Chapter 4

Zero-modified Poisson-Lindley distribution

4.1 Introduction

To serve the probabilistic description of an experiment with a slight modification at point zero, the zero-modified Poisson-Lindley (ZMPL) distribution is studied in this chapter.

In the probability model the Poisson distribution is usually assumed for count data; however, in many real applications it is likely to observe that the number of zeroes is greater than what would be expected for the Poisson model, which is called zero-modification or inflation. The major motivation force behind the development of zero-modified distribution is that, many distributions obtained in the course of experimental investigations often have an excess frequency of the observed event at zero point.

A combination of the original distribution with probability mass function (pmf) $P_x\,; x = 0,1,2,\ldots$ together with the degenerate distribution with all probabilities concentrated at the origin gives a zero-modified distribution with pmf

$$P[X = 0] = w + (1 - w)P_0 \qquad (4.4.1)$$

$$P[X = x] = (1 - w)P_x \;;\; x \geq 1 \qquad (4.4.2)$$

where w is a parameter assuming arbitrary value in the interval $0 < w < 1$. It is also possible to take $w < 0$, provided $w + (1 - w)P_0 \geq 0$. [cf. Johnson et al. (2005)]

Recently, the zero-modified formulations are widely used in many fields because of the low frequency of the events. Zero-modified distribution is also called zero- inflated. Mechanisms producing zero-modified distribution have been discussed by Heilbron (1994) in the context of generalized linear model. Dutta and Borah (2014) studied and fitted the ZMPL distribution to some well-known data for empirical comparison.

The organization of the chapter as follows: Section 4.2 is concerned with the probability mass function, probability generating function and expression for probabilities of the ZMPL distribution. Section 4.3 deals with the graphical representation of the ZMPL distribution. Attempt has been made to obtain some distributional properties including expressions for factorial and raw moments, cumulants and coefficient of variation, measure of skewness and kurtosis of this distribution in section 4.4. In section 4.5, problem of parameter estimation is discussed. To see the application of the ZMPL distribution in different fields, fit this distribution to some well-known data sets and it is observed that the proposed distribution gives better fit in all the cases in section 4.6.

4.2 Probability mass function and Probability generating function

The probability mass function (pmf) of the ZMPL distribution is given by

$$P[X = 0] = w + (1 - w)P_0$$

$$= w + (1 - w)\frac{\theta^2(\theta+2)}{(1+\theta)^3} \;\; ; \; x = 0 \qquad (4.2.1)$$

$$P[X = x] = (1 - w)P_x = (1 - w)\frac{\theta^2(x+\theta+2)}{(1+\theta)^{x+3}}; \;\; x \geq 1 \qquad (4.2.2)$$

where $P_0 = \frac{\theta^2(\theta+2)}{(1+\theta)^3}$ be the zero-order probability and $P_x = \frac{\theta^2(x+\theta+2)}{(1+\theta)^{x+3}}$ be the pmf of

Poisson-Lindley distribution mentioned in chapter 2. θ and w be the two parameters of the distribution i.e. $\theta > 0$ and $0 < w < 1$. It is also possible to take the parameter $w < 0$, provided $w + (1-w)P_0 \geq 0$ where $P_0 = P(X = 0)$.

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

$$G(t) = w + (1-w)g(t) \qquad (4.2.3)$$

where, $g(t) = \frac{\theta^2(\theta+2-t)}{(1+\theta)(\theta+1-t)^2}$ is the pgf of the Poisson-Lindley distribution [cf. Sankaran(1970)].

