Thai **J**ournal of **M**athematics : (2020) 183–192
Special Issue: The 14^{th} IMT-GT ICMSA 2018

http://thaijmath.in.cmu.ac.th
Online ISSN 1686-0209

A Modified Poisson Lindley Distribution with Application

Yupapin Atikankul[†], **Winai Bodhisuwan**[†] **and Ampai Thongteeraparp**[†,1]

[†]Department of Statistics, Faculty of Science,
Kasetsart University, Bangkok 10900, Thailand
e-mail : `yupapin.a@ku.th, fsciwnb@ku.ac.th and fsciamu@ku.ac.th`

Abstract : In this paper, a new weighted distribution is proposed. The distribution is a mixture of the size-biased Poisson distribution and the size-biased new weighted Lindley distribution. Some mathematical properties of the distribution are derived. The maximum likelihood estimators of the unknown parameters are investigated. Finally, the proposed distribution is applied to a real data set for illustrating of the usefulness of the distribution.

Keywords : size-biased distribution; new weighted Lindley distribution; maximum likelihood estimation
2010 Mathematics Subject Classification : 60E05; 62H10

1 Introduction

When observations are recorded with unequal probability, the original distribution may not be appropriate to describe these recorded observations. Weighted distributions can be applied in this situation. Moreover, they have been applied for various applications such as ecology, reliability, biomedicine, family data, etc. The concept of weighted distributions was proposed by Fisher [1]. Later, Rao [2] formulated the idea of weighted distributions and presented several sampling situations that these situations cannot be considered by original distributions.

Suppose that the original observation have the probability density function (pdf) $f_0(x)$, then the recorded observation with unequal probability may be assumed to have the pdf $f(x)$, that is of the form

$$f(x) = \frac{w(x)f_0(x)}{E(w(X))},$$

where $w(x)$ is a non-negative weighted function and $\mathrm{E}(w(X))$ exists.

Patio and Rao [3] studied some models leading to weighted functions when $w(x) = x$, called a size-biased function. This function is the best known weighted function. The size-biased distribution can be written as

$$f(x) = \frac{xf_0(x)}{E(X)}. \tag{1.1}$$

[0]This research was supported by Kasetsart University and Rajamangala University of Technology Phra Nakhon.

[1]`Corresponding author.`

Some size-biased discrete distributions have been presented in [3]. Ghitany and Al-Mutairi [4] proposed the size-biased Poisson-Lindley distribution. Their results showed that the size-biased Poisson-Lindley distribution is more flexible than the size-biased Poisson distribution. After that, size-biased mixed Poisson distributions have been occurred in many literatures. Shanker and Fesshaye studied the size-biased Poisson-Amarendra distribution [5] and the size-biased Poisson-Sujatha distribution [6]. Next, Shanker proposed the size-biased Poisson-Shanker distribution [7] and the size-biased Poisson-Akash distribution [8]. Moreover, Shanker and Mishra [9] also introduced the size-biased two parameter Poisson-Lindley distribution.

Asgharzadeh et al. [10] proposed the new weighted Lindley (NWL) distribution which is a mixture of the weighted exponential distribution and the weighted gamma distribution. They compared the NWL distribution with various distributions which the NWL distribtuion provide a better fit than others. The pdf of the NWL distribution is

$$g_0(\lambda) = \frac{\theta^2(1+\alpha)^2}{\alpha\theta(1+\alpha)+\alpha(2+\alpha)}(1+\lambda)(1-e^{-\theta\alpha\lambda})e^{-\theta\lambda}, \tag{1.2}$$

for $\lambda > 0$, $\theta > 0$, $\alpha > 0$.

In this paper, we propose the modified Poisson Lindley (MPL) distribution, which is a mixture of the size-biased Poisson distribution and the size-biased new weighted Lindley distribution. Furthermore, we apply the MPL distribution to a real data set.

2 The Distribution and Mathematical Properties

In this section, the probability mass function (pmf), cumulative distribution function (cdf), probability generating function (pgf), moment generating function (mgf) and moments of the MPL distribution are presented.

2.1 The Probability Mass Function

The size-biased new weighted Lindley (SBNWL) distribution can be obtained by substituting (1.2) in (1.1). Therefore, the pdf of the SBNWL can be derived as

$$g(\lambda) = \frac{\theta^3(1+\alpha)^3}{(1+\alpha)^3(\theta+2)-(1+\alpha)\theta-2}\lambda(1+\lambda)(1-e^{-\theta\alpha\lambda})e^{-\theta\lambda}, \tag{2.1}$$

for $\lambda > 0$, $\theta > 0$, $\alpha > 0$.

