

The Odd Power Lindley Generator of Probability Distributions: Properties, Characterizations and Regression Modeling

Mustafa Ç. Korkmaz¹, Emrah Altun², Haitham M. Yousof³ & G.G. Hamedani⁴

¹ Department of Measurement and Evaluation, Artvin Çoruh University, Artvin, Turkey

² Department of Statistics, Bartin University, Bartin, Turkey

³ Department of Statistics, Mathematics and Insurance, Benha University, Benha, Egypt

⁴ Department of Mathematics, Statistics and Computer Science, Marquette University, USA

Correspondence: Mustafa Ç. Korkmaz, Department of Measurement and Evaluation, Artvin Çoruh University, 08000, Artvin, Turkey.

Received: October 30, 2018 Accepted: December 17, 2018 Online Published: January 25, 2019

doi:10.5539/ijsp.v8n2p70 URL: https://doi.org/10.5539/ijsp.v8n2p70

Abstract

In this study, a new flexible family of distributions is proposed with its statistical properties as well as some useful characterizations. The maximum likelihood method is used to estimate the unknown model parameters by means of two simulation studies. A new regression model is proposed based on a special member of the proposed family called, the log odd power Lindley Weibull distribution. Residual analysis is conducted to evaluate the model assumptions. Four applications to real data sets are given to demonstrate the usefulness of the proposed model.

Keywords: mixture family, power Lindley distribution, regression modeling, Q-Q plot, sensitivity analysis

1. Introduction

As of late, there has been an extraordinary enthusiasm for introducing more flexible distributions through extending the classical distributions by incorporating additional shape parameters to the baseline model. Many generalized families of distributions have been proposed and studied over the last two decades for modeling data in many applied areas. So, several classes of distributions have been constructed by extending common families of continuous distributions. These generalized distributions give more flexibility via adding one (or more) shape parameters to the baseline model. They were pioneered by Gupta et al. (1998) who proposed the exponentiated-G (Exp-G) class, which consists of raising the cumulative distribution function (cdf) to a positive power parameter. Many other classes can be cited such as the T-X family by Alzaatreh et al. (2013), the Lomax-G by Cordeiro et al. (2014), Burr X generator by Yousof et al. (2016), the generalized two-sided class of distributions by Korkmaz and Genç (2017), the Burr XII generator by Cordeiro et al. (2018), among others.

Recently, Ghitany et al. (2013) introduced the power Lindley (PL) distribution with the following cdf and probability density function (pdf)

$$F_{PL}(x; \alpha, \beta) = 1 - [1 + \beta x^\alpha / (\beta + 1)] e^{-\beta x^\alpha}, x \geq 0, \tag{1}$$

and

$$f_{PL}(x; \alpha, \beta) = \alpha \beta^2 (1 + x^\alpha) x^{\alpha-1} e^{-\beta x^\alpha} / (\beta + 1), x > 0, \tag{2}$$

respectively, where $\alpha > 0$ is a shape parameter and $\beta > 0$ is a scale parameter. Using the T- X idea, we define the new family by taking $W \left[G(x; \underline{\psi}) \right] = \frac{G(x; \underline{\psi})}{1 - G(x; \underline{\psi})}$ and $r(t) = f_{PL}(t; \alpha, \beta)$, where $G(x; \underline{\psi})$ is the baseline cdf depending on the vector parameter $\underline{\psi}$. The cdf of the new family is given by

$$F(x) = F(x; \alpha, \beta, \underline{\psi}) = 1 - \left(1 + \frac{\beta}{\beta + 1} \left[\frac{G(x; \underline{\psi})}{1 - G(x; \underline{\psi})} \right]^\alpha \right) \exp \left\{ -\beta \left[\frac{G(x; \underline{\psi})}{1 - G(x; \underline{\psi})} \right]^\alpha \right\}, x \in \mathbb{R}. \tag{3}$$

The pdf corresponding to (3) is given by

$$F(x) = f(x; \alpha, \beta, \psi) = \frac{\alpha\beta^2 g(x, \psi) G(x, \psi)^{\alpha-1}}{\beta + 1 [1 - G(x, \psi)]^{\alpha+1}} \times \left(1 + \left[\frac{G(x, \psi)}{1 - G(x, \psi)} \right]^\alpha \right) \underbrace{\exp \left\{ -\beta \left[\frac{G(x, \psi)}{1 - G(x, \psi)} \right]^\alpha \right\}}_A, x \in \mathbb{R}, \tag{4}$$

where $g(x; \psi)$ is the baseline pdf and $\alpha > 0$ and $\beta > 0$ are two extra parameters. The hazard rate function (hrf) is given by

$$h(x; \alpha, \beta, \psi) = \frac{\alpha\beta^2 g(x; \psi) G(x; \psi)^{\alpha-1}}{(\beta + 1) [1 - G(x; \psi)]^{\alpha+1} \left(1 + \frac{\beta}{\beta+1} \left[\frac{G(x; \psi)}{1 - G(x; \psi)} \right]^\alpha \right)} \left(1 + \left[\frac{G(x; \psi)}{1 - G(x; \psi)} \right]^\alpha \right), x \in \mathbb{R}. \tag{5}$$

We call this new family the odd power Lindley-G family and denote it by $OPL - G(\alpha, \beta, \psi)$. For $\alpha = 1$, the $OPL - G$ family is reduced to the $OL - G$ family which was introduced by Silva et al. (2017).

Let T be a PL random variable in (1) and (2). The OPL-G random variable having cdf (3) can be derived as follows

$$P(X \leq x) = \Pr \left\{ T \leq G(x; \psi) / [1 - G(x; \psi)] \right\}.$$

Hence, the random variable $X = G^{-1} [T / (T + 1); \psi]$ has the OPL-G distribution, where $G^{-1}(\cdot; \psi)$ is the inverse of the baseline cdf.

We can also motivate the OPL-G family with the mixture family structure as follows. Let $F_1(x; \alpha, \beta, \psi)$ be the cdf of the odd Weibull-G family (OW-G) (Bourguignon et al., 2014) and $F_2(x; \alpha, \beta, \psi)$ be the cdf of a generalized gamma-G family (GG-G). We note that the cdfs of the OW-G and GG-G families are given by

$$F_1(x; \alpha, \beta, \psi) = 1 - \exp \left\{ -\beta \left[\frac{G(x; \psi)}{1 - G(x; \psi)} \right]^\alpha \right\}, x \in \mathbb{R},$$

and

$$F_2(x; \alpha, \beta, \psi) = 1 - \left(1 + \beta \left[\frac{G(x; \psi)}{1 - G(x; \psi)} \right]^\alpha \right) \exp \left\{ -\beta \left[\frac{G(x; \psi)}{1 - G(x; \psi)} \right]^\alpha \right\}, x \in \mathbb{R},$$

respectively. Then, the OPL-G family can be expressed as

$$OPL - G(x; \alpha, \beta, \psi) = p [OW-G(x; \alpha, \beta, \psi)] + (1 - p) [GG-G(x; \alpha, \beta, \psi)], \tag{6}$$

where $p = \beta / (\beta + 1)$ is the mixing proportion. Hence, we can say that the OPL-G family is a mixture family.

The rest of the paper is organized as follows. Useful expansions for pdf and cdf of OPL-G family are presented in Section 2. Some of its special cases are taken up in Section 3. In Section 4, we derive some of its mathematical properties. Section 5 deals with some characterizations of the new family. Section 6 offers the maximum likelihood estimation method. Two Monte Carlo simulation studies performed in Section 7. A new log-location regression model as well as residual analysis are presented in Section 8. Section 9 is devoted to applications to real data sets to illustrate empirically the importance of the new model. Finally, some conclusions and future work are given in Section 10.

2. Useful Expansions for Density of OPL-G Family

By Expanding the quantity A in power series, the OPL-G pdf in (4) can be expressed as

$$f(x) = \sum_{j=0}^{\infty} \frac{\alpha\beta^{2+j} (-1)^j}{\beta + 1 j!} \underbrace{\frac{G(x)^{\alpha j + \alpha - 1}}{[1 - G(x)]^{\alpha j + \alpha + 1}}}_{B_1} g(x) + \sum_{j=0}^{\infty} \frac{\alpha\beta^{2+j} (-1)^j}{\beta + 1 j!} \underbrace{\frac{G(x)^{\alpha j + 2\alpha - 1}}{[1 - G(x)]^{\alpha j + 2\alpha + 1}}}_{B_2} g(x).$$

Using the generalized binomial expansion for the quantities B_1 and B_2 , we can write

$$\begin{aligned} f(x) &= \sum_{j,k=0}^{\infty} \frac{\alpha\beta^{2+j}(-1)^j\Gamma(\alpha j + \alpha + k + 1)}{\beta + 1 j!k!\Gamma(\alpha j + \alpha + 1)} g(x) G(x)^{\alpha j + \alpha + k - 1} \\ &+ \sum_{j,k=0}^{\infty} \frac{\alpha\beta^{2+j}(-1)^j\Gamma(\alpha j + 2\alpha + k + 1)}{\beta + 1 j!k!\Gamma(\alpha j + 2\alpha + 1)} g(x) G(x)^{\alpha j + 2\alpha - 1} \\ &= \sum_{j,k=0}^{\infty} a_{j,k}(\alpha j + \alpha + k) g(x) G(x)^{\alpha j + \alpha + k - 1} \\ &+ \sum_{j,k=0}^{\infty} b_{j,k}(\alpha j + 2\alpha + k) g(x) G(x)^{\alpha j + 2\alpha - 1}, \end{aligned}$$

finally

$$f(x) = \sum_{j,k=0}^{\infty} [a_{j,k}\pi_{\alpha(j+1)+k}(x) + b_{j,k}\pi_{\alpha(j+2)+k}(x)], \tag{7}$$

where

$$\begin{aligned} a_{j,k} &= \frac{\alpha\beta^{2+j}(-1)^j\Gamma(\alpha j + \alpha + k + 1)}{j!k!(\beta + 1)\Gamma(\alpha j + \alpha + 1)(\alpha j + \alpha + k)}, \\ b_{j,k} &= \frac{\alpha\beta^{2+j}(-1)^j\Gamma(\alpha j + 2\alpha + k + 1)}{j!k!(\beta + 1)\Gamma(\alpha j + 2\alpha + 1)(\alpha j + 2\alpha + k)}, \end{aligned}$$

and $\pi_{\delta}(x) = \delta g(x, \psi) G(x, \psi)^{\delta-1}$ is the pdf of the Exp-G distribution with power parameter δ . The corresponding OPL-G cdf can be written as

$$F(x) = \sum_{j,k=0}^{\infty} [a_{j,k}\Pi_{\alpha(j+1)+k}(x) + b_{j,k}\Pi_{\alpha(j+2)+k}(x)], \tag{8}$$

where $\Pi_{\delta}(x) = G(x, \psi)^{\delta}$ is the cdf of Exp-G with power parameter δ . Equation (7) and (8) reveal that pdf of OPL-G is a linear combination of Exp-G densities. Thereby, some properties of the proposed family such as moments and generating function can be determined by means of Exp-G distribution. The properties of Exp-G distributions have been studied by many authors in recent years, see Shirke and Kakade (2006) for exponentiated log-normal and Nadarajah and Gupta (2007) for exponentiated gamma distributions, among others.

3. Some Members of the OPL-G Family

The OPL-G family can produce great flexible models by using any baseline models. Here, we present three special cases of this family since they extend some widely-known distributions. We give pdf and cdf of the new distributions. The hrfs of these distributions can be obtained from Equation (5).

