ISSN: 0974-5823 Vol. 6 No. 3 December, 2021

*International Journal of Mechanical Engineering*

# Discrete Garima Distribution and its use in lifetime data

**Anushree Shukla[1*], Ajit Paul[1], Pramendra Singh Pundir[2]**

[1]*Department of Mathematics and Statistics, SHUATS, PRAYAGRAJ, India*

[2]*Department of Statistics, University of Allahabad, Allahabad, India*

**Abstract**

A one parameter discrete Garima distribution is derived corresponding to the one parameter continuous Garima distribution using infinite series approach for discretization of continuous probability distribution. The fundamental properties of this distribution such as moments, estimate of parameter from method of moments and method of maximum likelihood has been derived. Also some distributional characteristics as well as reliability properties such as hazard rate, second rate of failure and survival behavior has been discussed. At the, end the suitability of the proposed distribution has been discussed through various real life datasets.

**Keywords:** Discrete lifetime models, Garima distribution, reliability, failure rate, Maximum likelihood estimation.

## 1. Introduction-

In statistics, probability distributions are categorized on the basis of the nature of the variables. There are so many cases of real life which are only measurable in discrete nature. Almost in all the cases observed values are discrete because they are measured to only a finite number of decimal places and cannot really constitute all points in a continuum. Even if the measurements are taken on a continuous scale the observations may be recorded in a way making discrete model more appropriate. It is therefore reasonable to consider the observations as coming from a discretized distribution generated from the underlying continuous model.

There are several methods available in Statistics literature to derive a discrete distribution from a continuous distribution. One of the first proposed discretization methods is based on the definition of pmf that depends on an infinite series. The method of discretization by an infinite series was firstly considered by *Good [1]* who has proposed the discrete Good distribution to model the population frequencies of species and the estimation of parameters. This method of discretization has been explored by several authors Kulasekara and Tonkyn [2], Sato et al. [3] Nekoukhou et al.[4]. They have also proposed a new version of this method of discretization when the support of continuous random variable defined only on $R_+$. Thus if the random variable $Y$ defined on $R_+$ then the probability mass function (pmf) of random variable $X$ is defined as

$$P(X=x)=P(x)=\frac{f_Y(x,\theta)}{\sum_{i=0}^{\infty} f_Y(i,\theta)};\ X\in Z_+ \qquad (1)$$

In most of the cases, the discrete distributions obtained from this method are not in compact form due to their normalizing constant. Apart from the infinite series method of discretization, there is one more method available in literature for the discretization of continuous random variable. If the random variable $X$ has the survival function defined as $S(x)=P(X>x)$ then the probability mass function of discrete distribution associated with this continuous distribution can be written as

$$P(X=x)=S(x)-S(x+1);\ x=0,1,2,... \qquad (2)$$

There are several discrete distributions derived by *Krishna and Pundir [5]*, *Chakraborty and Chakraborty [6]* are available using this method of discretization.

This paper deals with the discretization of one parameter Garima distribution using infinite series method of discretization. Section 2, consists the derivation of functional form of discrete Garima distribution along with its probability mass function (pmf) and cumulative distribution function (cdf) with its behavior. The generating functions and the moments of this distribution have been derived in next Section 3 along with the behavior of coefficient of variation, skewness and kurtosis. In Section 4, we have discussed its reliability properties such as survival function, hazard rate and second rate of failure. The estimate of its parameter using method of moments and method of maximum likelihood has been discussed in next Section 5 and the last, Section 6 consist of usefulness of this distribution for different real life datasets.

## 2. Discrete Garima distribution-

The pdf and cdf of continuous random variable $Y$ having Garima distribution with parameter $\theta$ is introduced by *Shanker[7]* are given by

$$g(y,\theta)=\frac{\theta}{2+\theta}(1+\theta+\theta y)e^{-\theta y} \quad ;y>0,\theta>0 \qquad (3)$$

$$G(y,\theta)=1-\left(1+\frac{\theta y}{2+\theta}\right)e^{-\theta y} \quad ;y>0,\theta>0 \qquad (4)$$

It has increasing hazard rate and decreasing mean residual life time. Using the definition 2, probability mass function (pmf) of a discrete random variable $X$ corresponding to the continuous distribution $Y$ following Garima distribution (3) can be obtained as

$$P(X=x)=P(x,\theta)=\frac{\left(e^{\theta}-1\right)^{2}}{e^{\theta}\left(\theta e^{\theta}+e^{\theta}-1\right)}\left(1+\theta+\theta x\right)e^{-\theta x}\ ;\ \theta>0, x=0,1,2,\ldots \quad (5)$$

We would know this distribution as discrete Garima distribution (DGD).