Differentiating (4.2.3) w.r.to't' and then equating the coefficient of 't^r' on both sides of the equation, we have the expression for probabilities as

$$P_{r+1} = \frac{1}{(1+\theta)^2}[2(1+\theta)P_r - P_{r-1}]\,; \quad r > 1. \qquad (4.2.4)$$

where, $$P_0 = w + (1-w)\frac{\theta^2(\theta+2)}{(1+\theta)^3}$$

$$P_1 = \frac{(1-w)\theta^2(\theta+3)}{(1+\theta)^4}$$

After some suitable transformation of equation (4.2.4), we may have general expression as

$$P_r = \frac{(1-w)\theta^2(\theta+2+r)}{(1+\theta)^{3+r}} \quad ; r=1, 2, 3, \ldots \qquad (4.2.5)$$

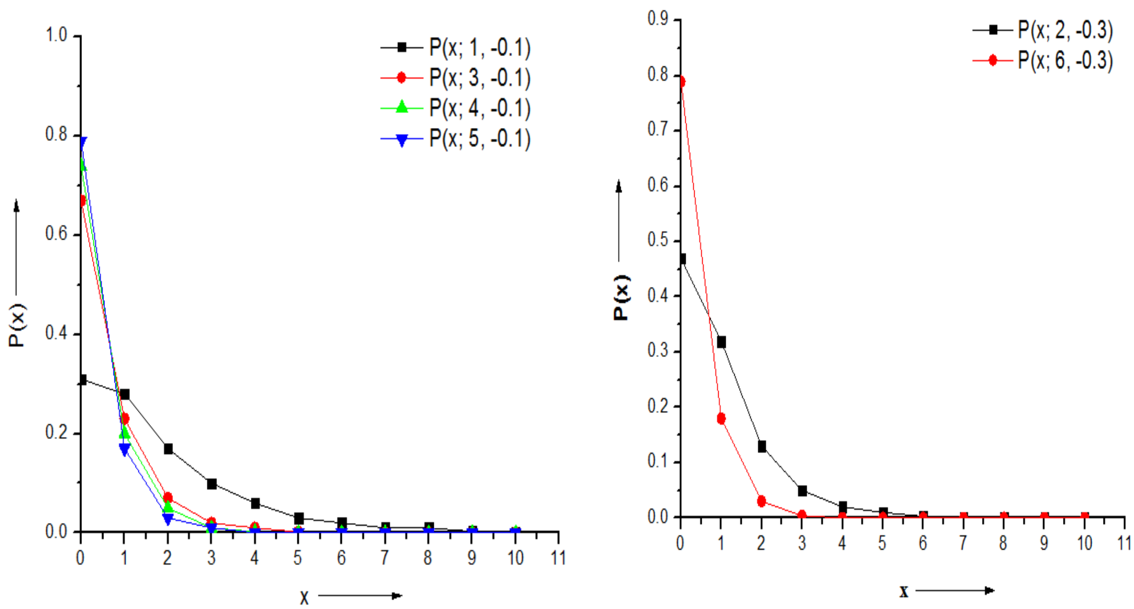
where $$P_0 = w + (1-w)\frac{\theta^2(\theta+2)}{(1+\theta)^3}$$

This result can be used for finding higher order probabilities of the ZMPL distribution very easily.

