The Exponentiated Power Shanker Distribution With Application

Nofiu Idowu Badmus[1*], Amaury de Souza[2], Olanrewaju Faweya[3]

[1]Department of Statistics, University of Lagos, Akoka, Yaba, Nigeria
[2]Federal University of Mato Grosso do Sul, Campo Grande, Brazil
[3]Department of Statistics, Ekiti State University, Ekiti State, Nigeria
*Email: nibadmus@unilag.edu.ng

Abstract

A new exponentiated power Shanker distribution is proposed, and its statistical properties are thoroughly discussed. The distribution parameters are estimated using the maximum likelihood estimation method. The new distribution is then extended to a regression model by applying the logarithmic transformation. The proposed regression model is applied to a marriage dissolution data set consisting of 568 entries, where the response variable is the number of years in the marriage and the predictor variables are the husband's education level, husband's race (black or not), marital mixing (mixed or not), and divorce status. The results reveal that some predictors significantly influence the duration of marriage. The estimated regression model is given by $year(\hat{x}) = 0.0809 - 1.6190(Heduc) + 0.4343(Heblack) - 1.0747(Mixed) + 1.1819(Divorce)$. The model selection criteria for the propose model yield the following values: Akaike information criterion = 8073.054, Bayesian information criterion = 8086.080, consistent Akaike information criterion = 8087.080, and Hannan-Quinn information criterion = 8078.137. Furthermore, a machine learning-based multiple regression model is conducted to estimate mean absolute error, mean squared error, and root mean squared error, which are 10.1138, 162.1605, and 12.7342, respectively. The results are compared with those obtained from a traditional linear regression model, and findings suggest that the machine learning-based multiple regression model outperforms the linear regression model, with smaller error values.

Keywords: transformation, machine learning, regression model, response, predictor, mean absolute error, mean squared error

Introduction

In statistical modeling, several methods exist for combining two or more distributions to create a hybrid distribution. This study adopts the exponentiated-G family of distributions introduced by Gupta et al. (1998) to develop a new distribution called the exponentiated power Shanker (EPS) distribution. The EPS distribution results from combining the power Shanker (PS) distribution, studied by Shanker and Shukla (2017a), with the exponential distribution. The motivation for proposing the EPS distribution is that it is a three-parameter distribution with sufficient flexibility to model lifetime data of diverse types, including those from medical sciences, engineering, economics, insurance, finance, and educat-ion. By contrast, Shanker and Hagos (2016) noted that several one- or two-parameter distributions, such as the Akash, Shanker, Lindley, and exponential distributions, are monot-onically decreasing and lack the flexibility needed to adequately handle and fit various lifetime data sets.

Several authors have applied the exponentiated-G distribution in combination with other distributions in the literature. For example, Ashour and Eltehiwy (2015) studied the exponentiated power Lindley distribution, Shanker and Prodhani (2022) discussed the exponentiated Sujatha distribution, Elgarhy and Shawki (2017) examined the exponentiated Sushila distribution, Aafaq and Subramania (2019) introduced the exponentiated Ishita distribution, and Badmus and Faweya (2022) proposed exponentiated new modified weighted Rayleigh distribution with application. In addition, Badmus and Aromolaran (2024) investiga-ted the exponentiated Sujatha distribution and its applications, while Badmus, Olufolabo, and Akingbade (2021) examined a two-parameter exponential Akash distribution. Similarly, Okereke and Uwaeme (2018) analyzed the exponentiated Akash distribution. Furthermore, Shanker and Shukla (2017a, 2017b) presented the PS distribution and the power Akash distribution, among others.

The study extends the exponentiated power Sujatha distribution to a regression model due to its strong ability to fit diverse lifetime data. This extension is achieved through a transformation approach, in which the logarithm of the response variable is taken and the parameters of the EPS distribution are re-parameterized. Many researchers have employed this transformation method to develop regression models from probability distributions. For instance, Nwankwo et al. (2023) studied the exponentiated power Akash distribution and its applications to health and disease data sets; Ferreira and Cordeiro (2023) investigated the exponentiated power Ishita distribution through properties, simulations, and regression analysis; and Shabani et al. (2018) examined the exponentiated power Lindley logarithmic distribution, among others.

The paper is organized as follows. Section 2 presents the materials and methods used to obtain the EPS distribution and its properties. Parameter estimation is discussed in Section 3. In Section 4, we introduce the logarithmic form of the EPS distribution. Section 5 applies the proposed model to marriage dissolution data and compares its fit with other competing models. Finally, Section 6 provides a comparison between machine learning and least-squares models, followed by discussion and conclusions.

Materials and Methods

The EPS Distribution

The probability density function (pdf) and cumulative density function (cdf) of the power Shanker according to Shanker and Shukla (2017a) are given by

$$f(y;\theta,\alpha)=\frac{\alpha\theta^2}{(\theta^2+1)}y^{\alpha-1}\left(\theta+y^{\alpha}\right)e^{-\theta y^{\alpha}},$$

$$y,\theta,\quad and\quad \alpha>0 \tag{1}$$

and

$$F(y;\theta,\alpha)=1-\left[1+\frac{\theta y^{\alpha}}{\theta^2+1}\right]e^{-\theta y^{\alpha}},$$

$$y,\theta,\quad and\quad \alpha>0 \tag{2}$$

Then, the cdf and pdf of the exponentiated-G distribution with power parameter $k>0$ are given by

$$F(y;k,\Sigma)=G\left(y;\Sigma\right)^{k} \tag{3}$$

and

$$f(y;k,\Sigma)=kg\left(y;\Sigma\right)G\left(y;\Sigma\right)^{k-1} \tag{4}$$

By substituting Equation 2 in Equation 3, and Equations 1 and 2 in Equation 4, we obtained both the cdf and pdf of the proposed EPS distribution respectively as follows:

$$F_{EPS}(y;k,\theta,\alpha) = \left\{1 - \left[1 + \frac{\theta y^{\alpha}}{\theta^2+1}\right] e^{-\theta y^{\alpha}}\right\}^k,$$

$$y, k, \theta \quad and \quad \alpha > 0 \tag{5}$$

and

$$f_{EPS}(y;\theta,\alpha) = \frac{k\alpha\theta^2}{(\theta^2+1)} y^{\alpha-1}\left(\theta + y^{\alpha}\right) e^{-\theta y^{\alpha}} \times$$

