

# Komal distribution with properties and application in survival analysis

## Abstract

The modeling and analysis of lifetime data are becoming a challenge for the statistician and policy makers because the lifetime data are in general stochastic in nature. During recent decades several one parameter lifetime distributions have been proposed by researchers but they are not suitable due to the nature of the distribution and the stochastic nature of the data. In this paper an attempt has been made to propose a new one parameter lifetime distribution named Komal distribution. The statistical properties, estimation of parameter and application of the distribution to a lifetime dataset have been presented.

**Keywords:** lifetime distributions, statistical properties, estimation of parameter, applications

Volume 12 Issue 2 - 2023

**Rama Shanker**

Department of Statistics, Assam University, Silchar, Assam, India

**Correspondence:** Rama Shanker, Department of Statistics, Assam University, Silchar, Assam, India, Email shankerrama2009@gmail.com

**Received:** April 05, 2023 | **Published:** April 25, 2023

## Introduction

In the present era, modeling of lifetime data is a serious challenge because the lifetime data are stochastic in nature. It has been observed that policy makers are struggling to find a suitable distribution for lifetime data. During recent decades several one parameter lifetime distributions have been proposed in Statistics literature but due to distributional nature or the nature of the lifetime data, these proposed distributions do not give proper fit. Several researchers in the field of distribution theory are trying to propose a new lifetime distribution as per the stochastic nature of lifetime data. Upto 1958, there was only one lifetime distribution named exponential distribution which was in use for the analysis and modeling of lifetime data. Lindley<sup>1</sup> proposed another lifetime distribution known as Lindley distribution and Ghitany et al.<sup>2</sup> after detailed study on its statistical properties and application observed that Lindley distribution gives much closure fit than exponential distribution. While working on the comparative study of exponential and Lindley distribution, Shanker et al.<sup>3</sup> observed that exponential and Lindley distributions are competing well and there were some datasets where these two distributions do not provide good fit. Shanker<sup>4,5</sup> proposed two new one parameter lifetime distributions namely Shanker distribution and Akash distribution which gave much better fit than both exponential and Lindley distribution. Shanker et al.<sup>6</sup> provides a comparative study on applications of exponential, Lindley and Akash distribution. Further, Shanker & Hagos<sup>7</sup> presented a detailed study on applications of exponential, Lindley, Shanker and Akash distribution and showed that still there are some datasets where these distributions did not provide better fit. Further, Shanker<sup>8</sup> introduced Sujatha distribution which provides much better fit than exponential, Lindley, Shanker and Akash distribution. Again, Shanker<sup>9</sup> proposed another one parameter lifetime distribution named Garima distribution to model data arising from behavioral sciences, but this also does not give good fit on several real lifetime datasets. Now the question is to search a distribution which is both flexible and tractable in nature to capture the variation in the datasets. When a distribution does not give good fit, then some researchers prefer to transform the dataset to satisfy the assumptions of the distribution but this is not a preferable method because the original nature of the dataset gets lost. Some researchers also prefer to modify the distribution by adding extra shape parameter or scale parameter to suit the nature of the dataset. But, instead of transforming the original dataset or modifying the distribution suiting to dataset, it is better to search a

new distribution which provides better fit for the given datasets when the existing distributions fails to provide good fit.

In the present paper an attempt has been made to propose a new one parameter lifetime distribution, named Komal distribution, which would provide a better fit over exponential, Lindley, Shanker, Akash and Sujatha distributions. Some of its statistical properties, estimation techniques of parameter and an application to a real lifetime dataset has been discussed and presented.

## Komal distribution

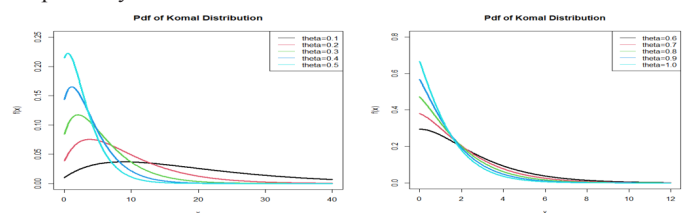
Taking the convex combination of exponential ( $\theta$ ) and gamma ( $2, \theta$ ) distributions with respective mixing proportions  $\frac{\theta(\theta+1)}{\theta^2 + \theta + 1}$  and  $\frac{1}{\theta^2 + \theta + 1}$ , a new probability density function (pdf) can be expressed as

$$f(x; \theta) = \frac{\theta^2}{\theta^2 + \theta + 1} (1 + \theta + x) e^{-\theta x}; x > 0, \theta > 0$$

