UNIT HALF LOGISTIC NORMAL DISTRIBUTION: PROPERTIES AND APPLICATIONS

Abstract

In this study, a new generalization of the normal distribution called the unit half logistic normal (UHLN) distribution has been proposed by introducing a shape parameter into the normal distribution to make it more flexible. Several statistical properties of the new distribution which include; the cumulative hazard function, reversed hazard function, hazard rate average function, quantile function, moments, moment generating function and order statistics has been derived. Estimators such as the maximum likelihood, ordinary least squares, weighted least squares and Cramér-von Mises were developed for the new model. The performances of the estimators were investigated via Monte Carlo simulation using six different sample sizes and replicated 5000 times. The maximum likelihood was observed to be the most consistent and the best technique, hence was used to estimate the parameters of the new distribution. The applications of the UHLN distribution was demonstrated using three different datasets and compared with the normal, transmuted normal, beta normal, McDonal normal and logistic distributions. The results revealed that the UHLN distribution performs better for the given datasets.

Keywords: Unit half logistic, normal, error function, simulation, order statistics.

1. Introduction

Researchers in the field of statistical distribution continue to develop new parametric distributions or generalize the existing ones in the quest to obtain more flexible distributions or improve upon the performances of the existing ones in modeling datasets. This has become necessary because most of the existing statistical distributions always fail to give best fit to the new forms of data evolving randomly on daily basis with varied characteristics. In statistical data analysis, the choice of an appropriate model for the analysis is done based on which of them provides the best fit for a given dataset.

The normal distribution is the most popular classical distribution with wider applications in several fields. However it is only best at modeling symmetric data and fails to provide best fit to asymmetric and heavy tailed datasets. As a result several researchers have proposed the generalization of the normal distribution to improve upon its flexibility in modeling real data sets with varying degrees of skewness and kurtosis. These include the lognormal [1], folded normal [2], skew normal [3], beta normal [4], the generalized normal [5], McDonald normal [6], gamma normal [7], odd-log logistic normal [8] and transmuted normal [9] distributions among others. This study therefore proposes another generalization of the normal distribution called the unit half logistic normal (UHLN) using the unit half logistic generated (UHL-G) family of distribution proposed by [10] by introducing a shape parameter into the normal distribution.

The remaining part of the article is organized as follows: Section 2 presents the UHLN distribution. Section 3 presents the statistical properties of the UHLN distribution. Section 4 presents the estimation methods for the parameters of the new distribution. Section 5 presents a simulation experiment to assess the performance of the estimators of the UHLN distribution. Empirical applications of the UHLN using three different datasets are presented in section 6. Section 7 presents the conclusion of the study.

2. UHLN Distribution

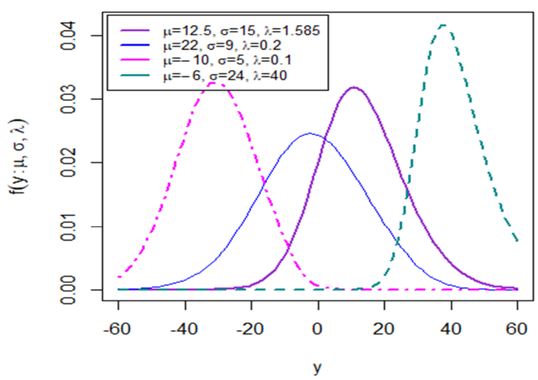
For any random variable Y that follows the half logistic distribution, [10] proposed that, a random variable $T = e^{-Y}$ follows the unit half logistic distribution with its PDF given as $f(t) = \frac{2\lambda t^{\lambda-1}}{(1+t^{\lambda})^2}, \lambda > 0, 0 < t < 1.$ They applied the T - X transformation technique proposed by [11], to transform the PDF of the unit half logistic distribution to obtain the unit half logistic-generated (UHL - G) family with CDF given as $F(x) = \int_0^{G(x;\xi)} \frac{2\lambda t^{\lambda-1}}{(1+t^{\lambda})^2} dt = \frac{2G(x;\xi)^{\lambda}}{1+G(x;\xi)^{\lambda}}, x \in \mathbb{R}, \lambda > 0,$ where, λ is an extra shape parameter, ξ is a $p \times 1$ vector of parameters and $G(x;\xi)$ is the CDF of the baseline distribution with corresponding PDF $g(x;\xi)$ of the UHL-G family is given as $f(x) = \frac{2\lambda g(x;\xi) G(x;\xi)^{\lambda-1}}{\left[1+G(x;\xi)^{\lambda}\right]^2}, x \in \mathbb{R}.$ Suppose that the cumulative distribution function (CDF) of normal distribution is given by $G\left(y;\mu,\sigma\right)=\Phi\left(\frac{y-\mu}{\sigma}\right), -\infty<y<\infty, -\infty<\mu<\infty, 0<\sigma$ and the corresponding probability density function (PDF) given by $g\left(y;\mu,\sigma\right)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right\}, -\infty<y<\infty.$ The CDF, PDF and Hazard rate function of the random variable Y that follow the UHLN distribution, using [10] UHL-G family of distribution are respectively given by

$$F\left(y;\mu,\sigma\right)=\frac{2\left[\Phi\left(\frac{y-\mu}{\sigma}\right)\right]^{\lambda}}{1+\left[\Phi\left(\frac{y-\mu}{\sigma}\right)\right]^{\lambda}}, y\in\mathbb{R}, \tag{1}$$

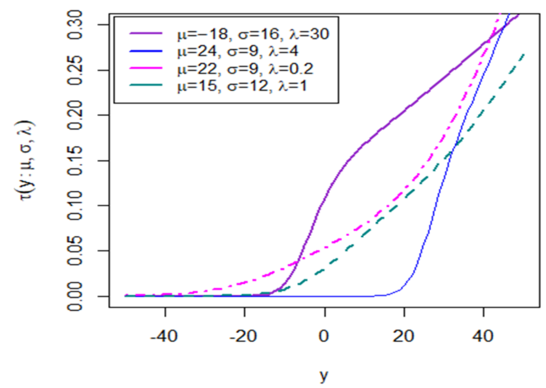
$$f\left(y;\mu,\sigma,\lambda\right)=\frac{2\lambda\exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right]\left[\Phi\left(\frac{y-\mu}{\sigma}\right)\right]^{\lambda-1}}{\sigma\sqrt{2\pi}\left[1+\left[\Phi\left(\frac{y-\mu}{\sigma}\right)\right]^{\lambda}\right]^2}, y\in\mathbb{R}, \tag{2}$$

$$\tau\left(y;\mu,\sigma,\lambda\right)=\frac{2\lambda\exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right]\left[\Phi\left(\frac{y-\mu}{\sigma}\right)\right]^{\lambda-1}}{\sigma\sqrt{2\pi}\left[1-\left[\Phi\left(\frac{y-\mu}{\sigma}\right)\right]^{2\lambda}\right]}, -\infty<y<\infty, \tag{3}$$

where, $-\infty<\mu<\infty$ is the location parameter, $\sigma>0$ is the scale parameter and $\lambda>0$ is the shape parameter. Henceforth, we represent a random variable Y that follows the UHLN distribution as $Y\sim\text{UHLN}\left(\mu,\sigma,\lambda\right)$.



(a)



(b)

Figure 1: (a) PDF and (b) Hazard plot of the UHLN distribution

Figure 1 gives the plot of the PDF and hazard rate functions respectively for different parameter values. From Figure 1, it can be observed that the density of the UHLN distribution exhibits different shapes such as right skewed, left skewed and nearly symmetric with varying degrees of kurtosis measures. The hazard rate function also exhibits different monotonically increasing shapes for different combination of parameter values.

3. Statistical Properties of the UHLN Distribution

3.1 Mixture Representation

The mixture representation of the density of the UHLN distribution is very helpful in deriving the properties of the UHLN distribution. The mixture representation is presented in this subsection.

Lemma 1. The PDF of the UHLN distribution can be expressed in a mixture form as

$$f\left(y;\mu,\sigma,\lambda\right)=2\lambda\sum_{i=0}^{\infty}\left(-1\right)^{i}\left(i+1\right)\phi\left(\frac{y-\mu}{\sigma}\right)\left[\Phi\left(\frac{y-\mu}{\sigma}\right)\right]^{\lambda(i+1)-1},-\infty<y<\infty, \tag{4}$$

where, $\phi\left(\frac{y-\mu}{\sigma}\right)$ is the PDF of the normal distribution.