$$\left\{1 - \left[1 + \frac{\theta y^{\alpha}}{\theta^2+1}\right] e^{-\theta y^{\alpha}}\right\}^{k-1},$$

$$y, k, \theta \quad and \quad \alpha > 0 \tag{6}$$

where k and α are shapes and θ is a scale parameter. In the same manner, the reliability and the failure rate functions are also obtained as follows:

$$Rel_{EPS}(y;k,\theta,\alpha) = 1 - G\left(y;\Sigma\right)^k$$

$$= 1 - \left\{1 - \left[1 + \frac{\theta y^{\alpha}}{\theta^2+1}\right] e^{-\theta y^{\alpha}}\right\}^k \tag{7}$$

and

$$Fal_{EPS}(y;k,\theta,\alpha) = \frac{kg\left(y;\Sigma\right) G\left(y;\Sigma\right)^{k-1}}{1 - G\left(y;\Sigma\right)^k}$$

$$Fal_{EPS}(y;k,\theta,\alpha) = \frac{f_{EPS}(y;k,\theta,\alpha)}{Rel_{EPS}(y;k,\theta,\alpha)}$$

$$= \frac{k\alpha\theta^2}{(\theta^2+1)} y^{\alpha-1}\left(\theta + y^{\alpha}\right) e^{-\theta y^{\alpha}} \times$$

$$\frac{\left\{1 - \left[1 + \frac{\theta y^{\alpha}}{\theta^2+1}\right] e^{-\theta y^{\alpha}}\right\}^{k-1}}{1 - \left\{1 - \left[1 + \frac{\theta y^{\alpha}}{\theta^2+1}\right] e^{-\theta y^{\alpha}}\right\}^k} \tag{8}$$

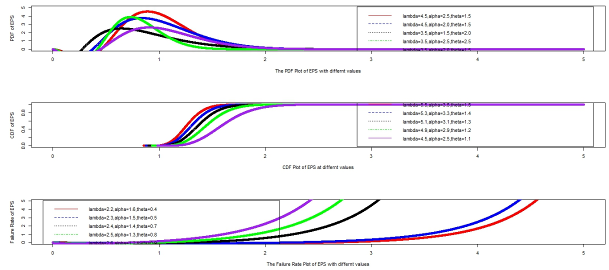

Figure 1. The density, distribution, and failure rate plots of the EPS distribution with different values of the parameters.

Some Properties of the EPS Distribution

Moment
The r^{th} moment of $Y \sim$ EPS distribution is given as follows:

$$\mu_r^{'} = \int_0^{\infty} y^r f_{EPS}(y) dy$$

$$\mu_r^{'} = \frac{k\alpha\theta^2}{(\theta^2+1)} \int_0^{\infty} y^r\left(\theta + y^{\alpha}\right) y^{\alpha-1} e^{-\theta y^{\alpha}}$$

$$\left\{1 - \left[1 + \frac{\theta y^{\alpha}}{\theta^2+1}\right] e^{-\theta y^{\alpha}}\right\}^{k-1} dy \tag{9}$$

Using binomial expansion, $\mu_r^{'}$ in Equation 11 is achieved from Equation 9 as follows:

$$(1-B)^{k-1} = \sum_{j=0}^{k-1} \binom{k-1}{j} (-1)^j B_j^j,$$

$$B = \left(1 + \frac{\theta y^{\alpha}}{\theta^2+1}\right) e^{-\theta y^{\alpha}}$$

Then, expand
$B^j = e^{-j\theta y^{\alpha}} \sum_{i=0}^{j} \binom{j}{i} \left(\frac{\theta y^{\alpha}}{\theta^2+1}\right)^i$
Also, setting $t = y^{\alpha} \Rightarrow dt = \alpha y^{\alpha-1} dy$ and

$y^r = t^{\frac{r}{\alpha}}$. The integral becomes a sum gamma-type integral:

$$\int_0^\infty (\theta+t)t^{i+\frac{r}{\alpha}}e^{-(j+1)\theta t}dt = \theta\frac{\Gamma\left(i+\frac{r}{\alpha}+1\right)}{\left[(j+1)\,\theta\right]^{i+\frac{r}{\alpha}+1}} + \frac{\Gamma\left(i+\frac{r}{\alpha}+2\right)}{\left[(j+1)\,\theta\right]^{i+\frac{r}{\alpha}+2}} \quad (10)$$

By simplifying Equation 10 using $\Gamma(x+1) = x\Gamma x$ and gathering constants, this results in a compact closed form, and Equation 11 becomes the r^{th} moment of the EPS distribution:

$$\mu_r^{'} = k\theta^{-\frac{r}{\alpha}}\sum_{j=0}^{k-1}(-1)^j\binom{k-1}{j}\sum_{i=0}^{j}\binom{j}{i} \frac{\Gamma\left(i+\frac{r}{\alpha}+1\right)\left[(j+1)\,\theta^2\left(i+\frac{r}{\alpha}+1\right)\right]}{\left(\theta^2+1\right)^{i+1}(j+1)^{i+\frac{r}{\alpha}+2}} \quad (11)$$

See Nwankwo et al. (2023) and Shanker and Shukla (2017a). The mean of the proposed distribution can be obtained from Equation 11 when $r = 1$ given by $E(Y) = \mu_1^{'}$, and when $r = 2$, the variance can be derived using the relation $Var(Y) = \mu_2^{'} - (\mu_1^{'})^2$.

