Available online at https://asric.africa/engineering-sciences
ASRIC Journal on Engineering Sciences Vol.4(1)(2023) 32-39

Simulation and Application to Survival Analysis of Neck cancer disease with New Probability Distribution

Saeed E. Hemeda[1], El-Basuony H. Mohamed[1] & Seyam M. Abd El-Aziez[1]

[1]Obour High Institute for Management & Informatics and Computer Sciences
Email: Saiedhemeda@oi.edu.eg, helbasuony@oi.edu.eg, Mohameds@oi.edu.eg

Abstract
In this paper, a new distribution depends on the hyperbolic sine family of distributions with inverted exponential distribution as will be generated. The hyperbolic sine family of distributions was introduced by Kharazmi and Saadatinik (2016). They studied some properties of this family and obtained the estimates of its parameters by different methods. Various properties of the proposed distribution including explicit expressions for the moments, quantiles, moment generating function, failure rate function, mean residual lifetime, order statistics and expressions of entropies are derived. Superiority of this model is proved in some simulations and application of Survival Analysis of Neck cancer disease.
Keywords: *Sinh inverted exponential distribution; Moments; Survival function; Entropy; Maximum likelihood estimates.*

1. Introduction

In many applied areas such as lifetime analysis and other fields, there is strong need to develop the classical distributions. So, different methods to generating new families of distributions are defined. These include; Azzalini's skew family by Azzalini (1985), beta-G by Eugene et al. (2002). Recently, Kharazmi and Saadatinik (2018) discussed the hyperbolic sine family of distributions, Chakraborty and Handique (2017) investigated the generalized Marshall-Olkin Kumaraswamy-G family and generalized inverted Weibull family by Hemeda et. al (2019) and more.

According to Kharazmi and Saadatinik (2018), the hyperbolic Sine (HS) family with cumulative CDF is

$$F(x)=\frac{2e^{\delta}}{(e^{\delta}-1)^2}\left(\cosh(\delta G(x))-1\right), \tag{1}$$

and probability density PDF is

$$f(x)=\frac{2\delta e^{\delta}}{(e^{\delta}-1)^2}g(x)\sinh(\delta G(x));x>0,\delta>0 . \tag{2}$$

Where, $G(x)$ and $g(x)$ are the CDF and PDF for any random variable, respectively and the hyperbolic sine function $(Sinh(x))$ is defined as

$$Sinh(x)=\frac{1}{2}\left(e^{x}-e^{-x}\right) . \tag{3}$$

Using series expansion theorem, $Sinh\ (x)$ takes the following formula;

$$Sinh(x)=\sum_{j=0}^{\infty}\frac{x^{2j+1}}{(2j+1)!} . \tag{4}$$

In our study, we will take $G(x)$ is the CDF of the inverted exponential distribution and $g(x)$ its PDF.
Dey (2007) studied the inverted exponential (IE) distribution with CDF and PDF are given by

$$G(x;\theta)=e^{\frac{-\theta}{x}}, \tag{5}$$

$$g(x;\theta)=\frac{\theta}{x^2}e^{\frac{-\theta}{x}}\ ;\ x>0,\ \theta>0. \tag{6}$$

2. The New Model

This section contributes the representation of Sinh inverted exponential (SIEDD) distribution. The CDF, reliability, hazard rate, cumulative hazard rate functions are deduced and discussed analytically. Above that; statistical measures will be discussed.

By substituting from (5) and (6) into (1) and (2), then the SIEDD CDF and PDF will be

$$F(x;\theta,\delta)=\frac{2e^{\delta}}{(e^{\delta}-1)^2}\left(\cosh\left(\delta e^{\frac{-\theta}{x}}\right)-1\right), \tag{7}$$

$$f(x;\theta,\delta)=\frac{2\delta\theta e^{\delta}}{(e^{\delta}-1)^2x^2}e^{\frac{-\theta}{x}}\sinh\left(\delta e^{\frac{-\theta}{x}}\right);x>0,\theta,\delta>0. \tag{8}$$

Using (1), the PDF will be in the following form

$$f(x;\theta,\delta)=\frac{\delta\theta e^{\delta}x^{-2}}{(e^{\delta}-1)^2}\left(e^{\delta e^{\frac{-\theta}{x}}-\frac{\theta}{x}}-e^{-\delta e^{\frac{-\theta}{x}}-\frac{\theta}{x}}\right). \tag{9}$$

