Journal of Statistics Applications & Probability Letters
An International Journal

http://dx.doi.org/10.18576/jsapl/090305

Epsilon-Skew Rayleigh: Estimation and Application

John M. Greene[1,*] *and Hassan Elsalloukh*[2]

[1]School of Business, Henderson State University, 1100 Henderson Ave., Arkadelphia, AR, 71999, U.S.
[2]Mathematics and Statistics Department, University of Arkansas at Little Rock, 2801 S University Ave., Little Rock, AR, 72204, U.S.

Received: 26 Oct. 2021, Revised: 22 Jan. 2022, Accepted: 12 Jul. 2022
Published online: 1 Sep. 2022

Abstract: In this paper, we develop the maximum likelihood estimators for the parameters of the Epsilon-Skew Rayleigh (ESR) distribution and demonstrate its ability to model population data. Maximum likelihood estimators are derived numerically and algebraically whenever it is possible. Application of the ESR is demonstrated by using numerical estimators to model 2015 demographic data from the Philippines. The fit of the model is evaluated and compared to the fit of a similar model.

Keywords: Rayleigh Distribution, Epsilon-Skew, Bimodal, Distribution theory

1 Introduction

Over the last two decades, there has been a growing interest in developing a flexible parametric asymmetric and bimodal distribution that can account for skewness and bimodality. In [1], [2], [3], and [4] a new parametric exponential asymmetric distribution called the Epsilon-Skew Exponential Power distribution was developed that can be used on skewed data. Also, in [5] and [6], a skew bimodal family distribution was developed and later used to analyze skewed and bimodal data in [7] and [8]. Subsequently, [9] and [10] developed the Skewed Double Inverted Weibull distribution and [11] The Epsilon-Skew Exponentiated Beta distribution.

Most recently, in [12] we introduced the Epsilon-Skew Rayleigh (ESR) distribution and developed its properties. It is a bimodal and skewed distribution with location, shape and skew parameters. In this paper, we will determine the log-likelihood function of the ESR and develop the maximum likelihood estimator (MLE) of each parameter. Applying numerical methods, we describe the R code used to generate ESR-distributed data sets and estimate the simulation parameters. We then use the ESR to model population age data from The Philippines 2015 census [13] and perform a comparison of the results to those of a similar model.

2 The Epsilon-Skew Rayleigh Distribution

2.1 The Probability Density Function of the ESR

Definition 1. *The random variable X has ESR distribution denoted by $X \sim ESR(\theta, \sigma, \varepsilon)$, if there exist parameters $\theta \in \mathbb{R}$, $\sigma > 0$, and $-1 < \varepsilon < 1$ such that the pdf of X is*

$$f(x;\theta,\sigma,\varepsilon) = \frac{1}{2\sigma^2} \begin{cases} \frac{x-\theta}{1-\varepsilon} \exp\left(-\frac{1}{2\sigma^2}\left(\frac{x-\theta}{1-\varepsilon}\right)^2\right) & x \geq \theta \\ \frac{\theta-x}{1+\varepsilon} \exp\left(-\frac{1}{2\sigma^2}\left(\frac{\theta-x}{1+\varepsilon}\right)^2\right) & x < \theta \end{cases}, \tag{1}$$

where θ, σ, and ε are location, scale, and skewness parameters respectively.

* Corresponding author e-mail: greenej@hsu.edu

3 Maximum Likelihood Estimation

3.1 Log-Likelihood Function of the ESR

Lemma 1. *The log-likelihood function of the ESR is given by*

$$\begin{aligned}\log L &= -nlog\,(2) - nlog\left(\sigma^2\right) - n_1 log\,(1-\varepsilon) - n_2 log\,(1+\varepsilon) + \sum_{i=1}^{n_1}\log\left(x_{1i}-\theta\right) \\ &+ \sum_{j=1}^{n_2}\log\left(\theta - x_{2j}\right) - \left(\frac{\sum_{i=1}^{n_1}\left(x_{1i}-\theta\right)^2}{2\sigma^2\left(1-\varepsilon\right)^2}\right) - \left(\frac{\sum_{j=1}^{n_2}\left(\theta-x_{2j}\right)^2}{2\sigma^2\left(1+\varepsilon\right)^2}\right) \quad ,\end{aligned} \tag{2}$$

where $x_{1i} \geq \theta$ *for* $i: 1,2,...,n_1$ *and* $x_{2j} < \theta$ *for* $j: 1,2,...,n_2$.