Let $X|\lambda$ be a size-biased Poisson (SBP) random variable with pmf

$$f(x|\lambda) = \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!}, \tag{2.2}$$

for $x = 1, 2, \ldots$, $\lambda > 0$, where λ is distributed as SBNWL; thus, X is a MPL random variable.

Theorem 2.1. *An unconditional random variable X is said to be the MPL distribution, if the pmf of X is*

$$f(x) = \frac{\theta^3(1+\alpha)^3}{(1+\alpha)^3(\theta+2)-(1+\alpha)\theta-2}\left[\frac{(\theta+x+2)x}{(\theta+1)^{x+2}} - \frac{(\alpha\theta+\theta+x+2)x}{(\alpha\theta+\theta+1)^{x+2}}\right], \tag{2.3}$$

for $x = 1, 2, \ldots,$ $\theta > 0$, $\alpha > 0$.

Proof. Let $X|\lambda \sim \text{SBP}(\lambda)$ and $\lambda \sim \text{SBNWL}(\theta, \alpha)$, the pmf of X can be derived as

$$\begin{aligned}
f(x) &= \int_0^\infty f(x|\lambda) g(\lambda) d\lambda \\
&= \int_0^\infty \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!} \frac{\theta^3(1+\alpha)^3}{(1+\alpha)^3(\theta+2)-(1+\alpha)\theta-2} \lambda(1+\lambda)(1-e^{-\theta\alpha\lambda})e^{-\theta\lambda} d\lambda \\
&= \frac{\theta^3(1+\alpha)^3}{\Gamma(x)[(1+\alpha)^3(\theta+2)-(1+\alpha)\theta-2]} \times \\
&\quad \left[\frac{\Gamma(x+2)}{(\theta+1)^{x+2}} + \frac{\Gamma(x+1)}{(\theta+1)^{x+1}} - \frac{\Gamma(x+2)}{(\alpha\theta+\theta+1)^{x+2}} - \frac{\Gamma(x+1)}{(\alpha\theta+\theta+1)^{x+1}}\right] \\
&= \frac{\theta^3(1+\alpha)^3}{(1+\alpha)^3(\theta+2)-(1+\alpha)\theta-2} \left[\frac{(\theta+x+2)x}{(\theta+1)^{x+2}} - \frac{(\alpha\theta+\theta+x+2)x}{(\alpha\theta+\theta+1)^{x+2}}\right].
\end{aligned}$$