Proof. Using the generalized binomial expansion,

$\left(y+a\right)^{-n}=\sum_{k=0}^{\infty}\left(-1\right)^{k}\binom{n+k-1}{k}y^{k}a^{-n-k},\left|y\right|<a$ and that $0<\Phi\left(\frac{y-\mu}{\sigma}\right)<1$. Then we have

$$\left[\left[\Phi\left(\frac{y-\mu}{\sigma}\right)\right]^{\lambda}+1\right]^{-2}=\sum_{i=0}^{\infty}\left(-1\right)^{i}\binom{2+i-1}{i}\left[\Phi\left(\frac{y-\mu}{\sigma}\right)\right]^{\lambda i} \tag{5}$$

Substituting equation (5) into the PDF of the UHLN distribution stated in equation (2), we get

$$f\left(y;\mu,\sigma,\lambda\right)=\frac{2\lambda\exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^{2}\right]}{\sigma\sqrt{2\pi}}\sum_{i=0}^{\infty}\left(-1\right)^{i}\binom{i+1}{i}\left[\Phi\left(\frac{y-\mu}{\sigma}\right)\right]^{\lambda(i+1)-1}$$

Hence

$$f\left(y;\mu,\sigma,\lambda\right)=2\lambda\sum_{i=0}^{\infty}\left(-1\right)^{i}\binom{i+1}{i}\phi\left(\frac{y-\mu}{\sigma}\right)\left[\Phi\left(\frac{y-\mu}{\sigma}\right)\right]^{\lambda(i+1)-1},\ y\in\mathbb{R}.$$

This completes the proof of lemma 1.

3.2 Cumulative Hazard Function

Cumulative hazard function (CHF) is the accumulated probability of failures up till time t. The CHF is very useful in survival and reliability data analysis in the field of biology and engineering. By definition, the CHF is given by

$$\mathrm{H}(y)=-\log S(y)=\int_{0}^{y}h(t)dt.$$

Hence the CHF $\mathrm{H}\left(y;\mu,\sigma,\lambda\right)$ of the UHLN distribution is given by

$$\mathrm{H}\left(y;\mu,\sigma,\lambda\right)=\log\left[\frac{\left[\Phi\left(\frac{y-\mu}{\sigma}\right)\right]^{\lambda}}{1-\left[\Phi\left(\frac{y-\mu}{\sigma}\right)\right]^{\lambda}}\right]. \tag{6}$$

3.3 Hazard Rate Average Function

Hazard rate average function (HRAF) is a function that is used to determine the average rate of increase or decrease of the hazard rate. By definition the HRAF is given by

$$\varpi(y)=\frac{\mathrm{H}(y)}{y},$$

where, $\mathrm{H}(y)$ is the CHF. Therefore the HRAF of the UHLN distribution is given by

$$\varpi(y;\mu,\sigma,\lambda)=\frac{\log\left[\dfrac{\left[\Phi\left(\dfrac{y-\mu}{\sigma}\right)\right]^{\lambda}}{1-\left[\Phi\left(\dfrac{y-\mu}{\sigma}\right)\right]^{\lambda}}\right]}{y}. \tag{7}$$

3.4 Reversed Hazard Function

Reversed hazard rate function (RHRF) also known in literature as the retro hazard is very crucial in survival data analysis. It is used in the estimation of survival function and the analysis of censored data. The reversed hazard rate function is the ratio of the PDF to the corresponding CDF. Let $r(y;\mu,\sigma,\lambda)$ denote the reversed hazard rate function of the UHLN distribution. By definition, the retro hazard is given by

$$r(y;\mu,\sigma,\lambda)=\frac{f(y;\mu,\sigma,\lambda)}{F(y;\mu,\sigma,\lambda)}.$$

Therefore the HRAF of the UHLN distribution is given by

$$r(y;\mu,\sigma,\lambda)=\frac{\dfrac{2\lambda\exp\left[-\dfrac{1}{2}\left(\dfrac{y-\mu}{\sigma}\right)^{2}\right]\left[\Phi\left(\dfrac{y-\mu}{\sigma}\right)\right]^{\lambda-1}}{\sigma\sqrt{2\pi}\left[1+\left[\Phi\left(\dfrac{y-\mu}{\sigma}\right)\right]^{\lambda}\right]^{2}}}{\dfrac{2\left[\Phi\left(\dfrac{y-\mu}{\sigma}\right)\right]^{\lambda}}{1+\left[\Phi\left(\dfrac{y-\mu}{\sigma}\right)\right]^{\lambda}}}. \tag{8}$$

3.5 Quantile Function

The quantile function is the inverse of the CDF and it is very useful in generating random numbers for a given probability distribution. It is also useful in describing some properties of a given distribution such as the median, quartiles, skewness and kurtosis [12]. Let the quantile function of UHLN distribution be denoted as $Q_y(u)$. The quantile function of the UHLN distribution is given by

$$Q_y(u) = \sigma\sqrt{2}\mathrm{erf}^{-1}\left[2\left(\frac{u}{2-u}\right)^{\frac{1}{\lambda}} - 1\right] + \mu, u \in [0,1], \tag{9}$$

where $\mathrm{erf}^{-1}(\cdot)$ is the inverse of the error function.

The mean is usually overlooked for the median as the most appropriate measure of central tendency required for data that contain outliers or extreme values [12]. The median of UHLN distribution $Q_y(0.5)$ is obtained by substituting $u = 0.5$ into the quantile function expressed in equation (9).

Therefore the median of the UHLN distribution is given by

$$Q_y(0.5) = \sigma\sqrt{2}erf^{-1}\left[2(3)^{-\frac{1}{\lambda}} - 1\right] + \mu. \tag{10}$$

Similarly, the lower quartile $Q_y(0.25)$ and the upper quartile $Q_y(0.75)$ are obtained by substituting $u = 0.25$ and $u = 0.75$ respectively into the quantile function. The inter quartile range (IQR) is $IQR = Q_y(0.75) - Q_y(0.25)$, the quartile deviation (QD) is $QD = \frac{\{Q_y(0.75) - Q_y(0.25)\}}{2}$ and the coefficient of quartile deviation (CQD) is $CQD = \frac{\{Q_y(0.75) - Q_y(0.25)\}}{\{Q_y(0.75) + Q_y(0.25)\}}$. Bowley's coefficient of skewness (BS) proposed by [13] and Moors' coefficient of kurtosis (MK) proposed by [14] are both useful measures of shapes of a distribution and are respectively calculated using the quantiles as follows;

$$BS=\frac{\left\{Q_y\left(0.75\right)+Q_y\left(0.25\right)-2Q_y\left(0.5\right)\right\}}{\left\{Q_y\left(0.75\right)-Q_y\left(0.25\right)\right\}},\text{ and}$$

$$MK=\frac{\left\{Q_y\left(0.375\right)-Q_y\left(0.125\right)+Q_y\left(0.875\right)-Q_y\left(0.625\right)\right\}}{\left\{Q_y\left(0.75\right)-Q_y\left(0.25\right)\right\}}.$$

Table 1 shows the lower quartile, median, upper quartile, inter quartile range, quartile deviation, coefficient of quartile deviation, Bowley's coefficient of skewness and Moor's coefficient of kurtosis for some chosen parameter values of the UHLN distribution. From Table 1, it can be observed that the Bowley's coefficient of skewness shows that the quantile of the UHLN distribution can be left skewed, right skewed or approximately symmetric and the Moor's coefficient of kurtosis shows the quantile of the UHLN distribution can only be leptokurtic.

Table 1: Quantiles and quantile measures of shapes

μ	σ	λ	$Q_y(0.25)$	$Q_y(0.5)$	$Q_y(0.75)$	*IQR*	*QD*	*CQD*	*BS*	*MK*
-10	**5**	**0.200**	-29.2409	-23.2123	-17.1015	12.1394	6.0697	-0.2620	0.0068	1.2276
		0.500	-20.2270	-16.1032	-11.7923	8.4347	4.2173	-0.2634	0.0221	1.2369
		0.100	-38.9451	-30.7288	-22.5471	16.3980	8.1990	-0.2667	-0.0021	1.2217
		0.600	-18.8097	-14.9672	-10.9223	7.8874	3.9437	-0.2653	0.0256	1.2389
12.5	**15**	**1.585**	4.3288	12.5003	21.4435	17.1147	8.5573	0.6641	0.0451	1.2492
		1.000	-3.5136	6.0391	16.3002	19.8138	9.9069	1.5496	0.0358	1.2443
		2.600	11.4882	18.4982	26.3236	14.8354	7.4177	0.3923	0.0550	1.2542
		10.500	26.8620	31.7796	37.5461	10.6841	5.3420	0.1659	0.0794	1.2666
0	**27**	**5.000**	12.4478	22.9893	35.0498	22.6020	11.301	0.4759	0.0672	1.2604
		14.000	30.4424	38.7775	48.6344	18.1920	9.0960	0.2301	0.0837	1.2688
		28.000	40.4314	47.7535	56.5693	16.1379	8.0690	0.1664	0.0926	1.2736
		9.000	23.2650	32.4220	43.1081	19.8432	9.9216	0.2990	0.0771	1.2654

3.6 Moments

Moments are very useful in statistical data analysis. They are used to obtain the measures of central tendencies, measures of dispersion and measures of shapes. The following proposition gives the r^{th} non-central moment of the UHLN distribution.

