The Log-Generalized Gamma Burr III Regression Model with Application to Predict Death of Breast Cancer

Olobatuyi Kehinde Ibukun[1]*, Asiribo Osebekwiin Ebenezer[2] and Talabi Olusola Adetunji[3]

[1]*University of Milano-Bicocca, Milan, Italy Department of Statistics, Italy*

[2]*Federal University of Agriculture, Abeokuta Department of Statistics, Nigeria*

[3]*Federal University of Agriculture, Abeokuta Health Centers, Nigeria*

*Corresponding author: Olobatuyi Kehinde Ibukun, University of Milano-Bicocca, Milan, Italy Department of Statistics, Italy

ARTICLE INFO

ABSTRACT

We study the properties of the called log-generalized gamma Burr III distribution defined by the logarithm of the generalized gamma Burr III random variable (Olobatuyi, et al. [1]). An advantage of the new distribution is that it includes as special sub-models classical distributions reported and has the ability to model unimodal HFs. We obtain formal expressions for the moments, moment generating function, quantile function and mean and median deviations. We constructed a regression model based on the new distribution to predict relief time of headache patients and death of breast cancer patients treated by mastectomy. It can be applied to censored data since it represents a parametric family of models that includes as special sub-models several widely known regression models. The regression model was fitted to a data set of 1207 eligible breast cancer patients. We predict survival probability after the mastectomy in terms of highly significant clinical and pathological explanatory variables associated with the death of the patients. The predicted probabilities of survival are calculated under two nested models.

Keywords: Generalized Gamma Burr III Distribution; Censored Data; Log- Generalized Gamma Burr III Distribution; Log- Generalized Gamma Burr III Regression Model; Survival Function

Introduction

Standard lifetime distributions usually present very strong restrictions to produce bathtub curves, and thus appear to be inappropriate for interpreting data with this characteristic. Some distributions were introduced to model this kind of data, as the exponential power family [2], the beta integrated model [3], and the generalized log-gamma distribution, among others. A good review of these models is described, for instance, in [4] In the last decade, new classes of distributions for modeling this type of data based on extensions of the Weibull distribution were developed. For example, the exponentiated Weibull (EW) (Mudholkar, et al. [5]), the additive Weibull [6], the modified Weibull (Lai, et al. [7]), the beta Weibull (BW) (Famoye, et al. [8,9]) and the generalized modified Weibull (Carrasco, et al. [10]) distributions. Further, (Carrasco, et al. [10]) investigated several mathematical properties of the BW geometric distribution, which is a highly flexible lifetime model to cope with different degrees of kurtosis and asymmetry. The Generalized Gamma Burr III (GGBIII) distribution, due to its flexibility in accommodating the different types of the risk function depending

on its parameters, can be used in a variety of problems in modeling survival data. The main motivation for the use of the GGBIII model is that it contains as special sub-models several distributions such as the generalized gamma Fisk, Zografos and Balakrishnan-Burr III, Zografos and Balakrishnan-Fisk, Burr III distribution among others. Also, it was reported that gamma model is the most effective model for analyzing highly skewed data such as survival data, [11].

Breast cancer presents a major risk to American women, who have a 1 in 8 lifetime chance of developing the disease. The estimated incidence of invasive breast cancer in the United States for 2010 was 207,090 women, making it the most common cancer after skin cancer in women. Although survival has improved because of advances in treatment and early diagnoses as a result of the increased use of mammographic screening, fatalities in 2010 have been put at 40,000. Mastectomy is surgery that removes the entire breast. All the breast tissue is removed, sometimes along with other nearby tissues. If just the breast is removed (and not lymph nodes under the arm) it is called a simple (or total) mastectomy. A simple mastectomy combined with an axillary lymph node dissection (discussed below) is called a modified radical mastectomy. The primary end point was survival (DFS), defined as time to the earliest of either death (all-cause), or last follow-up. The follow-up period was defined as time to the earliest of either death (all-cause), or last follow-up. For the first time, we propose a log-generalized gamma Burr III regression model to predict the t months survival probability after mastectomy in terms of highly significant clinical and pathological variables associated with death of the patient after surgery. The study cohort comprises 1207 patients with clinically localized cancer treated by mastectomy. The data consist of the random response variable given by the number of months (y_i) after mastectomy. Uncensored observations correspond to patients having death time computed. Censored observations correspond to patients who were not observed to have died at the time the data were collected. The numbers of censored and uncensored observations are 1135 and 72, respectively, of the total of 1207 patients.

In this article, we propose a location-scale regression model based on the LGGBIII distribution, referred to as the LGGBIII regression model, which is a feasible alternative for modeling the existing types of failure rate functions. Some inferential issues were carried out using the asymptotic distribution of the maximum likelihood estimators (MLEs). The sections are organized as follows. In Section 2, we define the LGGBIII distribution. Mathematical properties of this distribution are investigated in Section 3. In Section 4, we obtain the order statistics. We propose a LGGBIII regression

model for censored data and discuss inferential issues in Section 5. In Section 6, a breast cancer data set is analyzed to show the flexibility, practical relevance and applicability of our regression model. Section 7 ends with some concluding remarks.

Generalized Gamma Burr III Distribution

Most generalized Burr III distributions such as Beta Burr III distribution Antonio and Silva (2014) have been proposed in reliability literature to provide better fitting of certain data sets than the traditional two and three parameter Burr III models. The GGBIII density function (Olobatuyi, et al. [1]) with five parameters $\alpha > 0$, $\beta > 0$, $\delta > 0$, $k > 0$ and $\lambda > 0$ is given by ($t > 0$)

$$f(t) = \frac{\rho\beta\delta\lambda^\delta t^{-\delta-1}}{\Gamma(\alpha)} \left(1+(\frac{t}{\lambda})^{-\delta}\right)^{-\beta-1} \left[\left[-\log\left(1+(\frac{t}{\lambda})^{-\delta}\right)^{-\beta}\right]^p\right]^{\alpha-1}$$

$$\exp\left(-\left[-\log\left(1-\left(1+\frac{t}{\lambda}\right)^{-\delta}\right)^{-\beta}\right]^p\right)\left[1\left(1+\left(\frac{t}{\lambda}\right)^{-\delta}\right)^{-\beta}\right]^{-1} \quad (1)$$

where $\Gamma(.)$ is a gamma function. Here, α and k are two additional shape parameters to the Burr III distribution to model the skewness and kurtosis of the data. The important characteristic of the GGBIII distribution is that it contains as special sub-models. The hazard and survival rate functions corresponding to (1) are

$$h(t) = \frac{\rho\beta\lambda^\delta\partial t^{-\delta-1}\left(1+(\frac{t}{\lambda})^{-\delta}\right)^{-\beta-1}}{1-(1+\left(\frac{t}{\lambda}\right)^{-\delta})^{-\beta}} \frac{\left[\left[-\log\left(1-(\frac{t}{\lambda})^{-\delta}\right)^{-\beta}\right]\right]^{p\alpha-1}\exp\left(-\left[-\log\left(1-\left(1+\frac{t}{\lambda}\right)^{-\delta}\right)^{-\beta}\right]^p\right)}{\Gamma(\alpha)-\gamma\left[-\log[1-\left(1+\frac{t}{\lambda}\right)^{-\delta}]^{-\beta}\right]^p,\alpha\right)}$$

and (2)

$$s(t) = 1 - \frac{\gamma\left(\left[-\log\left[1-\left(1+\left(\frac{t}{\lambda}\right)^{-\delta}\right)^{-\beta}\right]\right]^p,\alpha\right)}{\Gamma(\alpha)} \quad (3)$$