## 3. Distributional Properties:

### 3.1 Behavior of pmf

The behavior and nature of pmf of DGD for the different values of $\theta$, is depicted in Fig 1. It can be seen that that the value of pmf is decreasing as the value of variable increases for the fixed values of parameter $\theta$ while pmf decreases for the increasing values of parameter $\theta$ at fixed values of random variable X. The behavior of cdf of DGD for varying value of parameter $\theta$ has been shown graphically in Fig 2.

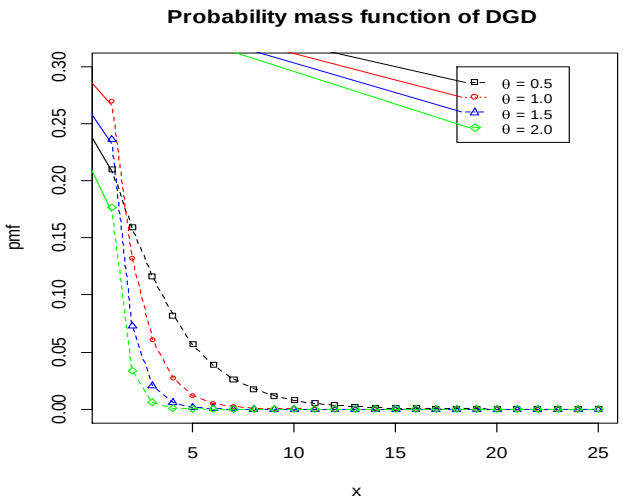


**Fig 1- Behavior of probability mass function of discrete Garima distribution**

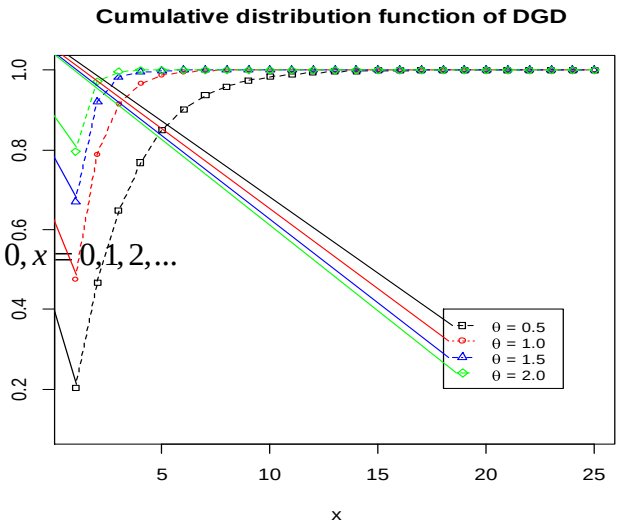


**Fig 2- Behavior of cumulative distribution function of discrete Garima distribution**

### 3.2 Generating functions and moments

The moments generating function (mgf) $M_X(t)$ can be obtained as

$$M_X(t)=E\left(e^{tx}\right)=\sum_{x=0}e^{tx}P(X=x)$$
$$=\frac{\left(e^{\theta}-1\right)^{2}}{\left(e^{\theta}+\theta e^{\theta}-1\right)}\frac{\left(\theta e^{\theta-t}+e^{\theta-t}-1\right)}{e^{t}\left(e^{\theta-t}-1\right)^{2}};t\neq\theta \quad (6)$$

With the help of mgf, one can easily obtain the Cumulants generating function as

$$K_X(t)=\log M_X(t) \quad (7)$$
$$=-t-2\log\left(e^{\theta-t}-1\right)-\log\left(\theta e^{\theta}+e^{\theta}-1\right)+\log\left(\theta e^{\theta-t}+e^{\theta-t}-1\right)+2\log\left(e^{\theta}-1\right)$$

