Journal of Statistical Theory and Applications
Volume 11, Number 3, 2012, pp. 265-291 ISSN 1538-7887

The Kumaraswamy-Log-Logistic Distribution

Tiago Viana Flor de Santana[*], Edwin M. M. Ortega[†], Gauss M. Cordeiro[‡]
and Giovana O. Silva[§]

Abstract

The log-logistic distribution is widely used in survival analysis when the failure rate function presents a unimodal shape. Based on the log-logistic and Kumaraswamy distributions, we introduce the called Kumaraswamy-log-logistic distribution. The new distribution contains several important distributions discussed in the literature as sub-models such as the log-logistic, exponentiated log-logistic and Burr XII distributions, among several others. The beauty and importance of the new distribution lies in its ability to model non-monotonic failure rate functions, which are quite common in lifetime data analysis and reliability. Some of its structural properties are studied. We propose the Kumaraswamy-logistic regression model which has, as sub-models, various widely-known regression models. We discuss the method of maximum likelihood to estimate the model parameters and determine the observed information matrix. Two real data sets illustrate the importance and flexibility of the proposed models.

Keywords: Censored data; Log-logistic distribution; Maximum likelihood estimation; Moment; Order statistic; Regression model.

[*]Department of Exact Sciences, University of São Paulo, Av. Pádua Dias 11, 13418-900 – Piracicaba, SP, Brazil (tiagodesantana@yahoo.com.br)

[†]Department of Exact Sciences, University of São Paulo, Av. Pádua Dias 11, 13418-900 – Piracicaba, SP, Brazil (edwin@esalq.usp.br)

[‡]Department of Statistics, University Federal of Pernambuco, Cidade Universitária, 50740-540 – Recife, PE, Brazil (gausscordeiro@uol.com.br)

[§]Department of Statistics, University Federal of Bahia, Av. Ademar de Barros s/n, 40170-110 – Salvador, BA, Brazil (giovana@ufba.br)

1 Introduction

The log-logistic distribution is widely used in practice and it is an alternative to the log-normal distribution since it presents a failure rate function that increases, reaches a peak after some finite period and then declines gradually. The properties of the log-logistic distribution make it an attractive alternative to the log-normal and Weibull distributions in the analysis of survival data (Collet, 2003). This distribution can exhibit a monotonically decreasing failure rate function for some parameter values. It shares some properties of the log-normal and normal distributions (Ahmad *et al.*, 1988), i.e., if T has a log-logistic distribution, then $Y = \log(T)$ has a logistic distribution. Some applications of the log-logistic distribution are discussed in economy to model the wealth and income (Kleiber and Kotz, 2003) and in hydrology to model stream flow data (Ashkar and Mahdi, 2006). Collet (2003) suggested the log-logistic distribution for modeling the time following a heart transplantation.

The distribution introduced by Kumaraswamy (1980), also refereed to as the “minimax” distribution, is not very common among statisticians and has been little explored in the literature, nor its relative interchangeability with the beta distribution has been widely appreciated. We use the term “Kum” distribution to denote the Kumaraswamy distribution. Its cumulative distribution function (cdf) has a simple form

$$F_{Kum}(x) = 1 - (1 - x^a)^{b-1}, \quad 0 < x < 1, \tag{1}$$

where $a > 0$ and $b > 0$. The density function corresponding to (1) is

$$f_{Kum}(x) = a\,b\,x^{a-1}\,(1 - x^a)^{b-1}, \quad 0 < x < 1, \tag{2}$$

which can be unimodal, increasing, decreasing or constant, depending on the parameter values. Jones (2009) advocated the Kum distribution as a generator instead of the beta generator, since its quantile function takes a simple form. He explored the background and genesis of the Kum distribution and, more importantly, made clear some similarities and differences between the beta and Kum distributions. Jones (2009) highlighted several advantages of the Kum distribution over the beta distribution: the normalizing constant is very simple, simple explicit formulae for the distribution and quantile functions which do not involve any special functions and a simple formula for random variate generation. However, the beta distribution has the following advantages over the Kum distribution: simpler formulae for moments and moment generating function (mgf), a one-parameter sub-family of symmetric distributions, simpler moment estimation and more ways of generating the distribution by means of physical processes.

If G denotes the cdf of a random variable, Cordeiro and de Castro (2011) defined the Kum-G distribution by

$$F(t) = 1 - [1 - G^a(t)]^b, \quad t > 0, \tag{3}$$

where $a > 0$ and $b > 0$ are two additional shape parameters to the G distribution, whose role is to govern skewness and tail weights. The probability density function (pdf) corresponding to (3) is

$$f(t) = a\, b\, g(t)\, G(t)^{a-1}\, [1 - G^a(t)]^{b-1}, \tag{4}$$

where $g(t) = dG(t)/dt$. Equation (4) does not involve any special function, such as the incomplete beta function, as is the case of the β-G distribution proposed by Eugene *et al.* (2002). So, the Kum-G distribution is obtained by adding two shape parameters a and b to the G distribution. The generalization (4) contains distributions with unimodal and bathtub shaped hazard rate functions. It also contemplates a broad class of models with monotonic hazard rate function. Clearly, the Kum density function (2) is a basic exemplar of (4) for $G(t) = t$. A physical interpretation of the Kum-G distribution given by (3) and (4) (for a and b positive integers) is as follows. Consider a system is formed by b independent components and that each component is made up of a independent subcomponents. Suppose the system fails if any of the b components fails and that each component fails if all of the a subcomponents fail. Let $T_{j1}, \ldots, T_{ja}$ denote the lifetimes of the subcomponents within the jth component, $j = 1, \ldots, b$, having a common cdf $G(t)$. Let T_j denote the lifetime of the jth component, for $j = 1, \ldots, b$, and let T denote the lifetime of the entire system. Then, the cdf of T is

$$\begin{aligned}
\Pr(T \le t) &= 1 - \Pr\left(T_1 > t, \ldots, T_b > t\right) = 1 - \Pr^b\left(T_1 > t\right) \\
&= 1 - \left\{1 - \Pr\left(T_1 \le t\right)\right\}^b = 1 - \left\{1 - \Pr\left(T_{11} \le t, \ldots, T_{1a} \le t\right)\right\}^b \\
&= 1 - \left\{1 - \Pr^a\left(T_{11} \le t\right)\right\}^b = 1 - \left\{1 - G^a(t)\right\}^b .
\end{aligned}$$

So, it follows that the Kum-G distribution given by (3) and (4) is precisely the time to failure distribution of the entire system.

In this note, we combine the works of Kumaraswamy (1980) and Cordeiro and de Castro (2011) to study the mathematical properties of a new four-parameter distribution refereed to as the Kumaraswamy-log-logistic (KumLL) distribution. The new model contains a large number of sub-models such as the log-logistic, exponentiated log-logistic, Burr XII, among others. The new distribution is very suitable for testing goodness of fit of these sub-models and for defining a widely regression model. We hope that this extension will attract wider applications in reliability, medicine and other areas of research.

The rest of the article is organized as follows. In Section 2, we define the new distribution. Some of its properties are presented in Section 3 including general expansions for the moments and mgf. Mean deviations and Bonferroni and Lorenz curves are addressed in Section 4. Maximum likelihood estimation of the model parameters is investigated in Section 5. In Section 6, we define another new model called the Kum-logistic (KumL) distribution and derive formal expansions for its moments and mgf. Further, we define the KumL regression model and use the method of maximum likelihood to estimate the model parameters. Two real lifetime data sets are analyzed in Section 7 to illustrate the usefulness of the KumLL and KumL models. Finally, concluding remarks are addressed in Section 8.

2 The Kumaraswamy-log-logistic distribution

The cdf and pdf of the log-logistic (LL) distribution are (for $t > 0$)

$$G_{\alpha,\gamma}(t) = 1 - \left[1 + \left(\frac{t}{\alpha}\right)^{\gamma}\right]^{-1} \quad \text{and} \quad g_{\alpha,\gamma}(t) = \frac{\gamma}{\alpha^{\gamma}}\, t^{\gamma-1} \left[1 + \left(\frac{t}{\alpha}\right)^{\gamma}\right]^{-2}, \tag{5}$$

respectively, where $\alpha > 0$ is scale parameter and $\gamma > 0$ is a shape parameter. Basic properties of the log-logistic distribution are given, for example, by Kleiber and Kotz (2003), Lawless (2003) and Ashkar and Mahdi (2006). The moments are easily derived as (Tadikamalla, 1980)

$$E(T^r) = \alpha^r\, B(1 - r\gamma^{-1}, 1 + r\gamma^{-1}) = \frac{r\,\pi\,\alpha^r\,\gamma^{-1}}{\sin(r\,\pi\,\gamma^{-1})}, \quad r < \gamma,$$

where $B(a,b) = \int_0^1 \omega^{a-1}\,(1-\omega)^{b-1} d\omega$ is the beta function. Hence,

$$E(T) = \frac{\pi\,\alpha\,\gamma^{-1}}{\sin(\pi\,\gamma^{-1})}, \quad \gamma > 1 \quad \text{and} \quad \mathrm{Var}(T) = \frac{2\,\pi\,\alpha^2\,\gamma^{-1}}{\sin(2\,\pi\,\gamma^{-1})} - \left[\frac{\pi\,\alpha\,\gamma^{-1}}{\sin(\pi\,\gamma^{-1})}\right]^2, \quad \gamma > 2.$$

By substituting (5) in (4), the KumLL density function with four parameters ($a > 0$, $b > 0$, $\alpha > 0$ and $\gamma > 0$) can be defined by

$$f(t) = \frac{a\,b\,\gamma}{\alpha^{a\gamma}}\, t^{a\gamma-1} \left[1 + \left(\frac{t}{\alpha}\right)^{\gamma}\right]^{-(a+1)} \left\{1 - \left[1 - \frac{1}{1 + (\frac{t}{\alpha})^{\gamma}}\right]^{a}\right\}^{b-1}, \quad t > 0, \tag{6}$$

where α is a scale parameter and the shape parameters a, b and γ govern the skewness of (6). The KumLL density is straightforward to compute using any statistical software with numerical facilities. The new distribution due to its flexibility in accommodating non-monotonic failure rate function may be an important distribution that can be used in a variety of problems in modeling survival data. The log-logistic and exponentiated log-logistic distributions are clearly the most important sub-models of (6) for $a = b = 1$ and $b = 1$, respectively. If $a = 1$, we obtain the Burr XII distribution. If T is a random variable with density function (6), we write $T \sim$KumLL(a, b, α, γ).

