Kuwait J.Sci., Vol.49, No.(2),April.2022,pp(1-14)

A flexible probability model for reliability data analysis: The extended half-normal distribution with further results

Eman Khorsheed[1,*], Hugo S. Salinas[2], Hassan S. Bakouch[3]
[1]*Dept. of Mathematics*, College of Science,
University of Bahrain, P.O.Box 32038, Sakhir, Kingdom of Bahrain
[2]*Dept. de Matemática, Facultad de Ingeniería, Universidad de Atacama,*
Copiapó, Chile
[3]*Dept. of Mathematics, Faculty of Science,*
Tanta University, Tanta, Egypt
**Corresponding author: ekhorsheed@uob.edu.bh*

Abstract

Recent years have shown growth in the potential applications of extensions of skew and half-normal distributions. In this paper, we provide and study, in detail, an extended class of such distributions. The presented distribution has a scale parameter and two shape parameters. The motivations behind the development of this distribution are: 1- the probability distribution function has skewed, unimodal and bathtub shapes with different styles and the hazard rate function is increasing with various shapes, which makes this distribution flexible enough to analyze reliability data; 2- it can be expressed as a mixture of the Generalized Half-Normal distribution and the new weighted Generalized Half-Normal distribution. A number of statistical results are derived. Monte Carlo simulation analysis is carried out with Maximum Likelihood Estimation to assess the performance of this technique for a variety of the distribution parameter values. The power of the distribution is demonstrated with real applications using two reliability datasets from the industry. The results reveal that the proposed distribution outperformed the Generalized Lindley, Exponentiated Exponential, Gamma, Alpha-Skew-Normal, the Half-Alpha-Skew-Normal, and the Modified Generalized Half-Normal distributions.

Keywords: Coefficients of asymmetry and kurtosis; half-normal; Maximum Likelihood Estimation; reliability; skew-normal

1. Introduction

In the era of Big data, the gathered observations exhibit different characteristics that reflect the complexity of our modern life. Some of these characteristics are obvious such as volume, variety (structured/unstructured), velocity and value as described in (Katal *et al.*, 2013). On the contrary, other features require analyzing and mining the information hidden behind the collected data, which may support decision-making in many industries. For example, modeling and analyzing reliability data may be utilized to determine the boundaries of safe use of a material, component, or a system. Therefore, estimating the threshold stress level or the probability distribution of failure strengths/times is crucial. Many related studies exist in the literature, for example, (Kishorilal and Mukhopadhyay, 2018; Ansell and Phillips, 1990). In real life applications, various statistical distributions have been utilized to analyze datasets under investigation, among those we mention the well-known Half-Normal (HN) distribution with zero mean. This truncated distribution has been used for risk analysis in several fields such as engineering (Krenek *et al.*, 2016), finance (Aharony *et al.*, 1980; Badia *et al.*, 2020), and medicine (Krause *et al.*, 2018). Despite being widely used, the HN distribution has a limited control on the shape of the produced models because it has a scale parameter only. Several authors proposed more general models than the HN distribution. Cooray and Aranda (2008) added a shape parameter to the HN distribution and developed the Generalized Half-Normal (GHN). This distribution has been extensively modified and recent generalizations have been proposed following the same methodology of the developers. For example, the Beta Generalized Half-Normal distribution was produced by (Pescim *et al.*, 2010). Azzalini (1985) proposed the Skew-Normal distribution. Elal-Olivero (2010) generated the Alpha-Skew-Normal (ASN) distribution, which has one asymmetric parameter. On the other side, Olmos and Venegas (2018) studied the Modified Generalized Half-Normal (MGHN) variable developed with the modulus of a ASN random variable, considering a scale parameter and a shape parameter. The Half-Alpha-Skew-Normal (HASN) distribution is another generalization of the HN distribution and was introduced in (Olmos and Venegas, 2018).

Despite the available distributions for modeling reliability data, this new era not only demands non-linear distributions to unveil the hidden relations but also flexible models that can capture most of the variations within the data. Here, we present a nonlinear distribution that may be suitable for modeling various reliability datasets, namely the Extended Half-Normal (EHN). The EHN distribution is formulated as an extension of the GHN, ASN, HASN, and the MGHN distributions and was initially introduced in (Khorsheed *et al.*, 2020). Unlike several existing one and two parameter-families, the EHN is a three parameter-family that incorporate two shape parameters, which makes it more flexible for distribution shape control. This work contains further EHN statistical properties and results, a simulation and sensitivity analysis study, and two practical applications using fiber stress and strength data.

The article will proceed as follows. In section 2 we explain the development of the EHN distribution, hazard function, some statistical properties, moments, skewness and kurtosis coefficients. In section 3 we make the inference by implementing the maximum likelihood estimation approach and present a simulation study. We demonstrate the relevance of the (EHN) distribution for two reliability data applications using goodness-of-fit statistics in section 4. In section 5 we list the limitations of the EHN distribution and finally, in section 6 we provide a discussion.

2. The Extended Half-Normal distribution and further statistical properties

Recall that if $X \sim EHN(\lambda, \beta, \sigma)$, then the pdf of X is

$$f_X(x;\lambda,\beta,\sigma) = \frac{2\lambda(2\sigma^{2\lambda}+\beta x^{2\lambda})}{\sigma^{2\lambda+1}(2+\beta)}\left(\frac{x}{\sigma}\right)^{\lambda-1}\phi\left(\left(\frac{x}{\sigma}\right)^{\lambda}\right), x>0 \tag{1}$$

and the corresponding cdf is

$$F_X(x;\lambda,\beta,\sigma) = 2\Phi\left(\left(\frac{x}{\sigma}\right)^{\lambda}\right) - \frac{2\beta\, x^{\lambda}}{\sigma^{\lambda}(2+\beta)}\phi\left(\left(\frac{x}{\sigma}\right)^{\lambda}\right) - 1, \tag{2}$$

where $\sigma, \lambda > 0$ and $\beta \geq 0$.