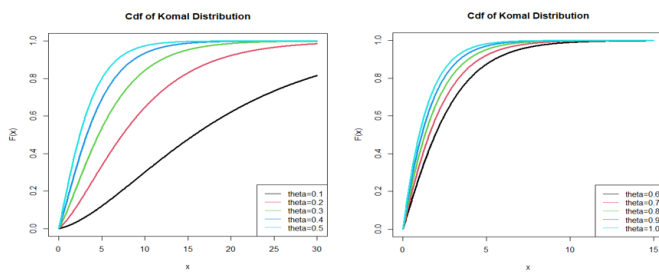
We would call this new distribution as 'Komal distribution'. Like other one parameter lifetime distributions, Komal distribution has been derived as a convex combination of exponential distribution and gamma distribution, it is expected to give better fit over exponential and other one parameter distributions derived using convex combinations of exponential distribution and gamma distribution. The cumulative distribution function (cdf) of Komal distribution can thus be obtained as

$$F(x; \theta) = 1 - \left[ 1 + \frac{\theta x}{\theta^2 + \theta + 1} \right] e^{-\theta x}; x > 0, \theta > 0$$

The behaviour of the pdf and the cdf of Komal distribution for varying values of parameter  $\theta$  have been presented in Figures 1 & 2 respectively.



**Figure 1** Graphs of the pdf of Komal distribution for selected values of the parameter.



**Figure 2** Graphs of the cdf of Komal distribution for selected values of the parameter.

### Descriptive measures of Komal distribution

As we know that moments are essential to know the descriptive nature such as coefficient of variation, skewness, kurtosis and index of dispersion of any distribution. Following the approach of obtaining the  $r$ th moment of Shanker distribution and Akash distribution by Shanker<sup>4,5</sup>, the  $r$ th moment about origin  $\mu_r'$  of Komal distribution can be obtained as

$$\begin{aligned} \mu_r' &= E(X^r) = \frac{\theta^2}{\theta^2 + \theta + 1} \int_0^\infty x^r (1 + \theta + x) e^{-\theta x} dx \\ &= \frac{r!(\theta^2 + \theta + r + 1)}{\theta^r (\theta^2 + \theta + 1)}; r = 1, 2, 3, \dots \end{aligned} \quad (3.1)$$

Substituting  $r = 1, 2, 3, 4$  in (3.1), the first four moments about origin of Komal distribution can be obtained as

$$\begin{aligned} \mu_1' &= \frac{\theta^2 + \theta + 2}{\theta(\theta^2 + \theta + 1)}, \quad \mu_2' = \frac{2(\theta^2 + \theta + 3)}{\theta^2(\theta^2 + \theta + 1)} \\ \mu_3' &= \frac{6(\theta^2 + \theta + 4)}{\theta^3(\theta^2 + \theta + 1)}, \quad \mu_4' = \frac{24(\theta^2 + \theta + 5)}{\theta^4(\theta^2 + \theta + 1)}. \end{aligned}$$

The moments about the mean of Komal distribution, using relationship between moments about the mean and the moments about the origin, can thus be obtained as

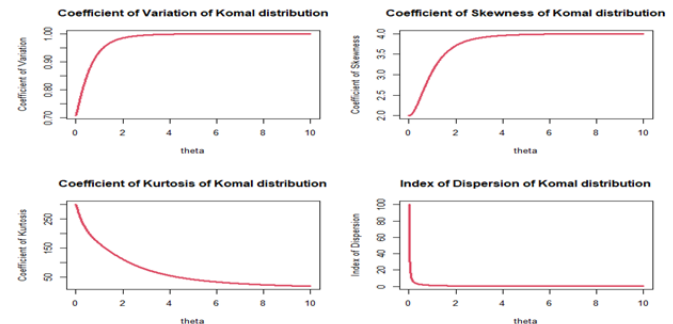
$$\begin{aligned} \mu_2 &= \frac{\theta^4 + 2\theta^3 + 5\theta^2 + 4\theta + 2}{\theta^2(\theta^2 + \theta + 1)^2} \\ \mu_3 &= \frac{2(\theta^6 + 3\theta^5 + 9\theta^4 + 13\theta^3 + 12\theta^2 + 6\theta + 2)}{\theta^3(\theta^2 + \theta + 1)^3} \\ \mu_4 &= \frac{3(3\theta^8 + 12\theta^7 + 434\theta^6 + 1260\theta^5 + 2471\theta^4 + 2856\theta^3 + 2428\theta^2 + 1208\theta + 400)}{\theta^4(\theta^2 + \theta + 1)^4}. \end{aligned}$$

