Revista Colombiana de Estadística - Applied Statistics
July 2024, volume 47, issue 2, pp. 453–473
http://doi.org/10.15446/rce.v47n2.112328

The Flexible Generalized Gamma Distribution With Applications to COVID-19 Data

La distribución gamma generalizada flexible con aplicaciones a los datos COVID-19

Alexsandro A. Ferreira[a], Gauss M. Cordeiro[b]

Department of Statistics, Centro de Ciências Exatas e da Natureza, Universidade Federal de Pernambuco, Recife, Brazil

Abstract

The article presents the flexible generalized gamma distribution, whose density function can be expressed as an infinite linear combination of generalized gamma densities. Some of its statistical properties are reported, and maximum likelihood estimation is also discussed. A regression model with two systematic components is constructed for censored data. Three applications to real COVID-19 data reveal that the new model provides adequate fit and outperforms some competing models.

***Key words*:** Acceptance-rejection; COVID-19; Flexible Weibull distribution; Regression model.

Resumen

El artículo presenta la distribución gamma generalizada flexible, cuya función de densidad puede expresarse como una combinación lineal infinita de densidades gamma generalizadas. Se describen algunas de sus propiedades estadísticas y se analiza la estimación por máxima verosimilitud. Se construye un modelo de regresión con dos componentes sistemáticos para datos censurados. Tres aplicaciones a datos reales de COVID-19 revelan que el nuevo modelo proporciona y supera a algunos modelos de la competencia.

***Palabras clave*:** Aceptación-rechazo; COVID-19; Distribución Weibull flexible; Modelo de regresión.

[a]Ph.D(c). E-mail: alexsandro.ferreira.aaf@gmail.com
[b]PhD. E-mail: Gauss@de.ufpe.br

1. Introduction

In recent years, there has been a growing interest in the development of flexible statistical distributions that can accurately model data sets across a variety of fields. The generalized gamma (GG) distribution (Stacy, 1962) has become a popular tool for analyzing asymmetric data due to its versatility in handling different types of hazard rate functions (hrfs). In addition, the GG distribution includes well-known submodels, such as the Weibull, gamma, Rayleigh, exponential, and Maxwell.

Recently, different generalizations and modifications of the GG distribution to model lifetimes have been proposed in the literature. Some examples include the four-parameter generalized gamma distribution geometric (Ortega et al., 2011), which extends a number of well-known special lifetime models such as the Weibull geometric, Rayleigh geometric, and exponential geometric, among others. The exponentiated generalized gamma distribution (Cordeiro et al., 2011) includes an extra shape parameter and has shown excellent fits to lifetime data. The beta generalized gamma distribution (Cordeiro et al., 2013) and Kumaraswamy generalized gamma distributions (Pascoa et al., 2011) both with two extra parameters, are also two important extensions.

Tahir et al. (2022) introduced the flexible generalized family (FGF), which has significant advantages over other generators. For example, it was not developed from any known parent model. Additionally, FGF does not involve the inclusion of extra parameters and allows the choice of any baseline model. The special models generated from FGF are free of unidentifiability problems, allowing both exponentiated and inverted models to be adopted. Another major advantage is that new special models based on FGF are able to compete with existing parent models or other alternative models. They can also produce flexible shapes for densities and hrfs, in some cases with more flexibility than the baseline models. Thus, the FGF emerges as an advantageous and parsimonious generator for building new distributions.

Motivated by the importance of the GG distribution and the FGF generator, this article combines the two, thus proposing the flexible generalized gamma (FGG) distribution. The new density can be expressed as an infinite linear combination of GG densities, thus allowing for easy derivation of some important properties. Maximum likelihood estimation is also discussed. Another contribution of this paper is the construction of a regression model based on the FGG distribution to handle censored observations.

The applicability of the FGG distribution is illustrated with three real COVID-19 data sets. Therefore, this paper provides a valuable contribution to the field by introducing greater flexibility to the GG distribution. The comprehensive analytical results, regression modeling, simulations, and real data analyses prove its wide potential.

This paper is structured as follows. Section 2 defines the FGG distribution. Section 3 establishes its properties, and Section 4 constructs a new regression model. The performance of the estimators is evaluated via simulations in Section 5.

Applications to COVID-19 data are illustrated in Section 6, and some conclusions are offered in Section 7.

2. The FGG Distribution

Let $G(x) = G(x; \boldsymbol{\xi})$ be the cumulative distribution function (cdf) for any baseline distribution, where $\boldsymbol{\xi}$ is the associated parameter vector. The cdf of the FGF without extra parameters (for $x \in \mathbb{R}$)

$$F(x; \boldsymbol{\xi}) = 1 - [1 - G(x)]^{G(x)}. \tag{1}$$

The probability density function (pdf) of the GG distribution with three parameters $\alpha, \tau, p > 0$ has the form (for $x > 0$)

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

where $\Gamma(\cdot)$ is the complete gamma function. The cdf corresponding to (2) is

$$G(x; \alpha, \tau, p) = \gamma_1\left(p, \left[\frac{x}{\alpha}\right]^{\tau}\right), \tag{3}$$

where $\gamma_1 = \gamma(p, x)/\Gamma(p)$, $\gamma(\cdot, \cdot)$ is the lower incomplete gamma function. Thus, substituting (3) in Equation (1), the cdf of the FGG distribution is

$$F(x; \alpha, \tau, p) = 1 - \left\{1 - \gamma_1\left(p, \left[\frac{x}{\alpha}\right]^{\tau}\right)\right\}^{\gamma_1\left(p, \left[\frac{x}{\alpha}\right]^{\tau}\right)}. \tag{4}$$

Let $X \sim \text{FGG}(\alpha, \tau, p)$ have the cdf (4). The pdf of X reduces to

$$\begin{aligned} f(x; \alpha, \tau, p) = {} & \frac{\tau}{\alpha\,\Gamma(p)} \left(\frac{x}{\alpha}\right)^{\tau p - 1} \exp\left[-\left(\frac{x}{\alpha}\right)^{\tau}\right] \left\{1 - \gamma_1\left(p, \left[\frac{x}{\alpha}\right]^{\tau}\right)\right\}^{\gamma_1\left(p, \left[\frac{x}{\alpha}\right]^{\tau}\right)} \\ & \times \left\{\frac{\gamma_1\left(p, \left[\frac{x}{\alpha}\right]^{\tau}\right)}{1 - \gamma_1\left(p, \left[\frac{x}{\alpha}\right]^{\tau}\right)} - \log\left[1 - \gamma_1\left(p, \left[\frac{x}{\alpha}\right]^{\tau}\right)\right]\right\}. \end{aligned} \tag{5}$$

The hrf of X becomes

$$\begin{aligned} h(x; \alpha, \tau, p) = {} & \frac{\tau}{\alpha\,\Gamma(p)} \left(\frac{x}{\alpha}\right)^{\tau p - 1} \exp\left[-\left(\frac{x}{\alpha}\right)^{\tau}\right] \left\{\frac{\gamma_1\left(p, \left[\frac{x}{\alpha}\right]^{\tau}\right)}{1 - \gamma_1\left(p, \left[\frac{x}{\alpha}\right]^{\tau}\right)}\right. \\ & \left. - \log\left[1 - \gamma_1\left(p, \left[\frac{x}{\alpha}\right]^{\tau}\right)\right]\right\}. \end{aligned}$$

Figure 1(a) reports the density of X for some parameter values, which reveals the flexibility of the new model in dealing with right-skewed, left-skewed, and symmetric data. Figure 1(b) shows that the hrf can model increasing, decreasing, unimodal, and bathtub shapes. Table 1 lists some important sub-models of the FGG distribution.