Shape of GGBIII Distribution

Plots of the density function of the Generalized Gamma Burr III distribution for selected parameters values are given in Figure 1. The plot indicates that the GGBIII distribution can be decreasing or right skewed.
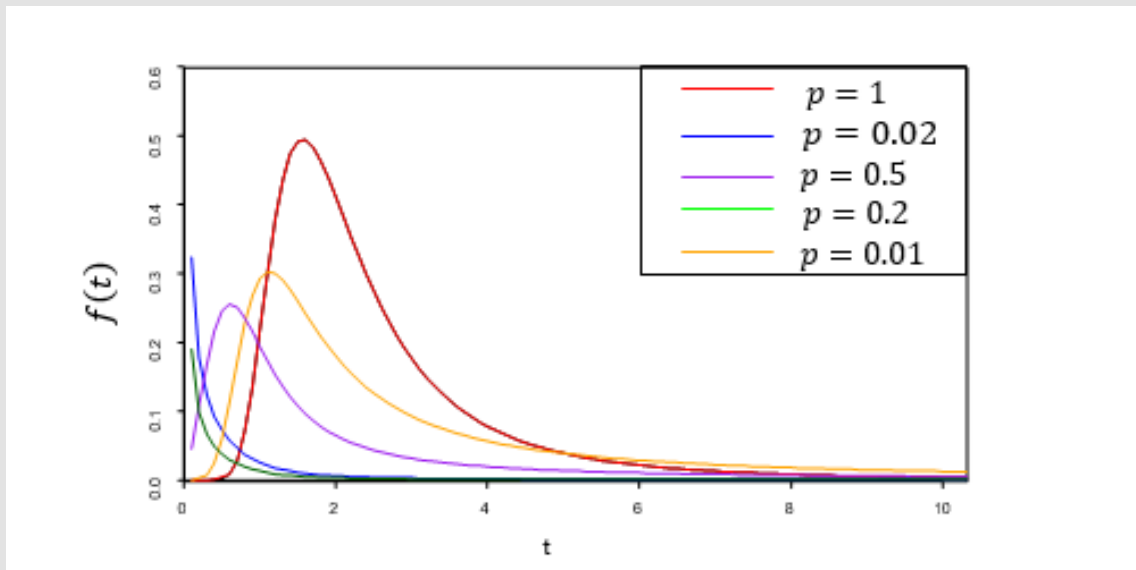
Figure 1: Shape of GGBIII distribution for selected parameters.

The Log-Generalized Gamma Burr III Distribution

In this section, log-generalized gamma Burr III distribution is introduced. It is based on the logarithm of the continuous GGBIII distribution that is presented above. The log-generalized gamma Burr III distribution is proposed and denoted as LGGBIII. Some of its mathematical properties are studied, estimation by the method of maximum likelihood is discussed, and applications to two real datasets are described. The new distribution is shown to outperform at least two models which are the log-ZBD and Cox model. Let T be a random variable having the GGBIII density function (1), The random variable $Y = \log(T)$ has a log-generalized gamma-Burr II(L-GGBIII) distribution, whose density function is reparametrized in

$$\delta = \frac{1}{\sigma}$$

terms of and $\mu = \log \lambda$ can be expressed as

$$g(y;\alpha,p,\beta,\mu,\sigma) = \frac{p\beta}{\sigma\Gamma(\alpha)}\exp\left\{-\left(\frac{y-\mu}{\sigma}\right)-\left[-\log[1-\left(1+\exp-\left(\frac{y-\mu}{\sigma}\right)\right)^{-\beta}\right]^{p}\right\}$$

$$[-\log(1-(1+\exp-(\tfrac{y-\mu}{\sigma}))^{-\beta})]^{p\alpha-1}(1+\exp-(\tfrac{y-\mu}{\sigma}))^{-\beta-1}[1-(1+\exp-(\tfrac{y-\mu}{\sigma}))^{-\beta}]^{-1} \quad (4)$$

Where $-\infty < y < \infty$, $\sigma > 0$, and $-\infty < \mu < \infty$, $p > 0$, $\alpha > 0$, $\beta > 0$. We refer to the new model (4) as the LGGBIII distribution, say $LGGBIII(\alpha, \beta, p, \mu, \sigma)$ where μ is a location parameter, σ is a dispersion parameter and α, β and k are shape parameters. The following results hold if $T \sim GGBIII(\alpha, \beta, p, \delta, \lambda)$ then $Y = \log(T) \sim LGGBIII(\alpha, \beta, p, \mu, \sigma)$. The standard random variable $Z = Y - \mu/\sigma$ with density function is defined as

$$\pi(z;p,\alpha,\beta) = \frac{p\beta}{\sigma\Gamma(\alpha)}(1+\exp(-z)^{-\beta-1}[-\log(1-(1+\exp(-z))^{-\beta})]^{p\alpha-1}$$

$$\exp\{(-z)-[-\log(1-(1+\exp(-z))^{-\beta}]^{p}\ (1-(1-(1+\exp(-z))^{-\beta})]^{-1}\} \quad (5)$$

The special case of the model lead to a standard log-ZBD (LZBD new) distribution for $p = 1$. For $\beta = p = 1$, we obtain the log-Zografos and Balakrishnana Fisk (LZB-F new). The survival functions corresponding to (4) and (5) are

$$s(y;\alpha,\beta,p,\mu,\sigma) = 1 - \frac{\gamma([-\log(1-\left[-\log[1-\left(1+\exp-\left(\frac{y-\mu}{\sigma}\right)\right)^{-\beta}\right]\right]^{p},\alpha)}{\sigma\Gamma(\alpha)} \quad (6)$$

$$s(z;\alpha,\beta,p) = 1 - \frac{\gamma([-\log(1-\left[-\log[1-\left(1+\exp(-z)\right)^{-\beta}\right]\right]^{p},\alpha)}{\sigma\Gamma(\alpha)} \quad (7)$$

Expansions of Density Functions

If a random variable Z has the LGGBIII density, we say $Z\sim LGGBIII(\alpha, \beta, k)$. Let $u = (1 + e^{-z})^{-\beta}$, and then using the series representation from Gradshteyn and Ryzhik (2000).

$$e^{-[-\log(1-u)]^{p}} = \sum_{h=o}^{\infty}(-1)^{h}\frac{[-\log(1-u)]^{ph}}{h!},1.212no3$$

$$\frac{1}{1-u} = \sum_{h=o}^{\infty}u^{i},0.234no1,$$

$$[-\log(1-u)]^{p\alpha-1} = u^{p\alpha-1}\left[\sum_{j=o}^{\infty}\binom{p\alpha-1}{j}u^{j}\left(\sum_{s=o}^{\infty}\frac{u^{s}}{s+2}\right)^{j}\right],$$

and applying the result on the power series raised to a positive integer, with $cs = (s + 2)-1$, that is,

$$\left(\sum_{s=o}^{\infty} c_{s,j} u^s\right)^j = \left(\sum_{s=o}^{\infty} a_{s,j} u^s\right)^{,j}$$
0.314

Where $a_{s,j} = (sc_0)^{-1}\sum_{l=1}^{s}[j(l+1)-s]c_l a_{s-l,j},$ and $a_{o,j} = c_0^j$, [12], the LGGBIII PDF can be written as

$$\pi(z) = \frac{p\beta[1+e^{-z}]^{-\beta-1}}{\Gamma(\alpha)} u^{p\alpha-1} \sum_{j=o}^{\infty}\sum_{s=o}^{\infty}\binom{p\alpha-1}{j} a_{s,j} u^{j+s} \sum_{h=o}^{\infty}(-1)^h \frac{[-\log(1-u)]^{ph}}{h!} \sum_{h=o}^{\infty} u^i$$

$$\frac{p\beta[1+e^{-z}]^{-\beta-1}}{\Gamma(\alpha)} \sum_{j=o}^{\infty}\sum_{s,=o}^{\infty}\sum_{h=o}^{\infty}\sum_{i=o}^{\infty} \binom{p\alpha-1}{j}(-1)^h \frac{[-\log(1-u)]^{ph}}{h!} a_{s,j} u^{p\alpha+s+j+i-1} \quad (8)$$

$$\sum_{j=o}^{\infty}\sum_{s,=o}^{\infty}\sum_{h=o}^{\infty}\sum_{i=o}^{\infty}(-1)^h\binom{p\alpha-1}{j}\frac{p\beta[1+e^{-z}]^{-\beta-1}}{\Gamma(\alpha)}\frac{[-\log(1-u)]^{ph}}{h!} a_{s,j}[1+e^{-z}]^{-\beta p\alpha+\beta s+\beta j+\beta i+\beta} \quad (9)$$

