www.iiste.org

Length Biased Gumbel Distribution And Its Application To Wind Speed Data.

Olawale Basheer Akanbi [1*] Amidu Akinola Oyebanjo [2]
[1, 2]Department of Statistics, University of Ibadan, Nigeria
[*]Email of the corresponding author: muhdbashola@yahoo.com

Abstract

The concept of length biased distribution can be employed in development of proper models for life time data. Its method is adjusting the original probability density function from real data and the expectation of those data. This adjustment can bring about correct conclusions on the models. In this research, a new class of length biased approach was introduced to gumbel distribution which is called length baised gumbel distribution (LBGD). The theoretical properties of this distribution were derived and the model parameters were also estimated by Maximum Likelihood Estimation (MLE). This procedure is to adjust the original probability density function from real and ex-pected data which leads to good conclusion on the model. This distribution was later applied to wind speed data (extreme value data).

Keywords: Extreme Value Theory, Weighted Distribution, Length biased Gumbel Distribution, Alkaine Information Criterion, Wald Test

1. Introduction

The theory of weighted distributions provides a collective access for the problems of model specification and data interpretation. It provides a technique for fitting models to the un- known weight functions when samples can be taken both from the original distribution and the developed distribution. Weighted distributions take into account the method of ascer- tainment, by adjusting the probabilities of the actual occurrence of events to arrive at a specification of the probabilities of those events as observed and recorded. The weighted distributions occur frequently in the studies related to reliability, analysis of family data, Meta analysis and analysis of intervention data, biomedicine, ecology and other areas, for the improvement of proper statistical models.

The Concept of weighted distributions can be traced to the work of Fisher (1934) in connec- tion with his studies on how methods of ascertainment can influence the form of distribution of recorded observations. Later it was introduced and formulated in general terms by Rao (1965) in connection with modeling statistical data where the usual practice of using stan- dard distributions for the purpose was not found to be appropriate. In Rao's paper, he identified various situations that can be modeled by weighted distributions. These situations refer to instances where the recorded observations cannot be considered as a random sample from the original distributions. This may occur due to non observability of some events or damage caused to the original observation resulting in a reduced value, or adoption of a sampling procedure which gives unequal chances to the units in the original. Characteristics of many length biased distributions, preservation stability results and com- parisons for weighted and length biased distributions were presented by Khatree (1989). Rao (1965) extended the idea of the methods of ascertainment upon estimation of frequencies and introduced the concept of weighted distributions. However, as far as we have gathered, no study has been carried out that provide a forum for the existing works and future re- search direction on weighted distributions for applied researchers and practitioners. This, we believe, is a big gap in the literature and the current article is an attempt to fill this gapto some extent.

One of the basic problems when one use weighted distributions as a tool in the selection of suitable models for observed data is the choice of the weight function that fits the data. Depending upon the choice of weight function w(t), we have different weighted models. For example, when the weight function depends on the lengths of units of interest ($w(x) = x$), the resulting distribution is called length-biased. Other approaches are Area Biased and Cubic Biased. Some researchers have worked on Length Biased for other distributions and or with different weighted functions: Oluyede and George [2002] looked at the characteristics of many length biased distributions, Gao, et al. (2011) found that the length-biased Poisson (events) data arise in bioinformatics, Ratnaparkhi and Naik-Nimbalkar (2012) regarding the length biased data arising in oil filed exploration and seeks to associate length-biased data with a random sampling procedure, Kareema Abed Alkadim (2016) present the Double Weighted Inverse Weibull (DWIW) using different weight functions. Hesham M Reyad, et al. (2017) introduced a new class of length-biased weighted Frechet (LBWF) distribution. Sofi Mudasir and Ahmad S. P (2017) derived two parameter length biased Weibull distribution. Salman et al, (2019) derived a new generalization of weighted

Weibull distribution using Topp Leone family of distributions.

2. Length Biased Distribution

When the weight function depends on the lengths of units of interest (i.e. $w(x)$), the resulting distribution is called length-biased. In this case, the pdf of a length-biased random variable X_L is defined as

$$f_w(x) = \frac{xf(x)}{\mu} \qquad a < x < b \tag{1}$$

$$\mu = E[w(x)] = \int_{-\infty}^{\infty} w(x) f(x) dx$$

Fisher (1934) introduced these distributions to model ascertainment bias and formalized in a unifying theory by Rao (1965). These distributions arise in practice when observations from a sample are recorded with unequal probability and provide a unifying approach for the problems where the observations fall in the non-experimental.

Patil and Rao (1978) examined some general models leading to weighted distributions with weight functions not necessarily bounded by unity and studied length biased (size biased) sampling with applications to wildlife populations and human families.

3. The Length biased gumbel distribution

The weighted distributions arise when the observations generated from a stochastic process are not given equal chance of being recorded; instead they are recorded according to some weighted function. When the weight function depends on the lengths of the units of interest, the resulting distribution is called length biased.

This is the first time in the history of applying weighted distribution to Gumbel distribution and appropriate justification will be made on it.

First introduced by Fisher (1934) to model ascertainment bias, these were later formalized in a unifying theory by Rao (1965). These distributions arise in practice when observations from a sample are recorded with unequal probability and provide unifying approach for the problems when the observations fall in the non experimental, non replicated and non random categories.

The probability density function of Gumbel distribution is given as :

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

and the weighted distribution approach is given as:

$$g(x) = \frac{w(x)f(x)}{\int_{-\infty}^{\infty} xf(x)dx} \quad (3)$$

When $w(x) = x$, then it is known as Length Biased Gumbel Distribution. To solve for the denominator of equation (2)

$$\frac{1}{\sigma}\int_{-\infty}^{\infty} xe^{-\left[\left(\frac{x-\mu}{\sigma}\right)+e^{\left(\frac{x-\mu}{\sigma}\right)}\right]}dx \quad (4)$$

Let

$$z = \left(\frac{x-\mu}{\sigma}\right) \quad (5)$$

Differentiate equation (18) with respect to x.

$$\frac{dz}{dx} = \frac{1}{\sigma}$$
$$dx = \sigma dz \quad (6)$$

Make x the subject of the formula from equation (4)

$$z = \frac{x-\mu}{\sigma})$$
$$z\sigma = x - \mu \quad (7)$$
$$x = z\sigma + \mu$$

Substitute equation (6) and (7) into equation (4)

$$\frac{1}{\sigma}\int_{-\infty}^{\infty}(z\sigma+\mu)\,e^{-\left[z+e^{-z}\right]}\sigma dz$$

$$\int_{-\infty}^{\infty} z\sigma e^{-\left[z+e^{-z}\right]}dz+\int_{-\infty}^{\infty}\mu e^{-\left[z+e^{-z}\right]}dz$$

$$\sigma\int_{-\infty}^{\infty} ze^{-\left[z+e^{-z}\right]}dz+\mu\int_{-\infty}^{\infty}e^{-\left[z+e^{-z}\right]}dz$$

