MedCrave
Step into the World of Research
Biometrics & Biostatistics International Journal

Research Article
Open Access
CrossMark

Size-biased poisson-garima distribution with applications

Abstract

In this paper, a size-biased Poisson-Garima distribution (SBPGD) has been obtained by size-biasing the Poisson-Garima distribution (PGD) introduced by Shanker (2017). The moments about origin and moments about mean have been obtained and hence expressions for coefficient of variation (C.V.), skewness and Kurtosis have been obtained. The estimation of its parameter using the method of moment and the method of maximum likelihood estimation has been discussed. The goodness of fit of SBPGD has been discussed for two real data sets using maximum likelihood estimate and the fit shows quite satisfactory over size-biased Poisson distribution (SBPD) and size-biased Poisson-Lindley distribution (SBPLD).

Keywords: garima distribution, poisson-garima distribution, size-biasing, moments, estimation of parameter, goodness of fit

Volume 6 Issue 3 - 2017

Rama Shanker, Kamlesh Kumar Shukla
Department of Statistics, Eritrea Institute of Technology, Eritrea

Correspondence: Rama Shanker, Department of Statistics, Eritrea Institute of Technology, Asmara, Eritrea, Email shankerrama2009@gmail.com

Received: August 23, 2017 | **Published:** August 28, 2017

Introduction

Shanker[1] has obtained Poisson-Garima distribution (PGD) for modeling count data having probability mass function (p.m.f.)

$$P_0(x;\theta)=\frac{\theta}{\theta+2}\frac{\theta x+\left(\theta^2+3\theta+1\right)}{(\theta+1)^{x+2}}\ ;x=0,1,2,...,\ \theta>0 \tag{1.1}$$

The first four moments about origin and the variance of PGD obtained by Shanker[1] are as follows:

$$\mu_1'=\frac{\theta+3}{\theta(\theta+2)},\ \mu_2'=\frac{\theta^2+5\theta+8}{\theta^2(\theta+2)},\ \mu_3'=\frac{\theta^3+9\theta^2+30\theta+30}{\theta^3(\theta+2)}$$

$$\mu_4'=\frac{\theta^4+17\theta^3+92\theta^2+204\theta+144}{\theta^4(\theta+2)},\ \text{and}\ \mu_2=\frac{\theta^3+6\theta^2+12\theta+7}{\theta^2(\theta+2)^2}$$

The detailed discussion about its properties, estimation of parameter, and applications has been discussed by Shanker[1] and it has been shown that it is better than Poisson and Poisson-Lindley distributions for modeling count data in various fields of knowledge. The PGD arises from the Poisson distribution when its parameter λ follows Garima distribution introduced by Shanker[2] having probability density function (p.d.f.)

$$f_0(\lambda;\theta)=\frac{\theta}{\theta+2}(1+\theta+\theta\lambda)e^{-\theta\lambda}\ ;\lambda>0,\ \theta>0 \tag{1.2}$$

Size-biased distributions arise in practice when observations from a sample having probability proportional to some measure of unit size. Fisher[3] firstly introduced these distributions to model ascertainment biases which were later formalized by Rao[4] in a unifying theory. Size-biased observations occur in many research areas and its fields of applications includes medical science, sociology, psychology, ecology, geological sciences etc. The applications of size-biased distribution theory in fitting distributions of diameter at breast height (DBH) data arising from horizontal point sampling (HPS) has been discussed by Van Deusen.[5] Further, Lappi and Bailey[6] have applied size-biased distributions to analyze HPS diameter increment data. The statistical applications of size-biased distributions to the analysis of data relating to human population and ecology can be found in Patil and Rao.[7,8] Some of the recent results on size-biased distributions pertaining to parameter estimation in forestry with special emphasis on Weibull family have been discussed by Gove.[9] Ducey and Gove[10] discussed size-biased distributions in the generalized beta distribution family, with applications to forestry.

Let a random variable X has the original probability distribution $P_0(x;\theta)$, then a simple size-biased distribution is given by its probability function

$$P_1(x;\theta)=\frac{x\cdot P_0(x;\theta)}{\mu_1'} \tag{1.3}$$

Where $\mu_1'=E(X)$ is the mean of the original probability distribution.