Furthermore, the r^{th} incomplete moment of the EPS distribution is

$$w,(u) = \int_0^u y^r f_{EPS}(y)dy)$$

We reuse the binomial steps above by letting $t = y^\alpha, dt = \alpha y^{\alpha-1}$ with $B = \left(1+\frac{\theta y^\alpha}{\theta^2+1}\right)e^{-\theta y^\alpha}$,

$$(1-B)^{k-1} = \sum_{j=0}^{k-1}\binom{k-1}{j}(-1)^j\sum_{i=0}^{j}\binom{j}{i}\times \left(\frac{\theta y^\alpha}{\theta^2+1}\right)^i e^{-j\theta y^\alpha}.$$

Putting $\alpha = i+\frac{r}{\alpha}$ and $v = u^\alpha$, the truncated integral becomes

$$\int_0^\infty (\theta+t)e^{-(j+1)\theta t}dt = \theta\left[(j+1)\,\theta\right]^{-(\alpha+1)}\gamma(\alpha+1)\times ((j+1)\theta v) + \left[(j+1)\,\theta\right]^{-(\alpha+2)}\gamma(\alpha+2),((j+1)\theta v) \quad (12)$$

where $\gamma(.,.)$ is the lower incomplete gamma, and fixing them together results in the incomplete r^{th} moment of the EPS distribution in Equation 13:

$$w(u) = k\sum_{j=0}^{k-1}(-1)^j\binom{k-1}{j}\sum_{i=0}^{j}\binom{j}{i}\frac{1}{(\theta^2+1)^{i+1}}\times \left[\frac{\Gamma(\alpha+1)}{(j+1)^{\alpha+1}\theta^{i+\frac{r}{\alpha}}}Q(\alpha+1,s) + \frac{\Gamma(\alpha+2)}{(j+1)^{\alpha+2}\theta^{i+\frac{r}{\alpha+2}}}Q(\alpha+2,s)\right] \quad (13)$$

where $s = (j+1)\theta y u^\alpha, Q(v,s) = \dfrac{\gamma(v,s)}{\Gamma(s)}$ is the regularized incomplete gamma.

Estimation of the Parameters

Let $y_1, y_2, y_3, ..., y_n$ be a random variable sample of size n from EPS distribution $\varphi = (k,\theta,\alpha)$, and the log-likelihood function is given by

$$\ell(\varphi) = \prod_{i=1}^{n} f_{EPS}(y,k,\alpha,\theta)$$

$$\ell(\varphi) = n\left[log(k)+log(\alpha)+2log(\theta)-log(\theta^2+1)\right] +(\alpha-1)\sum_{i=1}^{n}log(y_i)+\sum_{i=1}^{n}log(\theta+y_i^\alpha)-\theta\sum_{i=1}^{n}y_i^\alpha + (k-1)\sum_{i=1}^{n}log\left\{1-\left[1+\frac{\theta {y_i}^\alpha}{\theta^2+1}\right]e^{-\theta {y_i}^\alpha}\right\} \quad (14)$$

Furthermore, the maximum likelihood estimation of EPS distribution is obtained as

$$\frac{\partial \ell}{\partial k} = \frac{n}{k} + \sum_{i=1}^{n} log\left\{1 - \left[1 + \frac{\theta {y_i}^{\alpha}}{\theta^2 + 1}\right] e^{-\theta {y_i}^{\alpha}}\right\} \tag{15}$$

$$\frac{\partial \ell}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n} log y_i + \sum_{i=1}^{n} \frac{y_i^{\alpha} log y_i}{\theta + y_i^{\alpha}} - \theta \sum_{i=1}^{n} y_i^{\alpha} log y_i$$

$$+(k-1) \sum_{i=1}^{n} \frac{y_i^{\alpha} log(y_i) e^{-\theta y_i^{\alpha}} \theta^2 (\theta + y_i^{\alpha})}{(\theta^2 + 1)\left\{1 - \left[1 + \frac{\theta y_i^{\alpha}}{\theta^2 + 1}\right] e^{-\theta y_i^{\alpha}}\right\}} \tag{16}$$

and

$$\frac{\partial \ell}{\partial \theta} = \frac{2n}{\theta} - \frac{2n\theta}{\theta^2 + 1} + \sum_{i=1}^{n} \frac{1}{\theta + y_i^{\alpha}} - \sum_{i=1}^{n} y_i^{\alpha} \quad +$$

$$(k-1) \sum_{i=1}^{n} \frac{e^{-\theta y_i^{\alpha}} \left(\frac{\theta y_i^{\alpha}}{\theta^2 + 1} y_i^{\alpha} - \frac{\partial(\theta y_i^{\alpha}/\theta^2 + 1)}{\partial \theta}\right)}{\left\{1 - \left[1 + \frac{\theta {y_i}^{\alpha}}{\theta^2 + 1}\right] e^{-\theta {y_i}^{\alpha}}\right\}} \tag{17}$$

Therefore, Equations 16 and 17 cannot be solved in a closed form, rather the study used R code R 4.5.0 version (2025; Nash, 2014) to obtain their optimum function. However, the full proof of the Equations 15, 16, and 17 are in the appendix.

Logarithmic of the EPS Distribution

Let $T = log(Y)$, where $Y \sim EPS(k, \alpha, \theta)$ is its pdf defined in Equation 6. Here, we reparameterize by letting $\alpha = \frac{1}{\sigma}$, $\mu = x^T \beta$, and $\theta = e^{\frac{\mu}{\sigma}}$. Also, $A(t) = e^{\left(\frac{t-\mu}{\sigma}\right)}, D = e^{\frac{2\mu}{\sigma}}, P = 1 + D$, and $DA = \theta y^{\alpha} = e^{\left(\frac{t-\mu}{\sigma}\right)}$. Then, the log-exponentiated power Shanker (Log-EPS) density for $y \in \mathbb{R}$ is

$$f_{LEPS}(t; k, \sigma, \mu) = \frac{k}{\sigma} \frac{D^2 A\,(1 + A)}{P} e^{(-DA)} \times$$

$$\left[1 - \left(1 + \frac{DA}{P}\right) e^{(-DA)}\right]^{k-1} \tag{18}$$

where $k, \sigma > 0$ and $\mu \in \mathbb{R}$. Suppose $Y \sim EPS(k, \alpha, \theta)$ also, $Xlog(Y) \sim LEPS(k, \mu, \sigma)$. Meanwhile, the reliability and failure rate function are

$$Rel_{LEPS}(z; k, \mu, \sigma) = 1 - \left[1 - \left(1 + \frac{z}{P}\right) e^{(-z)}\right]^k \tag{19}$$

and

$$fr_{LEPS}(z; k, \mu, \sigma) = \frac{f_{LEPS}(z; k, \sigma, \mu)}{Rel_{LEPS}(z; k, \mu, \sigma)}$$

$$= \frac{\frac{k}{\sigma} \frac{D^2 A\,(1 + A)}{P} e^{(-z)} \left[1 - \left(1 + \frac{z}{P}\right) e^{(-z)}\right]^{k-1}}{1 - \left[1 - \left(1 + \frac{z}{P}\right) e^{(-z)}\right]^k};$$

$$z \in \mathbb{R} \tag{20}$$

respectively, where $D^2 A = e^{\frac{z + 3\mu}{\sigma}}$ and $z = \dfrac{x - \mu}{\sigma}$ (Alkarni 2016; Asgharzadeh et al., 2018; Marthin & Rao, 2020).

