Publications

Available Online

JOURNAL OF SCIENTIFIC RESEARCH

J. Sci. Res. **10**(2), 145-157 (2018) www.banglajol.info/index.php/JSR

A Generalized Size-Biased Poisson-Lindley Distribution and Its Applications to Model Size Distribution of Freely-Forming Small Group

R. Shanker, K. K. Shukla*

Department of Statistics, College of Science, Eritrea Institute of Technology, Asmara, Eritrea

Received 16 December 2017, accepted in final revised form 13 March 2018

Abstract

In this paper, generalized size-biased Poisson-Lindley distribution (GSBPLD) which includes size-biased Poisson-Lindley distribution (SBPLD) as particular case, has been proposed and studied. Its moments based measures including coefficients of variation, skewness, kurtosis, and index of dispersion have been derived and their nature and behavior have been discussed with varying values of the parameters. The estimation of its parameter has been discussed using maximum likelihood estimation. Some applications of the proposed distribution have been explained through datasets relating to size distribution of freely-forming and the goodness of fit has been found satisfactory over SBPLD and size-biased Poisson distribution (SBPD).

Keywords: Size-biased distribution; Poisson-Lindley distribution, Kurtosis; Maximum likelihood estimation.

doi: http://dx.doi.org/10.3329/jsr.v10i2.34905 J. Sci. Res. **10** (2), 145-157 (2018)

1. Introduction

A size-biased Poisson Lindley distribution (SBPLD), introduced by Ghitany and Al-Mutairi [1], having parameter θ is defined by its probability mass function (pmf)

$$P_1(x;\theta)=\frac{\theta^3}{\theta+2}\frac{x(x+\theta+2)}{(\theta+1)^{x+2}}\ ;x=1,2,3,...,\ \theta>0 \tag{1.1}$$

The first four moments about origin and the variance of the SBPLD are given by

$$\mu_1'=\frac{\theta^2+4\theta+6}{\theta(\theta+2)}$$

$$\mu_2'=\frac{\theta^3+8\theta^2+24\theta+24}{\theta^2(\theta+2)}$$

* *Corresponding author:* kkshukla22@gmail.com

$$\mu_3' = \frac{\theta^4 + 16\theta^3 + 78\theta^2 + 168\theta + 120}{\theta^3(\theta+2)}$$

$$\mu_4' = \frac{\theta^5 + 32\theta^4 + 240\theta^3 + 840\theta^2 + 1320\theta + 720}{\theta^4(\theta+2)}$$

$$\mu_2 = \mu_2' - \left(\mu_1'\right)^2 = \frac{2\left(\theta^3 + 6\theta^2 + 12\theta + 6\right)}{\theta^2(\theta+2)^2}$$

A detailed study on SBPLD and its applications to model data relating to thunderstorms are done by Shanker *et al.* [2] and found that SBPLD is a suitable model for thunderstorms data. It would be noted that SBPLD is a simple size-biased version of Poisson-Lindley distribution (PLD) [3] having pmf

$$P_2(x;\theta) = \frac{\theta^2(x+\theta+2)}{(\theta+1)^{x+3}} \;;x = 0,1,2,3,...,\ \theta > 0 \tag{1.2}$$

Sankaran [3] has obtained its moments, discussed some of statistical properties, estimation of parameter and applications to model count data. A detailed and critical study on applications of PLD for count data relating to biological sciences has been done by Shanker and Hagos [4] and found that PLD gives much closer fit than Poisson distribution.

Note that the PLD is a Poisson mixture of Lindley distribution [5] and defined by its probability density function (pdf)

$$f_1(x;\theta) = \frac{\theta^2}{\theta+1}(1+x)e^{-\theta x} \;;x > 0,\ \theta > 0 \tag{1.3}$$

Ghitany *et al.* [6] discussed statistical properties including moments based coefficients, hazard rate function, mean residual life function, mean deviations, stochastic ordering, Renyi entropy measure, order statistics, Bonferroni and Lorenz curves, stress- strength reliability, along with estimation of parameter and application to model waiting time data in a bank. Shanker *et al.* [7] have critical and comparative study on applications of Lindley and exponential distributions for modeling lifetime data from biological sciences and engineering and observed that there are many lifetime data where exponential distribution gives better fit than Lindley distribution.

