

Modification of the Probability Model using the Alpha Power Transformation Technique and its Effect on Diabetes Survival Time Data

Rio Perdi Tinando¹, Rado Yendra², Muhammad Marizal³, Ari Pani Desvina⁴

^{1,2,3,4}Universitas Islam Negeri Sultan Syarif Kasim Riau

ABSTRACT: Along with the changes in the patterns of data regarding the survival time of individuals with diabetes, often influenced by the consumption of unhealthy fast food. It is essential to update the appropriate probability model for this data. Two-parameter probability models, such as the Weibull, Gamma, and Log-Normal distributions, require modification by increasing the number of parameters in the probability model. The Alpha Power Transformation (APT) technique will be employed for this purpose. Three-parameter probability models generated from the APT technique, specifically, the Alpha Power Weibull Distribution (APWB), the Alpha Power Gamma Distribution (APGM), and the Alpha Power Log-Normal Distribution (APLN) will be employed to improve the accuracy of the probability model for diabetes survival time data. All probability models in this study will employ the maximum likelihood method for parameter estimation. The optimal model will be identified based on the goodness-of-fit test, which will incorporate both graphical methods (such as density graphs and cumulative distribution functions) and numerical methods (including the Akaike Information Criterion (AIC) and negative log-likelihood). The results of the goodness-of-fit tests indicate that the modified model obtained using the APT technique, particularly the APWB probability model, yields a superior probability model compared to the other models.

KEYWORDS: Survival Time, Weibull, Gamma, Alpha Power Transformation, Alpha Power Weibull Distribution (APWB), the Alpha Power Gamma Distribution (APGM), and the Alpha Power Log-Normal Distribution (APLN).

INTRODUCTION

Probability models are essential tools across various academic disciplines, enabling the precise quantification of uncertainties. These models are particularly crucial in sectors such as finance, engineering, healthcare, and environmental sciences. The application of these distributions in medical science may vary. The log-normal distribution is often used to model positive parameters, such as tumor sizes or durations of treatment responses. Generalized extreme value distributions may be relevant for analyzing extreme patient responses. The Weibull distribution could be pertinent in studies concerning the reliability of medical devices, while the generalized Pareto distribution may be advantageous for addressing rare, extreme events or outliers. Understanding the survival times of diabetic patients is crucial for estimating the risk of mortality associated with diabetes. Survival time studies can be conducted using statistical techniques to develop models that accurately represent survival patterns. Various studies have sought to identify the optimal probability model for diabetic survival time data. Ummu et al. [1] estimated the duration of diabetes survival using the Weibull, Gamma, and Log-Normal distributions. The results indicated that the Weibull model best fit the observational data. This conclusion was further supported by numerical criteria such as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), which yielded the

lowest values among the models compared. Furthermore, Manda Lisa Usvita et al. [2] compared three types of distributions: Exponential (E), Weibull (W), and Rayleigh-Lomax (RL), as applied to the survival times of diabetes patients. The Method of Moments was employed to estimate the parameters. Based on the smallest values of the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), along with graphical inspection of the probability density function (pdf) of the survival times, the study demonstrated that the Rayleigh-Lomax distribution is the best fit for modeling the survival times of diabetes patients at Mandau RSUD in Bengkalis Regency, Riau Province. Sutriana et al. [3] utilized the Lindley (LIN) distribution along with three modified Lindley distributions: the Weighted Lindley Exponential (WLE), the Power Modified Lindley (PML), and the Lindley Half-Cauchy (LHC), as well as the Rayleigh-Lomax distribution (RL). The optimal fit was determined by selecting the distribution with the lowest values of the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), and the negative log-likelihood (-l). Overall, the Rayleigh-Lomax (RL) distribution was identified as the best model. Nevertheless, they have demonstrated limitations in their capabilities to effectively model complex failure rate data [4]. These models often, it is challenging to accurately represent the complex patterns found in the data. Including increasing, decreasing,

or bathtub shaped failure rates. Limitations have prompted researchers to pursue more flexible and adaptable approaches. Robust modeling alternatives have led to the development of more effective solutions. Generalized and generated (G-class) models. One of the processes that makes the distributions richer and more flexible to represent real-life

$$F_{APT}(x) = \frac{\alpha^{G(x)} - 1}{\alpha - 1}, \text{ if } \alpha > 0, \alpha \neq 1$$
$$= G(x), \text{ if } \alpha = 1$$