The survival and hazard rate functions corresponding to (6) are

$$S(t) = 1 - F(t) = \left\{1 - \left[1 - \frac{1}{1 + (\frac{t}{\alpha})^{\gamma}}\right]^{a}\right\}^{b} \tag{7}$$

and

$$h(t) = \frac{a\,b\,\gamma}{\alpha^{a\gamma}}\; t^{a\gamma-1}\left[1+\left(\frac{t}{\alpha}\right)\right]^{-(a+1)}\left\{1-\left[1-\frac{1}{1+(\frac{t}{\alpha})^{\gamma}}\right]^{a}\right\}^{-1},$$

respectively.

Plots of the KumLL density function for some parameter values are displayed in Figure 1. A characteristic of the proposed distribution is that its failure rate function accommodates increasing, decreasing, unimodal and bathtub shaped forms, that depend basically on the values of the shape parameters. Moreover, it is quite flexible for modeling survival data.

Alternatively, other works had introduced new distributions for modeling bathtub shaped failure rate. For example, the exponentiated Weibull (EW) distribution introduced by Mudholkar et al. (1995, 1996), the additive Weibull distribution presented by Xie and Lai (1995), the extended Weibull distribution (Xie et al. 2002), the modified Weibull (MW) distribution proposed by Lai et al. (2003), the beta exponential (BE) distribution introduced by Nadarajah and Kotz (2006), the extended flexible Weibull distribution defined by Bebbington et al. (2007), the beta Weibull (BW) distribution studied by Lee et al. (2007), the generalized modified Weibull (GMW) distribution proposed by Carrasco et al. (2008) and, more recently, the beta modified Weibull (BMW) distribution investigated by Silva et al. (2010).

Plots of the KumLL hazard rate function for selected parameter values are given in Figure 2.

The KumLL quantile function is determined by inverting $F(t)$ in (7)

$$t = Q(u) = \alpha\left\{\left[1-\{1-(1-u)^{1/b}\}^{1/a}\right]^{-1/\gamma}-1\right\}. \tag{8}$$

Thus, the new distribution is easily simulated as $T = Q(U)$, where U has the uniform $U(0,1)$ distribution. This compares extremely favorably with the sophisticated algorithms preferred to simulate random variates from the beta generated distributions.

3 Basic Properties

We provide simple expansions for the KumLL density function depending on whether the parameter b (or a) is real non-integer or integer. Further, we derive infinite sums for its moments and mgf.

3.1 Expansions for the KumLL density function

Using the binomial expansion in (4), the KumLL density function can be expressed as

$$f(t) = g_{\alpha,\gamma}(t)\sum_{i=0}^{\infty} w_i\, G_{\alpha,\gamma}(t)^{a(i+1)-1} \tag{9}$$

(a) (b) (c)

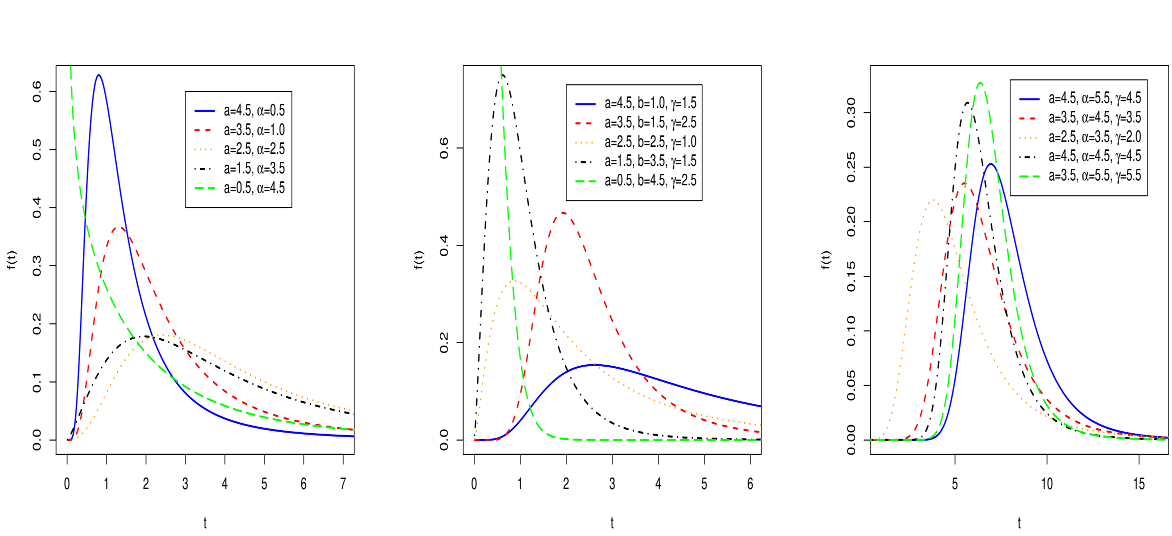

Figure 1: The KumLL density function for some parameter values. (a) For different values of a and α with $b = 1.5$ and $\gamma = 1.5$. (b) For different values of a, b and γ with $\alpha = 1.5$. (c) For different values of a, α γ with $b = 1.5$.

(a) (b) (c)

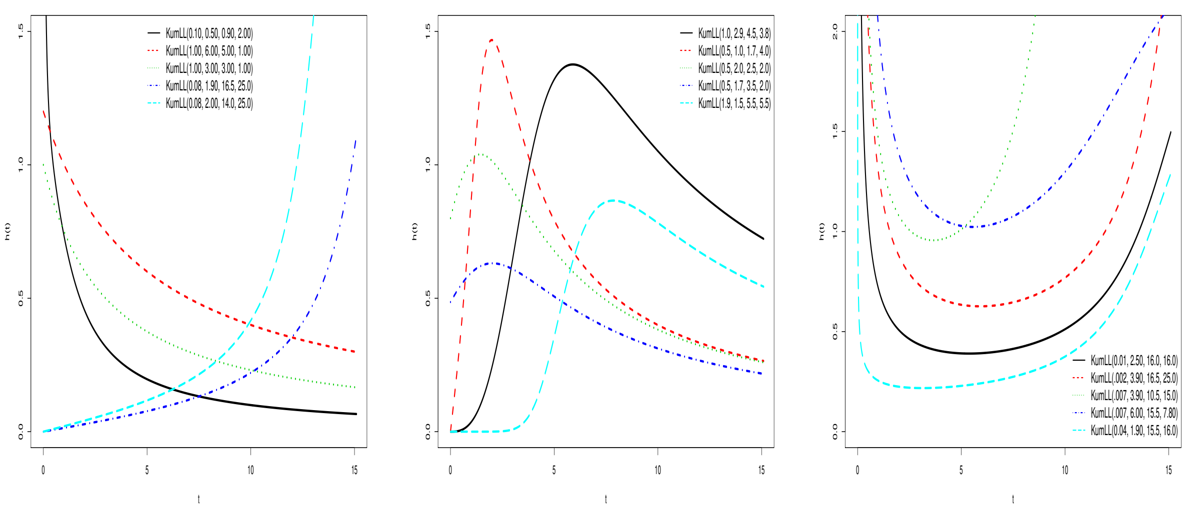

Figure 2: The KumLL hazard rate function. (a) Increasing and decreasing shapes. (b) Unimodal shaped. (c) Bathtub shaped.

where $w_i = w_i(a, b) = (-1)^i \, a \, b \binom{b-1}{i}$. If b is an integer, the index i in the previous sum stops at $b - 1$. If a is an integer, equation (9) reveals that the KumLL density function equals the

LL density function multiplied by an infinite weighted power series of $G_{\alpha,\gamma}(t)$. Otherwise, if a is real non-integer, we can use the expansion

$$G(t)^a = \sum_{r=0}^{\infty} s_r(a)\, G_{\alpha,\gamma}(t)^r, \tag{10}$$

where

$$s_r(a) = \sum_{j=r}^{\infty} (-1)^{r+j} \binom{a}{j} \binom{j}{r}. \tag{11}$$

So, we can write

$$f(t) = g_{\alpha,\gamma}(t) \sum_{r=0}^{\infty} t_r\, G_{\alpha,\gamma}(t)^r, \tag{12}$$

where the coefficients are $t_r = t_r(a,b) = \sum_{i=0}^{\infty} w_i\, s_r(a(i+1)-1)$ and the quantity $s_r(a(i+1)-1)$ can be calculated from (11).

3.2 Moments and Generating function

The probability weighted moments (PWMs) of a random variable T with cdf $G(t)$ are defined by $\tau_{k,n} = E[T^k\, G(T)^n]$ for k and n positive integers. Here, the PWMs of the LL distribution are used to compute the ordinary moments of the KumLL distribution.

Theorem 1: If $T \sim$LL(α, γ), for $k < \gamma$, we obtain

$$\begin{aligned}\tau_{k,n} &= \frac{\gamma}{\alpha^\gamma} \int_0^\infty t^{k+\gamma-1} \left[1 + \left(\frac{t}{\alpha}\right)^\gamma\right]^{-2} \left\{1 - \left[1 + \left(\frac{t}{\alpha}\right)^\gamma\right]^{-1}\right\}^n dt \\ &= \alpha^k \int_0^1 w^{-k\gamma^{-1}} (1-w)^{n+k\gamma^{-1}} dw = \alpha^k\, B(n+1+k\,\gamma^{-1}, 1-k\,\gamma^{-1}).\end{aligned}$$

Theorem 2: If $T \sim$KumLL(a, b, α, γ), for $k < \gamma$, the kth moment of T can be expressed as:

- For $a > 0$ integer,

$$\mu'_k = \alpha^k \sum_{i=0}^{\infty} w_i\, B(a(i+1) + k\,\gamma^{-1}, 1 - k\gamma^{-1}). \tag{13}$$

- For $a > 0$ real non-integer,

$$\mu'_k = \alpha^k \sum_{r=0}^{\infty} t_r\, B(r+1+k\,\gamma^{-1}, 1-k\,\gamma^{-1}). \tag{14}$$

The proof is given in Appendix A. The skewness and kurtosis measures can be calculated from the ordinary moments using well-known relationships. Plots of the skewness and kurtosis for some choices of the parameter b as function of a, and for some choices of the parameter a as function of b, are displayed in Figures 3 and 4, respectively. These plots immediately reveal that the skewness and kurtosis depend on both parameters a and b.

