Estimating the Survival Function of Mixture Komal Distribution Using a New Hybrid Metaheuristic Algorithm

Shahad Kareem Taher [1, a)], Bayda Atiya Kalaf [2, b)]

[1,2]Department of Mathematics, College of Education for Pure Sciences- Ibn Al-Haitham, University of Baghdad, Baghdad, Iraq

[b)] Corresponding author: baydaa.a.k@ihcoedu.uobaghdad.edu.iq
[a)]Shahad.Taher2203@ihcoedu.uobaghdad.edu.iq

.

Abstract: This paper introduces a new mixture of the Komal distribution with a single parameter β and derives key statistical properties, such as the survival function, probability density function, hazard function, and cumulative distribution function. In addition, the study proposes a new hybrid algorithm (PSOMO) by combining the Particle Swarm Optimization (PSO) algorithm with the Monkey (MO) algorithm to estimate the survival function based on the two distribution parameters. The simulation was used to compare the performance of the proposed algorithm with the standard algorithm (PSO and MO). The results showed that the proposed algorithm (PSOMO) achieves near perfect accuracy under simulated conditions for the survival function while achieving a lower mean square error than other estimation methods.

Keywords: Mixture Distribution, Particle Swarm Optimization, Monkey Algorithm, Survival Function

INTRODUCTION

Recently, survival analysis has gained significant importance in various fields of science. Survival analysis refers to the study of the probability of events associated with failure or time of death after treatment [1]. For many years, statisticians have been interested in estimating and analyzing survival functions, as these functions are essential for understanding the nature of data related to time of failure or death [2-7]. Statistical distributions are vital tools in survival analysis, as they help accurately describe and interpret a data set. Over the past decades, many researchers have focused on developing more flexible probability distributions to accommodate the diversity and complexity of data [7-11]. Therefore, several more flexible distributions have recently been developed using innovative data analysis techniques. .[15] used an exponential Weibull distribution for the model of maximum and significant wave heights. [16] presented the Kumaraswamy-Weibull distribution, while [9] presented the Poisson-Weibull distribution and [17] proposed an extension of the Weibull distribution. [18] defined a three-parameter model for the lifetime, the so-called Weibull-Rayleigh distribution. [19] derived a Bayesian estimate of the Weibull mixture. [20] developed a new generalized class of distributions, the Burr-Weibull Power Series. [21] introduced one-parameter Lindley and Weibull distributions. [22] studied reliability analysis for mixed Weibull. With data's increasing complexity and models' nonlinearity, estimating distributions' parameters has become a major challenge [23]. For this reason, researchers have turned to using meta-heuristic algorithms that have proven effective in improving parameter estimation. For example, [24] used the log-likelihood maximization genetic algorithm (GA) to estimate the mixture's normal distribution parameters. [25] Also, it relied on the harmony search algorithm to estimate the parameters of probability distributions. [26] proposed the Jack-Knife algorithm to estimate the parameters of the mixture Komal distribution, reflecting the growing trend towards using modern techniques to improve the accuracy of estimates. Despite the successful use of metaheuristic algorithms in attacking complex machine scheduling problems, we cannot adopt one algorithm for all real-world problems [27-29].

we cannot adopt one algorithm for all real-world problems, since, despite the successful use of metaheuristic algorithms in attacking complex problems [30, 31]. Accordingly, this paper aims to present a new mixture distribution and estimate its parameters based on the survival function using a new hybrid meta-heuristic algorithm.

MIXTURE OF KOMAL DISTRIBUTION

Shanker (2023) introduced a one parameter lifetime distribution named the Komal distribution, having a probability density function (pdf) and a cumulative distribution function (cdf)

$$f(x|\beta) = \frac{\beta_1^2}{\beta_1^2+\beta_1+1}(1+\beta_1+x)e^{-\beta x} \quad (1)$$

$$F(x|(1-\beta)) = 1-\frac{\beta_1 x}{\beta_1^2+\beta_1+1}e^{-\beta_1 x} \quad (2)$$