4.3 Graphical Representations

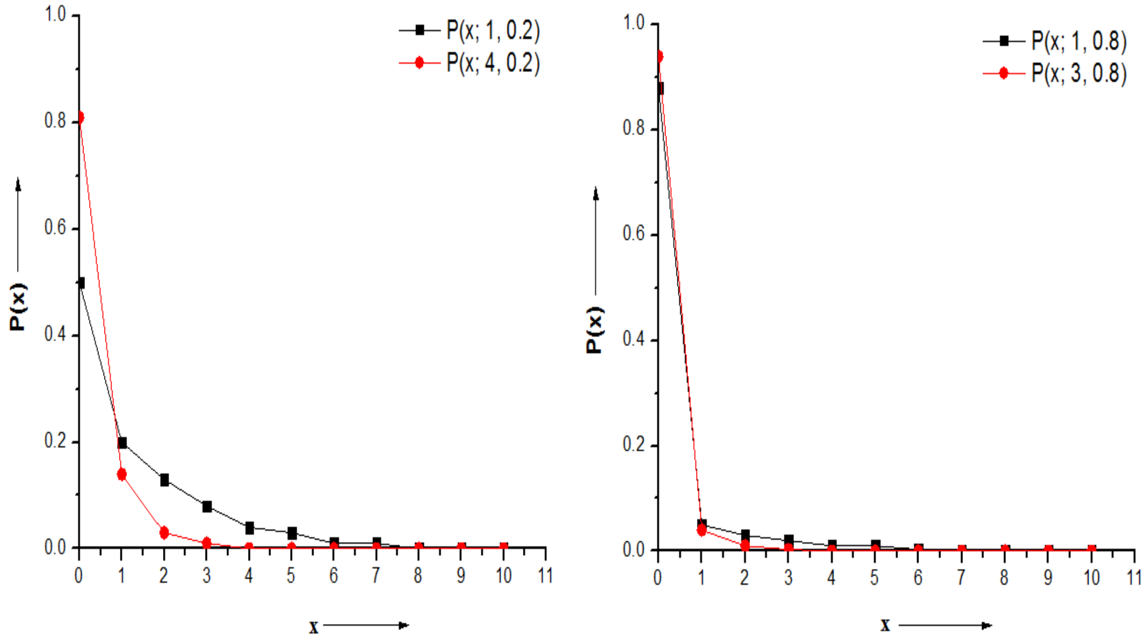
To study the behavior of the ZMPL distribution for varying values of the parameters θ and w, the probabilities for possible values of x are computed by using the above equation (4.2.5) and different graphs may be drawn for various values of the two parameters.

Fig 4.1 Plots of probability P(x; θ, w) for (i) $w = -0.1$ and $\theta = 1, 3, 4, 5$ (ii) $w = -0.3$ and $\theta = 2, 6$ respectively for the ZMPL distribution.



The graphs clearly indicate that for the changes in the values of parameter θ there are significant differences in the probability distribution. That is, as the value of θ increases, the value of P(x; θ, w) is maximum at $x = 0$ and decreases for other values of x. For further study of the effect of the change in the values of parameter w on the behavior of the ZMPL distribution, some graphs are drawn in **Fig 4.2**, **4.3** and **4.4.**

Fig. 4.2 Plots of probability P(x; θ, w) for (i) $w = 0.2$ and $\theta = 1, 4$ (ii) $w = 0.8$ and $\theta = 1, 3$ respectively for the ZMPL distribution.



From the above figure it is clear that when the value of w increases, the value of P(x; θ, w) is maximum at $x = 0$ and decreases sharply to a certain point then slowly decreases. The model looks like L-shaped.

.

Fig. 4.3 Plots of probability P(x; θ, w) for $\theta = 2$ and $w = -0.8, -0.6$ respectively for the ZMPL distribution.

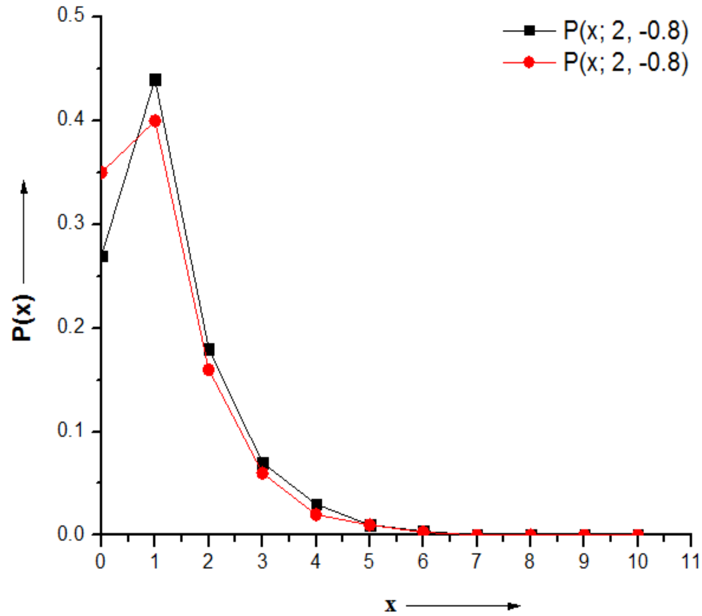
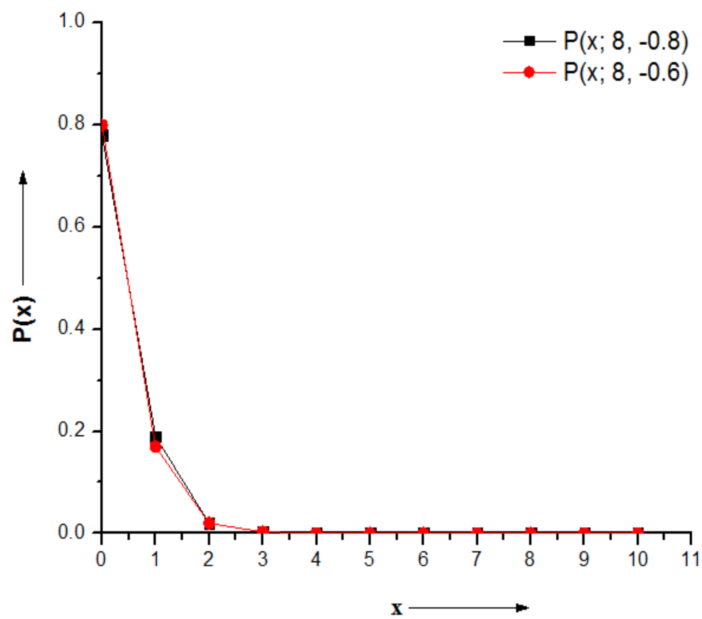