The mgf of T can be expressed as

$$M(-s) = \int_0^\infty \exp(-st)\, f(t)dt = \sum_{k=0}^\infty \frac{(-s)^k}{k!} \int_0^\infty t^k f(t)dt = \sum_{k=0}^\infty \frac{\mu'_k\,(-s)^k}{k!}.$$

Now, we provide an alternative representation for $M(-s)$. Let $A(t) = (t\alpha^{-1})^\gamma$. We can write

$$M(-s) = \frac{a\,b\,\gamma}{\alpha^{a\gamma}} \int_0^\infty \exp(-st)\, t^{a\gamma-1}\, [1 + A(t)]^{-(a+1)} \{1 - [1 - [1 + A(t)]^{-1}]^a\}^{b-1} dt$$

Since $0 < [1 + A(t)]^{-1} < 1$, we obtain

$$\{1 - [1 - [1 + A(t)]^{-1}]^a\}^{b-1} = \sum_{i,j=0}^\infty (-1)^{i+j} \binom{b-1}{i} \binom{ai}{j} [1 + A(t)]^{-j}$$

and then

$$M(-s) = \frac{a\,b\,\gamma}{\alpha^{a\gamma}} \sum_{i,j=0}^\infty (-1)^{i+j} \binom{b-1}{i} \binom{ai}{j} \int_0^\infty \exp(-st)\, t^{a\gamma-1}\, [1 + A(t)]^{-j} dt,$$

where the integral $I_j(s)$ reduces to

$$I_j(s) = \alpha^{a\gamma} \int_0^\infty \frac{x^{a\gamma-1}\, \exp(-s\alpha x)}{(1 + x^\gamma)^j}\, dx.$$

The quantity $I_j(s)$ can be calculated provided that $\gamma = p/q$, where p and q are co-primes natural numbers. In this case, using integral (2.3.1) in Prudnikov *et al.* (1986), we obtain

$$I_j(s) = \alpha^{a\gamma}\,(P + Q).$$

Here,

$$P = \sum_{h=0}^{p-1} \frac{(-s\alpha)^h}{h!\gamma}\, B(a_1, 1 - \rho - a_1)\ {}_{q+1}F_{q+p}(1, \Delta(q, a_1); \Delta(p, 1 + h), \Delta(q, b_1); z),$$

$$Q = \sum_{k=1}^{q-1} \frac{(-1)^k\, j}{(k-1)!}\, (s\alpha)^{v_1}\, \Gamma(a\gamma - 2k\gamma)\ {}_{q+1}F_{q+p}(1, \Delta(q, j + k); \Delta(p, b_2), \Delta(q, 1 + k); z),$$

where $v_1 = \gamma[j+k-a]$, $a_1 = (a\gamma+h)/\gamma$, $b_1 = a_1+1-j$, $b_2 = 1+\gamma(j+k-a)$, $z = (-1)^{p+q}\left(\frac{s\alpha}{q}\right)^p$,

$\Delta(k,a) = (a/k, (a+1)/k, \cdots, (a+k-1)/k)$,

$$_pF_q(a_1, a_2, \cdots, a_p; b_1, b_2, \cdots, b_q; z) = \sum_{k=0}^{\infty} \frac{(a_1)_k (a_2)_k \ldots (a_p)_k z^k}{(b_1)_k (b_2)_k \ldots (b_p)_k k!}$$

is the generalized hypergeometric function and $(f)_k = f(f+1)\ldots(f+k-1)$ denotes the ascending factorial. The condition $\gamma = p/q$ is not restrictive since every real number can be approximated by a rational number.

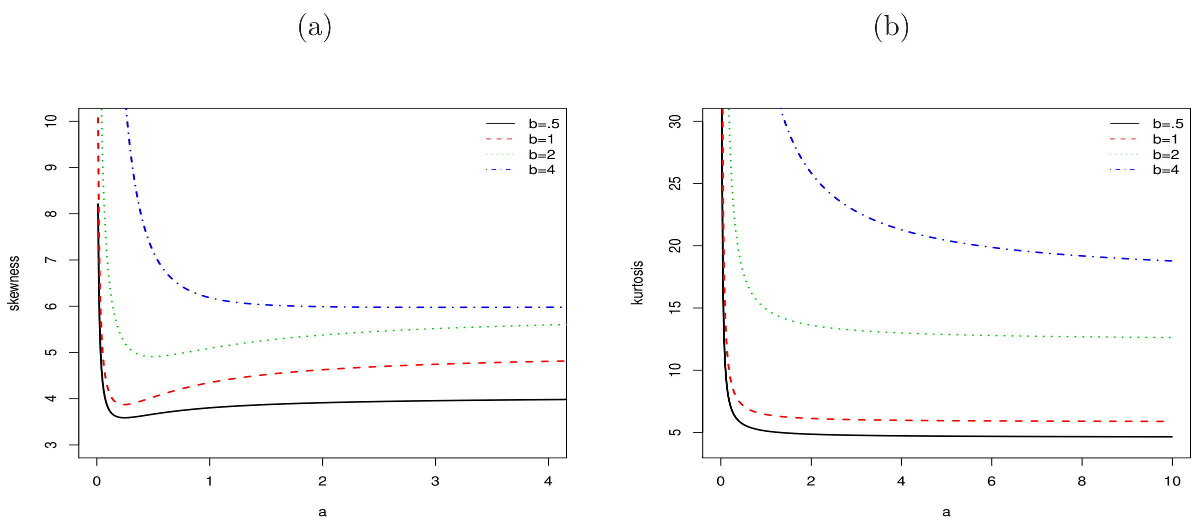

Figure 3: Skewness and kurtosis of the KumLL distribution as a function of a for some values of b and $\alpha = 2$ and $\gamma = 1$.

A second alternative expansion for the mgf follows by expanding the binomial terms in the quantile function (8). We have

$$Q(u) = \alpha \left(\sum_{k=0}^{\infty} q_k \, u^k - 1 \right),$$

where

$$q_k = (-1)^k \binom{j/b}{k} \sum_{i,j=0}^{\infty} (-1)^{i+j} \binom{-\gamma^{-1}}{i} \binom{i/a}{j}.$$

Hence, the mgf of T reduces to

$$M(s) = \int_0^1 \exp\left\{ s\,\alpha \left(\sum_{k=0}^{\infty} q_k \, u^k - 1 \right) \right\} du = \sum_{m=0}^{\infty} \frac{s^m \, \alpha^m}{m!} \int_0^1 \left(\sum_{k=0}^{\infty} q_k \, u^k - 1 \right)^m du.$$

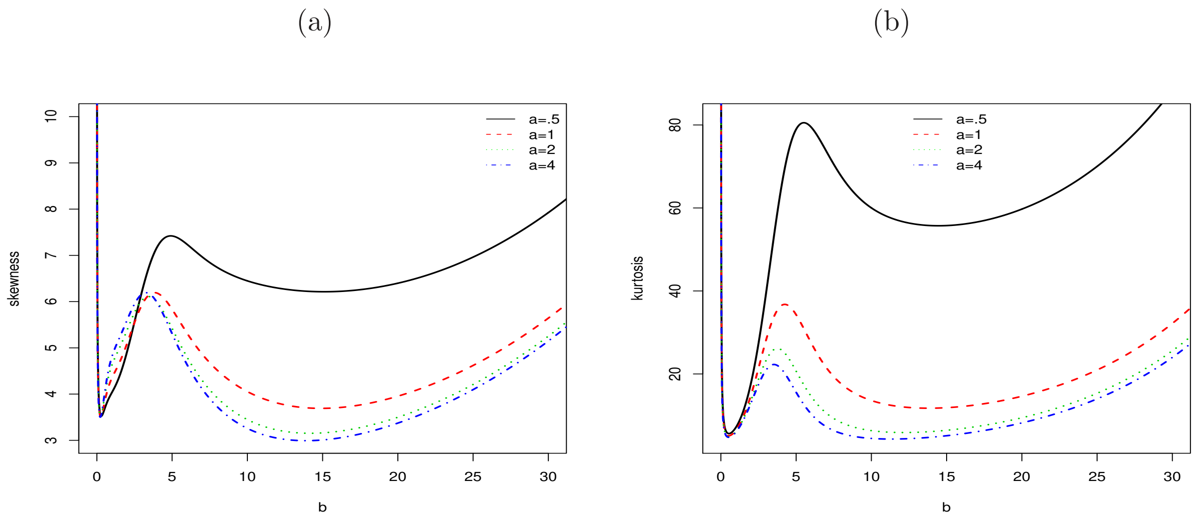

Figure 4: Skewness and kurtosis of the KumLL distribution as a function of b for some values of a and $\alpha = \gamma = 2$.

We have

$$M(s) = \sum_{m=0}^{\infty} \sum_{r=0}^{m} \binom{m}{r} \frac{(-1)^{m-r} s^m \alpha^m}{m!} \int_0^1 \left(\sum_{k=0}^{\infty} q_k u^k \right)^r du.$$

We use an equation in Section 0.314 of Gradshteyn and Ryzhik (2000) for a power series raised to a positive integer j given by

$$\left(\sum_{k=0}^{\infty} q_k u^k \right)^r = \sum_{k=0}^{\infty} c_{r,k} u^k, \tag{15}$$

where the coefficients $c_{r,k}$ can be obtained from the recurrence equation (for $k = 1, 2, \ldots$)

$$c_{r,k} = (k q_0)^{-1} \sum_{t=1}^{k} [t(r+1) - k] q_t c_{r,k-t} \tag{16}$$

and $c_{r,0} = q_0^r$. Hence, the coefficients $c_{r,k}$ are directly determined from $c_{r,0}, \ldots, c_{r,k-1}$ and, then, from the quantities $q_0, \ldots, q_k$. They can be given explicitly in terms of the q_k's, although it is not necessary for programming numerically our expansions in any algebraic or numerical software. Hence, we obtain

$$M(s) = \sum_{m=0}^{\infty} b_m \frac{s^m}{m!}, \tag{17}$$

where

$$b_m = \alpha^m \sum_{k=0}^{\infty} \sum_{r=0}^{m} \binom{m}{r} \frac{(-1)^{m-r}\, c_{r,k}}{(k+1)}.$$

Clearly, the moments of T follow by simple differentiation of (17) as $\mu'_m = E(T^m) = b_m$.

4 Other Measures

Here, we compute the means deviations, Bonferroni and Lorenz curves. 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 KumLL distribution with density function (6), we can derive the mean deviations about the mean $\mu'_1 = E(T)$ (which follows from Theorem 2) and about the median m from the relations

$$\delta_1 = 2\mu'_1\, F(\mu'_1) - 2T(\mu'_1) \qquad \text{and} \qquad \delta_2 = \mu'_1 - 2T(m),$$

where $T(z) = \int_0^z t\, f(t)$. The median is $m = \alpha \left\{ \left[1 - \{1 - 2^{-1/b}\}^{1/a}\right]^{-1/\gamma} - 1 \right\}$. From equation (4) and by integration by parts, we have

$$T(z) = a\, b\, \gamma\, \alpha^{-1} \sum_{k=0}^{\infty} (-1)^k \binom{b-1}{k} \int_0^z t \left(\frac{t}{\alpha}\right)^{-\gamma-1} \left[1 + \left(\frac{t}{\alpha}\right)^{-\gamma}\right]^{-1-a(k+1)} dt.$$

Using Theorem 1, we obtain

$$T(z) = a\, b\, \alpha \sum_{k=0}^{\infty} (-1)^k \binom{b-1}{k} B_{G_{\alpha,\gamma}(z)} \left(a(k+1) + \gamma^{-1}, 1 - \gamma^{-1}\right), \tag{18}$$

where $B_w(a,b) = \int_0^w x^{a-1}\,(1-x)^{b-1} dx$ is the incomplete beta ratio function and $\gamma > 1$. From equation (18), we can obtain δ_1 and δ_2.