$$=\sum_{j=o}^{\infty}\sum_{s,=o}^{\infty}\sum_{h=o}^{\infty}\sum_{i=o}^{\infty}(-1)^h\binom{p\alpha-1}{j}\frac{p\beta[1+e^{-z}]^{-\beta-1}}{\Gamma(\alpha)}\frac{[-\log(1-u)]^{ph}}{h!} a_{s,j} u^{p\alpha+s+j+i-1} \quad (10)$$

$$=\sum_{j=o}^{\infty}\sum_{s,=o}^{\infty}\sum_{h=o}^{\infty}\sum_{i=o}^{\infty}(-1)^h\binom{p\alpha-1}{j}\frac{p\beta}{\Gamma(\alpha)} a_{s,j}[1+e^{-z}]^{-\beta(p\alpha+s+j+i)-1}\frac{[-\log(1-u)]^{ph}}{h!} \quad (11)$$

$$=\sum_{j=o}^{\infty}\sum_{s,=o}^{\infty}\sum_{h=o}^{\infty}\sum_{i=o}^{\infty}(-1)^h\binom{p\alpha-1}{j}\frac{p\beta^{(p\alpha+s+i)}}{\Gamma(\alpha)^{(p\alpha+s+j+i)}} a_{s,j}^{-\beta(p\alpha+s+j+i)-1}\frac{[-\log(1-u)]^{ph}}{h!} \quad (12)$$

$$=\sum_{j=o}^{\infty}\sum_{s,=o}^{\infty}\sum_{h=o}^{\infty}\sum_{i=o}^{\infty}\frac{(-1)^h}{h!}\binom{p\alpha-1}{j}\frac{p\beta^{(p\alpha+s+i)}}{\Gamma(\alpha)^{(p\alpha+s+j+i)}} a_{s,j}[1+e^{-z}]^{-\beta(p\alpha+s+j+i)-1}[-\log(1-(1+e^{-z})^{-\beta})]^{ph} \quad (13)$$

Where $\pi(z;p,\beta)(p\alpha+j+s+i),\alpha)$ is the Burr III PDF with parameters $p,\beta(p\alpha+j+s+i),\alpha.$ Let $C = \{(j, s, h, i) \in Z^3 +\}$

$$\sum_{h=o}^{\infty}\psi_v = \frac{(-1)^h}{h!}\binom{k\alpha-1}{j}\frac{a_{s,j}}{\Gamma(\alpha)(k\alpha+s+j+i)},$$

$$\pi_{LGGBIII}(z) = \sum_{V\varepsilon c}\psi_v f(z;k,\beta(k\alpha+j+s+i),\alpha). \quad (14)$$

Shape of LGGBIII Distribution Function

The plot (4) in Figure 1 for selected parameter values show great flexibility of the density function in terms of the parameters α, β, *p* in Figure 2a β = 3, *p* = 2 and in Figure 1b, α = 2.5, *p* = 2 and in Figure 1d, α = 2.5, β = 3.
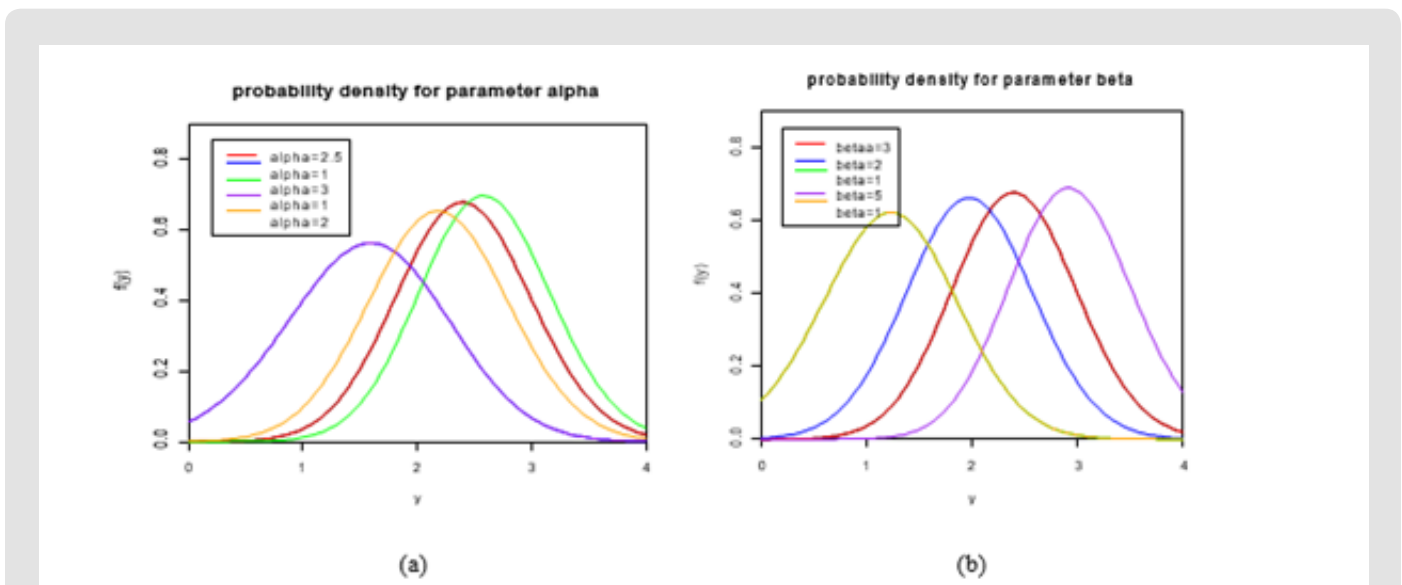
Figure 2: Plots of the LGGBIII density for some parameter values, (a), and (b), for constant mu and sigma.

LGGBIII Quantile Function

We now give an expansion for the quantile function $q = F^{-1}(k)$ (given k) of the LGGBIII distribution. First, we have k = F(q). It is possible to obtain as function of p from some expansions for the inverse of the gamma incomplete function $Q_{GGBIII}(q) = k$, $0 < q < 1$.

$$\frac{\gamma\left([-\log[1-(1+e^{-q})^{-\beta}]^p,\alpha\right)}{\Gamma(\alpha)} = k$$

$$\gamma\left([-\log[1-(1+e^{-q})^{-\beta}]^p,\alpha\right) = k\Gamma(\alpha)$$

Let $[-\log[1-(1+e^{-q})^{-\beta}]^{p}$ then,

$$\left([-\log[1-(1+e^{-q}]^{-\beta}\right) = u^{1/p}$$

$$[1-(1+e^{-q}]^{-\beta}) = e^{-u1/p}$$

$$1-e^{-u1/p} = (1=(1+e^{-q}]^{-\beta}$$

$$(1-e^{-u1/p})^{-1/\beta}-1 = e^{-q}$$

$$q = -\log[(1-e^{-u1/p})^{-1/\beta}-1] \quad (15)$$

Moments, Moment Generating Function Mean and Median Deviations

In this section, we present the moments, moment generating function, mean and median deviations for the GGBIII distribution.

$$Q_{LGGBIII}(q) = -\log[(1-e^{-[\gamma-1(K\Gamma(\alpha),\alpha)1/p})^{-1/\beta}-1] \quad (16)$$

Moments and Moment Generating Function

As with any other distribution, many of the interesting characteristics and features of the LGGBIII distribution can be studied through the moments. Let $\beta* = \beta(p\alpha+j+s+i)$, and $X\sim LBIII(\mu,\sigma,\beta*)$. Then the rth moment of the random variable Y is

$$E(X^{r}) = \beta^{*}(\exp\mu)^{r}B(\beta^{*}+\gamma\sigma,1\text{-}\gamma\sigma), \text{ r}<1/\sigma \quad (32)$$

So that the rth raw moment is thus given by the following

$$\underset{E}{E(Y^{r})} = \underset{V\varepsilon c}{\sum}\psi_{v}\beta^{*}(\exp\mu)^{r}B(\beta^{*}+r\sigma,1\text{-}r\sigma), \text{ r}<1/\sigma, \quad (33)$$