Recall that

$$\int_{-\infty}^{\infty}e^{-\left[z+e^{-z}\right]}dz=1$$

Then,

$$=\mu+\sigma\int_{-\infty}^{\infty}ze^{-\left[z+e^{-z}\right]}dz$$

Note that

$$\gamma=\int_{-\infty}^{\infty}ze^{-\left[z+e^{-z}\right]}dz$$

$$=\mu+\sigma\gamma$$

Now,

$$g(x)=\frac{xf(x)}{\int_{-\infty}^{\infty}xf(x)dx}$$

$$g(x)=\frac{\frac{1}{\sigma}xe^{-\left[\left(\frac{x-\mu}{\sigma}\right)+e^{\left(\frac{x-\mu}{\sigma}\right)}\right]}}{\mu+\sigma\gamma}$$

$$g(x)=\frac{xe^{-\left[\left(\frac{x-\mu}{\sigma}\right)+e^{\left(\frac{x-\mu}{\sigma}\right)}\right]}}{\sigma\mu+\sigma^{2}\gamma} \qquad (8)$$

3.1 Properities of length biased gumbel distribution

Properities of length biased gumbel distribution are verification of probability density function, cummulative density function, Survival function, Hazard function and so on.

3.1.1 Verification of probability density function

$$\int_{-\infty}^{\infty} g(x)dx = 1 \qquad (9)$$

Proof:

$$= \int_{-\infty}^{\infty} \frac{xe^{-\left[\left(\frac{x-\mu}{\sigma}\right)+e^{-\left(\frac{x-\mu}{\sigma}\right)}\right]}}{\sigma u + \sigma^2 \gamma} dx$$

Let

$$z = \frac{x-\mu}{\sigma} \qquad (10)$$

Differentiate equation (10) with respect to z

$$\frac{dz}{dx} = \frac{1}{\sigma}$$
$$dx = \sigma dz \qquad (11)$$

Make x the subject of the formula from equation (10)

$$z\sigma = x - \mu$$

$$x = z\sigma + \mu \qquad (12)$$

Substitute equation (10), (11) and (12) into equation (9)

$$\frac{1}{\mu\sigma+\sigma^2\gamma}\int_{-\infty}^{\infty}(z\sigma+\mu)e^{-[z+e^{-z}]}\sigma dz$$

$$\frac{\sigma}{\mu\sigma+\sigma^2\gamma}\left[\int_{-\infty}^{\infty}(z\sigma)e^{-[z+e^{-z}]}dz+\int_{-\infty}^{\infty}(\mu)e^{-[z+e^{-z}]}dz\right]$$

$$\frac{\sigma}{\mu\sigma+\sigma^2\gamma}\left[\sigma\int_{-\infty}^{\infty}(z)e^{-[z+e^{-z}]}dz+\mu\int_{-\infty}^{\infty}e^{-[z+e^{-z}]}dz\right]$$

$$\int_{-\infty}^{\infty}e^{-[z+e^{-z}]}dz=\int_{0}^{\infty}e^{-[z+e^{-z}]}dz=1$$

$$\int_{-\infty}^{\infty}(z)e^{-[z+e^{-z}]}dz=\gamma,$$

$$\frac{\sigma}{\mu\sigma+\sigma^2\gamma}[\sigma\gamma+\mu]$$

$$\frac{\sigma^2\gamma+\mu\sigma}{\mu\sigma+\sigma^2\gamma}$$

$$=1$$

3.1.2 Cummulative density function of length biased gumbel distribution

The cummulative density function of a random variable X is a function given by

$$\begin{aligned}G(x) &= \int_{-\infty}^{x} g(x)dx\\ &= \int_{-\infty}^{x}\frac{xe^{-\left(\frac{x-\mu}{\sigma}\right)}e^{-e^{-\frac{(x-\mu)}{\sigma}}}}{\sigma\mu+\sigma^2\gamma}dx\\ &= \frac{1}{\sigma\mu+\sigma^2\gamma}\int_{-\infty}^{x}xe^{-\left(\frac{x-\mu}{\sigma}\right)}e^{-e^{\left(\frac{x-\mu}{\sigma}\right)}}dx \qquad (13)\end{aligned}$$

Let

$$z = \frac{x-\mu}{\sigma}$$

Differentiate *z* with respect to *x*

$$\frac{dz}{dx} = \frac{1}{\sigma}$$

$$dx = \sigma dz \qquad (14)$$

$$z\sigma = x - \mu$$

$$x = z\sigma + \mu \tag{15}$$

substitute equation 15, equation 14 into equation 13

$$= \frac{1}{\sigma\mu + \sigma^2\gamma}\int_{-\infty}^{z}(z\sigma+\mu)e^{-z}\times e^{-e^{z}}\sigma dz$$

$$= \frac{1}{\mu+\sigma\gamma}\left[\int_{-\infty}^{z} z\sigma e^{-z}.e^{-e^{e^{-z}}}dz + \mu\int_{0}^{z} e^{-z}e^{-e^{-z}}dz\right] \tag{16}$$

Let

$$y = e^{-z}$$

$$dz = \frac{-dy}{y} \tag{17}$$

substitute equation (17) into equation (16)

$$= \frac{1}{\mu+\sigma\gamma}\left[\sigma\int_0^y -lny.ye^{-y}.\frac{dy}{y} + \mu\int_0^y ye^{-y}.\frac{-dy}{y}\right]$$

$$= \frac{1}{\mu+\sigma\gamma}\left[\sigma\int_0^y lnye^{-y}dy + \mu\int_0^y -e^{-y}dy\right]$$

$$= \frac{1}{\mu+\sigma\gamma}\left[\sigma\int_0^y lnye^{-y}dy - \mu\int_0^y e^{-y}dy\right]$$

Recall that

$$\int_0^y e^{-y}dy = 1 - e^{-y} \tag{18}$$

$$= \frac{1}{\mu+\sigma\gamma}\left[\sigma\int_0^y lnye^{-y}dy - \mu(1-e^{-y})\right]$$

Using Integration by part for $\int_0^y lnye^{-y}dy$

Let $u = lny$ and $du = e^{-y}$

$$\frac{du}{dy} = \frac{1}{y}$$ and $v = e^{-y}$

$$du = \frac{dy}{y}$$ and $-e^{-y}$

$$uv - \int vdu$$
$$lny. - e^{-y} - \int -e^{-y}.\frac{dy}{y}$$
$$-lnye^{-y} + \int e^{-y}dy$$
$$-lnye^{-y} + \frac{(1-e^{-y})}{y}$$
$$\frac{1 - ylnye^{-y} - e^{-y}}{y}$$

By integration by part

$$\begin{aligned} \int udv &= uv - \int vdu \\ &= lny. - e^{-y} - \int -e^{-y}.\frac{dy}{y} \\ &= -lnye^{-y} + \int e^{-y}dy \\ &= -lnye^{-y} + \frac{(1-e^{-y})}{y} \\ &= \frac{1 - ylnye^{-y} - e^{-y}}{y} \end{aligned} \quad (19)$$