Hence, a parametric regression model for response variable X_i and a vector of predictor variables $U_i^{'} = (u_{i1}, u_{i2}, ..., u_{ip})$ is obtained as

$$x_i = U^{'} \beta + \sigma z_i, \qquad i = 1, 2, 3, ..., n \tag{21}$$

where $\mu_i = U^{'} \beta$ is the location of $x_i, \beta = (\beta_1, ..., \beta_p)$, a vector of unknown regression coefficients, and z is the error term which has density function of Equation 18 (Ortega et al., 2013; Badmus, Akinyemi, & Onyeka-Ubaka, 2021; Nwankwo et al., 2023). Meanwhile, the reliability and density function of $X_i|U'$ are

$$Rel(x|U') = 1 - \left[1 - \left(1 + \frac{z_i}{P}\right) e^{(-z_i)}\right]^k \quad (22)$$

and

$$f(x|U') = \frac{k}{\sigma} \frac{D^2 A\,(1 + A)}{P} e^{(-z_i)} \times$$

$$\left[1 - \left(1 + \frac{z_i}{P}\right) e^{(-z_i)}\right]^{k-1} \quad (23)$$

where $D^2 A = e^{\frac{z_i + 3\mu_i}{\sigma}}$.

Application to Marriage Dissolution Data

In this section, parameter estimates and the goodness of fit of the EPS distribution are obtained using the method of maximum likelihood. The model is applied to a real data set consisting of the number of years in marriage for 568 couples from the U.S. marriage dissolution study. For comparative purposes, the fit of the EPS distribution is evaluated against several alternative two-parameter and three-parameter models. Specifically, the comparisons include the three-parameter exponentiated-power distributions (Akash, Ishita, and Lindley), the two-parameter power distributions (Shanker and Akash), and the exponentiated Akash distribution.

The estimated parameters are denoted as $(\hat{k}, \hat{\alpha}, \hat{\theta})$ for the three-parameter models and $(\hat{\alpha}, \hat{\theta})$ for the two-parameter models. The goodness-of-fit assessment is based on the Shapiro–Wilk statistic (W*), Anderson–Darling statistic (A*), Kolmogorov–Smirnov statistic (KS), and their corresponding p-values, as summarized in Table 1. Furthermore, Table 2 presents the overall performance and model selection results, highlighting the comparative behavior of the EPS distribution relative to the competing models.

Obviously from Tables 1 and 2 above, the proposed distribution yields better fit (due to its ability to produce smaller values from the data used) than any other compared distributions with three and two parameters and EPS distribution can be seen as an important model for modeling social and education data which are right skewed (non-normal data). Meanwhile, the nature of the data (years in marriage) is illustrated in Figure 2 using scatter, line-scatter, histogram, and density plots. The plots clearly reveal that the data are right skewed.

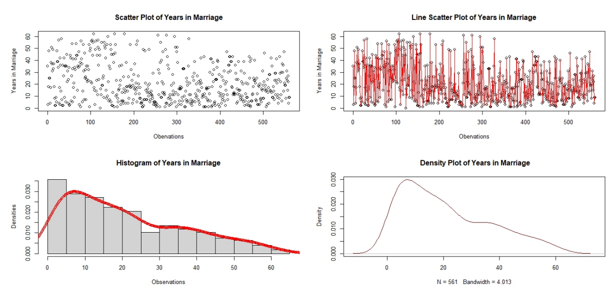

Figure 2. The scatter, line-scatter, and density plots and histogram of the data.

However, the Log-EPS distribution is compared with other distributions in this section. The variables considered are as follows: years: duration of marriage; heduc: education of the husband coded 0 = less than 12 years, 1 = 12 to 15 years, and 2 = 16 or more years; heblack: coded 1 if the husband is black and 0 otherwise; mixed: coded 1 if the husband and wife have different ethnicity, 0 otherwise, and div: the failure indicator, coded 1 for divorce and 0 for censoring (Lillard & Panis, 2000). The estimates of parameter $(\hat{k}, \hat{\sigma}, \hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3 \;\; \& \;\; \hat{\beta}_4)$, standard error and $p - value$ in Table 3, and model performance criteria (-2Log-likelihood, Akaike information criterion [AIC], consistent Akaike Information criterion [CAIC], Bayesian information criterion [BIC], and Hannan-Quinn information criterion [HQIC]) are presented in Table 4. The

Table 1. MLEs and Statistics: (W*), (A*), (KS), and P

Distribution	Estimate	W*	A*	KS
EPS (proposed)	$\hat{k} = 0.3654(0.1001^*)$ $\hat{\alpha} = 0.1619(0.0084^*)$ $\hat{\theta} = 2.5548(0.5363^*)$	0.2928	1.9186	0.0416 0.2807^P
EPA (Nwankwo et al., 2023)	$\hat{k} = 29.3254(6.5405^*)$ $\hat{\alpha} = 0.2321(0.0111^*)$ $\hat{\theta} = 3.2035(0.1843^*)$	1.4460	9.3746	0.6328 $< 2.2e-16^P$
EPI (Ferreira & Cordeiro, 2023)	$\hat{k} = 8.4644(0.9192^*)$ $\hat{\alpha} = 0.3094(0.0074^*)$ $\hat{\theta} = 2.0262(0.0551^*)$	1.1925	7.8513	0.1898 $< 2.2e-16^P$
EPL (Ashour & Eltehiwy, 2015)	$\hat{k} = 113.7877(60.6813^*)$ $\hat{\alpha} = 4.1003(0.5104^*)$ $\hat{\theta} = 0.1725(0.0174^*)$	NaN	NaN	0.6822 $< 2.2e-16^P$
PS (Shanker & Shukla, 2017)	$--$ $\hat{\alpha} = 0.8440(0.0254^*)$ $\hat{\theta} = 0.1574(0.0134^*)$	0.2810	2.1317	0.4325 $< 2.2e-16^P$
PA (Shanker & Shukla, 2017b)	$--$ $\hat{\alpha} = 0.7155(0.0187^*)$ $\hat{\theta} = 0.3480(0.0210^*)$	0.2843	2.1462	0.6102 $< 2.2e-16^P$
EA (Okereke & Uwaeme, 2018)	$\hat{k} = 0.7051(0.0326^*)$ $--$ $\hat{\theta} = 0.1287(0.0037^*)$	0.2883	2.1847	0.3028 $< 2.2e-16^P$

Note. MLEs = maximum likelihood estimations, P = p-value, EPS = exponentiated power Shanker, EPA = exponentiated power Akash, EPI = exponentiated power Ishita, EPL = exponentiated power Lindley, PS = power Shanker, PA = power Akash, EA = exponentiated Akash, NaN = not a number.