Proposition 1. The r^{th} non-central moment of the UHLN distribution is given by

$$\mu_r' = 2\lambda \sum_{i=0}^{\infty}\sum_{j=0}^{r}(-1)^i (i+1)\left(\frac{\sigma}{\mu}\right)^j \sigma\mu^r \binom{r}{j}\left[\sum_{k=0}^{\infty}(-1)^k \binom{\lambda(i+1)-1}{k} I_{j,k} + (-1)^j I_{j,\lambda(i+1)-1}\right], \qquad (11)$$

where, $I_{j,k} = \int_0^{\infty} z^j \phi(z)\left[1-\Phi(z)\right]^k dz$ and $I_{j,\lambda(i+1)-1} = \int_0^{\infty} z^j \phi(z)\left[1-\Phi(z)\right]^{\lambda(i+1)-1} dz$.

Proof. By definition

$$\mu_r' = \int_{-\infty}^{\infty} y^r dF(y).$$

Substituting the mixture form of the PDF from Lemma 1, into the definition and letting $z = \left(\frac{y-\mu}{\sigma}\right)$, then $y = \sigma z + \mu$, which also implies $\frac{dy}{dz} = \sigma$ or $dy = \sigma dz$.

Thus

$$\mu_r' = 2\lambda \sum_{i=0}^{\infty}(-1)^i (i+1)\sigma \int_{-\infty}^{\infty} (\mu + \sigma z)^r \phi(z)\left[\Phi(z)\right]^{\lambda(i+1)-1} dz. \qquad (12)$$

Simplifying using binomial expansion and further evaluating the integral in equation (12) thus

$$\int_{-\infty}^{\infty} z^j \phi(z)\left[\Phi(z)\right]^{\lambda(i+1)-1} dz = \sum_{k=0}^{\infty}(-1)^k \binom{\lambda(i+1)-1}{k} \int_0^{\infty} z^j \phi(z)\left[1-\Phi(z)\right]^k dz + (-1)^j \int_0^{\infty} z^j \phi(z)\left[1-\Phi(z)\right]^{\lambda(i+1)-1} dz. \qquad (13)$$

Combining equations (12) and (13) we have

$$\mu_r' = 2\lambda \sum_{i=0}^{\infty}\sum_{j=0}^{r}(-1)^i (i+1)\left(\frac{\sigma}{\mu}\right)^j \sigma\mu^r \binom{r}{j}\left[\sum_{k=0}^{\infty}(-1)^k \binom{\lambda(i+1)-1}{k} I_{j,k} + (-1)^j I_{j,\lambda(i+1)-1}\right].$$

This completes the proof of proposition 1.

Table 2 shows the first six moments, standard deviation (SD), coefficient of variation (CV), coefficient of skewness (CS) and coefficient of kurtosis (CK) for $\text{I}: \mu = 5, \sigma = 40$ and $\lambda = 34$, $\text{II}: \mu = 27, \sigma = 16.3$ and $\lambda = 0.5$, $\text{III}: \mu = 3, \sigma = 20$ and $\lambda = 33$
and $\text{IV}: \mu = 25, \sigma = 24$ and $\lambda = 20$ parameter values of the UHLN distribution. The SD, CV, CS and CK values are calculated using;

$$\text{SD} = \sqrt{\mu_2' - \mu^2}, \qquad \text{CV} = \frac{\sigma}{\mu} = \sqrt{\frac{\mu_2'}{\mu^2} - 1}, \qquad \text{CS} = \frac{\mu_3' - 3\mu\mu_2' + 2\mu^3}{\left(\mu_2' - \mu^2\right)^{3/2}}, \text{ and}$$

$$\text{CK} = \frac{\mu_4' - 4\mu\mu_3' + 6\mu^2\mu_2' - 3\mu^4}{\left(\mu_2' - \mu^2\right)^2}, \text{ respectively.}$$

From Table 2, the CS values show that the UHLN distribution can be left skewed, right skewed or nearly symmetric. Also, the values of the CK show that the UHLN distribution can be leptokurtic, platykurtic and closely mesokurtic. This implies the UHLN distribution can model heavy or light tailed data as well as left, right or symmetric datasets. These characteristics of the UHLN distribution signify it is highly flexible.

Table 2: First six moments, SD, CV, CS and CK

μ_r'	**I**	**II**	**III**	**IV**
μ_1'	5.116436	7.204978	3.861621×10^{1}	1.952992×10^{1}
μ_2'	2.793911×10^{2}	4.472179×10^{2}	1.607290×10^{3}	1.055911×10^{3}
μ_3'	1.536615×10^{4}	8.924934×10^{3}	6.933251×10^{4}	5.749423×10^{4}
μ_4'	8.503834×10^{5}	5.628270×10^{5}	3.091270×10^{6}	3.150428×10^{6}
μ_5'	4.731781×10^{7}	1.619216×10^{7}	1.420577×10^{8}	1.736130×10^{8}
μ_6'	2.645545×10^{9}	1.043645×10^{9}	6.709021×10^{9}	9.616622×10^{9}
SD	1.591267×10^{1}	1.988231×10^{1}	1.077397×10^{1}	2.597101×10^{1}
CV	3.110109	2.759524	2.790011×10^{-1}	1.329806
CS	2.815767	8.135320×10^{-4}	-1.359280	6.009370×10^{-1}
CK	9.010588	2.795356	6.794911	1.404617

3.7 Moment Generating Function

MGF is an alternative representation of the PDF of a random variable. If the MGF of a random variable exist, it can be used to obtain all the moments of the random variable. The MGF of the UHLN distribution is given by the following proposition.

Proposition 2. The MGF of the UHLN distribution is given by

$$\mathrm{M}_Y(t)=2\lambda\sum_{r=0}^{\infty}\sum_{i=0}^{\infty}\sum_{j=0}^{r}\frac{t^r}{r!}(-1)^i(i+1)\binom{r}{j}\left(\frac{\sigma}{\mu}\right)^j\sigma\mu^r\times\left[\sum_{k=0}^{\infty}(-1)^k\binom{\lambda(i+1)-1}{k}I_{j,k}+(-1)^j I_{j,\lambda(i+1)-1}\right]. \tag{14}$$

Proof. By definition, the MGF is given by

$$\mathrm{M}_Y(t)=\mathrm{E}\left(e^{tY}\right)=\int_{-\infty}^{\infty}e^{ty}dF(y).$$

Using the Taylors series,

$$\mathrm{M}_Y(t)=\sum_{r=0}^{\infty}\frac{t^r}{r!}\mu_r^{'}.$$

Hence

$$\mathrm{M}_Y(t)=2\lambda\sum_{r=0}^{\infty}\sum_{i=0}^{\infty}\sum_{j=0}^{r}\frac{t^r}{r!}(-1)^i(i+1)\binom{r}{j}\left(\frac{\sigma}{\mu}\right)^j\sigma\mu^r\left[\sum_{k=0}^{\infty}(-1)^k\binom{\lambda(i+1)-1}{k}I_{j,k}+(-1)^j I_{j,\lambda(i+1)-1}\right].$$

This completes the proof of proposition 2.

3.8 Order Statistics

Order statistics is an elementary tool but vital in non-parametric statistics and statistical inference. It helps in computing the minimum and maximum values as well as the range of a sampled random variable. It is also useful in estimating the sample median and other quantiles. Suppose that $Y_1, Y_2, ..., Y_n$ is an independent identically distributed random sample from the UHLN distribution, that is $Y \sim \mathrm{UHLN}(\mu,\sigma,\lambda)$, where, $Y_i^{'s}$ are sorted in increasing order of magnitude with $Y_i < Y_{i+1} < ... < Y_n, \forall Y_i^{'s}, i = 1,2,...,n.$ The PDF of the p^{th} order statistics of the UHLN distribution for $p = 1,2,...,n$ is given by

$$f_{p:n}(y)=\frac{1}{B(p,n-p+1)}f(y)\left[F(y)\right]^{p-1}\left[1-F(y)\right]^{n-p}, \qquad (15)$$

where, $f(y)$ and $F(y)$ are the PDF and the CDF of the UHLN distribution respectively, and $B(\cdot,\cdot)$ is the beta function. Since $0 < F(y) < 1$ for $-\infty < y < \infty$. Using binomial series to expand and evaluate the CDF in equation (15) the PDF of the p^{th} order statistics of the UHLN distribution in equation (15) becomes

$$f_{p:n}(y)=\frac{1}{B(p,n-p+1)}f(y)\sum_{j=0}^{n-p}(-1)^j\binom{n-p}{j}\left[F(y)\right]^{p+j-1}. \qquad (16)$$

Further substituting the PDF and the CDF of the UHLN distribution into equation (16), yields the PDF of the p^{th} order statistics of the UHLN distribution.

$$f_{p:n}(y)=\frac{2\lambda}{B(p,n-p+1)}\sum_{j=0}^{n-p}\sum_{i=0}^{\infty}(-1)^{i+j}(i+1)\binom{n-p}{j}\phi\left(\frac{y-\mu}{\sigma}\right)\left[\Phi\left(\frac{y-\mu}{\sigma}\right)\right]^{\lambda(i+1)-1}\times\left[\frac{2\left[\Phi\left(\frac{y-\mu}{\sigma}\right)\right]^{\lambda}}{1+\left[\Phi\left(\frac{y-\mu}{\sigma}\right)\right]^{\lambda}}\right]^{p+j-1},-\infty<y<\infty \tag{17}$$

4. Parameter Estimation

This section presents the parameter estimation methods for the UHLN distribution. Four estimation methods are used to estimate the parameters. These include: Maximum likelihood estimation (MLE), ordinary least squares (OLS), weighted least squares (WLS) and Cramér-von Mises (CVM).