The classical HN, GHN, HASN and MGHN distributions can be obtained as submodels of the EHN distribution. More specifically, if $X \sim EHN(\lambda, \beta, \sigma)$, then

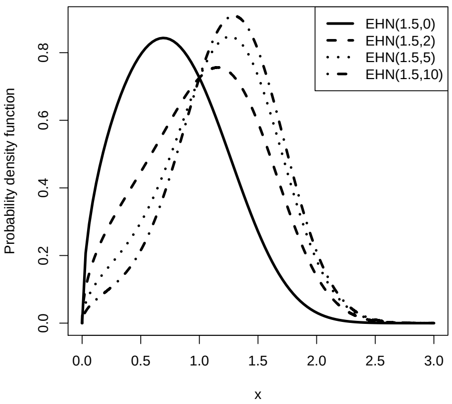
1. For $\beta = 0$ we obtain the Generalized Half-Normal distribution (GHN) with the pdf

$$f(x;\lambda,\sigma) = \frac{2\lambda}{\sigma^{\lambda}}x^{\lambda-1}\phi\left(\left(\frac{x}{\sigma}\right)^{\lambda}\right) \tag{3}$$

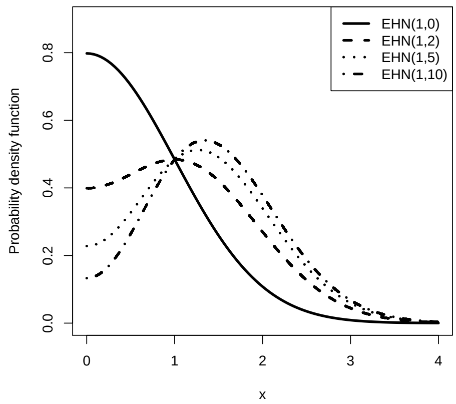
2. For $\beta = 25$ we obtain

$$f(x;\sigma,\lambda) = \frac{2\lambda x^{\lambda-1}(2\sigma^{2\lambda}+25x^{2\lambda})}{27\sigma^{3\lambda}}\phi\left(\left(\frac{x}{\sigma}\right)^{\lambda}\right), \tag{4}$$

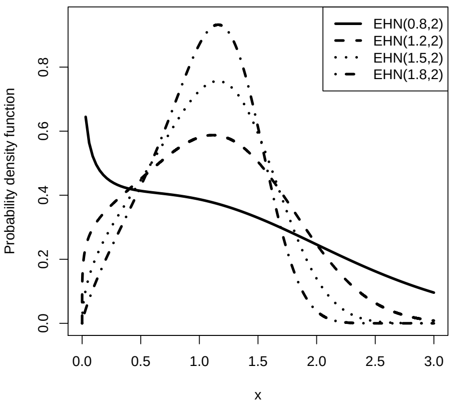
which is the pdf of the MGHN distribution.



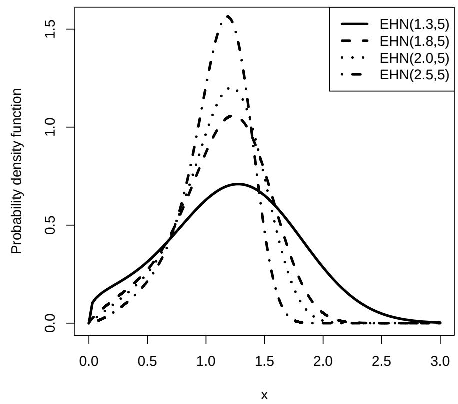
$(a) X \sim EHN(\lambda = 1.5, \beta, \sigma = 1)$



$(b) X \sim EHN(\lambda = 1, \beta, \sigma = 1)$



$(c) X \sim EHN(\lambda, \beta = 2, \sigma = 1)$



$(d) X \sim EHN(\lambda, \beta = 5, \sigma = 1)$

Fig. 1. EHN pdf for different values of λ and β.

3. For $\lambda = 1$ and $\sigma = 1$ we obtain the HASN distribution with pdf

$$f(x;\beta) = \frac{2(2+\beta x^2)}{(2+\beta)}\phi(x) \tag{5}$$

4. For $\beta = 0$ and $\lambda = 1$ we obtain

$$f(x;\sigma) = \frac{2}{\sigma}\phi\left(\frac{x}{\sigma}\right), \tag{6}$$

that is the Half-Normal distribution of Johnson and Balakrishnan (1995).

Figure 1 displays the Extended Half-Normal pdf for a selected set of values of the parameters λ and β with $\sigma = 1$. Owing to the two shape parameters of the proposed distribution, its pdf has skewed, unimodal and bathtub shapes with various styles, as can be seen in Figure 1. The wide range of the EHN shapes and skewness levels demonstrated in this figure provides a strong evidence of the distribution high degree of flexibility, which is an important characteristics required in this age. Moreover, the hazard rate (hr) function of the EHN is increasing with different shapes which adapts the nature of reliability data hazard rates. Curves of the probability density function and hazard rate of the developed EHN model motivate reliability data analysis will be illustrated later using two reliability datasets. A final motivation is that the EHN model can be expressed as a mixture of the GHN and the new weighted Generalized Half-Normal distribution as follows:

$$f_X(x;\lambda,\beta,\sigma) = pf(x;\lambda,\sigma) + (1-p)g(x;\lambda,\sigma),$$

where $p = \frac{2}{2+\beta}$, $f(x;\lambda,\sigma)$ is defined in equation (3) and $g(x;\lambda,\sigma) = \frac{2\lambda}{\sigma^{3\lambda}}x^{3\lambda-1}\phi((\frac{x}{\sigma})^{\lambda}) = \frac{x^{2\lambda}}{\sigma^{2\lambda}}f(x;\lambda,\sigma)$ is a weighted version of the Generalized Half-Normal distribution with the weight function $w(x) = x^{2\lambda}$ and a normalizing constant $\sigma^{2\lambda} = \mathbf{E}(w(X))$.

In many practical situations, variations across items within a population of interest may exist. Therefore, for an accurate data analysis a mixture model that can take into account the underlying statistical heterogeneity is recommended. Below are some main further results.