An application of $T(z)$ refers to the Bonferroni and Lorenz curves. These curves have applications in fields like reliability, demography, economics, insurance and medicine. They are defined by

$$B(\pi) = \frac{T(q)}{\pi\, \mu'_1} \qquad \text{and} \qquad L(\pi) = \frac{T(q)}{\mu'_1},$$

respectively, where the quantile function $q = Q(\pi)$ is determined from (8) to yield q for a given probability π.

5 Maximum Likelihood Estimation

Let T_i be a random variable following (6) with the vector of parameters $\boldsymbol{\theta} = (a, b, \alpha, \gamma)^T$. The data encountered in survival analysis and reliability studies are often censored. A very simple random censoring mechanism that is often realistic is one in which each individual i is assumed to have a lifetime T_i and a censoring time C_i, where T_i and C_i are independent random variables. Suppose that the data consist of n independent observations $t_i = \min(T_i, C_i)$ for $i = 1, \ldots, n$. The distribution of C_i does not depend on any of the unknown parameters of T_i. Parametric inference for such data are usually based on likelihood methods and their asymptotic theory. The censored log-likelihood $l(\boldsymbol{\theta})$ for the model parameters is

$$\begin{aligned} l(\boldsymbol{\theta}) &= r\log\left(\frac{a\,b\,\gamma}{\alpha^{a\gamma}}\right) + (a\gamma-1)\sum_{i\in F}\log(t_i) - (a+1)\sum_{i\in F}\log\left[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}\right] + \\ & (b-1)\sum_{i\in F}\log\left\{1-\left[1-\frac{1}{1+(\frac{t_i}{\alpha})^{\gamma}}\right]^{a}\right\} + b\sum_{i\in C}\log\left\{1-\left[1-\frac{1}{1+(\frac{t_i}{\alpha})^{\gamma}}\right]^{a}\right\}, \quad (19) \end{aligned}$$

where r is the number of failures and F and C denote the uncensored and censored sets of observations, respectively. The score functions corresponding to the components in $\boldsymbol{\theta}$ are

$$\begin{aligned} U_a(\boldsymbol{\theta}) &= \frac{r}{a}[1-a\gamma\log(\alpha)] + \gamma\sum_{i\in F}\log(t_i) - \sum_{i\in F}\log\left[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}\right] - (b-1)\sum_{i\in F}\frac{u_i^a\log(u_i)}{(1-u_i^a)} \\ & -b\sum_{i\in C}\frac{u_i^a\log(u_i)}{(1-u_i^a)}, \end{aligned}$$

$$U_b(\boldsymbol{\theta}) = \frac{r}{b} + \sum_{i\in F}\log(1-u_i^a) + \sum_{i\in C}\log(1-u_i^a),$$

$$\begin{aligned} U_\alpha(\boldsymbol{\theta}) &= -\frac{r\,a\,\gamma}{\alpha} + \frac{(a+1)\gamma}{\alpha}\sum_{i\in F}\frac{(\frac{t_i}{\alpha})^{\gamma}}{(1+\frac{t_i}{\alpha})^{\gamma}} + \frac{(b-1)a\gamma}{\alpha}\sum_{i\in F}\frac{u_i^a(\frac{t_i}{\alpha})^{\gamma}}{[1+(\frac{t_i}{\alpha})^{\gamma}]^2u_i(1-u_i^a)} \\ & +\frac{b\,a\,\gamma}{\alpha}\sum_{i\in C}\frac{u_i^a(\frac{t_i}{\alpha})^{\gamma}}{[1+(\frac{t_i}{\alpha})^{\gamma}]^2u_i(1-u_i^a)} \end{aligned}$$

and

$$\begin{aligned} U_\gamma(\boldsymbol{\theta}) &= \frac{r}{\gamma}[1-a\gamma\log(\alpha)] + a\sum_{i\in F}\log(t_i) - (a+1)\sum_{i\in F}\frac{(\frac{t_i}{\alpha})^{\gamma}\log(\frac{t_i}{\alpha})}{1+(\frac{t_i}{\alpha})^{\gamma}} \\ & -(b-1)a\sum_{i\in F}\frac{u_i^a(\frac{t_i}{\alpha})^{\gamma}\log(\frac{t_i}{\alpha})}{[1+(\frac{t_i}{\alpha})^{\gamma}]^2u_i(1-u_i^a)} - ba\sum_{i\in C}\frac{u_i^a(\frac{t_i}{\alpha})^{\gamma}\log(\frac{t_i}{\alpha})}{[1+(\frac{t_i}{\alpha})^{\gamma}]^2u_i(1-u_i^a)}, \end{aligned}$$

where $u_i = 1 - [1 + (\frac{t_i}{\alpha})^{\gamma}]^{-1}$.

For interval estimation and hypothesis tests on the model parameters, we require the 4×4 unit observed information matrix

$$J = J(\boldsymbol{\theta}) = \begin{pmatrix} J_{aa} & J_{ab} & J_{a\alpha} & J_{a\gamma} \\ & J_{bb} & J_{b\alpha} & J_{b\gamma} \\ & & J_{\alpha\alpha} & J_{\alpha\gamma} \\ & & & J_{\gamma\gamma} \end{pmatrix},$$

where the elements are defined in Appendix B. Consequently, the MLE $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ can be obtained numerically from the nonlinear equations $U_a(\boldsymbol{\theta}) = 0$, $U_b(\boldsymbol{\theta}) = 0$, $U_\alpha(\boldsymbol{\theta}) = 0$ and $U_\gamma(\boldsymbol{\theta}) = 0$.

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 $N_4(0, I(\boldsymbol{\theta})^{-1})$, where $I(\boldsymbol{\theta})$ is the unit expected information matrix. This approximated distribution holds when $I(\boldsymbol{\theta})$ is replaced by $J(\widehat{\boldsymbol{\theta}})$, i.e., the observed information matrix evaluated at $\widehat{\boldsymbol{\theta}}$. The multivariate normal $N_4(0, J(\widehat{\boldsymbol{\theta}})^{-1})$ distribution for $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ can be used to construct approximate confidence intervals for the individual parameters and for the hazard and survival functions. In fact, an asymptotic confidence interval with significance level γ for each parameter θ_r is given by

$$ACI(\theta_r, 100(1-\gamma)\%) = \left(\hat{\theta}_r - z_{\gamma/2}\sqrt{\hat{\bar{J}}^{\theta_r,\theta_r}}, \hat{\theta}_r + z_{\gamma/2}\sqrt{\hat{\bar{J}}^{\theta_r,\theta_r}}\right),$$

where $\hat{\bar{J}}^{r,r}$ is the estimated rth diagonal element of $J(\widehat{\boldsymbol{\theta}})^{-1}$ for $r = 1, \ldots, 4$ and $z_{\gamma/2}$ is the quantile $1 - \gamma/2$ of the standard normal distribution.

The likelihood ratio (LR) statistic is useful for testing the goodness-of-fit of the KumLL distribution and for comparing it with some of its sub-models (see Section 2). We can compute the maximum values of the unrestricted and restricted log-likelihoods to construct LR statistics. For example, we may use the LR statistic to check if the fit using the KumLL distribution is statistically "superior" to the fits using the exponentiated-log-logistic and log-logistic distributions for a given data set. In any case, hypothesis tests of the type $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus $H : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, where $\boldsymbol{\theta}_0$ is a specified vector, can be performed using LR statistics. For example, the test of $H_0 : b = 1$ versus $H : H_0$ *not true* is equivalent to compare the exponentiated-log-logistic and KumLL distributions, for which the LR statistic is $w = 2\big[\ell(\widehat{a}, \widehat{b}, \widehat{\alpha}, \widehat{\gamma}) - \ell(\widetilde{a}, 1, \widetilde{\alpha}, \widetilde{\gamma})\big]$, where $\widehat{a}$, $\widehat{b}$, $\widehat{\alpha}$ and $\widehat{\gamma}$ are the MLEs under H and $\widetilde{a}$, $\widetilde{\alpha}$ and $\widetilde{\gamma}$ are the estimates under H_0.

6 The Kumaraswamy-Logistic Regression Model

In many practical applications, lifetimes are affected by variables, which are referred to as explanatory variables or covariates, such as the cholesterol level, blood pressure and many others. So it is important to explore the relationship between the lifetime and explanatory variables. An approach based on a regression model can be used. The vector of explanatory variables is denoted by $\mathbf{x} = (x_1, \ldots, x_p)^T$, which is related to response variable $Y = \log(T)$ through a regression model. We require the distribution of $Y = \log(T)$ referred to as the Kum-logistic (KumL) distribution.

6.1 The KumL distribution

Let T be a random variable having the KumLL density function (6). The random variable $Y = \log(T)$ has a KumL density function, parameterized in terms of $\gamma = \sigma^{-1}$ and $\alpha = \exp(\mu)$, given by

$$f(y) = \frac{a\,b}{\sigma} \exp\left[a\Big(\frac{y-\mu}{\sigma}\Big)\right]\left[1+\exp\Big(\frac{y-\mu}{\sigma}\Big)\right]^{-(a+1)}\left\{1-\left[1-\frac{1}{1+\exp(\frac{y-\mu}{\sigma})}\right]^{a}\right\}^{b-1}, \quad (20)$$

where $-\infty < y,\ \mu < \infty$, $a > 0$, $b > 0$ and $\sigma > 0$. A random variable Y following (20) is denoted by $Y \sim$KumL(a, b, μ, σ), where μ is the location parameter, σ is the dispersion parameter and a and b are shape parameters. Thus,

$$\text{if} \quad T \sim \text{KumLL}(a, b, \alpha, \gamma) \quad \text{then} \quad Y = \log(T) \sim \text{KumL}(a, b, \mu, \sigma).$$

An important characteristic of this distribution is that it contains some important sub-models. It becomes the logistic distribution when $a = b = 1$. If $b = 1$, it gives the exponentiated logistic (exp-logistic) distribution. If $\sigma = \gamma^{-1}$, $a = \alpha$ and $x = y + \mu$, it leads to the generalized logistic distribution (Alkasasbeh and Raqab, 2009). Further, if $a = 2$ and $b = 1$, it reduces to the skew logistic distribution, see, for example, Nadarajah (2009).