4.1 Maximum Likelihood Estimation

Supposed that $y_1, y_2, ..., y_n$ are independent identically distributed random observations of size n obtained from $\text{UHLN}(y;\mu,\sigma,\lambda)$. If $\vartheta=(\mu,\sigma,\lambda)'$ is the vector of parameters, then total log-likelihood function is given by

$$\ell(y;\vartheta)=n\log(2\lambda)-\left[n\log\left(\sigma\sqrt{2\pi}\right)\right]-\left[\frac{1}{2}\sum_{i=0}^{n}\left(\frac{y_i-\mu}{\sigma}\right)^2\right]+\left[(\lambda-1)\sum_{i=0}^{n}\log\left[\frac{1}{2}\left\{1+\text{erf}\left(\frac{y_i-\mu}{\sigma\sqrt{2}}\right)\right\}\right]\right]-\left[2\sum_{i=1}^{n}\log\left[1+\left[\frac{1}{2}\left\{1+\text{erf}\left(\frac{y_i-\mu}{\sigma\sqrt{2}}\right)\right\}\right]^{\lambda}\right]\right] \tag{18}$$

The maximum likelihood estimates of the parameters can be obtained by maximizing directly the total log-likelihood function in equation (18). Alternatively, the score functions can be equated to zero and solving the system of equations to obtain the maximum likelihood estimates. The score functions are obtained by differentiating equation (18) with respect to each of the parameters. The score functions are:

$$\frac{\partial\ell(y;\vartheta)}{\partial\mu}=(1-\lambda)\sum_{i=1}^{n}\frac{\sqrt{\frac{2}{\pi}}\exp\left[-\frac{1}{2}\left(\frac{y_i-\mu}{\sigma}\right)^2\right]}{\sigma\left\{1+\text{erf}\left(\frac{y_i-\mu}{\sigma\sqrt{2}}\right)\right\}}+2\sum_{i=1}^{n}\frac{\lambda\left\{1+\text{erf}\left(\frac{y_i-\mu}{\sigma\sqrt{2}}\right)\right\}^{\lambda-1}\times\exp\left[-\frac{1}{2}\left(\frac{y_i-\mu}{\sigma}\right)^2\right]}{\sigma\sqrt{\pi}\left(\sqrt{2}\right)^{\lambda}\left[\left(1+2^{-\lambda}\right)\left\{1+\text{erf}\left(\frac{y_i-\mu}{\sigma\sqrt{2}}\right)\right\}^{\lambda}\right]}+\frac{1}{2}\sum_{i=1}^{n}\frac{2(y_i-\mu)}{\sigma^2}, \tag{19}$$

$$\frac{\partial\ell(y;\vartheta)}{\partial\sigma}=-\frac{n}{\sigma}+(1-\lambda)\sum_{i=1}^{n}\frac{(y_i-\mu)\sqrt{\frac{2}{\pi}}\exp\left[-\frac{1}{2}\left(\frac{y_i-\mu}{\sigma}\right)^2\right]}{\sigma^2\left\{1+\operatorname{erf}\left(\frac{y_i-\mu}{\sigma\sqrt{2}}\right)\right\}}+2\sum_{i=1}^{n}\frac{(y_i-\mu)\exp\left[-\frac{1}{2}\left(\frac{y_i-\mu}{\sigma}\right)^2\right]\lambda\left\{1+\operatorname{erf}\left(\frac{y_i-\mu}{\sigma\sqrt{2}}\right)\right\}^{\lambda-1}}{\sigma^2\sqrt{\pi}\left(\sqrt{2}\right)^{\lambda}\left[\left(1+2^{-\lambda}\right)\left\{1+\operatorname{erf}\left(\frac{y_i-\mu}{\sigma\sqrt{2}}\right)\right\}^{\lambda}\right]}+\frac{1}{2}\sum_{i=1}^{n}\frac{2}{\sigma}\left(\frac{y_i-\mu}{\sigma}\right)^2, \tag{20}$$

and

$$\frac{\partial\ell(y;\vartheta)}{\partial\lambda}=\frac{n}{\lambda}+\sum_{i=1}^{n}\log\left[\frac{1}{2}\left\{1+\operatorname{erf}\left(\frac{y_i-\mu}{\sigma\sqrt{2}}\right)\right\}\right]+2\sum_{i=1}^{n}\frac{2^{-\lambda}\left\{1+\operatorname{erf}\left(\frac{y_i-\mu}{\sigma\sqrt{2}}\right)\right\}^{\lambda}\left[\log\left\{\frac{2}{1+\operatorname{erf}\left(\frac{y_i-\mu}{\sigma\sqrt{2}}\right)}\right\}\right]}{1+2^{-\lambda}\left\{1+\operatorname{erf}\left(\frac{y_i-\mu}{\sigma\sqrt{2}}\right)\right\}^{\lambda}}. \tag{21}$$

The score functions are then solved simultaneously by equating equations (19), (20) and (21) to zero to obtain the maximum likelihood estimates of each parameter. That is, $\frac{\partial\ell(y;\vartheta)}{\partial\mu}=0$, $\frac{\partial\ell(y;\vartheta)}{\partial\sigma}=0$ and $\frac{\partial\ell(y;\vartheta)}{\partial\lambda}=0$. However the score functions do not have closed form, therefore the resulting system of equations are solved numerically to obtain the parameter estimates $\hat{\mu},\hat{\sigma}$ and $\hat{\lambda}$.

4.2 Least Squares Estimation

[15] developed the least squares estimation method for estimating the parameters of a distribution which comprise of the ordinary least squares (OLS) and the weighted least squares (WLS). Supposed that $y_{(1)},y_{(2)},...,y_{(n)}$ are independent identically distributed random observations of size n obtained from $\text{UHLN}(y;\mu,\sigma,\lambda)$, where, $y_{(i)}^{'s}$ are sorted in order statistics. If $\vartheta=(\mu,\sigma,\lambda)'$ is the vector of parameters, then the OLS parameter estimates $\hat{\mu}_{OLS},\hat{\sigma}_{OLS}$ and $\hat{\lambda}_{OLS}$ of the UHLN distribution are obtained by minimizing the function:

$$L(y;\vartheta)=\sum_{i=1}^{n}\left[F\left(y_{(i)};\vartheta\right)-\frac{i}{n+1}\right]^2, \tag{22}$$

with respect to the parameters μ, σ and λ. Similarly, the WLS estimates $\hat{\lambda}_{WLS}, \hat{\mu}_{WLS}$ and $\hat{\sigma}_{WLS}$ are obtained by minimizing the function

$$WL(y;\vartheta) = \sum_{i=1}^{n} \frac{(n+1)^2 (n+2)}{i(n-i+1)} \left[F\left(y_{(i)};\vartheta\right) - \frac{i}{n+1} \right]^2, \tag{23}$$

with respect to the parameters μ, σ and λ to obtain the following equations.

4.3 Cramér-von Mises Estimation

The CVM approach is an alternative parameter estimation technique that is based on the minimum distance estimation methods and observed to have the least bias relative to the other minimum distance estimation techniques. For independent identically distributed random observations, $y_{(1)}, y_{(2)}, ..., y_{(n)}$ of size n sorted in order statistics obtained from $\text{UHLN}(y;\mu,\sigma,\lambda)$. If $\vartheta = (\mu,\sigma,\lambda)'$ is the vector of parameters, then CVM parameter estimates $\hat{\lambda}_{CVM}, \hat{\mu}_{CVM}$ and $\hat{\sigma}_{CVM}$. of the UHLN distribution are obtained by minimizing the function:

$$\text{CVM}(y;\vartheta) = \frac{1}{12n} + \sum_{i=1}^{n} \left[F\left(y_{(i)};\vartheta\right) - \frac{2i-1}{2n} \right]^2, \tag{24}$$

with respect to the parameters μ, σ and λ.