Theorem 2.1 *If $Z \sim ASN(\alpha)$, then $Y = |Z| \sim HASN(\beta)$ where $\beta = \alpha^2$.*

Proof. By using the cdf of Z displayed in (Elal-Olivero, 2010), we have

P(Y $\leq y) = P(|Z| \leq y)$
= P(Z$\leq y) - P(Z \leq -y)$
= $\Phi(y) + \alpha\left(\dfrac{2-\alpha y}{2+\alpha^2}\right)\phi(y) - \Phi(-y) - \alpha\left(\dfrac{2+\alpha y}{2+\alpha^2}\right)\phi(-y)$
= $2\Phi(y) - \frac{2\alpha^2 y}{2+\alpha^2}\phi(y) - 1$
= $2\Phi(y) - \frac{2\beta y}{2+\beta}\phi(y) - 1.$

Therefore, Y has the $HASN(\beta)$ distribution.

Theorem 2.2 *If $Y \sim HASN(\beta)$, then $X = \sigma Y^{1/\lambda} \sim EHN(\lambda,\beta,\sigma)$.*

Proof. As given in (Khorsheed *et al.*, 2020).

Corollary 2.3 *The hazard rate function for the random variable $X \sim EHN(\lambda,\beta,\sigma)$ is*

$$h_X(x;\lambda,\beta,\sigma) = \frac{\lambda x^{\lambda-1}(2\sigma^{2\lambda}+\beta x^{2\lambda})\phi\left(\left(\frac{x}{\sigma}\right)^{\lambda}\right)}{\sigma^{2\lambda}\left[\sigma^{\lambda}(2+\beta)\Phi\left(-\left(\frac{x}{\sigma}\right)^{\lambda}\right)+\beta x^{\lambda}\phi\left(\left(\frac{x}{\sigma}\right)^{\lambda}\right)\right]}. \tag{7}$$

Figure 2 demonstrates the hazard rate function curves for a range of values of λ and β. All these curves are increasing with different styles which suits the realistic behavior of the hazard rate for real life data.

Proposition 2.4 *For a random variable $X \sim EHN(\lambda,\beta,\sigma)$ and integers $r = 1, 2, ...$, the corresponding r-th moments are given by*

$$\mu_r = \frac{2^{r/2\lambda}\sigma^r(\beta r + \lambda(2+\beta))}{\sqrt{\pi}\,\lambda(2+\beta)}\Gamma\left(\frac{r+\lambda}{2\lambda}\right), \tag{8}$$

where $\Gamma(x)$ corresponds to the gamma function.

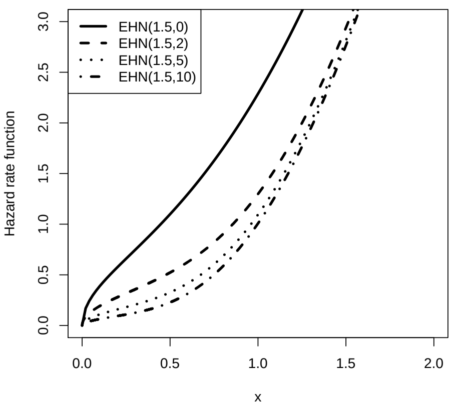
Proof. By Theorem 2.2 and the pdf of the EHN given in equation (1),
E(X$^r) = \sigma^r\,\mathbf{E}(Y^{r/\lambda})$

$= \sigma^r\int_0^{\infty} y^{r/\lambda}\frac{2(2+\beta y^2)}{\beta}\phi(y)dy$

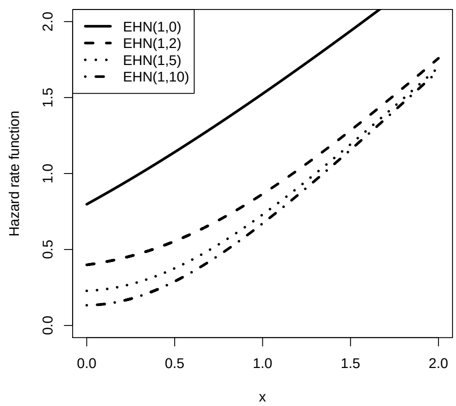
$= \frac{2\sigma^r}{2+\beta}\left[\int_0^{\infty} 2y^{r/\lambda}\phi(y)dy + \frac{\beta}{2}\int_0^{\infty} 2y^{r/\lambda+2}\phi(y)dy\right]$

$= \frac{\sigma^r}{2+\beta}\left[2\mathbf{E}(U^{r/\lambda}) + \beta\mathbf{E}(U^{r/\lambda+2})\right],$

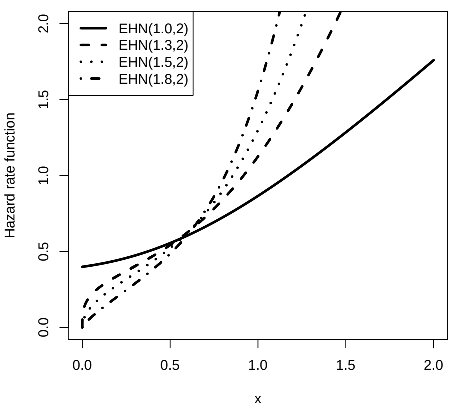
where U is a HN random variable such that $\mathbf{E}(U^{r/\lambda}) = 2^{r/2\lambda}\Gamma(\frac{r+\lambda}{2\lambda})/\sqrt{\pi}$. What follows is achieved with a little algebraic manipulation and some simplifications.



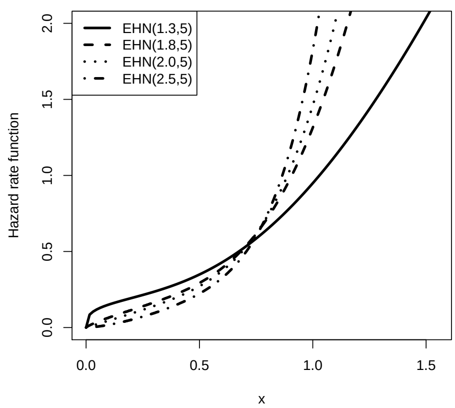
$(a) X \sim EHN(\lambda = 1.5, \beta, \sigma = 1)$



$(b) X \sim EHN(\lambda = 1, \beta, \sigma = 1)$



$(c) X \sim EHN(\lambda, \beta = 2, \sigma = 1)$



$(d) X \sim EHN(\lambda, \beta = 5, \sigma = 1)$

Fig. 2. EHN hazard rate function with a range of values of λ and β.