Differentiating equation (7) w.r. to $t$, $'r'$ times and put $t=0$, we get $r^{th}$ cumulant as

$$\kappa_r=\left|\frac{d^r}{dt^r}K_X(t)\right|_{t=0}$$

$$\kappa_1=\left|\frac{d}{dt}K_X(t)\right|_{t=0}=\frac{2\theta e^{\theta}+e^{\theta}-1}{\left(e^{\theta}-1\right)\left(\theta e^{\theta}+e^{\theta}-1\right)}=mean\ldots(8)$$

$$\kappa_2=\left|\frac{d^2}{dt^2}K_X(t)\right|_{t=0}=\frac{\left\{e^{2\theta}\left(2\theta^2+3\theta+1\right)-e^{\theta}\left(2\theta+2\right)-\left(\theta-1\right)\right\}e^{\theta}}{\left(e^{\theta}-1\right)^{2}\left(\theta e^{\theta}+e^{\theta}-1\right)^{2}}=\mu_2(\text{variance}) \quad (9)$$

$$\kappa_3 = \left|\frac{d^3}{dt^3}K_X(t)\right|_{t=0} = \frac{\left\{\begin{matrix}(e^\theta-1)(\theta e^\theta+e^\theta-1)(2\theta^2e^{2\theta}+3\theta e^{2\theta}-3\theta-4\theta e^\theta+e^{2\theta}-4e^\theta+3)\\+2(\theta e^\theta+2e^\theta-2)(2\theta^2e^{2\theta}+3\theta e^{2\theta}-\theta-2\theta e^\theta+e^{2\theta}-2e^\theta+1)\end{matrix}\right\}e^\theta}{(e^\theta-1)^3(\theta e^\theta+e^\theta-1)^3} = \mu_3 \quad \text{.....(10)}$$

$$\kappa_4 = \left|\frac{d^4}{dt^4}K_X(t)\right|_{t=0} = \frac{-6(\theta+1)^4e^{4\theta}}{(\theta e^\theta+e^\theta-1)^4} + \frac{12(\theta+1)^3e^{3\theta}}{(\theta e^\theta+e^\theta-1)^3} - \frac{7(\theta+1)^2e^{2\theta}}{(\theta e^\theta+e^\theta-1)^2} + \frac{(\theta+1)e^\theta}{(\theta e^\theta+e^\theta-1)} - \frac{2e^\theta}{(e^\theta-1)} + \frac{14e^{2\theta}}{(e^\theta-1)^2} - \frac{24e^{3\theta}}{(e^\theta-1)^3} + \frac{12e^{4\theta}}{(e^\theta-1)^4}$$

(11)

And

$$\mu_4 = \kappa_4 + 3\kappa_2^2 = \frac{-6(\theta+1)^4e^{4\theta}}{(\theta e^\theta+e^\theta-1)^4} + \frac{12(\theta+1)^3e^{3\theta}}{(\theta e^\theta+e^\theta-1)^3} - \frac{7(\theta+1)^2e^{2\theta}}{(\theta e^\theta+e^\theta-1)^2} + \frac{(\theta+1)e^\theta}{(\theta e^\theta+e^\theta-1)} - \frac{2e^\theta}{(e^\theta-1)} + \frac{14e^{2\theta}}{(e^\theta-1)^2} - \frac{24e^{3\theta}}{(e^\theta-1)^3} + \frac{12e^{4\theta}}{(e^\theta-1)^4} + 3\left[\frac{\{e^{2\theta}(2\theta^2+3\theta+1)-e^\theta(2\theta+2)-(\theta-1)\}e^\theta}{(e^\theta-1)^2(\theta e^\theta+e^\theta-1)^2}\right]^2$$

(12)

All the central moments $(\mu_r; r=2,3,4)$ have been obtained from the cumulant generating function further it can be used to calculate the other descriptive measures of the distribution to characterize its properties.

### 3.3 Coefficient of dispersion, skewness and kurtosis

The coefficient of variation (CV) of the distribution can be obtained as the ration of standard deviation by its mean

$$CV = \frac{\sqrt{\mu_2}}{mean} = \frac{\sqrt{\{e^{2\theta}(2\theta^2+3\theta+1)-e^\theta(2\theta+2)-(\theta-1)\}e^\theta}}{2\theta e^\theta+e^\theta-1} \quad (13)$$

The Pearson's coefficients has been calculated to obtain the expression of skewness as kurtosis