(1)

The corresponding probability density function (PDF) of the APT family is

$$f_{APT}(x) = \frac{\log(\alpha)}{\alpha - 1} \alpha^{G(x)} g(x), \text{ if } \alpha > 0, \alpha \neq 1$$
$$= g(x), \text{ if } \alpha = 1$$

(2)

The primary objective of this study is to present an extension of the Weibull, Gamma, and Log-Normal distributions by utilizing the Alpha Power Transformation defined by Mahdavi and Kundu [5]. Mahdavi and Kundu [5] proposed a novel class of distributions known as the Alpha Power Transformation (APT) family. This family is specifically designed to analyze lifetime data derived from systems that exhibit a range of monotonic and non-monotonic failure patterns. Applications of this method are available in [6]. Transformation of several distributions has been studied, including alpha power-transformed Weibull [7], alpha power-transformed generalized exponential [8], alpha power-transformed Lindley [9], alpha power-transformed extended exponential [10], alpha power-transformed inverse Lindley [11], alpha power-transformed inverse Lomax [12], new alpha power-transformed power Lindley [13], and alpha power-transformed extended power Lindley distribution [14]. This study aims to identify the most suitable distribution for the survival times of diabetic patients based on various goodness-of-fit criteria. A preliminary investigation was conducted on the survival times of diabetes in 50 patients from the Bengkalis region. The objective is to propose six distributions: the Weibull (GM) distribution, the Gamma (GM) distribution, the Log-Normal (LN) distribution, and three modified three-parameter distributions: the Alpha Power Weibull Distribution (APWB), the Alpha Power

data is the alpha power transformation (APT). The method has been proposed by Mahdavi and Kundu [5] with the parameter alpha to incorporate skewness into the base distribution. The cumulative distribution function (cdf) and probability density function (pdf) of random variables are, respectively, specified by

Gamma Distribution (APGM), and the Alpha Power Log-Normal Distribution (APLN). The proposed distributions are compared with existing distribution functions to evaluate their effectiveness in characterizing diabetes-related data. Unknown parameter estimates were calculated using the Maximum Likelihood Method. Graphical methods, such as probability density function (pdf) and cumulative distribution function (cdf) plots, along with numerical criteria like the Akaike Information Criterion (AIC) and negatif log-likelihood (-l), were employed to determine the distribution that best fits the diabetes data. The next section presents the distributions selected for modeling the duration of diabetes survival time data.

Materials

In this investigation, an independent sample of 50 diabetes patients was examined at Mandau Regional General Hospital (RSUD) in Bengkalis Regency, Riau Province. The following table presents preliminary data regarding the survival durations of these patients. Descriptive statistics related to the duration of diabetes, including the mean, variance, minimum, and maximum values, are detailed in Table 1. The observed consistency between the variations in the data and the means suggests that the survival times of diabetes patients exhibit a relatively stable pattern.

Table 1. The descriptive statistics for duration of diabetes (years)

Statistics	Mean	Variation	Minimum	Maximum
	3.936	4.828882	0.300	9.300

Methods: Probability Density Function (PDF), Distribution Function (CDF) and Maximum Likelihood

This study examines six probability density functions (PDFs) associated with modeling the duration of diabetes: WB, GM, LN, APWB, APGM, and APLN. The equations defining the PDFs for the various distributions of interest are presented below, where x (duration of diabetes) represents the observed values of the random variable for the event of interest. To fit a specific theoretical distribution to the observed distribution of diabetes duration, parameters are estimated using the

maximum likelihood method. The maximum likelihood function for this model is implicit and complex; a detailed discussion of this function is beyond the scope of this paper. The nonlinear equation generated by the maximum log-likelihood function (l(ω)), where (ω) represents the parameters, requires a numerical method, specifically, the Newton–Raphson method to find the solution. This method is applied iteratively to obtain the solution. Various initial values have been tested, and if they converge to the same value, that value is considered the estimated parameter. The

procedure for conducting goodness-of-fit tests for model selection, both numerically and graphically, is also discussed.