In the present paper, a size-biased Poisson-Garima distribution (SBPGD) has been proposed. It s raw and central moments and central moments based properties including coefficient of variation, skewness, kurtosis and index of dispersion have been obtained and discussed. Some of its statistical properties have been discussed. The method of moment and the method of maximum likelihood estimation have been discussed for estimating the parameter of SBPGD. The goodness of fit of SBPGD has also been presented.

Size-biased poisson-garima distribution

Using (1.1) and (1.3), the p.m.f. of the size-biased Poisson-Garima distribution (SBPGD) with parameter θ can be obtained as

$$P_1(x;\theta)=\frac{x\cdot P_0(x;\theta)}{\mu_1'}=\frac{\theta^2}{\theta+3}\frac{x^2\theta+x\left(\theta^2+3\theta+1\right)}{(\theta+1)^{x+2}};x=1,2,3,..,\theta>0 \tag{2.1}$$

where $\mu_1' = \dfrac{\theta+3}{\theta(\theta+2)}$ is the mean of the PGD (1.1).

The SBPGD can also be obtained from the size-biased Poisson distribution (SPBD) with p.m.f.

$$g(x|\lambda) = \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!}\ ;x=1,2,3,...,\lambda>0 \tag{2.2}$$

when its parameter λ follows size-biased Garima distribution (SBGD) with p.d.f.

$$h(\lambda;\theta) = \frac{\theta^2}{\theta+3}\lambda(1+\theta+\theta\lambda)e^{-\theta\lambda};\ x>0,\ \theta>0 \tag{2.3}$$

Thus the p.m.f of SBPGD can be obtained as

$$P(X=x) = \int_0^\infty g(x|\lambda)\cdot h(\lambda;\theta)\,d\lambda$$

$$= \int_0^\infty \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!}\frac{\theta^2}{\theta+3}\lambda(1+\theta+\theta\lambda)e^{-\theta\lambda}\,d\lambda \tag{2.4}$$

$$= \frac{\theta^2}{(\theta+3)(x-1)!}\int_0^\infty e^{-(\theta+1)\lambda}\left[(1+\theta)\lambda^x+\theta\lambda^{x+1}\right]d\lambda$$

$$= \frac{\theta^2}{\theta+3}\frac{x^2\theta+x\left(\theta^2+3\theta+1\right)}{(\theta+1)^{x+2}}\ ;x=1,2,3,..,\theta>0$$

which is the p.m.f of SBPGD with parameter θ.

Graphs of SBPGD for varying values of parameter θ are shown in figure 1. It is obvious from the graphs of SBPGD that as the value of parameter θ increases, the initially the graphs shift upward and decreases fast for increasing values of x. Also the graphs become convex for values of $\theta \ge 2$.

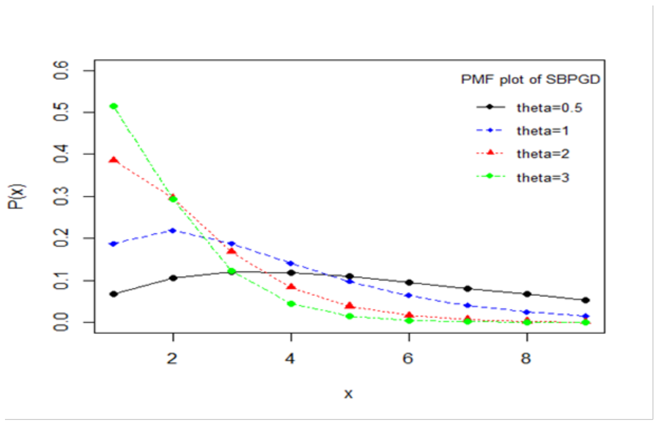

Figure 1 Graphs of SBPGD for varying values of parameter θ.