□

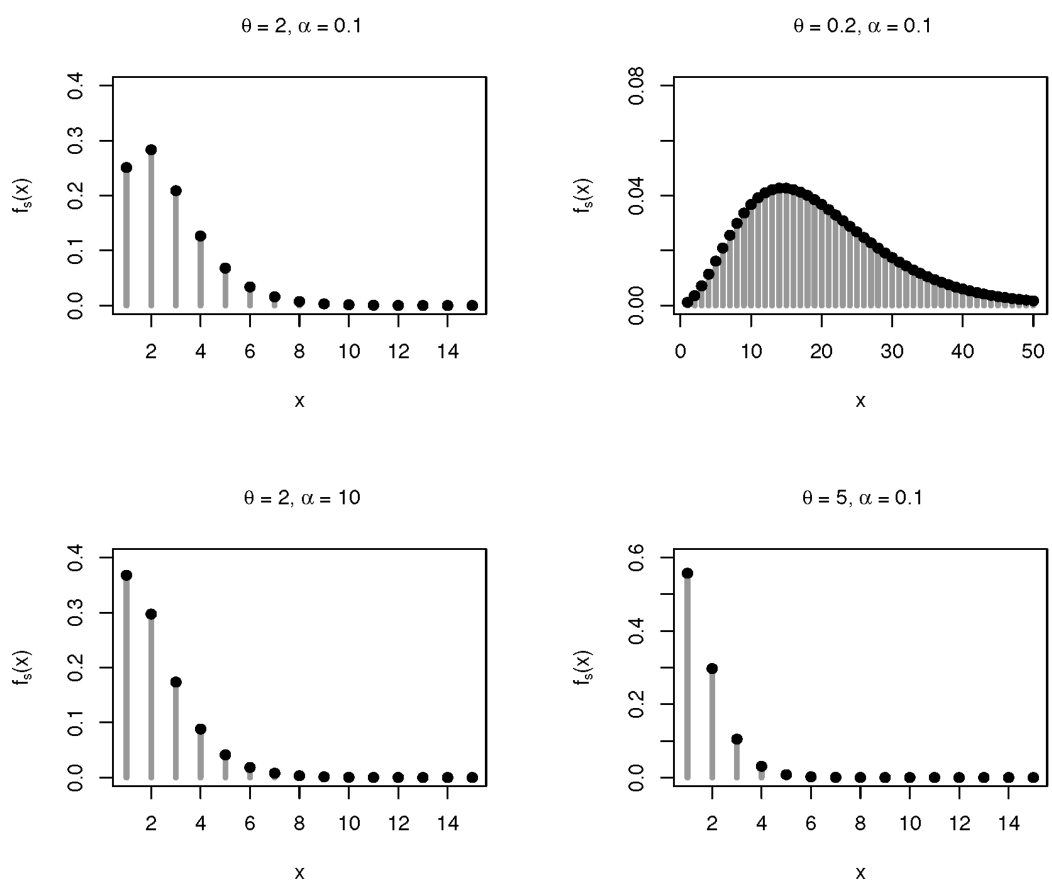

Figure 1: Some pmf plots of the MPL distribution with different parameter values

Figure 1 shows various pmf plots of the MPL distribution with specified parameter values. The pmf of MPL distribution is unimodal.

The cdf of the MPL distribution is expressed as

$$\begin{aligned} F(x) &= P(X \le x) \\ &= 1 - \Big[\frac{\theta^3(1+\alpha)^3}{(1+\alpha)^3(\theta+2)-(1+\alpha)\theta-2}\Big(\frac{\Gamma(x+3)}{\Gamma(x+1)(\theta+1)^{x+3}}\ {}_2F_1\left(1, x+3; x+1; \frac{1}{\theta+1}\right) \\ &\quad + \frac{x+1}{(\theta+1)^{x+2}}\ {}_2F_1\Big(1, x+2; x+1; \frac{1}{\theta+1}\Big) \\ &\quad - \frac{\Gamma(x+3)}{\Gamma(x+1)(\alpha\theta+\theta+1)^{x+3}}\ {}_2F_1\Big(1, x+3; x+1; \frac{1}{\alpha\theta+\theta+1}\Big) \\ &\quad - \frac{x+1}{(\alpha\theta+\theta+1)^{x+2}}\ {}_2F_1\Big(1, x+2; x+1; \frac{1}{\alpha\theta+\theta+1}\Big)\Big)\Big], \end{aligned}$$

where ${}_2F_1(a, b; c; z) = \sum_{n=0}^{\infty} \frac{(a)_n(b)_n}{(c)_n}\frac{z^n}{n!}$, called the hypergeometric function; and $(a)_n = \frac{\Gamma(a+n)}{\Gamma(a)} = a(a+1)\ldots(a+n-1)$, called Pochhammer's symbol [11].

2.2 Mathematical Properties

Theorem 2.2. *Let X be a MPL random variable with parameters (θ, α), the pgf of X is*

$$G(z) = \frac{\theta^3(1+\alpha)^3 z}{(1+\alpha)^3(\theta+2)-(1+\alpha)\theta-2}\left[\frac{\theta-z+3}{(\theta-z+1)^3} - \frac{(\alpha+1)\theta-z+3}{((\alpha+1)\theta-z+1)^3}\right].$$

Proof. Let $X|\lambda \sim \text{SBP}(\lambda)$ and $\lambda \sim \text{SBNWL}(\theta, \alpha)$, the pgf of X is obtained by

$$\begin{aligned} G(z) &= E(z^X) \\ &= \sum_{x=1}^{\infty} z^x f(x) \\ &= \sum_{x=1}^{\infty} z^x \int_0^{\infty} \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!}\frac{\theta^3(1+\alpha)^3}{(1+\alpha)^3(\theta+2)-(1+\alpha)\theta-2}\lambda(1+\lambda)(1-e^{-\theta\alpha\lambda})e^{-\theta\lambda}d\lambda \\ &= \int_0^{\infty} ze^{(z-1)\lambda}\frac{\theta^3(1+\alpha)^3}{(1+\alpha)^3(\theta+2)-(1+\alpha)\theta-2}\lambda(1+\lambda)(1-e^{-\theta\alpha\lambda})e^{-\theta\lambda}d\lambda \\ &= \frac{\theta^3(1+\alpha)^3 z}{(1+\alpha)^3(\theta+2)-(1+\alpha)\theta-2}\left[\frac{\theta-z+3}{(\theta-z+1)^3} - \frac{(\alpha+1)\theta-z+3}{((\alpha+1)\theta-z+1)^3}\right]. \end{aligned}$$

□

Theorem 2.3. *Let X be a MPL random variable with parameters (θ, α), the mgf of X is*

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

Proof. The mgf is obtained from the pgf as $M(t) = G(e^t)$ □

The first and the second raw moments of X are

$$\begin{aligned} \mu &= \delta_1\left(\frac{(\theta+2)^2+2}{\theta^4} - \frac{(\delta_2+2)^2+2}{\delta_2^4}\right), \\ \mu_2' &= \delta_1\left(\frac{\theta^3+8\theta^2+24\theta+24}{\theta^5} - \frac{\delta_2^3+8\delta_2^2+24\delta_2+24}{\delta_2^5}\right), \end{aligned}$$

where $\delta_1 = \frac{\theta^3(1+\alpha)^3}{(1+\alpha)^3(\theta+2)-(1+\alpha)\theta-2}$ and $\delta_2 = \alpha\theta + \theta$.