Table 2. Performance and Selection Criteria for Fitted Distributions

Distribution	-2LogL	AIC	CAIC	BIC	HQIC	Rank
EPS	2259.372	4524.745	4524.788	4537.761	4529.825	1st
EPA	2647.686	5301.372	5301.415	5314.388	5306.452	5th
EPI	2700.281	5406.562	5406.605	5419.578	5411.642	6th
EPL	2626.067	5258.135	5258.177	5271.150	5263.215	4th
PS	2274.488	4552.976	4552.997	4561.653	4556.363	2nd
PA	2279.753	4563.505	4563.526	4572.182	4566.892	3rd
EA	4194.584	8393.168	8393.189	8401.845	8396.554	7th

Note. -2LogL = -2x log-likelihood, AIC = Akaike information criterion, CAIC = consistent Akaike information criterion, BIC = Bayesian information criterion, HQIC = Hannan- Quinn information criterion.

regression model is given by

$$x_i = \beta_0 + \beta_1 y_{i1} + \beta_2 y_{i2} + \beta_3 y_{i3} +, ..., +\beta_n y_{in} + \epsilon_i$$

$$i = 1, 2, ..., 568. \quad (24)$$

$$year(x_i) = \beta_0 + \beta_1(Heduc) + \beta_2(Heblack) + \beta_3(Mixed) + \beta_n(Divorce) + \epsilon_i \quad (25)$$

$$year(\hat{x}) = 0.0809 - 1.6190(Heduc) + 0.4343 (Heblack) - 1.0747(Mixed) + 1.1819(Divorce) \quad (26)$$

The estimated (β_s) parameters of the Log-EPS regression model in Table 3 have both positive and negative values. This means that husband education (–1.62%) and mixture (–1.07%) have a negative effect in marriage. Then, husband being black (0.43%) and divorce (1.18%) have a positive effect in marriage according to the data used. Table 4 contains information on model selection criteria and reveals that the Log-EPS model has the smallest value among other models. It is noticed that the Log-EPS model is flexible and capable of handling and modeling such data compared to other models considered. All outputs in Tables 1, 2, 3, and 4 are made from R-code.

Furthermore, the study extends the work to compare machine learning and least square models, and the output generated is presented in Table 5, which consists of the estimated $\hat{\beta}_s$, mean absolute error (MAE), mean square error (MSE), and root mean square error (RMSE). The results from machine learning are superb over the least square model. The outputs are generated using Python programming. Figure 3 illustrates the density and empirical distribution function plots of the distributions examined in this study. Likewise, Figure 4 depicts the line scatter plots comparing the proposed distribution with the competing models. All figures were generated using the R statistical programming environment.

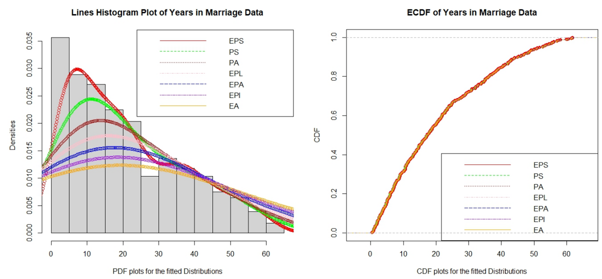

Figure 3: The densities and empirical cumulative distribution Function (ECDF) plots of the considered distribution.

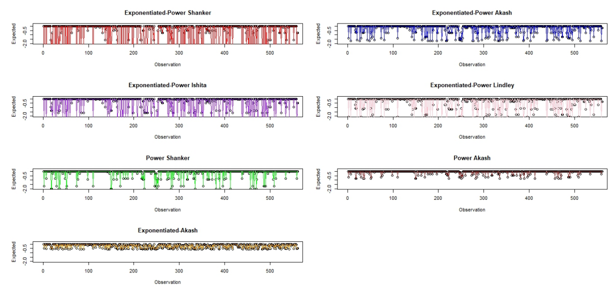

Figure 4. The line-scatter plot of the considered distribution.

Discussions and Concluding Remark

This article introduces a new three-parameter distribution, namely, the EPS distribution, and further extends it to the Log-exponentiated power Shanker (Log-EPS) model. Several important statistical properties of these distributions have been derived, including their probability density functions, cumulative distribution functions, survival functions, and moments. The method of maximum likelihood estimation (MLE) has also been presented for parameter estimation, ensuring practical applicability of the models.

To assess the performance of the EPS and Log-EPS distributions, both theoretical results and empirical analyses were conducted. In particular, a real-life data set was used to compare

Table 3. Model Parameter Estimates, Standard Errors (in Parentheses), the β(s), and p-values (in Brackets)