The moment generating function of the LGGBIII distribution is given by

$$M_{Y}(t) = E(e^{tY}) = E\left[1+tY+\frac{(tY)^{3}}{3!}+...\right], \quad (34)$$

$$= \sum_{r=1}^{\infty}\frac{t^{r}}{r!}E(Y^{r}),$$

Than we have,

Mean and Median Deviations

Mean Deviation: If Y has the LGGBIII distribution, we derive the mean deviation about the mean μ by

$$\delta_{1} = \int_{0}^{\infty}|y-\mu|gGD(y)dy = 2\mu G_{LGGBIII}(\mu)-2\mu+2T(\mu), \quad (36)$$

Where $\mu = E(Y)$ and T

$$T(\mu) = \int_{\mu}^{\infty}y\cdot gLGGBIII(y)dy.Let\beta^{*} = \beta(p\alpha+j+s+i), \text{ than}$$

$$T(\mu) = \sum_{veC}\varphi vT_{LBIII}(\beta^{*},\mu,\sigma)(\mu) \quad (37)$$

$$= \sum_{veC}\varphi_{v}\beta^{*}\exp(\mu)[B(\beta^{*}+\sigma,1-\sigma)-B(t(\mu)\beta^{*}+\sigma,1-\sigma)] \quad (38)$$

$$\delta_{1} = 2\sum_{veC}\varphi_{v}\beta^{*}\exp(\mu)[B(\beta^{*}+\sigma,1-\sigma)]\left[\frac{\gamma(-\log(1-(1-\exp(\frac{y-\mu}{\sigma}))^{-\beta})],\alpha)}{\Gamma(\alpha)}-1\right]$$

$$+2\sum_{veC}\varphi_{v}\beta^{*}\exp(\mu)[B(\beta^{*}+\sigma,1-\sigma)-B(t(\mu);\beta^{*}+\sigma,1-\sigma)]$$

$$\delta_{1} = +2\sum_{veC}\varphi_{v}\beta^{*}\exp(\mu)[B(\beta^{*}+\sigma,1-\sigma)]\left[\frac{\gamma(-\log(1-(1+\exp(-z))^{-\beta})]^{p},\alpha)}{\Gamma(\alpha)}-1\right] \quad (39)$$

Where $t(\mu) = (1+\exp(-z))^{-1}$ and

$$B(Y;A,B) = \int_{0}^{y}t^{a-1}(1-t)^{b-1}dt$$

Median Deviation: If Y has the LGGBIII distribution, we derive the median deviation about the median M by

$$\delta_{2} = \int_{0}^{\infty}|y-M|g_{GD}(X)dx = 2T(M)-\mu, \quad (40)$$

$$T(M) = \int_{0}^{\infty}y^{\bullet}g_{LGGBIII}(y)dyM|g_{GD}$$

where

$$M = \underset{LGGBIII(0.5)=-\log\left[\left(1-e^{-[\gamma-1(0.5\Gamma(\alpha),\alpha)]}\right)^{-1/\beta}-1\right]}{Q} \quad (41)$$

$$T(M) = \sum_{\upsilon\varepsilon C}\varphi T_{LBIII}(\beta^{*},\sigma,\mu)(M)$$

$$= \sum_{\upsilon\varepsilon C}\varphi_{v}\beta^{*}\exp(\mu)[B(\beta^{*}+\sigma,1-\sigma)-B(t(M);\beta^{*+\sigma,1-\sigma)]} \quad (42)$$

$$2\sum_{\upsilon\varepsilon C}\varphi_{v}\beta^{*}\exp(\mu)[B(\beta^{*}+\sigma,1-\sigma)-B(t(M);\beta^{*+\sigma,1-\sigma)]}$$

$$-\sum_{\upsilon\varepsilon C}\varphi_{v}\beta^{*}\exp(\mu)[B(\beta^{*}+\sigma,1-\sigma)] \quad (43)$$

Order Statistics of LGGBIII Distribution

Order statistics make their appearance in many areas of statistical theory and practice. The density $f_{i:n}(x)$ of the ith order statistic

$(Z_{i:n})$ for $i = 1, \ldots, n$ from i.i.d. LGGBIII random variables $Z_1, \ldots, Z_n$ is simply given by

$$f_{i;n}(Z) = \frac{n!\pi(Z)}{(n-i)!(i-1)!}[\Pi(z)^{i-1}[1-\Pi(Z)^{n-i},$$

By using the binomial expression to the second factor we obtain

$$f_{i;n}(Z) = \frac{n!\pi(Z)}{(n-i)!(i-1)!}\sum_{j-0}^{n-i}(-1)^j \binom{n-i}{j}[\pi(z)]^{i+j-1}$$

$$f_{i;n}(Z) = \frac{n!\pi(Z)}{(n-i)!(i-1)!}\sum_{j-0}^{n-i}(-1)^j \binom{n-i}{j}\left[\frac{\gamma\left(\left[-\log(1-(1+e^{-z})^{-\beta}\right]\right)^p,\alpha\right)}{\Gamma(\alpha)}\right]^{i+j-1}$$

$$\gamma(z,\alpha) = \sum_{i=0}^{\infty}\frac{(-1)^{l_z l+2}}{(l+\alpha)l!},$$

$$d_l = (-1)^l/((l+\alpha)l!), \text{ then}$$

$$f_{i;n}(Z) = \frac{n!\pi(Z)}{(n-i)!(i-1)!}\sum_{j-0}^{n-1}\binom{n-i}{j}\frac{(-1)^j}{[\Gamma(\alpha)]^{i+j-1}}([-\log(1-u)]^p)^{\alpha(i+j-1)}\left[\sum_{i=0}^z\frac{(-1)^i([-\log(1-(1+e^{-z}]^{-\beta}]^p)^i)}{(l+\alpha)l!}\right]^{i+j-1}$$

$$= \frac{n!p([-\log(1-(1+e^{-z})^{-\beta})])^{p\alpha-1}e^{-[-\log(1-(1+e^{-z})^{-\beta})]^p}[1-(1+e^{-z})^{-\beta)}]^{-1}\pi(z)}{(n-l)!(i-1)!\Gamma(\alpha)}$$

$$\sum_{j-0}^{n-i}\sum_{l-0}^{\infty}\binom{n-i}{j}\frac{(-1)^j d_{m,i+j-1}}{[\Gamma(\alpha)]^{i+j-1}}\times([-\log(1-(1+e^{-z})^{-\beta})]^p)^{\alpha(i+j-1)+l}$$

where $f(z; (\alpha(i+j)+l), p, \beta,)$ is the LGGBIII pdf with parameters β, p and shape parameter $\alpha^* = \alpha(i+j) + l$.

$$= \frac{n!}{(n-l)!(i-1)!}\sum_{j-0}^{n-i}\sum_{l-0}^{\infty}\binom{n-i}{j}\frac{(-1)^j d_{m,i+j-1}\Gamma((\alpha(i+j)+l))}{[\Gamma(\alpha)]^{i+j-1}}\frac{p\beta(1+e^{-z})^{-\beta-1}}{\Gamma((\alpha(i+j)+l])}[-\log(1-(1+e^{-z})^{-\beta})]^{p\alpha(i+j-1)-1}$$

$$\times\log e^{-[-\log(1-(1+e^{-z})^{-\beta})]^p}[(1-(1+e^{-z})^{-\beta}]^{-1}$$