Corollary 2.5 *For $r = 1, 2, 3, 4$ and $X \sim EHN(\lambda, \beta, \sigma)$, the mean, variance, skewness (γ_X) and kurtosis (κ_X) coefficients are, respectively, given as* $\mu = \sigma\rho_1$, $Var(X) = \sigma^2\left(\rho_2 - \rho_1^2\right)$, $\gamma_X = \frac{\rho_3 - 3\rho_1\rho_2 + 2\rho_1^3}{(\rho_2 - \rho_1^2)^{3/2}}$ *and* $\kappa_X = \frac{\rho_4 - 4\rho_1\rho_3 + 6\rho_1^2\rho_2 - 3\rho_1^4}{(\rho_2 - \rho_1^2)^2}$, *where* $\rho_r = \rho_r(\lambda, \beta) = \frac{2^{r/2\lambda}(\beta r + \lambda(2+\beta))}{\sqrt{\pi}\,\lambda(2+\beta)}\Gamma\left(\frac{r+\lambda}{2\lambda}\right)$.

Proof. These formulas are derived by substituting into the definitions of the variance, skewness and kurtosis coefficients directly using the results of proposition 2.4 with $r = 1, 2, 3$ and 4.

3. Estimation with Inference and a Simulation Study

In this section, we discuss the Maximum Likelihood Estimation for the EHN distribution parameters σ, λ and β and provide a simulation and sensitivity analysis study to gain insight on the obtained estimators. To find the MLE for each parameter, we used the R software (R core Team 2014) with the machine learning tool of (Byrd and Zhu, 1995). Here, we also present the observed information matrix for the distribution.

3.1 The Maximum Likelihood Estimation

The log-likelihood based on a random sample $X_1, X_2, \ldots, X_n$ from the $EHN(\lambda, \beta, \sigma)$ is given by

$$\begin{aligned} l(\theta) &= n\log\lambda - 3n\lambda\log\sigma - n\log(2+\beta) \\ &+ \sum_{i=1}^{n}\log(2\sigma^{2\lambda} + \beta x_i^{2\lambda}) \\ &+ (\lambda - 1)\sum_{i=1}^{n}\log x_i - \frac{1}{2\sigma^{2\lambda}}\sum_{i=1}^{n} x_i^{2\lambda}, \end{aligned} \tag{9}$$

where $\theta = (\lambda, \beta, \sigma)'$.

The 2^{nd} order derivatives of the log-likelihood function given in equation (9) with regard to the parameters of interest are obtained as follows:

$l_{\lambda\lambda} = \frac{-n}{\lambda^2} - \frac{2}{\sigma^{2\lambda}}\log^2\sigma\sum_{i=1}^{n} x_i^{2\lambda} + \frac{4}{\sigma^{2\lambda}}\log\sigma\sum_{i=1}^{n} x_i^{2\lambda}\log x_i - \frac{2}{\sigma^{2\lambda}}\sum_{i=1}^{n} x_i^{2\lambda}\log^2 x_i$

$+ 8\beta\sigma^{2\lambda}\sum_{i=1}^{n}\frac{x_i^{2\lambda}(\log\sigma - \log x_i)^2}{(2\sigma^{2\lambda} + \beta x_i^{2\lambda})^2}$,

$l_{\lambda\beta} = I_{\beta\lambda} = -4\sigma^{2\lambda}\sum_{i=1}^{n}\frac{x_i^{2\lambda}(\log\sigma - \log x_i)}{(2\sigma^{2\lambda} + \beta x_i^{2\lambda})^2}$,

$l_{\lambda\sigma} = l_{\sigma\lambda} = \frac{-3n}{\sigma} + \frac{1}{\sigma^{2\lambda+1}}\sum_{i=1}^{n} x_i^{2\lambda} - \frac{2\lambda}{\sigma^{2\lambda+1}}\log\sigma\sum_{i=1}^{n} x_I^{2\lambda} + \frac{2\lambda}{\sigma^{2\lambda+1}}\sum_{i=1}^{n} x_i^{2\lambda}\log x_i$

$+ \frac{1}{\sigma^{2\lambda+1}}\sum_{i=1}^{n}\frac{8\sigma^{2\lambda} + 4\beta(1 + 2\lambda\log\sigma - 2\lambda\log x_i)x_i^{2\lambda}}{(2\sigma^{2\lambda} + \beta x_i^{2\lambda})^2}$,

$l_{\beta\beta} = \frac{n}{(2+\beta)^2} - \sum_{i=1}^{n}\frac{x_i^{4\lambda}}{(2\sigma^{2\lambda} + \beta x_i^{2\lambda})^2}$,

$l_{\beta\sigma} = l_{\sigma\beta} = -4\lambda\sigma^{2\lambda-1}\sum_{i=1}^{n}\frac{x_i^{2\lambda}}{(2\sigma^{2\lambda} + \beta x_i^{2\lambda})^2}$,

$l_{\sigma\sigma} = \frac{3n\lambda}{\sigma^2} - \frac{\lambda(2\lambda+1)}{\sigma^{2\lambda+2}}\sum_{i=1}^{n} x_i^{2\lambda} - \frac{4\lambda}{\sigma^{-2\lambda+2}}\sum_{i=1}^{n}\frac{2\sigma^{2\lambda} + \beta(1-2\lambda)x_i^{2\lambda}}{(2\sigma^{2\lambda} + \beta x_i^{2\lambda})^2}$.