5. Simulation Study

A simulation experiment was carried out with three different sets of parameter values: $\text{I}: \mu = 4.4, \sigma = 9.5, \lambda = 0.5$, $\text{II}: \mu = -0.5, \sigma = 12.5, \lambda = 2.9$ and $\text{III}: \mu = 0.3, \sigma = 14.5, \lambda = 6.6$ using six different sample sizes: $n = 25, 50, 75, 100, 125, 150$ and the process was replicated $\text{N} = 5000$. The average absolute bias (AB) and root mean square error (RMSE) of the estimators of the parameters are computed and compared to identify the best estimation techniques for the parameters. The outcome of the simulation experiments are presented in Tables 3, 4 and 5. The AB and RMSE of the parameters are calculated using,

$\text{AB} = \left| \frac{1}{\text{N}} \sum_{i=1}^{\text{N}} \left(\hat{\vartheta} - \vartheta \right) \right|$, and $\text{RMSE} = \sqrt{\frac{1}{\text{N}} \sum_{i=1}^{\text{N}} \left(\hat{\vartheta} - \vartheta \right)^2}$, respectively, where, $\vartheta = (\mu \quad \sigma \quad \lambda)'$. From Tables 3, 4 and 5 it can be seen that, the MLE is the most efficient and the only consistent estimator of the UHLN distribution since it has the least AB and RMSE values and its AB and RMSE values reduces as the sample sizes increases. This implies that the MLE has clearly shown dominance over all the other proposed estimators and hence it is considered the best techniques to estimate the parameters of the UHLN distribution.

Table 3: Monte Carlo simulation results for $\mu = 4.4, \sigma = 9.5$ and $\lambda = 0.5$

Parameter	n	AB				RMSE			
		MLE	OLS	WLS	CVM	MLE	OLS	WLS	CVM
$\mu = 4.4$	25	3.1556	3.1404	2.9104	3.4385	5.2935	5.2847	5.0927	5.5667
	50	2.3897	2.3966	2.3940	2.8500	4.6202	4.6506	4.5758	4.9582
	75	2.2459	2.5698	2.4143	2.7920	4.4389	4.6463	4.1180	4.8767
	100	1.8392	1.9023	1.7993	2.0502	4.1478	4.2082	4.1085	4.4132
	125	1.7267	2.2199	1.7790	2.1378	4.0080	4.3804	4.0568	4.3929
	150	1.2970	1.9545	1.8018	2.0307	3.6721	4.1623	4.5434	4.2575
$\sigma = 9.5$	25	0.4495	0.7372	0.7968	0.2435	2.0452	2.1297	2.0611	2.2416
	50	0.4768	0.8414	0.8850	0.6318	1.8849	2.0355	1.9713	2.0124
	75	0.5401	0.9046	0.8503	0.7325	1.7429	1.8695	1.8112	1.8806
	100	0.4622	0.7919	0.6873	0.6372	1.6949	1.7890	1.7302	1.7867
	125	0.4236	0.8266	0.6655	0.6420	1.6015	1.7644	1.6582	1.7443
	150	0.3000	0.7869	0.6778	0.6705	1.4984	1.7090	1.6358	1.6916
$\lambda = 0.5$	25	0.2128	0.2069	0.2044	0.2078	0.3541	0.3351	0.3296	0.3542
	50	0.1716	0.1814	0.1835	0.1958	0.3231	0.3162	0.3146	0.3338
	75	0.1637	0.1935	0.1873	0.2003	0.3131	0.3181	0.3165	0.3324
	100	0.1348	0.1526	0.1439	0.1547	0.2929	0.2900	0.2884	0.3005
	125	0.1258	0.1723	0.1432	0.1607	0.2817	0.3051	0.2868	0.3043
	150	0.0983	0.1603	0.1454	0.1601	0.2598	0.2958	0.2867	0.3014

Table 4: Monte Carlo simulation results for $\mu=-5, \sigma=12.4$ and $\lambda=2.9$

Parameter	n	AB				RMSE			
		MLE	**OLS**	**WLS**	**CVM**	**MLE**	**OLS**	**WLS**	**CVM**
$\mu=-5$	**25**	2.6338	3.0224	3.1895	2.8604	8.5629	8.7407	8.8564	8.7900
	50	2.6297	2.9503	3.0833	3.0112	8.4652	8.9002	8.8170	8.9713
	75	2.3788	2.3827	2.2474	2.5721	8.2035	8.6286	8.3631	8.6837
	100	2.6922	2.8917	2.5788	3.2448	8.1241	8.6692	8.3836	8.9348
	125	2.7140	3.0816	2.8845	3.0943	8.0561	8.8038	8.4965	8.9503
	150	2.2135	2.6031	2.4478	2.6544	7.4250	8.2325	7.9472	8.5210
$\sigma=12.4$	**25**	1.5032	0.9033	1.0070	1.5416	3.4348	3.3142	3.3091	3.6513
	50	1.2299	1.0242	1.1010	1.3690	3.1738	3.2232	3.2390	3.3825
	75	1.1087	0.9298	0.8792	1.2200	2.9478	3.0840	2.9856	3.1994
	100	1.1518	1.0398	0.9743	1.3168	2.8993	3.0985	2.9224	3.2046
	125	1.1436	1.1689	1.1357	1.3236	2.8791	3.1288	3.0296	3.2389
	150	0.8929	0.9927	0.9274	1.1411	2.5722	2.9056	2.7628	3.0372
$\lambda=2.9$	**25**	0.0043	0.2018	0.2065	0.0812	1.8137	1.7131	1.7364	1.8050
	50	0.0912	0.1852	0.2149	0.1516	1.7460	1.7328	1.7384	1.7893
	75	0.0462	0.0434	0.0282	0.0535	1.7211	1.7417	1.7143	1.7737
	100	0.1735	0.2155	0.1508	0.2623	1.6718	1.7023	1.6721	1.7463
	125	0.1905	0.2340	0.2053	0.2006	1.6655	1.7333	1.7030	1.7785
	150	0.1341	0.1605	0.1496	0.1255	1.5606	1.6662	1.6323	1.7217

Table 5: Monte Carlo simulation results for $\mu = 0.3, \sigma = 14.5$ and $\lambda = 6.6$

Parameter	n	AB				RMSE			
		MLE	OLS	WLS	CVM	MLE	OLS	WLS	CVM
$\mu = 0.3$	**25**	0.4355	0.4911	0.4765	0.4533	1.0367	1.0577	1.0540	1.0406
	50	0.4581	0.5516	0.4946	0.5364	1.0565	1.0767	1.0644	1.0770
	75	0.4530	0.4719	0.4734	0.4871	1.0598	1.0621	1.0696	1.0762
	100	0.3636	0.4208	0.3668	0.4429	1.0444	1.0536	1.0346	1.0624
	125	0.3403	0.4351	0.3659	0.4441	1.0354	1.0546	1.0444	1.0610
	150	0.2842	0.4156	0.3026	0.4189	1.0212	1.0577	1.0174	1.0562
$\sigma = 14.5$	**25**	0.4821	0.4220	0.3451	0.7723	1.5395	1.5392	1.4975	1.7488
	50	0.1492	0.0296	0.0235	0.2637	1.1624	1.1438	1.1520	1.2175
	75	0.0453	0.0255	0.0035	0.1962	0.9529	1.0274	0.9746	1.0876
	100	0.0046	0.0690	0.0757	0.0286	0.8673	0.9325	0.8963	0.9566
	125	0.0237	0.0697	0.0501	0.0578	0.7831	0.8743	0.8238	0.8852
	150	0.0167	0.1080	0.0686	0.0064	0.7423	0.8318	0.7697	0.8539
$\lambda = 6.6$	**25**	0.9231	0.7344	0.6581	1.1059	1.8265	1.7940	1.7389	2.0037
	50	0.6323	0.4751	0.4790	0.6832	1.3837	1.3310	1.2846	1.4275
	75	0.5022	0.3610	0.4274	0.5635	1.1869	1.1543	1.1693	1.2920
	100	0.3432	0.2809	0.2563	0.3935	1.0191	1.0125	0.9549	1.0728
	125	0.3332	0.3038	0.2830	0.4105	0.9599	0.9858	0.9346	1.0347
	150	0.2759	0.2865	0.2322	0.3837	0.8477	0.9184	0.8414	0.9702

6 Applications

This section presents the application of the UHLN distribution using three empirical datasets. The flexibility of the UHLN distribution is also demonstrated in this section and compared with normal (N), transmuted normal (TN) [9], beta-normal (BN) [4], McDonald normal (McDN) [6] and logistic (L) distribution based on their log-likelihood (ℓ), Akaike Information Criterion (AIC), Corrected Akaike Information Criterion (AICc) and Bayesian Information Criterion (BIC) selection criteria and also on Anderson-Darling (A-D) and Cramér-von Mises (CVM) goodness-of-fit tests. When comparing candidate statistical distributions for any given data, the distribution with the highest log-likelihood (ℓ), whiles having the least AIC, AICc, BIC, A-D and CVM is considered the best for that data.

6.1 Dataset I: Deep Groove Ball Bearings

Table 6 presents the descriptive statistics of the number of million revolutions before failure of ball bearings. It can be seen from Table 6 that the minimum number of million revolutions before failure of ball bearings is 17.88 million and the maximum is 173.4 million. Also, the median number of million revolutions before failure is 67.8 million. The mean and standard deviation are 72.2296 million and 37.4804 million respectively. From Table 6, it is observed that the coefficient of skewness is 0.8812 and the excess kurtosis is 0.1921 which indicates that the data is right skewed and leptokurtic. The deep groove ball bearings data is a life test involving 23 balls, it is found in [16] and it was used previously by [17].