The Log-Generalized Gamma Burr III Regression Model

In many practical applications, the lifetimes are affected by explanatory variables such as the cholesterol level, blood pressure, weight and many others. Parametric regression models to estimate univariate survival functions for censored data regression problems are widely used. A parametric model that provides a good fit to lifetime data tends to yield more precise estimates of the quantities of interest. Based on the LGGBIII density function, we propose a linear location-scale regression model linking the response variable y_i and the explanatory variable vector $V_T = (v_{i1}, \ldots, v_{ip})$ as follows

$$y_i = V_i^T\gamma + \sigma z_i, i = 1, \ldots\ldots, n, \quad (44)$$

where the random error z_i has density function (40), $\boldsymbol{\gamma} = (B_1, \ldots, Bp)$, $\sigma > 0, p > 0, \alpha > 0, \beta > 0$ are unknown parameters. $\mu i = V^T\boldsymbol{\gamma}$ is the location of y_i. The location parameter vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$ T is represented by a linear model $\boldsymbol{\mu} = \boldsymbol{V}_{\gamma}$, where $\boldsymbol{V} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ T is a known model matrix. Let F and C be the sets of individuals for which y_i is the log-lifetime and log-censoring, respectively. The log-likelihood function for the vector $\boldsymbol{\theta} = (\alpha, \beta, p, \sigma, \gamma_T)T$ of parameters from model (39) has the form

$$l(\theta) = \sum_{i\varepsilon F}\log[f(y_i)] + \sum_{i\varepsilon C}\log[S(y_i)] \quad (45)$$

where $f(y_i)$ is the density function (4) and $S(y_i)$ is the survival function (5) of Y_i. The log-likelihood function for $\boldsymbol{\theta}$ reduces to

$$l(\theta) = r[\log(p)+\log(\beta)-\log(\sigma)-\log(\Gamma(\alpha))]-(\beta+1)\sum_{i\varepsilon F}\log[1+e^{-z_i}]+(p\alpha-1)\sum_{i\varepsilon F}\log[-\log(1-)1+e^{-z_i})^{-\beta}]$$

$$-\sum_{i\varepsilon F}z_i - \sum_{i\varepsilon F}[-\log(1-(1+e^{-z_i})^{-\beta})]^p - \sum_{i\varepsilon F}\log(1-(1+e^{-z_i})^{-\beta})$$

$$+\sum_{i\varepsilon F}\log(1-\frac{\gamma([-\log(1-(1+e^{z_i})^{-\beta})]^p,\alpha)}{\Gamma(\alpha)}) \quad (46)$$

where r is the number of uncensored observations (failures) and $z_i = (yi - \boldsymbol{V}^T{}_{\gamma})/\sigma$. The score functions for the parameters p, α, β, σ given by

$$V_p(\theta) = \frac{r}{p} + \alpha\sum_{i\varepsilon F}\log(v_i) - \sum_{i\varepsilon F}(v_i)^p\log(v_i) - \sum_{i\varepsilon F}\frac{\left[\frac{\gamma(V_l^P,\alpha)}{\Gamma(\alpha)}\right]_p}{1-\frac{\gamma(v_i^p,\alpha)}{\Gamma(\alpha)}}$$

$$V_\alpha(\theta) = \frac{r\Gamma^\cdot(\alpha)}{\Gamma(\alpha)} + p\sum_{i\varepsilon F}\log(v_i) - \sum_{i\varepsilon F}\frac{\left[\frac{\gamma(V_l^P,\alpha)}{\Gamma(\alpha)}\right]_\alpha}{1-\frac{\gamma(v_i^p,\alpha)}{\Gamma(\alpha)}}$$

$$V_\beta(\theta) = \frac{r^\cdot}{\beta} - \sum_{i\varepsilon F}\log(1+e^{-z_i}) + (p\alpha-1)\sum_{i\varepsilon F}\frac{(1+e^{-z_i})^{-\beta}\log(1+e^{-z_i}))}{(1-(1+e^{-z_i})^{-\beta})[v_i]} - \frac{\left[\frac{\gamma(V_l^P,\alpha)}{\Gamma(\alpha)}\right]_\beta}{1-\frac{\gamma(v_i^p,\alpha)}{\Gamma(\alpha)}}$$

$$-\frac{1}{\sigma}\sum_{i\varepsilon F}\frac{z_i(1+e^{-z_i})^{-\beta}\log(1+e^{-z_i}))}{(1-(1+e^{-z_i})^{-\beta})} - \sum_{i\varepsilon F}\frac{\left[\frac{\gamma(V_l^P,\alpha)}{\Gamma(\alpha)}\right]_\alpha}{1-\frac{\gamma(v_i^p,\alpha)}{\Gamma(\alpha)}}$$

$$V_\sigma(\theta) = \frac{1}{\sigma}\sum_{i\varepsilon F}u_{ij} - (\beta+1)\frac{1}{\sigma}\sum_{i\varepsilon F}\frac{u_{ij}e^{-z_i}}{1+e^{-z_i})} - \frac{(p\alpha-1)}{\sigma}\sum_{i\varepsilon F}-\frac{u_{ij}(1+e^{-z_i})^{-\beta}}{(1-(1+e^{-z_i})^{-\beta})[v_i]} - \frac{1}{\sigma}\frac{u_{ij}(1+e^{-z_i})^{-\beta}}{[v_i]^k}$$

$$-\frac{1}{\sigma}\sum_{i\varepsilon F}\frac{z_i(1+e^{-z_i})^{-\beta}\log(1+e^{-z_i}))}{(1-(1+e^{-z_i})^{-\beta})}-\sum_{i\varepsilon F}\frac{\left[\frac{\gamma(V_i^P,\alpha)}{\Gamma(\alpha)}\right]_\sigma}{1-\frac{\gamma(v_i^p,\alpha)}{\Gamma(\alpha)}}$$

$$V_{\gamma j}(\theta)=\frac{1}{\sigma}\sum_{i\varepsilon F}u_{ij}-(\beta+1)\frac{1}{\sigma}\sum_{i\varepsilon F}\frac{u_{ij}e^{-z_i}}{1+e^{-z_i}}-\frac{(p\alpha-1)}{\sigma}\sum_{i\varepsilon F}\frac{u_{ij}(1+e^{-z_i})^{-\beta}}{(1-(1+e^{-z_i})^{-\beta})[v_i]}-\frac{1}{\sigma}\sum_{i\varepsilon F}\frac{u_{ij}(1+e^{-z_i})^{-\beta}}{[v_i]^p}$$

$$-\frac{1}{\sigma}\sum_{i\varepsilon F}\frac{u_{ij}(1+e^{-z_i})^{-\beta}}{(1-(1+e^{-z_i})^{-\beta})}-\sum_{i\varepsilon F}\frac{\left[\frac{\gamma(V_i^P,\alpha)}{\Gamma(\alpha)}\right]_{\gamma j}}{1-\frac{\gamma(v_i^p,\alpha)}{\Gamma(\alpha)}}$$

Here

$$\left[\frac{\gamma(v_i^P,\alpha)}{\Gamma(\alpha)}\right]_p=\partial\frac{\gamma(V_i^P,\alpha)}{\Gamma(\alpha)}/\partial p,\left[\frac{\gamma(V_i^P,\alpha)}{\Gamma(\alpha)}\right]_\alpha=\partial\frac{\gamma(V_i^P,\alpha)}{\Gamma(\alpha)}/\partial\alpha,\left[\frac{\gamma(V_i^P,\alpha)}{\Gamma(\alpha)}\right]_\beta=\partial\frac{\gamma(V_i^P,\alpha)}{\Gamma(\alpha)}/\partial\beta,\left[\frac{\gamma(V_i^P,\alpha)}{\Gamma(\alpha)}\right]_\sigma=\partial\frac{\gamma(V_i^P,\alpha)}{\Gamma(\alpha)}/\partial\sigma,$$

$$\left[\frac{\gamma(v_i^P,\alpha)}{\Gamma(\alpha)}\right]_{\gamma j}=\partial\frac{\gamma(V_i^P,\alpha)}{\Gamma(\alpha)}/\partial\gamma_j,v_i=-\log(1-(1+e^{-z_i})^{-\beta}),and\ j=1,\ldots\ldots p.$$

The MLE $\boldsymbol{\theta}$ of the vector $\boldsymbol{\theta}$ of unknown parameters can be calculated by maximizing the log-likelihood (46). We use the subroutine NLMixed in SAS to calculate $\boldsymbol{\theta}$. Initial values for and can be taken from the fit of the log-Zog Fisk (LZFisk) regression model with $\beta = p = 1$. The fitted LGGBIII model gives the estimated survival function of Y for any individual with explanatory vector x

$$s(y;\alpha,\beta,p,\sigma,\gamma^T)=1-\frac{\gamma([-\log(1-(1+\exp(\frac{y-X^T\gamma}{\sigma}))^{-\beta})]^P,\alpha)}{\Gamma(\alpha)}$$

The invariance property of the MLEs yields the survival function for $T = \exp(Y)$

$$s(t;\alpha,\beta,p,\delta,\lambda)=1-\frac{\gamma([-\log(1-(1+(t/\lambda)^{-\delta})^{-\beta}]^P,\alpha)}{\Gamma(\alpha)}$$

