Available online at http://scik.org
J. Math. Comput. Sci. 11 (2021), No. 1, 414-426
https://doi.org/10.28919/jmcs/5146
ISSN: 1927-5307

A NEW TWO PARAMETER GAMMA-EXPONENTIAL MIXTURE

SAURAV SARMA[1,*], INZAMUL AHMED[2], ANJANA BEGUM[3]

[1]Department of Statistics, Gauhati University, Guwahati, Assam, India
[2]Department of Agricultural Statistics, Assam Agricultural University, Jorhat, India
[3]Mathematical Sciences Division, Institute of Advanced Study in Science and Technology, Guwahati 781035, Assam, India

Abstract. In this study, a new continuous distribution, which is a two-component finite mixture of exponential and gamma distribution, is suggested. Various properties of the proposed distribution, such as survival function, hazard rate function, moments, moment generating function, and characteristic function have been discussed. The maximum likelihood estimation is used for estimating the unknown parameters involved in this work. Finally, the proposed distribution is fitted to real-life data to illustrate its application, and a comparison against some existing distributions are drawn.

Keywords: finite mixture; gamma distribution; exponential distribution.

2010 AMS Subject Classification: 62E15, 62F10, 62F99, 62P30, 62P35.

1. Introduction

Defining new distribution to meet the need of explaining complex situations is a common practice in statistical theory. Often it is seen that a population may be a mix of two or more distinct distributions, and hence a single distribution may not explain the population accurately.

*Corresponding author
E-mail address: sarmasaurav@gauhati.ac.in
Received October 28, 2020

To explore such a population we take two or more known distributions in a definite proportion so that it can explain the population better.

Lindley (1958)(1965) [1] [2] introduced a two-component finite mixture taking one part exponential distribution and one part gamma, which later named after him as Lindley distribution. However, due to the simplicity and popularity of exponential distribution, Lindley distribution didn't grab attention for quite some time. Ghitany et al. (2007)[3] did a detailed study of Lindley's distribution and its statistical properties. They also showed that Lindley distribution performs better in describing the waiting time of bank customers compared to Exponential distribution.

Sattayatham and Talangtam [4] introduced the finite mixture lognormal distributions for the fitting of motor insurance claims data. Satsayamon et al. [5] presented a new family of generalized gamma distribution as a mixture of the generalized gamma and length biased generalized gamma distributions. Recently, Shanker et al. [6] [7] derived several distributions based on a finite mixture of gamma and exponential distribution.

In this study, a two-component finite mixture of gamma and exponential distributions is suggested. For convenience, the suggested distribution will be referred to as TPGEM distribution. The various statistical properties of TPGEM distribution are discussed. The parameter of the proposed mixture is estimated under the maximum likelihood method. Finally, to demonstrate its applicability, the proposed TPGEM distribution is fitted to a real-life data set related to the survival time of breast cancer patients.

2. Two Parameter Gamma-Exponential Distribution

Let us consider a two-component finite mixture of Exponential distribution with scale parameter 'k'and Gamma distribution with shape parameter 's'and scale parameter 'k'with their mixing proportions as $\frac{k}{1+k}$ and $\frac{1}{1+k}$ respectively.

The probability density function of the mixture can be written as,

$$f(x;s,k) = \frac{k^2}{1+k}\left[1+\frac{k^{s-2}}{\Gamma(s)}x^{s-1}\right]e^{-kx}; x \geq 0, k > 0, s > 1. \tag{1}$$

