A WEIGHTED NADARAJAH AND HAGHIGHI DISTRIBUTION

Muhammad Nauman Khan[1*], Abdus Saboor[1], Gauss M. Cordeiro[2], Mamoona Nazir[1], Rodrigo R. Pescim[3]

We propose and study a two-parameter weighted Nadarajah and Haghighi distribution. The new distribution can be viewed as an alternative model to some of the classical two-parameter distributions such as the Weibull, gamma, exponentiated half-logistic and exponentiated exponential distributions. We explore some of its mathematical properties. The maximum likelihood estimation method is adopted to estimate the model parameters. A Monte Carlo simulation study is performed to assess the adequacy of the estimates. We compare the fits of the proposed distribution and other competitive models to three real data sets.

Keywords: Maximum likelihood estimation, Monte Carlo simulation, Weibull distribution, Weighted distribution.

MSC2010: 60E05, 62E15, 62F10.

1. Introduction

For a non-negative random variable X with probability density function (pdf) $f(x)$ and a non-negative weight function $w(x)$ with finite non-zero expectation, the pdf of the weighted random variable X_w pioneered by Patil and Rao [1], say $f_w(x)$, is given by

$$f_w(x) = \frac{w(x)f(x)}{E[w(X)]}, \tag{1}$$

where $E[w(X)]$ is the expected value of $w(X)$ and represents a normalizing constant. For more details on the weighted distributions we refer the interested reader to [2].

Nadarajah and Haghighi [3] defined the Nadarajah-Haghighi (NH) density given by

$$g(x) = \alpha\,\lambda(1+\lambda x)^{\alpha-1}\mathrm{e}^{1-(1+\lambda\,x)^{\alpha}},\ x>0,\ \alpha,\ \lambda>0. \tag{2}$$

[1]Department of Mathematics, Kohat University of Science & Technology, Kohat, Pakistan, e-mail: *zaybasdf@gmail.com

[2]Departamento de Estatística, Universidade Federal de Pernambuco, 50740-540, Recife, PE, Brazil

[3]Departamento de Estatística, Universidade Estadual de Londrina, 86057-970, Londrina, PR, Brazil

The survival function corresponding to (2) is given by

$$\bar{G}(x) = \mathrm{e}^{1-(1+\lambda\, x)^{\alpha}}.$$

The NH distribution is an alternate model to the gamma, Weibull and exponentiated exponential distributions and has been extended successfully in order to provide more accurate statistical models and inferential procedures. For example, the exponentiated NH distribution pioneered by [4] and Marshall-Olkin NH distribution introduced by [5].

Due to the importance of the NH distribution and the concept of weighted distributions, we define a weighted version of the NH distribution called the *weighted Nadarajah and Haghighi* (WNH) distribution with a particular weight function. The WNH distribution can be viewed as an alternate model to some of the classical two-parameter distributions like Weibull, gamma, etc. and can have wider applications in reliability, survival analysis, forestry and ecological areas.

Definition: A random variable X is said to follow the WNH distribution, if its pdf (for $x > 0$) has the form

$$f(x) = \frac{2\alpha\lambda(1+\lambda x)^{\alpha-1}\mathrm{e}^{1-(1+\lambda x)^{\alpha}}}{\left[1+\mathrm{e}^{1-(1+\lambda x)^{\alpha}}\right]^{2}}, \quad \alpha, \lambda > 0, \tag{3}$$

where α and λ are shape and scale parameters, respectively.

Then, the cumulative distribution function (cdf) of the WNH model is given by

$$F(x) = \frac{1-\mathrm{e}^{1-(1+\lambda x)^{\alpha}}}{1+\mathrm{e}^{1-(1+\lambda x)^{\alpha}}}. \tag{4}$$

Henceforth, we denote a random variable X having the density function (3) by $X \sim \mathrm{WNH}(\alpha, \lambda)$. The hazard rate function (hrf) of X is given by

$$h(x) = \frac{\alpha\lambda(1+\lambda x)^{\alpha-1}}{1+\mathrm{e}^{1-(1+\lambda x)^{\alpha}}}. \tag{5}$$