A r.v. x is considered to have a finite mixture of Komal distribution (MKD) as its PDF and CDF can be written as, respectively:

$$f(x,\beta_1,\beta_2) = \alpha f(x,\beta_1) + (1-\alpha) f(x,\beta_2)$$

$$f(x,\beta_1,\beta_2) = \frac{\beta_2^2}{\beta_2^2+\beta_2+1}(1+\beta_2+x)e^{-\beta_2 t} + \alpha\left(\frac{\beta_1^2}{\beta_1^2+\beta_1+1}(1+\beta_1+x)e^{-\beta_1 x} - \frac{\beta_2^2}{\beta_2^2+\beta_2+1}(1+\beta_2+x)e^{-\beta_2 t}\right) \quad (3)$$

$$F(x,\beta_1,\beta_2) = 1-\alpha\left(e^{-\beta_1} + \frac{\beta_1 x}{\beta_1^2+\beta_1+1}e^{-\beta_1 x}\right) - (1-\alpha)\left(e^{-\beta_2} + \frac{\beta_2 x}{\beta_2^2+\beta_2+1}e^{-\beta_2 x}\right) \quad (4)$$

The survival and hazard function of the combined is given as:

$$S(x) = 1-\left[1-\alpha\left(e^{-\beta_1} + \frac{\beta_1 x}{\beta_1^2+\beta_1+1}e^{-\beta_1 x}\right) - (1-\alpha)\left(e^{-\beta_2} + \frac{\beta_2 x}{\beta_2^2+\beta_2+1}e^{-\beta_2 x}\right)\right]$$

$$S(x) = \alpha\left(e^{-\beta_1} + \frac{\beta_1 x}{\beta_1^2+\beta_1+1}e^{-\beta_1 x}\right) + (1-\alpha)\left(e^{-\beta_2} + \frac{\beta_2 x}{\beta_2^2+\beta_2+1}e^{-\beta_2 x}\right) \quad (5)$$

$$h(x) = \frac{\alpha\frac{\beta_1^2}{\beta_1^2+\beta_1+1}(1+\beta_1+x)e^{-\beta_1 x} + (1-\alpha)\frac{\beta_2^2}{\beta_2^2+\beta_2+1}(1+\beta_2+1)e^{-\beta_2 x}}{\alpha\left(e^{-\beta_1} + \frac{\beta_1 x}{\beta_1^2+\beta_1+1}e^{-\beta_1 x}\right) + (1-\alpha)\left(e^{-\beta_2} + \frac{\beta_2 x}{\beta_2^2+\beta_2+1}e^{-\beta_2 x}\right)} \quad (6)$$