It would be recalled that the p.m.f of size-biased Poisson-Lindley distribution (SBPLD) given by

$$P_2(X=x) = \frac{\theta^3}{\theta+2}\frac{x(x+\theta+2)}{(\theta+1)^{x+2}}\ ;x=1,2,3,...,;\theta>0 \tag{1.7}$$

has been introduced by Ghitany and Mutairi,[11] which is a size-biased version of Poisson-Lindley distribution (PLD) introduced by Sankaran.[12] Ghitany and Mutairi[11] have discussed its various mathematical and statistical properties, estimation of the parameter using maximum likelihood estimation and the method of moments, and goodness of fit. Shanker et al.,[13] has detailed study on the applications of size-biased Poisson-Lindley distribution (SBPLD) for modeling data on thunderstorms and observed that in most data sets, SBPLD gives better fit than size-biased Poisson distribution (SBPD).

Moments and moments based measures

Using (2.4), the r^{th} factorial moment about origin of the SBPGD (2.1) can be obtained as

$$\mu_{(r)}' = E\left[E\left(X^{(r)}|\lambda\right)\right] = \int_0^\infty\left[\sum_{x=1}^\infty x^{(r)}\frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!}\right]\frac{\theta^2}{\theta+3}\lambda(1+\theta+\theta\lambda)e^{-\theta\lambda}\,d\lambda$$

$$= \frac{\theta^2}{\theta+3}\int_0^\infty\left[\lambda^{r-1}\sum_{x=r}^\infty x\frac{e^{-\lambda}\lambda^{x-r}}{(x-r)!}\right]\lambda(1+\theta+\theta\lambda)e^{-\theta\lambda}\,d\lambda$$

Taking $y=x-r$, we get

$$\mu_{(r)}' = \frac{\theta^2}{\theta+3}\int_0^\infty\left[\lambda^{r-1}\sum_{y=0}^\infty(y+r)\frac{e^{-\lambda}\lambda^y}{y!}\right]\lambda(1+\theta+\theta\lambda)e^{-\theta\lambda}\,d\lambda$$

$$= \frac{\theta^2}{\theta+3}\int_0^\infty\lambda^{r-1}(\lambda+r)\lambda(1+\theta+\theta\lambda)e^{-\theta\lambda}\,d\lambda$$

$$= \frac{r!\{(\theta+1)(r\theta+r+1)+(r+1)(r\theta+r+2)\}}{\theta^r(\theta+3)};r=1,2,3,... \tag{3.1}$$

Substituting $r=1,2,3,$ and 4, the first four factorial moments about origin can be obtained and using the relationship between factorial moments about origin and moments about origin, the first four moments about origin of SBPGD can be obtained as

$$\mu_1' = \frac{\theta^2+5\theta+8}{\theta(\theta+3)}$$

$$\mu_2' = \frac{\theta^3+9\theta^2+30\theta+30}{\theta^2(\theta+3)}$$

$$\mu_3' = \frac{\theta^4+17\theta^3+92\theta^2+204\theta+144}{\theta^3(\theta+3)}$$

$$\mu_4' = \frac{\theta^5+33\theta^4+270\theta^3+990\theta^2+1560\theta+840}{\theta^4(\theta+3)}$$

Using the relationship between moments about mean and the moments about origin, the moments about mean of the SBPGD are thus obtained as

$$\mu_2 = \frac{2\left(\theta^3+8\theta^2+20\theta+13\right)}{\theta^2(\theta+3)^2}$$

Citation: Shanker R, Shukla KK. Size-biased poisson-garima distribution with applications. *Biom Biostat Int J.* 2017;6(3):335–340.
DOI: 10.15406/bbij.2017.06.00167

$$\mu_3=\frac{2\left(\theta^5+13\theta^4+68\theta^3+171\theta^2+195\theta+80\right)}{\theta^3\left(\theta+3\right)^3}$$

$$\mu_4=\frac{2\left(\theta^7+26\theta^6+269\theta^5+1435\theta^4+4230\theta^3+6819\theta^2+5520\theta+1740\right)}{\theta^4\left(\theta+3\right)^4}$$