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{\left[\begin{matrix}\left\{\begin{matrix}(e^\theta-1)(\theta e^\theta+e^\theta-1)(2\theta^2e^{2\theta}+3\theta e^{2\theta}-3\theta-4\theta e^\theta+e^{2\theta}-4e^\theta+3)\\+2(\theta e^\theta+2e^\theta-2)(2\theta^2e^{2\theta}+3\theta e^{2\theta}-\theta-2\theta e^\theta+e^{2\theta}-2e^\theta+1)\end{matrix}\right\}e^\theta\end{matrix}\right]^2}{\left[\{e^{2\theta}(2\theta^2+3\theta+1)-e^\theta(2\theta+2)-(\theta-1)\}e^\theta\right]^3} \quad (14)$$

and

$$\gamma_2 = \beta_2 - 3 = \frac{\mu_4}{\mu_2^2} - 3 \quad (15)$$

Using (14) and (15), one can calculate the values of skewness and kurtosis. From the Table-1, it can be seen that the discrete Garima distribution is over dispersive as its variance is greater than mean. The value of mean and variance are decreasing for increasing value of parameter $\theta$, while coefficient of variation (CV), skewness and kurtosis are increasing as the value of parameter increases. From the values of skewness, it can be interpreted that DGD is positive skewed distribution. As the value of kurtosis is highly positive, it can be said that is leptokurtic distribution and suitable for peaked over dispersive datasets.

**Table 1- Values of descriptive statistics of DGD for various values of parameter $\theta$**

| Value of $\theta$ | Mean | Variance | CV | Skewness | Kurtosis |
|---|---|---|---|---|---|
| 0.2 | 6.885929 | 43.07503 | 0.953126 | 1.587056 | 3.890781 |
| 0.4 | 3.14784 | 10.57209 | 1.032923 | 1.588768 | 4.011516 |
| 0.6 | 1.910652 | 4.597185 | 1.122185 | 1.637683 | 4.165034 |
| 0.8 | 1.299262 | 2.520193 | 1.221857 | 1.748809 | 4.359449 |
| 1.0 | 0.938554 | 1.565143 | 1.332962 | 1.940285 | 4.604327 |
| 1.2 | 0.703403 | 1.049785 | 1.45662 | 2.234553 | 4.911039 |
| 1.4 | 0.540107 | 0.741258 | 1.594062 | 2.659949 | 5.293194 |
| 1.6 | 0.421751 | 0.542651 | 1.746644 | 3.25279 | 5.767171 |
| 1.8 | 0.333328 | 0.407827 | 1.915869 | 4.060091 | 6.352773 |
| 2.0 | 0.265792 | 0.312556 | 2.103398 | 5.143092 | 7.074006 |

## 4. Reliability characteristics:

### 4.1 Survival function

The survival function can be calculated as

$$S(x) = P(X>x) = \frac{\{(\theta x+1)(e^\theta-1)+\theta(2e^{2\theta}-1)\}e^{-\theta x}}{e^\theta(\theta e^\theta+e^\theta-1)}; \theta>0, x=0,1,2,... \quad (16)$$

From the figure- is can be seen that the survival of discrete Garima distribution

### 4.2 Hazard rate or mortality rate

The hazard rate is the measure of failure or mortality in indefinite small time interval and can be

calculated as

$$h(x) = \frac{P(X=x)}{S(x)} = \frac{(e^\theta-1)^2(1+\theta+\theta x)}{(\theta x+1)(e^\theta-1)+\theta(2e^{2\theta}-1)}; \theta>0, x=0,1,2,... \quad (17)$$

From the Fig 3 is can be observed that DGD has increasing hazard rate or mortality rate for any value of parameter. As the value of parameter increases, the hazard rate increases rapidly.

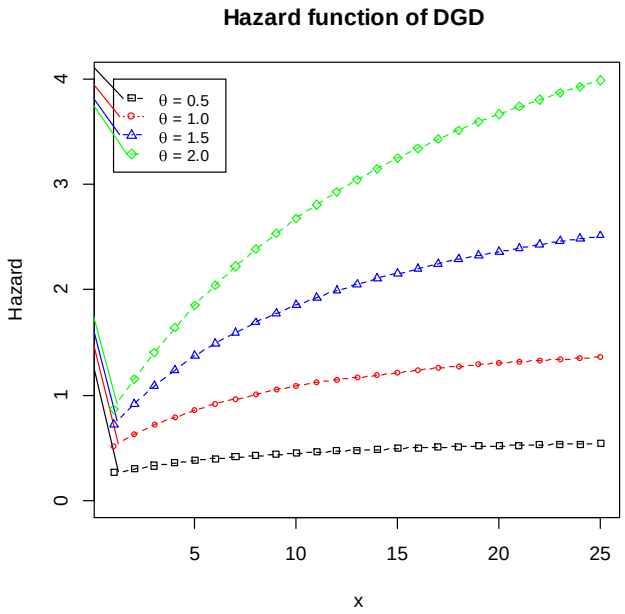


**Fig 3- Behavior of hazard function of discrete Garima distribution**

### 4.3 Second rate of failure

The hazard rate defined in (17) is bounded and hence cannot be convex also it is not additive for series system. The second rate of failure (SRF) defined as $SRF(x)=\log\left(S(x)/S(x+1)\right)$ was introduced by Roy and Gupta [8] (also used by Xie et al. [9]) to overcome these inherent problems of the failure rate.