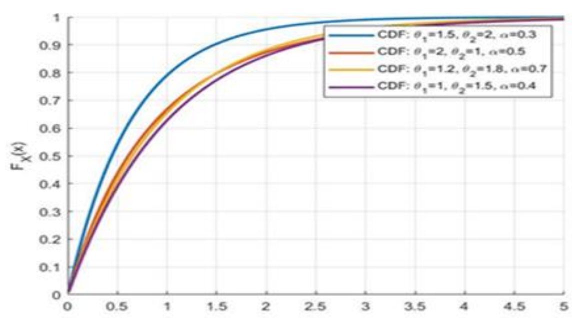


FIGURE1. CDF for the MKD

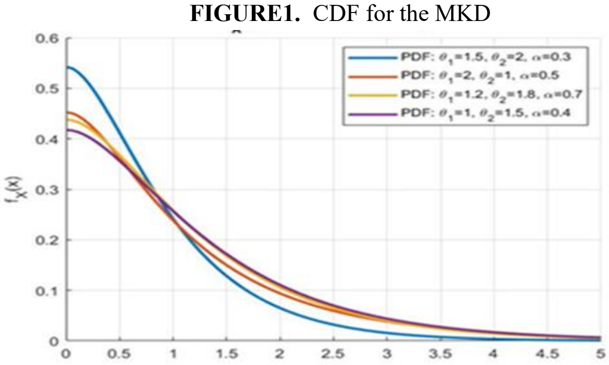


FIGURE 2. PDF for the MKD

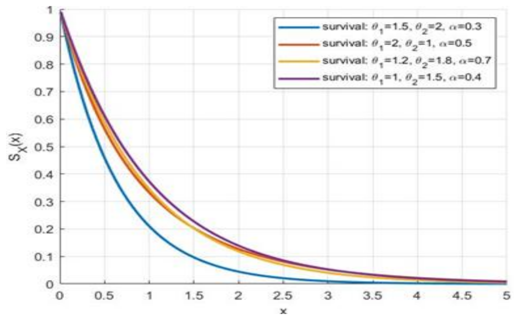


FIGURE3. Survival function for MKD

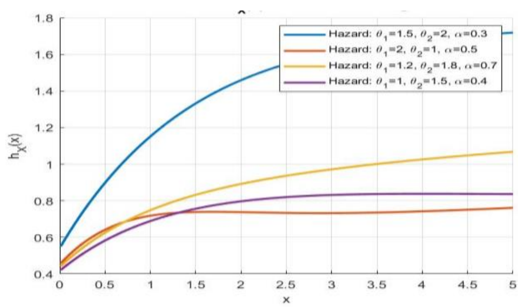


FIGURE 4. Hazard function for the MKD

HYBRID META-HEURISTIC ALGORITHMS

Many algorithms are used in optimization processes, and each algorithm has advantages and disadvantages that affect its performance in solving different problems. While some algorithms perform highly in a particular class of issues, they may be less effective when applied to other problems. This difference is not limited to problem classes but extends to specific cases within a single problem. In light of this, modern heuristics focus on two main components: exploration and exploitation. Exploration aims to search for new solutions in the solution space, while exploitation focuses on improving the discovered solutions to reach the optimal one. A new hybrid algorithm, PSOMO, has been developed to balance these two components, combining the Particle Swarm Optimization Algorithm (PSO) and the Monkey Algorithm (MO). The PSOMO algorithm is designed to estimate the parameters of a mixture of distributions based on the survival function. This approach improves computational efficiency, as diversity in the search is essential to avoid falling into local solutions, especially when the algorithm converges towards optimal solutions. Steps of the PSOMO algorithm: Generate initial solutions: A random set of solutions is generated as an initial step. Implement MO algorithm: It improves exploration by moving solutions towards the best solutions based on light intensity. Implement the PSO: It expands the search by replacing some solutions with new solutions according to the nesting strategy. Evaluate solutions: The updated solutions are evaluated using the objective function associated with the likelihood function.

$$\sum_{i=1}^{n} \log\left(\alpha \frac{\beta_1^2}{\beta_1^2+\beta_1+1}(1+\beta_1+x)e^{-\beta_1 x} \quad + (1-\alpha)\frac{\beta_2^2}{\beta_2^2+\beta_2+1}(1+\beta_2+x)e^{-\beta_2 t}\right)$$

The PSOMO algorithm updates solutions by combining the strengths of PSO and MO to balance exploration and exploitation. The stopping criterion was primarily a fixed number of iterations, determined through preliminary experiments that showed stable results beyond a certain point; in some cases, relative changes in the objective function were also considered. Parameter tuning for both PSO and MO components was conducted experimentally by testing various values (e.g., swarm size, inertia, moves), and selecting those that minimized the estimation error while maintaining stability.

SIMULATION STUDY

The effectiveness of the proposed method for parameter estimation was evaluated using simulation, where 1000 iterations were performed to generate each simulation case. To verify the effect of sample size on performance, different sample sizes were tested: 15, 25, 50, 75, 100, and 120. The performance was simulated based on the mean square error (MSE) criteria according to the following steps:

Step 1: Algorithm parameter initialization, all parameters of the PSO, MO, and hybrid algorithm (PSOMO)are adjusted.

Step 2: Generating Random Samples from a continuous uniform distribution on the interval (0,1). These samples are then transformed into samples that follow the Komal distribution using the cumulative distribution function as follows: $F(x, \beta_1, \beta_2) = U$ to find x, The numerical method (Newton-Raphson) was used. Then, define a vector that is used for all required parameters, such as Ӽ = [β_1, β_2, α], and generates N solutions for Ӽ.

Step 3: Recall the S from equation (5).

Step 4: The optimal value of $\hat{s}$ is determined using the PSO, MO, and PSOMO algorithms.

Step 5: Calculating the mean squared error (MSE). Based on L=1000 trials, the MSE is calculated as follows: $\text{MSE} = \left(\frac{1}{L}\sum_{i=1}^{L}\left(\hat{S}_i - S\right)^2\right)$.