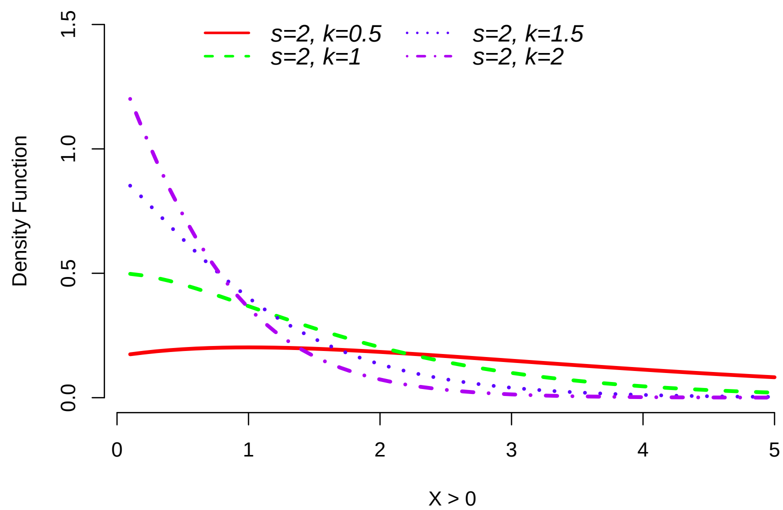

FIGURE 1. Density Plot of TPGEM Distribution for different Scale Parameters

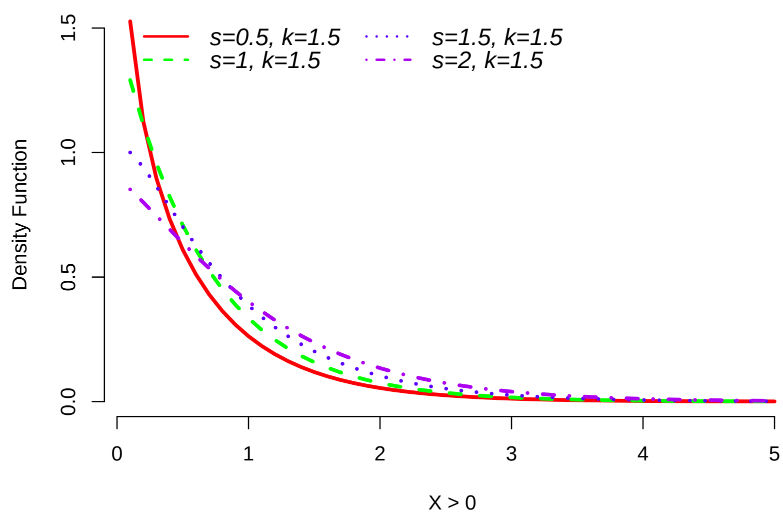

FIGURE 2. Density Plot of TPGEM Distribution for different Shape Parameters

The corresponding distribution function is given by,

$$F(x;s,k) = \frac{1}{1+k}\left[k(1-e^{-kx}) + P(s,kx)\right] \tag{2}$$

where, $P(a,b) = \frac{\gamma(a,b)}{\Gamma(a)}$ is the lower regularized gamma function.