The Hessian matrix can be written as

$$\mathrm{H}(\theta) = \begin{pmatrix} l_{\lambda\lambda} & l_{\lambda\beta} & l_{\lambda\sigma} \\ l_{\beta\lambda} & l_{\beta\beta} & l_{\beta\sigma} \\ l_{\sigma\lambda} & l_{\sigma\beta} & l_{\sigma\sigma} \end{pmatrix}$$

Under some regularity conditions, we have

$$\sqrt{n}(\widehat{\theta}-\theta) \rightarrow N_3(0, I^{-1}(\theta)),$$

where $I(\theta) = -\mathbf{E}(H(\theta))$ is the Fisher information matrix. Then asymptotically $\widehat{\theta}$ has a normal distribution with mean θ and variance $I^{-1}(\theta)$. This asymptotic behavior is also valid if $I(\theta)$ is replaced by $-H(\widehat{\theta})$ where $-H(\widehat{\theta})$ is the observed information matrix.

3.2 Simulation and Sensitivity Analysis

A Monte Carlo simulation study related to the estimation of the EHN parameters σ, β and λ is discussed next. The goal is to assess the quality of the maximum likelihood estimates. The result of Theorem 2.2 allows us to generate random numbers $X \sim EHN(\sigma_0, \beta_0, \lambda_0)$ through the steps of the following algorithm:

1. Choose the sample size n and the model parameters by the values λ_0, β_0 and σ_0;
2. Generate $W \sim \chi^2_3$, $Y \sim N(0,1)$ and S such that $P(S = 1) = P(S = -1) = \frac{1}{2}$;
3. Compute $R = \sqrt{W}\, S$;
4. Compute $Z = \sqrt{\frac{\beta_0}{2+\beta_0}}\, R + \sqrt{\frac{2}{2+\beta_0}}\, Y$;
5. Compute $X = \sigma_0 |Z|^{1/\lambda_0}$.

where Z is a ASN random variable (Elal-Olivero, 2010). We generated 1,000 random samples of different sizes $n = 100, 200, 300,$ and 400 from the EHN distribution with fixed parameter values $\lambda = 1$, $\beta = 4$ and $\sigma = 1$ considering the above algorithm. We computed the corresponding MLEs using R with the starting point (1, 1, 1). Table 1 represents the performance of the maximum likelihood estimators in terms of mean of Bias and MSE for the EHN parameters $\lambda, \beta,$ and σ. To test the stability of the results, we performed sensitivity analysis using different initial and parameter values. Due to lack of space, we will demonstrate the results only for the following EHN models: ($\lambda = 1, \beta = 4, \sigma = 1$) with initial values (0.8, 1.5, 0.5), ($\lambda = 1.5, \beta = 4, \sigma = 1$) with initial values (1, 1, 1), and ($\lambda = 2, \beta = 9, \sigma = 3$) with initial values (1, 0, 3). As Tables 2 to 4 display, the new MLEs are, in general, consistent with the primary analysis of the EHN(1, 4, 1). Convergence is confirmed with each case by the utilized machine learning algorithm. The MLEs returned adequate estimates for almost all parameter values and converged quite quickly therefore, most estimation errors are very small especially when "good" initial values were used to start the estimation process. Further, as n increases the errors decay significantly toward zero as Figures 3 and 4 reveal. The effect of increasing the sample size on the errors of estimation has been widely investigated and our findings in this regard are in line with many other studies such as (Amjad and Ismail, 2021). These simulation based results indicate that the EHN is a promising flexible distribution for analyzing reliability data, and that the MLE is an efficient and stable inference technique for the parameters of this new distribution.

4. Applications To Reliability Data

In this section, we demonstrate the importance of the EHN distribution in modeling reliability data using the method of Maximum Likelihood Estimation. Here, we provide two real applications to assess the Kolmogorov-Smirnov (K-S) and other goodness-of-fit statistics for the EHN distribution with respect to breaking stress of carbon fibers of 50 mm in length observations and strength of 1.5 cm glass fibers measured at the National Physical laboratory in England. Descriptive statistics for both datasets are displayed in Table 5. For more details on the these sets, the reader is referred to (Bakouch and Abd El-Bar, 2017).

The observed information in Table 5 and the corresponding data histograms displayed in Figures 5 and 6 suggest that each dataset may be modeled by negative-skewed and unimodal distributions with heavier tails than a normal distribution. Fits by several possible classic and relatively new distributions other than the EHN are presented for comparison purposes. These distributions are:

Table 1. Average bias and average MSE of MLEs for EHN model with true parameter values $\lambda = 1$, $\beta = 4$ and $\sigma = 1$ calculated when 1,000 replications for each n are considered. The initial values are (1, 1, 1).

Sample size n	Parameter	Bias	MSE
100	λ	0.0011	0.0013
	β	0.0246	0.6064
	σ	0.00098	0.00095
200	λ	0.00089	0.00079
	β	0.01524	0.2322
	σ	0.00081	0.00066
300	λ	0.00084	0.00071
	β	0.0138	0.189
	σ	0.00078	0.00061
400	λ	0.00075	0.00056
	β	0.0113	0.128
	σ	0.00075	0.00056

Table 2. Average bias and average MSE of MLEs for EHN model with true parameter values $\lambda = 1$, $\beta = 4$ and $\sigma = 1$ calculated when 1,000 replications for each n are considered. The initial values are (0.8, 1.5, 0.5).

Sample size n	Parameter	Bias	MSE
100	λ	0.0012	0.0014
	β	0.039	1.53
	σ	0.0014	0.002
200	λ	0.00102	0.00105
	β	0.0264	0.6952
	σ	0.00133	0.00176
300	λ	0.00091	0.00083
	β	0.016	0.2864
	σ	0.0012	0.001
400	λ	0.00082	0.00068
	β	0.01	0.26
	σ	0.001	0.0016

Table 3. Average bias and average MSE of MLEs for EHN model with true parameter values $\lambda = 1.5$, $\beta = 4$ and $\sigma = 1$ calculated when 1,000 replications for each n are considered. The initial values are (1, 1, 1).

Sample size n	Parameter	Bias	MSE
100	λ	0.0019	0.0036
	β	0.0227	0.516
	σ	0.000498	0.000248
200	λ	0.0016	0.0025
	β	0.0171	0.294
	σ	0.00048	0.00023
300	λ	0.0016	0.0025
	β	0.0137	0.188
	σ	0.00045	0.00025
400	λ	0.0014	0.0019
	β	0.012	0.154
	σ	0.00044	0.0002

Table 4. Average bias and average MSE of MLEs for EHN model with true parameter values $\lambda = 2$, $\beta = 9$ and $\sigma = 3$ calculated when 1,000 replications for each n are considered. The initial values are (1, 0, 3).

