WEIBULL LINDLEY DISTRIBUTION

Authors: A. ASGHARZADEH
– Department of Statistics, University of Mazandaran,
Babolsar, Iran

S. NADARAJAH
– School of Mathematics, University of Manchester,
Manchester M13 9PL, UK
mbbsssn2@manchester.ac.uk

F. SHARAFI
– Department of Statistics, University of Mazandaran,
Babolsar, Iran

Abstract:

• A new distribution is introduced based on compounding Lindley and Weibull distributions. This distribution contains Lindley and Weibull distributions as special cases. Several properties of the distribution are derived including the hazard rate function, moments, moment generating function, and Lorenz curve. An estimation procedure by the method of maximum likelihood and a simulation study to assess its performance are given. Three real data applications are presented to show that the new distribution fits better than known generalizations of the Lindley distribution.

1.   INTRODUCTION

The Lindley distribution was first proposed by Lindley [20] in the context of fiducial and Bayesian inference. In recent years, this distribution has been studied and generalized by several authors, see Ghitany *et al.* [15], Zalerzadeh and Dolati [29], Ghitany *et al.* [14], Bakouch *et al.* [4], Barreto-Souza and Bakouch [5] and Ghitany *et al.* [13].

In this paper, we introduce a new generalization of the Lindley distribution referred to as the Weibull Lindley (WL) distribution by compounding Lindley and Weibull distributions. The compounding approach gives new distributions that extend well-known families of distributions and at the same time offer more flexibility for modeling lifetime data. The flexibility of such compound distributions comes in terms of one or more hazard rate shapes, that may be decreasing or increasing or bathtub shaped or upside down bathtub shaped or unimodal.

Many recent distributions have been introduced by using a compounding approach. For example, Adamidis and Loukas [1] proposed a distribution by taking the minimum of N independent and identical exponential random variables, where N is a geometric random variable. But this distribution allows for only decreasing hazard rates. Kus [18] proposed a distribution by taking the minimum of N independent and identical exponential random variables, where N is a Poisson random variable. But this distribution also allows for only decreasing hazard rates. Barreto-Souza *et al.* [6] proposed a distribution by taking the minimum of N independent and identical Weibull random variables, where N is a geometric random variable. But this distribution does not allow for bathtub shaped hazard rates, the most realistic hazard rates. Morais and Barreto-Souza [22] proposed a distribution by taking the minimum of N independent and identical Weibull random variables, where N is a power series random variable. But this distribution also does not allow for bathtub shaped hazard rates. Asgharzadeh *et al.* [3] proposed a distribution by taking the minimum of N independent and identical Pareto type II random variables, where N is a Poisson random variable. But this distribution allows for only decreasing hazard rates. Silva *et al.* [27] proposed a distribution by taking the minimum of N independent and identical extended Weibull random variables, where N is a power series random variable. This distribution does allow for bathtub shaped hazard rates, but that is expected since the extended Weibull distribution contains as particular cases many generalizations of the Weibull distribution. Bourguignon *et al.* [7] proposed a distribution by taking the minimum of N independent and identical Birnbaum–Saunders random variables, where N is a power series random variable. But this distribution does not allow for bathtub shaped hazard rates.

The WL distribution introduced here is obtained by compounding just two random variables (Lindley and Weibull random variables). Besides the WL dis-

tribution has just three parameters, less than several of the distributions cited above.

Let Y denote a Lindley random variable with parameter $\lambda > 0$ and survival function $\overline{G}(y) = \frac{1+\lambda+\lambda y}{1+\lambda}e^{-\lambda y}$, $y > 0$. Let Z denote a Weibull random variable with parameters $\alpha > 0$ and $\beta > 0$, and survival function $\overline{Q}(z) = e^{-(\beta z)^\alpha}$, $z > 0$. Assume Y and Z are independent random variables. We define $X = \min(Y, Z)$ as a WL random variable and write $X \sim WL(\alpha, \beta, \lambda)$. The survival function of X is

$$\overline{F}(x) = \overline{G}(x)\overline{Q}(x).$$

The cumulative distribution function (cdf) of X can be written as

$$(1.1) \qquad F(x) = 1 - \frac{1 + \lambda + \lambda x}{1 + \lambda}e^{-\lambda x - (\beta x)^\alpha}$$

for $x > 0$, $\alpha > 0$, $\beta \geq 0$ and $\lambda \geq 0$. The probability density function (pdf) of X is

$$(1.2) \quad f(x) = \frac{1}{1+\lambda}\left[\alpha\lambda(\beta x)^\alpha + \alpha\beta(1+\lambda)(\beta x)^{\alpha-1} + \lambda^2(1+x)\right]e^{-\lambda x - (\beta x)^\alpha}$$

for $x > 0$, $\alpha > 0$, $\beta \geq 0$ and $\lambda \geq 0$.

Some special cases of the WL distribution are: the Weibull distribution with parameters α and β for $\lambda = 0$; the Rayleigh distribution with parameter α for $\lambda = 0$ and $\beta = 2$; the exponential distribution with parameter β for $\lambda = 0$ and $\alpha = 1$; the Lindley distribution with parameter λ for $\beta = 0$.

The WL distribution can be used very effectively for analyzing lifetime data. Some possible motivations for the WL distribution are:

- The WL distribution accommodates different hazard rate shapes, that may be decreasing or increasing or bathtub shaped, see Figure 2. Bathtub shaped hazard rates are very important in practice. None of the known generalizations of the Lindley distribution accommodate a bathtub shaped hazard rate function.

- The WL distribution has closed form expressions for survival and hazard rate functions, which is not the case for some generalizations of the Lindley distribution. Hence, the likelihood function for the WL distribution takes explicit forms for ordinary type-II censored data and progressively type-II censored data. Hence, the WL distribution could be a suitable model to analyse ordinary type-II censored data and progressively type-II censored data.

- The Lindley and Weibull distributions are special cases of the WL distribution.

- Suppose a system is composed of two independent components in series; let Y and Z denote their lifetimes; suppose Y is a Lindley random variable and Z is a Weibull random variable; then the lifetime of the system is a WL random variable.

- Suppose a system is composed of n independent components in series; let $Y, Z_1, Z_2, ..., Z_{n-1}$ denote their lifetimes; suppose Y is a Lindley random variable and $Z_1, Z_2, ..., Z_{n-1}$ are identical Weibull random variables; then the lifetime of the system is also a WL random variable.

- The pdf of the WL distribution can be bimodal, see Figure 1. This is not the case for the Weibull distribution or any generalization of the Lindley distribution. So, any bimodal data set (see Figure 8 for example) cannot be adequately modeled by any of the known generalizations of the Lindley distribution.

- Additive hazard rates arise in many practical situations, for example, event-history analysis (Yamaguchi [28]), modeling of excess mortalities (Gail and Benichou [12], page 391), modeling of breast cancer data (Cadarso-Suarez *et al.* [8]), modelling of hazard rate influenced by periodic fluctuations of temperature (Nair *et al.* [24], page 268), and "biologic" and "statistical" interactions in epidemiology (Andersen and Skrondal [2]). Hence, it is import to have distributions based on additive hazard rates. The WEL distribution is the first generalization of the Lindley distribution based on additive hazard rates.

The rest of this paper is organized as follows: various mathematical properties of the WL distribution are derived in Sections 2 to 4; estimation and simulation procedures for the WL distribution are derived in Section 5; three real data applications are illustrated in Section 6.

Some of the mathematical properties derived in Sections 2 to 4 involve infinite series: namely, (3.1), (3.2) and (4.1). Extensive computations not reported here showed that the relative errors between (3.1), (3.2) and (4.1) and their versions with the infinite series in each truncated at twenty did not exceed 10^{-20}. This shows that (3.1), (3.2) and (4.1) can be computed for most practical uses with their infinite sums truncated at twenty. The computations were performed using Maple. Maple took only a fraction of a second to compute the truncated versions of (3.1), (3.2) and (4.1). The computational times for the truncated versions were significantly smaller than those for the untruncated versions and those based on numerical integration.

Throughout this paper, we report conclusions on various properties of the WL distribution: the last four paragraphs of Section 2.1 reporting conclusions on the shape of the pdf of the WL distribution; the last paragraph of Section 2.2 reporting conclusions on the shape of the hazard rate function of the WL dis-

tribution; Section 2.3 reporting conclusions on the shape of the quartiles of the
WL distribution; the last paragraph of Section 3 reporting conclusions on the
mean, variance, skewness and kurtosis of the WL distribution; the last paragraph
of Section 4 reporting conclusions on the Lorenz curve of the WL distribution.
These conclusions are the result of extensive graphical analyses based on a wide
range of parameter values (although the graphics presented here are based on a
few choices of parameter values). However, we have no analytical proofs for these
conclusions.

2.   SHAPES

Here, we study the shapes of the pdf, (1.2), the corresponding hazard rate
function and the corresponding quartiles. Shape properties are important because
they allow the practitioner to see if the distribution can be fitted to a given data
set (this can be seen, for example, by comparing the shape of the histogram of
the data with possible shapes of the pdf). Shape properties of the hazard rate
function are useful to see if the distribution can model increasing failure rates,
decreasing failure rates or bathtub shaped failure rates. Shape properties of the
hazard rate function has implications, for example, to the design of safe systems
in a wide variety of applications. Quartiles are fundamental for estimation (for
example, quartile estimators) and simulation.