The descriptive constants including coefficient of variation (CV), coefficient of skewness (CS), coefficient of kurtosis (CK) and the index of dispersion (ID) of Komal distribution are thus obtained as

$$\begin{aligned} CV &= \frac{\sqrt{\mu_2}}{\mu_1'} = \frac{\sqrt{\theta^4 + 2\theta^3 + 5\theta^2 + 4\theta + 2}}{\theta^2 + \theta + 2} \\ CS &= \frac{\mu_3^2}{\mu_2^3} = \frac{4(\theta^6 + 3\theta^5 + 9\theta^4 + 13\theta^3 + 12\theta^2 + 6\theta + 2)^2}{(\theta^4 + 2\theta^3 + 5\theta^2 + 4\theta + 2)^3} \\ CK &= \frac{\mu_4}{\mu_2^2} = \frac{3(3\theta^8 + 12\theta^7 + 434\theta^6 + 1260\theta^5 + 2471\theta^4 + 2856\theta^3 + 2428\theta^2 + 1208\theta + 400)}{(\theta^4 + 2\theta^3 + 5\theta^2 + 4\theta + 2)^2} \end{aligned}$$

$$ID = \frac{\mu_2}{\mu_1'} = \frac{\theta^4 + 2\theta^3 + 5\theta^2 + 4\theta + 2}{\theta(\theta^2 + \theta + 1)(\theta^2 + \theta + 2)}$$

Behaviour of coefficient of variation (CV), coefficient of skewness (CS), coefficient of kurtosis (CK) and index of dispersion (ID) of Komal distribution for changing values of parameter are shown in the Figure 3. The coefficient of variation and the coefficient of skewness are non-decreasing whereas the coefficient of kurtosis and the index of dispersion are non-increasing.



**Figure 3** Graph of CV, CS, CK and ID of Komal distribution for values of the parameter.

### Reliability properties of Komal distribution

#### Hazard rate function

The hazard rate function of a random variable  $X$  having pdf  $f(x; \theta)$  and cdf  $F(x; \theta)$  is defined as

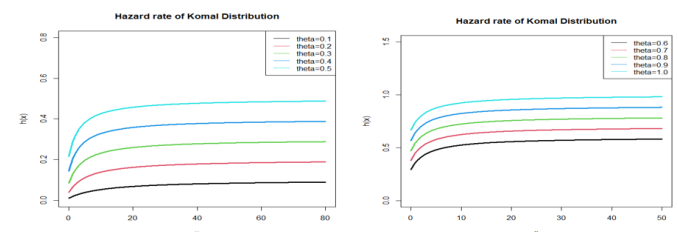
$$h(x, \theta) = \lim_{\Delta x \rightarrow 0} \frac{P(X < x + \Delta x | X > x)}{\Delta x} = \frac{f(x; \theta)}{1 - F(x; \theta)}$$

Thus, the hazard rate function of Komal distribution can be obtained as

$$h(x, \theta) = \frac{\theta^2(1 + \theta + x)}{(\theta^2 + \theta + 1 + \theta x)}$$

This gives  $h(0, \theta) = \frac{\theta^2(\theta + 1)}{\theta^2 + \theta + 1} = f(0, \theta)$ . The behaviour of the

hazard rate function of Komal distribution for various values of parameter  $\theta$  is shown in the following Figure 4. The hazard rate of Komal distribution is monotonically non-decreasing. Further, as the values of parameter increases, the hazard rate of Komal distribution scaled up.



**Figure 4** Graphs of the hazard rate function of Komal distribution for selected values of the parameter.

#### Mean residual life function

Let  $X$  be a random variable over the support  $(0, \infty)$  representing the lifetime of a system. Mean Residual life (MRL) function measures the expected value of the remaining lifetime of the system, provided it has survived up to time  $x$ . Let us consider the conditional random

variable  $X_x = (X - x | X > x); x > 0$ . Then, the MRL function, denoted by  $m(x)$ , is defined as