We motivate the WNH distribution by the following facts:

- The WNH density (3) is obtained by taking the NH density (2) for the baseline in (1), the weight function as $w(x) = \left[1+\mathrm{e}^{1-(1+\lambda x)^{\alpha}}\right]^{-2}$ with expectation $1/2$.
- If X has the WNH density (3) and using the transformation $Y = (1+\lambda\, x)^{\alpha} - 1$, the distribution of Y follows the standard half-logistic distribution with cdf $F(y) = \frac{1-\mathrm{e}^{-y}}{1+\mathrm{e}^{-y}}$.
- If X has the WNH density (3) and using the transformation $Y = X + \lambda^{-1}$, the distribution of Y follows the zero truncated weighted Weibull distribution with cdf $F(y) = \frac{1-\mathrm{e}^{1-(\lambda\, y)^{\alpha}}}{1+\mathrm{e}^{1-(\lambda\, y)^{\alpha}}}$.

The quantile functions (qfs) are in widespread use in general statistics to obtain mathematical properties of a distribution and often find representations in terms of lookup tables for key percentiles. For simulating the WNH model, let $p \sim U(0,1)$. Then, for $p \in (0,1)$, the qf of X becomes

$$Q(p) = \frac{1}{\lambda}\left[\left\{1 - \log\left(\frac{1-p}{1+p}\right)\right\}^{\frac{1}{\alpha}} - 1\right]. \tag{6}$$

So, if U is a uniform distribution in $(0,1)$, then $X = Q(U)$ has density (3). In particular, the median of X is $M = \lambda^{-1}\left[\{1+\log(3)\}^{1/\alpha} - 1\right]$.

The plots in Figure 1 reveal that the pdf of X is positively skewed.

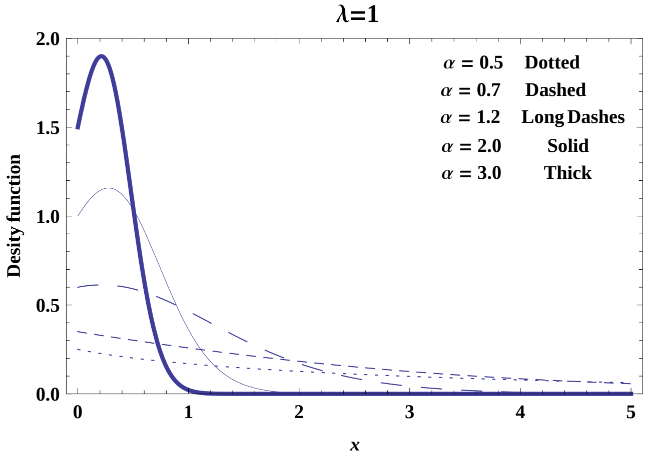

FIGURE 1. Plots of the WNH density

The rest of this paper is organized as follows. In Section 2, we study some statistical functions of the WNH distribution. In Section 3, we discuss the maximum likelihood estimation procedure. A simulation study is performed in Section 4. In Section 5, we present three applications to real data sets. Finally, Section 6 offers some conclusions.

2. Statistical properties

In this section, we derive computable representations for some statistical functions associated with the WNH distribution, whose pdf admits a simple linear representation.

2.1. Linear representation

Let $G(x) = 1 - e^{1-(1+\lambda x)^{\alpha}}$ be the NH cdf. Then, we can write the WNH cdf given by (4) as

$$F(x) = G(x)/[2 - G(x)].$$

By expanding the denominator in power series, we have

$$F(x) = \sum_{i=1}^{\infty} a_i\, G(x)^i = \sum_{i=1}^{\infty} a_i\, H_i(x),$$

where $a_i = 2^{-i}$ and $H_i(x) = G(x)^i$ (for $i \geq 1$) is the cdf of the exponentiated NH (ENH) distribution with power parameter "i". By differentiating the last equation, we can write the pdf of X as

$$f(x) = \sum_{i=1}^{\infty} a_i\, h_i(x), \tag{7}$$

where $h_i(x) = i\, G(x)^{i-1} g(x)$ (for $i \geq 1$) is the ENH density with power parameter i.