Proof. Let X $\sim$ ESR$(\theta,\sigma,\varepsilon)$ from (1). Let n_1 be the number of observations such that $x_{1i} \geq \theta$ and n_2 be the number of observations such that $x_{2j} < \theta$, with $n_1 + n_2 = n$. The likelihood function, L, of this sample is given by

$$L = \prod_{i=1}^{n_1}\frac{1}{2\sigma^2}\frac{x_{1i}-\theta}{1-\varepsilon}\exp\left(-\frac{1}{2\sigma^2}\left(\frac{x_{1i}-\theta}{1-\varepsilon}\right)^2\right)\prod_{j=1}^{n_2}\frac{1}{2\sigma^2}\frac{\theta-x_{2j}}{1+\varepsilon}\exp\left(-\frac{1}{2\sigma^2}\left(\frac{\theta-x_{2j}}{1+\varepsilon}\right)^2\right).$$

Applying the logarithm and simplifying completes the proof.

3.2 Estimation of the Skew Parameter

Lemma 2. *The maximum likelihood estimator of the skew parameter is given by*

$$\widehat{\varepsilon}_{x\geq\theta} = 1 - \sqrt{\frac{\sum_{i=1}^{n_1}\left(x_{1i}-\theta\right)^2}{n_1\sigma^2}} \text{ and } \widehat{\varepsilon}_{x<\theta} = -1 + \sqrt{\frac{\sum_{j=1}^{n_2}\left(\theta-x_{2j}\right)^2}{n_2\sigma^2}} \tag{3}$$

Proof. Let X $\sim$ ESR$(\theta,\sigma,\varepsilon)$ from (1). Let n_1 be the number of observations such that $x_{1i} \geq \theta$ and n_2 be the number of observations such that $x_{2j} < \theta$, with $n_1 + n_2 = n$. Differentiating (2) with respect to ε for each case, we obtain

$$\frac{\partial}{\partial\varepsilon}\log L_{x\geq\theta} = \frac{n_1}{1-\varepsilon} - \frac{\sum_{i=1}^{n_1}\left(x_{1i}-\theta\right)^2}{\sigma^2}\left(\frac{1}{\left(1-\varepsilon\right)^3}\right) \text{ and } \frac{\partial}{\partial\varepsilon}\log L_{x<\theta} = \frac{n_2}{1+\varepsilon} - \frac{\sum_{j=1}^{n_2}\left(\theta-x_{2j}\right)^2}{\sigma^2}\left(\frac{1}{\left(1+\varepsilon\right)^3}\right).$$

Equating each expression to zero and solving for ε, the lemma is proved for both cases.

3.3 Estimation of the Scale Parameter

Lemma 3. *The maximum likelihood estimator of the scale parameter is given by*

$$\widehat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n_1}\left(x_i-\theta\right)^2}{2n(1-\varepsilon)^2} + \frac{\sum_{i=1}^{n_2}\left(\theta-x_i\right)^2}{2n\left(1+\varepsilon\right)^2}} \quad . \tag{4}$$

Proof. Let $x_1, x_2, \ldots, x_n$ be a sample of n observations of X $\sim$ ESR$(\theta,\sigma,\varepsilon)$. Differentiating (2) with respect to σ, we obtain

$$\frac{\partial}{\partial\sigma}\log L = -n\frac{\partial}{\partial\sigma}log\left(\sigma^2\right) - \frac{\sum_{i=1}^{n_1}\left(x_i-\theta\right)^2}{2\left(1-\varepsilon\right)^2}\frac{\partial}{\partial\sigma}\left(\frac{1}{\sigma^2}\right) - \frac{\sum_{i=1}^{n_2}\left(\theta-x_i\right)^2}{2\left(1+\varepsilon\right)^2}\frac{\partial}{\partial\sigma}\left(\frac{1}{\sigma^2}\right).$$

Equating this expression to zero and solving for σ completes the proof.

3.4 Estimation of the Location Parameter

Differentiating (2) with respect to θ and setting the expression equal to zero, we arrive at the following equation, which lacks a closed-form solution.

$$0 = \sum_{i=1}^{n_1} \frac{1}{x_{1i} - \theta} + \sum_{j=1}^{n_2} \frac{1}{\theta - x_{2j}} + \left(\frac{1}{\sigma^2 (1-\varepsilon)^2} \right) \sum_{i=1}^{n_1} (x_{1i} - \theta) - \left(\frac{1}{\sigma^2 (1+\varepsilon)^2} \right) \sum_{j=1}^{n_2} (\theta - x_{2j})$$