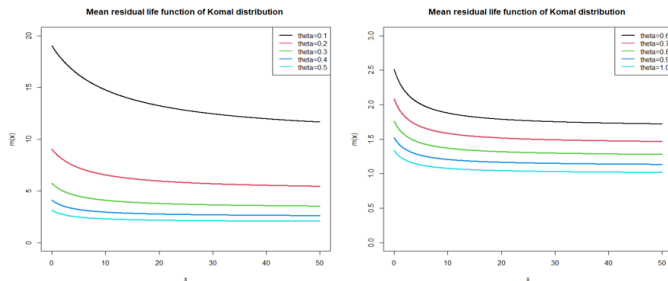
$$m(x) = E(X_x) = \frac{1}{S(x)} \int_x^\infty [1 - F(t)] dt$$

The MRL function of Komal distribution can thus be obtained as

$$m(x) = \frac{1}{\{\theta^2 + \theta + 1 + \theta x\} e^{-\theta x}} \int_x^\infty t(\theta^2(1 + \theta + t)) e^{-\theta t} dt - x = \frac{\theta^2 + \theta + 2 + \theta x}{\theta(\theta^2 + \theta + 1 + \theta x)}$$

This gives  $m(0) = \frac{\theta^2 + \theta + 2}{\theta(\theta^2 + \theta + 1)} = \mu_1'$ . The behaviour of the mean

residual life function of Komal distribution for various values of parameter  $\theta$  is shown in the following Figure 5. It is clear that the mean residual life function of Komal distribution is monotonically non-increasing.



**Figure 5** Graphs of the mean residual life function of Komal distribution for values of parameter.

### Stochastic ordering

In Probability theory and statistics, a stochastic order quantifies the concept of one random variable being bigger than another. A random variable  $X$  is said to be smaller than a random variable  $Y$  in the

- i. Stochastic order ( $X \leq_{st} Y$ ) if  $F_X(x) \geq F_Y(y)$  for all  $x$
- ii. Hazard rate order ( $X \leq_{hr} Y$ ) if  $h_X(x) \geq h_Y(y)$  for all  $x$
- iii. Mean residual life order ( $X \leq_{mrl} Y$ ) if  $m_X(x) \geq m_Y(y)$  for all  $x$
- iv. Likelihood ratio order ( $X \leq_{lr} Y$ ) if  $\frac{f_X(x)}{f_Y(y)}$  decrease in  $x$

The following results due to Shaked & Shantikumar<sup>10</sup> are well known for establishing stochastic ordering of distributions

$$X <_{lr} Y \Rightarrow X <_{hr} Y \Rightarrow X <_{mrl} Y \Rightarrow X <_{st} Y$$

**Theorem:** Let  $X \sim$  Komal distribution ( $\theta_1$ ) and  $Y \sim$  Komal ( $\theta_2$ ). If  $\theta_1 > \theta_2$ , then  $X <_{lr} Y$  hence  $X <_{hr} Y$ ,  $X <_{mrl} Y$  and  $X <_{st} Y$ .

**Proof:** We have

$$\frac{f_X(x; \theta_1)}{f_Y(x; \theta_2)} = \left[ \frac{\theta_1^2(\theta_2^2 + \theta_2 + 1)}{\theta_2^2(\theta_1^2 + \theta_1 + 1)} \right] \left( \frac{1 + \theta_1 + x}{1 + \theta_2 + x} \right)^{-(\theta_1 - \theta_2)x}$$

This gives

$$\log \left[ \frac{f_X(x; \theta_1)}{f_Y(x; \theta_2)} \right] = \log \left[ \frac{\theta_1^2(\theta_2^2 + \theta_2 + 1)}{\theta_2^2(\theta_1^2 + \theta_1 + 1)} \right] + \log \left( \frac{1 + \theta_1 + x}{1 + \theta_2 + x} \right)^{-(\theta_1 - \theta_2)x}$$

Therefore,  $\frac{d}{dx} \log \left[ \frac{f_X(x; \theta_1)}{f_Y(x; \theta_2)} \right] = \frac{\theta_2 - \theta_1}{(1 + \theta_1 + x)(1 + \theta_2 + x)} - (\theta_1 - \theta_2)$

Thus, for  $\theta_1 > \theta_2$ ,  $\frac{d}{dx} \log \left[ \frac{f_X(x; \theta_1)}{f_Y(x; \theta_2)} \right] < 0$ . this means that  $X <_{lr} Y$

hence  $X <_{hr} Y$ ,  $X <_{mrl} Y$  and  $X <_{st} Y$ .