Sample size n	Parameter	Bias	MSE
100	λ	0.0033	0.011
	β	0.045	2.07
	σ	0.0012	0.015
200	λ	0.0029	0.0084
	β	0.021	0.46
	σ	0.0011	0.0012
300	λ	0.0026	0.0069
	β	0.020	0.43
	σ	0.0010	0.00108
400	λ	0.0025	0.0063
	β	0.015	0.22
	σ	0.00096	0.00093

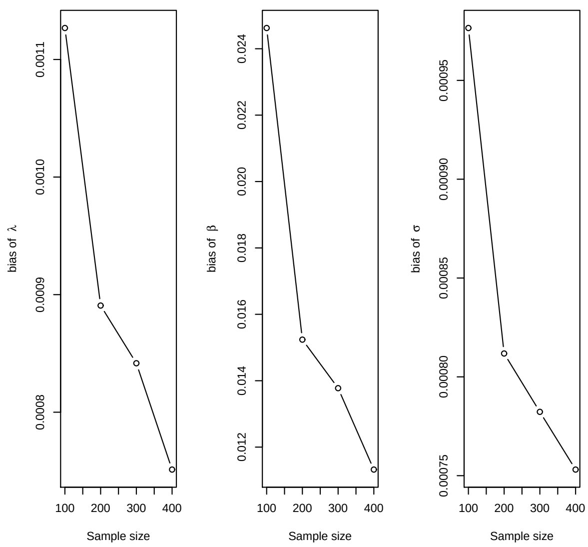


Fig. 3. The mean bias(n) versus $n = 100, 200, 300, 400$ associated with the EHN MLEs for $\lambda = 1$, $\beta = 4$ and $\sigma = 1$ computed by bias(n)= $\frac{1}{1000}\sum_{i=1}^{1000}(\widehat{\theta} - \theta)$ for $\theta \in \{\lambda, \beta, \sigma\}$

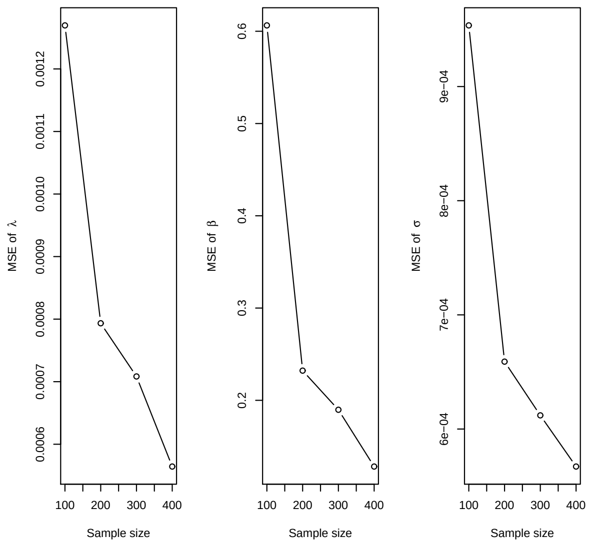


Fig. 4. The mean MSE(n) versus $n = 100, 200, 300, 400$ associated with the EHN MLEs for $\lambda = 1$, $\beta = 4$ and $\sigma = 1$ computed by MSE(n)= $\frac{1}{1000}\sum_{i=1}^{1000}(\widehat{\theta} - \theta)^2$ for $\theta \in \{\lambda, \beta, \sigma\}$

1. Generalized Lindley (GL) distribution with pdf

$$f(x) = \frac{\gamma\lambda^2}{1+\lambda}(1+x)e^{-\lambda x}1 - \frac{1+\lambda+\lambda x}{1+\gamma}e^{-\lambda x}{}^{\gamma-1}, \tag{10}$$

where $x > 0, \lambda, \gamma > 0$.

2. Exponentiated Exponential (EE) distribution with pdf

$$f(x) = \alpha\lambda e^{-\lambda x}(1 - e^{-\lambda x})^{\alpha-1}, \tag{11}$$

where $x > 0, \alpha, \sigma > 0$.

3. Modified Generalized Half-Normal (MGHN) distribution with pdf

$$f(x) = \frac{2\gamma x^{\gamma-1}}{27\beta^{3\gamma}}(2\beta^{2\gamma} + 25x^{2\gamma})\phi\left(\left(\frac{x}{\beta}\right)^{\gamma}\right), \tag{12}$$

where $x > 0, \gamma, \beta > 0$.

4. Gamma distribution (Ga) with pdf

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}}x^{\alpha-1}e^{-x/\beta}, \tag{13}$$

where $x > 0, \alpha, \beta > 0$.

5. Alpha-Skew-Normal Distribution (ASN) with pdf

$$f(x) = \frac{(1-\alpha x)^2 + 1}{2+\alpha^2}\phi(x) \tag{14}$$

where $x \in \mathbf{R}, \alpha \in \mathbf{R}$.

6. Half-Alpha-Skew-Normal distribution presented in equation (5).

Table 5. Summary statistics of the breaking stress of the 50 mm carbon fibers and the 1.5 cm glass fibers data sets, where a_1 and a_2 represent the coefficients of skewness and kurtosis respectively.