Weibull Distribution (WB)

The Weibull distribution with shape parameter λ and scale parameter β has PDF given by

$$g(x; \lambda, \beta) = \frac{\lambda}{\beta} \left(\frac{x}{\beta}\right)^{\lambda-1} \exp\left(-\left(\frac{x}{\beta}\right)^{\lambda}\right), x > 0, \lambda, \beta > 0 \quad (3)$$

The CDF is

$$G(x; \lambda, \beta) = 1 - \exp\left(-\left(\frac{x}{\beta}\right)^{\lambda}\right) \quad (4)$$

Gamma Distribution (GM)

The gamma distribution is defined on $(0, \infty)$ by the PDF,

$$g(x; \gamma, \kappa) = \frac{1}{\Gamma(\gamma)\kappa^{\gamma}} x^{\gamma-1} \exp\left(-\frac{x}{\kappa}\right), x > 0, \gamma, \kappa > 0 \quad (5)$$

where $\gamma > 0$ and $\kappa > 0$ are parameters. The CDF is

$$G(x; \gamma, \kappa) = \frac{1}{\Gamma(\gamma)\kappa^{\gamma}} \int_0^x t^{\gamma-1} \exp\left(-\frac{t}{\kappa}\right) dt \quad (6)$$

Log-Normal Distribution (LN)

The Log-Normal with two parameters, μ and σ denote the location and scale parameters, respectively. The PDF is

$$g(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\log(x)-\mu}{\sigma}\right)^2\right), x > 0, \mu, \sigma > 0 \quad (7)$$

Thus, the CDF is:

$$G(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_0^x \frac{1}{t} \exp\left(-\frac{1}{2}\left(\frac{\log(t)-\mu}{\sigma}\right)^2\right) dt \quad (8)$$

Alpha Power Transformed Weibull Distribution (APTW)

This section introduces a three-parameter APTW distribution. By substituting equations (3) and (4) into equation (2), we obtain the probability density function (PDF) for the three-parameter APTW distribution, which is expressed as follows:

$$f_{APTW}(x) = \frac{\log(\alpha)}{\alpha-1} \left(\alpha^{1-\exp\left(-\left(\frac{x}{\beta}\right)^{\lambda}\right)} \right) \left(\frac{\lambda}{\beta} \left(\frac{x}{\beta}\right)^{\lambda-1} \exp\left(-\left(\frac{x}{\beta}\right)^{\lambda}\right) \right) \quad (9)$$

where $x > 0$; $\alpha > 0$ is the additional shape parameter. The cumulative distribution function (CDF) corresponding to the given parameters is obtained by substituting equation (6) into equation (1). This results in the following expression:

$$F_{APTW}(x) = \frac{\left(\alpha^{1-\exp\left(-\left(\frac{x}{\beta}\right)^{\lambda}\right)} \right)_{-1}}{\alpha-1} \quad (10)$$

Alpha Power Transformed Gamma Distribution (APTG)

Using the alpha power transformation technique developed by Mahdavi and Kundu [13], we can insert equations (5) and (6) into equation (2) to derive the three-parameter APTG probability density function (PDF), which is expressed as follows:

$$f_{APTG}(x) = \frac{\log(\alpha)}{\alpha-1} \left(\alpha^{\left(\frac{1}{\Gamma(\gamma)\kappa^{\gamma}} \int_0^x t^{\gamma-1} \exp\left(-\frac{t}{\kappa}\right) dt \right)} \right) \left(\frac{1}{\Gamma(\gamma)\kappa^{\gamma}} x^{\gamma-1} \exp\left(-\frac{x}{\kappa}\right) \right) \quad (11)$$

where $x > 0$; $\alpha > 0$ is the additional shape parameter. The corresponding cumulative distribution function (CDF) is obtained by substituting equation (4) into equation (1), resulting in the following expression:

$$F_{APTG}(x) = \frac{\left(\alpha^{\left(\frac{1}{\Gamma(\gamma)\kappa^{\gamma}} \int_0^x t^{\gamma-1} \exp\left(-\frac{t}{\kappa}\right) dt \right)} \right)_{-1}}{\alpha-1} \quad (12)$$