2.1. Shape of probability density function

We can see from (1.2) that

$$
\lim_{x \to 0} f(x) = \begin{cases} \infty, & \alpha < 1, \\ \dfrac{\beta(1+\lambda) + \lambda^2}{1+\lambda}, & \alpha = 1, \\ \dfrac{\lambda^2}{1+\lambda}, & \alpha > 1 \end{cases}
$$

and

$$
f(x) \sim \begin{cases} \dfrac{\lambda^2}{1+\lambda} x e^{-\lambda x - (\beta x)^\alpha}, & \alpha < 1, \\ \dfrac{(\beta + \lambda)\lambda}{1+\lambda} x e^{-\lambda x - \beta x}, & \alpha = 1, \\ \dfrac{\alpha \beta^\alpha \lambda}{1+\lambda} e^{-\lambda x - (\beta x)^\alpha}, & \alpha > 1 \end{cases}
$$

as $x \to \infty$.

Note also that $f(x)$ can be written as

$$f(x) = g(x)\overline{Q}(x) + q(x)\overline{G}(x),$$

where

$$g(x) = \frac{\lambda^2(1+x)}{1+\lambda}e^{-\lambda x}$$

and

$$q(x) = \alpha\beta^\alpha x^{\alpha-1}e^{-(\beta x)^\alpha}.$$

So, the first derivative of $f(x)$ is

$$f'(x) = g'(x)\overline{Q}(x) + q'(x)\overline{G}(x) - 2g(x)q(x).$$

Therefore, $f(x)$ is decreasing if $g'(x) < 0$ and $q'(x) < 0$. This is possible if $\lambda \geq 1$ and $\alpha \leq 1$.

The first derivative of $f(x)$ is

$$f'(x) = \frac{e^{-\lambda x - (\beta x)^\alpha}}{1+\lambda}\left[(1+\lambda+\lambda x)\left[\alpha\beta^2(\alpha-1)(\beta x)^{\alpha-2} - (\alpha\beta)^2(\beta x)^{2\alpha-2}\right]\right.$$
$$\left. -2\alpha\beta\lambda^2(\beta x)^{\alpha-1} + \lambda^2(1-\lambda-\lambda x)\right].$$

So, the modes of $f(x)$ at say $x = x_0$ are the roots of

$$(2.1) \quad (1+\lambda+\lambda x)\left[\alpha\beta^2(\alpha-1)(\beta x)^{\alpha-2} - (\alpha\beta)^2(\beta x)^{2\alpha-2}\right] =$$
$$= 2\alpha\beta\lambda^2(\beta x)^{\alpha-1} - \lambda^2(1-\lambda-\lambda x).$$

The roots of (2.1) are difficult to find in general. However, if $\beta = 0$ then $x_0 = \frac{1-\lambda}{\lambda}$, the mode of the Lindley distribution, for $0 < \lambda < 1$.

We now study (2.1) graphically. Figure 1 shows possible shapes of the pdf of the WL distribution for selected (α, β, λ).

The left plot in Figure 1 shows bimodal shapes of the pdf with a maximum followed by a minimum. The x coordinates of the (local minimum, local maximum) are $(0.395, 0.933)$ for $\lambda = 1.2$, $(0.451, 0.921)$ for $\lambda = 1.4$, $(0.500, 0.906)$ for $\lambda = 1.6$ and $(0.547, 0.889)$ for $\lambda = 1.8$. The location of the minimum moves more to the right with increasing values of λ. The location of the maximum moves more to the left with increasing values of λ.

The right plot in Figure 1 shows unimodal shapes of the pdf. The x coordinates of the mode are 0.370 for $\lambda = 1.2$, 0.298 for $\lambda = 1.4$, 0.227 for $\lambda = 1.6$ and 0.157 for $\lambda = 1.8$. The location of the mode moves more to the left with increasing values of λ.
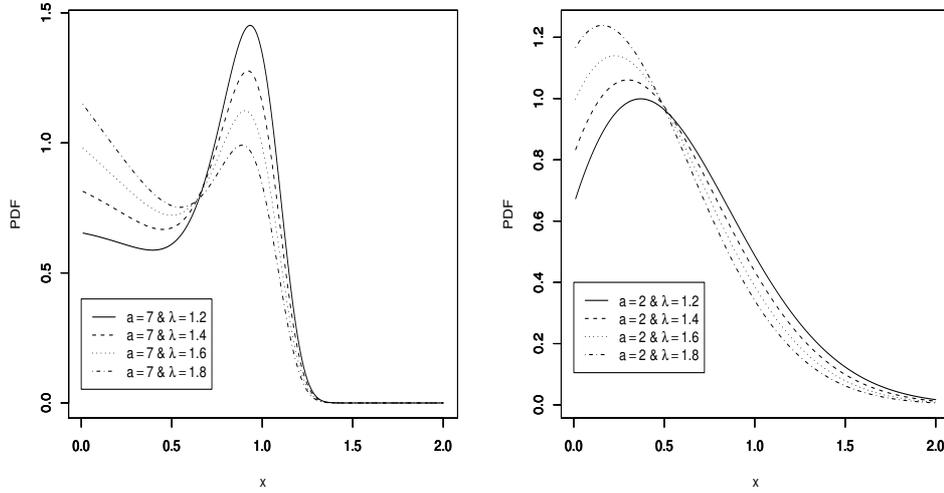
Figure 1:   Pdfs of the WL distribution for $\beta = 1$.

Monotonically decreasing shapes and monotonically decreasing shapes containing an inflexion point are also possible for the pdf.

In each plot, the upper tails of the pdf become lighter with increasing values of λ. The lower tails of the pdf become heavier with increasing values of λ.

2.2. Shape of hazard rate function

Using (1.1) and (1.2), the hazard rate function of the WL distribution can be obtained as

$$(2.2) \qquad h(x) = \frac{\lambda^2(1+x)}{1+\lambda+\lambda x} + \alpha\beta(\beta x)^{\alpha-1}.$$

It is obvious that the hazard rate functions of Lindley and Weibull distributions are contained as particular cases for $\beta = 0$ and $\lambda = 0$, respectively. Also, (2.2) can be expressed as

$$h_X(x) = h_Y(x) + h_Z(x),$$

i.e., the hazard rate function of the WL distribution is the sum of the hazard rate functions of Lindley and Weibull distributions. As a result, the hazard rate function of the WL distribution can exhibit monotonically increasing, monotonically decreasing and bathtub shapes, see Figure 2.

We can see from (2.2) that

$$\lim_{x \to 0} h(x) = \begin{cases} \infty, & \alpha < 1, \\ \dfrac{\beta(1+\lambda) + \lambda^2}{1+\lambda}, & \alpha = 1, \\ \dfrac{\lambda^2}{1+\lambda}, & \alpha > 1 \end{cases}$$

and

$$\lim_{x \to \infty} h(x) = \begin{cases} \lambda, & \alpha < 1, \\ \lambda + \beta, & \alpha = 1, \\ \infty, & \alpha > 1. \end{cases}$$

Bathtub shapes of the the hazard rate function appear possible when α is close enough to 0 or α is close enough to 1, see Figure 2. Monotonically decreasing shapes are possible for all values of α in between (i.e., in between α being close enough to 0 and α being close enough to 1). Monotonically increasing shapes are possible for all other values of α.
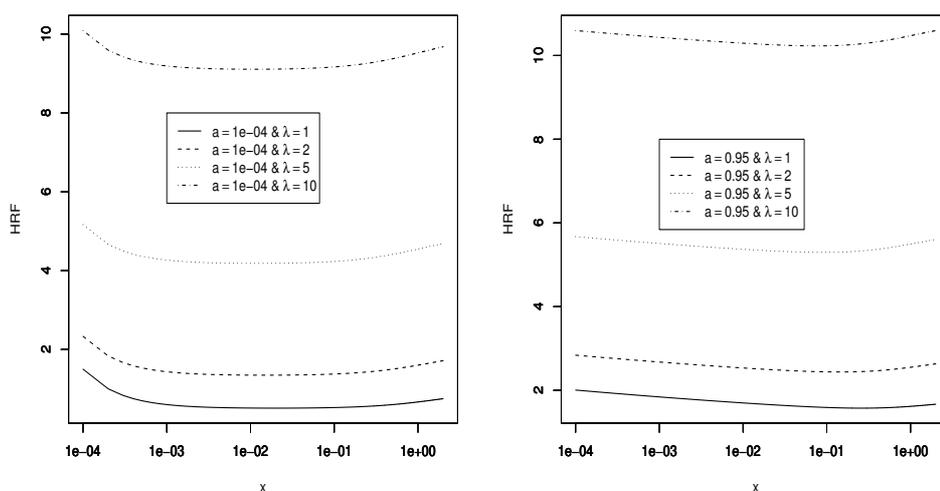

Figure 2: Hazard rate function of the WL distribution for $\beta = 1$.