The generalized Poisson-Lindley distribution (GPLD) [8] having parameters θ and α is defined by its pmf

$$P_3(x;\theta) = \frac{\Gamma(x+\alpha)}{\Gamma(x+1)\Gamma(\alpha+1)}\frac{\theta^{\alpha+1}}{(\theta+1)^{x+\alpha+1}}\left(\frac{x+\alpha\theta+2\alpha}{\theta+1}\right);x = 0,1,2,3,...,\ \theta > 0, \alpha > 0 \tag{1.4}$$

The first four moments about origin and the variance of GPLD obtained by Mahmoudi and Zakerzadeh [8] are given by

$$\mu_1' = \frac{\alpha(\theta+1)+1}{\theta(\theta+1)}$$

$$\mu_2' = \frac{\alpha\theta^2 + \left(\alpha^2 + 2\alpha + 1\right)\theta + \left(\alpha^2 + 3\alpha + 2\right)}{\theta^2\left(\theta+1\right)}$$

$$\mu_3' = \frac{\alpha\theta^3 + \left(3\alpha^2 + 4\alpha + 1\right)\theta^2 + \left(\alpha^3 + 6\alpha^2 + 11\alpha + 6\right)\theta + \left(\alpha^3 + 6\alpha^2 + 11\alpha + 6\right)}{\theta^3\left(\theta+1\right)}$$

$$\mu_4' = \frac{\alpha\theta^4 + \left(7\alpha^2 + 8\alpha + 1\right)\theta^3 + \left(6\alpha^3 + 25\alpha^2 + 33\alpha + 14\right)\theta^2 + \left(\alpha^4 + 12\alpha^3 + 47\alpha^2 + 72\alpha + 36\right)\theta + \left(\alpha^4 + 10\alpha^3 + 35\alpha^2 + 50\alpha + 24\right)}{\theta^4\left(\theta+1\right)}$$

$$\mu_2 = \frac{\alpha\theta^3 + \left(3\alpha + 1\right)\theta^2 + 3\left(\alpha+1\right)\theta + \left(\alpha+1\right)}{\theta^2\left(\theta+1\right)^2}$$

Mahmoudi and Zakerzadeh [8] have obtained its moments and discussed its statistical properties, estimation of parameters and applications to model count data. Note that GPLD is a Poisson mixture of two-parameter generalized Lindley distribution (GLD), proposed by Zakerzadeh and Dolati [9], having parameters θ and α defined by its pdf

$$f_2\left(x;\theta,\alpha\right) = \frac{\theta^{\alpha+1}}{\theta+1}\frac{x^{\alpha-1}}{\Gamma\left(\alpha+1\right)}\left(\alpha+x\right)e^{-\theta x} \quad ; x>0,\ \theta>0, \alpha>0 \tag{1.5}$$

In fact the distribution in (1.5) is a particular case (β=1) of a three parameter generalized Lindley distribution (GLD) [9] having pdf

$$f_3\left(x;\theta,\alpha,\beta\right) = \frac{\theta^{\alpha+1}}{\theta+\beta}\frac{x^{\alpha-1}}{\Gamma\left(\alpha+1\right)}\left(\alpha+\beta x\right)e^{-\theta x} \quad ; x>0, \theta>0, \alpha>0, \beta>0 \tag{1.6}$$

Lindley distribution, gamma distribution and weighted Lindley distribution (WLD) proposed by Ghitany *et al.* [6] are particular cases of (1.6) at (α = β =1), (β = 0) and ((β = α), respectively. Shanker [10] obtained various raw moments and central moments of GLD and discussed properties based on moments including coefficient of variation, skewness, kurtosis and index of dispersion of GLD and its comparative study with generalized gamma distribution (GGD) introduced by Stacy [11] to model various lifetime data from engineering and biomedical sciences and concluded that in many cases GGD gives much better fit than GLD. A detailed comparative study on modelling of real lifetime data from engineering and biomedical sciences using GLD and GGD have been done by Shanker and Shukla [12] and found that there are several lifetime data where GGD gives much closer fit than GLD.

Suppose a random variable X has probability distribution P_0 ($x;\theta$); x=0, 1, 2, … θ >0.. If sample units are weighted or selected with probability proportional to x^{α}, then the corresponding size-biased distribution of order α is defined by its probability mass function

$$P_4(x;\theta) = \frac{x^\alpha \cdot P_0(x;\theta)}{\mu'_\alpha} \tag{1.7}$$

where $\mu'_\alpha = E(X^\alpha) = \sum_{x=0}^{\infty} x^\alpha P_0(x;\theta)$. When $\alpha = 1$, (1.7) is known as simple size-biased distribution and is applicable for size-biased sampling and for $\alpha = 2$, (1.7) is known as area-biased distribution and is applicable for area-biased sampling. Size-biased distributions are a particular class of weighted distributions which arise naturally in practice when observations from a sample are recorded with probability proportional to some measure of unit size. In field applications, size-biased distributions can arise either because individuals are sampled with unequal probability by design or because of unequal detection probability. Size-biased distributions come into play when organisms occur in groups, and group size influences the probability of detection. Fisher [13] firstly introduced these distributions to model ascertainment biases which were later formalized by Rao [14] in a unifying theory for problems where the observations fall in non-experimental, non-replicated and non-random categories. Size-biased distributions have applications in environmental science, econometrics, social science, biomedical science, human demography, ecology, geology, forestry etc.

