Chapter 2

Some Properties and Application of Size-biased Poisson-Lindley Distribution

2.1 Introduction

Poisson-Lindley distribution is a well-known one parameter mixture Poisson distribution which has wide applications in the theory of accident proneness with probability mass function (p.m.f)

$$f(x;\theta) = \frac{\theta^2(x+\theta+2)}{(1+\theta)^{x+3}}, \quad x = 0,\ 1\ , \ldots\ , \quad \theta > 0. \tag{2.1.1}$$

The distribution arises from the Poisson distribution when its parameter λ follows a continuous Lindley distribution due to Lindley (1958). Sankaran (1970) investigated this distribution with application to errors and accidents data. Some of the difficulties in obtaining maximum likelihood estimator (MLE) of the parameter θ of Poisson-Lindley distribution was pointed out by him. Ghitany and Al-Mutairi (2008) obtained size-biased form of the Poisson-Lindley distribution along with its applications. Some properties and application of the size-biased Poisson-Lindley distribution are reviewed in this chapter. Some results of the chapter were presented in Dutta and Borah (2014).

Size-biased distributions are a special case of weighted distributions. The concept of weighted distributions can be traced to the study of the effect of methods of ascertainment upon estimation of frequencies by fisher (1934), weighted distributions were later formalized in a unifying theory by Rao (1965). Such distributions arise naturally when observations generated from a stochastic process are not given equal chance of being recorded; instead they are recorded according to some weighted function.

If the random variable X has distribution $f(x;\theta)$, with unknown parameter θ, then the corresponding weighted distribution is of the form

$$f^{w}(x;\theta) = \frac{w(x)f(x;\theta)}{E[w(x)]} \qquad (2.1.2)$$

where $w(x)$ is a non-negative weight function such that $E[w(x)]$ exists. When the weight function of the weighted distribution depends on the lengths of the units of the interest; the resulting distribution is called length-biased. The weighted distribution with the weight function $w(x) = x$ is known as size-biased distribution.

A study of size-biased sampling and related form- invariant weighted distributions was made by Patil and Ord (1975). Van Deusen (1986) arrived at size biased distribution theory independently and applied it to fitting distributions of diameter at breast height (DBH) data arising from horizontal point sampling (HPS) (Grosenbaugh) inventories. Gove (2003) reviewed some of the more recent results on size-biased distributions pertaining to parameter estimation in forestry. Mir (2009) also discussed some of the discrete size-biased distributions.

The organization of this chapter is as follows: Section 2.2 deals with the materials and methods used for study the size-biased Poisson-Lindley (SBPL) distribution. In section 2.3, the derivation of size-biased Poisson-Lindley distribution is presented. Some distributional properties of the distribution including distribution function, probability generating function, recursive relations of probabilities, moments and cumulants, coefficient of variation etc. is derived in section 2.4. Some estimation methods for estimating the parameter are discussed in section 2.5. In section 2.6, the distribution is fitted to some reported data sets for empirical comparison and to see the application part. Also the graphical representations of the fitting of the distribution have been discussed in section 2.7.

2.2 Materials and Methods

To introduce the concept of a weighted distribution, suppose X is a non-negative random variable with its natural probability mass function (pmf)$f(x;\theta)$, where the natural parameter is θ. Let the weight function be $w(x)$ is a non-negative function. Then the weighted pmf is obtained as

$$f^{w}(x;\theta) = \frac{w(x)f(x;\theta)}{E[w(x)]}, \quad x = 1,2,3,\dots \tag{2.2.1}$$

Assuming that $E[w(x)]$ exists.

By taking weight $w(x) = x$ we obtain size-biased distribution. For example, when $w(x)$ =x, in that case $E(x)$=μ. Then the distribution is called size-biased distribution with pmf

$$f^{*}(x;\theta) = \frac{xf(x;\theta)}{\mu} \tag{2.2.2}$$

2.3 Size-biased Poisson-Lindley distribution

A size-biased Poisson-Lindley distribution is obtained by applying weights $w(x) = x$ to the Poisson-Lindley distribution. The pmf of the size-biased Poisson-Lindley distribution can be obtained as

$$f^{*}(x;\theta) = \frac{xf(x;\theta)}{\mu} = \frac{\theta^3}{\theta+2}\frac{x(x+\theta+2)}{(1+\theta)^{x+2}} \quad ,x = 1,\ 2,\ \dots, \quad \theta > 0 \tag{2.3.1}$$

where $\mu = \frac{\theta+2}{\theta(1+\theta)}$ is the mean of the Poisson-Lindley distribution with pmf (2.1.1).

