, New York, 2002).

[33] Lawless, J.F. *Statistical Models and Methods for Lifetime Data.* (John Wiley, New York, 2003).

[34] Malhotra, J., Sharma, A.K. and Kaler, R.S. *On the performance analysis of wireless receiver using generalized-gamma fading model*, Annals of Telecommunications **64** (1), 147-153, 2009.

[35] Mudholkar, G.S., Srivastava, D.K. and Freimer, M. *The exponentiated Weibull family: a reanalysis of the bus-motor-failure data*, Technometrics **37** (4), 436-445, 1995.

[36] Nadarajah, S. and Gupta, A.K. *A generalized gamma distribution with application to drought data*, Mathematics and Computers in Simulation **74** (1), 1-7 2007.

[37] Nadarajah, S. *On the use of the generalised gamma distribution*, International Journal of Electronic **95** (10), 1029-1032, 2008a.

[38] Nadarajah, S. *Explicit expressions for moments of order statistics*, Statistics and Probability Letters **78** (2), 196-205, 2008.

[39] Ortega, E.M.M., Bolfarine, H. and Paula, G.A. *Influence diagnostics in generalized log-gamma regression models*, Computational Statistics and Data Analysis **42** (1-2), 165-186, 2003.

[40] Ortega, E.M.M., Paula, G.A. and Bolfarine, H. *Deviance residuals in generalized log-gamma regression models with censored observations*, Journal of Statistical Computation and Simulation **78** (8), 747-764, 2008.

[41] Ortega, E.M.M., Cancho, V.G. and Paula, G.A. *Generalized log-gamma regression models with cure fraction*, Lifetime Data Analysis **15** (1), 79-106, 2009.

[42] Ortega, E.M.M., Cordeiro, G.M. and Pascoa, M.A.R. *The generalized gamma geometric distribution*, Journal of Statistical Theory and Applications **10**, 433-454, 2011.

[43] Ortega, E.M.M., Cordeiro, G.M., Pascoa, M.A.R. and Couto, E.V. *The log-exponentiated generalized gamma regression model for censored data*, Journal of Statistical Computation and Simulation **82**, 1169-1189, 2012.

[44] Ortega, E.M.M., Barriga, G.D.C., Hashimoto, E.M., Cancho, V.G. and Cordeiro, G.M. *A New Class of Survival Regression Models with Cure Fraction*, Journal of Data Science **12**, 107-136, 2014.

[45] Pascoa, M.A.R., Ortega, E.M.M. and Cordeiro, G.M. *The Kumaraswamy generalized gamma distribution with application in survival analysis.* Statistical Methodology **8** (5), 411-433, 2011.

[46] Pascoa, M.A.R., de Paiva, C.M.M., Cordeiro, G.M. and Ortega, E.M.M. *The Log-Kumaraswamy generalized gamma regression model with application to chemical dependency data.* Journal of Data Science **11**, 781-818, 2013.

[47] Prudnikov, A.P., Brychkov, Y.A. and Marichev, O.I. *Integrals and Series.* (Gordon and Breach Science Publishers, Amsterdam, 1986).

[48] Silva, R.B., Barreto-Souza, W. and Cordeiro, G.M. *A new distribution with decreasing, increasing and upside-down bathtub failure rate*, Computational Statistics and Data Analysis **54** (4), 935-944, 2010.

[49] Silva, G.O., Ortega, E.M.M. and Cordeiro, G.M. *The beta modified Weibull distribution*, Lifetime Data Analysis **16**, 409-430, 2010.

[50] Stacy, E.W. *A generalization of the gamma distribution*, Annals of Mathematical Statistics **33** (3), 1187-1192, 1962.

[51] Stacy, E.W. and Mihram, G.A. *Parameter estimation for a generalized gamma distribution*, Technometris **7** (3), 349-358, 1965.

[52] Trott, M. *The mathematica guidebook for symbolics.* (Springer, New York, 2006).

[53] Wright, E.M. *The asymptotic expansion of the generalized Bessel function.* Proceedings of the London Mathematical Society **38**, 257-270, 1935.

[54] Xie, X. and Liu, X. *Analytical three-moment autoconversion parameterization based on generalized gamma distribution*, Journal of Geophysical Research **114** (D17), 1984-2012, 2009.

CONSIDERAÇÕES GERAIS

Propor novas distribuições de probabilidade é útil em aplicações da Estatística, em especial aquelas que envolvem dados complexos, tais como observações censuradas, comportamento bimodal e assimétrico. Com essa motivação, foram propostas três distribuições de probabilidade e um modelo de regressão paramétrico.

A primeira distribuição apresentada é a gama generalizada geométrica estendida (ExGGG, do inglês *extended generalized gamma geometric*) com função risco crescente, decrescente, unimodal, em forma de banheira e outras formas não-padronizadas. A grande vantagem dessa distribuição em relação aos seus submodelos é o fato de possuir densidade bimodal, o que não ocorre com nenhum deles. Algumas propriedades da nova distribuição foram discutidas, bem como a expansão da função densidade, momentos, função geradora de momentos, desvios médios, confiabilidade e estatísticas de ordem. A estimação dos parâmetros foi abordada pelo método da máxima verossimilhança e *bayesiano*. A utilidade da nova distribuição é ilustrada na análise de um conjunto de dados reais, mostrando-se um ajuste mais adequado do que os seus sub-modelos.

O segundo modelo é uma extensão da distribuição ExGGG, em que um dos parâmetros de forma pode assumir qualquer valor real ao invés de apenas positivo, oferecendo ainda mais flexibilidade à distribuição já proposta.

E ainda, obteve-se a distribuição do logaritmo da ExGGG e expansões para os seus momentos ordinários. Com base nesta terceira distribuição, definiu-se o modelo de regressão log-ExGGG. A estimação dos parâmetros do modelo de regressão foi discutida pelo método da máxima verossimilhança e pela abordagem *bayesiana*. A aplicabilidade desse novo modelo também foi verificada na modelagem de um conjunto de dados reais.

Ao final, depreende-se que as distribuições propostas e o novo modelo de regressão podem ser bastante úteis na análise de dados de sobrevivência, em especial os assimétricos e bimodais.

Como perspectiva de pesquisas futuras, pretende-se implementar as principais funções das distribuições propostas em programa R. Além disso, almeja-se desenvolver uma análise de diagnóstico para o novo modelo de regressão.