The main motivation of this paper is to introduce a generalized size-biased Poisson-Lindley distribution (GSBPLD), a size-biased version of generalized Poisson-Lindley distribution (GPLD), to model count data excluding zero counts because two-parameter GSBPLD have enough flexibility than one parameter SBPLD. Various moments and moments based measures have been obtained. The nature and behavior of coefficients of variation, skewness, kurtosis, index of dispersion have been explained graphically. Maximum likelihood estimation has been discussed for estimating the parameters of the distribution. Applications and the goodness of fit of the proposed distribution have been explained through datasets relating to size distribution of freely-forming small group and compared with other size-biased distributions.

2. Generalized Size-Biased Poisson-Lindley Distribution

Using pmf (1.4) and its mean in (1.7), the pmf of generalized size-biased Poisson-Lindley distribution (GSBPLD) can be obtained as

$$P_5(x;\theta,\alpha) = \frac{\Gamma(x+\alpha)}{\Gamma(x)\Gamma(\alpha+1)} \frac{\theta^{\alpha+2}}{(\alpha\theta+\alpha+1)} \frac{x+\alpha\theta+2\alpha}{(\theta+1)^{x+\alpha+1}}; \theta > 0, \alpha > 0, x = 1,2,3,\ldots \cdots \tag{2.1}$$

The graphs of the pmf of GSBPLD for varying values of parameters θ and α are shown in Fig. 1.

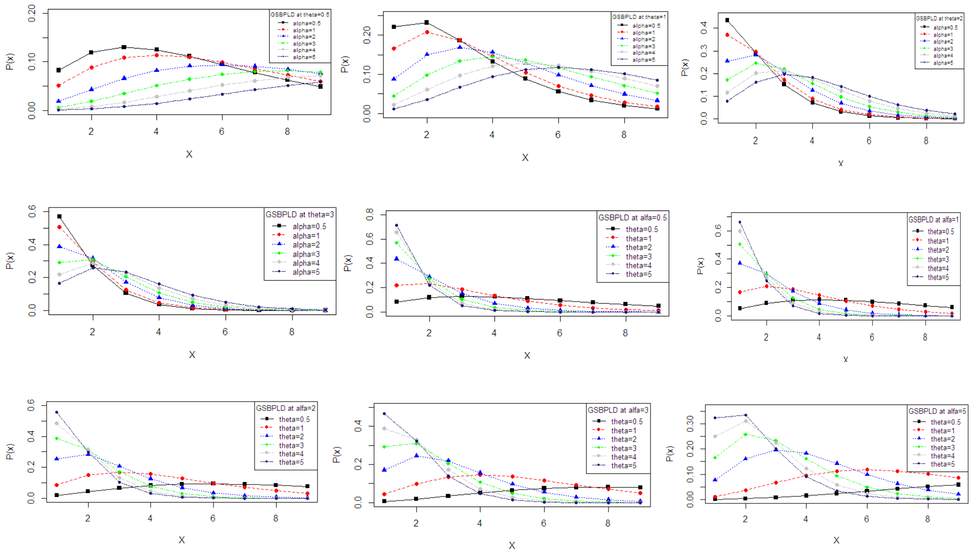

Fig. 1. Graphs of pmf of GSBPLD for varying values of parameters θ and α.

Since it is difficult and complicated to obtain the moments of GSBPLD directly, an attempt has been made to derive the pmf of GSBPLD as a size-biased Poisson mixture of size-biased generalized Lindley distribution (SBGLD) which is very much helpful in deriving the moments. Suppose the parameter λ of size-biased Poisson distribution (SBPD) with pmf

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

follows size-biased generalized Lindley distribution (SBGLD) with pdf

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

Thus the SBPD mixture of GSBPLD can be obtained as

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

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

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

$$=\frac{\theta^{\alpha+2}}{(\alpha\theta+\alpha+1)\Gamma(\alpha+1)\Gamma(x)}\left[\frac{\alpha\Gamma(x+\alpha)}{(\theta+1)^{x+\alpha}}+\frac{\Gamma(x+\alpha+1)}{(\theta+1)^{x+\alpha+1}}\right]$$

$$= \frac{\Gamma(x+\alpha)}{\Gamma(x)\Gamma(\alpha+1)} \frac{\theta^{\alpha+2}}{(\alpha\theta+\alpha+1)} \frac{x+\alpha\theta+2\alpha}{(\theta+1)^{x+\alpha+1}} ; x = 1, 2, 3, ...$$

which is the pmf of GSBPLD obtained in (2.1).