Substitute equation 19 into equation 18

Now $G(x) = \frac{1}{\mu + \sigma\gamma}\left[\sigma\left(\frac{1 - ylnye^{-y} - e^{-y}}{y}\right) - \mu(1 - e^{-y})\right]$

3.1.3 Survival function of length biased gumbel distribution

Survival function, also known as Survivor function or Reliability function is a property of any random variable that maps a set of events, usually associated with mortality or failure of some system. It captures the probability that the system will survived beyond a specified time. The mathematical expression of survival function of length biased gumbel distribution is denoted as S(x) and given as:

ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.11, No.5, 2021

$$
\begin{aligned}
S(x) &= 1 - G(x) \\
&= 1 - \frac{1}{\mu+\sigma\gamma}\left[\sigma\left(\frac{1 - ylnye^{-y} - e^{-y}}{y}\right) - \mu(1-e^{-y})\right] \\
&= 1 - \frac{1}{\mu+\sigma\gamma}\left[\frac{\sigma\left(1 - ylnye^{-y} - e^{-y}\right) - \mu y(1-e^{-y})}{y}\right] \\
S(x) &= \frac{\mu y + \sigma y\gamma - \sigma(1 - ylnye^{-y} - e^{-y}) + \mu y(1-e^{-y})}{y(\mu+\sigma\gamma)}
\end{aligned}
$$

recall that $y = e^{-z}$ and $z = (\frac{x-\mu}{\sigma})$

Define survival function in term of x

Now,

$$
S(x) = \frac{\mu e^{-(\frac{x-\mu}{\sigma})} + \gamma\sigma e^{-(\frac{x-\mu}{\sigma})} - \sigma(1 + (\frac{x-\mu}{\sigma})e^{-(\frac{x-\mu}{\sigma})}e^{-e^{-(\frac{x-\mu}{\sigma})}} - e^{-e^{-(\frac{x-\mu}{\sigma})}}) + \mu e^{-(\frac{x-\mu}{\sigma})}(1 - e^{-e^{-(\frac{x-\mu}{\sigma})}})}{(\mu+\sigma\gamma)e^{-(\frac{x-\mu}{\sigma})}}
$$

Activate Windows

3.1.4 Hazard function of length biased gumbel distribution

Hazard function is the probability of failure in an infinitesimally small period between y and δy given that the subject has survived up to time y. It measures risk. The greater the hazard between times y_1 and y_2, the greater the risk of failure in the time interval.

The mathematical expression of hazard function of length biased gumbel distribution is denoted as $h(x)$ and given as:

$$\begin{aligned}
h(x) &= \frac{g(x)}{S(x)} \\
&= \frac{g(x)}{1-G(x)} \\
&= \frac{\frac{xe^{-\left(\frac{x-\mu}{\sigma}\right)} \times e^{-e^{-\frac{(x-\mu)}{\sigma}}}}{\sigma\mu+\sigma^2\gamma}}{\frac{y\mu+\sigma y\gamma-\sigma(1-ylnye^{-y}-e^{-y})+\mu y(1-e^{-y})}{y(\mu+\sigma\gamma)}} \\
&= \frac{xe^{-\left(\frac{x-\mu}{\sigma}\right)}e^{-e^{-\left(\frac{x-\mu}{\sigma}\right)}}}{\sigma(\mu+\sigma\gamma)} \frac{y(\mu+\sigma\gamma)}{\mu+\sigma\gamma-\sigma(1-ylnye^{-y})-e^{-y}+\mu y(1-e^{-y})} \\
h(x) &= \frac{xye^{-\left(\frac{x-\mu}{\sigma}\right)}e^{-e^{-\left(\frac{x-\mu}{\sigma}\right)}}}{\sigma\left[\mu+\sigma\gamma-\sigma\left(1-ylnye^{-y}-e^{-y}\right)+\mu y\left(1-e^{-y}\right)\right]} \\
h(x) &= \frac{xye^{-\left(\frac{x-\mu}{\sigma}\right)}e^{-e^{-\left(\frac{x-\mu}{\sigma}\right)}}}{\sigma\mu+\sigma^2\gamma-\sigma^2\left(1-ylnye^{-y}-e^{-y}\right)+\sigma\mu y\left(1-e^{-y}\right)}
\end{aligned}$$

(20)

3.1.5 Median of length biased gumbel distribution

$$\int_0^m g(x)dx = 0.5$$

$$\int_0^m \frac{xe^{-\left(\frac{x-\mu}{\sigma}\right)}e^{-e^{\left(\frac{x-\mu}{\sigma}\right)}}}{\sigma\mu+\sigma^2\gamma} = 0.5 \quad (21)$$

Let

$$z = \frac{x-\mu}{\sigma} \quad (22)$$

Differentiate equation (22) with respect to *x*

$$\frac{dz}{dx} = \frac{1}{\sigma}$$

$$dx = \sigma dz \tag{23}$$

make *x* the subject of the formula from equation (22)

$z\sigma \quad = \quad x - \mu x \;=\quad z\sigma + \mu$ (24)

$$\int_0^m \frac{(z\sigma+\mu))\,e^{\frac{-z\sigma+\mu-\mu}{\sigma}}e^{-e^{-z}}}{\sigma\mu+\sigma^2\gamma}\sigma dz = 0.5$$

$$\int_0^m \frac{(z\sigma+\mu)e^{-z}e^{-e^{-z}}}{\sigma(\mu+\sigma\gamma)}\sigma dz = 0.5$$

$$\frac{1}{\mu+\sigma\gamma}\left[\int_m^0 z\sigma e^{-z}e^{-e^{-z}}dz + \int_0^m \mu e^{-z}e^{-e^{-z}}dz\right] = 0.5$$

$$\frac{1}{\mu+\sigma\gamma}\left[\sigma\int_0^m ze^{-z}e^{-e^{-z}}dz + \mu\int_0^m e^{-z}e^{-e^{-z}}dz\right] = 0.5$$

Let $m = e^{-z}$

$$\frac{1}{\mu+\sigma\gamma}\left[\sigma\frac{(1-\mathrm{m}\,ln\;\mathrm{m}e^{-m}.e^{-m})}{m} - \mu\;\;1-e^{-m})\right] = \frac{1}{2}$$

$$\frac{2}{\mu+\sigma\gamma}\left[\sigma\;\;1-m\,lnme^{-m}.e^{-m} - \mu m(1-e^{-m}))\right] = 1$$

$$\frac{\mu+\sigma\gamma}{2} = \sigma - \sigma mlnme^{-m}.\sigma e^{-m} - \mu m(1-e^{-m})$$