PERFORMANCE COMPARISON

To determine the best method for the proposed estimation algorithms (PSO, MO, and PSOMO algorithms) of the mixture Komal Distribution, six sample sets (15, 25, 50, 75, 100, and 120) were used to estimate the Survival function based on parameters (β_1, β_2,α) of distributions. A simulation study was used to compare the proposed algorithms. Then, the simulation results for all methods, as shown in tables (1-3), depend on the MSE Survival analyses. The hybrid algorithm consistently provides lower values than MA and PSO, suggesting better optimization performance, especially at higher iterations. MA shows relatively stable values but lacks the precision of PSO or Hybrid methods. Hybrid tends to outperform MA and PSO in some cases (e.g., at n=25). This suggests the hybrid approach leverages the strengths of both methods effectively. Scalability: As n increases, the differences between the methods diminish slightly, but PSOMO still shows superior performance, particularly for n=75. PSO's performance increases with n, indicating potential robustness for more significant problems.

TABLE 1. MSR values of $\hat{s}$ when $\beta_1 = 1.5$, $\beta_2 = 2.0$ and $\alpha = 0.5$

n	PSO	MA	Hybrid
15	4.9841e-06	0.0059582	2.9806e-08
25	5.9321e-07	0.036868	6.3533e-08
50	2.1238e-08	0.013567	1.924e-07
75	7.0756e-0	0.027343	1.6455e-07
100	9.0393e-06	0.020765	4.8399e-07
120	1.6213e-07	0.023684	4.7027e-08

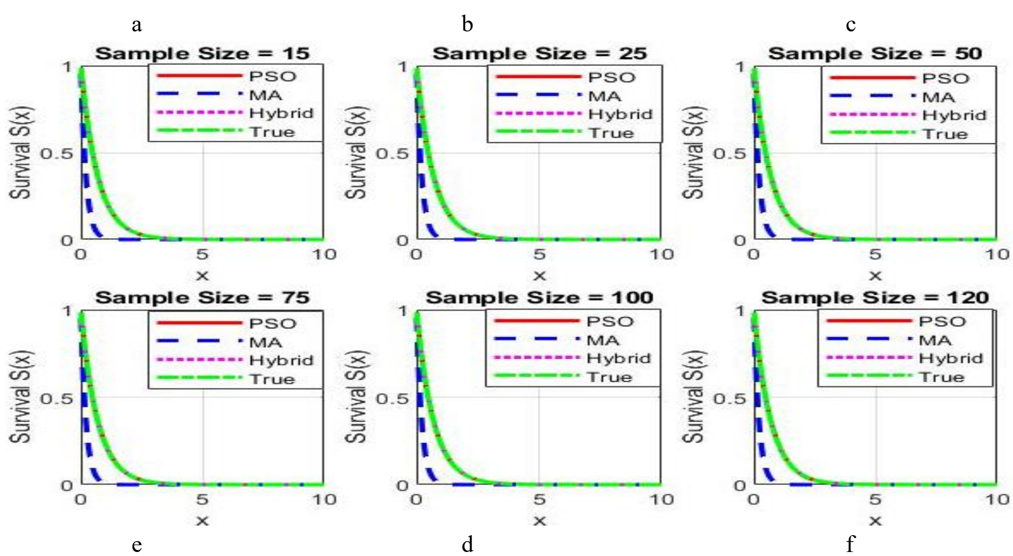
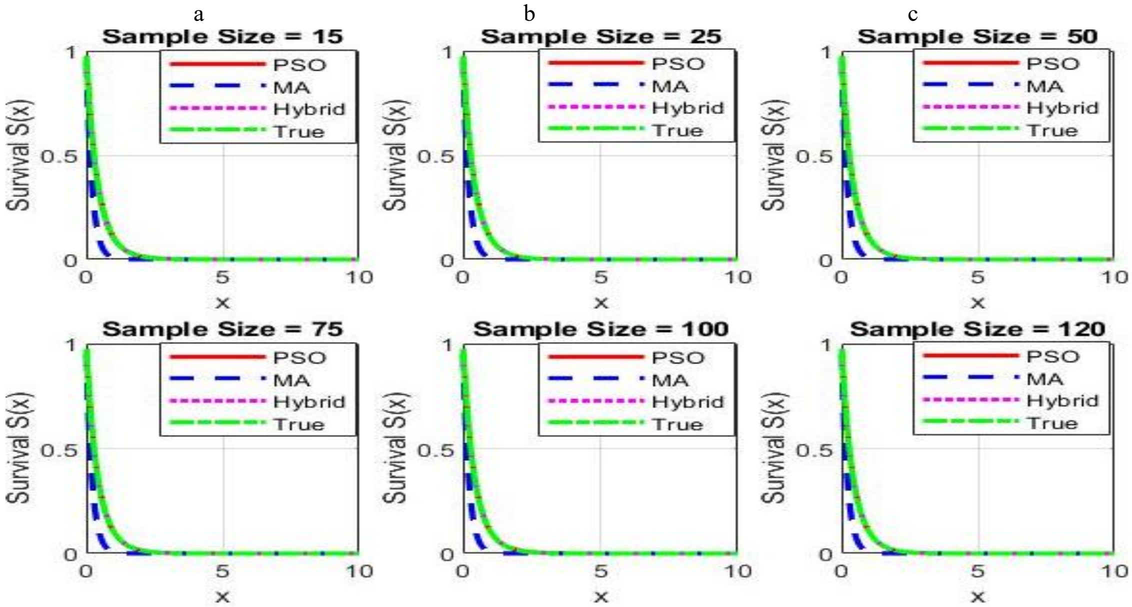