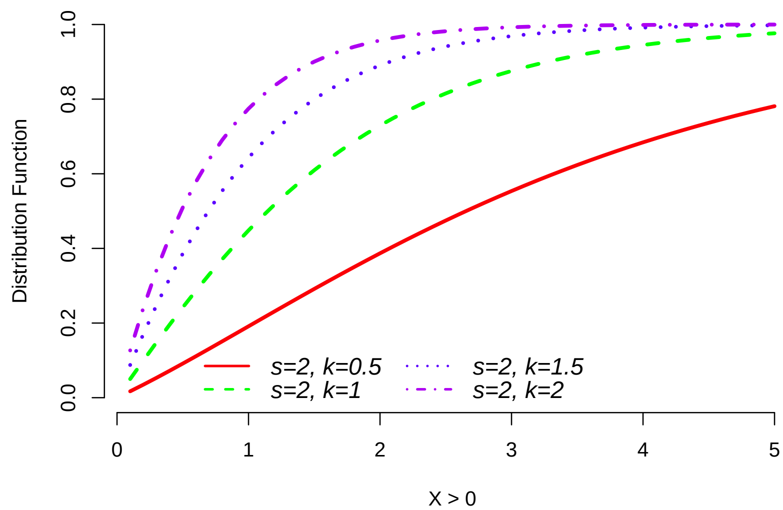

FIGURE 3. Density Plot of TPGEM Distribution for different Scale Parameter

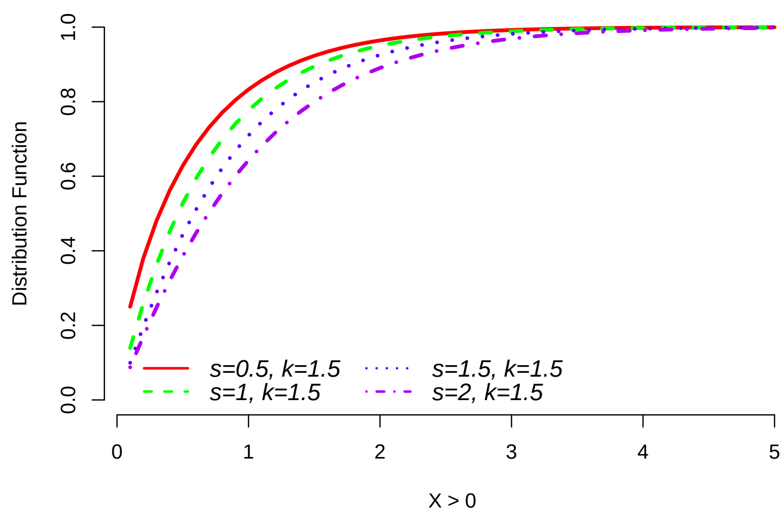

FIGURE 4. Density Plot of TPGEM Distribution for different Shape Parameter

3. MOMENTS OF TPGEM DISTRIBUTION

3.1. Raw Moments. The r^{th} raw moment is given by,

$$\mu_r^{'} = \frac{1}{k^r(1+k)}\left[k\Gamma(r+1) + \frac{\Gamma(r+s)}{\Gamma(s)}\right]. \tag{3}$$

Replacing $r = 1, 2, 3, 4$ in (3) we get the first four raw moments as,

$$\mu_1^{'} = mean = \frac{k+s}{k(1+k)}$$

$$\mu_2^{'} = \frac{2k^2 + s(s+1)}{k^2(1+k)}$$

$$\mu_3^{'} = \frac{6k + s(s+1)(s+2)}{k^3(1+k)}$$

$$\mu_4^{'} = \frac{24k + s(s+1)(s+2)(s+3)}{k^4(1+k)}.$$

3.2. Central Moments.

The corresponding central moments are,

$$\mu_2 = \frac{ks^2 - ks + s + 2k^3 + k^2}{k^2(1+k)^2}$$

$$\mu_3 = \frac{(k^2-k)s^3 + 6ks^2 + (-6k^3 - k^2 + k + 2)s - 6k^4 + 2k^3 + 12k^2 + 6k}{k^3(1+k)^3}$$

$$\mu_4 = \frac{\begin{array}{c}(k^3 - k^2 + k)s^4 + (2k^3 + 10k^2 - 4k)s^3 + (17k^3 + 13k^2 + 17k + 3)s^2 + \\ (24k^4 - 8k^3 - 40k^2 - 14k + 6)s + 12k^5 + 9k^4 + 24k^3 + 48k^2 + 24k\end{array}}{k^4(1+k)^4}$$

3.3. Skewness.

$$\begin{aligned}\beta_1 &= \frac{\mu_3^2}{\mu_2^3} \\ &= \frac{((k^2-k)s^3 + 6ks^2 + (-6k^3 - k^2 + k + 2)s - 6k^4 + 2k^3 + 12k^2 + 6k)^2}{(ks^2 - ks + s + 2k^3 + k^2)^3}\end{aligned}$$

3.4. Kurtosis.

$$\begin{aligned}\beta_2 &= \frac{\mu_4}{\mu_2^2} \\ &= \frac{\begin{array}{c}(k^3 - k^2 + k)s^4 + (2k^3 + 10k^2 - 4k)s^3 + (17k^3 + 13k^2 + 17k + 3)s^2 \\ + (24k - 8k^3 - 40k^2 - 14k + 6)s + 12k^5 + 9k^4 + 24k^3 + 40k^2 + 24k\end{array}}{(ks^2 - ks + s + 2k^3 + k^2)^3}\end{aligned}$$

Table (1) represent the values of mean, standard deviation, skewness, and kurtosis for different values of the shape parameter and scale parameter of TPGEM distribution. We see that the

TABLE 1. Summary Statistics of TPGEM distribution for Different Parameter Values

s	k	Mean	SD	Skewness	Kurtosis
0.5	0.5	1.33	1.56	80.50	35.30
	1.0	0.75	0.81	5.31	10.85
	1.5	0.53	0.65	0.04	4.66
	2.0	0.42	0.56	0.21	2.73
1.0	0.5	2.00	2.67	30.38	15.75
	1.0	1.00	1.00	4.00	9.00
	1.5	0.67	0.71	0.01	4.92
	2.0	0.50	0.58	0.31	3.12
1.5	0.5	2.67	4.22	14.70	9.68
	1.0	1.25	1.31	3.19	7.75
	1.5	0.80	0.83	0.03	5.16
	2.0	0.58	0.64	0.28	3.53
2.0	0.5	3.33	6.22	8.31	7.06
	1.0	1.50	1.75	2.62	6.80
	1.5	0.93	1.00	0.10	5.25
	2.0	0.67	0.72	0.15	3.87

variance of our proposed distribution is greater than mean. Hence we can say that our proposed TPGEM distribution is over dispersed.