Estimation of θ can be performed visually from a histogram. Furthermore, in [12] we showed that the modes of the distribution are given by

$$m_{X<\theta} = \theta - (1+\varepsilon)\sigma \;\; and \;\; m_{X\geq\theta} = \theta + (1-\varepsilon)\sigma \, .$$

Combining these equations and solving for θ we arrive at an estimator which is a function of the skew parameter, as derived above, and the modes, which can be easily measured from the data.

$$\widehat{\theta} = \left(\frac{1-\widehat{\varepsilon}}{2} \right) m_{X<\theta} + \left(\frac{1+\widehat{\varepsilon}}{2} \right) m_{X\geq\theta} \tag{5}$$

For each parameter, the second derivative of the log-likelihood function is negative at the solution of the first derivative, hence the maximum for each MLE exists.

4 Numerical Methods

4.1 Introduction

We chose R for coding the density, probability, quantile, and random vector functions for the ESR distribution as well as for the estimation functions for the skew, scale and location parameters. The density and random vector functions are displayed simultaneously in Figure 1.

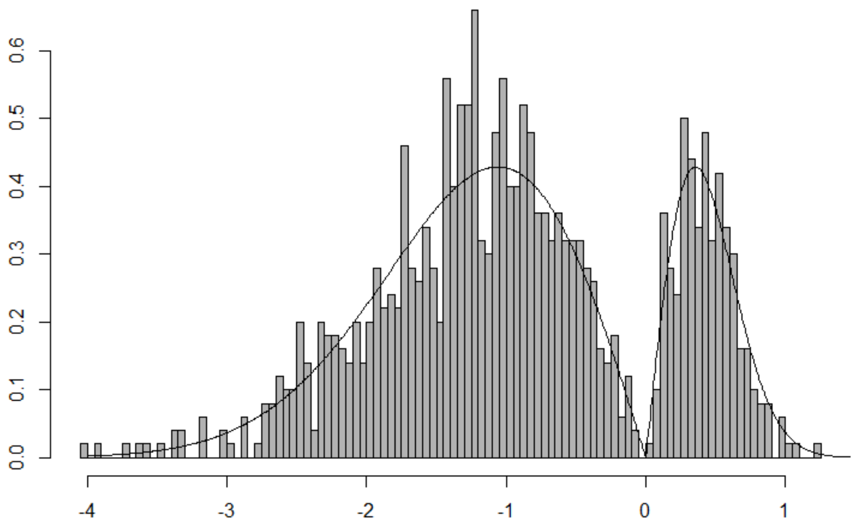

Fig. 1: ESR pdf and random vector function with n = 1000, θ = 0, σ = 0.7, ε = 0.5 values.

While no closed-form system of MLE equations for the ESR parameter vector exists, the values can be numerically estimated from the data.

4.2 Numerical Estimation of Parameters

An initial estimation for the location parameter, θ, can be made visually or numerically obtained from the histogram. We showed in [12] that the probabilities to the left and right of the location parameter are given by $(1+\varepsilon)/2$ and $(1-\varepsilon)/2$ respectively, where ε is the skewness parameter. Therefore, the proportion of data values on either side of $\hat{\theta}$ can be used to determine $\hat{\varepsilon}$. The location and skew estimators allow us to use (4) to estimate the scale parameter, σ. At this point (5) can be used to improve upon our initial estimate of θ. Finally, we use (3) to refine our estimate of ε. Further iterations quickly produce stable values for all of the estimators.

4.3 Simulation

ESR-distributed data sets were created using the random vector function with n=1000 and the numerical estimation functions were applied. Two of these data sets and the resulting estimators are displayed in Table 1.

Table 1: A comparison of simulated ESR data to numerically estimated parameters.

	n	Location	Shape	Skew
Simulation 1	1000	$\theta = 0$	$\sigma = 0.5$	$\varepsilon = 0.5$
Estimation 1		$\widehat{\theta} = 0.020$	$\widehat{\sigma} = 0.490$	$\widehat{\varepsilon} = 0.490$
Simulation 2	1000	$\theta = 1.6$	$\sigma = 0.42$	$\varepsilon = -0.34$
Estimation 2		$\widehat{\theta} = 1.588$	$\widehat{\sigma} = 0.419$	$\widehat{\varepsilon} = -0.372$

5 Application

5.1 Philippine Census Data

The Philippines performs a national census every five years. The national population has surpassed 100 million according to the 2015 census which produced a count of 100,979,303 people. This was up from the count of 92,337,852 in 2010 [13]. The detailed information which the government collects includes the parameters of gender and age, reported with a class width of four years. The 2015 data was presented graphically using an opposing bar chart with male numbers on the left and female numbers on the right. Each side appears similar in shape to that of the Rayleigh distribution. Figure 2 contains a relative frequency histogram of the data with male age classes labeled with negative integers and female age classes with positive integers. So displayed, it appears to be approximately ESR distributed.