3. Moments, Skewness, Kurtosis and Index of Dispersion

Using (2.4), the rth factorial moment about origin of the GSBPLD (2.1) can be obtained as

$$\mu_{(r)}' = E\left[E\left(X^{(r)} \mid \lambda\right)\right], \text{ where } X^{(r)} = X(X-1)(X-2)...(X-r+1)$$

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

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

Taking $x - r = y$, we get

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

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

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

$$= \frac{\theta^{\alpha+2}}{(\alpha\theta+\alpha+1)\Gamma(\alpha+1)} \int_0^\infty \lambda^{r+\alpha-1} \left[\lambda^2 + (\alpha+r)\lambda + r\alpha\right] e^{-\theta\lambda} d\lambda$$

Using gamma integral and a little algebraic simplification, the rth factorial moment about origin of GSBPLD (2.1) can be obtained as

$$\mu_{(r)}' = \frac{\Gamma(\alpha+r)}{\Gamma(\alpha+1)} \frac{\left\{r\alpha\theta^2 + (\alpha+r)^2\theta + (\alpha+r)(\alpha+r+1)\right\}}{\theta^r(\alpha\theta+\alpha+1)} \quad ; r = 1, 2, 3, ... \tag{3.1}$$

Taking $r = 1, 2, 3, \text{and } 4$ in (3.1), the first four factorial moments about origin of GSBPLD (2.1) can be obtained

$$\mu_{(1)}' = \frac{\alpha\theta^2 + (\alpha+1)^2\theta + (\alpha+1)(\alpha+2)}{\theta(\alpha\theta+\alpha+1)}$$

$$\mu_{(2)}'=\frac{(\alpha+1)\left\{2\alpha\theta^2+(\alpha+2)^2\theta+(\alpha+2)(\alpha+3)\right\}}{\theta^2(\alpha\theta+\alpha+1)}$$

$$\mu_{(3)}'=\frac{(\alpha+1)(\alpha+2)\left\{3\alpha\theta^2+(\alpha+3)^2\theta+(\alpha+3)(\alpha+4)\right\}}{\theta^3(\alpha\theta+\alpha+1)}$$

$$\mu_{(4)}'=\frac{(\alpha+1)(\alpha+2)(\alpha+3)\left\{4\alpha\theta^2+(\alpha+4)^2\theta+(\alpha+4)(\alpha+5)\right\}}{\theta^4(\alpha\theta+\alpha+1)}.$$

Now using the relationship between factorial moments about origin and the moments about origin, the first four moments about origin of GSBPLD (2.1) can be obtained as

$$\mu_1'=\frac{\alpha\theta^2+(\alpha+1)^2\theta+(\alpha+1)(\alpha+2)}{\theta(\alpha\theta+\alpha+1)}$$

$$\mu_2'=\frac{\alpha\theta^3+(3\alpha^2+4\alpha+1)\theta^2+(\alpha^3+6\alpha^2+11\alpha+6)\theta+(\alpha^3+6\alpha^2+11\alpha+6)}{\theta^2(\alpha\theta+\alpha+1)}$$

$$\mu_3'=\frac{\left\{\begin{array}{l}\alpha\theta^4+(7\alpha^2+8\alpha+1)\theta^3+(6\alpha^3+25\alpha^2+33\alpha+14)\theta^2+(\alpha^4+12\alpha^3+47\alpha^2+72\alpha+36)\theta\\+(\alpha^4+10\alpha^3+35\alpha^2+50\alpha+24)\end{array}\right\}}{\theta^3(\alpha\theta+\alpha+1)}$$

$$\mu_4'=\frac{\left\{\begin{array}{l}\alpha\theta^5+(15\alpha^2+16\alpha+1)\theta^4+(25\alpha^3+90\alpha^2+95\alpha+30)\theta^3+(10\alpha^4+85\alpha^3+260\alpha^2+335\alpha+150)\theta^2\\+(\alpha^5+20\alpha^4+135\alpha^3+400\alpha^2+524\alpha+240)\theta+(\alpha^5+15\alpha^4+85\alpha^3+225\alpha^2+274\alpha+120)\end{array}\right\}}{\theta^4(\alpha\theta+\alpha+1)}$$

Now, using the relationship $\mu_r=E\left(Y-\mu_1'\right)^r=\sum_{k=0}^{r}\binom{r}{k}\mu_k'\left(-\mu_1'\right)^{r-k}$ between moments about mean and the moments about origin, the moments about mean of the GSBPLD (2.1) can be obtained as

$$\mu_2=\frac{(\alpha+1)\alpha^2\theta^3+3(\alpha^2+2\alpha+1)\alpha\theta^2+(3\alpha^3+10\alpha^2+9\alpha+2)\theta+(\alpha^3+4\alpha^2+5\alpha+2)}{\theta^2(\alpha\theta+\alpha+1)^2}$$

$$\mu_3=\frac{\left\{\begin{array}{l}(\alpha+1)\alpha^3\theta^5+2(3\alpha^2+5\alpha+2)\alpha^2\theta^4+(14\alpha^3+37\alpha^2+14\alpha+5)\alpha\theta^3\\+(16\alpha^4+59\alpha^3+66\alpha^2+25\alpha+2)\theta^2+(9\alpha^4+39\alpha^3+57\alpha^2+33\alpha+6)\theta\\+(2\alpha^4+10\alpha^3+18\alpha^2+14\alpha+4)\end{array}\right\}}{\theta^3(\alpha\theta+\alpha+1)^3}$$