Model	$\hat{k}$	$\hat{\sigma}$	Estimated Parameter
Log-EPS (proposed)	0.2019 (0.2050)	0.5342 (0.0142)	$\hat{\beta}_0 = 0.0809(0.0112)$ $\hat{\beta}_1 = -1.6190(0.2853)[1.39e-08^{***}]$ $\hat{\beta}_2 = 0.4343(3.1482)[0.8900]$ $\hat{\beta}_3 = -1.0747(2.2891)[0.6390]$ $\hat{\beta}_4 = 1.1819(0.2810)[< 2.61e-05^{***}]$
Log-EPA (Nwankwo et al., 2023)	0.9998 (0.0005)	1.0428 (0.0073)	$\hat{\beta}_0 = 1.5680(1.9431)$ $\hat{\beta}_1 = -8.0102(0.9081)[< 2e-16^{***}]$ $\hat{\beta}_2 = 2.3129(2.1798)[0.2890]$ $\hat{\beta}_3 = -3.6466(2.3342)[0.1180]$ $\hat{\beta}_4 = -0.1963(1.2094)[0.8710]$
Log-EPI (Ferreira & Cordeiro, 2023)	0.9353 (0.0485)	1.0561 (0.0052)	$\hat{\beta}_0 = -0.0845(0.0517)$ $\hat{\beta}_1 = -8.8736(0.2135)[< 2e-16^{***}]$ $\hat{\beta}_2 = -10.8228(1.6074)[1.66e-11^{***}]$ $\hat{\beta}_3 = -6.4112(2.7700)[0.0180^{*}]$ $\hat{\beta}_4 = -9.7332(0.1323)[< 2e-16^{***}]$
Log-EPL (Shabani et al., 2018)	1.0691 (0.1695)	0.8108 (0.0060)	$\hat{\beta}_0 = 0.0381(0.0644)$ $\hat{\beta}_1 = -5.4548(0.2060)[< 2e-16^{***}]$ $\hat{\beta}_2 = -0.4618(1.6774)[0.7831]$ $\hat{\beta}_3 = 5.6307(1.0740)[1.71e-07^{***}]$ $\hat{\beta}_4 = -0.0030(0.0015)[0.0534]$
Log-PS	– –	0.5000 (0.0277)	$\hat{\beta}_0 = 3.5000(1.1295)$ $\hat{\beta}_1 = 1.4000(1.0525)[0.1835]$ $\hat{\beta}_2 = 3.5000(8.7371)[0.6887]$ $\hat{\beta}_3 = 1.5000(7.6354)[0.8443]$ $\hat{\beta}_4 = 3.5000(1.1173)[0.0017^{**}]$
Log-PA	– –	0.7443 (0.0108)	$\hat{\beta}_0 = 2.4175(0.3786)$ $\hat{\beta}_1 = -2.3586(0.0909)[< 2e-16^{***}]$ $\hat{\beta}_2 = 2.8324(2.5483)[0.2660]$ $\hat{\beta}_3 = -0.2427(1.0744)[0.8210]$ $\hat{\beta}_4 = -2.5748(NAN)[NAN]$
Log-EA	– –	0.8790 (0.0164)	$\hat{\beta}_0 = 3.2540(1.4812)$ $\hat{\beta}_1 = 1.7326(0.8243)[0.0356^{*}]$ $\hat{\beta}_2 = 2.6045(1.0500)[0.0131^{*}]$ $\hat{\beta}_3 = 1.6370(0.7667)[0.0328^{*}]$ $\hat{\beta}_4 = 0.0391(1.2337)[0.9747]$

Note. Log-EPS = Log-exponentiated power Shanker, Log-EPA = Log-exponentiated power Akash, Log-EPI = Log-exponentiated power Ishita, Log-EPL = Log-exponentiated power Lindley, Log-PS = Log-power Shanker, Log-PA = Log-power Akash, Log-EA = Log-exponentiated Akash, $^{*}p < 0.05$, $^{**}p < 0.01$. $^{***}p < 0.001$.

Table 4. Model Selection Criteria

Distribution	-2LogL	AIC	CAIC	BIC	HQIC
Log-EPS	4433.527	8073.054	8087.080	8086.080	8078.137
Log-EPA	5061.834	10129.670	10143.690	10142.690	10134.750
Log-EPI	4319.140	8644.280	8658.306	8657.306	8649.363
Log-EPL	4096.548	8199.096	8213.122	8212.122	8204.178
Log-PS	4979.949	9963.898	9923.582	9972.582	9957.287
Log-PA	4667.380	9338.760	9348.444	9347.444	9342.149
Log-EA	5314.537	10633.070	10642.760	10641.760	10636.460

Note. -2LogL = -2x log-likelihood, AIC = Akaike information criterion, CAIC = consistent Akaike information criterion, BIC = Bayesian information criterion, HQIC = Hannan- Quinn information criterion, Log-EPS = Log-exponentiated power Shanker, Log-EPA = Log-exponentiated power Akash, Log-EPI = Log-exponentiated power Ishita, Log-EPL = Log-exponentiated power Lindley, Log-PS = Log-power Shanker, Log-PA = Log-power Akash, Log-EA = Log-exponentiated Akash.

Table 5. Machine Learning and Least Square Model Using Marriage Dissolution Data

Model	Estimate	MAE	MSE	RMSE
Machine learning	$\hat{\beta}_0 = 30.7352$ $\hat{\beta}_1 = -7.1823$ $\hat{\beta}_2 = -1.8977$ $\hat{\beta}_3 = -14.7729$ $\hat{\beta}_4 = -13.5405$	10.1138	162.1605	12.7342
Least square	$\hat{\beta}_0 = 30.3480$ $\hat{\beta}_1 = -6.3250$ $\hat{\beta}_2 = -1.4380$ $\hat{\beta}_3 = -14.7100$ $\hat{\beta}_4 = -13.8860$	20.9668	483.3172	21.9845

Note. MAE = mean absolute error, MSE = mean square error, RMSE = root mean square error.

the proposed models with other well-known two-parameter and three-parameter distributions. Model evaluation was carried out using standard statistical tools such as goodness-of-fit measures and model selection criteria.
The findings reveal that the EPS and Log-EPS distributions provide superior fits to the data compared with the competing models. This demonstrates their enhanced flexibility in modeling lifetime and reliability data. The results highlight the practical relevance of the proposed models, particularly in contexts where tail behavior and hazard rate patterns require additional flexibility.
In conclusion, the EPS and Log-EPS distributions constitute valuable contributions to statistical modeling. Their ability to outperform existing alternatives underscores their robustness and suitability for real-world applications. Future studies may extend these models to regression frameworks, Bayesian estimation approaches, multivariate generalizations, and applications in diverse fields such as reliability engineering, actuarial science, medical statistics, and risk analysis.

Funding

This research did not receive any funding from external sources.

Acknowledgement

The authors would like to acknowledge the contributions of those (including the reviewers) who helped make this research successful.

Conflict of Interest

The authors have no conflicts of interest to disclose.

References

Aafaq, A. R., & Subramania, C. (2019). Exponentiated Ishita distribution: Properties and applications. *International Journal of Management Technology and Engineering*, *9*(5), 2473–2484.