FIGURE 5a,b,c,d,e,f. Comparison of Survival Function for Different Sample Sizes

TABLE 2, MSR values of $\hat{s}$when $\beta_1 = 2 , \beta_1 = 3.5 , and\ \alpha = 0.5$

n	PSO	MA	Hybrid
15	3.4863e-07	0.0072613	1.4298e-07
25	6.0923e-08	0.0080372	5.2167e-09
50	1.2073e-07	0.011118	4.297e-07
75	5.4453e-07	0.0099121	1.0451e-07
100	1.3654e-07	0.0068967	4.6841e-08
120	9.6192e-07	0.0078046	1.1269e-07

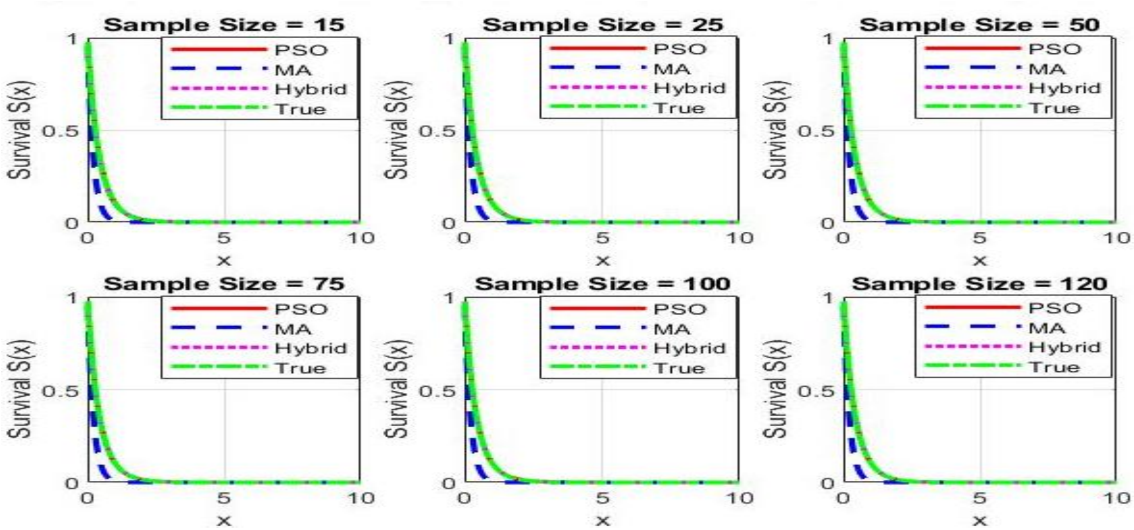
a b c



e d f

FIGURE 6a,b,c,d,e. Comparison of Survival Function for Different Sample Sizes

a b c



e d f

FIGURE 7. Comparison of Survival Function for Different Sample Sizes

TABLE 3. MSR values of $\hat{s}$when $\beta_1 = 1, \beta_1 = 2, and\ \alpha = 0.5$

n	PSO	MA	Hybrid
15	5.9779e-07	0.026817	8.779e-07
25	4.4703e-06	0.037547	1.7115e-06
50	8.3128e-07	0.045056	1.9904e-07
75	3.3542e-07	0.043468	3.1097e-08
100	1.4456e-07	0.051461	8.2879e-08
120	3.0152e-05	0.039987	2.2119e-06