The coefficient of variation $\left(C.V\right)$, coefficient of Skewness $\left(\sqrt{\beta_1}\right)$, coefficient of Kurtosis $\left(\beta_2\right)$ and the index of dispersion $\left(\gamma\right)$ of the SBPGD are thus obtained as

$$C.V=\frac{\sigma}{\mu_1'}=\frac{\sqrt{2\left(\theta^3+8\theta^2+20\theta+13\right)}}{\theta^2+5\theta+8}$$

$$\sqrt{\beta_1}=\frac{\mu_3}{\mu_2^{3/2}}=\frac{\theta^5+13\theta^4+68\theta^3+171\theta^2+195\theta+80}{\sqrt{2}\left(\theta^3+8\theta^2+20\theta+13\right)^{3/2}}$$

$$\beta_2=\frac{\mu_4}{\mu_2^{\,2}}=\frac{\left(\theta^7+26\theta^6+269\theta^5+1435\theta^4+4230\theta^3+6819\theta^2+5520\theta+1740\right)}{2\left(\theta^3+8\theta^2+20\theta+13\right)^2}$$

$$\gamma=\frac{\sigma^2}{\mu_1'}=\frac{2\left(\theta^3+8\theta^2+20\theta+13\right)}{\theta\left(\theta+3\right)\left(\theta^2+5\theta+8\right)}$$

Graphs of coefficient of variation, coefficient of skewness, coefficient of kurtosis and index of dispersion of SBPGD for varying values of parameter θ are shown in figure 2. It is obvious from the graphs that C.V and the index of dispersion are monotonically decreasing while the coefficient of skewness and coefficient of kurtosis are decreasing for increasing value of the parameter θ.

The condition under which SBPGD and SBPLD are over-dispersed, equi-dispersed or under-dispersed are presented in table 1.