4. GENERATING FUNCTIONS

4.1. Moment Generating Function. The moment generating function of TPGEM distribution is obtained by the expected value of e^{tX} where X follows TPGEM distribution given by (1) as

$$
\begin{aligned}
M_X(t) &= E[e^{tx}] \\
&= \int_0^\infty e^{tx} \frac{k^2}{1+k}\left[1+\frac{k^{s-2}}{\Gamma(s)}x^{s-1}\right]e^{-kx}dx \\
&= \frac{1}{1+k}\left[\frac{k^2}{k-t}+\frac{k^s}{(k-t)^s}\right]
\end{aligned}
$$

4.2. Characteristic Function. The corresponding characteristic function can also be obtain in a similar way as

$$
\begin{aligned}
\Phi_X(t) &= E[e^{itx}] \\
&= \int_0^\infty e^{itx} \frac{k^2}{1+k}\left[1+\frac{k^{s-2}}{\Gamma(s)}x^{s-1}\right]e^{-kx}dx \\
&= \frac{1}{1+k}\left[\frac{k^2}{k-it}+\frac{k^s}{(k-it)^s}\right]
\end{aligned}
$$

5. Survival Function

The survival function of the proposed TPGEM distribution is obtained as

$$
\begin{aligned}
S(x;k,s) &= 1-F(x;k,s) \\
&= 1-\left\{\frac{1}{1+k}\left[k(1-e^{-kx})+P(s,kx)\right]\right\}
\end{aligned}
\tag{4}
$$

6. Hazard Function

Similarly, we can obtain the hazard rate function of TPGEM distribution as

$$
\begin{aligned}
h(x;s,k) &= \frac{f(x;k,s)}{1-F(x;k,s)} \\
&= \frac{\frac{k^2}{1+k}\left[1+\frac{k^{s-2}}{\Gamma(s)}x^{s-1}\right]e^{-kx}}{1-\left\{\frac{1}{1+k}\left[k(1-e^{-kx})+P(s,kx)\right]\right\}}
\end{aligned}
\tag{5}
$$

7. Maximum Likelihood Estimation

Let $x_1, x_{21}x_3, \ldots, x_n$ be a random sample from TPGEM distribution with shape parameter s and scale parameter k, then the likelihood function corresponding to (1) is,

$$
L(x;k,s) = \frac{k^{2n}}{(1+k)^n}\prod_{i=1}^{n}\left\{1+\frac{k^{s-2}}{\Gamma(s)}x_i^{s-1}\right\}e^{-k\sum_{i=1}^{n}x_i} \tag{6}
$$