The survival and hazard rate (hr) functions are

$$S(x;\theta,\delta)=1-\frac{2e^{\delta}}{(e^{\delta}-1)^2}\left(\cosh\left(\delta e^{\frac{-\theta}{x}}\right)-1\right), \tag{10}$$

$$h(x;\theta,\delta)=\frac{2\delta\theta e^{\delta-\frac{\theta}{x}}\sinh\left(\delta e^{\frac{-\theta}{x}}\right)}{\left[1-\frac{2e^{\delta}}{(e^{\delta}-1)^2}\left(\cosh\left(\delta e^{\frac{-\theta}{x}}\right)-1\right)\right](e^{\delta}-1)^2x^2}. \tag{11}$$

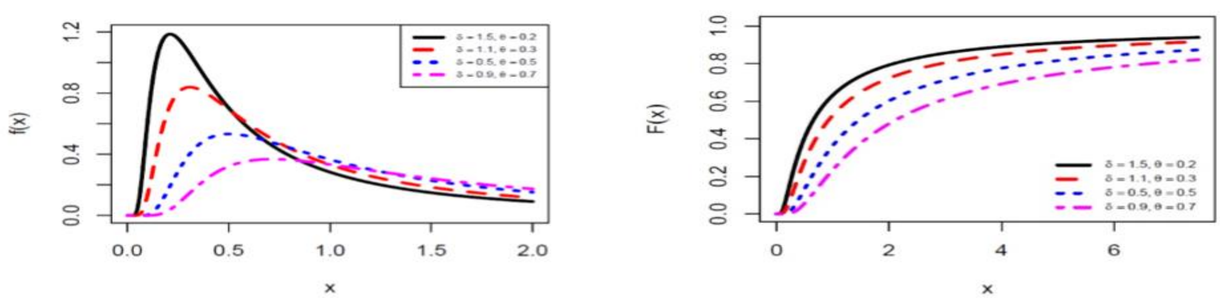

Figure 1: Probability density and cumulative functions of the new model.

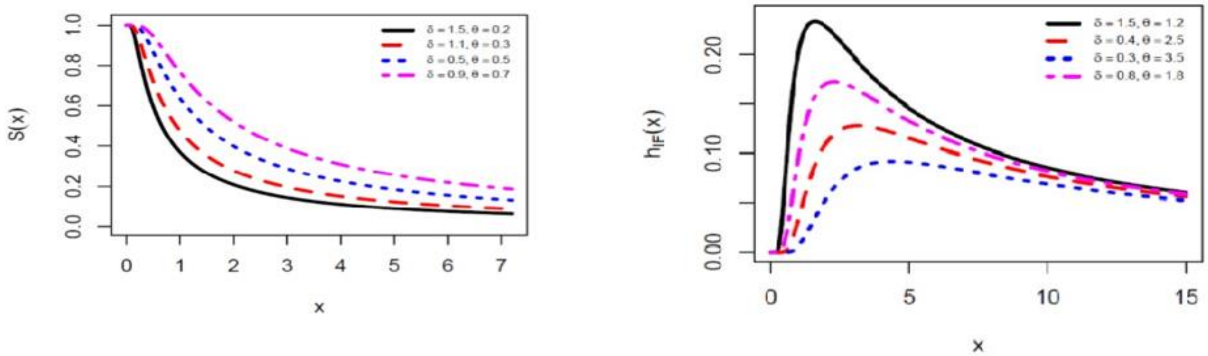

Figure 2: Survival and Hazard rate functions of the new model.

3. Useful Statistical Properties

Various statistical measures will be deduced such as moments, moment generating function, incomplete moments and mean residual life time, quantile function, median, mode, entropies, skewness and kurtosis of SIEDD distribution in this section.

3.1 The Moments About Mean

From equations (3), (4), we can obtain the probability density function in the series form as the following;

$$f_{SIE}(x)=\sum_{j=0}^{\infty} A_j\, x^{-2}\, e^{\frac{-2\theta(j+1)}{x}};x>0,\delta,\theta>0 \quad . \tag{12}$$

Where, $A_j=\dfrac{2\theta\,\delta^{2j+2}\,e^{\delta}}{(e^{\delta}-1)^2(2j+1)!}$.