Fig. 4.4 Plots of probability P(x; θ, w) for $\theta = 8$ and $w = -0.8, -0.6$ respectively for the ZMPL distribution.



.   .

From **Fig 4.3**, it is observed that the model loose its L-shaped form when θ and w both are small. But from **Fig 4.4**, it is seen that, when the value of θ is large and the value of w is increased slowly, then the value of P(x; θ, w) is maximum at $x = 0$ and the probabilities decreases sharply to a certain point then slowly decreases for other increasing values of x sharply giving a L-shaped.

4.4 Distributional Properties

This section deals with derivation of some distributional properties of the ZMPL distribution given as follows:

4.4.1 Expression for factorial and raw moments

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

$$g(t+1) = w + (1-w)G(t+1) \tag{4.4.1}$$

where $G(t+1) = \frac{\theta^2(\theta+1-t)}{(1+\theta)(\theta-t)^2}$ is the fmgf of the Poisson-Lindley distribution [cf. Borah and Deka Nath (2001a)]. Expanding (4.4.1) and equating the coefficient of $\frac{t^r}{r!}$, we get the expression for factorial moments recurrence relation for ZMPL distribution as follows

$$\mu'_{(r)} = \frac{r}{\theta^2}\left[2\theta\mu'_{(r-1)} - (r-1)\mu'_{(r-2)}\right]; \quad r > 2 \tag{4.4.2}$$

where $\mu'_{(1)} = \frac{1!(1-w)(\theta+2)}{\theta(1+\theta)}$

$\mu'_{(2)} = \frac{2!(1-w)(\theta+3)}{\theta^2(1+\theta)}$

where $\mu'_{(r)}$ is the r^{th} factorial moment of ZMPL distribution.

The general expression for r^{th} factorial moment may be written as

$$\mu'_{(r)} = \frac{r!(1-w)(\theta+r+1)}{\theta^r(1+\theta)} \;\; ; r = 1, 2, ... \qquad (4.4.3)$$

The moment generating function (mgf) of the ZMPL distribution may be written as

$$m(t) = w + (1-w)\frac{\theta^2(\theta+2-e^t)}{(1+\theta)(1+\theta-e^t)^2} \qquad (4.4.4)$$