2.3. Shape of quartiles

The p-th quartile say x_p of a WL random variable defined by $F(x_p) = p$ is the root of

$$x_p = \frac{1+\lambda}{\lambda}\left[(1-p)e^{\lambda x_p + (\beta x_p)^\alpha} - 1\right]$$

for $0 < p < 1$. Numerical investigations showed that x_p are monotonic decreasing functions of λ and monotonic increasing functions of α except for high quartiles.

3. MOMENT GENERATING FUNCTION AND MOMENTS

The moment generating function is fundamental for computing moments, factorial moments and cumulants of a random variable. It can also be used for estimation (for example, estimation methods based on empirical moment generating functions).

Moments are fundamental for any distribution. For instance, the first four moments can be used to describe any data fairly well. Moments are also useful for estimation (for example, the method of moments).

Several interesting characteristics of a distribution can be studied by moments and moment generating function. Let $X \sim WL(\alpha, \beta, \lambda)$. Then the moment generating function of X can be expressed as

$$
\begin{aligned}
M_X(t) \;=\; & E\left(e^{tX}\right) \\
=\; & \sum_{i=0}^{\infty} \frac{(-1)^i \beta^{i\alpha}}{i!(\lambda-t)^{i\alpha}} \left\{ \frac{\lambda^2 \Gamma(i\alpha+1)}{(1+\lambda)(\lambda-t)} + \frac{\lambda^2 \Gamma(i\alpha+2)}{(1+\lambda)(\lambda-t)^2} \right. \\
& \left. + \frac{\alpha\beta^\alpha \Gamma\left(\alpha(i+1)\right)}{(\lambda-t)^\alpha} + \frac{\alpha\lambda\beta^\alpha \Gamma\left(\alpha(i+1)+1\right)}{(1+\lambda)(\lambda-t)^{\alpha+1}} \right\},
\end{aligned}
$$

(3.1)

where $\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt$ denotes the gamma function. The r-th raw moment of X can be expressed as

$$
\begin{aligned}
\mu_r' = E\left(X^r\right) \;=\; & \frac{\lambda^2}{1+\lambda} \sum_{i=0}^{\infty} \frac{(-1)^i \beta^{i\alpha}}{i!\lambda^{i\alpha+r+1}} \left[\Gamma(i\alpha+r+1) + \frac{\Gamma(i\alpha+r+2)}{\lambda} \right] \\
& + \alpha\beta^\alpha \sum_{i=0}^{\infty} \frac{(-1)^i \beta^{i\alpha}}{i!\lambda^{i\alpha+\alpha+r}} \left[\Gamma(i\alpha+\alpha+r) + \frac{\Gamma(i\alpha+\alpha+r+1)}{1+\lambda} \right].
\end{aligned}
$$

(3.2)

The central moments, coefficient of variation, skewness and kurtosis of X can be readily obtained using the raw moments of X. Numerical investigations of the behavior of the mean, variance, skewness and kurtosis versus α and λ showed the following: i) mean is a monotonic decreasing function of λ; ii) mean is either a monotonic increasing function of α or initially decreases before increasing with respect to α; iii) variance is either a monotonic decreasing function of λ or initially increases before decreasing with respect to λ; iv) variance is either a monotonic decreasing function of α or a monotonic increasing function of α; v) skewness is either a monotonic decreasing function of λ or initially decreases before increasing with respect to λ; vi) skewness is either a monotonic decreasing function of α or initially increases before decreasing with respect to α; vii) kurtosis is either a monotonic decreasing function of λ, a monotonic increasing function of λ, initially decreases before increasing with respect to λ, or initially increases, then decreases

before increasing with respect to λ; viii) kurtosis is either a monotonic decreasing function of α, initially increases before decreasing with respect to α, initially decreases before increasing with respect to α, or initially increases, then decreases before increasing with respect to α.

4.   LORENZ CURVE

The Lorenz curve for a positive random variable X is defined as the graph of the ratio

$$L\left(F(x)\right) = \frac{E(X \mid X \le x)F(x)}{E(x)} = \frac{\int_0^x xf(x)dx}{\int_0^{+\infty} xf(x)dx}$$

against $F(x)$ with the property $L(p) \le p$, $L(0) = 0$ and $L(1) = 1$. If X represents annual income, $L(p)$ is the proportion of total income that accrues to individuals having the $100p$ percent lowest incomes. If all individuals earn the same income then $L(p) = p$ for all p. The area between the line $L(p) = p$ and the Lorenz curve may be regarded as a measure of inequality of income, or more generally, of the variability of X.

The Lorenz curve has also received applications in areas other than income modeling: hierarchy theory for digraphs (Egghe [9]); depression and cognition (Maldonado *et al.* [21]); disease risk to optimize health benefits under cost constraints (Gail [11]); seasonal variation of environmental radon gas (Groves-Kirkby *et al.* [16]); statistical nonuniformity of sediment transport rate (Radice [26]).

For the WL distribution, we have

$$\begin{aligned}
\int_0^x xf(x)dx &= \frac{1}{1+\lambda} \sum_{i=0}^{\infty} \frac{(-1)^i \beta^{i\alpha}}{i!\lambda^{i\alpha}} \gamma(i\alpha + 2, \lambda x) \\
&+ \frac{1}{1+\lambda} \sum_{i=0}^{\infty} \frac{(-1)^i \beta^{i\alpha}}{i!\lambda^{i\alpha+1}} \gamma(i\alpha + 3, \lambda x) \\
&+ \alpha\beta \sum_{i=0}^{\infty} \frac{(-1)^i \beta^{i\alpha+\alpha-1}}{i!\lambda^{i\alpha+\alpha+1}} \gamma(i\alpha + \alpha + 1, \lambda x) \\
&+ \frac{\alpha}{1+\lambda} \sum_{i=0}^{\infty} \frac{(-1)^i \beta^{i\alpha+\alpha}}{i!\lambda^{i\alpha+\alpha+1}} \gamma(i\alpha + \alpha + 2, \lambda x),
\end{aligned}$$

where $\gamma(a, x) = \int_0^x t^{a-1}e^{-t}dt$ denotes the incomplete gamma function. So, the

Lorenz curve for the WL distribution is

$$
\begin{aligned}
L(p) \;=\; \frac{1}{\mu}\Bigg[& \sum_{i=0}^{\infty}\frac{(-1)^i\beta^{i\alpha}}{i!\lambda^{i\alpha}}\gamma(i\alpha+2,\lambda x) + \frac{1}{1+\lambda}\sum_{i=0}^{\infty}\frac{(-1)^i\beta^{i\alpha}}{i!\lambda^{i\alpha+1}}\gamma(i\alpha+3,\lambda x) \\
& +\alpha\beta\sum_{i=0}^{\infty}\frac{(-1)^i\beta^{i\alpha+\alpha-1}}{i!\lambda^{i\alpha+\alpha+1}}\gamma(i\alpha+\alpha+1,\lambda x)
\end{aligned}
$$

$$
(4.1)\qquad +\frac{\alpha}{1+\lambda}\sum_{i=0}^{\infty}\frac{(-1)^i\beta^{i\alpha+\alpha}}{i!\lambda^{i\alpha+\alpha+1}}\gamma(i\alpha+\alpha+2,\lambda x)\Bigg].
$$

Possible shapes of (4.1) versus α and λ are shown in Figure 3. When $\alpha = 0.5$, the curves bend further towards the diagonal line as λ increases. When $\alpha = 1$, the curves bend further away from the diagonal line as λ increases. For each fixed λ, the curves bend further towards the diagonal line as α increases.
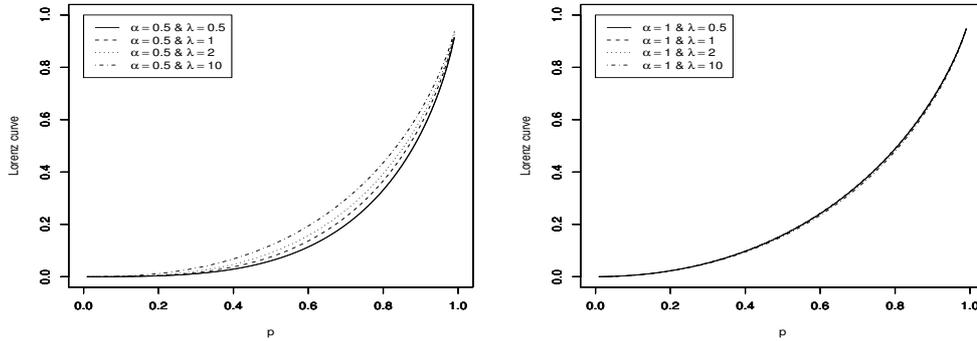

Figure 3:   Lorenz curve of the WL distribution for $\beta = 1$.

5.   ESTIMATION AND SIMULATION

Maximum likelihood estimation of the three parameters of the WL distribution is considered in Section 5.1. An assessment of the performance of the maximum likelihood estimators is performed in Section 5.2. Maximum likelihood estimation of the three parameters in the presence of censored data is considered in Section 5.3. A scheme for simulating from the WL distribution is given in Section 5.4.