The plots of the density function (20) for selected parameter values are given in Figure 5. These plots show great flexibility of the new parameters a and b, which can be useful in several practical situations. The associated survival function is

$$S(y) = \left\{1-\left[1-\frac{1}{1-\exp(\frac{y-\mu}{\sigma})}\right]^{a}\right\}^{b}.$$

The random variable $Z = (Y - \mu)/\sigma$ has density function

$$\pi(z) = a\,b\,\exp(az)\,[1+\exp(z)]^{-(a+1)}\left\{1-\left[1-\frac{1}{1+\exp(z)}\right]^{a}\right\}^{b-1}, \; -\infty < z < \infty. \quad (21)$$

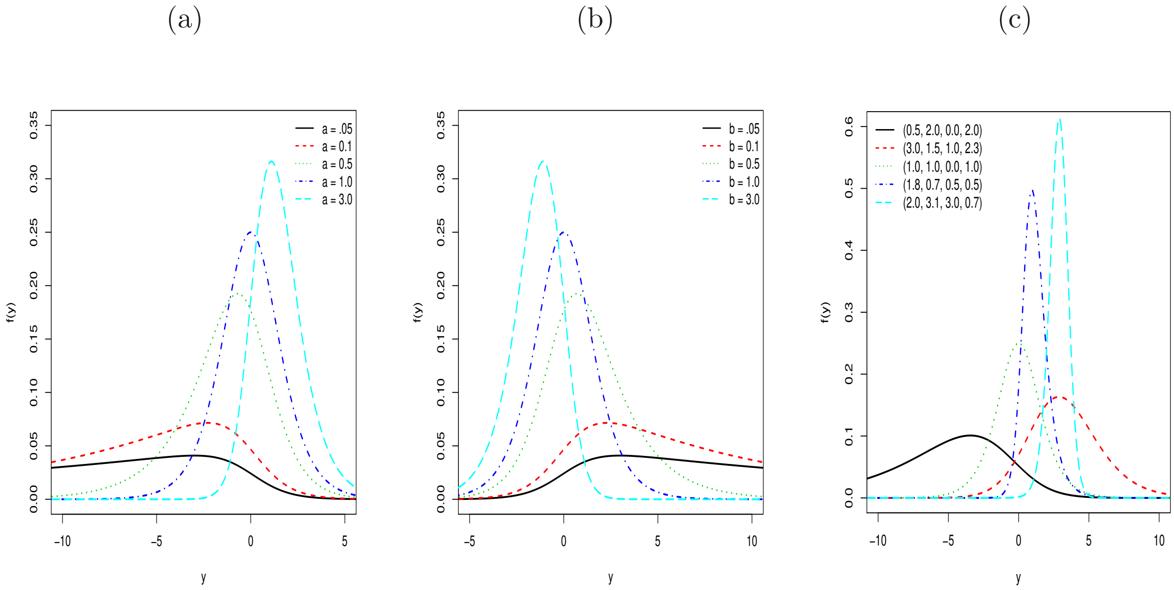

Figure 5: Plots of the KumL density function for some parameter values. (a) For $b = 1$, $\mu = 0$ and $\sigma = 1$. (b) For $a = 1$, $\mu = 0$ and $\sigma = 1$. (c) For values different of a, b, μ and σ.

Now, we can expand the binomial term in (21) to obtain

$$\left\{1 - \sum_{i=0}^{\infty}(-1)^i \binom{a}{i} \{1 + \exp(z)\}^{-i}\right\}^{b-1} = \sum_{j=0}^{\infty}(-1)^j \binom{b-1}{j} \left\{\sum_{i=0}^{\infty} f_i \, [1 + \exp(z)]^{-i}\right\}^j,$$

where $f_i = (-1)^i \binom{a}{i}$. Then, using (15) and (16), we can obtain

$$\pi(z) = a\, b \sum_{i,j=0}^{\infty} (-1)^j \binom{b-1}{j} d_{j,i} \exp(az)\, [1 + \exp(z)]^{-(a+i+1)}. \quad (22)$$

where the quantities $d_{j,i}$ can be followed from the recurrence equation (for $j = 1, 2, \ldots$)

$$d_{j,i} = (i\, f_0)^{-1} \sum_{t=1}^{i} [(j+1)\, t - i]\, f_t\, d_{j,i-t}.$$

The kth moment of Z is given by

$$E(Z^k) = a\, b \sum_{i,j=0}^{\infty} (-1)^j \binom{b-1}{j} d_{j,i} \int_{-\infty}^{\infty} z^k \exp(az)\, [1 + \exp(z)]^{-(a+i+1)} dz.$$

Setting $x = \exp(z)$, we obtain

$$E(Z^k) = a\,b \sum_{i,j=0}^{\infty} (-1)^j \binom{b-1}{j} d_{j,i} \int_0^{\infty} x^{a-1}\,(1+x)^{-(a+i+1)}\,\log^k(x)dx.$$

Using equation (2.6.4.6) in Prudnikov et al.(1986), we calculate the integral as

$$\int_0^{\infty} x^{a-1}\,(1+x)^{-(a+i+1)}\,\log^k(x)dx = \left(\frac{\partial}{\partial a}\right)^k [B(a, i+1)]$$

and then

$$E(Z^k) = a\,b \sum_{i,j=0}^{\infty} (-1)^j \binom{b-1}{j} d_{j,i} \left(\frac{\partial}{\partial a}\right)^k [B(a, i+1)]\,. \tag{23}$$

The skewness and kurtosis measures can be calculated from the ordinary moments (23) using well-known relationships. Plots of these measures for some choices of b as function of a, and for some choices of a as function of b, with $\mu = 1$ and $\sigma = 2$, are displayed in Figures 6 and 7, respectively.

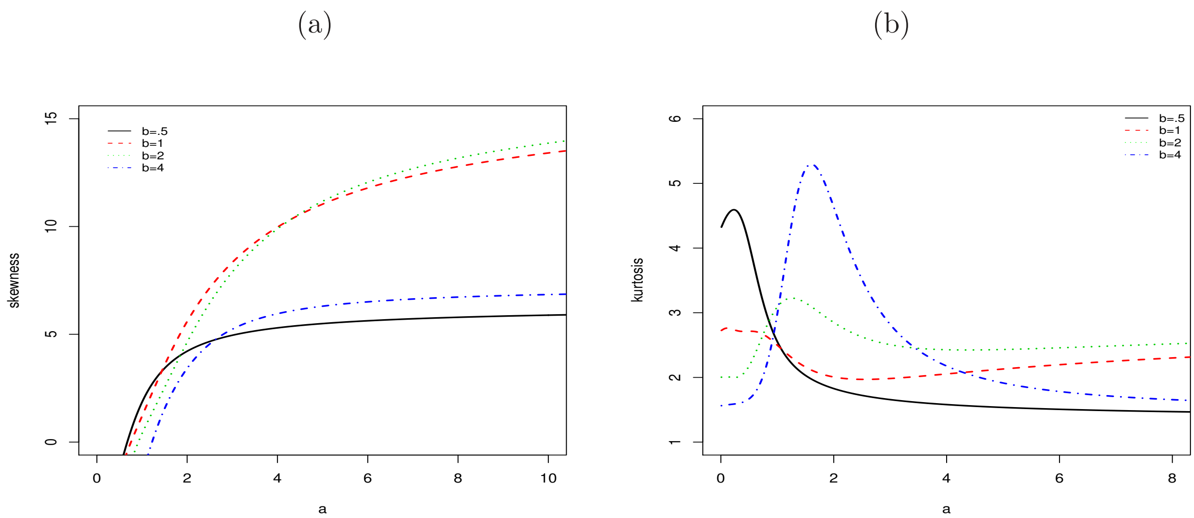

Figure 6: Skewness and kurtosis of the KumL distribution as a function of a for some values of b.

6.2 Model and Estimation

Let $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^T$ be the explanatory variable vector associated with the ith response variable y_i for $i = 1, \ldots, n$. Consider a sample $(y_1, \mathbf{x}_1), \ldots, (y_n, \mathbf{x}_n)$ of n independent observations, where each random response is defined by $y_i = \min\{\log(t_i), \log(c_i)\}$. We assume non-

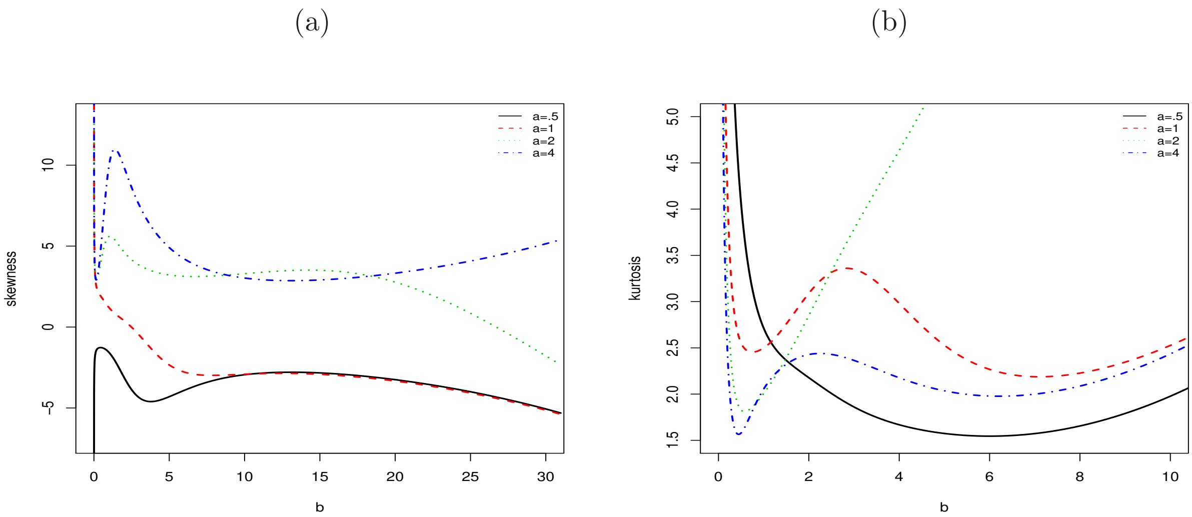

Figure 7: Skewness and kurtosis of the KumL distribution as a function of b for some values of a.

informative censoring and that the observed lifetimes and censoring times are independent. We propose a log-linear regression model for the response variable y_i based on the KumL density function given by

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \sigma z_i, \;\; i = 1, \ldots, n, \tag{24}$$

where the random error z_i follows the distribution (21), $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$, $\sigma > 0$, $a > 0$ and $b > 0$ are unknown scale parameters and $\mathbf{x}_i$ is the explanatory variable vector modeling the location parameter $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$. Hence, the location parameter vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^T$ of the KumL model has a linear structure $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, where $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$ is a known model matrix. The logistic regression model can be defined by equation (24) with $a = b = 1$.

