ARCSINE RAYLIEGH PARETO DISTRIBUTION: PROPERTIES AND APPLICATION TO CARBON FIBERS DATA SETS

1Usman Abubakar, 1Abdulhameed Ado Osi, 1Iliyasu Abubakar Salisu, 2Hassan Muhammad, 3Yusuf Ibrahim Muhammad and 3Abbas Abubakar

1Department of Statistics, Aliko Dangote University of Science and Technology Wudil, Kano state. Nigeria.  
2Department of Mathematics, Federal university of Technology Babura, jiggawa state. Nigeria.  
3Department of Statistics, Jigawa State Polytechnic Dutse, Jigawa State. Nigeria.

*Corresponding authors’ email: Usmanabubakar@Jigpoly.edu.ng

ABSTRACT
In this paper, we introduce a new modified distribution called arcsine Rayleigh Pareto (ASRP) Distribution. We derived its mathematical and statistical properties, including survival function, hazard function, entropy, moment, moment generating function, and order statistics. We also used maximum likelihood estimation for estimating the parameters of the distribution. The plots of the cdf, pdf, hazard rate function, and survival function were illustrated with right skewed probability density function, cumulative distribution function with monotone increasing function, which converge at one. The value of some goodness of fit measure (i.e. AIC, AICc, and BIC) were computed, as well as the KS, A, and W statistic. Finally, we suggested that the new modified model outperform better than the other standard distribution using Carbon fibers data sets.

Keywords: Arcsine Rayleigh Pareto distribution, Properties, Maximum likelihood estimation, Application

INTRODUCTION
In general, modeling data from financial, biomedical, engineering, and lifelong testing studies is a significant area of study (Ahmad et al., 2022). (Abubakar et al., 2024) claims that since there are still many scenarios in which none of the traditional or classical probability models can sufficiently describe the actual data, it is imperative to propose new models that can more faithfully capture the real-life phenomena found in a given dataset. In many situations, some researchers focus on modifying an existing probability model in order to increase its performance, versatility, and flexibility to handle the data that an existing distribution cannot fit well. This happens, especially when the data is skewed either to the right or to the left.

The proposed Arcsine rayleigh pareto (ASRP) distribution is a modification to the Rayleigh pareto distribution introduced by (Al-Kadim and Mohammed, 2018), and similar methods in many researches, including Composite Rayleigh Pareto Distribution by (Benatmane et al., 2021), Arcsine Kumaraswamy generalized family by (Emam and Tashkandy, 2022), Sine topp-Leone generalized family of distribution by (Al-Baitain et al., 2020), cosine Topp-Leone family of distributions by (Nanga et al., 2023), Tangent Topp-Leone G family of distributions by (Nanga et al., 2022), among others.

(Ali-Kadim and Mohammed, 2018) provide several mathematical properties, quantile, reliability function, and parameter estimation using maximum likelihood methods. The cumulative distribution function and probability density function of the Rayleigh Pareto distributions were given by (Al-Kadim and Mohammed, 2018) as;

\[ F(y) = 1 - \exp \left( -\frac{z^2}{2b^2} \right) \]  

The above distribution can be modified as ASRP distribution, and Such modification was used by (Emam and Tashkandy, 2022). The cumulative distribution of Arcsine Rayleigh Pareto Distribution is;

\[ G(z,a,b,c) = \frac{z}{\pi} \arcsine \left( 1 - \exp \left( -\frac{z^2}{2b^2} \right) \right) \]  

After differentiating equation (2) and some further simplification, the probability density function is derived as;

\[ g(z,a,b,c) = \frac{az^{a-1}\exp\left(\frac{-z^2}{2b^2}\right)}{\left(1 - \exp\left(\frac{-z^2}{2b^2}\right)\right)^2} \]  

The plot of the cdf and pdf of ASRP Distribution for some selected parameters is illustrated in the figures below.

Figure 1: The plot of the pdf of ASRP Distribution for some selected parameters.
Some Statistical Properties of ASRP Distribution

Some properties of ASRP Distribution, which comprises the survival function, Hazard function, Entropy and quantile.

\[ S(z) = 1 - \frac{2}{\pi} \arcsine \left( 1 - \exp \left( \frac{-(z \cdot c)}{2b^2} \right) \right) \] (4)

\[ H(z) = \frac{a \pi^2 z - 1}{\exp \left( \frac{- (z \cdot c)^2}{2b^2} \right) \left( 1 - \frac{\pi^2 z \arcsine \left( 1 - \exp \left( \frac{-(z \cdot c)^2}{2b^2} \right) \right)}{2b^2} \right)} \] (5)

\[ G^{-1}(m) = C \left( -2b^2 \log \left( 1 - \sin \left( \frac{\pi m}{2} \right) \right) \right)^{\frac{1}{2}} \] (6)