Since the nth moments is defined as

$$\mu_n'=\int_0^{\infty} x^n f(x;\theta,\delta)dx \quad .$$

By substituting from (13) into the last equation, the nth moment is written as

$$\mu_n'=\sum_{j=0}^{\infty} A_j\int_0^{\infty} x^{n-2}e^{\frac{-2\theta(j+1)}{x}}dx \quad ,$$

$$\mu_n'=\sum_{j=0}^{\infty} A_j\,\frac{\Gamma(n-1)}{\left(\theta(2\,\mathrm{j}+1)\right)^{n-1}};m=2,3,... \tag{13}$$

From the last equation, the 2nd and 3rd and other moments can be calculated.

The 2nd and 3rd moments are

$$\mu_2=\sum_{j=0}^{\infty} A_j=\sum_{j=0}^{\infty}\frac{2\delta^{2j+2}\,e^{\delta}}{(e^{\delta}-1)^2(2j+2)!}. \tag{14}$$

$$\mu_3=\sum_{j=0}^{\infty}\frac{2\;\delta^{2j+2}\,e^{\delta}}{\theta(2\,\mathrm{j}+1)(e^{\delta}-1)^2(2j+2)!}. \tag{15}$$

3.2 Moment Generating Function and Mean Residual Life Time

The moment generating function of a probability distribution can be derived as follows

$$M_{SIE}(t)=\sum_{m=0}^{\infty}\frac{t^m\,\mu_m'}{m!},$$

where, μ'_m is the mth moment about origin. Using (13) then $M_{SIE}(t)$ will be

$$M_{SIE}(t)=\frac{t^m}{m!}\sum_{j=0}^{\infty}A_j\frac{\Gamma(m-1)}{[\theta(2j+1)]^{m-1}};m=2,3,... \tag{16}$$

The mean residual of SIEDD distribution $m_{SIE}(t)$ is determined by

$$m(t)=\left[\frac{1}{S(t)}\int_t^{\infty}x\,f(x)\,dx\right]-t$$ (Gupta and Gupta (1983)). (17)

Substituting from PDF (13) then $m_{SIE}(t)$ will be calculated by

$$m_{SIE}(t)=\frac{\sum_{j=0}^{\infty}A_j\,\Gamma(2,t)}{1-\frac{2e^{\delta}}{(e^{\delta}-1)^2}\left(\cosh\left(\delta e^{\frac{-\theta}{x}}\right)-1\right)}-t \tag{18}$$

3.3 Quantile Function

The quantile function of SIED, ($x(p)=F^{-1}(p)$) is determined by converting (7) as follows:

$$x(p)=\frac{-\theta}{\ln\left[\frac{1}{\delta}\cosh^{-1}\left(\frac{p(e^{\delta}-1)^2}{2e^{\delta}}+1\right)\right]} \tag{19}$$

Equation (19) can be solved numerically, the SIEDD random variable X can be generated where p has the uniform distribution on the interval [0,1].

3.4 Skewness and kurtosis

The skewness (Λ) and kurtosis (Υ) coefficients based on quantiles are computed from the following formulas:

$$\Lambda=\frac{q(0.75)-2Q(0.25)+Q(0.25)}{Q(0.75)-Q(0.25)}$$ (Kenney and Keeping (1962)). (20)

$$\Upsilon=\frac{Q(0.875)-Q(0.625)+Q(0.375)-Q(0.125)}{Q(0.75)-Q(0.25)}$$, (Moors (1988)). (21)

Substituting from (19) into (20) and (21) respectively, we can get some values to the skewness and kurtosis coefficients of SIED as represented in Figure 3.