Table 6: Descriptive statistics for dataset I

Min.	Max.	Median	Mean	Std.	Skewness	Kurtosis
17.8800	173.4000	67.8000	72.2296	37.4804	0.8812	0.1921

Table 7 presents the maximum likelihood estimates of the parameters of fitted distributions with their corresponding standard errors and *p-values* both in brackets using dataset I. The significance of the parameter estimates were tested using their *p-values*. From Table 7, it is observed that the UHLN, N and L distributions have all their parameters to be significant at 5% level. One each of the parameters of TN and McDN are observed not to be significant at 5% level. Table 7 also indicated that all the four parameters of the BN are not significant at 5% level.

Table 7: Maximum likelihood estimates of parameters for data set I.

Model	$\hat{a}$	$\hat{b}$	$\hat{c}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\lambda}$
UHLN				-405.0937	127.3305	10000
SE				(77.0664)	(21.6921)	(0.4853)
p-value				(<0.0001)	(<0.0001)	(<0.0001)
N				72.2343	36.6585	
SE				(7.6438)	(5.4054)	
p-value				(<0.0001)	(<0.0001)	
TN				82.4460	37.6004	0.5226
SE				(12.5428)	(6.2907)	(0.4786)
p-value				(<0.0001)	(<0.0001)	(0.2744)
BN	4.1288	0.1792		8.2819	22.6891	
SE	(2.1855)	(0.1993)		(3.4488)	(12.5002)	
p-value	0.0589)	(0.3687)		(0.0163)	(0.0695)	
McDN	139.7333	364.2013	2.1053	-13.3090	783.9879	
SE	(21.4472)	(47.9309)	(0.0572)	(18.4637)	(54.1208)	
p-value	(<0.0001)	(<0.0001)	(<0.0001)	(0.471)	(<0.0001)	
L				68.3494	20.4664	
SE				(7.4551)	(3.5584)	
p-value				(<0.0001)	(<0.0001)	

Table 8 presents the log-likelihood, model selection criteria and goodness-of-fit test for all the six distributions under consideration using dataset I. From Table 8, it is observed that all the six distributions have passed the goodness-of-fit test for dataset I. Even though the goodness-of-fit test indicated that data follows all the distributions under consideration, but the UHLN distribution has the least chances of committing type I error. Results from Table 8 shows that UHLN distribution has the highest value of ℓ and the least values of AIC, AICc, BIC, A-D, and CVM which indicate that it has shown absolute dominance since the model with the highest values of ℓ and the least values of AIC, AICc, BIC, A-D, and CVM is the best. Therefore the UHLN distribution is the best model for dataset I.

Table 8: Log-likelihood and goodness-of-fit statistics for dataset I.

Model	ℓ	AIC	AICc	BIC	A-D	CVM
UHLN	**-113.18**	**232.3556**	**233.6187**	**235.7620**	**0.2296**	**0.0403**
p-value					(0.9801)	(0.9357)
N	-115.47	234.9445	235.5445	237.2155	0.6122	0.1070
p-value					(0.6347)	(0.5552)
TN	-115.05	236.1054	237.3685	239.5119	1.8969	0.3970
p-value					(0.1052)	(0.0728)
BN	-113.38	234.7671	236.9893	239.3091	0.3615	0.0704
p-value					(0.8848)	(0.7537)
McDN	-115.51	241.0101	244.5395	246.6875	0.6166	0.1077
p-value					(0.6305)	(0.552)
L	-115.35	234.6992	235.2992	239.3091	0.5122	0.0752
p-value					(0.7326)	(0.7246)

The probability probability plot gives the display of the goodness-of-fit tests of the fitted distributions in describing dataset II. From Figure 2, it can be seen that the UHLN distribution fit the data better than the others since it matches the diagonal line closely.

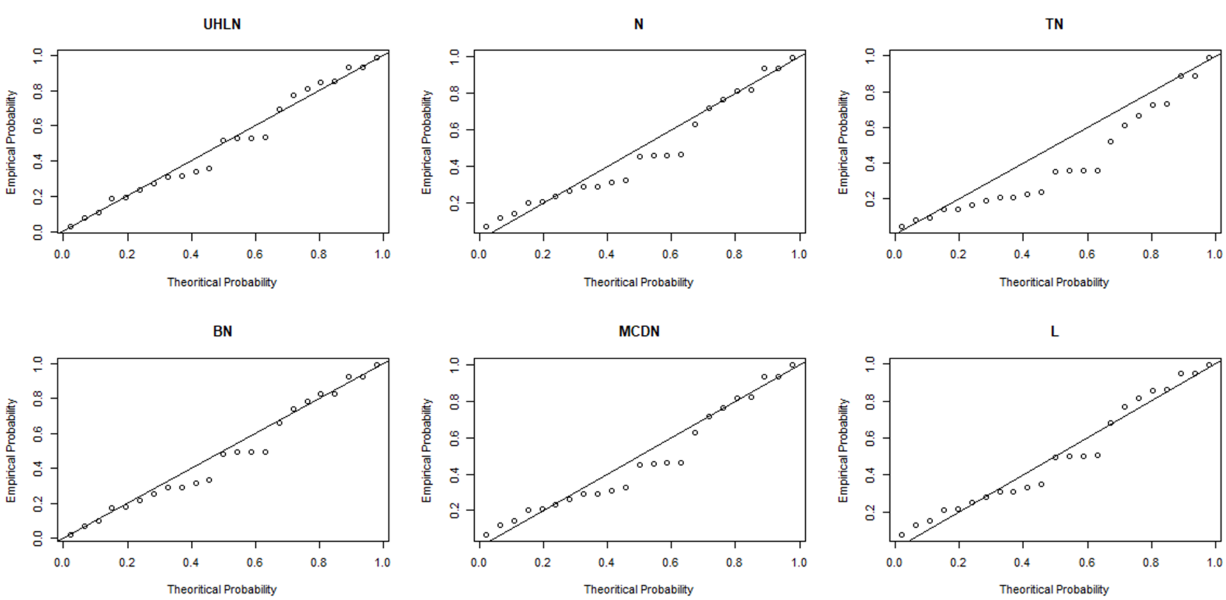


Figure 2: Probability probability plot of the fitted distributions for dataset I

6.2 Datasets II: Counts of White Blood Cell

Table 9 presents the descriptive statistics of white blood cell counts per litre of Australian Sports Athletes. From Table 9, the minimum count of white blood cell is 3.3 and the maximum is 14.3. The median count is 6.85 and the average and standard deviation of the counts of white cell are 7.1089 and 1.8003 respectively. It is also observed that the coefficient of skewness and excess kurtosis of the data are 0.8290 and 1.4055 respectively. This means that the data is skewed to the right and heavy tailed. Data II is the counts of white blood cell per litre of 202 Australian sports athletes and it is found in [18].

Table 9: Descriptive statistics for dataset II.

Min.	Max.	Median	Mean	Std.	Skewness	Kurtosis
3.3000	14.3000	6.8500	7.1089	1.8003	0.8290	1.4055

Table 10 presents the maximum likelihood estimates of the parameters of fitted distributions with their corresponding standard errors and *p-values* both in brackets using dataset II. From Table 10, it is observed that all the parameters of the UHLN and the other five distributions under consideration are significant except $\hat{a}$ of BN distribution.

Table 10: Maximum likelihood estimates of parameters for data set II.

Model	$\hat{a}$	$\hat{b}$	$\hat{c}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\lambda}$
UHLN				-5.7348	4.9556	198.2556
SE				(0.6449)	(0.2705)	(0.0110)
p-value				(<0.0001)	(<0.0001)	(<0.0001)
N				7.1089	1.7959	
SE				(0.1264)	(0.0893)	
p-value				(<0.0001)	(<0.0001)	
TN				7.8249	1.9154	0.6865
SE				(0.2409)	(0.1228)	(0.1747)
p-value				(<0.0001)	(<0.0001)	(<0.0001)
BN	224.1906	1195.3563		52.1533	44.8738	
SE	(218.7320)	(3.9220)		(11.6640)	(12.7910)	
p-value	(0.3053)	(<0.0001)		(<0.0001)	(0.0004)	
McDN	276.1720	611.5821	12.8034	-101.2062	79.7462	
SE	(0.0693)	(0.1007)	(1.1734)	(0.3869)	(2.5425)	
p-value	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	
L				6.9965	0.9905	
SE				(0.1212)	(0.05802)	
p-value				(<0.0001)	(<0.0001)	

Table 11 presents the log-likelihood, model selection criteria and goodness-of-fit test for all the six distributions under consideration using dataset II. The results of the goodness-of-fit tests contained in Table 11 shows that dataset II follows all the distributions under consideration except TN distribution. It is also observed that UHLN distribution has the least values of AIC, AICc, BIC, A-D and CVM and the highest of ℓ. This simply means that UHLN distribution is the best model for dataset II.