$$\mu_4 = \frac{\left\{\begin{array}{l}(\alpha+1)\alpha^4\theta^7 + \left(3\alpha^3+17\alpha^2+19\alpha+5\right)\alpha^3\theta^6 + \left(18\alpha^4+100\alpha^3+157\alpha^2+84\alpha+9\right)\alpha^2\theta^5 \\ +\left(45\alpha^5+286\alpha^4+594\alpha^3+487\alpha^2+141\alpha+7\right)\alpha\,\theta^4 \\ +\left(60\alpha^6+443\alpha^5+1136\alpha^4+1256\alpha^3+602\alpha^2+101\alpha+2\right)\theta^3 \\ +\left(45\alpha^6+379\alpha^5+1140\alpha^4+1574\alpha^3+1045\alpha^2+303\alpha+26\right)\theta^2 \\ +\left(18\alpha^6+168\alpha^5+576\alpha^4+948\alpha^3+798\alpha^2+324\alpha+48\right)\theta \\ +\left(3\alpha^6+30\alpha^5+114\alpha^4+216\alpha^3+219\alpha^2+114\alpha+24\right)\end{array}\right\}}{\theta^4\left(\alpha\theta+\alpha+1\right)^4}$$

The coefficient of variation (*C. V*), coefficient of Skewness $\left(\sqrt{\beta_1}\right)$ and the coefficient of Kurtosis $\left(\beta_2\right)$ and index of dispersion (γ) of the GSBPLD (2.1) and are thus obtained as

$$C.V = \frac{\sigma}{\mu_1'} = \frac{\sqrt{(\alpha+1)\alpha^2\theta^3 + 3\left(\alpha^2+2\alpha+1\right)\alpha\,\theta^2 + \left(3\alpha^3+10\alpha^2+9\alpha+2\right)\theta + \left(\alpha^3+4\alpha^2+5\alpha+2\right)}}{\alpha\,\theta^2 + (\alpha+1)^2\theta + (\alpha+1)(\alpha+2)}$$

$$\sqrt{\beta_1} = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\left\{\begin{array}{l}(\alpha+1)\alpha^3\theta^5 + 2\left(3\alpha^2+5\alpha+2\right)\alpha^2\theta^4 + \left(14\alpha^3+37\alpha^2+14\alpha+5\right)\alpha\theta^3 \\ +\left(16\alpha^4+59\alpha^3+66\alpha^2+25\alpha+2\right)\theta^2 + \left(9\alpha^4+39\alpha^3+57\alpha^2+33\alpha+6\right)\theta \\ +\left(2\alpha^4+10\alpha^3+18\alpha^2+14\alpha+4\right)\end{array}\right\}}{\left\{(\alpha+1)\alpha^2\theta^3 + 3\left(\alpha^2+2\alpha+1\right)\alpha\,\theta^2 + \left(3\alpha^3+10\alpha^2+9\alpha+2\right)\theta + \left(\alpha^3+4\alpha^2+5\alpha+2\right)\right\}^{3/2}}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{\left\{\begin{array}{l}(\alpha+1)\alpha^4\theta^7 + \left(3\alpha^3+17\alpha^2+19\alpha+5\right)\alpha^3\theta^6 + \left(18\alpha^4+100\alpha^3+157\alpha^2+84\alpha+9\right)\alpha^2\theta^5 \\ +\left(45\alpha^5+286\alpha^4+594\alpha^3+487\alpha^2+141\alpha+7\right)\alpha\,\theta^4 \\ +\left(60\alpha^6+443\alpha^5+1136\alpha^4+1256\alpha^3+602\alpha^2+101\alpha+2\right)\theta^3 \\ +\left(45\alpha^6+379\alpha^5+1140\alpha^4+1574\alpha^3+1045\alpha^2+303\alpha+26\right)\theta^2 \\ +\left(18\alpha^6+168\alpha^5+576\alpha^4+948\alpha^3+798\alpha^2+324\alpha+48\right)\theta \\ +\left(3\alpha^6+30\alpha^5+114\alpha^4+216\alpha^3+219\alpha^2+114\alpha+24\right)\end{array}\right\}}{\left\{(\alpha+1)\alpha^2\theta^3 + 3\left(\alpha^2+2\alpha+1\right)\alpha\,\theta^2 + \left(3\alpha^3+10\alpha^2+9\alpha+2\right)\theta + \left(\alpha^3+4\alpha^2+5\alpha+2\right)\right\}^2}$$

$$\gamma = \frac{\sigma^2}{\mu_1'} = \frac{(\alpha+1)\alpha^2\theta^3 + 3\left(\alpha^2+2\alpha+1\right)\alpha\,\theta^2 + \left(3\alpha^3+10\alpha^2+9\alpha+2\right)\theta + \left(\alpha^3+4\alpha^2+5\alpha+2\right)}{\theta\left(\alpha\theta+\alpha+1\right)\left\{\alpha\,\theta^2 + (\alpha+1)^2\theta + (\alpha+1)(\alpha+2)\right\}}$$

Nature and behavior of coefficient of variation, coefficient of skewness, coefficient of kurtosis and index of dispersion of GSBPLD for varying values of parameters θ and α are shown in Fig. 2.