2.2. Moments

The rth ordinary moment of the WNH random variable, say $\mu_r' = E(X^r)$, follows from (7) as

$$\mu_r' = \sum_{i=1}^{\infty} a_i \int_0^{\infty} x^r\, h_i(x)\, \mathrm{d}x.$$

Using the binomial theorem, we obtain (for $x > 0$)

$$\left[1 - \mathrm{e}^{1-(1+\lambda\, x)^{\alpha}}\right]^{i-1} = \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} \mathrm{e}^{j[1-(1+\lambda\, x)]}.$$

By noting that $0 < \mathrm{e}^{1-(1+\lambda\, x)^{\alpha}} < 1$, we can write

$$\mu_r' = \alpha\, \lambda \sum_{i=1}^{\infty} \sum_{j=0}^{i-1} (-1)^j\ \mathrm{e}^{j+1}\, a_i \binom{i-1}{j} A_{r,j}\ \ , \tag{8}$$

where

$$A_{r,j} = \int_0^{\infty} x^r\ (1+\lambda\, x)^{\alpha-1}\, \mathrm{e}^{-(j+1)(1+\lambda\, x)^{\alpha}} \mathrm{d}x.$$

For $r \in \Re$, and setting $u = (j+1)(1+\lambda x)^{\alpha}$, we have

$$x = \left\{\lambda^{-1}\left[\left(\frac{u}{j+1}\right)^{1/\alpha} - 1\right]\right\}$$

and then, after some algebra, we can write

$$A_{r,j} = \frac{\lambda^{-r-1}}{\alpha\,(j+1)} \int_{j+1}^{\infty} \left[\left(\frac{u}{j+1}\right)^{1/\alpha} - 1\right]^r \mathrm{e}^{-u}\, \mathrm{d}u. \tag{9}$$

The most general case of the binomial theorem is the identity

$$(x+a)^{\nu} = \sum_{k=0}^{\infty} \binom{\nu}{k} x^k\, a^{\nu-k}, \tag{10}$$

where $\binom{\nu}{k}$ is a binomial coefficient and ν is a real number. This power series converges when $\nu \geq 0$ is an integer or $|x/a| < 1$. By using (10) in equation (9), since $\left|[u/(j+1)]^{1/\alpha}\right| < 1$, it follows by interchanging the sum and the integral

$$A_{r,j} = \frac{\lambda^{-r-1}}{\alpha\,(j+1)} \sum_{k=0}^{\infty}(-1)^k \binom{r}{k} \int_{j+1}^{\infty} \left(\frac{u}{j+1}\right)^{(r-k)/\alpha} \mathrm{e}^{-u}\,\mathrm{d}u.$$

Then, we can write from (8)

$$\mu_r' = \frac{1}{\lambda^r} \sum_{i=1}^{\infty}\sum_{j=0}^{i-1}\sum_{k=0}^{\infty} \frac{(-1)^{j+k}\,\mathrm{e}^{j+1}\,a_i}{(j+1)^{\frac{r+\alpha-k}{\alpha}}} \binom{i-1}{j} \binom{r}{k} \Gamma\left(\frac{r+\alpha-k}{\alpha}, j+1\right). \quad (11)$$

Equation (11) is an extension of expression (9) given by [4].

Next, we define a lemma.

Lemma 2.1. *The rth incomplete moment of X, say $J(x; r, \theta) = \int_0^x y^r f(y)\,\mathrm{d}y$, is given by*

$$J(x;r,\theta) = \alpha\,\lambda \sum_{i=1}^{\infty}\sum_{j=0}^{i-1} (-1)^j\ \mathrm{e}^{j+1}\,a_i \binom{i-1}{j}$$
$$\times \int_0^x y^r\ (1+\lambda\,y)^{\alpha-1}\,\mathrm{e}^{-(j+1)(1+\lambda\,y)^{\alpha}}\,\mathrm{d}y, \qquad r = 1, 2, \ldots,$$

where $\theta = (\lambda,\, \alpha)$. Then, we have

$$J(x;r,\theta) = \frac{1}{\lambda^r} \sum_{i=1}^{\infty}\sum_{j=0}^{i-1}\sum_{k=0}^{\infty} \frac{(-1)^{j+k}\,a_i\,\mathrm{e}^{j+1}}{(j+1)^{\frac{r+\alpha-k}{\alpha}}} \binom{i-1}{j} \binom{r}{k}$$
$$\times \left[\Gamma\left(\frac{r+\alpha-k}{\alpha}, j+1\right) - \Gamma\left(\frac{r+\alpha-k}{\alpha}, (j+1)\,(1+\lambda\,y)^{\alpha}\right)\right].$$