Ghitany and Al-Mutairi (2008) investigated this distribution with application and also investigated some distributional properties of size-biased Poisson-Lindley distribution. They also showed that the size-biased Poisson-Lindley (SBPL) distribution also arises from the size-biased Poisson (SBP) distribution with pmf

$$g(x\backslash\lambda) = e^{-\lambda}\frac{\lambda^{x-1}}{(x-1)!}, \quad x = 1, 2, 3, \ldots, \quad \lambda > 0. \qquad (2.3.2)$$

when its parameter λ follows a size-biased Lindley (SBL) model with pdf

$$h(\lambda\,;\,\theta) = \frac{\theta^3}{\theta+2}\lambda\,(1+\lambda)e^{-\theta\lambda}, \quad \lambda > 0, \quad \theta > 0. \qquad (2.3.3)$$

They proved that, the method of moments and the maximum likelihood estimators of size-biased Poisson-Lindley distribution are positively biased (asymptotically unbiased) for small (large) sample size, consistent and asymptotically normal with almost equal asymptotic variances.

2.4 Distributional properties of size-biased Poisson-Lindley distribution

The pmf of the size-biased Poisson-Lindley (SBPL) distribution is given by

$$P_x(\theta) = \frac{\theta^3}{\theta+2}\frac{x(x+\theta+2)}{(1+\theta)^{x+2}}, \quad x = 1,2,3, \ldots, \quad \theta > 0. \qquad (2.4.1)$$

where θ is known as parameter of the distribution and denote the distribution by SBPL (θ). [cf. Ghitany and Al-Mutairi (2008)]

Note that, the total probability must be equal to 1. That is,

$$\sum_{x=1}^{\infty} P_x(\theta) = \frac{\theta^3}{(\theta+2)(1+\theta)^2}\sum_{x=1}^{\infty}\frac{x(x+\theta+2)}{(1+\theta)^x} = 1.$$

2.4.1 Distribution function

The corresponding distribution function of pmf (2.4.1) is given as

$$F(x) = P(X \leq x) = \frac{\theta^3}{(\theta+2)(1+\theta)^2} \sum_{m=1}^{x} \frac{m(m+\theta+2)}{(1+\theta)^m} \quad , x = 1,2,3, \ldots \qquad (2.4.2)$$

2.4.2 Probability generating function (pgf)

The probability generating function (pgf) of the SBPL distribution may be written as

$$g(t) = \frac{\theta^3 t(3+\theta-t)}{(2+\theta)(1+\theta-t)^3}, \quad \theta > 0 \qquad (2.4.3)$$

Expanding the equation (2.4.3) and then equating the co-efficient of 't^r' on both sides, we get the expression for probability to obtain the higher order probabilities of the distribution as

$$P_{r+1} = \frac{(4r+2\theta r+4)P_r - rP_{r-1}}{r(3+\theta)(1+\theta)}, \qquad r \geq 1 \qquad (2.4.4)$$

where, $$P_1 = \frac{\theta^3(\theta+3)}{(\theta+2)(\theta+1)^3} \qquad (2.4.5)$$

The general expression for 'r^{th}' order probability of the distribution is given as

$$P_r = \frac{\theta^3 r(\theta+r+2)}{(\theta+2)(1+\theta)^{r+2}} \quad ; \quad r = 1,2, \ldots, \theta > 0 \qquad (2.4.6)$$

The higher order probabilities can be obtained very easily by using relation (2.4.6).

2.4.3 Factorial moment generating function (fmgf)

The factorial moment generating function (fmgf) of the SBPL distribution may be written as

$$g(\mathrm{t}+1)=\frac{\theta^3}{(\theta+2)}\frac{(1+t)(\theta+2-t)}{(\theta-t)^3}, \quad \theta>0 \tag{2.4.7}$$

Differentiating the equation (2.4.7) with respect to 't' and equating the coefficients of $\frac{t^r}{r!}$, the following expression for factorial moments may be obtained as

$$\mu_{(r)}=\frac{1}{\theta^3}r\left[3\theta^2\mu_{(r-1)}-(r-1)\{3\theta\mu_{(r-2)}+(r-2)\mu_{(r-3)}\}\right], \quad r\geq 3. \tag{2.4.8}$$

where, $$\mu_{(1)}=\frac{\theta^2+4\theta+6}{\theta(\theta+2)} \tag{2.4.9}$$

$$\mu_{(2)}=\frac{2(2\theta^2+9\theta+12)}{\theta^2(\theta+2)} \tag{2.4.10}$$

where '$\mu_{(r)}$'stands for the r^{th} order factorial moment.