Let $\quad k \quad = \quad e{-}m \qquad n = lnm$

$$\begin{aligned}
\text{Then} \quad \frac{\mu+\sigma\gamma}{2} &= m(\mu k-\mu-k\sigma n)+\sigma(1-k)\\
\frac{\mu+\sigma\gamma}{2} &= \frac{m(\mu k-\mu-k\sigma n)}{\mu}+\frac{\sigma(1-k)}{m}\\
\frac{\sigma(1-k)}{m} &= \frac{\mu+\sigma\gamma}{2}-(\mu k-\mu-k\sigma n)\\
\frac{\sigma(1-k)}{m} &= \frac{\mu+\sigma\gamma-2(\mu k-k\sigma n)}{2}\\
\frac{1}{m} &= \frac{\mu+\sigma\gamma-2(\mu k-\mu-k\sigma n)}{2\sigma(1-k)}\\
m &= \frac{2\sigma(1-k)}{\mu+\sigma\gamma-2(\mu k-\mu-k\sigma n)}\\
\text{Recall that} \quad m &= \frac{x-\mu}{\sigma}\\
\frac{x-\mu}{\sigma} &= \frac{2\sigma(1-k)}{\mu+\sigma\gamma-2(\mu k-\mu-k\sigma n)}\\
x-\mu &= \frac{2\sigma^2(1-k)}{\mu+\sigma\gamma-2(\mu k-\mu-k\sigma n)}\\
x &= \frac{2\sigma^2(1-k)}{\mu+\sigma\gamma-2(\mu k-\mu-k\sigma n)}+\mu
\end{aligned}$$

Quantile of 3.1.6 Quantile of the length biased gumbel distribution

$$\begin{aligned}
G(x) &= U\\
&= \frac{1}{\mu+\sigma\gamma}\left[\frac{\sigma\left(1-zlnze^{-z}-e^{-e^{-z}}\right)-\mu z\left(1-e^{-z}\right)}{z}\right]
\end{aligned} \tag{25}$$

Let $k = lny$ and $n = e^{-z}$

$$\begin{aligned}
\text{then} \quad \frac{1}{\mu+\sigma\gamma}\left[\frac{\sigma-\sigma zkn-\sigma n-\mu z+\mu zn}{z}\right] &= U\\
U(\mu+\sigma\gamma) &= \frac{(\mu zn-\mu z-\sigma zkn)}{z}+\frac{\sigma(1-n)}{z}\\
U(\mu+\sigma\gamma) &= \frac{z(\mu n-\mu-\sigma kn)}{z}+\frac{\sigma(1-n)}{z}\\
U(\mu+\sigma\gamma)-\mu n+\mu+\sigma kn &= \frac{\sigma(1-n)}{z}
\end{aligned}$$

$$\frac{z}{\sigma(1-n)} = \frac{1}{U(\mu+\sigma\gamma)-\mu n+\mu+\sigma kn}$$
$$z = \frac{\sigma(1-n)}{U(\mu+\sigma\gamma)-\mu n+\mu+\sigma kn}$$

Recall that $z = \frac{x-\mu}{\sigma}$

$$\frac{x-\mu}{\sigma} = \frac{\sigma(1-n)}{U(\mu+\sigma\gamma)-\mu n+\mu+\sigma kn}$$
$$x = \frac{\sigma^2(1-n)}{U(\mu+\sigma\gamma)-\mu n+\mu+\sigma kn}+\mu$$

3.1.7 Mean of length biased gumbel distribution

$$g(x) = \frac{xe^{-\left[\left(\frac{x-\mu}{\sigma}\right)+e^{-\left(\frac{x-\mu}{\sigma}\right)}\right]}}{\sigma\mu+\sigma^2\gamma}; -\infty < x < \infty \quad (26)$$

$$E(x) = \int_{-\infty}^{\infty} xg(x)dx$$
$$= \frac{1}{\sigma\mu+\sigma^2\gamma}\int_{-\infty}^{\infty} x^2e^{-\left[\left(\frac{x-\mu}{\sigma}\right)+e^{-\left(\frac{x-\mu}{\sigma}\right)}\right]} \quad (27)$$

Let

$$z = \frac{x-\mu}{\sigma} \quad (28)$$

Differentiate equation (28) with respect to *x*

$$\frac{dz}{dx} = \frac{1}{\sigma}$$
$$dx = \sigma dz \quad (29)$$

put equation 29 in to equation 27

$$= \frac{1}{\sigma\mu+\sigma^2\gamma}\int_{-\infty}^{\infty}(z+\mu)^2+e^{-\left[z+e^{-z}\right]}\sigma dz$$
$$= \frac{\sigma}{\sigma\left(\mu+\sigma\gamma\right)}\int_{-\infty}^{\infty}(z+\mu)^2+e^{-\left[z+e^{-z}\right]}dz$$
$$= \frac{1}{\mu+\sigma\gamma}\int_{-\infty}^{\infty} z^2\sigma^2+2z\sigma\mu+\mu^2)\,e^{-\left[z+e^{-z}\right]}dz$$
$$= \frac{1}{\mu+\sigma\gamma}\left[\sigma^2\int_{-\infty}^{\infty} z^2e^{-\left[z+e^{-z}\right]}+2\sigma\mu\int_{-\infty}^{\infty} ze^{-\left[z+e^{-z}\right]}+\mu^2\int_{-\infty}^{\infty} ze^{-\left[z+e^{-z}\right]}\right]dz$$

$$\int_{-\infty}^{\infty} e^{-\left[z+e^{-z}\right]}dz = 1 \quad \int_{-\infty}^{\infty} ze^{-\left[z+e^{-z}\right]}dz = \gamma$$

Recall that $\int_{-\infty}^{\infty} e^{-\left[z+e^{-z}\right]}dz = 1$ $\int_{-\infty}^{\infty} ze^{-\left[z+e^{-z}\right]}dz = \gamma$ and $\int_{-\infty}^{\infty} z^2e^{-\left[z+e^{-z}\right]}dz = \gamma^2$

Then

$$\begin{aligned} E(x) &= \frac{1}{\mu+\sigma\gamma}\left[\sigma^2\gamma^2 + 2\sigma\mu\gamma + \mu^2\right] \\ &= \frac{(\sigma\gamma+\mu)^2}{(\mu+\sigma\gamma)} \\ E(x) &= \frac{(\sigma\gamma+\mu)(\mu+\sigma\gamma)}{(\mu+\sigma\gamma)} \\ &= (\mu+\sigma\gamma) \end{aligned}$$

(30)

3.1.8 Variance of length biased gumbel distribution

$$\begin{aligned} Var(x) = E(x-\mu)^2 &= \int_{-\infty}^{\infty} x^2 g(x)dx - \left[\int_{-\infty}^{\infty} xg(x)dx\right]^2 \\ &= E(x^2) - [E(x^2)]^2 \end{aligned} \quad (31)$$

$$\begin{aligned} E(x^2) &= \int_{-\infty}^{\infty} x^2 g(x)dx \\ &= \int_{-\infty}^{\infty} x^2 \frac{xe^{-\left[\left(\frac{x-\mu}{\sigma}\right)+e^{-\left(\frac{x-\mu}{\sigma}\right)}\right]}}{\sigma\mu+\sigma^2\gamma}dx \\ &= \int_{-\infty}^{\infty} \frac{x^3e^{-\left[\left(\frac{x-\mu}{\sigma}\right)+e^{-\left(\frac{x-\mu}{\sigma}\right)}\right]}}{\sigma\mu+\sigma^2\gamma}dx \end{aligned} \quad (32)$$