Proof. The proof follows easily by changing variables in the integration. □

3. Maximum likelihood estimation

Given the observed values $x_1, \cdots, x_n$ from the WNH distribution, the MLEs of the unknown parameters in $\theta = (\alpha, \lambda)^T$ can be determined by maximizing directly the log-likelihood function given by

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

The above log-likelihood can be maximized numerically by using the `R` (`optim` function), `SAS` (`PROC NLMIXED`), `Ox` program (sub-routine `MaxBFGS`), *Nmaximize* command in **Mathematica**, among others.

The adequacy of the fitted models to a data set can be verified using some goodness-of-fit statistics. The MLEs can be evaluated using *Nmaximize* command in **Mathematica** as well as the goodness-of-fit statistics including the Anderson-Darling (A^*) and Cramér-von Mises (W^*) statistics. These statistics can determine how closely a specific density fits the histogram of a given data set. In general, the distribution with smallest values for these statistics yields a better fit.

4. **Simulation study**

We perform a Monte Carlo simulation study to assess the finite sample behavior of the MLEs of λ and α. The results are obtained from 2,000 Monte Carlo replications and the simulations are carried out using the `R` software. In each replication, a random sample of size n is drawn from the WNH(λ, α) distribution and the parameters are estimated by the maximum likelihood method. The WNH random variable X is generated using the inversion method. We consider two setups with the following values for the parameters of the model: $\lambda = 1.5$, and $\alpha = 2.0$. The mean estimates of the model parameters and their root mean squared errors (RMSEs) for the sample sizes $n = 50$, 100, 200 and 500 are given in Table 1. We note that the RMSEs of the estimates of λ and α decrease toward zero when the sample size n increases, which reveals the consistency of the MLEs. These estimates have small biases.

Table 1. Mean estimates and RMSEs of the MLEs of λ and α.

n	Parameter	Mean	RMSE
	α	1.5246	1.2224
50	λ	4.0768	4.9927
	α	1.4997	0.8067
100	λ	2.8499	2.7316
	α	1.4994	0.5628
200	λ	2.2810	1.0427
	α	1.4993	0.3515
500	λ	2.0886	0.4145

5. **Applications**

In this section, we fit some well-known distributions and the WNH distribution to three data sets from different fields of scientific investigation, and then select the best fitted model among them.

5.1. **Data fitting**

In the applications, the WNH model is compared with gamma, Weibull, exponentiated exponential(EE) [6] and exponentiated half-logistic(EHL) [7] distributions.

TABLE 2. Descriptive Statistics

Data Set	Mean	Median	Standard Deviation	Kurtosis	Skewness
D1	59.6	22	71.8848	1.6936	4.96666
D2	92.0744	54	107.916	2.13921	8.02311
D3	12.1594	9.6	11.631	1.57966	5.58786

The applications are to the "Air conditioning system 1", "Air conditioning system 2" and "Telephone call" data sets. The data are described below.

i) Air conditioning system 1 data set (D1)

The first data set containing 30 observations represents the numbers of failures for the air conditioning system of an air plane as reported in [8].

ii) Air conditioning system 2 data set (D2)

The second data set containing 188 observations represents the number of successive failures for the air conditioning system of each member in a fleet of 13 Boeing 720 jet airplanes as reported in [9].

iii) The telephone call data set (D3)

The third data set represents times between 35 consecutive telephone calls (in seconds) as reported in [10].