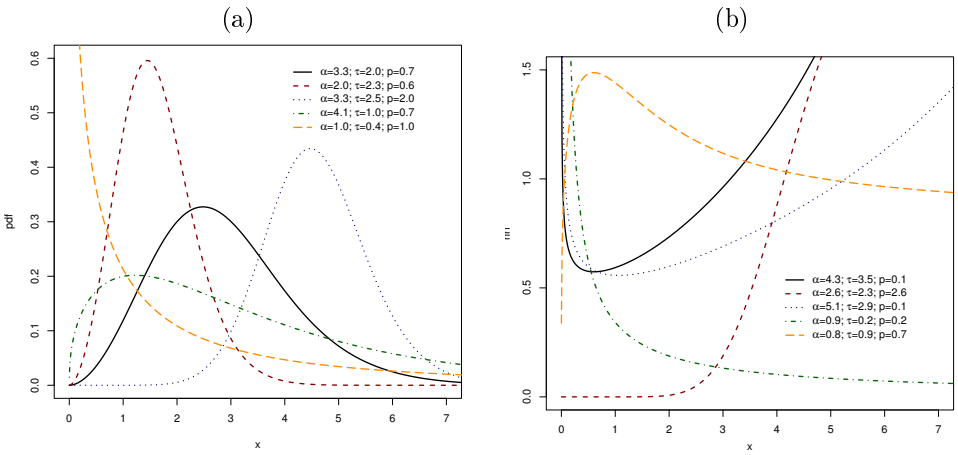

FIGURE 1: Density and hrf of X.

TABLE 1: Special models of the FGG distribution.

Distribution	α	τ	p
Flexible gamma	α	1	p
Flexible chi-square	2	1	$n/2$
Flexible exponential	α	1	1
Flexible Weibull	α	τ	1
Flexible Rayleigh	α	2	1
Flexible Maxwell	α	2	3/2
Flexible half normal	$\sigma\sqrt{2}$	2	1/2
Flexible log normal	α	τ	$\to\infty$
Flexible generalized half normal	$2^{1/(2\gamma)}\sigma$	2γ	1/2
Flexible Nakagami distribution	$\sqrt{w/\mu}$	2	μ

3. Properties

3.1. Useful Expansion

Following the findings of Tahir et al. (2022), the FGG density (5) can be written as

$$f(x;\alpha,\tau,p)=\sum_{k=1}^{\infty}\eta_{k+1}\,\frac{(k+1)\tau}{\alpha\,\Gamma(p)}\left(\frac{x}{\alpha}\right)^{\tau p-1}\exp\left[-\left(\frac{x}{\alpha}\right)^{\tau}\right]\gamma_1\left(p,\left[\frac{x}{\alpha}\right]^{\tau}\right)^{k}, \qquad (6)$$

where $\eta_2=2$, $\eta_3=3/2$, $\eta_4=-2/3$, $\eta_5=-5/4$, $\eta_6=-11/20$, etc. These coefficients are determined by differentiating the Taylor series of Equation (1).

Using the result (A4) (given in Appendix A) in Equation (6) leads to

$$f(x;\alpha,\tau,p)=\sum_{k=1}^{\infty}\eta_{k+1}\,\frac{(k+1)\tau}{\alpha\,\Gamma(p)}\exp\left[-\left(\frac{x}{\alpha}\right)^{\tau}\right]\sum_{i=0}^{\infty}\frac{b_{k,i}}{\Gamma(p)^k}\left(\frac{x}{\alpha}\right)^{[\,i+(k+1)p\,]\tau-1}. \qquad (7)$$

From Equation (7), one can write

$$f(x;\alpha,\tau,p)=\sum_{k=1}^{\infty}\sum_{i=0}^{\infty}\varphi_{k,i}\,g_{\alpha,\tau,p^*}(x), \tag{8}$$

where

$$\varphi_{k,i}=\frac{(k+1)\,\eta_{k+1}\,b_{k,i}\,\Gamma(p^*)}{\Gamma(p)^{k+1}},$$

$p^*=i+(k+1)p$, and $g_{\alpha,\tau,p^*}(x)$ is the $\mathrm{GG}(\alpha,\tau,p^*)$ density. According to Equation (8), the FGG density is a linear combination of GG densities. Thus, the properties of the new distribution are easily derived by knowing.

3.2. Moments and Generating Function

The rth ordinary moment of the $\mathrm{GG}(\alpha,\tau,p)$ distribution can be expressed as

$$\mu'_{r,\mathrm{GG}}=\frac{\alpha^r}{\Gamma(p)}\Gamma\left(\frac{r}{\tau}+p\right). \tag{9}$$

From Equations (8) and (9), the rth moment of X becomes

$$\mu'_r=\sum_{k=1}^{\infty}\sum_{i=0}^{\infty}\frac{(k+1)\,\eta_{k+1}\,b_{k,i}\,\alpha^r}{\Gamma(p)^{k+1}}\,\Gamma\left(\frac{r}{\tau}+p^*\right).$$

Another way to obtain μ'_r is to calculate it directly from (5)

$$\begin{aligned}\mu'_r=\frac{\tau}{\alpha\,\Gamma(p)}\int_0^{\infty}x^r\left(\frac{x}{\alpha}\right)^{\tau p-1}\exp\left[-\left(\frac{x}{\alpha}\right)^{\tau}\right]\left\{1-\gamma_1\left(p,\left[\frac{x}{\alpha}\right]^{\tau}\right)\right\}^{\gamma_1\left(p,\left[\frac{x}{\alpha}\right]^{\tau}\right)}\\ \times\left\{\frac{\gamma_1\left(p,\left[\frac{x}{\alpha}\right]^{\tau}\right)}{1-\gamma_1\left(p,\left[\frac{x}{\alpha}\right]^{\tau}\right)}-\log\left[1-\gamma_1\left(p,\left[\frac{x}{\alpha}\right]^{\tau}\right)\right]\right\}\mathrm{d}x.\end{aligned}$$

Setting $y=(x/\alpha)^{\tau}$ in the last equation

$$\begin{aligned}\mu'_r=\frac{\alpha^r}{\Gamma(p)}\int_0^{\infty}y^{p+\frac{r}{\tau}-1}\exp(-y)\left[1-\gamma_1(p,y)\right]^{\gamma_1(p,y)}\\ \times\left\{\frac{\gamma_1(p,y)}{1-\gamma_1(p,y)}-\log\left[1-\gamma_1(p,y)\right]\right\}\mathrm{d}y.\end{aligned} \tag{10}$$

The power expansion of $(1-z)^z$ at $z=0$ holds

$$(1-z)^z=\sum_{i=0}^{\infty}c_i\,z^i,\ \ |z|<1, \tag{11}$$

where $c_0 = 1$, $c_1 = 0$, $c_2 = -1$, $c_3 = -1/2$, $c_4 = 1/6$, $c_5 = 1/4$, etc. Another power series is taken into account

$$\frac{z}{1-z} - \log(1-z) = \sum_{k=1}^{\infty} \frac{(k+1)}{k} z^k, \ |z| < 1. \tag{12}$$

Applying the expansion (11) in Equation (10)

$$\mu'_r = \frac{\alpha^r}{\Gamma(p)} \int_0^{\infty} y^{p+\frac{r}{\tau}-1} \exp(-y) \sum_{i=0}^{\infty} c_i \, \gamma_1(p,y)^i \left\{ \frac{\gamma_1(p,y)}{1-\gamma_1(p,y)} - \log\left[1-\gamma_1(p,y)\right] \right\} \mathrm{d}y.$$

By using the expansion (12),

$$\mu'_r = \frac{\alpha^r}{\Gamma(p)} \int_0^{\infty} y^{p+\frac{r}{\tau}-1} \exp(-y) \sum_{i=0}^{\infty} c_i \, \gamma_1(p,y)^i \sum_{j=1}^{\infty} \frac{(j+1)}{j} \gamma_1(p,y)^j \, \mathrm{d}y.$$

Hence,

$$\mu'_r = \sum_{i=0}^{\infty} \sum_{j=1}^{\infty} \frac{(j+1)\,\alpha^r c_i}{j\,\Gamma(p)} I\left(p + \frac{r}{\tau}, i+j\right),$$

where

$$I\left(p + \frac{r}{\tau}, i+j\right) = \int_0^{\infty} y^{p+\frac{r}{\tau}-1} \exp(-y) \ \gamma_1(p,y)^{i+j} \, \mathrm{d}y.$$

This integral can be derived from Equations (24) and (25) of Nadarajah (2008), expressed in terms of the Lauricella function of type A, as defined by Exton (1978), and Aarts (2000) in the form

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

where $(a)_i = a(a+1)\cdots(a+i-1)$ is the ascending factorial (with $(a)_0 = 1$). Thus,

$$\begin{aligned} I\left(p + \frac{r}{\tau}, i+j\right) &= p^{-(i+j)} \, \Gamma\left[\frac{r}{\tau} + (i+j+1)p\right] \\ &\quad \times F_A^{(i+j)}\left(\frac{r}{\tau} + (i+j+1)p; p, \ldots, p; p+1, \ldots, p+1; -1, \ldots, -1\right). \end{aligned}$$

Plots of the skewness and kurtosis of X varying α and τ with fixed p are displayed in Figure 2, which indicates that the skewness increases and kurtosis decreases when τ grows.

