Songklanakarin J. Sci. Technol.
42 (1), 152-162, Jan. - Feb. 2020

http://www.sjst.psu.ac.th

*Original Article*

# The new Poisson mixed weighted Lindley distribution with applications to insurance claims data

Yupapin Atikankul, Ampai Thongteeraparp, and Winai Bodhisuwan*

*Department of Statistics, Faculty of Science,*
*Kasetsart University, Chatuchak, Bangkok, 10900 Thailand*

Received: 8 June 2017; Revised: 6 October 2018; Accepted: 17 October 2018

## Abstract

Mixed Poisson distributions have been applied for overdispersed count data analysis. In this paper, an alternative mixed Poisson distribution is proposed. The proposed distribution is derived by mixing the Poisson distribution with the new weighted Lindley distribution, named as the new Poisson mixed weighted Lindley distribution. Some special cases and several statistical properties have been derived including shape, factorial moments, probability generating function, moment generating function and moments. The maximum likelihood estimators of the parameters are obtained. Finally, two automobile insurance claims data sets are analyzed to compare the performance of the proposed distribution with some competitive distributions.

**Keywords:** count data, overdispersion, new weighted Lindley distribution, mixed Poisson distribution, maximum likelihood estimation

## 1. Introduction

Count data models are utilized in various fields such as public health, insurance, and agriculture. Some applications of count data have been studied, for example, the daily number of seizures of patients with epilepsy (Albert, 1991), the number of automobile insurance claims in Germany (Tröbliger, 1961), and the number of roots of types of apple rootstock (Ridout, Demétrio, & Hinde, 1998). The Poisson distribution is a classic distribution for describing count data with the property of equality of variance and mean, i.e. equidispersion. Unfortunately, in practical count data, the variance is usually larger than the mean which is referred to as overdispersion, and rarely, the variance may be smaller than the mean which is referred to as underdispersion. The Poisson distribution cannot be applied to account for these phenomena.

Overdispersion occurs in almost all count data. An approach that is widely applied to deal with overdispersed count data is mixed Poisson distributions (Grandell, 1997; Gupta & Ong, 2005; Karlis & Xekalaki, 2005;). The best known distribution of the Poisson type is the negative binomial distribution (Greenwood & Yule, 1920) which arises from a mixture of Poisson and gamma distributions.

The Lindley distribution (Lindley, 1958) arises from mixing the exponential distribution with the gamma distribution. Ghitany, Atieh, and Nadarajah (2008) applied the Lindley distribution to a data set of waiting times for bank customers. The result showed that the Lindley distribution provided a better fit than the exponential distribution. Thus in various papers, the Lindley distribution and some of its modifications have been studied and developed as mixing distributions for a mixed Poisson distribution. Sankaran (1970) proposed the discrete Poisson-Lindley (PL) distribution, which arises from the Poisson distribution and the Lindley mixing distribution (Lindley, 1958). Mahmoudi and Zakerzadeh (2010) introduced a mixture of the Poisson and generalized Lindley distributions (Zakerzadeh & Dolati, 2009). Shanker, Sharma, and Shanker (2012) proposed a mixed Poisson distribution with the two-parameter Lindley mixing distribution (Shanker, Sharma, & Shanker, 2013). Shanker and Mishra (2014) introduced the two-parameter Lindley distribution (Shanker & Mishra, 2013) as a mixing distribution. In the same year, the Poisson-weighted Lindley (PWL) distribution was proposed by El-Monsef and Sohsah (2014) and Manesh, Hamzah, and Zamani (2014). Wongrin

*Corresponding author
Email address: fsciwnb@ku.ac.th

and Bodhisuwan (2016) proposed a mixture of the Poisson and new generalized Lindley distributions (Elbatal, Merovci, & Elgarhy, 2013).

The new weighted Lindley (NWL) distribution was proposed by Asgharzadeh, Bakouch, Nadarajah, and Sharafi (2016). The NWL distribution is obtained by mixing the weighted exponential distribution (Gupta & Kundu, 2009) with the weighted gamma distribution. The probability density function (pdf) is

$$g(x) = \frac{\theta^2(1+\alpha)^2}{\alpha\theta(1+\alpha)+\alpha(2+\alpha)}(1+x)(1-e^{-\theta\alpha x})e^{-\theta x}, \quad (1)$$

for $x>0, \theta>0$ and $\alpha>0$.

The Lindley and weighted Lindley distributions are special cases of the NWL distribution. The pdf is log-concave and unimodal. A data set of the amount of carbon in leaves from the different mountainous areas of Navarra, Spain was modeled with the NWL distribution. The result showed that the NWL distribution is a better fit than the compared distributions.

In this paper, the new Poisson mixed weighted Lindley (NPWL) distribution is proposed which is a mixture of the Poisson and the NWL distributions. The rest of this paper is arranged as follows. In Section 2, the NPWL distribution is introduced. Some important statistical properties such as shape, factorial moments, probability generating function, moment generating function and moments are exhibited in Section 3. In Section 4, the steps for random variate generation are presented. In Section 5, the maximum likelihood estimators of the parameters are discussed. In Section 6, the NPWL distribution is applied to some automobile insurance claims data sets. The conclusions are presented in Section 7.