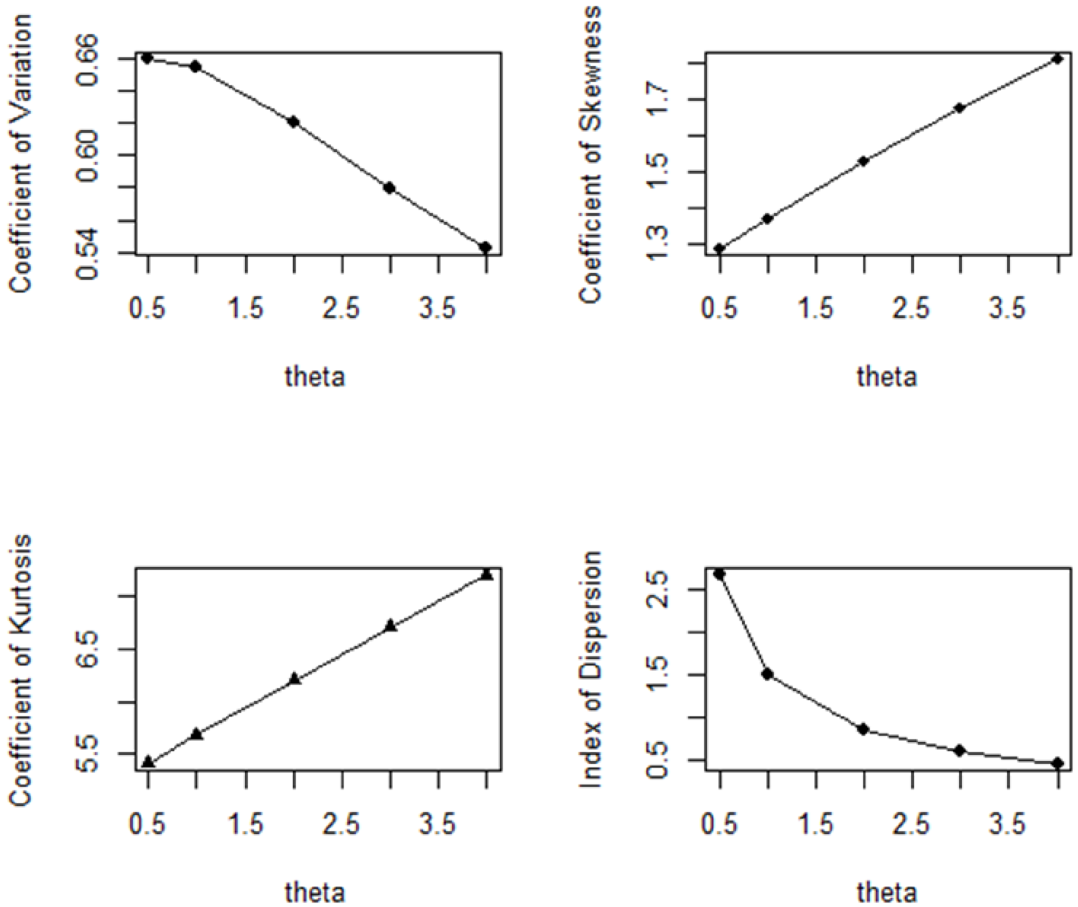

Figure 2 Graphs of coefficient of variation, coefficient of skewness, coefficient of kurtosis and index of dispersion of SBPGD for varying values of parameter θ.

Citation: Shanker R, Shukla KK. Size-biased poisson-garima distribution with applications. *Biom Biostat Int J.* 2017;6(3):335–340. DOI: 10.15406/bbij.2017.06.00167

Table 1 Over-dispersion, equi-dispersion and under-dispersion of SBPGD and SBPLD

Distributions	Over-dispersion $\left(\mu<\sigma^2\right)$	Equi-dispersion $\left(\mu=\sigma^2\right)$	Under-dispersion $\left(\mu>\sigma^2\right)$
SBPGD	$\theta<1.671162$	$\theta=1.671162$	$\theta>1.671162$
SBPLD	$\theta<1.636061$	$\theta<1.636061$	$\theta<1.636061$

Statistical properties of SBPGD

Unimodality and increasing failure rate

Since

$$\frac{P_1(x+1;\theta)}{P_1(x;\theta)}=\left(\frac{1}{\theta+1}\right)\left[1+\frac{2x\theta+\left(\theta^2+4\theta+1\right)}{x^2\theta+x\left(\theta^2+3\theta+1\right)}\right]$$

is a deceasing function of x, $P_1(x;\theta)$ is log-concave. Therefore, SBPGD is unimodal, has an increasing failure rate (IFR), and hence increasing failure rate average (IFRA). It is new better than used in expectation (NBUE) and has decreasing mean residual life (DMRL). Detailed discussion about definitions and interrelationships between these aging concepts are available in Barlow and Proschan.[14]

Generating functions

Probability Generating Function: The probability generating function of the SBPGD (2.1) can be obtained as

$$P_X(t)=E\left(t^X\right)=\frac{\theta^2}{(\theta+3)(\theta+1)^2}\left[\theta\sum_{x=1}^{\infty}x^2\left(\frac{t}{\theta+1}\right)^x+\left(\theta^2+3\theta+1\right)\sum_{x=1}^{\infty}x\left(\frac{t}{\theta+1}\right)^x\right]$$

$$=\frac{\theta^2}{(\theta+3)(\theta+1)^2}\left[\frac{\theta t(\theta+1+t)(\theta+1)}{(\theta+1-t)^3}+\frac{t\left(\theta^2+3\theta+1\right)(\theta+1)}{(\theta+1-t)^2}\right]$$

$$=\frac{\theta^2 t}{(\theta+3)(\theta+1)}\left[\frac{\theta(\theta+1+t)}{(\theta+1-t)^3}+\frac{\theta^2+3\theta+1}{(\theta+1-t)^2}\right]$$

Moment generating function: The moment generating function of the SBPGD (2.1) can be given by

$$M_X(t)=E\left(e^{tX}\right)=\frac{\theta^2 e^t}{(\theta+3)(\theta+1)}\left[\frac{\theta\left(\theta+1+e^t\right)}{\left(\theta+1-e^t\right)^3}+\frac{\theta^2+3\theta+1}{\left(\theta+1-e^t\right)^2}\right]$$

Estimation of parameter

Method of moment estimate (MOME): Let $x_1,x_2,...,x_n$ be a random sample of size n from the SBPGD (2.1). Equating the population to the corresponding sample mean, the MOME $\tilde{\theta}$ of θ of SBPGD (2.1) can be obtained as

$$\tilde{\theta}=\frac{-(3\bar{x}-5)+\sqrt{9\bar{x}^2+2\bar{x}-7}}{2(\bar{x}-1)}$$

where $\bar{x}$ is the sample mean.