CONCLUSION

This paper introduced a new statistical model based on the Komal distribution. Several statistical properties of the distribution are derived, including the survival function, probability density function, hazard function, cumulative distribution function, moments of rank, mean, variance, median, moment-generating function, and mode. In addition, the study proposed a new hybrid algorithm (PSOMO), which was developed by integrating the particle swarm optimization algorithm and the monkey algorithm. This algorithm is designed to enhance the estimation accuracy of the survival function using distribution parameters, thereby improving it over traditional methods. The simulation results showed that the hybrid algorithm (PSOMO) is accurate and significantly reduces the mean square error, making it a more efficient and effective option for estimating these functions. Despite these encouraging results, some limitations are worth noting, such as the potential difficulty in scaling up to large data sets and the dependence on assumptions about the shape of the distribution used. Future research suggests testing the algorithm using real data from medical or industrial applications, as well as comparing it with other probability distributions to verify its flexibility and accuracy in diverse contexts.

REFERENCES

1. A. Ibrahim and B. A. Kalaf, “Estimation of the survival function based on the log-logistic distribution,” International Journal of Nonlinear Analysis and Applications, vol. 13, no. 1, pp. 127–141, Mar. 2022. doi:10.22075/ijnaa.2022.5466
2. S. J. Almalki and J. Yuan, “A new modified Weibull distribution,” Reliability Engineering and System Safety, vol. 111, pp. 164–170, Jan. 2013. doi:10.1016/j.ress.2012.10.018
3. S. Al-Marzouki, “Investigation of nonlinear vibration in spacecraft structures,” Journal of Nonlinear Sciences and Applications, vol. 12, no. 8, pp. 543–551, Aug. 2019. doi:10.22436/jnsa.012.08.05
4. M. M. A. Almazah and M. Ismail, “Selection of efficient parameter estimation method for two-parameter Weibull distribution,” Mathematical Problems in Engineering, vol. 2021, Article 5806068, 2021. doi:10.1155/2021/5806068
5. N. H. Al-Noor, M. A. Khaleel, and G. J. Mohammed, “Theory and applications of Marshall–Olkin–Weibull distribution,” Journal of Physics: Conference Series, vol. 1999, p. 012101, Sep. 2021. doi:10.1088/1742-6596/1999/1/012101
6. J. A. Altawil, “[0,1] truncated Lomax–Lomax distribution with properties,” Journal of Kufa for Mathematics and Computer, vol. 8, no. 1, pp. 1–8, Mar. 2021. doi:10.31642/JoKMC/2018/080101
7. M. A. Shamran, A. Abdalhussain, I. H. Hussein, A. A. Mohammed, and I. H. Alkanani, “Comparison between Modified Weighted Pareto Distribution and Many Other Distributions,” Baghdad Science Journal, vol. 20, no. 3, pp. 816–988, Jul. 2023. doi:10.21123/bsj.2023.816988
8. G. J. M. Mahdi, B. A. Kalaf, and M. A. Khaleel, “Enhanced Supervised Principal Component Analysis for Cancer Classification,” Iraqi Journal of Science, vol. 62, no. 4, pp. 1321–1333, Apr. 2021. doi:10.24996/ijs.2021.62.4.28
9. B. A. Kalaf, S. H. Raheem, and A. N. Salman, “Estimation of the reliability system in model of stress-strength according to distribution of inverse Rayleigh,” Journal of Physics: Conference Series, vol. 1897, p. 012011, 2021.