The expression for raw moment recurrence relation obtained from equation (4.4.4) for the ZMPL distribution is given as

$$\mu'_r = \frac{(1-w)(\theta+3-2^r)}{\theta(1+\theta)} + \sum_{j=0}^{r-1}\frac{\{3(1+\theta)^2-3*2^{j+1}(1+\theta)+2^{j+1}\}}{\theta^3}\binom{r}{j+1}\mu'_{r-j}\,, \;\; r > 1 \qquad (4.4.5)$$

where $$\mu'_1 = \frac{(1-w)(\theta+2)}{\theta(1+\theta)} \qquad (4.4.6)$$

The second order raw moment obtained from relation (4.4.5) is given by

$$\mu'_2 = \frac{(1-w)(\theta^2+4\theta+6)}{\theta^2(1+\theta)} \qquad (4.4.7)$$

where μ'_r stands for the r^{th} raw moment of the ZMPL distribution. The higher order raw moments can be obtained from equation (4.4.5) very easily. The raw moments can be also obtained from the factorial moments of the ZMPL distribution by using the relation between raw and factorial moments.

Now, the central moments of the ZMPL distribution which can be obtained from the raw moments are given below as

$$\mu_2 = \frac{(1-w)\{\theta^3+4\theta^2+6\theta+2+w(\theta+2)^2\}}{\theta^2(1+\theta)^2} \qquad (4.4.8)$$

$$\mu_3 = \frac{(1-w)\{\theta^5+7\theta^4+22\theta^3+32\theta^2+18\theta+4+w\left(3\theta^4+17\theta^3+36\theta^2+30\theta+4\right)+2w^2(\theta^3+6\theta^2+12\theta+18)\}}{\theta^3(1+\theta)^3} \quad (4.4.9)$$

$$\mu_4 = \frac{(1-w)}{\theta^4(1+\theta)^4}[\theta^7+5\theta^6+97\theta^5+258\theta^4+406\theta^3+314\theta^2+48\theta+24+w$$

$$(4\theta^6+36\theta^5+145\theta^4+312\theta^3+400\theta^2+384\theta+48)+3w^2$$

$$(2\theta^5+15\theta^4+44\theta^3+52\theta^2)+3w^3(\theta^4+8\theta^3+24\theta^2+32\theta+16)] \quad (4.4.10)$$

Remark 4.1 Note that, if $w \to 0$ in the above equations then the raw and central moment of the ZMPL distribution reduces to the moments of the Poisson-Lindley distribution of Sankaran (1970).

4.4.2 Expression for Coefficients of Skewness and Kurtosis

The expression for the coefficients of skewness and kurtosis of the ZMPL can be written as follows

$$\sqrt{\beta}_{\mathbf{1}} = \frac{\mu_3}{{\mu_2}^2} = \frac{A+wB+2w^2C}{\sqrt{(1-w)D^3}} \quad (4.4.11)$$

where $A = \theta^5+7\theta^4+22\theta^3+32\theta^2+18\theta+4$

$B = 3\theta^4+17\theta^3+36\theta^2+30\theta+4$

$C = \theta^3+6\theta^2+12\theta+18$

$D = \theta^3+4\theta^2+6\theta+2+w(\theta+2)^2$

and $$\beta_2 = \frac{\mu_4}{{\mu_2}^2} = \frac{P + wQ + 3w^2R + 3w^3S}{(1-w)D^2} \tag{4.4.12}$$

where $P = \theta^7 + 5\theta^6 + 97\theta^5 + 258\theta^4 + 406\theta^3 + 314\theta^2 + 48\theta + 24$

$Q = 4\theta^6 + 36\theta^5 + 145\theta^4 + 312\theta^3 + 400\theta^2 + 384\theta + 48$

$R = 2\theta^5 + 15\theta^4 + 44\theta^3 + 52\theta^2$

$S = \theta^4 + 8\theta^3 + 24\theta^2 + 32\theta + 16$

From the above expression of $\sqrt{\beta_1}$, it is seen that for any given value of $\theta > 0$ and w is close to unity, the skewness is infinitely large and it becomes smaller as the value of w decreases. The value of β_2 is positive for all values of $\theta > 0$ and $0 < w < 1$. Hence, the ZMPL distribution is leptokurtic.