Alpha Power Transformed Log-Normal Distribution (APTL)

This section introduces a three-parameter APTW distribution. Inserting equation (7) and (8) in (2), we get the three parameter APTW, PDF which is given by,

$$f_{APTL}(x) = \frac{\log(\alpha)}{\alpha-1} \left(\alpha^{\left(\frac{1}{\sigma\sqrt{2\pi}} \int_0^x \frac{1}{t} \exp\left(-\frac{1}{2}\left(\frac{\log(t)-\mu}{\sigma}\right)^2\right) dt \right)} \right) \left(\frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\log(x)-\mu}{\sigma}\right)^2\right) \right) \quad (13)$$

where $x > 0$; $\alpha > 0$ is the additional shape parameter. And the corresponding CDF is obtained by inserting (8) in (1) and is given by,

$$F_{APTL}(x) = \frac{\alpha \left(\frac{1}{\sigma\sqrt{2\pi}} \int_0^x \frac{1}{t} \exp\left(-\frac{1}{2}\left(\frac{\log(t)-\mu}{\sigma}\right)^2\right) dt \right)}{\alpha-1} \quad (14)$$

The most suitable distribution was identified using results from several goodness-of-fit (GOF) tests. These tests relied on graphical inspection, such as probability density function (PDF) plots and distribution function (CDF) plots, as well as numerical criteria, including Akaike's Information Criterion (AIC) and log-likelihood (-l), to assess the GOF of the

distributions. In most instances, the outcomes of the graphical inspections were consistent; however, the AIC results varied. The distribution with the lowest AIC value was selected as the best fit. The formulas for the numerical methods, including AIC, are presented in Table 2.

Table 2 presents the formulas of numerical criteria for model evaluation.

Numerical Criteria	Formula
AIC	$-2l + 2p$
-l	- log likelihood

l = log likelihood, p = Number of parameters

RESULTS AND DISCUSSION

In this section, we analyze a dataset of diabetes survival times to demonstrate the performance of the Weibull (WB), Gamma (GM), Log-Normal (LN), Asymmetric Weibull (APWB), Asymmetric Gamma (APGM), and Asymmetric

Log-Normal (APLN) distributions in practice. The fitting of these distributions was evaluated using the data. The computed parameter values for the various probability density functions are presented in Table 3.

Table 3. Computed parameter values of different probability density functions

	WB	GM	LN	APWB	APGM	APLN
α				2.135022		
λ	1.855466			1.705072		
β	4.424057			3.920597		
α					6.478968	
γ		2.573773			1.959773	
κ		0.653933			0.672140	
α						15.06112
μ			1.163241			0.664828
σ			0.724779			0.766772

On the graphical presentation of the modeling of diabetes survival time data, specifically in the duration of diabetes survival time histogram, the probability density function (PDF) and cumulative distribution function (CDF) plots for the Weibull (WB), Gamma (GM), Log-Normal (LN), Adjusted Weibull (APWB), Adjusted Gamma (APGM), and Adjusted Log-Normal (APLN) distribution models are illustrated in Figures 1, 2, 3, 4, 5, and 6, respectively. Upon examining the PDF and CDF plots, it was determined that some distributions yield similar results. From these figures, the WB and APWB distribution models demonstrate strong performance in modeling the duration of diabetes survival

time data. However, rather than relying solely on graphical evaluation, Table 4 offers a more meaningful comparison using Akaike Information Criterion (AIC) and negative log-likelihood (-l) values. Table 4 presents the AIC and -l test statistics for the goodness-of-fit assessment of the duration of diabetes survival time data across the WB, GM, LN, APWB, APGM, and APLN distributions. According to these results, although similar outcomes are observed for all six distributions, the lowest AIC and -l values are associated with the WB and APWB distributions. In conclusion, it is evident that the WB and APWB distributions provide superior modeling in terms of numerical criteria.