2.4.4 Moment generating function (mgf)

The moment generating function (mgf) may be written as

$$m(t)=\frac{\theta^3e^t(3+\theta-e^t)}{(2+\theta)(1+\theta-e^t)^3}, \quad \theta>0 \tag{2.4.11}$$

The expression for recursive relation of raw moments of the distribution is given as

$$\mu'_{r+1}=\frac{1}{A}\left[B+(1+\theta)(3+\theta)\mu'_r+\sum_{j=1}^{r}\left\{\binom{r}{j}(4+2\theta-2^j)+\binom{r}{j-1}(4-2^{j-1})\right\}\mu'_{r+1-j}\right], \quad r>1 \tag{2.4.12}$$

where $A = \theta(\theta + 2)$

$B = (4 - 2^r)$

$$\mu_1^{'} = \frac{\theta^2+4\theta+6}{\theta(\theta+2)} \tag{2.4.13}$$

The second and third order raw moments of the distribution obtained from relation (2.4.12) given as

$$\mu_2^{'} = \frac{\theta^4+10\theta^3+40\theta^2+72\theta+48}{\theta^2(1+\theta)^2} \tag{2.4.14}$$

$$\mu_3^{'} = \frac{\theta^6+20\theta^5+146\theta^4+564\theta^3+1104\theta^2+1052\theta+480}{\theta^3(1+\theta)^3} \tag{2.4.15}$$

Hence, the mean and variance of the SBPL distribution is given as

$$\mu = \frac{\theta^2+4\theta+6}{\theta(\theta+2)} \tag{2.4.16}$$

and $$\sigma^2 = \frac{2(\theta^3+6\theta^2+12\theta+6)}{\theta^2(\theta+2)^2} \tag{2.4.17}$$

The third and fourth order central moments of the SBPL distribution given as [cf. Ghitany and Al-Mutairi (2008)]

$$\mu_3 = \frac{2(\theta^5+10\theta^4+42\,\theta^3+72\theta+24)}{\theta^3(1+\theta)^3} \tag{2.4.18}$$

$$\mu_4 = \frac{2(\theta^7+22\theta^6+184\theta^5+780\theta^4+1800\theta^3+225\theta^2+1440\theta+360)}{\theta^4(1+\theta)^4} \tag{2.4.19}$$

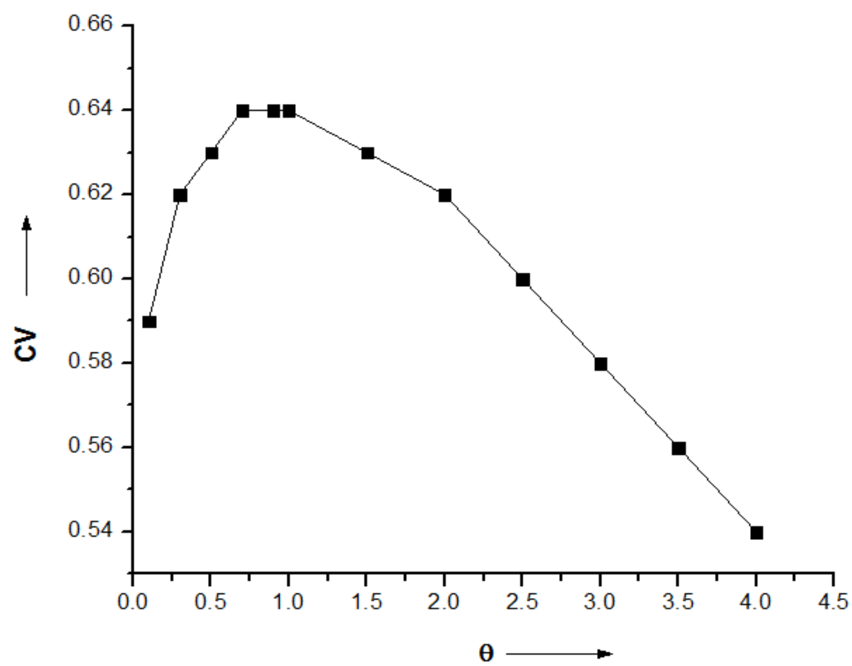
2.4.5 Coefficient of variation

The coefficient of variation (CV) is the ratio of the standard deviation to the mean. The higher is the coefficient of variation, the greater the level of dispersion around the mean. It is generally expressed as a percentage. Without units, it allows for comparison between distributions of values whose scales of measurement are not comparable.