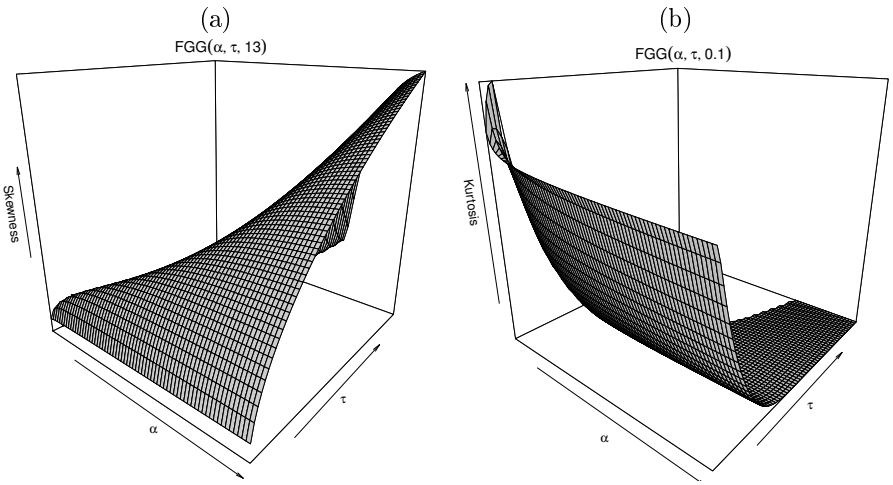

FIGURE 2: Skewness and kurtosis of X.

The rth incomplete moment of X, say $m_r(s) = \int_0^s x^r f(x)\,\mathrm{d}x$, follows from Equation (8) as

$$m_r(s) = \sum_{k=1}^{\infty}\sum_{i=0}^{\infty} \varphi_{k,i}\,\frac{\tau}{\alpha\,\Gamma(p^*)}\int_0^s x^r \left(\frac{x}{\alpha}\right)^{\tau p^*-1} \exp\left[-\left(\frac{x}{\alpha}\right)^{\tau}\right]\mathrm{d}x.$$

Thus, calculating the integral above, it comes

$$m_r(s) = \sum_{k=1}^{\infty}\sum_{i=0}^{\infty} \frac{(k+1)\,\eta_{k+1}\,b_{k,i}\,\alpha^r}{\Gamma(p)^{k+1}}\,\gamma\left(\frac{r}{\tau}+p^*, \left[\frac{s}{\alpha}\right]^{\tau}\right).$$

An application of incomplete moments is to the Bonferroni and Lonrez curves, which are (for a given probability ν) $B(\nu) = m_1(q)/\nu\mu_1'$ and $L(\nu) = m_1(q)/\mu_1'$, respectively, where q is the solution of $F(x;\alpha,\tau,p) = \nu$. Plots of these curves versus ν for specified values of τ and p ($\alpha = 0.1$) are reported in Figure 3.

A representation for the generating function (gf) of X, say $M_X(t) = \mathbb{E}[\mathrm{e}^{tX}]$, follows from Equation (8)

$$M_X(t) = \sum_{k=1}^{\infty}\sum_{i=0}^{\infty} \varphi_{k,i}\,M_{\alpha,\tau,p^*}(t), \tag{13}$$

where $M_{\alpha,\tau,p^*}(t)$ denotes the gf of the GG model with these parameters, namely

$$M_{\alpha,\tau,p^*}(t) = \frac{\tau}{\alpha\,\Gamma(p^*)}\int_0^{\infty} \exp(tx)\left(\frac{x}{\alpha}\right)^{\tau p^*-1} \exp\left[-\left(\frac{x}{\alpha}\right)^{\tau}\right]\mathrm{d}x.$$

By expanding the first exponential in a Taylor series and setting $u = x/\alpha$,

$$M_{\alpha,\tau,p^*}(t) = \frac{\tau}{\Gamma(p^*)}\sum_{j=0}^{\infty}\frac{(t\alpha)^j}{j!}\int_0^{\infty} u^{\tau p^*+j-1}\exp(-u)^{\tau}\mathrm{d}u.$$

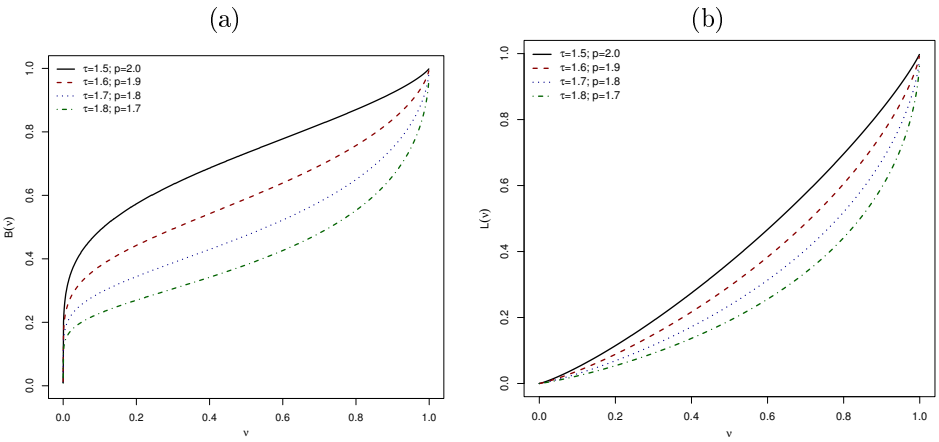

FIGURE 3: Bonferroni and Lorenz curves of X.

By solving the previous integral

$$M_{\alpha,\tau,p^*}(t) = \frac{1}{\Gamma(p^*)} \sum_{j=0}^{\infty} \frac{(t\alpha)^j}{j!} \Gamma\left(\frac{j}{\tau} + p^*\right). \tag{14}$$

However, for $\tau > 1$, it can be simplified using the Wright generalized hypergeometric function

$${}_p\Psi_q \left[\begin{array}{c} (\alpha_1, A_1), \ldots, (\alpha_p, A_p) \\ (\beta_1, B_1), \ldots, (\beta_q, B_q) \end{array} ; x \right] = \sum_{n=0}^{\infty} \frac{\prod_{j=1}^{p} \Gamma(\alpha_j + A_j n)\, x^n}{\prod_{j=1}^{q} \Gamma(\beta_j + B_j n)\, n!}. \tag{15}$$

We can rewrite (14) from Equation (15) as

$$M_{\alpha,\tau,p^*}(t) = \frac{1}{\Gamma(p^*)}\, {}_1\Psi_0 \left[\begin{array}{c} (1, \tau^{-1}) \\ - \end{array} ; \alpha t \right]. \tag{16}$$

Finally, the combination of (16) and (13) gives the gf of X.