Alkarni, S. H. (2016). A class of Lindley and Weibull distributions. *Open Journal of Statistics, 6*(4), 685–700. https://doi.org/10.4236/ojs.2016.64058

Asgharzadeh, A., Nadarajah, S., & Sharafi, F. (2018). Weibull Lindley distribution. *REVSTAT-Statistical Journal, 16*(1), 87–113. https://doi.org/10.57805/revstatv16i1.234

Ashour, S. K., & Eltehiwy, M. A. (2015). Exponentiated power Lindley distribution. *Journal of Advanced Research, 6*(6), 895–905.

Badmus, N. I., & Aromolaran, A. D. (2024). On the exponentiated-Sujatha distribution and its applications. *International Journal of Applied Mathematics, Computational Sciences and System Engineering*, *6*, 44–54. https://doi.org/10.37394/232026.2024.6.5

Badmus, N. I., & Faweya, O. (2022). Exponentiated new modified weighted Rayleigh distribution with application. *Journal of e-Science Letters*, *3*(5), 1–8. http://dx.doi.org/10.46890/SL.2022.v03i05.001

Badmus, N. I., Akinyemi, M. I., & Onyeka-Ubaka, J. N. (2021). A Log-beta Rayleigh Lomax regression model. *Journal of Afrika Statistika*, *16*(4), 2993–3007. https://doi.org/10.16929/as/2021.2993.192

Badmus, N. I., Olufolabo, O. O., & Akingbade, A. A. (2021). A two-parameter exponential-Akash distribution: Theory and application. *International Journal of Mathematical and Computational Method*, *6*, 20–27. http://www.iaras.org/iaras/journals/ijmcm

Elgarhy, M., & Shawki, A. W. (2017). Exponentiated Sushila distribution. *International Journal of Scientific Engineering and Science*, *1*(7), 9–12.

Ferreira, A. A., & Cordeiro, G. M. (2023). The exponentiated-power Ishita distribution: Properties, simulations, regression and applications. *Chilean Journal of Statistics (ChJS)*, *14*(1), 65–81. https://doi.org/10.32372/chjs.14-01-05

Gupta, R. C., Gupta, P. I., & Gupta, R. D. (1998). Modeling failure time data by Lehmann alternatives. *Communications in Statistics: Theory & Methods, 27*(4), 887–904.

Lillard, L. A., & Panis, C. W. A. (2000). *ML Multilevel Multiprocess Statistical Software* (Release 1.0). EconWare. divorce.dta.

Marthin, P., & Rao, G. S. (2020). Generalized Weibull-Lindley (GWL) distribution in modeling lifetime data. *Journal of Mathematics, 2020*, 1–15. https://doi.org/10.-1155/2020/2049501

Nash, J. C. (2014). *Nonlinear parameter optimization using R tools.* John Wiley and Sons Ltd., https://doi.org/10.1002/9781118884003

Nwankwo, M. P., Bright, C. N., & Okechukwu, J. O. (2023). The exponentiated power Akash distribution: properties, regression and applications to infant mortality rate and COVID-19 patients' life cycle. *Annals of Biometrics & Biometric Applications, 5*(4), 1–13.

Okereke, E. W., & Uwaeme, O. R. (2018). Exponentiated Akash distribution and its applications. *Journal of the Nigerian Statistical Association, 30*, 1–13.

Ortega, E. M. M., Cordeiro, G. M., & Kattan, M. W. (2013). The Log-beta Weibull regression model with application to predict recurrence of prostate cancer. *Regular Article, Statistical Paper*, *54*(1), 113–132

R (Version 4.5.0). (2025). *"How about a twenty-six" was Released on 11th April, 2025.* https://cran.r-project.org/doc/manuals/r.devel/NEWS.html

Shanker, R., & Hagos, K. K. (2016). On modeling of lifetime data using Akash, Shanker, Lindley and exponential distribution. *Biometric and Biostatistics International Journal, 3*(6), 1–9.

Shanker, R., & Prodhani, H. R. (2022). *Exponentiated Sujatha distribution.* Eighth International Conference on Statistics for Twenty-First Century 2022, Trivandrum, Kerala, India.

Shanker, R., & Shukla, K. K. (2017a). Power Shanker distribution and its application. *Turkiye Klinikleri Journal of Biostatistics, 9*(3), 175–187. https://doi.org/10.5336/biostatic.2017-56601

Shanker, R., & Shukla, K. K. (2017b). Power-Akash distribution and its applications. *International Journal of Applied Quantitative Methods*, *12*(3), 1–1012.

Shabani, A., Moghadam, M. K., & Moradi, E. (2018). Exponentiated-power Lindley logarithmic: Models properties and applications. *Annals of Data Science*, *5*, 583–613.

Appendix

The partial derivatives from the log-likelihood in Equation 14 with respect to k, α, and θ are detailed as follows:

Recall that

$$\ell(\varphi) = \prod_{i=1}^{n} f_{EPS}(y, k, \alpha, \theta)$$

$$\ell(\varphi) = n\left[log(k) + log(\alpha) + 2log(\theta) - log(\theta^2+1)\right]$$

$$+(\alpha-1)\sum_{i=1}^{n} log(y_i) + \sum_{i=1}^{n} log(\theta + y_i^{\alpha}) -$$

$$\theta\sum_{i=1}^{n} y_i^{\alpha} + (k-1)\sum_{i=1}^{n} log\left\{1 - \left[1 + \frac{\theta {y_i}^{\alpha}}{\theta^2+1}\right] e^{-\theta {y_i}^{\alpha}}\right\}$$

Setting $H_i = \left[1 + \frac{\theta {y_i}^{\alpha}}{\theta^2+1}\right] e^{-\theta {y_i}^{\alpha}}$, then $log\left(1 - H_i\right)$, and $L = \theta^2 + 1$

Differentiating with respect to k is given as

$$\frac{\partial \ell}{\partial k} = \frac{n}{k} + \sum_{i=1}^{n} log\left\{1 - \left[1 + \frac{\theta {y_i}^{\alpha}}{\theta^2+1}\right] e^{-\theta {y_i}^{\alpha}}\right\} \quad \blacksquare$$

This is Equation 15.