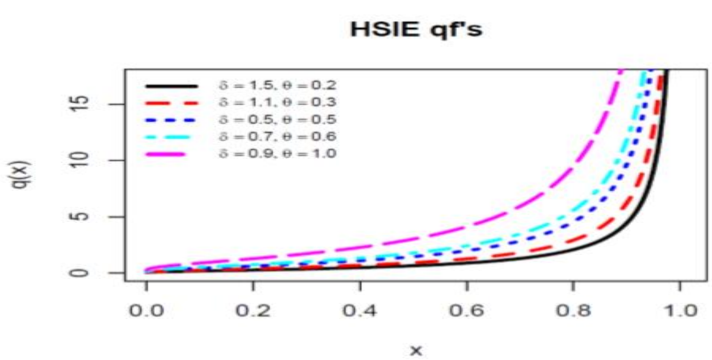

Figure 3: Quantile function of SIED with selected values of its parameters.

3.5 The Rényi Measure

The entropy of a random variable X is an important measure, it is defined as a measure of variation of the uncertainty (see, Rényi (1961)). In this subsection, we discuss Rényi measure.

The Rényi entropy $E_{\mathrm{Re}n}(\xi)$ of a random variable X is defined as

$$E_{\mathrm{Re}n}(\xi)=\frac{1}{1-\xi}Log\left[\int_0^\infty f^{\xi}(x)\,dx\right],$$

where $\xi>0$ and $\xi\neq 1$. Based on PDF (13) using binomial expansion theorem and after some simplifications, we obtain

$$f_{SIE}^{\xi}(x)=\sum_{n=0}^{\infty}\sum_{m=0}^{n}\sum_{k=0}^{\xi}c_{n,m,k}\;x^{-m-2\xi}\,e^{\frac{-\theta(n-m)}{x}}, \tag{22}$$

where

$$c_{n,m,k}=\frac{(-1)^{k+m}\theta^{m}(k-2\xi)^{n}\delta^{(n-m)}\,\xi!}{k!m!(\xi-k)!(n-m)!}\left(\frac{\delta\,\theta\,e^{\delta}}{(e^{\delta}-1)^{2}}\right)^{\xi}.$$

The Rényi entropy will be

$$E_{\mathrm{Re}n}(\xi)=\frac{1}{1-\xi}Log\left[\sum_{n=0}^{\infty}\sum_{m=0}^{n}\sum_{k=0}^{\xi}c_{n,m,k}\;\frac{\Gamma(m+2\xi-1)}{\left(\theta(n-m)\right)^{m+2\xi-1}}\right];\ \theta\neq 0,\ n\neq m. \tag{23}$$

4. Maximum likelihood Estimation of The Parameters

In this section, the maximum likelihood estimators of the model parameters $\Im=(\delta,\theta)$ of SIED from complete samples are deduced. Assume $X_1,\ X_2,\dots,X_n$ be a simple random sample from SIED with observed values $x_1,\ x_1,\dots x_n$, the log likelihood function of (8) is obtained as follows

$$L(\Im)=\prod_{j=1}^{n}f_{SIE}(x_j),$$

$$lnL(\Im)=\sum_{j=1}^{n}\ln f_{SIE}(x_j),$$

Based on the PDF (9), then

$$lnL(\Im)=\sum_{j=1}^{n}\ln\left[\frac{\delta\,\theta\,e^{\delta}\,x_j^{-2}}{(e^{\delta}-1)^{2}}\left(e^{\delta e^{\frac{-\theta}{x_j}}-\frac{\theta}{x_j}}-e^{-\delta e^{\frac{-\theta}{x_j}}-\frac{\theta}{x_j}}\right)\right],$$

$$lnL(\Im)=n\ln\left(\frac{\delta\theta e^{\delta}}{(e^{\delta}-1)^{2}}\right)-2\ln\sum_{j=1}^{n}x_j-\frac{\theta}{\sum_{j=1}^{n}x_j}+\ln\left(\sum_{j=1}^{n}e^{\delta e^{\frac{-\theta}{x_j}}}-\sum_{j=1}^{n}e^{-\delta e^{\frac{-\theta}{x_j}}}\right).$$