## 2. The New Poisson Mixed Weighted Lindley Distribution

The proposed distribution is a mixture of the Poisson distribution and the NWL distribution (Asgharzadeh, Bakouch, Nadarajah, & Sharafi, 2016). Moreover; some special cases, the distribution function and the survival function of the proposed distribution are shown in this section.

Let $X \mid \lambda$ be a Poisson random variable with probability mass function (pmf)

$$f(x \mid \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad (2)$$

for $x = 0, 1, 2, \ldots$ and $\lambda>0$, written as $X \mid \lambda \sim$ Pois $(\lambda)$. Now we assume that $\lambda$ is distributed as the NWL distribution with pdf

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

for $\lambda>0$, $\theta>0$ and $\alpha>0$, written as $\lambda \sim$ NWL $(\theta,\alpha)$; then the unconditional discrete random variable *X* follows the NPWL distribution.

**Proposition 1.** A discrete random variable *X* is distributed as the NPWL distribution with parameters $\theta$ and $\alpha$, denoted as $X \sim$ NPWL $(\theta,\alpha)$, the pmf of *X* is

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

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

**Proof.** Let $X \mid \lambda \sim$ Pois $(\lambda)$ and $\lambda \sim$ NWL $(\theta,\alpha)$, the pmf of *X* is derived as the following

$$f(x) = \int_0^\infty f(x \mid \lambda)g(\lambda)d\lambda$$

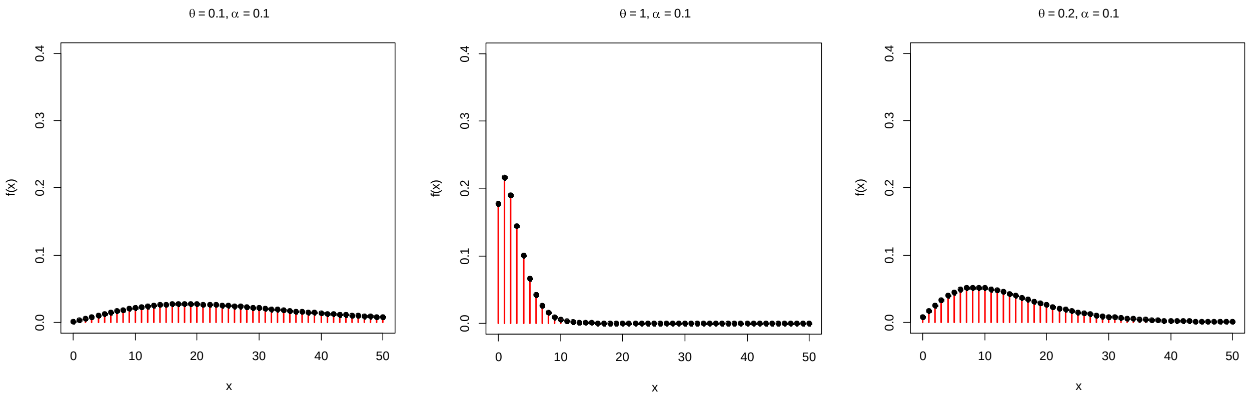
$$= \int_0^\infty \frac{e^{-\lambda}\lambda^x}{x!}\frac{\theta^2(1+\alpha)^2}{\left(\alpha\theta(1+\alpha)+\alpha(2+\alpha)\right)}(1+\lambda)(1-e^{-\theta\alpha\lambda})e^{-\theta\lambda}d\lambda$$

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

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

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

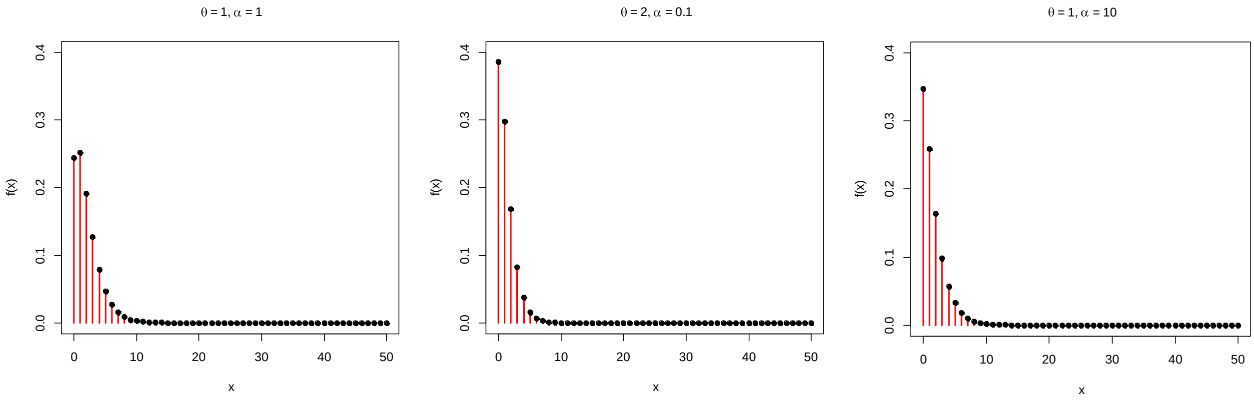
$$= \frac{\theta^2(1+\alpha)^2}{\alpha\theta(1+\alpha)+\alpha(2+\alpha)}\left(\frac{\theta+x+2}{(\theta+1)^{x+2}}-\frac{\alpha\theta+\theta+x+2}{(\alpha\theta+\theta+1)^{x+2}}\right).$$




(a)
ID = 10.735, skewness = 1.165,
kurtosis = 5.034

(b)
ID = 2.039, skewness = 1.33,
kurtosis = 5.517

(c)
ID = 5.932, skewness = 1.177,
kurtosis = 5.064




(d)
ID = 1.95, skewness = 1.48,
kurtosis = 6.193

(e)
ID = 1.527, skewness = 1.512,
kurtosis = 6.122

(f)
ID = 2.102, skewness = 1.744,
kurtosis = 7.347

Figure 1. Some probability mass function plots of the NPWL distribution with different values of $\theta$ and $\alpha$.