3.3. Estimation

The log-likelihood function for the parameter vector $\boldsymbol{\theta} = (\alpha, \tau, p)^\top$ based on an independent and identically distributed (iid) random sample $x_1, \ldots, x_n$ from the FGG distribution is

$$\begin{aligned} \ell(\boldsymbol{\theta}) = {} & n\left[\log(\tau) - \log(\alpha) - \log\Gamma(p)\right] + (\tau p - 1) \sum_{i=1}^{n} \log\left(\frac{x_i}{\alpha}\right) - \sum_{i=1}^{n} \left(\frac{x_i}{\alpha}\right)^{\tau} \\ & + \sum_{i=1}^{n} \gamma_1\left(p, \left[\frac{x_i}{\alpha}\right]^{\tau}\right) \log\left\{1 - \gamma_1\left(p, \left[\frac{x_i}{\alpha}\right]^{\tau}\right)\right\} \\ & + \sum_{i=1}^{n} \log\left\{ \frac{\gamma_1\left(p, \left[\frac{x_i}{\alpha}\right]^{\tau}\right)}{1 - \gamma_1\left(p, \left[\frac{x_i}{\alpha}\right]^{\tau}\right)} - \log\left[1 - \gamma_1\left(p, \left[\frac{x_i}{\alpha}\right]^{\tau}\right)\right] \right\}. \end{aligned} \tag{17}$$

The maximum likelihood estimate (MLE) of $\boldsymbol{\theta}$ can be determined numerically by maximizing (17). The AdequacyModel package (Marinho et al., 2019) in `R` facilitates this process by offering a variety of maximization methods, such as Broyden-Fletcher-Goldfarb-Shannon (BFGS), Nelder-Mead, and Simulated Annealing (SANN).

4. FGG Regression Model

In various real-world situations, the response variable can be affected by explanatory variables in both linear and nonlinear ways. The effects of these explanatory variables on the response variable can be evaluated through the parameters of position, scale, and shape. Recent research has introduced new regressions, such as those proposed by Prataviera et al. (2018), Prataviera et al. (2021), Biazatti et al. (2022), and Cordeiro et al. (2023). Following a similar approach, an FGG regression model is constructed for censored samples, where the parameters α and τ are related to the explanatory variables by the systematic components $\alpha_i = \exp(\boldsymbol{v}_i^\top \boldsymbol{\beta}_1)$, and $\tau_i = \exp(\boldsymbol{v}_i^\top \boldsymbol{\beta}_2)$, $i = 1, \ldots, n$, where $\boldsymbol{v}_i^\top = (v_{i1}, \ldots, v_{ip})^\top$ is the vector of explanatory variables, $\boldsymbol{\beta}_1 = (\beta_{11}, \ldots, \beta_{1p})^\top$, and $\boldsymbol{\beta}_2 = (\beta_{21}, \ldots, \beta_{2p})^\top$ represents the vectors of unknown parameters. Thus, the survival function of $X_i|\boldsymbol{v}_i$ follows as

$$S(x|\boldsymbol{v}_i) = \left\{1 - \gamma_1\left(p, \left[\frac{x}{\alpha_i}\right]^{\tau_i}\right)\right\}^{\gamma_1\left(p, \left[\frac{x}{\alpha_i}\right]^{\tau_i}\right)}.$$

Consider iid observations $(x_1, \boldsymbol{v}_1), \ldots, (x_n, \boldsymbol{v}_n)$, where $x_i = \min(X_i, C_i)$. Here, X_i represents the lifetime, and C_i represents the non-informative censoring time (assuming independence). For right-censored data, the log-likelihood function for the parameter vector $\boldsymbol{\zeta} = (p, \boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$ can be expressed as

$$\begin{aligned}
\ell(\boldsymbol{\zeta}) = {} & d\left[\log(\tau_i) - \log(\alpha_i) - \log\Gamma(p)\right] + (\tau_i p - 1)\sum_{i\in F}\log\left(\frac{x_i}{\alpha_i}\right) - \sum_{i\in F}\left(\frac{x_i}{\alpha_i}\right)^{\tau_i} \\
& + \sum_{i\in F}\gamma_1\left(p, \left[\frac{x_i}{\alpha_i}\right]^{\tau_i}\right)\log\left\{1 - \gamma_1\left(p, \left[\frac{x_i}{\alpha_i}\right]^{\tau_i}\right)\right\} \\
& + \sum_{i\in F}\log\left\{\frac{\gamma_1\left(p, \left[\frac{x_i}{\alpha_i}\right]^{\tau_i}\right)}{1 - \gamma_1\left(p, \left[\frac{x_i}{\alpha_i}\right]^{\tau_i}\right)} - \log\left[1 - \gamma_1\left(p, \left[\frac{x_i}{\alpha_i}\right]^{\tau_i}\right)\right]\right\} \\
& + \sum_{i\in C}\gamma_1\left(p, \left[\frac{x_i}{\alpha_i}\right]^{\tau_i}\right)\log\left\{1 - \gamma_1\left(p, \left[\frac{x_i}{\alpha_i}\right]^{\tau_i}\right)\right\}, \qquad (18)
\end{aligned}$$

where d is the number of failures, while F and C are the sets of lifetimes and censoring times, respectively. The MLE of $\boldsymbol{\zeta}$ can be obtained by numerically maximizing the function (18). Several numerical methods can be used for this task, such as BFGS, Nelder-Mead, and SANN.

5. Simulations

To evaluate the MLEs of the FGG distribution, three scenarios with sample sizes of $n = 50$, 100, 200, and 500 are evaluated using 1 000 Monte Carlo replicates. The Newton-Raphson algorithm is used to generate random samples since the FGG quantile function does not have an analytical solution. The average estimates (AEs), biases, and mean squared errors (MSEs) can be calculated. The steps involved in the Newton-Raphson algorithm to generate variates from the FGG distribution are:

1. Set α, τ, p, and x^0.
2. Generate $u \sim \text{Uniform}(0, 1)$.
3. Update x^0 by using Newton's formula
$$x^* = x^0 - \frac{F(x^0; \alpha, \tau, p) - u}{f(x^0; \alpha, \tau, p)}.$$
4. While $\mid x^0 - x^* \mid > \epsilon$, where ϵ is a small tolerance limit, set $x^0 = x^*$ and return to step 3. Otherwise, set $x^0 = x^*$ as a variate from the FGG distribution.
5. Repeat steps 2-4 n times to generate $x_1, \ldots, x_n$ variates from the FGG distribution.

The findings in Table 2 confirm the consistency of the estimators, as evidenced by the reduction of biases and MSEs (in all scenarios), as well as the convergence of the AEs to the true values when the sample size increases.

Table 2: Simulation findings.

		(0.1, 1.5, 0.8)			(0.4, 0.5, 1.8)			(1.7, 0.9, 0.6)		
n	$\boldsymbol{\theta}$	AE	Bias	MSE	AE	Bias	MSE	AE	Bias	MSE
50	α	0.0997	-0.0002	0.0026	1.2939	0.8939	3.8036	1.8119	0.1119	1.7317
	τ	2.0980	0.5980	8.7245	0.6927	0.1927	1.1392	1.3851	0.4851	7.3450
	p	1.0331	0.2331	0.6559	2.0903	0.2903	2.0192	0.9165	0.3165	0.8986
100	α	0.0975	-0.0024	0.0017	0.9208	0.5208	1.6257	1.7202	0.0202	1.0381
	τ	1.6500	0.1500	0.4861	0.5654	0.0654	0.0603	1.0010	0.1010	0.4528
	p	0.9792	0.1792	0.3992	2.0949	0.2949	1.5251	0.7879	0.1879	0.4086
200	α	0.0973	-0.0026	0.0009	0.6782	0.2782	0.6812	1.6810	-0.0189	0.5640
	τ	1.5480	0.0480	0.1702	0.5250	0.0250	0.0244	0.9255	0.0255	0.0612
	p	0.9105	0.1105	0.1960	2.0486	0.2486	1.0054	0.6909	0.0909	0.1381
500	α	0.0990	-0.0009	0.0004	0.5329	0.1329	0.2417	1.6977	-0.0022	0.2482
	τ	1.5208	0.0208	0.0634	0.5092	0.0092	0.0097	0.9109	0.0109	0.0215
	p	0.8438	0.0438	0.0696	1.9398	0.1398	0.6303	0.6303	0.0303	0.0344

To evaluate the MLEs of the regression model, 1 000 Monte Carlo replicates are performed with sample $n = 50$, 100, 200, and 500 using the acceptance-rejection method. The real parameters are: $p = 0.7$, $\beta_{10} = 0.9$, $\beta_{11} = 0.6$, $\beta_{20} = 0.3$, and $\beta_{21} = 4.5$. The censoring times $c_1, \ldots, c_n$ are generated from a uniform distribution $(0, b)$, resulting in approximately 0%, 10%, and 30% censoring.