Let

$$z = \frac{x-\mu}{\sigma} \quad (33)$$

Differentiate equation (32) with respect to *x*

$$\frac{dz}{dx} = \frac{1}{\sigma}$$
$$dx = \sigma dz \quad (34)$$

make *x* the subject of the formula from equation (32)

$$z\sigma = x - \mu$$
$$x = z\sigma + \mu \quad (35)$$

put equation 33 and equation 34 in to equation 31

$$E(x^2) = \frac{1}{\sigma\mu + \sigma^2\gamma}\int_{-\infty}^{\infty} (z\sigma + \mu)^3\, e^{-\left[z+e^{-z}\right]}\sigma dz$$
$$= \frac{\sigma}{\sigma(\mu + \sigma\gamma)}\int_{-\infty}^{\infty}\ z\sigma^3 + 3z^2\sigma^2\mu + 3z\sigma\mu^2 + \mu^3)\ e^{-\left[z+e^{-z}\right]}$$

(36)

$$= \frac{1}{\mu + \sigma\gamma}\left[\sigma^3\int_{-\infty}^{\infty} z^3 e^{-\left[z+e^{-z}\right]}dz + 3\sigma^2\mu\int_{-\infty}^{\infty} z^2 e^{-\left[z+e^{-z}\right]}dz + 3\sigma\mu^2\int_{-\infty}^{\infty} z e^{-\left[z+e^{-z}\right]}dz + \mu^3\int_{-\infty}^{\infty} e^{-\left[z+e^{-z}\right]}d\right.$$

Recall that $\int_{-\infty}^{\infty} e^{-\left[z+e^{-z}\right]} dz = 1, \quad \int_{-\infty}^{\infty} z e^{-\left[z+e^{-z}\right]} dz = \gamma, \quad \int_{-\infty}^{\infty} z^2 e^{-\left[z+e^{-z}\right]} dz = \gamma^2,$

and $\int_{-\infty}^{\infty} z^3 e^{-\left[z+e^{-z}\right]} dz = \gamma^3$

$$\begin{aligned} E(x^2) &= \frac{1}{\mu+\sigma\gamma}\left[\sigma^3\gamma^3 + 3\sigma^2\gamma^2\mu + 3\sigma\gamma\mu^2 + \mu^3\right] \\ &= \frac{(\mu+\sigma\gamma)^3}{(\mu+\sigma\gamma)} \\ &= \frac{(\mu+\sigma\gamma)^2(\mu+\sigma\gamma)}{\mu+\gamma} \\ &= (\mu+\sigma\gamma)^2 \end{aligned}$$

$$\begin{aligned} \therefore \quad Var(x) &= E(x^2) - [E(x^2)]^2 \\ &= (\mu+\sigma\gamma)^2 - [(\mu+\sigma\gamma)]^2 \\ &= 0 \end{aligned}$$

3.1.9 Maximum likelihood of length biased gumbel distribution

Maximum likelihood provides a consistent approach to parameter estimation problems, this means that maximum likelihood estimates can be developed for a large variety of estimation situations. It begins with the mathematical expression known as a likelihood function of the sample data. The likelihood of a set of data is the probability of obtaining that particular set of data given the chosen probability model. The probability density function of length biased gumbel distribution is given as:

$$f(x) = \frac{x e^{-[z+e^{z}}}{\sigma\mu + \sigma^2\gamma} \qquad (37)$$

$$z = \frac{x-\mu}{\sigma}$$

where

The likelihood function is given as:

$$L\left(x_1, x_2, \cdots\cdots, x_n\right) = \prod_{i=1}^{n} f\left(x, \mu, \sigma\right)$$

$$L\left(x_1, x_2, \cdots\cdots, x_n\right) = \prod_{i=1}^{n} \frac{xe^{-[z+e^z]}}{\sigma\mu + \sigma^2\gamma}$$

$$= \prod_{i=1}^{n} \frac{xe^{-\left[\left(\frac{x-\mu}{\sigma}\right)+e^{\left(\frac{x-\mu}{\sigma}\right)}\right]}}{\sigma\mu + \sigma^2\gamma}$$

$$= \frac{x_1 e^{-\left[\left(\frac{x_1-\mu}{\sigma}\right)+e^{\left(\frac{x_1-\mu}{\sigma}\right)}\right]}}{\sigma\mu + \sigma^2\gamma} \cdot \frac{x_2 e^{-\left[\left(\frac{x_2-\mu}{\sigma}\right)+e^{\left(\frac{x_2-\mu}{\sigma}\right)}\right]}}{\sigma\mu + \sigma^2\gamma} \cdots\cdots \frac{x_n e^{-\left[\left(\frac{x_n-\mu}{\sigma}\right)+e^{\left(\frac{x_n-\mu}{\sigma}\right)}\right]}}{\sigma\mu + \sigma^2\gamma} \quad (38)$$

Take ln of both side

$$\ln L\left(x_1, x_2, \cdots\cdots, x_n\right) = \left(\frac{1}{\sigma\mu + \sigma^2\gamma}\right)^n \sum x_i e^{-\left[\sum\left(\frac{x-\mu}{\sigma}\right)+e^{\sum\left(\frac{x-\mu}{\sigma}\right)}\right]}$$

$$= n\ln \ \sigma\mu + \sigma^2\gamma) + \sum \ln x_i - \sum\left(\frac{x-\mu}{\sigma}\right) - e^{\sum\left(\frac{x-\mu}{\sigma}\right)} \quad (39)$$

Differentiate with respect to each parameter and equate to zero

$$\frac{\delta \ln L\left(x, \mu, \sigma\right)}{\delta\mu} = -n\left(\frac{\sigma}{\sigma\mu}\right) + \frac{n}{\sigma} - \frac{n}{\sigma}e^{\sum\left(\frac{x-\mu}{\sigma}\right)} = 0$$

$$= \frac{-n}{\mu} + \frac{n}{\sigma} - \frac{n}{\sigma}e^{\sum\left(\frac{x-\mu}{\sigma}\right)} = 0$$

$$\frac{n}{\mu} = \frac{n}{\sigma} - \frac{n}{\sigma}e^{\sum\left(\frac{x-\mu}{\sigma}\right)}$$

$$\frac{\mu}{n} = \frac{\sigma}{n} - \frac{\sigma}{ne^{\sum\left(\frac{x-\mu}{\sigma}\right)}} \quad (40)$$