3.1 The OPL-Normal (OPL-N) Distribution

The normal distribution, symmetrical unimodal pdf shaped and increasing hrf shaped, is well-known in statistics and other areas. So, the OPL-N distribution is defined from (3) and (4) by taking $G(x)$ and $g(x)$ to be the cdf and pdf of the normal distribution. Its pdf and cdf are given by

$$\begin{aligned} f_{OPL-N}(x; \Theta) &= \frac{\alpha\beta^2\phi\left(\frac{x-\mu}{\sigma}\right)\left[\Phi\left(\frac{x-\mu}{\sigma}\right)\right]^{\alpha-1}}{(\beta + 1)\left[1 - \Phi\left(\frac{x-\mu}{\sigma}\right)\right]^{\alpha+1}} \left[1 + \left(\frac{\Phi\left(\frac{x-\mu}{\sigma}\right)}{1 - \Phi\left(\frac{x-\mu}{\sigma}\right)}\right)^{\alpha}\right] \\ &\times \exp\left\{-\beta\left(\frac{\Phi\left(\frac{x-\mu}{\sigma}\right)}{1 - \Phi\left(\frac{x-\mu}{\sigma}\right)}\right)^{\alpha}\right\}, x \in \mathbb{R}, \end{aligned}$$

and

$$F_{OPL-N}(x; \Theta) = 1 - \left[1 + \frac{\beta}{\beta + 1} \left(\frac{\Phi\left(\frac{x-\mu}{\sigma}\right)}{1 - \Phi\left(\frac{x-\mu}{\sigma}\right)}\right)^{\alpha}\right] \exp\left\{-\beta\left(\frac{\Phi\left(\frac{x-\mu}{\sigma}\right)}{1 - \Phi\left(\frac{x-\mu}{\sigma}\right)}\right)^{\alpha}\right\}, x \in \mathbb{R},$$

respectively, where $\Theta = (\alpha, \beta, \mu, \sigma)^T$ is the parameter vector, $\mu \in \mathbb{R}$, $\alpha, \beta, \sigma > 0$, $\phi(\cdot)$ and $\Phi(\cdot)$ are pdf and cdf of the standard normal distribution respectively. We denote it by $OPL - N(\Theta)$. For $\alpha = 1$, the OL-normal (OLN) distribution is

obtained. Plots of the OPL-N density and hazard functions for selected parameter values are displayed in Figure 1. From this Figure, we see that we obtain bi-modal shaped, firstly uni-modal shaped and then increasing shaped, left skewed and right skewed distributions. Also, the plots indicate that the hrf of the OPL-N distribution is increasing and then bathtub shaped.

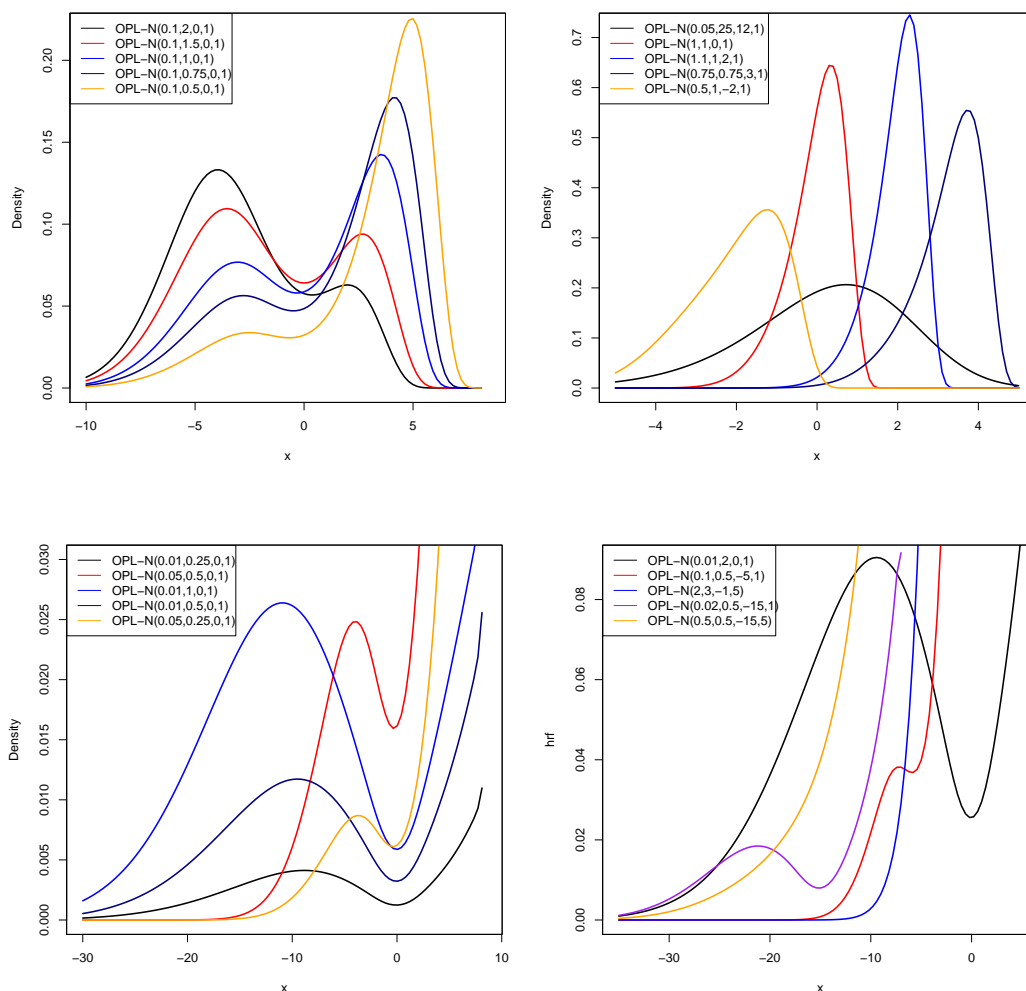


Figure 1. The pdf and hrf of the OPL-N distribution for selected parameter values

3.2 The OPL-Weibull (OPL-W) Distribution

The Weibull cdf with the shape $\gamma > 0$ and the scale parameters $\theta > 0$ is $G(x; \theta, \gamma) = 1 - \exp(-(\theta x)^\gamma)$ (for $x > 0$). The pdf and cdf of a random variable X with OPL-W distribution, say $X \sim OPL - W(\Theta)$ are, respectively, given by

$$f_{OPL-W}(x; \Theta) = \frac{\alpha \gamma \theta^\gamma \beta^2 x^{\gamma-1} (1 - e^{-(\theta x)^\gamma})^{\alpha-1}}{(\beta + 1) e^{-\alpha(\theta x)^\gamma}} \left[1 + (e^{(\theta x)^\gamma} - 1)^\alpha \right] \exp \left\{ -\beta (e^{(\theta x)^\gamma} - 1)^\alpha \right\}, x > 0,$$

and

$$F_{OPL-W}(x; \Theta) = 1 - \left[1 + \frac{\beta}{\beta + 1} (e^{(\theta x)^\gamma} - 1)^\alpha \right] \exp \left\{ -\beta (e^{(\theta x)^\gamma} - 1)^\alpha \right\}, x \geq 0$$

where $\Theta = (\alpha, \beta, \theta, \gamma)^T$ is the parameter vector and $\alpha, \beta, \theta, \gamma > 0$. For $\alpha = 1$, the OL-Weibull (OLW) distribution (Silva et al., 2017) is obtained. Plots of the OPL-W density and hazard functions for selected parameter values are displayed in Figure 2. From this Figure, we can say that the OPL-W distribution has very flexible pdf shapes such as uni-modal, decreasing, U-shaped, firstly U-shaped and then decreasing shaped. Also, its hrf can be increasing, decreasing and bathtub shaped.

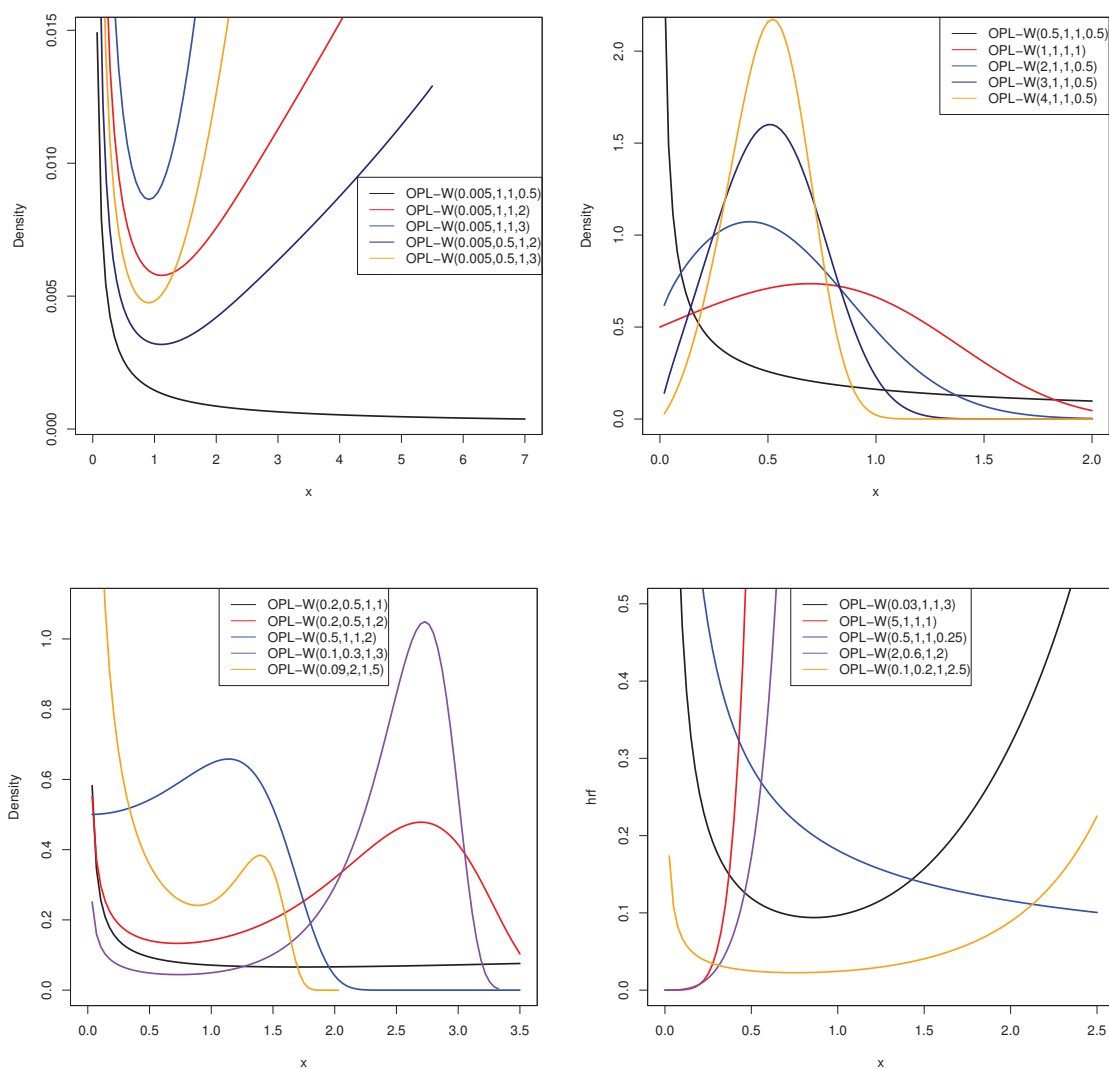


Figure 2. The pdf and hrf of the OPL-W distribution for selected parameter values