Let F and C be the sets of individuals for which y_i is the log-lifetime or log-censoring, respectively. The total log-likelihood function for the model parameters $\boldsymbol{\theta} = (a, b, \sigma, \boldsymbol{\beta}^T)^T$ can be expressed from equations (21) and (24) as

$$\begin{aligned} l(\boldsymbol{\theta}) \;\; = \;\; & r \log\left(\frac{a\, b}{\sigma}\right) + a \sum_{i \in F} z_i - (a+1) \sum_{i \in F} \log[1 + \exp(z_i)] \\ & + (b-1) \sum_{i \in F} \log\left\{1 - \left[1 - \frac{1}{1+\exp(z_i)}\right]^a\right\} + b \sum_{i \in C} \log\left\{1 - \left[1 - \frac{1}{1-\exp(z_i)}\right]^a\right\} \end{aligned} \tag{25}$$

where $z_i = (y_i - \mathbf{x}_i^T \boldsymbol{\beta})/\sigma$ and r is the observed number of failures. The maximum likelihood estimate (MLE) $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ can be obtained by maximizing the log-likelihood function (25) using the procedure NLMixed in SAS. From the fitted model (24), the survival function for y_i can be estimated by

$$S(y_i;\hat{a},\hat{b},\hat{\sigma},\widehat{\boldsymbol{\beta}}^T)=\left\{1-\left[1-\frac{1}{1-\exp\left(\frac{y_i-\mathbf{x}^T\widehat{\boldsymbol{\beta}}}{\hat{\sigma}}\right)}\right]^{\hat{a}}\right\}^{\hat{b}} \tag{26}$$

or

$$S(t_i;\hat{a},\hat{b},\hat{\alpha},\hat{\gamma})=\left\{1-\left[1-\frac{1}{1+(\frac{t_i}{\hat{\alpha}})^{\hat{\gamma}}}\right]^{\hat{a}}\right\}^{\hat{b}}, \tag{27}$$

where $\hat{\gamma}=\hat{\sigma}^{-1}$ and $\hat{\alpha}=\exp(\mathbf{x}^T\widehat{\boldsymbol{\beta}})$.

Under general regularity conditions, the asymptotic distribution of $\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})$ is multivariate normal $N_{p+3}(0,K(\boldsymbol{\theta})^{-1})$, where $K(\boldsymbol{\theta})$ is the expected information matrix. The asymptotic covariance matrix $K(\boldsymbol{\theta})^{-1}$ of $\widehat{\boldsymbol{\theta}}$ can be approximated by the inverse of the $(p+3)\times(p+3)$ observed information matrix $J(\boldsymbol{\theta})$ and then the asymptotic inference for the parameter vector $\boldsymbol{\theta}$ can be based on the normal approximation $N_{p+3}(0,J(\boldsymbol{\theta})^{-1})$ for $\widehat{\boldsymbol{\theta}}$. The observed information matrix is

$$J(\boldsymbol{\theta})=\begin{pmatrix} J_{aa} & J_{ab} & J_{a\sigma} & J_{a\beta_j} \\ . & J_{bb} & J_{b\sigma} & J_{b\beta_j} \\ . & . & J_{\sigma\sigma} & J_{\sigma\beta_j} \\ . & . & . & J_{\beta_j\beta_s} \end{pmatrix},$$

whose elements can be computed numerically.

The asymptotic multivariate normal $N_{p+3}(0,J(\boldsymbol{\theta})^{-1})$ distribution can be used to construct approximate confidence regions for some parameters in $\boldsymbol{\theta}$. In fact, an $100(1-\alpha)\%$ asymptotic confidence interval for each parameter θ_r is given by

$$ACI_r=\left(\hat{\theta}_r-z_{\alpha/2}\sqrt{\widehat{\ddot{J}}^{r,r}},\hat{\theta}_r+z_{\alpha/2}\sqrt{\widehat{\ddot{J}}^{r,r}}\right),$$

where $\widehat{\ddot{J}}^{r,r}$ denotes the rth diagonal element of the inverse of the estimated observed information matrix $J(\widehat{\boldsymbol{\theta}})^{-1}$ and $z_{\alpha/2}$ is the quantile $1-\alpha/2$ of the standard normal distribution.

The interpretation of the estimated coefficients could be based on the ratio of median times (see Hosmer and Lemeshow, 1999) which holds for continuous or categorical explanatory variables. When the explanatory variable is binary (0 or 1), and considering the ratio of median times with $x=1$ in the numerator, if $\hat{\beta}$ is negative (positive), it implies that the individuals with $x=1$ present reduced (increased) median survival time in $[\exp(\hat{\beta})\times 100\%]$ as compared to those individuals in the group with $x=0$, assuming that the other explanatory variables are fixed.

We are also interested to investigate if the KumL model is a good model to fit the data under investigation. Clearly, the LR statistic can be used to discriminate between the exponentiated logistic and KumL models since they are nested models. In this case, the hypotheses to be tested are $H_0 : b = 1$ versus $H_1 : b \neq 1$ and the LR statistic reduces to $w = 2\{l(\widehat{\boldsymbol{\theta}}) - l(\widetilde{\boldsymbol{\theta}})\}$, where $\widetilde{\boldsymbol{\theta}}$ is the MLE of $\boldsymbol{\theta}$ under H_0. The null hypothesis is rejected if $w > \chi_1^2(1-\alpha)$, where $\chi_1^2(1-\alpha)$ is the quantile of the chi-square distribution with one degree of freedom.

7 Applications

We provide two applications to real data with right censored to demonstrate the usefulness of the KumLL distribution. We consider an application with no covariates to AIDS data reported by Silva (2004) and other application to malanoma data (Ibrahim *et al.*, 2001) using the new regression model.

7.1 AIDS data

Aids is a pathology that mobilizes the sufferers because of the implications for their interpersonal relationships and reproduction. Therapeutic advances have enabled seropositive women to bear children safely. In this respect, the pediatric immunology outpatient service and social service of the Hospital das Clínicas have a special program for care of newborns of seropositive mothers, to provide orientation and support for antiretroviral therapy to allow these women and their babies to live as normally as possible. We analyze a data set on the time to serum reversal of 143 children exposed to HIV by vertical transmission, born at Hospital das Clínicas (associated with the Ribeirão Preto School of Medicine) from 1995 to 2001, where the mothers were not treated (Silva, 2004; Perdoná, 2006). Vertical HIV transmission can occur during gestation in around 35% of cases, during labor and birth itself in some 65% of cases, or during breast feeding, varying from 7% to 22% of cases. Serum reversal or serological reversal can occur in children of HIV-contaminated mothers. It is the process by which HIV antibodies disappear from the blood in an individual who tested positive for HIV infection. As the months pass, the maternal antibodies are eliminated and the child ceases to be HIV positive. The exposed newborns were monitored until definition of their serological condition, after administration of Zidovudin (AZT) in the first 24 hours and for the following 6 weeks.

Table 1 displays the MLEs (and the corresponding standard errors in parentheses) of the model parameters and the values for some models of the following statistics: AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion) and CAIC (Consistent Akaike Information Criterion). The computations were done using the subroutine NLMixed in SAS. These results indicate that the KumLL 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.

The LR statistic for testing the hypotheses H_0: $a = b = 1$ versus H_1: H_0 is not true, i.e. to compare the log-logistic and KumLL models, is $w = 37.9$ (p-value <0.0001), which yields

Table 1: MLEs of the model parameters for the the AIDS data, the corresponding SE (given in parentheses) and the measures AIC, BIC and CAIC.

Modelo	a	b	α	γ	AIC	CAIC	BIC
KumLL	0.08 (0.02)	0.28 (0.07)	499.37 (29.34)	15.46 (3.78)	1616.5	1616.8	1628.4
Log-logistic	1	1	454.87 (20.58)	3.29 (0.26)	1650.4	1650.5	1656.4
			α_1	γ_2			
Weibull	-	-	537.46 (21.41)	2.35 (0.16)	1630.6	1630.7	1636.5
			μ	σ			
Log-normal	-	-	6.03 (0.05)	0.71 (0.05)	1704.1	1704.2	1710.0

favorable indications toward to the KumLL model. In order to assess if the model is appropriate, the estimated survival function for the KumLL distribution and some of its sub-models and the empirical survival function are given in Figure 8. These plots indicate that the KumLL model provides a good fit for the current data.

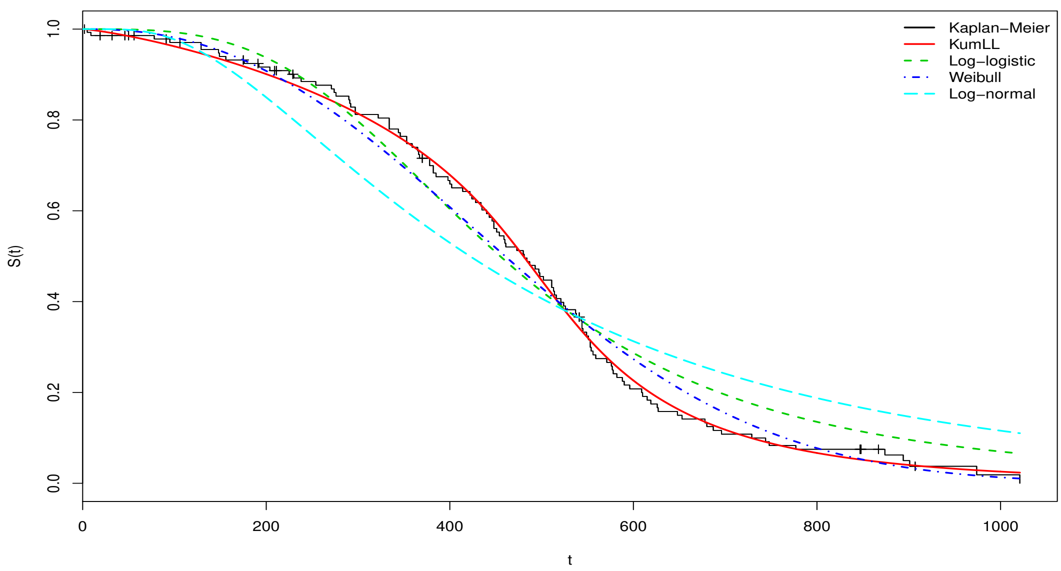

Figure 8: Estimated survival functions and the empirical survival function for AIDS data.