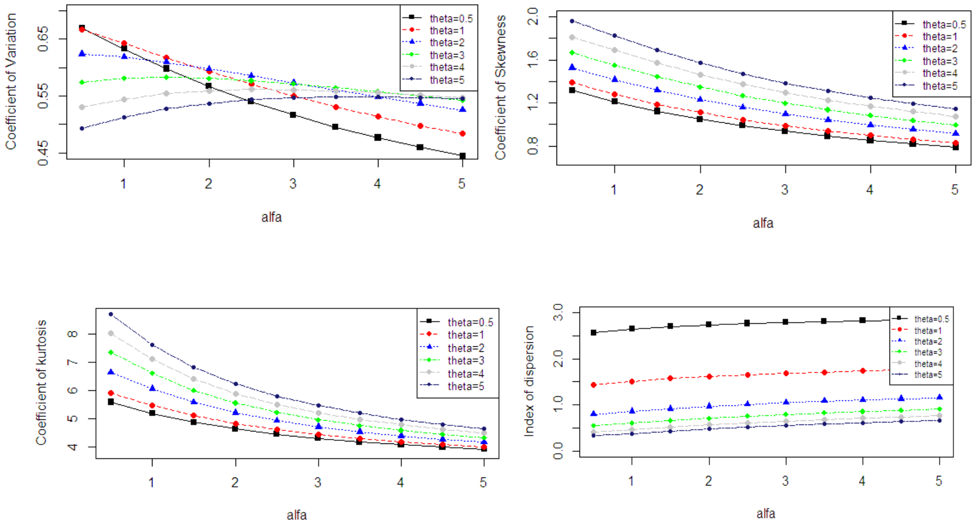

Fig. 2. Coefficient of variation, coefficient of skewness, coefficient of kurtosis and index of dispersion of GSBPLD for varying values of parameters θ and α.

4. Maximum Likelihood Estimation

Let $x_1, x_2, ..., x_n$ be a random sample of size n from the GSBPLD (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 GSBPLD (2.1) is given by

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

The log likelihood function can be obtained as

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

The maximum likelihood estimates $(\hat{\theta},\hat{\alpha})$ of (θ,α) of GSBPLD (2.2) is the solutions of the following log likelihood equations

$$\frac{\partial \log L}{\partial \theta}=\frac{n(\alpha+2)}{\theta}-\frac{n\alpha}{\alpha\theta+\alpha+1}-\sum_{x=1}^{k}\frac{(x+\alpha+1)f_x}{\theta+1}+\sum_{x=1}^{k}\frac{\alpha f_x}{x+\alpha\theta+2\alpha}=0$$

$$\frac{\partial \log L}{\partial \alpha}=n\log\theta-\frac{n(\theta+1)}{\alpha\theta+\alpha+1}-n\psi(\alpha+1)-\sum_{x=1}^{k}f_x\log(\theta+1)\ ,$$

$$+\sum_{x=1}^{k}f_x\psi(x+\alpha)+\sum_{x=1}^{k}\frac{(\theta+2)f_x}{x+\alpha\theta+2\alpha}=0$$

where $\psi(\alpha)=\frac{d}{d\alpha}\ln\Gamma(\alpha)$ is the digamma function.

These two log likelihood equations do not seem to be solved directly. However, the Fisher's scoring method can be applied to solve these equations. We have

$$\frac{\partial^2 \log L}{\partial \theta^2}=-\frac{n(\alpha+2)}{\theta^2}+\frac{n\alpha^2}{(\alpha\theta+\alpha+1)^2}+\sum_{x=1}^{k}\frac{(x+\alpha+1)f_x}{(\theta+1)^2}-\sum_{x=1}^{k}\frac{\alpha^2 f_x}{(x+\alpha\theta+2\alpha)^2}$$

$$\frac{\partial^2 \log L}{\partial \theta\,\partial \alpha}=\frac{n}{\theta}-\frac{n}{(\alpha\theta+\alpha+1)^2}-\sum_{x=1}^{k}\frac{(x+1)f_x}{\theta+1}+\sum_{x=1}^{k}\frac{x\,f_x}{(x+\alpha\theta+2\alpha)^2}=\frac{\partial^2 \log L}{\partial \alpha\,\partial \theta}$$

$$\frac{\partial^2 \log L}{\partial \alpha^2}=\frac{n(\theta+1)^2}{(\alpha\theta+\alpha+1)^2}-n\psi'(\alpha+1)+\sum_{x=1}^{k}f_x\psi'(x+\alpha)-\sum_{x=1}^{k}\frac{(\theta+2)^2 f_x}{(x+\alpha\theta+2\alpha)^2}\ ,$$

where $\psi'(\alpha)=\frac{d}{d\alpha}\psi(\alpha)$ is the trigamma function.