3.3 The OPL-Gamma (OPL-Ga) Distribution

Consider the gamma distribution with the shape parameter $\gamma > 0$ and the scale parameter $\theta > 0$, where the pdf and cdf (for $x > 0$) are given by $G(x; \lambda, \theta) = \gamma(\lambda, \theta x) \Gamma^{-1}(\lambda)$, where $\gamma(\lambda, \theta x) = \int_0^{\theta x} t^{\lambda-1} e^{-t} dt$ is the upper incomplete gamma function and $\Gamma(\cdot)$ is complete gamma function. The pdf and cdf of OPL-Ga are given by

$$f_{OPL-Ga}(x; \Theta) = \frac{\alpha \beta^2 \theta^\lambda x^{\lambda-1} e^{-\theta x} [\gamma(\lambda, \theta x) \Gamma^{-1}(\lambda)]^{\alpha-1}}{(\beta + 1) \Gamma(\lambda) [1 - \gamma(\lambda, \theta x) \Gamma^{-1}(\lambda)]^{\alpha+1} \left[1 + \left(\frac{\gamma(\lambda, \theta x)}{\Gamma(\lambda) - \gamma(\lambda, \theta x)} \right)^\alpha \right]} \times \exp \left\{ -\beta \left(\frac{\gamma(\lambda, \theta x)}{\Gamma(\lambda) - \gamma(\lambda, \theta x)} \right)^\alpha \right\}$$

and

$$F_{OPL-Ga}(x; \Theta) = 1 - \left[1 + \frac{\beta}{\beta + 1} \left(\frac{\gamma(\lambda, \theta x)}{\Gamma(\lambda) - \gamma(\lambda, \theta x)} \right)^\alpha \right] \exp \left\{ -\beta \left(\frac{\gamma(\lambda, \theta x)}{\Gamma(\lambda) - \gamma(\lambda, \theta x)} \right)^\alpha \right\},$$

respectively, where $\Theta = (\alpha, \beta, \theta, \lambda)^T$ is the parameter vector and $\alpha, \beta, \theta, \lambda > 0$. We denote it by $OPL - Ga(\Theta)$. For $\alpha = 1$, the OL-gamma (OLGa) distribution is obtained. Plots of the OPL-Ga density and hazard functions for selected parameter values are displayed in Figure 3. From this Figure, we observe decreasing, uni-modal shaped, U-shaped and firstly U-shaped and then decreasing shaped distributions. Also, the plots point out that the OPL-Ga distribution has decreasing, increasing, bathtub shaped.

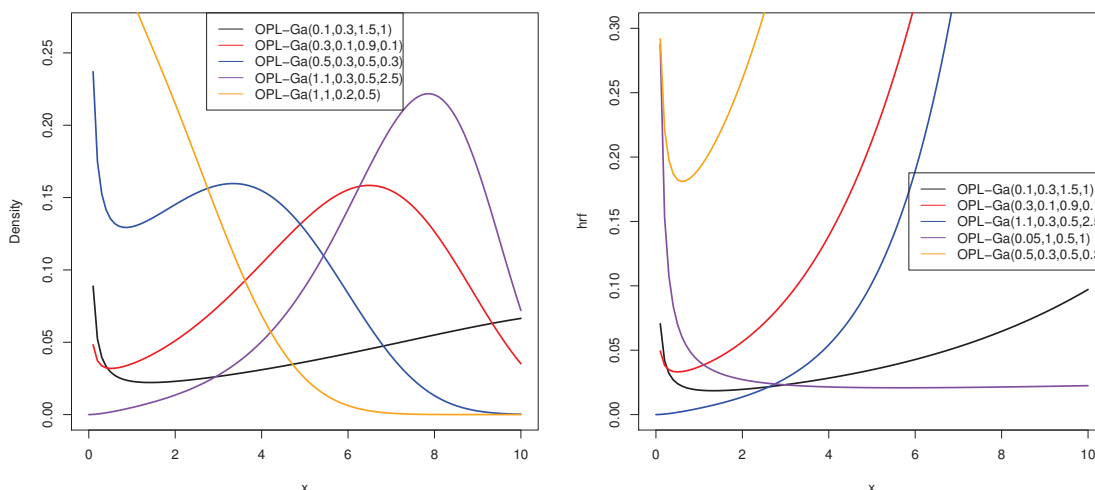


Figure 3. The pdf and hrf of the OPL-Ga distribution for selected parameter values

4. Mathematical Properties

4.1 Moments, Incomplete Moments and Generating Function

The r^{th} ordinary moment of X is given by $\mu'_r = E(X^r) = \int_{-\infty}^{\infty} x^r f(x) dx$. Then we obtain

$$\mu'_r = \sum_{j,k=0}^{\infty} [a_{j,k} Y_{\alpha(j+1)+k}^r + b_{j,k} Y_{\alpha(j+2)+k}^r]. \tag{9}$$

Henceforth, Y_{δ} denotes the Exp-G model with power parameter δ . For $\delta > 0$, we have $E(Y_{\delta}^r) = \delta \int_{-\infty}^{\infty} x^r g(x; \psi) G(x; \psi)^{\delta-1} dx$, which can be computed numerically in terms of the baseline quantile function (qf) $Q_G(u; \psi) = G^{-1}(u; \psi)$ as $E(Y_{\delta}^r) = \delta \int_0^1 Q_G(u; \psi)^r u^{\delta-1} du$. Setting $r = 1$ in (9), we have the mean of X . The last integration can be computed numerically for most parent distributions. The skewness and kurtosis measures can be calculated from the ordinary moments using well-known relationships. The n^{th} central moment of X , say M_n , is

$$M_n = E(X - \mu)^n = \sum_{h=0}^n (-1)^h \binom{n}{h} (\mu'_1)^n \mu'_{n-h}.$$

The r^{th} incomplete moment, say $I_r(t)$, of X can be expressed from (6) as

$$I_r(t) = \int_{-\infty}^t x^r f(x) dx = \sum_{j,k=0}^{\infty} \left\{ \int_{-\infty}^t [a_{j,k} x^r \pi_{\alpha(j+1)+k}(x) + b_{j,k} x^r \pi_{\alpha(j+2)+k}(x)] dx \right\}. \tag{10}$$

The mean deviations about the mean [$\Upsilon_1 = E(|X - \mu'_1|)$] and about the median [$\Upsilon_2 = E(|X - M|)$] of X are given by $\Upsilon_1 = 2\mu'_1 F(\mu'_1) - 2I_1(\mu'_1)$ and $\Upsilon_2 = \mu'_1 - 2I_1(M)$, respectively, where $\mu'_1 = E(X)$, $M = Median(X) = Q(0.5)$ is the median, $F(\mu'_1)$ is easily calculated from (3) and $I_1(t)$ is the first incomplete moment given by (10) with $r = 1$. A general equation for $I_1(t)$ can be derived from (10) as

$$I_1(t) = \sum_{j,k=0}^{\infty} [a_{j,k} J_{\alpha(j+1)+k}(t) + b_{j,k} J_{\alpha(j+2)+k}(t)],$$

where $J_{\delta}(x) = \int_{-\infty}^x x \pi_{\delta}(x) dx$ is the first incomplete moment of the Exp-G model. The moment generating function (mgf) $M_X(t) = E(e^{tX})$ of X can be derived using equation (9) as

$$M_X(t) = \sum_{j,k=0}^{\infty} [a_{j,k} M_{\alpha(j+1)+k}(t) + b_{j,k} M_{\alpha(j+2)+k}(t)],$$

where $M_{\delta}(t)$ is the mgf of Y_{δ} . Hence, $M_X(t)$ can be determined from the Exp-G generating function.

4.2 Quantile Function (qf) and Random Number Generation

The OPL-G family can easily be simulated by inverting (3) as follows: if $U \sim U(0, 1)$, then the random variable X_U can be obtained from the baseline qf, say $Q_G(u) = G^{-1}(u)$. In fact, the random variable

$$X_U = G^{-1} \left(\frac{\left(-\frac{1}{\beta} - \frac{W_{-1}((\beta+1)(U-1)e^{-\beta-1})}{\beta} - 1 \right)^{1/\alpha}}{1 + \left(-\frac{1}{\beta} - \frac{W_{-1}((\beta+1)(U-1)e^{-\beta-1})}{\beta} - 1 \right)^{1/\alpha}} \right)$$

has cdf (3), where $G^{-1}(\cdot)$ is the inverse of the baseline cdf and $W_{-1}(\cdot)$ denotes the negative branch of the Lambert W function. X_U can be used as a random number generator for OPL-G distribution. Also, we can obtain random number from OPL-G by using mixture structure in (6). We can give this procedure with the following an algorithm.

Algorithm(Mixture Form)

1. Generate $U \sim Uniform(0, 1)$;
2. Generate $Y \sim OW - G(\alpha, \beta, \underline{\psi})$;
3. Generate $Z \sim GG - G(\alpha, \beta, \underline{\psi})$;
4. if $U \leq \frac{\beta}{\beta+1}$, then $X = Y$ otherwise, set $X = Z$.

By using packet programme, the random variates from the OLP-G distribution can be generated by the transformation method. For example, we first generate a random variate Y from the PL distribution by using the rplindley function in the LindleyR package in R program, then set $X = G^{-1} \left(\frac{Y}{Y+1} \right)$.

5. Characterizations

This section deals with various characterizations of OPL-G distribution. These characterizations are presented in two directions: (i) based on a simple relationship between two truncated moments and (ii) in terms of the hazard function. It should be pointed out that due to the nature of the OPL-G distribution our characterizations may be the only possible ones for this distribution.

We present our characterizations (i) and (ii) in two subsections.

5.1 Characterizations Based on Truncated Moments

We employ a theorem due to Glänzel (1987), see Theorem 1 of Appendix A .The result, however, holds also when the interval H is not closed since the condition of Theorem 1 is on the interior of H . We like to mention that this kind of characterization based on a truncated moment is stable in the sense of weak convergence (see, Glänzel 1990).

Proposition 5.1. Let $X : \Omega \rightarrow \mathbb{R}$ be a continuous random variable and let $q_1(x) = \left(1 + \left[\frac{G(x, \underline{\psi})}{1-G(x, \underline{\psi})} \right]^\alpha \right)^{-1}$ and $q_2(x) = q_1(x) \exp \left\{ -\beta \left[\frac{G(x, \underline{\psi})}{1-G(x, \underline{\psi})} \right]^\alpha \right\}$ for $x \in \mathbb{R}$. The random variable X belongs to the family (4) if and only if the function η defined in Theorem1 has the form

$$\eta(x) = \frac{1}{2} \exp \left\{ -\beta \left[\frac{G(x, \underline{\psi})}{1-G(x, \underline{\psi})} \right]^\alpha \right\}, \quad x \in \mathbb{R}.$$

Proof. Let X be a random variable with pdf (4), then

$$(1 - F(x)) E [q_1(X) | X \geq x] = \frac{\beta}{\beta + 1} \exp \left\{ -\beta \left[\frac{G(x, \underline{\psi})}{1-G(x, \underline{\psi})} \right]^\alpha \right\}$$

and

$$(1 - F(x)) E [q_2(X) | X \geq x] = \frac{\beta}{2(\beta + 1)} \exp \left\{ -2\beta \left[\frac{G(x, \underline{\psi})}{1-G(x, \underline{\psi})} \right]^\alpha \right\}.$$

Further,

$$\eta(x) q_1(x) - q_2(x) = -\frac{q_1(x)}{2} \exp \left\{ -\beta \left[\frac{G(x, \underline{\psi})}{1 - G(x, \underline{\psi})} \right]^\alpha \right\} < 0 \text{ for } x \in \mathbb{R}.$$

Conversely, if η is given as above, then

$$s'(x) = \frac{\eta'(x) q_1(x)}{\eta(x) q_1(x) - q_2(x)} = \frac{\alpha \beta g(x, \underline{\psi}) G(x, \underline{\psi})^{\alpha-1}}{[1 - G(x, \underline{\psi})]^{\alpha+1}}, \quad x \in \mathbb{R},$$

and hence

$$s(x) = \beta \left[\frac{G(x, \underline{\psi})}{1 - G(x, \underline{\psi})} \right]^\alpha, \quad x \in \mathbb{R}.$$

Now, according to Theorem 1, X has density (4).