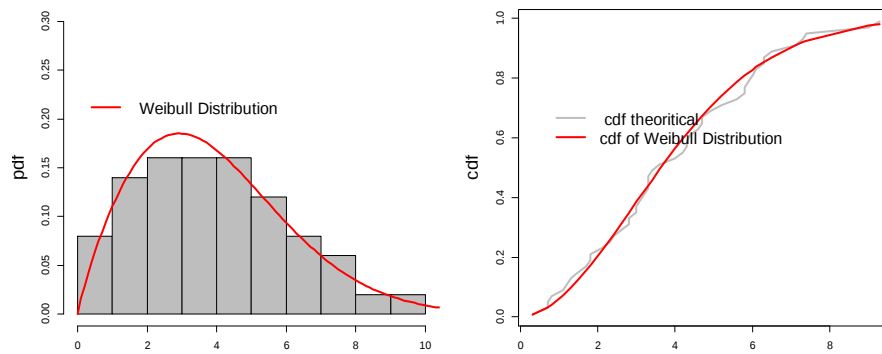


Figure 1. fitted PDF and CDF plot for WB distributions, respectively

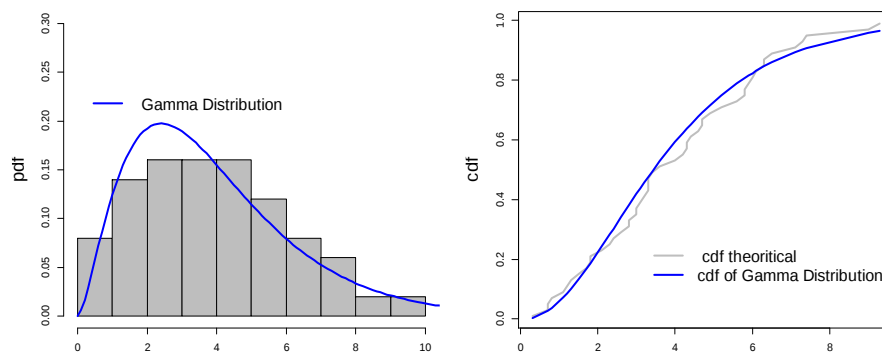


Figure 2. fitted PDF and CDF plot for GM distributions, respectively

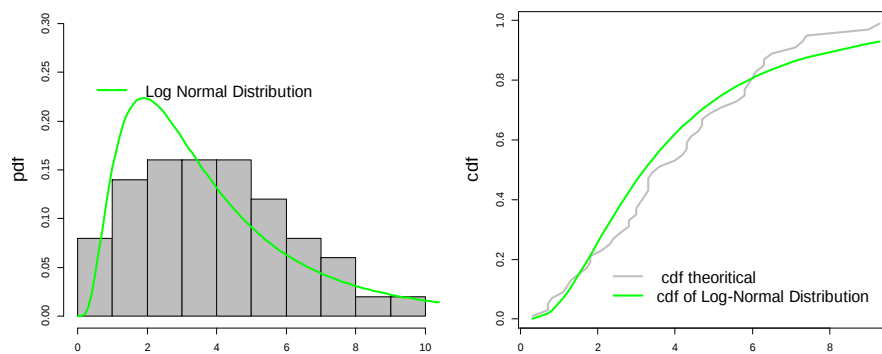


Figure 3. fitted PDF and CDF plot for LN distributions, respectively

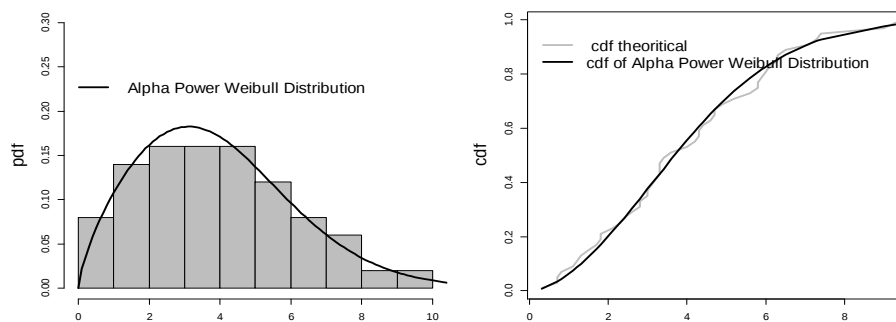


Figure 4. fitted PDF and CDF plot for APWB distributions, respectively

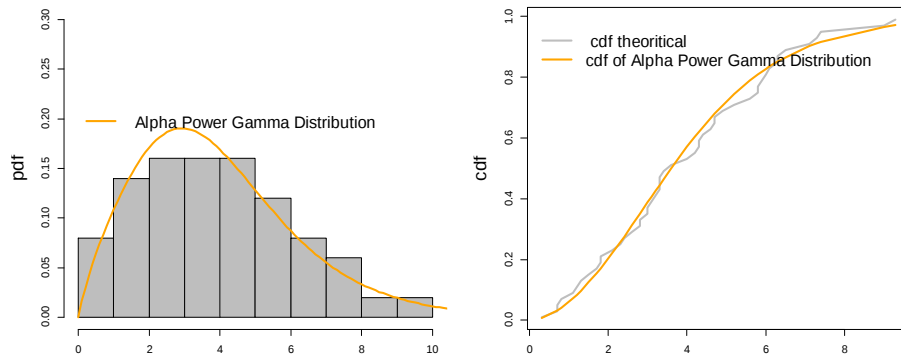


Figure 5. fitted PDF and CDF plot for APGM distributions, respectively

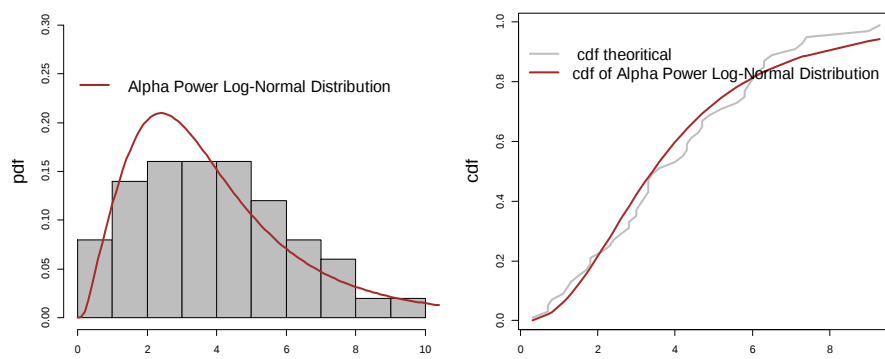


Figure 6. fitted PDF and CDF plot for APLN distributions, respectively

Table 4. AIC and – Log Likelihood (l) function values

	WB	GM	LN	APWB	APGM	APLN
-log	107.285	108.691	113.032	107.185	107.858	110.797
AIC	218.571	221.382	230.065	220.371	221.716	227.594

CONCLUSION

In this research, the focus is on determining the most effective statistical model for analyzing the duration of diabetes survival time data. The study examines six distributions: the Two-Parameter Weibull Distribution (WB), the Two-Parameter Gamma Distribution (GM), the Two-Parameter Log-Normal Distribution (LN), the Three-Parameter Alpha Power Transformed Weibull Distribution (APWB), the Three-Parameter Alpha Power Transformed Gamma Distribution (APGM), and the Three-Parameter Alpha Power Transformed Log-Normal Distribution (APLN). The findings indicate that conventional probability density functions, such as the Log-Normal distribution, are inadequate; therefore, extended functions are employed to model the observed survival duration distributions more effectively. The results clearly demonstrate that the proposed extended density function, APWB, provides a viable alternative to other probability density functions in describing the duration of diabetes survival time.