Table 11: Log-likelihood and goodness-of-fit statistics for dataset II.

Model	ℓ	AIC	AICc	BIC	A-D	CVM
UHLN	**-395.75**	**797.5013**	**797.6225**	**807.4261**	**0.3635**	**0.047182**
p-value					(0.8838)	(0.8937)
N	-404.90	813.7902	813.8505	820.4067	1.3570	0.22581
p-value					(0.2146)	(0.2226)
TN	-401.47	808.9469	809.0681	818.8717	19.161	3.9836
p-value					(<0.0001)	(<0.0001)
BN	-405.49	818.9858	819.1888	832.2188	1.4285	0.23737
p-value					(0.1945)	(0.2054)
McDN	-404.66	819.3199	819.626	835.8612	1.3236	0.22049
p-value					(0.2248)	(0.2311)
L	-401.58	807.1620	807.2223	813.7786	0.8736	0.10902
p-value					(0.4307)	(0.5429)

Figure 3 gives the probability probability plot of the fitted distributions using dataset II. The probability probability plot is used to check how well the fitted distributions respectively describe dataset II. Figure 3 shows clearly the UHLN distribution gives the best fit than the other five fitted distributions.

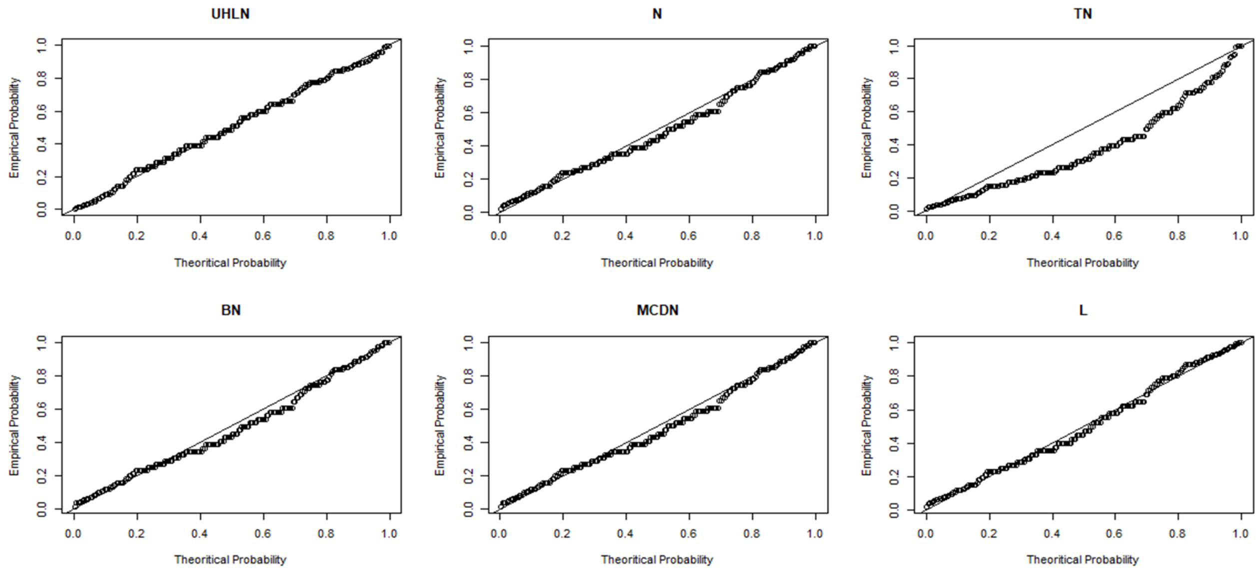


Figure 3: Probability probability plot of the fitted distributions for dataset II.

6.3 Dataset III: Weight of Pregnant Women

Table 12 presents the descriptive statistics of the weights of the 311 pregnant women at War Memorial Hospital-Navrongo. Table 12 shows that the minimum weight of the pregnant women is 46kg and the maximum weight is 101kg. Also, the average weight of the pregnant women is 73.49587kg. The median and standard deviation are 75kg and 13.5459kg respectively. From Table 12, it is observed that the coefficient of skewness is -0.1598 and the excess kurtosis is -0.8602 which indicates that the data is skewed to the left and less peaked relative to the kurtosis of the normal distribution. Data III is the weights in kilograms of 311 pregnant women during their last antenatal visit to the Navrongo War Memorial Hospital before delivery and it was sourced from the Navrongo War Memorial Hospital in January, 2022. The data on the weights of pregnant women obtained from the Navrongo War Memorial Hospital is presented in Appendix.

Table 12: Descriptive statistics for dataset III

Min.	Max.	Median	Mean	Std.	Skewness	Kurtosis
46	101	75	73.4958	13.5459	-0.1598	-0.8602

Table 13 presents the maximum likelihood estimates of the parameters of fitted distributions with their corresponding standard errors and *p-values* both in brackets using dataset III. The significance of the parameter estimates were tested using their *p-values*. From Table 13, it is observed that all the parameters of the UHLN, N, and L distributions are significant at 5% level.

Two parameters, $\hat{\mu}$ and $\hat{\sigma}$ for both BN and TN are significant at 5% level but the parameters $\hat{a}$, $\hat{b}$ and $\hat{\sigma}$ for BN, $\hat{\lambda}$ for TN and all the parameters of the McDN distribution are not significant at 5% level.

Table 13: Maximum likelihood estimates of parameters for data set III.

Model	$\hat{a}$	$\hat{b}$	$\hat{c}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\lambda}$
UHLN				101.0219	2.2149	0.0138
SE				(0.4514)	(0.5284)	(0.0061)
p-value				(<0.0001)	(<0.0001)	(0.0247)
N				73.4958	13.5242	
SE				(0.7669)	(0.5423)	
p-value				(<0.0001)	(<0.0001)	
TN				72.4902	13.5558	-0.1375
SE				(1.6582)	(0.5578)	(0.1996)
p-value				(<0.0001)	(<0.0001)	(0.4909)
BN	2.2940	21.2789		123.9667	36.5054	
SE	(3.0513)	(48.9181)		(60.9410)	(30.4714)	
p-value	(0.4522)	(0.6636)		(0.0419)	(0.2309)	
McDN	1.3768	36.4615	13.0726	19.8469	76.4386	
SE	(1.6966)	(158.0893)	(59.0268)	(346.7416)	(198.4442)	
p-value	(0.4171)	(0.8176)	(0.8247)	(0.9544)	(0.7001)	
L				73.7911	8.0608	
SE				(0.8119)	(0.3716)	
p-value				(<0.0001)	(<0.0001)	

Table 14 presents the log-likelihood, information selection criteria and goodness of fit statistics for the six distributions under consideration using dataset III. The A-D test in Table 14 fail to reject the hypotheses that dataset III follows the UHLN, N BN and McDN distributions. The CVM fail to reject that dataset III follows UHLN, N, BN, McDN and L distributions. The results contained in Table 14 indicate that the UHLN distribution is the best model for the data. The superiority of the UHLN distribution over the other models is very obvious since it has the highest value of the log-likehood and the least values of AIC, BIC, AICc, A-D and CVM as shown in Table 14.

Table 14: Log-likelihood and goodness-of-fit statistics for dataset III

Model	ℓ	AIC	AICc	BIC	A-D	CVM
UHLN	**-1245.94**	**2497.884**	**2497.962**	**2509.104**	**1.8348**	**0.3192**
p-value					(0.1135)	(0.1194)
N	-1251.28	2506.564	2506.603	2514.044	2.2035	0.3647
p-value					(0.0712)	(0.0897)
TN	-1251.05	2508.092	2508.171	2519.312	3.3454	0.6075
p-value					(0.0184)	(0.0214)
BN	-1250.47	2508.945	2509.076	2523.904	1.9926	0.3385
p-value					(0.0927)	(0.1056)
McDN	-1250.55	2511.103	2511.300	2529.802	1.9913	0.3397
p-value					(0.0929)	(0.1049)
L	-1261.79	2527.573	2527.612	2535.053	2.7346	0.4258
p-value					(0.0374)	(0.0617)

Figure 4 displays the probability probability plot of the fitted distributions using dataset III. The probability probability plot is a plot used to assess how well a distribution describes a given dataset. From Figure 4, the UHLN distribution provides the best fit for the data than the other five fitted distributions.