Differentiating $lnL(\Im)$ with respect to θ, δ and setting the result equals to zero, the maximum likelihood estimators will be gotten. The partial derivatives of $lnL(\Theta)$ with respect to each parameter are given as

$$\frac{\partial lnL(\Im)}{\partial\theta}=\frac{n}{\theta}-\frac{1}{\sum_{j=1}^{n}x_j}-\frac{\delta\sum_{j=1}^{\infty}\frac{1}{x_j}e^{\left(\delta e^{\frac{-\theta}{x_j}}\frac{-\theta}{x_j}\right)}}{\sum_{i=1}^{\infty}\left(e^{\delta e^{\frac{-\theta}{x_j}}}-e^{-\delta e^{\frac{-\theta}{x_j}}}\right)}. \tag{24}$$

$$\frac{\partial lnL(\Im)}{\partial\delta}=\frac{n}{\delta}+n-\frac{2ne^{\delta}}{(e^{\delta}-1)}+\frac{\sum_{j=1}^{n}\left(e^{\left(\delta e^{\frac{-\theta}{x_j}}-\frac{\theta}{x_j}\right)}+e^{\left(-\delta e^{\frac{-\theta}{x_j}}-\frac{\theta}{x_j}\right)}\right)}{\sum_{j=1}^{n}\left(e^{\delta e^{\frac{-\theta}{x_j}}}-e^{-\delta e^{\frac{-\theta}{x_j}}}\right)}. \tag{25}$$

The maximum likelihood estimators of the model parameters are determined by solving the non-linear equations (24) and (25) simultaneously. These equations can be solved numerically using iterative technique. For interval estimation of the parameters, the 2×2 observed information matrix $I(\Omega)=\{I_{uv}\}$ for (θ,δ). Under the regularity conditions, the known asymptotic properties of the maximum likelihood method ensure that: $\sqrt{n}(\hat{\Omega}-\Omega) \overset{d}{\rightarrow} N_2(0, I^{-1}(\Omega))$ as $n \rightarrow \infty$, where $\overset{d}{\rightarrow}$ means the convergence in distribution, with mean $O=(0,0)^T$ and 2×2 covariance matrix $I^{-1}(\Omega)$ then, the $100(1-\beta)\%$ confidence intervals for θ and δ are given, respectively, as follows

$\hat{\delta} \pm Z_{\beta/2}\sqrt{var\ (\hat{\delta})}$ and $\hat{\theta} \pm Z_{\beta/2}\sqrt{var\ (\hat{\theta})}$, where $Z_{\beta/2}$ is the standard normal at $\beta/2$. The significance level is $\beta/2$ and the variances of θ, δ are the diagonal elements of $I^{-1}(\Omega)$ corresponding to the model parameters.

5. Simulation Study

A simulation study is carried out to investigate the performance of estimators for SIEDD distribution in terms of their bias (bias), mean square error (MSE) using maximum Likelihood estimation (MLE) method. Simulated procedures can be described as follows:

Generated samples of sizes n = 30, 50, 100 from SIEDD distribution are generated and parameters are estimated using the maximum likelihood estimation method. 10000 such repetitions are made to calculate the bias and mean square error (MSE) of these estimates using the formula of estimates for any parameter η by

Bias$_\psi$ ($\hat{\psi}$)= $\frac{1}{1000}\sum_{i=1}^{1000}(\hat{\psi}-\psi)$ and MSE$_\psi$ ($\hat{\psi}$)= $\frac{1}{1000}\sum_{i=1}^{1000}(\hat{\psi}-\psi)^2$ respectively.

From Table 1, it is observed that;

i. As sample size n increases, bias decreases. That shows accuracy of the MLE of the parameters.

ii. As sample size n increases, MSE decreases. That shows consistency (or preciseness) of the MLE of the parameters as shown in figure 4.