REFERENCES

1. Ummu Athifah, Rado Yendra, Muhammad Marizal, and Rahmadeni, Survival Time Data Modeling for Diabetics During the Covid 19 Pandemic Using Several Sensitive Distributions (Weibull, Gamma, and Normal Logs), Applied Mathematical Sciences, 15(2021), 717-724
2. Manda Lisa Usvita, Arisman Adnan, Rado Yendra (2021). The Modelling Survival Times for Diabetes Patient Using Exponential, Weibull and Rayleigh-Lomax Distribution *International Journal of Mathematics Trends and Technology*, 109-112.
3. Sutriana, Rado Yendra, Rahmadeni, Muhammad Marizal (2023). The Comparison Duration Diabetes Survival Times Modelling Using Lindley (LIN), Weighted Lindley Exponential (WLE), Power Modified Lindley (PML), Lindley Half-Cauchy (LHC) and Rayleigh Lomax (RL) Distributions. *International Journal Of Mathematics And Computer Research*, 11(12), 3894-3898

4. S. J. Dugasa , A. T. Goshu, and B. G. Arero (2024). Alpha Power Transformation of the Lindley Probability Distribution, *Wiley Journal of Probability and Statistics*, 2024, 1-18
5. A. Mahdavi, D. Kundu (2017). A new method for generating distributions with an application to exponential distribution. *Communications in Statistics- Theory and Methods*, 46 (13), 6543-6557
6. A. Khalil, A. A. H. Ahmadini, M. Ali, W. K. Mashwani, S. S. Alshqaq, and Z. Salleh, “A novel method for developing efficient probability distributions with applications to engineering and life science data,” *Journal of Mathematics*, vol. 2021, Article ID 4479270, 13 pages, 2021.
7. S. Dey, V. K. Sharma, and M. Mes:oui, “A new extension of Weibull distribution with application to lifetime data,” *Annals of Data Science*, vol. 4, no. 1, pp. 31–61, 2017.
8. S. Dey, A. Alzaatreh, C. Zhang, and D. Kumar, “A new extension of generalized exponential distribution with application to ozone data,” *Ozone: Science and Engineering*, vol. 39, no. 4, pp. 273–285, 2017.
9. S. Dey, I. Ghosh, and D. Kumar, “Alpha power transformed Lindley distribution: properties and associated inference with application to earthquake data,” *Annals of Data Science*, vol. 6, no. 4, pp. 623–650, 2019.
10. A. S. Hassan, R. E. Mohamd, M. Elgarhy, and A. Fayomi, “Alpha power transformed extended exponential distribution: properties and applications,” (*e Journal of Nonlinear Science and Applications*, vol. 12, no. 04, pp. 239–251, 2018.
11. S. Dey, M. Nassar, and D. Kumar, “Alpha power transformed inverse Lindley distribution: a distribution with an upside-down bathtub-shaped hazard function,” *Journal of Computational and Applied Mathematics*, vol. 348, pp. 130–145, 2019.
12. M. Ahsan-ul-Haq, S. M. Choudhary, A. H. Al-Marshadi, and M. Aslam, “A new generalization of Lindley distribution for modeling of wind speed data,” *Energy Reports*, vol. 8, no. 1, pp. 1–11, 2022.
13. R. A. ZeinEldin, M. A. U. Haq, S. Hashmi, and M. Elsehety, “Alpha power transformed inverse Lomax distribution with different methods of estimation and applications,” *Complexity*, vol. 27, pp. 1–5, 2020.
14. F. Y. Eissa and C. D. Sonar, “Alpha power transformed extended power Lindley distribution,” *Journal of Statistical Theory and Applications*, vol. 22, no. 1-2, pp. 1–18, 2022.