7.2 Melanoma Data

Here, the application of the regression model to a real data set on cancer recurrence is discussed. The data are part of a study on cutaneous melanoma (a type of malignant cancer) for the evaluation of postoperative treatment performance with a high dose of a certain drug (interferon alfa-2b) in order to prevent recurrence. Patients were included in the study from 1991 to 1995 and follow-up was conducted until 1998. These data (Ibrahim *et al.*, 2001) present the survival times, T, defined as the time until the patient's death. The original sample size was $n = 427$ patients, 10 of whom did not present a value for the explanatory variable tumor thickness. When such cases were removed, a sample of size $n = 417$ patients was retained. The percentage of censored observations was 56%. The following variables were associated with each participant, $i = 1, \ldots, 417$: y_i: observed time (in years); c_i: censoring indicator (0=censoring, 1=lifetime observed); x_{i1}: nodule (nodule category: 1 to 4). The variable nodule category from 1 to 4 is coded from the number of lymph nodes involved in the disease $(0, 1, 2-3, \text{and} \geq 4)$.

We analyze these data using the KumL regression model. First, we consider the structure

$$y_i = \beta_0 + \beta_1 x_{i1} + \sigma z_i, \qquad i = 1, \ldots, 417, \tag{28}$$

where the errors $z_1, \ldots, z_{417}$ are independent random variables having the density function (21).

Table 2 lists the MLEs of the parameters for the KumL and logistic regression models fitted to the melanoma data (using the procedure NLMixed in SAS) and the values of the AIC, BIC and CAIC statistics to compare the KumL and logistic regression models. The KumL regression model outperforms the logistic model irrespective of the criteria and can be used effectively in the analysis of these data. The new model involves two extra parameters which gives more flexibility to fit these data.

Table 2: MLEs of the parameters from the KumL regression model fitted to the melanoma data, the corresponding SEs (given in parentheses), p-values in [.] and the AIC, CAIC and BIC measures.

Model	a	b	σ	β_0	β_1	AIC	CAIC	BIC
KumL	0.53	0.11	0.26	0.87	-0.33	892.3	892.4	912.5
	(0.39)	(0.05)	(0.12)	(0.19)	(0.06)			
				[<0.0001]	[<0.0001]			
logistic	1	1	0.75	2.43	-0.37	918.5	918.7	930.7
	(-)	(-)	(0.04)	(0.19)	(0.07)			
				[<0.0001]	[<0.0001]			

The LR statistic for testing the hypotheses H_0: $a = b = 1$ versus H_1: H_0 is not true, i.e. to compare the logistic and KumL regression models, is $w = 30.30$ (p-value <0.0001) supporting the KumL model. Further, we note from the fitted KumL regression model that

x_1 is significant at 1% and that there is a significant difference between the nodule levels 1, 2, 3 and 4 for the failure times. A graphical comparison between the KumL and logistic models (see Figure 9) reveals that the larger model provides a superior fit. The plots of the estimated survival function (27) for the KumL and logistic regression models and the empirical survival function are displayed in Figure 9.

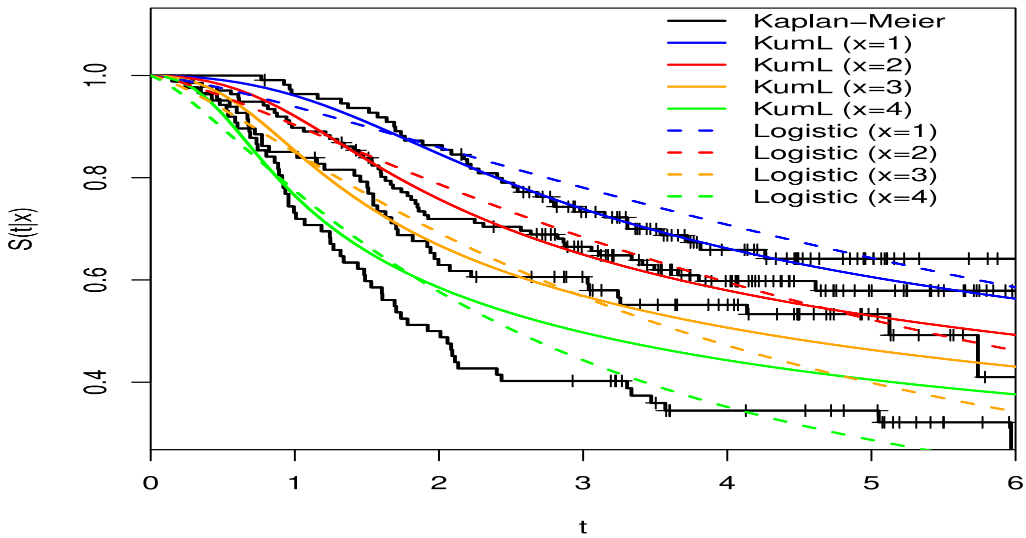

Figure 9: Estimated survival functions for the fitted KumL and logistic models and the empirical survival function for each level of the melanoma data.

8 Conclusions

A four parameter lifetime distribution called "the Kumaraswamy-log-logistic (KumLL) distribution" is proposed. It is a simple generalization of the log-logistic distribution. The new model extends several distributions widely used in the lifetime literature and it is more flexible than the exponentiated log-logistic and log-logistic distributions. The new distribution could have increasing, decreasing, bathtub and unimodal hazard rate functions. It is then very versatile to model lifetime data with a bathtub shaped hazard rate function and also to model a variety of uncertainty situations. Explicit expressions for the moments and moment generating function are provided. The application of the new distribution is straightforward. We discuss maximum likelihood estimation and hypothesis tests for the model parameters. Further, we define the called Kumaraswamy-logistic (KumL) distribution and derive an expansion for its moments. Based on this new distribution, we define the KumL regression model which is very suitable for modeling censored and uncensored lifetime data. The new regression model allows to test the goodness of fit of some known regression models as special models. Hence, the proposed regression model serves as a good alternative for lifetime data analysis. We use the procedure NLMixed in SAS to obtain the maximum likelihood estimates and perform asymptotic likelihood ratio tests for the model parameters. We demonstrate in two applications to real data that the KumLL and KumL distributions can produce better fits than their sub-models.

Appendix A: Proof of Theorem 2

For the KumLL distribution (6), the moments are

$$\mu'_k = \int_0^\infty t^k f(t)dt,\ t > 0.$$

For a integer, from (9) and (12), we obtain

$$\mu'_k = \sum_{i=0}^{\infty} w_i \int_0^\infty t^k g_{\alpha,\gamma}(t) G_{\alpha,\gamma}(t)^{a(i+1)-1} dt.$$

If a is real non-integer, we have

$$\mu'_k = \sum_{r=0}^{\infty} t_r \int_0^\infty t^k g_{\alpha,\gamma}(t) G_{\alpha,\gamma}(t)^r dt$$

Hence,

$$\mu'_k = \begin{cases} \alpha^k \sum\limits_{i=0}^{\infty} w_i\ B(a(i+1) + k\gamma^{-1}, 1 - k\gamma^{-1}), & \text{for } a \text{ integer,} \\ \alpha^k \sum\limits_{r=0}^{\infty} t_r\ B(r + 1 + k\gamma^{-1}, 1 - k\gamma^{-1}), & \text{for } a \text{ real non-integer.} \end{cases} \blacksquare$$

9 *

Appendix B

By differentiating (19), the elements of the observed information matrix $J(\boldsymbol{\theta})$ for the parameters (a, b, α, γ) are:

$$J_{aa} = -\frac{r}{a^2} - (b-1)\sum_{i\in F}[\log(u_i)]^2 \frac{u_i^a + u_i^{2a}}{[1-u_i^a]^2} - b\sum_{i\in C}[\log(u_i)]^2 \frac{u_i^a + u_i^{2a}}{[1-u_i^a]^2},$$

$$J_{ab} = -\sum_{i\in F} \frac{u_i^a \log(u_i)}{1-u_i^a} - \sum_{i\in C} \frac{u_i^a \log(u_i)}{1-u_i^a},$$

$$J_{a\alpha} = -\frac{r\gamma}{\alpha} + \frac{\gamma}{\alpha}\sum_{i\in F}\frac{\left(\frac{t_i}{\alpha}\right)^{\gamma}}{1+\left(\frac{t_i}{\alpha}\right)^{\gamma}} + \frac{(b-1)\gamma}{\alpha}\sum_{i\in F}\frac{\left(\frac{t_i}{\alpha}\right)^{\gamma}}{[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}]^2}\frac{u_i^{a-1}}{1-u_i^a}\left\{1+\frac{a\log(u_i)}{1-u_i^a}\right\} + $$
$$+ \frac{b\gamma}{\alpha}\sum_{i\in C}\frac{\left(\frac{t_i}{\alpha}\right)^{\gamma}}{[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}]^2}\frac{u_i^{a-1}}{1-u_i^a}\left\{1+\frac{a\log(u_i)}{1-u_i^a}\right\},$$

$$J_{a\gamma} = -r\log(\alpha) + \sum_{i\in F}\log(t_i) - \sum_{i\in F}\frac{\left(\frac{t_i}{\alpha}\right)^{\gamma}}{1+\left(\frac{t_i}{\alpha}\right)^{\gamma}}\log\left(\frac{t_i}{\alpha}\right) -$$
$$- (b-1)\sum_{i\in F}\frac{\left(\frac{t_i}{\alpha}\right)^{\gamma}}{[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}]^2}\frac{u_i^{a-1}}{1-u_i^a}\log\left(\frac{t_i}{\alpha}\right)\left\{1+\frac{a\log(u_i)}{1-u_i^a}\right\} -$$
$$- b\sum_{i\in C}\frac{\left(\frac{t_i}{\alpha}\right)^{\gamma}}{[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}]^2}\frac{u_i^{a-1}}{1-u_i^a}\log\left(\frac{t_i}{\alpha}\right)\left\{1+\frac{a\log(u_i)}{1-u_i^a}\right\},$$

$$J_{bb} = -\frac{r}{b^2},$$

$$J_{b\alpha} = \frac{a\gamma}{\alpha}\sum_{i\in F}\frac{\left(\frac{t_i}{\alpha}\right)^{\gamma}}{[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}]^2}\frac{u_i^{a-1}}{1-u_i^a} + \frac{a\gamma}{\alpha}\sum_{i\in C}\frac{\left(\frac{t_i}{\alpha}\right)^{\gamma}}{[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}]^2}\frac{u_i^{a-1}}{1-u_i^a},$$