Data	n	$\bar{x}$	s	a_1	a_2
Carbon	66	2.759	0.891	-0.132	3.223
Glass	50	1.441	0.331	-0.629	3.774

Maximum likelihood estimates of the unknown parameters of the seven distributions are produced using the machine learning tool developed by (Byrd and Zhu, 1995), available in R. The corresponding Kolmogorov-Smirnov statistics and the associated p-values are displayed in Tables 6 and 7. For more comparisons, other goodness-of-fit statistics that have been widely used in many applications as presented in (Khorsheed and Razzaghi, 2020) are considered here. These statistics are: Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Consistent Akaike Information Criterion (CAIC) and Hannan-Quinn Information Criterion (HQIC). The corresponding values of these measures are presented in the same tables mentioned above. The results reveal that the EHN distribution has the lowest goodness-of-fit measures, whereas the highest values are associated with the related ASN and HASN distributions. From Figures 5 & 6, we noticed that the type of skewness of both datasets is reversely estimated by almost all models except the EHN and the MGHN distributions. Moreover, the corresponding ML estimates of the ASN and HASN distributions are extremely large and almost approach to ∞. According to (Elal-Olivero, 2010), this phenomenon may suggests bimodal-normal fits rather than unimodal. To investigate this assumption, we conducted the Hartigan's dip test for unimodality & multimodality (Maechler, 2016) using the corresponding R package. With each dataset, we obtained a p-value > 0.9 for the null hypothesis of unimodality. The results derived from the Hartigan's test and also the K-S statistics indicate that both ASN and HASN distributions did not fit the datasets well enough at 5% significance level.

Furthermore, the MGHN and the EHN models performed almost in the same manner when fitting the carbon observations as seen in Figure 5. This is due to the fact that the MGHN is a special case of the EHN distribution with $\beta = 25$ (equation 4) and the corresponding ML estimated scale and shape parameters of both distributions are not significantly different for this particular set of data as Table 6 reveals. Also, the recorded skewness level of this dataset is small (-0.132), i.e., the underlying distribution is approximately symmetric. When the skewness is higher, as observed in the glass dataset (-0.629), the MGHN model did not fit the data adequately at the peak and the right tail of the distribution as corroborated by Figure 6. Obviously, the EHN distribution outperformed the MGHN in this case. These results indicate that the EHN model provides the most flexible fits among all compared distributions especially when the data are moderately to highly skewed.

5. Concluding Remarks

In this article, we present a promising flexible distribution with three parameters called the Extended Half-Normal (EHN) distribution. According to particular values of its parameters, it includes few special cases known in the literature: Generalized Half-Normal, Modified Generalized Half-Normal, Half-Alpha-Skew-Normal and the Half-Normal distributions. Various theoretical properties of the EHN model, namely the moments, asymmetry and kurtosis are derived. A demonstrative Monte Carlo and sensitivity analysis study is presented. Applying the Maximum Likelihood approach, the estimation results show a good retrieval of the EHN parameters. Moreover, two reliability datasets are used to demonstrate the practical importance of the EHN model. The EHN distribution appears to be a good competitor to several existing distributions for lifetime and reliability skewed data.

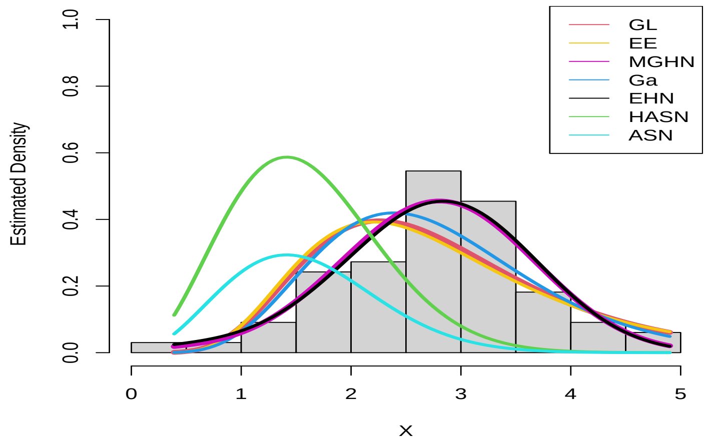


Fig. 5. The fitted densities for the competing distributions for breaking stress of carbon fibers data.

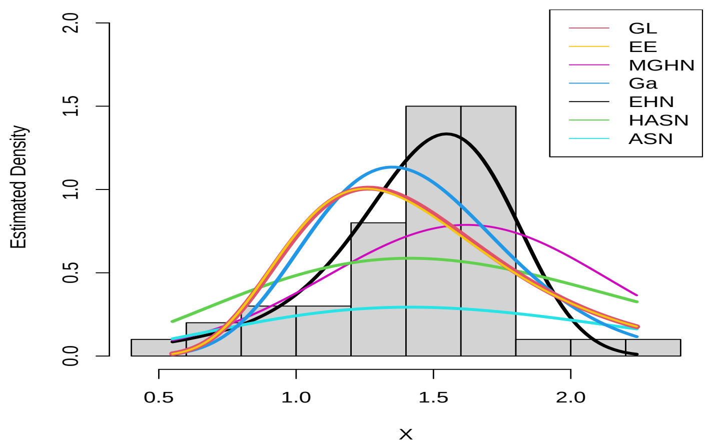


Fig. 6. The fitted densities for the competing distributions for strength of 1.5 cm. glass data.

Table 6. The estimated parameters for the models fitted to the breaking stress of carbon fibers dataset and the values of AIC, BIC, CAIC, HQIC, K-S statistics with the corresponding p-value.

Distribution	Estimates	AIC	BIC	CAIC	HQIC	K-S	p-value
$GL(\gamma,\lambda)$	$\widehat{\gamma}=7.0412$	191.5939	195.9733	191.3235	193.3244	0.1319	0.1839
	$\widehat{\lambda}=1.2461$						
$EE(\lambda,\alpha)$	$\widehat{\lambda}=1.0077$	194.7447	199.1241	194.4743	196.4752	0.1550	0.0839
	$\widehat{\alpha}=9.2009$						
$MGHN(\beta,\gamma)$	$\widehat{\beta}=2.1394$	177.2419	183.8109	176.8362	179.8376	0.0744	0.8316
	$\widehat{\gamma}=1.5176$						
$Ga(\alpha,\beta)$	$\widehat{\alpha}=7.4885$	186.3351	190.7144	186.0646	188.0656	0.1328	0.1948
	$\widehat{\beta}=0.3685$						
$ASN(\alpha)$	$\widehat{\alpha}=9.9983\times10^{9}$	427.5930	429.7826	427.4577	428.4582	0.8055	8.881×10^{-16}
$HASN(\beta)$	$\widehat{\beta}=9.9997\times10^{9}$	336.0977	338.2872	335.9623	336.963	0.5908	8.882×10^{-16}
$EHN(\lambda,\beta,\sigma)$	$\widehat{\lambda}=1.5705$	176.4462	183.0152	176.0405	179.0419	0.0731	0.9342
	$\widehat{\beta}=16.0108$						
	$\widehat{\sigma}=2.1888$						

Table 7. The estimated parameters for the models fitted to the strength 1.5cm glass fibers measurements and the values of AIC, BIC, CAIC, HQIC, K-S statistics with the corresponding p-value.