5.1. Maximum likelihood estimation

Suppose $x_1, ..., x_n$ is a random sample from the WL distribution. The log-likelihood function is

$$\ell(\alpha, \beta, \lambda) = \sum_{i=0}^{n} \log \left[\lambda^2 (1 + x_i) + \alpha\lambda (\beta x_i)^\alpha + \alpha\beta(1 + \lambda) (\beta x_i)^{\alpha-1} \right] - \lambda \sum_{i=0}^{n} x_i$$

$$(5.1) \qquad - \sum_{i=0}^{n} (\beta x_i)^\alpha - n \log(1 + \lambda).$$

The maximum likelihood estimators of (α, β, λ) can be obtained by solving the likelihood equations

$$\frac{\partial \ell(\alpha, \beta, \lambda)}{\partial \alpha} = \sum_{i=0}^{n} \frac{1}{W(x_j)} \left\{ [\beta(1 + \lambda) + \alpha\beta(1 + \lambda) \log (\beta x_i)] (\beta x)^{\alpha-1} \right.$$

$$(5.2) \qquad \left. + [\lambda + \alpha\lambda \log (\beta x_i)] (\beta x_i)^\alpha \right\} - \lambda \sum_{i=0}^{n} (\beta x)^\alpha \log (\beta x_i) = 0,$$

$$\frac{\partial \ell(\alpha, \beta, \lambda)}{\partial \beta} = \sum_{i=0}^{n} \frac{1}{\beta W(x_i)} \left[\alpha\beta (1 + x_i) (\beta x_i)^{\alpha-1} + \alpha^2 \lambda (\beta x_i)^\alpha \right]$$

$$(5.3) \qquad - \frac{\alpha}{\beta} \sum_{i=0}^{n} (\beta x_i)^\alpha = 0$$

and

$$\frac{\partial \ell(\alpha, \beta, \lambda)}{\partial \lambda} = \sum_{i=0}^{n} \frac{1}{W(x_i)} \left[\alpha (\beta x_i)^\alpha + \alpha\beta (\beta x_i)^{\alpha-1} + 2\lambda (1 + x_i) \right]$$

$$(5.4) \qquad - \sum_{i=0}^{n} x_i - \frac{n}{1 + \lambda} = 0,$$

where $W(x) = \lambda^2(1+x) + \alpha\lambda(\beta x)^\alpha + \alpha\beta(1+\lambda)(\beta x)^{\alpha-1}$. Alternatively, the MLEs can be obtained by maximizing (5.1) numerically. We shall use the latter approach in Sections 5.2 and 6. The maximization was performed by using the nlm function in R (R Development Core Team [25]). In Sections 5.2 and 6, the function nlm was executed with the initial values taken to be:

(i) The true parameter values (applicable for Section 5.2 only);

(ii) $\alpha = 0.01, 0.02, ..., 10$, $\beta = 0.01, 0.02, ..., 10$ and $\lambda = 0.01, 0.02, ..., 10$;

(iii) The moments estimates, i.e., the solutions $E(X) = (1/n) \sum_{i=1}^{n} x_i$, $E(X^2) = (1/n) \sum_{i=1}^{n} x_i^2$ and $E(X^3) = (1/n) \sum_{i=1}^{n} x_i^3$, where $E(X)$,

$E\left(X^2\right)$ and $E\left(X^3\right)$ are given by (3.2). These equations do not give explicit solutions. They were solved numerically using a quasi-Newton algorithm. Numerical investigations showed that this involved roughly the same amount of time as solving of $\frac{\partial \ell(\alpha,\beta,\lambda)}{\partial \alpha} = 0$, $\frac{\partial \ell(\alpha,\beta,\lambda)}{\partial \beta} = 0$ and $\frac{\partial \ell(\alpha,\beta,\lambda)}{\partial \lambda} = 0$ using a quasi-Newton algorithm.

In the cases of i) and iii), the function nlm converged all the time and the MLEs were unique. In the case of ii), the MLEs were unique whenever the function nlm converged. In the case of ii), the function nlm did not converge about five percent of the time.

For interval estimation of (α, β, λ), we consider the observed Fisher information matrix given by

$$I_F(\alpha, \beta, \lambda) = - \begin{pmatrix} I_{\alpha\alpha} & I_{\alpha\beta} & I_{\alpha\lambda} \\ I_{\beta\alpha} & I_{\beta\beta} & I_{\beta\lambda} \\ I_{\lambda\alpha} & I_{\lambda\beta} & I_{\lambda\lambda} \end{pmatrix},$$

where $I_{\phi_1\phi_2} = \partial^2 \ell / \partial\phi_1 \partial\phi_2$.

Under certain regularity conditions (see, for example, Ferguson [10]) and Lehmann and Casella [19], pages 461-463) and for large n, the distribution of $\sqrt{n}\left(\widehat{\alpha} - \alpha, \widehat{\beta} - \beta, \widehat{\lambda} - \lambda\right)$ can be approximated by a trivariate normal distribution with zero means and variance-covariance matrix given by the inverse of the observed information matrix evaluated at the maximum likelihood estimates. This approximation can be used to construct approximate confidence intervals, confidence regions, and testing hypotheses for the parameters. For example, an asymptotic confidence interval for α with level $1 - \gamma$ is $\left(\widehat{\alpha} \mp z_{1-\gamma/2}\sqrt{I^{\widehat{\alpha},\widehat{\alpha}}}\right)$, where $I^{\widehat{\alpha},\widehat{\alpha}}$ is the $(1,1)$-th element of the inverse of $I_F\left(\widehat{\alpha}, \widehat{\beta}, \widehat{\lambda}\right)$ and $z_{1-\gamma/2}$ is the $(1 - \gamma/2)$-th quartile of the standard normal distribution.

5.2. Simulation study

Here, we assess the performance of the maximum likelihood estimators given by (5.2)–(5.4) with respect to sample size n. The assessment was based on a simulation study:

1. Generate ten thousand samples of size n from (1.2). The inversion method was used to generate samples, i.e., variates of the WL distribution were generated using

$$U = \frac{1 + \lambda}{\lambda}\left[(1 - p)e^{\lambda X + (\beta X)^\alpha} - 1\right],$$

where $U \sim U(0, 1)$ is a uniform variate on the unit interval.

2. Compute the maximum likelihood estimates for the ten thousand samples, say $\left(\widehat{\alpha}_i, \widehat{\beta}_i, \widehat{\lambda}_i\right)$ for $i = 1, 2, ..., 10000$.

3. Compute the biases and mean squared errors given by

$$\text{bias}_h(n) = \frac{1}{10000} \sum_{i=1}^{10000} \left(\widehat{h}_i - h\right), \quad \text{MSE}_h(n) = \frac{1}{10000} \sum_{i=1}^{10000} \left(\widehat{h}_i - h\right)^2$$

for $h = \alpha, \beta, \lambda$.

We repeated these steps for $n = 10, 11, ..., 100$ with $\alpha = 1$, $\beta = 1$ and $\lambda = 1$, so computing $\text{bias}_h(n)$ and $\text{MSE}_h(n)$ for $h = \alpha, \beta, \lambda$ and $n = 10, 11, ..., 100$.
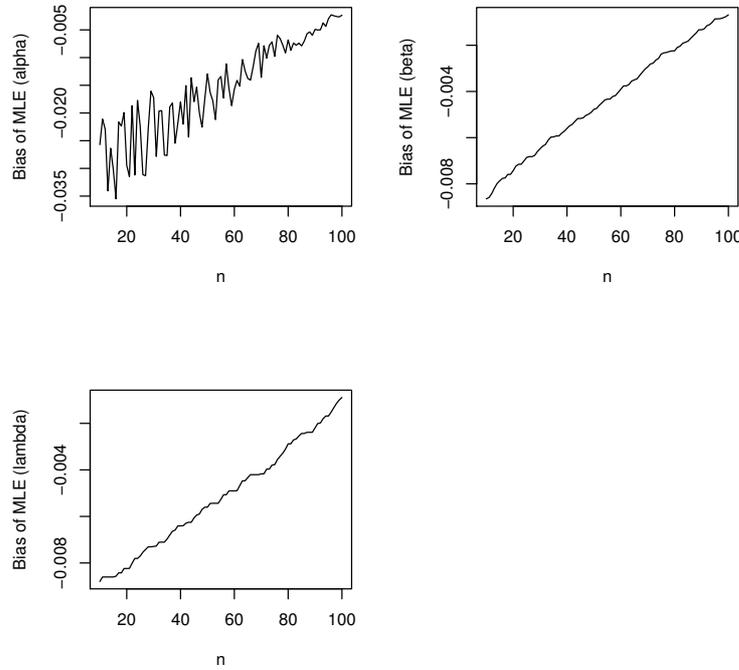

Figure 4: Biases of $\left(\widehat{\alpha}, \widehat{\beta}, \widehat{\lambda}\right)$ versus n.