By listing out all the terms that involve α_i $nlog\alpha \Rightarrow \frac{n}{\alpha}$, $(\alpha-1)\sum_{i=1}^{n} logy_i \Rightarrow \sum_{i=1}^{n} logy_i$, $\sum_{i=1}^{n} log\left(\theta + y_i^{\alpha}\right) \Rightarrow \sum_{i=1}^{n} \frac{y_i^{\alpha} lny_i}{\theta + y_i^{\alpha}}$ $-\theta\sum_{i=1}^{n} y_i^{\alpha} \Rightarrow -\theta\sum_{i=1}^{n} y_i^{\alpha} lny_i$, $(k-1)\sum_{i=1}^{n} log \left\{1 - \left[1 + \frac{\theta {y_i}^{\alpha}}{\theta^2+1}\right] e^{-\theta {y_i}^{\alpha}}\right\}$, and this gives $(k-1)\sum_{i=1}^{n} \frac{\frac{-\partial H_i}{\partial \alpha}}{\left\{1 - \left[1 + \frac{\theta {y_i}^{\alpha}}{\theta^2+1}\right] e^{-\theta {y_i}^{\alpha}}\right\}}$. Then, we need $\frac{-\partial H_i}{\partial \alpha}$; we have

$$H_i = (1 + Ry_i^{\alpha})e^{-\theta y_i^{\alpha}}, \quad R = \frac{\theta}{\theta^2+1} = \frac{\theta}{L}$$

$$\frac{\partial H_i}{\partial \alpha} = e^{-\theta y_i^{\alpha}} Ry_i^{\alpha} ln\left(y_i\right) + \left(1 + Ry_i^{\alpha}\right) e^{-\theta y_i^{\alpha}} \left[-\theta y_i^{\alpha} ln\left(y_i\right)\right]$$

$$\frac{\partial H_i}{\partial \alpha} = y_i^{\alpha} lny_i e^{-\theta y_i^{\alpha}} \left[R - \theta\left(1 + Ry_i^{\alpha}\right)\right]$$

By factorizing this $y_i^{\alpha} lny_i e^{-\theta y_i^{\alpha}}$, and simplify by substituting $R = \frac{\theta}{L}$

$$R - \theta\left(1 + Ry_i^{\alpha}\right) = \frac{\theta}{L} - \theta - \frac{\theta^2}{L} y_i^{\alpha} = \frac{\theta^2\left(\theta + y_i^{\alpha}\right)}{L}$$

Hence, a compact simplified form can be written as

$$\frac{\partial H_i}{\partial \alpha} = y_i^{\alpha} lny_i e^{-\theta y_i^{\alpha}} \frac{\theta^2\left(\theta + y_i^{\alpha}\right)}{L}$$

Putting it into the contribution from the $(k-1)$ term,

$$(k-1)\sum_{i=1}^{n} \frac{\frac{-\partial H_i}{\partial \alpha}}{1 - H_i} = (k-1)\sum_{i=1}^{n} \frac{y_i^{\alpha} lny_i e^{-\theta y_i^{\alpha}} \theta^2\left(\theta + y_i^{\alpha}\right)}{L\left[1 - H_i\right]}$$

This becomes Equation 16:

$$\frac{\partial \ell}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n} lny_i + \sum_{i=1}^{n} \frac{y_i^{\alpha} lny_i}{\theta + y_i^{\alpha}} - \theta\sum_{i=1}^{n} y_i^{\alpha} lny_i +$$

$$(k-1)\sum_{i=1}^{n} \frac{y_i^{\alpha} ln(y_i) e^{-\theta y_i^{\alpha}} \theta^2(\theta + y_i^{\alpha})}{(\theta^2+1)\left\{1 - \left[1 + \frac{\theta y_i^{\alpha}}{\theta^2+1}\right] e^{-\theta y_i^{\alpha}}\right\}} \quad \blacksquare$$

Differentiating with respect to θ

From $n\left[2log\theta - log\left(\theta^2+1\right)\right]$:

$$\frac{d}{d\theta}\left[2log\theta - log\left(\theta^2+1\right)\right] = \frac{2}{\theta} - \frac{2\theta}{\theta^2+1}$$

From $\sum_{i=1}^{n} log\left(\theta + y_i^{\alpha}\right)$: $\sum_{i=1}^{n} \frac{1}{\theta + y_i^{\alpha}}$, $-\theta \sum_{i=1}^{n} y_i^{\alpha}$ − $\sum_{i=1}^{n} y_i^{\alpha}$. Also, from the $(k-1)\sum_{i=1}^{n} log\left(1 - H_i\right)$ term, we have

$$(k-1)\sum_{i=1}^{n} \frac{\frac{-\partial H_i}{\partial \theta}}{1 - H_i}$$

However, we need $\frac{\partial H_i}{\partial \theta}$, and by using $H_i = \left(1 + R y_i^{\alpha}\right) e^{-\theta y_i^{\alpha}}$ with $R = \frac{\theta}{L}$, and $L = \theta^2 + 1$, we get two contributions (derivative of R and the exponential):

$$\frac{\partial R}{\partial \theta} = \frac{1 - \theta^2}{L^2}$$

Therefore,

$$\frac{\partial H_i}{\partial \theta} = \left(y_i^{\alpha} \frac{1 - \theta^2}{L^2}\right) e^{-\theta y_i^{\alpha}} + \left(1 + R y_i^{\alpha}\right) e^{-\theta y_i^{\alpha}} \left(-y_i^{\alpha}\right)$$

$$= y_i^{\alpha} e^{-\theta y_i^{\alpha}} \left[\frac{1 - \theta^2}{L^2} - \left(1 + R y_i^{\alpha}\right)\right]$$

A compact form can be written as

$$\frac{\partial H_i}{\partial \theta} = y_i^{\alpha} e^{-\theta y_i^{\alpha}} \left[\frac{1 - \theta^2}{\left(\theta^2 + 1\right)^2} - 1 - \frac{\theta y_i^{\alpha}}{\theta^2 + 1}\right]$$

Putting all together yields Equation 17.

$$\frac{\partial l}{\partial \theta} = n\left(\frac{2}{\theta} - \frac{2\theta}{\theta^2 + 1}\right) + \sum_{i=1}^{n} \frac{1}{\theta + y_i^{\alpha}} - \sum_{i=1}^{n} y_i^{\alpha} +$$

$$(k-1)\sum_{i=1}^{n} \frac{\frac{-\partial H_i}{\partial \theta}}{1 - H_i} \quad \blacksquare$$