10. A. F. Daghestani, K. S. Sultan, and A. S. Al-Moisheer, “Mixture of Lindley and Weibull distributions: Properties and estimation,” Journal of Statistics Applications and Probability, vol. 10, pp. 301–313, 2021. doi:10.18576/jsap/100203
11. R. A. Fisher, “The correlation between relatives on the supposition of Mendelian inheritance,” Annals of Eugenics, vol. 6, pp. 13–25, 1934. doi:10.1111/j.1469-1809.1934.tb02105.x
12. S. Abbas, M. Farooq, J. A. Darwish, S. H. Shahbaz, and M. Q. Shahbaz, “Analysis of COVID-19 dynamics using ARIMA and LSTM models,” Scientific Reports, vol. 13, 2023. doi:10.1038/s41598-023-48288-x
13. M. J. Mohammed and I. H. Hussein, “Statistical characterization of mechanical properties of composite materials,” IOP Conference Series: Materials Science and Engineering, vol. 571, p. 012014, 2019.
14. H. M. Yousof, M. M. Ali, K. Aidi, G. G. Hamedani, and M. A. Ibrahim, “A new family of skew-t distributions and its applications in Bayesian modeling,” Statistics, Optimization and Information Computing, vol. 10, pp. 519–547, 2022.
15. G. Muraleedharan, A. D. Rao, P. G. Kurup, N. U. Nair, and S. Mourani, “Wave transformation and sediment transport at coastal structures,” Coastal Engineering, vol. 54, pp. 630–638, 2007.
16. G. M. Cordeiro, E. M. Ortega, and S. Nadarajah, “A new lifetime distribution with decreasing and upside-down bathtub shapes,” Journal of the Franklin Institute, vol. 347, pp. 1399–1429, 2010.
17. E. M. Bereta, F. Louzada, and M. A. Franco, “The zero-inflated generalized Poisson distribution: Bayesian estimation and its applications,” Advances and Applications in Statistics, vol. 22, pp. 107–118, 2011.
18. S. B. Provost, S. Abdus, and A. Munir, “Modelling rainfall using the Pearson system,” Pakistan Journal of Statistics, vol. 27, pp. 111–131, 2011.
19. F. Merovci and I. Elbatal, “On the exponentiated generalized Rayleigh distribution: Properties and applications,” Applied Mathematics and Information Sciences, vol. 9, pp. 21–27, 2015.
20. D. Florence and P. Patrick, “Reliability evaluation of power grids under cascading failures,” Reliability Engineering and System Safety, vol. 180, pp. 453–462, 2018.
21. O. B. Oluyede, P. Mdlongwa, B. Makubate, and S. Huang, “Fitting generalized gamma distributions to rainfall data in South Africa,” Australian Journal of Statistics, vol. 48, pp. 1–13, 2019.
22. Z. A. Alkareem and B. A. Kalaf, “On the estimation of mixed Rayleigh–Weibull models,” Iraqi Journal of Science, pp. 1662–1680, 2022.
23. A.-J. An, S. Rahiem, and B. Kalaf, “Evaluation of educational outcomes based on statistical quality control,” Journal of College of Education, vol. 6, 2010.
24. K. Chi, J. Kang, K. Wu, and X. Wang, “Energy-efficient routing in intelligent networks,” in Proceedings of the International Conference on Intelligent Networking and Collaborative Systems, pp. 411–414, 2016. IEEE.
25. X. Wen, X. Yu-xia, S. Gui-xiang, and Z. Ying-zhi, “Dynamic modeling of electrical transport mechanisms in smart vehicles,” in Proceedings of the International Conference on Transportation, Mechanical, and Electrical Engineering, pp. 336–338, 2011. IEEE.
26. K. Al Abdullah, M. A. Batal, R. Hamdan, T. Khalil, and C. Salame, AIP Conf. Proc. 1968, 030079 (2018).
27. F.-K. Wang and P.-R. Huang, “Efficient simulation techniques for large-scale queuing networks,” Statistics and Computing Simulation, vol. 84, pp. 1975–1989, 2014.
28. B. A. Kalaf, S. H. Raheem, and A. N. Salman, “Estimation of the reliability system in model of stress–strength according to distribution of inverse Rayleigh,” Periodicals of Engineering and Natural Sciences (PEN), vol. 9, no. 2, pp. 524–533, 2021. doi: 10.21533/pen.v9i2.2067
29. S. H. Raheem, B. A. Kalaf, and A. N. Salman, “Comparison of some estimation methods of stress–strength model: R = P (Y < X < Z),” Baghdad Science Journal, vol. 18, no. 2 (Suppl.), pp. 1103–1103, 2021. doi: 10.21123/bsj.2021.18.2(Suppl.).1103
30. H. H. Örkcü, V. S. Özsoy, E. Aksoy, and M. I. Dogan, “Characterization of stress-strength models using epsilon-skew distributions,” Applied Mathematics and Computation, vol. 268, pp. 201–226, 2015.
31. B. A. Kalaf, G. J. Mohammed, and M. D. Salman, “Reliability assessment in inverse Rayleigh stress-strength models,” Journal of Physics: Conference Series, vol. 1897, p. 012011, 2021.