4.4.3 **Expression for Coefficient of Variation** (CV)

The coefficient of variation for the ZMPL distribution may be written as

$$CV = \frac{\sqrt{v(x)}}{\mu} = \frac{\sqrt{(1-w)\{\theta^3 + 4\theta^2 + 6\theta + 2 + w(\theta+2)^2\}}}{(1-w)(\theta+2)} \tag{4.4.11}$$

It is seen from the above expression that, coefficient of variation of the ZMPL distribution is increased if the value of θ and w is increased in (4.4.11). For study the effect of the changes in the values of CV of the ZMPL distribution for different values of θ and w, in **Table 4.1** we have been calculated CV for different values of θ and w.

It is observed from the **Table 4.1** that the CV of the ZMPL distribution is increased as the values of w is increased when θ is fixed. Similarly, if w is fixed and θ is increased the CV is also increased. It may be noted from the above table that the values of w seems to be more sensitive parameter than the other parameter θ of the distribution.

Table 4.1 The coefficient of variation of the ZMPL distribution for different values of θ and w.

$\theta \rightarrow$ / w ↓.	1.0	1.5	2.0	2.5	3.0	3.5	4.0
-0.4	0.86	1.04	1.19	1.32	1.44	1.56	1.66
-0.1	1.10	1.28	1.44	1.58	1.71	1.83	1.95
0.1	1.30	1.49	1.66	1.81	1.95	2.08	2.21
0.5	1.97	2.19	2.40	2.59	2.76	2.93	3.09
0.9	4.84	5.29	5.72	6.12	6.47	6.85	7.20

In **Fig. 4.5** and **4.6**, we have shown the graphical representations of the CV of the ZMPL distribution for varying values of the parameters θ and w. In **Fig. 4.5**, we take different fixed values of θ along the horizontal axis and corresponding coefficient of variations along the vertical axis. Similarly, we take different fixed values of w along the horizontal axis and corresponding coefficient of variations along the vertical axis in **Fig. 4.6.**

Fig. 4.5 The CV increases slowly as θ increases for different values of w for the ZMPL distribution.

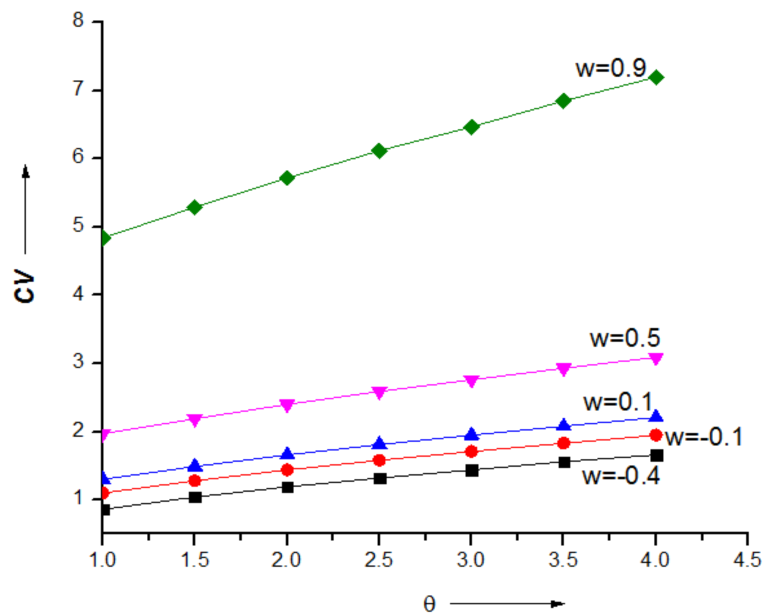
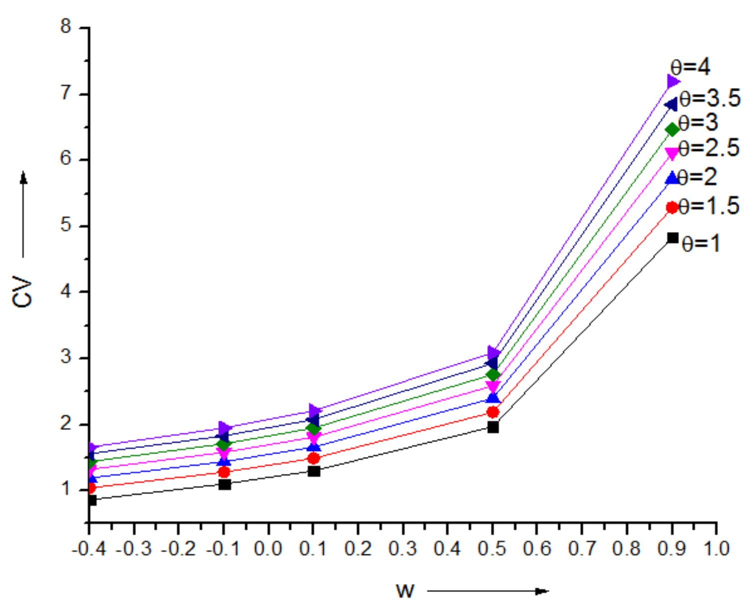


Fig. 4.5 The CV increases as w increases for different values of θ for the ZMPL distribution.



It is clear from the above figures that, the CV of the ZMPL distribution is increased as the values of θ and w is increased and also it is seen that, w be the more sensitive than the other parameter θ.

4.5 Estimation of Parameters

The estimation of parameters of zero-modified distribution other than w can be carried out by ignoring the observed frequency in the zero-class and then using a technique appropriate to the original distribution by omission of zero-class. After the other parameters have been estimated, parameter w can then be estimated by equating the observed and expected frequencies in the zero-class.

The parameters θ and w of the ZMPL distribution can be estimated by using the following methods.

4.5.1 Method of moments

The parameters θ and w of the ZMPL distribution may be estimated from the first two raw moments μ_1' and μ_2' from equations (4.4.6) and (4.4.7). Then, we have

$$\hat{\theta} = \frac{(2\mu_1'-\mu_2')+\sqrt{(\mu_2'-2\mu_1'+2\mu_1'\mu_2'}}{(\mu_2'-\mu_1')} \quad ; \quad \text{where} \quad \mu_2' = \mu_2 + {\mu_1'}^2 \tag{4.5.1}$$

and $$\hat{w} = 1 - \frac{\theta(1+\theta)\mu_1'}{(\theta+2)} \tag{4.5.2}$$

4.5.2 Method based on the first two relative frequencies

Eliminating w between the first two frequencies $\frac{f_1}{N} = \frac{(1-w)\theta^2(\theta+3)}{(1+\theta)^4}$ and $\frac{f_2}{N} = \frac{(1-w)\theta^2(\theta+4)}{(1+\theta)^5}$ of the ZMPL distribution, we get the estimate of parameter θ as

$$\hat{\theta} = \frac{(f_1 - 4f_2) + \sqrt{{f_1}^2 + 8f_1f_2 + 4{f_2}^2}}{2f_2} \qquad (4.5.3)$$

and w may be estimated from the following equation

$$\hat{w} = 1 - \frac{\theta(1+\theta)\bar{x}}{(\theta+2)} \qquad (4.5.4)$$

where, $\bar{x}$ be the sample mean.

4.6 Applications

As the ZMPL distribution has a simple form and has only two parameters so it may be applied in different fields such as biology and ecology, social information, genetic and so on which are discussed below.

For the fitting of the ZMPL distribution to see its suitability and applicability in the field of biology and ecology, in **Table 4.2**, we have considered Student's historic data on Haemocytometer counts of yeast cells of Plunkett and Jain(1975) for which Gegenbauer distribution (GD) was fitted by Borah (1984), using method of moments. In **Table 4.3**, we have considered data set of Beall (1940) , for which generalized Poisson distribution was fitted by Jain (1975) (using MLE)). It is observed from **Table 4.2** and **4.3** that, method of moment gives better results than the method based on the first two frequencies in all the cases. It is also clear from the values of the expected ZMPL frequencies that there is some improvement, however small it may be, in fitting of the ZMPL distribution over the other distributions considered earlier. In case of **Table 4.2**, the method of ratio of first two relative frequencies does not give satisfactory fit, as the computed χ^2 value is quite large.