TABLE 3: Simulations for the FGG regression.

		0%			10%			30%		
n	$\boldsymbol{\zeta}$	AE	Bias	MSE	AE	Bias	MSE	AE	Bias	MSE
50	p	1.4762	0.7762	6.8759	1.6032	0.9032	8.6607	1.7113	1.0113	12.0441
	β_{10}	0.4123	-0.4876	2.3670	0.3208	-0.5791	3.1721	0.2543	-0.6456	4.2259
	β_{11}	0.5719	-0.0280	0.0076	0.5679	-0.0320	0.0092	0.5596	-0.0403	0.0222
	β_{20}	0.3887	0.0887	0.6976	0.3817	0.0817	0.7239	0.4687	0.1687	1.0851
	β_{21}	4.5176	0.0176	0.0098	4.5209	0.0209	0.0119	4.5338	0.0338	0.0240
100	p	1.1043	0.4043	2.7734	1.1492	0.4492	3.4557	1.3736	0.6736	6.6134
	β_{10}	0.6272	-0.2727	1.0877	0.5996	-0.3003	0.4585	0.3138	-0.4414	2.3207
	β_{11}	0.5855	-0.0144	0.0030	0.5840	-0.0159	0.0034	0.5763	-0.0236	0.0064
	β_{20}	0.2889	-0.0110	0.1860	0.2942	-0.0057	0.2205	0.2883	-0.0116	0.3175
	β_{21}	4.5044	0.0044	0.0021	4.5075	0.0075	0.0066	4.5150	0.0150	0.0145
200	p	0.8179	0.1179	0.3836	0.8341	0.1341	0.5225	0.8797	0.1797	1.0639
	β_{10}	0.8089	-0.0910	0.2095	0.7974	-0.1025	0.2605	0.7652	-0.1347	0.4337
	β_{11}	0.5951	-0.0048	0.0005	0.5945	-0.0054	0.0007	0.5927	-0.0072	0.0012
	β_{20}	0.3063	0.0063	0.0787	0.3052	0.0052	0.0872	0.3090	0.0090	0.1166
	β_{21}	4.5007	0.0007	0.0002	4.5010	0.0010	0.0003	4.5045	0.0045	0.0113
500	p	0.7376	0.0376	0.0461	0.7428	0.0428	0.0537	0.7688	0.0688	0.1017
	β_{10}	0.8681	-0.0318	0.0373	0.8636	-0.0364	0.0428	0.8408	-0.0591	0.0769
	β_{11}	0.5981	-0.0018	0.0001	0.5979	-0.0020	0.0001	0.5968	-0.0032	0.0002
	β_{20}	0.2992	-0.0007	0.0257	0.2989	-0.0010	0.0308	0.2957	-0.0042	0.0515
	β_{21}	4.5000	0.0000	0.0000	4.5000	0.0000	0.0000	4.4998	-0.0001	0.0001

The simulation process is described as follows (for $i = 1, \ldots, n$):

1. Generate $v_{i1} \sim \text{Uniform}\,(0, 1)$, and calculate $\alpha_i = \exp(\beta_{10} + \beta_{11} v_{i1})$, and $\tau_i = \exp(\beta_{20} + \beta_{21} v_{i1})$.
2. Generate y_i from $w(y_i) = \tau_i\, \alpha_i^{-\tau_i}\, y_i^{\tau_i - 1} \exp\left[-(y_i/\alpha_i)^{\tau_i}\right]$.
3. Generate $u \sim \text{Uniform}\,(0, 1)$.
4. If $u \leq f(y_i)/M w(y_i)$, then $x_i = y_i$, where $f(\cdot)$ comes from (5) setting $\tau = \tau_i$, and $\alpha = \alpha_i$, and $M = \max\left[f(y_i)/w(y_i)\right]$. Otherwise, return to step 2.
5. Generate $c_i \sim \text{Uniform}\,(0, b)$.
6. The censoring indicator $\delta_i = 1$ if $x_i \leq c_i$ and $\delta_i = 0$ otherwise, and $x_i^* = \min(x_i, c_i)$ provides the observed times.

Table 3 shows that the biases and MSEs decrease as n increases for all censoring levels, indicating consistent estimators. There is a slight increase in biases and MSEs as the censoring rate increases, but they remain low for large samples ($n \geq 200$) even with 30% censoring. The estimates of β_{11} and β_{21} are less affected by censoring and present more accurate estimates, while p and β_{10} are most impacted.

Equations (17) and (18) are optimized using the `BFGS` numerical method implemented in the `optim` function in R with numerical derivatives. The initial values of the parameters for both optimizations are set to their actual values. These simulation scripts are carried out without the use of any package in R.

6. Applications

The new models are evaluated on three real data sets, and their fits are compared using the Cramér-von Mises (W^*) and Anderson-Darling (A^*) statistics (Chen & Balakrishnan, 1995), along with the Akaike information criterion (AIC), consistent Akaike information criterion (CAIC), Bayesian information criterion (BIC), Hannan-Quinn information criterion (HQIC), and Kolmogorov-Smirnov (KS) with its p-value. Lower values of these measures indicate a better fit. The statistical analyses address in Sections 6.1 and 6.2 use the AdequacyModel package in R with the `BFGS` method. One of the main advantages of this package is its ability to calculate the MLEs, their standard errors (SEs), and various adequacy measures without the need to define the log-likelihood function. All that is required is to provide the pdf and cdf of the distribution that fits the data set. In Section 6.3, an R script is executed to obtain the results. This script requires the `GenSA` package to determine the initial values and employs the SANN numerical method via the `optim` function, whose output provides the MLEs, SEs, and adequacy measures for the chosen regressions fitted to the data.

6.1. COVID-19 Data (Recife)

The first data set contains information on the recovery times (in days) of 257 individuals diagnosed with COVID-19 through laboratory testing in the city of Recife, Brazil, between 2021 and 2022. These data, available at `https://opendatasus.saude.gov.br/en/dataset/notificacoes-de-sindrome-gripal-leve-2022`, reveal that the average recovery period is 48.167 days, with a standard deviation of 24.137. The skewness (0.116) and kurtosis (2.035) indicate a right-skewed and platykurtic distribution of the data.

The FGG distribution is compared with some distributions listed in Table 1, thus including the flexible generalized half normal (FGHN), flexible half normal (FHN), flexible Weibull (FW), and flexible exponential (FE) distributions. Additionally, well-known distributions in the literature such as GG, exponentiated Weibull (EW) (Mudholkar & Srivastava, 1993), and exponentiated Fréchet (EF) (Nadarajah & Kotz, 2003) are also considered for comparison.

The MLEs and SEs in parentheses from the models fitted to COVID-19 data in Recife are reported in Table 4. All distributions produce accurate estimates, with the FGG distribution showing the best fit, as indicated by the lowest values of the adequacy measures in Table 5 and the highest p-value in the KS test.

Table 4: Estimates of the models fitted to COVID-19 (Recife).