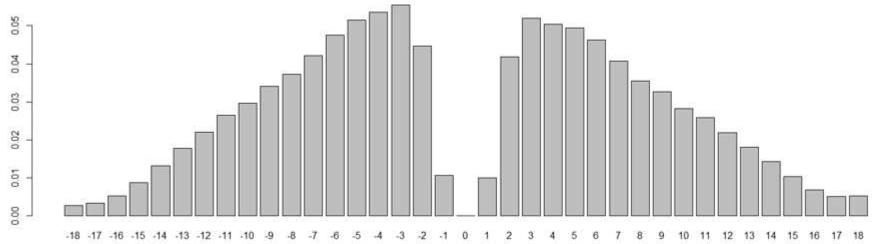

Fig. 2: Relative Frequencies of the 2015 Philippine population by reported age group and gender.

In order to evaluate the effectiveness of applying the ESR model to real data, a comparison was made with another bimodal, three-parameter model, the Bimodal Normal distribution as it was presented in [14] as a further development of its original presentation in [15]. According to [14], a real-valued random variable X is said to have a Bimodal Normal (BN) distribution if it has density

$$f(x;\mu,\sigma,\alpha) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 - \frac{\alpha^2}{2}\right] \cosh\left[\alpha\left(\frac{x-\mu}{\sigma}\right)\right],$$

where $\mu \in \mathbb{R},\ \sigma > 0,\ \alpha \in \mathbb{R}$ are location, scale, and a parameter that controls the uni- or bimodality of the distribution, respectively. The results of the ML approach for fitting the two models to the data are given in Table 2.

Table 2: The ML estimates with Cramér–von Mises and Anderson–Darling statistics for the two models.

Model	ML Estimates	Cramér–von Mises CV(.05,37)=0.218	Anderson–Darling CV(.05,37)=0.757
BN	$\hat{\mu} = 0.053,\ \hat{\sigma} = 4.131,\ \hat{\alpha} = 1.5595$	0.0091	0.1597
ESR	$\hat{\theta} = 0.053,\ \hat{\sigma} = 5.499,\ \hat{\varepsilon} = 0.0042$	0.0062	0.0935

The ESR model presents the best fit, having resulted in smaller Cramér–von Mises and Anderson–Darling values. This finding is also corroborated by the graph in Figure 3, which displays the densities of both fitted models with marks representing the relative frequency values of the data classes.

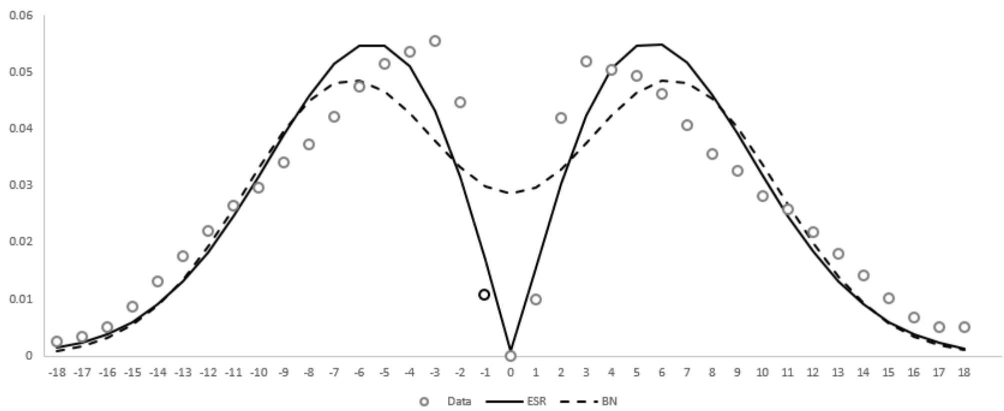

Fig. 3: Models fitted by the ML approach to the 2015 Philippine population data: ESR(solid line) and BN(dashed line).

We now examine the sum of the squared ordinary residuals for each model. The sum is less for the ESR that for the BN, supporting our previous findings. An F test of the sums is presented in Table 3.

Table 3: F test of the sums of squared ordinary residuals of the models.