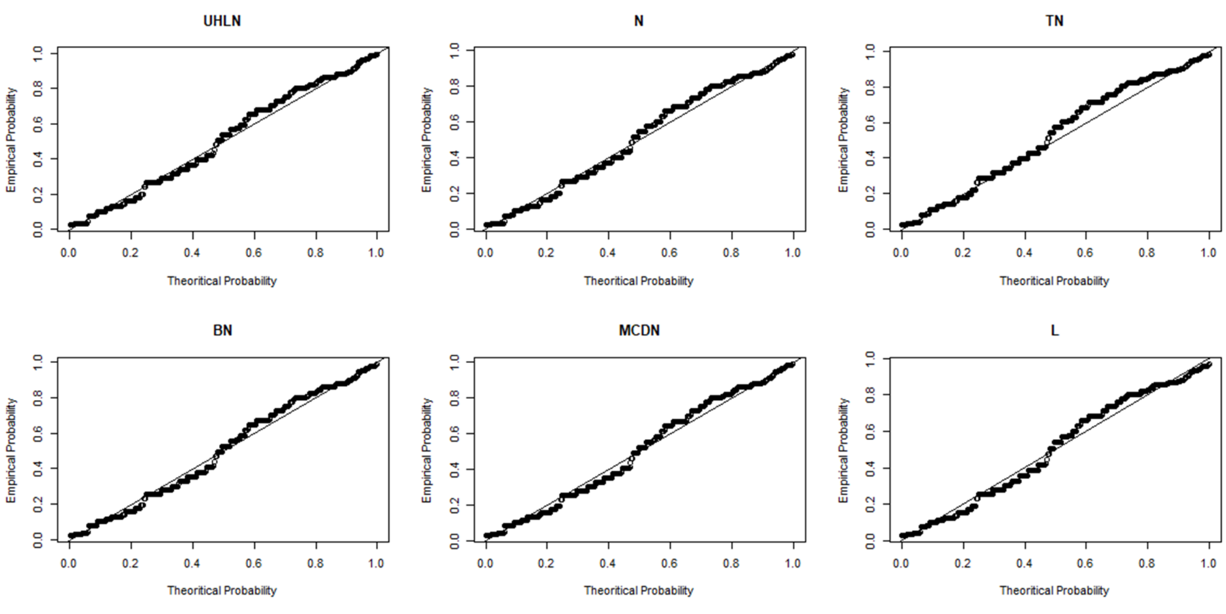


Figure 4: Probability probability plot of the fitted distributions for dataset III

7. Conclusions

This study has developed a new generalization of the normal distribution called the unit half logistic normal distribution (UHLN) by introducing a shape parameter λ into the normal distribution. The UHLN distribution turns out to be very flexible when it was fitted with three empirical datasets and compared to the modeling abilities of normal, transmuted normal, beta normal, McDonald normal and logistic distributions. The density and quantile functions of the UHLN distribution has shown that the distribution is highly flexible and capable of modeling left, right or symmetric datasets as well as leptokurtic, platykurtic or mesokurtic data. The maximum likelihood estimation technique was used in estimating the parameters of the UHLN distribution. Finally, three datasets were used to demonstrate the flexibility of the UHLN distribution and it was revealed that the new distribution provides the best fit for all the three datasets than several other competing models.

References

[1] Galton, F. (1879). The geometric mean in vital and society statistics. *Proceedings of the Royal Society of London*, **29**:365-367.

[2] Leone, F. C., Nelson, L. S., and Nottingham, R. B. (1961). The folded normal distribution. *Technometrics*, **3**(4): 543-550.

[3] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian journal of statistics*, 171-178.

[4] Eugene, N., Lee, C., and Famoye, F. (2002). Beta-normal distribution and its applications. *Communications in Statistics-Theory and methods*, **31**(4): 497-512.

[5] Nadarajah, S. (2005). A generalized normal distribution. *Journal of Applied statistics*, **32**(7): 685-694.

[6] Cordeiro, G. M., Cintra, R. J., Rêgo, L. C., and Ortega, E. M. (2012). The McDonald normal distribution. *Pakistan Journal of Statistics and Operation Research*, 301-329.

[7] Alzaatreh, A., Famoye, F., and Lee, C. (2014). The gamma-normal distribution: Properties and applications. *Computational Statistics and Data Analysis*, **69**: 67-80.

[8] da Braga, A. S., Cordeiro, G. M., Ortega, E. M., and da Cruz, J. N. (2016). The odd log-logistic normal distribution: Theory and applications in analysis of experiments. *Journal of Statistical Theory and Practice*, **10**(2): 311-335.

[9] Ieren, T. G., & Abdullahi, J. (2020). A transmuted Normal distribution: properties and applications. *Equity journal of science and technology*, ***7***(1):16-34.

[10] Abubakari, A. G., Nasiru, S., and Anzagra, L. (2021). Unit half logistic-generated family of distributions: Properties and applications. *Taru Journal of Organizational Behavior & Analytics* (accepted).

[11] Alzaatreh, A., Lee, C., and Famoye, F. (2013). A new method for generating families of continuous distributions. *Metron*, **71**(1): 63-79.

[12] Chowdhury, M. H. (2018). *Quantile Regression Approach for Analyzing Gene Expression Data.* Unpublished Doctoral dissertation. The University of Regina, Canada.

[13] Bowley, A. L. (1920). Elements of statistics (4th d.). *New York: Charles Scribner's Sons*.

[14] Moors, J. J. A. (1988). A quantile alternative for kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician)*, **37**(1): 25-32.

[15] Swain, J. J., Venkatraman, S., & Wilson, J. R. (1988). Least-squares estimation of distribution functions in Johnson's translation system. *Journal of Statistical Computation and Simulation*, *29*(4): 271-297.

[16] Lawless, J. F., (1982), Statistical Models and Methods for Lifetime Data. John Wiley & Sons, New York, USA.

[17] Ieren, T. G., Oyamakin, S. O., Yahaya, A., Chukwu, A. U., Umar, A. A. and Kuje, S., (2018b):On Making an Informed Choice between Two Lomax-Based Continuous Probability Distributions Using Lifetime Data. *Asian Journal of Probability & Statistics*, **2**(2): 1-11.

[18] Telford, R. D., & Cunningham, R. B. (1991). Sex, sport, and body-size dependency of hematology in highly trained athletes. *Medicine and science in sports and exercise*, **23**(7): 788-794.

Appendix

Table A1.1: Dataset I: Weights of 311 pregnant women at War Memorial Hospital-Navrongo.

74.0	47.0	65.0	71.0	78.2	82.0	96.0	87.0	89.5	53.7	71.0	89.0
74.0	86.0	64.0	77.0	95.0	88.0	82.0	76.0	78.0	56.0	70.0	59.0
91.0	64.0	72.0	99.0	85.0	85.0	83.0	82.0	88.0	62.0	65.0	60.0
88.5	89.0	94.0	88.0	81.0	93.0	75.0	85.0	68.0	47.0	60.0	85.0
87.0	89.0	77.0	76.0	88.0	82.0	75.0	74.0	80.0	62.0	46.0	54.0
97.0	76.0	73.0	87.9	81.0	68.0	79.0	79.3	47.0	57.0	58.0	56.0
85.0	88.0	76.0	90.4	95.0	80.0	80.0	75.0	80.0	71.0	46.0	67.0
69.0	82.3	83.2	82.0	95.0	86.3	76.2	89.0	56.0	69.0	60.0	75.0
90.0	95.0	87.6	79.0	76.0	75.0	85.2	82.0	69.0	62.0	53.7	61.0
86.7	68.0	88.0	74.0	85.0	96.0	99.0	82.0	68.0	47.0	70.0	53.7
88.0	83.0	70.0	91.0	86.0	99.0	80.0	48.0	48.0	49.5	57.0	57.0
84.6	80.0	58.0	79.0	76.2	84.0	80.0	65.0	46.0	58.0	56.0	60.0
100.5	74.0	84.0	90.0	85.0	86.0	91.0	70.0	69.0	80.0	80.0	61.0
89.0	84.0	86.0	98.0	101	85.2	83.0	58.0	57.0	47.0	67.0	67.0
84.0	96.0	73.0	77.0	76.3	84.5	94.0	60.0	78.0	48.0	88.0	88.0
99.0	85.5	69.0	86.0	85.0	88.5	69.1	89.5	54.0	57.0	71.0	54.0
81.0	65.2	94.0	76.0	83.0	79.0	67.5	89.0	80.0	48.0	78.0	67.0
80.0	81.2	76.3	70.1	89.0	92.0	79.0	59.0	89.0	65.0	75.0	80.0
85.0	99.0	71.3	85.0	90.0	82.0	87.0	58.0	53.7	80.0	46.0	58.0
86.2	85.0	76.9	80.1	86.0	81.0	79.0	70.0	69.0	56.0	60.0	71.0
92.0	61.0	77.0	66.0	53.7	66.0	47.0	67.0	66.0	65.0	71.0	60.0
66.0	58.0	79.0	65.0	68.0	66.0	58.0	70.0	56.0	65.0	60.0	70.0
68.0	58.0	56.0	69.0	53.7	69.0	62.0	46.0	66.0	65.0	89.0	66.0
70.0	56.0	65.0	60.0	70.0	88.0	48.0	47.0	65.0	89.5	65.0	48.0
66.0	66.0	88.0	65.0	58.0	59.0	88.0	68.0	71.0	75.0	67.0	60.0
61.0	65.0	58.0	68.0	61.0	75.0	57.0	67.0	65.0	66.0	66.0	