Maximum likelihood estimate (MLE): Let $x_1,x_2,...,x_n$ be a random sample of size n from the SBPGD (2.1) and let f_x be the observed frequency in the sample corresponding to $X=x$ $(x=1,2,3,...,k)$ such that $\sum_{x=1}^{k}f_x=n$, where k is the largest observed value having non-zero frequency. The likelihood function L of the SBPGD (2.1) is given by

$$L=\left(\frac{\theta^2}{\theta+3}\right)^n\frac{1}{(\theta+1)^{\sum_{x=1}^{k}f_x(x+2)}}\prod_{x=1}^{k}\left[x^2\theta+x\left(\theta^2+3\theta+1\right)\right]^{f_x}$$

The log likelihood function is obtained as

$$\log L=n\log\left(\frac{\theta^2}{\theta+3}\right)-\sum_{x=1}^{k}f_x(x+2)\log(\theta+1)+\sum_{x=1}^{k}f_x\log\left[x^2\theta+x\left(\theta^2+3\theta+1\right)\right]$$

The first derivative of the log likelihood function is given by

$$\frac{d\log L}{d\theta}=\frac{n(\theta+6)}{\theta(\theta+3)}-\frac{n(\bar{x}+2)}{\theta+1}+\sum_{x=1}^{k}\frac{(x+2\theta+3)f_x}{x\theta+\left(\theta^2+3\theta+1\right)}$$

where $\bar{x}$ is the sample mean.

The maximum likelihood estimate (MLE), $\hat{\theta}$ of θ is the solution of the equation $\frac{d\log L}{d\theta}=0$ and is given by the solution of the non-linear equation

$$\sum_{x=1}^{k}\frac{(x+2\theta+3)f_x}{x\theta+\left(\theta^2+3\theta+1\right)}-\frac{n(\bar{x}+2)}{\theta+1}+\frac{n(\theta+6)}{\theta(\theta+3)}=0$$

This non-linear equation can be solved by any numerical iteration methods such as Newton- Raphson method, Bisection method, Regula –Falsi method etc. note that in this paper, we have solved above equation using Newton-Raphson method where the initial value of θ is the value given by the method of moment estimate.

Data analysis

In this section, we fit SBPGD using maximum likelihood estimate to test its goodness of fit over SBPD and SBPLD. The first data-set is the immunogold assay data of Cullen et al.,[15] regarding the distribution of number of counts of sites with particles from immunogold assay data, the second data-set is the number of European red mites on apple leaves, reported by Garman[16] (Tables 2&3).

It is obvious from above tables that SBPGD gives better fit than both SBPD and SBPLD

Table 2 Distribution of number of counts of sites with particles from immunogold data

No. of sites with particles	Observed frequency	Expected frequency		
		SBPD	SBPLD	SBPGD
1	122	111.3	119.0	119.1
2	50	64.1	53.8	53.7
3	18	18.5	18.0	18.0
4	4	3.5	5.3	5.3
5	4	0.6	1.9	1.9
Total	198	198.0	198.0	198.0
ML estimate		$\hat{\theta}$=0.576	$\hat{\theta}$=4.051	$\hat{\theta}$=2.0992
χ^2		4.642	0.51	0.40
d.f.		1	2	2
p-value		0.031	0.7749	0.8187

Table 3 Number of European red mites on apple leaves, reported by Garman (1923)