$$J_{b\gamma} = -a\sum_{i\in F}\frac{\left(\frac{t_i}{\alpha}\right)^{\gamma}}{[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}]^2}\frac{u_i^{a-1}}{1-u_i^a}\log\left(\frac{t_i}{\alpha}\right) - a\sum_{i\in C}\frac{\left(\frac{t_i}{\alpha}\right)^{\gamma}}{[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}]^2}\frac{u_i^{a-1}}{1-u_i^a}\log\left(\frac{t_i}{\alpha}\right),$$

$$J_{\alpha\alpha} = \frac{ra\gamma}{\alpha^2} - \frac{(a+1)\gamma}{\alpha^2}\sum_{i\in F}\frac{\left(\frac{t_i}{\alpha}\right)^{\gamma}}{1+\left(\frac{t_i}{\alpha}\right)^{\gamma}}\left[1+\frac{\gamma}{1+\left(\frac{t_i}{\alpha}\right)^{\gamma}}\right] - \frac{a(b-1)\gamma}{\alpha^2}\sum_{i\in F}\frac{u_i^{a-1}}{1-u_i^a}\frac{\left(\frac{t_i}{\alpha}\right)^{\gamma}}{[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}]^2} -$$
$$- \frac{a(b-1)\gamma^2}{\alpha^2}\sum_{i\in F}\left\{\frac{\left(\frac{t_i}{\alpha}\right)^{2\gamma}}{[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}]^4}\frac{(a-1)u_i^{a-2}+u_i^{2a-2}}{[1-u_i^a]^2} + \frac{u_i^{a-1}}{1-u_i^a}\left(\frac{t_i}{\alpha}\right)^{\gamma}\frac{1-\left(\frac{t_i}{\alpha}\right)^{\gamma}}{[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}]^3}\right\} -$$
$$- \frac{ab\gamma}{\alpha^2}\sum_{i\in C}\frac{u_i^{a-1}}{1-u_i^a}\frac{\left(\frac{t_i}{\alpha}\right)^{\gamma}}{[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}]^2} - \frac{ab\gamma^2}{\alpha^2}\sum_{i\in C}\left\{\frac{\left(\frac{t_i}{\alpha}\right)^{2\gamma}}{[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}]^4}\frac{(a-1)u_i^{a-2}+u_i^{2a-2}}{[1-u_i^a]^2} + \right.$$
$$+ \left.\frac{u_i^{a-1}}{1-u_i^a}\left(\frac{t_i}{\alpha}\right)^{\gamma}\frac{1-\left(\frac{t_i}{\alpha}\right)^{\gamma}}{[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}]^3}\right\},$$

$$
\begin{aligned}
J_{\alpha\gamma} &= -\frac{ra}{\alpha}+\frac{a+1}{\alpha}\sum_{i\in F}\frac{\left(\frac{t_i}{\alpha}\right)^{\gamma}}{1+\left(\frac{t_i}{\alpha}\right)^{\gamma}}+\frac{(a+1)\gamma}{\alpha}\sum_{i\in F}\frac{\left(\frac{t_i}{\alpha}\right)^{\gamma}}{[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}]^2}\log\left(\frac{t_i}{\alpha}\right)+\\
&+\frac{a(b-1)}{\alpha}\sum_{i\in F}\frac{u_i^{a-1}}{1-u_i^a}\frac{\left(\frac{t_i}{\alpha}\right)^{\gamma}}{[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}]^2}+\frac{a(b-1)\gamma}{\alpha}\sum_{i\in F}\frac{\left(\frac{t_i}{\alpha}\right)^{\gamma}}{[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}]^2}\log\left(\frac{t_i}{\alpha}\right)\left[\frac{(a-1)u_i^{a-2}}{1-u_i^a}+\right.\\
&+\left.\frac{au_i^{2a-2}}{[1-u_i^a]^2}\right]+\frac{a(b-1)\gamma}{\alpha}\sum_{i\in F}\frac{u_i^{a-1}}{1-u_i^a}\log\left(\frac{t_i}{\alpha}\right)\left[\frac{\left(\frac{t_i}{\alpha}\right)^{\gamma}}{[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}]^2}-\frac{2\left(\frac{t_i}{\alpha}\right)^{\gamma}}{[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}]^3}\right]+\\
&+\frac{ab}{\alpha}\sum_{i\in F}\frac{u_i^{a-1}}{1-u_i^a}\frac{\left(\frac{t_i}{\alpha}\right)^{\gamma}}{[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}]^2}+\frac{ab\gamma}{\alpha}\sum_{i\in F}\frac{\left(\frac{t_i}{\alpha}\right)^{\gamma}}{[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}]^2}\log\left(\frac{t_i}{\alpha}\right)\left[\frac{(a-1)u_i^{a-2}}{1-u_i^a}+\right.\\
&+\left.\frac{au_i^{2a-2}}{[1-u_i^a]^2}\right]+\frac{ab\gamma}{\alpha}\sum_{i\in F}\frac{u_i^{a-1}}{1-u_i^a}\log\left(\frac{t_i}{\alpha}\right)\left[\frac{\left(\frac{t_i}{\alpha}\right)^{\gamma}}{[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}]^2}-\frac{2\left(\frac{t_i}{\alpha}\right)^{\gamma}}{[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}]^3}\right]
\end{aligned}
$$

and

$$
\begin{aligned}
J_{\gamma\gamma} &= -\frac{r}{\gamma^2}-(a+1)\sum_{i\in F}\left[\log\left(\frac{t_i}{\alpha}\right)\right]^2\frac{\left(\frac{t_i}{\alpha}\right)^{\gamma}}{[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}]^2}-\\
&- a(b-1)\sum_{i\in F}\left[\frac{\left(\frac{t_i}{\alpha}\right)^{\gamma}}{[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}]^2}\log\left(\frac{t_i}{\alpha}\right)\right]^2\left[\frac{(a-1)u_i^{a-2}}{1-u_i^a}+\frac{au_i^{2a-2}}{1-u_i^a}\right]-\\
&- a(b-1)\sum_{i\in F}\frac{u_i^{a-1}}{1-u_i^a}\left[\log\left(\frac{t_i}{\alpha}\right)\right]^2\left[\frac{\left(\frac{t_i}{\alpha}\right)^{\gamma}}{[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}]^2}-\frac{2\left(\frac{t_i}{\alpha}\right)^{2\gamma}}{[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}]^3}\right]-\\
&- ab\sum_{i\in C}\left[\frac{\left(\frac{t_i}{\alpha}\right)^{\gamma}}{[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}]^2}\log\left(\frac{t_i}{\alpha}\right)\right]^2\left[\frac{(a-1)u_i^{a-2}}{1-u_i^a}+\frac{au_i^{2a-2}}{1-u_i^a}\right]-\\
&- ab\sum_{i\in C}\frac{u_i^{a-1}}{1-u_i^a}\left[\log\left(\frac{t_i}{\alpha}\right)\right]^2\left[\frac{\left(\frac{t_i}{\alpha}\right)^{\gamma}}{[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}]^2}-\frac{2\left(\frac{t_i}{\alpha}\right)^{2\gamma}}{[1+\left(\frac{t_i}{\alpha}\right)^{\gamma}]^3}\right].
\end{aligned}
$$

References

[1] Ahmad, M. I., Sinclair, C. D. and Werritty, A. (1988). Log-Logistic Flood Frequency Analysis. *Journal of Hydrology*, **98**, 205-224.

[2] Alkasasbeh, M. R. and Raqab, M. Z. (2009). Estimation of the generalized logistic distribution parameters: Comparative study. *Statistical Methodology*, **6**, 262-279.

[3] Ashkar, F. and Mahdi, S. (2006). Fitting the log-logistic distribution by generalized moments. *Journal of Hydrology*, **328**, 694-703.

[4] Bebbington, M., Lai, C. D. and Zitikis, R. (2007). A flexible Weibull extension. *Reliability Engineering and System Safety*, **92**, 719-726.

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

[6] Collet, D. (2003). *Modelling Survival data in medical research.* Chapman and Hall: London.

[7] Cordeiro, G. M. and de Castro M. (2011). A new family of generalized distributions. *Journal of Statistical Computation and Simulation*, **81**, 883-898.

[8] Eugene, N., Lee, C. and Famoye, F. (2002). Beta-normal distribution and its applications. *Communications in Statistics, Theory and Methods*, **31**, 497-512.

[9] Gradshteyn, I.S. and Ryzhik, I. M. (2000). *Table of Integrals, Series, and Products.* Academic Press: New York.

[10] Hosmer, D. W. and Lemeshow, S. (1999). *Applied Survival Analysis.* John Wiley: New York.

[11] Ibrahim, J. G., Chen, M. H. and Sinha, D. (2001). *Bayesian survival analysis.* Springer-Verlag: New York.

[12] Jones, M. C. (2009). Kumaraswamy's distribution: A beta-type distribution with some tractability advantages. *Statistical Methodology*, **6**, 70-81.

[13] Kleiber, C. and Kotz, S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences.* John Wiley: New York.

[14] Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes. *Journal of Hydrology*, **46**, 79-88.

[15] Lai, C. D., Xie, M. and Murthy, D. N. P. (2003). A modified Weibull distribution. *Transactions on Reliability*, **52**, 33-37.

[16] Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data.* John Wiley: New York.

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

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

[19] Mudholkar, G.S., Srivastava, D. K. and Kollia, G. D. (1996). A generalization of the Weibull distribution with application to the analysis of survival data. *Journal of the American Statistical Association*, **91**, 1575-1583.

[20] Nadarajah, S. and Kotz, S. (2006). The beta exponential distribution. *Reliability Engineering and System Safety*, **91**, 689-697.

[21] Nadarajah, S. (2009). The skew logistic distribution. *Advances in Statistical Analysis*, **93**, 187-203.

[22] Perdoná, G.S.C. (2006). *Modelos de riscos aplicados à análise de sobrevivência.* Doctoral Thesis, Institute of Computer Science and Mathematics, University of São Paulo, Brazil.

[23] Prudnikov, A. P., Brychkov, Y. A. and Marichev, O. I. (1986). *Integrals and Series*, volume 1. Gordon and Breach Science Publishers: Amsterdam.

[24] Silva, A. N. F. (2004). *Estudo evolutivo das crianças expostas ao HIV e notificadas pelo núcleo de vigilância epidemiológica do HCFMRP-USP*. M.Sc. Thesis. University of São Paulo, Brazil.

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

[26] Tadikamalla, P. R. (1980). A look at the Burr and related distributions. *International Statistical Review*, **48**, 337-344.

[27] Xie, M. and Lai, C. D. (1995). Reliability analysis using an additive Weibull model with bathtub-shaped failure rate function. *Reliability Engineering and System Safety*, **52**, 87-93.

[28] Xie, M., Tang Y. and Goh, T. N. (2002). A modified Weibull extension with bathtub failure rate function. *Reliability Engineering and System Safety*, **76**, 279-285.