The approximate multivariate normal distribution $Np+5(0, \mathbf{L}(\boldsymbol{\theta})-1)$ for $\boldsymbol{\theta}$ can be used in the classical way to construct approximate confidence regions for some parameters in $\boldsymbol{\theta}$. We can use the likelihood ratio LR statistic for comparing some special sub-models with the LGGBIII model. We consider the partition $\boldsymbol{\theta} = (\boldsymbol{\theta}_T, \boldsymbol{\theta}_T)$ T where $\boldsymbol{\theta}$ 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 $\boldsymbol{H_0}:\boldsymbol{\theta}_1=\boldsymbol{\theta}_1^{(0)}$ versus the alternative hypothesis $\boldsymbol{H_0}:\boldsymbol{\theta}_1\neq\boldsymbol{\theta}_1^{(0)}$ is given by $\omega=2\{\ell(\boldsymbol{\theta})-\ell(\boldsymbol{\theta})\}$ where where $\boldsymbol{\theta}$ and $\boldsymbol{\theta}$ are the estimates under the null and alternative hypotheses, respectively. The statistic ω is asymptotically (as $n\to\infty$) distributed as χ_h^2 where h is the

dimension of the subset of parameters of interest.

Results and Discussion

Predicting the Time to Headache Relief

Thirty-eight patients are divided into two groups of equal size, and different pain relievers are assigned to each group. The outcome reported is the time in minutes until headache relief. The variable censor indicates whether relief was observed during the observation period (censor = 0) or whether the observation is censored (censor=1).

The variables involved in the study are

- t_i – survival time to Headache relief (in minutes);

- $cens_i$ – censoring indicator (0 or 1);

- $group_j$ – (j = 1,2);

Now, by fitting the model

$$y_i = \beta_0 - \beta_1 * (group_{ij1} - 2) + \sigma z_i,$$

$$H_0: grp\ 1\ time = grp\ 2time\ vs\ H1: grp\ 1\ time < grp\ 2\ time$$

The random variable z_i follows the LGGBIII distribution (5) for $i = 1, \ldots ,38$. We are interested in modelling which group recovers faster. The MLEs of the model parameters are calculated using the procedure NLMIXED in SAS. Iterative maximization of the logarithm of the likelihood function (46) starts with initial values for $\beta0 = \beta1 = 1$, $\alpha = 1$, $\beta = 1$, $p = 1$, $\sigma = 1$, to fit the regression model. Table 1 lists the Maximum Likelihood Estimation of the model parameters. The value of Akaike Information Criterion (AIC), Corrected Akaike Information Criterion (AICc), and Bayesian Information Criterion (BIC) statistics are smaller for LGGBIII regression model. A comparison of the new model with one of its sub-models which is LZB-D model and most useful model for fitting survival data which is Log-Weibull (LW) model using LR statistic is presented in Table 2 together with p-values. From the values of these statistics, LGG-BIII distribution provided a good fit for this data. The LGGBIII regression model outperforms the other models irrespective of the criteria and it can be used effectively in the analysis of these data. So, the proposed model is a great alternative to model survival data. The model to know which group has a high rate of Headache relief that is the group that has a short time to Headache relief was fitted, recall that $\log(\lambda) = \mu$ then, since $\mu = v^TQ$ say $\log(\lambda) = b_0 - b_1 \times (group - 2)$ as the mean of the y_i. Now, from the result provided in Table 3 then substituting the value of $b0$ and $b1$ into the regression model as follows. For group 1

Table 1: MLEs of the Model Parameters for the Time –To- Headache Relief.

Model	a	β	p	σ	b0	b1
LGGBIII	2.1801 -0.2355	0.06645 -0.0149	0.3863 -0.044	0.0117 -0.0007	3.1337 -0.0163 [< 0.0001]	−0.1265 -0.0227 [< 0.0001]
LZB-D	6.2949 -2.8683	0.012 -0.0265	—	0.0687 -0.02	3.091 -0.0773 [< 0.0001]	−0.1487 -0.0638 [< 0.0252]
LW	—	—	—	0.2122 -0.0304	3.3091 -0.0589 [< 0.0001]	−0.1933 -0.0227 [0.0185]

Table 2: The -2L, AIC, AICC, BIC of Time of Headache Relief.

MODEL	-2Log-Like-lihood	AIC	AICC	BIC
LGGBIII	4.4	16.4	19.1	26.2
LZB-D	15.3	25.3	27.2	33.5
LW	18.8	24.8	25.5	29.7

Table 3: The Likelihood Ratio Test Statistic.

MODEL	Hypothesis	LR Statistic	P-Value
LGGBIII vs LZB-D	H0: LZBD vs H1: LGGBIII	10.9	0.0017
LGGBIII vs LW	H0: LW vs H1: LGGBIII	14.4	0.0007

While for group 2

$$\log \lambda = 3.1337 - (-0.1265) \times (-1)$$

$$\lambda_1 = \exp(3.0072)20.231,$$

The probabilities of headache relief by t *minutes* are calculated as follows

$$G_1(Y) = \frac{\gamma\left(-\log(1-(1-\exp(\frac{y-3.0072}{0.01165})^{-0.06645]^{0.3863}}\right)}{\Gamma(2.1801)} ,2.1801)$$

Note that $y = \log(minutes)$.

$$G_2(Y) = \frac{\gamma\left(-\log(1-(1-\exp(\frac{y-3.1337}{0.01165})^{-0.06645]^{0.3863}}\right)}{\Gamma(2.1801)} ,2.1801)$$

These probabilities calculated at the observed times are shown for the two groups in Table 4.

Since the slope estimate is negative i.e. $b_1 = -0.1265$, $SE = 0.02272$ then t value = -5.57 with p-value of < 0.0001, it was seen that group 1 has a shorter time to Headache relief than group 2 that is pain reliever 1 leads to overall significantly faster relief, but the estimated probabilities give no information about patient-to-patient variation within and between groups. For example, while pain reliever 1 provides faster relief overall, some patients in group 2 may respond more quickly than other patients in group 1. Provided below in Figures 3 & 4 is a graphical representation of the predicted value for the two groups.

Table 4: The predicted value of the Time to Headache Relief.

Obs	Group	Censor	Patient	Minutes	Prediction
1	1	0	1	11	0.01909
2	1	0	2	12	0.02815
3	1	0	3	19	0.2425
4	1	0	4	19	0.2425
5	1	0	5	19	0.2425
6	1	0	6	19	0.2425
7	1	0	7	21	0.53836
8	1	0	8	20	0.36158
9	1	0	9	21	0.53836
10	1	0	10	21	0.53836
11	1	0	11	20	0.36158
12	1	0	12	21	0.53836

13	1	0	13	20	0.36158
					0.53836
					0.80058
					0.84721
					0.88929
					0.53836
					0.76627
14	1	0	14	21	0.03185
15	1	0	15	25	0.36158
					0.53836
					0.80058
					0.84721
					0.88929
					0.53836
					0.76627
16	1	0	16	27	0.03185
17	1	0	17	30	0.36158
					0.53836
					0.80058
					0.84721
17	1	0	17	30	0.88929
					0.53836
					0.76627
18	1	1	18	21	0.03185
19	1	1	19	24	0.36158
					0.53836
					0.80058
					0.84721
					0.88929
					0.53836
					0.76627
20	2	0	20	14	0.03185
21	2	0	21	16	0.05748
22	2	0	22	16	0.05748
23	2	0	23	21	0.20697
24	2	0	24	21	0.20697
25	2	0	25	23	0.41004
26	2	0	26	23	0.41004
27	2	0	27	23	0.41004
28	2	0	28	23	0.41004
29	2	1	29	25	65269
30	2	0	30	23	41004
31	2	0	31	24	0.5591
32	2	0	32	24	0.5591
33	2	1	33	26	0.71406
34	2	1	34	32	0.8666