Corollary 5.1. Let $X : \Omega \rightarrow \mathbb{R}$ be a continuous random variable and let q_1 be as in Proposition 5.1. Then, X has pdf (4) if and only if there exist functions q_2 and η defined in Theorem 1 satisfying the differential equation

$$\frac{\eta'(x) q_1(x)}{\eta(x) q_1(x) - q(x)} = \frac{\alpha \beta g(x, \underline{\psi}) G(x, \underline{\psi})^{\alpha-1}}{[1 - G(x, \underline{\psi})]^{\alpha+1}}, \quad x \in \mathbb{R}.$$

The general solution of the differential equation in Corollary 5.1 is

$$\eta(x) = \exp \left\{ \beta \left[\frac{G(x, \underline{\psi})}{1 - G(x, \underline{\psi})} \right]^\alpha \right\} \times \left[- \int \frac{\alpha \beta g(x, \underline{\psi}) G(x, \underline{\psi})^{\alpha-1}}{[1 - G(x, \underline{\psi})]^{\alpha+1}} \exp \left\{ -\beta \left[\frac{G(x, \underline{\psi})}{1 - G(x, \underline{\psi})} \right]^\alpha \right\} (q_1(x))^{-1} q_2(x) dx + D \right],$$

where D is a constant. Note that a set of functions satisfying the above differential equation is given in Proposition 5.1 with $D = 0$. Clearly, there exist other triplet of functions (q_1, q_2, η) satisfying the conditions of Theorem 1.

5.2 Characterization Based on Hazard Function

It is known that the hazard function, h_F , of a twice differentiable distribution function, F , satisfies the first order differential equation

$$\frac{f'(x)}{f(x)} = \frac{h'_F(x)}{h_F(x)} - h_F(x).$$

For many univariate continuous distributions, this is the only characterization available in terms of the hazard function. The following characterization establish a non-trivial characterization of OPL-G distribution, for $\alpha = 1$, in terms of the hazard function, which is not of the above trivial form.

Proposition 5.2. Let $X : \Omega \rightarrow \mathbb{R}$ be a continuous random variable. For $\alpha = 1$, the pdf of X is (4) if and only if its hazard function $h_F(x)$ satisfies the differential equation

$$h'_F(x) - \frac{g'(x, \underline{\psi})}{g(x, \underline{\psi})} h_F(x) = \beta^2 g(x, \underline{\psi}) \frac{d}{dx} \left\{ \frac{[\beta + 1 - G(x, \underline{\psi})]^{-1}}{[1 - G(x, \underline{\psi})]^2} \right\}, \quad x \in \mathbb{R},$$

with the initial condition $\lim_{x \rightarrow -\infty} h_F(x) = \frac{\beta^2}{\beta + 1} \lim_{x \rightarrow -\infty} g(x, \underline{\psi})$.

Proof. If X has pdf (4), then clearly the above differential equation holds. Now, if the differential equation holds, then

$$\frac{d}{dx} \left\{ g(x, \underline{\psi})^{-1} h_F(x) \right\} = \beta^2 \frac{d}{dx} \left\{ \frac{[\beta + 1 - G(x, \underline{\psi})]^{-1}}{[1 - G(x, \underline{\psi})]^2} \right\},$$

or

$$h_F(x) = \beta^2 g(x, \underline{\psi}) \left\{ \frac{[\beta + 1 - G(x, \underline{\psi})]^{-1}}{[1 - G(x, \underline{\psi})]^2} \right\},$$

which is the hazard function of the OPL-G distribution for $\alpha = 1$.

Remark 5.1. It is easy to see that

$$\begin{aligned} & \frac{d}{dx} \left\{ \frac{[\beta + 1 - G(x, \underline{\psi})]^{-1}}{[1 - G(x, \underline{\psi})]^2} \right\} \\ &= g(x, \underline{\psi}) \left(\frac{[\beta + 1 - G(x, \underline{\psi})]^{-2}}{[1 - G(x, \underline{\psi})]^2} \right) \left\{ 3 + \frac{2\beta}{1 - G(x, \underline{\psi})} \right\}. \end{aligned}$$

6. Estimation and Inference

Here, we consider estimation of the unknown parameters of the OPL-G distribution by the maximum likelihood method. Let x_1, \dots, x_n be a random sample from the OPL-G distribution with a $(q + 2) \times 1$ parameter vector $\Psi = (\alpha, \beta, \underline{\psi})^T$, where $\underline{\psi}$ is a $q \times 1$ baseline parameter vector. The log-likelihood function for Ψ is given by

$$\begin{aligned} \ell(\Psi) &= n \log \alpha - 2n \log \beta - n \log(\beta + 1) + \sum_{i=1}^n \log g(x_i; \underline{\psi}) + (\alpha - 1) \sum_{i=1}^n \log G(x_i; \underline{\psi}) \\ &\quad - (\alpha + 1) \sum_{i=1}^n \log [1 - G(x_i; \underline{\psi})] + \sum_{i=1}^n \log(1 + \eta_i^\alpha) - \beta \sum_{i=1}^n \eta_i^\alpha, \end{aligned}$$

where $\eta_i = G(x_i; \underline{\psi}) / [1 - G(x_i; \underline{\psi})]$. The components of the score vector, $U(\Psi) = \left(\frac{\partial \ell}{\partial \alpha}, \frac{\partial \ell}{\partial \beta}, \frac{\partial \ell}{\partial \underline{\psi}} \right)^T$, are given as

$$\begin{aligned} \frac{\partial \ell}{\partial \alpha} &= \frac{n}{\alpha} + \sum_{i=1}^n \log G(x_i; \underline{\psi}) - \sum_{i=1}^n \log [1 - G(x_i; \underline{\psi})] + \sum_{i=1}^n \frac{\eta_i^\alpha \log(\eta_i)}{1 + \eta_i^\alpha} - \beta \sum_{i=1}^n \eta_i^\alpha \log(\eta_i), \\ \frac{\partial \ell}{\partial \beta} &= -\frac{2n}{\beta} - \frac{n}{\beta + 1} - \sum_{i=1}^n \eta_i^\alpha, \end{aligned}$$

and (for $r = 1, \dots, q$)

$$\begin{aligned} \frac{\partial \ell}{\partial \underline{\psi}_r} &= \sum_{i=1}^n \frac{\partial g(x_i; \underline{\psi}) / \partial \underline{\psi}_r}{g(x_i; \underline{\psi})} + (\alpha - 1) \sum_{i=1}^n \left[\frac{\partial G(x_i; \underline{\psi})}{\partial \underline{\psi}_r} \right] \\ &\quad + (\alpha + 1) \sum_{i=1}^n \frac{\partial G(x_i; \underline{\psi}) / \partial \underline{\psi}_r}{1 - G(x_i; \underline{\psi})} + \alpha \sum_{i=1}^n \frac{\zeta_i \eta_i^{\alpha-1}}{1 + \eta_i^\alpha} - \alpha \beta \sum_{i=1}^n \zeta_i \eta_i^{\alpha-1}, \end{aligned}$$

where

$$\zeta_i = \left[\frac{\partial G(x_i; \underline{\psi})}{\partial \underline{\psi}_r} \right] [1 - G(x_i; \underline{\psi})]^{-2}.$$

Setting the nonlinear system of equations $U_\alpha = U_\beta = U_{\underline{\psi}_r} = 0$ (for $r = 1, \dots, q$) and solving them simultaneously yields the MLEs $\widehat{\Psi} = (\widehat{\alpha}, \widehat{\beta}, \widehat{\underline{\psi}}^T)^T$. To solve these equations, it is more convenient to use nonlinear optimization methods such as the quasi-Newton algorithm to numerically maximize $\ell(\Psi)$. For interval estimation of the parameters, we can evaluate

numerically the elements of the $(q + 2) \times (q + 2)$ observed information matrix $J(\Psi) = \{-\frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j}\}$. Under standard regularity conditions when $n \rightarrow \infty$, the distribution of $\widehat{\Psi}$ can be approximated by a multivariate normal $N_p(0, J(\widehat{\Psi})^{-1})$ distribution to construct approximate confidence intervals for the parameters. Here, $J(\widehat{\Psi})$ is the total observed information matrix evaluated at $\widehat{\Psi}$. The method of the re-sampling bootstrap can be used for correcting the biases of the MLEs of the model parameters. Good interval estimates may also be obtained using the bootstrap percentile method.

We can compute the maximum values of the unrestricted and restricted log-likelihoods to obtain likelihood ratio (LR) statistics for testing sub-model of the OPL-G distribution. Hypothesis tests of the type $H_0 : \omega = \omega_0$ versus $H_1 : \omega \neq \omega_0$, where ω is a vector formed with some components of Ψ and ω_0 is a specified vector, can be performed using LR statistics. For example, the test of $H_0 : \alpha = 1$ versus $H_1 : H_0$ is not true is equivalent to comparing the OPL-G and PL-G distributions and the LR statistic is given by $w = 2\{\ell(\widehat{\alpha}, \widehat{\beta}, \widehat{\psi}) - \ell(1, \widehat{\beta}, \widehat{\psi})\}$, where $\widehat{\alpha}, \widehat{\beta}$ and $\widehat{\psi}$ are the MLEs under H and $\widehat{\psi}$ is the estimate under H_0 .

We can compute the maximum values of the unrestricted and restricted log-likelihoods to obtain likelihood ratio (LR) statistics for testing some sub-models of the OPL-G distribution.

7. Log-OPL-W Regression Model

Let X be a random variable having the OPL-W density function with four parameters $\alpha > 0, \beta > 0, \theta > 0$ and $\gamma > 0$, given in Section 3. The density function of $Y = \log(X)$, replacing $\gamma = 1/\sigma$ and $\theta = \exp(-\mu)$, is given by (for $y \in \mathfrak{R}$)

$$f(y) = \frac{\alpha\beta^2}{\sigma(\beta+1)} \frac{\exp\left[\left(\frac{y-\mu}{\sigma}\right) - \exp\left(\frac{y-\mu}{\sigma}\right)\right] \left\{1 - \exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]\right\}^{\alpha-1}}{\left[2 - \exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]\right]^{\alpha+1}} \times \left(1 + \left[\frac{1 - \exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]}{2 - \exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]}\right]^\alpha\right) \exp\left\{-\beta \left[\frac{1 - \exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]}{2 - \exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]}\right]^\alpha\right\}, \tag{11}$$

where $\mu \in \mathfrak{R}$ is the location parameter, $\sigma > 0$ is the scale parameter and $\alpha > 0, \beta > 0$ are the shape parameters. LOPLW distribution is denoted as $Y \sim \text{LOPLW}(\alpha, \beta, \sigma, \mu)$. Figure 5 displays density plots of LOPLW distribution for some parameter values. As seen from Figure 4, LOPLW distribution can be very flexible for modeling left skewed data.