The Renyi entropy, moment, moment generating function, and Order Statistics of ASRP Distribution are:

\[ I_\phi(z) = \frac{1}{1 - \phi} \log \int_{-\infty}^{\infty} g(z)^\phi \, dz \] (7)

\[ g(z)^\phi = (\mu \eta)^\phi \] (8)

Where \[ \eta = \frac{a \pi^2 z - 1}{\exp \left( \frac{- (z \cdot c)^2}{2b^2} \right) \left( 1 - \frac{\pi^2 z \arcsine \left( 1 - \exp \left( \frac{-(z \cdot c)^2}{2b^2} \right) \right)}{2b^2} \right)} \]
And
\[ \mu = \frac{a}{\pi b^2 c^a} \]
Implies that, the entropy of ASRP Distribution is
\[ I_s(z) = \frac{1}{1-q} \log \int_{0}^{\infty} (\eta)^q \, d\eta \]
\[ = \frac{1}{1-q} [\log \mu + \log \int_{0}^{\infty} \eta^q \, d\eta] \]  

The nth moment of a continuous distribution is given by;
\[ \mu^n = E(x^n) = \int_{-\infty}^{\infty} x^n g(x, a, b, c) \, dx \]  

The moment generating function \( M(t) \) is, 
\[ M(t) = E(e^{tx}) = \int_{0}^{\infty} e^{tx} g(x, a, b, c) \, dx \]  

For Order Statistics, Let \( z_1, z_2, \ldots, z_n \) be a random sample from the ASRP distribution and let \( z(1), \ldots, z(n) \) be the corresponding order statistics. The nth order statistic's pdf is written as
\[ g_{(i,n)}(z) = \frac{n(i-1)}{(i-1)(n-i-1)!} \sum_{j=0}^{n-i} g(z(j))^j \frac{n!}{(n-j)!} \]  

Where
\[ g(z) = \frac{az^{\alpha+1-1} \exp(-\frac{z}{\beta})}{\beta} \left( 1 - \exp(-\frac{z}{\beta}) \right) \]  

\[ \frac{\alpha}{\beta} \]  

Parameter Estimation Using Maximum Likelihood
To estimate the parameters of the ASRP distribution using a complete sample, let \( z_1, z_2, \ldots, z_n \) be a random sample from the ASRP distribution. The log-likelihood of the parameter vector is written as;
\[ \log(l(a, b, c)) = \log \prod_{i=1}^{n} g(z(i), a, b, c) \]  

\[ \log l(a, b, c) = n \log a + (a - 1) \log z + \frac{1}{2 \beta^2} \sum_{i=0}^{\infty} \left( \frac{z}{\beta} \right)^a - n \log \pi - 2 n \log b - \alpha \log c + \]
\[ \frac{1}{2} \sum_{i=0}^{\infty} \log \left( 1 - \left( 1 - \exp \left( -\frac{z}{\beta} \right) \right) \right) \]  

\[ \frac{d}{dz} \log \pi = \frac{a}{n} \sum_{i=0}^{\infty} \log z - \frac{1}{2 \beta^2} \sum_{i=0}^{\infty} x^a \log \left( \frac{z}{\beta} \right) - \log c + \]
\[ \sum_{i=0}^{\infty} z^i \left( 1 - \exp \left( \frac{-c^n}{2b^2} \right) \right) \frac{\log(z)}{z} \exp \left( \frac{-c^n}{2b^2} \right) \] (24)

\[ \frac{dL}{db} = \frac{1}{b^2c^2} \sum_{i=0}^{\infty} z^i \frac{2n}{b} = \frac{2n}{b} \sum_{i=0}^{\infty} \frac{z^i \left( 1 - \exp \left( \frac{-c^n}{2b^2} \right) \right) \exp \left( \frac{-c^n}{2b^2} \right)}{b^2c^2} \] (25)

### Application to Carbon Fibers Data Sets

The real data set was used by (Selim and Badr, 2016). The data represent the strength data measured in GPA, for single carbon fibers were tested under tension at gauge lengths of 1, 10, 20 and 50 mm. For illustrative purpose, we consider only the data set consisting the single fibers of 20 mm, with a sample of size 63. The data are:


### Table 1: Summary of the carbon fibers data sets

<table>
<thead>
<tr>
<th>Min</th>
<th>Q(1)</th>
<th>Median</th>
<th>Mean</th>
<th>Q(3)</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.901</td>
<td>2.554</td>
<td>2.996</td>
<td>3.059</td>
<td>3.421</td>
<td>5.020</td>
</tr>
</tbody>
</table>

To compare the fitted model with some standard models using the carbon fibers and bladder cancer Data sets, we used methods of goodness of fit, which comprises the log-likelihood function evaluated at the MLEs, the Akaike information criterion (AIC), the Bayesian information criterion (BIC), the Akaike information corrected criterion (AICc), Anderson-Darling (A*), Cramer-von Mises (W*), and Kolmogorov-Smirnov (K-S*) (Aryal et al., 2017).

Generally speaking, the smaller the values of these statistics, the better the fit (Al-Kadim and Mohammed, 2018).

\[ AIC = -2L + 2k \] (26)

\[ BIC = -2L + k \log(n) \] (27)

And \[ AIC_c = -2L + \frac{2kn}{n-k-1} \] (28)
Table 2: Competitors model

<table>
<thead>
<tr>
<th>Distributions</th>
<th>Author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rayleigh Pareto</td>
<td>(Al-Kadim and Mohammed, 2018)</td>
</tr>
<tr>
<td>Exponentiated Weibull</td>
<td>(Shawky and Abu-Zinadah, 2009)</td>
</tr>
<tr>
<td>Exponentiated Pareto</td>
<td>(Nadarajah, 2005)</td>
</tr>
</tbody>
</table>

Table 3: Goodness of fit measure and estimate of the parameters using MLE

<table>
<thead>
<tr>
<th>Distr.</th>
<th>(a^*)</th>
<th>(b^*)</th>
<th>(c^*)</th>
<th>LL</th>
<th>AIC</th>
<th>AICc</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASRP</td>
<td>5.330</td>
<td>2.114</td>
<td>1.970</td>
<td>-18.29134</td>
<td>42.58267</td>
<td>42.98945</td>
<td>49.01208</td>
</tr>
<tr>
<td>RP</td>
<td>0.002471</td>
<td>4.801721</td>
<td>93.189154</td>
<td>-24.88366</td>
<td>55.76733</td>
<td>56.17411</td>
<td>62.19673</td>
</tr>
<tr>
<td>EXPW</td>
<td>0.7878</td>
<td>1.54616</td>
<td>28.96688</td>
<td>-56.31814</td>
<td>118.6363</td>
<td>119.0431</td>
<td>125.0657</td>
</tr>
<tr>
<td>EXPP</td>
<td>34.2529</td>
<td>2.7867</td>
<td>-92.27645</td>
<td>188.5529</td>
<td>188.7529</td>
<td>192.8392</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Normality test for the fitted models

<table>
<thead>
<tr>
<th>Distribution</th>
<th>KS</th>
<th>(A)</th>
<th>(W)</th>
<th>(P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASRP</td>
<td>0.085799</td>
<td>0.87734</td>
<td>0.12667</td>
<td>0.7426</td>
</tr>
<tr>
<td>RP</td>
<td>0.97151</td>
<td>0.32514</td>
<td>0.056579</td>
<td>2.2 \times 10^{16}</td>
</tr>
<tr>
<td>EXPW</td>
<td>0.079744</td>
<td>0.032222</td>
<td>0.060305</td>
<td>0.8179</td>
</tr>
<tr>
<td>EXPP</td>
<td>0.2703</td>
<td>0.32758</td>
<td>0.062385</td>
<td>0.0002009</td>
</tr>
</tbody>
</table>

The summary of carbon fibers data sets in Table 1 and Figure 5 illustrates the right-skewness nature of the data. By considering the values in Table 1, we can observe that the mean is greater than the median of the distribution. Likewise, in Table 3, the likelihood of the modified Arcsine Rayleigh Pareto distribution is higher than for other standard distributions. Also, the goodness of fit measure for AIC, BIC, and AICC for ASRP is smaller than that of other distributions, which shows the better performance of ASRP.

**CONCLUSION**

In this paper, we introduce a new modified distribution called the Arcsine Rayleigh pareto distribution, which is a modification of the Rayleigh pareto distribution. We examined some of its statistical and mathematical properties, as well as the estimation of parameters using the maximum likelihood method. Lastly, we suggested that the ASRP distribution is a better one for modeling right-skewed data sets, which also perform better when compared with other competitors models.

**REFERENCES**


©2024 This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license viewed via [https://creativecommons.org/licenses/by/4.0/](https://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is cited appropriately.