For DGD

$$SRF(x)=\log\left(\frac{S(x)}{S(x+1)}\right)=\log\left(\frac{(\theta x+1)\left(e^{\theta}-1\right)+\theta\left(2e^{2\theta}-1\right)}{\left\{(\theta x+\theta+1)\left(e^{\theta}-1\right)+\theta\left(2e^{2\theta}-1\right)\right\}e^{-\theta}}\right)$$

## 5. Parameter estimation:

### 5.1 Method of moments estimation

In this method one makes the equation of sample moments and population moments and solve for the estimates of parameters. For DGD, equating the sample mean equals to population mean to get the moments estimator, $\hat{\theta}$ of parameter $\theta$,

$$E(X)=\bar{X}=\frac{2\theta e^{\theta}+e^{\theta}-1}{\left(e^{\theta}-1\right)\left(\theta e^{\theta}+e^{\theta}-1\right)}$$

$$(\theta\bar{x}+\bar{x})e^{2\theta}-(2\bar{x}+\bar{x}+2\theta+1)e^{\theta}+(\bar{x}+1)=0 \quad (18)$$

Solving the equation (18), one can get the moment estimator $\hat{\theta}$ of parameter $\theta$.

### 5.2 Method of maximum likelihood

The method of maximum likelihood consists of maximizing the likelihood function to get the estimator of parameter. For the given sample $x_1, x_2, ..., x_n$ of size $n$, the Likelihood function of the parameter of DGD is as follows

$$L=\prod_{i=1}^{n}\mathrm{P}(x_i,\theta)$$
$$=\prod_{i=1}^{n}\frac{\left(e^{\theta}-1\right)^2}{e^{\theta}\left(\theta e^{\theta}+e^{\theta}-1\right)}\left(1+\theta+\theta x_i\right)e^{-\theta x_i} \quad (19)$$

Taking log both side for log likelihood function

$$\log L=n\log\left(\frac{\left(e^{\theta}-1\right)^2}{e^{\theta}\left(\theta e^{\theta}+e^{\theta}-1\right)}\right)+\sum_{i=1}^{n}\log\left(1+\theta+\theta x_i\right)-\theta\sum_{i=1}^{n}x_i \quad (20)$$

Differentiating equation (20) wrt to $\theta$ and equating equals to zero can get the log likelihood equation

$$\frac{-n\left(\theta e^{\theta}+2e^{\theta}\right)}{\theta e^{\theta}+e^{\theta}-1}+\frac{2ne^{\theta}}{e^{\theta}-1}+\sum_{i=1}^{n}\left(\frac{x_i+1}{\theta x_i+\theta+1}\right)-\sum_{i=1}^{n}x_i=0 \quad (21)$$

Solving the non-linear equation (21), one can get the ML estimate of the parameter $\theta$.

## 6. Goodness of fit on real data

As DGD is a over-dispersed model, so it can be used for the modeling of over-dispersed data. In general biological data are over-dispersed hence applicability of this model has been shown for two real biological data sets. The first dataset is the data regarding number of Homocytometer yeast cell counts per square, available in Gosset [10] and the second data is regarding frequencies of the observed number of days that experienced X thunderstorm events at Cape kennedy, Florida for the 11-year period of record in the summer, January 1957 to December 1967 [11-12]. The applicability of DGD has been compared with Poisson distribution (PD) which is equi-dispersed model and Poisson-Lindley distribution (PLD) [13] which is over-dispersed distribution. The chi-square distribution has been used to test the significance of goodness of fit. From Table 3 and Table 4, is been observed that DGD is better fit than PD and PLD hence it can be used for the modeling of over-dispersed data.