The central moments of X are $\mu_k = E(X-\mu)^k = \sum_{j=0}^{k} \binom{k}{j} (-1)^{k-j} \mu'_j \mu^{k-j}$. Hence, the variance, index of dispersion (ID), skewness and kurtosis of the MPL distribution are, respectively

$$\begin{aligned} V(X) &= \mu_2 = \mu'_2 - \mu^2, \\ ID(X) &= \frac{\mu_2}{\mu}, \\ \text{skewness}(X) &= \frac{\mu_3}{\mu_2^{3/2}}, \\ \text{kurtosis}(X) &= \frac{\mu_4}{\mu_2^2}. \end{aligned}$$

Figure 2 presents some plots of the mean, ID, skewness and kurtosis of the MPL distribution versus the parameters (θ, α).

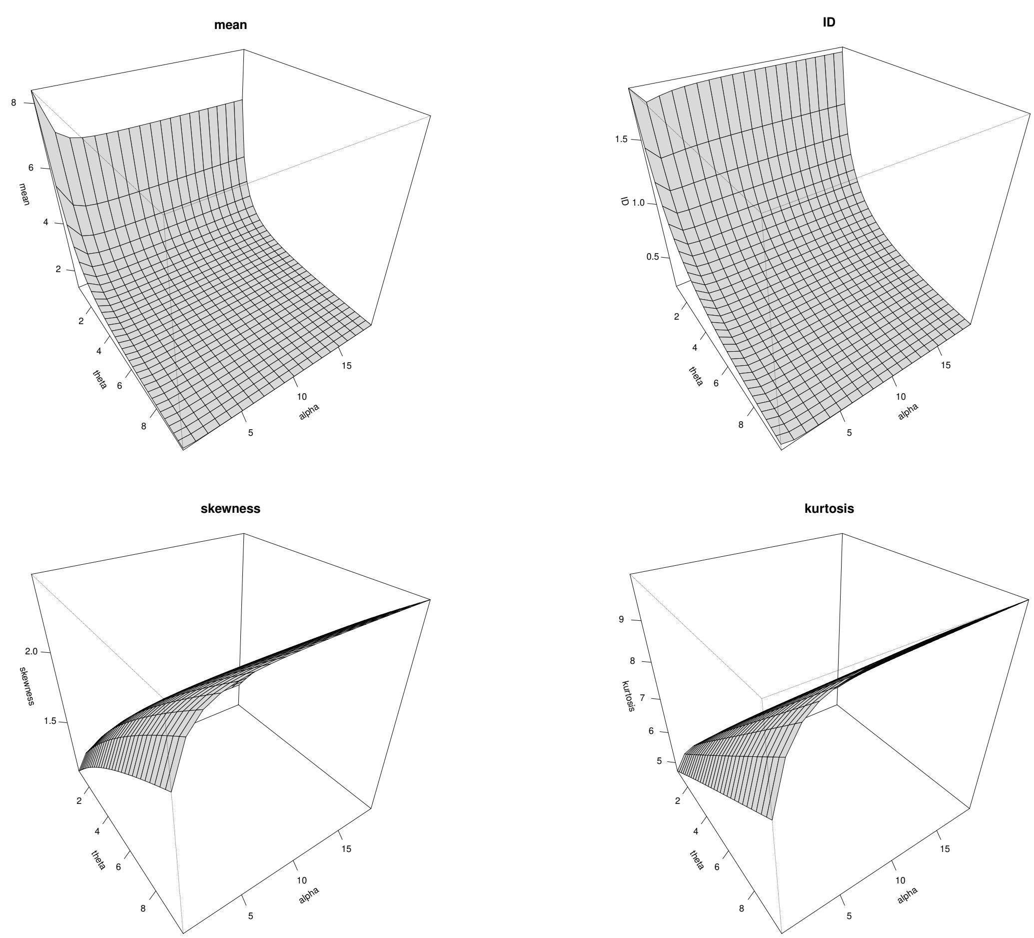

Figure 2: Mean, ID, skewness and kurtosis plots of the MPL distribution

Figure 2 illustrates that the mean is decreasing as θ and α increase. The ID is decreasing as θ increases but it is increasing as α increases. Moreover, the ID plot shows that the MPL distribution is both overdispersed and underdispersed distributions. On the other hand, the skewness and kurtosis are increasing as θ and α increase.