The coefficient of variation of the SBPL distribution may be written as

$$CV = \frac{\sqrt{V(X)}}{\mu} = \frac{\sqrt{2(\theta^3+6\theta^2+12\theta+6)}}{\theta^2+4\theta+6} \qquad (2.4.20)$$

Fig.2.1 The coefficient of variation of the SBPL distribution for different values of θ .



From the above figure it is observed that, the coefficient of variation increases to a certain point, then it again decreases as θ increases which is not exactly a bell-shaped curve.

2.4.6 Cumulant generating function (cgf)

The cumulant generating function of the SBPL distribution may be written as

$$K(t) = \log m(t) = log\left[\frac{\theta^3 e^t(3+\theta-e^t)}{(2+\theta)(1+\theta-e^t)^3}\right], \theta > 0 \tag{2.4.21}$$

The expression for recursive relation of cumulants is given by

$$K_{r+1} = \frac{1}{\theta(\theta+2)}\left[(1+\theta)(3+\theta)+4-2^r+\sum_{j=1}^{r}(4+2\theta-2^j)\binom{r}{j}K_{r+1-j}\right],\ r > 1 \tag{2.4.22}$$

where $K_1 = \frac{\theta^2+4\theta+6}{\theta(\theta+2)}$ (Mean)

be the first order cumulants of the distribution.

2.4.7 Harmonic mean

The harmonic mean of the SBPL distribution may be obtained as

$$\frac{1}{H} = \sum_{x=1}^{\infty}\frac{1}{x}P_x(\theta) = \sum_{x=1}^{\infty}\frac{1}{x}\left(\frac{\theta^3}{\theta+2}\right)\frac{x(x+\theta+2)}{(1+\theta)^{x+2}}, x = 1, 2, \dots..$$

$$\Rightarrow \frac{1}{H} = \frac{\theta^3}{(\theta+2)(1+\theta)}\frac{(1+2\theta)}{\theta^2}$$

$$\Rightarrow H = \frac{(\theta+2)(1+\theta)}{\theta(1+2\theta)}, \tag{2.4.23}$$

where $P_x(\theta)$ be the probability mass function (2.4.1) of the SBPL distribution.

Remark 2.4.1 Note that, the harmonic mean of the size-biased Poisson-Lindley (SBPL) distribution is less than the mean of this distribution for all values of $\theta > 0$.

2.4.8 Survival function or reliability function

The reliability function of the SBPL distribution may be written as

$$R(x) = 1 - F(x) = \frac{(\theta+2)(1+\theta) - \theta^3 \sum_{m=1}^{x} \frac{m(m+\theta+2)}{(1+\theta)^m}}{(\theta+2)(1+\theta)} \; ; \; x = 1,2, \dots, \theta > 0 \qquad (2.4.24)$$

2.4.9 Hazard rate function

The hazard rate function of the SBPL distribution is given by

$$h(x) = \frac{\theta^3 x(x+\theta+2)}{(1+\theta)^x [(\theta+2)(1+\theta) - \theta^3 \sum_{m=1}^{x} \frac{m(m+\theta+2)}{(1+\theta)^m}]} \; ; \; x = 1, 2, 3, \dots, \theta > 0 \qquad (2.4.25)$$

2.4.10 Generalized form of size-biased Poisson-Lindley distribution

Let X is a random variable with generalized Poisson-Lindley (GPL) distribution with parameter α and θ . Then the generalized size-biased Poisson-Lindley distribution of X is given by

$$f^*(x;\alpha,\theta) = \frac{x}{\mu} f(x;\alpha,\theta)$$

$$= \frac{\Gamma(x+\alpha)}{(x-1)!\,\Gamma(\alpha+1)} \frac{\theta^{\alpha+2}}{(1+\theta)^{x+\alpha}\{1+\alpha(1+\theta)\}} \left(\alpha + \frac{x+\alpha}{1+\theta}\right), \quad x = 1,2,\dots, \qquad (2.4.26)$$

where $\theta > 0$ and $\alpha > 0$ be the two parameter of the distribution and $\mu = \frac{1+\alpha(1+\theta)}{\theta(1+\theta)}$ be the mean of generalized Poisson-Lindley distribution of Mohmoudi and Zakerzadeh (2010).