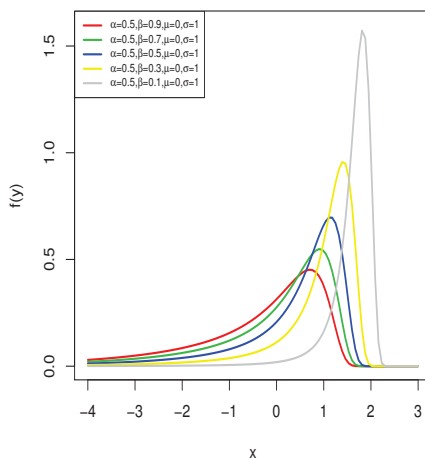


Figure 4. Plots of the LOPLW density for selected parameter values

The corresponding survival function is

$$S(y) = \left(1 + \frac{\beta}{\beta + 1} \left[\frac{1 - \exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]}{2 - \exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]}\right]^\alpha\right) \exp\left\{-\beta \left[\frac{1 - \exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]}{2 - \exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]}\right]^\alpha\right\}. \tag{12}$$

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

$$f(z) = \frac{\alpha\beta^2}{(\beta+1)} \frac{\exp\left[z - \exp(z)\right] \left\{1 - \exp\left[-\exp(z)\right]\right\}^{\alpha-1}}{\left[2 - \exp\left[-\exp(z)\right]\right]^{\alpha+1}} \left(1 + \left[\frac{1 - \exp\left[-\exp(z)\right]}{2 - \exp\left[-\exp(z)\right]}\right]^\alpha\right) \times \exp\left\{-\beta \left[\frac{1 - \exp\left[-\exp(z)\right]}{2 - \exp\left[-\exp(z)\right]}\right]^\alpha\right\}. \tag{13}$$

Based on the LOPLW density, given in (11), the log-linear location-scale regression model is proposed by linking the response variable y_i and the explanatory variable vector $\mathbf{v}_i^\top = (v_{i1}, \dots, v_{ip})$ by

$$y_i = \mathbf{v}_i^\top \boldsymbol{\beta} + \sigma z_i, \quad i = 1, \dots, n, \tag{14}$$

where the random error z_i has density function (13), $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$, $\sigma > 0$, $\alpha > 0$ and $\beta > 0$ are unknown parameters. The parameter $\mu_i = \mathbf{v}_i^\top \boldsymbol{\beta}$ is the location of y_i . The location parameter vector $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^\top$ is represented by a linear model $\boldsymbol{\mu} = V\boldsymbol{\beta}$, where $V = (v_1, \dots, v_n)^\top$ is a known model matrix.

Let $(y_1, v_1), \dots, (y_n, v_n)$ be a sample of n independent observations, where each random response is defined by $y_i = \min\{\log(x_i), \log(c_i)\}$. Let F and C be the sets of individuals for which y_i is the log-lifetime or log-censoring, respectively. The log-likelihood function for the vector of parameters $\boldsymbol{\tau} = (\alpha, \beta, \sigma, \boldsymbol{\beta}^\top)^\top$ from model (14) has the form $l(\boldsymbol{\tau}) = \sum_{i \in F} l_i(\boldsymbol{\tau}) +$

$\sum_{i \in C} l_i^{(c)}(\boldsymbol{\tau})$, where $l_i(\boldsymbol{\tau}) = \log[f(y_i)]$, $l_i^{(c)}(\boldsymbol{\tau}) = \log[S(y_i)]$, $f(y_i)$ is the density (11) and $S(y_i)$ is the survival function (12) of Y_i . Then, the total log-likelihood function for $\boldsymbol{\tau}$ is given by

$$\begin{aligned} \ell(\boldsymbol{\tau}) &= r \log\left(\frac{\alpha\beta^2}{\sigma(\beta+1)}\right) + \sum_{i \in F} (z_i - u_i) + (\alpha - 1) \sum_{i \in F} \log\{1 - \exp[-u_i]\} \\ &\quad - (\alpha + 1) \sum_{i \in F} \log\{2 - \exp[-u_i]\} \\ &\quad + \sum_{i \in F} \log\left(1 + \left[\frac{\{1 - \exp[-u_i]\}}{2 - \exp[-u_i]}\right]^\alpha\right) - \beta \sum_{i \in F} \left[\frac{\{1 - \exp[-u_i]\}}{2 - \exp[-u_i]}\right]^\alpha \\ &\quad + \sum_{i \in C} \log\left(1 + \frac{\beta}{\beta+1} \left[\frac{\{1 - \exp[-u_i]\}}{2 - \exp[-u_i]}\right]^\alpha\right) - \beta \sum_{i \in C} \left[\frac{\{1 - \exp[-u_i]\}}{2 - \exp[-u_i]}\right]^\alpha, \end{aligned} \tag{15}$$

where $u_i = \exp(z_i)$, $z_i = (y_i - v_i^\top \boldsymbol{\beta})/\sigma$ and r is the number of uncensored observations and c is the number of censored observations. The MLE $\widehat{\boldsymbol{\tau}}$ of the vector of unknown parameters can be evaluated by maximizing the log-likelihood (15). The optim fuction of R software is used to minimize the negative log-likelihood function, given in (15). The asymptotic distribution of $(\widehat{\boldsymbol{\tau}} - \boldsymbol{\tau})$ is multivariate normal $N_{p+2}(0, K(\boldsymbol{\tau})^{-1})$, where $K(\boldsymbol{\tau})$ is the expected information matrix. The asymptotic covariance matrix $K(\boldsymbol{\tau})^{-1}$ of $\widehat{\boldsymbol{\tau}}$ can be approximated by the inverse of the $(p+2) \times (p+2)$ observed information matrix $-\ddot{L}(\boldsymbol{\tau})$, whose elements can be evaluated numerically. The approximate multivariate normal distribution $N_{p+2}(0, -\ddot{L}(\boldsymbol{\tau})^{-1})$ for $\widehat{\boldsymbol{\tau}}$ can be used to construct approximate confidence intervals for the parameters of $\boldsymbol{\tau}$.

8. Simulation Studies

In this Section, we perform two simulation studies by using the OPL-W and OPL-N distributions to illustrate the performance of MLEs for the parameters of these distribution. The random numbers generation is obtained with rplindley function in the LindleyR package in R program. This method is given by qf subsection. MLEs results were obtained using optim-CG routine in the R programme.

In the first simulation study, we obtain the graphical results and generate $N = 1000$ samples of size $n = 20, 30, \dots, 1000$ from OPL-W distribution with true parameters values $\alpha = 0.5, \beta = 10, \theta = 1$ and $\gamma = 2$. In this simulation study, we calculate the empirical mean, standard deviations (sd), bias and mean square error (MSE) of the MLEs. The bias and MSE are calculated by $\widehat{Bias}_h = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{h}_i - h)$ and $\widehat{MSE}_h = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{h}_i - h)^2$, respectively for $h = \alpha, \beta, \theta, \gamma$. We give results of this simulation study in Figure 5.

In the second simulation study, we generate 1,000 samples of sizes 20,50 and 100 from selected OPL-N distributions. For this simulation study, we obtain the empirical means and sd's of the parameters. The results of this simulation study are reported in Table 1.

From Figure 5 and Table 1, we observe that when the sample size increases, the empirical means approach the true parameter value for both distributions. For the same case, the standard deviations, biases and MSEs decrease in all the cases. The above results are as expected.

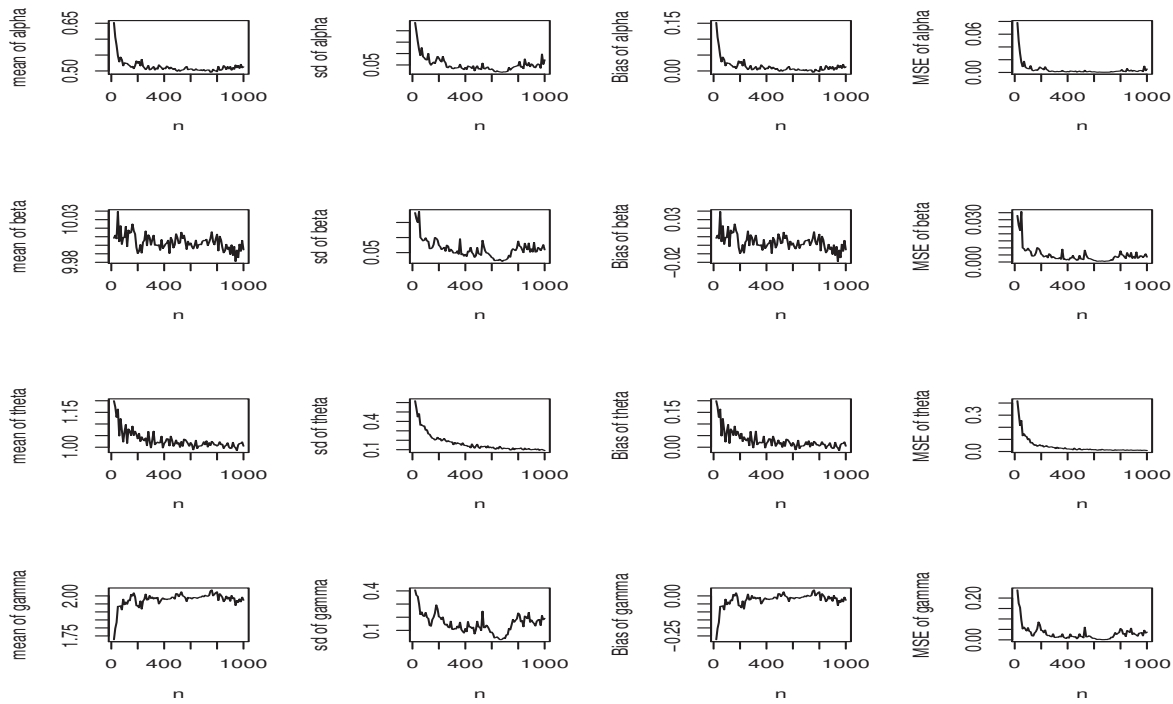


Figure 5. Simulation results of the special OPL-W distribution

Table 1. Empirical means and standard deviations (in parentheses) for the special OPL-N distributions