Table 1: Bias and MSE of MLEs for SIEDD distribution

	$\hat{\theta} = 0.3$			$\hat{\delta} = 0.1$		
n	30	50	100	30	50	100
BIAS	-0.10987	-0.1078	-0.1006	-0.0963	-0.0712	-0.0160
MSE	0.0142	0.0123	0.0022	0.0693	0.0293	0.0053
	$\hat{\theta} = 0.1$			$\hat{\delta} = 0.2$		
n	30	50	100	30	50	100
BIAS	-0.1099	-0.1037	-0.0097	-0.1963	-0.1960	-0.1960
MSE	0.1123	0.0184	0.0112	0.0385	0.0274	0.0080
	$\hat{\theta} = 0.2$			$\hat{\delta} = 0.2$		
n	30	50	100	30	50	100
BIAS	-0.2099	-0.2099	-0.2098	-0.1961	-0.1959	-0.1959
MSE	0.0443	0.0441	0.0440	0.0384	0.0384	0.0384
	$\hat{\theta} = 0.2$			$\hat{\delta} = 0.3$		
n	30	50	100	30	50	100
BIAS	-0.2099	-0.2099	-0.2098	-0.2962	-0.2960	-0.2959
MSE	0.0443	0.0441	0.0441	0.0877	0.0876	0.0876
	$\hat{\theta} = 0.5$			$\hat{\delta} = 0.1$		
n	30	50	100	30	50	100
BIAS	-0.1400	-0.1100	-0.1100	-0.0905	-0.0453	-0.0150
MSE	0.1601	0.1601	0.1600	0.0092	0.0092	0.0092
	$\hat{\theta} = 0.1$			$\hat{\delta} = 0.5$		
n	30	50	100	30	50	100
BIAS	-0.1109	-0.1107	-0.1107	-0.4953	-0.4952	-0.4949
MSE	0.0123	0.0123	0.0123	0.2453	0.2452	0.2449
	$\hat{\theta} = 0.1$			$\hat{\delta} = 0.1$		
n	30	50	100	30	50	100
BIAS	-0.1102	-0.1100	-0.1100	-0.0962	-0.0959	-0.0959
MSE	0.0122	0.0122	0.0121	0.0093	0.0092	0.0092
	$\hat{\theta} = 0.5$			$\hat{\delta} = 0.5$		
n	30	50	100	30	50	100
BIAS	-0.5099	-0.5099	-0.5099	-0.4960	-0.4959	-0.4952
MSE	0.2601	0.2600	0.2600	0.2460	0.2495	0.2453

6. Neck cancer Disease Application

In this section, the SIED distribution is fitted for a real data. The real data represents the survival times of patients suffering from Neck cancer disease. The patients in this group were treated using a combined radiotherapy and chemotherapy (CT&RT). The data are

12.2	23.56	23.74	25.78	31.98	37	41.35	47.38	55.46	58.36	63.47	68.46
78.26	74.47	81.43	84	92	94	110	112	119	127	130	133
140	146	155	159	173	179	194	195	209	249	281	319
339	432	469	519	633	725	817	776				

Kumar et al. (2015) fitted this data to the inverted Lindley distribution. We have fitted this data set with SIEDD distribution compared with Weibull (W) and inverted exponential (IE) probability distributions. The results of estimated values of the parameters (Log-likelihood, AIC, BIC and KS) are listed in Table 2. The Q-Q plot, histogram, fitted PDF and estimated CDF of the SIEDD curve to this data have been shown in Figures 5 and 6 respectively. The selection criterion is that the lowest Log-likelihood and AIC correspond to the best model fitted. The MLEs, AIC, BIC and KS are shown in Table 2. From the Table, we can observe that the SIEDD model shows the smaller Log-likelihood, AIC, BIC and KS than other competing distributions.

Table 2: Statistical measures of fitted models using survival times of patients suffering from Neck cancer disease data

Distribution	Estimators	LL	AIC	BIC	KS
SIEDD	$\hat{\theta} = 1.53\,, \hat{\delta} = 7.91$	-279.32	564.64	409.13	0.1840
W	$\hat{\beta} = 3.07\,, \hat{\alpha} = 11.26$	-288.79	597.43	532.02	0.1752
IE	$\hat{\delta} = 5.47$	-480.35	862.71	813.00	0.0637