## Parameter estimation of Komal distribution

### Method of moment estimate

Since Komal distribution has one parameter, equating the population mean to the corresponding sample mean ( $\bar{x}$ ), we get the third-degree polynomial equation of parameter  $\theta$  in the form

$$\bar{x} \theta^3 + (\bar{x} - 1) \theta^2 + (\bar{x} - 1) \theta - 2 = 0.$$

Solving this third-degree polynomial equation using Newton-Raphson method, we can easily get the moment estimate of the parameter.

### Maximum likelihood estimate

Suppose  $(x_1, x_2, x_3, \dots, x_n)$  be a random sample of size  $n$  from Komal distribution. The log likelihood function,  $\log L$  of Komal distribution is given by

$$\log L = \sum_{i=1}^n \log f(x_i; \theta) = n \{ 2 \log \theta - \log(\theta^2 + \theta + 1) \} + \sum_{i=1}^n \log(1 + \theta + x_i) - n \theta \bar{x}$$

The maximum likelihood estimate (MLE) ( $\hat{\theta}$ ) of the parameters ( $\theta$ ) of Komal distribution is the solution of the following log likelihood equation

$$\frac{d \log L}{d \theta} = \frac{2n}{\theta} - \frac{n(2\theta + 1)}{\theta^2 + \theta + 1} + \sum_{i=1}^n \frac{1}{1 + \theta + x_i} - n \bar{x} = 0$$

This gives

$$\sum_{i=1}^n \frac{1}{1 + \theta + x_i} + \frac{n(\theta + 2)}{\theta(\theta^2 + \theta + 1)} - n \bar{x} = 0.$$

This is a non-linear equation in  $\theta$ . This can be solved using Newton-Raphson method available in R software to get the maximum likelihood estimate (MLE) of the parameter  $\theta$  by taking the moment estimate of the parameter as the initial value. It should be noted that the method of moment estimate of the parameter will not be the same as that of the MLE.

## Application of Komal distribution

The application and the goodness of fit of Komal distribution has been discussed with a failure time dataset. Following failure time dataset has been considered.

Data set: The following skewed to right dataset relating to the failure times of 20 electric bulbs discussed by Murthy et al.<sup>11</sup> is considered and the observations are:

1.32, 12.37, 6.56, 5.05, 11.58, 10.56, 21.82, 3.60, 1.33, 12.62, 5.36, 7.71, 3.53, 19.61, 36.63,

0.39, 21.35, 7.22, 12.42, 8.92.

The values ML estimates of parameter and its standard error in parenthesis,  $-2 \log L$ , AIC (Akaike Information Criterion), AICC (Akaike Information Criterion corrected), BIC (Bayesian Information criterion), K-S (Kolmogorov-Smirnov) for the considered distributions

for the given dataset have been computed and presented in table 2. The formulae for computing AIC, AICC, BIC and K-S Statistics are as follows:

$$AIC = -2\log L + 2k, AICC = AIC + \frac{2k(k+1)}{n-k-1},$$

$$BIC = -2\log L + k \log n, D = \sup_x |F_n(x) - F_0(x)|$$

where  $k$  = number of parameter,  $n$  = sample size .

The pdf and the cdf of the fitted distributions are given in the Table 1.

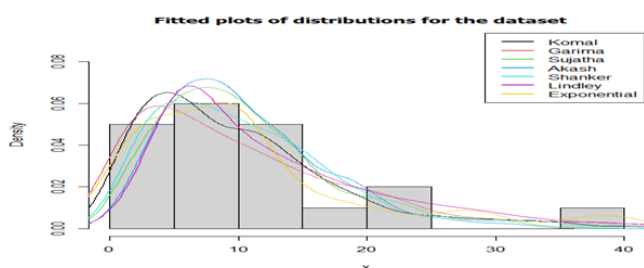
The fitted plots of considered distributions for the given datasets have been presented in Figure 6. The goodness of fit in Table 2 and the fitted plots of distributions for the dataset in figure 6 show that Komal distribution provides best fit over exponential, Lindley, Shanker, Akash and Sujatha distributions and therefore Komal distribution can be considered as the most suitable lifetime distribution for modeling lifetime data from biomedical science and engineering.