Make μ the subject of the formular

$$\hat{\mu} = n\left(\frac{\sigma}{n} - \frac{\sigma}{ne^{\sum\left(\frac{x-\mu}{\sigma}\right)}}\right)$$

$$\hat{\mu} = \sigma - \frac{\sigma}{e^{\sum\left(\frac{x-\mu}{\sigma}\right)}}$$

$$\hat{\mu} = \sigma - \sigma e^{-\sum\left(\frac{x-\mu}{\sigma}\right)}$$

$$\hat{\mu} = \ln\sigma - \ln\sigma\sum_{i=1}^{n} e^{-\frac{x_i}{\sigma}}$$

$$\frac{\delta \ln L(x,\mu,\sigma)}{\delta\sigma} = -n\left(\frac{\mu}{\sigma\mu} + \frac{2\sigma\gamma}{\sigma^2\gamma}\right) + \frac{\sum x_i}{\sigma^2} - \frac{\mu n}{\sigma^2} + e^{\sum\left(\frac{x-\mu}{\sigma}\right)}\left(\frac{n\mu - \sum x_i}{\sigma^2}\right)$$

$$= -n\left(\frac{1}{\sigma} + \frac{2}{\sigma}\right) + \frac{\sum x_i}{\sigma^2} - \frac{\mu n}{\sigma^2} + \left(\frac{n\mu - \sum x_i}{\sigma^2}\right)e^{\sum\left(\frac{x-\mu}{\sigma}\right)}$$

$$\frac{-3n}{\sigma} + \frac{\sum x_i}{\sigma^2} - \frac{\mu n}{\sigma^2} + \left(\frac{n\mu - \sum x_i}{\sigma^2}\right)e^{\sum\left(\frac{x-\mu}{\sigma}\right)} = 0$$

$$-3\sigma n - \sum x_i - \mu n + \left(n\mu - \sum x_i\right)e^{\sum\left(\frac{x-\mu}{\sigma}\right)} = 0$$

$$3\sigma n = \left(n\mu - \sum x_i\right)e^{\sum\left(\frac{x-\mu}{\sigma}\right)} - \sum x_i - \mu n$$

$$\hat{\sigma} = \frac{\left(n\mu - \sum x_i\right)e^{\sum\left(\frac{x-\mu}{\sigma}\right)} - \sum x_i - \mu n}{3n}$$

$$\hat{\sigma} = \frac{\left(\mu - \bar{x}\right)e^{\sum\left(\frac{x-\mu}{\sigma}\right)} - \bar{x} - \mu}{3}$$

4 Data Analysis

Our primary objective is to fit length biased gumbel distribution to wind data of different sample sizes by applying the concepts EVT to model the tails of the distribution for peak level. We describe the historical data for wind data on environment, the preliminary tests undertaken on the data and exploratory techniques, determine of the size of the data, the fitting of the Length Biased Gumbel distribution, and the examination of better model using model selection criterion (AIC). The empirical analysis has been undertaken by writing program code that was executed using the R package. The Length Biased Gumbel distribution was used to fit the extreme value data and investigate the likelihood-based confidence intervals for quantile of the fitted distribution.

Preliminary Analysis of Data:

The descriptive analysis performed in this section gives the important features of the data obtained. Table 1 gives quick descriptive summary of all the parameters under study. The data used in this study is a secondary data which is a record of peak value of wind speed in Ibadan.

Table 1: Descriptive statistics for wind speed data

Number of	730	SE Mean	0.15

Observation			
Minimum	3.5685	LCL Mean	9.5955
Maximum	16.2936	UCL Mean	10.1846
1. Quartile	7.1317	Variance	16.432
3. Quartile	12.9135	Stdev	4.0536
Mean	9.8901	Skewness	-0.0066
Median	9.599	Kurtosis	-1.1281
Sum	7219.78		

Since the mean value is greater than standard deviation value, it means that more of the data is clustered about the mean. i.e. the data points tend to be close to the mean. This indicates that the wind speed data has fairly symmetric pattern which can be as a result lack of outliers.

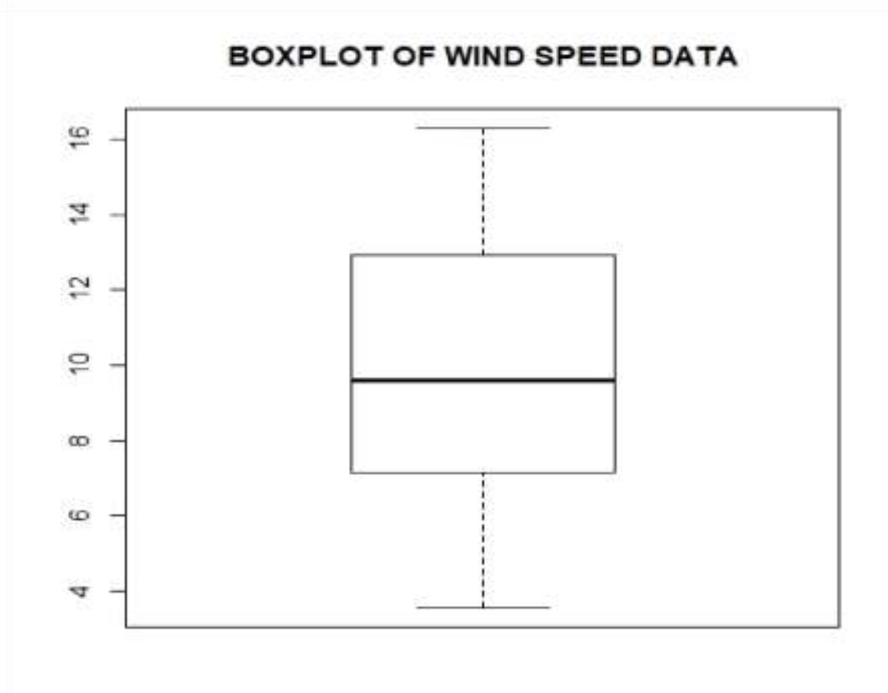

Figure 1: Boxplot of Wind speed data

The box plot (box and whisker diagram) is a standardized way of displaying the distribution of data based on the summaries: the central rectangle spans the first quartile to the third quartile (the interquartile range). A segment inside the rectangle shows the median and “whiskers” above and below the box show the locations of the minimum and maximum. Although a boxplot can tell you whether a data set is symmetric (when the median is in the center of the box) or not..