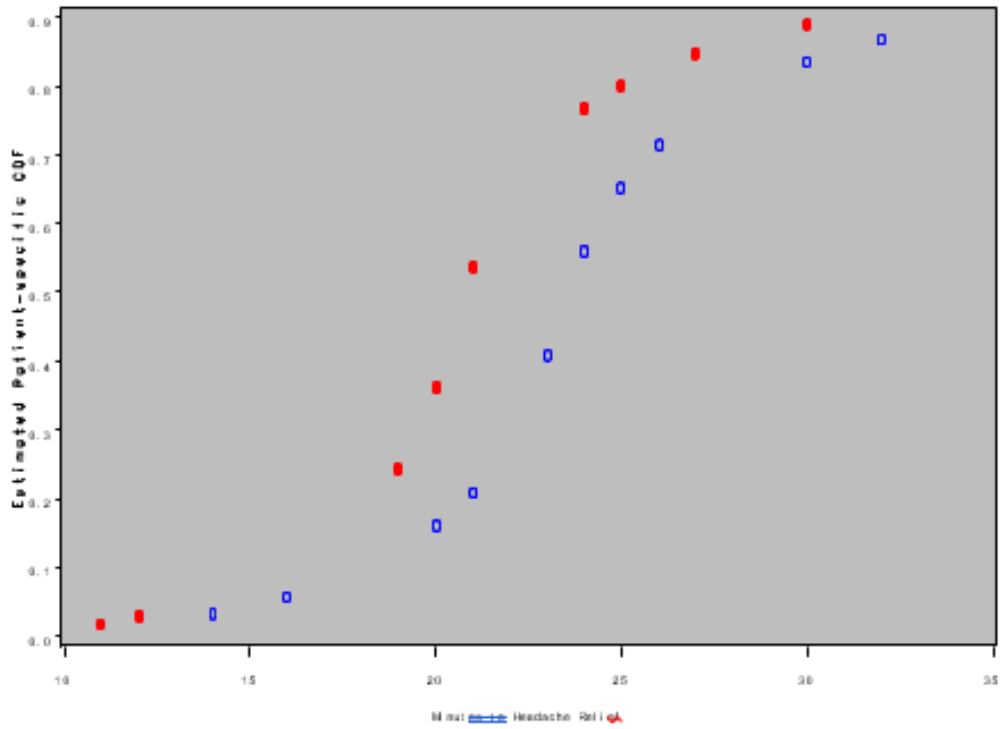| 35	2	1	35	30	0.83619
36	2	0	36	30	0.83619
37	2	1	37	32	0.8666
38	2	1	38	20	0.15989

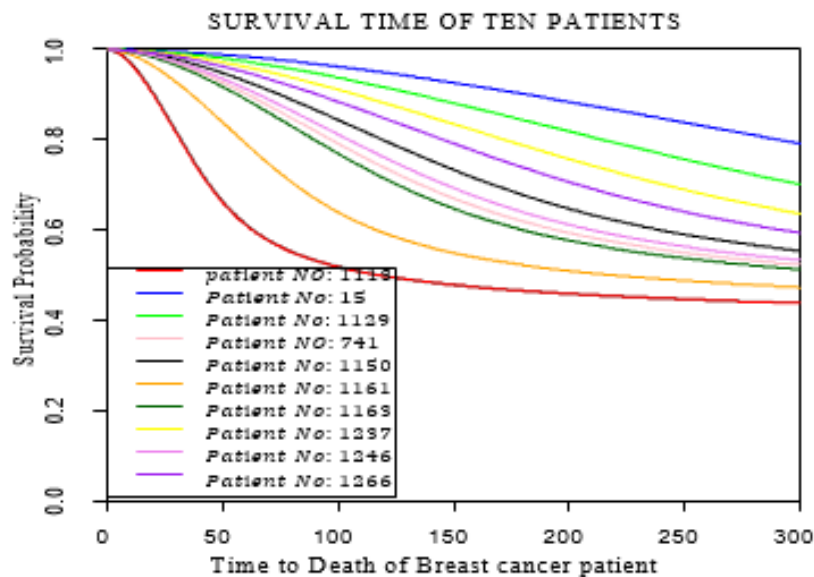**Figure 3:** The plot of predicted values of Time to headache relief.

Figure 4: The survival probability of the breast cancer patient after mastectomy.

Predicting Time-to-Death of Breast Cancer Patient

The study cohort comprises 1207 patients with cancer treated by mastectomy. The data consist of the random response variable given by the number of months (y_i) after mastectomy. Uncensored observations correspond to patients having death time computed. Censored observations correspond to patients who were not observed to have died at the time the data were collected. The numbers of censored and uncensored observations are 1135 and 72, respectively, of the total of 1207 patients. The following explanatory variables were associated with each patient (for $i = 1, ... , 1207$):

- δ_i: is the event indicator where 1 represents the event (death) and 0 is censored;

- *pathcat*: is the size of the pathologic tumor

- *est*: is the estrogen receptor status (0=negative, 1=positive, unknown= 2)

- *pr*: is the progesterone receptor status (negative =0, positive = 1, unknown = 2)

- *patholo*: is the tumor size category in centimeter (0, <=2, 2-5, >5)

- *lni*: is the lymph node involvement (0 =no, 1= yes)

- *hg*: is the histologic grade (grade 1, grade 2, grade 3, unknown = 4)

- *inpos*: is the positive axillary lymph node.

Now, we present the result by fitting the model

$$y_i = b_o + b_1 pats_i + b_2 est_i + b_3 prt_1 + b_4 tsct_i + b_5 \ln it_i + b_6 hgt_i + b_7 pa \ln t_i + \sigma z_i,$$

where the dependent variable *yi* follows the LGGBIII density function (4) for $i = 1, ... , 1207$. The MLEs of the model parameters was calculated using routine NLMIXED in SAS package. Iterative maximization of the logarithm of the likelihood function starts with initial values for and taken from the fit of the L regression model with $p = 1$. Tables 5 & 6 lists the MLEs of the parameters for the LGGBIII and LZBD regression models fitted to the current data. The LR statistic for testing the hypothesis that $H_0 : p = 1$ versus $H_1: H_0$ is not true, i.e., to compare the LGGBIII and LZBD regression models, is $w = 2\{542.5 - 524.8\} = 17.70$ (p-value < 0.0001), which gave favorable indications toward to the LGGBIII model. A comparison of the new model with its sub-model using AIC, AICc, and BIC criteria was performed in Table 7. From the values of these statistics, LGGBIII distribution provides a good fit for these data. The fitted LGGBIII regression model indicates that not all explanatory variables are significant at 5%. [13] proposed a very useful regression model for analyzing censoring failure times, where the random variable of interest represents failure time and the failures times are assumed identically distributed in some specified form. He noted that if the proportional hazards assumption holds (or, is assumed to hold) then it is possible to estimate the selected parameter(s) without any consideration of the hazard function (non-parametric approach). This approach to survival data is called proportional haz-

ards model. The Cox model may be specialized if a reason exists to assume that the baseline hazard follows a parametric form. In this case, the baseline hazard can be replaced by a parametric density. Typically, we can then maximize the full likelihood which greatly simplifies model fitting and provides interpretability at the cost of flexibility. From the result of the regression model, it was observed that not all the explanatory variables are significant that is size of the pathologic tumor, positive auxiliary lymph node, and the tumor size category in centimeter are significant at 5% and they are used to predict the survival time of the patients. Based on these three significant variables, MLEs for this new model were presented [14,15].

Table 5: MLEs of the parameters for the LGGBII and LZBD regression models fitted to the Breast Cancer data.