Parameters	$n = 20$				$n = 50$				$n = 100$			
	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\mu}$	$\hat{\sigma}$
1,5,0,1	1.3281 (0.8949)	4.8099 (0.2851)	-0.0861 (0.1737)	1.1638 (0.6879)	1.1606 (0.6665)	4.9131 (0.1495)	-0.0355 (0.1259)	1.1049 (0.4826)	1.1311 (0.5544)	4.9434 (0.1147)	-0.0011 (0.0904)	1.1002 (0.4434)
0.5,5,0,1	0.8338 (0.7375)	4.6636 (0.5231)	-0.2680 (0.3835)	1.2233 (0.7881)	0.6435 (0.3751)	4.7598 (0.4942)	-0.0974 (0.2725)	1.1409 (0.4535)	0.5745 (0.2864)	4.8830 (0.2862)	-0.0492 (0.1897)	1.0686 (0.3523)
0.5,5,-1,2	0.7784 (0.6150)	4.6867 (0.6412)	-1.4902 (0.6716)	2.4721 (1.4254)	0.6129 (0.3555)	4.8485 (0.4751)	-1.2292 (0.4523)	2.1792 (0.8643)	0.5931 (0.2506)	4.9133 (0.2900)	-1.1668 (0.3041)	2.1780 (0.6474)
2,10,1,2	2.0651 (0.9860)	9.9460 (0.1015)	0.9624 (0.2657)	1.8793 (0.7970)	1.9587 (0.9088)	9.9682 (0.0725)	1.0095 (0.1921)	1.9271 (0.7541)	2.0398 (0.7284)	9.9793 (0.0641)	1.0094 (0.1210)	2.0674 (0.6208)
0.5,5,5,5	0.7379 (0.5591)	5.0713 (1.0634)	4.0969 (1.1493)	5.8290 (2.6787)	0.5406 (0.1798)	5.2049 (0.6566)	4.7107 (0.5602)	5.0259 (1.0739)	0.5354 (0.1677)	4.9965 (0.3931)	4.8673 (0.3994)	5.1401 (0.9710)
10,3,-5,3	10.0295 (0.1425)	3.0020 (0.2560)	-5.0111 (0.0485)	2.8605 (0.4680)	9.9966 (0.0880)	2.9923 (0.1340)	-5.0042 (0.0290)	2.9983 (0.2995)	10.0048 (0.0628)	3.0262 (0.2219)	-4.9996 (0.0244)	2.9731 (0.2054)
1,3,0,5,0,5	1.0454 (0.7997)	2.7555 (0.3340)	0.4449 (0.0871)	0.4712 (0.3249)	1.0145 (0.5744)	2.8821 (0.3081)	0.4864 (0.0698)	0.4955 (0.2415)	0.9811 (0.3642)	2.9258 (0.1839)	0.4885 (0.0358)	0.4833 (0.1549)
0.5,5,-0.5,0.5	0.7696 (0.6217)	4.8540 (0.2594)	-0.6163 (0.1805)	0.6145 (0.3963)	0.6007 (0.3560)	4.8989 (0.3125)	-0.5351 (0.1222)	0.5475 (0.2162)	0.5761 (0.2624)	4.9232 (0.2932)	-0.5304 (0.1057)	0.5333 (0.1525)

9. Real Data Applications

In this section, we present three applications to real data sets to illustrate the usefulness of the OPL-N, OPL-W and OPL-Ga distributions. We compare these distribution model with the odd Burr normal (OBN) (Alizadeh et al., 2017), power normal (PN) (Gupta and Gupta, 2008), OLN, normal (N), Lomax-Weibull (LxW) (Cordeiro et al., 2014), OLW, Weibull (W), exponentiated generalized gamma (EGGa) (Cordeiro et al., 2011), OLGa, gamma (Ga) and PL models. The pdfs of the OBN, PN, LxW and EGGa are given by

$$f_{OBN}(x; \alpha, \beta, \mu, \sigma) = \frac{\alpha\beta\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\frac{x-\mu}{\sigma}\right)^{\alpha-1}\left[1-\Phi\left(\frac{x-\mu}{\sigma}\right)\right]^{\alpha\beta-1}}{\sigma\left\{\Phi\left(\frac{x-\mu}{\sigma}\right)^{\alpha} + \left[1-\Phi\left(\frac{x-\mu}{\sigma}\right)^{\alpha}\right]\right\}^{\beta+1}}, x, \mu \in \mathfrak{R}, \alpha, \beta, \sigma > 0,$$

$$f_{PN}(x; \beta, \mu, \sigma) = \beta\sigma^{-1}\phi\left(\frac{x-\mu}{\sigma}\right)\left[\Phi\left(\frac{x-\mu}{\sigma}\right)\right]^{\beta-1}, x, \mu \in \mathfrak{R}, \beta, \sigma > 0,$$

$$f_{LxW}(x; \alpha, \beta, \theta, \gamma) = \alpha\beta^{\alpha}\gamma\theta^{\gamma}x^{\gamma-1}[\beta + (x\theta)^{\gamma}]^{-\alpha-1}, x > 0, \alpha, \beta, \theta, \gamma > 0$$

and

$$f_{EGGa}(x; \alpha, \beta, \theta, \lambda) = \alpha\lambda\theta(\theta x)^{\lambda-1}\Gamma^{-1}(\beta)\left[\gamma(\beta, (\theta x)^{\lambda})\Gamma^{-1}(\beta)\right]^{\alpha-1}e^{-(\theta x)^{\lambda}}, x > 0, \alpha, \beta, \theta, \lambda > 0,$$

respectively.

To determine the best model, we also computed the estimated log-likelihood values $\hat{\ell}$, Akaike Information Criteria (AIC), corrected Akaike information criterion (CAIC), Bayesian information criterion (BIC), Hannan Quinn information criterion (HQIC), Kolmogorov-Smirnov (K-S), Cramer von Mises (W^*) and Anderson-Darling (A^*) goodness of-fit statistics for all distribution models. We note that the statistics W^* and A^* are described in detail in Chen and Balakrishnan (1995). In general, it can be chosen as the best model which has the smaller the values of the AIC, CAIC, BIC, HQIC, K-S, W^* and A^* statistics and the larger the values of $\hat{\ell}$ and p-values. All computations are performed by the maxLike routine in the R programme. The details are the followings.

9.1 Data Modelling for Three Sub-models

The first real data set is breaking strengths of 100 yarn given by Duncan (1974). The data are: 66, 117, 132, 111, 107, 85, 89, 79, 91, 97, 138, 103, 111, 86, 78, 96, 93, 101, 102, 110, 95, 96, 88, 122, 115, 92, 137, 91, 84, 96, 97, 100, 105, 104, 137, 80, 104, 104, 106, 84, 92, 86, 104, 132, 94, 99, 102, 101, 104, 107, 99, 85, 95, 89, 102, 100, 98, 97, 104, 114, 111, 98, 99, 102, 91, 95, 111, 104, 97, 98, 102, 109, 88, 91, 103, 94, 105, 103, 96, 100, 101, 98, 97, 97, 101, 102, 98, 94, 100, 98, 99, 92, 102, 87, 99, 62, 92, 100, 96, 98. This data also analyzed by Tahir et al. (2017).

The second data, studied by Meeker and Escobar (1998, p. 383), gives the times of failure and running times for a sample of devices from a field-tracking study of a larger system. The data are: 275, 13, 147, 23, 181, 30, 65, 10, 300, 173, 106, 300, 300, 212, 300, 300, 300, 2, 261, 293, 88, 247, 28, 143, 300, 23, 300, 80, 245, 266. This data also analyzed by Cordeiro et al. (2010) and Alexander et al. (2012).

The third data set refers to the lifetimes of 50 industrial devices put on life test at time zero given by Aarset (1987). The data are: 0.1, 0.2, 1.0, 1.0, 1.0, 1.0, 2.0, 3.0, 6.0, 7.0, 11.0, 12.0, 18.0, 18.0, 18.0, 18.0, 18.0, 21.0, 32.0, 36.0, 40.0, 45.0, 45.0, 47.0, 50.0, 55.0, 60.0, 63.0, 63.0, 67.0, 67.0, 67.0, 67.0, 72.0, 75.0, 79.0, 82.0, 82.0, 83.0, 84.0, 84.0, 84.0, 85.0, 85.0, 85.0, 85.0, 86.0, 86.0. These data are also analyzed by Cordeiro et al. (2017).

The results are reported in Table 2 and 3. These Tables clearly show that the distribution models of the OPL-G family model have the smallest values of the AIC, CAIC, BIC, HQIC, K-S, W^* and A^* statistics and have the largest values of the $\hat{\ell}$ and all p-values among the fitted models. Then, OPL-N, OPL-W and OPL-Ga models could be chosen as the best models for the three data sets under the above criteria.

The histograms of the all data sets and the plots of the fitted pdfs and cdfs for all models are shown in Figures 6-8. From these Figures, we see that the OPL-N model fits to first data set as bi-modal shaped. For other data sets, the OPL-W and OPL-Ga models fit to the histograms of the data sets with a more adequate fitting than other models.

The results of LR statistics are shown in Table 4 for three data sets. We can say that the additional parameter of the all OPL-G distribution is essential because we reject the null hypotheses of all three LR tests in favour of the OPL-N, OPL-W and OPL-Ga distributions. Hence, these models provide a better representation of the data sets than the OLN, OLW and OLGa models based on the LR test at the 5% significance level.

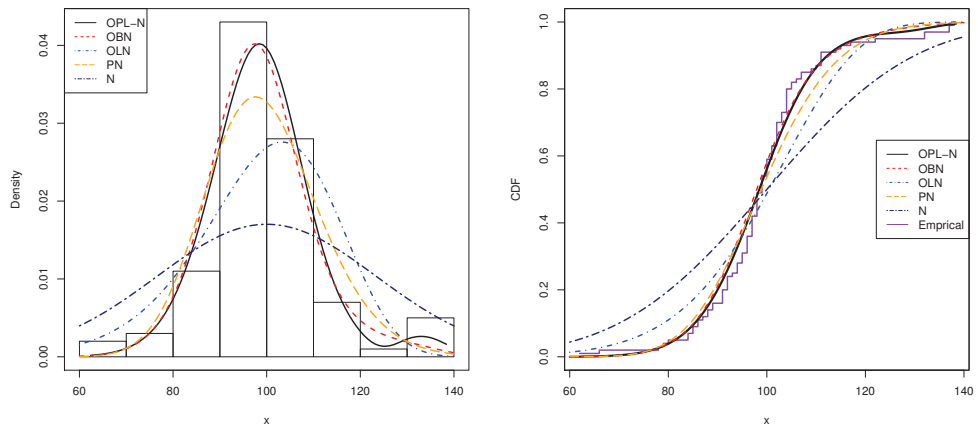


Figure 6. The fitted pdfs and cdfs for the first data sets

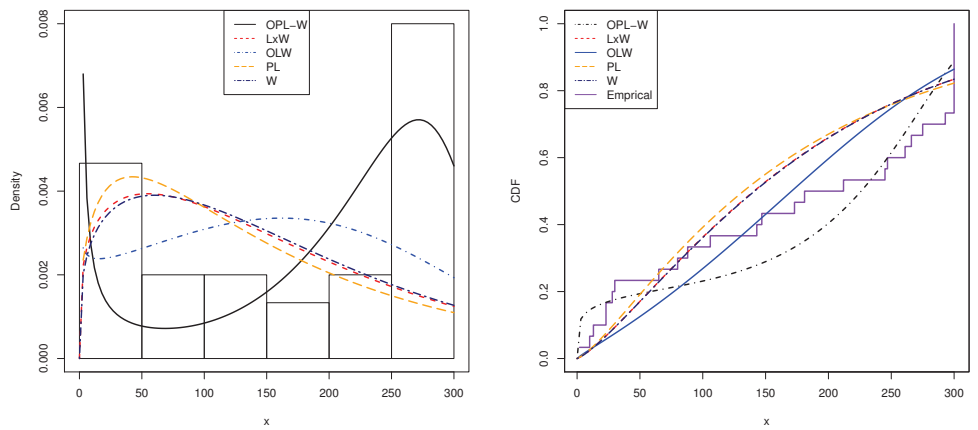


Figure 7. The fitted pdfs and cdfs for the second data sets

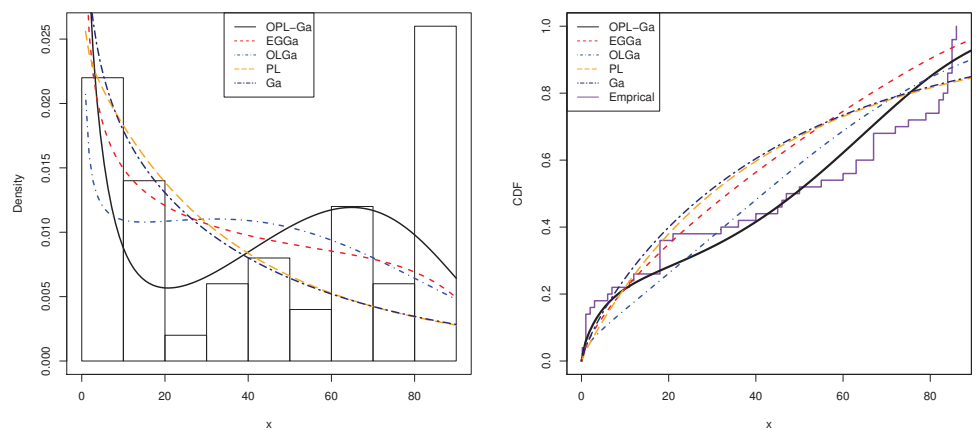


Figure 8. The fitted pdfs and cdfs for the third data sets

Table 2. MLEs, standard errors of the estimates (in parentheses), estimated log-likelihood values ($\hat{\ell}$) and, K-S statistics ([·] denotes its p-value)