Table 2 summarizes some descriptive statistics for the three data sets. It can be noted from Table 2 that the three data sets are highly right skewed. The MLEs for all fitted models and the goodness-of-fit statistics A^* and W^* are reported in Table 3. Based on the figures in these tables, the WNH model provides the best fit to the three data sets compared to the gamma, Weibull, EE and EHL distributions. The estimates for the proposed model are very precise.

6. **Conclusions**

In this paper, we propose a new two-parameter distribution, namely the *weighted Nadarajah and Haghighi* (WNH) distribution. We study some of its statistical functions. The estimation of the model parameters is performed by the maximum likelihood method. We analyze three real data sets and the WNH distribution provides an adequate fit to each data set. In conclusion, the WNH distribution provides a rather flexible mechanism for fitting a wide spectrum of positive real data sets with shape property of being right skewed.

Acknowledgments

The authors are grateful to the editor and two anonymous reviewers for their valuable feedback which have improved the presentation of the paper. The authors wish to express their gratitude to Prof. Hassan S. Bakouch (Tanta University, Egypt) for his valuable suggestions.

TABLE 3. MLEs of the parameters and goodness-of-fit statistics

	For D1			
Distributions	Estimates		A^*	W^*
WNH(α, λ)	0.479957	0.112027	**0.46092**	**0.08318**
Gamma(k, θ)	0.811911	73.4070	0.65338	0.117212
Weibull(λ, k)	54.613446	0.853587	0.552232	0.097430
EE(α, λ)	0.809287	0.014542	0.672897	0.121041
EHL(σ, λ)	58.676392	0.650296	0.883707	0.156466
	For D2			
Distributions	Estimates		A^*	W^*
WNH(α, λ)	0.541764	0.052968	**0.561922**	**0.067041**
Gamma(k, θ)	0.904733	101.769	1.31096	0.231627
Weibull(λ, k)	87.75651	0.910896	0.94684	0.144364
EE(α, λ)	0.910031	0.010199	1.39972	0.252818
EHL(σ, λ)	83.26937	0.724783	2.78309	0.52523
	For D3			
Distributions	Estimates		A^*	W^*
WNH(α, λ)	0.756136	0.189029	**0.197489**	**0.030507**
Gamma(k, θ)	1.15775	10.5025	0.210751	0.035416
Weibull(λ, k)	12.53077	1.078352	0.20206	0.032676
EE(α, λ)	1.172184	0.090953	0.211606	0.035771
EHL(σ, λ)	9.337973	0.917498	0.296068	0.048410

REFERENCES

[1] *G. P. Patil and C. R. Rao*, Weighted distributions and size-biased sampling with applications to wild life populations and human families, Biometrics, **34** (1978), 179–189.

[2] *G. P. Patil*, Weighted distributions, John Wiley & Sons, (2005).

[3] *S. Nadarajah and F. Haghighi*, An extension of the exponential distribution, Statistics, **45** (2011), 543–558.

[4] *A. J. Lemonte*, A new exponential-type distribution with constant, decreasing, increasing, upside-down bathtub and bathtub-shaped failure rate function, Computational Statistics and Data Analysis, **62** (2013), 149–170.

[5] *A. J. Lemonte, G. M. Cordeiro and G. MorenoArenas*, A new useful three-parameter extension of the exponential distribution, Statistics, **50** (2016), 312–337.

[6] *I. B. A. Moniem*, The Exponentiated Exponential distribution, International Journal of Mathematical Analysis and Applications, **2** (2015), 68–73.

[7] *G. M. Cordeiro, M. Alizadeh and E. M. M. Ortega*, The Exponentiated Half-Logistic Family of Distributions: Properties and Applications, Journal of Probability and Statistics, **2014**, Article ID 864396, pages 21.

[8] *W. Zucchini and H. Linhart*, Model Selection, Wiley, New York, (1986).

[9] *F. Proschan*, Theoritical Explanation of Observed Decreasing Failure Rate, Technometrics, **5** (1963), 375–383.

[10] *A. S. Hadi, E. Castillo, J. M. Sarabia and N. Balakrishnan*, Extreme Value and Related Models with Applications in Engineering and Science, Newyork: John Wiley and Sons, (2005).