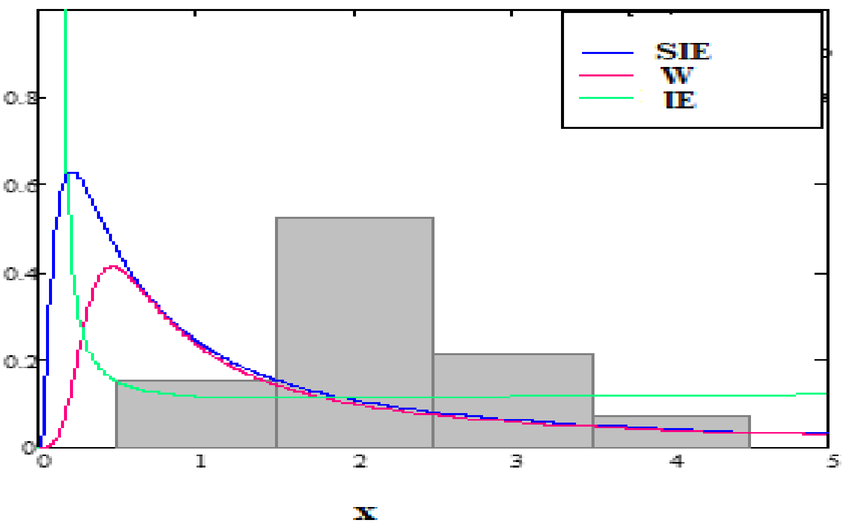

Figure 4: The Histogram and fitted models PDF for the data.

Figure 4 shows that; the Neck cancer disease application of the SIEDD distribution provides a better fit than other alternative distributions.

7. Conclusion

In this article, we have introduced and studied a new probability distribution called sinh inverted exponential distribution based on hyperbolic sin generator. The estimates of the parameters are obtained by different methods. Various properties of the proposed distribution moments including quantiles, moment generating function, failure rate function, mean residual lifetime, order statistics and expressions of entropies are derived. The structural and reliability properties of this distribution have been studied and inference on parameters have also been mentioned.The estimation of parameters is approached by maximum likelihood methode.We presented a simulation study to exhibit the performance and accuracy of maximum likelihood estimates of the SIEDD model parameters. The Neck cancer disease real data application is applied to illustrate the efficience and applicability of the SIEDD distribution. The application of the SIEDD distribution shows that it could provide a better fit than other alternative distributions. Superiority of this model is proved in some simulations and application of Survival Analysis of Neck cancer disease.

References

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6), 716-723.

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

Chakraborty, S. and Handique, L. (2017). The generalized Marshall-Olkin Kumaraswamy-G family of distributions. Journal of Data Science, forthcoming

Cordeiro, G.M. and de Castro, M. (2011). A new family of generalized distributions. J. Stat. Comput. Simul., 81(7), 883-893.

Dey, S. (2007). Inverted exponential distribution as a life distribution model from a Bayesian viewpoint. Data science journal, 6, 107-113.

Eugene, N., Lee, C. and Famoye, F. (2002). Beta-normal distribution and its applications. Communications in Statistics - Theory and methods, 31(2), 497 - 512.

Gupta, P. L. and Gupta, R. C. (1983). On the moments of residual life in reliability and some characterization results. Communications in Statistics-Theory and Methods, 12(2), 449-461.

Gupta, R. C., Gupta, P. L., & Gupta, R. D. (1998). Modeling failure time data by Lehman alternatives. Communications in Statistics-Theory and methods, 27(2), 887-904.

Gupta RD, Kundu D (2009) A new class of weighted exponential distributions. Statistics 43(6):621 634.

Kenney, J. F. and Keeping, E. (1962). Mathematics of Statistics. D. Van Nostrand Company.

Kharazmi, O., Saadatinik, A. (2016). Hyperbolic cosine-F family of distributions with an application to exponential distribution. Gazi Univ J Sci 29(2):811–829.

Kharazmi, O, Saadatinik, A. (2018). Hyperbolic sine-F families of distributions with an application to exponential distribution. submitted. Gazi Univ J Sci.

Kumar, P., Sharma, R., Ray, S., Mehariya, S., Patel, S. K., Lee, J. K., and Kalia, V. C. (2015). Dark fermentative bioconversion of glycerol to hydrogen by Bacillus thuringiensis. Bioresource technology, 182, 383-388.

Marshall, A. W., & Olkin, I. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. Biometrika, 84(1), 641-652.

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

Nadarajah, S., Cancho, V. G., & Ortega, E. M. (2013). The geometric exponential Poisson distribution. Statistical Methods & Applications, 22(3), 355-380.

Saeed E. Hemeda, Ali M. Abdallah, Sudhansu S. Maiti and Sukanta Pramanik (2019). Generalized Inverted Weibull-G Family of Distributions with Application to Life Time Data, Far East Journal of Mathematical Sciences (FJMS) 112 (2), 237-266.

Rényi, A. (1961). On measures of entropy and information, Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability.