Chapter 7

A Comparative Study on Zero-truncated Poisson-Lindley and Quasi Poisson-Lindley Distributions

7.1 Introduction

Truncation of a distribution occurs when a range of possible variate values either is ignored or is impossible to observe. In most common form of truncation, the zeroes are not recorded. In this case, the zero-truncated distributions can be used as a distribution for the sizes of groups. This situation occurs in applications such as the number of claims per claimant, the number of occupants per car etc.

When the data to be modeled originate from a generating mechanism that structurally excludes zero counts, the Poisson-Lindley distribution must be adjusted to count for the missing zeros. Ghitany, Al-Mutairi and Nadarajah (2008) obtained zero-truncated Poisson-Lindley (ZTPL) distribution to model count data by considering the zero-truncated form of Poisson-Lindley distribution as

$$f(x;\ \theta) = \frac{f_0(x;\theta)}{1-f_0(0;\theta)} = \frac{\theta^2}{\theta^2+3\theta+1}\frac{(x+\theta+2)}{(1+\theta)^x}\ ;\quad x = 1, 2, 3, \dots;\ \theta > 0. \qquad (7.1.1)$$

They also showed that, the ZTPL distribution is unimodal and has an increasing failure rate, i.e. since

$$\frac{f(x+1;\theta)}{f(x;\theta)} = \frac{1}{(1+\theta)}\left(1 + \frac{1}{x+\theta+2}\right) \tag{7.1.2}$$

is a decreasing function in x, the ZTPL distribution is unimodal. Also, since

$$\frac{f(x+2;\theta)f(x;\theta)}{f^2(x+1;\theta)} = \frac{1+(1/x+\theta+3)}{1+(1/x+\theta+2)} < 1 \tag{7.1.3}$$

$f(x;\ \theta)$ is log-concave and hence the ZTPL distribution has an increasing failure rate. [cf. Ghitany, Al-Mutairi and Nadarajah (2008)]

The mean and variance of ZTPL distribution obtained by Ghitany et al. (2008) are given by

$$\mu = \frac{(1+\theta)^2(\theta+2)}{\theta(\theta^2+3\theta+1)} \tag{7.1.4}$$

and

$$\sigma^2 = \frac{(1+\theta)^2(\theta^3+6\theta^2+10\theta+2)}{\theta^2(\theta^2+3\theta+1)^2} \tag{7.1.5}$$

In this chapter, we have introduced the zero-truncated quasi Poisson–Lindley (ZTQPL) distribution for analyzing different types of count data along with its applications. An attempt has been made to find its statistical properties and then compared them with zero-truncated Poisson-Lindley (ZTPL) distribution investigated by Ghitany et al. (2008). Finally, the introduced distribution is fitted to some reported data sets and the fit is compared with earlier fitted zero-truncated Poisson-Lindley distribution for empirical comparison. It is seen that, the ZTQPL distribution is more flexible than the ZTPL distribution for analyzing different types of count data.

7.2 Zero-truncated Quasi Poisson-Lindley distribution

The quasi Poisson-Lindley distribution is a two parameter discrete mixture distribution with pmf

$$P_1(x;\alpha,\theta)=\frac{\theta}{\alpha+1}\frac{[\alpha+\theta(1+\alpha)+\theta x]}{(1+\theta)^{x+2}}\,;\,x=0,1,2,\ldots,\theta>0,\alpha>-1 \qquad (7.2.1)$$

arises from the Poisson distribution when its parameter $\boldsymbol{\lambda}$ follows a quasi Lindley distribution of Shanker and Mishra (2013a) with probability density function (pdf)

$$f(x;\alpha,\theta)=\frac{\theta}{\alpha+1}(\alpha+\theta x)e^{-\theta x}\;,\quad x>0,\;\theta>0,\alpha>\;-1 \qquad (7.2.2)$$

It can be seen that, the Poisson-Lindley distribution of Sankaran (1970) already mentioned in chapter1 is a particular case of QPL distribution (7.2.1) at $\alpha=\theta$. The model (7.2.1) is a more generalized and more flexible than the Poisson-Lindley distribution. [See chapter 6]

The probability mass function (pmf) of the zero-truncated quasi Poisson-Lindley (ZTQPL) distribution with parameters α and θ is obtained by considering its zero-truncated form as follows

$$P_T(x;\;\alpha,\theta)=\frac{P_1(x;\,\alpha,\theta)}{1-p_1(0;\,\alpha,\theta)}=\frac{\theta}{[\alpha+\theta(\alpha+2)+1]}\frac{[\alpha+\theta(1+\alpha)+\theta x]}{(1+\theta)^{x}}\;;\;x=1,2,3,\ldots;\;\theta>0;\;\alpha>-1 \qquad (7.2.3)$$

where, $P_x\,(x;\,\theta,\alpha)$ be the probability mass function (pmf) and $P_1(0;\,\alpha,\,\theta)$ be the probability at $x=0$, of QPL distribution already discussed in chapter 6.While θ is the scale and α be the shape parameters of the distribution. Simply denote it by ZTQPL (α,θ). It is also seen that, zero-truncated Poisson-Lindley (ZTPL) distribution (7.1.1) is a particular case of ZTQPL distribution at $\alpha=\theta$.

Since,

$$\frac{P_T(x+1;\alpha,\theta)}{P_T(x;\alpha,\theta)} = \frac{1}{1+\theta}\left[1+\frac{\theta}{\{\alpha+\theta(1+\alpha)+\theta x\}}\right]$$

is a decreasing function in x, the ZTQPL distribution is unimodal. [cf. Johnson et al.2005)]

7.3 Graphical Representations

To study the behavior of the ZTQPL distribution for varying values of the parameters θ and α, the probabilities for possible values of x are computed by using the above equation (7.2.3) and different graphs may be drawn for various values of the two parameters.

Fig. 7.1 Plots of probability P(x; θ, α) for fixed $\alpha = -0.5$ and $\theta = 2, 4, 6$ respectively for the ZTQPL distribution.

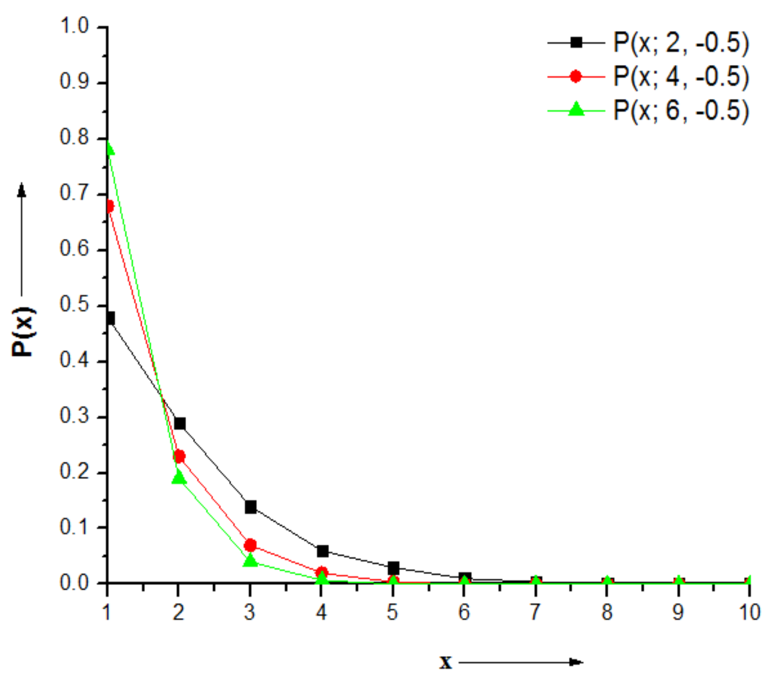


Fig. 7.2 Plots of probability P(x; θ, α) for fixed $\alpha = -0.8$ and $\theta = 2, 4, 6$ respectively for the ZTQPL distribution.

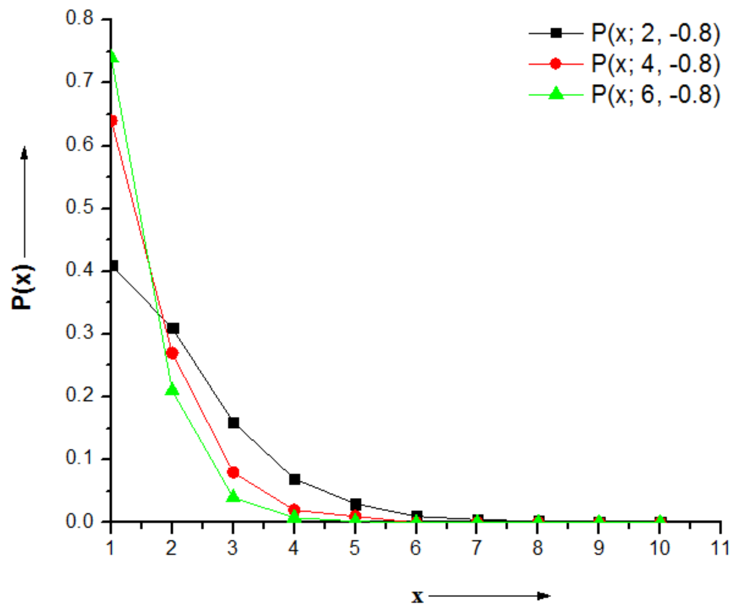


Fig. 7.3 Plots of probability P(x; θ, α) for fixed $\alpha = 1$ and $\theta = 2, 4, 6$ respectively for the ZTQPL distribution.

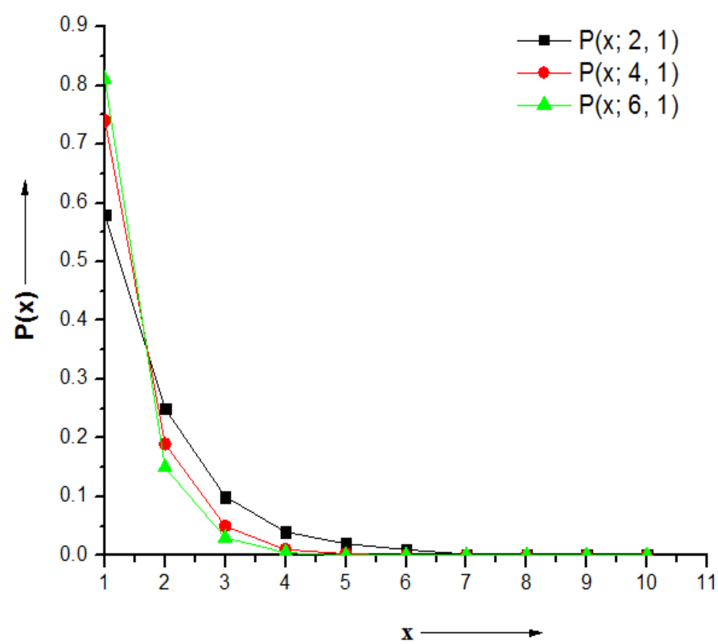
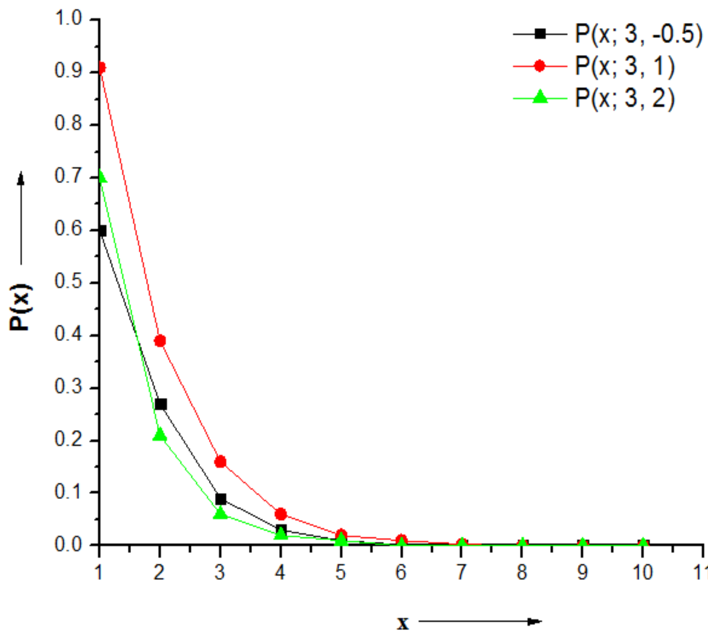


Fig. 7.4 Plots of probability P(x; θ, α) for fixed $\theta = 3$ and $\alpha = -0.5, 1, 2$ respectively for the ZTQPL distribution.



It is clear from the above **Fig. 7.1, 7.2**, and **7.3** that when the value of θ is large and the value of α increases slowly, then value of P(x; θ, α) is maximum at $x = 1$ and decreases sharply to a certain point then slowly decreases giving a L-shaped. Again, from **Fig. 7.4** it is observed that for fixed $\theta = 3$, when α takes values -0.5, 1, and 2 respectively, then the value of P(x; θ, α) is maximum at $x = 1$ and decreases for other increasing values of x also giving a L-shaped.

7.4 Statistical Properties

In this Section, some important statistical properties and its related measures for the ZTQPL distribution have been derived.

7.4.1 Probability generating function

If x follows ZTQPL (α, θ), then the probability generating function (pgf) of x is obtained as

$$g(t) = \sum_{x=1}^{\infty} t^x P_T(x;\ \alpha, \theta) \ = \frac{\theta}{\alpha+\theta(\alpha+2)+1} \sum_{x=1}^{\infty} \left(\frac{t}{1+\theta}\right)^x [\alpha + \theta(1+\alpha) + \theta x]$$

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

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

$$= \frac{\theta t[(1+\theta)(\theta+2\theta+\theta\alpha)-t(\alpha+\theta+\theta\alpha)]}{[\alpha+\theta(\alpha+2)++1](1+\theta-t)^2} \ ; t > 0 \qquad (7.4.1)$$

Note that, for $\alpha \rightarrow \theta$ this $g(t)$ reduces to the pgf of the ZTPL distribution of Ghitany et al. (2008) which is given as

$$g(t) = \frac{\theta^2 t[(1+\theta)(\theta+3)-(\theta+2)t]}{(\theta^2+3\theta+1)(1+\theta-t)^2} \ ;\ \theta > 0, t > 0. \qquad (7.4.2)$$

The recursive relation for probabilities of the ZTQPL distribution obtained from pgf (7.4.1) is given as

$$P_r = \frac{1}{(1+\theta)^2} [2(1+\theta)P_{r-1} - P_{r-2}] \ ;\ \ r > 2 \qquad (7.4.3)$$

where, $$P_1 = \frac{\theta(\alpha+2\theta+\theta\alpha)}{(1+\theta)[\alpha+\theta(\alpha+2)+1]}$$

$$P_2 = \frac{\theta(\alpha+3\theta+\theta\alpha)}{(1+\theta)^2[\alpha+\theta(\alpha+2)+1]}$$

If $\alpha \to \theta$, in the probabilities of the ZTQPL distribution then these probabilities are same as that of the ZTPL distribution of Ghitany et al. (2008) which are given below as

$$P_1 = \frac{\theta^2(\theta+3)}{(1+\theta)(\theta^2+3\theta+1)} \text{ and } P_2 = \frac{\theta^2(\theta+4)}{(1+\theta)^2(\theta^2+3\theta+1)} \text{ etc.}$$

The general expression for probability $P(X = r)$ of ZTQPL distribution is given by

$$P_r = \frac{\theta[\alpha+(r+1)\theta+\theta\alpha]}{(1+\theta)^r[\alpha+\theta(\alpha+2)+1]} \; ; r = 1, 2, 3, \dots \qquad (7.4.4)$$

The higher order probabilities can be obtained very easily by using either relation (7.4.3) or (7.4.4).

7.4.2 Moments and related measures

In this subsection, the r^{th}order factorial moment and moments about origin and central moments of the ZTQPL distribution can be obtained as follows:

The factorial moment generating function of the ZTQPL distribution is written as

$$m(t) = \frac{\theta(1+\theta)(1+t)(\alpha+2\theta+\theta\alpha)-\theta(1+t)^2(\alpha+\theta+\alpha\theta)}{(\theta-t)^2[\alpha+\theta(\alpha+2)+1]} \; ; \; \theta > 0, \alpha > -1 \qquad (7.4.5)$$

Note that, the factorial moment generating function of the ZTPL distribution is a particular form of factorial moment generating function of the ZTQPL distribution, i.e. if $\alpha \to \theta$ in (7.4.5) then it will be reduces to the factorial moment generating function of the ZTPL distribution, which is given by

$$m(t) = \frac{\theta^2(1+t)[(1+\theta)(\theta+3)-(\theta+2)(1+t)]}{(\theta-t)^2(\theta^2+3\theta+1)} \; ; \; \theta > 0 \qquad (7.4.6)$$

The factorial moment recursive relation of the ZTQPL distribution is given as

$$\mu'_{(r)} = \frac{1}{\theta^2}\left[2\theta r\mu'_{(r-1)} - r(r-1)\mu'_{(r-2)}\right];\ r > 2 \tag{7.4.7}$$

where, $\mu'_{(1)} = \frac{(1+\theta)^2(\alpha+2)}{\theta[\alpha+\theta(\alpha+2)+1]}$ (Mean)

$$\mu'_{(2)} = \frac{(1+\theta)^2 2(\alpha+3)}{\theta^2[\alpha+\theta(\alpha+2)+1]}$$

be the first two factorial moments of the ZTQPL distribution.

The general expression for r^{th} order factorial moment is given by

$$\mu'_{(r)} = \frac{(1+\theta)^2 r!(\alpha+r+1)}{\theta^r[\alpha+\theta(\alpha+2)+1]};\ r = 1, 2, 3, \ldots \tag{7.4.8}$$

After obtaining the first four factorial moments and then using the relationship between factorial moments and moments about origin, the first four moments about origin of the ZTQPL distribution are given as follows

$$\mu'_1 = \frac{(1+\theta)^2(\alpha+2)}{\theta[\alpha+\theta(\alpha+2)+1]} \tag{7.4.9}$$

$$\mu'_2 = \frac{(1+\theta)^2[(2+\alpha)(\theta+2)+2]}{\theta^2[\alpha+\theta(\alpha+2)+1]} \tag{7.4.10}$$

$$\mu'_3 = \frac{(1+\theta)^2[(2+\alpha)\theta^2+(6\alpha+18)\theta+(6\alpha+24)]}{\theta^3[\alpha+\theta(\alpha+2)+1]} \tag{7.4.11}$$

$$\mu'_4 = \frac{(1+\theta)^2[(2+\alpha)\theta^3+14(\alpha+3)\theta^2+36(\alpha+4)\theta+24(\alpha+5)]}{\theta^4[\alpha+\theta(\alpha+2)+1]} \tag{7.4.12}$$

Note that, if $\alpha \to \theta$ these raw moments will be same as that of the ZTPL distribution of Ghitany et al. (2008).

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

$$\mu = \frac{(1+\theta)^2(\alpha+2)}{\theta[\alpha+\theta(\alpha+2)+1]} \tag{7.4.13}$$

and
$$\sigma^2 = \frac{(1+\theta)^2[(1+\theta)(\alpha^2+4\alpha)+\theta\alpha+6\theta+2]}{\theta^2[\alpha+\theta(\alpha+2)+1]^2} \tag{7.4.14}$$

If $\alpha \to \theta$ in (7.4.13) and (7.4.14), then the mean and variance of the ZTQPL distribution will be same as that of the ZTPL distribution which is mentioned in above introduction section.

The index of dispersion for the ZTQPL distribution is given by

$$\gamma = \frac{\sigma^2}{\mu} = \frac{\alpha(1+\theta)(\alpha+4)+\theta\alpha+6\theta+2}{\theta(\alpha+2)[\alpha+\theta(\alpha+2)+1]} \tag{7.4.15}$$

it follows that the ZTQPL distribution is over-dispersed ($\sigma^2 > \mu$) for all values of (α, θ), and equi-dispersed ($\sigma^2 = \mu$) for large amount of θ. For $\alpha = \theta$, the index of dispersion of ZTQPL distribution reduces to index of dispersion of the ZTPL distribution [cf. Ghitany et al. (2008)], which is given by

$$\gamma = \frac{\theta^3+6\theta^2+10\theta+2}{\theta(\theta+2)(\theta^2+3\theta+1)} \tag{7.4.16}$$

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 CV of the ZTQPL distribution is given as

$$CV = \frac{(1+\theta)[\alpha(\alpha+4)+2]+\theta(\alpha+4)}{\theta(\alpha+2)[\alpha+\theta(\alpha+2)+1]}, \theta > 0, \alpha > -2 \tag{7.4.17}$$

If $\alpha \to \theta$, then it will be same as that of the ZTPL distribution. In the above expression as the values of θ and α increases, the CV for the values decreases slowly.

The distribution of a random variable is often characterized in terms of its moment generating function (mgf), a real function whose derivatives at zero are equal to the moments of the random variable. Moment generating functions have great practical relevance not only because they can be used to easily derive moments, but also because a probability distribution is uniquely determined by its mgf. The mgf of the ZTQPL distribution can be derived as

$$M_x(t) = \sum_{x=1}^{\infty} e^{tx}\, P_T(x;\ \alpha, \theta) == \frac{\theta e^t\left[(1+\theta)(\theta+2\theta+\theta\alpha)-e^t(\alpha+\theta+\theta\alpha)\right]}{[\alpha+\theta(\alpha+2)++1](1+\theta-e^t)^2} \qquad (7.4.18)$$

The ZTQPL distribution is more generalized and more flexible than the one parameter ZTPL distribution of Ghitany et al (2008) for analyzing different types of count data.

7.5 Estimation of Parameter

One of the most important property of a distribution is the problem of parameter estimation. In case of ZTPL distribution, the single parameter θ was estimated by using method of moment [cf. Ghitany et al. (2008)]. To estimate the parameters of the ZTQPL distribution we have been considered an ad-hoc method, i.e; estimate the two parameters in terms of ratio of the first two probabilities and mean of the distribution.

In case of ZTQPL distribution,

$$P_1 = \frac{\theta(\alpha+2\theta+\theta\alpha)}{(1+\theta)[\alpha+\theta(\alpha+2)+1]}$$

$$P_2 = \frac{\theta(\alpha+3\theta+\theta\alpha)}{(1+\theta)^2[\alpha+\theta(\alpha+2)+1]}$$

The ratio of these two probabilities gives

$$\frac{p_2}{p_1} = \frac{1}{(1+\theta)} \frac{(\alpha+3\theta+\theta\alpha)}{(\alpha+2\theta+\theta\alpha)}$$

$$\Rightarrow (1+\theta)\alpha = \left[\frac{\theta p_1}{(1+\theta)p_2 - p_1} - 2\theta\right] \qquad \text{[after calculation]} \qquad (7.5.1)$$

Again, the mean of the ZTQPL distribution gives

$$\mu = \frac{(1+\theta)^2(\alpha+2)}{\theta[\alpha+\theta(\alpha+2)+1]}$$

$$\Rightarrow (1+\theta)\alpha = \frac{2(1+\theta)^2 - \theta(1+2\theta)\mu}{[\theta\mu - (1+\theta)]} \qquad \text{[after calculation]} \qquad (7.5.2)$$

Now equating the both sides of (7.5.1) and (7.5.2), we have a quadratic equation in terms of θ as

$$A\theta^2 + B\theta + C = 0 \qquad (7.5.3)$$

where, $A = p_1(\mu - 1) + p_2(\mu - 2)$

$B = p_2(\mu - 4) + p_1(1 - \mu)$

$C = 2(p_1 - p_2)$

which gives the estimator of the parameter θ as

$$\hat{\theta} = \frac{-B + \sqrt{B^2 - 4AC}}{2A} \qquad (7.5.4)$$

After obtaining the estimator $\hat{\theta}$ of θ, the other parameter α can be estimated either from

$$\hat{\alpha} = \frac{1}{(1+\theta)}\left[\frac{\theta p_1}{(1+\theta)p_2 - p_1} - 2\theta\right] \qquad (7.5.5)$$

or,

$$\hat{\alpha} = \frac{2(1+\theta)^2 - \theta(1+2\theta)\mu}{(1+\theta)[\theta\mu - (1+\theta)]} \quad (7.5.6)$$

7.6 Applications and goodness of fit

In order to examine the flexibility and to see the applications of the ZTQPL distribution, two sets of reported data taken from Ghitany and Al-Mutairi (2008) have been considered in **Table 7.1** and **7.2**. The first data set represents the immunogold assay data of Cullen et al. (1990) for which the ZTPL distribution was fitted by Ghitany, Al-Mutairi and Nadarajah (2008). 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, 4. The second data set represents animal abundance data of Keith and Meslow (1968). 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.

We are interested in testing the null hypothesis H_0: "Number of attached particles and number of snowshoe hares is a ZTQPL random variable" verses the alternative hypothesis H_1: "Number of attached particle and number of snowshoe hares is not a ZTQPL random variable".

The expected frequencies of ZTQPL distribution along with the fitted ZTPL distribution estimated parameters and computed χ^2 and p −value are shown in **Table 7.1** and **Table 7.2**. It is clear from the tables that, the null hypothesis H_0 cannot be rejected; indeed, the close agreement between the observed and expected frequencies suggests that the ZTQPL distribution provides a "good fit" to the two data sets as compared to the zero-truncated Poisson-Lindley distribution fitted earlier.

Table 7.1 Observed vs. expected frequencies of the ZTQPL distribution for immunogold assay data. [data Cullen et al. (1990)]

Number of attached particles	Observed Frequency $\bar{x} = 1.576$	ZTPL Ghitany et al.(2008) $\hat{\theta} = 2.185$	ZTQPL $\hat{\theta} = 2.890$ $\hat{\alpha} = -0.242$
1	122	124.8	121.9
2	50	46.8	50.2
3	18	17.1	17.9
4	4	6.1	5.9
5	4	3.2	2.1
Total	198	198.0	198.0
χ^2		0.511	0.017
$d.f$		2	2
$p - value$		0.76	0.99

Note: ZTPL: Zero-truncated Poisson-Lindley distribution.

ZTQPL: Zero-truncated quasi Poisson-Lindley distribution.

Table 7.2 Observed vs. expected frequencies of the ZTQPL distribution for animal abundance data. [data Keith and Meslow (1968)]

Number of snowshoe hare	Observed frequency	ZTPL $\hat{\theta} = 3.101$	ZTQPL $\hat{\theta} = 2.606$ $\hat{\alpha} = 7.647$
1	184	187.4	183.1
2	55	53.5	54.8
3	14	14.8	16.4
4	4	4.2	4.9
5	4	1.1	1.8
Total	261	261.0	261.0
χ^2		1.522	0.608
$d.f$		2	2
$p - value$		0.47	0.74

Note: ZTPL: Zero-truncated Poisson-Lindley distribution.

ZTQPL: Zero-truncated quasi Poisson-Lindley distribution.

****