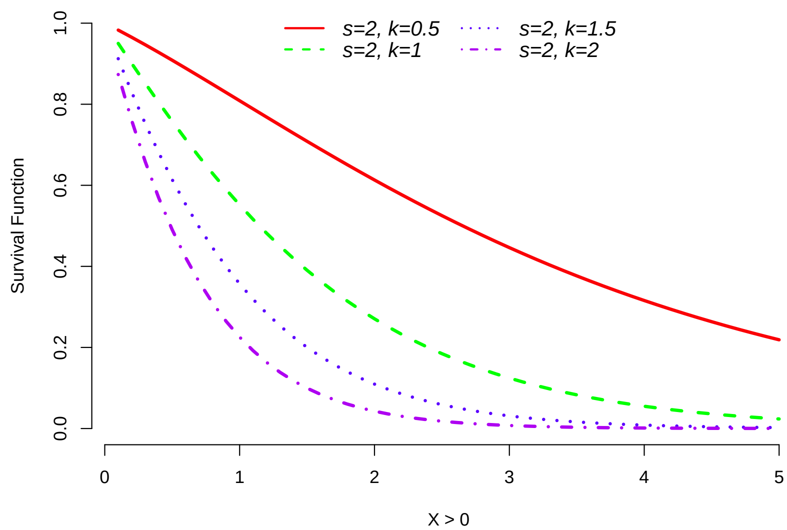

FIGURE 5. Survival Function of TPGEM Distribution for different Scale Parameter

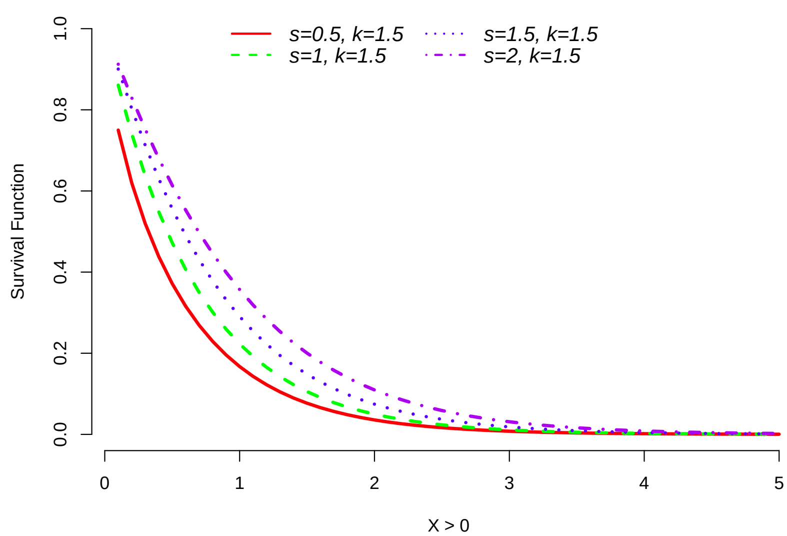

FIGURE 6. Survival Function of TPGEM Distribution for different Shape Parameter

The corresponding log likelihood function is,

$$\log L(x;k,s) = 2n\log k - n\log(1+k) + \sum_{i=1}^{n} \log\{\Gamma(s) + k^{s-2}\alpha_i^{s-1}\} - n\log\Gamma(s) - k\sum_{i=1}^{n} x_i. \tag{7}$$