θ	LGGBIII Regression Model				LZBD Regression Model			
	Estimate	SE	p-value 95% CI	Estimate	SE	p-value	95%CI
p	0.258	0.488	0.597 (−0.699,1.215)	1	-	-	-
α	1.905	0.849	0.025 (0.240,3.571)	1.436	0.711	0.044	(0.041,2.831)
β	0.833	1.251	0.505 (−1.620,3.287)	1.494	1.282	0.244	(−1.022,4.009)
σ	0.217	0.501	0.666 (−0.766,1.199)	0.969	0.33	0.003	(0.323,1.616)
$\beta 0$	5.601	0.675	< .0001 (4.276,6.926)	5.041	0.814	< .0001	(3.443,6.639)
$\beta 1$	-0.521	0.105	< .0001 (−0.726, −0.315)	−0.504	-	< .0001	-
$\beta 2$	-0.063	0.029	0.035 (−0.116, −0.004)	−0.093	0.029	0.002	(−0.151, −0.035)
$\beta 3$	-0.08	0.09	0.375 (−0.257,0.097)	−0.090	0.089	0.315	(−0.265,0.085)
$\beta 4$	-0.042	0.189	0.823 (−0.413,0.328)	0.031	0.196	0.872	(−0.353, −0.416)
$\beta 5$	0.153	0.18	0.397 (−0.201,0.507)	0.168	0.188	0.372	(−0.201,0.537)
$\beta 6$	0.518	0.104	< .0001 (0.314,0.722)	0.093	0.03	0.002	(−0.035, −0.151)
$\beta 7$	-0.211	0.248	0.395 (−0.698,0.276)	−0.145	0.224	0.517	(−0.585,0.294)

Table 6: MLEs of the parameters for the Cox regression models fitted to Breast Cancer data.

Parameter	Estimate	SE	P-value	95%CI
Q1	0.764	0.145	< .0001	(−0.726, −0.315)
Q2	0.081	0.029	0.005	(−0.116, −0.004)
Q3	0.092	0.137	0.501	(−0.257,0.097)
Q4	0.011	0.294	0.969	(−0.413,0.328)
Q5	−0.256	0.281	0.362	(−0.201,0.507)
Q6	−0.758	0.144	< .0001	(0.314,0.722)
Q7	0.342	0.294	1.354	(−0.698,0.276)

Table 7: AIC, BIC and AICC statistics for comparing the LGGBIII, LZBD and Cox models.

MODEL	-2Log-Likelihood	AIC	AICC	BIC
LGGBIII	524.8	548.8	549	609.9
LZB-D	542.5	564.5	564.8	620.6
COX	868.8	882.8	–	-

$$y_i = b_o + b_1 * patholo + b_2 * inpos + b_6 * pathcat + \sigma z_i$$

For a breast cancer patient treated by mastectomy with explanatory vector x, the survival probability was estimated, say $P(T > t; \alpha, \beta, p, \sigma, Q, \boldsymbol{X}) = S(t; \alpha, \beta, p, \sigma, Q, \boldsymbol{X})$. If the survival time of Patient 1113 was calculated a follows, see Table 8:

$$S(t) = 1 - \frac{\left(\left[-\log\left[1 - (1 + (\frac{t}{\lambda})^{-\delta})^{-\beta}\right]\right]^p, \alpha\right)}{\Gamma(\alpha)}$$

Table 8: Survival for randomly selected patients' probability.

Patients	Patholo	Inpos	Status	Pathcat
15	0.15	0	0	99
742	2	1	0	1
1113	99	20	1	99
1129	0.7	0	0	1
1150	1.2	4	1	1
1161	1.8	9	1	1
1163	2	2	0	1
1237	1.1	0	1	1
1236	3	0	0	2
1266	2.4	0	0	2

Table 9: The Significant value of Breast Cancer data LGGBIII Regression Model.

θ	Estimate	Standard.E	P-value
α	1.499	0.44	–
p	0.156	0.06	–
β	1.734	0.88	–
σ	0.226	0.15	–
$b0$	5.766	0.311	< 0.0001
$b1$	−0.554	0.093	< 0.0001
$b2$	−0.078	0.019	< 0.0001
$b6$	0.552	0.093	< 0.0001

From Table 9, we have $\alpha = 1.499$, $\beta = 1.734$, $p = 0.156$, $\delta = 1/\delta = 4.425$,

$$\log\lambda = \mu = X^T\beta = 5.766 - 99 \times 0.554 - 0.078 \times 20 + 99 \times 0.552$$

$$= 4.008, then \ \lambda = \exp(4.008) = 55.036.$$

So, the survival probability of this patient say $t = 200 \ months$ by inserting the parameters in equation above we have

$$P(T \geq 200) = 0.4599784$$

That is the patient 1113 has chance of 46% of surviving past these months. Evidently, the survival probability converges to zero

when the linear predictor $\mu_i = \exp(XTQ)$ tends to $-\infty$ and converges to one when the linear predictor goes to $+\infty$. In other words, the death of patients with breast cancer treated by mastectomy for a fixing time t after the surgery, approaches one (zero) when the linear predictor increases to a very large negative (positive) number. We consider ten hypothetical patients who underwent mastectomy having fixed values for the explanatory variables given below

Conclusion

A new class of generalized Burr III distribution called the generalized gamma-Burr III distribution was proposed and studied. The GGBIII distribution has the family of Zografos and Balakrishnan distribution as special cases. The density of this new class of distributions was expressed as a linear combination of Burr III density functions. The GGBIII distribution was established to possess hazard function with flexible behavior. We also obtained closed form expressions for the moments, mean and median deviations, and distribution of order statistics. Maximum likelihood estimation technique was used to estimate the model parameters. To further the previous work, in this work, we built on the log transformation of GGBIII distribution proposed in our previous work. The main motivation was to predict the survival probability of breast cancer patients after surgery called mastectomy. The name of the log transformation is Log-Generalized Gamma Burr III (LGGBIII) model. We compared the proposed log-transform model with existing models such as Log-Zografos-Balakrishnan model Log- Weibull model and Cox model. Before the prediction, it has been established in the previous work that GBBIII distribution fitted better than its competitors. This work only confirms our curiosity in predicting time – to – death of breast cancer patients. We randomly selected 10 patients and calculated their survival probability which was also visualized in the figure above. Additionally, we analyzed two groups of patients with headaches. Also, predicted the time – to – headache relief as shown in the table. In general, the proposed LGGBIII model has higher predictive power compared to its competitors as established by different goodness of fit tests.

Acknowledgement

References

1. Olobatuyi K I, Asiribo O E, Dawodu G A (2018) A New Class of Generalized Burr III Distribution for Lifetime Data. International Journal of Statistical Distributions and Applications. Vol. 4(1): 6-21.

2. Smith R, Bain L (1975) An Exponential Power Life-Testing Distribution. Communications in Statistics-Theory and Methods 4: 469-481.

3. Hjorth U (1980) A reliability distribution with increasing, decreasing, constant and bathtub failure rates. Human and Ecological Risk Assessment Technometrics 22: 99-107.

4. Rajarshi S M B (1988) Bathtub distributions. Communications in Statistics, Theory and Methods 17: 2597-2521.

5. Mudholkar G S, Srivastava D K, Friemer M (1995) The exponentiated Weibull family: A reanalysis of the bus-motor-failure data. Technometrics 37(4): 436-445.

6. Xie M, Lai C D (1996) Reliability analysis using an additive Weibull model with bathtub shaped failure rate function. Reliability Engineering and System Safety 52(1): 87-93.

7. Lai C D, Xie M, Murthy D N (2003) A modified Weibull distribution. Transactions on Reliability, 52: 33-37 Lawless, J. F. 2003. Statistical Models and Methods for Lifetime data. Wiley, New York.

8. Famoye F, Lee C, Olumolade O (2005) The beta-Weibull distribution. Journal of Statistical Theory and Applications 4: 121-136.

9. Lee C, Famoye F, Olumolade O (2007) Beta-Weibull distribution: some properties and applications to censored data. Journal of Modern Applied Statistical Methods 6: 173-186.

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

11. Marcelino P, Ortega EM, Cordeiro G M (2011) The Kumaraswamy generalized gamma distribution with application in survival analysis. Statistical methodology 8.5: 411-433.

12. Gradshteyn I S, Ryzhik I M (2000) Table of Integrals, Series, And Products (7th Edn.).,

13. Cox D R (1972) Regression models and life tables (with discussion). Journal of the Royal Statistical Society: Series B (Statistical Methodology) 34(2): 187-220.

14. Antonio E G, da-Silva C Q (2014) The Beta Burr III Model for Lifetime Data. 1 – 33. Brazil.

15. Cordeiro GM, Giovana O S, Ortega E M (2011) The Beta Extended Weibull Family, p. 1-29.

Assets of Publishing with us

- Global archiving of articles
- Immediate, unrestricted online access
- Rigorous Peer Review Process
- Authors Retain Copyrights
- Unique DOI for all articles

https://biomedres.us/