The probability generating function (pgf) of size-biased Poisson-Lindley distribution is given as

$$G_{SBGPL}(t) = \left(\frac{\theta}{\theta-t+1}\right)^{\alpha+2} \frac{t\{\alpha(\theta-t+2)+1\}}{\alpha(1+\theta)+1} \quad (2.4.27)$$

Remark 2.4.2: Note that, in (2.4.26) as $\alpha \to 0$ then it reduces to one parameter size-biased Poisson-Lindley (SBPL) distribution with parameter θ as a limiting form.

2.5 Estimation of parameter

In this section, the single parameter θ of the SBPL distribution can be estimated by employing the following methods.

2.5.1 Method of moment

Let $x_1, x_2, \ldots, x_n$ be a random sample of size n from the size-biased Poisson-Lindley(SBPL) distribution with pmf (2.4.1), then the moment estimator of $\hat{\theta}$ is obtained by setting the mean of the distribution equal to the sample mean, that is $E(x) = \mu$.

The MoM estimate $\hat{\theta}$ of θ is obtained as

$$\bar{x} = \mu$$

$$\Rightarrow \bar{x} = \frac{\theta^2 + 4\theta + 6}{\theta(\theta + 2)}$$

$$\Rightarrow \hat{\theta} = \frac{2-\bar{x}+\sqrt{\bar{x}^2+2\bar{x}-2}}{\bar{x}-1}, \quad \bar{x} > 1 \quad (2.5.1)$$

[cf. Ghitany and Al-Mutairi (2008)]

2.5.2 Method based on the first two relative frequencies

For the SBPL distribution, θ may be estimated by taking ratio of the first two relative frequencies as

$$\hat{\theta} = \frac{-(4f_2-2f_1)+\sqrt{(4f_2-2f_1)^2-4f_2(3f_2-8f_1)}}{2f_2} \qquad (2.5.2)$$

where $\frac{f_1}{N} = \frac{\theta^3(\theta+3)}{(\theta+2)(\theta+1)^3}$ and $\frac{f_2}{N} = \frac{\theta^3 2(\theta+4)}{(\theta+2)(1+\theta)^4}$ be the first relative frequencies of size-biased Poisson-Lindley distribution.

2.6 Applications

It is believed that the size-biased Poisson-Lindley (SBPL) distribution should give a reasonably good fit to some numerical data for which various distributions was fitted earlier. Therefore, to illustrate the applications and to justify suitability of size-biased Poisson-Lindley distribution in a practical application, fitting this distribution to some data sets and have compared them with size-biased Poisson distribution as measured by χ^2 criterion. In getting the χ^2 criterion for goodness of fit, tail frequencies are grouped to obtain 5 or slightly greater than 5 for the expected frequency in each group.

In **Table** 2.1, we have considered immunogold assay data [data from Cullen et al. (1990)] and animal abundance data in **Table** 2.2 [data from Keith and Meslow (1968)] for fitting of the size-biased Poisson-Lindley distribution.

Fitting of distribution to immunogold assay data

Cullen et al. (1990) gave counts of sites with 1, 2, 3, 4 and 5 particles from immunogold assay data. The counts were 122, 50, 18, 4, and 4. For the problem chosen, $N = 198\ and\ \bar{x} = 1.576.$

Fitting of distribution to animal abundance data

In a study carried out by Keith and Meslow (1968), snowshoe hares were captured over 7 days. There were 261 hares caught over 7 days. Of these, 188 were caught once, 55 were caught twice, 14 were caught three times, 4 were caught four times, and 4 were five times. For the problem chosen, $N = 261 \; and \; \bar{x} = 1.425$.

The null hypothesis H_0: "Distribution of the data is size-biased Poisson-Lindley." Against the alternative hypothesis H_1: "Distribution of data is not size-biased Poisson-Lindley."

The observed frequencies together with the expected frequencies of size-biased Poisson [Ghitany and Al-Mutairi (2008)] and size-biased Poisson-Lindley distributions are shown in **Table** 2.1 and **Table** 2.2, by using different method of estimation.