Model	MLEs (SEs)		
FGG (α, τ, p)	93.901	8.197	0.086
	(2.654)	(2.531)	(0.030)
GG (α, τ, p)	86.866	6.969	0.198
	(4.201)	(2.361)	(0.080)
EW (α, τ, β)	81.277	6.365	0.218
	(1.827)	(0.147)	(0.016)
EF (α, θ, γ)	40.072	0.451	917.172
	(11.607)	(0.026)	(287.595)
FW (α, τ, 1)	42.002	1.304	
	(1.581)	(0.066)	
FGHN ($2^{1/2\gamma}$, 2γ, 1/2)	47.929	1.028	
	(1.654)	(0.053)	
FHN ($\sigma\sqrt{2}$, 2, 1/2)	47.598		
	(1.562)		
FE (α, 1, 1)	38.925		
	(1.740)		

The generalized likelihood ratio (GLR) tests (Vuong, 1989) reveal the superiority of the FGG model in terms of fit compared to alternative models, GG (GLR = 12.161), EW (GLR = 7.979), EF (GLR = 4.430), FW (GLR = 11.591), FGHN (GLR = 15.129), FHN (GLR = 14.283), and FE (GLR = 16.514), at the 5% significance level. The plots in Figure 4 confirm the superior fit of the FGG model to COVID-19 data, thus supporting the numerical results presented. The new distribution provides an excellent fit to the real COVID-19 data set and outperforms alternative distributions, thus making it a valuable tool for analyzing similar data sets.

Table 5: Adequacy measures of the models fitted to COVID-19 data (Recife).

Model	W^*	A^*	AIC	CAIC	BIC	HQIC	KS	p-value
FGG	0.049	0.451	2341.699	2341.794	2352.346	2345.981	0.038	0.844
GG	0.078	0.626	2342.882	2342.977	2353.529	2347.164	0.047	0.614
EW	0.074	0.603	2342.386	2342.480	2353.033	2346.667	0.045	0.648
EF	0.754	4.655	2414.680	2414.775	2425.327	2418.962	0.113	0.002
FW	0.441	2.885	2381.526	2381.573	2388.624	2384.380	0.073	0.125
FGHN	0.261	1.805	2366.333	2366.380	2373.431	2369.187	0.063	0.252
FHN	0.271	1.863	2364.615	2364.631	2368.164	2366.042	0.064	0.229
FE	0.537	3.441	2404.313	2404.329	2407.862	2405.741	0.103	0.008

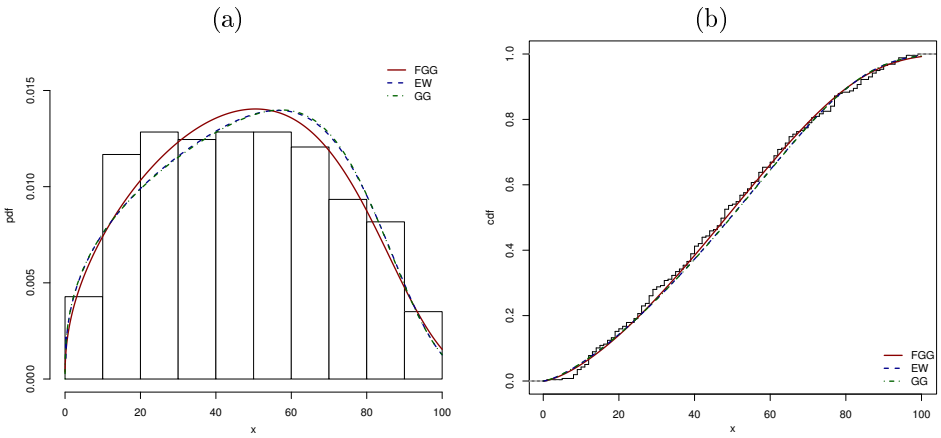

FIGURE 4: Pdfs and cdfs of the FGG, EW, and GG models fitted to COVID-19 data (Recife).

6.2. COVID-19 Data (Alagoas)

The second data set contains the lifetimes in days for 143 individuals diagnosed with COVID-19 in the state of Alagoas, Brazil, during 2020-2021, extracted from the link: `https://dados.al.gov.br/catalogo/dataset/painel-covid-19-em-alagoas`. The average lifetime is approximately 9.685 days, with a standard deviation of 5.562. The skewness (0.128) and kurtosis (2.001) indicate that the data are right-skewed and platykurtic.

The MLEs and SEs of the models fitted to the current data are given in Table 6. The competing distributions evaluated are the same as those mentioned in Section 6.1, and all of them provide accurate estimates except for the GG and EF distributions. Among these models, the FGG distribution is the best one to fit the data, with the lowest values of W^*, A^*, and other adequacy measures as shown in Table 7.

The GLR tests indicate that the FGG model has a statistically superior fit compared to its competitors, the GG (GLR = 2.629), EW (GLR = 40.687), EF (GLR = 13.397), FW (GLR = 14.505), FGHN (GLR = 7.415), FHN (GLR = 8.118) and FE (GLR = 15.724) models at the 5% level. Figure 5 shows that the estimated pdf and cdf of the FGG model are the closest to the histogram and empirical cdf of the data, thus validating its flexibility in fitting COVID-19 data in Alagoas.

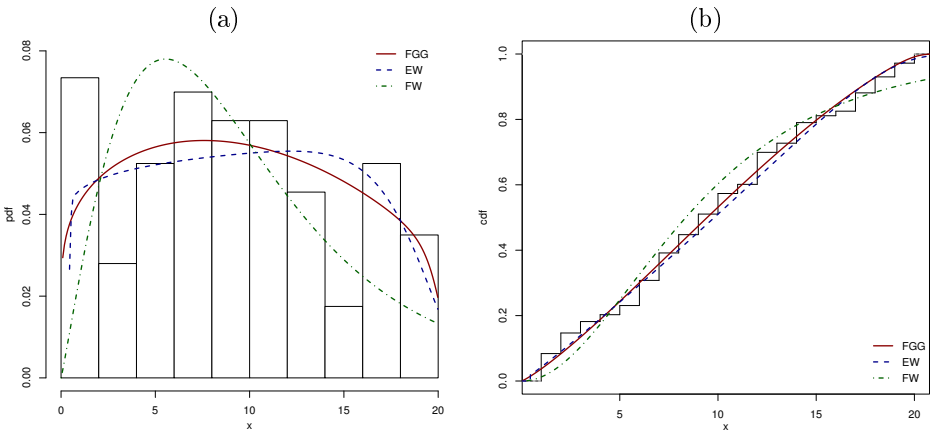

FIGURE 5: Pdfs and cdfs of the FGG, EW, and FW models fitted to COVID-19 data (Alagoas).

TABLE 6: Estimates of the models fitted to COVID-19 (Alagoas).

Model	MLEs (SEs)		
FGG (α, τ, p)	20.382	42.807	0.013
	(0.325)	(20.872)	(0.006)
GG (α, τ, p)	20.150	40.587	0.025
	(0.287)	(30.738)	(0.019)
EW (α, τ, β)	18.659	10.008	0.108
	(0.025)	(0.026)	(0.009)
EF (α, θ, γ)	37.466	0.369	326.279
	(14.228)	(0.028)	(164.968)
FW (α, τ, 1)	7.905	1.047	
	(0.496)	(0.072)	
FGHN ($2^{1/2\gamma}$, 2γ, 1/2)	9.318	0.841	
	(0.526)	(0.060)	
FHN ($\sigma\sqrt{2}$, 2, 1/2)	9.801		
	(0.436)		
FE (α, 1, 1)	7.780		
	(0.473)		

TABLE 7: Adequacy measures of the models fitted to COVID-19 data (Alagoas).

Model	W^*	A^*	AIC	CAIC	BIC	HQIC	KS	p-value
FGG	0.100	0.809	866.751	866.923	875.639	870.363	0.069	0.489
GG	0.131	0.947	868.116	868.289	877.004	871.728	0.101	0.107
EW	0.121	0.947	871.588	871.760	880.476	875.200	0.078	0.341
EF	0.805	5.309	927.807	927.979	936.695	931.418	0.147	0.003
FW	0.493	3.515	909.718	909.804	915.644	912.126	0.103	0.095
FGHN	0.317	2.453	898.236	898.322	904.162	900.644	0.097	0.130
FHN	0.256	2.070	902.646	902.674	905.608	903.849	0.118	0.035
FE	0.510	3.618	908.161	908.189	911.124	909.365	0.108	0.068

6.3. COVID-19 Data (Ceará)

The third data set contains information on the lifetimes (in days) of 368 individuals diagnosed with COVID-19 in the state of Ceará, Brazil, between 2021 and 2022. The data are available at `https://github.com/integrasus/api-covid-ce`. Here, the response variable x_i represents the survival time from symptom onset to death due to COVID-19 (failure).