Data Set	Model	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\ell}$	$K - S$
I	OPL-N	0.0121 (0.0014)	3.9929 (0.4091)	124.9099 (0.0742)	1.6778 (0.0743)	-380.5249	0.0929 [0.3534]
	OBN	0.1580 (0.0849)	5.1797 (0.9733)	124.2684 (2.7426)	5.9323 (1.8944)	-384.1239	0.1120 [0.1630]
	OLN	1	21.7148 (3.5901)	155.2424 (2.7683)	29.6520 (1.6163)	-404.3403	0.2023 [0.0006]
	PN	–	5.2988 (2.2521)	77.4016 (6.3475)	18.4740 (1.9113)	-391.5384	0.1383 [0.0437]
	N	–	–	99.4300 (1.2672)	12.3969 (0.8961)	-393.6391	0.1562 [0.0152]
II	OPL-W	0.0656 (0.0008)	0.5289 (0.1091)	0.0127 (7e-07)	2.4919 (3e-05)	-177.6348	0.1534 [0.4803]
	LxW	71.3342 (9.7238)	80.9065 (4.1035)	0.0059 (0.0013)	1.2688 (0.2048)	-184.4178	0.2191 [0.1123]
	OLW	1	0.6738 (0.4413)	0.0062 (0.0034)	0.8592 (0.2521)	-179.8946	0.1980 [0.1901]
	PL	0.7837 (0.1077)	0.0357 (0.0208)	–	–	-185.8193	0.2172 [0.1178]
	W	–	–	0.0053 (0.0008)	1.2650 (0.2045)	-184.3138	0.2194 [0.1112]
III	OPL-Ga	0.0946 (0.0104)	0.6416 (0.1002)	0.3736 (0.0090)	6.9166 (0.0059)	-222.3725	0.1288 [0.3773]
	EGGa	0.0066 (0.0003)	11.3194 (1.3198)	0.0130 (0.0009)	9.2635 (1.7648)	-225.3919	0.2133 [0.0211]
	OLGa	1	0.4698 (0.2714)	0.0108 (0.0026)	0.3795 (0.1923)	-230.6980	0.1562 [0.1741]
	PL	0.6640 (0.0685)	0.1612 (0.0448)	–	–	-242.0873	0.1961 [0.0428]
	Ga	–	–	0.0175 (0.0040)	0.7991 (0.1379)	-240.1902	0.2022 [0.0335]

Table 3. Information criteria results, A^* and W^* statistics ($[\cdot]$ and $\{\cdot\}$ denote their p-values)

Data Set	Model	AIC	CAIC	BIC	HQIC	A^*	W^*
I	OPL-N	769.0498	769.4708	779.4705	773.2672	0.8802	0.1514
						[0.4264]	{0.3859}
	OBN	776.2479	776.6689	786.6686	780.4653	1.2697	0.2163
						[0.2424]	{0.2382}
	OLN	814.6805	814.9305	822.4960	817.8436	5.2421	0.9078
					[0.0022]	{0.0039}	
	PN	789.0767	789.3267	796.8922	792.2398	2.1550	0.3624
						[0.0757]	{0.0908}
	N	791.2783	791.4020	796.4886	793.3870	2.6855	0.4426
						[0.0398]	{0.0557}
II	OPL-W	363.2695	364.8695	368.8743	365.0626	1.4369	0.2046
						[0.1923]	{0.2593}
	LxW	376.8356	378.4356	382.4404	378.6286	2.1105	0.3315
						[0.0804]	{0.1101}
	OLW	365.7893	366.7123	369.9928	367.1340	1.9824	0.2750
					[0.0943]	{0.1591}	
	PL	375.6386	376.0830	378.4410	376.5351	2.1183	0.3499
						[0.0796]	{0.0978}
	W	372.6277	373.0721	375.4301	373.5242	2.1098	0.3315
						[0.0804]	{0.1101}
III	OPL-Ga	452.7450	453.6339	460.3931	455.6575	1.2226	0.1613
						[0.2589]	{0.3577}
	EGGa	458.7839	459.6728	466.4320	461.6963	4.0595	0.6168
						[0.0082]	{0.0199}
	OLGa	467.3961	467.9178	473.1321	469.5804	2.3971	0.3402
					[0.0563]	{0.1043}	
	PL	488.1747	488.4300	491.9987	489.6309	3.3958	0.5350
						[0.0175]	{0.0321}
	Ga	484.3804	484.6358	488.2045	485.8367	3.2502	0.5566
						[0.0206]	{0.0282}

Table 4. LR statistics for the data sets

Data Set	Model	Hypothesis	w	p-value
I	OPL – N vs OLN	$H_0 : \alpha = 1, H_1 : H_0$ false	47.6308	5.1e-12
II	OPL – W vs OLW	$H_0 : \alpha = 1, H_1 : H_0$ false	4.5199	0.0335
III	OPL – Ga vs OLGa	$H_0 : \alpha = 1, H_1 : H_0$ false	16.6510	4.5e-05

9.2 Regression Modelling: Stanford Heart Transplant Data

Recently, Brito et al. (2017) introduced the Log-Topp-Leone odd log-logistic-Weibull (Log-TLOLL-W) regression model. Brito et al. (2017) used the Stanford heart transplant data set to prove the usefulness of Log-TLOLL-W regression model. Here, we use the same data set to demonstrate the flexibility of LOPLW regression model against to Log-TLOLL-W regression model. These data set is available in *p3state.msm* package of R software. The sample size is $n = 103$, the percentage of censored observations is 27%. The aim of this study is to relate the survival times (t) of patients with the following explanatory variables: x_1 - year of acceptance to the program; x_2 - age of patient (in years); x_3 - previous surgery status (1 = yes, 0 = no); x_4 - transplant indicator (1 = yes, 0 = no); c_i - censoring indicator (0 =censoring, 1 =lifetime observed).

The regression model fitted to the voltage data set is given by

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \sigma z_i , \tag{16}$$

respectively, where the random variable y_i follows the LOPLW distribution given in (11).

The results for above regression models are presented in Table 5. The MLEs of the model parameters and their SEs, p values and $-\ell$, AIC and BIC statistics are listed in Table 5. Based on the figures in Table 5, LOPLW model has the lowest values of the $-\ell$, AIC and BIC statistics. Therefore, it is clear that LOPLW regression model outperforms among others for these data set. According to results of LOPLW regression model, β_1 and β_2 are statistically significant at 5% level.

Table 5. MLEs of the parameters to Stanford Heart Transplant Data for Log-Weibull, Log-TLOLL-W and LOPLW regression models with corresponding SEs, p -values and $-\ell$, AIC and BIC statistics

Parameters	Models								
	Log-Weibull			Log-TLOLL-W			LOPLW		
	Estimate	S.E.	p-value	Estimate	S.E.	p-value	Estimate	S.E.	p-value
α	-	-	-	2.34	3.546	-	4.713	2.809	-
β	-	-	-	24.029	3.015	-	0.032	0.021	-
σ	1.478	0.133	-	9.68	12.526	-	16.327	8.381	-
β_0	1.639	6.835	0.811	-0.645	8.459	0.939	-0.443	3.334	0.894
β_1	0.104	0.096	0.279	0.074	0.097	0.448	0.253	0.092	0.006
β_2	-0.092	0.02	<0.001	-0.053	0.02	0.009	-0.076	0.018	<0.001
β_3	1.126	0.658	0.087	1.676	0.597	0.005	0.191	0.528	0.717
β_4	2.544	0.378	<0.001	2.394	0.384	<0.001	0.215	0.373	0.564
$-\ell$	171.2405			164.684			161.237		
AIC	354.481			345.368			338.474		
BIC	370.289			366.445			359.552		

Residual Analysis of LOPLW model for Stanford heart transplant data set

Residual analysis is conducted to evaluate the adequacy of the fitted model. For this goal, two type residual are considered: martingale and modified deviance residuals. The martingale residuals is defined in counting process and takes values between +1 and $-\infty$. The martingale residuals for LOPLW model is,

$$r_{M_i} = \begin{cases} 1 + \log \left(1 + \frac{\beta}{\beta+1} \left[\frac{\{1-\exp[-u_i]\}}{2-\exp[-u_i]} \right]^\alpha \right) \exp \left\{ -\beta \left[\frac{\{1-\exp[-u_i]\}}{2-\exp[-u_i]} \right]^\alpha \right\} & \text{if } i \in F, \\ \log \left(1 + \frac{\beta}{\beta+1} \left[\frac{\{1-\exp[-u_i]\}}{2-\exp[-u_i]} \right]^\alpha \right) \exp \left\{ -\beta \left[\frac{\{1-\exp[-u_i]\}}{2-\exp[-u_i]} \right]^\alpha \right\} & \text{if } i \in C, \end{cases} \tag{17}$$

where $u_i = \exp(z_i)$, $z_i = (y_i - v_i^T \beta) / \sigma$. The modified deviance residual, proposed by Therneau et al. (1990), is given by

$$r_{D_i} = \begin{cases} \text{sign}(r_{M_i}) \{ -2 [r_{M_i} + \log(1 - r_{M_i})] \}^{1/2}, & \text{if } i \in F \\ \text{sign}(r_{M_i}) \{ -2r_{M_i} \}^{1/2}, & \text{if } i \in C, \end{cases} \tag{18}$$

where \hat{r}_{M_i} is the martingale residual.

Figure 9 displays the index plot of the modified deviance residuals and its Q-Q plot against to $N(0, 1)$ quantiles for Stanford heart transplant data set. Based on Figure 9, we conclude that none of observed values appears as possible outliers. Therefore, the fitted model is appropriate for these data set.