**Table 1** The pdf and the Cdf of fitted distributions

Distributions	Pdf	Cdf
Garima	$f(x; \theta) = \frac{\theta}{\theta + 2}(1 + \theta + \theta x)e^{-\theta x}$	$F(x; \theta) = 1 - \left(1 + \frac{\theta x}{\theta + 2}\right)e^{-\theta x}$
Sujatha	$f(x; \theta) = \frac{\theta^3}{\theta^2 + \theta + 2}(1 + x + x^2)e^{-\theta x}$	$F(x; \theta) = 1 - \left[1 + \frac{\theta x(\theta x + \theta + 2)}{\theta^2 + \theta + 2}\right]e^{-\theta x}$
Akash	$f(x; \theta) = \frac{\theta^3}{\theta^2 + 2}(1 + x^2)e^{-\theta x}$	$F(x; \theta) = 1 - \left[1 + \frac{\theta x(\theta x + 2)}{\theta^2 + 2}\right]e^{-\theta x}$
Shanker	$f(x; \theta) = \frac{\theta^2}{\theta^2 + 1}(\theta + x)e^{-\theta x}$	$F(x; \theta) = 1 - \left(1 + \frac{\theta x}{\theta^2 + 1}\right)e^{-\theta x}$
Lindley	$f(x; \theta) = \frac{\theta^2}{\theta + 1}(1 + x)e^{-\theta x}$	$F(x; \theta) = 1 - \left[1 + \frac{\theta x}{\theta + 1}\right]e^{-\theta x}$
Exponential	$f(x; \theta) = \theta e^{-\theta x}$	$F(x; \theta) = 1 - e^{-\theta x}$

**Table 2** ML estimates,  $-2\log L$ , AIC, AICC, BIC, K-S of the distribution for the dataset

Distributions	$\hat{\theta}$	$-2\log L$	AIC	AICC	BIC	K-S	P-value
Komal	0.1745 (0.0275)	133.33	135.33	135.90	135.52	0.0992	0.9914
Garima	0.1408 (0.0273)	133.18	135.18	135.75	135.37	0.1255	0.9218
Sujatha	0.2689 (0.0345)	137.54	139.54	140.11	139.73	0.1294	0.9037
Akash	0.2786 (0.0355)	138.47	140.47	141.04	140.66	0.1607	0.6434
Shanker	0.1885 (0.0292)	134.65	136.65	137.22	136.84	0.1172	0.9539
Lindley	0.1762 (0.0280)	133.44	135.44	136.01	135.63	0.1122	0.9684
Exponential	0.0952 (0.0212)	134.04	136.04	136.61	136.23	0.1255	0.9218



**Figure 6** Fitted plots of distributions considered for dataset.

### Conclusions and future works

In this paper a new lifetime distribution named Komal distribution for analysing and modeling lifetime data from biomedical science and engineering has been proposed. Some of its important statistical properties, estimation of parameter and application to a real lifetime dataset from survival analysis has been presented. Since the distribution is new one, it would be of great hope and expectation that

this will capture the attention of researchers working in biomedical science, engineering and insurance to model lifetime data in their respective fields. As the distribution has flexibility, tractability and practicability, future of the distribution would be quite bright among researchers in biomedical sciences and engineering.

### Acknowledgements

The author is thankful to the editor in chief of the journal and the anonymous reviewer of the paper for fruitful comments.

### Conflicts of interest

The author declares that there are no conflicts of interest.

### Funding

None.

### References

1. Lindley DV. Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society, Series B*. 1958;20(1):102-107.

2. Ghitany, M.E., Atieh, B, Nadarajah S. Lindley distribution and its application, mathematics computing and simulation. 2008;78:493–506.
3. Shanker R, Hagos F, Sujatha S. On modeling of lifetimes data using exponential and Lindley distributions. *Biom Biostat Int J.* 2015;2(5):140–147.
4. Shanker R. Shanker distribution and its applications. *International Journal of Statistics and Applications.* 2015a;5(6):338–348.
5. Shanker R. Akash distribution and its applications. *International Journal of Probability and Statistics.* 2015b;4(3):65–75.
6. Shanker R, Hagos F, Sujatha, S. On modeling of lifetime data using one parameter Akash, Lindley and exponential distributions. *Biom Biostat Int J.* 2016;3(2):54–62.
7. Shanker R, Hagos F. On modeling of lifetime data using Akash, Shanker, Lindley and exponential distributions. *Biom Biostat Int J.* 2016;3(6):214–224.
8. Shanker R. Sujatha distribution and its applications. *Statistics in Transition-New Series.* 2016a;17(3):391–410.
9. Shanker R. Garima distribution and Its application to model behavioral science data. *Biom Biostat Int J.* 2016b;4(7):275–281.
10. Shaked M, Shanthikumar JG. *Stochastic orders and their applications.* New York: Academic Press; 1994.
11. Murthy DNP, Xie M, Jiang R. *Weibull models.* New York: John Wiley & Sons; 2004.