3 Parameter Estimation

This section, the parameter estimates of the MPL distribution are derived by the method of maximum likelihood. Moreover, the aymptotic distribution of the parameters are obtained.

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the MPL distribution with parameters (θ, α). The likelihood function of the MPL distribution is

$$L(\theta,\alpha)=\prod_{i=1}^{n}\frac{\theta^3(1+\alpha)^3}{(1+\alpha)^3(\theta+2)-(1+\alpha)\theta-2}\left[\frac{(\theta+x_i+2)x_i}{(\theta+1)^{x_i+2}}-\frac{(\alpha\theta+\theta+x_i+2)x_i}{(\alpha\theta+\theta+1)^{x_i+2}}\right].$$

The log-likelihood function can be written as

$$\begin{aligned} l(\theta,\alpha) &= 3n\log(\theta)+3n\log(1+\alpha)-n\log\big((1+\alpha)^3(\theta+2)-\theta(1+\alpha)-2\big) \\ &\quad +\sum_{i=1}^{n}\left[\log\big(x_i(\theta+x_i+2)(\alpha\theta+\theta+1)^{x_i+2}-x_i(\alpha\theta+\theta+x_i+2)(\theta+1)^{x_i+2}\big)\right] \\ &\quad -\left(\sum_{i=1}^{n}x_i+2n\right)\log(\theta+1)-\left(\sum_{i=1}^{n}x_i+2n\right)\log(\alpha\theta+\theta+1). \end{aligned}$$

The score functions are

$$\begin{aligned} \frac{\partial l(\theta,\alpha)}{\partial\theta} &= \frac{3n}{\theta}-\frac{n\left((\alpha+1)^3-\alpha-1\right)}{(\alpha+1)^3(\theta+2)-(\alpha+1)\theta-2} \\ &\quad +\sum_{i=1}^{n}\left[\frac{x_i(\alpha\theta+\theta+1)^{x_i+2}-(\alpha+1)x_i(\theta+1)^{x_i+2}}{x_i(\theta+x_i+2)(\alpha\theta+\theta+1)^{x_i+2}-x_i(\alpha\theta+\theta+x_i+2)(\theta+1)^{x_i+2}}\right. \\ &\quad -\frac{x_i(x_i+2)(\theta+1)^{x_i+1}(\alpha\theta+\theta+x_i+2)}{x_i(\theta+x_i+2)(\alpha\theta+\theta+1)^{x_i+2}-x_i(\alpha\theta+\theta+x_i+2)(\theta+1)^{x_i+2}} \\ &\quad \left.+\frac{(\alpha+1)x_i(x_i+2)(\theta+x_i+2)(\alpha\theta+\theta+1)^{x_i+1}}{x_i(\theta+x_i+2)(\alpha\theta+\theta+1)^{x_i+2}-x_i(\alpha\theta+\theta+x_i+2)(\theta+1)^{x_i+2}}\right] \\ &\quad -\frac{\left(\sum_{i=1}^{n}x_i+2n\right)}{\theta+1}-\frac{\left(\sum_{i=1}^{n}x_i+2n\right)(\alpha+1)}{\alpha\theta+\theta+1}, \end{aligned}$$

and

$$\begin{aligned} \frac{\partial l(\theta,\alpha)}{\partial\alpha} &= \frac{3n}{\alpha+1}-n\left(\frac{3(\theta+2)(1+\alpha)^2-\theta}{(\theta+2)(\alpha+1)^3-\theta(\alpha+1)-2}\right) \\ &\quad +\sum_{i=1}^{n}\frac{\theta x_i(x_i+2)(x_i+\theta+2)(\alpha\theta+\theta+1)^{x_i+1}-\theta(\theta+1)^{x_i+2}x_i}{x_i(x_i+\theta+2)(\alpha\theta+\theta+1)^{x_i+2}-x_i(\theta+1)^{x_i+2}(\alpha\theta+\theta+x_i+2)}-\frac{\left(\sum_{i=1}^{n}x_i+2n\right)\theta}{\alpha\theta+\theta+1}. \end{aligned}$$

The maximum likelihood estimators of θ and α are $\hat{\theta}$ and $\hat{\alpha}$, respectively. The maximum likelihood estimates can be obtained by setting the score functions to zero and solving them. Although they are complicate, we can get the maximum likelihood estimates by the numerical methods. In this paper, the Nelder-Mead method in optimx function [12] for R language [13] is applied to estimate the parameters of the MPL distribution.