Model	SS	Critical $F_{.05}(34,34)$	F = SS1/SS2	p-value
1) Bimodal Normal	0.00307	1.7721	2.1379	0.0149
2) Epsilon-Skew Rayleigh	0.00144			

The null hypothesis, model 1 is at least as appropriate as model 2, is rejected. The Epsilon-Skew Rayleigh distribution is therefore a significantly better model for this population data than is the Bimodal Normal distribution.

6 Conclusion

In this paper we have studied further properties of the previously introduced Epsilon-Skew Rayleigh Distribution. We successfully developed closed-form maximum likelihood estimators for the shape and skew parameters. We demonstrated the use of R coded functions for the generation of values from an ESR-distributed random variable and for obtaining numerically derived estimations for the location, shape, and skew parameters from a given data set. The effectiveness of these numerical methods were applied to a real population data set and the results were compared to those produced by a different bimodal three-parameter distribution. Both distributions provided satisfactory models according to their Cramér–von Mises and Anderson–Darling statistics, with the ESR performing better than the compared model. The results of an F test on the sum of squared ordinary residuals concluded that the ESR provided a significantly better fit for the data than did the compared model. This comparative analysis demonstrated the effective applicability of the Epsilon-Skew Rayleigh Distribution for modeling skewed bimodal data.

Future work will include development of the Epsilon-Skew Exponentiated Rayleigh Distribution which will provide improved flexibility for fitting both uni-modal and bimodal skewed data, allowing for unequal modal values. We would like to gratefully acknowledge the referees' suggestions which helped to improve this work.

References

[1] Elsalloukh, H., Guardiola, J., and Young, M. D. (2005). The Epsilon-Skew Exponential Power Distribution Family. Far East Journal of Theoretical Statistics, 16, No. 1, 97-112.
[2] Elsalloukh, H. (2008). The Epsilon-Skew Laplace Distribution. JSM Proceedings, Biometrics Section: The American Statistical Association Conference, 1174-1180.
[3] Elsalloukh, H. (2009), Further Results on the Epsilon-Skew Exponential Power Distribution Family. Far East Journal of Theoretical Statistics, 28, No. 1, 201-212.
[4] Al-Mousawi, H., Elsalloukh, H., and Guardiola, J. (2012), Further Analysis of the Epsilon Skew Laplace Distribution. Universal Journal of Mathematics and Mathematical Sciences, 2, No. 2, 181-194.
[5] Abdulah, E. and Elsalloukh, H. (2012), The Epsilon Skew Gamma Distribution. JSM Proceedings, Biometrics Section: The American Statistical Association Conference, 50-59.
[6] Abdulah, E. and Elsalloukh, H. (2012), The Epsilon Skew Inverted Gamma Distribution. JSM Proceedings, Biometrics Section: The American Statistical Association Conference, 42-49.
[7] Abdulah, E. and Elsalloukh, H. (2013), Analyzing Skewed Data with the Epsilon Skew Gamma Distribution, J. Stat. Appl. Pro. 2, No. 3, 195-202.
[8] Abdulah, E. and Elsalloukh, H. (2014), Bimodal Class Based on the Inverted Symmetrized Gamma Distribution with Applications, J. Stat. Appl. Pro. 3, No. 1, 1-7.
[9] Flaih, A., Elsalloukh, H., Mendi, E., and Milanova, M. (2012), The Skewed Double Inverted Weibull Distribution. Applied Mathematics and Information Sciences, 6, No.2, 269-274.
[10] Flaih, A., Elsalloukh, H., Mendi, E., and Milanova, M. (2012), The Exponentiated Inverted Weibull Distribution. Applied Mathematics and Information Sciences, 6, No.2, 167-171.
[11] Akmyradov, C. and Elsalloukh, H. (2019), The Epsilon-Skew Exponentiated Beta Distribution, J. Stat. Appl. Pro. Lett. 6, No. 3, 101-108.
[12] Greene, J.M. and Elsalloukh, H. (2021), The Epsilon-Skew Rayleigh Distribution, J. Stat. Appl. Pro. Lett. 8, No. 3, 153-158.
[13] Philippine Statistics Authority (2017), 2015 Census of Population, Report No. 2 – Demographic and Socioeconomic Characteristics Philippines.
[14] Vila, R., Saulo, H., and Roldan, J. (2021). On some properties of the bimodal normal distribution and its bivariate version. arXiv preprint arXiv:2106.00097.
[15] Gómez-Déniz, E., Sarabia, J. M., and Calderín-Ojeda, E. (2021). Bimodal normal distribution: Extensions and applications. Journal of Computational and Applied Mathematics, 388, 113292.