Figures 4 and 5 show how the three biases and the three mean squared errors vary with respect to n. The following observations can be made: the biases for each parameter are negative; the biases appear largest for the parameter, α; the biases appear smallest for the parameters, β and λ; the biases for each parameter increase to zero as $n \to \infty$; the mean squared errors for each parameter decrease to zero as $n \to \infty$; the mean squared errors appear largest for the parameter, α; the mean squared errors appear smallest for the parameter, λ. These observations are for only one choice for (α, β, λ), namely that $(\alpha, \beta, \lambda) = (1, 1, 1)$. But the results were similar for a wide range of other values of (α, β, λ). In particular,

the biases for each parameter always increased to zero as $n \to \infty$ and the mean squared errors for each parameter always decreased to zero as $n \to \infty$.
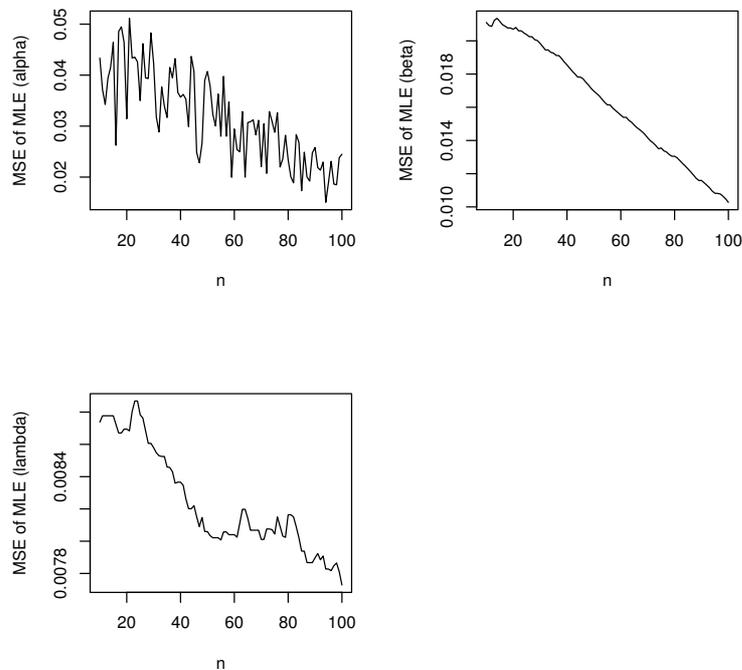

Figure 5:   Mean squared errors of $\left(\widehat{\alpha}, \widehat{\beta}, \widehat{\lambda}\right)$ versus n.

Section 6 presents three real data applications. The sample size for the first data set is eight hundred and seventy seven. The sample size for the second data set is twenty six. The sample size for the third data set is two hundred and ninety five. We shall see later in Section 6 that the WL distribution provides good fits to the three data sets. Based on this fact, the biases for $\widehat{\alpha}$, $\widehat{\beta}$ and $\widehat{\lambda}$ can be expected to be less than 0.025, 0.007 and 0.0075, respectively, for all of the data sets. The mean squared errors for $\widehat{\alpha}$, $\widehat{\beta}$ and $\widehat{\lambda}$ can be expected to be less than 0.04, 0.02 and 0.0088, respectively, for all of the data sets. Hence, the point estimates given in Section 6 for all data sets can be considered accurate enough.

5.3. Censored maximum likelihood estimation

Often with lifetime data, we encounter censored data. There are different forms of censoring: type I censoring, type II censoring, etc. Here, we consider the general case of multi-censored data: there are n subjects of which

- n_0 are known to have failed at the times $x_1, ..., x_{n_0}$;

- n_1 are known to have failed in the interval $[s_{j-1}, s_j]$, $j = 1, ..., n_1$;

- n_2 survived to a time r_j, $j = 1, ..., n_2$ but not observed any longer.

Note that $n = n_0 + n_1 + n_2$ and that type I censoring and type II censoring are contained as particular cases of multi-censoring. The log-likelihood function of the model parameters for this multi-censoring data is:

$$\ell(\alpha, \beta, \lambda) = \sum_{i=0}^{n_0} \log\left[\lambda^2 (1 + x_i) + \alpha\lambda (\beta x_i)^\alpha + \alpha\beta(1 + \lambda) (\beta x_i)^{\alpha-1}\right] - \lambda \sum_{i=0}^{n_0} x_i$$

$$- \sum_{i=0}^{n_0} (\beta x_i)^\alpha - n_0 \log(1 + \lambda)$$

$$+ \sum_{i=1}^{n_1} \log\left\{F(s_i) - F(s_{i-1})\right\}$$

(5.5) $$+ \sum_{i=1}^{n_2} \log\left\{1 - F(r_i)\right\},$$

where $F(\cdot)$ is given by (1.1). The MLEs can be obtained by maximizing (5.5) numerically. The maximization can be performed by using the nlm function in R.

5.4. Generating data

Section 5.2 gave an inversion method for simulating from the WL distribution. Here, we present an alternative method for simulation.

We know that a WL random variable is the minimum of independent Weibull and Lindley random variables. So, to generate a random sample from the WL distribution, the following algorithm can also be used:

1. First generate a random sample $v_1, ..., v_n$ from Weibull(α, β);

2. Independently, generate a random sample $w_1, ..., w_n$ from Lindley(λ);

3. Set $x_i = \min(v_i, w_i)$ for $i = 1, ..., n$.

Then $x_1, x_2, ..., x_n$ will be a random sample from WL(α, β, λ).

6.   APPLICATIONS

In this section, we fit the WL distribution to three real data sets. We compare the fits of the WL distribution to the fits of some related distributions: the extended Lindley (EL) distribution due to Bakouch *et al.* [4] with the pdf

$$f(x) = \frac{\lambda(1 + \lambda + \lambda x)^{\alpha-1}}{(1 + \lambda)^{\alpha}} \left[\beta(1 + \lambda + \lambda x)(\lambda x)^{\beta-1} - \alpha \right] e^{-(\lambda x)^{\beta}}$$

for $x > 0$, $\alpha \in (-\infty, 0) \cup \{0, 1\}$, $\beta > 0$ and $\lambda > 0$; the weighted Lindley (WEL) distribution due to Ghitany *et al.* [14] with the pdf

$$f(x) = \frac{\theta^{c+1}}{(\theta + c)\Gamma(c)} x^{c-1}(1 + x)e^{-\theta x}$$

for $x > 0$, $c > 0$ and $\theta > 0$; the exponential Poisson Lindley (EPL) distribution due to Barreto-Souza and Bakouch [5] with the pdf

$$f(x) = \frac{\beta\theta^2(1 + \theta)^2 e^{-\beta x} \left(3 + \theta - e^{-\beta x} \right)}{(1 + 3\theta + \theta^2) \left(1 + \theta - e^{-\beta x} \right)^3}$$

for $x > 0$, $\theta > 0$ and $\beta > 0$; the Lindley distribution with the pdf

$$f(x) = \frac{\lambda^2}{\lambda + 1}(1 + x)e^{-\lambda x}$$

for $x > 0$ and $\lambda > 0$; the generalized Lindley (GL1) distribution due to Zalerzadeh and Dolati [29] with the pdf

$$f(x) = \frac{\theta^2(\theta x)^{\alpha-1}(\alpha + \gamma x)e^{-\theta x}}{(\gamma + \theta)\Gamma(\alpha + 1)}$$

for $x > 0$, $\alpha > 0$, $\theta > 0$ and $\gamma > 0$; the power Lindley (PL) distribution due to Ghitany *et al.* [13] with the pdf

$$f(x) = \frac{\alpha\beta^2}{\beta + 1}(1 + x^{\alpha})x^{\alpha-1}e^{-\beta x^{\alpha}}$$

for $x > 0$, $\alpha > 0$ and $\beta > 0$; and, the generalized Lindley (GL2) distribution due to Nadarajah *et al.* [23] with the pdf

$$f(x) = \frac{\alpha\lambda^2}{1 + \lambda}(1 + x) \left[1 - \frac{1 + \lambda + \lambda x}{1 + \lambda} e^{-\lambda x} \right]^{\alpha-1} e^{-\lambda x}$$

for $x > 0$, $\alpha > 0$ and $\lambda > 0$.

6.1. Data set 1

The first data are times to reinfection of STD for eight hundred and seventy seven patients. The data were taken from Section 1.12 of Klein and Moeschberger [17]. We fitted the eight distributions to the data. Table 1 gives the parameter estimates, standard errors obtained by inverting the observed information matrix, log-likelihood values, values of Akaike information criterion (AIC), values of Bayesian information criterion (BIC), p values based on the Kolmogorov–Smirnov (KS) statistic, p values based on the Anderson–Darling (AD) statistic, and p values based on the Cramér–von Mises (CVM) statistic. The fitted pdfs of the three best fitting distributions as well as the empirical histogram are shown in Figure 6. The corresponding probability plots are shown in Figure 7.

Table 1: Parameter estimates, standard errors, log-likelihoods, AICs, BICs and goodness-of-fit measures for data set 1.

Distribution	Parameter estimates (s.e)	$-\log L$	AIC	BIC	KS	AD	CVM
EL	$\widehat{\lambda} = 8.806 \times 10^{-1}(1.302 \times 10^{-2})$, $\widehat{\alpha} = -9.804 \times 10^{-1}(3.034 \times 10^{-2})$, $\widehat{\beta} = 9.935 \times 10^{-7}(8.098 \times 10^{-3})$	9203.4	18412.9	18427.2	4.080×10^{-4}	2.700×10^{-5}	1.979×10^{-4}
WEL	$\widehat{\theta} = 2.878 \times 10^{-3}(1.076 \times 10^{-4})$, $\widehat{c} = 9.359 \times 10^{-2}(1.324 \times 10^{-2})$	6082.4	12168.7	12178.3	9.474×10^{-3}	2.879×10^{-4}	6.021×10^{-2}
EPL	$\widehat{\theta} = 2.326(7.568 \times 10^{-1})$, $\widehat{\beta} = 2.190 \times 10^{-3}(1.854 \times 10^{-4})$	6055.1	12114.1	12123.7	8.151×10^{-2}	1.395×10^{-2}	1.174×10^{-1}
Lindley	$\widehat{\lambda} = 5.397 \times 10^{-3}(1.313 \times 10^{-4})$	6413.0	12828.1	12832.9	1.996×10^{-3}	1.762×10^{-4}	5.579×10^{-4}
GL1	$\widehat{\theta} = 7.872 \times 10^{-2}(4.339 \times 10^{-3})$, $\widehat{\alpha} = 1.453 \times 10^{-6}(4.164 \times 10^{-2})$, $\widehat{\gamma} = 7.595 \times 10^{-1}(6.225 \times 10^{-2})$	27827.1	55660.1	55674.4	1.864×10^{-5}	2.149×10^{-5}	1.697×10^{-4}
PL	$\widehat{\alpha} = 5.696 \times 10^{-1}(1.364 \times 10^{-2})$, $\widehat{\beta} = 7.671 \times 10^{-2}(6.420 \times 10^{-3})$	6056.3	12116.7	12126.2	6.872×10^{-2}	1.232×10^{-3}	7.611×10^{-2}
GL2	$\widehat{\lambda} = 2.980 \times 10^{-3}(1.391 \times 10^{-4})$, $\widehat{\alpha} = 3.660 \times 10^{-1}(1.509 \times 10^{-2})$	6031.8	12067.5	12077.1	1.695×10^{-1}	7.568×10^{-2}	1.480×10^{-1}
WL	$\widehat{\lambda} = 2.331 \times 10^{-3}(2.714 \times 10^{-4})$, $\widehat{\alpha} = 6.435 \times 10^{-1}(3.870 \times 10^{-2})$, $\widehat{\beta} = 1.740 \times 10^{-3}(2.792 \times 10^{-4})$	6022.9	12051.7	12066.0	3.131×10^{-1}	8.243×10^{-2}	2.735×10^{-1}

We can see that the WL distribution gives the smallest AIC value, the smallest BIC value, the largest p value based on the KS statistic, the largest p value based on the AD statistic, and the largest p value based on the CVM statistic. The second smallest AIC, BIC values and the second largest p values are given by the GL2 distribution. The third smallest AIC, BIC values and the third largest p values are given by the EPL distribution. The fourth smallest AIC, BIC values and the fourth largest p values are given by the PL distribution. The fifth smallest AIC, BIC values and the fifth largest p values are given by the WEL distribution. The sixth smallest AIC, BIC values and the sixth largest p values are given by the Lindley distribution. The seventh smallest AIC, BIC values and the seventh largest p values are given by the EL distribution. The largest AIC, BIC values and the smallest p values are given by the GL1 distribution.

Hence, the WL distribution provides the best fit based on the AIC values, BIC values, p values based on the KS statistic, p values based on the AD statistic, and p values based on the CVM statistic. The density and probability plots also show that the WL distribution provides the best fit.
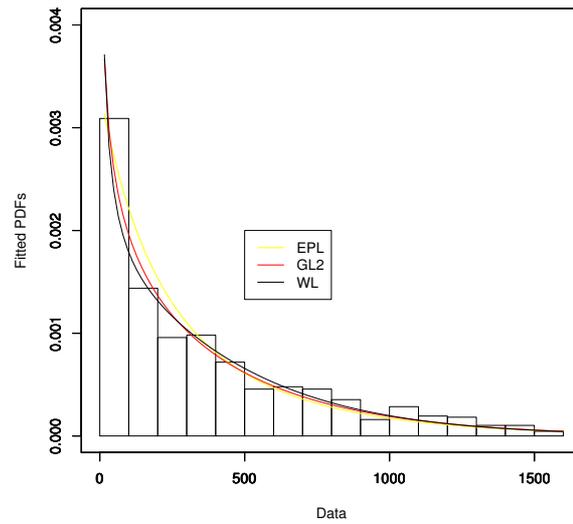

Figure 6:   Fitted pdfs of the three best fitting distributions for data set 1.
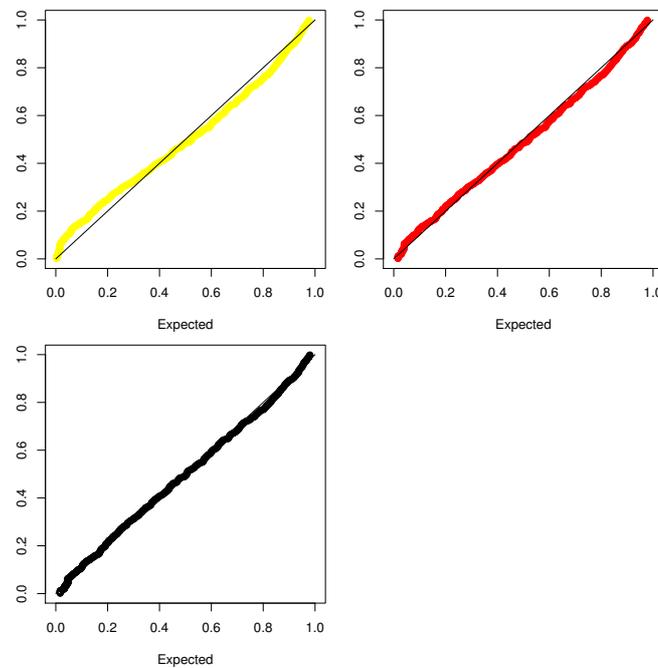

Figure 7:   PP plots of the three best fitting distributions for data set 1
            (yellow for the EPL distribution, red for the GL2 distribution
            and black for the WL distribution).

6.2. Data set 2

The second data are times to death of twenty six psychiatric patients. The data were taken from Section 1.15 of Klein and Moeschberger [17]. The eight distributions were fitted to this data. The parameter estimates, standard errors and the various measures are given in Table 2. The corresponding density and probability plots are shown in Figures 8 and 9, respectively.

Table 2: Parameter estimates, standard errors, log-likelihoods, AICs, BICs and goodness-of-fit measures for data set 2.

Distribution	Parameter estimates (s.e)	$-\log L$	AIC	BIC	KS	AD	CVM
EL	$\widehat{\lambda} = 7.510 \times 10^{-1} (9.604 \times 10^{-2})$, $\widehat{\alpha} = -8.534 \times 10^{-1} (2.414 \times 10^{-1})$, $\widehat{\beta} = 2.050 \times 10^{-6} (1.197 \times 10^{-1})$	164.9	335.8	339.5	1.163×10^{-3}	3.780×10^{-5}	9.083×10^{-5}
WEL	$\widehat{\theta} = 7.727 \times 10^{-2} (2.090 \times 10^{-2})$, $\widehat{c} = 1.107 (4.659 \times 10^{-1})$	107.7	219.3	221.8	4.691×10^{-1}	1.392×10^{-3}	3.064×10^{-2}
EPL	$\widehat{\theta} = 1.267 \times 10^4 (4.662 \times 10^5)$, $\widehat{\beta} = 3.784 \times 10^{-2} (7.442 \times 10^{-3})$	111.1	226.3	228.8	1.840×10^{-2}	4.964×10^{-5}	1.024×10^{-4}
Lindley	$\widehat{\lambda} = 7.311 \times 10^{-2} (1.016 \times 10^{-2})$	107.7	217.4	218.6	7.924×10^{-1}	1.126×10^{-2}	4.895×10^{-2}
GL1	$\widehat{\theta} = 8.282 \times 10^{-2} (2.152 \times 10^{-2})$, $\widehat{\alpha} = 1.420 (5.569 \times 10^{-1})$, $\widehat{\gamma} = 2.742 \times 10^{-1} (3.199 \times 10^{-1})$	107.1	220.3	224.0	2.920×10^{-2}	1.163×10^{-4}	7.787×10^{-4}
PL	$\widehat{\alpha} = 1.225 (2.069 \times 10^{-1})$, $\widehat{\beta} = 3.452 \times 10^{-2} (2.470 \times 10^{-2})$	106.9	217.8	220.3	6.957×10^{-1}	8.737×10^{-3}	3.715×10^{-2}
GL2	$\widehat{\lambda} = 7.547 \times 10^{-2} (1.398 \times 10^{-2})$, $\widehat{\alpha} = 1.069 (2.874 \times 10^{-1})$	107.7	219.3	221.8	1.081×10^{-1}	1.582×10^{-4}	4.384×10^{-3}
WL	$\widehat{\lambda} = 4.340 \times 10^{-2} (1.051 \times 10^{-2})$, $\widehat{\alpha} = 9.901 (2.822)$, $\widehat{\beta} = 2.832 \times 10^{-2} (1.014 \times 10^{-3})$	93.4	192.8	196.5	8.998×10^{-1}	8.372×10^{-1}	2.849×10^{-1}

We can see again that the WL distribution gives the smallest AIC value, the smallest BIC value, the largest p value based on the KS statistic, the largest p value based on the AD statistic, and the largest p value based on the CVM statistic. The second smallest AIC, BIC values and the second largest p values are given by the Lindley distribution. The third smallest AIC, BIC values and the third largest p values are given by the PL distribution. The fourth smallest AIC, BIC values and the fourth largest p values are given by the WEL distribution. The fifth smallest AIC, BIC values and the fifth largest p values are given by the GL2 distribution. The sixth smallest AIC, BIC values and the sixth largest p values are given by the GL1 distribution. The seventh smallest AIC, BIC values and the seventh largest p values are given by the EPL distribution. The largest AIC, BIC values and the smallest p values are given by the EL distribution.

Hence, the WL distribution again provides the best fit based on the AIC values, BIC values, p values based on the KS statistic, p values based on the AD

statistic, and p values based on the CVM statistic. The density and probability plots again show that the WL distribution provides the best fit.
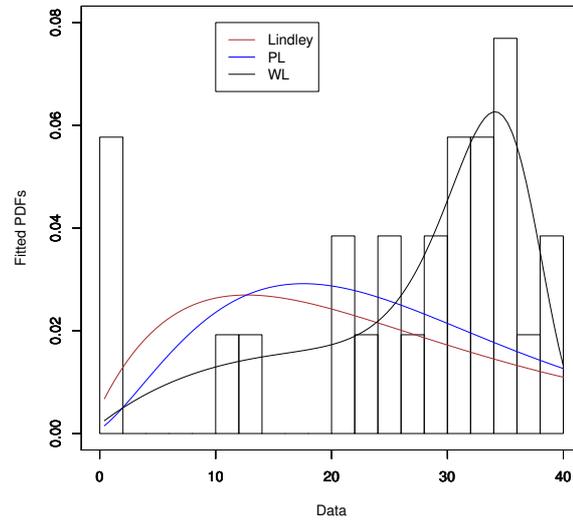

Figure 8:   Fitted pdfs of the three best fitting distributions for data set 2.
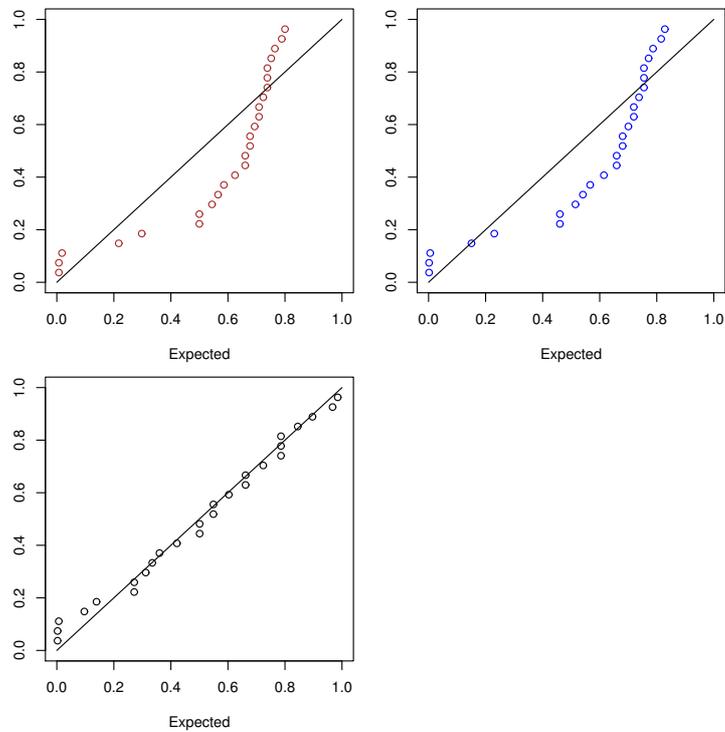

Figure 9:   PP plots of the three best fitting distributions for data set 2 (brown for the Lindley distribution, blue for the PL distribution and black for the WL distribution).

6.3. Data set 3

The third data are times to infection for AIDS for two hundred and ninety five patients. The data were taken from Section 1.19 of Klein and Moeschberger [17]. The eight distributions were fitted to this data. The parameter estimates, standard errors and the various measures are given in Table 3. The corresponding density and probability plots are shown in Figures 10 and 11, respectively.

Table 3: Parameter estimates, standard errors, log-likelihoods, AICs, BICs and goodness-of-fit measures for data set 3.

Distribution	Parameter estimates (s.e)	$-\log L$	AIC	BIC	KS	AD	CVM
EL	$\hat{\lambda} = 2.066 \times 10^{-1}(5.340 \times 10^{-3})$, $\hat{\alpha} = -1.425 \times 10^{-1}(8.097 \times 10^{-2})$, $\hat{\beta} = 3.503(2.776 \times 10^{-1})$	537.5	1080.9	1092.0	7.775×10^{-1}	9.719×10^{-1}	1.866×10^{-1}
WEL	$\hat{\theta} = 1.396(1.130 \times 10^{-1})$, $\hat{c} = 5.025(4.472 \times 10^{-1})$	563.3	1130.6	1138.0	2.072×10^{-1}	1.269×10^{-1}	9.133×10^{-3}
EPL	$\hat{\theta} = 1.145 \times 10^{4}(1.091 \times 10^{5})$, $\hat{\beta} = 2.403 \times 10^{-1}(1.402 \times 10^{-2})$	713.2	1430.4	1437.8	6.394×10^{-2}	2.022×10^{-2}	7.102×10^{-4}
Lindley	$\hat{\lambda} = 4.106 \times 10^{-1}(1.731 \times 10^{-2})$	659.7	1321.4	1325.1	1.113×10^{-1}	8.193×10^{-2}	1.443×10^{-3}
GL1	$\hat{\theta} = 1.402(1.157 \times 10^{-1})$, $\hat{\alpha} = 5.099(5.533 \times 10^{-1})$, $\hat{\gamma} = 3.872(4.373)$	563.3	1132.5	1143.6	1.744×10^{-1}	9.239×10^{-2}	3.008×10^{-3}
PL	$\hat{\alpha} = 2.099(8.683 \times 10^{-2})$, $\hat{\beta} = 8.357 \times 10^{-2}(1.176 \times 10^{-2})$	544.7	1093.5	1100.9	5.437×10^{-1}	1.664×10^{-1}	3.248×10^{-2}
GL2	$\hat{\lambda} = 7.544 \times 10^{-1}(3.321 \times 10^{-2})$, $\hat{\alpha} = 4.536(4.812 \times 10^{-1})$	571.4	1146.9	1154.2	1.136×10^{-1}	9.136×10^{-2}	1.616×10^{-3}
WL	$\hat{\lambda} = 1.595 \times 10^{-1}(3.235 \times 10^{-2})$, $\hat{\alpha} = 4.036(4.329 \times 10^{-1})$, $\hat{\beta} = 1.949 \times 10^{-1}(6.412 \times 10^{-3})$	535.7	1077.4	1088.4	8.059×10^{-1}	9.908×10^{-1}	8.666×10^{-1}

We can see yet again that the WL distribution gives the smallest AIC value, the smallest BIC value, the largest p value based on the KS statistic, the largest p value based on the AD statistic, and the largest p value based on the CVM statistic. The second smallest AIC, BIC values and the second largest p values are given by the EL distribution. The third smallest AIC, BIC values and the third largest p values are given by the PL distribution. The fourth smallest AIC, BIC values and the fourth largest p values are given by the WEL distribution. The fifth smallest AIC, BIC values and the fifth largest p values are given by the GL1 distribution. The sixth smallest AIC, BIC values and the sixth largest p values are given by the GL2 distribution. The seventh smallest AIC, BIC values and the seventh largest p values are given by the Lindley distribution. The largest AIC, BIC values and the smallest p values are given by the EPL distribution.

Hence, the WL distribution yet again provides the best fit based on the AIC values, BIC values, p values based on the KS statistic, p values based on the AD

statistic, and p values based on the CVM statistic. The density and probability plots yet again show that the WL distribution provides the best fit.
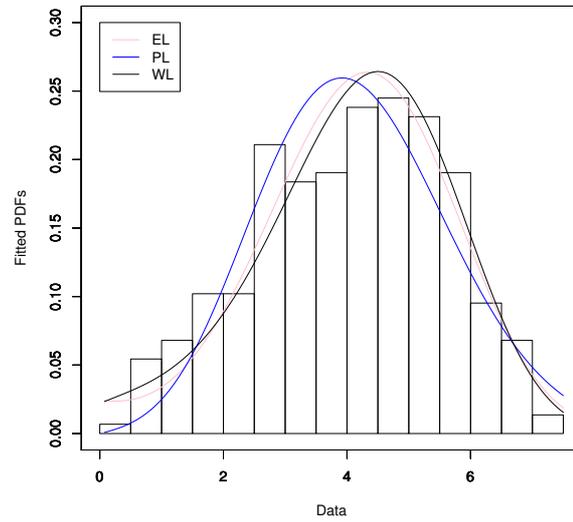

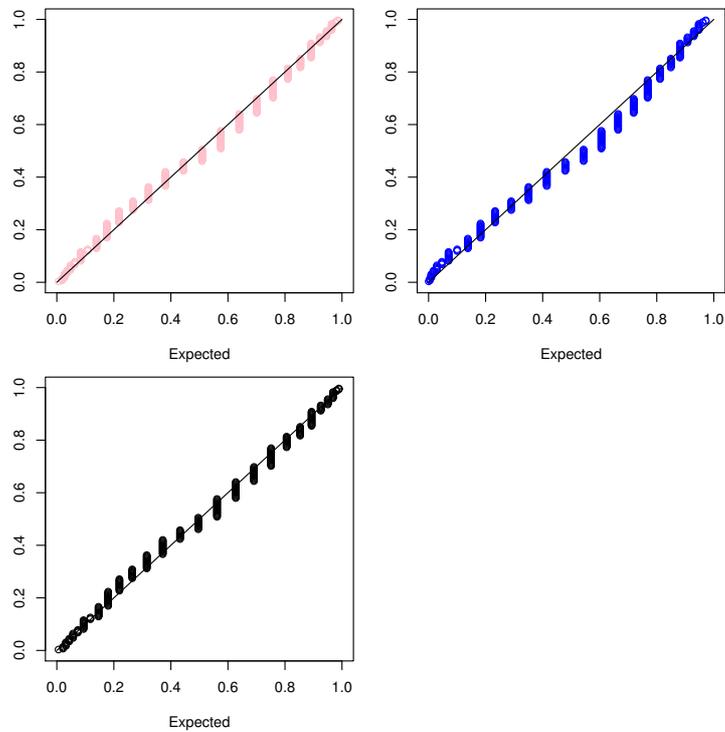**Figure 10**: Fitted pdfs of the three best fitting distributions for data set 3.

Figure 11: PP plots of the three best fitting distributions for data set 3 (pink for the EL distribution, blue for the PL distribution and black for the WL distribution).

7. CONCLUSIONS

We have introduced a three-parameter generalization of the Lindley distribution referred to as the Weibull Lindley distribution. We have provided at least seven possible motivations for this new distribution. We have studied shapes of probability density and hazard rate functions, moments, moment generating function, Lorenz curve, maximum likelihood estimators in the presence of complete data and maximum likelihood estimators in the presence of censored data. We have assessed the finite sample performance of the maximum likelihood estimators by simulation. We have provided three real data applications.

We have seen that the probability density function can be bimodal, unimodal, monotonically decreasing or monotonically decreasing with an inflexion point. The hazard rate function can be monotonically increasing, monotonically decreasing or bathtub shaped. The maximum likelihood estimators appear to be regular for sample sizes larger than twenty. The data applications show that the Weibull Lindley distribution provides better fits than all known generalizations of the Lindley distribution for at least three data sets.

ACKNOWLEDGMENTS

REFERENCES

[1]   ADAMIDIS, K. and LOUKAS, S. (1998). A lifetime distribution with decreasing failure rate, *Statistics and Probability Letters*, **39**, 35–42.

[2]   ANDERSEN, P.K. and SKRONDAL, A. (2015). A competing risks approach to "biologic" interaction, *Lifetime Data Analysis*, **21**, 300–314.

[3]   ASGHARZADEH, A.; BAKOUCH, H.S. and ESMAEILI, L. (2013). Pareto Poisson–Lindley distribution with applications, *Journal of Applied Statistics*, **40**, 1717–1734.

[4]   BAKOUCH, H.S.; AL-ZAHRANI, B.M.; AL-SHOMRANI, A.A.; MARCHI, V.A.A. and LOUZADA, F. (2012). An extended Lindley distribution, *Journal of the Korean Statistical Society*, **41**, 75–85.

[5]   BARRETO-SOUZA, W. and BAKOUCH, H.S. (2013). A new lifetime model with decreasing failure rate, *Statistics*, **47**, 465–476.

[6]   BARRETO-SOUZA, W.; MORAIS, A.L. and CORDEIRO, G.M. (2011). The Weibull-geometric distribution, *Journal of Statistical Computation and Simulation*, **81**, 645–657.

[7]   BOURGUIGNON, M.; SILVA, R.B. and CORDEIRO, G.M. (2014). A new class of fatigue life distributions, *Journal of Statistical Computation and Simulation*, **84**, 2619–2635.

[8]   CADARSO-SUAREZ, C.; MEIRA-MACHADO, L.; KNEIB, T. and GUDE, F. (2010). Flexible hazard ratio curves for continuous predictors in multi-state models: An application to breast cancer data, *Statistical Modelling*, **10**, 291–314.

[9]   EGGHE, L. (2002). Development of hierarchy theory for digraphs using concentration theory based on a new type of Lorenz curve, *Mathematical and Computer Modelling*, **36**, 587–602.

[10] FERGUSON, T.S. (1996). *A Course in Large Sample Theory*, Chapman and Hall, London.

[11] GAIL, M.H. (2009). Applying the Lorenz curve to disease risk to optimize health benefits under cost constraints, *Statistics and Its Interface*, **2**, 117–121.

[12] GAIL, M.H. and BENICHOU, J. (2000). *Encyclopedia of Epidemiologic Methods*, John Wiley and Sons.

[13] GHITANY, M.E.; AL-MUTAIRI, D.K.; BALAKRISHNAN, N. and AL-ENEZI, L.J. (2013). Power Lindley distribution and associated inference, *Computational Statistics and Data Analysis*, **64**, 20–33.

[14] GHITANY, M.E.; ALQALLAF, F.; AL-MUTAIRI, D.K. and HUSAIN, H.A. (2011). A two-parameter weighted Lindley distribution and its applications to survival data, *Mathematics and Computers in Simulation*, **81**, 1190–1201.

[15] GHITANY, M.E.; ATIEH, B. and NADARAJAH, S. (2008). Lindley distribution and its application, *Mathematics and Computers in Simulation*, **78**, 493–506.

[16] GROVES-KIRKBY, C.J.; DENMAN, A.R. and PHILLIPS, P.S. (2009). Lorenz curve and Gini coefficient: Novel tools for analysing seasonal variation of environmental radon gas, *Journal of Environmental Management*, **90**, 2480–2487.

[17] KLEIN, J.P. and MOESCHBERGER, M.L. (1997). *Survival Analysis Techniques for Censored and Truncated Data*, Springer Verlag, New York.

[18] KUS, C. (2007). A new lifetime distribution, *Computational Statistics and Data Analysis*, **51**, 4497–4509.

[19] LEHMANN, L.E. and CASELLA, G. (1998). *Theory of Point Estimation*, second edition, Springer Verlag, New York.

[20] LINDLEY, D.V. (1958). Fiducial distributions and Bayes' theorem, *Journal of the Royal Statistical Society*, A, **20**, 102–107.

[21] MALDONADO, A.; PEREZ-OCON, R. and HERRERA, A. (2007). Depression and cognition: New insights from the Lorenz curve and the Gini index, *International Journal of Clinical and Health Psychology*, **7**, 21–39.

[22] MORAIS, A.L. and BARRETO-SOUZA, W. (2011). A compound class of Weibull and power series distributions, *Computational Statistics and Data Analysis*, **55**, 1410–1425.

[23] NADARAJAH, S.; BAKOUCH, H.S. and TAHMASBI, R. (2011). A generalized Lindley distribution, *Sankhyā*, B, **73**, 331–359.

[24] NAIR, U.N.; SANKARAN, P.G. and BALAKRISHNAN, N. (2010). *Quantile-Based Reliability Analysis*, Birkhauser, Boston.

[25] R DEVELOPMENT CORE TEAM (2014). *A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

[26] RADICE, A. (2009). Use of the Lorenz curve to quantify statistical nonuniformity of sediment transport rate, *Journal of Hydraulic Engineering*, **135**, 320–326.

[27] SILVA, R.B.; BOURGUIGNON, M.; DIAS, C.R.B. and CORDEIRO, G.M. (2013). The compound class of extended Weibull power series distributions, *Computational Statistics and Data Analysis*, **58**, 352–367.

[28] YAMAGUCHI, K. (1993). Modeling time-varying effects of covariates in event-history analysis using statistics from the saturated hazard rate model, *Sociological Methodology*, **23**, 279–317.

[29] ZAKERZADEH, H. and DOLATI, A. (2009). Generalized Lindley distribution, *Journal of Mathematical Extension*, **3**, 13–25.