The asymptotic variances and covariances of maximum likelihood estimators are calculated by the elements of the inverse of the Fisher information matrix, $I^{-1}(\theta,\alpha)$, but they are complicated to obtain. Thus, we can replace the Fisher information matrix with the observed information matrix. The observed information matrix of the maximum likelihood estimators of the parameters can be written as

$$J(\theta,\alpha)=\begin{pmatrix} J_{11} & J_{12}\\ J_{21} & J_{22}\end{pmatrix},$$

where $J_{11}=-\frac{\partial^2 l(\theta,\alpha)}{\partial\theta^2}$, $J_{12}=J_{21}=-\frac{\partial^2 l(\theta,\alpha)}{\partial\theta\partial\alpha}$ and $J_{22}=-\frac{\partial^2 l(\theta,\alpha)}{\partial\alpha^2}$.

The second partial derivative of the log-likelihood with respect to the parameters are

$$\begin{aligned}
\frac{\partial^2 l(\theta,\alpha)}{\partial\theta^2} &= -\frac{3n}{\theta^2}+\frac{n\left((\alpha+1)^3-\alpha-1\right)^2}{((\alpha+1)^3(\theta+2)-(\alpha+1)\theta-2)^2}\\
&+\sum_{i=1}^{n}\Big[\frac{2(\alpha+1)x_i(x_i+2)(\alpha\theta+\theta+1)^{x_i+1}-2(\alpha+1)x_i(x_i+2)(\theta+1)^{x_i+1}}{x_i(\theta+x_i+2)(\alpha\theta+\theta+1)^{x_i+2}-x_i(\alpha\theta+\theta+x_i+2)(\theta+1)^{x_i+2}}\\
&-\frac{x_i(x_i+1)(x_i+2)(\theta+1)^{x_i}(\alpha\theta+\theta+x_i+2)}{x_i(\theta+x_i+2)(\alpha\theta+\theta+1)^{x_i+2}-x_i(\alpha\theta+\theta+x_i+2)(\theta+1)^{x_i+2}}\\
&+\frac{(\alpha+1)^2x_i(x_i+1)(x_i+2)(\alpha\theta+\theta+1)^{x_i}(\theta+x_i+2)}{x_i(\theta+x_i+2)(\alpha\theta+\theta+1)^{x_i+2}-x_i(\alpha\theta+\theta+x_i+2)(\theta+1)^{x_i+2}}\Big]\\
&-\sum_{i=1}^{n}\Big[\frac{\Big(x_i(\alpha\theta+\theta+1)^{x_i+2}-(\alpha+1)x_i(\theta+1)^{x_i+2}-x_i(x_i+2)(\theta+1)^{x_i+1}(\alpha\theta+\theta+x_i+2)}{(x_i(\theta+x_i+2)(\alpha\theta+\theta+1)^{x_i+2}-x_i(\theta+1)^{x_i+2}(\alpha\theta+\theta+x_i+2))^2}\\
&+\frac{(\alpha+1)x_i(x_i+2)(\theta+x_i+2)(\alpha\theta+\theta+1)^{x_i+1}\Big)^2}{(x_i(\theta+x_i+2)(\alpha\theta+\theta+1)^{x_i+2}-x_i(\theta+1)^{x_i+2}(\alpha\theta+\theta+x_i+2))^2}\Big]\\
&\frac{\left(\sum_{i=1}^{n}x_i+2n\right)}{(\theta+1)^2}+\frac{\left(\sum_{i=1}^{n}x_i+2n\right)(\alpha+1)^2}{(\alpha\theta+\theta+1)^2},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 l(\theta,\alpha)}{\partial\theta\partial\alpha} &= -\frac{2n(2\alpha+3)}{((\theta+2)\alpha^2+(3\theta+6)\alpha+2\theta+6)^2}\\
&+\sum_{i=1}^{n}\Big[\frac{x_i(x_i+2)(x_i+\theta+2)(\alpha\theta+\theta+1)^{x_i+1}+\theta x_i(x_i+2)(\alpha\theta+\theta+1)^{x_i+1}}{x_i(x_i+\theta+2)(\alpha\theta+\theta+1)^{x_i+2}-x_i(\theta+1)^{x_i+2}(\alpha\theta+\theta+x_i+2)}\\
&+\frac{\theta x_i(x_i+1)(x_i+2)(x_i+\theta+2)(\alpha+1)(\alpha\theta+\theta+1)^x_i}{x_i(x_i+\theta+2)(\alpha\theta+\theta+1)^{x_i+2}-x_i(\theta+1)^{x_i+2}(\alpha\theta+\theta+x_i+2)}\\
&-\frac{\theta(\theta+1)^{x_i+1}x_i(x_i+2)-(\theta+1)^{x_i+2}x_i}{x_i(x_i+\theta+2)(\alpha\theta+\theta+1)^{x_i+2}-x_i(\theta+1)^{x_i+2}(\alpha\theta+\theta+x_i+2)}\\
&-\Big(\Big(\frac{\theta x_i(x_i+2)(x_i+\theta+2)(\alpha\theta+\theta+1)^{x_i+1}-\theta x_i(\theta+1)^{x_i+2}}{(x_i(\theta+x_i+2)(\alpha\theta+\theta+1)^{x_i+2}-x_i(\theta+1)^{x_i+2}(\alpha\theta+\theta+x_i+2))^2}\Big)\\
&\times\Big(x_i(\alpha\theta+\theta+1)^{x_i+2}+x_i(x_i+2)(\theta+x_i+2)(\alpha+1)(\alpha\theta+\theta+1)^{x_i+1}\\
&-(\theta+1)^{x_i+1}x_i(x_i+2)(\alpha\theta+\theta+x_i+2)+(\theta+1)^{x_i+2}x_i(-\alpha-1)\Big)\Big)\Big]-\frac{\left(\sum_{i=1}^{n}x_i+2n\right)}{(\alpha\theta+\theta+1)^2},
\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 l(\theta,\alpha)}{\partial\alpha^2} &= -\frac{3n}{(\alpha+1)^2}+\frac{n(3(\theta+2)(1+\alpha)^2-\theta)^2}{((\theta+2)(\alpha+1)^3-\theta(\alpha+1)-2)^2}-\frac{6(\theta+2)n(\alpha+1)}{(\theta+2)(\alpha+1)^3-\theta(\alpha+1)-2}\\ &+\sum_{i=1}^{n}\Big[\frac{\theta^2 x_i(x_i+1)(x_i+2)(x_i+\theta+2)(\alpha\theta+\theta+1)^{x_i}}{x_i(x_i+\theta+2)(\alpha\theta+\theta+1)^{x_i+2}-(\theta+1)^{x_i+2}x_i(\alpha\theta+\theta+x_i+2)}\\ &-\frac{\left(\theta x_i(x_i+2)(x_i+\theta+2)(\alpha\theta+\theta+1)^{x_i+1}-\theta(\theta+1)^{x_i+2}x_i\right)^2}{\left(x_i(x_i+\theta+2)(\alpha\theta+\theta+1)^{x_i+2}-x_i(\theta+1)^{x_i+2}(\alpha\theta+\theta+x_i+2)\right)^2}\Big]\\ &+\frac{\left(\sum_{i=1}^{n}x_i+2n\right)\theta^2}{(\alpha\theta+\theta+1)^2}.\end{aligned}$$