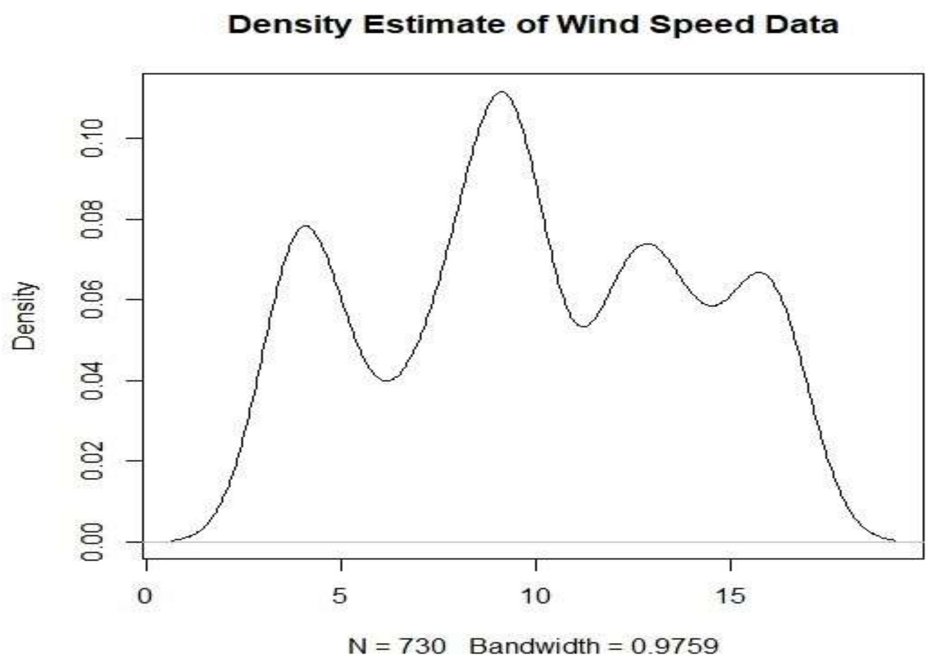

Figure 2: Density Function of Wind speed data

Density Function of a continuous random variable, is a function that describes the relative likelihood for this random variable to take on a given value. The probability of the random variable falling within a particular range of values is given by the integral of this variables density over that range. That is, it is given by the area under the density function but above the horizontal axis and between the lowest and greatest values of the range.

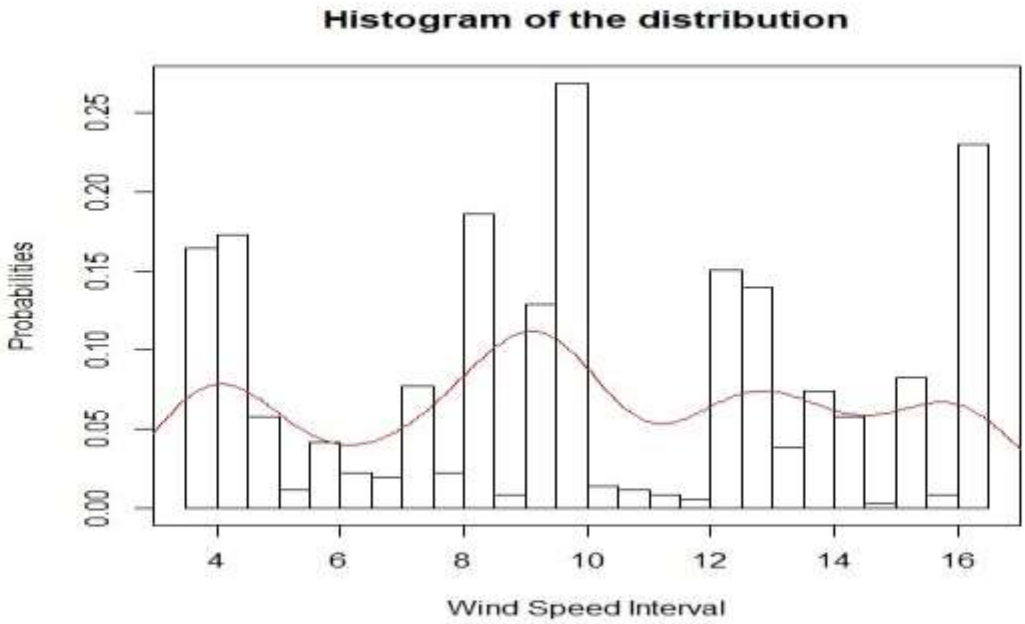

Figure 3: Histogram of Wind speed data

In figure 4, Histogram is a graphical method of presenting a large amount of data by way of bars, to reflect the distribution frequency and proportion or density of each class interval as a data set. Since a histogram provides planners and analysts with information presented in a compact and organized manner. The histogram confirms the skewed data (right) and more of the wind speed data fall between 10 and 16.

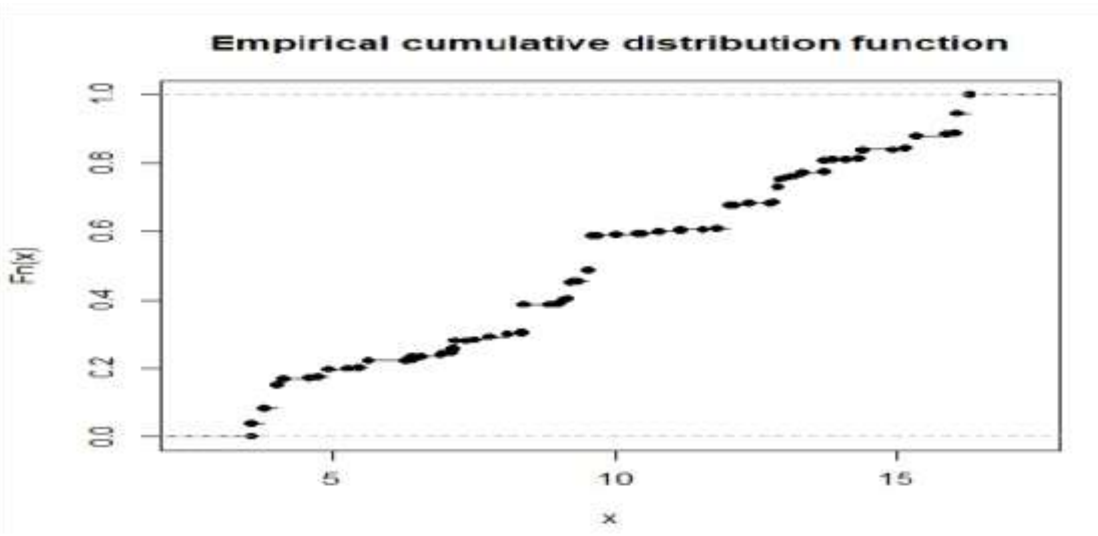

Figure 4: Cumulative function of Wind speed data

4.1: Maximum Likelihood Estimation Approach

The maximum likelihood estimator ($\hat{\theta}$) of the parameter θ is obtained by maximizing $L(\hat{\theta})$. Usually, for mathematical convenience, we rather work with $l(\hat{\theta}) = logL(\hat{\theta})$, which interestingly gives us no loss of information in using $l(\hat{\theta})$ because the log is a one to one function.

Table 2: LBGD Estimates for Wind Speed Data

LBGD	PARAMETER VALUE	AIC
μ	-1.264	-50460
Σ	8.7088	

Table 3: Parameters etimations for LBGD Simulated Data of different Sizes

DISTRIBUTION	SAMPLE SIZES	PARAMETER VALUE	AIC
LBGD	n=200	μ= 0.3228	**666.6433**
		σ= 0.2621	
	n=500	μ=0.2871	1729.112
		σ= 0.2505	
	n=1000	μ= 0.3112	3375.669
		σ= 0.2583	

The above Tables show that the parameter estimations of Length Biased Gumbel Distribution. and the distribution with the smallest AIC, is selected as the better one. From table 3, it can be seen that the smaller the sample size, the better the estimate. Correspondingly, the AIC value of Length Biased Gumbel Distribution for when n = 200 is lower compared to when the sample size n =1000.