Number of European red mites	Observed frequency	Expected frequency		
		SBPD	SBPLD	SBPGD
1	38	28.7	31.7	31.9
2	17	25.7	23.9	23.8
3	10	15.3	13.2	13.1
4	9			
5	3			
6	2			
7	1			
8	0			
Total	80	80.0	80.0	80.0
ML Estimate		$\hat{\theta}$=1.791615	$\hat{\theta}$=2.163462	$\hat{\theta}$=2.08381
χ^2		9.827	5.30	5.11
d.f.		2	2	2
P-value		0.0073	0.0706	0.0777

Conclusions

A size-biased Poisson-Garima distribution (SBPGD) has been obtained by size-biasing the Poisson-Garima distribution (PGD) introduced by Shanker.[1] Its raw moments and central moments have been obtained and moments and hence expressions for coefficient of variation (C.V.), skewness, kurtosis and index of dispersion have been obtained. The method of moment and the method of maximum likelihood estimation have been discussed for estimating its parameter. The goodness of fit of SBPGD has been presented for two real data sets and the fit shows quite satisfactory over size-biased Poisson distribution (SBPD) and size-biased Poisson-Lindley distribution (SBPLD).[17] Therefore, SBPGD can be considered an important distribution for modeling data which structurally excludes zero-counts.

Acknowledgements

None.

Conflicts of interest

None.

References

1. Shanker R. The Discrete Poisson–Garima Distribution. *Biom Biostat Int J*. 2017;5(2):00127.
2. Shanker R. Garima Distribution and its Application to Model Behavioral Science Data. *Biom Biostat Int J*. 2016;4(7):00116.
3. Fisher RA. The effects of methods of ascertainment upon the estimation of frequencies. *Ann Eugenics*. 1934;6(1):13–25.
4. Rao CR. On discrete distributions arising out of methods of ascertainment In: Patil GP (Ed.), Classical and Contagious Discrete Distributions. *Statistical Publishing Society*, Calcutta. 1965;pp. 320–332.
5. Van Deusen PC. Fitting assumed distributions to horizontal point sample diameters. *For Sci*. 1986;32(1):146–148.
6. Lappi J, Bailey RL. Estimation of diameter increment function or other tree relations using angle–count samples. *Forest science*. 1987;33:725–739.
7. Patil GP, Rao CR. The Weighted distributions: A survey and their applications. In applications of Statistics (Ed P.R. Krishnaiah, North Holland Publications Co., Amsterdam. 1977;pp. 383–405.
8. Patil GP, Rao CR. Weighted distributions and size–biased sampling with applications to wild–life populations and human families. *Biometrics*. 1978;34(2):179–189.
9. Gove JH. Estimation and applications of size–biased distributions in forestry. In Modeling Forest Systems. In: Amaro A, et al. (Eds), CABI Publishing, USA. 2003;pp. 201–212.
10. Ducey MJ, Gove JH. Size–biased distributions in the generalized beta distribution family, with applications to forestry. *Forestry– An International Journal of Forest Research*. 2015;88:143–151.
11. Ghitany ME, Al–Mutairi DK. Size–biased Poisson–Lindley distribution and Its Applications. *Metron – International Journal of Statistics LXVI*. 2008;(3):299–311.
12. Sankaran M. The discrete Poisson–Lindley distribution. *Biometrics*. 1970;26(1):145–149.
13. Shanker R, Hagos F, Abrehe Y. On Size –Biased Poisson–Lindley Distribution and Its Applications to Model Thunderstorms. *American Journal of Mathematics and Statistics*. 2015;5(6):354–360.
14. Barlow RE, Proschan F. Statistical Theory of Reliability and Life Testing, Silver Spring, MD. 1981.
15. Cullen MJ, Walsh J, Nicholson LV, et al. Ultrastructural localization of dystrophin in human muscle by using gold immunolabelling. *Proc R Soc Lond B Biol Sci.* 240(1297):197–210.
16. Garman P. The European red mites in Connecticut apple orchards. *Connecticut Agri Exper Station Bull*. 1923;252:103–125.
17. Lindley DV. Fiducial distributions and Bayes theorem. *Journal of the Royal Statistical Society*. 1958;20(1):102–107.