As $n\to\infty$, the distribution of $\sqrt{n}(\hat{\theta}-\theta,\hat{\alpha}-\alpha)$ is asymptotically bivariate normal with zero means and variance covariance matrix (see Lehmann and Casella [14]), which it can be approximated as $J^{-1}(\theta,\alpha)$.

4 Application

The MPL distribution is fitted with a real data set to compare with the competitive distributions. The sized-biased Poisson (SBP) distribution [3] and the size-biased Poisson Lindley (SBPL) distribution [4] are employed as the compared distributions. The criteria for selected model are the Anderson-Darling (AD) goodness of fit test for discrete distributions [15], the Kolmogorov-Smirnov (KS) goodness of fit test for discrete distributions [15], the negative log-likelihood (-LL) and the Akaike Information Criterion (AIC).

In this paper, we consider a data set of groups of pedestrians on a spring afternoon in Portland, Oregon [16]. The mean and variance of this data set are 1.5118 and 0.5605, respectively.

Table 1: Groups of pedestrians on a spring afternoon in Portland, Oregon

Size of groups	Observed frequencies	Expected frequencies		
		SBP	SBPL	MPL
1	1486	1452.502	1532.037	1508.301
2	694	743.273	630.762	662.991
3	195	190.173	192.078	192.809
4	37	32.438	51.417	46.426
5	10	4.150	12.786	10.006
6	1	0.425	3.029	2.003
Estimated parameters (Standard error)		$\hat{\lambda}=0.512$ (0.015)	$\hat{\theta}=4.505$ (0.131)	$\hat{\theta}=6.466$ (1.279) $\hat{\alpha}=0.003$ (0.394)
-LL		2308.697	2311.183	2305.705
AIC		4619.393	4624.365	4615.409
AD statistic		1.325	2.236	0.557
p-value		0.144	0.044	0.419
KS statistic		0.014	0.019	0.009
p-value		0.743	0.346	0.986

Table 1 shows that the p-value based on the AD test for discrete distributions of the SBPL distribution is less than the 5% significance level; thus, this data set should not be fitted with the SBPL distribution. The MPL distribution has the largest p-value based on the AD test and KS test for discrete distributions and also has the smallest negative log-likelihood and AIC.