Distribution	Estimates	AIC	BIC	CAIC	HQIC	K-S	p-value
$GL(\gamma,\lambda)$	$\widehat{\gamma}=24.3069$	50.9599	54.7839	50.7238	52.4162	0.2068	0.0236
	$\widehat{\lambda}=3.0704$						
$EE(\lambda,\alpha)$	$\widehat{\lambda}=2.6850$	52.0331	55.8571	51.7968	53.4892	0.2171	0.0180
	$\widehat{\alpha}=28.9652$						
$MGHN(\beta,\gamma)$	$\widehat{\beta}=1.2366$	33.4234	39.1595	33.0691	35.6077	0.1648	0.1179
	$\widehat{\gamma}=2.3012$						
$Ga(\alpha,\beta)$	$\widehat{\alpha}=15.9115$	41.9366	45.7607	41.7005	43.3929	0.2157	0.0191
	$\widehat{\beta}=0.0906$						
$ASN(\alpha)$	$\widehat{\alpha}=9.9982\times10^{9}$	136.3403	138.2524	136.2222	137.0682	0.5907	2.221×10^{-16}
$HASN(\beta)$	$\widehat{\beta}=9.9973\times10^{9}$	67.0257	68.9377	66.9076	67.7538	0.3543	4.02×10^{-6}
$EHN(\lambda,\beta,\sigma)$	$\widehat{\lambda}=2.6312$	32.4929	38.2291	32.1740	34.6773	0.1601	0.1379
	$\widehat{\beta}=7.4592$						
	$\widehat{\sigma}=1.3084$						

References

Aharony, J., Jones, C., & **Swary, I. (1980)**. An Analysis of Risk and Return Characteristics of Corporate Bankruptcy Using Capital Market data. *The Journal of Finance*, 35, 1001-1016.

Amjad, S., & **Ismail, M. (2021)**. Efficient transformed ratio-type estimator using single auxiliary information. *Kuwait Journal of Science*, 42(2), 40-50.

Ansell, J., & **Phillips, M. (1990)**. Practical reliability data analysis. *Reliability Engineering & System Safety*, 28(3), 337-356.

Azzalini, A. (1985). A class of distribution which includes the normal ones. *Scandinavian Journal of Statistics*, 12, 171-178.

Badia, M., Serrats, J., & **Rodon, M. (2020)**. Risk Analysis through the Half-Normal Distribution. *mathematics*, 8(11):2080. https://doi.org/10.3390/math8112080.

Bakouch, H., & **Abd El-Bar, A.(2017)**. A New weighted Gompertz Distribution with applications to reliability data. *Applications of Mathematics*, 62(3), 269-296.

Byrd Lpnj, R. H., & **Zhu, C. (1995)**. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5), 1190-1208.

Cooray, K., & **Ananda, M.(2008)**. A generalization of the half-normal distribution with applications to lifetime data. *Communications in Statistics Theory and Methods*, 37(3), 1323-1337.

Elal-Olivero, D. (2010). Alpha–skew–normal distribution. *Proyecciones Journal of Mathematics*, 29(3), 224-240.

Johnson KS, N. L., & **Balakrishnan, N. (1995)**. Continuous Univariate Distribution, 2, 2nd edition. New York, NY: John Wiley New York.

Katal, A., Wazid, M., & **Goudar, R. (2013)**. Big Data: Issues, Challenges, Tools and Good Practices. In *Procedures of the 2013 Sixth International Conference on Contemporary Computing*, (pp. 404-409). Noida: IEEE.

Khorsheed, E., & **Razzaghi, M. (2020)**. Model Averaged Benchmark Dose Analysis for Multiple Outcomes in Developmental Toxicity Experiments. *Applied Mathematics and Information Sciences*, 14(4), 733-742.

Khorsheed, E., Salinas, H. S., & **Bakouch, H. (2020)**. A New family of Skew-Normal lifetime Distributions for Industrial Applications. In *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)* (pp. 1-4). Sakheer, Bahrain: IEEE Xplore, doi: 10.1109/ICDABI51230.2020.9325590.

Kishorilal, D., & **Mukhopadhyay, A. (2018)**. Reliability investigation of diesel engines used in dumpers by the Bayesian approach. *Kuwait Journal of Science*, 45(4), 15-25.

Krause, A.L, Kurowski, L., Yawar, K., & **Van Gorder, R.A. (2018)**. Stochastic epidemic metapopulation models on Networks: SIS dynamics and control strategies. *Journal of Theoretical Biology*, 449, 35-52.

Krenek, R., Cha, J., & **Cho, B. (2016)**. Development of the convolutions of truncated normal random variables with three different quality characteristics in engineering applications. *Computers & Industrial Engineering*, 94, 125-137.

Maechler, M. (2016). Diptest: Hartigan's Dip Test Statistic for Unimodality—Corrected: R Package Version 0.75-7.2016. Available online: https://CRAN.R-project.org/.

Olmos, N., & **Venegas, O. (2018)**. Modified Generalized Half-Normal Distribution with Application to Lifetimes. *Applied Mathematics & Information Sciences*, 12(3), 637-643.

Pescim, R. R., Demertrio, C. G. B., Cordeiro, G. M., Ortega, E. M. M., & **Urbano, M. R. (2010)**. The beta generalized half-normal distribution. *Computational Statistics & Data Analysis*, 54(4), 945-957.

R Core Team (2020). R: A language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL http://www.R-project.org/.

Submitted: 21/02/2021
Revised: 25/05/2021
Accepted: 23/06/2021
DOI: 10.48129/kjs.12615