It is observed from **Table** 2.1 and **Table** 2.2, the expected frequencies of fitted size-biased Poisson-Lindley distribution by using both of the estimation procedure provides an excellent fit to the observed data as compared to the other one. So, we accept the hypothesis that the given data came from a size-biased Poisson-Lindley distribution. Also it is seen that, method of moment estimation gives better result in fitting of the distribution as compared to the method based on the first two relative frequencies.

It is also clear from following tables that, size-biased Poisson-Lindley distribution gives a "good fit" to immunogold assay data and animal abundance data. Thus, SBPL distribution found to be suitable use in forestry and bio-medical research.

Table 2.1 Comparison of observed frequencies for immunogold assay data with fitted size-biased Poisson (SBP) and size-biased Poisson-Lindley (SBPL) distributions and Chi-square goodness-of-fit test. [data Cullen et al. (1990)]

No. of Particles	Observed frequencies (O_i)	Expected frequencies (E_i)		
		SBP(MoM) Ghitany et al.(2008)	SBPL(MoM)	SBPL(RF)
1	122	111.3	118.9	125.5
2	50	64.1	54.0	51.4
3	18	18.5	18.2	15.7
4	4	3.5	5.4	4.3
5	4	0.6	1.5	1.1
Total	198	198.0	198.0	198.0
Estimated parameters		$\hat{\lambda}$= 0.576	$\hat{\theta}$= 4.048	$\hat{\theta}$ = 4.528
χ^2		4.642	0.554	2.067
d.f.= $(n-p-1)$		1	2	2
$p-value$		0.031	0.758	0.356

Note: SBP: Size-biased Poisson distribution.

SBPL: Size-biased Poisson-Lindley distribution.

RF: Based on the first two relative frequencies.

MoM: Method of moments.

Table 2.2 Comparison of observed frequencies for animal abundance data with fitted size-biased Poisson (SBP) and size-biased Poisson-Lindley (SBPL) distributions and Chi-square goodness-of-fit test. [data Keith and Meslow (1968)]

No. of Counts	Observed frequencies (O_i)	Expected frequencies (E_i)		
		SBP(MoM) Ghitany et al.(2008)	SBPL(MoM)	SBPL(RF)
1	184	170.6	177.2	188.9
2	55	72.5	62.6	56.4
3	14	15.4	16.3	12.5
4	4	2.2	3.8	2.4
5	4	0.3	1.1	0.5
Total	261	261.0	261.0	260.7
Estimated parameters		$\hat{\lambda}$= 0.425	$\hat{\theta}$=5.343	$\hat{\theta}$= 6.402
χ^2		6.216	1.214	2.991
$d.f.$= $(\mathrm{n}-\mathrm{p}-1)$		1	1	1
$p-value$		0.013	0.271	0.084

Note: RF: Based on the first two relative frequencies.

MoM: Method of moments.

2.7 Graphical representations of fitting of distribution

For graphical representation of fitting of the distribution, some graphs are plotted in **Fig. 2.2** and **Fig. 2.3** by taking number of observations along the horizontal axis and the corresponding frequencies (f_i ; $i = 1, 2, \dots, 5$) along the vertical axis. In **Fig**. **2.2** taking no. of particles, i.e. x_i ; $i = 1,2,..,5$ along the horizontal axis and the corresponding frequencies 122, 50, 18,4 and 4 along the vertical axis four graphs are plotted for observed frequencies (O_i), expected frequencies (e_i) of fitted size-biased Poisson and size-biased Poisson-Lindley distributions given in the above **Table 2.1**.

Similarly, in **Fig.2.3** taking no. of counts, i.e. x_i ; $i = 1,2,..,5$ along the horizontal axis and the corresponding frequencies 184, 55, 14, 4 and 4 along the vertical axis it contains four graphs for observed frequencies (O_i) and expected frequencies (e_i) of size-biased Poisson distribution (Method of moment), size-biased Poisson-Lindley distribution (Method of moment and method based on the first two relative frequencies) given in **Table 2.2**.

Fig. 2.2 The observed frequencies (o_i) for immunogold assay data and expected frequencies (e_i) of fitted size-biased Poisson (SBP) and size-biased Poisson - Lindley (SBPL) distributions.