In **Table 4.4**, we have considered the problem of accidents to 647 working women on high explosive shells in 5 week period [data from Greenwood and Yule (1920)] for which Poisson distribution is fitted earlier and Poisson-Lindley distribution was fitted by Sankaran (1970). It is clearly observed from **Table 4.4** that, the ZMPL distribution describe the data very well. The result obtained by using the method of ratio of first two relative frequencies does not give good fit to the data set for which the results is not reported in **Table 4.4**.

Table 4.2 Fit of the ZMPL distribution to data on Haemocytometer Counts of Yeast Cells. [data of Plunkett and Jain(1975)]

No. of Yeast cells per square	Observed frequencies	Fitted Distribution		GD [Borah (1984)] $\hat{\alpha} = 0.198$ $\hat{\beta} = 0.004$
		ZMPL(MoM) $\hat{\theta}$ =2.669 $\hat{w}$ = -0.431	ZMPL (RF) $\hat{\theta}$=3.0328 $\hat{w}$= -0.6586	
0	213	213.00	204.00	214.15
1	128	127.00	139.18	123.00
2	37	40.91	40.23	44.88
3	18	12.82	11.39	13.36
4	3	3.95	3.18	3.55
5	1	1.20	0.88	0.86
6	0	0.53	0.34	0.20
Total	400	400.0	400.0	400.0
χ^2		1.037	3.93	2.834
$d.f$		2	2	2
$p - value$		0.59	0.14	0.24

Note: ZMPL: Zero-modified Poisson-Lindley distribution.

. MoM: Method of moments.

RF: Method based on the first two relative frequencies.

GD: Gegenbauer distribution.

Table 4.3 Fit of the ZMPL distribution on Pyrausta nubilslis in 1937. [data of Beall (1940)]

No. of Insects	Observed frequencies	Fitted Distribution		GPD[Jain (1975)] $\hat{\lambda}_1 = 0.549$ $\hat{\lambda}_2 = 1.358$
		ZMPL(MoM) $\hat{\theta}$ =1.719 $\hat{w}$ = 0.0573	ZMPL (RF) $\hat{\theta}$=1.449 $\hat{w}$= 0.228	
0	33	32.07	34.08	32.46
1	12	13.47	11.23	13.47
2	6	6.00	5.61	5.60
3	3	2.59	2.71	2.42
4	1	1.09	1.28	1.08
5	1	0.77	1.09	0.97
Total	56	56.0	56.0	56.0
χ^2		0.215	0.096	0.25
$d.f$		1	2	1
$p-value$		0.64	0.95	0.61

.

Note: ZMPL: Zero-modified Poisson-Lindley distribution.

MoM: Method of moments.

RF: Method base on the first two relative frequencies.

GPD: Generalized Poisson distribution.

Table 4.4 Comparison of observed frequencies for accidents to 647 women on high explosive shells in 5 weeks with fitted the Poisson, Poisson-Lindley (PL) and zero-modified Poisson-Lindley (ZMPL) distributions. [data from Greenwood and Yule (1920)]

No. of accidents	Observed frequencies	Poisson $\hat{\theta} = 0.465$	PLD(MoM) Sankaran(1970) $\hat{\theta} = 2.729$	ZMPL(MoM) $\hat{\theta} = 2.451$ $\hat{w} = 0.116$
0	447	406	439.5	447.1
1	132	189	142.8	132.1
2	42	45	45.0	39.0
3	21	7	13.9	13.0
4	3	1	4.2	3.4
≥ 5	2	0.1	1.3	1.0
Total	647	647.0	647.0	647.0
χ^2		61.08	4.82	4.25
$d.f$		2	3	2
$p-value$		-	0.18	0.12

Note : ZMPL: Zero-modified Poisson-Lindley distribution.

. PLD : Poisson-Lindley distribution.

MoM : Method of moments.

RF : Method based on the first two relative frequencies.

****