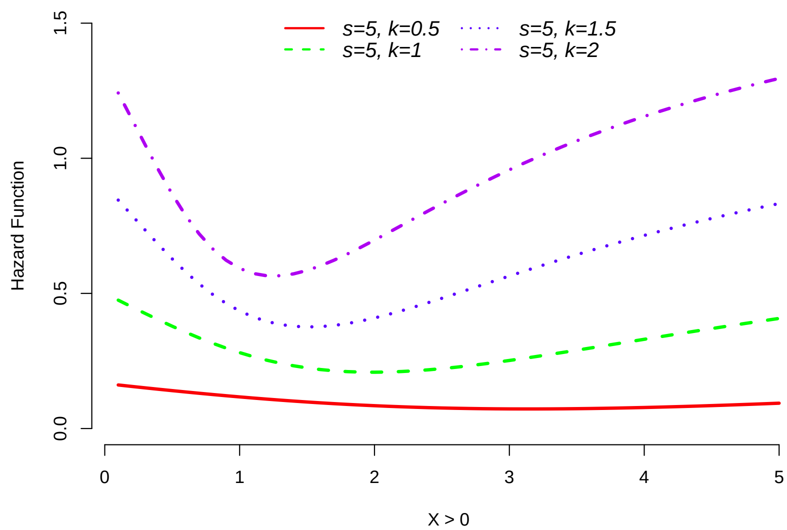

FIGURE 7. Hazard Function of TPGEM Distribution for different Scale Parameter

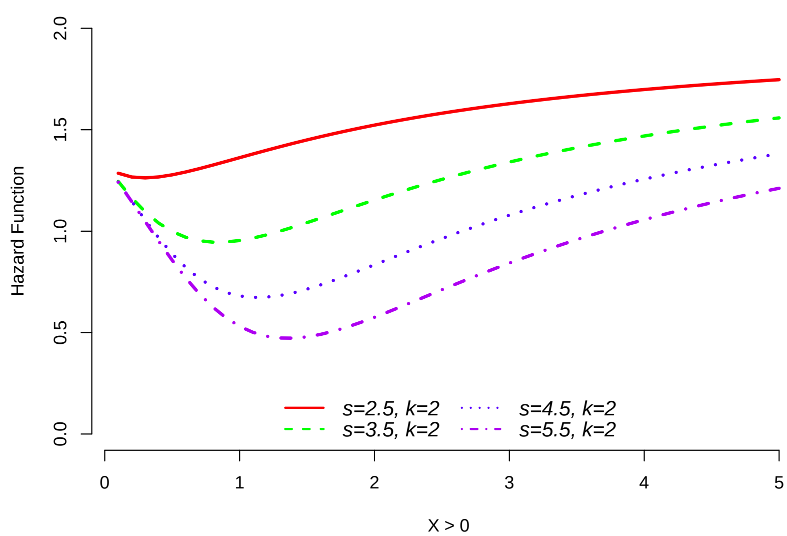

FIGURE 8. Hazard Function of TPGEM Distribution for different Shape Parameter

Simultaneous solution of $\frac{\partial \log L(x;k,s)}{\partial k} = 0$ and $\frac{\partial \log L(x;k,s)}{\partial s} = 0$ will give the MLE estimate of the unknown parameters k and s. However, clearly they does not have explicit analytical solution. Hence it can be solved numerically. The numerical estimates of $\hat{k}$ and $\hat{s}$ are obtained using R

TABLE 2. Goodness of Fit Statistics

Statistics	TPGEM	Exponential	Gamma	Lindley
KS	0.071	0.121	0.124	0.091
CVM	0.056	0.459	0.114	0.113
AD	0.403	2.692	0.532	1.011

statistical software.

8. FITTING TO REAL-LIFE DATA

To illustrate the application of our proposed TPGEM distribution in real life, the distribution is fitted to a data set of 121 breast cancer patients taken from Lee [8]. To compare the efficiency of the proposed distribution, it is compared with Exponential, Gamma and Lindley distribution in fitting the dataset. Table (2) presents the values of goodness-of-fit statistics Kolmogorov-Smirnov (KS), Cramer-von Mises and Anderson-Darling. Table (3) presents the values of goodness-of-fit criteria such as Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Corrected Akaike Information Criterion (AICC), Hannan-Quinn Information Criterion (HQIC), and Consistent Akaike Information Criterion, CAIC.

0.3, 0.3, 4.0, 5.0, 5.6, 6.2, 6.3, 6.6, 6.8, 7.4, 7.5, 8.4, 8.4, 10.3,11.0, 11.8, 12.2, 12.3, 13.5, 14.4, 14.4, 14.8, 15.5, 15.7, 16.2, 16.3, 16.5, 16.8, 17.2, 17.3, 17.5, 17.9, 19.8, 20.4, 20.9, 21.0, 21.0, 21.1, 23.0, 23.4, 23.6, 24.0, 24.0, 27.9, 28.2, 29.1, 30.0, 31.0, 31.0, 32.0, 35.0, 35.0, 37.0, 37.0, 37.0, 38.0, 38.0, 38.0, 39.0, 39.0, 40.0, 40.0, 40.0, 41.0, 41.0, 41.0, 42.0, 43.0, 43.0, 43.0, 44.0, 45.0, 45.0, 46.0, 46.0, 47.0, 48.0, 49.0, 51.0, 51.0, 51.0, 52.0, 54.0, 55.0, 56.0, 57.0, 58.0, 59.0, 60.0, 60.0, 60.0, 61.0, 62.0, 65.0, 65.0, 67.0, 67.0, 68.0, 69.0, 78.0, 80.0,83.0, 88.0, 89.0, 90.0, 93.0, 96.0, 103.0, 105.0, 109.0, 109.0, 111.0, 115.0, 117.0, 125.0, 126.0, 127.0, 129.0, 129.0, 139.0, 154.0

From Table (2) and Table (3) we can clearly state that, our proposed TPGEM distribution is explaining the dataset better than Exponential, Gamma and Lindley distributions.

TABLE 3. Goodness of Fit Criteria

Criteria	TPGEM	Exponential	Gamma	Lindley
AIC	1161.160	1172.255	1163.609	1163.863
BIC	1167.751	1175.051	1169.201	1166.659
AICC	1161.262	1172.286	1163.711	1163.897
HQIC	1166.495	1174.923	1168.945	1166.531
CAIC	1168.752	1176.051	1171.201	1167.659