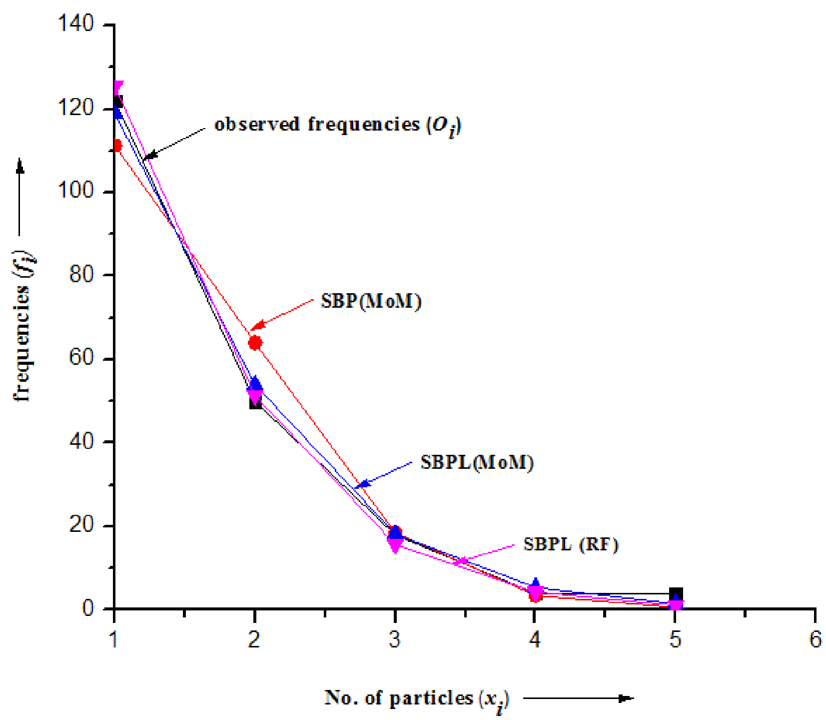
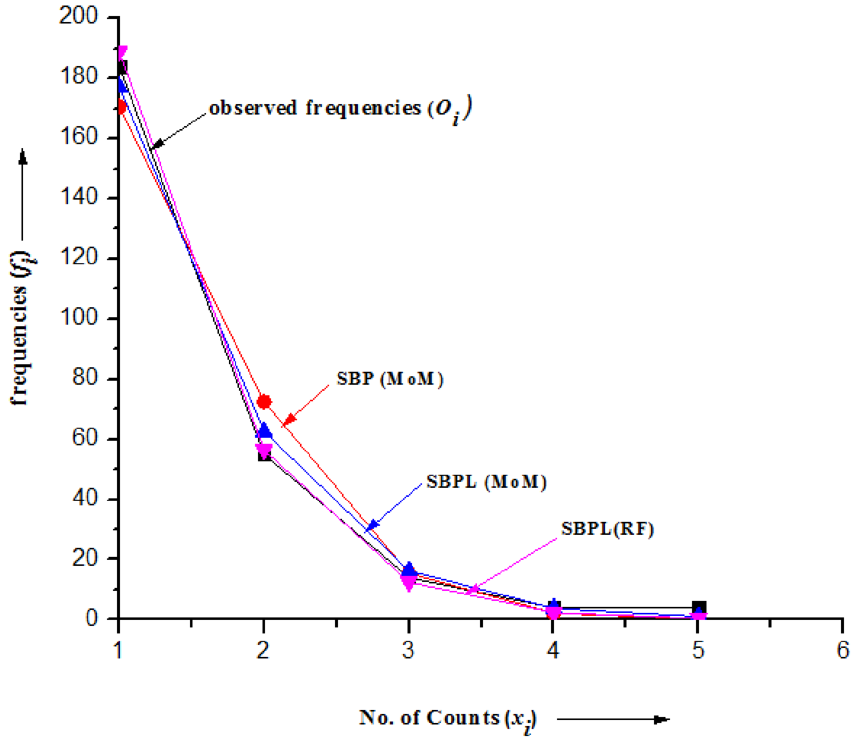


Fig. 2.3 The observed frequencies (o_i) for animal abundance data and expected frequencies (e_i) of fitted size-biased Poisson (SBP) and size-biased Poisson - Lindley (SBPL) distributions.



It is observed from **Fig. 2.2** and **2.3** that, the graphs plotted for expected frequencies of size-biased Poisson-Lindley (SBPL) distribution for both of the method of estimation gives close fit to the observed frequencies than the fit of the size-biased Poisson (SBP) distribution for both the cases. But, the graph plotted for expected frequencies of size-biased Poisson-Lindley (SBPL) distribution by using method of moment gives very close fit than the graph plotted by using the method based on the first two relative frequencies in both cases.

****