**Special cases**

$(i)$ If $\alpha \to \infty$, the NPWL distribution becomes the PL distribution (Sankaran, 1970).

$(ii)$ If $\alpha \to 0$, the NPWL distribution becomes the PWL distribution (El-Monsef & Sohsah, 2014; Manesh, Hamzah, & Zamani, 2014) with $c = 2$.

The cumulative distribution and survival functions of the NPWL distribution are given by

$$F(x) = P(X \le x)$$

$$= 1 - \left[ \frac{\theta^2(1+\alpha)^2}{\alpha\theta(1+\alpha)+\alpha(2+\alpha)} \left( \frac{x+2}{(\theta+1)^{x+3}} \, {}_2F_1\left(1, x+3; x+2; \frac{1}{\theta+1}\right) + \frac{1}{(\theta+1)^{x+2}} \, {}_2F_1\left(1, x+2; x+2; \frac{1}{\theta+1}\right) \right.\right.$$

$$\left.\left. - \frac{x+2}{(\alpha\theta+\theta+1)^{x+3}} \, {}_2F_1\left(1, x+3; x+2; \frac{1}{\alpha\theta+\theta+1}\right) - \frac{1}{(\alpha\theta+\theta+1)^{x+2}} \, {}_2F_1\left(1, x+2; x+2; \frac{1}{\alpha\theta+\theta+1}\right) \right) \right], \quad (4)$$

$$S(x) = P(X > x) = 1 - F(x),$$

where ${}_2F_1(a,b;c;z) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!}$, is the hypergeometric function; and $(a)_n = \frac{\Gamma(a+n)}{\Gamma(a)} = a(a+1)\ldots(a+n-1)$, is Pochhammer's symbol (Johnson, Kemp, & Kotz, 2005).

## 3. Properties

This section presents some statistical properties such as shape, factorial moments, probability generating function, moment generating function and moments of the NPWL ($\theta$, $\alpha$) distribution.

### 3.1 Shape

Holgate's theorem states that if $g(\lambda)$ is the pdf of mixing distribution, which is unimodal and absolutely continuous distribution, then the pmf of the mixed Poisson distribution is unimodal (Holgate, 1970). According to this theorem, the pmf of NPWL distribution is unimodal because the pdf of NWL distribution is unimodal (Asgharzadeh, Bakouch, Nadarajah, & Sharafi, 2016).

The NPWL distribution is log-concave $\left( \frac{f(x+2)f(x)}{\left(f(x+1)\right)^2} < 1 \right)$, therefore it has an increasing failure rate (Johnson, Kemp, & Kotz, 2005).

### 3.2 Factorial moments

**Proposition 2.** Let $X \sim$ NPWL $(\theta, \alpha)$, then the factorial moments of $X$ are

$$\mu'_{[k]} = \frac{\theta^2(1+\alpha)^2}{\alpha\theta(1+\alpha)+\alpha(2+\alpha)} \left( \frac{\Gamma(k+2)}{\theta^{k+2}} + \frac{\Gamma(k+1)}{\theta^{x+1}} - \frac{\Gamma(k+2)}{(\alpha\theta+\theta)^{k+2}} - \frac{\Gamma(k+1)}{(\alpha\theta+\theta)^{k+1}} \right).$$

**Proof.** The $k$th factorial moment of mixed Poisson distribution can be written in the form

$$\mu'_{[k]} = E\left[(X)_k\right] = \int_0^{\infty} \lambda^k g(\lambda) d\lambda.$$

If $X \sim$ NPWL $(\theta, \alpha)$, then the factorial moments of $X$ is given by

$$\mu'_{[k]} = \int_0^{\infty} \lambda^k \frac{\theta^2(1+\alpha)^2}{\alpha\theta(1+\alpha)+\alpha(2+\alpha)} (1+\lambda)(1-e^{-\theta\alpha\lambda}) e^{-\theta\lambda} d\lambda$$

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

$$= \frac{\theta^2(1+\alpha)^2}{\alpha\theta(1+\alpha)+\alpha(2+\alpha)} \left( \frac{\Gamma(k+2)}{\theta^{k+2}} + \frac{\Gamma(k+1)}{\theta^{x+1}} - \frac{\Gamma(k+2)}{(\alpha\theta+\theta)^{k+2}} - \frac{\Gamma(k+1)}{(\alpha\theta+\theta)^{k+1}} \right).$$

### 3.3 Probability generating function

**Proposition 3.** Let $X\sim$ NPWL $(\theta,\alpha)$, then the probability generating function (pgf) of *X* is

$$G(s) = \frac{\theta^2(1+\alpha)^2}{\alpha\theta(1+\alpha)+\alpha(2+\alpha)}\left(\frac{\theta-s+2}{(\theta-s+1)^2}-\frac{\alpha\theta+\theta-s+2}{(\alpha\theta+\theta-s+1)^2}\right), \quad s<\theta+1.$$