5 Summary and Conclusion

The study was carried out to establish the properties of the proposed Length Biased Gumbel Distribution and its application to wind speed data. The main objective of this study was to derive the statistical properties of the proposed distribution and its application to environmental event. The estimations and applications of weighted distributions to real life data and simulated data from gumbel distribution having different sample sizes to the length biased gumbel distribution demonstrates that it has both solid theoretical underpinning and practical use to real Life data. AIC fit measures the length biased models to offer substantial improvement as the sample size increases for the data generated from the gumbel distribution. Also the fitting in these tables reveal that the length biased gumbel distributions provide us better fits in the situations when the sample size is small compared to when we have large sample size. High volume of wind speed in recent years, resulting to life losses and huge damages then, demand urgent actions. The emergency is also stressed by the fact that people face the threat of climate change. Success can only be reached if an interdisciplinary approach is adopted.

6 Recommendation

This study emphasizes the urgent need to facilitate effective management for anticipated wind to prevent the re-occurrence of the damage and losses life. Arising from this study, based on the findings in the field, as well as the need to protect the environment, the human population and livelihoods from unmitigated wind disasters in future, the following recommendations are hereby suggested: Any fields of study using natural events as variables should make use of weighted distributions which are milestone for efficient modeling of statistical data and prediction when the standard distributions are not appropriate; Weighted approach provides a technique for fitting models to the unknown weight functions when samples can be taken both from the original distribution and the developed distribution; Weighted distributions take into account the method of ascertainment, by adjusting the probabilities of the actual occurrence of events to arrive at a specification of the probabilities of those events as observed and recorded which all researchers must be encourages to use when dealing with environmental data; Weighted distribution theory gives a unified approach to modeling the biased data and should be taught in probability and mathematical statistics courses; It provides consolidating formulation for correction of biases that exist in unequally weighted sample data. Also, the theory provides a means of fitting models to the unknown weighting function when samples can be taken both from the original distribution and the resulting ”biased” distribution

References

Abu Bakar, M. R., Salah, K. A., Ibrahim, N. A., Haron, K. (2009). "Bayesian Approach for Joint Longitudinal and Time-to-Event Data with Survival Fraction". Bull. Malays. *Math. Sci. Soc.* 32, 75-100.

Aditi Datta (2013). "A Study of the Cure Rate Model with Case Weights and Time-Dependent Weights". *Guelph, Ontario, Canada.*

Ayesha A. (2017). "Sized biased lindley distribution and its properties; A special case of weighted distribution". *Applied Mathematics* 08(06), 808-819.doi:10.4236/am.2017.86063

Bader Ahmad I. Aljawadi, Mohd Rizam A. Bakar1, Noor Akma Ibrahim, and Mohamad Al-Omari (2012). "Parametric Maximum Likelihood Estimation of Cure Fraction Using Interval-Censored Data.". *Science Alert.*

Berkson J. and Gage R.P. (1952). Survival curve for cancer patients following treatment. Journal of the American Statistical Association, 47:501–515.

Betensky, R. A. and Schoenfeld, D.A. (2001). "Nonparametric Estimation in a cure model with Random cure times." *Biometrics* 57, 282-286.

Binbing Y, Tiwari R, Cronin K, et al. (2004). "Cure fraction estimation from the mixture cure models for grouped survival data." *Stat Med*; 23: 1733–47.

Boag J.W (1949). "Maximum likelihood estimates of the proportion of patients cured by cancer therapy." *Journal of the Royal Statistical Society. Series B (Methodological),* 11(1):15–53..

Chao Cai (2013). "Advanced Methodology Developments in Mixture Cure Models". *University of South Carolina , Ph.D. Thesis Scholar Commons*.

Chen, M.-H., Ibrahim, J. G., and Sinha, D. (1999). "A new Bayesian model for survival data with a surviving fraction." *Journal of the American Statistical Association,* 94(447) , 909–919.

Cooner, F., Banerjee, S., Carlin, B. P. and Sinha, D. (2007). "Flexible Cure Rate Modelling under Latent Activation Schemes." *J. Amer. Statist. Assoc.* 102 560–572.

Fauzia Taweab, Noor Akma Ibrahim and Jayanthi Arasan (2015). *A Bounded Cumulative Hazard Model with A change- Point According to a Threshold in a Covariate for Right-Censored Data*.

Hazard Model with A change- Point According to a Threshold in a Covariate for Right-Censored Data.

Fisher, R. A. (1934). "The effects of methods of ascertainment upon the estimation of frequencies." *Annals of Eugenics*, 6, 13-25.

Fisher, R. and Tippett, L. (1928). “Limiting forms of the frequency distribution of the largest and smallest member of a sample.” *Proc. Cambridge Phil. Soc.*, 24:180–190

Francisci S, Capocaccia R, Grande E, et al.(2009). “The Cure of Cancer: a European Perspective.” *Eur J Cancer*; 45:1067–79.

FrBchet, M. (1927). “Sur la loi de probabilitC de 1’6cart maximum.” *Ann. SOC. Polon. Math. Cracovie* 6, 93-116.

Gao, L., et al. (2011). “Length bias correction for RNA-seq gene set analyses.” *Bioinformtics*, 27 662-669.

Hesham M Reyad, et al.(2017). “The length biased weighted frechet distribution: Properties and estimation.” *International journal of statistics and applied mathematics* 3(1): 189-200.

Johnson, N.L., S. Kotz and N. Balakrishnan, (1995). *Continuous Univariate Distributions*. (2nd Edn)., John Wiley and Sons, New York, ISBN-10: 0471584940, pp: 752

Kareema Abed Al kadim.(2016). “Double Weighted Distributions.” *Basrah journal of science* (A) Vol. 34(3); 51−63, 2016.

Kinnison, R. P. (1985). *Applied Extreme Value Statistics*, Battelle Press, Macmillan.

Oluyede, B.O. and E.O. George, 2002. “On stochastic inequalities and comparisons of reliability measures for weighted distributions.” *Math. Prob. Eng.*, 8: 1-13. DOI: 10.1080/10241230211380

Patil, G.P.(1991). “Encountered Data, Statistical Ecology, Environmental Statistics and Weighted Distribution Methods,” *Environmetrics*, 2(4), 377-423.

Patil, G.P. and C.R. Rao, (1978). Weighted distributions and size biased sampling with applications to wildlife populations and human families. Biometrics, 34: 179-189

Ratnaparkhi, et al.(2017). “Weighted Distributions: A brief review, perspective and characterizations.” *International journal of statistics and probability* 2017; ISSN 1927-7032.

Rao, C. R. (1965), *On discrete distributions arising out of method of ascertainment, in classical and Contagious Discrete, G.P. Patiled;* Pergamum Press and Statistical publishing Society, Calcutta. 320-332.

Salman A, et al.(2019). “A new generalized weighted weibull distributions,” *Pakistan Journal of Statistics and Operation Research*. Vol. 15 No. 1 2019 pp.161−178.