Approximately 50% of the observations are censored, referring to individuals who died from non-COVID-19 causes or survived the study period. The variables considered (for $i = 1, \ldots, 368$) include: δ_i : censoring indicator (0 = censored, 1 = observed lifetime), v_{i1} : age (in years), and v_{i2} : immunodeficiency (1 = yes, 0 = no).

Figure 6(a) reveals a peak around 60 years, thus indicating that COVID-19 mainly affects middle-aged and older adults. Figure 6(b) shows a significantly lower probability of survival over time among immunodeficient patients compared to those without immunodeficiency, highlighting its negative impact on COVID-19 prognosis.

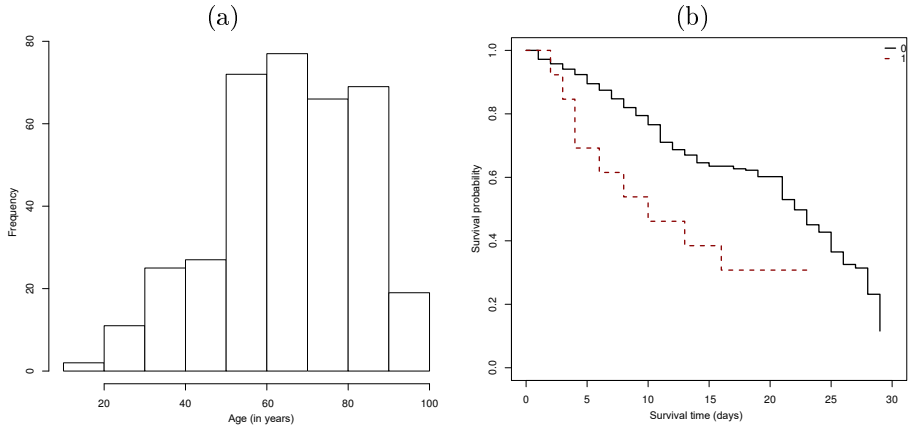

FIGURE 6: Histogram for age (a) and Kaplan-Meier curves for immunodeficiency (b) for COVID-19 data (Ceará).

In the FGG model and its sub-models, the explanatory variables are associated with systematic components (for $i = 1, \ldots, 368$)

$$\alpha_i = \exp(\beta_{10} + \beta_{11} v_{i1} + \beta_{12} v_{i2}) \text{ and } \tau_i = \exp(\beta_{20} + \beta_{21} v_{i1} + \beta_{22} v_{i2}).$$

The logarithmic link function is used to relate the explanatory variables for two parameters of the FGG and FW distributions and one parameter of the FR, FE, and FM distributions. Among the fitted regression models, the FGG regression offers the best fit for COVID-19 data in Ceará, as indicated by the results in Table 8. Additionally, likelihood ratio (LR) tests confirm the superiority of the FGG model over the alternatives, because the null hypotheses (Table 9) are strongly rejected in favor of the FGG model.

According to the numbers in Table 10, the explanatory variable v_{i1} (β_{11}) is significant at the 5% level, indicating that age is a factor that can shorten the time to failure. Moreover, age and immunodeficiency are significant for the variability of survival times (β_{21}) and β_{22}). However, the explanatory variable immunodeficiency (β_{12}) does not significantly affect the time to failure, leading to a reduced regression model in the form

$$\alpha_i = \exp(\beta_{10} + \beta_{11} v_{i1}) \text{ and } \tau_i = \exp(\beta_{20} + \beta_{21} v_{i1} + \beta_{22} v_{i2}).$$

The estimation results are shown in Table 11. A new LR test confirms the superior fit of the reduced FGG regression over the FW (LR = 21.575), FE (LR = 30.726), FR (LR = 299.367), and FM (LR = 603.548) regressions, and the p-values align with those in Table 9. The updated adequacy measures for the reduced regressions FGG (AIC = 1526.882, CAIC = 1527.283, BIC = 1550.331, HQIC = 1536.198), FW (AIC = 1546.076, CAIC = 1546.387, BIC = 1565.617, HQIC = 1553.840), FE (AIC = 1549.227, CAIC = 1549.338, BIC = 1557.043, HQIC = 1552.333), FR (AIC = 1817.868, CAIC = 1817.978, BIC = 1825.684, HQIC = 1820.974), and FM (AIC = 2122.049, CAIC = 2122.159, BIC = 2129.865, HQIC = 2125.154) further corroborate this claim. Residual analysis for the reduced FGG regression fit is conducted using the quantile residuals (qrs) (Dunn & Smyth, 1996), namely

$$qr_i = \Phi^{-1}\left(1 - \left\{1 - \gamma_1\left(\hat{p}, \left[\frac{x_i}{\hat{\alpha}_i}\right]^{\hat{\tau}_i}\right)\right\}^{\gamma_1\left(\hat{p}, \left[\frac{x_i}{\hat{\alpha}_i}\right]^{\hat{\tau}_i}\right)}\right),$$

where $\Phi(\cdot)^{-1}$ is the inverse cdf of the standard normal distribution, $\hat{\alpha}_i = \exp(\boldsymbol{v}_i^\top \hat{\boldsymbol{\beta}}_1)$, and $\hat{\tau}_i = \exp(\boldsymbol{v}_i^\top \hat{\boldsymbol{\beta}}_2)$.

Figure 7 shows that the qrs have random behavior and approximately follow a standard normal distribution. This indicates that there is no evidence against the assumptions of the FGG regression model for COVID-19 data in Ceará. Further, the close correspondence between the empirical and estimated survival functions for the explanatory variable immunodeficiency (Figure 8) confirms the adequacy of the fitted model.

TABLE 8: Adequacy measures of the regressions fitted to COVID-19 data (Ceará).

Model	AIC	CAIC	BIC	HQIC
FGG	1529.084	1529.586	1556.400	1539.952
FW	1544.554	1544.955	1568.002	1553.870
FE	1545.023	1545.188	1556.747	1549.680
FR	1808.341	1808.507	1820.065	1812.999
FM	2107.194	2107.359	2118.918	2111.852

Table 9: LR tests for COVID-19 data (Ceará).

Model	Hypotheses	LR statistic	p-value
FGG vs FW	$H_0 : p = 1$ vs $H_1 : H_0$ is false	17.470	< 0.0001
FGG vs FE	$H_0 : p = \tau = 1$ vs $H_1 : H_0$ is false	23.938	< 0.0001
FGG vs FR	$H_0 : \tau = 2, p = 1$ vs $H_1 : H_0$ is false	287.257	< 0.0001
FGG vs FM	$H_0 : \tau = 2, p = 3/2$ vs $H_1 : H_0$ is false	586.110	< 0.0001

Table 10: Complete FGG regression estimates for COVID-19 data (Ceará).

Parameter	MLEs	SEs	p-value
p	0.011	0.003	0.001
β_{10}	4.266	0.069	<0.001
β_{11}	-0.009	0.000	<0.001
β_{12}	-0.079	0.308	0.796
β_{20}	4.702	0.372	<0.001
β_{21}	-0.009	0.002	0.001
β_{22}	-0.745	0.329	0.024

Table 11: Reduced FGG regression estimates for COVID-19 data (Ceará).