**Proof.** The pgf of mixed Poisson distribution can be defined as

$$G(s) = E(s^X) = \int_0^\infty e^{(s-1)\lambda} g(\lambda)d\lambda .$$

If $X\sim$ NPWL $(\theta,\alpha)$, then the pgf of *X* is obtained by

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

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

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

$$= \frac{\theta^2(1+\alpha)^2}{\alpha\theta(1+\alpha)+\alpha(2+\alpha)}\left(\frac{\theta-s+2}{(\theta-s+1)^2}-\frac{\alpha\theta+\theta-s+2}{(\alpha\theta+\theta-s+1)^2}\right), s<\theta+1.$$

### 3.4 Moment generating function

**Proposition 4.** Let $X\sim$ NPWL $(\theta,\alpha)$, then the moment generating function (mgf) of *X* is

$$M_X(t) = \frac{\theta^2(1+\alpha)^2}{\alpha\theta(1+\alpha)+\alpha(2+\alpha)}\left(\frac{\theta-e^t+2}{(\theta-e^t+1)^2}-\frac{\alpha\theta+\theta-e^t+2}{(\alpha\theta+\theta-e^t+1)^2}\right), t<\log(\theta+1).$$

The mgf is obtained trivially from the pgf as $M_X(t)=G(\mathrm{e}^t)$.

### 3.5 Moments

The *k*th raw moment is obtained by taking the *k*th derivative of mgf with respect to *t* and setting *t* to zero.

Let $\delta_1 = \frac{\theta^2(1+\alpha)^2}{\alpha\theta(1+\alpha)+\alpha(2+\alpha)}$, $\delta_2 = \alpha\theta+\theta$ and $\delta_3 = \frac{\theta+2}{\theta^3}-\frac{\alpha\theta+\theta+2}{(\alpha\theta+\theta)^3}$, if $X\sim$ NPWL $(\theta,\alpha)$ then the first four raw moments of *X* are

$$\mu_1' = \delta_1\delta_3,$$

$$\mu_2' = \delta_1\left(\frac{\theta^2+4\theta+6}{\theta^4}-\frac{\delta_2^2+4\delta_2+6}{\delta_2^4}\right),$$

$$\mu_3' = \delta_1\left(\frac{\theta^3+8\theta^2+24\theta+24}{\theta^5}-\frac{\delta_2^3+8\delta_2^2+24\delta_2+24}{\delta_2^5}\right),$$

$$\mu_4' = \delta_1\left(\frac{\theta^4+16\theta^3+78\theta^2+168\theta+120}{\theta^6}-\frac{\delta_2^4+16\delta_2^3+78\delta_2^2+168\delta_2+120}{\delta_2^6}\right).$$

The central moments of $X$ can be written in term of raw moments as

$$\mu_k = E(X-\mu)^k = \sum_{j=0}^{k}\binom{k}{j}(-1)^{k-j}\mu_j'\mu^{k-j}.$$

Thus, the first four central moments of $X$ are

$$\mu_1 = 0,$$

$$\mu_2 = \delta_1\left(\frac{\theta^2+4\theta+6}{\theta^4}-\frac{\delta_2^2+4\delta_2+6}{\delta_2^4}-\delta_1\delta_3^2\right),$$

$$\mu_3 = \delta_1\left(\frac{\theta^3+8\theta^2+24\theta+24}{\theta^5}-\frac{\delta_2^3+8\delta_2^2+24\delta_2+24}{\delta_2^5}\right)+\delta_3\delta_1^2\left(-\frac{3(\theta^2+4\theta+6)}{\theta^4}+\frac{3(\delta_2^2+4\delta_2+6)}{\delta_2^4}+2\delta_3^2\delta_1\right),$$

$$\mu_4 = \delta_1\left(\frac{\theta^4+16\theta^3+78\theta^2+168\theta+120}{\theta^6}-\frac{\delta_2^4+16\delta_2^3+78\delta_2^2+168\delta_2+120}{\delta_2^6}\right)$$
$$-4\delta_1^2\delta_3\left(\frac{\theta^3+8\theta^2+24\theta+24}{\theta^5}-\frac{\delta_2^3+8\delta_2^2+24\delta_2+24}{\delta_2^5}\right)+6\delta_3^2\delta_1^3\left(\frac{\theta^2+4\theta+6}{\theta^4}-\frac{\delta_2^2+4\delta_2+6}{\delta_2^4}\right)-3(\delta_1\delta_3)^4.$$

The skewness, kurtosis, and index of dispersion (ID) of $X \sim$ NPWL $(\theta,\alpha)$ are

$$\text{skewness}(X) = \frac{\mu_3}{\mu_2^{3/2}}, \quad \text{kurtosis}(X) = \frac{\mu_4}{\mu_2^2},$$

$$\text{ID}(X) = \frac{V(X)}{E(X)} = \frac{\mu_2}{\mu_1'} = \frac{-\theta^4\delta_3^2\delta_1\delta_2^4-\theta^4\delta_2^2-4\theta^4\delta_2-6\theta^4+\theta^2\delta_2^4+4\theta\delta_2^4+6\delta_2^4}{\theta^4\delta_3\delta_2^4}.$$

The mean, ID, skewness and kurtosis plots of the NPWL distribution are shown in Figure 2 which illustrates that the mean decreases as $\theta$ and $\alpha$ increase. For fixed $\theta$, as $\alpha$ increases, the ID increases but the ID decreases as $\theta$ increases for a fixed $\alpha$. Moreover, we can see that the ID is greater than one, thus the NPWL distribution is overdispersed, while the skewness and kurtosis increase as $\theta$ and $\alpha$ increase.

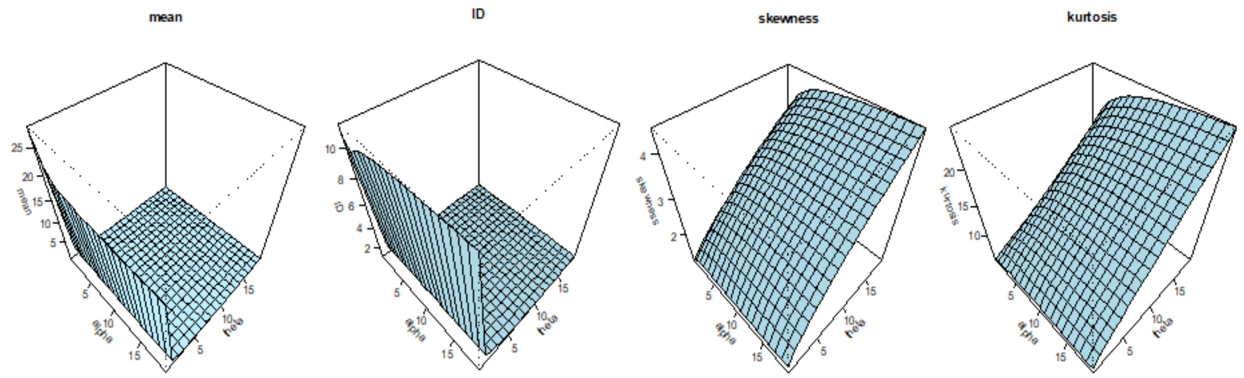



Figure 2. Mean, ID, and skewness and kurtosis plots of the NPWL distribution.

## 4. Random Variate Generation

In this section, a NPWL random variate generation is indicated. Let $X \sim$ Pois $(\lambda)$ and $\lambda \sim$ NWL $(\theta,\alpha)$, then random variables from the NPWL $(\theta,\alpha)$ distribution can be generated by the following algorithm.

1. Generate $\lambda_i$ using the rnwlindley function in the LindleyR package (Mazucheli, Fernandes, & de Oliveira, 2016) in the R programming language (R Core Team, 2018).
2. Generate $X_i$ from Pois $(\lambda_i)$, i = 1, 2, …, n.

## 5. Parameter Estimation

In this section, the parameter estimates of the NPWL distribution are obtained using maximum likelihood.

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed as NPWL distribution with $\Theta = (\theta,\alpha)^T$, the parameter vector. Then the likelihood function from (3) is

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

The associated log-likelihood function can be expressed as:

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

The score functions are found to be

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

and

$$\begin{aligned} \frac{\partial l(\Theta)}{\partial\alpha} = {} & \frac{2n}{\alpha+1} - \frac{n\big(\theta(1+2\alpha)+2(1+\alpha)\big)}{\alpha\big(\theta(1+\alpha)+2+\alpha\big)} - \theta\frac{\sum_{i=1}^{n} x_i + 2n}{\alpha\theta+\theta+1} \\ & + \sum_{i=1}^{n}\frac{\theta(x_i+2)(\theta+x_i+2)(\alpha\theta+\theta+1)^{x_i+1} - \theta(\theta+1)^{x_i+2}}{(\theta+x_i+2)(\alpha\theta+\theta+1)^{x_i+2} - (\alpha\theta+\theta+x_i+2)(\theta+1)^{x_i+2}}. \end{aligned}$$

The score functions are set equal to zero in order to obtain the parameter estimates of the NPWL distribution. Although these equations are non-linear, the maximum likelihood estimates can be solved by the numerical methods. In this paper, the method of moment estimators are given as the initial values for the BFGS method in the optimx function of the optimx package (Nash & Varadhan, 2011) in the R programming language (R Core Team, 2018).

## 6. Applications

We consider the application of the NPWL distribution to two automobile insurance claim data sets. These data sets are overdispersed count data with excess zeros. The competitive distributions are as follows:

$(i)$ The Poisson distribution (Poisson, 1837), with pmf

$$f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \quad \lambda > 0$$

$(ii)$ The PL distribution (Sankaran, 1970), with pmf

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

$(iii)$ The PWL distribution (El-Monsef & Sohsah, 2014; Manesh, Hamzah, & Zamani, 2014), with pmf

$$f(x) = \frac{\Gamma(x+c)}{x!\Gamma(c)}\frac{\theta^{1+c}}{(\theta+c)}\frac{(1+\theta+x+c)}{(\theta+1)^{1+x+c}}, \quad c > 0 \text{ and } \theta > 0.$$

The mixed Poisson distributions have proportions of zeros higher than the Poisson distribution. Thus, they have been applied to overdispersed and zero-inflated data. Puig (2006) proposed the zero-inflation index (*zi*), which is a measure to detect zero inflation. If $X$ is a non-negative integer random variable with mean and proportion of zeros are $\mu$ and $p_0$, respectively. The *zi* is

$$zi(X) = 1+\frac{\log(p_0)}{\mu}. \tag{5}$$

If $X$ is a Poisson random variable, the *zi* is equal to 0 but if $X$ is zero-inflated, then the *zi* is greater than 0. Figure 3 shows the *zi* versus the index of dispersion for the PL, PWL, and NPWL distributions. The *zi* values of the three distributions are similar. That is, the *zi* increases as the index of dispersion increases.

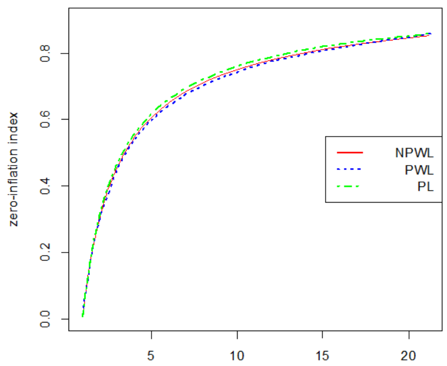



Figure 3. Zero-inflation index versus index of dispersion.

In this paper, the distribution with minimum the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), the negative log-likelihood (-LL), and the maximum P-value based on the Anderson-Darling (AD) goodness of fit test for a discrete distribution (Choulakian, Lockhart, & Stephens, 1994) is recommended as the most appropriate distribution for fitting overdispersed count data. Moreover, the likelihood ratio (LR) test is employed to compare special cases of the NPWL distribution.

**Application 1.** Lemaire (1985) presented the number of claims from third-party automobile liability in Belgium. This data set is overdispersed with a mean of 0.101, variance of 0.107, ID of 1.063 and the *zi* of 0.029.

Table 1 illustrates that the P-values based on the AD test for the PL and Poisson distributions are less than the 5% significance level; hence, this data set cannot be described by the PL and Poisson distributions. The means of all distributions are equal to the mean of the data set but the variance, ID and *zi* of the PWL and NPWL distributions are nearest the variance, ID and *zi* of the data set. However, the NPWL distribution provides the lowest -LL, AIC, BIC, and the highest P-value based on the AD test for a discrete distribution.

The LR statistic for testing $H_0$ : PWL vs. $H_1$ : NPWL is 0.36 with a P-value 0.549. Thus, there is no statistically significant difference between the PWL and NPWL distributions.

**Application 2.** Tröbliger (1961) as cited in Klugman, Panjer and Willmot (2012) presented the count of motor vehicle insurance claims per policy in Germany during 1960. It had a mean of 0.144, variance of 0.164, ID of 1.136 and the *zi* of 0.058; therefore, this data set is overdispersed.

Table 2 shows that the Poisson distribution also cannot describe this data set because its P-value based on the AD test for a discrete distribution is less than the 5% significance level. The mean of all distributions are also equal to the mean of the data set. The variance and ID of the PL distribution and the data set are equal but the *zi* of the PWL distribution is closest the *zi* of the data set. The NPWL

Table 1. Distribution of number of claims from third-party automobile liability in Belgium.

| Number of claims | Observed frequencies | Expected frequencies | | | |
|---|---|---|---|---|---|
| | | Poisson | PL | PWL | NPWL |
| 0 | 96978 | 96689.53 | 97147.15 | 96980.98 | 96981.22 |
| 1 | 9240 | 9773.440 | 8928.854 | 9230.643 | 9231.137 |
| 2 | 704 | 493.953 | 816.306 | 708.754 | 706.911 |
| 3 | 43 | 16.643 | 74.286 | 50.019 | 50.884 |
| 4 | 9 | 0.421 | 6.733 | 3.372 | 3.583 |
| 5 | 0 | 0.009 | 0.608 | 0.221 | 0.250 |
| Estimated parameters | | $\hat{\lambda}$ = 0.101 | $\hat{\theta}$ = 10.73 | $\hat{c}$ = 1.62 | $\hat{\theta}$ = 14.10 |
| | | | | $\hat{\theta}$ = 16.89 | $\hat{\alpha}$ =1.835 |
| Mean | | 0.101 | 0.101 | 0.101 | 0.101 |
| Variance | | 0.101 | 0.111 | 0.107 | 0.107 |
| ID | | 1 | 1.098 | 1.061 | 1.061 |
| *zi* | | 0 | 0.046 | 0.029 | 0.029 |
| -LL | | 36188.25 | 36122.53 | 36104.11 | 36103.93 |
| AIC | | 72378.51 | 72247.06 | 72212.22 | 72211.85 |
| BIC | | 72388.09 | 72256.64 | 72231.38 | 72231.01 |
| AD statistic | | 10.301 | 2.667 | 0.005 | 0.005 |
| P-value | | <0.01 | 0.022 | 0.991 | 0.992 |

Abbreviations: PL, Poisson-Lindley; PWL, Poisson-weighted Lindley; NPWL, new Poisson mixed weighted Lindley; ID, index of dispersion; *zi*, zero-inflation index; -LL, negative log-likelihood; AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion; AD, Anderson-Darling.

Table 2. Distribution of number of automobile insurance claims in Germany 1960.

| Number of claims | Observed frequencies | Expected frequencies | | | |
|---|---|---|---|---|---|
| | | Poisson | PL | PWL | NPWL |
| 0 | 20592 | 20420.94 | 20612.13 | 20596.48 | 20593.11 |
| 1 | 2651 | 2945.103 | 2604.386 | 2631.839 | 2637.904 |
| 2 | 297 | 212.371 | 326.210 | 318.143 | 315.640 |
| 3 | 41 | 10.209 | 40.562 | 37.588 | 37.367 |
| 4 | 7 | 0.368 | 5.013 | 4.380 | 4.397 |
| 5 | 0 | 0.011 | 0.616 | 0.505 | 0.515 |
| 6 | 1 | 0 | 0.075 | 0.058 | 0.060 |
| 7 | 0 | 0 | 0.009 | 0.007 | 0 |
| Estimated parameters | | $\hat{\lambda}$ = 0.144 | $\hat{\theta}$ = 7.728 | $\hat{c}$ = 1.103 | $\hat{\theta}$ = 8.145 |
| | | | | $\hat{\theta}$ = 8.453 | $\hat{\alpha}$ =14.24 |
| Mean | | 0.144 | 0.144 | 0.144 | 0.144 |
| Variance | | 0.144 | 0.164 | 0.162 | 0.161 |
| ID | | 1.00 | 1.136 | 1.123 | 1.121 |
| *zi* | | 0 | 0.063 | 0.057 | 0.056 |
| -LL | | 10297.84 | 10223.88 | 10223.49 | 10223.28 |
| AIC | | 20597.69 | 20449.76 | 20450.97 | 20450.56 |
| BIC | | 20605.75 | 20457.82 | 20467.11 | 20466.69 |
| AD statistic | | 10.565 | 0.198 | 0.051 | 0.035 |
| P-value | | <0.01 | 0.618 | 0.883 | 0.922 |

Abbreviations: PL, Poisson-Lindley; PWL, Poisson-weighted Lindley; NPWL, new Poisson mixed weighted Lindley ID, index of dispersion; *zi*, zero-inflation index; -LL, negative log-likelihood; AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion; AD, Anderson-Darling.

distribution also provides the lowest -LL and the highest P-value based on the AD test for a discrete distribution but the PL distribution provides the lowest AIC and BIC. However, the number of parameters of the PL distribution is less than the NPWL distribution.

In order to test $H_0$ : PL vs $H_1$ : NPWL, the LR statistic is 1.2 with a P-value 0.273. Hence, there is no statistically significant difference between the PL and NPWL distributions. The LR statistic for testing $H_0$ : PWL vs $H_1$ : NPWL is 0.42 with a P-value 0.517. There is also no statistically significant difference between the PWL and NPWL distributions.

The log expected values of the distributions and the log observed values are shown in Figure 4 which shows that the log expected values of the NPWL and PWL distributions are close to the observed values in both data sets. However, the log expected values of the Poisson distribution are far from observed values. Thus, the Poisson distribution cannot be applied to describe the overdispersed count data.

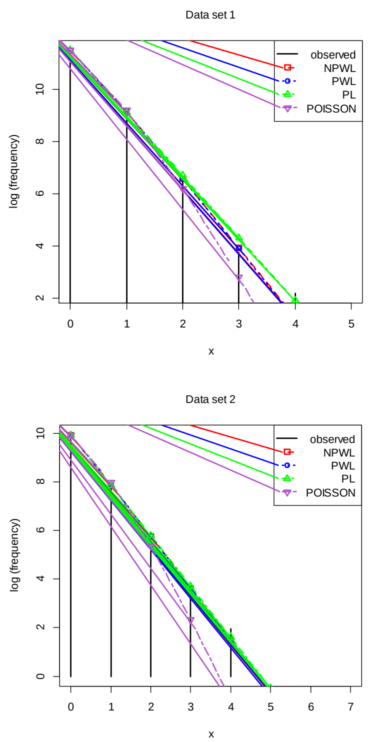



Figure 4. Log plots of the expected values and the observed values.

## 7. Conclusions

In this paper, an alternative mixed Poisson distribution, namely the NPWL distribution, is introduced. The NPWL distribution is derived from the Poisson distribution where the parameter follows the NWL distribution. The PL and PWL distributions are special cases of the NPWL distribution. Also, the pmf of NPWL distribution is log-concave and unimodal. Some statistical properties were studied such as shape, factorial moments, probability generating function, moment generating function and moments. Parameter estimation was derived by the maximum likelihood estimation. Moreover, two real overdispersed count data sets that were analyzed showed that the NPWL distribution provides a satisfactory fit in both data sets. Therefore, it is an alternative distribution for modeling overdispersed count data.

## Acknowledgements

The authors express their gratitude to the Department of Statistics, Faculty of Science, Kasetsart University and Rajamangala University of Technology Phra Nakhon. Further, the authors would like to thank two reviewers for valuable comments and suggestions which improved this paper.

## References


Albert, P. S. (1991). A two-state Markov mixture model for a time series of epileptic seizure counts. *Biometrics*, *47*(4), 1371-1381.

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

Arnold, T. B., & Emerson, J. W. (2011). Nonparametric goodness-of-fit tests for discrete null distributions. *R Journal*, *3*(2), 34-39.

Bhati, D., Sastry, D. V. S., & Qadri, P. M. (2015). A new generalized Poisson-Lindley distribution: Applications and Properties. *Austrian Journal of Statistics*, *44*(4), 35-51.

Choulakian, V., Lockhart, R. A., & Stephens, M. A. (1994). Cramér‐von Mises statistics for discrete distributions. *Canadian Journal of Statistics*, *22*(1), 125-137.

Elbatal, I., Merovci, F., & Elgarhy, M. (2013). A new generalized Lindley distribution. *Mathematical Theory and Modeling*, *3*(13), 30-47.

El-Monsef, M. A., & Sohsah N. M. (2014). Poisson–weighted Lindley distribution. *Jokull Journal*, *64*(5), 192-202.

Ghitany, M. E., Alqallaf, F., Al-Mutairi, D. K., & Husain, H. A. (2011). A two-parameter weighted Lindley distribution and its applications to survival data. *Mathematics and Computers in Simulation*, *81*(6), 1190-1201.

Ghitany, M. E., Atieh, B., & Nadarajah, S. (2008). Lindley distribution and its application. *Mathematics and Computers in Simulation*, *78*(4), 493-506.

Grandell, J. (1997). *Mixed poisson processes*. London, England: CRC Press.

Greenwood, M., & Yule, G. U. (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *Journal of the Royal Statistical Society*, *83*(2), 255-279.

Gupta, R. C., & Ong, S. H. (2005). Analysis of long-tailed count data by Poisson mixtures. *Communications in Statistics—Theory and Methods*, *34*(3), 557-573.

Gupta, R. D., & Kundu, D. (2009). A new class of weighted exponential distributions. *Statistics*, *43*(6), 621-634.

Holgate, P. (1970). The modality of some compound Poisson distributions. *Biometrika*, *57*(3), 666-667.

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

Karlis, D., & Xekalaki, E. (2005). Mixed Poisson distributions. *International Statistical Review*, *73*(1), 35-58.

Klugman, S. A., Panjer, H. H., & Willmot, G. E. (2012). *Loss models: from data to decisions* (Volume 715). New Jersey, NJ: John Wiley and Sons.

Lemaire, J. (1985). *Automobile insurance*. Berlin, Germany: Springer Science and Business Media. doi:10.1007/978-94-015-7708-3

Lindley, D. V. (1958). Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society. Series B (Methodological)*, *20*(1),102-107.

Low, Y. C., Ong, S. H., & Gupta, R. C. (2017). Generalized Sichel distribution and associated inference. *Journal of Statistical Theory and Applications*, *16*(3), 322-336.

Mahmoudi, E., & Zakerzadeh, H. (2010). Generalized Poisson–Lindley distribution. *Communications in Statistics—Theory and Methods*, *39*(10), 1785-1798.

Manesh, S. N., Hamzah, N. A., & Zamani, H. (2014). Poisson-weighted Lindley distribution and its application on insurance claim data. *AIP Conference Proceedings*, *1605*, 834-839. Penang, Malaysia: American Institute of Physics. doi:10.1063/1.4887698

Mazucheli, J., Fernandes, L. B., & de Oliveira, R. P. (2016). LindleyR: The Lindley distribution and its modifications. R package version 1.1.0. Retrieved from https://CRAN.R-project.org/package=LindleyR

Nash, J. C., & Varadhan, R. (2011). Unifying optimization algorithms to aid software system users: optimx for R. *Journal of Statistical Software*, *43*(9), 1-14.

Poisson, S. D. (1837). *Recherches sur la probabilité des jugements en matière criminelle et en matière civile precédées des règles générales du calcul des probabilités par SD Poisson*. Bachelier.

Puig, P., & Valero, J. (2006). Count data distributions: some characterizations with applications. *Journal of the American Statistical Association*, *101*(473), 332-340.

R Core Team (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Ridout, M., Demétrio, C. G., & Hinde, J. (1998). Models for count data with many zeros (Invited Paper). *In Proceedings of the XIXth International Biometric Conference* (Volume 19, No. 19, pp. 179-192). Cape Town, South Africa: International Biometric Society.

Sankaran, M. (1970). The discrete Poisson-Lindley distribution. *Biometrics*, *26*(1), 145-149.

Shanker, R., & Mishra, A. (2013). A two-parameter Lindley distribution. *Statistics in Transition New Series*, *1*(14), 45-56.

Shanker, R., & Mishra, A. (2014). A two-parameter Poisson-Lindley distribution. *International Journal of Statistics and Systems*, *9*(1), 79-85.

Shanker, R., Sharma, S., & Shanker, R. (2012). A discrete two-parameter Poisson Lindley distribution. *Journal of the Ethiopian Statistical Association*, *21,* 15-22.

Shanker, R., Sharma, S., & Shanker, R. (2013). A two-parameter Lindley distribution for modeling waiting and survival times data. *Applied Mathematics*, *4*(2), 363-368.

Tröbliger, A. (1961). Mathematische Untersuchungen zur Beitragsrückgewähr in der Kraftfahrversiche rung. *Blätter der DGVFM, 5*(3), 327-348.

Wongrin, W., & Bodhisuwan, W. (2016). The Poisson-generalised Lindley distribution and its applications. *Songklanakarin Journal of Science and Technology*, *38*(6), 645-656.

Zakerzadeh, H., & Dolati, A. (2009). Generalized Lindley Distribution. *Journal of Mathematical Extension*, *3*(2), 1-7.