The expected values of the MPL, SBPL, SBP distributions and the observed values are showed in Figure 3. It shows that the expected values of the MPL distribution are nearest the observed values.

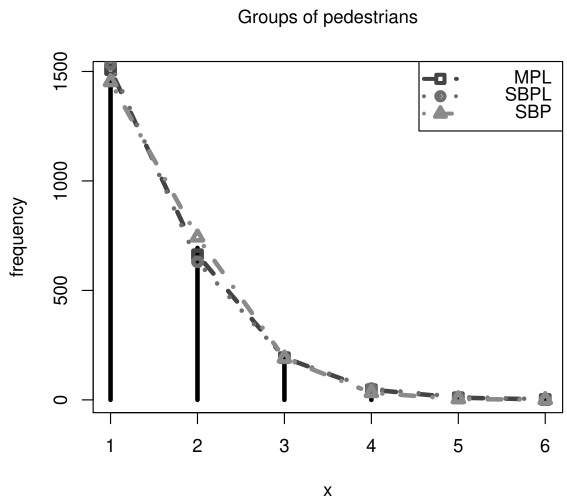

Figure 3: Plots of fitting distributions with groups of pedestrians

In general, this type of data is described by the SBP distribution. Now, we apply the proposed distribution to this data set. The results from Table 1 and Figure 3 show that the proposed distribution provides a better fit than the compared distributions.

5 Conclusions

In this paper, we have proposed a new mixture size-biased distribution, namely modified Poisson Lindley (MPL) distribution. The cummulative distribution function, probability generating function, moment generating function and moments have been studied. Parameter estimation has been derived by maximum likelihood estimation. Finally, a real data set was analyzed to show the performance of the MPL distribution. The results indicate that the MPL distribution is more efficient than the SBP the SBPL distributions.

Acknowledgement(s) : The authors would like to thank Department of Statistics, Faculty of Science, Kasetsart University. We would also like to thank Rajamangala University of Technology Phra Nakhon for supporting the first author.

References

[1] R. A. Fisher, The effect of methods of ascertainment upon the estimation of frequencies, Annals of eugenics, 6 (1934) 13-25.

[2] C. R. Rao, On discrete distributions arising out of methods of ascertainment, Sankhyā: The Indian Journal of Statistics, Series A, (1965) 311-324.

[3] G. P. Patil and C. R. Rao, Weighted distributions and size-biased sampling with applications to wildlife populations and human families, Biometrics, (1978) 179-189.

[4] M. E. Ghitany and D. K. Al-Mutairi, size-biased Poisson-Lindley distribution and its application, METRON-International Journal of Statistics, 66 (2008) 299-311.

[5] R. Shanker and H. Fesshaye, A size-biased Poisson-Amarendra distribution and its applications, International Journal of Statistics and Applications, 6 (2016) 376-385.

[6] R. Shanker and H. Fesshaye, Size-biased Poisson-Sujatha distribution with applications, American Journal of Mathematics and Statistics, 6 (2016) 145-154.

[7] R. Shanker, A size-biased Poisson-Shanker distribution and its applications, International Journal of Probability and Statistics, 6 (2017) 33-44.

[8] R. Shanker, Size-biased Poisson-Akash distribution and its applications, International Journal of Statistics and Applications, 7 (2017) 289-297.

[9] R. Shanker and A. Mishra, On size-biased two parameter Poisson-Lindley distribution and its applications, American Journal of Mathematics and Statistics, 7 (2017) 99-107.

[10] A. Asgharzadeh, H. S. Bakouch, S. Nadarajah, and F. Sharafi, A new weighted Lindley distribution with application, Brazilian Journal of Probability and Statistics, 30 (2016) 1-27.

[11] N. L. Johnson, A. W. Kemp and S. Kotz, Univariate discrete distributions, 3rd ed., John Wiley & Sons, New Jersey (2005).

[12] J. C. Nash and R. Varadhan, Unifying optimization algorithms to aid software system users: optimx for R, Journal of Statistical Software, 43 (2011) 1-14.

[13] R Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Austria, 2018, URL https://www.R-project.org/.

[14] L.E. Lehmann and G. Casella, Theory of Point Estimation, 2nd ed., Springer, New York (1998) 461-463.

[15] T. B. Arnold and J. W. Emerson, Nonparametric goodness-of-fit tests for discrete null distributions, R Journal. 3 (2011) 34-39.

[16] J. S. Coleman and J. James, The equilibrium size distribution of freely forming groups, Sociometry, 24 (1961) 36–45.