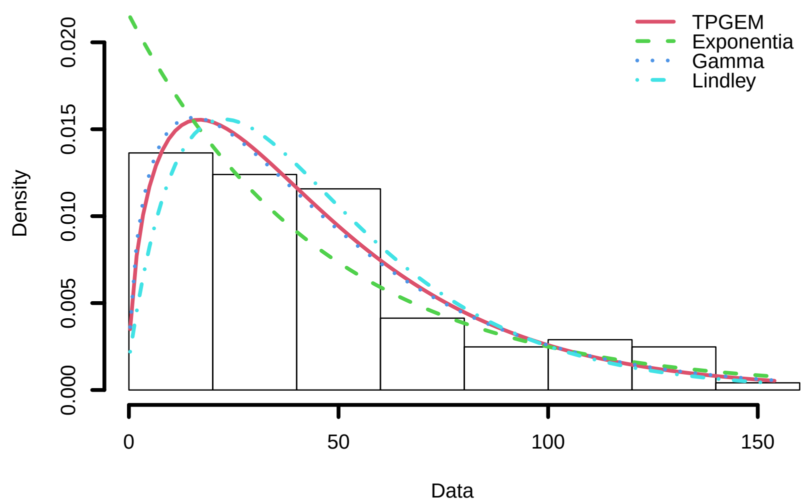

FIGURE 9. Density comparison between the fitted distributions in explaining the data

9. CONCLUSION

In this article, we have proposed a new continuous distribution called “TPGEM distribution”. The proposed TPGEM distribution is a two-component finite mixture of Exponential distribution with scale parameter k and Gamma distribution with shape parameter s and scale parameter k. The proposed distribution has Lindley distribution as its special case. Some of the statistical properties such as moments, generating functions, survival function, and hazard rate function was studied. The Maximum likelihood estimation of the unknown parameters involved in the model was studied and it was found out that a direct solution to the likelihood equations does

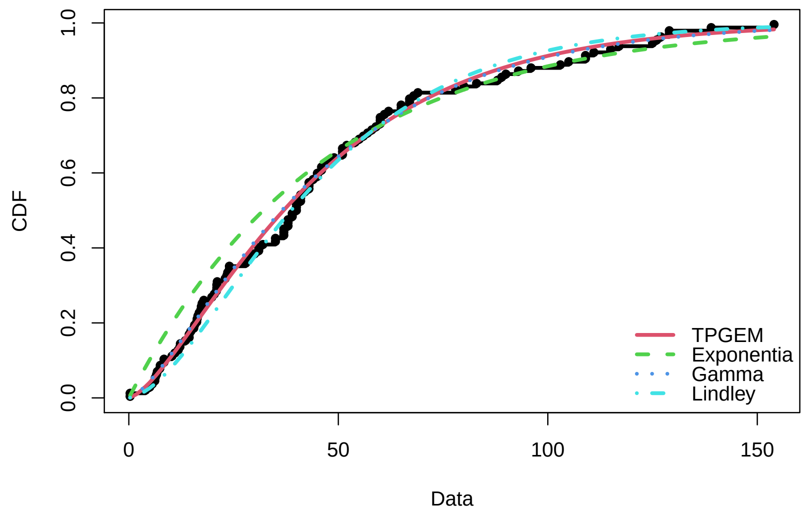

FIGURE 10. Distribution comparison between the fitted distributions in explaining the data

not exist and hence numerical method was used to estimate the parameters. Finally, applying the distribution to fit real-life data demonstrates its superiority compared to Exponential, Gamma, and Lindley distribution.

CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interests.

REFERENCES

[1] D.V. Lindley, Fiducial Distributions and Bayes' Theorem, J. R. Stat. Soc. Series B. 20 (1) (1958), 102-107.

[2] D.V. Lindley, Introduction to Probability and Statistics from a Bayesian Viewpoint, Part II: Inference, Cambridge University Press, New York, 1965.

[3] M. E. Ghitany, B. Atieh, and S. Nadarajah, Lindley distribution and its application, Math. Comput. Simul. 78 (2008), 493-506.

[4] P. Sattayatham, T. Talangtam, Fitting of Finite Mixture Distributions to Motor Insurance Claims, J. Math. Stat. 8 (1) (2012), 49-56.

[5] S. Suksaengrakcharoen, W. Bodhisuwan, A New Family of Generalized Gamma Distribution and Its Application, J. Math. Stat. 10 (2) (2014), 211-220.

[6] R. Shanker, Akash Distribution and its Applications, Int. J. Prov. Stat. 4 (3) (2015), 65-75.

[7] R. Shanker, K. Shukla, On two-parameter Akash Distribution, Biom. Biostat. Int. J. 6 (5) (2017), 416-425.

[8] R. A. Elton, E. T. Lee, Statistical Methods for Survival Data Analysis, Biometrics, 1995.