Parameter	MLEs	SEs	p-value
p	0.011	0.004	0.008
β_{10}	4.263	0.070	<0.001
β_{11}	-0.008	0.000	<0.001
β_{20}	4.672	0.431	<0.001
β_{21}	-0.009	0.002	0.001
β_{22}	-0.592	0.250	0.018

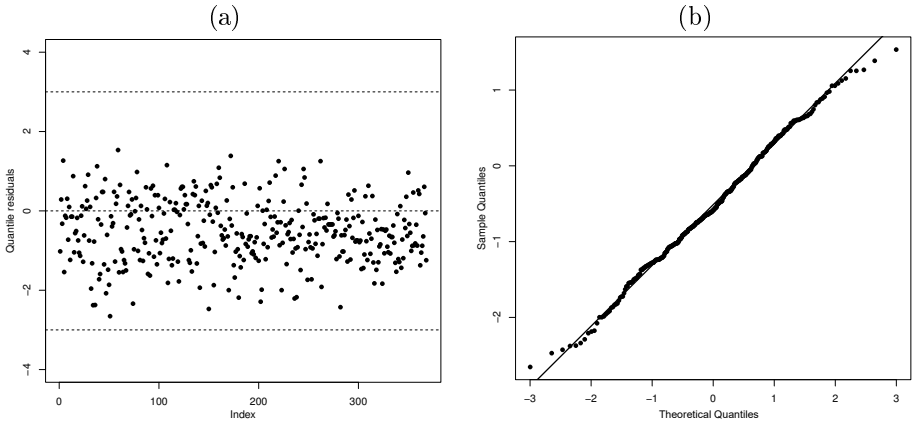

Figure 7: Index plot (a) and normal probability plot (b) for COVID-19 data (Ceará).

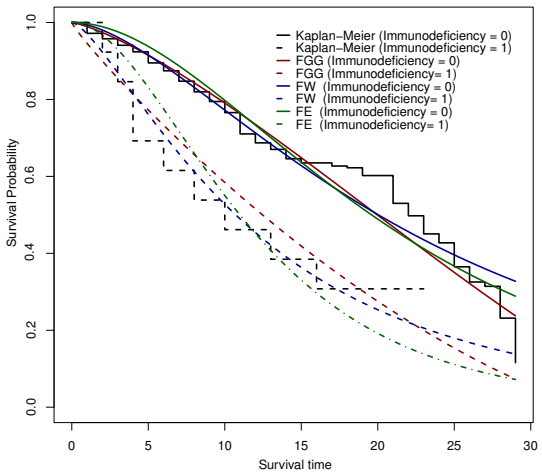

FIGURE 8: Estimated survival functions and the empirical survival for immunodeficiency.

7. Conclusions

This article proposed the flexible generalized gamma distribution, whose density is expressed as a linear combination of generalized gamma densities. This formulation facilitates the derivation of the main properties of this new distribution. Simulation results validate the consistency of the maximum likelihood estimators. The new model provides good fits and outperforms alternative models for real COVID-19 data. Regression analysis identifies significant effects of age and immunodeficiency on lifetime. Future research could explore Bayesian inference methods and applications in other fields, such as economics and engineering.

Acknowledgements

Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE) [IBPG-1448-1.02/20] supports this work.

[Received: January 2024 — Accepted: May 2024]

References

Aarts, R. M. (2000), 'Lauricella functions', `http://www.mathworld.wolfram.com/LauricellaFunctions.html`. From MathWorld - A Wolfram Web Resource, created by Eric W. Weisstein.

Biazatti, E. C., Cordeiro, G. M., Rodrigues, G. M., Ortega, E. M. M. & de Santana, L. H. (2022), 'A Weibull-beta prime distribution to model COVID-19 data with the presence of covariates and censored data', *Stats* **5**(4), 1159–1173.

Chen, G. & Balakrishnan, N. (1995), 'A general purpose approximate goodness-of-fit test', *Journal of Quality Technology* **27**(2), 154–161.

Cordeiro, G. M., Castellares, F., Montenegro, L. C. & de Castro, M. (2013), 'The beta generalized gamma distribution', *Statistics* **47**(4), 888–900.

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

Cordeiro, G. M., Rodrigues, G. M., Ortega, E. M. M., de Santana, L. H. & Vila, R. (2023), 'An extended Rayleigh model: Properties, regression and covid-19 application', *Chilean Journal of Statistics* **14**(1), 1–25.

Dunn, P. K. & Smyth, G. K. (1996), 'Randomized quantile residuals', *Journal of Computational and Graphical Statistics* **5**(3), 236–244.

Exton, H. (1978), *Handbook of hypergeometric integrals: theory, applications, tables, computer programs*, Ellis Horwood.

Gradshteyn, I. S. & Ryzhik, I. M. (2007), *Table of integrals, series, and products*, seventh edn, Elsevier/Academic Press, Amsterdam. Translated from the Russian, Translation edited and with a preface by Alan Jeffrey and Daniel Zwillinger, With one CD-ROM (Windows, Macintosh and UNIX).

Marinho, P. R. D., Silva, R. B., Bourguignon, M., Cordeiro, G. M. & Nadarajah, S. (2019), 'AdequacyModel: An R package for probability distributions and general purpose optimization', *PLOS ONE* **14**(8), 1–30.

Mudholkar, G. S. & Srivastava, D. K. (1993), 'Exponentiated Weibull family for analyzing bathtub failure-rate data', *IEEE Transactions on Reliability* **42**(2), 299–302.

Nadarajah, S. (2008), 'Explicit expressions for moments of order statistics', *Statistics & Probability Letters* **78**(2), 196–205.

Nadarajah, S. & Kotz, S. (2003), 'The exponentiated Fréchet distribution', *Interstat Electronic Journal* **14**, 01–07.

Ortega, E. M. M., Cordeiro, G. M. & Pascoa, M. A. R. (2011), 'The generalized gamma geometric distribution', *Journal of Statistical Theory and Applications* **10**(4), 433–54.

Pascoa, M. A. R., Ortega, E. M. M. & Cordeiro, G. M. (2011), 'The Kumaraswamy generalized gamma distribution with application in survival analysis', *Statistical methodology* **8**(5), 411–433.

Prataviera, F., Ortega, E. M. M., Cordeiro, G. M., Pescim, R. R. & Verssani, B. A. W. (2018), 'A new generalized odd log-logistic flexible Weibull regression model with applications in repairable systems', *Reliability Engineering & System Safety* **176**, 13–26.

Prataviera, F., Silva, A. M. M., Cardoso, E. J. B. N., Cordeiro, G. M. & Ortega, E. M. M. (2021), 'A novel generalized odd log-logistic Maxwell-based regression with application to microbiology', *Applied Mathematical Modelling* **93**, 148–164.

Stacy, E. W. (1962), 'A generalization of the gamma distribution', *The Annals of Mathematical Statistics* **33**, 1187–1192.

Tahir, M. H., Hussain, M. A. & Cordeiro, G. M. (2022), 'A new flexible generalized family for constructing many families of distributions', *Journal of applied statistics* **49**(7), 1615–1635.

Vuong, Q. H. (1989), 'Likelihood ratio tests for model selection and non-nested hypotheses', *Econometrica* **57**(2), 307–333.

Appendix A.

Applying the findings of Cordeiro et al. (2011), one can write

$$\gamma_1\left(p,\left[\frac{x}{\alpha}\right]^{\tau}\right)^k=\frac{1}{\Gamma(p)^k}\left(\frac{x}{\alpha}\right)^{kp\tau}\left\{\sum_{i=0}^{\infty}\left[-\left(\frac{x}{\alpha}\right)^{\tau}\right]^i\frac{1}{(p+i)i!}\right\}^k. \tag{A1}$$

By employing the findings of Section 0.314 of Gradshteyn & Ryzhik (2007), one can represent a power series raised to a positive integer as

$$\left[\sum_{i=0}^{\infty}a_i\left(\frac{x}{\alpha}\right)^{i\tau}\right]^k=\sum_{i=0}^{\infty}b_{k,i}\left(\frac{x}{\alpha}\right)^{i\tau}, \tag{A2}$$

where $a_i=(-1)^i/[(p+i)i!]$, $b_{k,0}=a_0^k$, and (for $i\geq 1$)

$$b_{k,i}=\frac{1}{ia_0}\sum_{j=1}^{i}[\,j(k+1)-i\,]a_j\,b_{k,i-j}. \tag{A3}$$

Thus, by combining (A1) and (A2) yields

$$\gamma_1\left(p,\left[\frac{x}{\alpha}\right]^{\tau}\right)^k=\sum_{i=0}^{\infty}\frac{b_{k,i}}{\Gamma(p)^k}\left(\frac{x}{\alpha}\right)^{(i+kp)\tau}, \tag{A4}$$

where the coefficients $b_{k,i}$ are calculated from (A3).