**Table 3-** Observed and expected number of homocytometer yeast cell counts per square observed by gusset

| Number of red mites | Observed Frequency | Expected frequency | | |
|---|---|---|---|---|
| | | PD | PLD | DGD |
| 0 | 213 | 202.1 | 234 | 230.06 |
| 1 | 128 | 138 | 99.4 | 104.66 |
| 2 | 37 | 47.1 | 40.5 | 41.6 |
| 3 | 18 | 10.7 | 16 | 15.4 |
| 4 | 3 | 1.8 | 6.2 | 5.46 |
| 5 | 1 | 0.2 | 2.4 | 1.88 |
| 6 | 0 | 0.1 | 1.5 | 0.63 |
| ML estimate | | 0.6825 | 1.9502 | 1.2265 |
| Chi-square | | 10.09 | 11.06 | 9.56 |
| p-value | | 0.0178 | 0.0114 | 0.1443 |

Table 4- Frequencies of the observed number of days that experienced X thunderstorm events at Cape kennedy, Florida for the 11-year period of record in the summer, January 1957 to December 1967.

| Number of red mites | Observed Frequency | Expected frequency | | |
|---|---|---|---|---|
| | | PD | PLD | DGD |
| 0 | 549 | 449.2 | 511.8 | 534.02 |
| 1 | 246 | 364.9 | 295.7 | 271.6 |
| 2 | 117 | 148.2 | 128.1 | 121.84 |
| 3 | 67 | 40.1 | 49.3 | 51.09 |
| 4 | 25 | 8.1 | 17.8 | 20.53 |
| 5 | 7 | 1.3 | 6.2 | 8.01 |
| 6 | 1 | 0.2 | 3.1 | 3.06 |
| ML estimate | | 0.812253 | 1.24195 | 1.09685 |
| Chi-square | | 142.57 | 21.47 | 10.47 |
| p-value | | 0.000 | 0.0007 | 0.106 |

## 7. Conclusions

In this article we have discussed a new one parameter discrete Garima distribution (DGD), which is an over-dispersed model and discrete analogues of continuous Garima distribution. Its generating function such as moment generating function, cumulants generating function have been derived and using this we have also obtained its moments based descriptive statistics such as mean, variance along with skewness and kurtosis. The parameter estimation has been performed using method of moments and maximum likelihood estimation techniques. The applicability of this model has been shown using two real dataset and found that it is better model than PD and PLD for over-dispersed dataset.

**References-**


1. Good LJ. The population frequencies of species and the estimation of population parameters. *Biometrika*. 1953;40:237–264. https://doi.org/10.2307/2333344
2. Kulasekara KB, Tonkyn DW. A new discrete distribution with application to survival, dispersal and dispersion. Commun Stat Simul Comput. 1992;21:499–518. https://www.tandfonline.com/doi/abs/10.1080/03610919208813032
3. Sato H, Ikota M, Aritoshi S. A new defect distribution in meteorology with a consistent discrete exponential formula and its applications. IEEE Trans SemicondManufactur.1999;12(4):409–418. **DOI:** 10.1109/66.806118
4. Nekoukhou VM, Alamatsaz MH, Bidram H. Discrete Generalized exponential distribution. Communications

in Statistics-Theory & Methods. 2012;41:2000–2013 https://doi.org/10.1080/03610926.2011.555044
5. Krishna, H., & Pundir, P. S. (2009). Discrete Burr and discrete Pareto distributions. Statistical Methodology, 6(2), 177-188. DOI: 10.1016/j.stamet.2008.07.001
6. Chakraborty S, Chakravarty D. (2012) Discrete Gamma Distributions: Properties and Parameter Estimations, Communications in Statistics - Theory and Methods, 41:18, 3301-3324, DOI: 10.1080/03610926.2011.563014
7. Shanker R. Garima distribution and its application to model behavioral science data. Biom Biostat Int J. 2016;4(7):275–281. DOI: 10.15406/bbij.2016.04.00116
8. Roy, D. (2002). Discretization of continuous distributions with an application to stress-strength reliability. Calcutta Statistical Association Bulletin, 52(1-4), 297-314.
9. Xie, M., Gaudoin, O., Bracwuemond, C. (2002). Redefining failure rate function for discrete distribution. Int. J. Reliab. Quali. Safety Eng. 9(3):275–285.
10. Gosset WS. The probable error of the mean. Biometrika. 1908;6(1):1–25.
11. Falls LW, Wilford WO, Carter MC. Probability distribution for thunderstorm activity at Cape Kennedy, Florida. Journal of Applied, Meteorology 1971;10:97-104.
12. Carter MC. A model for thunderstorm activity: use of the compound negative binomial-positive binomial distribution. Journal of the Royal Statistical Society, Series C (Applied Statistics) 2001;2:196-201.
13. Sankaran M. The discrete Poisson-Lindley distribution. Biometrics. 1970;26:145–149.