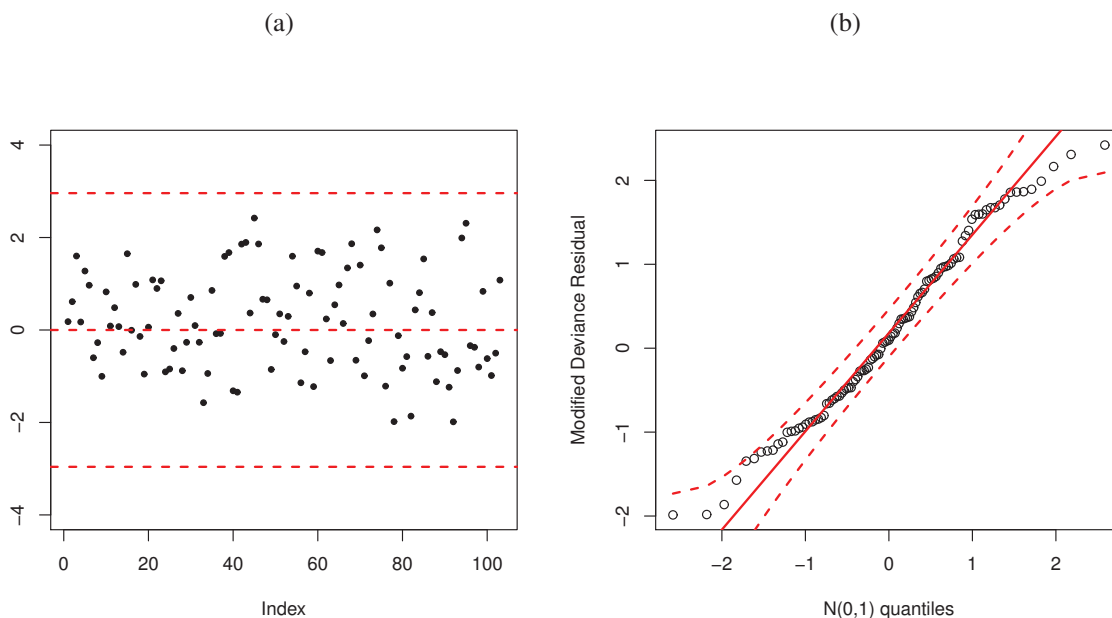


Figure 9. (a) Index plot of the modified deviance residual and (b) Q-Q plot for modified deviance residual

10. Conclusions

In this paper, we proposed a new flexible class of distributions and provided a comprehensive treatment its mathematical properties as well as some useful characterizations. The maximum likelihood method is used to estimate the model parameters, we assess the performance of the maximum likelihood estimators by means of two simulation studies. Also we introduce a new regression model based on a special member of the new family called the log odd power Lindley Weibull distribution. We show that the new log location-scale regression model for lifetime data can be very useful in analysing real data and provide more realistic fits than other regression models. Index plot of the modified deviance residual and Q-Q plot for modified deviance residual are presented to illustrate that our new model is more appropriate to Stanford heart transplant data set than other competitive models like log-Weibull and log-Topp-Leone odd log-logistic-Weibull model. We hope that the results given in this paper will be useful for practitioners and researchers.

Appendix A

Theorem 1. Let $\Omega, \mathcal{F}, \mathbf{P}$ be a given probability space and let $H = [a, b]$ be an interval for some $d < b$ $a = -\infty, b = \infty$ might as well be allowed. Let $X : \Omega \rightarrow H$ be a continuous random variable with the distribution function F and let q_1 and q_2 be two real functions defined on H such that

$$\mathbf{E}[q_2(X) | X \geq x] = \mathbf{E}[q_1(X) | X \geq x] \eta(x), \quad x \in H,$$

is defined with some real function η . Assume that $q_1, q_2 \in C^1(H), \eta \in C^2(H)$ and F is twice continuously differentiable and strictly monotone function on the set H . Finally, assume that the equation $\eta q_1 = q_2$ has no real solution in the interior of H . Then F is uniquely determined by the functions q_1, q_2 and η , particularly

$$F(x) = \int_a^x C \left| \frac{\eta'(u)}{\eta(u) q_1(u) - q_2(u)} \right| \exp(-s(u)) du,$$

where the function s is a solution of the differential equation $s' = \frac{\eta' q_1}{\eta q_1 - q_2}$ and C is the normalization constant, such that $\int_H dF = 1$.

We like to mention that this kind of characterization based on the ratio of truncated moments is stable in the sense of weak convergence (see, Glänzel [2]), in particular, let us assume that there is a sequence $\{X_n\}$ of random variables with distribution functions $\{F_n\}$ such that the functions q_{1n}, q_{2n} and η_n ($n \in \mathbb{N}$) satisfy the conditions of Theorem 1 and let $q_{1n} \rightarrow q_1, q_{2n} \rightarrow q_2$ for some continuously differentiable real functions q_1 and q_2 . Let, finally, X be a random variable with distribution F . Under the condition that $q_{1n}(X)$ and $q_{2n}(X)$ are uniformly integrable and the family $\{F_n\}$ is relatively compact, the sequence X_n converges to X in distribution if and only if η_n converges to η , where

$$\eta(x) = \frac{E[q_2(X) | X \geq x]}{E[q_1(X) | X \geq x]}.$$

This stability theorem makes sure that the convergence of distribution functions is reflected by corresponding convergence of the functions q_1 , q_2 and η , respectively. It guarantees, for instance, the 'convergence' of characterization of the Wald distribution to that of the Lévy-Smirnov distribution if $\alpha \rightarrow \infty$.

A further consequence of the stability property of Theorem 1 is the application of this theorem to special tasks in statistical practice such as the estimation of the parameters of discrete distributions. For such purpose, the functions q_1 , q_2 and, specially, η should be as simple as possible. Since the function triplet is not uniquely determined it is often possible to choose $\xi\eta$ as a linear function. Therefore, it is worth analyzing some special cases which helps to find new characterizations reflecting the relationship between individual continuous univariate distributions and appropriate in other areas of statistics.

References

- Aarset, M. V. (1987). How to identify a bathtub hazard rate. *IEEE Transactions on Reliability*, 36(1), 106-108. <https://doi.org/10.1109/TR.1987.5222310>
- Alexander, C., Cordeiro, G. M., Ortega, E. M., & Sarabia, J. M. (2012). Generalized beta-generated distributions. *Computational Statistics & Data Analysis*, 56(6), 1880-1897. <https://doi.org/10.1016/j.csda.2011.11.015>
- Alizadeh, M., Cordeiro, G. M., Nascimento, A. D., Lima, M. D. C. S., & Ortega, E. M. (2017). Odd-Burr generalized family of distributions with some applications. *Journal of Statistical Computation and Simulation*, 87(2), 367-389. <https://doi.org/10.1080/00949655.2016.1209200>
- Alzaatreh, A., Lee, C., & Famoye, F. (2013). A new method for generating families of continuous distributions. *Metron*, 71, 63-79. <https://doi.org/10.1007/s40300-013-0007-y>
- Bourguignon, M., Silva, R. B., & Cordeiro, G. M. (2014). The Weibull-G family of probability distributions. *Journal of Data Science*, 12(1), 53-68.
- Brito, E., Cordeiro, G. M., Yousof, H. M., Alizadeh, M., & Silva, G. O. (2017). The Topp-Leone odd log-logistic family of distributions. *Journal of Statistical Computation and Simulation*, 1-19. <https://doi.org/10.1080/00949655.2017.1351972>
- Chen, G., & Balakrishnan, N. (1995). A general purpose approximate goodness-of-fit test, *J. Qual. Technol.*, 27, 154-161. <https://doi.org/10.1080/00224065.1995.11979578>
- Cordeiro, G. M., Alizadeh, M., Ozel, G., Hosseini, B., Ortega, E. M. M., & Altun, E. (2017). The generalized odd log-logistic family of distributions: properties, regression models and applications. *Journal of Statistical Computation and Simulation*, 87(5), 908-932. <https://doi.org/10.1080/00949655.2016.1238088>
- Cordeiro, G. M., Ortega, E. M., & Nadarajah, S. (2010). The Kumaraswamy Weibull distribution with application to failure data. *Journal of the Franklin Institute*, 347(8), 1399-1429. <https://doi.org/10.1016/j.jfranklin.2010.06.010>
- Cordeiro, G. M., Ortega, E. M. M., Popovic, B. V., & Pescim, R. R. (2014). The Lomax generator of distributions: Properties, minification process & regression model, *Applied Mathematics and Computation*, 247, 465-486. <https://doi.org/10.1016/j.amc.2014.09.004>
- Cordeiro, G. M., Ortega, E. M., & Silva, G. O. (2011). The exponentiated generalized gamma distribution with application to lifetime data. *Journal of statistical computation and simulation*, 81(7), 827-842. <https://doi.org/10.1080/00949650903517874>
- Cordeiro, G. M., Yousof, H. M., Ramires, T. G., & Ortega, E. M. M. (2018). The Burr XII system of densities: properties, regression model and applications. *Journal of Statistical Computation and Simulation*, 88(3), 432-456. <https://doi.org/10.1080/00949655.2017.1392524>
- Duncan, A. J. (1974). *Quality Control and Industrial Statistics*, Irwin Homewood, USA.
- Ghitany, M. E., Al-Mutairi, D. K., Balakrishnan, N., & Al-Enezi, L. J. (2013). Power Lindley distribution and associated inference. *Computational Statistics & Data Analysis*, 64, 20-33. <https://doi.org/10.1016/j.csda.2013.02.026>
- Glänzel, W. (1987). A characterization theorem based on truncated moments and its application to some distribution families, *Mathematical Statistics and Probability Theory (Bad Tatzmannsdorf, 1986)*, Vol. B, Reidel, Dordrecht, 75-84.

- Glänzel, W. (1990). Some consequences of a characterization theorem based on truncated moments, *Statistics: A Journal of Theoretical and Applied Statistics*, 21(4), 613–618.
- Gupta, R. C., Gupta, P. L., & Gupta, R. D. (1998). Modeling failure time data by Lehmann alternatives. *Commun. Stat. Theory Methods*, 27, 887-904. <https://doi.org/10.1080/03610929808832134>
- Gupta, R. D., & Gupta, R. C. (2008). Analyzing skewed data by power normal model. *Test*, 17(1), 197-210. <https://doi.org/10.1007/s11749-006-0030-x>
- Korkmaz, M. Ç., & Genç, A. I. (2017). A new generalized two-sided class of distributions with an emphasis on two-sided generalized normal distribution. *Communications in Statistics - Simulation and Computation*, 46(2), 1441-1460. <https://doi.org/10.1080/03610918.2015.1005233>
- Nadarajah, S., & Gupta, A. K. (2007). The exponentiated gamma distribution with application to drought data. *Calcutta Statistical Association Bulletin*, 59, 29-54.
- Silva, F. S., Percontini, A., de Brito, E., Ramos, M. W., Venancio, R., & Cordeiro, G. M. (2017). The Odd Lindley-G Family of Distributions. *Austrian Journal of Statistics*, 46(1), 65-87. <https://doi.org/10.17713/ajs.v46i1.222>
- Shirke, D. T., & Kakade, C. S. (2006). On exponentiated lognormal distribution. *International Journal of Agricultural and Statistical Sciences*, 2, 319-326.
- Therneau, T. M., Grambsch, P. M., & Fleming, T. R. (1990). Martingale-based residuals for survival models. *Biometrika*, 77(1), 147-160. <https://doi.org/10.1093/biomet/77.1.147>
- Tahir, M. H., Zubair, M., Cordeiro, G. M., Alzaatreh, A., & Mansoor, M. (2017). The Weibull-Power Cauchy Distribution: Model, Properties and Applications. *Hacetatepe Journal of Mathematics and Statistics*.
- Yousof, H. M., Afify, A. Z., Hamedani, G. G., & Aryal, G. (2016). the Burr X generator of distributions for lifetime data. *Journal of Statistical Theory and Applications*, 16, 1-19.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).