The maximum likelihood estimates $(\hat{\theta},\hat{\alpha})$ of (θ,α) of GSBPLD (2.2) is the solution of the following equations

$$\begin{bmatrix}\frac{\partial^2 \log L}{\partial \theta^2} & \frac{\partial^2 \log L}{\partial \theta\,\partial \alpha}\\ \frac{\partial^2 \log L}{\partial \theta\,\partial \alpha} & \frac{\partial^2 \log L}{\partial \alpha^2}\end{bmatrix}_{\substack{\hat{\theta}=\theta_0\\ \hat{\alpha}=\alpha_0}}\begin{bmatrix}\hat{\theta}-\theta_0\\ \hat{\alpha}-\alpha_0\end{bmatrix}=\begin{bmatrix}\frac{\partial \log L}{\partial \theta}\\ \frac{\partial \log L}{\partial \alpha}\end{bmatrix}_{\substack{\hat{\theta}=\theta_0\\ \hat{\alpha}=\alpha_0}}$$

Where θ_0 and α_0 are the initial values of θ and α, respectively. These equations are solved iteratively till sufficiently close values of θ and $\hat{\alpha}$ are obtained.

5. Applications

As we know that size-biased distributions are very much useful for modeling data relating to situation when organisms occur in groups and group size influences the probability of detection. In this section an attempt has been made to test the goodness of fit of GSBPLD with data relating to the size distribution of freely-forming small groups at various public places, available in James [15] and Coleman and James [16].

Based on the values of chi-square $\left(\chi^2\right)$, p-value, -2logL and AIC (Akaike Information Criterion), it is obvious that GSBPLD gives much closer fit than SBPD and SBPLD and hence it can be considered an important distribution for modeling size distribution of freely-forming small groups at various public places and other data which structurally excludes zero counts. Note that AIC has been calculated using the formula $AIC = -2\log L + 2k$, where k is the number of parameters involved in the distribution.

Table 1. Pedestrians-Eugene, Spring, Morning.

Group size	Observed frequency	Expected frequency		
		SBPD	SBPLD	GSBPLD
1	1486	1452.4	1532.5	1486.1
2	694	743.3	630.6	694.5
3	195	190.2	191.9	192.4
4	37	32.4	51.3	41.1
5	10	4.1	12.8	7.5
6	1	0.6	3.9	1.5
Total	2423	2423.0	2423.0	2423.0
ML Estimate		$\hat{\theta}=0.5118$	$\hat{\theta}=4.5082$	$\hat{\theta}=10.7002$ $\hat{\alpha}=4.3729$
χ^2		7.370	13.760	0.880
d.f.		2	3	2
p-value		0.0251	0.003	0.644
$-2\log L$		10445.34	4622.36	4607.70
AIC		10447.34	4624.36	4611.70

Table 2. Play Groups-Eugene, Spring, Public Playground A.

Group size	Observed frequency	Expected frequency		
		SBPD	SBPLD	GSBPLD
1	306	292.2	309.4	303.7
2	132	155.2	131.2	138.9
3	47	41.2	41.1	41.4
4	10	7.3	11.3	10.1
5	2	1.1	4.0	2.9
Total	497	497.0	497.0	497.0
ML Estimate		$\hat{\theta}=0.5312$	$\hat{\theta}=4.3548$	$\hat{\theta}=6.4729$ $\hat{\alpha}=2.2561$
χ^2		6.479	0.932	1.194
d.f.		2	2	1
p-value		0.039	0.6281	0.2753
$-2\log L$		2142.03	971.86	970.90
AIC		2144.03	973.86	974.90

Table 3. Play Groups-Eugene, Spring, Public Playground D.

Group size	Observed frequency	Expected frequency		
		SBPD	SBPLD	GSBPLD
1	305	296.5.2	314.4	304.2
2	144	159.0	134.4	148.0
3	50	42.6	42.5	42.9
4	5	7.6	11.8	9.6
5	2	1.0	3.1	1.8
6	1	0.3	0.8	0.5
Total	507	507.0	507.0	507.0
ML Estimate		$\hat{\theta}=0.5365$	$\hat{\theta}=4.3179$	$\hat{\theta}=9.9326$ $\hat{\alpha}=4.2180$
χ^2		3.035	6.415	2.56
d.f.		2	2	1
p-value		0.219	0.040	0.1095
$-2\log L$		2376.75	993.10	989.93
AIC		2378.75	995.1	993.93

The fitted probability plots of the distributions for the considered distributions has been shown in figure 3 and it is obvious that GSBPLD gives much closer fit than SBPD and SBPLD.

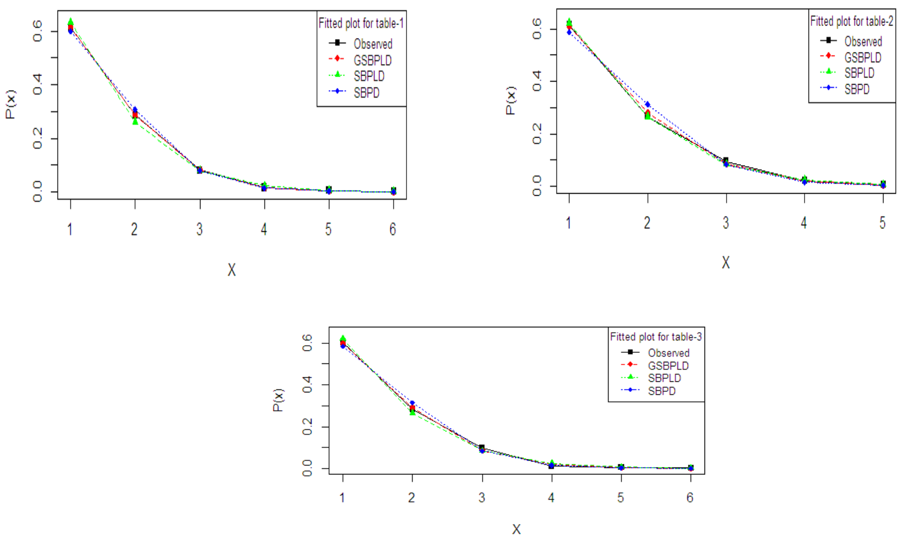

Fig. 3. Fitted probability plot of distributions for table-1, 2, 3 respectively.

6. Conclusion

A generalized size-biased Poisson-Lindley distribution (GSBPLD), which includes size-biased Poisson-Lindley distribution (SBPLD) as particular case, has been suggested and detailed study has been conducted about its nature and behavior for varying values of parameters. The moments based measures including coefficients of variation, skewness, kurtosis, and index of dispersion have been derived and their nature and behavior have been explained graphically with varying values of the parameters. Maximum likelihood estimation has been discussed for estimating its parameters. Applications and goodness of fit of the distribution have been explained through datasets relating to size distribution of freely-forming small group and fit has been found satisfactory over SBPD and SBPLD.

References

1. M. E. Ghitany and D. K. Al-Mutairi, Metron – Int. J. Stat. LXVI (3), 299 (2008).
2. R. Shanker, F. Hagos, and Y. Abrehe, Am. J. Math. Stat. **5(6)**, 354 (2015).
3. M. Sankaran, *Biometrics* **26**, 145 (1970). https://doi.org/10.2307/2529053
4. R. Shanker and F. Hagos, Biomet. Biostat. Int. J. **2(4)**, 1 (2015).
5. D. V. Lindley, J. Royal Stat. Soc. Ser. *B* **20**, 102 (1958).
6. M. E. Ghitany, B. Atieh, and S. Nadarajah, Math. Comput. Simulation **78(4)**, 493 (2008). https://doi.org/10.1016/j.matcom.2007.06.007
7. R, Shanker, F. Hagos, and S. Sujatha, Biomet. Biostat. Int. J. **2(5)**, 1 (2015).
8. E. Mahmoudi and H. Zakerzadeh, Communications in Statistics-Theory Methods **39**, 1785 (2010). https://doi.org/10.1080/03610920902898514
9. H. Zakerzadeh and A. Dolati, J. Math. Extension **3(2)**, 13 (2009).
10. R. Shanker, Insights Biomed. **1(2)**, 1 (2016).
11. E. W. Stacy, Ann. Math. Stat. **33**, 1187 (1962). https://doi.org/10.1214/aoms/1177704481
12. R. Shanker and K. K. Shukla, Biomet. Biostat. Int. J. **4(7)**, 1 (2016).
13. R. A. Fisher, Ann. Human Genetics **6**, 13 (1934). https://doi.org/10.1111/j.1469-1809.1934.tb02105.x
14. C. R. Rao, On Discrete Distributions Arising out of Methods of Ascertainment, ed. G. P. Patil, In: Classical and Contagious Discrete Distributions (Statistical Publishing Society, Calcutta, 1965) pp. 320 – 332.
15. J. James, Am. Sociol. Rev. **18**, 569 (1953). https://doi.org/10.2307/2087444
16. J. S. Coleman and J. James, Sociometry **24**, 36 (1961). https://doi.org/10.2307/2785927